Forum of Mathematics, Sigma (2019), Vol. 7, e22, 29 pages
doi:10.1017/fms.2019.21

SINGULARITY OF RANDOM SYMMETRIC MATRICES—A COMBINATORIAL APPROACH TO IMPROVED BOUNDS

ASAF FERBER and VISHESH JAIN
Massachusetts Institute of Technology, Department of Mathematics, USA;
email: ferbera@mit.edu, visheshj@mit.edu

Received 6 February 2019; accepted 8 July 2019

Abstract

Let M_n denote a random symmetric $n \times n$ matrix whose upper-diagonal entries are independent and identically distributed Bernoulli random variables (which take values 1 and -1 with probability $1/2$ each). It is widely conjectured that M_n is singular with probability at most $(2+o(1))^{-n}$. On the other hand, the best known upper bound on the singularity probability of M_n, due to Vershynin (2011), is 2^{-n^c}, for some unspecified small constant $c > 0$. This improves on a polynomial singularity bound due to Costello, Tao, and Vu (2005), and a bound of Nguyen (2011) showing that the singularity probability decays faster than any polynomial. In this paper, improving on all previous results, we show that the probability of singularity of M_n is at most $2^{-n^{1/4}\sqrt{\log n}/1000}$ for all sufficiently large n. The proof utilizes and extends a novel combinatorial approach to discrete random matrix theory, which has been recently introduced by the authors together with Luh and Samotij.

2010 Mathematics Subject Classification: 60B20

1. Introduction

The invertibility problem for Bernoulli matrices is one of the most well-studied problems in discrete random matrix theory. Letting A_n denote a random $n \times n$ matrix, whose entries are independent and identically distributed (i.i.d.) Bernoulli random variables which take values ± 1 with probability $1/2$ each, this problem asks for the value of c_n, which is the probability that A_n is singular. By considering the event that two rows or two columns of A_n are equal (up to a sign), it is clear that

$$c_n \geqslant (1 + o(1))n^2 2^{1-n}.$$

It has been widely conjectured that this bound is, in fact, tight. On the other hand, perhaps surprisingly, it is nontrivial even to show that c_n tends to 0 as n goes to infinity—this was first accomplished in 1967 by Komlós [8], who showed using the classical Erdős–Littlewood–Offord anticoncentration inequality that

$$c_n = O(n^{-1/2}).$$

Subsequently, a breakthrough result due to Kahn, Komlós, and Szemerédi in 1995 [7] showed that

$$c_n = O(0.999^n).$$

After intermediate improvements in the base of the exponent due to Tao and Vu [15] and Bourgain, Vu, and Wood [1], this conjecture has been settled up to lower order terms recently (in fact, a few months after the appearance of the present work) in a very impressive work of Tikhomirov [17], showing that

$$c_n \leqslant (2 + o(1))^{-n}.$$

Another widely studied model of random matrices is that of random *symmetric* matrices; apart from being important for applications, it is also very interesting from a technical perspective as it is one of the simplest models with nontrivial correlations between the entries of the matrix. Formally, let M_n denote a random $n \times n$ symmetric matrix, whose upper-diagonal entries are i.i.d. Bernoulli random variables which take values ± 1 with probability $1/2$ each, and let q_n denote the probability that M_n is singular. Despite its similarity to c_n, much less is known about q_n, as we discuss below.

The problem of determining whether q_n tends to 0 as n goes to infinity was first posed by Weiss in the early 1990s and only settled in 2005 by Costello, Tao, and Vu [2], who showed that

$$q_n = O(n^{-1/8+o(1)}).$$

In order to do this, they introduced and studied a quadratic variant of the Erdős-Littlewood–Offord inequality. Subsequently, Nguyen [9] developed a quadratic variant of *inverse* Littlewood–Offord theory to show that

$$q_n = O_C(n^{-C})$$

for any $C > 0$, where the implicit constant in $O_C(\cdot)$ depends only on C. This so-called quadratic inverse Littlewood–Offord theorem in [9] builds on previous work of Nguyen and Vu [10], which is itself based on deep Freiman-type theorems in additive combinatorics (see [16] and the references therein). The current best known upper bound on q_n is due to Vershynin [18], who used a sophisticated and

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

technical geometric framework pioneered by Rudelson and Vershynin [13, 14] to show that

$$q_n = O(2^{-n^c})$$

for some unspecified small constant $c > 0$.

As far as lower bounds on q_n are concerned, once again, by considering the event that the first and last rows of M_n are equal (up to a sign), we see that $q_n \geqslant (2 + o(1))^{-n}$. It is commonly believed that this lower bound is tight.

CONJECTURE 1.1 [2, 19]. *We have*

$$q_n = (2 + o(1))^{-n}.$$

In this paper, we obtain a much stronger upper bound on q_n, thereby making progress towards Conjecture 1.1.

THEOREM 1.2. *There exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$,*

$$q_n \leqslant 2^{-n^{1/4}\sqrt{\log n}/1000}.$$

REMARK 1.3. While the constant 1000 in the above theorem is somewhat arbitrary, the leading order term $n^{1/4}\sqrt{\log n}$ in the exponent is optimal for the argument in this paper. We believe that improving the exponent to even $n^{(1/2)+\epsilon}$ (for some absolute constant $\epsilon > 0$) will likely require new ideas beyond those in the present work, since even in the case of i.i.d. Rademacher random matrices, the combinatorial techniques from [3] that we build upon here are only able to obtain an upper bound of $2^{-\tilde{\Omega}(\sqrt{n})}$ on the singularity probability.

Apart from providing a stronger conclusion, our proof of the above theorem is considerably shorter than previous works, and introduces and extends several novel combinatorial tools and ideas in discrete random matrix theory (some of which are based on joint work of the authors with Luh and Samotij [3]). We believe that these ideas allow for a unified approach to the singularity problem for many different discrete random matrix models, which have previously been handled in an *ad hoc* manner (see also the discussion at the end of the next subsection).

1.1. Outline of the proof and comparison with previous work.

In this subsection, we provide a very brief, and rather imprecise, outline of our proof, and compare it to previous works of Nguyen [9] and Vershynin [18]; for further comparison with the work of Costello, Tao, and Vu, see [9].

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

Let $\boldsymbol{x} := (x_1, \ldots, x_n)$ be the first row of M_n, let M^1_{n-1} denote the bottom-right $(n-1) \times (n-1)$ submatrix of M_n, and for $2 \leqslant i, j \leqslant n$, let c_{ij} denote the cofactor of M^1_{n-1} obtained by removing its $(i-1)^{st}$ row and $(j-1)^{st}$ column. Then, Laplace's formula for the determinant gives

$$\det(M_n) = x_1 \det(M_{n-1}) - \sum_{i,j=2}^{n} c_{ij} x_i x_j,$$

so that our goal is to bound the probability (over the randomness of x and c_{ij}) that this polynomial is zero. By a standard reduction due to [2] (see Lemmas 2.1 and 2.3 and Corollary 2.4), we may further assume that M^1_{n-1} has rank either $n-2$ or $n-1$. In this outline, we will only discuss the case when M^1_{n-1} has rank $n-1$; the other case is easier, and is handled exactly as in [9] (see Lemma 2.5 and equation (8)).

A decoupling argument due to [2] (see Lemma 2.10) further reduces the problem (albeit in a manner incurring a loss) to bounding from above the probability that

$$\sum_{i \in U_1} \sum_{j \in U_2} c_{ij}(x_i - x'_i)(x_j - x'_j) = 0,$$

where $U_1 \sqcup U_2$ is an arbitrary nontrivial partition of $[n-1]$, and x'_i, x'_j are independent copies of x_i, x_j (see Corollary 2.11). For the remainder of this discussion, the reader should think of $|U_2|$ as 'small'(more precisely, $|U_2| \sim n^{1/4}\sqrt{\log n}$). We remark that a similar decoupling based reduction is used in [18] as well, whereas [9] also uses a similar decoupling inequality in proving the so-called quadratic inverse Littlewood–Offord theorem. The advantage of decoupling is that for any given realization of the variables $(c_{ij})_{2 \leqslant i,j \leqslant n}$ and $(x_j - x'_j)_{j \in U_2}$, the problem reduces to bounding from above the probability that the *linear sum*

$$\sum_{i \in U_1} R_i(x_i - x'_i) = 0,$$

where $R_i := \sum_{j \in U_2} c_{ij}(x_j - x'_j)$. Problems of this form are precisely the subject of standard (linear) Littlewood–Offord theory.

Broadly speaking, Littlewood–Offord theory applied to our problem says that the less 'additive structure' the $|U_1|$-dimensional vector $(R_i)_{i \in U_1}$ possesses, the smaller the probability of the above sum being zero. Quantifying this in the form of 'Littlewood–Offord type theorems' has been the subject of considerable research over the years; we refer the reader to [11, 14] for general surveys on the Littlewood–Offord problem with a view towards random matrix theory. Hence, our goal is to show that with very high probability, the vector $(R_i)_{i \in U_1}$ is additively

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

'very unstructured'. This is the content of our structural theorem (Theorem 3.2), which is at the heart of our proof.

The statement (and usefulness) of our structural theorem is based on the following simple, yet powerful, observations.

- The $(n-1)$-dimensional vector $\boldsymbol{R} := (R_2, \ldots, R_n)$, where recall that $R_i = \sum_{j \in U_2} c_{ij}(x_j - x'_j)$, is zero if and only if $x_j = x'_j$ for all $j \in |U_2|$, which happens with probability exponentially small in $|U_2|$; the if and only if statement holds since the matrix $(c_{ij})_{2 \leqslant i,j \leqslant n}$ is proportional to the matrix $(M^1_{n-1})^{-1}$, which is assumed to be invertible.

- The vector $\boldsymbol{R}$ is orthogonal to at least $n-1-|U_2|$ rows of M^1_{n-1} (Lemma 2.12). This follows since for any $2 \leqslant j_0 \leqslant n$, the $n-1$ dimensional vector $(c_{ij_0})_{2 \leqslant i \leqslant n}$ is orthogonal to all but the j_0th row of M^1_{n-1}, again since the matrix $(c_{ij})_{2 \leqslant ij \leqslant n}$ is proportional to the matrix $(M^1_{n-1})^{-1}$.

- The probability of the linear sum $\sum_{i \in U_1} R_i(x_i - x'_i)$ being zero is 'not much more' than the probability of the linear sum $\sum_{2 \leqslant i \leqslant n} R_i(x_i - x'_i)$ being zero (Lemma 2.9).

Taken together, these observations show that it suffices to prove a structural theorem of the following form: *every* nonzero integer vector which is orthogonal to 'most' rows of M^1_{n-1} is 'very unstructured'. In [9], a structural theorem along similar lines is also proven. However, it suffers from two drawbacks. First, the notion of 'very unstructured' in the conclusion there is much weaker, leading to the bound $O_C(n^{-C})$ for any constant $C > 0$, as opposed to our bound from Theorem 1.2. Second, such a conclusion is not obtained for every nonzero integer vector, but only for those nonzero integer vectors for which 'most' coefficients satisfy the additional additive constraint of being contained in a 'small' generalized arithmetic progression (GAP) of 'low complexity'. Consequently, the simple observations mentioned above no longer suffice, and the rest of the proof in [9] is necessarily more complicated.

