Singularity of discrete random matrices

SINGULARITY OF DISCRETE RANDOM MATRICES

VISHESH JAIN, ASHWIN SAH, AND MEHTAAB SAWHNEY

Abstract. Let ξ be a non-constant real-valued random variable with finite support and let $M_n(\xi)$ denote an $n \times n$ random matrix with entries that are independent copies of ξ.

For ξ which is not uniform on its support, we show that

$$\mathbb{P}[M_n(\xi) \text{ is singular}] = \mathbb{P}[\text{zero row or column}] + (1 + o_n(1))\mathbb{P}[\text{two equal (up to sign) rows or columns}],$$

thereby confirming a folklore conjecture.

As special cases, we obtain:

- For $\xi = \text{Bernoulli}(p)$ with fixed $p \in (0, 1/2)$,
$$\mathbb{P}[M_n(\xi) \text{ is singular}] = 2n(1-p)^n + (1 + o_n(1))n(n-1)(p^2 + (1-p)^2)^n,$$
which determines the singularity probability to *two* asymptotic terms. Previously, no result of such precision was available in the study of the singularity of random matrices. The first asymptotic term confirms a conjecture of Litvak and Tikhomirov.
- For $\xi = \text{Bernoulli}(p)$ with fixed $p \in (1/2, 1)$,
$$\mathbb{P}[M_n(\xi) \text{ is singular}] = (1 + o_n(1))n(n-1)(p^2 + (1-p)^2)^n.$$
Previously, only the much weaker upper bound of $(\sqrt{p} + o_n(1))^n$ was known due to the work of Bourgain-Vu-Wood.

For ξ which is uniform on its support:

- We show that
$$\mathbb{P}[M_n(\xi) \text{ is singular}] = (1 + o_n(1))^n \mathbb{P}[\text{two rows or columns are equal}].$$
- Perhaps more importantly, we provide a sharp analysis of the contribution of the 'compressible' part of the unit sphere to the lower tail of the smallest singular value of $M_n(\xi)$.

1. Introduction

Let $M_n(\xi)$ be an $n \times n$ random matrix, each of whose entries is an independent copy of a random variable ξ. We will restrict attention to when ξ is a real-valued random variable whose support is finite and contains at least two points (which we call *discrete*). What is the probability that $M_n(\xi)$ is singular? This question, which has been studied since the 1960s, has attracted considerable attention over the years. A well-known folklore conjecture is that the dominant contribution to the probability of singularity is from the events that a row or column is zero, or that two rows or two columns are equal (possibly up to a sign). In order to facilitate discussion, let us introduce some notation. For a vector $v \in \mathbb{R}^n$, we define the event

$$\mathcal{E}_v := \{M_n(\xi)v = 0\}.$$

We will also denote the canonical basis vectors of $\mathbb{R}^n$ by $e_1, \ldots, e_n$. Then, the aforementioned conjecture may be stated as follows.

Conjecture 1.1. *Let ξ be a discrete random variable, and let $M_n(\xi)$ be an $n \times n$ random matrix whose entries are independent copies of ξ. Then*

$$\mathbb{P}[M_n(\xi) \text{ is singular}] = (1 + o_n(1))\bigg(2n\mathbb{P}[\mathcal{E}_{e_1}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1 - e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1 + e_2}]\bigg).$$

In this paper, as our first main result, we confirm a stronger version of Conjecture 1.1 for *all* discrete distributions which are not uniform on their support. Let $s_n(M_n)$ denote the least singular value of an $n \times n$ matrix M_n; recall that $s_n(M_n) = \inf_{x \in \mathbb{S}^{n-1}} \|M_n x\|_2$, where $\mathbb{S}^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$ and $\|\cdot\|_2$ denotes the standard Euclidean norm on $\mathbb{R}^n$.

Theorem 1.2. *Let ξ be a discrete random variable which is not uniform on its support. There exist $c_\xi, C_\xi > 0$ so that for all sufficiently large n, and for all $t \geq 0$,*

$$\mathbb{P}[s_n(M_n(\xi)) \leq t/\sqrt{n}] \leq C_\xi t + 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1+O(\exp(-c_\xi n)))\left(n(n-1)\mathbb{P}[\mathcal{E}_{e_1+e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1-e_2}]\right).$$

By applying Theorem 1.2 with $t = 0$ for the upper bound, and considering the probability that a row or column is zero, or that two rows or two columns are the same (up to a sign) for the lower bound (cf. the corresponding calculation in [15, Section 3.2]), we thus establish the following strengthening of Conjecture 1.1 for discrete distributions which are not uniform on their support.

Theorem 1.3. *Let ξ be a discrete random variable which is not uniform on its support. There exists $c_\xi > 0$ such that*

$$\mathbb{P}[M_n(\xi) \text{ is singular}] = 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1 + O(\exp(-c_\xi n)))\left(n(n-1)\mathbb{P}[\mathcal{E}_{e_1+e_2}] + n(n-1)\mathbb{P}[\mathcal{E}_{e_1-e_2}]\right).$$

Let us record the consequence of this theorem for the special case of $\xi = \mathrm{Ber}(p)$, which has attracted considerable attention.

Theorem 1.4. *Fix $p \in (0, 1/2)$. There exists $c_p > 0$ such that*

$$\mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular}] = 2n(1-p)^n + (1 + O(\exp(-c_p n)))n(n-1)(p^2 + (1-p)^2)^n.$$

Remark. As discussed in Section 1.1, the above theorem resolves a conjecture of Litvak and Tikhomirov [15, Problem 8.2], thereby completing the program of determining the (dominant) mechanism leading to the singularity of sparse Bernoulli matrices. In fact, the above theorem provides the first *two* terms in the asymptotic expansion of the singularity probability of $M_n(\mathrm{Ber}(p))$; a result of this precision was not available before in any context.

Theorem 1.5. *Fix $p \in (1/2, 1)$. There exists $c_p > 0$ such that*

$$\mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular}] = (1 + O(\exp(-c_p n)))n(n-1)(p^2 + (1-p)^2)^n.$$

Remark. The above theorem provides the leading term in the asymptotic expansion of the singularity probability of $M_n(\mathrm{Ber}(p))$. Prior to this work, even the correct value of

$$\lim_{n \to \infty} n^{-1} \log \mathbb{P}[M_n(\mathrm{Ber}(p)) \text{ is singular})]$$

had not been determined; compared to the true value of $(p^2 + (1-p)^2)$ for this quantity, the previous best-known result of Bourgain, Vu, and Wood [3] provides a weaker upper bound of $\sqrt{p}$. The reason that the case $p \in (1/2, 1)$ is more challenging than $p \in (0, 1/2)$ (treated in [1,5,15], see the discussion below) is that in the former case, the dominant contribution to the probability of singularity comes from the event of *two* rows or columns being equal to each other, whereas in the latter case, the dominant contribution comes from the much simpler event of a single row or column being zero.

For general discrete distributions, we determine the value of $\lim_{n\to\infty} n^{-1} \log \mathbb{P}[M_n(\xi) \text{ is singular}]$. The only case not covered by Theorem 1.2 is that of uniform distributions, which we handle with a non-exact main term.

Theorem 1.6. *Let ξ be a discrete random variable. There exists $C_\xi > 0$ such that for any fixed $\epsilon > 0$ and for all sufficiently large n and all $t \geq 0$,*

$$\mathbb{P}[s_n(M_n) \leq t/\sqrt{n}] \leq C_\xi t + 2n\mathbb{P}[\mathcal{E}_{e_1}] + (1+\epsilon)^n \mathbb{P}[\mathcal{E}_{e_1-e_2}].$$

Remark. For non-uniform discrete distributions, Theorem 1.2 is strictly stronger.

1.1. **Previous work.** Let us put Theorems 1.2 and 1.6 in the context of known results. For convenience, we will use $q_n(\xi)$ to denote $\mathbb{P}[M_n(\xi) \text{ is singular}]$. The work of Komlós [13] was the first to show that $q_n(\mathrm{Ber}(1/2)) = o_n(1)$. Much later, an exponential bound on $q_n(\mathrm{Ber}(1/2))$ was obtained by Kahn, Komlós, and Szemerédi [11]. Subsequently, the base of the exponent was improved to 0.939 and $3/4 + o_n(1)$ in a series of works by Tao and Vu [23, 24], and later to $1/\sqrt{2} + o_n(1)$ by Bourgain, Vu, and Wood [3]. Finally, a truly breakthrough result of Tikhomirov [25] in 2018 established that $q_n(\mathrm{Ber}(p)) = (1 - p + o_n(1))^n$ for fixed $p \in (0, 1/2]$. As mentioned earlier, for fixed $p \in (1/2, 1)$, the analogous result was not known prior to this work.

Conjecture 1.1 has been most accessible for sparse Bernoulli distributions, in which case, the right hand side simplifies considerably to $(1 + o_n(1)) \cdot 2n\mathbb{P}[\mathcal{E}_{e_1}]$. Here, by the Bernoulli distribution with parameter p, which we will henceforth denote by $\mathrm{Ber}(p)$, we mean the two point distribution which attains the value 1 with probability p and the value 0 with probability $1 - p$. Basak and Rudelson [1] confirmed the conjecture for $\xi = \mathrm{Ber}(p_n)$ for p_n in a certain range of sparsity limited to $n^{-1} \ln n - n^{-1} g(n) \le p_n \le n^{-1} \ln n + o_n(n^{-1} \ln \ln n)$, where $g(n)$ is some function which grows slowly with n. Subsequently, Litvak and Tikhomirov showed that the conjecture also holds for $\xi = \mathrm{Ber}(p_n)$ for $Cn^{-1} \ln n \le p_n \le c$, where $c > 0$ is a small absolute constant and $C > 0$ is a large absolute constant. Recent work of Huang [5] was able to bridge the gap between the regimes covered in [1] and [15], leaving open the regime $p \in (c, 1/2)$. Establishing Conjecture 1.1 (as opposed to the stronger Theorem 1.4) in this case does not require the full strength of the ideas in this paper; in particular the treatment of the 'compressible' part of the unit sphere is substantially simpler. Since this is a case of particular interest (see [15, Problem 8.2]), we have isolated the proof (given Theorem 2.1) of Conjecture 1.1 for sparse Bernoulli random variables in Section 3, which also serves as a 'warm-up' to subsequent sections.

For general discrete distributions ξ, the only previous systematic study in the literature is the work of Bourgain, Vu, and Wood [3]. They show [3, Corollary 1.2] that if ξ is a discrete distribution with $\sup_{r \in \mathbb{R}} \mathrm{P}[\xi = r] =: p$, then $q_n(\xi) \le (\sqrt{p} + o_n(1))^n$, which is far from optimal (the true bound is never more than $(p + o_n(1))^n$, although it may be much smaller). On the other hand, up to a possible $o_n(1)$ term, Theorem 1.6 in this work always obtains the correct base of the exponent.

For certain specific distributions, Bourgain, Vu, and Wood obtain the correct base of the exponent (again, up to a $o_n(1)$ term). Specifically, they show [3, Corollaries 3.1, 3.2] that if $\xi_{1,\mu}$ is a random variable taking on the value 0 with probability $1 - \mu$ and ± 1 with probability $\mu/2$ each, and if $\xi_{2,\mu}$ is a random variable taking on the value 0 with probability $1 - \mu$ and $\pm 1, \pm 2$ with probability $\mu/4$ each, then $q_n(\xi_{1,\mu}) = (1 - \mu + o_n(1))^n$ for all $\mu \in (0, 1/2)$ and $q_n(\xi_{2,\mu}) = (1 - \mu + o_n(1))^n$ for all $\mu \in (0, 16/25)$. For these random variables, Theorem 1.6 determines the correct base of the singularity probability for all fixed $\mu \in (0, 1)$, and Theorem 1.2 determines the leading order in the asymptotic expansion for $\mu \in (0, 1), \mu \neq 2/3$ in the first case, and $\mu \in (0, 1), \mu \neq 4/5$ in the second case. In fact, for $\mu \in (0, 2/3)$ in the first case, and $\mu \in (0, 4/5)$ in the second case, Theorem 1.2 determines the first two terms in the asymptotic expansion.

We remark that the results of [3] such as [3, Corollary 1.2] are also applicable to discrete random variables valued in the complex numbers, and settings where the entries of $M_n(\xi)$ are not identically distributed, and a small number of rows of $M_n(\xi)$ are possibly deterministic; we have not pursued these extensions.

Finally, we remark that there was a recent paper of Irmatov [6] which claimed to resolve Conjecture 1.1 for Rademacher random matrices. Experts have informed us that there are some unresolved issues in that work that its author is aware of, including [6, Theorem 3]. Furthermore, upon slight modification, the proof in [6] would appear to give impossibly good error terms.

1.2. **Additional results.** The next result addresses the main question left open by our work, namely, the resolution of Conjecture 1.1 for discrete distributions ξ which are uniform on their support. In this direction, we provide a sharp analysis of the contribution of a certain low-entropy part of the unit sphere; in fact, it is this contribution which forms the leading term of the conjectured asymptotic expansion of the singularity probability. This theorem is also central to the proofs of Theorems 1.2 and 1.6.

Theorem 1.7. *Fix a discrete distribution ξ. There exist $\delta, \rho, \eta > 0$ depending on ξ such that for all sufficiently large n and $t \le 1$,*

$$\mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t\right] \le n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + (t + \mathbb{P}[\mathcal{E}_{e_1-e_2}])e^{-\eta n}.$$

The set $\mathrm{Cons}(\delta, \rho)$ appearing above is the set of unit vectors which have at least $(1-\delta)n$ coordinates within distance $\rho/\sqrt{n}$ of each other (see Definition 3.2), although a trivial modification shows this result holds for any sufficiently low-entropy subset of the unit sphere.

Our techniques also lend themselves naturally to studying a certain model of random matrices with combinatorially dependent entries. Let Q_n denote a random matrix with independent rows, each of which is chosen uniformly from among those vectors in $\{0,1\}^n$ which have sum exactly $\lfloor n/2 \rfloor$. In [18], Nguyen showed that $\mathbb{P}[Q_n \text{ is singular}] = O_C(n^{-C})$ for any $C > 0$, and conjectured [18, Conjecture 1.4] that $\mathbb{P}[Q_n \text{ is singular}] = (1/2 + o_n(1))^n$. After intermediate work [4, 7], an exponential upper bound on the singularity probability was only very recently obtained in work of Tran [26]. Our next result settles [18, Conjecture 1.4].

Theorem 1.8. *For every $\epsilon > 0$, there exists C_ϵ depending on ϵ such that for all sufficiently large n, and for all $t \ge 0$,*

$$\mathbb{P}[s_n(Q_n) \le t/\sqrt{n}] \le C_\epsilon t + (1/2 + \epsilon)^n.$$

1.3. **Overview of the techniques.** As in many works in this area, we use the high-level strategy (going back to Kašin [12] and subsequently used in [14, 20–22]) of dividing the unit sphere into 'structured' and 'unstructured' components, and estimating the contribution of each part separately. However, compared to previous works, the treatment of both components require overcoming significant obstacles which unavoidably arise in the sharp analysis of the invertibility of random matrices in any amount of generality.

For instance, in the analysis of structured vectors, we need to additionally capture the event that two rows/columns of the matrix are equal (up to a sign) whereas previous considerations of sharp invertibility only addressed scenarios where the dominant contribution to the probability of singularity is due to a single row or column being zero. As discussed in the remark after Theorem 1.3, this is a fundamental issue. Moreover, in the analysis of unstructured vectors, we need precise metric entropy estimates for the anti-concentration problem with respect to random vectors on general multi-slices. Obtaining partial estimates of this nature (which are not sufficient to prove Conjecture 1.1) even for the special case of the Boolean slice is already a highly non-trivial endeavor which is at the heart of the recent work of Litvak and Tikhomirov [15], where it is accomplished using the substantially more involved notion of the 'UDLCD'.

Structured vectors: The structured vectors in our work are 'almost-constant vectors' i.e. those vectors on $\mathbb{S}^{n-1}$ which have $(1-\delta)n$ coordinates within distance $\rho/\sqrt{n}$ of each other, where $\delta, \rho > 0$ are sufficiently small constants. This class of structured vectors arises naturally in the consideration of the anti-concentration property of a sequence of numbers with respect to a random vector constrained to lie in a 'slice'. Moreover, since vectors which are close to the standard basis vectors e_i or to $e_i \pm e_j$ clearly play a special role in the problem under consideration, it is natural to separately handle 'elementary' and 'non-elementary' structured vectors.

Our treatment of structured vectors, culminating in Theorem 1.7, requires significant innovations compared to previous works on the sharp invertibility of sparse random Bernoulli matrices [1, 5, 15] – in the sparse Bernoulli case, the corresponding class of elementary vectors only needs to consist of those vectors which are close to some e_i, and the largest atom of the the random variable $\mathrm{Ber}(p)$ is conveniently at 0.

In the present work, in order to handle non-elementary vectors, we need to develop novel sharp anticoncentration estimates Propositions 6.2 and 6.3. (In contrast, the essentially standard estimate Lemma 3.7 is sufficient for the case of sparse Bernoulli random variables at the corresponding step). Even more involved is the analysis of elementary vectors, for which we develop a new technique. Let us begin by discussing this technique for $\xi = \mathrm{Ber}(p)$ for fixed $p \in (0, 1/2)$, in which case, the elementary vectors are those which are close to some standard basis vector. For concreteness, consider vectors which are sufficiently close to e_1. We show that, if any such vector has exponentially small image, then either the first column of the matrix is the zero vector, or it must belong to a universal subset of nonzero vectors of $\{0,1\}^n$ of measure at most $(1-p+\epsilon)^n$. The first case corresponds to the term $\mathbb{P}[\mathcal{E}_{e_1}]$ in Conjecture 1.1; for the second case, we leverage the seminal work of Rudelson and Vershynin to show that, on our event, the probability that any vector in this universal subset appears as the first column of the matrix is at most $\exp(-4\epsilon n)$, at which point we can conclude using the union bound.

Of course, for general discrete random variables ξ, one must enlarge the class of elementary vectors to include unit vectors which are close to $(e_i \pm e_j)/\sqrt{2}$ and unit vectors which are close to e_i. In the first case (Propositions 6.5 and 6.7), we use a rotation trick to reduce to a situation where we can use an analysis similar to (but more complicated than) the one outlined in the previous paragraph. The second case requires a very careful treatment since we are aiming for a leading term of the form $(\mathbb{P}[\xi = 0])^n$ (as opposed to $(\sup_{r\in\mathbb{R}} \mathbb{P}[\xi = r])^n$), and moreover, the desired error is $(\mathbb{P}[\xi = \xi'] - \eta)^n$ which may be very small. To accomplish this, we first prove a version of Theorem 1.6 with an estimate on the singularity probability of the form $(\sup_{r\in\mathbb{R}} \mathbb{P}[\xi = r] + o_n(1))^n$ (Proposition 5.4 and Theorem 5.5), and then leverage these preliminary estimates to obtain the desired bound.

We emphasize that our treatment of structured vectors, as captured by Theorem 1.7, is not sensitive to the non-uniformity of the distribution ξ. In particular, given Theorems 1.2 and 1.7, the only missing case in the complete resolution of Conjecture 1.1 (in fact, in a stronger form) is a sharp analysis of unstructured vectors in the case when ξ is uniform on its support.

Unstructured vectors: The unstructured vectors are the complement of the structured vectors i.e. those which do not have a $(1-\delta)$-fraction of their coordinates within $\rho/\sqrt{n}$ of each other. Our treatment of these vectors relies on the non-uniformity of ξ by exploiting the gap between $\mathbb{P}[\xi = \xi']$ and the entropy of ξ; the idea to exploit such a gap to prove sharp invertibility results (in the case of Bernoulli random variables) is due to Litvak and Tikhomirov [15].

The main ingredient in our work for handling such vectors is Theorem 2.1, which is an extension of [25, Theorem B] to a (real) multislice, i.e., the set of vectors in $\{a_1, \dots, a_k\}^n$ which have a prescribed number of coordinates taking on each of the values $a_1, \dots, a_k$. Such a result was previously not known even for the Boolean slice; indeed, the work [15] uses a rather involved notion of arithmetic structure to study anti-concentration on Boolean slices, which is not powerful enough to handle slices that are *not* very far from the central slice. We remark that in general, even establishing much less precise versions of [25, Theorem B] on the Boolean slice has been very challenging, despite much work due to the natural connection to certain combinatorial models of random matrices (cf. [10] and the references therein).

Compared to [25, Theorem B], we need to overcome two challenges. The first, as mentioned above, is the lack of independence between the coordinates of a vector uniformly distributed on the multi-slice. The second challenge is that $a_1, \dots, a_k$ are now arbitrary real numbers (corresponding to the support of ξ), and hence, certain arguments tailored for integers no longer apply. Overcoming these challenges requires additional ideas, which we discuss in Section 2.

1.4. **Notation.** For a positive integer N, $\mathbb{S}^{N-1}$ denotes the set of unit vectors in $\mathbb{R}^N$, and if $x \in \mathbb{R}^N$ and $r \geq 0$ then $\mathbb{B}_2^N(x,r)$ denotes the radius r Euclidean ball in $\mathbb{R}^N$ centered at x. $\|\cdot\|_2$ denotes the standard Euclidean norm of a vector, and for a matrix $A = (a_{ij})$, $\|A\|$ is its spectral norm (i.e., $\ell^2 \to \ell^2$ operator norm).

We will let $[N]$ denote the interval $\{1,\ldots,N\}$. For nonnegative integers $m \leq n$, we let $\{0,1\}^n_m$ be the set of vectors in $\{0,1\}^n$ with sum m.

Since it is essential throughout the paper, we formally record the definition of a discrete random variable and the corresponding random matrix.

Definition 1.9. We say that a random variable ξ is a discrete random variable (equivalently, has a discrete distribution) if it is real-valued, its support is finite, and the support contains at least two distinct points. $M_n(\xi)$ denotes the $n \times n$ random matrix, with independent entries that are copies of ξ.

For ξ a discrete random variable with $k = |\operatorname{supp}(\xi)|$ (so that $k \geq 2$), we will denote its support by $\vec{a} = (a_1,\ldots,a_k)$, and the (nonzero) probabilities of attaining $a_1,\ldots,a_k$ by $\vec{p} = (p_1,\ldots,p_k)$. Note that $\|\vec{p}\|_1 = 1$, and $\|\vec{p}\|_2^2 \leq \|\vec{p}\|_\infty$ with equality if and only if ξ is uniform on its support. We will use $H(\xi)$ to denote the natural-logarithmic entropy of ξ, i.e., $H(\xi) = H(\vec{p}) = \sum_{i=1}^k -p_i \log(p_i)$. We will (somewhat abusively) use p_0 to denote $\mathbb{P}[\xi = 0]$.

For a random variable ξ and a real number $r \geq 0$, we let $\mathcal{L}(\xi, r) := \sup_{z\in\mathbb{R}} \mathbb{P}[|\xi - z| \leq r]$. We will use $\ell^1(\mathbb{Z})$ to denote the set of functions $f\colon \mathbb{Z} \to \mathbb{R}$ for which $\sum_{z\in\mathbb{Z}} |f(z)| < \infty$.

We will also make use of asymptotic notation. For functions f, g, $f = O_\alpha(g)$ (or $f \lesssim_\alpha g$) means that $f \leq C_\alpha g$, where C_α is some constant depending on α; $f = \Omega_\alpha(g)$ (or $f \gtrsim_\alpha g$) means that $f \geq c_\alpha g$, where $c_\alpha > 0$ is some constant depending on α, and $f = \Theta_\alpha(g)$ means that both $f = O_\alpha(g)$ and $f = \Omega_\alpha(g)$ hold. For parameters ϵ, δ, we write $\epsilon \ll \delta$ to mean that $\epsilon \leq c(\delta)$ for a sufficient function c.

Finally, we will omit floors and ceilings where they make no essential difference.

1.5. **Organization.** The remainder of this paper is organized as follows. In Section 2, we prove our key inversion of randomness estimate for conditional thresholds on the multislice (Theorem 2.1). In Section 3, we combine this with a much simpler analysis of the structured vectors (compared to the proof of Theorem 1.7) in order to complete the proof of Conjecture 1.1 for the special case of $\xi = \operatorname{Ber}(p)$ for a fixed $p \in (0, 1/2)$. This section also serves as a 'warm-up' to the subsequent sections. In Section 4, we use the results of Section 2 to prove the necessary invertibility estimate for unstructured vectors (Theorem 4.1). In Section 5, we prove a weaker version of Theorem 1.6; this is used in our significantly more involved treatment of structured vectors in general case (i.e., the proof of Theorem 1.7), which is the content of Section 6. In Section 7, we quickly combine Theorems 1.7 and 4.1 to prove Theorems 1.2 and 1.6. Finally, in Section 8, we prove Theorem 1.8 for the combinatorial model of random matrices discussed earlier.

