

## SINGULARITY OF DISCRETE RANDOM MATRICES

VISHESH JAIN, ASHWIN SAH, AND MEHTAAB SAWHNEY

ABSTRACT. Let  $\xi$  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  $\xi$ .

For  $\xi$  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.

- 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  $\xi$  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  $\xi$ . We will restrict attention to when  $\xi$  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  $e_1, \dots, e_n$ . Then, the aforementioned conjecture may be stated as follows.

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

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

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  $\xi$  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 [\[13, Section 3.2\]](#)), we thus establish the following strengthening of [Conjecture 1.1](#) for discrete distributions which are not uniform on their support.

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

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

In particular, for fixed  $p \in (0, 1), p \neq 1/2$ , there exists  $c_p > 0$  for which

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

**Remark.** Let us discuss the implications of this corollary for  $\xi = \text{Ber}(p)$ .

- For fixed  $p \in (1/2, 1)$ , [Corollary 1.3](#) provides the leading term in the asymptotic expansion of the singularity probability of  $M_n(\text{Ber}(p))$ . Prior to this work, even the correct value of  $\lim_{n \rightarrow \infty} n^{-1} \log \mathbb{P}[M_n(\text{Ber}(p)) \text{ is singular}]$  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 [\[2\]](#) 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 [\[1, 4, 13\]](#), 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.
- However, even for fixed  $p \in (0, 1/2)$ , we obtain more precise results than those in the recent work of Litvak and Tikhomirov [\[13\]](#) as well as [\[7, 23\]](#). Namely, [Corollary 1.3](#) provides the first *two* terms in the asymptotic expansion of the singularity probability of  $M_n(\text{Ber}(p))$ ; a result of this precision was not available before in any context.

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

**Theorem 1.4.** *Let  $\xi$  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 [1.4](#) 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 [\[11\]](#) was the first to show that  $q_n(\text{Ber}(1/2)) = o_n(1)$ . Much later, an exponential bound on  $q_n(\text{Ber}(1/2))$  was obtained by Kahn, Komlós, and Szemerédi [\[9\]](#). 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 [\[21, 22\]](#), and later to  $1/\sqrt{2} + o_n(1)$  by Bourgain,

Vu, and Wood [2]. Finally, a truly breakthrough result of Tikhomirov [23] in 2018 established that  $q_n(\text{Ber}(p)) = (1 - p + o_n(1))^n$  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  $\text{Ber}(p)$ , 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  $\xi = \text{Ber}(p_n)$  for  $p_n$  in a certain range of sparsity limited to  $n^{-1} \ln n - \omega_n(n^{-1}) \leq p_n \leq n^{-1} \ln n + o_n(n^{-1} \ln \ln n)$ . Subsequently, Litvak and Tikhomirov showed that the conjecture also holds for  $\xi = \text{Ber}(p_n)$  for  $Cn^{-1} \ln n \leq p_n \leq c$ , where  $c > 0$  is a small absolute constant and  $C > 0$  is a large absolute constant. Recent work of Huang [4] was able to bridge the gap between the regimes covered in [1] and [4], leaving open the regime  $p \in (c, 1/2)$ . Establishing [Conjecture 1.1](#) (as opposed to the stronger [Theorem 1.2](#)) in this regime does not require the full strength of the ideas in this paper – since this is a case of substantial interest (see Problem 8.2 in the recent work of Litvak and Tikhomirov [13]), we have provided the much simpler details in the self-contained companion note [7], which also serves as a gentle introduction to the present work and contains complete details of some relatively standard arguments which are only sketched in the present work.

For general discrete distributions  $\xi$ , the only previous systematic study in the literature is the work of Bourgain, Vu, and Wood [2]. They show [2, Corollary 1.2] that if  $\xi$  is a discrete distribution with  $\sup_{r \in \mathbb{R}} \mathbb{P}[\xi = r] =: p$ , then  $q_n(\xi) \leq (\sqrt{p} + o_n(1))^n$ , 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 1.4](#) 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 [2, Corollaries 3.1, 3.2] that if  $\xi_{1,\mu}$  is a random variable taking on the value 0 with probability  $1 - \mu$  and  $\pm 1$  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 1.4](#) 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 [2] such as [2, 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 [5] 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 [5, Theorem 3]. Furthermore, upon slight modification, the proof in [5] 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  $\xi$  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 [1.4](#).

**Theorem 1.5.** *Fix a discrete distribution  $\xi$ . There exist  $\delta, \rho, \eta > 0$  depending on  $\xi$  such that for all sufficiently large  $n$  and  $t \leq 1$ ,*

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

The set  $\text{Cons}(\delta, \rho)$  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 3.1](#)), although a trivial modification shows this result holds for any sufficiently low-entropy subset of the unit sphere.

**Remark.** In the companion note [\[7\]](#), as a natural consequence of our study of anti-concentration on slices, we are able to resolve a conjecture of Nguyen [\[16, Conjecture 1.4\]](#) on the probability of singularity for a certain combinatorial model of random matrices. Namely, let  $Q_n$  be an  $n \times n$  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 [\[16\]](#), Nguyen showed that  $\mathbb{P}[Q_n \text{ is singular}] = O_C(n^{-C})$  for any  $C > 0$ , and conjectured [\[16, Conjecture 1.4\]](#) that  $\mathbb{P}[Q_n \text{ is singular}] = (1/2 + o_n(1))^n$ . After intermediate work [\[3, 6\]](#), an exponential upper bound on the singularity probability was only very recently obtained in work of Tran [\[24\]](#), the key difficulty being establishing even relatively coarse inverse Littlewood–Offord type results for the slice. In [\[7\]](#), we settle [\[16, Conjecture 1.4\]](#) using the techniques developed in this work.

**Theorem 1.6 ([\[7, Theorem 1.4\]](#)).** *For every  $\epsilon > 0$ , there exists  $C_\epsilon$  depending on  $\epsilon$  such that for all sufficiently large  $n$ , and for all  $t \geq 0$ ,*

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

**1.3. Overview of the techniques.** As in many works in this area, we use the high-level strategy (going back to Kašin [\[10\]](#) and subsequently used in [\[12, 18–20\]](#)) 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 [Corollary 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 [\[13\]](#), 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 1.5](#), requires significant innovations compared to previous works on the sharp invertibility of sparse random Bernoulli matrices [\[1, 4, 13\]](#) – 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  $\text{Ber}(p)$  is conveniently at 0. We refer the reader to [7] for the much simpler argument in this case.

In the present work, in order to handle non-elementary vectors, we need to develop novel sharp anticoncentration estimates [Propositions 5.2 and 5.3](#) (in contrast, the essentially standard estimate [7, Lemma 2.6] was sufficient for the corresponding step in [7]). Even more involved is the analysis of elementary vectors, for which we develop a new technique. Let us begin by discussing this technique for  $\xi = \text{Ber}(p)$  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  $\xi$ , 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 5.5 and 5.7](#)), 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 1.4](#) with an estimate on the singularity probability of the form  $(\sup_{r \in \mathbb{R}} \mathbb{P}[\xi = r] + o_n(1))^n$  ([Proposition 4.4](#) and [Theorem 4.5](#)), and then leverage these preliminary estimates to obtain the desired bound.

We emphasize that our treatment of structured vectors, as captured by [Theorem 1.5](#), is not sensitive to the non-uniformity of the distribution  $\xi$ . In particular, given [Theorems 1.2 and 1.5](#), 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  $\xi$  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  $\xi$  by exploiting the gap between  $\mathbb{P}[\xi = \xi']$  and the entropy of  $\xi$ ; 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 [13].

The main ingredient in our work for handling such vectors is [Theorem 2.1](#), which is an extension of [23, 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 [13] 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 [23, 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. [8] and the references therein).