The structural theorem in [18] is perhaps closer in spirit to ours, although there are many key differences, of which we mention here the most important one. Roughly speaking, both [18] and the present work prove the respective structural theorems by taking the union bound, over the choice of a nonzero (integer) vector which is not 'very unstructured', that the matrix–vector product of M^1_{n-1} with this vector is contained in a small prescribed set. *A priori*, this union bound is over an infinite collection of vectors. In order to overcome this obstacle, [13, 18] adopt a geometric approach of grouping vectors on the unit sphere into a finite number of clusters based on Euclidean distances; using the union bound and a nontrivial estimate of the number of clusters to show that with very high probability, the

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

matrix–vector product of M^1_{n-1} with a representative of each cluster is 'far' from the small prescribed set; and then, using estimates on the operator norm of M^1_{n-1} to deduce a similar result for all other vectors in each cluster. Naturally, this geometric approach is very involved, and leads to additional losses at various steps (which is why [18] obtains a worse bound on q_n than Theorem 1.2).

In contrast, we overcome this obstacle with a completely novel and purely combinatorial approach of clustering vectors based on the residues of their coordinates modulo a large prime, and using a combinatorial notion due to Halász [4] to quantify the amount of additive structure in a vector (Proposition 3.3). In particular, with our approach, the analogue of the problem of 'bounding the covering number of sublevel sets of regularized LCD'—which constitutes a significant portion of [18] (see Section 7.1 there), is one of the key contributions of that work, and is also a major contributor to the suboptimality of the final result—can be solved more efficiently and with a short double-counting argument (see Theorem 3.10, which is based on joint work of the authors with Luh and Samotij in [3], and Corollary 3.11).

It is worth mentioning that [18] provides bounds not just for the probability of singularity of M_n, but also for the probability that the 'least singular value' of M_n (as well as random matrices with more general entries) is 'very small'. Very recent work [5, 6] of the second author shows how to develop the combinatorial ideas introduced in [3] (which we use here) in order to obtain quantitative control on the lower tail of the least singular value for a variety of random matrix models. We anticipate that the ideas in the present work can be combined with those in [5, 6] to control the lower tail of the least singular value of symmetric random matrices as well.

The rest of this paper is organized as follows. In Section 2, we discuss in detail the overall proof strategy leading to the reduction to the structural theorem; in Section 3, we state and prove our structural theorem; and in Section 4, we put everything together to quickly complete our proof.

Notation: Throughout the paper, we will omit floors and ceilings when they make no essential difference. For convenience, we will also say 'let $p = x$ be a prime', to mean that p is an odd prime between x and $2x$; again, this makes no difference to our arguments. As is standard, we will use $[n]$ to denote the discrete interval $\{1, \ldots, n\}$. All logarithms are natural unless noted otherwise.

2. Proof strategy: reduction to the structural theorem

In this section, we discuss the strategy underlying our proof of Theorem 1.2. The key conclusions are equations (2), (8), and (12), which show that it suffices to prove the structural theorem in Section 3 in order to prove Theorem 1.2.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

2.1. Preliminary reductions. For any $n \in \mathbb{N}$ and $k \in [n]$, let $\mathcal{R}k_k(n)$ denote the event that M_n has rank exactly k, and let $\mathcal{R}k_{\leqslant k}(n)$ denote the event that M_n has rank at most k. Thus, our goal is to bound the probability of $\mathcal{R}k_{\leqslant n-1}(n)$. The next lemma, which is due to Nguyen [9], shows that it suffices to bound the probability of $\mathcal{R}k_{n-1}(n)$.

LEMMA 2.1 [9, Lemma 2.1]. *For any $\ell \in [n-2]$,*

$$\Pr\left[\mathcal{R}k_\ell(n)\right] \leqslant 0.1 \times \Pr\left[\mathcal{R}k_{2n-\ell-2}(2n-\ell-1)\right].$$

The proof of this lemma uses the following simple observation due to Odlyzko [12]:

OBSERVATION 2.2. *Let V be any subspace of $\mathbb{R}^n$ of dimension at most ℓ. Then, $|V \cap \{\pm 1\}^n| \leqslant 2^\ell$.*

Proof of Lemma 2.1. It suffices to show that for any $\ell \leqslant n-2$,

$$\Pr\left[\mathcal{R}k_{\ell+2}(n+1) \mid \mathcal{R}k_\ell(n)\right] \geqslant 1 - 2^{-n+\ell}. \tag{1}$$

Indeed, iterating this equation shows that

$$\begin{aligned}
&\Pr[\mathcal{R}k_{2n-\ell-2}(2n-\ell-1) \mid \mathcal{R}k_\ell(n)] \\
&\quad \geqslant \prod_{j=1}^{n-\ell-1} \Pr\left[\mathcal{R}k_{\ell+2j}(n+j) \mid \mathcal{R}k_{\ell+2j-2}(n+j-1)\right] \\
&\quad \geqslant \prod_{j=1}^{n-\ell-1} (1 - 2^{-n+\ell+j}) \geqslant 0.1,
\end{aligned}$$

which gives the desired conclusion.

In order to prove equation (1), consider the coupling of M_n and M_{n+1} where M_n is the top-left $n \times n$ submatrix of M_{n+1}. Suppose M_n has rank ℓ, and let $V(M_n)$ be the (ℓ-dimensional) subspace spanned by its rows. By Observation 2.2, $|V(M_n) \cap \{\pm 1\}^n| \leqslant 2^\ell$. Therefore, the probability that the vector formed by the first n coordinates of the last row of M_{n+1} lies in $V(M_n)$ is at most $2^{-n+\ell}$. If this vector does not lie in $V(M_n)$, then the symmetry of the matrix also shows that the last column of M_{n+1} does not lie in the span of the first n columns of M_{n+1}, so that the rank of M_{n+1} exceeds the rank of M_n by 2. □

The following lemma, also due to Nguyen, allows us to reduce to the case where the rank of the $(n-1) \times (n-1)$ symmetric matrix obtained by removing the first row and the first column of M_n is at least $n-2$.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

LEMMA 2.3 [9, Lemma 2.3]. *Assume that M_n has rank $n-1$. Then, there exists $i \in [n]$ such that the removal of the ith row and the ith column of M_n results in a symmetric matrix M_{n-1} of rank at least $n-2$.*

Proof. Without loss of generality, we can assume that the last $n-1$ rows of M_n are independent. Therefore, the matrix M_{n-1}, which is obtained by removing the first row and first column of M_n has rank at least $n-2$. □

As a simple corollary of the above lemma, we obtain the following:

COROLLARY 2.4. *For $i \in [n]$, let $\mathcal{R}k^{i}_{n-1}(n)$ denote the event that M_n has rank $n-1$, and the symmetric matrix obtained by removing the ith row and the ith column of M_n has rank at least $n-2$. Then,*

$$\Pr\left[\mathcal{R}k_{n-1}(n)\right] \leqslant n \Pr\left[\mathcal{R}k^{1}_{n-1}(n)\right].$$

Proof. Suppose that M_n has rank $n-1$. By Lemma 2.3, there exists an $i \in [n]$ for which the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and ith column has rank at least $n-2$. Moreover, by symmetry,

$$\Pr[\mathcal{R}k^{i}_{n-1}(n)] = \Pr[\mathcal{R}k^{1}_{n-1}(n)] \quad \text{for all } i \in [n].$$

Therefore, by the union bound,

$$\Pr[\mathcal{R}k_{n-1}(n)] = \Pr\left[\bigcup_{i=1}^{n} \mathcal{R}k^{i}_{n-1}(n)\right] \leqslant \sum_{i=1}^{n} \Pr[\mathcal{R}k^{i}_{n-1}(n)] = n \Pr[\mathcal{R}k^{1}_{n-1}(n)].$$

□

Let M^{1}_{n-1} denote the $(n-1)\times(n-1)$ symmetric matrix obtained by deleting the first row and first column of M_n. Let $\mathcal{D}(n-1)$ denote the 'degenerate' event that M^{1}_{n-1} has rank $n-2$, and let $\mathcal{ND}(n-1)$ denote the 'nondegenerate' event that M^{1}_{n-1} has full rank $n-1$. By definition,

$$\mathcal{R}k^{1}_{n-1}(n) = \left(\mathcal{R}k^{1}_{n-1}(n) \cap \mathcal{D}(n-1)\right) \sqcup \left(\mathcal{R}k^{1}_{n-1}(n) \cap \mathcal{ND}(n-1)\right),$$

and hence,

$$\Pr\left[\mathcal{R}k^{1}_{n-1}(n)\right] = \Pr\left[\mathcal{R}k^{1}_{n-1}(n) \cap \mathcal{D}(n-1)\right] + \Pr\left[\mathcal{R}k^{1}_{n-1}(n) \cap \mathcal{ND}(n-1)\right]. \quad (2)$$

It is thus enough to bound each of the above two summands.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

2.2. Bounding $\Pr\left[\mathcal{R}k_{n-1}^1(n) \cap \mathcal{D}(n-1)\right]$.

Let $\boldsymbol{x} := (x_1, \ldots, x_n)$ denote the first row of M_n. It follows from Laplace's formula for the determinant that

$$\det(M_n) = x_1 \det\left(M_{n-1}^1\right) - \sum_{2 \leqslant i,j \leqslant n} c_{ij} x_i x_j, \tag{3}$$

where c_{ij} denotes the cofactor of M_{n-1}^1 obtained by removing its $(i-1)^{st}$ row and $(j-1)^{st}$ column. In order to deal with $M_n \in \mathcal{R}k_{n-1}^1(n) \cap \mathcal{D}(n-1)$, we use the following observation due to Nguyen (see [9, Section 9]).

LEMMA 2.5. *For every* $M_n \in \mathcal{R}k_{n-1}^1(n) \cap \mathcal{D}(n-1)$, *there exists some* $\lambda := \lambda\left(M_{n-1}^1\right) \in \mathbb{Q} \setminus \{0\}$ *and some* $\boldsymbol{a} := \boldsymbol{a}\left(M_{n-1}^1\right) = (a_2, \ldots, a_n) \in \mathbb{Z}^{n-1} \setminus \{\boldsymbol{0}\}$ *such that*

$$M_{n-1}^1 \boldsymbol{a} = \boldsymbol{0}, \tag{4}$$

and

$$\det(M_n) = \lambda \left(\sum_{2 \leqslant i \leqslant n} a_i x_i \right)^2. \tag{5}$$

Proof. Let $\operatorname{adj}\left(M_{n-1}^1\right)$ denote the adjugate matrix of M_{n-1}^1; note that this is an integer-valued symmetric matrix since M_{n-1}^1 is an integer-valued symmetric matrix. Since M_{n-1}^1 is of rank $n-2$, its kernel is of rank 1. Moreover, the equation

$$M_{n-1}^1 \operatorname{adj}\left(M_{n-1}^1\right) = \det\left(M_{n-1}^1\right) I_{n-1} \tag{6}$$

shows that every column of $\operatorname{adj}\left(M_{n-1}^1\right)$ is in the kernel of M_{n-1}^1 as $\det(M_{n-1}^1) = 0$ by assumption. It follows that the matrix $\operatorname{adj}\left(M_{n-1}^1\right)$ is an integer-valued symmetric matrix of rank 1, which cannot be zero since M_{n-1}^1 is of rank $n-2$. Hence, there exists some $\lambda \in \mathbb{Q} \setminus \{0\}$ and a vector $\boldsymbol{a} = (a_2, \ldots, a_n)^T \in \mathbb{Z}^{n-1} \setminus \{\boldsymbol{0}\}$ such that

$$\operatorname{adj}\left(M_{n-1}^1\right) = \lambda \boldsymbol{a}\boldsymbol{a}^T. \tag{7}$$

In particular, every column of $\operatorname{adj}\left(M_{n-1}^1\right)$ is equal to a multiple of the vector $\boldsymbol{a}$. By considering any column which is a nonzero multiple of $\boldsymbol{a}$, equation (6) along with $\det\left(M_{n-1}^1\right) = 0$ gives equation (4). Moreover, by writing the entries of the adjugate matrix in terms of the cofactors, we see that equation (7) is equivalent to the following: for all $2 \leqslant i, j \leqslant n$:

$$c_{ij} = \lambda a_i a_j.$$

Substituting this in equation (3) and using $\det\left(M_{n-1}^1\right) = 0$ gives equation (5). □

Before explaining how to use Lemma 2.5, we need the following definition.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

DEFINITION 2.6 (Atom probability). Let $\mathfrak{R}$ be an arbitrary ring (with a unit element). For a vector $\boldsymbol{a} := (a_1, \ldots, a_n) \in \mathfrak{R}^n$, we define its μ-atom probability by

$$\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}) := \sup_{c \in \mathfrak{R}} \Pr_{x_1^\mu, \ldots, x_n^\mu} \left[a_1 x_1^\mu + \cdots + a_n x_n^\mu = c\right],$$

where the x_i^μ's are i.i.d. random variables taking on the value 0 with probability μ and the values ± 1, each with probability $(1-\mu)/2$.