1.6. **Acknowledgements.** We thank Mark Rudelson, Konstantin Tikhomirov, and Yufei Zhao for comments on the manuscript. We are grateful to an anonymous referee for their careful reading of our paper and for numerous comments which have improved the presentation. A.S. and M.S. were supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. This work was done when V.J. was participating in a program at the Simons Institute for the Theory of Computing.

2. Inversion of randomness on the multislice

In this section, we prove our key inversion of randomness result, Theorem 2.1. We will focus on the non-independent "multislice" version as its deduction is strictly harder than the independent version, Theorem 2.14 (which we will only use to establish the preliminary estimate Theorem 5.5).

The proof of Theorem 2.1 follows a direction introduced by Tikhomirov [25]. In this approach, the relevant Lévy concentration function of a random vector is replaced with certain random averages of functions. One then shows that the random vectors with large values of the Lévy concentration function are super-exponentially rare, by first demonstrating a weaker notion of anticoncentration after revealing $(1-\epsilon)n$ coordinates of the random vector, and then iterating a smoothing procedure on linear-sized pieces of the vector which allows one to bootstrap the strength of anticoncentration considered.

Our major challenges lie in (i) the non-independence of the coordinates of a vector on the multislice, as the arguments in [25] rely strongly on the independence structure of the considered model, and (ii) the freedom to allow the support of ξ to consist of arbitrary real numbers, as certain arguments in [25] rely on the integrality of the support. For a more gentle introduction to the techniques in this section, we refer the reader to the expository paper [9, Theorem 3.1] where we record the proof for the Boolean slice, a setting which encounters the first challenge but not the second. We note that the presentation here is entirely self-contained and familiarity with [9] is not assumed.

2.1. **Statement and preliminaries.** Let $N, n \geq 1$ be integers and let $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$ be real parameters. We say that $\mathcal{A} \subseteq \mathbb{Z}^n$ is $(N, n, K_1, K_2, K_3, \delta)$-admissible if

- $\mathcal{A} = A_1 \times \cdots \times A_n$, where each A_i is a subset of $\mathbb{Z}$,
- $|A_1| \cdots |A_n| \leq (K_3 N)^n$,
- $\max_i \max\{|a| : a \in A_i\} \leq nN$,
- A_i is an integer interval of size at least $2N+1$ for $i > 2\delta n$, and either (P1) and (P2) hold, or (Q1) and (Q2) hold:

(P1) A_{2i} is an integer interval of size at least $2N+1$ contained in $[-K_1 N, K_1 N]$ for $i \leq \delta n$,
(P2) A_{2i-1} is symmetric about 0, is a union of two integer intervals of total size at least $2N$, and satisfies $A_{2i-1} \cap [-K_2 N, K_2 N] = \emptyset$ for $i \leq \delta n$.
(Q1) A_{2i} is an integer interval of size at least $2N+1$ contained in $[K_1 N, K_2 N]$ for $i \leq \delta n$,
(Q2) A_{2i-1} is an integer interval of size at least $2N+1$ contained in $[-K_2 N, -K_1 N]$ for $i \leq \delta n$.

Recall at this point that ξ, which has (nonzero) probabilities $\vec{p} = (p_1, \ldots, p_k)$ on atoms $\vec{a} = (a_1, \ldots, a_k)$, is fixed. Let $\mathcal{A} = A_1 \times \cdots \times A_n$ be an $(N, n, K_1, K_2, K_3, \delta)$-admissible set, and let $(X_1, \ldots, X_n)$ be the random vector uniformly distributed on $\mathcal{A}$. For any $f : \mathbb{R} \to \mathbb{R}$, any $0 \leq \ell \leq n$, and any $\vec{s} \in \mathbb{Z}_{\geq 0}^k$ with $\|\vec{s}\|_1 = \ell$, define the random function (depending on the randomness of $X_1, \ldots, X_n$):

$$f_{\mathcal{A},\vec{s},\ell}(t) := \mathbb{E}_b\bigg[f\bigg(t + \sum_{i=1}^{\ell} b_i X_i\bigg)\bigg|\#\{b_i = a_j\} = s_j \ \forall j \in [k]\bigg],$$

where $\mathbb{E}_b$ denotes the expectation over a random vector $b = (b_1, \ldots, b_\ell) \in \mathbb{R}^\ell$ with coordinates independently distributed as ξ. The conditioning encodes that for all $j \in [k]$, there are exactly s_j coordinates (out of ℓ) where b hits the atom a_j.

Theorem 2.1. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \geq 1$, there are $L_{2.1} = L_{2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$, and $\gamma_{2.1} = \gamma_{2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and $n_{2.1} = n_{2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$ and $\eta_{2.1} = \eta_{2.1}(\xi, \epsilon, \delta, K_1, K_2, K_3, M)$ such that the following holds.*

Let $n \geq n_{2.1}$, $1 \leq N \leq \exp((H(\vec{p}) - \epsilon)n)$, $f \in L^1(\mathbb{R})$ be a nonnegative function such that $\|f\|_1 = 1$ and $\log_2 f$ is $\eta_{2.1}$-Lipschitz, and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Suppose also that $\|\vec{\gamma}\|_\infty \leq \gamma_{2.1}$. Then, for any $\vec{m} \in \mathbb{Z}_{\geq 0}^k$ such that $\|\vec{m}\|_1 = n$ and $\|\vec{m} - \vec{p}n\|_\infty \leq \gamma_{2.1} n$,

$$\mathbb{P}[\|f_{\mathcal{A},\vec{m},n}\|_\infty \geq L_{2.1}(N\sqrt{n})^{-1}] \leq \exp(-Mn).$$

Given this we can deduce the following corollary which is crucial in our application.

Definition 2.2. Fix a discrete distribution ξ. Let $\vec{\gamma}$ be a nonnegative vector with $\|\vec{\gamma}\|_\infty \in (0, \min(\vec{p}))$ and let $r \geq 0$. For a vector $(x_1, \dots, x_n) \in \mathbb{R}^n$, we define

$$\mathcal{L}_{\xi,\vec{\gamma}}\bigg(\sum_{i=1}^n b_i x_i, r\bigg) := \sup_{z\in\mathbb{R}} \mathbb{P}\bigg[\bigg|\sum_{i=1}^n b_i x_i - z\bigg| \leq r \bigg| \#\{b_i = a_j\} \in [p_j n - \gamma_j n, p_j n + \gamma_j n]\ \forall j \in [k]\bigg],$$

where $b_1, \dots, b_n$ are independent ξ random variables. We also define

$$\mathcal{L}_{\xi}\bigg(\sum_{i=1}^n b_i x_i, r\bigg) = \sup_{z\in\mathbb{R}} \mathbb{P}\bigg[\bigg|\sum_{i=1}^n b_i x_i - z\bigg| \leq r\bigg].$$

Corollary 2.3. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \geq 1$, there are $L_{2.3} = L_{2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$ and $\gamma_{2.3} = \gamma_{2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and $n_{2.3} = n_{2.3}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$ such that the following holds.*

Let $n \geq n_{2.3}$, $1 \leq N \leq \exp((H(\vec{p}) - \epsilon)n)$ and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Suppose also that $\|\vec{\gamma}\|_\infty \leq \gamma_{2.3}$. Then

$$\bigg|\bigg\{x \in \mathcal{A} : \mathcal{L}_{\xi,\vec{\gamma}}\bigg(\sum_{i=1}^n b_i x_i, \sqrt{n}\bigg) \geq L_{2.3} N^{-1}\bigg\}\bigg| \leq e^{-Mn}|\mathcal{A}|.$$

Proof sketch. This is essentially the same as the deduction in [25, Corollary 4.3]. We apply Theorem 2.1 to $f(t) := 2^{-|t|/\sqrt{n}}/\iota$, where $t \in \mathbb{R}$ and ι is an appropriate normalization, separately for all $\vec{m} \in \mathbb{Z}^k_{\geq 0}$ such that $\|\vec{m} - \vec{p}n\|_\infty \leq \gamma_{2.1} n$, and then conclude using a union bound. □

The proof of Theorem 2.1 makes use of an anticoncentration estimate on the multislice, which we record below (Lemmas 2.5 and 2.6), and is ultimately a consequence of the following standard anticoncentration inequality due to Kolmogorov-Lévy-Rogozin.

Lemma 2.4 ([19])**.** *Let $\xi_1, \dots, \xi_n$ be independent random variables. Then, for any real numbers $r_1, \dots, r_n > 0$ and any real $r \geq \max_{i\in[n]} r_i$, we have*

$$\mathcal{L}\bigg(\sum_{i=1}^n \xi_i, r\bigg) \leq \frac{C_{2.4} r}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i)) r_i^2}},$$

where $C_{2.4} > 0$ is an absolute constant.

Lemma 2.5. *Fix $(a_1, \dots, a_k) \in \mathbb{R}^k$ with distinct coordinates. Let $\sigma, \lambda \in (0, 1/3)$ and $r > 0$. Let $Z = \{z_1, \dots, z_n\}$ be a set of real numbers for which there exist disjoint subsets $Z_1, Z_2 \subseteq Z$ such that $|Z_1|, |Z_2| \geq \sigma n$ and such that $|z_i - z_j| \geq r$ for all $z_i \in Z_1, z_j \in Z_2$. Then, there exists $C_{2.5} = C_{2.5}(\lambda, \sigma, k)$ such that for any $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ with $\|\vec{s}\|_1 = n$ and with $s_\ell \in [\lambda n, (1-\lambda)n]$ for some $\ell \in [k]$, we have*

$$\mathcal{L}\bigg(\sum_{i=1}^n z_i b_i, r \cdot \min_{i<j} |a_i - a_j|\bigg) \leq \frac{C_{2.5}}{\sqrt{n}},$$

where $(b_1, \dots, b_n)$ is a random vector uniformly chosen from among those with s_j coordinates equal to a_j for all $j \in [k]$.

Proof. By reindexing the coordinates of Z, we may assume that for $i \in [\sigma n]$, $z_{2i-1} \in Z_1$ and $z_{2i} \in Z_2$. In particular, for $i \in [\sigma n]$, we have $|z_{2i} - z_{2i-1}| \geq r$. Furthermore, by the pigeonhole principle, there exists some $\ell' \neq \ell$ such that $s_{\ell'} \geq \lambda n/k$. We will now use the randomness within the atoms a_ℓ and $a_{\ell'}$ in order to derive the anticoncentration result. Note that $\sum_{i=1}^n b_i z_i$ has the same distribution as

$$\sum_{i>2\sigma n} z_i b_i + \sum_{j\leq\sigma n} \bigg(z_{2j-1} b_{2j-1} + z_{2j} b_{2j} + b'_j (b_{2j} - b_{2j-1})(z_{2j-1} - z_{2j})\bigg),$$

where $b'_1, \ldots, b'_{\sigma n}$ are i.i.d. Ber(1/2) random variables. Next, note that by a standard large deviation estimate, we have

$$\mathbb{P}[|\{j \in [\sigma n] : \{b_{2j-1}, b_{2j}\} = \{a_\ell, a_{\ell'}\}| \le c(\sigma, \lambda, k)n] \le \exp(-c(\sigma, \lambda, k)n), \tag{2.1}$$

where $c(\sigma, \lambda, k) > 0$ is a constant depending only on σ, λ, and k. On the other hand, on the complement of this event, we may conclude by applying Lemma 2.4 to (2.1), using only the randomness in $b'_1, \ldots, b'_{\sigma n}$. □

Lemma 2.6. *Fix a discrete distribution ξ, $\lambda \in (0, 1/3)$, $\delta_0 \in (0, 1/4)$. Let $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible for some integer parameters N, n and real parameters $\delta \in [\delta_0, 1/4)$, $K_3 > K_2 > K_1 > 1$. Suppose that $n > n_{2.6}(\lambda, \delta_0, K_1, K_2, K_3)$, $\ell \ge \delta_0 n$, and $\vec{s} \in \mathbb{Z}^k_{\ge 0}$ with $\|s\|_1 = \ell$ and $s_{j_0} \in [\lambda \ell, (1-\lambda)\ell]$ for some $j_0 \in [k]$. Then, for any interval J,*

$$\int_{t \in J} f_{\mathcal{A}, \vec{s}, \ell}(t)dt \le \frac{C_{2.6}(\lambda, \xi, \delta_0, K_1, K_2) \max(|J|, N)}{N\sqrt{n}}.$$

Proof. The proof is nearly identical to that in [25, Lemma 4.4] though we provide details as we are in the slightly different setting of $L^1(\mathbb{R})$. Fix $X_1, \ldots, X_\ell$. Then

$$\begin{aligned}
\int_{t \in J} f_{\mathcal{A}, \vec{s}, \ell}(t)dt &= \int_{t \in J} \mathbb{E}_b\left[f\left(t + \sum_{i=1}^{\ell} b_i X_i\right) \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] dt \\
&= \mathbb{E}_b\left[\int_{t \in J} f\left(t + \sum_{i=1}^{\ell} b_i X_i\right) dt \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] \\
&= \mathbb{E}_b\left[\int_{t \in \mathbb{R}} f(t) \mathbb{1}_{J + \sum_{i=1}^{\ell} b_i X_i}(t) dt \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] \\
&= \int_{t \in \mathbb{R}} f(t) \mathbb{E}_b\left[\mathbb{1}_{J + \sum_{i=1}^{\ell} b_i X_i}(t) \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] dt \\
&= \int_{t \in \mathbb{R}} f(t) \mathbb{P}_b\left[\sum_{i=1}^{\ell} b_i X_i \in J - t \middle| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] dt \\
&\le \mathcal{L}\left(\sum_{i=1}^{\ell} b_i X_i, |J|\right) \int_{t \in \mathbb{R}} |f(t)| dt \le \mathcal{L}\left(\sum_{i=1}^{\ell} b_i X_i, |J|\right),
\end{aligned}$$

where $(b_1, \ldots, b_\ell)$ is uniformly chosen from vectors which have s_j coordinates equal to a_j for all $j \in [k]$, and we have used that $\|f\|_1 = 1$. The required estimate now follows immediately from Lemma 2.5 applied with $r = (K_2 - K_1)N$, which is possible due to the admissibility of $\mathcal{A}$. □

2.2. **Preprocessing on real-valued multislices.** As in [25], we first prove a version of Theorem 2.1 in which L is allowed to depend on M.

Proposition 2.7. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \min(\vec{p})$, and a given parameter $M \ge 1$, there is $\gamma_{2.7} = \gamma_{2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3) \in (0, \epsilon)$ independent of M and there are $L_{2.7} = L_{2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) > 0$ and $n_{2.7} = n_{2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \ge 1$ such that the following holds.*

Let $n \ge n_{2.7}$, $1 \le N \le \exp((H(\vec{p}) - \epsilon)n)$, and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Let f be a nonnegative function in $L^1(\mathbb{R})$ with $\|f\|_1 = 1$ such that $\log_2 f$ is 1-Lipschitz. Then, for all $\ell \in [(1 - \gamma_{2.7})n, n]$ and $\vec{s} \in \mathbb{Z}^k_{\ge 0}$ with $\|\vec{s}\|_1 = \ell$ and $\|\vec{s} - \vec{p}\ell\|_\infty \le \gamma_{2.7}\ell$, we have

$$\mathbb{P}\left[\|f_{\mathcal{A}, \vec{s}, \ell}\|_\infty \ge L_{2.7}(N\sqrt{n})^{-1}\right] \le \exp(-Mn).$$

Proposition 2.7 should be seen as an analogue of [25, Proposition 4.5] for the multislice. As mentioned earlier, compared to [25], our situation is much more delicate since we are working with a vector with non-independent coordinates and need to extract a term corresponding to the entropy of the multislice. (Such complications are already encountered when working with a Boolean slice.) Working on real multislices presents additional difficulties (along with significant notational complications), owing to the fact that we are working on $L^1(\mathbb{R})$; this extension is handled by using the log-Lipschitz condition on f. We note that the corresponding statements in [25] do not need to use any log-Lipschitz assumption at this stage of the argument since they are proved for $\ell^1(\mathbb{Z})$. We also note that, while the constant 1 in 1-log-Lipschitz is arbitrary, some condition of this nature is necessary to rule out f being very close to a Dirac mass ([25]).

We first note the trivial recursive relation

$$f_{\mathcal{A},\vec{s},\ell}(t) = \sum_{i=1}^{k} \frac{s_i}{\ell} f_{\mathcal{A},\vec{s}-e_i,\ell-1}(t + a_i X_\ell)$$

for all $1 \le \ell \le n$ and $\vec{s} \in \mathbb{Z}_{\ge 0}^k$ with $\|\vec{s}\|_1 = \ell$. If any coordinate of $\vec{s}$ is zero, note that the corresponding term (which would be undefined) has a coefficient of 0, and drops out. Note also that, by definition, $f_{\mathcal{A},\vec{0},0} = f$.

Definition 2.8 (Step record and averaging sequence)**.** Fix $f, \mathcal{A}, \vec{s}, \ell$, a point $t \in \mathbb{R}$, and a choice of $X = (X_1, \ldots, X_n)$. For such a choice, we define the *averaging sequence* $(t_i)_{i=0}^{\ell}$ and *step record* $(w_i)_{i=1}^{\ell}$ as follows:

- $t_\ell := t$,
- Since

$$h_\ell := f_{\mathcal{A},\vec{s},\ell}(t_\ell) = \sum_{j=1}^{k} \frac{s_j}{\ell} f_{\mathcal{A},\vec{s}-e_j,\ell-1}(t_\ell + a_j X_\ell),$$

  at least one of the k terms $f_{\mathcal{A},\vec{s}-e_j,\ell-1}(t_\ell + a_j X_\ell)$ has a positive coefficient and is at least h_ℓ. If it is index j, set $w_\ell = j$.
- Set $t_{\ell-1} := t_\ell + a_{w_\ell} X_\ell$, $h_{\ell-1} := f_{\mathcal{A},\vec{s}-e_{w_\ell},\ell-1}(t_{\ell-1})$, and repeat with $t_{\ell-1}, \vec{s} - e_{w_\ell}, \ell - 1$.

It will be convenient to write

- $W_i(j) := \#\{u \in [i] : w_u = j\}$ and $\overline{W}_i(j) := W_i(j)/i$ for all $i \in [\ell]$ and $j \in [k]$. We will view $W_i = (W_i(1), \ldots, W_i(k))$ as a vector in $\mathbb{Z}^k$.

We note some straightforward consequences of these definitions.

- $W_\ell = \vec{s}$.
- $W_{i-1} = W_i - e_{w_i}$ for $1 \le i \le \ell$, where we assume $W_0 = \vec{0}$.
- $\|W_i\|_1 = i$.
- $t_{i-1} = t_i + a_{w_i} X_i$ for all $i \in [\ell]$.
- $f_{\mathcal{A},W_i,i}(t_i) = \sum_{j=1}^k \overline{W_i}(j) f_{\mathcal{A},W_i - e_j, i-1}(t_i + a_j X_i)$.
- $h_i = f_{\mathcal{A},W_i,i}(t_i)$.
- $f(t_0) = h_0 \ge h_1 \ge \cdots \ge h_\ell = f_{\mathcal{A},\vec{s},\ell}(t)$.

Definition 2.9 (Drops and robust steps)**.** With notation as above, given $i \in [\ell]$:

- For $\lambda \in (0, 1)$, we say that step i is λ-*robust* if

$$\overline{W}_i(w_i) \in (\lambda, 1 - \lambda)$$

- For $R > 0$, we say that there is an R-*drop* at step i if

$$f_{\mathcal{A},W_i - e_j, i-1}(t_{i-1} + z X_i) \le \frac{R}{N\sqrt{n}}$$

for all $j \in [k]$ such that $W_i(j) > 0$ and for all $z \in \mathrm{supp}(\xi - \xi') \setminus \{0\}$.

Next we show that if $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty$ is large in an appropriate sense, then there is a step record and averaging sequence with linearly many robust steps which do not participate in an R-drop.

Lemma 2.10. *Let $\xi, \mathcal{A}, f, N, \epsilon$ be as in Proposition 2.7, and let $L \geq 1$. Then, there exist $\lambda_{2.10} = \lambda_{2.10}(\xi, \epsilon) \in (0, 1/3)$, $\gamma_{2.10} = \gamma_{2.10}(\xi, \epsilon) \in (0, 1)$, and $n_{2.10} = n_{2.10}(\xi, \epsilon)$ for which the following holds.*

Let $n \geq n_{2.10}$, $R = \gamma_{2.10} L$, let $\ell \in [(1-\gamma_{2.10})n, n]$ and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ satisfy $\|\vec{s}\|_1 = \ell$ and $\|\vec{s} - \vec{p}\ell\|_\infty \leq \gamma_{2.10}\ell$. Then, for $(X_1, \ldots, X_n) \in \mathcal{A}$,

$$\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L(N\sqrt{n})^{-1}$$

implies that there exists some $t \in \mathbb{R}$ with $f_{\mathcal{A},\vec{s},\ell}(t) \geq L(N\sqrt{n})^{-1}$ so that its averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$ satisfy

$$\#\{i \in [\ell] : \text{step } i \text{ is } \lambda_{2.10}\text{-robust and is not an } R\text{-drop}\} \geq \gamma_{2.10} n.$$

Proof. Consider $(X_1, \ldots, X_n) \in \mathcal{A}$ satisfying $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L(N\sqrt{n})^{-1}$. Then, there is some $t \in \mathbb{R}$ such that $f_{\mathcal{A},\vec{s},\ell}(t) \geq L(N\sqrt{n})^{-1}$. We will show that the conclusion of the lemma is satisfied for this t, for suitable choice of $\gamma_{2.10}, \lambda_{2.10}$. Below, we will make extensive use of the notation and relations in Definitions 2.8 and 2.9. Let $(t_i)_{i=0}^{\ell}$ and $(w_i)_{i=1}^{\ell}$ denote, respectively, the averaging sequence and step record of t. Note that

$$L(N\sqrt{n})^{-1} \leq f_{\mathcal{A},\vec{s},\ell}(t) = h_0 \prod_{i=1}^{\ell} \frac{h_i}{h_{i-1}} \leq h_{\ell-1} \leq \cdots \leq h_0.$$

We begin by controlling the ratios h_i/h_{i-1} at steps i which are R-drops. Hence, suppose that step i is an R-drop. If $w_i = u$, then $W_i = W_{i-1} + e_u$ and $t_i = t_{i-1} - a_u X_i$. Hence

$$\begin{aligned}
\frac{h_i}{h_{i-1}} &= \sum_{j=1}^{k} \overline{W}_i(j) \frac{f_{\mathcal{A},W_i - e_j, i-1}(t_i + a_j X_i)}{f_{\mathcal{A},W_{i-1},i-1}(t_{i-1})} \\
&= \overline{W}_i(u) + \sum_{j \neq u} \overline{W}_i(j) \frac{f_{\mathcal{A},W_i - e_j, i-1}(t_{i-1} + (a_j - a_u)X_i)}{h_{i-1}} \\
&\leq \overline{W}_i(u) + \sum_{j \neq u} \overline{W}_i(j) \frac{R(N\sqrt{n})^{-1}}{L(N\sqrt{n})^{-1}} \\
&= \overline{W}_i(u) + (1 - \overline{W}_i(u))\gamma_{2.10}.
\end{aligned}$$

The inequality uses the definition of R-drops (this is applicable since $a_j - a_u \in \mathrm{supp}(\xi - \xi') \setminus \{0\}$) along with $h_i \geq L(N\sqrt{n})^{-1}$. Note that if the condition $W_i(j) > 0$ in the definition of R-drops is not satisfied, then the jth term already drops out in the first line. Thus, we see that if step i is an R-drop, then

$$\frac{h_i}{h_{i-1}} \leq \overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.10}. \tag{2.2}$$

Note that if step i is $\lambda_{2.10}$-robust, the right-hand side is at least $\lambda_{2.10}$. Therefore, for any step i which is $\lambda_{2.10}$-robust, we have

$$\lambda_{2.10} \leq \overline{W}_i(w_i) + (1 - \overline{W}_i(w_i))\gamma_{2.10} \leq \overline{W}_i(w_i)\left(1 + \frac{\gamma_{2.10}}{\lambda_{2.10}}\right), \tag{2.3}$$

where the final inequality uses $(1 - \overline{W}_i(w_i))/\overline{W}_i(w_i) \leq 1/\lambda_{2.10}$ at any $\lambda_{2.10}$-robust step i.

Now, let $I \subseteq [\ell]$ denote the steps i which are $\lambda_{2.10}$-robust, and let $J \subseteq I$ denote the steps i which are *not* R-drops (so that $I \setminus J$ is the set of $\lambda_{2.10}$-robust R-drops). Our goal is to provide a lower bound on $|J|$.