Compared to [23, 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  $\xi$ ), and hence, certain arguments tailored for integers no longer apply. Overcoming these challenges requires additional ideas, which we discuss in [Section 2](#). The reader is also referred to [7, Theorem 3.1] where we record the proof for the Boolean slice, a setting which encounters the first challenge but not the second.

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.,  $\ell^2 \rightarrow \ell^2$  operator norm).

We will let  $[N]$  denote the interval  $\{1, \dots, N\}$ . For nonnegative integers  $m \leq n$ , we let  $\{0, 1\}_m^n$  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 1.7.** We say that a random variable  $\xi$  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  $\xi$ .

For  $\xi$  a discrete random variable with  $k = |\text{supp}(\xi)|$  (so that  $k \geq 2$ ), we will denote its support by  $\vec{a} = (a_1, \dots, a_k)$ , and the (nonzero) probabilities of attaining  $a_1, \dots, a_k$  by  $\vec{p} = (p_1, \dots, p_k)$ . Note that  $\|\vec{p}\|_1 = 1$ , and  $\|\vec{p}\|_2^2 \leq \|\vec{p}\|_\infty$  with equality if and only if  $\xi$  is uniform on its support. We will use  $H(\xi)$  to denote the natural-logarithmic entropy of  $\xi$ , 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  $\xi$  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  $f: \mathbb{Z} \rightarrow \mathbb{R}$  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  $C_\alpha$  is some constant depending on  $\alpha$ ;  $f = \Omega_\alpha(g)$  (or  $f \gtrsim_\alpha g$ ) means that  $f \geq c_\alpha g$ , where  $c_\alpha > 0$  is some constant depending on  $\alpha$ , and  $f = \Theta_\alpha(g)$  means that both  $f = O_\alpha(g)$  and  $f = \Omega_\alpha(g)$  hold. For parameters  $\epsilon, \delta$ , 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 use this to prove the invertibility estimate for unstructured vectors ([Theorem 3.8](#)). In the short [Section 4](#), we prove a weaker version of [Theorem 1.4](#); this is used in our treatment of structured vectors (i.e., the proof of [Theorem 1.5](#)), which is the content of [Section 5](#). In [Section 6](#), we quickly combine [Theorems 1.5](#) and [3.8](#) to prove [Theorems 1.2](#) and [1.4](#).

1.6. **Acknowledgements.** We thank Mark Rudelson, Konstantin Tikhomirov, and Yufei Zhao for comments on the manuscript. 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 4.5](#)).

The proof of [Theorem 2.1](#) follows a direction introduced by Tikhomirov [[23](#)]. 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 multi-slice, as the arguments in [23] rely strongly on the independence structure of the considered model, and (ii) the freedom to allow the support of  $\xi$  to consist of arbitrary real numbers, as certain arguments in [23] rely on the integrality of the support.

The reader may benefit from first perusing a simpler version of this argument, for slices of the Boolean hypercube, which is presented in the companion note [7, Section 3]. We will often note the corresponding statements in [7] for the reader's convenience.

**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  $\xi$ , which has (nonzero) probabilities  $\vec{p} = (p_1, \dots, p_k)$  on atoms  $\vec{a} = (a_1, \dots, a_k)$ , 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  $(X_1, \dots, X_n)$  be the random vector uniformly distributed on  $\mathcal{A}$ . For any  $f : \mathbb{R} \rightarrow \mathbb{R}$ , 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  $X_1, \dots, X_n$ ):

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

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

**Theorem 2.1.** *Fix a discrete distribution  $\xi$ . 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  $\|\vec{m} - \vec{p}n\|_\infty \leq \gamma_{2.1}n$ ,*

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

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

**Definition 2.2.** Fix a discrete distribution  $\xi$ . 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

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

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

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

**Corollary 2.3.** *Fix a discrete distribution  $\xi$ . 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*

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

*Proof sketch.* This is essentially the same as the deduction in [23, Corollary 4.3]. We apply Theorem 2.1 to  $f(t) := 2^{-|t|/\sqrt{n}}/\iota$ , where  $t \in \mathbb{R}$  and  $\iota$  is an appropriate normalization, separately for all  $\vec{m} \in \mathbb{Z}_{\geq 0}^k$  such that  $\|\vec{m} - \vec{p}n\|_\infty \leq \gamma_{2.1}n$ , and then conclude using a union bound.  $\square$

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** ([17]). *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*

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

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  $\vec{s} \in \mathbb{Z}_{\geq 0}^k$  with  $\|\vec{s}\|_1 = n$  and with  $s_\ell \in [\lambda n, (1 - \lambda)n]$  for some  $\ell \in [k]$ , we have*

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

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  $a_\ell$  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

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

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

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



where  $c(\sigma, \lambda, k) > 0$  is a constant depending only on  $\sigma$ ,  $\lambda$ , 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  $b'_1, \dots, b'_{\sigma n}$ .  $\square$

**Lemma 2.6.** *Fix a discrete distribution  $\xi$ ,  $\lambda \in (0, 1/3)$ ,  $\delta_0 \in (0, 1/4)$ . 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  $n > n_{2.6}(\lambda, \delta_0, K_1, K_2, K_3)$ ,  $\ell \geq \delta_0 n$ , and  $\vec{s} \in \mathbb{Z}_{\geq 0}^k$  with  $\|\vec{s}\|_1 = \ell$  and  $s_{j_0} \in [\lambda\ell, (1-\lambda)\ell]$  for some  $j_0 \in [k]$ . Then, for any interval  $J$ ,*

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

*Proof.* The proof is nearly identical to that in [\[23, Lemma 4.4\]](#) though we provide details as we are in the slightly different setting of  $L^1(\mathbb{R})$ . Fix  $X_1, \dots, X_\ell$ . Then

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

where  $(b_1, \dots, b_\ell)$  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  $\mathcal{A}$ .  $\square$

**2.2. Preprocessing on real-valued multislices.** As in [\[23\]](#), 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  $\xi$ . 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 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  $n_{2.7} = n_{2.7}(\xi, \epsilon, \delta, K_1, K_2, K_3, M) \geq 1$  such that the following holds.*

*Let  $n \geq n_{2.7}$ ,  $1 \leq N \leq \exp((H(\vec{p}) - \epsilon)n)$ , 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  $\vec{s} \in \mathbb{Z}_{\geq 0}^k$  with  $\|\vec{s}\|_1 = \ell$  and  $\|\vec{s} - \vec{p}\ell\|_\infty \leq \gamma_{2.7}\ell$ , we have*

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

For a simpler version of this argument, we refer the reader to the proof of [\[7, Proposition 3.6\]](#). [Proposition 2.7](#) should be seen as an analogue of [\[23, Lemma 4.6\]](#) for the multislice. As mentioned earlier, compared to [\[23\]](#), 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, as in [7, Proposition 3.6]. 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 [7, 23] 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 ([23]).

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  $1 \leq \ell \leq n$  and  $\vec{s} \in \mathbb{Z}_{\geq 0}^k$  with  $\|\vec{s}\|_1 = \ell$ . If any coordinate of  $\vec{s}$  is zero, note 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  $X = (X_1, \dots, X_n)$ . 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  $h_\ell$ . 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  $W_i = (W_i(1), \dots, W_i(k))$  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  $1 \leq i \leq \ell$ , where we assume  $W_0 = \vec{0}$ .
- $\|W_i\|_1 = i$ .
- $t_{i-1} = t_i + a_{w_i} X_i$  for all  $i \in [\ell]$ .
- $f_{\mathcal{A}, W_i, i}(t_i) = \sum_{j=1}^k \overline{W}_i(j) f_{\mathcal{A}, W_i - e_j, i-1}(t_i + a_j X_i)$ .
- $h_i = f_{\mathcal{A}, W_i, i}(t_i)$ .
- $f(t_0) = h_0 \geq h_1 \geq \dots \geq h_\ell = f_{\mathcal{A}, \vec{s}, \ell}(t)$ .