REMARK 2.7. We will often refer to the 0-atom probability simply as the atom probability, and denote it by $\rho^{\mathfrak{R}}(\boldsymbol{a})$ instead of $\rho_0^{\mathfrak{R}}(\boldsymbol{a})$. Similarly, we will denote x_i^0 simply as x_i.

Although we will not need them in this subsection, we will later make use of the following two simple lemmas about the atom probability. The first lemma shows that the μ-atom probability of a vector is bounded above by the μ-atom probability of any of its restrictions.

LEMMA 2.8. *Let $\boldsymbol{a} \in \mathfrak{R}^n$, and let $\boldsymbol{a}|_{U_1}$ denote the restriction of $\boldsymbol{a}$ to $U_1 \subseteq [n]$. Then,*

$$\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}) \leqslant \rho_\mu^{\mathfrak{R}}\big(\boldsymbol{a}|_{U_1}\big).$$

Proof. Let $c^* := \arg\max_{c \in \mathfrak{R}} \Pr_{\boldsymbol{x}^\mu}\left[\sum_{i \in [n]} a_i x_i^\mu = c\right]$. Then,

$$\begin{aligned}\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}) &= \Pr_{\boldsymbol{x}^\mu}\left[\sum_{i \in [n]} a_i x_i^\mu = c^*\right] = \Pr_{\boldsymbol{x}^\mu}\left[\sum_{i \in [U_1]} a_i x_i^\mu = c^* - \sum_{i \in [\overline{U_1}]} a_i x_i^\mu\right] \\ &= \mathbb{E}_{(x_i^\mu)_{i \in \overline{U_1}}}\left[\Pr_{(x_i^\mu)_{i \in [U_1]}}\left[\sum_{i \in [U_1]} a_i x_i^\mu = c^* - \sum_{i \in [\overline{U_1}]} a_i x_i^\mu\right]\right] \\ &\leqslant \mathbb{E}_{(x_i^\mu)_{i \in \overline{U_1}}}\left[\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1})\right] = \rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1}),\end{aligned}$$

where the third equality follows from the law of total probability, and the fourth inequality follows from the definition of $\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1})$. □

The second lemma complements Lemma 2.8, and shows that the μ-atom probability cannot increase too much if, instead of the original vector, we work with its restriction to a sufficiently large subset of coordinates.

LEMMA 2.9. *Let $\boldsymbol{a} \in \mathfrak{R}^n$, and let $\boldsymbol{a}|_{U_1}$ denote the restriction of $\boldsymbol{a}$ to U_1. Then,*

$$\rho_\mu^{\mathfrak{R}}\big(\boldsymbol{a}|_{U_1}\big) \leqslant \max\left\{\mu, \frac{1-\mu}{2}\right\}^{-|U_2|} \rho_\mu^{\mathfrak{R}}(\boldsymbol{a}).$$

Proof. Let $c_0 := \arg\max_{c \in \mathfrak{R}} \Pr_{\boldsymbol{x}^\mu}\left[\sum_{i \in U_1} a_i x_i^\mu = c\right]$ where the x_i^μ's are as in Definition 2.6, and let $c_1 := c_0 + \sum_{i \in U_2} a_i$. Then,

$$\Pr_{\boldsymbol{x}^\mu}\left[\sum_{i\in[n]} a_i x_i^\mu = c_0\right] \geqslant \Pr_{(x_i^\mu)_{i\in U_1}}\left[\sum_{i\in U_1} a_i x_i^\mu = c_0\right] \prod_{j\in U_2} \Pr_{x_j^\mu}\left[x_j^\mu = 0\right]$$
$$\geqslant \rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1})\mu^{|U_2|},$$

and

$$\Pr_{\boldsymbol{x}^\mu}\left[\sum_{i\in[n]} a_i x_i^\mu = c_1\right] \geqslant \Pr_{(x_i^\mu)_{i\in U_1}}\left[\sum_{i\in U_1} a_i x_i^\mu = c_0\right] \prod_{j\in U_2} \Pr_{x_j^\mu}\left[x_j^\mu = 1\right]$$
$$\geqslant \rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1})\left(\frac{1-\mu}{2}\right)^{|U_2|}.$$

Taking the maximum of the two expressions gives

$$\rho_\mu^{\mathfrak{R}}(\boldsymbol{a}) \geqslant \max\left\{\mu, \frac{1-\mu}{2}\right\}^{|U_2|} \rho_\mu^{\mathfrak{R}}(\boldsymbol{a}|_{U_1}),$$

and by rearranging we obtain the desired conclusion. □

Returning to the goal of this subsection, for $0 < \rho \leqslant 1$, let $\mathcal{N}ull_\rho(n-1)$ denote the event—depending only on M_{n-1}^1—that *every* nonzero integer null vector of M_{n-1}^1 has atom probability (in $\mathbb{Z}$) at most ρ. Then, we have

$$\Pr_{M_n}\left[\mathcal{R}k_{n-1}^1(n) \cap \mathcal{D}(n-1)\right] \leqslant \Pr_{M_n}\left[\mathcal{R}k_{n-1}^1(n) \cap \mathcal{D}(n-1) \cap \mathcal{N}ull_\rho(n-1)\right]$$
$$+ \Pr_{M_{n-1}^1}\left[\overline{\mathcal{N}ull_\rho(n-1)}\right]$$
$$\leqslant \Pr_{M_{n-1}^1, \boldsymbol{x}}\left[\left(\sum_{2\leqslant i\leqslant n} a_i(M_{n-1}^1)x_i = 0\right) \cap \mathcal{N}ull_\rho(n-1)\right]$$
$$+ \Pr_{M_{n-1}^1}\left[\overline{\mathcal{N}ull_\rho(n-1)}\right]$$
$$\leqslant \sum_{A_{n-1}\in\mathcal{N}ull_\rho(n-1)} \Pr_{\boldsymbol{x}}\left[\left(\sum_{2\leqslant i\leqslant n} a_i(A_{n-1})x_i = 0\right)\right]$$
$$\times \Pr_{M_{n-1}^1}\left[M_{n-1}^1 = A_{n-1}\right]$$
$$+ \Pr_{M_{n-1}^1}\left[\overline{\mathcal{N}ull_\rho(n-1)}\right]$$

$$\leqslant \rho + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{N}ull_\rho(n-1)}\right], \tag{8}$$

where the second line follows from equation (5); the third line is trivial; and the last line follows from the definition of $\mathcal{N}ull_\rho(n-1)$. Theorem 3.2 shows that 'typically', every nonzero integer null vector of M^1_{n-1} has 'small' atom probability, and will be used to bound the right-hand side of equation (8).

2.3. Bounding $\Pr\left[\mathcal{R}k^1_{n-1}(n) \cap \mathcal{ND}(n-1)\right]$.

Once again, we start with equation (3). However, for $M_{n-1} \in \mathcal{ND}(n-1)$, adj $\left(M^1_{n-1}\right)$ is invertible, and we no longer have the factorization of the determinant in Lemma 2.5 available to us. In this case, in order to reduce to a problem involving the anticoncentration of a linear form, we will follow an idea by Costello, Tao and Vu [2]. The basic tool is the following decoupling inequality from [2].

LEMMA 2.10 [2, Lemma 4.7]. *Let Y and Z be independent random variables, and $E = E(Y, Z)$ be an event depending on Y and Z. Then,*

$$\Pr[E(Y,Z)]^4 \leqslant \Pr[E(Y,Z) \cap E(Y',Z) \cap E(Y,Z') \cap E(Y',Z')],$$

where Y' and Z' denote independent copies of Y and Z, respectively.

Next, we explain how to use the above decoupling lemma for our purpose. For this discussion, recall equation (3). Fix a nontrivial partition $[n] = U_1 \sqcup U_2$. Let $Y := (x_i)_{i \in U_1}$ and $Z := (x_i)_{i \in U_2}$. Let $E_{\alpha,\mathbf{c}} := E_{\alpha,\mathbf{c}}(Y,Z)$ denote the event that

$$Q_{\alpha,\mathbf{c}}(Y,Z) := \alpha - \sum_{2 \leqslant i,j \leqslant n} c_{ij} x_i x_j = 0,$$

where α and $\mathbf{c} := (c_{ij})_{2 \leqslant i,j \leqslant n}$ are fixed. Then, the previous lemma shows that

$$\Pr\left[E_{\alpha,\mathbf{c}}(Y,Z)\right]^4 \leqslant \Pr\left[E_{\alpha,\mathbf{c}}(Y,Z) \cap E_{\alpha,\mathbf{c}}(Y',Z) \cap E_{\alpha,\mathbf{c}}(Y,Z') \cap E_{\alpha,\mathbf{c}}(Y',Z')\right].$$

On the other hand, whenever the event on the right holds, we also have

$$Q_{\alpha,\mathbf{c}}(Y,Z) - Q_{\alpha,\mathbf{c}}(Y',Z) - Q_{\alpha,\mathbf{c}}(Y,Z') + Q_{\alpha,\mathbf{c}}(Y,Z) = 0.$$

Direct computation shows that the left hand side equals

$$R_{\mathbf{c}} := \sum_{i \in U_1} \sum_{j \in U_2} c_{ij}(x_i - x_i')(x_j' - x_j) = \sum_{i \in U_1} R_i(x_i - x_i'),$$

where x_i' denotes an independent copy of x_i, and R_i denotes the random sum $\sum_{j \in U_2} c_{ij}(x_j' - x_j)$. To summarize, we have deduced the following.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