Since $h_0 \le \|f\|_\infty \le \|f\|_1 = 1$ (this uses the 1-Lipschitz condition on $\log_2 f$), we have

$$
\begin{aligned}
L(N\sqrt{n})^{-1} &\le \prod_{i\in I\setminus J} \frac{h_i}{h_{i-1}} \le \prod_{i\in I\setminus J} (\overline{W}_i(w_i) + (1-\overline{W}_i(w_i))\gamma_{2.10}) \\
&= \frac{\prod_{i\in I}(\overline{W}_i(w_i) + (1-\overline{W}_i(w_i))\gamma_{2.10})}{\prod_{i\in J}(\overline{W}_i(w_i) + (1-\overline{W}_i(w_i))\gamma_{2.10})} \\
&\le \frac{(1+\gamma_{2.10}/\lambda_{2.10})^{|I|}\prod_{i\in I}\overline{W}_i(w_i)}{\lambda_{2.10}^{|J|}} \\
&= (1+\gamma_{2.10}/\lambda_{2.10})^{|I|}\lambda_{2.10}^{-|J|}\prod_{i\in[\ell]}\overline{W}_i(w_i)\prod_{i\in[\ell]\setminus I}\overline{W}_i(w_i)^{-1} \\
&= (1+\gamma_{2.10}/\lambda_{2.10})^{|I|}\cdot\lambda_{2.10}^{-|J|}\cdot\binom{\ell}{\vec{s}}^{-1}\cdot\prod_{i\in[\ell]\setminus I}\overline{W}_i(w_i)^{-1}; \qquad (2.4)
\end{aligned}
$$

here, the first line uses $h_i/h_{i-1} \le 1$ and (2.2), the third line uses (2.3), and the last line uses the identity

$$\prod_{i\in[\ell]}\overline{W}_i(w_i) = \binom{\ell}{\vec{s}}^{-1} := \binom{\ell}{s_1,\ldots,s_k}^{-1}.$$

This follows since both sides are equal to the probability that a uniformly random sample from $[k]^\ell$, conditioned on having s_j copies of j for each $j \in [k]$, returns $(w_1,\ldots,w_\ell)$.

Note that the first and the third terms in the final product in (2.4) are easy to suitably control (by taking $\gamma_{2.10}$ and $\lambda_{2.10}$ to be sufficiently small). As we will see next, these parameters also allow us to make the last term at most $\exp(c\epsilon n)$ for any constant $c > 0$.

Let $K \subseteq [\ell] \setminus I$ denote those indices i such that $\overline{W}_i(w_i) \ge 1 - \lambda_{2.10}$. Then,

$$\prod_{i\in K}\overline{W}_i(w_i)^{-1} \le (1-\lambda_{2.10})^{-|K|}. \qquad (2.5)$$

It remains to bound

$$\prod_{i\in[\ell]\setminus(I\cup K)}\overline{W}_i(w_i)^{-1}.$$

Note that for every $i \in [\ell] \setminus (I \cup K)$, we have $\overline{W}_i(w_i) \le \lambda_{2.10}$. Let J_j for $j \in [k]$ be the set of $i \in [\ell] \setminus (I \cup K)$ with $w_i = j$.

The following is the key point: let $i_1,\ldots,i_{u_j} \in J_j$ be all elements of J_j in order. Then, for all $y \in [u_j]$, we have

$$y \le W_{i_y}(j) \le \lambda_{2.10}\ell.$$

Hence,

$$u_j \le \lambda_{2.10}\ell \quad\text{and}\quad \overline{W}_{i_y}(w_{i_y})^{-1} \le i_y/y \le \ell/y.$$

We derive

$$\prod_{i\in[\ell]\setminus(I\cup K)}\overline{W}_i(w_i)^{-1} = \prod_{j=1}^{k}\prod_{i\in J_j}\overline{W}_i(w_i)^{-1} \le \left(\prod_{u=1}^{\lceil\lambda_{2.10}\ell\rceil}\frac{\ell}{u}\right)^{k} \le \left(\frac{e}{\lambda_{2.10}}\right)^{2k\lambda_{2.10}\ell}. \qquad (2.6)$$

Substituting (2.5) and (2.6) in (2.4), we have

$$Ln^{-1/2}\exp((\epsilon - H(\vec{p}))n) \le \lambda_{2.10}^{-|J|}\cdot\left(1+\frac{\gamma_{2.10}}{\lambda_{2.10}}\right)^{\ell}\cdot\binom{\ell}{\vec{s}}^{-1}\cdot(1-\lambda_{2.10})^{-\ell}\cdot\left(\frac{e}{\lambda_{2.10}}\right)^{2k\lambda_{2.10}\ell}. \tag{2.7}$$

We will first choose $\lambda_{2.10}$, and then choose some $\gamma_{2.10} < \lambda_{2.10}^2$. Note that, by enforcing the constraint $\gamma_{2.10} < \lambda_{2.10}^2$, we can choose $\lambda_{2.10}$ sufficiently small depending on ϵ and ξ so that the second term, the fourth term, and the fifth term in the product in (2.7) are each bounded above by $\exp(\epsilon n/10)$ and so that (using Stirling's approximation) the third term is bounded above by $\exp(\epsilon n/10 - H(\vec{p})n)$. Hence, we can choose $\lambda_{2.10}$ depending on ϵ and ξ such that

$$n^{-1/2}\exp(\epsilon n/2) \le \lambda_{2.10}^{-|J|}.$$

Now, for all n sufficiently large depending on ϵ, we can find $\gamma_{2.10}$ sufficiently small depending on $\epsilon, \lambda_{2.10}$ such that $|J| \ge \gamma_{2.10}n$. This completes the proof. □

We are now ready to prove Proposition 2.7.

Proof of Proposition 2.7. We use Lemma 2.10 along with a union bound. For controlling individual events in the union, we will use the following. Consider a step record $(w_i)_{i=1}^{\ell}$. We write $A_i = A_{i,0} \cup A_{i,1}$, where each of these is an integer interval of size at least N (this is possible by the admissibility of $\mathcal{A}$). Now suppose step i is $\lambda_{2.10}$-robust with respect to $(w_i)_{i=1}^{\ell}$. If $i > \delta_0 n$, then for any $t \in \mathbb{R}$, $j \in [k]$ and $z \in \operatorname{supp}(\xi - \xi') \setminus \{0\}$, by Lemma 2.6, we have

$$\begin{aligned}
\mathbb{E}[f_{\mathcal{A},W_i-e_j,i-1}(t+zX_i)|X_1,\dots,X_{i-1}] &= \frac{1}{|A_i|}\sum_{\tau\in t+zA_i} f_{\mathcal{A},W_i-e_j,i-1}(\tau)\\
&\le \max_{y\in\{0,1\}}\frac{1}{|A_{i,y}|}\sum_{\tau\in t+zA_{i,y}} f_{\mathcal{A},W_i-e_j,i-1}(\tau)\\
&\le \max_{y\in\{0,1\}}\frac{2^{|z|}}{|A_{i,y}|}\left|\int_{t+z\min A_{i,y}}^{t+z\max A_{i,y}} f_{\mathcal{A},W_i-e_j,i-1}(\tau)d\tau\right|\\
&\le \frac{2^{|z|+1}C_{2.6}(\lambda_{2.10}/2,\xi,\delta_0,K_1,K_2)\max(|z||A_{i,y}|,N)}{|A_{i,y}|N\sqrt{n}}\\
&\le \frac{4^{|z|+1}C_{2.6}(\lambda_{2.10}/2,\xi,\delta_0,K_1,K_2)}{N\sqrt{n}}.
\end{aligned}$$

Here, we have used that $i-1 \ge \delta_0 n$, that $W_i - e_j$ has at least one coordinate in $[\lambda_{2.10}(i-1)/2, (1-\lambda_{2.10}/2)(i-1)]$ (since W_i satisfies a similar property with coordinate w_i), and that each $A_{i,y}$ is length at least N. We also used that $\log_2 f$ is 1-Lipschitz in the second inequality (where the absolute values are put just in case $z < 0$ and the limits of integration are in the wrong direction).

Now, consider $t \in \mathbb{R}$ with averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$. Note that, given the 'starting point' t_0 of the averaging sequence, the points $t_1, \dots, t_{i-1}$ are determined by $X_1, \dots, X_{i-1}$. In particular, the event that step i is not an R-drop is determined by $t_0, X_1, \dots, X_i, w_1, \dots, w_i$. Therefore, by Markov's inequality, we see that for any $\lambda_{2.10}$-robust step i with $i > \delta_0 n$, given the step record $(w_i)_{i=1}^{\ell}$ and the starting point t_0 of the averaging sequence $(t_i)_{i=0}^{\ell}$,

$$\mathbb{P}[\text{step } i \text{ is not an } R\text{-drop}|X_1,\dots,X_{i-1}] \le \frac{k^3 4^{2\|\vec{a}\|_\infty+1}C_{2.6}(\lambda_{2.10}/2,\xi,\delta_0,K_1,K_2)}{R}. \tag{2.8}$$

This follows from a union bound over the at most k^3 possible conditions for an R-drop and the fact that all $z \in \operatorname{supp}(\xi - \xi') \setminus \{0\}$ have magnitude at most $2\|\vec{a}\|_\infty$.

From here on, the proof closely follows the proof of [25, Proposition 4.5]. Fix parameters as given in the proposition statement. Let $\lambda_{2.10} = \lambda_{2.10}(\xi, \epsilon)$. We choose $\gamma_{2.7} = \gamma_{2.10}(\xi, \epsilon)$. Further, we set $R' = \gamma_{2.10}L/2$, where $L \geq 1$ will be chosen later.

Let $\mathcal{E}_L$ denote the event that $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \geq L(N\sqrt{n})^{-1}$. For $(X_1, \ldots, X_n) \in \mathcal{E}_L$, by Lemma 2.10, there exists $t \in \mathbb{R}$ with $f_{\mathcal{A},\vec{s},\ell}(t) \geq L(N\sqrt{n})^{-1}$ with averaging sequence $(t_i)_{i=0}^{\ell}$ and step record $(w_i)_{i=1}^{\ell}$ such that

$$\#\{i \in [\ell] : \text{step } i \text{ is } \lambda_{2.10}\text{-robust and is not a } 2R'\text{-drop in } (t_i)_{i=0}^{\ell}\} \geq \gamma_{2.10}n.$$

We then shift t_0 to the nearest integer $\widetilde{t_0}$. We also shift $(t_i)_{i=1}^{\ell}$ by the same amount to obtain points $(\widetilde{t_i})_{i=1}^{\ell}$ (note that these points are not necessarily integers). We call the sequence $(\widetilde{t_i})_{i=0}^{\ell}$, which technically may no longer be an averaging sequence, a *witnessing sequence*. We see that every index which is not a $2R'$-drop in $(t_i)_{i=0}^{\ell}$ will not be an R'-drop in $(\widetilde{t_i})_{i=0}^{\ell}$ as $\log_2 f$ is 1-Lipschitz.

Taking a union bound over the choice of the step record is not costly, and note that given $(X_1, \ldots, X_n)$ and the step record, the witnessing sequence is completely determined by its starting point $\widetilde{t_0}$. Furthermore, the definition of the witnessing sequence and the definition of $f_{\mathcal{A},\vec{s},\ell}$ easily show that

$$\widetilde{t_0} \in \{\tau \in \mathbb{Z} : f(\tau) > (2N\sqrt{n})^{-1}\} =: \mathcal{D}.$$

Note that $\mathcal{D}$ is a deterministic set depending only on f. Further, since $\|f\|_1 = 1$ and $\log_2 f$ is 1-Lipschitz, we see that

$$|\mathcal{D}| \leq 4N\sqrt{n}.$$

To summarize, we have shown that if $(X_1, \ldots, X_n) \in \mathcal{E}_L$, then there exists a witnessing sequence $(\widetilde{t_i})_{i=0}^{\ell}$ with step record $(w_i)_{i=1}^{\ell}$ such that $\widetilde{t_0} \in \mathcal{D}$, and such that

$$\#\{i \in [\ell] : \text{step } i \text{ is } \lambda_{2.10}\text{-robust and is not an } R'\text{-drop in } (\widetilde{t_i})_{i=0}^{\ell}\} \geq \gamma_{2.10}n.$$

Therefore, by the union bound and since $N \leq k^n$ (as $H(\vec{p}) \leq \log k$), it follows that

$$\mathbb{P}[\mathcal{E}_L] \leq (2k^2)^n \sup_{\substack{I \subseteq [\ell], |I| = \lceil \gamma_{2.10}n \rceil \\ \widetilde{t_0} \in \mathcal{D}, (w_i)_{i=1}^{\ell} \in [k]^{\ell}}} \mathbb{P}[\text{The witnessing sequence starts at } \widetilde{t_0}, \text{ has step record } (w_i)_{i=1}^{\ell}, \text{ and}$$
$$\text{every } i \in I \text{ is } \lambda_{2.10}\text{-robust and is not an } R'\text{-drop}],$$

where the supremum is only over those $(w_i)_{i=1}^{\ell}$ which have s_j coordinates equal to j for all $j \in [k]$.

From (2.8), taking $\delta_0 = \gamma_{2.10}/2$, it follows that the probability appearing on the right hand side above is bounded by

$$\left(\frac{2k^3 4^{2\|\vec{a}\|_\infty + 1} C_{2.6}(\lambda_{2.10}/2, \xi, \gamma_{2.10}/2, K_1, K_2)}{\gamma_{2.10}L}\right)^{\gamma_{2.10}n/2},$$

since there are at least $\gamma_{2.10}n/2$ values of $i \in I$ with $i > \delta_0 n$ and since $R' = \gamma_{2.10}L/2$ by definition. Therefore, taking L and n sufficiently large depending on M and the parameters appearing above gives the desired conclusion. □

2.3. **Refining the initial estimate.** We now need to remove the dependence of L on M. This is accomplished by the main result of this subsection, Proposition 2.11, which is a multislice and $L^1(\mathbb{R})$ analogue of [25, Proposition 4.10]. Even though we are working in the much more complicated setting of real multislices, remarkably, our proof of Proposition 2.11 is able to use [25, Proposition 4.10] as a black box: roughly, we first use a re-randomization procedure to reduce smoothing on the multislice for $L^1(\mathbb{R})$ to smoothing on the hypercube, also for $L^1(\mathbb{R})$. At this juncture, the necessary smoothing estimate on the hypercube for $L^1(\mathbb{R})$ can in fact be lifted from the smoothing estimate for the hypercube for $\ell^1(\mathbb{Z})$, proved in [25]. In particular, we reduce the smoothing estimate for general log-Lipschitz functions in $L^1(\mathbb{R})$ to that of a simpler class of "step" functions, which in turn is equivalent to $\ell^1(\mathbb{Z})$.

Proposition 2.11. *Fix a discrete distribution ξ. There exists $h = h(\xi) \geq 1$ so that the following holds. For any $\epsilon \in (0,1)$, $\widetilde{R} \geq 1$, $L_0 \geq h\widetilde{R}$, and $M \geq 1$, there is $\gamma_{2.11} = \gamma_{2.11}(\xi)$ and there are $n_{2.11} = n_{2.11}(\xi, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.11} = \eta_{2.11}(\xi, \epsilon, L_0, \widetilde{R}, M) \in (0,1)$ with the following property. Let $L_0 \geq L \geq h\widetilde{R}$, let $n \geq n_{2.11}$, $N \in \mathbb{N}$, and let $g \in L^1(\mathbb{R})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,
(B) $\log_2 g$ *is* $\eta_{2.11}$*-Lipschitz,*
(C) $\int_{t\in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any interval I of size N, and*
(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq 2\lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \ldots, Y_{2\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in L^1(\mathbb{R})$ by

$$\widetilde{g}(t) = \mathbb{E}_b\bigg[g\bigg(t + \sum_{i=1}^{2\lfloor \epsilon n \rfloor} b_i Y_i\bigg)\bigg| \#\{b_i = a_j\} = s_j \ \forall j \in [k]\bigg]$$

where $b = (b_1, \ldots, b_{2\lfloor \epsilon n \rfloor})$ is a vector of independent ξ components and $\vec{s} \in \mathbb{Z}^k_{\geq 0}$ satisfies $\|\vec{s}\|_1 = 2\lfloor \epsilon n \rfloor$ and

$$\bigg\|\frac{\vec{s}}{2\lfloor \epsilon n \rfloor} - \vec{p}\bigg\|_\infty \leq \gamma_{2.11}.$$

Then

$$\mathbb{P}\bigg[\|\widetilde{g}\|_\infty > \frac{19L/20}{N\sqrt{n}}\bigg] \leq \exp(-Mn).$$

We now state an analogue of Proposition 2.11 for independent scaled Bernoulli random variables, which in fact is strong enough to imply Proposition 2.11.

Proposition 2.12. *Fix $h \geq 1$, and let $z \in [h^{-1}, h]$. For any $\epsilon \in (0,1)$, $\widetilde{R} \geq 1$, $L_0 \geq 64h^2\widetilde{R}$, and $M \geq 1$, there are $n_{2.12} = n_{2.12}(h, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.12} = \eta_{2.12}(h, \epsilon, L_0, \widetilde{R}, M) \in (0,1)$ with the following property. Let $L_0 \geq L \geq 64h^2\widetilde{R}$, let $n \geq n_{2.12}$, $N \in \mathbb{N}$, and let $g \in L^1(\mathbb{R})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,
(B) $\log_2 g$ *is* $\eta_{2.12}$*-Lipschitz,*
(C) $\int_{t\in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any interval I of size N, and*
(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq \lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \ldots, Y_{\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in L^1(\mathbb{R})$ by

$$\widetilde{g}(t) = \mathbb{E}_b g\bigg(t + z\sum_{i=1}^{\lfloor \epsilon n \rfloor} b_i Y_i\bigg)$$

where b is a vector of independent $\mathrm{Ber}(1/2)$ *components. Then*

$$\mathbb{P}\bigg[\|\widetilde{g}\|_\infty \geq \frac{9L/10}{N\sqrt{n}}\bigg] \leq \exp(-Mn).$$

This follows almost immediately from an $\ell^\infty(\mathbb{Z})$ decrement result established by Tikhomirov [25].

Proposition 2.13 ([25, Proposition 4.10]). *For any $p \in (0, 1/2]$, $\epsilon \in (0, 1)$, $\widetilde{R} \geq 1$, $L_0 \geq 16\widetilde{R}$, and $M \geq 1$ there are $n_{2.13} = n_{2.13}(p, \epsilon, L_0, \widetilde{R}, M) > 0$ and $\eta_{2.13} = \eta_{2.13}(p, \epsilon, L_0, \widetilde{R}, M) \in (0, 1)$ with the following property. Let $L_0 \geq L \geq 16\widetilde{R}$, let $n \geq n_{2.13}$, $N \in \mathbb{N}$, and let $g \in \ell^1(\mathbb{Z})$ be a nonnegative function satisfying*

(A) $\|g\|_1 = 1$,
(B) $\log_2 g$ *is* $\eta_{2.13}$*-Lipschitz,*
(C) $\sum_{t \in I} g(t) \leq \widetilde{R}/\sqrt{n}$ *for any integer interval I of size N, and*
(D) $\|g\|_\infty \leq L/(N\sqrt{n})$.

For each $i \leq \lfloor \epsilon n \rfloor$, let Y_i be a random variable uniform on some disjoint union of integer intervals of cardinality at least N each, and assume that $Y_1, \ldots, Y_{\lfloor \epsilon n \rfloor}$ are mutually independent. Define a random function $\widetilde{g} \in \ell^1(\mathbb{Z})$ by

$$\widetilde{g}(t) = \mathbb{E}_b g\left(t + \sum_{i=1}^{\lfloor \epsilon n \rfloor} b_i Y_i\right)$$

where b is a vector of independent $\mathrm{Ber}(p)$ *components. Then*

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{(1 - p(1 - 1/\sqrt{2}))L}{N\sqrt{n}}\right] \leq \exp(-Mn).$$

Remark. In [25, Proposition 4.10], there is a condition $N \leq 2^n$ which is not necessary (indeed, it is not used anywhere in the proof) and so has been dropped. In fact, we will only need values $N \leq k^n$, in which case one can actually replace n by kn and ϵ by ϵ/k (and adjust other parameters appropriately) in order to deduce what we need directly from the statement as written in [25]. We will only need this statement for $p = 1/2$.

We first prove Proposition 2.12.

Proof of Proposition 2.12. Consider the operator $\mathcal{O} : L^1(\mathbb{R}) \to \ell^1(\mathbb{Z})$ given by

$$(\mathcal{O}\omega)(t) = \int_{-z/2}^{z/2} \omega(zt + u)\, du.$$

We note that $\|\omega\|_1 = \|\mathcal{O}\omega\|_1$ and if ω is nonnegative and $\log_2 \omega$ is η-Lipschitz, then

$$z2^{-\eta h/2}\|\omega\|_\infty \leq \|\mathcal{O}\omega\|_\infty \leq z\|\omega\|_\infty.$$

Given $g \in L^1(\mathbb{R})$ satisfying the given conditions, we consider $g' \in \ell^1(\mathbb{Z})$ defined via

$$g' = \mathcal{O}g.$$

We see that g' satisfies properties (A), (B), (C), (D) of Proposition 2.13 with log-Lipschitz constant slightly changed (depending on z, hence h), L changed to zL, and $\widetilde{R}$ increased to $4h\widetilde{R}$. These last changes are responsible for the condition $L_0 \geq 64h^2\widetilde{R}$.

Since $zL \geq h^{-1}L \geq 16(4h\widetilde{R})$, we may apply Proposition 2.13 to g' to deduce that $\|\widetilde{g'}\|_\infty$ is small, except with superexponentially small probability. Here $\widetilde{g'}$ is averaged in the sense of Proposition 2.13 with respect to the same $Y_1, \ldots, Y_{\lfloor \epsilon n \rfloor}$.

Now, by Fubini's theorem, note that

$$\widetilde{g'} = \mathcal{O}\widetilde{g},$$

where $\widetilde{g}$ is averaged in the sense of Proposition 2.12. Therefore,

$$\mathbb{P}\left[\|\mathcal{O}\widetilde{g}\|_\infty > \frac{(2 + \sqrt{2})zL/4}{N\sqrt{n}}\right] \leq \exp(-Mn),$$

so that

$$\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{(2+\sqrt{2})2^{\eta_{2.12}h/2}L/4}{N\sqrt{n}}\right] \le \exp(-Mn).$$

Finally, if $\eta_{2.12}$ is appropriately small, we deduce the desired as

$$\frac{2+\sqrt{2}}{4} < \frac{9}{10}. \qquad \square$$

Finally, we are able to deduce Proposition 2.11.