**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  $\lambda$ -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

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

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

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  $\vec{s} \in \mathbb{Z}_{\geq 0}^k$  satisfy  $\|\vec{s}\|_1 = \ell$  and  $\|\vec{s} - \vec{p}\ell\|_\infty \leq \gamma_{2.10}^\ell$ . Then, for  $(X_1, \dots, X_n) \in \mathcal{A}$ ,*

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

*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  $(X_1, \dots, X_n) \in \mathcal{A}$  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

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

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

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

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

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

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

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

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

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

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  $(w_1, \dots, w_\ell)$ .

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

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

It remains to bound

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

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

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

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

Hence,

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

We derive

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

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

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

We will first choose  $\lambda_{2.10}$ , and then choose some  $\gamma_{2.10} < \lambda_{2.10}^2$ . Note that, enforcing the constraint  $\gamma_{2.10} < \lambda_{2.10}^2$ , we can choose  $\lambda_{2.10}$  sufficiently small depending on  $\epsilon$  and  $\xi$  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  $\epsilon$  and  $\xi$  such that

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

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

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  $z \in \text{supp}(\xi - \xi') \setminus \{0\}$ , by Lemma 2.6, we have

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

Here, we have used that  $i-1 \geq \delta_0 n$ , 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  $(t_i)_{i=0}^\ell$ ,

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

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

From here on, the proof closely follows the proofs of [23, 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  $(X_1, \dots, X_n) \in \mathcal{E}_L$ , 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  $\tilde{t}_0$ . We also shift  $(t_i)_{i=1}^\ell$  by the same amount to obtain points  $(\tilde{t}_i)_{i=1}^\ell$  (note that these points are not necessarily integers). We call the sequence  $(\tilde{t}_i)_{i=0}^\ell$ , 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  $(\tilde{t}_i)_{i=0}^\ell$  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  $(X_1, \dots, X_n)$  and the step record, the witnessing sequence is completely determined by its starting point  $\tilde{t}_0$ . Furthermore, the definition of the witnessing sequence and the definition of  $f_{\mathcal{A}, \vec{s}, \ell}$  easily show that

$$\tilde{t}_0 \in \{\tau \in \mathbb{Z} : f(\tau) > (2N\sqrt{n})^{-1}\} =: \mathcal{D}.$$

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  $(X_1, \dots, X_n) \in \mathcal{E}_L$ , then there exists a witnessing sequence  $(\tilde{t}_i)_{i=0}^\ell$  with step record  $(w_i)_{i=1}^\ell$  such that  $\tilde{t}_0 \in \mathcal{D}$ , and such that

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

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

$$\mathbb{P}[\mathcal{E}_L] \leq (2k^2)^n \sup_{\substack{I \subseteq [\ell], |I| = \lceil \gamma_{2.10}n \rceil \\ \tilde{t}_0 \in \mathcal{D}, (w_i)_{i=1}^\ell \in [k]^\ell}} \mathbb{P}[\text{The witnessing sequence starts at } \tilde{t}_0, \text{ has step record } (w_i)_{i=1}^\ell, \text{ and}$$

$$\text{every } i \in I \text{ is } \lambda_{2.10}\text{-robust and is not an } R'\text{-drop}],$$

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.  $\square$

**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 [23, 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 [23, 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 [23]. 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  $\ell^1(\mathbb{Z})$ .

A simpler version of this argument, for Boolean slices, is presented in [7, Proposition 3.10].

**Proposition 2.11.** *Fix a discrete distribution  $\xi$ . There exists  $h = h(\xi) \geq 1$  so that the following holds. For any  $\epsilon \in (0, 1)$ ,  $\tilde{R} \geq 1$ ,  $L_0 \geq h\tilde{R}$ , and  $M \geq 1$ , there is  $\gamma_{2.11} = \gamma_{2.11}(\xi)$  and there are  $n_{2.11} = n_{2.11}(\xi, \epsilon, L_0, \tilde{R}, M) > 0$  and  $\eta_{2.11} = \eta_{2.11}(\xi, \epsilon, L_0, \tilde{R}, M) \in (0, 1)$  with the following property. Let  $L_0 \geq L \geq h\tilde{R}$ , 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)  $\int_{t \in I} g(t) \leq \tilde{R}/\sqrt{n}$  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  $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}$  are mutually independent. Define a random function  $\tilde{g} \in L^1(\mathbb{R})$  by

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

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

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

Then

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

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  $\epsilon \in (0, 1)$ ,  $\tilde{R} \geq 1$ ,  $L_0 \geq 64h^2\tilde{R}$ , and  $M \geq 1$ , there are  $n_{2.12} = n_{2.12}(h, \epsilon, L_0, \tilde{R}, M) > 0$  and  $\eta_{2.12} = \eta_{2.12}(h, \epsilon, L_0, \tilde{R}, M) \in (0, 1)$  with the following property. Let  $L_0 \geq L \geq 64h^2\tilde{R}$ , 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)  $\int_{t \in I} g(t) \leq \tilde{R}/\sqrt{n}$  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  $Y_1, \dots, Y_{\lfloor \epsilon n \rfloor}$  are mutually independent. Define a random function  $\tilde{g} \in L^1(\mathbb{R})$  by

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

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

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

This follows almost immediately from an  $\ell^\infty(\mathbb{Z})$  decrement result established by Tikhomirov [23].

**Proposition 2.13** ([23, Proposition 4.10]). *For any  $p \in (0, 1/2]$ ,  $\epsilon \in (0, 1)$ ,  $\tilde{R} \geq 1$ ,  $L_0 \geq 16\tilde{R}$ , and  $M \geq 1$  there are  $n_{2.13} = n_{2.13}(p, \epsilon, L_0, \tilde{R}, M) > 0$  and  $\eta_{2.13} = \eta_{2.13}(p, \epsilon, L_0, \tilde{R}, M) \in (0, 1)$  with the following property. Let  $L_0 \geq L \geq 16\tilde{R}$ , 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)  $\sum_{t \in I} g(t) \leq \tilde{R}/\sqrt{n}$  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  $Y_1, \dots, Y_{\lfloor \epsilon n \rfloor}$  are mutually independent. Define a random function  $\tilde{g} \in \ell^1(\mathbb{Z})$  by

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

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

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

**Remark.** In [23, 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  $\epsilon$  by  $\epsilon/k$  (and adjust other parameters appropriately) in order to deduce what we need directly from the statement as written in [23]. We will only need this statement for  $p = 1/2$ .

Now we first prove [Proposition 2.12](#).

*Proof of Proposition 2.12.* Consider the operator  $\mathcal{O} : L^1(\mathbb{R}) \rightarrow \ell^1(\mathbb{Z})$  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  $\omega$  is nonnegative and  $\log_2 \omega$  is  $\eta$ -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  $\tilde{R}$  increased to  $4h\tilde{R}$ . These last changes are responsible for the condition  $L_0 \geq 64h^2\tilde{R}$ .