COROLLARY 2.11. *Let $U_1 \sqcup U_2$ be an arbitrary nontrivial partition of $[n]$. Let $\boldsymbol{w} = (w_1, \ldots, w_{|U_1|})$ be the random vector with coordinates $w_i := x_i - x_i'$. Then, with notation as above, and for any $(n-1)\times(n-1)$ symmetric matrix A_{n-1}, we have*

$$\Pr_{M_n}\left[\mathcal{R}k^1_{n-1}(n)\middle|M^1_{n-1} = A_{n-1}\right] \leqslant \Pr_{\boldsymbol{x},\boldsymbol{x}'}\left[\sum_{i\in U_1} R_i w_i = 0\middle|M^1_{n-1} = A_{n-1}\right]^{1/4}.$$

Using this corollary, we thus see that

$$\begin{aligned}
&\Pr_{M_n}\left[\mathcal{R}k^1_{n-1}(n)\cap\mathcal{ND}(n-1)\right]^4\\
&= \left(\sum_{A_{n-1}\in\mathcal{ND}(n-1)} \Pr_{M_n}\left[\mathcal{R}k^1_{n-1}(n)|M^1_{n-1} = A_{n-1}\right]\Pr\left[M^1_{n-1} = A_{n-1}\right]\right)^4\\
&\leqslant \sum_{A_{n-1}\in\mathcal{ND}(n-1)} \Pr_{M_n}\left[\mathcal{R}k^1_{n-1}(n)|M^1_{n-1} = A_{n-1}\right]^4\Pr\left[M^1_{n-1} = A_{n-1}\right]\\
&\leqslant \sum_{A_{n-1}\in\mathcal{ND}(n-1)} \Pr_{\boldsymbol{x},\boldsymbol{x}'}\left[\sum_{i\in U_1} R_i w_i = 0|M^1_{n-1} = A_{n-1}\right]\Pr\left[M^1_{n-1} = A_{n-1}\right]\\
&= \Pr_{\boldsymbol{x},\boldsymbol{x}',M^1_{n-1}}\left[\left(\sum_{i\in U_1} R_i w_i = 0\right)\cap\mathcal{ND}(n-1)\right],
\end{aligned} \tag{9}$$

where the second line follows from Jensen's inequality. Hence, we have reduced the problem of bounding $\Pr\left[\mathcal{R}k^1_{n-1}(n)\cap\mathcal{ND}(n-1)\right]$ to a linear anticoncentration problem.

In order to use equation (9) profitably, we will rely on the following simple, but crucial, observation about the vector $\boldsymbol{R} := (R_2, \ldots, R_n) \in \mathbb{Z}^{n-1}$, where R_i is defined as above.

LEMMA 2.12. *$\boldsymbol{R}$ is orthogonal to at least $n-1-|U_2|$ rows of M^1_{n-1}.*

Proof. Observe that $\boldsymbol{R}$ is a linear combination of the columns of $\operatorname{adj}\left(M^1_{n-1}\right)$ corresponding to the indices in U_2. By equation (6), each of these columns is orthogonal to each of the rows with indices in $[n-1]\cap U_1$; therefore, the same is true for $\boldsymbol{R}$. Since $|[n-1]\cap U_1| \geqslant n-1-|U_2|$, we are done. □

For $0 < \delta, \gamma \leqslant 1$, let $\mathcal{O}rth_{\delta,\gamma n}(n-1)$ denote the event—depending only on M^1_{n-1}—that *every* integer nonzero vector which is orthogonal to at least $(1-\gamma)n$ rows of M^1_{n-1} has μ-atom probability (in $\mathbb{Z}$) at most δ, uniformly for all $0 \leqslant \mu \leqslant 1/2$. Let $U_1 \sqcup U_2$ be a partition of $[n]$ where $U_2 := [\gamma n - 1]$. Then, with the vector

$\boldsymbol{R}$ defined as above, we have

$$\begin{aligned}
&\Pr_{\boldsymbol{x},\boldsymbol{x}',M^1_{n-1}}\left[\left(\sum_{i\in U_1} R_i w_i = 0\right)\cap \mathcal{ND}(n-1)\right]\\
&\leqslant \Pr_{\boldsymbol{x},\boldsymbol{x}',M^1_{n-1}}\left[\left(\sum_{i\in U_1} R_i w_i = 0\right)\cap \mathcal{O}rth_{\delta,\gamma n}(n-1)\cap \mathcal{ND}(n-1)\right]\\
&\quad + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{O}rth_{\delta,\gamma n}(n-1)}\right]\\
&\leqslant \sum_{A_{n-1}\in \mathcal{O}rth_{\delta,\gamma n}(n-1)\cap\mathcal{ND}(n-1)} \Pr_{\boldsymbol{w}}\left[\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right]\\
&\quad \times \Pr_{M^1_{n-1}}\left[M^1_{n-1} = A_{n-1}\right]\\
&\quad + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{O}rth_{\delta,\gamma n}(n-1)}\right]. \qquad (10)
\end{aligned}$$

As in Section 2.2, we will provide an upper bound on $\Pr_{\boldsymbol{w}}\left[\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right]$ which is uniform in the choice of $A_{n-1}\in \mathcal{O}rth_{\delta,\gamma n}(n-1)\cap\mathcal{ND}(n-1)$. We start by observing that

$$\begin{aligned}
&\Pr_{\boldsymbol{w}}\left[\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right]\\
&\leqslant \Pr_{\boldsymbol{w}}\left[\left(\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right)\cap\left(\boldsymbol{R}(A_{n-1})\neq \boldsymbol{0}\right)\right]\\
&\quad + \Pr_{\boldsymbol{w}}\left[\boldsymbol{R}(A_{n-1}) = \boldsymbol{0}\right]\\
&= \Pr_{\boldsymbol{w}}\left[\left(\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right)\cap\left(\boldsymbol{R}(A_{n-1})\neq \boldsymbol{0}\right)\right] + 2^{-|U_2|}\\
&\leqslant \Pr_{\boldsymbol{w}}\left[\left(\sum_{i\in U_1} R_i(A_{n-1}) w_i = 0\right)\cap\left(\boldsymbol{R}(A_{n-1})\neq \boldsymbol{0}\right)\right] + 2^{-\gamma n+1}. \qquad (11)
\end{aligned}$$

To see why the second equality holds, observe as before that

$$\boldsymbol{R}(A_{n-1}) := \sum_{j\in U_2} w_j \mathbf{col}_j\left(\operatorname{adj}\left(M^1_{n-1}\right)\right),$$

where $\mathbf{col}_j\left(\operatorname{adj}\left(M^1_{n-1}\right)\right)$ denotes the jth column of $\operatorname{adj}\left(M^1_{n-1}\right)$. Since $A_{n-1}\in \mathcal{ND}(n-1)$, it follows that these columns are linearly independent, and hence

$\boldsymbol{R}(A_{n-1}) = \boldsymbol{0}$ if and only if $w_j = 0$ for all $j \in |U_2|$, which happens precisely with probability $2^{-|U_2|}$.

It remains to bound the first summand in equation (11). For this, note that since $A_{n-1} \in \mathcal{O}rth_{\delta,\gamma n}(n-1)$ and $|U_2| = \gamma n - 1$, Lemma 2.12, together with $\boldsymbol{R}(A_{n-1}) \neq \boldsymbol{0}$, shows that $\rho^{\mathbb{Z}}_{1/2}(\boldsymbol{R}(A_{n-1})) \leqslant \delta$. Then, by Lemma 2.9, it follows that $\rho^{\mathbb{Z}}_{1/2}(\boldsymbol{R}(A_{n-1})|_{U_1}) \leqslant 2^{|U_2|}\delta \leqslant 2^{\gamma n}\delta$. Finally, combining this with equations (9) and (10), we have

$$\Pr_{M_n}\left[\mathcal{R}k^1_{n-1}(n) \cap \mathcal{ND}(n-1)\right] \leqslant \left(2^{\gamma n}\delta + 2^{-\gamma n+1} + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{O}rth_{\delta,\gamma n}(n-1)}\right]\right)^{1/4}. \tag{12}$$

3. The structural theorem

This section is devoted to the proof of our structural theorem, which is motivated by equations (8) and (12).

3.1. Statement and initial reductions.

In order to state the structural theorem, we need the following definition.

DEFINITION 3.1. For $0 \leqslant \alpha := \alpha(n)$, $\beta := \beta(n) \leqslant 1$, let $\mathcal{O}rth_{\alpha,\beta n}(n)$ denote the event that every integer nonzero vector which is orthogonal to at least $(1-\beta)n$ many rows of M_n has μ-atom probability (in $\mathbb{Z}$) at most α, uniformly for all $0 \leqslant \mu \leqslant 1/2$.

THEOREM 3.2. *Let* $\alpha(n) = 2^{-n^{1/4}\sqrt{\log n}/64}$, $\beta(n) = n^{-3/4}\sqrt{\log n}/128$, *and* $n \in \mathbb{N}$ *be sufficiently large. Then,*

$$\Pr_{M_n}\left[\overline{\mathcal{O}rth_{\alpha,\beta n}(n)}\right] \leqslant 2^{-n/32}.$$

Roughly, we will prove Theorem 3.2 by taking a union bound, over the choice of the nonzero integer vector with large μ-atom probability, of the probability that this vector is orthogonal to at least $(1-\beta)n$ many rows of M_n. However, there is an obstacle since, *a priori*, this union bound is over an infinite collection of vectors. In order to overcome this, we will work instead with the coordinate-wise residues of the vector modulo a suitably chosen prime $p(n)$.

In the next proposition, we make use of the event $\mathcal{O}rth^p_{\alpha,\beta n}(n)$, which is defined exactly as $\mathcal{O}rth_{\alpha,\beta n}(n)$, except that we work over $\mathbb{F}_p$ instead of the integers.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

PROPOSITION 3.3. *Let $\alpha(n) = 2^{-n^{1/4}\sqrt{\log n}/64}$ and $\beta(n) = n^{-3/4}\sqrt{\log n}/128$. Let $p(n) = 2^{n^{1/4}\sqrt{\log n}/32}$ be a prime, and let $n \in \mathbb{N}$ be sufficiently large. Then,*

$$\Pr_{M_n}\left[\overline{\mathcal{O}rth^{p}_{\alpha,\beta n}(n)}\right] \leqslant 2^{-n/32}.$$

Before proving Proposition 3.3, let us quickly show how to deduce Theorem 3.2 from it.

Proof of Theorem 3.2 given Proposition 3.3. It suffices to show that

$$\overline{\mathcal{O}rth_{\alpha,\beta n}(n)} \subseteq \overline{\mathcal{O}rth^{p}_{\alpha,\beta n}(n)}$$

for any prime p. To see this, suppose $M_n \in \overline{\mathcal{O}rth_{\alpha,\beta n}(n)}$. So, there exists an integer nonzero vector $\boldsymbol{a}$ which is orthogonal to at least $(1-\beta)n$ many rows of M_n and has μ-atom probability (in $\mathbb{Z}$) greater than α, for some $0 \leqslant \mu \leqslant 1/2$. Furthermore, by rescaling $\boldsymbol{a}$ if necessary, we may assume that $\gcd(a_1, \ldots, a_n) = 1$. Therefore, letting $\boldsymbol{a}_p$ be the image of $\boldsymbol{a}$ under the natural map from $\mathbb{Z}^n \to \mathbb{F}_p^n$, we see that $\boldsymbol{a}_p \in \mathbb{F}_p^n \setminus \{\boldsymbol{0}\}$ and is orthogonal (over $\mathbb{F}_p$) to (at least) the same $(1-\beta)n$ rows of M_n. Finally, $\rho_\mu^{\mathbb{F}_p}(\boldsymbol{a}_p) \geqslant \rho_\mu^{\mathbb{Z}}(\boldsymbol{a}) > \beta$, since for any $c \in \mathbb{Z}$, every solution $\boldsymbol{x} \in \{-1, 0, 1\}^n$ of $a_1x_1 + \cdots + a_nx_n = c$ over the integers is also a solution of the same equation in $\mathbb{F}_p$. Thus, the vector $\boldsymbol{a}_p$ witnesses that $M_n \in \overline{\mathcal{O}rth^{p}_{\alpha,\beta n}(n)}$. □

The next lemma is the first step towards the proof of Proposition 3.3 and motivates the subsequent discussion. In its statement, the support of a vector $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{F}_p^n$, denoted by $\mathrm{supp}(\boldsymbol{a})$, refers to the set of indices $i \in [n]$ such that $\boldsymbol{a}_i \neq 0 \mod p$.