Proof of Proposition 2.11. Similar to the proof of Lemma 2.5, we can use an equivalent method of sampling from the $\vec{s}$-multislice to rewrite $\widetilde{g}(t)$ as

$$\begin{aligned}\widetilde{g}(t) &= \mathbb{E}_b\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n\rfloor}(b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i})\right)\middle|\#\{b_i = a_j\} = s_j \ \forall j \in [k]\right] \\ &= \mathbb{E}_{b,b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n\rfloor} b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i} + b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i})\right)\middle|\#\{b_i = a_j\} = s_j \ \forall j \in [k]\right],\end{aligned}$$

where b' is an $\lfloor \epsilon n\rfloor$-dimensional vector with *independent* $\mathrm{Ber}(1/2)$ components. Below, we will fix b and use only the randomness in b'. In order to do this, let

$$B_0 := \left\{b_1, \ldots, b_{2\lfloor \epsilon n\rfloor} : \#\{i : b_{2i-1} = a_1, b_{2i} = a_2\} \ge \min(\vec{p})^2\epsilon n/8\right\}.$$

Then, provided that $\gamma_{2.11}$ is chosen sufficiently small depending on ξ, and n is sufficiently large depending on ξ and ϵ, we have

$$\mathbb{E}_b[1_{B_0}|\#\{b_i = a_j\} = s_j \ \forall j \in [k]] > \frac{1}{2}.$$

Let $\mathcal{E}_L$ denote the event (depending on $Y_1, \ldots, Y_{2\lfloor \epsilon n\rfloor}$) that $\|\widetilde{g}\|_\infty > 19L/(20N\sqrt{n})$. Now, suppose $Y_1, \ldots, Y_{2\lfloor \epsilon n\rfloor} \in \mathcal{E}_L$, and suppose further that $\|\widetilde{g}\|_\infty$ is attained at $t \in \mathbb{R}$. Let

$$B_1 := \left\{b_1, \ldots, b_{2\lfloor \epsilon n\rfloor} : \mathbb{E}_{b'}\left[g\left(t+\sum_{i=1}^{\lfloor \epsilon n\rfloor}(b_{2i-1}Y_{2i-1}+b_{2i}Y_{2i}+b'_i(b_{2i}-b_{2i-1})(Y_{2i-1}-Y_{2i})\right)\middle| b\right] \ge \frac{9L/10}{N\sqrt{n}}\right\}.$$

Since $\|g\|_\infty \le L/(N\sqrt{n})$, it follows from the reverse Markov inequality that

$$\mathbb{E}_b[1_{B_1}|\#\{b_i = a_j\} = s_j \ \forall j \in [k]] > \frac{1}{2}.$$

Thus, we see that for every $(Y_1, \ldots, Y_{2\lfloor \epsilon n\rfloor}) \in \mathcal{E}_L$, there exists some $b \in B_0 \cap B_1$. Hence, taking a union bound, we see that

$$\begin{aligned}\mathbb{P}\left[\|\widetilde{g}\|_\infty > \frac{19L/20}{N\sqrt{n}}\right] &\le \mathbb{P}[\exists b \in B_0 : b \in B_1] \\ &\le |B_0| \sup_{b\in B_0} \mathbb{P}\left[\exists t : \mathbb{E}_{b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n\rfloor}(b_{2i-1}Y_{2i-1} + b_{2i}Y_{2i} + b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i}))\right)\right] \ge \frac{9L/10}{N\sqrt{n}}\right] \\ &\le |B_0| \sup_{b\in B_0} \mathbb{P}\left[\exists t : \mathbb{E}_{b'}\left[g\left(t + \sum_{i=1}^{\lfloor \epsilon n\rfloor} b'_i(b_{2i} - b_{2i-1})(Y_{2i-1} - Y_{2i})\right)\right] \ge \frac{9L/10}{N\sqrt{n}}\right]. \qquad (2.9)\end{aligned}$$

We now bound the probability appearing on the right hand side of the above equation uniformly for $b \in B_0$. We fix $b \in B_0$, and note that, by definition, there is a set $I = \{i_1, \ldots, i_m\} \subseteq \lfloor \epsilon n \rfloor$ such that $|I| = m \geq \min(\vec{p})^2 \epsilon n/8$ and such that for all $j \in [m]$,

$$b'_{i_j}(b_{2i_j} - b_{2i_j-1})(Y_{2i_j-1} - Y_{2i_j}) = b'_{i_j}(a_2 - a_1)(Y_{2i_j-1} - Y_{2i_j}).$$

For $j \in [k]$, let $Y^b_j := Y_{2i_j} - Y_{2i_j-1}$. Let $Y_{-2\cdot I}$ denote all components of $Y_1, \ldots, Y_{2\lfloor \epsilon n \rfloor}$, except those corresponding to indices in $2 \cdot I$, and let $Y_{2\cdot I}$ denote the remaining components. Then, for $b \in B_0$ and a choice of $Y_{-2\cdot I}$, we define the random function (depending on $Y_{2\cdot I}$),

$$\widetilde{g}_{b,Y_{-2\cdot I}}(t) := \mathbb{E}_{b'} g\bigg(t + (a_1 - a_2) \sum_{j=1}^{\lfloor \min(\vec{p})^2 \epsilon n/8 \rfloor} b'_j Y^b_j\bigg).$$

Thus, we see that for any $b \in B_0$ and $Y_{-2\cdot I}$, the probability appearing on the right hand side of (2.9) is bounded by

$$\mathbb{P}\bigg[\|\widetilde{g}_{b,Y_{-2\cdot I}}\|_\infty \geq \frac{9L/10}{N\sqrt{n}}\bigg],$$

where the probability is over the choice of $Y_{2\cdot I}$.

At this point, we can apply Proposition 2.12 to $\widetilde{g}_{b,Y_{-2\cdot I}}$. Let us quickly check that the hypotheses of Proposition 2.12 are satisfied. The assumptions on g needed in Proposition 2.12 are satisfied because the same properties are assumed in Proposition 2.11 (see below for the log-Lipschitz condition). Moreover, $b'_1, \ldots, b'_{\lfloor \min(\vec{p})^2 \epsilon n/8 \rfloor}$ are independent Ber(1/2) random variables. Finally, notice that, given $Y_{-2\cdot I}$, each Y^b_j is a random variable uniform on some disjoint intervals of cardinality at least N each (since Y^b_j is a translation of Y_{2i_j} which is assumed to satisfy this property). Also, $a_1 - a_2$ is bounded away from 0 (in terms of ξ).

Thus, Proposition 2.12 shows that the expression on the right hand side of (2.9) is bounded above by

$$|B_0| \sup_{b \in B_0, Y_{-2\cdot I}} \mathbb{P}\bigg[\|\widetilde{g}_{b,Y_{-2\cdot I}}\|_\infty > \frac{9L/10}{N\sqrt{n}}\bigg] \leq k^n \exp(-M \min(\vec{p})^2 n/8),$$

provided that we choose $\eta_{2.11}$ sufficiently small compared to $\eta_{2.12}(d, \min(\vec{p})^2\epsilon/8, L_0, \tilde{R}, M)$, where $d = \max(|a_2 - a_1|, |a_2 - a_1|^{-1})$. The desired result now follows after rescaling M by a constant factor (depending on ξ). □

2.4. **Deriving the final result.** We now prove the main result of this section, Theorem 2.1. The proof of Theorem 2.1 given Propositions 2.7 and 2.11 is similar to the derivation in [25, Theorem 4.2] however we record the argument in full detail below.

Proof of Theorem 2.1. Fix ξ and any admissible parameters $\delta, K_1, K_2, K_3, \epsilon, N$, and the given parameter $M \geq 1$. We need to choose $L_{2.1}, \gamma_{2.1}, \eta_{2.1}, n_{2.1}$, where the first two quantities are allowed to depend on all the parameters *except* M, and the last two quantities are allowed to depend on all the parameters.

We let

$$L' := L_{2.7}(\xi, \epsilon/2, \delta, K_1, K_2, K_3, 2M); \quad \gamma_{2.1} := \gamma = \min\{\gamma_{2.7}(\xi, \epsilon/2, \delta, K_1, K_2, K_3), \gamma_{2.11}(\xi)\}/4;$$

note that $\gamma_{2.1} = \gamma$ does not depend on M. We choose

$$\widetilde{R} := C_{2.6}(1/4, \delta/2, K_1, K_2); \quad L_{2.1} := 16\widetilde{R};$$

note that $L_{2.1}$ does not depend on M, as desired. We choose q to be the smallest positive integer for which

$$0.95^q L' \leq 16\widetilde{R}.$$

Now, let

$$\eta_{2.1} = \eta_{2.11}(p, \gamma n/2q, \max\{L', 16\widetilde{R}\}, \widetilde{R}, 2M),$$

and suppose that $f \in \ell^1(\mathbb{Z})$ with $\|f\|_1 = 1$, and that $\log_2 f$ is $\eta_{2.1}$-Lipschitz.

Step 1: Let $\ell := \lceil (1-\gamma)n \rceil$. Since $N \le \exp((H(\vec{p}) - \epsilon)n)$, it follows from Proposition 2.7 and the choice of parameters that, as long as $\|\vec{s} - \vec{p}\ell\|_\infty \le 4\gamma\ell$, then for all sufficiently large n,

$$\mathbb{P}[\|f_{\mathcal{A},\vec{s},\ell}\|_\infty \ge L'(N\sqrt{n})^{-1}] \le \exp(-2Mn).$$

Let $\mathcal{E}_0$ be the event that $\|f_{\mathcal{A},\vec{s},\ell}\|_\infty < L'(N\sqrt{n})^{-1}$ simultaneously for all $\vec{s}$ satisfying $\|\vec{s}-\vec{p}\ell\|_\infty \le 4\gamma\ell$. Then, by the union bound, we see that

$$\mathbb{P}[\mathcal{E}_0^c] \le n^k \exp(-2Mn).$$

Step 2: We split the interval $[\ell+1, n]$ into q subintervals of size $\gamma n/q$ each, which we denote by $I_1, \dots, I_q$. Note that

$$f_{\mathcal{A},\vec{s},n}(t) = \mathbb{E}_b f_{\mathcal{A},\vec{s'},\ell}\bigg(t + \sum_{i=1}^{q}\sum_{j\in I_i} b_j X_j\bigg);$$

here, we have sampled a uniform point in the multislice $\#\{b_i = a_j\} = s_j \ \forall j \in [k]$ by first sampling from the distribution of its last γn coordinates, which we denote by $b = (b_{\ell+1}, \dots, b_n)$, and then sampling the remaining coordinates, subject to the constraint that the amounts of each value a_j is in total equal to s_j. For each $j \in [k]$ let s'_j be the number of values in b equal to a_j (which is therefore a random variable).

Note that if $s_j \in [p_j n - \gamma n, p_j n + \gamma n]$, then we always have $s'_j \in [p_j\ell - 2\gamma\ell, p_j\ell + 2\gamma\ell]$. In particular, on the event $\mathcal{E}_0$, we have for all $\vec{s}$ satisfying $\|\vec{s} - \vec{p}n\|_\infty \le \gamma n$ and for all possible realizations of $\vec{s'}$ that

$$\|f_{\mathcal{A},\vec{s'},\ell}\|_\infty < L'(N\sqrt{n})^{-1}.$$

Step 3: For $i \in [q]$, let $\vec{s}^{(i)}$ be the vector of values $s_j^{(i)} = \#\{u \in I_i : b_u = a_j\}$ for $j \in [k]$. Let $\mathcal{G}$ be the event that $\|\vec{s}^{(i)} - \vec{p}|I_i|\|_\infty \le \gamma|I_i|$ for all $i \in [q]$. Then, by a standard large deviation estimate, it follows that for all n sufficiently large, $\mathbb{P}[\mathcal{G}^c] \le n^{-1/2}$ (say). Hence, using the conclusion of the previous step, conditioning on the values $\vec{s}^{(1)}, \dots, \vec{s}^{(q)}$, and using the law of total probability, we see that on the event $\mathcal{E}_0$,

$$f_{\mathcal{A},\vec{s},n}(t) \le L'n^{-1/2} \cdot (N\sqrt{n})^{-1} + \sup_{\substack{\|\vec{s}^{(i)}/|I_i| - \vec{p}\|_\infty \le \gamma \\ \text{for all } i \in [q]}} \mathbb{E}_b f_{\mathcal{A},\vec{s'},\ell}\bigg(t + \sum_{i=1}^{q}\sum_{u\in I_i} b_u X_j\bigg), \tag{2.10}$$

where each vector $(b_u)_{u\in I_i}$ is independently sampled uniformly from the multislice corresponding to $\vec{s}^{(i)}$. Fix vectors $\vec{s}^{(1)}, \dots, \vec{s}^{(q)}$ such that $\|\vec{s}^{(i)}/|I_i| - \vec{p}\|_\infty \le \gamma$ and $\|\vec{s}^{(i)}\|_1 = |I_i|$. In particular, this fixes $\vec{s'} = \vec{s} - \vec{s}^{(1)} - \dots - \vec{s}^{(q)}$. We define the sequence of functions $(g_r)_{r=0}^{q}$ by

$$g_0(t) = f_{\mathcal{A},\vec{s'},\ell}(t)$$

$$g_r(t) = \mathbb{E}_b\bigg[f_{\mathcal{A},\vec{s'},\ell}\bigg(t + \sum_{i=1}^{r}\sum_{j\in I_i} b_j X_j\bigg)\bigg|\#\{u \in I_i : b_u = a_j\} = \vec{s}_j^{(i)} \ \forall j \in [k], i \in [r]\bigg] \text{ for } r \ge 1.$$

Note that for all $r \in [q]$,

$$g_r(t) = \mathbb{E}_b\bigg[g_{r-1}\bigg(t + \sum_{j\in I_k} b_j X_j\bigg)\bigg|\#\{u \in I_r : b_u = a_j\} = \vec{s}_j^{(r)} \ \forall j \in [k]\bigg].$$

Step 4: We now wish to apply Proposition 2.11 to $g_0, \dots, g_{q-1}$ successively with parameters $L_0, \dots, L_{q-1}$ given by

$$L_r := L' \cdot (19/20)^r.$$

More precisely, for $r \geq 1$, let $\mathcal{E}_r$ denote the event that $\|g_r\|_\infty \leq L_r(N\sqrt{n})^{-1}$. We claim that

$$\mathbb{P}[\mathcal{E}_r | \mathcal{E}_{r-1}] \geq 1 - \exp(-2Mn) \text{ for all } r \in [q]. \tag{2.11}$$

Let us quickly check that on the event $\mathcal{E}_{r-1}$, the hypotheses of Proposition 2.11 are satisfied for g_{r-1}. We have $\|g_{r-1}\|_1 = 1$ and $\log_2 g_{r-1}$ is $\eta_{2.1}$-Lipschitz since it is a convex combination of functions satisfying the same properties. The condition $\|g_{r-1}\|_\infty \leq L_{r-1}(N\sqrt{n})^{-1}$ holds on $\mathcal{E}_{r-1}$ by definition. Moreover, the condition that for any interval I of size N,

$$\int_{t \in I} g_{r-1}(t) \leq \widetilde{R}/\sqrt{n}$$

follows since, by Lemma 2.6 and our choice of $\widetilde{R}$, the analogous property holds for $f_{\mathcal{A},\vec{s'},\ell}$, and hence for g_{r-1}, which is a convex combination of translates of $f_{\mathcal{A},\vec{s'},\ell}$. Proposition 2.11 now justifies (2.11). In particular, by the union bound, we have

$$\mathbb{P}[\mathcal{E}_q \mid \mathcal{E}_0] \geq 1 - q\exp(-2Mn).$$

Combine this with the estimate on $\mathcal{P}[\mathcal{E}_0^c]$ (with an at most n^{kq} sized union bound to account for the choice of $\vec{s}^{(1)}, \dots, \vec{s}^{(q)}$), then use (2.10), and finally take n sufficiently large (so that all the quoted results hold). This gives the desired conclusion. $\square$

2.5. **Independent model.** We conclude this section with an analogue of Corollary 2.3 in the independent case.

Theorem 2.14. *Fix a discrete distribution ξ. For $0 < \delta < 1/4$, $K_3 > K_2 > K_1 > 1$, $\epsilon \ll \|\vec{p}\|_\infty$, and a given parameter $M \geq 1$, there is $L_{2.14} = L_{2.14}(\xi, \epsilon, \delta, K_1, K_2, K_3) > 0$ independent of M and $n_{2.14} = n_{2.14}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$ such that the following holds.*

Let $n \geq n_{2.14}$, $1 \leq N \leq \|\vec{p}\|_\infty^{-n} \exp(-\epsilon n)$ and $\mathcal{A}$ be $(N, n, K_1, K_2, K_3, \delta)$-admissible. Then

$$\left| \left\{ x \in \mathcal{A} : \mathcal{L}_\xi\left(\sum_{i=1}^n b_i x_i, \sqrt{n} \right) \geq L_{2.14} N^{-1} \right\} \right| \leq e^{-Mn}|\mathcal{A}|.$$

The proof of Theorem 2.14 is analogous to that of Theorem 2.1 followed by Corollary 2.3, except that the random variables b_i are now independent copies of ξ. This independence simplifies matters dramatically, as one can derive an analogue of Proposition 2.7 by simply considering drops (as in [25, Proposition 4.5]) instead of "well-conditioned" drops, and then using subsampling arguments similar to those appearing above to prove analogues of Proposition 2.11 and Theorem 2.1. We leave the details to the interested reader.

3. Sharp invertibility of sparse Bernoulli matrices

In this section, we use Theorem 2.1 to confirm Conjecture 1.1 for the case $\xi = \text{Ber}(p)$ for fixed $p \in (0, 1/2)$, thereby resolving [15, Problem 8.2]. More precisely, we will show the following.

Theorem 3.1. *Fix $p \in (0, 1/2)$. There exist constants $C_p, \epsilon_p, n_p > 0$ such that for all $n \geq n_p$ and $t \geq 0$,*

$$\mathbb{P}[s_n(B_n(p)) \leq t/\sqrt{n}] \leq C_p t + (2 + (1 - \epsilon_p)^n) n (1 - p)^n.$$

Compared to the proof of Theorem 1.2, the main difference in this section is the substantially simpler treatment of structured vectors, due to the reasons mentioned in the introduction. At the same time, the arguments used in this section form the basis of developments in subsequent sections, and we hope that encountering them in this simpler setting will clarify their role later in the paper.

3.1. **Almost-constant and almost-elementary vectors.** We recall the usual notion of almost-constant vectors, a modification of compressible vectors (see, e.g., [26]).

Definition 3.2 (Almost-constant vectors)**.** For $\delta, \rho \in (0,1)$, we define $\mathrm{Cons}(\delta,\rho)$ to be the set of $x \in \mathbb{S}^{n-1}$ for which there exists some $\lambda \in \mathbb{R}$ such that $|x_i - \lambda| \le \rho/\sqrt{n}$ for at least $(1-\delta)n$ values $i \in [n]$.

The main result of this subsection is the following.

Proposition 3.3. *For any $c > 0$, there exist $\delta, \rho, \epsilon, n_0 > 0$ depending only on c, such that for all $n \ge n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta,\rho) : \|B_n(p)x\|_2 \le 2^{-cn}] \le n(1-p)^n + (1-p-\epsilon)^n.$$

For later use, we record the following simple property of non-almost-constant vectors.

Lemma 3.4. *For $\delta, \rho \in (0, 1/4)$, there exist $\nu, \nu' > 0$ depending only on δ, ρ, and a finite set $\mathcal{K}$ of* positive *real numbers, also depending only on δ, ρ, such that if $x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta,\rho)$, then at least one of the following two conclusions is satisfied.*

(1) There exist $\kappa, \kappa' \in \mathcal{K}$ such that

$$|x_i| \le \frac{\kappa}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n], \text{ and}$$

$$\frac{\kappa + \nu'}{\sqrt{n}} < |x_i| \le \frac{\kappa'}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n].$$

(2) There exist $\kappa, \kappa' \in \mathcal{K}$ such that

$$\frac{\kappa}{\sqrt{n}} < x_i < \frac{\kappa'}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n], \text{ and}$$

$$-\frac{\kappa'}{\sqrt{n}} < x_i < -\frac{\kappa}{\sqrt{n}} \text{ for at least } \nu n \text{ indices } i \in [n].$$

Proof. Let $I_0 := \{i \in [n] : |x_i| \le 4/\sqrt{\delta n}\}$. Since $\|x\|_2 = 1$, it follows that $|I_0| \ge (1-\delta/16)n$. We consider the following cases.

Case I: $|\{i \in I_0 : |x_i| < \rho/(10\sqrt{n})\}| \ge \delta n/16$. Since $x \notin \mathrm{Cons}(\delta,\rho)$, there are at least δn indices $j \in [n]$ such that $|x_j| \ge \rho/\sqrt{n}$. Moreover, at least $\delta n/8$ of these indices satisfy $|x_j| \le 4/\sqrt{\delta n}$. Hence, in this case, the first conclusion is satisfied for suitable choice of parameters.

Case II: $|\{i \in I_0 : |x_i| < \rho/(10\sqrt{n})\}| < \delta n/16$ and $|\{i \in I_0 : x_i \le -\rho/(10\sqrt{n})\}| \ge \delta n/16$ and $|\{i \in I_0 : x_i \ge \rho/(10\sqrt{n})\}| \ge \delta n/16$. In this case, the second conclusion is clearly satisfied for suitable choice of parameters.

Case III: Either $|\{i \in I_0 : x_i \ge \rho/(10\sqrt{n})\}| \ge (1-\delta/4)n$ or $|\{i \in I_0 : x_i \le -\rho/(10\sqrt{n})\}| \ge (1-\delta/4)n$. We assume that we are in the first sub-case; the argument for the second sub-case is similar. We decompose

$$[0, 4/\sqrt{\delta n}] = \cup_{\ell=1}^{L} J_\ell,$$

where $J_\ell := [(\ell-1)\cdot\rho/(10\sqrt{n}), \ell\cdot\rho/(10\sqrt{n}))$, and $L = O_{\delta,\rho}(1)$. Let

$$\ell_0 := \min\{\ell \in L : |\{i \in [n] : x_i \in J_1 \cup \cdots \cup J_\ell\}| \ge \delta n/16\}.$$

Since $x \notin \mathrm{Cons}(\delta,\rho)$, there are at least δn indices $j \in [n]$ such that $|x_j - \ell_0 \cdot \rho/(10\sqrt{n})| > \rho/\sqrt{n}$. On the other hand, by the assumption of this case and the definition of ℓ_0, there exist at least $\delta n - (\delta n/4) - (\delta n/16) - (\delta n/16) > \delta n/2$ indices $j \in I_0$ for which $x_j > (\ell_0 + 2)\cdot\rho/(10\sqrt{n})$. Thus, the first conclusion is satisfied for suitable choice of parameters. $\square$

We also isolate the following set of almost-elementary vectors.

Definition 3.5 (Almost-elementary vectors)**.** For $\delta > 0$ and $i \in [n]$, let

$$\mathrm{Elem}_i(\delta) := \mathbb{S}^{n-1} \cap \mathbb{B}_2^n(e_i, \delta) = \{x \in \mathbb{S}^{n-1} : \|x - e_i\|_2 \le \delta\}.$$

We define the set of δ-almost-elementary vectors by

$$\mathrm{Coord}(\delta) := \bigcup_{i=1}^{n} \mathrm{Elem}_i(\delta).$$

We will need a standard concentration estimate for the operator norm of a random matrix with independent centered sub-Gaussian entries.

Lemma 3.6 ([27, Lemma 4.4.5])**.** *There exists an absolute constant $C > 0$ such that the following holds. Let A be an $m \times n$ i.i.d. matrix with mean 0, sub-Gaussian entries with sub-Gaussian norm at most K. Then for any $t \ge 0$ we have*

$$\mathbb{P}[\|A\| \le C(\sqrt{m} + \sqrt{n} + t)] \le 2\exp(-t^2/K^2).$$

To prove Proposition 3.3, we will handle almost-elementary vectors and non-almost-elementary vectors separately. We begin with the easier case of non-almost-elementary vectors.

3.1.1. *Invertibility on non-almost-elementary vectors.* Using Lemma 2.4, we can prove the following elementary fact about sums of independent $\mathrm{Ber}(p)$ random variables.

Lemma 3.7. *Fix $p, \delta \in (0, 1/2)$. There exists $\theta = \theta(\delta, p) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Coord}(\delta)$,*

$$\mathcal{L}(b_1x_1 + \cdots + b_nx_n, \theta) \le 1 - p - \theta,$$

where $b = (b_1, \ldots, b_n)$ is a random vector whose coordinates are independent $\mathrm{Ber}(p)$ random variables.

Proof. Since $\mathrm{Coord}(\delta)$ is increasing with δ, it suffices to prove the statement for all sufficiently small δ (depending on p). We may assume that $|x_1| \ge |x_2| \ge \cdots \ge |x_n|$. The desired conclusion follows by combining the following two cases.

Case 1: Suppose $|x_2| > \delta^4$. We claim that there is some $\theta = \theta(\delta, p)$ for which

$$\mathcal{L}(b_1x_1 + \cdots + b_nx_n, \theta) \le \mathcal{L}(b_1x_1 + b_2x_2, \theta) \le 1 - p - \theta.$$

We borrow elements from [15, Proposition 3.11]. The first inequality is trivial. For the second inequality, we note that the random variable $b_1x_1 + b_2x_2$ is supported on the four points $\{0, x_1, x_2, x_1 + x_2\}$. Moreover, the sets $\{0, x_1 + x_2\}$ and $\{x_1, x_2\}$ are δ^4-separated, and each of these two sets is attained with probability at most $\max\{p^2 + (1-p)^2, 2p(1-p)\} < 1 - p - \theta$, where the final inequality uses $p < 1/2$.

Case 2: $|x_2| \le \delta^4$. Note that we must have $\|(x_2, \ldots, x_n)\|_2 \ge \delta/2$, since otherwise, we would have $\|x - e_1\|_2 < \delta$, contradicting $x \notin \mathrm{Coord}(\delta)$. We claim that there is some $\theta = \theta(\delta, p) > 0$ such that

$$\mathcal{L}(b_1x_1 + \cdots + b_nx_n, \theta) \le \mathcal{L}(b_2x_2 + \cdots + b_nx_n, \theta) \le 1 - p - \theta.$$

Once again, the first inequality is trivial. The second inequality follows from Lemma 2.4 applied with $\xi_i = b_ix_i$ for $i = 2, \ldots, n$, $r_i = |x_i|/4$, and $r = \delta^4$, and from our assumption that δ was small enough in terms of p. ∎

By combining the preceding statement with a standard net argument exploiting the low metric entropy of almost-constant vectors and the well-controlled operator norm of random matrices with independent centered subgaussian entries (Lemma 3.6), we obtain invertibility on non-almost-elementary vectors.