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

Now, by Fubini's theorem, note that

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

where  $\tilde{g}$  is averaged in the sense of [Proposition 2.12](#). Therefore,

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



so that

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

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

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

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  $\tilde{g}(t)$  as

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

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

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

Then, provided that  $\gamma_{2.11}$  is chosen sufficiently small depending on  $\xi$ , and  $n$  is sufficiently large depending on  $\xi$  and  $\epsilon$ , 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  $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}$ ) that  $\|\tilde{g}\|_\infty > 19L/(20N\sqrt{n})$ . Now, suppose  $Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor} \in \mathcal{E}_L$ , and suppose further that  $\|\tilde{g}\|_\infty$  is attained at  $t \in \mathbb{R}$ . Let

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

Since  $\|g\|_\infty \leq L/(N\sqrt{n})$ , 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  $(Y_1, \dots, Y_{2\lfloor \epsilon n \rfloor}) \in \mathcal{E}_L$ , there exists some  $b \in B_0 \cap B_1$ . Hence, taking a union bound, we see that

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

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  $I = \{i_1, \dots, i_m\} \subseteq [en]$  such that  $|I| = m \geq \min(\bar{p})^2 \epsilon n / 8$  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  $Y_j^b := Y_{2i_j} - Y_{2i_j-1}$ . Let  $Y_{-2.I}$  denote all components of  $Y_1, \dots, Y_{2\lfloor en \rfloor}$ , except those corresponding to indices in  $2 \cdot I$ , and let  $Y_{2.I}$  denote the remaining components. Then, for  $b \in B_0$  and a choice of  $Y_{-2.I}$ , we define the random function (depending on  $Y_{2.I}$ ),

$$\tilde{g}_{b, Y_{-2.I}}(t) := \mathbb{E}_{b'} g \left( t + (a_1 - a_2) \sum_{j=1}^{\lfloor \min(\bar{p})^2 \epsilon n / 8 \rfloor} b'_j Y_j^b \right).$$

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

$$\mathbb{P} \left[ \|\tilde{g}_{b, Y_{-2.I}}\|_\infty \geq \frac{9L/10}{N\sqrt{n}} \right],$$

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

At this point, we can apply Proposition 2.12 to  $\tilde{g}_{b, Y_{-2.I}}$ . 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,  $b'_1, \dots, b'_{\lfloor \min(\bar{p})^2 \epsilon n / 8 \rfloor}$  are independent  $\text{Ber}(1/2)$  random variables. Finally, notice that, given  $Y_{-2.I}$ , each  $Y_j^b$  is a random variable uniform on some disjoint intervals of cardinality at least  $N$  each (since  $Y_j^b$  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  $\xi$ ).

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

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

provided that we choose  $\eta_{2.11}$  sufficiently small compared to  $\eta_{2.12}(d, \min(\bar{p})^2 \epsilon / 8, L_0, \tilde{R}, M)$ , 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  $\xi$ ).  $\square$

**2.4. Deriving the final result.** Theorem 2.1 is almost immediate given the previous propositions.

*Proof of Theorem 2.1.* The proof of Theorem 2.1 given Propositions 2.7 and 2.11 is similar to the derivation in [23, Theorem 4.2]; we refer the reader to the companion note [7, Theorem 3.1] for a leisurely account of the complete details. In words, we use Proposition 2.7 to show that any revelation of the first  $(1 - \gamma)n$  random variables (for suitably chosen  $\gamma$ ) will give an  $L^\infty$  bound at roughly the right scale  $L'(N\sqrt{n})^{-1}$ , where  $L'$  depends on  $M$ . Then, we divide the remaining random variables into a constant number of pieces (this number depends on  $L'$  and  $C_{2.5}$  with appropriate parameters), and use Proposition 2.11 iteratively on these pieces. As in Steps 2 and 3 in the proof of [7, Theorem 3.1], we may restrict our attention to the case when each of the remaining pieces is a well-conditioned multislice (this automatically implies that the first  $(1 - \gamma)n$  coordinates are also a well-conditioned multislice), noting that the complement of this event occurs with very low probability, and the product of this probability with  $L'(N\sqrt{n})^{-1}$  is much smaller than the bound on the essential sup-norm that we are aiming for.  $\square$

**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  $\xi$ . 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  $\xi$ . This independence simplifies matters dramatically, as one can derive an analogue of [Proposition 2.7](#) by simply considering drops (as in [\[23, Proposition 4.5\]](#)) 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. NON-ALMOST-CONSTANT VECTORS

**3.1. Preliminaries.** We recall the following basic definition of almost-constant vectors.

**Definition 3.1.** We define  $\text{Cons}(\delta, \rho)$  to be the set of  $x \in \mathbb{S}^{n-1}$  such that there is  $\lambda \in \mathbb{R}$  so that  $|x_i - \lambda| \leq \rho/\sqrt{n}$  for at least  $(1 - \delta)n$  values  $i \in [n]$ . We let  $\text{Noncons}(\delta, \rho) = \mathbb{S}^{n-1} \setminus \text{Cons}(\delta, \rho)$ .

The following is a standard fact regarding non-almost-constant vectors. For the reader’s convenience, we provide complete details of the proof in [\[7\]](#).

**Lemma 3.2** ([\[7, Lemma 2.3\]](#)). *For  $\delta, \rho \in (0, 1/4)$ , there exist  $\nu, \nu' > 0$  depending only on  $\delta, \rho$ , and a finite set  $\mathcal{K}$  of positive real numbers, also depending only on  $\delta, \rho$ , such that the following holds.*

*If  $x \in \text{Noncons}(\delta, \rho)$ , then at least one of the following two conclusions is satisfied.*

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

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

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

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

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

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

We record two useful tensorization statements.

**Lemma 3.3** ([\[23, Lemma 3.2\]](#)). *Let  $\chi_1, \dots, \chi_m$  be independent random variables.*

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

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

*Then for  $\epsilon \geq \epsilon_0$*

$$\mathbb{P}[\|(\chi_1, \dots, \chi_m)\|_2 \leq \epsilon\sqrt{m}] \leq (CK\epsilon)^m,$$

*where  $C$  is an absolute constant.*

- Assume that for some  $\eta, \tau > 0$ ,

$$\mathbb{P}[|\chi_i| \leq \eta] \leq \tau.$$

Then for  $\epsilon \in (0, 1]$ ,

$$\mathbb{P}[\|(\chi_1, \dots, \chi_m)\|_2 \leq \eta\sqrt{\epsilon m}] \leq (e/\epsilon)^{\epsilon m} \tau^{m-\epsilon m}.$$

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

**Lemma 3.4** ([25, 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  $t \geq 0$  we have*

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

Finally, we will need the following lemma, proved using randomized rounding (cf. [14]), which is a straightforward generalization of [23, Lemma 5.3]. We omit details since the proof is identical.

**Lemma 3.5.** *Let  $y = (y_1, \dots, y_n) \in \mathbb{R}^n$  be a vector, and let  $\mu > 0$ ,  $\lambda \in \mathbb{R}$  be fixed. Let  $\Delta$  denote a probability distribution which is supported in  $[-s, s]^n$ . There exist constants  $c_{3.5}$  and  $C_{3.5}$ , 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  $(b_1, \dots, b_n)$  is distributed according to  $\Delta$ . Then, there exists a vector  $y' = (y'_1, \dots, y'_n) \in \mathbb{Z}^n$  satisfying

- (R1)  $\|y - y'\|_\infty \leq 1$ ,
- (R2)  $\mathbb{P}[\left|\sum_{i=1}^n b_i y'_i - \lambda\right| \leq t] \leq C_{3.5} \mu t$  for all  $t \geq \sqrt{n}$ ,
- (R3)  $\mathcal{L}(\sum_{i=1}^n b_i y'_i, \sqrt{n}) \geq c_{3.5} \mathcal{L}(\sum_{i=1}^n b_i y_i, \sqrt{n})$ ,
- (R4)  $|\sum_{i=1}^n y_i - \sum_{i=1}^n y'_i| \leq C_{3.5} \sqrt{n}$ .