LEMMA 3.4. *Let $1 \leqslant d \leqslant n$ be an integer, and let p be a prime. Let $\mathcal{S}pt^{p}_{\geqslant d,\beta n}(n)$ denote the event that every vector in $\mathbb{F}_p^n \setminus \{\boldsymbol{0}\}$ which is orthogonal (over $\mathbb{F}_p$) to at least $(1-\beta)n$ many rows of M_n has support of size at least d. Suppose further that $\beta \leqslant 1/2$, $d \leqslant n/2$, $p^{\beta n} \leqslant 2^{n/2}$, $p^d \leqslant 2^{n/8}$, $H(\beta) \leqslant 1/4$, and $H(d/n) \leqslant 1/16$ (where $H(x) := -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function for $x \in [0, 1]$). Then,*

$$\Pr_{M_n}\left[\overline{\mathcal{S}pt^{p}_{\geqslant d,\beta n}(n)}\right] \leqslant 2^{-n/16}.$$

The proof of this lemma will use the following simple, yet powerful, observation.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

OBSERVATION 3.5. *Let Σ be an $n \times n$ permutation matrix. Then, for a uniformly random $n \times n$ symmetric $\{\pm 1\}$-matrix M_n, the random matrix $\Sigma^{-1} M_n \Sigma$ is also a uniformly distributed $n \times n$ symmetric $\{\pm 1\}$-matrix.*

Proof. It is clear that $\Sigma^{-1} M_n \Sigma$ is an $n \times n$ $\{\pm 1\}$-matrix. That it is symmetric follows from $\Sigma^{-1} = \Sigma^T$ and $M_n^T = M_n$. Finally, $\Sigma^{-1} M_n \Sigma$ is uniformly distributed since conjugation by Σ is manifestly a bijection from the set of $n \times n$ $\{\pm 1\}$ symmetric matrices to itself. □

Proof of Lemma 3.4. Let d be as in the statement of the lemma, and for $1 \leqslant s \leqslant d$, let $\mathbf{Supp}_{=s}(n)$ denote the set of all vectors in $\mathbb{F}_p^n$ which have support of size exactly s. Observe that $|\mathbf{Supp}_{=s}(n)| \leqslant \binom{n}{s} p^s$. We will now bound the probability that any given $\boldsymbol{a} \in \mathbf{Supp}_{=s}(n)$ is orthogonal to at least $(1-\beta)n$ rows of a uniformly chosen M_n.

For this, let $\Sigma = \Sigma(\boldsymbol{a})$ denote a fixed, but otherwise arbitrary, permutation matrix for which $\Sigma \mathbb{1}_{\mathrm{supp}(\boldsymbol{a})} = \mathbb{1}_{[n-s+1,n]}$. In other words, Σ permutes the vector $\boldsymbol{a}$ so that its nonzero entries are placed in the last s coordinates. Since Observation 3.5 shows that $\Sigma^{-1} M_n \Sigma$ is a uniformly random $n \times n$ $\{\pm 1\}$-symmetric matrix, it follows that

$$\begin{aligned}
&\Pr_{M_n}[\boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } M_n] \\
&\quad = \Pr_{M_n}\left[\boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } \Sigma^{-1} M_n \Sigma\right] \\
&\quad = \Pr_{M_n}\left[\Sigma^{-1} M_n \Sigma \boldsymbol{a} = \boldsymbol{v} \text{ for some } \boldsymbol{v} \in \bigcup_{t=0}^{\beta n} \mathbf{Supp}_{=t}(n)\right] \\
&\quad \leqslant \sum_{t=0}^{\beta n} \Pr_{M_n}\left[\Sigma^{-1} M_n \Sigma \boldsymbol{a} = \boldsymbol{v} \text{ for some } \boldsymbol{v} \in \mathbf{Supp}_{=t}(n)\right] \\
&\quad = \sum_{t=0}^{\beta n} \Pr_{M_n}\left[M_n \Sigma \boldsymbol{a} = \boldsymbol{v} \text{ for some } \boldsymbol{v} \in \mathbf{Supp}_{=t}(n)\right] \\
&\quad \leqslant \sum_{t=0}^{\beta n} \sum_{\boldsymbol{v} \in \mathbf{Supp}_{=t}(n)} \Pr_{M_n}\left[M_n \Sigma \boldsymbol{a} = \boldsymbol{v}\right], \qquad (13)
\end{aligned}$$

where the third line follows by the union bound; the fourth line follows since the size of the support of a vector is invariant under the action of Σ; and the last line follows again by the union bound.

Next, we provide a (crude) upper bound on $\Pr_{M_n}\left[M_n(\Sigma \boldsymbol{a}) = \boldsymbol{v}\right]$ for any fixed $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{F}_p^n$. For this, we isolate the last column of the matrix M_n by

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

rewriting the system of equations $M_n(\Sigma \boldsymbol{a}) = \boldsymbol{v}$ as

$$m_{in} = (\Sigma \boldsymbol{a})_n^{-1}\left(v_i - \sum_{j=1}^{n-1} m_{ij}(\Sigma \boldsymbol{a})_j\right) \quad \text{for all } i \in [n], \tag{14}$$

where m_{ij} denotes the (i, j)th entry of the matrix M_n, and the equation makes sense since $(\Sigma \boldsymbol{a})_n \neq 0$ by our choice of Σ. Note that the right-hand side of the equation is completely determined by the top-left $(n-1)\times(n-1)$ submatrix of M_n. Further, the entries m_{in}, $i \in [n]$ are mutually independent even after conditioning on any realization of the top-left $(n-1) \times (n-1)$ submatrix of M_n. Since m_{in} takes on any value with probability at most $1/2$, it follows that conditioned on any realization of the top-left $(n-1) \times (n-1)$ submatrix of M_n, equation (14) is satisfied with probability at most $(1/2)^n$. Hence, by the law of total probability, $\Pr_{M_n}[M_n \Sigma \boldsymbol{a} = \boldsymbol{v}] \leqslant 2^{-n}$. Substituting this in equation (13), we see that

$$\begin{aligned}\Pr_{M_n}[\boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } M_n] &\leqslant 2^{-n}\sum_{t=0}^{\beta n} |\mathbf{Supp}_{=t}(n)| \\ &\leqslant 2^{-n}\sum_{t=0}^{\beta n}\binom{n}{t}p^t \\ &\leqslant 2^{-n}p^{\beta n}\sum_{t=0}^{\beta n}\binom{n}{t} \\ &\leqslant 2^{-n/2}2^{nH(\beta)} \leqslant 2^{-n/4}, \end{aligned} \tag{15}$$

where the fourth inequality follows by the assumption on $p^{\beta n}$ and the standard inequality $\sum_{t=0}^{\beta n}\binom{n}{t} \leqslant 2^{nH(\beta)}$ for $\beta \leqslant 1/2$, and the last inequality follows by the assumption on $nH(\beta)$. Finally, we have

$$\begin{aligned}\Pr_{M_n}\left[\overline{\mathcal{S}pt^p_{\geqslant d,\beta n}(n)}\right] &\leqslant \sum_{s=1}^{d}\sum_{\boldsymbol{a}\in \mathbf{Supp}_{=s}(n)} \Pr_{M_n}[\boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } M_n] \\ &\leqslant 2^{-n/4}\sum_{s=1}^{d}|\mathbf{Supp}_{=s}(n)| \leqslant 2^{-n/4}\sum_{s=1}^{d}\binom{n}{s}p^s \\ &\leqslant 2^{-n/4}p^d\sum_{s=1}^{d}\binom{n}{s} \leqslant 2^{-n/8}2^{nH(d/n)} \leqslant 2^{-n/16},\end{aligned}$$

where the fifth inequality follows by the assumption on p^d and d, and the last inequality follows by the assumption on $H(d/n)$. □

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

3.2. Tools and auxiliary results. Following Lemma 3.4, we will bound

$$\Pr_{M_n}\left[\overline{\mathcal{O}rth^{p}_{\alpha,\beta n}(n)} \cap \mathcal{S}pt^{p}_{\geqslant d,\beta n}(n)\right]$$

for suitably chosen parameters. Our proof of this bound will be based on the following two key ingredients. The first is a classical anticoncentration inequality due to Halász, which bounds the atom probability of a vector in terms of the 'arithmetic structure' of its coordinates. In order to state it, we need the following definition.

DEFINITION 3.6. Let $\boldsymbol{a} \in \mathbb{F}_p^n$ and let $k \in \mathbb{N}$. We define $R_k(\boldsymbol{a})$ to be the number of solutions to

$$\pm a_{i_1} \pm a_{i_2} \pm \cdots \pm a_{i_{2k}} = 0 \mod p,$$

where repetitions are allowed in the choice of $i_1, \ldots, i_{2k} \in [n]$.

THEOREM 3.7 (Halász, [4]). *Let p be any odd prime and let $\boldsymbol{a} := (a_1, \ldots, a_n) \in \mathbb{F}_p^n \setminus \{\boldsymbol{0}\}$. Then,*

$$\sup_{0\leqslant\mu\leqslant\frac{1}{2}} \max_{q\in\mathbb{F}_p} \Pr\left[\sum_i a_i x_i^{\mu} = q\right] \leqslant \frac{1}{p} + \frac{C R_k(\boldsymbol{a})}{2^{2k} n^{2k} f(|\mathrm{supp}(\boldsymbol{a})|)^{1/2}} + e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2},$$

where C is an absolute constant (which we may assume is at least 1), and $f(|\mathrm{supp}(\boldsymbol{a})|)$ is a positive real number which is at most $\min\{|\mathrm{supp}(\boldsymbol{a})|/100, n/k\}$.

Halász's inequality is typically stated and proved over the integers, but the version over $\mathbb{F}_p$ stated above easily follows using the same ideas. For the reader's convenience, we provide a complete proof in Appendix A.

The second ingredient is a 'counting lemma' due to the authors together with Luh and Samotij [3], which bounds the number of vectors in $\mathbb{F}_p^n$ with a slightly different (but practically equivalent) notion of 'rich additive structure'.

DEFINITION 3.8. Let $\boldsymbol{a} \in \mathbb{F}_p^n$ and let $k \in \mathbb{N}$. We define $R_k^*(\boldsymbol{a})$ to be the number of solutions to

$$\pm a_{i_1} \pm a_{i_2} \cdots \pm a_{i_{2k}} = 0 \mod p$$

that satisfy $|\{i_1, \ldots, i_{2k}\}| \geqslant 1.01k$.