Proposition 3.8. *Fix $p \in (0,1/2)$. Then, for any $\delta' > 0$, there exist $\delta, \rho, \epsilon', n_0 > 0$ (depending on p, δ') such that for all $n \geq n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta,\rho) \setminus \mathrm{Coord}(\delta') : \|B_n(p)x\|_2 \leq \epsilon' \sqrt{n}] \leq (1-p-\epsilon')^n.$$

Proof. The argument is closely related to the proof of [25, Proposition 3.6], and we omit the standard details. The point is that we can first choose ϵ' depending on δ', p, then choose ρ sufficiently small depending on δ', p, ϵ', and finally, choose δ sufficiently small depending on all prior choices. □

3.1.2. *Invertibility on almost-elementary vectors.* We now prove the much more delicate claim that $\mathrm{Coord}(\delta')$ contributes the appropriate size to the singularity probability.

Proposition 3.9. *Fix $p \in (0,1/2)$. Given $\theta' > 0$, there exist $\delta', \theta, n_0 > 0$ depending on p and θ' such that for all $n \geq n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Coord}(\delta') : \|B_n(p)x\|_2 \leq \exp(-\theta' n)] \leq n \cdot \bigg((1-p)^n + (1-p-\theta)^n\bigg).$$

Before proceeding to the proof, we need the following preliminary fact, which essentially follows from the seminal work of Rudelson and Vershynin [21] (although we were not able to locate the precise statement needed here in the literature).

Lemma 3.10. *Fix $p \in (0,1)$. For any $c > 0$, there exist $c', n_0 > 0$ depending on c and p for which the following holds. For all $n \geq n_0$ and for any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} : \|Ax - v\|_2 \leq 2^{-cn}] \leq 2^{-c'n},$$

where A is a random $n \times (n-1)$ matrix with independent $\mathrm{Ber}(p)$ entries.

Proof. By reindexing the coordinates, we may write

$$A = \begin{bmatrix} R \\ B \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v' \end{bmatrix},$$

where B is an $(n-1) \times (n-1)$ matrix, $v' \in \mathbb{R}^{n-1}$ and $\|v'\|_2 \geq 1/2$. Let $\mathcal{E} = \{s_{n-1}(B) \leq 2^{-cn/2}\}$. Then, by an extension of the main result of Rudelson and Vershynin [21] (see, e.g., [8, Theorem 1.3] for a concrete reference), $\mathbb{P}[\mathcal{E}] \leq 2^{-c_1 n}$ for some $c_1 > 0$ depending on c and p.

Moreover, on the event $\mathcal{E}^c$, if there exists some $x \in \mathbb{R}^{n-1}$ such that $\|Ax - v\|_2 \leq 2^{-cn}$, then

$$\|Bx - v'\|_2 \leq 2^{-cn} \implies \|x - B^{-1}v'\|_2 \leq 2^{-cn/2}, \quad \text{and}$$
$$|Rx - v_1| \leq 2^{-cn} \implies |R(B^{-1}v') - v_1| \leq 2^{-cn} + n2^{-cn/2}.$$

Let $x_0 := B^{-1}v'/\|B^{-1}v'\|_2$; this is a random vector depending on B. It follows from a straightforward modification of the argument of Rudelson and Vershynin that there exists a constant $c_2 > 0$, depending on c and p, such that with probability at least $1 - 2^{-c_2 n}$, $\mathcal{L}(\sum_{i=1}^{n-1} b_i \cdot (x_0)_i, 2^{-cn/4}) \leq c_2^{-1} 2^{-c_2 n}$, where $b_1, \dots, b_n$ are independent $\mathrm{Ber}(p)$ random variables (again, one has to take into account that $\mathrm{Ber}(p)$ does not have mean 0, but this is not an issue, see, e.g., [8]). Let $\mathcal{G}$ denote the event (depending on B) that this occurs.

Finally, note that since $\|B\| \leq n$ deterministically, we have $\|B^{-1}v'\|_2 \geq 1/(2n)$. Hence, we see that for all n sufficiently large depending on c and p, we have

$$\begin{aligned}\mathbb{P}[\mathcal{E}^c \wedge \|Ax - v\|_2 \leq 2^{-cn}] &\leq \mathbb{P}[\mathcal{E}^c \wedge |R(B^{-1}v') - v_1| \leq 2n \cdot 2^{-cn/2}] \\ &\leq \mathbb{P}[\mathcal{E}^c \wedge \mathcal{G} \wedge |R(B^{-1}v') - v_1| \leq 2n \cdot 2^{-cn/2}] + 2^{-c_2 n} \\ &\leq \mathcal{L}\bigg(\sum_{i=1}^{n-1} b_i \cdot (x_0)_i, 4n^2 \cdot 2^{-cn/2}\bigg) + 2^{-c_2 n} \leq 2^{-c_2 n/2},\end{aligned}$$

which completes the proof. □

Now we are ready to prove Proposition 3.9.

Proof of Proposition 3.9. By taking a union bound, and exploiting the permutation invariance of the distribution of the matrix, it suffices to prove the statement for $\mathrm{Elem}_1(\delta')$ and without the additional factor of n on the right hand side. Let $\mathcal{G}_K$ denote the event that $\|B_n(p) - pJ_{n\times n}\| \leq K\sqrt{n}$, where $J_{n\times n}$ denotes the $n \times n$ all ones matrix. We fix a choice of K such that $\mathbb{P}[\mathcal{G}_K^c] \leq \exp(-2n)$, which is possible by Lemma 3.6.

For $\delta' \in (0, 1/4)$, let $\mathcal{E}_{\delta'}$ denote the event that there exists some $x \in \mathrm{Elem}_1(\delta')$ such that $\|B_n(p)x\|_2 \leq \exp(-\theta' n)$. By rescaling, we see that on the event $\mathcal{E}_{\delta'}$, there exists $y = e_1 + u \in \mathbb{R}^n$ with $u_1 = 0$ and $\|u\|_2 \leq 4\delta'$ for which $\|B_n(p)y\|_2 \leq 2\exp(-\theta' n)$. For convenience, let $u' := (u_2, \dots, u_n) \in \mathbb{R}^{n-1}$. Writing

$$B_n(p) = \begin{bmatrix} b_{11} & R \\ S & B_{n-1} \end{bmatrix},$$

we see that

$$|Ru' + b_{11}| \leq 2\exp(-\theta' n), \qquad \|B_{n-1}u' + S\|_2 \leq 2\exp(-\theta' n).$$

Let $B^{(1)}$ denote the first column of $B_n(p)$ and let $B^{(-1)}$ denote the $n \times (n-1)$ matrix formed by excluding the first column of $B_n(p)$. Then, on the event $\mathcal{E}_{\delta'} \wedge \mathcal{G}_K$, we have

$$\begin{aligned}\|B^{(1)} + pJ_{n\times n-1}u'\|_2 &= \|B^{(1)} + B^{(-1)}u' - B^{(-1)}u' + pJ_{n\times n-1}u'\|_2 \\ &\leq \|B^{(1)} + B^{(-1)}u'\|_2 + \|(B^{(-1)} - pJ_{n\times n-1})u'\|_2 \\ &\leq 4\exp(-\theta' n) + K\sqrt{n}\cdot 4\delta' \\ &\leq 8K\delta'\sqrt{n}.\end{aligned}$$

The key point is the following. Let $\mathcal{C} := \{x \in \{0,1\}^n : \exists \lambda \in \mathbb{R} \text{ such that } \|x - \lambda 1_n\|_2 \leq 8K\delta'\sqrt{n}\}$, where 1_n denotes the n-dimensional all ones vector. Then, it is readily seen that for any $\epsilon > 0$, there exists $\delta' > 0$ sufficiently small so that

$$\mathbb{P}[B^{(1)} \in \mathcal{C}] \leq (1 - p + \epsilon)^n.$$

To summarize, we have shown that for any $\theta', \epsilon > 0$, there exists $\delta' \in (0, 1/4)$ such that

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{\delta'}] &\leq \mathbb{P}[\mathcal{E}_{\delta'} \wedge \mathcal{G}_K] + \exp(-2n) \\ &\leq \sum_{a\in\mathcal{C}} \mathbb{P}[B^{(1)} = a]\cdot\mathbb{P}[\exists u' \in \mathbb{R}^{n-1}, \|u'\|_2 \leq 4\delta' : \|B^{(-1)}u' + a\| \leq 2\exp(-\theta' n)]. \end{aligned} \tag{3.1}$$

If we only wanted a bound of the form $(1 - p + \epsilon)^n$ on the right hand side, then we would be done. However, since we want a more precise bound, we need to perform a more refined analysis based on whether or not $a = 0$.

Case I: $a = 0$. The contribution of this term to the sum in (3.1) is exactly $(1 - p)^n$.

Case II: $a \neq 0$. Since $a \in \{0,1\}^n$, we have $\|a\|_2 \geq 1$. In this case, we can apply Lemma 3.10 with $c = \theta'$ to see that, for all n sufficiently large,

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1}, \|u'\|_2 \leq 4\delta' : \|B^{(-1)}u' + a\| \leq 2\exp(-\theta' n)] \leq 2^{-c'n},$$

where $c' > 0$ depends only on θ' and p. Thus, we see that the contribution of $a \in \mathcal{C}$, $a \neq 0$ to the sum in (3.1) is at most

$$(1 - p + \epsilon)^n \times 2^{-c'n}.$$

Since c' does not depend on δ', we can fix c' (depending on θ' and p), and then choose δ' sufficiently small so that $\epsilon > 0$ is small enough to make the above product at most $(1 - p - \theta)^n$, for some $\theta > 0$ depending on the previous parameters. ∎

We now put everything together to prove Proposition 3.3.

Proof of Proposition 3.3. Let $c > 0$ be as in the statement of the proposition, and choose $\theta' > 0$ such that $\exp(-\theta' n) = 2^{-cn}$. Then, applying Proposition 3.9 with this choice of θ', we find δ', θ, n_0 such that for all $n \geq n_0$, we have the desired estimate for $x \in \mathrm{Coord}(\delta')$ (provided that ϵ is chosen small enough). Then, we apply Proposition 3.8 with this choice of δ' to find $\delta, \rho, n_0 > 0$ such that for all $n \geq n_0$, we have the desired estimate for $x \in \mathrm{Cons}(\delta, \rho) \setminus \mathrm{Coord}(\delta')$, provided again that we choose $\epsilon > 0$ sufficiently small. □

3.2. **The structure theorem for Boolean slices.** The following is a natural extension of the threshold function appearing in [25] to the Boolean slice.

Definition 3.11. Let $p \in (0, 1/2]$, $\gamma \in (0, p)$, and $L \geq 1$. Then, for any integer $n \geq 1$ and $x \in \mathbb{S}^{n-1}$, we define

$$\mathcal{T}_{p,\gamma}(x, L) := \sup\left\{t \in (0,1) : \mathcal{L}_{p,\gamma}\left(\sum_{i=1}^n b_i x_i, t\right) > Lt\right\}.$$

For $p \in (0, 1/2)$, let $H = H(p)$ denote an $(n-1) \times n$ random matrix, each of whose entries is an independent copy of a $\mathrm{Ber}(p)$ random variable. We fix a function $v(H)$ which takes as input an $(n-1) \times n$ matrix and outputs a unit vector in its right kernel. The goal of this subsection is to prove the following result about the threshold function of $v(H)$.

Proposition 3.12. *Let $\delta, \rho, \epsilon \in (0, 1)$. Let $p \in (0, 1/2)$ and let $H = H(p)$ denote an $(n-1) \times n$ random matrix as above. There exist $L_{3.12} = L_{3.12}(\delta, \rho, p, \epsilon)$, $\gamma_{3.12} = \gamma_{3.12}(\delta, \rho, p, \epsilon)$ and $n_{3.12} = n_{3.12}(\delta, \rho, p, \epsilon)$ such that for all $n \geq n_{3.12}$, with probability at least $1 - 4^{-n}$, exactly one of the following holds.*

- *$v(H) \in \mathrm{Cons}(\delta, \rho)$, or*
- *$\mathcal{T}_{p,\gamma_{3.12}}(v(H), L_{3.12}) \leq \binom{n}{pn}^{-1} \exp(\epsilon n)$.*

Remark. We note that $p \in (0, 1/2)$ is not actually necessary for this statement or its proof, and this is crucial in the more general setting of all discrete random variables.

We will need the following lemma, proved using randomized rounding (cf. [16]), which is a straightforward generalization of [25, Lemma 5.3]. We omit details since the proof is identical.

Lemma 3.13. *Let $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ be a vector, and let $\mu > 0$, $\lambda \in \mathbb{R}$ be fixed. Let Δ denote a probability distribution which is supported in $[-s, s]^n$. There exist constants $c_{3.13}$ and $C_{3.13}$, depending only on s, for which the following holds. Suppose that for all $t \geq \sqrt{n}$,*

$$\mathbb{P}\left[\left|\sum_{i=1}^n b_i y_i - \lambda\right| \leq t\right] \leq \mu t,$$

where $(b_1, \ldots, b_n)$ is distributed according to Δ. Then, there exists a vector $y' = (y_1', \ldots, y_n') \in \mathbb{Z}^n$ satisfying

(R1) $\|y - y'\|_\infty \leq 1$,
(R2) $\mathbb{P}[|\sum_{i=1}^n b_i y_i' - \lambda| \leq t] \leq C_{3.13}\mu t$ *for all* $t \geq \sqrt{n}$,
(R3) $\mathcal{L}(\sum_{i=1}^n b_i y_i', \sqrt{n}) \geq c_{3.13}\mathcal{L}(\sum_{i=1}^n b_i y_i, \sqrt{n})$,
(R4) $|\sum_{i=1}^n y_i - \sum_{i=1}^n y_i'| \leq C_{3.13}\sqrt{n}$.

We also record two useful tensorization statements.

Lemma 3.14 ([25, Lemma 3.2])**.** *Let $\chi_1, \ldots, \chi_m$ be independent random variables.*

- *Assume that for all $\epsilon \geq \epsilon_0$,*

$$\mathbb{P}[|\chi_i| \leq \epsilon] \leq K\epsilon.$$

Then for $\epsilon \geq \epsilon_0$
$$\mathbb{P}[\|(\chi_1, \ldots, \chi_m)\|_2 \leq \epsilon\sqrt{m}] \leq (CK\epsilon)^m,$$
where C is an absolute constant.

- *Assume that for some $\eta, \tau > 0$,*
$$\mathbb{P}[|\chi_i| \leq \eta] \leq \tau.$$
Then for $\epsilon \in (0,1]$,
$$\mathbb{P}[\|(\chi_1, \ldots, \chi_m)\|_2 \leq \eta\sqrt{\epsilon m}] \leq (e/\epsilon)^{\epsilon m}\tau^{m-\epsilon m}.$$

Proof of Proposition 3.12. For lightness of notation, we will often denote $v(H)$ simply by v. If $v \notin \mathrm{Cons}(\delta, \rho)$, it follows from Lemma 2.5 that for all $\gamma < p/4$, there exists some $L_0 = L_0(\delta, \rho, p)$ and $C_0 = C_0(\delta, \rho, p)$ such that $\mathcal{T}_{p,\gamma}(v, L) \leq C_0 \cdot n^{-1/2}$ for all $L \geq L_0$. Fix K_0 such that the event $\mathcal{E}_{K_0} := \{\|H - pJ_{n-1\times n}\| \leq K_0\sqrt{n}\}$ holds with probability at least $1 - 2^{-1729n}$.

Let $L > L_0$ be a parameter to be chosen later, depending on $\delta, \rho, p, \epsilon$. Let $\gamma < 1/4$ be a parameter to be chosen later depending on $\delta, \rho, p, \epsilon$. Fix some $N \in [C_0^{-1} \cdot \sqrt{n}, \binom{n}{pn}\exp(-\epsilon n)]$, and let $\mathcal{U}_N$ denote the event that $\mathcal{T}_{p,\gamma}(v, L) \in [1/N, 2/N]$. We proceed to bound $\mathbb{P}[\mathcal{U}_N \wedge \mathcal{E}_{K_0}]$.

Let $D := C_1\sqrt{n}N$, where $C_1 = C_1(\delta, \rho) \geq 1$ will be an integer chosen later. Let $y := Dv$. Since for all $t \geq \sqrt{n}$,
$$\begin{aligned}\mathbb{P}\left[\left|\sum_{i=1}^n b_i y_i\right| \leq t\right] &= \mathbb{P}\left[\left|\sum_{i=1}^n b_i v_i\right| \leq \frac{t}{C_1\sqrt{n}N}\right] \\ &\leq \mathbb{P}\left[\left|\sum_{i=1}^n b_i v_i\right| \leq \frac{t}{\sqrt{n}N}\right] \leq \frac{L}{N} \cdot \frac{2t}{\sqrt{n}},\end{aligned}$$
it follows that by applying Lemma 3.13 to y, with $\mu := 2L/(N\sqrt{n})$, $\lambda = 0$, and the distribution on $\mathbb{R}^n$ coinciding with that of n independent $\mathrm{Ber}(p)$ random variables conditioned to have sum in $[pn - \gamma n, pn + \gamma n]$, we see that for all sufficiently large n, there exists some $y' \in \mathbb{Z}^n$ satisfying the conclusions of Lemma 3.13 (note that $C_{3.13}, c_{3.13}$ in this case are absolute constants). By (R3), we have
$$\begin{aligned}\mathcal{L}\bigg(\sum_{i=1}^n b_i y_i', \sqrt{n}\bigg) &\geq c_{3.13}\mathcal{L}\bigg(\sum_{i=1}^n b_i v_i, 1/(C_1N)\bigg) \\ &\geq (2C_1)^{-1} \cdot c_{3.13}\mathcal{L}\bigg(\sum_{i=1}^n b_i v_i, 2/N\bigg) \\ &\geq (2C_1)^{-1} \cdot c_{3.13} \cdot 2LN^{-1}. \end{aligned} \tag{3.2}$$
Moreover, by (R1) and (R4), we have on the event $\mathcal{E}_{K_0}$ that
$$\begin{aligned}\|Hy'\|_2 &= \|H(y' - y)\|_2 \\ &\leq \|(H - pJ_{n-1\times n})(y' - y)\|_2 + \|pJ_{n-1\times n}(y' - y)\|_2 \\ &\leq K_0'n, \end{aligned} \tag{3.3}$$
where K_0' is a constant depending only on K_0.

We claim that there is an absolute constant $C_2 > 0$ and a constant $C_3 = C_3(\delta, \rho)$, a collection of real numbers (depending on δ, ρ) $(K_3)_j > (K_2)_j > (K_1)_j > 1$ for $j \in [C_3 \cdot C_2^n]$, a positive real number $\delta' > 0$ (depending on δ, ρ), and a collection of $(N, n, (K_1)_j, (K_2)_j, (K_3)_j, \delta')$-admissible sets $\mathcal{A}_j$ (depending on δ, ρ) for $j \in [C_3 \cdot C_2^n]$ such that $y' \in \mathcal{A}_j$ for some $j \in [C_3 \cdot C_2^n]$. Let $\widetilde{v} := y'/D$. By (R1), it follows that $\|\tilde{v} - v\|_\infty \leq D^{-1}$. Moreover, by Lemma 3.4, there exist ν, ν' depending on δ, ρ, and a finite set $\mathcal{K}$ of positive real numbers, also depending on δ, ρ, such that either the first conclusion or the second conclusion of Lemma 3.4 is satisfied for v. Since $D^{-1} \le C_1^{-1}C_0/n$, we see that there exists n_0 depending on δ, ρ, p such that for all $n \ge n_0$, $\widetilde{v}$ satisfies either the first conclusion or the second conclusion of Lemma 3.4, with $\nu/2, \nu'/2$ and $2^{-1} \cdot \mathcal{K} \cup 2 \cdot \mathcal{K}$. After paying an overall factor of at most 2^n, we may assume that the νn coordinates of $\widetilde{v}$ satisfying this conclusion are the first νn coordinates. The remaining $(1-\nu)n$ coordinates of $\widetilde{v}$ lie in the $(1-\nu)n$-dimensional ball of radius 1. By a volumetric argument, we see that this ball can be covered by at most 100^n translates of $[0, n^{-1/2}]^{(1-\nu)n}$. By paying an overall factor of 100^n, we may fix the translate of $[0, n^{-1/2}]^{(1-\nu)n}$ that the remaining $(1-\nu)n$ coordinates lie in. Note that each such translate contains at most $(2D/\sqrt{n})^{(1-\nu)n}$ points in $(1/D)\mathbb{Z}^n$. Finally, taking $C_1(\delta, \rho)$ sufficiently large so that $C_1(\delta, \rho) \cdot \min(2^{-1} \cdot \mathcal{K}) > 1$ and rescaling by D proves the claim.

To summarize, we have so far shown the following. For parameters L and γ depending on $\delta, \rho, p, \epsilon$ (to be chosen momentarily), on the event $\mathcal{U}_N \wedge \mathcal{E}_{K_0}$, the event $\mathcal{B}_j$ holds for some $j \in [C_2 \cdot C_3^n]$, where $\mathcal{B}_j$ is the event that there exists some $y' \in \mathcal{A}_j$ satisfying (3.2), (3.3), and (by (R2)),

$$\mathbb{P}\left[\left|\sum_{i=1}^n b_i y'_i\right| \le t\right] \le C_{3.13}\mu t \text{ for all } t \ge \sqrt{n}, \tag{3.4}$$

where recall that $\mu = 2L/(N\sqrt{n})$.

We are now ready to specify the parameters L and γ. First, let

$$L' := \max_j L_{2.3}(\mathrm{Ber}(p), \epsilon, \delta', (K_1)_j, (K_2)_j, (K_3)_j); \gamma' := \min_j \gamma_{2.3}(\mathrm{Ber}(p), \epsilon, \delta', (K_1)_j, (K_2)_j, (K_3)_j)/4.$$

Then, let

$$L := (2C_1) \cdot c_{3.13}^{-1} \cdot L' + L_0; \gamma := \gamma'.$$

Our goal is to bound $\mathbb{P}[\cup_j \mathcal{B}_j]$. Let $H_1, \dots, H_{n-1}$ denote the rows of H. By a standard large deviation estimate, we can find an absolute constant $Q \ge 1$ such that the event

$$\mathcal{W}_Q := \{|\{i \in [n-1] : \sum_{j=1}^n H_{i,j} \notin [pn - \gamma n, pn + \gamma n]\}| \le Q\}$$

holds with probability at least $1 - 2^{-1729n}$. Then, it suffices to bound $\mathbb{P}[\cup_j(\mathcal{B}_j \wedge \mathcal{W}_Q)]$. We will provide a uniform (in j) upper bound on $\mathbb{P}[\mathcal{B}_j \wedge \mathcal{W}_Q]$, and then conclude using the union bound. Note that on the event $\mathcal{B}_j$, y' belongs to the set $\mathcal{D}_j$ defined by

$$\mathcal{D}_j := \left\{x \in \mathcal{A}_j : \mathcal{L}_{p,\gamma}\left(\sum_{i=1}^n b_i x_i, \sqrt{n}\right) \ge LN^{-1}\right\}.$$

By the choice of L, it follows from Corollary 2.3 that for any $M \ge 1$, for all sufficiently large n,

$$|\mathcal{D}_j| \le e^{-Mn}|\mathcal{A}_j| \le e^{-Mn}(K_3 N)^n, \tag{3.5}$$

where $K_3 := \max_j (K_3)_j$. Moreover, it follows from (3.4) and Lemma 3.14 that

$$\mathbb{P}[\{\|Hy'\|_2 \le K'_0 n\} \wedge \mathcal{W}_Q] \le \left(\frac{C_4 L K'_0}{N}\right)^{n-Q}, \tag{3.6}$$

where $C_4 \ge 1$ is an absolute constant. Here, we have used that on the event $\mathcal{W}_Q$, the entries of at least $n - Q$ rows have sum in $[pn - \gamma n, pn + \gamma n]$.

Finally, from (3.5) and (3.6), we see that first taking M to be sufficiently large (compared to various constants depending on $\delta, \rho, p, \epsilon$), and then taking n sufficiently large, gives the desired conclusion. □

3.3. **Proof of Theorem 3.1.** We now have all the ingredients needed to prove Theorem 3.1. The proof uses the insight from [15] of exploiting the exponential gap between $\binom{n}{pn}$ and $(1-p)^n$ for $p < 1/2$ by using a 'row boosting' argument to reduce to an anticoncentration problem on a well-conditioned slice.