**3.2. The structure theorem.** Let  $\xi$  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  $\xi$  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. A key ingredient in the proof of the main result of this section, [Theorem 3.8](#), 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 3.6.** Fix a discrete distribution  $\xi$ . Let  $\vec{\gamma} \in \mathbb{R}_{\geq 0}^k$  with  $\|\vec{\gamma}\|_\infty < \min(\vec{p})$ , and let  $L \geq 1$ . Then, for any integer  $n \geq 1$  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 3.7.** *Let  $\delta, \rho, \epsilon \in (0, 1)$ . There exist  $L_{3.7} = L_{3.7}(\delta, \rho, \xi, \epsilon)$ ,  $\gamma_{3.7} = \gamma_{3.7}(\delta, \rho, \xi, \epsilon)$  and  $n_{3.7} = n_{3.7}(\delta, \rho, \xi, \epsilon)$  such that for all  $n \geq n_{3.7}$ , with probability at least  $1 - k^{-2n}$ , exactly one of the following holds.*

- $v(A) \in \text{Cons}(\delta, \rho)$ , or
- $\mathcal{T}_{\xi, \gamma_{3.7} \mathbf{1}_k}(v(A), L_{3.7}) \leq \exp((\epsilon - H(\vec{p}))n)$ .

The proof of [Proposition 3.7](#) follows from [Corollary 2.3](#) and [Lemma 3.5](#). At a high level, we consider dyadic intervals for the threshold function of  $v(A)$ , round  $v(A)$  to an appropriate integer lattice (using [Lemma 3.5](#)), and then use [Corollary 2.3](#). In order to be in a setting where we can apply [Corollary 2.3](#), we use an idea of Litvak and Tikhomirov [[13](#)] – using the Chernoff bound, we restrict ourselves to the event that all but  $O_{\xi, \gamma 2.3}(1)$  rows belong to a well-conditioned multislice corresponding to  $\xi$ . We refer the reader to [[7](#), [Proposition 4.2](#)] for further details.

**3.3. Invertibility on the bulk.** We are now ready to state and prove the key result of this section.

**Theorem 3.8.** *Fix a discrete distribution  $\xi$ . For any  $\delta, \rho, \epsilon > 0$ , there exists  $C_{3.8} = C_{3.8}(\xi, \delta, \rho, \epsilon) > 0$  and  $n_{3.8}(\xi, \delta, \rho, \epsilon) \geq 1$  such that for all  $n \geq n_{3.8}$  and  $t \geq 0$ ,*

$$\mathbb{P} \left[ \inf_{x \in \text{Noncons}(\delta, \rho)} \|M_n(\xi)x\|_2 \leq t/\sqrt{n} \wedge \inf_{y \in \text{Cons}(\delta, \rho)} \|yM_n(\xi)\|_2 > C_{3.8}t \right] \leq C_{3.8}t + \exp((\epsilon - H(\vec{p}))n).$$

*Proof.* Let  $M := M_n(\xi)$  for simplicity, and let  $\delta, \rho, \epsilon > 0$  be as in the statement of the theorem. 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  $\gamma = \gamma_{3.7}(\delta, \rho, \xi, \epsilon)$ . Let  $W_\gamma \subseteq \text{supp}(\xi)^n$  denote the set of vectors  $x \in \text{supp}(\xi)^n$  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

$$\mathbb{P} \left[ \inf_{x \in \text{Noncons}(\delta, \rho)} \|Mx\|_2 \leq t/\sqrt{n} \wedge \inf_{y \in \text{Cons}(\delta, \rho)} \|yM\|_2 > Ct \wedge \mathcal{W}_Q \right]. \quad (3.1)$$

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  $\text{Noncons}(\delta, \rho)$  certifying the event  $\mathcal{E}_R$ , so that

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

Using [Lemma 3.2](#), 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  $i \in [n]$ ,  $\|x_1 M^{(1)} + \dots + x_n M^{(n)}\|_2 \geq |x_i| \text{dist}(M^{(i)}, M^{(-i)})$ , it follows that

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

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

$$(3.1) = \mathbb{P}[\mathcal{E}_R \wedge \mathcal{E}_L \wedge \mathcal{W}_Q] \leq \frac{2}{\nu n} \sum_{i=1}^n \mathbb{P}[\text{dist}(M^{(i)}, M^{(-i)}) \leq t/\kappa \wedge \mathcal{E}_L \wedge M^{(i)} \in W_\gamma].$$

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

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

Let  $v(M^{(-1)})$  be a unit vector normal to  $M^{(-1)}$ . Then, by [Proposition 3.7](#), except with probability  $k^{-2n}$  (over the randomness of  $M^{(-1)}$ ), exactly one of the following holds.

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

where  $L := L_{3.7}(\delta, \rho, \xi, \epsilon)$ . If the first possibility occurs, then  $\mathcal{M}_1$  cannot hold as  $v(M^{(-1)}) \in \text{Cons}(\delta, \rho)$  satisfies

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

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

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

#### 4. PRELIMINARY INVERTIBILITY ESTIMATES

In this short section, we will prove a version of [Theorem 1.4](#) with the weaker singularity estimate  $(\|\vec{p}\|_\infty + o_n(1))^n$ . This estimate, which generalizes [[23](#), 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  $\xi$  random variables.

**Lemma 4.1.** *Fix a discrete distribution  $\xi$ . There is  $\theta = \theta(\xi) > 0$  such that for all  $x \in \mathbb{S}^{n-1}$ ,*

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

*Proof.* This is essentially identical to the proof given in [[23](#), Lemma 3.5]. Briefly, if  $\|x\|_\infty \geq \delta$ , then we can choose  $\theta$  small enough (depending on  $\delta$  and  $\xi$ ) 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  $\delta$  is sufficiently small depending on  $\xi$ .  $\square$

Combining the above estimate with the second part of [Lemma 3.3](#), we have the following.

**Corollary 4.2.** *Fix a discrete distribution  $\xi$ . For every  $\epsilon > 0$ , there exists  $c > 0$  depending on  $\epsilon$  and  $\xi$  such that for any  $x \in \mathbb{S}^{n-1}$  and  $y \in \mathbb{R}^n$ , we have*

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

Moreover, combining this corollary with the low metric entropy of  $\text{Cons}(\delta, \rho)$  and [Lemma 3.4](#), we obtain the following (weak) estimate for invertibility on almost-constant vectors.

**Corollary 4.3.** *Fix a discrete distribution  $\xi$ . For every  $\epsilon > 0$ , there exist  $\delta, \rho, c > 0$  depending on  $\epsilon$  and  $\xi$  such that for any  $y \in \mathbb{R}^n$ ,*

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

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

**Proposition 4.4.** *Fix a discrete distribution  $\xi$ . For every  $\epsilon, \eta > 0$ , there exist  $\delta, \rho, L > 0$  depending on  $\epsilon, \eta, \xi$  such that for any  $y \in \mathbb{R}^n$ ,*

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

*Proof.* This follows essentially from combining [Corollary 4.3](#) with a cruder analogue of [Proposition 3.7](#), 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 3.4](#)) a sufficiently large constant  $K$  so that  $\mathcal{E}_K = \{\|M_n(\xi) - \mathbb{E}[\xi]J_{n \times n}\| \leq K\sqrt{n}\}$  satisfies  $\mathbb{P}[\mathcal{E}_K^c] \leq \|\vec{p}\|_\infty^{2n}$ . Then, it suffices to restrict to

$\mathcal{E}_K$ . Moreover, by the triangle inequality, we see that on the event  $\mathcal{E}_K$ ,  $\|M_n(\xi)\| \leq K\sqrt{n} + \mathbb{E}[\xi]n$ , 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  $|t| \leq K\sqrt{n} + \mathbb{E}[\xi]n$ .

Now, for the treatment of vectors in  $\text{Cons}(\delta, \rho)$ , we can divide the range of  $t$  into  $n^3$  uniformly spaced intervals, apply [Corollary 4.3](#) with  $y$  equal to the mid-point of an interval times  $y_0$ , and use the union bound. For the treatment of vectors in  $\text{Noncons}(\delta, \rho) \wedge \mathcal{T}(x, L) \leq (\|\vec{p}\|_\infty + \epsilon)^n$ , we divide the range of  $t$  into  $\|\vec{p}\|_\infty^{-2n}$  equally spaced intervals, use a slight generalization of the argument in the proof [[7](#), Proposition 4.2] with  $M$  sufficiently large (depending on  $\xi$ ) also for  $y$  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.  $\square$

Using [Corollary 4.3](#) and [Proposition 4.4](#), we can prove the following weaker version of [Theorem 1.4](#).