As mentioned above, $R_k(\boldsymbol{a})$ and $R_k^*(\boldsymbol{a})$ are practically equivalent. This is made precise by the following lemma.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

LEMMA 3.9 [3, Lemma 1.6]. *For all positive integers k, n with $k \leqslant n/2$ and any vector $\boldsymbol{a} \in \mathbb{F}_p^n$,*

$$R_k(\boldsymbol{a}) \leqslant R_k^*(\boldsymbol{a}) + (40k^{0.99}n^{1.01})^k.$$

Proof. By definition, $R_k(\boldsymbol{a})$ is equal to $R_k^*(\boldsymbol{a})$ plus the number of solutions to $\pm a_{i_1} \pm a_{i_2} \pm \cdots \pm a_{i_{2k}} = 0$ that satisfy $|\{i_1, \ldots, i_{2k}\}| < 1.01k$. The latter quantity is bounded from above by the number of sequences $(i_1, \ldots, i_{2k}) \in [n]^{2k}$ with at most $1.01k$ distinct entries times 2^{2k}, the number of choices for the ± signs. Thus

$$R_k(\boldsymbol{a}) \leqslant R_k^*(\boldsymbol{a}) + \binom{n}{1.01k}(1.01k)^{2k}2^{2k} \leqslant R_k^*(\boldsymbol{a}) + \left(4e^{1.01}k^{0.99}n^{1.01}\right)^k,$$

where the final inequality follows from the well-known bound $\binom{a}{b} \leqslant (ea/b)^b$. Finally, noting that $4e^{1.01} \leqslant 40$ completes the proof. □

We can now state the 'counting lemma' from [3]. In the following statement, the notation $\boldsymbol{b} \subset \boldsymbol{a}$ for $\boldsymbol{a} \in \mathbb{F}_p^n$ means that $\boldsymbol{b}$ is a subvector of $\boldsymbol{a}$, that is, an element of $\bigcup_{s=1}^n \mathbb{F}_p^s$ formed by retaining some of the entries of $\boldsymbol{a}$; the dimension of $\boldsymbol{b}$ is denoted by $|\boldsymbol{b}|$.

THEOREM 3.10 [3, Theorem 1.7]. *Let p be a prime and let $k \in \mathbb{N}$, $s \in [n]$, $t \in [p]$. Let*

$$\mathbf{B}_{k,s,\geqslant t}(n) := \left\{\boldsymbol{a} \in \mathbb{F}_p^n \mid \forall \boldsymbol{b} \subset \boldsymbol{a} \text{ s.t. } |\boldsymbol{b}| \geqslant s \text{ we have } R_k^*(\boldsymbol{b}) \geqslant t \cdot \frac{2^{2k} \cdot |\boldsymbol{b}|^{2k}}{p}\right\}$$

denote the set of '$k, s, \geqslant t$-bad vectors'. Then,

$$|\mathbf{B}_{k,s,\geqslant t}(n)| \leqslant \left(\frac{s}{n}\right)^{2k-1} p^n(0.01t)^{-n+s}.$$

The above theorem shows that there are very few vectors for which every sufficiently large subset has rich additive structure. However, in order to use the strategy in the proof of Lemma 3.4 effectively, we require that there are very few vectors for which every *moderately sized* subset has rich additive structure (see the proof of Corollary 3.13). This is accomplished by the following corollary.

COROLLARY 3.11. *Let p be a prime and let $k, s_1, s_2, d \in [n]$, $t \in [p]$ such that $s_1 \leqslant s_2$. Let*

$$\mathbf{B}_{k,s_1,s_2,\geqslant t}^d(n)$$
$$:= \left\{\boldsymbol{a} \in \mathbb{F}_p^n \middle| |\mathrm{supp}(\boldsymbol{a})| = d \text{ and } \forall \boldsymbol{b} \subset \boldsymbol{a}|_{\mathrm{supp}(\boldsymbol{a})} \text{ s.t. } s_2 \geqslant |\boldsymbol{b}| \geqslant s_1 : R_k^*(\boldsymbol{b}) \geqslant t \cdot \frac{2^{2k} \cdot |\boldsymbol{b}|^{2k}}{p}\right\}.$$

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

Then,

$$|\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n)| \leqslant \binom{n}{d} p^{d+s_2}(0.01t)^{-d+(s_1/s_2)d}.$$

Proof. At the expense of an overall factor of $\binom{n}{d}$, we may restrict our attention to those vectors in $\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n)$ whose support is $[d]$. In order to count the number of such vectors, we begin by decomposing $[d]$ into the intervals $I_1, \ldots, I_{m+1}$, where $m := \lfloor d/s_2 \rfloor$, $I_j := \{(j-1)s_2+1, \ldots, js_2\}$ for $j \in [m]$, and $I_{m+1} := \{ms_2+1, \ldots, d\}$. For a vector with support $[d]$ to be in $\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n)$, it must necessarily be the case that the restriction of the vector to each of the intervals $I_1, \ldots, I_m$ is in $\mathbf{B}_{k,s_1,\geqslant t}(s_2)$. Since there are at most $p^{|I_{m+1}|} \leqslant p^{s_2}$ many choices for the restriction of the vector to I_{m+1}, it follows from Theorem 3.10 that

$$\begin{aligned}|\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n)| &\leqslant \binom{n}{d}|\mathbf{B}_{k,s_1,\geqslant t}(s_2)|^m p^{s_2} \leqslant \binom{n}{d}\left\{\left(\frac{s_1}{s_2}\right)^{2k-1} p^{s_2}(0.01t)^{-s_2+s_1}\right\}^m p^{s_2}\\ &\leqslant \binom{n}{d}(p^{s_2}(0.01t)^{-s_2+s_1})^{d/s_2}p^{s_2} = \binom{n}{d}p^{d+s_2}(0.01t)^{-d+(s_1/s_2)d}.\end{aligned}$$

□

We conclude this subsection with a few corollaries of Theorem 3.7 and Corollary 3.11. Let $\boldsymbol{a} \in \mathbf{Supp}_{=d}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)$ for $s_1 \leqslant d \leqslant n$. Then, by definition, there exists $\Lambda = \Lambda(\boldsymbol{a}) \subseteq \mathrm{supp}(\boldsymbol{a})$ such that $s_1 \leqslant |\Lambda| = |\mathrm{supp}(\boldsymbol{a}|_\Lambda)| \leqslant s_2$ and $R^*_k(\boldsymbol{a}|_\Lambda) < (t+1)\cdot 2^{2k}|\Lambda|^{2k}/p$. From now on, fix such a subset $\Lambda(\boldsymbol{a})$ for every such vector $\boldsymbol{a}$.

COROLLARY 3.12. *Let p be a prime and let $\boldsymbol{a} \in \mathbf{Supp}_{=d}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)$ for $1 \leqslant s_1 \leqslant d \leqslant n$. Suppose $p^{-1} \geqslant \max\left\{e^{-s_1/2k}, (50k/s_1)^{0.99k}\right\}$ and $t \geqslant s_1 \geqslant k \geqslant 100$. Then,*

$$\sup_{0\leqslant\mu\leqslant 1/2} \rho^{\mathbb{F}_p}_\mu(\boldsymbol{a}|_{\Lambda(\boldsymbol{a})}) \leqslant \frac{2Ct\sqrt{k}}{p\sqrt{s_1}},$$

where $C \geqslant 1$ is an absolute constant.

Proof. For convenience of notation, let $\boldsymbol{b} := \boldsymbol{a}|_{\Lambda(\boldsymbol{a})}$. By applying Theorem 3.7 to the vector $\boldsymbol{b}$ with $f(|\mathrm{supp}(\boldsymbol{b})|) := |\mathrm{supp}(\boldsymbol{b})|/k = |\boldsymbol{b}|/k =: f(|\boldsymbol{b}|)$ (which is a valid choice for f since $k \geqslant 100$ by assumption), we get

$$\begin{aligned}\sup_{0\leqslant\mu\leqslant\frac{1}{2}} \rho^{\mathbb{F}_p}_\mu(\boldsymbol{b}) &\leqslant \frac{1}{p} + \frac{C(R^*_k(\boldsymbol{b}) + (40k^{0.99}|\boldsymbol{b}|^{1.01})^k)}{2^{2k}|\boldsymbol{b}|^{2k}\sqrt{|\boldsymbol{b}|/k}} + e^{-|\boldsymbol{b}|/2k}\\ &\leqslant \frac{1}{p} + \frac{C(t+1)}{p\sqrt{|\boldsymbol{b}|/k}} + \frac{C(40k^{0.99})^k}{|\boldsymbol{b}|^{0.99k}\sqrt{|\boldsymbol{b}|/k}} + e^{-|\boldsymbol{b}|/2k}\end{aligned}$$

$$\begin{aligned}
&\leqslant \frac{1}{p} + \frac{C(t+1)\sqrt{k}}{p\sqrt{|\boldsymbol{b}|}} + \frac{C(40k^{0.99})^k}{|\boldsymbol{b}|^{0.99k}} + e^{-|\boldsymbol{b}|/2k}\\
&\leqslant \frac{1}{p} + \frac{C(t+1)\sqrt{k}}{p\sqrt{s_1}} + C\left(\frac{50k}{s_1}\right)^{0.99k} + e^{-s_1/2k}\\
&\leqslant \frac{(2+C)}{p} + \frac{C(t+1)\sqrt{k}}{p\sqrt{s_1}} \leqslant \frac{2Ct\sqrt{k}}{p\sqrt{s_1}},
\end{aligned}$$

where the first line follows from Theorem 3.7, Lemma 3.9, and the choice of $\Lambda(\boldsymbol{a})$, the fifth line follows by the assumption on p, and the last line follows since $t \geqslant s_1 \geqslant 100$. □

COROLLARY 3.13. *Let p be a prime and let $\boldsymbol{a} \in \mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)$. Suppose $p^{-1} \geqslant \max\left\{e^{-s_1/2k}, (50k/s_1)^{0.99k}\right\}$, $n \geqslant d \geqslant s_1$, and $t \geqslant s_1 \geqslant k \geqslant 100$. Then, for $0 \leqslant \beta := \beta(n) \leqslant 1/2$,*

$$\Pr_{M_n}[\boldsymbol{a} \textit{ is orthogonal to} \geqslant (1-\beta)n \textit{ rows of } M_n] \leqslant 2^{nH(\beta)} p^{\beta n} \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-s_2},$$

where $C \geqslant 1$ is an absolute constant.