Proof of Theorem 3.1. Throughout, we fix functions $x(A), y(A)$ which take as input a matrix A and output (fixed, but otherwise arbitrary) right and left least singular unit vectors, respectively. Let $B = B_n(p)$ for simplicity. Fix $\epsilon > 0$ such that $\binom{n}{pn}\exp(\epsilon n) \le (1-p-\epsilon)^n$.

Step 1: By the work of Rudelson and Vershynin [21], there is some $c_p > 0$ so that for all $t > 2^{-2c_p n}$,

$$\mathbb{P}[s_n(B) \le t/\sqrt{n}] \le C_p t;$$

note that there is a slight complication since $\mathrm{Ber}(p)$ is not centered, but this can be handled using standard techniques (see, e.g., [8, Theorem 1.3]) Therefore, it suffices to consider the case $t \le 2^{-2c_p n}$.

Step 2: For $\delta, \rho \in (0,1)$, we define

$$\mathcal{E}_L(\delta,\rho) = \{\exists y \in \mathrm{Cons}(\delta,\rho) : \|y(B)^T B\|_2 \le 2^{-c_p n}\},$$
$$\mathcal{E}_R(\delta,\rho) = \{\exists x \in \mathrm{Cons}(\delta,\rho) : \|Bx(B)\|_2 \le 2^{-c_p n}\}.$$

Applying Proposition 3.3 with $c_p > 0$, we find that there exist $\delta, \rho, \epsilon' > 0$ such that for all sufficiently large n,

$$\mathbb{P}[s_n(B) \le t/\sqrt{n}] \le 2n(1-p)^n + 2(1-p-\epsilon')^n + \mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L(\delta,\rho)^c \wedge \mathcal{E}_R(\delta,\rho)^c].$$

Here, we have used that the distribution of B is invariant under transposition.

Step 3: Let $\gamma = \gamma_{3.12}(\delta,\rho,p,\epsilon)$. Let $W_\gamma \subseteq \{0,1\}^n$ denote the set of vectors $x \in \{0,1\}^n$ such that $\sum_{i=1}^n x_i \in [pn - \gamma n, pn + \gamma n]$. As in the proof of Proposition 3.12, let $Q \ge 1$ be a constant such that the event

$$\mathcal{W}_Q := \{|\{i \in [n] : B_i \notin W_\gamma\}| \le Q\}$$

holds with probability at least $1 - 2^{-1729n}$. Then, it suffices to bound $\mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge \mathcal{W}_Q]$, where for simplicity, we have omitted the parameters δ, ρ fixed in the previous step.

Let $B_1, \dots, B_n$ denote the rows of B, and for simplicity, let $y = y(B)$. On the event that $s_n(B) \le t/\sqrt{n}$, we have

$$\|y_1 B_1 + \cdots + y_n B_n\|_2 \le t/\sqrt{n}.$$

Moreover, on the event $\mathcal{E}_L^c$, using Lemma 3.4, there is a set $I \subseteq [n]$ such that $|I| \ge \nu n$ and such that for all $i \in I$, $|y_i| \ge \kappa/\sqrt{n}$, for some $\kappa := \kappa(\delta,\rho) > 0$. In particular, since for any $i \in [n]$, $\|y_1 B_1 + \cdots + y_n B_n\|_2 \ge |y_i|\,\mathrm{dist}(B_i, H_i)$, where H_i denotes the span of rows $B_1, \dots, B_{i-1}, B_{i+1}, \dots B_n$, it follows that

$$\mathrm{dist}(B_i, H_i) \le \frac{t}{\kappa} \text{ for all } i \in I.$$

Also, on the event $\mathcal{W}_Q$, there are at least $\nu n/2$ indices $i \in I$ such that $B_i \in W_\gamma$. Thus, we see that

$$\mathbb{P}[s_n(B) \le t/\sqrt{n} \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge \mathcal{W}_Q] \le \frac{2}{\nu n}\sum_{i=1}^n \mathbb{P}[\mathrm{dist}(B_i, H_i) \le t/\kappa \wedge \mathcal{E}_L^c \wedge \mathcal{E}_R^c \wedge B_i \in W_\gamma].$$

Step 4: By symmetry, it suffices to bound $\mathbb{P}[\mathcal{B}_1]$, where

$$\mathcal{B}_1 := \mathrm{dist}(B_1, H_1) \le t/\kappa \wedge \mathcal{E}_R^c \wedge B_1 \in W_\gamma.$$

Let $v(H_1)$ be a unit vector normal to H_1. Then, by Proposition 3.12, except with probability 4^{-n} (over the randomness of H_1), exactly one of the following holds.

- $v(H_1) \in \mathrm{Cons}(\delta,\rho)$, or
- $\mathcal{T}_{p,\gamma}(v(H_1), L) \le \binom{n}{pn}^{-1}\exp(\epsilon n)$,

where $L := L_{3.12}(\delta, \rho, p, \epsilon)$. If the first possibility occurs, then $\mathcal{B}_1$ cannot hold, since then, $v(H_1) \in \mathrm{Cons}(\delta, \rho)$ satisfies

$$\|Bv(H_1)\|_2 = |\langle B_1, v(H_1)\rangle| \le \mathrm{dist}(B_1, H_1) \le t/\kappa \le 2^{-2c_p n}/\kappa,$$

which contradicts $\mathcal{E}_R^c$ for all n sufficiently large. Hence, the second possibility must hold. But then, using $\mathrm{dist}(B_1, H_1) \ge |\langle B_1, v(H_1)\rangle|$, we have that (over the randomness of B_1),

$$\begin{aligned}\mathbb{P}[\mathrm{dist}(B_1, H_1) \le t/\kappa \wedge B_1 \in W_\gamma] &\le \mathbb{P}[|\langle B_1, v(H_1)\rangle| \le t/\kappa \mid B_1 \in W_\gamma] \\ &\le \frac{Lt}{\kappa} + \binom{n}{pn}^{-1}\exp(\epsilon n) \\ &\le \frac{Lt}{\kappa} + (1-p-\epsilon)^n.\end{aligned}$$

This completes the proof. □

4. Non-almost-constant vectors

In this short section, we prove the following result, which controls the invertibility of random matrices with i.i.d. discrete entries on the bulk of the unit sphere. We note that this is a generalization of the discussion in Sections 3.2 and 3.3.

Theorem 4.1. *Fix a discrete distribution ξ. For any $\delta, \rho, \epsilon > 0$, there exists $C_{4.1} = C_{4.1}(\xi, \delta, \rho, \epsilon) > 0$ and $n_{4.1}(\xi, \delta, \rho, \epsilon) \ge 1$ such that for all $n \ge n_{4.1}$ and $t \ge 0$,*

$$\mathbb{P}\left[\inf_{x \in \mathbb{S}^{n-1}\setminus \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n} \wedge \inf_{y \in \mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 > C_{4.1}t\right] \le C_{4.1}t + \exp((\epsilon - H(\vec{p}))n).$$

Let ξ be a discrete distribution, and let $A = A(\xi)$ denote an $(n-1)\times n$ random matrix, each of whose entries is an independent copy of a ξ random variable. We fix a function $v(A)$ which takes as input an $(n-1)\times n$ matrix and outputs a unit vector in its right kernel. As in Section 3, a key ingredient in the proof of Theorem 4.1 is a structure theorem for kernel vectors of A, which encodes the fact that (with very high probability) non-almost-constant kernel vectors of A must be maximally unstructured in the relevant sense.

Definition 4.2. Fix a discrete distribution ξ. Let $\vec{\gamma} \in \mathbb{R}^k_{\ge 0}$ with $\|\vec{\gamma}\|_\infty < \min(\vec{p})$, and let $L \ge 1$. Then, for any integer $n \ge 1$ and $x \in \mathbb{S}^{n-1}$, we define

$$\mathcal{T}_{\xi,\vec{\gamma}}(x, L) := \sup\left\{t \in (0,1) : \mathcal{L}_{\xi,\vec{\gamma}}\left(\sum_{i=1}^n b_i x_i, t\right) > Lt\right\}.$$

We also define

$$\mathcal{T}_{\xi}(x, L) := \sup\left\{t \in (0,1) : \mathcal{L}_{\xi}\left(\sum_{i=1}^n b_i x_i, t\right) > Lt\right\}.$$

Proposition 4.3. *Let $\delta, \rho, \epsilon \in (0,1)$ and $k = |\operatorname{supp}(\xi)|$. There exist $L_{4.3} = L_{4.3}(\delta, \rho, \xi, \epsilon)$, $\gamma_{4.3} = \gamma_{4.3}(\delta, \rho, \xi, \epsilon)$ and $n_{4.3} = n_{4.3}(\delta, \rho, \xi, \epsilon)$ such that for all $n \ge n_{4.3}$, with probability at least $1 - k^{-2n}$, exactly one of the following holds.*

- *$v(A) \in \mathrm{Cons}(\delta, \rho)$, or*
- *$\mathcal{T}_{\xi,\gamma_{4.3}1_k}(v(A), L_{4.3}) \le \exp((\epsilon - H(\vec{p}))n)$.*

Proposition 4.3 follows from Corollary 2.3 and Lemma 3.13 in an identical fashion to the proof given in Section 3.2, so we do not repeat it here. The only minor difference is that in the last part of the proof, we now restrict ourselves to the event that all but $O_{\xi,\gamma_{2.3}}(1)$ rows belong to a well-conditioned multislice corresponding to ξ (instead of simply restricting to well-conditioned slices).

Given this, we can complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Let $M := M_n(\xi)$ for simplicity, and let $\delta, \rho, \epsilon > 0$ be as in the statement of the theorem. Let $k = |\operatorname{supp}(\xi)|$ and denote the points in the support of ξ by $a_1, \dots, a_k$. We will denote the columns of M by $M^{(1)}, \dots, M^{(n)}$. Also, for each $i \in [n]$, $M^{(-i)}$ denotes the subspace spanned by all columns of M except for $M^{(i)}$.

Step 1: Let $\gamma = \gamma_{4.3}(\delta, \rho, \xi, \epsilon)$. Let $W_\gamma \subseteq \operatorname{supp}(\xi)^n$ denote the set of vectors $x \in \operatorname{supp}(\xi)^n$ such that $\#\{x_i = a_j\} \in [p_j n - \gamma n, p_j n + \gamma n]$ for all $j \in [k]$. Let $Q \geq 1$ be a constant such that the event

$$\mathcal{W}_Q := \{|\{i \in [n] : M^{(i)} \notin W_\gamma| \leq Q\}$$

holds with probability at least $1 - k^{-1729n}$. Then, it suffices to bound

$$\mathbb{P}\left[\inf_{x \in \mathbb{S}^{n-1} \setminus \operatorname{Cons}(\delta,\rho)} \|Mx\|_2 \leq t/\sqrt{n} \wedge \inf_{y \in \operatorname{Cons}(\delta,\rho)} \|yM\|_2 > Ct \wedge \mathcal{W}_Q\right]. \tag{4.1}$$

Let us denote the first of the three events in the equation above by $\mathcal{E}_R$, and the second event by $\mathcal{E}_L$.

Let $x = x(M)$ denote a vector in $\mathbb{S}^{n-1} \setminus \operatorname{Cons}(\delta, \rho)$ certifying the event $\mathcal{E}_R$, so that

$$\|x_1 M^{(1)} + \cdots + x_n M^{(n)}\|_2 \leq t/\sqrt{n}.$$

Using Lemma 3.4, there is a set $I \subseteq [n]$ such that $|I| \geq \nu n$ and such that for all $i \in I$, $|x_i| \geq \kappa/\sqrt{n}$, for some $\kappa := \kappa(\delta, \rho) > 0$. In particular, since for any $i \in [n]$, $\|x_1 M^{(1)} + \cdots + x_n M^{(n)}\|_2 \geq |x_i| \operatorname{dist}(M^{(i)}, M^{(-i)})$, it follows that

$$\operatorname{dist}(M^{(i)}, M^{(-i)}) \leq \frac{t}{\kappa} \text{ for all } i \in I.$$

Also, on the event $\mathcal{W}_Q$, there are at least $\nu n/2$ indices $i \in I$ such that $M^{(i)} \in W_\gamma$. Thus, we see that

$$(4.1) = \mathbb{P}[\mathcal{E}_R \wedge \mathcal{E}_L \wedge \mathcal{W}_Q] \leq \frac{2}{\nu n} \sum_{i=1}^{n} \mathbb{P}[\operatorname{dist}(M^{(i)}, M^{(-i)}) \leq t/\kappa \wedge \mathcal{E}_L \wedge M^{(i)} \in W_\gamma].$$

Step 2: By symmetry, it suffices to bound $\mathbb{P}[\mathcal{M}_1]$, where

$$\mathcal{M}_1 := \operatorname{dist}(M^{(1)}, M^{(-1)}) \leq t/\kappa \wedge \mathcal{E}_L \wedge M^{(1)} \in W_\gamma.$$

Let $v(M^{(-1)})$ be a unit vector normal to $M^{(-1)}$. Then, by Proposition 4.3, except with probability k^{-2n} (over the randomness of $M^{(-1)}$), exactly one of the following holds.

- $v(M^{(-1)}) \in \operatorname{Cons}(\delta, \rho)$, or
- $\mathcal{T}_{\xi, \gamma 1_k}(v(M^{(-1)}), L) \leq \exp((\epsilon - H(\vec{p}))n)$,

where $L := L_{4.3}(\delta, \rho, \xi, \epsilon)$. If the first possibility occurs, then $\mathcal{M}_1$ cannot hold as $v(M^{(-1)}) \in \operatorname{Cons}(\delta, \rho)$ satisfies

$$\|v(M^{(-1)})M\|_2 = |\langle M^{(1)}, v(M^{(-1)})\rangle| \leq \operatorname{dist}(M^{(1)}, M^{(-1)}) \leq t/\kappa \leq Ct,$$

(choosing C appropriately), which contradicts $\mathcal{E}_L$. Hence, the second possibility must hold. But then, using $\operatorname{dist}(M^{(1)}, M^{(-1)}) \geq |\langle M^{(1)}, v(M^{(-1)})\rangle|$, we have that (over the randomness of $M^{(1)}$),

$$\begin{aligned}\mathbb{P}[\operatorname{dist}(M^{(1)}, M^{(-1)}) \leq t/\kappa \wedge M^{(1)} \in W_\gamma] &\leq \mathbb{P}[|\langle M^{(1)}, v(M^{(-1)})\rangle| \leq t/\kappa \mid M^{(1)} \in W_\gamma] \\ &\leq \frac{Lt}{\kappa} + \exp((\epsilon - H(\vec{p}))n)).\end{aligned}$$

□

5. Preliminary invertibility estimates

In this section, we will prove a version of Theorem 1.6 with the weaker singularity estimate $(\|\vec{p}\|_\infty + o_n(1))^n$. This estimate, which generalizes [25, Theorem A], will be used crucially in our refined treatment of invertibility for almost-constant vectors in the next section. The techniques in this section also serve as a gentle warm-up to the next section, where much more involved versions of the arguments are presented.

We begin with the following elementary fact regarding sums of ξ random variables.

Lemma 5.1. *Fix a discrete distribution ξ. There is $\theta = \theta(\xi) > 0$ such that for all $x \in \mathbb{S}^{n-1}$,*

$$\mathcal{L}_\xi(b_1x_1 + \cdots + b_nx_n, \theta) \le \|\vec{p}\|_\infty.$$

Proof. This is essentially identical to the proof given in [25, Lemma 3.5]. Briefly, if $\|x\|_\infty \ge \delta$, then we can choose θ small enough (depending on δ and ξ) so the claim is immediate. Otherwise $\|x\|_\infty < \delta$ and $\|x\|_2 = 1$, in which case the claim follows from Lemma 2.4 as long as δ is sufficiently small depending on ξ. □

Combining the above estimate with the second part of Lemma 3.14, we have the following.

Corollary 5.2. *Fix a discrete distribution ξ. For every $\epsilon > 0$, there exists $c > 0$ depending on ϵ and ξ such that for any $x \in \mathbb{S}^{n-1}$ and $y \in \mathbb{R}^n$, we have*

$$\mathbb{P}[\|M_n(\xi)x - y\|_2 \le c\sqrt{n}] \le (\|\vec{p}\|_\infty + \epsilon)^n.$$

Moreover, combining this corollary with the low metric entropy of $\mathrm{Cons}(\delta,\rho)$ and Lemma 3.6, we obtain the following (weak) estimate for invertibility on almost-constant vectors.

Corollary 5.3. *Fix a discrete distribution ξ. For every $\epsilon > 0$, there exist $\delta, \rho, c > 0$ depending on ϵ and ξ such that for any $y \in \mathbb{R}^n$,*

$$\mathbb{P}\left[\inf_{x \in \mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x - y\|_2 \le c\sqrt{n}\right] \le (\|\vec{p}\|_\infty + \epsilon)^n.$$

Next, we show that with very high probability, the inverse of any fixed vector is unstructured.

Proposition 5.4. *Fix a discrete distribution ξ. For every $\epsilon, \eta > 0$, there exist $\delta, \rho, L > 0$ depending on ϵ, η, ξ such that for any $y \in \mathbb{R}^n$,*

$$\mathbb{P}\left[\exists x \in \mathbb{S}^{n-1} : M_n(\xi)x \parallel y \wedge x \in \mathrm{Cons}(\delta,\rho) \vee \mathcal{T}_\xi(x, L) \ge (\|\vec{p}\|_\infty + \eta)^n\right] \le (\|\vec{p}\|_\infty + \epsilon)^n.$$

Proof. This follows essentially from combining Corollary 5.3 with a cruder analogue of Proposition 4.3, the only difference being that we are considering $M_n(\xi)x \parallel y$ for arbitrary $y \in \mathbb{R}^n$ as opposed to only $y = 0$.

To handle this last point, we begin by choosing (using Lemma 3.6) a sufficiently large constant K so that $\mathcal{E}_K = \{\|M_n(\xi) - \mathbb{E}[\xi]J_{n\times n}\| \le K\sqrt{n}\}$ satisfies $\mathbb{P}[\mathcal{E}_K^c] \le \|\vec{p}\|_\infty^{2n}$. Then, it suffices to restrict to $\mathcal{E}_K$. Moreover, by the triangle inequality, we see that on the event $\mathcal{E}_K$, $\|M_n(\xi)\| \le K\sqrt{n} + \mathbb{E}[\xi]n$, so that in particular, on the event in the proposition (intersected with $\mathcal{E}_K$), we have that $M_n(\xi)x = ty_0$ with $y_0 \in \mathbb{S}^{n-1}$ fixed and for some $t \in \mathbb{R}$ with $|t| \le K\sqrt{n} + \mathbb{E}[\xi]n$.

Now, for the treatment of vectors in $\mathrm{Cons}(\delta,\rho)$, we can divide the range of t into n^3 uniformly spaced intervals, apply Corollary 5.3 with y equal to the mid-point of an interval times y_0, and use the union bound. For the treatment of vectors x satisfying $(x \in \mathbb{S}^{n-1} \setminus \mathrm{Cons}(\delta,\rho)) \wedge \mathcal{T}_\xi(x, L) \le (\|\vec{p}\|_\infty + \eta)^n$, we divide the range of t into $\|\vec{p}\|_\infty^{-2n}$ equally spaced intervals, use a slight generalization of the argument in the proof of Proposition 3.12 with M sufficiently large (depending on ξ) for each y equal to the mid-point of an interval times y_0, and finally use the union bound. We leave the details to the interested reader. □

Using Corollary 5.3 and Proposition 5.4, we can prove the following weaker version of Theorem 1.6.

Theorem 5.5. *Let ξ be a discrete random variable. For any $\epsilon > 0$, there exist $C, n_0 > 0$ depending on ξ, ϵ such that for all $n \geq n_0$ and $t \geq 0$,*

$$\mathbb{P}[s_n(M_n) \leq t/\sqrt{n}] \leq Ct + (\|\vec{p}\|_\infty + \epsilon)^n.$$

Proof. The deduction of this theorem follows from the argument in [25, Section 5] with the application of Corollary 5.3 and Proposition 5.4 at the appropriate steps. A similar deduction appears in Section 3.3 and a more complicated version of this deduction also appears in Section 7, so we omit the details. □

6. Almost-constant vectors

The goal of this section is to prove Theorem 1.7. The proof is presented at the end of the section and needs a few intermediate steps.

For the proof, we will need to isolate the following natural class of almost-elementary vectors.

Definition 6.1. (Almost-elementary vectors) For $\delta > 0$ and $i, j \in [n]$, $i \neq j$, let

$$\begin{aligned} \mathrm{Elem}_i(\delta) &:= \{x \in \mathbb{S}^{n-1} : \|x - e_i\|_2 \leq \delta\}, \\ \mathrm{Elem}_{i,j}(\delta) &:= \{x \in \mathbb{S}^{n-1} : \|x - (e_i - e_j)/\sqrt{2}\|_2 \leq \delta\}, \\ \mathrm{Elem}'_{i,j}(\delta) &:= \{x \in \mathbb{S}^{n-1} : \|x - (e_i + e_j)/\sqrt{2}\|_2 \leq \delta\}. \end{aligned}$$

Also, let

$$\begin{aligned} \mathrm{Elem}(\delta) &:= \bigcup_{i \in [n]} \mathrm{Elem}_i(\delta) \cup \bigcup_{i,j \in [n], i \neq j} \mathrm{Elem}_{i,j}(\delta), \\ \mathrm{Elem}'(\delta) &:= \mathrm{Elem}(\delta) \cup \bigcup_{i,j \in [n], i \neq j} \mathrm{Elem}'_{i,j}(\delta). \end{aligned}$$

Note that $\bigcup_{i \in [n]} \mathrm{Elem}_i(\delta)$ is exactly the set $\mathrm{Coord}(\delta)$ defined in Section 3.1.

For excluding almost-constant vectors which are not almost-elementary, we will need to develop sharp results regarding the Lévy concentration function of discrete random variables.

Proposition 6.2. *Fix a discrete distribution ξ and $\delta \in (0, 1/2)$. There exists $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \mathrm{Elem}'(\delta)$,*

$$\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \leq \|\vec{p}\|_2^2 - \theta.$$

Proof. Since $\mathrm{Elem}'(\delta)$ is increasing with δ, it suffices to prove the statement for sufficiently small δ (depending on ξ), which will be chosen during the course of the proof. Moreover, we may assume that $|x_1| \geq |x_2| \geq \cdots \geq |x_n|$.

Since $x \notin \mathrm{Elem}_1(\delta)$, we must have $\|(x_2, \ldots, x_n)\|_2 \geq \delta/2$. In case $|x_2| \leq \delta^4$, then we are done using Lemma 2.4 (cf. the proof of Lemma 3.7) for all sufficiently small δ. Similarly, if $\|(x_3, \ldots, x_n)\|_2 \geq \delta/4$ and $|x_3| \leq \delta^4$, we are done. We now analyze the remaining situations via case analysis.

Case I: $\delta^4 \leq |x_2| < (1 - \delta^5)|x_1|$. Since $\mathcal{L}_\xi(b_1 x_1 + \cdots + b_n x_n, \theta) \leq \mathcal{L}_\xi(b_1 x_1 + b_2 x_2, \theta)$, it suffices to bound the latter. Let ξ' be an independent copy of ξ. For any $s \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{P}[x_1 \xi + x_2 \xi' \in [s - c, s + c]]^2 &= \left(\sum_a \mathbb{P}[\xi' = a] \mathbb{P}[|\xi - x_1^{-1}(s - x_2 a)| \leq c|x_1|^{-1}] \right)^2 \\ &\leq \left(\sum_a \mathbb{P}[\xi' = a]^2 \right) \left(\sum_a \mathbb{P}[|\xi - x_1^{-1}(s - x_2 a)| \leq c|x_1|^{-1}]^2 \right) \leq \|\vec{p}\|_2^4, \end{aligned}$$

where the sum is over $a \in \operatorname{supp}(\xi)$. Here, the equality is by definition, the first inequality is Cauchy-Schwarz, and the last inequality holds as long as $c > 0$ is chosen small enough in terms of δ, ξ. Let us elaborate on this final point. We choose $c > 0$ small enough so that $c|x_1|^{-1} \le c\delta^{-4}$ is smaller than $|x_2/x_1|$ times half the minimum gap in $\operatorname{supp}(\xi)$, which is possible since $|x_2/x_1| \ge \delta^4$. Now, such a choice of c clearly implies that each summand in $\sum_a \mathbb{P}[|\xi - x_1^{-1}(s - x_2 a)| \le c|x_1|^{-1}]^2$ covers at most a single atom in $\operatorname{supp}(\xi)$, and that different choices of $a, a' \in \operatorname{supp}(\xi)$ cover distinct atoms in $\operatorname{supp}(\xi)$.