**Theorem 4.5.** *Let  $\xi$  be a discrete random variable. For any  $\epsilon > 0$ , there exist  $C, n_0 > 0$  depending on  $\xi, \epsilon$  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 [[23](#), Section 5] with the application of [Corollary 4.3](#) and [Proposition 4.4](#) at the appropriate steps. A more complicated version of this deduction also appears in [Section 6](#), so we omit details.  $\square$

## 5. ALMOST-CONSTANT VECTORS

The goal of this section is to prove [Theorem 1.5](#). 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 5.1.** (Almost-elementary vectors) For  $\delta > 0$  and  $i, j \in [n]$ ,  $i \neq j$ , let

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

Also, let

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

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

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

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

Since  $x \notin \text{Elem}_1(\delta)$ , we must have  $\|(x_2, \dots, x_n)\|_2 \geq \delta/2$ . In case  $|x_2| \leq \delta^4$ , then we are done using [Lemma 2.4](#) (cf. [[7](#), Lemma 2.6]) for all sufficiently small  $\delta$ . Similarly, if  $\|(x_3, \dots, x_n)\|_2 \geq \delta/4$  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  $\mathcal{L}_\xi(b_1x_1 + \dots + b_nx_n, \theta) \leq \mathcal{L}_\xi(b_1x_1 + b_2x_2, \theta)$ , it suffices to bound the latter. Let  $\xi'$  be an independent copy of  $\xi$ . For any  $s \in \mathbb{R}$ , we have

$$\begin{aligned} \mathbb{P}[x_1\xi + x_2\xi' \in [s - c, s + c]]^2 &= \left( \sum_a \mathbb{P}[\xi' = a] \mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}] \right)^2 \\ &\leq \left( \sum_a \mathbb{P}[\xi' = a]^2 \right) \left( \sum_a \mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}]^2 \right) \leq \|\vec{p}\|_2^4, \end{aligned}$$

where the sum is over  $a \in \text{supp}(\xi)$ . 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  $\delta, \xi$ . Let us elaborate on this final point. We choose  $c > 0$  small enough so that  $c|x_1|^{-1} \leq c\delta^{-4}$  is smaller than  $|x_2/x_1|$  times half the minimum gap in  $\text{supp}(\xi)$ , which is possible since  $|x_2/x_1| \geq \delta^4$ . Now, such a choice of  $c$  clearly implies that each summand in  $\sum_a \mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}]^2$  covers at most a single atom in  $\text{supp}(\xi)$ , and that different choices of  $a, a' \in \text{supp}(\xi)$  cover distinct atoms in  $\text{supp}(\xi)$ .

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

$$\mathbb{P}[\xi' = \sigma(a)] = \mathbb{P}[|\xi - x_1^{-1}(s - x_2a)| \leq c|x_1|^{-1}].$$

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

$$\text{supp}(\xi) \subseteq \bigcup_{j=1}^k [x_1^{-1}(s - x_2a) - c|x_1|^{-1}, x_1^{-1}(s - x_2a) + c|x_1|^{-1}],$$

so that in particular,  $\text{supp}(\xi)$  is contained in an interval of length at most  $|x_2/x_1|m_\xi + 2c|x_1|^{-1}$ , where  $m_\xi = \max\text{supp}(\xi) - \min\text{supp}(\xi)$ . But since  $|x_2/x_1| \leq 1 - \delta^5$  and  $c|x_1|^{-1} \leq c\delta^{-4}$ , we see (by taking  $c > 0$  sufficiently small) that  $\text{supp}(\xi)$  is contained in an interval of length at most  $(1 - \delta^5/2)m_\xi$ , which contradicts the definition of  $m_\xi$ . 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

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

which completes the analysis in this case, noting that the choice of  $c$  depends only on  $\xi, \delta$ .

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

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

**Case IV:**  $(1 - \delta^5)|x_1| \leq |x_3| \leq |x_2|$  and  $|x_2| \geq \delta^4$ . It suffices to bound  $\mathcal{L}_\xi(b_1x_1 + b_2x_2 + b_3x_3)$ . Let  $u_i \in \{\pm 1\}$  be defined via  $u_i = \text{sgn}(x_i) = x_i/|x_i|$ . Let  $m'_\xi > 0$  be the smallest positive real such that  $\text{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$ .

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

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

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

Since  $|x_1| \geq |x_2| \geq \delta^4$ , 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  $\delta$  is sufficiently small based on  $\xi$ . By first choosing  $\delta > 0$  sufficiently small depending on  $\xi$ , and then choosing  $c > 0$



sufficiently small depending on  $\delta$  and  $\xi$ , we may assume that  $2(c\delta^{-4} + 3\delta m'_\xi)$  is smaller than the minimum distance between two distinct atoms in both  $\text{supp}(\xi + \xi' + \xi'')$  and  $\text{supp}(\xi + \xi' - \xi'')$ , where  $\xi, \xi', \xi''$  are independent copies of  $\xi$ . 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  $\delta$  and  $c$ , we see that it suffices to show that for all  $s \in \mathbb{R}$ ,

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

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

$$\begin{aligned} \mathbb{P}[\xi_1 + \xi_2 + u_3\xi_3 = s]^2 &= \left( \sum_a \mathbb{P}[\xi_3 = a] \mathbb{P}[\xi_1 + \xi_2 = s - u_3a] \right)^2 \\ &\leq \left( \sum_a \mathbb{P}[\xi_3 = a]^2 \right) \left( \sum_a \mathbb{P}[\xi_1 + \xi_2 = s - u_3a]^2 \right) \\ &\leq \left( \sum_a \mathbb{P}[\xi_3 = a]^2 \right) \left( \sum_{a' \in \text{supp}(\xi_1 + \xi_2)} \mathbb{P}[\xi_1 + \xi_2 = a']^2 \right) \\ &\leq \|\vec{p}\|_2^4, \end{aligned}$$

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  $\text{supp}(\xi_1 + \xi_2)$  has strictly more positive atoms than  $\text{supp}(\xi)$  (since  $\xi$  is supported on at least 2 points), and this leads to the desired improvement since  $\xi$  has finite support.  $\square$

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

**Proposition 5.3.** *Fix a discrete distribution  $\xi$  and  $\delta \in (0, 1/2)$ . Suppose that  $\xi$  is not a translate of any origin-symmetric distribution. Then, there exists  $\theta = \theta(\delta, \xi) > 0$  such that for all  $x \in \mathbb{S}^{n-1} \setminus \text{Elem}(\delta)$ ,*

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

*Proof.* As before, since  $\text{Elem}(\delta)$  is increasing with  $\delta$ , it suffices to prove the statement for sufficiently small  $\delta$  depending on  $\xi$ . By [Proposition 5.2](#), we can choose  $\theta = \theta(\delta, \xi) > 0$  such that for all  $x \in \mathbb{S}^{n-1} \setminus \text{Elem}(\delta)$ ,

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

Hence, it remains to prove the result for  $x \in \text{Elem}'(\delta) \setminus \text{Elem}(\delta)$ . By symmetry, it suffices to consider  $x \in \text{Elem}'_{1,2}(\delta)$ . 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 5.2](#). Let  $m'_\xi > 0$  be the smallest positive real for which  $\text{supp}(\xi) \subseteq [-m'_\xi, m'_\xi]$ . We have

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

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

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

We have

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

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  $\xi$  is a shift (by  $s/2$ ) of an origin-symmetric random variable. Since we have assumed that  $\xi$  is not a shift of an origin-symmetric random variable, we see that equality cannot hold, and using that the support of  $\xi$  is finite, we can conclude.  $\square$

Using the preceding lemmas, and exploiting the low metric entropy of  $\text{Cons}(\delta, \rho)$  along with [Lemma 3.4](#), we obtain the following corollary. Note that since  $\xi$  may not have mean 0, one must perform the standard trick of densifying the net of these vectors along the direction  $1_n$  (see [\[23, Proposition 3.6\]](#)). We refer the reader to [\[7, Proposition 2.7\]](#) for the (standard) details.