Proof. The proof is very similar to the proof of Lemma 3.4. Let $\Lambda := \Lambda(\boldsymbol{a})$ and $\boldsymbol{b} := \boldsymbol{a}|_{\Lambda}$. As in the proof of Lemma 3.4, let Σ denote a fixed, but otherwise arbitrary, permutation matrix for which $\Sigma \mathbb{1}_{\Lambda} = \mathbb{1}_{[n-|\Lambda|+1,n]}$. Then, by equation (13),

$$\Pr_{M_n}[\boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } M_n] = \sum_{t=0}^{\beta n} \sum_{\boldsymbol{v} \in \mathbf{Supp}_{=t}} \Pr_{M_n}\left[M_n \Sigma \boldsymbol{a} = \boldsymbol{v}\right].$$

Next, we provide an upper bound on $\Pr_{M_n}\left[M_n(\Sigma \boldsymbol{a}) = \boldsymbol{v}\right]$ for any fixed $\boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{F}_p^n$. For this, note that the system of equations $M_n(\Sigma \boldsymbol{a}) = \boldsymbol{v}$ implies in particular that

$$\sum_{j=1}^{|\Lambda|} m_{i,n-|\Lambda|+j} b_j = v_i - \sum_{j=1}^{n-|\Lambda|} m_{i,j}(\Lambda \boldsymbol{a})_j \quad \text{for all } i \in [n-|\Lambda|]. \tag{16}$$

Note that the right-hand side is completely determined by the top-left $(n-|\Lambda|) \times (n-|\Lambda|)$ submatrix of M_n, and the entries of M_n appearing on the left are mutually independent even after conditioning on any realization of the top-left $(n-|\Lambda|) \times (n-|\Lambda|)$ submatrix of M_n. In particular, after conditioning on any realization

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

of the top-left submatrix of this size, each of the $n - |\Lambda|$ equations above is satisfied with probability which is at most $\rho^{\mathbb{F}_p}(\boldsymbol{b})$, and the satisfaction of different equations is mutually independent. Hence, by the law of total probability, the system equation (16) is satisfied with probability at most

$$\left(\rho^{\mathbb{F}_p}(\boldsymbol{b})\right)^{n-|\Lambda|} \leqslant \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-|\Lambda|} \leqslant \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-s_2},$$

where the middle bound follows from Corollary 3.12, and the right-hand bound follows since $|\Lambda| \leqslant s_2$. Finally, substituting this in equation (13) and proceeding as in equation (15) gives the desired conclusion. □

COROLLARY 3.14. *Let p be a prime and $k, s_1, s_2, d \in [n]$, $t \in [p]$ be such that $1 \leqslant s_1 \leqslant s_2 \leqslant n/2$, $s_1 \leqslant d \leqslant n$, $p^{-1} \geqslant \max\left\{e^{-s_1/2k}, (50k/s_1)^{0.99k}\right\}$, and $t \geqslant s_1 \geqslant k \geqslant 100$. Then, for $0 \leqslant \beta := \beta(n) \leqslant 1/2$,*

$$\Pr_{M_n}[\exists \boldsymbol{a} \in \mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n) : \boldsymbol{a} \text{ is orthogonal to} \geqslant (1-\beta)n \text{ rows of } M_n]$$
$$\leqslant (500C)^n p^{\beta n + 2s_2 + (s_1/s_2)d} \left(\frac{k}{s_1}\right)^{n/4},$$

where $C \geqslant 1$ is an absolute constant.

Proof. Using Corollary 3.13 to bound the probability that any given

$$\boldsymbol{a} \in \mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)$$

is orthogonal to at least $(1-\beta)n$ rows of M_n, and taking the union bound over all $|\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)|$ such vectors $\boldsymbol{a}$, we see that the desired probability is at most

$$|\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^d_{k,s_1,s_2,\geqslant(t+1)}(n)| \cdot 2^{nH(\beta)} p^{\beta n} \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-s_2}$$
$$\leqslant |\mathbf{B}^d_{k,s_1,s_2,\geqslant t}(n)| \cdot 2^n p^{\beta n} \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-s_2}$$
$$\leqslant 2^n \binom{n}{d} p^{d+s_2} (0.01t)^{-d+(s_1/s_2)d} p^{\beta n} \left(\frac{2Ct\sqrt{k}}{p\sqrt{s_1}}\right)^{n-s_2}$$
$$\leqslant (500C)^n p^{\beta n + s_2 + (s_1/s_2)d} \left(\frac{t}{p}\right)^{n-d-s_2} \left(\frac{k}{s_1}\right)^{n/4}$$
$$\leqslant (500C)^n p^{\beta n + 2s_2 + (s_1/s_2)d} \left(\frac{k}{s_1}\right)^{n/4},$$

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

where the second inequality follows from Corollary 3.11, and the third inequality follows from $s_2 \leqslant n/2$. □

3.3. Proof of Proposition 3.3.

By combining the results of the previous subsection, we can now prove Proposition 3.3.

Proof of Proposition 3.3. Consider the following choice of parameters: $k = n^{1/4}$, $s_1 = n^{1/2}\log n$, $s_2 = n^{3/4}\sqrt{\log n}$, $\beta n = n^{1/4}\sqrt{\log n}/128$, $d = n^{2/3}$, $\alpha = 2^{-n^{1/4}\sqrt{\log n}/64}$, and $p = 2^{n^{1/4}\sqrt{\log n}/32}$. Throughout, we will assume that n is sufficiently large for various inequalities to hold, even if we do not explicitly mention this.

Step 1: It is readily seen that the assumptions of Lemma 3.4 are satisfied, so that $\Pr\left[\mathcal{S}pt^{p}_{\geqslant d,\beta n}(n)\right] \leqslant 2^{-n/16}$. In other words, except with probability at most $2^{-n/16}$, every vector in $\mathbb{F}_p^n \setminus \{\mathbf{0}\}$ which is orthogonal to at least $(1-\beta)n$ rows of M_n has support of size at least $d = n^{2/3}$.

Step 2: Let $\boldsymbol{a} \in \mathbf{Supp}_{=s}(n) \setminus \mathbf{B}^{s}_{k,s_1,s_2,\geqslant\sqrt{p}}(n)$ for any $s \geqslant d$. Since the assumptions of Corollary 3.12 are satisfied for our choice of parameters, it follows from Corollary 3.12 and Lemma 2.8 that for any $0 \leqslant \mu \leqslant 1/2$,

$$\rho_{\mu}^{\mathbb{F}_p}(\boldsymbol{a}) \leqslant \rho_{\mu}^{\mathbb{F}_p}(\boldsymbol{a}|_{\Lambda(\boldsymbol{a})}) \leqslant \frac{2C\sqrt{k}}{\sqrt{ps_1}} \leqslant \alpha,$$

for all n sufficiently large.

Step 3: Therefore, it suffices to bound the probability that for some $s \geqslant d$, there exists some vector in $\mathbf{B}^{s}_{k,s_1,s_2,\geqslant\sqrt{p}}(n)$ which is orthogonal to at least $(1-\beta)n$ rows of M_n. By writing

$$\mathbf{B}^{s}_{k,s_1,s_2,\geqslant\sqrt{p}}(n) := \bigcup_{t=\sqrt{p}}^{p} \mathbf{B}^{s}_{k,s_1,s_2,\geqslant t}(n) \setminus \mathbf{B}^{s}_{k,s_1,s_2,\geqslant(t+1)}(n),$$

noting that the assumptions of Corollary 3.14 are satisfied, and taking the union bound over the choice of s and t, it follows that this event has probability at most

$$\begin{aligned} np(500C)^n p^{\beta n + 2s_2 + (s_1/s_2)s}\left(\frac{k}{s_1}\right)^{n/4} &\leqslant np(500C)^n p^{4s_2} 2^{-(n\log n)/16} \\ &\leqslant np(500C)^n 2^{-(n\log n)/32} \leqslant 2^{-(n\log n)/64}, \end{aligned}$$

for all n sufficiently large.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

Combining these steps, it follows that

$$\Pr_{M_n}\left[\overline{\mathcal{O}rth^{p}_{\alpha,\beta n}}\right] \leqslant 2^{-n/16} + 2^{-(n\log n)/64} \leqslant 2^{-n/32},$$

as desired. □

4. Proof of Theorem 1.2

Our main result is now immediate.

Proof of Theorem 1.2. By definition, $\overline{\mathcal{N}ull_{\rho}(n-1)} \subseteq \overline{\mathcal{O}rth_{\rho,\beta n}(n-1)}$ for every $\beta \geqslant 0$. Therefore, from equations (2), (8), and (12), it follows that

$$\begin{aligned}\Pr_{M_n}\left[\mathcal{R}k^{1}_{n-1}\right] \leqslant \alpha + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{O}rth_{\alpha,\beta n}(n-1)}\right] \\ + \left(2^{\beta n}\alpha + 2^{-\beta n+1} + \Pr_{M^1_{n-1}}\left[\overline{\mathcal{O}rth_{\rho,\beta n}(n-1)}\right]\right)^{1/4},\end{aligned}$$

where α and β are as in the statement of Theorem 3.2. From Theorem 3.2, it follows that the right-hand side of the above equation is at most $2^{-n^{1/4}\sqrt{\log n}/600}$ for all n sufficiently large. Finally, Lemma 2.1 and Corollary 2.4 give the desired conclusion. □

Acknowledgements

A.F. is partially supported by NSF 6935855. V.J. is partially supported by NSF CCF 1665252, NSF DMS-1737944 and ONR N00014-17-1-2598. V.J. would like to thank MIT Libraries for their financial support in making this article open access.

Appendix A. Proof of Halász's inequality over $\mathbb{F}_p$

In this appendix, we prove Theorem 3.7. The proof follows Halász's original proof in [4].

Proof of Theorem 3.7. Let e_p be the canonical generator of the Pontryagin dual of $\mathbb{F}_p$, that is, the function $e_p\colon \mathbb{F}_p \to \mathbb{C}$ defined by $e_p(x) = \exp(2\pi i x/p)$. Recall the following discrete Fourier identity in $\mathbb{F}_p$:

$$\delta_0(x) = \frac{1}{p}\sum_{r\in\mathbb{F}_p} e_p(rx),$$

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

where $\delta_0(0) = 1$ and $\delta_0(x) = 0$ if $x \neq 0$. Note that for any $q \in \mathbb{F}_p$,

$$\begin{aligned}\Pr_{x^\mu}\left[\sum_{i=1}^n a_i x_i^\mu = q\right] &= \mathbb{E}_{x^\mu}\left[\delta_0\left(\sum_{i=1}^n a_i x_i^\mu - q\right)\right] \\ &= \mathbb{E}_{x^\mu}\left[\frac{1}{p}\sum_{r\in\mathbb{F}_p} e_p\left(r\left(\sum_{j=1}^n a_j x_j^\mu - q\right)\right)\right] \\ &= \mathbb{E}_{x^\mu}\left[\frac{1}{p}\sum_{r\in\mathbb{F}_p}\prod_{j=1}^n e_p(ra_j x_j^\mu) e_p(-rq)\right] \\ &\leqslant \frac{1}{p}\sum_{r\in\mathbb{F}_p}\prod_{j=1}^n \left|\mu + (1-\mu)\cos\left(\frac{2\pi r a_j}{p}\right)\right| \\ &= \frac{1}{p}\sum_{r\in\mathbb{F}_p}\prod_{j=1}^n \left|\mu + (1-\mu)\cos\left(\frac{\pi r a_j}{p}\right)\right|,\end{aligned}$$

where the equality holds because the map $\mathbb{F}_p \ni r \mapsto 2r \in \mathbb{F}_p$ is a bijection (as p is odd) and (since $x \mapsto |\cos(\pi x)|$ has period 1 and it is therefore well defined for $x \in \mathbb{R}/\mathbb{Z}$) because $|\cos(2\pi x/p)| = |\cos(\pi(2x)/p)|$ for every $x \in \mathbb{F}_p$.