Moreover, for such a choice of c, equality in the final inequality holds if and only if there is a permutation σ on $\operatorname{supp}(\xi)$ such that for all $a \in \operatorname{supp}(\xi)$,

$$\mathbb{P}[\xi' = \sigma(a)] = \mathbb{P}[|\xi - x_1^{-1}(s - x_2 a)| \le c|x_1|^{-1}].$$

Summing over all the atoms in $\operatorname{supp}(\xi)$, we see that if equality holds in the final inequality, then

$$\operatorname{supp}(\xi) \subseteq \bigcup_{j=1}^{k} [x_1^{-1}(s - x_2 a) - c|x_1|^{-1}, x_1^{-1}(s - x_2 a) + c|x_1|^{-1}],$$

so that in particular, $\operatorname{supp}(\xi)$ is contained in an interval of length at most $|x_2/x_1| m_\xi + 2c|x_1|^{-1}$, where $m_\xi = \max \operatorname{supp}(\xi) - \min \operatorname{supp}(\xi)$. But since $|x_2/x_1| \le 1 - \delta^5$ and $c|x_1|^{-1} \le c\delta^{-4}$, we see (by taking $c > 0$ sufficiently small) that $\operatorname{supp}(\xi)$ is contained in an interval of length at most $(1 - \delta^5/2) m_\xi$, which contradicts the definition of m_ξ. Hence, we see that equality cannot hold in the final inequality.

Since equality does not hold, it follows from the above discussion that (for $c > 0$ sufficiently small), we have the stronger inequality

$$\mathbb{P}[x_1 \xi + x_2 \xi' \in [s - c, s + c]]^2 \le \|\vec{p}\|_2^2 (\|\vec{p}\|_2^2 - (\min \vec{p})^2),$$

which completes the analysis in this case, noting that the choice of c depends only on ξ, δ.

Case II: $|x_2| \ge (1 - \delta^5)|x_1|, \|(x_3, \ldots, x_n)\|_2 \le \delta/4$. This implies that $x \in \mathrm{Elem}'_{1,2}(\delta) \cup \mathrm{Elem}_{1,2}(\delta)$, thereby violating our assumption.

Case III: $\delta^4 \le |x_3| \le (1 - \delta^5)|x_1|$. This can be treated in exactly the same way as Case I.

Case IV: $(1 - \delta^5)|x_1| \le |x_3| \le |x_2|$ and $|x_2| \ge \delta^4$. It suffices to bound $\mathcal{L}_\xi(b_1 x_1 + b_2 x_2 + b_3 x_3)$. Let $u_i \in \{\pm 1\}$ be defined via $u_i = \operatorname{sgn}(x_i) = x_i/|x_i|$. Let $m'_\xi > 0$ be the smallest positive real such that $\operatorname{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$.

We begin by noting that for any $s \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}[x_1 \xi_1 + x_2 \xi_2 + x_3 \xi_3 \in [s - c, s + c]] \\
&\qquad = \mathbb{P}[|x_1|(u_1 \xi_1 + |x_1|^{-1}|x_2| u_2 \xi_2 + |x_1|^{-1}|x_3| u_3 \xi_3) \in [s - c, s + c]] \\
&\qquad \le \mathbb{P}[|x_1|(u_1 \xi_1 + u_2 \xi_2 + u_3 \xi_3) \in [s - c - 3\delta^5 m'_\xi, s + c + 3\delta^5 m'_\xi]],
\end{aligned}$$

where the inequality uses $(1 - \delta^5) \le |x_1|^{-1}|x_3| \le |x_1|^{-1}|x_2| \le 1$, $|x_1| \le 1$, and the definition of m'_ξ.

Since $|x_1| \ge |x_2| \ge \delta^4$, this localizes the value of $u_1 \xi_1 + u_2 \xi_2 + u_3 \xi_3$ to an interval of length at most $2(c\delta^{-4} + 3\delta m'_\xi)$. As discussed at the beginning, we can assume that δ is sufficiently small based on ξ. By first choosing $\delta > 0$ sufficiently small depending on ξ, and then choosing $c > 0$ sufficiently small depending on δ and ξ, we may assume that $2(c\delta^{-4} + 3\delta m'_\xi)$ is smaller than the minimum distance between two distinct atoms in both $\operatorname{supp}(\xi + \xi' + \xi'')$ and $\operatorname{supp}(\xi + \xi' - \xi'')$, where ξ, ξ', ξ'' are independent copies of ξ. Note that, after possibly multiplying by an overall negative sign, $u_1 \xi_1 + u_2 \xi_2 + u_3 \xi_3$ is distributed as either $\xi + \xi' + \xi''$ or $\xi + \xi' - \xi''$.

Therefore, by our choice of δ and c, we see that it suffices to show that for all $s \in \mathbb{R}$,

$$\mathbb{P}[\xi_1 + \xi_2 + \xi_3 = s] \le \|\vec{p}\|_2^2 - c_\xi, \quad \mathbb{P}[\xi_1 + \xi_2 - \xi_3 = s] \le \|\vec{p}\|_2^2 - c_\xi,$$

for some $c_\xi > 0$ depending only on ξ. Now for $u_3 \in \{\pm 1\}$, we have

$$\begin{aligned}
\mathbb{P}[\xi_1 + \xi_2 + u_3\xi_3 = s]^2 &= \left(\sum_a \mathbb{P}[\xi_3 = a]\mathbb{P}[\xi_1 + \xi_2 = s - u_3 a]\right)^2 \\
&\le \left(\sum_a \mathbb{P}[\xi_3 = a]^2\right)\left(\sum_a \mathbb{P}[\xi_1 + \xi_2 = s - u_3 a]^2\right) \\
&\le \left(\sum_a \mathbb{P}[\xi_3 = a]^2\right)\left(\sum_{a' \in \operatorname{supp}(\xi_1+\xi_2)} \mathbb{P}[\xi_1 + \xi_2 = a']^2\right) \\
&\le \|\vec{p}\|_2^4,
\end{aligned}$$

where the first line is by definition, the second line is Cauchy-Schwarz, and the last line follows by Young's convolution inequality. To obtain the inequality with a positive constant $c_\xi > 0$, we note that equality cannot hold in the third line since $\operatorname{supp}(\xi_1 + \xi_2)$ has strictly more positive atoms than $\operatorname{supp}(\xi)$ (since ξ is supported on at least 2 points), and this leads to the desired improvement since ξ has finite support. ☐

When ξ is not a translate of an origin-symmetric distribution, the above result can be strengthened.

Proposition 6.3. *Fix a discrete distribution ξ and $\delta \in (0, 1/2)$. Suppose that ξ is not a translate of any origin-symmetric distribution. Then, there exists $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \operatorname{Elem}(\delta)$,*

$$\mathcal{L}_\xi(b_1x_1 + \cdots + b_nx_n, \theta) \le \|\vec{p}\|_2^2 - \theta.$$

Proof. As before, since $\operatorname{Elem}(\delta)$ is increasing with δ, it suffices to prove the statement for sufficiently small δ depending on ξ. By Proposition 6.2, we can choose $\theta = \theta(\delta, \xi) > 0$ such that for all $x \in \mathbb{S}^{n-1} \setminus \operatorname{Elem}'(\delta)$,

$$\mathcal{L}_\xi(b_1x_1 + \cdots + b_nx_n, \theta) \le \|\vec{p}\|_2^2 - \theta.$$

Hence, it remains to prove the result for $x \in \operatorname{Elem}'(\delta) \setminus \operatorname{Elem}(\delta)$. By symmetry, it suffices to consider $x \in \operatorname{Elem}'_{1,2}(\delta)$. We will bound $\mathcal{L}_\xi(b_1x_1 + b_2x_2, \theta)$.

We use an argument similar to **Case IV** of the proof of Proposition 6.2. Let $m'_\xi > 0$ be the smallest positive real for which $\operatorname{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$. We have

$$\mathbb{P}[x_1\xi_1 + x_2\xi_2 \in [s - c, s + c]] \le \mathbb{P}\left[\frac{1}{\sqrt{2}}(\xi_1 + \xi_2) \in [s - c - 2m'_\xi\delta, s + c + 2m'_\xi\delta]\right].$$

Once again, by choosing δ and c sufficiently small (depending on ξ), we may assume that $2(c + 2m'_\xi\delta)$ is smaller than the minimum distance between two distinct atoms in $\operatorname{supp}(\xi + \xi')$, where ξ, ξ' are independent copies of ξ. With this choice of δ and c, the problem reduces to showing that there exists some $c_\xi > 0$ depending only on ξ such that for all $s \in \mathbb{R}$,

$$\mathbb{P}[\xi_1 + \xi_2 = s] \le \|\vec{p}\|_2^2 - c_\xi.$$

We have

$$\mathbb{P}[\xi_1 + \xi_2 = s] = \sum_a \mathbb{P}[\xi_1 = a]\mathbb{P}[\xi_2 = s - a] \le \left(\sum_a \mathbb{P}[\xi_1 = a]^2\right)^{1/2}\left(\sum_a \mathbb{P}[\xi_2 = s - a]^2\right)^{1/2} \le \|\vec{p}\|_2^2,$$

where the first inequality is Cauchy-Schwarz. To obtain the improved inequality with $c_\xi > 0$, we note that equality can hold in both inequalities if and only if $\mathbb{P}[\xi_1 = a] = \mathbb{P}[\xi_2 = s - a]$, which implies that ξ is a shift (by $s/2$) of an origin-symmetric random variable. Since we have assumed that ξ is not a shift of an origin-symmetric random variable, we see that equality cannot hold, and using that the support of ξ is finite, we can conclude. ☐

Using the preceding lemmas, and exploiting the low metric entropy of $\mathrm{Cons}(\delta, \rho)$ along with Lemma 3.6, we obtain the following corollary. Note that since ξ may not have mean 0, one must perform the standard trick of densifying the net of these vectors along the direction 1_n (see [25, Proposition 3.6]). We omit the (standard) proof; as in Proposition 3.8, it is closely related to the proof of [25, Proposition 3.6].

Corollary 6.4. *Fix a discrete distribution ξ. For all $\delta' > 0$, there exist $\delta, \rho, \epsilon', n_0 > 0$, depending on ξ and δ', such that for all $n \geq n_0$,*

$$\mathbb{P}[\exists x \in \mathrm{Cons}(\delta, \rho) \setminus \mathrm{Elem}'(\delta') : \|M_n(\xi)x\|_2 \leq \epsilon' \sqrt{n}] \leq (\|\vec{p}\|_2^2 - \epsilon')^n.$$

Further, if ξ is not a shift of any origin-symmetric random variable, then the same conclusion holds with $\mathrm{Elem}(\delta')$ *instead of* $\mathrm{Elem}'(\delta')$.

Given the previous corollary, it remains to analyze vectors in $\mathrm{Elem}'(\delta')$ (or only in $\mathrm{Elem}(\delta')$ if ξ is not a shift of any origin-symmetric random variable), which is the content of the remainder of this section.

6.1. **Two columns.** We first handle vectors in $\mathrm{Elem}_{i,j}(\delta')$. By the invariance of the distribution of $M_n(\xi)$ under permuting columns, it suffices to analyze vectors in $\mathrm{Elem}_{1,2}(\delta')$. We show the following.

Proposition 6.5. *Fix a discrete distribution ξ. There exist $\delta', \eta, n_0 > 0$ depending on ξ such that for all $n \geq n_0$ and $t \leq 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}_{1,2}(\delta') : \|M_n(\xi)x\|_2 \leq t] \leq \|\vec{p}\|_2^{2n} + (\|\vec{p}\|_2^2 - \eta)^n + t\exp(-\eta n).$$

We will need the following preliminary lemma, which essentially follows from the seminal work of Rudelson and Vershynin [21]. Since we were not able to locate the statement we need in the literature, we provide details below.

Lemma 6.6. *Fix $S, s > 0$. There exist $C', c', n_0 > 0$ depending on s, S such that the following holds. For all $n \geq n_0$, any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, any $\kappa \in (0,1)$, and all $t \leq 1$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} : \|Ax - v\|_2 \leq t] \leq C'n^3\sqrt{t}\exp(\kappa n) + \exp(-c'n)\exp(\kappa n),$$

where A is an $n \times (n-1)$ random matrix, each of whose entries is an independent random variable with sub-Gaussian norm at most S, and such that all but a collection of κn specified entries have variance at least s.

Proof. By the law of total probability, it suffices to assume that the κn specified entries are deterministic, and take the values $a_1, \ldots, a_{\kappa n}$. Consider the $n \times (n-1)$ random matrix A', which has the same distribution as A, except for the κn specified entries, which are now replaced by $a_1 + b_1, \ldots, a_{\kappa n} + b_{\kappa n}$, where $b_1, \ldots, b_{\kappa n}$ are independent $\mathrm{Ber}(1/2)$ random variables.

From a slight generalization of Lemma 3.10 (specifically, one should replace the application of [21] with an inhomogeneous version due to [17] and replace 2^{-cn} by t, see the proof of Lemma 6.9), we get that there exist C', c', n_0 depending on s, S such that for all $n \geq n_0$, for any $v \in \mathbb{R}^n$ with $\|v\|_2 \geq 1$, and for all $t \geq 1$, we have

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} : \|A'x - v\|_2 \leq t] \leq C'n^3\sqrt{t} + \exp(-c'n).$$

The conclusion now follows since, with probability $2^{-\kappa n}$, $b_1 = \cdots = b_{\kappa n} = 0$. ☐

We now prove Proposition 6.5.

Proof of Proposition 6.5. By Lemma 3.6, we can choose $K > 0$ depending on ξ such that $\mathbb{P}[\mathcal{E}_K] \le \|\vec{p}\|_2^{3n}$, where

$$\mathcal{E}_K := \{\|M_n(\xi) - \mathbb{E}[\xi]J_{n\times n}\| \le K\sqrt{n}\}.$$

For $\delta' \in (0, 1/4)$, which will be chosen later in terms of ξ, let

$$\mathcal{E} := \{\exists x \in \mathbb{B}_2^n(e_1, \delta') \cap \mathbb{S}^{n-1} : \|M_n(\xi)Qx\|_2 \le t\},$$

where Q is the rotation matrix whose bottom-right $(n-2)\times(n-2)$ minor is the identity matrix, and the top-left 2×2 minor is the rotation matrix given by

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Up to scaling δ' by a constant factor, this is clearly equivalent to the event that we wish to bound.

Note that on the event $\mathcal{E}$, there exists some vector $y = e_1 + u \in \mathbb{R}^n$ with $u_1 = 0$ and $\|u\|_2 \le 4\delta'$ such that

$$\|M_n(\xi)Qy\|_2 \le 2t.$$

Let $u' = (u_2, \dots, u_n) \in \mathbb{R}^{n-1}$, let $\widetilde{M}^{(1)}$ be the first column of $M_n(\xi)Q$, and let $\widetilde{M}^{(-1)}$ denote the $n \times (n-1)$ matrix obtained by removing this column. Then, on the event $\mathcal{E} \wedge \mathcal{E}_K$, we have

$$\begin{aligned} \|\widetilde{M}^{(1)} - \mathbb{E}[\xi]J_{n\times n-1}u'\|_2 &\le \|\widetilde{M}^{(1)} + \widetilde{M}^{(-1)}u'\|_2 + \|(\widetilde{M}^{(-1)} - \mathbb{E}[\xi]J_{n\times n-1})u'\|_2 \\ &\le 2t + K\sqrt{n}\cdot 4\delta' \\ &\le 8K\delta'\sqrt{n} \end{aligned}$$

for all sufficiently large n, since $t \le 1$.

The key point is the following. Let $\Xi := \operatorname{supp}(\xi - \xi')/\sqrt{2} \subseteq \mathbb{R}$. Let

$$\mathcal{C} := \{a \in \Xi^n : \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \le 8K\delta'\sqrt{n}\},$$

and for $\kappa = \kappa(\delta', \xi) > 0$, to be chosen later depending on δ', ξ, and for $z \in \Xi$, let

$$\mathcal{C}_z := \mathcal{C} \cap \{a \in \mathbb{R}^n : |\operatorname{supp}(a - z1_n)| \le \kappa n\}.$$

It is easy to see that

$$\mathcal{C} \subseteq \bigcup_{z\in\Xi} \mathcal{C}_z$$

for an appropriate choice of κ which goes to 0 as δ' goes to 0. Furthermore,

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}_z] \le \mathbb{P}[(\xi - \xi')/\sqrt{2} = z]^n \exp(c_{\kappa,\delta',\xi}n),$$

where $c_{\kappa,\delta',\xi} > 0$ goes to 0 as κ, δ' go to 0. Therefore, we have

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}] \le \|\vec{p}\|_2^{2n} \exp(2c_{\kappa,\delta',\xi}n), \text{ and}$$

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C} \setminus \mathcal{C}_0] \le (\|\vec{p}\|_2^2 - c_\xi)^n$$

for some $c_\xi > 0$ depending only on ξ, provided that δ' (hence κ) is chosen sufficiently small. Here, for the second inequality, we have used that by Cauchy-Schwarz (as in the proof of Proposition 6.2), the unique most probable atom of $(\xi - \xi')/\sqrt{2}$ is at 0, and is $\|\vec{p}\|_2^2$, so that any other atom in Ξ has probability at most $\|\vec{p}\|_2^2 - 2c_\xi$ for some $c_\xi > 0$.

So far, we have shown that for all κ and δ' sufficiently small (depending on ξ), we have

$$\begin{aligned} \mathbb{P}[\mathcal{E}] &\le \|\vec{p}\|_2^{3n} + \sum_{a\in\mathcal{C}} \mathbb{P}[\widetilde{M}^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|\widetilde{M}^{(-1)}u' + a\|_2 \le 2t|\widetilde{M}^{(1)}] \\ &\le \|\vec{p}\|_2^{3n} + (\|\vec{p}\|_2^2 - c_\xi)^n + \sum_{a\in\mathcal{C}_0} \mathbb{P}[\widetilde{M}^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|\widetilde{M}^{(-1)}u' + a\|_2 \le 2t|\widetilde{M}^{(1)}]. \end{aligned}$$

We proceed to bound the third term in the above sum.

Case I: If $a = 0$, we have $\mathbb{P}[\widetilde{M}^{(1)} = 0] = \|p\|_2^{2n}$.

Case II: If $a \neq 0$, we have in particular that $\|a\|_2 \geq h_\xi > 0$. The crucial observation is the following. Given $\widetilde{M}^{(1)} = a$, the entries of the first column of $\widetilde{M}^{(-1)}$ are independent random variables, each of which is distributed as the sum of two i.i.d. copies of $\xi/\sqrt{2}$, conditioned on knowing their difference. In particular, for the coordinates $i \in [n]$ for which $a_i = 0$, the corresponding coordinate of the first column of $\widetilde{M}^{(-1)}$ is distributed as $\sqrt{2} \cdot \xi^*$, where ξ^* has the same support as ξ but takes on atom a_i with probability proportional to p_i^2. Thus, we see that conditioned on $\widetilde{M}^{(1)} = a \in \mathcal{C}_0$, all entries of $\widetilde{M}^{(-1)}$ are independent with sub-Gaussian norm at most S_ξ, and all but at most κn entries have variance at least $s_\xi > 0$. Hence, by Lemma 6.6, and by using the lower bound $\|a\|_2 \geq h_\xi$, we find that there exist C', c', n_1 depending on ξ such that for all $n \geq n_1$,

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|\widetilde{M}^{(-1)} u' + a\|_2 \leq 2t | \widetilde{M}^{(1)}] \leq C' n^3 \sqrt{t} \exp(\kappa n) + \exp(-c'n) \exp(\kappa n).$$

Thus, the contribution of this case is at most

$$\|\vec{p}\|_2^{2n} \exp(2c_{\kappa,\delta',\xi} n) \exp(\kappa n) \left(C' n^3 \sqrt{t} + 2\exp(-c'n) \right).$$

By the AM-GM inequality, we have $\|\vec{p}\|_2^{2n} \sqrt{t} \leq t\|\vec{p}\|_2^n + \|\vec{p}\|_2^{3n}$. The desired conclusion now follows by taking $\eta > 0$ sufficiently small so that $\|\vec{p}\|_2^n \leq \exp(-2\eta n)$, and then taking δ' (hence κ) sufficiently small so that $2c_{\kappa,\delta',\xi} + \kappa < \min(c'/2, \eta/2)$. □

The preceding proposition handles vectors in $\mathrm{Elem}_{i,j}(\delta)$. If the distribution ξ is a translate of an origin-symmetric distribution, we also need to handle vectors in $\mathrm{Elem}'_{i,j}(\delta)$. In case the distribution ξ is itself an origin-symmetric distribution, the desired bound follows immediately from the previous proposition, using that the distribution of any column of $M_n(\xi)$ is invariant under negation in this case. Therefore, it remains to handle vectors in $\mathrm{Elem}'_{i,j}(\delta)$ when ξ is a nonzero translate of an origin-symmetric distribution, which is done by the next proposition.

Proposition 6.7. *Fix a discrete distribution ξ that is a* nonzero *translate of an origin-symmetric distribution. There exist $\delta', \eta, n_0 > 0$ depending on ξ such that for all $n \geq n_0$ and $t \leq 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}'_{1,2}(\delta') : \|M_n(\xi)x\|_2 \leq t] \leq (\|\vec{p}\|_2^2 - \eta)^n + t\exp(-\eta n).$$

Proof. The proof is essentially the same as that of Proposition 6.5. The lack of the "main term" $\|\vec{p}\|_2^2$ comes from the fact that $e_1 + e_2$ is unlikely to be a kernel vector since ξ is not origin-symmetric.

We quickly discuss the main modifications to the proof of Proposition 6.5. Throughout, $s \neq 0$ denotes a real number such that ξ and $s - \xi$ have the same distribution (such an s exists by our assumption about ξ). First, the top-left 2×2 minor of Q is now

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Next, we let $\Xi = \mathrm{supp}(\xi + \xi')/\sqrt{2}$ and as before, let

$$\mathcal{C} := \{a \in \Xi^n : \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \leq 8K\delta'\sqrt{n}\}$$

and for $z \in \Xi$,

$$\mathcal{C}_z := \mathcal{C} \cap \{a \in \mathbb{R}^n : |\mathrm{supp}(a - z1_n)| \leq \kappa n\},$$

where $\kappa = \kappa(\delta', \xi) > 0$ is chosen as in the previous argument. For such a choice of κ, we have

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C}] \leq \|\vec{p}\|_2^{2n} \exp(2c_{\kappa,\delta',\xi} n), \text{ and}$$

$$\mathbb{P}[\widetilde{M}^{(1)} \in \mathcal{C} \setminus \mathcal{C}_{s/\sqrt{2}}] \leq (\|\vec{p}\|_2^2 - c_\xi)^n.$$

This time the inequalities are derived as follows. We note that, by Cauchy-Schwarz, for any $z \in \Xi$, $\mathbb{P}[\xi + \xi' = z\sqrt{2}] \le \|\vec{p}\|_2^2$, with equality holding if and only if $\mathbb{P}[\xi = a] = \mathbb{P}[\xi' = z\sqrt{2} - a]$ for all $a \in \operatorname{supp}(\xi)$, which happens if and only if $z = s/\sqrt{2}$.

Using this, we have as before that

$$\mathbb{P}[\mathcal{E}] \le \|\vec{p}\|_2^{3n} + (\|\vec{p}\|_2^2 - c_\xi)^n + \sum_{a \in \mathcal{C}_{s/\sqrt{2}}} \mathbb{P}[\widetilde{M}^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|\widetilde{M}^{(-1)}u' + a\|_2 \le 2t|\widetilde{M}^{(1)}].$$

The most important detail is that for $\kappa \le 1/2$ (say), every $a \in \mathcal{C}_{s/\sqrt{2}}$ is nonzero, since it has at least $(1-\kappa)n$ coordinates equal to $s/\sqrt{2}$. Since s is a nonzero constant depending only on ξ, we can now use the analysis in **Case II** of the proof of Proposition 6.5. The final thing to note is that the distribution of the random variable $(\xi - \xi')/\sqrt{2}$, conditioned on $(\xi + \xi')/\sqrt{2} = s/\sqrt{2}$ coincides with the distribution of $(2\xi^* - s)/\sqrt{2}$, where ξ^* has the same support as ξ, but takes on atom a_i with probability proportional to p_i^2. The remaining details of the proof are essentially the same. □

6.2. **One column.** We now handle vectors in $\mathrm{Elem}_i(\delta')$. Once again, by permutation invariance, it suffices to handle $\mathrm{Elem}_1(\delta')$. We will prove the following.