**Corollary 5.4.** *Fix a discrete distribution  $\xi$ . For all  $\delta' > 0$ , there exist  $\delta, \rho, \epsilon', n_0 > 0$ , depending on  $\xi$  and  $\delta'$ , such that for all  $n \geq n_0$ ,*

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

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

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

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

**Proposition 5.5.** *Fix a discrete distribution  $\xi$ . There exist  $\delta', \eta, n_0 > 0$  depending on  $\xi$  such that for all  $n \geq n_0$  and  $t \leq 1$ ,*

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

We will need the following preliminary lemma, which essentially follows from the seminal work of Rudelson and Vershynin [\[19\]](#). Since we were not able to locate the statement we need in the literature, we provide details below and in [\[7, Lemma 2.9\]](#).

**Lemma 5.6.** *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  $\kappa n$  specified entries have variance at least  $s$ .*

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

From a slight generalization of [\[7, Lemma 2.9\]](#) (specifically, one should replace the application of [\[19\]](#) with an inhomogeneous version due to [\[15\]](#) and replace  $2^{-cn}$  by  $t$ , see the proof of [Lemma 5.9](#)),

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  $2^{-\kappa n}$ ,  $b_1 = \dots = b_{\kappa n} = 0$ .  $\square$

We now prove [Proposition 5.5](#).

*Proof of Proposition 5.5.* By [Lemma 3.4](#), we can choose  $K > 0$  depending on  $\xi$  such that  $\mathbb{P}[\mathcal{E}_K] \leq \|\vec{p}\|_2^{3n}$ , where

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

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

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

where  $Q$  is the rotation matrix whose bottom-right  $(n-2) \times (n-2)$  minor is the identity matrix, and the top-left  $2 \times 2$  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  $\delta'$  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  $\|u\|_2 \leq 4\delta'$  such that

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

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

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

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

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

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

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

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

It is easy to see that

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

for an appropriate choice of  $\kappa$  which goes to 0 as  $\delta'$  goes to 0. Furthermore,

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

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

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

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

for some  $c_\xi > 0$  depending only on  $\xi$ , provided that  $\delta'$  (hence  $\kappa$ ) is chosen sufficiently small. Here, for the second inequality, we have used that by Cauchy-Schwarz (as in the proof of [Proposition 5.2](#)),

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  $\Xi$  has probability at most  $\|\vec{p}\|_2^2 - 2c_\xi$  for some  $c_\xi > 0$ .

So far, we have shown that for all  $\kappa$  and  $\delta'$  sufficiently small (depending on  $\xi$ ), we have

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

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

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

**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  $\xi^*$  has the same support as  $\xi$  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  $S_\xi$ , and all but at most  $\kappa n$  entries have variance at least  $s_\xi > 0$ . Hence, by [Lemma 5.6](#), and by using the lower bound  $\|a\|_2 \geq h_\xi$ , we find that there exist  $C', c', n_1$  depending on  $\xi$  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  $\|\vec{p}\|_2^{2n}\sqrt{t} \leq t\|\vec{p}\|_2^{2n} + \|\vec{p}\|_2^{3n}$ . The desired conclusion now follows by taking  $\eta > 0$  sufficiently small so that  $\|\vec{p}\|_2^{2n} \leq \exp(-2\eta n)$ , and then taking  $\delta'$  (hence  $\kappa$ ) sufficiently small so that  $2c_{\kappa, \delta', \xi} + \kappa < \min(c'/2, \eta/2)$ .  $\square$

The preceding proposition handles vectors in  $\text{Elem}_{i,j}(\delta)$ . If the distribution  $\xi$  is a translate of an origin-symmetric distribution, we also need to handle vectors in  $\text{Elem}'_{i,j}(\delta)$ . In case the distribution  $\xi$  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  $\text{Elem}'_{i,j}(\delta)$  when  $\xi$  is a nonzero translate of an origin-symmetric distribution, which is done by the next proposition.

**Proposition 5.7.** *Fix a discrete distribution  $\xi$  that is a nonzero translate of an origin-symmetric distribution. There exist  $\delta', \eta, n_0 > 0$  depending on  $\xi$  such that for all  $n \geq n_0$  and  $t \leq 1$ ,*

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

*Proof.* The proof is essentially the same as that of [Proposition 5.5](#). 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  $\xi$  is not origin-symmetric.

We quickly discuss the main modifications to the proof of [Proposition 5.5](#). Throughout,  $s \neq 0$  denotes a real number such that  $\xi$  and  $s - \xi$  have the same distribution (such an  $s$  exists by our assumption about  $\xi$ ). First, the top-left  $2 \times 2$  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  $\Xi = \text{supp}(\xi + \xi')/\sqrt{2}$  and as before, let

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

and for  $z \in \Xi$ ,

$$\mathcal{C}_z := \mathcal{C} \cap \{a \in \mathbb{R}^n : |\text{supp}(a - z\mathbf{1}_n)| \leq \kappa n\},$$

where  $\kappa = \kappa(\delta', \xi) > 0$  is chosen as in the previous argument. For such a choice of  $\kappa$ , 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  $z \in \Xi$ ,  $\mathbb{P}[\xi + \xi' = z\sqrt{2}] \leq \|\vec{p}\|_2^2$ , with equality holding if and only if  $\mathbb{P}[\xi = a] = \mathbb{P}[\xi' = z\sqrt{2} - a]$  for all  $a \in \text{supp}(\xi)$ , which happens if and only if  $z = s/\sqrt{2}$ .

Using this, we have as before that

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

The most important detail is that for  $\kappa \leq 1/2$  (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  $\xi$ , we can now use the analysis in **Case II** of the proof of [Proposition 5.5](#). 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  $\xi^*$  has the same support as  $\xi$ , but takes on atom  $a_i$  with probability proportional to  $p_i^2$ . The remaining details of the proof are essentially the same.  $\square$

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

**Proposition 5.8.** *Fix a discrete distribution  $\xi$ . There exist  $C', \delta', \eta, n_0 > 0$  depending on  $\xi$  such that for all  $n \geq n_0$  and  $t \leq 1$ ,*

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

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 [\[7, Proposition 2.8\]](#), except that we need to replace the application of the results of Rudelson and Vershynin [\[19\]](#) with the much sharper [Proposition 4.4](#) and [Theorem 4.5](#).

We begin with the following proposition. The analogue in the sparse Bernoulli case is [\[7, Lemma 2.9\]](#).

**Lemma 5.9.** *Fix a discrete distribution  $\xi$ . For any  $\eta \in (0, 1)$ , there exist  $C, n_0 > 0$  depending on  $\xi, \eta$  for which the following holds. For any  $v \in \mathbb{R}^n$  with  $\|v\|_2 \geq 1$ , we have*

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

where  $A$  is a random  $n \times (n - 1)$  matrix with independent  $\xi$  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  $\|v'\|_2 \geq 1/2$ . Let  $\mathcal{E}_S = \{s_{n-1}(A_{n-1}) \leq \sqrt{t}\}$ . By [Theorem 4.5](#), we have that for all sufficiently large  $n$ , there exists a constant  $C'$  depending on  $\xi$  and  $\eta$  such that

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

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  $\delta, \rho, L$  to be chosen momentarily, let

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

By [Proposition 4.4](#), we can choose  $\delta, \rho, L > 0$  depending on  $\xi$  and  $\eta$  so that

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

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

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

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

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

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  $\|y\|_2 \geq 1/C''n^2$ . Now, we use the independence of  $R$  and  $A_{n-1}$  and the fact that  $\mathcal{E}$  implies

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

Since

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

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

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

Now we are ready to conclude [Proposition 5.8](#).