Proposition 6.8. *Fix a discrete distribution ξ. There exist $C', \delta', \eta, n_0 > 0$ depending on ξ such that for all $n \ge n_0$ and $t \le 1$,*

$$\mathbb{P}[\exists x \in \mathrm{Elem}_1(\delta') : \|M_n(\xi)x\|_2 \le t] \le p_0^n + C't\exp(-\eta n) + (\|\vec{p}\|_2^2 - \eta)^n.$$

The analysis is more delicate than the two column case, since (i) we may have $p_0 < \|\vec{p}\|_\infty$, but we still want to isolate p_0 as the major contribution coming from these events, and (ii) we are aiming for an error term of $(\|\vec{p}\|_2^2 - \eta)^n$, which may be smaller than $(p_0 - \eta)^n$. However, given the preparation above, the rest of the proof is similar to the proof in the sparse Bernoulli case, isolated in Proposition 3.9, except that we need to replace the application of the results of Rudelson and Vershynin [21] with the much sharper Proposition 5.4 and Theorem 5.5.

We begin with the following proposition.

Lemma 6.9. *Fix a discrete distribution ξ. For any $\eta \in (0,1)$, there exist $C, n_0 > 0$ depending on ξ, η for which the following holds. For any $v \in \mathbb{R}^n$ with $\|v\|_2 \ge 1$, $n \ge n_0$, and $t \ge 0$, we have*

$$\mathbb{P}[\exists x \in \mathbb{R}^{n-1} : \|Ax - v\|_2 \le t] \le C \cdot n^3 t^{1/2} + (\|\vec{p}\|_\infty + \eta)^n,$$

where A is a random $n \times (n-1)$ matrix with independent ξ entries.

Proof. Fix $\eta > 0$, and let $\mathcal{E}$ be the event whose probability we are trying to control. After potentially reindexing the coordinates, we may write

$$A = \begin{bmatrix} R \\ A_{n-1} \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v' \end{bmatrix}$$

where A_{n-1} is an $(n-1) \times (n-1)$ matrix and $v' \in \mathbb{R}^{n-1}$ satisfies $\|v'\|_2 \ge 1/2$. Let $\mathcal{E}_S = \{s_{n-1}(A_{n-1}) \le \sqrt{t}\}$. By Theorem 5.5, we have that for all sufficiently large n, there exists a constant C' depending on ξ and η such that

$$\mathbb{P}[\mathcal{E}_S] \le C'\sqrt{nt} + (\|\vec{p}\|_\infty + \eta/2)^n.$$

It therefore suffices to bound the probability of $\mathcal{E} \wedge \mathcal{E}_S^c$. In such a situation, we see that $y := (A_{n-1})^{-1}v'$ is unique. Let $y_0 := y/\|y\|_2$, and for δ, ρ, L to be chosen momentarily, let

$$\mathcal{E}_U = \{y_0 \in \mathrm{Cons}(\delta, \rho) \vee \mathcal{T}_\xi(y_0, L) \ge (\|\vec{p}\|_\infty + \eta/2)^n\}.$$

By Proposition 5.4, we can choose $\delta, \rho, L > 0$ depending on ξ and η so that

$$\mathbb{P}[\mathcal{E}_U] \le (\|\vec{p}\|_\infty + \eta/2)^n.$$

Hence, it suffices to bound the probability of $\mathcal{E} \wedge \mathcal{E}_S^c \wedge \mathcal{E}_U^c$. Let $x \in \mathbb{R}^{n-1}$ be a vector certifying this event. Then, we have for all sufficiently large n that

$$\|A_{n-1}x - v'\|_2 \le t \implies \|x - y\|_2 \le t^{1/2}, \text{ and}$$

$$|Rx - v_1| \le t \implies |Ry - v_1| \le t + nt^{1/2}.$$

Furthermore, since $\|v\|_2 \ge 1$, we have $\|y\|_2 \ge 1/C''n^2$, for some constant C'' depending on ξ.

We now fix a realization of A_{n-1} satisfying $\mathcal{E}_S^c \wedge \mathcal{E}_U^c$. In particular, this fixes y, y_0 satisfying the conditions in $\mathcal{E}_U^c$ and with $\|y\|_2 \ge 1/C''n^2$. Now, we use the independence of R and A_{n-1} and the fact that $\mathcal{E}$ implies

$$|Ry - v_1| \le t + nt^{1/2} \le 2nt^{1/2}.$$

Since

$$\mathcal{T}_\xi(y_0, L) < (\|\vec{p}\|_\infty + \eta/2)^n$$

and $\|y\|_2 \ge 1/C''n^2$, we find that the desired probability is bounded by

$$2LC''n^3t^{1/2} + L(\|\vec{p}\|_\infty + \eta/2)^n. \qquad \square$$

Now we are ready to conclude Proposition 6.8.

Proof of Proposition 6.8. A completely identical argument to the proof of Proposition 6.5 shows that for a sufficiently large constant K depending on ξ, and for $\Xi := \operatorname{supp}(\xi) \subseteq \mathbb{R}$,

$$\mathbb{P}[\mathcal{E}] \le p_0^{3n} + \sum_{a \in \mathcal{C}} \mathbb{P}[M^{(1)} = a]\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|M^{(-1)}u' + a\|_2 \le 2t],$$

where $M^{(1)}$ denotes the first column of M_n, $M^{(-1)}$ denotes the $n \times (n-1)$ matrix formed by excluding this column, and

$$\mathcal{C} = \{a \in \Xi^n : \exists \lambda \in \mathbb{R} \text{ with } \|a - \lambda 1_n\|_2 \le 8K\delta'\sqrt{n}\}.$$

We want to bound the contribution of the sum on the right hand side.

Case I: If $a = 0$, $\mathbb{P}[M^{(1)} = a] = p_0^n$.

Case II: If $a \neq 0$, then $\|a\|_2 \ge h_\xi > 0$. Hence, by Lemma 6.9, there is a constant $C > 0$ depending on ξ, η such that

$$\mathbb{P}[\exists u' \in \mathbb{R}^{n-1} : \|M^{(-1)}u' + a\|_2 \le 2t] \le Cn^3\sqrt{t} + (\|\vec{p}\|_\infty + \eta/2)^n.$$

Moreover, a similar (but easier) argument as in the proof of Proposition 6.5 shows that

$$\mathbb{P}[M^{(1)} \in \mathcal{C}] \le \|\vec{p}\|_\infty^n \exp(c_{\xi,\delta'}n),$$

where $c_{\xi,\delta'}$ goes to 0 as δ' goes to 0.

Hence, we see that the contribution to the sum from this case is bounded by

$$\|\vec{p}\|_\infty^n \exp(c_{\xi,\delta'}n) \cdot \Big(Cn^3\sqrt{t} + (\|\vec{p}\|_\infty + \eta/2)^n\Big).$$

By choosing δ' sufficiently small depending on ξ and η, and using $\|\vec{p}\|_\infty^2 \le \|\vec{p}\|_2^2 - c_\xi$ for some $c_\xi > 0$, we see as before (using the AM-GM inequality) that the above quantity is at most

$$t\exp(-\eta' n) + (\|\vec{p}\|_2^2 - \eta')^n$$

for a sufficiently small η' depending on ξ and η. This completes the proof. $\square$

The proof of Theorem 1.7 is now immediate.

Proof of Theorem 1.7. First, assume that ξ is not a shift of an origin-symmetric random variable. We choose δ' small enough so that the conclusions of Propositions 6.5 and 6.8 are satisfied. By the union bound, this shows that the contribution of $\mathrm{Elem}(\delta')$ to the probability is at most

$$n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}\mathbb{P}[\mathcal{E}_{e_1-e_2}] + (t + \|\vec{p}\|_2^{2n})e^{-\eta n},$$

for a sufficiently small $\eta > 0$ depending on ξ, and for all sufficiently large n depending on ξ. Now, we can conclude using Corollary 6.4.

Next, if ξ is a nonzero shift of an origin-symmetric random variable, we do the same, except we require Propositions 6.5, 6.7, and 6.8 and then conclude with Corollary 6.4.

Finally, we consider the case when ξ is an origin-symmetric random variable. As before, we begin by using Propositions 6.5 and 6.8. The only thing to note is that, by the symmetry of ξ about the origin, for all $i \neq j$, $\mathbb{P}[\mathcal{E}_{e_i-e_j}] = \mathbb{P}[\mathcal{E}_{e_i+e_j}]$. Hence, by the union bound, the contribution of $\mathrm{Elem}'(\delta')$ to the probability is at most

$$n\mathbb{P}[\mathcal{E}_{e_1}] + \binom{n}{2}(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + (t + \|\vec{p}\|_2^{2n})e^{-\eta n}.$$

Now, we can conclude using Corollary 6.4. ☐

7. Deduction of Theorems 1.2 and 1.6

Given the results in Sections 4 and 6, the deduction of Theorems 1.2 and 1.6 is immediate. Fix a discrete distribution ξ, and let $\delta, \rho, \eta, n_0 > 0$ be parameters depending on ξ coming from Theorem 1.7. Then, for the proof of Theorem 1.6, let $\epsilon > 0$ be as in the statement of the theorem (it suffices to assume that $\epsilon < 1$), and for the proof of Theorem 1.2, let $\epsilon > 0$ be such that

$$\exp(2\epsilon - H(\vec{p})) < \|\vec{p}\|_2^2,$$

which is possible since, by the weighted AM-GM inequality, we have

$$\exp(-H(\vec{p})) = \prod_i p_i^{p_i} \le \sum_i p_i^2 = \|\vec{p}\|_2^2,$$

and equality holds if and only if ξ is uniform on its support.

Let $C = C_{4.1}(\xi, \delta, \rho, \epsilon/2)$. By taking $C_{\xi,\epsilon}$ in Theorem 1.6 and C_ξ in Theorem 1.2 to be at least C, we may restrict our attention to $0 \le t \le 1/C$ (since for $t \ge 1/C$, the right-hand sides of Theorems 1.2 and 1.6 are at least 1). By Theorem 1.7 and Theorem 4.1, for all $0 \le t \le 1/C$, we have

$$\begin{aligned}
\mathbb{P}[s_n(M_n(\xi)) \le t/\sqrt{n}] &\le \mathbb{P}\left[\inf_{x\in\mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n}\right] + \mathbb{P}\left[\inf_{y\in\mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 \le Ct\right] \\
&\quad + \mathbb{P}\left[\inf_{x\in\mathbb{S}^{n-1}\setminus\mathrm{Cons}(\delta,\rho)} \|M_n(\xi)x\|_2 \le t/\sqrt{n} \wedge \inf_{y\in\mathrm{Cons}(\delta,\rho)} \|yM_n(\xi)\|_2 > Ct\right] \\
&\le 2n\mathbb{P}[\mathcal{E}_{e_1}] + (n^2 - n)(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + 2(Ct + \|\vec{p}\|_2^{2n})e^{-\eta n} \\
&\quad + Ct + \exp((\epsilon/2 - H(\vec{p}))n)
\end{aligned}$$

for all sufficiently large n. Here, we have used that $M_n(\xi)$ and $M_n(\xi)^\intercal$ have the same distribution.

For Theorem 1.2, we are done by our choice of ϵ.

For Theorem 1.6, we note that by Cauchy-Schwarz (as in Proposition 6.2), $\mathbb{P}[\mathcal{E}_{e_1+e_2}] \le \mathbb{P}[\mathcal{E}_{e_1-e_2}]$ and recall from above that $\exp(-H(\vec{p})) \le \|\vec{p}\|_2^2$. Using this, we can bound the right hand side of the above computation by

$$2Ct + 2n\mathbb{P}[\mathcal{E}_{e_1}] + 2n^2\exp(\epsilon n/2)\mathbb{P}[\mathcal{E}_{e_1-e_2}].$$

The desired conclusion follows since $2n^2\exp(\epsilon n/2) \le (1+\epsilon)^n$ for all $\epsilon < 1$ and n sufficiently large.

8. Singularity of random combinatorial matrices

In this section, we discuss the proof of Theorem 1.8. Given Theorem 4.1, the only ingredient we need is the following estimate for invertibility on almost constant vectors.

Proposition 8.1. *For any $\epsilon > 0$, there exist δ, ρ, c, n_0 depending on ϵ such that for all $n \geq n_0$,*

$$\mathbb{P}\left[\inf_{x\in \mathrm{Cons}(\delta,\rho)} \|Q_n x\|_2 \leq c\sqrt{n} \vee \inf_{y\in \mathrm{Cons}(\delta,\rho)} \|yQ_n\|_2 \leq c\sqrt{n}\right] \leq \left(\frac{1}{2}+\epsilon\right)^n.$$

We begin with the easier case of $\|yQ_n\|_2$.

Lemma 8.2. *For any $\epsilon > 0$, there exists c, n_0 depending on ϵ such that for all $n \geq n_0$ and for any $y \in \mathbb{S}^{n-1}$,*

$$\mathbb{P}[\|yQ_n\|_2 \leq c\sqrt{n}] \leq (1/2+\epsilon)^n.$$

Proof. Without loss of generality, we may assume that $|y_1| \geq \cdots \geq |y_n|$. We divide the proof into two cases depending on $|y_1|$. Let $\delta > 0$ be a constant to be chosen at the end of the proof.

Case I: $|y_1| < \delta$. Note that any entry in the first $n/4$ columns of Q_n, conditioned on all the remaining entries in the first $n/4$ columns of Q_n, is distributed as $\mathrm{Ber}(p)$ for some $p \in [1/3, 2/3]$. Moreover, by Lemma 2.4, it follows that for independent random variables $\xi_1, \dots, \xi_n$, where $\xi_i \sim \mathrm{Ber}(p_i)$ for some $p_i \in [1/3, 2/3]$,

$$\mathcal{L}(y_1\xi_1 + \cdots + y_n\xi_n, \delta) \leq 3C_{2.4}\delta.$$

Therefore, a slight conditional generalization of the second part of Lemma 3.14 (which has the same proof) shows that

$$\mathbb{P}[\|yQ_n\|_2 \leq \delta\sqrt{n/8}] \leq (20C_{2.4}\delta)^{n/8} \leq (1/4)^n,$$

provided that δ is chosen sufficiently small depending on $C_{2.4}$.

Case II: $|y_1| \geq \delta$. Let $R_1, \dots, R_n$ denote the rows of Q_n. Then,

$$\begin{aligned}\mathbb{P}[\|yQ_n\|_2 \leq \delta c\sqrt{n}] &\leq \sup_{R_2,\dots,R_n} \mathbb{P}[\|y_1R_1 + y_2R_2 + \cdots + y_nR_n\|_2 \leq \delta c\sqrt{n} | R_2, \dots, R_n] \\ &\leq \sup_{v\in\mathbb{R}^n} \mathbb{P}[\|R_1 - v\|_2 \leq c\sqrt{n}] \leq \binom{n}{n/2}^{-1}\binom{n}{2c^2n} \leq \left(\frac{1}{2}+\epsilon\right)^n,\end{aligned}$$

provided that $c > 0$ is chosen to be sufficiently small depending on $\epsilon > 0$. This completes the proof. □

Next, we deal with the harder case of $\|Q_nx\|_2$. We will need the following analogue of Lemma 5.1.

Lemma 8.3. *For any $\epsilon \in (0, 1/8)$, there exist $\theta = \theta(\epsilon) > 0$ and n_0 depending on ϵ for which the following holds. For all $n \geq n_0$ and for all $x \in \mathbb{S}^{n-1}$ such that $|\langle x, 1_n/\sqrt{n}\rangle| \leq 1/2$, we have*

$$\mathcal{L}(q \cdot x, \theta) \leq 1/2 + \epsilon,$$

where q is distributed uniformly on $\{0,1\}^n_{n/2}$.

Proof. Without loss of generality, we may assume that $|x_1| \geq \cdots \geq |x_n|$. Again, we divide the proof into two cases depending on $|x_1|$. Let $\delta > 0$ be a constant to be chosen at the end of the proof.

Case I: $|x_1| < \delta$. Let $\mu := \mathbb{E}[q \cdot x]$ and $\sigma^2 := \mathrm{Var}(q \cdot x)$. Since $\langle x, 1_n\rangle \leq \sqrt{n}/2$, a direct computation shows that $\sigma^2 \geq 3/16$. Moreover, a quantitative combinatorial central limit theorem due to Bolthausen [2] shows that the L^∞ distance between the cumulative distribution function of $(q \cdot x - \mu)/\sigma$ and that of the standard Gaussian is at most $C\delta$, where C is an absolute constant. Hence, for all δ sufficiently small, we have $\mathcal{L}(q \cdot x, \delta) \leq 1/4$ whenever $|x_1| < \delta$.

Case II: $|x_1| \geq \delta$. Let $\mathcal{G}$ denote the event (depending on q) that

$$(n-\epsilon^2 n-1)^{-1}\sum_{i=2}^{n-\epsilon^2 n} q_i \in [1/2-\epsilon^4/2, 1/2+\epsilon^4/2].$$

Then for all sufficiently large n, we have

$$\begin{aligned}\sup_{r\in\mathbb{R}}\mathbb{P}[|q\cdot x-r|\leq\theta] &\leq \sup_{r\in\mathbb{R}}\mathbb{P}[|q\cdot x-r|\leq\theta\wedge\mathcal{G}]+\mathbb{P}[\mathcal{G}^c]\\ &\leq \sup_{r\in\mathbb{R}}\mathbb{P}[|q\cdot x-r|\leq\theta\wedge\mathcal{G}]+2\exp(-\epsilon^8 n/128),\end{aligned}$$

where the final inequality is by a standard large deviation estimate. It remains to control $\mathbb{P}[|q\cdot x-r|\leq\theta\wedge\mathcal{G}]$. For this, fix any realization $q' := (q_2,\dots,q_{n-\epsilon^2 n})$ satisfying $\mathcal{G}$. Note that

$$1/2-2\epsilon^2 \leq \inf_{q'\in\mathcal{G}}\mathbb{P}[q_1=0 \mid q'] \leq \sup_{q'\in\mathcal{G}}\mathbb{P}[q_1=0\mid q'] \leq 1/2+2\epsilon^2.$$

Note also that, since $\sum_{i\geq n-\epsilon^2 n} x_i^2 \leq \epsilon^2$ (this uses $\|x\|_2=1$ and $|x_1|\geq|x_2|\geq\cdots\geq|x_n|$), it follows that

$$\sup_{q'\in\mathcal{G},q_1}\mathrm{Var}\left[\sum_{i\geq n-\epsilon^2 n} q_i x_i \middle| q_1, q'\right] \leq \epsilon^2,$$

so that by Markov's inequality,

$$\sup_{q'\in\mathcal{G},q_1}\mathbb{P}\left[\left|\sum_{i\geq n-\epsilon^2 n} q_i x_i - f(q',q_1)\right| \geq \frac{\delta}{8}\middle| q', q_1\right] \leq \frac{32\epsilon^2}{\delta^2},$$

where $f(q',q_1)$ denotes the mean of $\sum_{i\geq n-\epsilon^2 n} q_i x_i$ conditioned on q', q_1. Finally, since $|x_1|\geq\delta$, and since

$$\sup_{q'\in\mathcal{G}}|f(q',0)-f(q',1)| \leq |x_{n-\epsilon^2 n}| \leq 2/\sqrt{n},$$

it follows by putting everything together that

$$\sup_{r\in\mathbb{R}}\mathbb{P}[|q\cdot x-r|\leq\theta\wedge\mathcal{G}] \leq \sup_{r\in\mathbb{R}}\sup_{q'\in\mathcal{G}}\mathbb{P}[|q\cdot x-r|\leq\theta \mid q'] \leq 1/2+2\epsilon^2+64\epsilon^2/\delta^2,$$

provided that θ is chosen sufficiently small compared to δ, and n is sufficiently large. Indeed, the two values of q_1x_1 (for $q_1=1$ and $q_1=0$) differ by $|x_1|$, which is at least δ by assumption, and the above discussion shows that given q' and q_1, $\sum_{i\geq n-\epsilon^2 n} q_i x_i$ is localized in an interval of length $\delta/2+2/\sqrt{n}$ except with probability at most $32\epsilon^2/\delta^2$. Since δ is an absolute constant coming from **Case I**, this gives the desired conclusion for all sufficiently small ϵ, which completes the proof. □

Given the previous two lemmas, the proof of Proposition 8.1 is by now standard.

Proof of Proposition 8.1. The estimate for $\inf_{y\in\mathrm{Cons}(\delta,\rho)}\|yQ_n\|_2 \leq c\sqrt{n}$ (for a suitable choice of δ,ρ,c) follows immediately by combining Lemma 8.2 with the low metric entropy of $\mathrm{Cons}(\delta,\rho)$. To exploit the latter, one could either use a randomized rounding based net construction due to Livshyts [16, Theorem 4], which uses that $\|Q_n\|_{\mathrm{HS}}^2 \leq n^2$, or one could use the fact that there exists a constant K such that with probability at least $1-4^{-n}$, all singular values of Q_n except for the top singular value are at most $K\sqrt{n}$ (see [26, Proposition 2.8]).

For the estimate on $\inf_{x\in\mathrm{Cons}(\delta,\rho)}\|Q_n x\|_2 \leq c\sqrt{n}$ (for suitable δ,ρ,c), we begin by using the fact [26, Proposition 2.8] noted above that there exists a constant $K>0$ such that with probability at least $1-4^{-n}$, the operator norm of Q_n restricted to the subspace perpendicular to 1_n is at most $K\sqrt{n}$. Let us denote this event by $\mathcal{E}_K$. Then, on $\mathcal{E}_K$, for any $x\in\mathbb{S}^{n-1}$ such that $\langle x, 1_n/\sqrt{n}\rangle \geq 1/2$, we have $\|Q_n x\|_2 \geq n/4 - K\sqrt{n}$. Hence, on the event $\mathcal{E}_K$, and for all n sufficiently large, it suffices to consider the infimum over those vectors $x\in\mathrm{Cons}(\delta,\rho)$ which also satisfy $\langle x, 1_n/\sqrt{n}\rangle < 1/2$. For this, we can use Lemma 2.6 followed by the tensorization lemma Lemma 3.14, and then exploit the low metric entropy of $\mathrm{Cons}(\delta, \rho)$ as above. We leave the details to the interested reader. □

Finally, given Proposition 8.1, the proof of Theorem 1.8 follows exactly as in Theorem 1.6.

References

[1] Anirban Basak and Mark Rudelson, *Sharp transition of the invertibility of the adjacency matrices of sparse random graphs*, Probab. Theory Related Fields **180** (2021), 233–308.

[2] Erwin Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **66** (1984), 379–386.

[3] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), 559–603.

[4] Asaf Ferber, Vishesh Jain, Kyle Luh, and Wojciech Samotij, *On the counting problem in inverse Littlewood–Offord theory*, arXiv:1904.10425.

[5] Han Huang, *Rank of sparse Bernoulli matrices*, arXiv:2009.13726.

[6] Anwar Irmatov, *Asymptotics of the number of threshold functions and the singularity probability of random* $\{\pm 1\}$*-matrices*, Doklady Mathematics **101** (2020), 247–249.

[7] Vishesh Jain, *Approximate Spielman-Teng theorems for the least singular value of random combinatorial matrices*, arXiv:1904.10592.

[8] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *On the smoothed analysis of the smallest singular value with discrete noise*, arXiv:2009.01699.

[9] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Sharp invertibility of random Bernoulli matrices*, arXiv:2010.06553.

[10] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *The smallest singular value of dense random regular digraphs*, arXiv:2008.04755.

[11] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), 223–240.

[12] Boris S. Kašin, *Diameters of some finite-dimensional sets and classes of smooth functions*, Izvestiya: Mathematics **11** (1977), 317–333.

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

[14] Alexander E. Litvak, Alain Pajor, Mark Rudelson, and Nicole Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), 491–523.

[15] Alexander E. Litvak and Konstantin Tikhomirov, *Singularity of sparse Bernoulli matrices*, arXiv:2004.03131.

[16] Galyna V. Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, arXiv:1811.07038.

[17] Galyna V. Livshyts, Konstantin Tikhomirov, and Roman Vershynin, *The smallest singular value of inhomogeneous square random matrices*, arXiv:1909.04219.

[18] Hoi H. Nguyen, *On the singularity of random combinatorial matrices*, SIAM J. Discrete Math. **27** (2013), 447–458.

[19] Boris A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** (1961), 106–108.

[20] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), 575–600.

[21] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), 600–633.

[22] Gideon Schechtman, *Special orthogonal splittings of* L_1^{2k}, Israel J. Math. **139** (2004), 337–347.

[23] Terence Tao and Van H. Vu, *On random* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), 1–23.

[24] Terence Tao and Van H. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), 603–628.

[25] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), 593–634.

[26] Tuan Tran, *The smallest singular value of random combinatorial matrices*, arXiv:2007.06318.

[27] Roman Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018, An introduction with applications in data science, With a foreword by Sara van de Geer.

Department of Statistics, Stanford University, Stanford, CA 94305, USA
Email address: visheshj@stanford.edu

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Email address: {asah,msawhney}@mit.edu