*Proof of Proposition 5.8.* A completely identical argument to the proof of [Proposition 5.5](#) shows that for a sufficiently large constant  $K$  depending on  $\xi$ , and for  $\Xi := \text{supp}(\xi) \subseteq \mathbb{R}$ ,

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

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

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

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  $\|a\|_2 \geq h_\xi > 0$ . Hence, by [Lemma 5.9](#), there is a constant  $C > 0$  depending on  $\xi, \eta$  such that

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

Moreover, a similar (but easier) argument as in the proof of [Proposition 5.5](#) shows that

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

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

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

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

By choosing  $\delta'$  sufficiently small depending on  $\xi$  and  $\eta$ , and using  $\|\vec{p}\|_\infty^2 \leq \|\vec{p}\|_2^2 - c_\xi$  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  $\eta'$  depending on  $\xi$  and  $\eta$ . This completes the proof.  $\square$

The proof of [Theorem 1.5](#) is now immediate.

*Proof of [Theorem 1.5](#).* First, assume that  $\xi$  is not a shift of an origin-symmetric random variable. We choose  $\delta'$  small enough so that the conclusions of [Propositions 5.5](#) and [5.8](#) are satisfied. By the union bound, this shows that the contribution of  $\text{Elem}(\delta')$  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  $\xi$ , and for all sufficiently large  $n$  depending on  $\xi$ . Now, we can conclude using [Corollary 5.4](#).

Next, if  $\xi$  is a nonzero shift of an origin-symmetric random variable, we do the same, except we require [Propositions 5.5](#), [5.7](#), and [5.8](#) and then conclude with [Corollary 5.4](#).

Finally, we consider the case when  $\xi$  is an origin-symmetric random variable. As before, we begin by using [Propositions 5.5](#) and [5.8](#). The only thing to note is that, by the symmetry of  $\xi$  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  $\text{Elem}'(\delta')$  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 5.4](#).  $\square$

## 6. DEDUCTION OF [THEOREMS 1.2](#) AND [1.4](#)

Given the results in [Sections 3](#) and [5](#), the deduction of [Theorems 1.2](#) and [1.4](#) is immediate. Fix a discrete distribution  $\xi$ , and let  $\delta, \rho, \eta, n_0 > 0$  be parameters depending on  $\xi$  coming from [Theorem 1.5](#). Then, for the proof of [Theorem 1.4](#), 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

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

and equality holds if and only if  $\xi$  is uniform on its support.

Let  $C = C_{3.8}(\xi, \delta, \rho, \epsilon/2)$ . By taking  $C_{\xi, \epsilon}$  in [Theorem 1.4](#) and  $C_\xi$  in [Theorem 1.2](#) to be at least  $C$ , we may restrict our attention to  $0 \leq t \leq 1/C$  (since for  $t \geq 1/C$ , the right-hand sides of [Theorems 1.2](#) and [1.4](#) are at least 1). By [Theorem 1.5](#) and [Theorem 3.8](#), for all  $0 \leq t \leq 1/C$ , we have

$$\begin{aligned} \mathbb{P}[s_n(M_n(\xi)) \leq t/\sqrt{n}] &\leq \mathbb{P}\left[ \inf_{x \in \text{Cons}(\delta, \rho)} \|M_n(\xi)x\|_2 \leq t/\sqrt{n} \right] + \mathbb{P}\left[ \inf_{y \in \text{Cons}(\delta, \rho)} \|yM_n(\xi)\|_2 \leq Ct \right] \\ &+ \mathbb{P}\left[ \inf_{x \in \text{Noncons}(\delta, \rho)} \|M_n(\xi)x\|_2 \leq t/\sqrt{n} \wedge \inf_{y \in \text{Cons}(\delta, \rho)} \|yM_n(\xi)\|_2 > Ct \right] \end{aligned}$$

$$\begin{aligned} &\leq 2n\mathbb{P}[\mathcal{E}_{e_1}] + (n^2 - n)(\mathbb{P}[\mathcal{E}_{e_1-e_2}] + \mathbb{P}[\mathcal{E}_{e_1+e_2}]) + 2(Ct + \|\vec{p}\|_2^{2n})e^{-\eta n} \\ &\quad + Ct + \exp((\epsilon/2 - H(\vec{p}))n) \end{aligned}$$

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

For [Theorem 1.2](#), we are done by our choice of  $\epsilon$ .

For [Theorem 1.4](#), we note that by Cauchy-Schwarz (as in [Proposition 5.2](#)),  $\mathbb{P}[\mathcal{E}_{e_1+e_2}] \leq \mathbb{P}[\mathcal{E}_{e_1-e_2}]$  and recall from above that  $\exp(-H(\vec{p})) \leq \|\vec{p}\|_2^2$ . 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  $2n^2 \exp(\epsilon n/2) \leq (1 + \epsilon)^n$  for all  $\epsilon < 1$  and  $n$  sufficiently large.

## REFERENCES

- [1] Anirban Basak and Mark Rudelson, *Sharp transition of the invertibility of the adjacency matrices of sparse random graphs*, arXiv:1809.08454.
- [2] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), 559–603.
- [3] Asaf Ferber, Vishesh Jain, Kyle Luh, and Wojciech Samotij, *On the counting problem in inverse Littlewood–Offord theory*, arXiv:1904.10425.
- [4] Han Huang, *Singularity of Bernoulli matrices in the sparse regime  $pn = O(\log n)$* , arXiv:2009.13726.
- [5] A. Irmatov, *Asymptotics of the number of threshold functions and the singularity probability of random  $\{\pm 1\}$ -matrices*, Doklady Mathematics **101** (2020), 247–249.
- [6] Vishesh Jain, *Approximate Spielman-Teng theorems for the least singular value of random combinatorial matrices*, arXiv:1904.10592.
- [7] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *Sharp invertibility of random Bernoulli matrices*, in preparation.
- [8] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney, *The smallest singular value of dense random regular digraphs*, arXiv:2008.04755.
- [9] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random  $\pm 1$ -matrix is singular*, J. Amer. Math. Soc. **8** (1995), 223–240.
- [10] Boris Sergeevich Kašin, *Diameters of some finite-dimensional sets and classes of smooth functions*, Izvestiya: Mathematics **11** (1977), 317–333.
- [11] J. Komlós, *On the determinant of  $(0, 1)$  matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21.
- [12] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), 491–523.
- [13] Alexander E Litvak and Konstantin E Tikhomirov, *Singularity of sparse Bernoulli matrices*, arXiv:2004.03131.
- [14] Galyna V Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, arXiv:1811.07038.
- [15] Galyna V Livshyts, Konstantin Tikhomirov, and Roman Vershynin, *The smallest singular value of inhomogeneous square random matrices*, arXiv:1909.04219.
- [16] Hoi H. Nguyen, *On the singularity of random combinatorial matrices*, SIAM J. Discrete Math. **27** (2013), 447–458.
- [17] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Veroyatnost. i Primenen **6** (1961), 106–108.
- [18] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Annals of Mathematics (2008), 575–600.
- [19] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), 600–633.
- [20] Gideon Schechtman, *Special orthogonal splittings of  $L_1^{2k}$* , Israel J. Math. **139** (2004), 337–347.
- [21] Terence Tao and Van Vu, *On random  $\pm 1$  matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), 1–23.
- [22] Terence Tao and Van Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), 603–628.
- [23] Konstantin Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), 593–634.
- [24] Tuan Tran, *The smallest singular value of random combinatorial matrices*, arXiv:2007.06318.



- [25] 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`