At this point, we record the useful inequality

$$\left|\mu + (1-\mu)\cos\left(\frac{\pi x}{p}\right)\right| \leqslant \exp\left(-\frac{1}{2}\left\|\frac{x}{p}\right\|^2\right),$$

which is valid for every real number x uniformly for all $0 \leqslant \mu \leqslant 1/2$, where $\|x\| := \|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance to the nearest integer. Thus, we arrive at

$$\max_{q\in\mathbb{F}_p}\Pr_{x^\mu}\left[\sum_{i=1}^n a_i x_i^\mu = q\right] \leqslant \frac{1}{p}\sum_{r\in\mathbb{F}_p}\exp\left(-\frac{1}{2}\sum_{j=1}^n \|ra_j/p\|^2\right). \tag{A.1}$$

Now, for each nonnegative real t, we define the following ‘level sets’

$$T_t := \left\{r \in \mathbb{F}_p : \sum_{j=1}^n \|ra_j/p\|^2 \leqslant t\right\},$$

and note that

$$\sum_{r\in\mathbb{F}_p}\exp\left(-\frac{1}{2}\sum_{j=1}^n \|ra_j/p\|^2\right) = \frac{1}{2}\int_0^\infty e^{-t/2}|T_t|\,dt. \tag{A.2}$$

We will now use a critical estimate due to Halász. First, note that for any $m \in \mathbb{N}$, the iterated sumset mT_t is contained in T_{m^2t}. Indeed, for $r_1, \ldots, r_m \in T_t$, we have

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

from the triangle inequality and the Cauchy–Schwarz inequality that

$$\sum_{j=1}^{n}\left\|\sum_{i=1}^{m} r_i a_j/p\right\|^2 \leqslant \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\|r_i a_j/p\|\right)^2 \leqslant \sum_{j=1}^{n} m \sum_{i=1}^{m}\|r_i a_j/p\|^2 \leqslant m^2 t.$$

Recall that the Cauchy–Davenport theorem states that every pair of nonempty A, $B \subseteq \mathbb{F}_p$ satisfies $|A+B| \geqslant \min\{p, |A|+|B|-1\}$. It follows that for every positive integer m and every $t \geqslant 0$, the iterated sumset mT_t satisfies $|mT_t| \geqslant \min\{p, m|T_t| - m\}$. Hence, $|T_{m^2 t}| \geqslant \min\{p, m|T_t| - m\}$.

Next, since the map $\mathbb{F}_p \ni r \mapsto ra \in \mathbb{F}_p$ is bijective for every nonzero $a \in \mathbb{F}_p$, we have that

$$\begin{aligned}\sum_{r\in\mathbb{F}_p}\sum_{j=1}^{n}\|ra_j/p\|^2 &\geqslant \sum_{j\in\operatorname{supp}(\boldsymbol{a})}\sum_{r\in\mathbb{F}_p}\|ra_j/p\|^2\\ &= |\operatorname{supp}(\boldsymbol{a})|\sum_{r\in\mathbb{F}_p}\|r/p\|^2\\ &= \frac{2|\operatorname{supp}(\boldsymbol{a})|}{p^2}\sum_{i=1}^{(p-1)/2} i^2\\ &\geqslant \frac{|\operatorname{supp}(\boldsymbol{a})|p}{50}.\end{aligned}$$

On the other hand, it follows from the definition of T_t that for every $t \geqslant 0$,

$$\sum_{r\in\mathbb{F}_p}\sum_{j=1}^{n}\|ra_j/p\|^2 \leqslant |T_t|\cdot t + \big(p - |T_t|\big)\cdot n.$$

In particular, we see that $|T_s| < p$ if $s \leqslant |\operatorname{supp}(\boldsymbol{a})|/100$. Therefore, if $t \leqslant f(|\operatorname{supp}(\boldsymbol{a})|)$ (as in the statement of the theorem), it follows by setting $m := \lfloor\sqrt{f(|\operatorname{supp}(\boldsymbol{a})|)/t}\rfloor \geqslant 1$ that $|T_{m^2 t}| < p$, and hence,

$$|T_t| \leqslant \frac{|T_{m^2 t}|}{m} + 1 \leqslant \frac{2\sqrt{t}|T_{f(|\operatorname{supp}(\boldsymbol{a})|)}|}{\sqrt{f(|\operatorname{supp}(\boldsymbol{a})|)}} + 1. \tag{A.3}$$

We now bound the size of $T_{f(|\operatorname{supp}(\boldsymbol{a})|)}$. Using the elementary inequality $1 - 100\|z\|^2 \leqslant \cos(2\pi z)$, which holds for all $z \in \mathbb{R}$, it follows that $|T_{f(|\operatorname{supp}(\boldsymbol{a})|)}| \leqslant |T'|$, where

$$T' := \left\{r \in \mathbb{F}_p : \sum_{j=1}^{n}\cos(2\pi r a_j/p) \geqslant n - 100 f(|\operatorname{supp}(\boldsymbol{a})|)\right\}.$$

In turn, we will bound the size of T' by computing the moments of the random variable (over the randomness of $r \in \mathbb{F}_p$) given by $\sum_{j=1}^{n}\cos(2\pi r a_j/p)$. More precisely, by Markov's inequality, we have for any $\ell \in \mathbb{N}$ that

$$|T'| \leqslant \frac{1}{(n - 100 f(|\mathrm{supp}(\boldsymbol{a})|))^{2\ell}} \sum_{r \in T'} \left| \sum_{j=1}^{n} \cos\left(\frac{2\pi r a_j}{p}\right) \right|^{2\ell}. \tag{A.4}$$

Moreover, we also have

$$\begin{aligned}
\sum_{r \in T'} \left| \sum_{j=1}^{n} \cos\left(\frac{2\pi r a_j}{p}\right) \right|^{2\ell} &\leqslant \frac{1}{2^{2\ell}} \sum_{r \in \mathbb{F}_p} \left| \sum_{j=1}^{n} (\exp(2i\pi r a_j/p) + \exp(-2i\pi r a_j/p)) \right|^{2\ell} \\
&= \frac{1}{2^{2\ell}} \sum_{\epsilon_1, \ldots, \epsilon_{2\ell}} \sum_{j_1, \ldots, j_{2\ell}} \sum_{r \in \mathbb{F}_p} \exp\left(2\pi i r \sum_{i=1}^{2\ell} \epsilon_i a_{j_i}\right) \\
&= \frac{1}{2^{2\ell}} \sum_{\epsilon_1, \ldots, \epsilon_{2\ell}} \sum_{j_1, \ldots, j_{2\ell}} p \mathbb{1}_{\sum_{i=1}^{2\ell} \epsilon_i a_{j_i} = 0} \\
&\leqslant \frac{p R_\ell(\boldsymbol{a})}{2^{2\ell}}.
\end{aligned}$$

Finally, combining this with equations (A.1)–(A.4), we get for any $0 \leqslant \mu \leqslant 1/2$ and $k \in \mathbb{N}$ as in the statement of the theorem that

$$\begin{aligned}
\max_{q \in \mathbb{F}_p} \Pr_{x^\mu} \left[\sum_{i=1}^{n} a_i x_i^\mu = q \right] &\leqslant \frac{1}{2p} \int_0^{f(|\mathrm{supp}(\boldsymbol{a})|)} e^{-t/2} |T_t| \, dt + \frac{1}{2} e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2} \\
&\leqslant \frac{1}{2p} \int_0^{f(|\mathrm{supp}(\boldsymbol{a})|)} e^{-t/2} \left(\frac{2\sqrt{t}|T'|}{\sqrt{f(|\mathrm{supp}(\boldsymbol{a})|)}} + 1 \right) dt \\
&\quad + \frac{1}{2} e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2} \\
&\leqslant \frac{|T'|}{p\sqrt{f(|\mathrm{supp}(\boldsymbol{a})|)}} \int_0^{f(|\mathrm{supp}(\boldsymbol{a})|)} e^{-t/2} \sqrt{t} \, dt + \frac{1}{p} \\
&\quad + \frac{1}{2} e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2} \\
&\leqslant \frac{C_1 |T'|}{p\sqrt{f(|\mathrm{supp}(\boldsymbol{a})|)}} + \frac{1}{p} + e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2} \\
&\leqslant \frac{1}{p} + \frac{C_1 R_k(\boldsymbol{a})}{2^{2k}(n - 100 f(|\mathrm{supp}(\boldsymbol{a})|))^{2k} \sqrt{f(|\mathrm{supp}(\boldsymbol{a})|)}} \\
&\quad + e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2} \\
&\leqslant \frac{1}{p} + \frac{C R_k(\boldsymbol{a})}{2^{2k} n^{2k} \sqrt{f(|\mathrm{supp}(\boldsymbol{a})|)}} + e^{-f(|\mathrm{supp}(\boldsymbol{a})|)/2},
\end{aligned}$$

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press

as desired, where the last inequality uses the assumption that $f(|\text{supp}(\boldsymbol{a})|) \leqslant n/k$. $\square$

References

[1] J. Bourgain, V. H. Vu and P. M. Wood, 'On the singularity probability of discrete random matrices', *J. Funct. Anal.* **258**(2) (2010), 559–603.

[2] K. P. Costello, T. Tao and V. H. Vu, 'Random symmetric matrices are almost surely nonsingular', *Duke Math. J.* **135**(2) (2006), 395–413.

[3] A. Ferber, V. Jain, K. Luh and W. Samotij, 'On the counting problem in inverse Littlewood–Offord theory', Preprint, 2019, arXiv:1904.10425.

[4] G. Halász, 'Estimates for the concentration function of combinatorial number theory and probability', *Period. Math. Hungar.* **8**(3-4) (1977), 197–211.

[5] V. Jain, 'Approximate Spielman–Teng theorems for random matrices with heavy tailed entries: a combinatorial view', Preprint, 2019, arXiv:1904.11108.

[6] V. Jain, 'Approximate Spielman–Teng theorems for the least singular value of random combinatorial matrices', Preprint, 2019, arXiv:1904.10592.

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

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

[9] H. H. Nguyen, 'Inverse Littlewood–Offord problems and the singularity of random symmetric matrices', *Duke Math. J.* **161**(4) (2012), 545–586.

[10] H. H. Nguyen and V. H. Vu, 'Optimal inverse Littlewood–Offord theorems', *Adv. Math.* **226**(6) (2011), 5298–5319.

[11] H. H. Nguyen and V. H. Vu, 'Small ball probability, inverse theorems, and applications', in *Erdős Centennial* (Springer, Berlin, Heidelberg, 2013), 409–463.

[12] A. M. Odlyzko, 'On subspaces spanned by random selections of {±}-1 vectors', *J. Combin. Theory Ser.* A **47**(1) (1988), 124–133.

[13] M. Rudelson and R. Vershynin, 'The Littlewood–Offord problem and invertibility of random matrices', *Adv. Math.* **218**(2) (2008), 600–633.

[14] M. Rudelson and R. Vershynin, 'Non-asymptotic theory of random matrices: extreme singular values', in *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes)*, Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures (Hindustan Book Agency, New Delhi, 2010), 1576–1602.

[15] T. Tao and V. H. Vu, 'On the singularity probability of random Bernoulli matrices', *J. Amer. Math. Soc.* **20**(3) (2007), 603–628.

[16] T. Tao and V. H. Vu, 'John-type theorems for generalized arithmetic progressions and iterated sumsets', *Adv. Math.* **219**(2) (2008), 428–449.

[17] K. Tikhomirov, 'Singularity of random Bernoulli matrices', Preprint, 2018, arXiv:1812.09016.

[18] R. Vershynin, 'Invertibility of symmetric random matrices', *Random Structures Algorithms* **44**(2) (2014), 135–182.

[19] V. H. Vu, 'Random discrete matrices', in *Horizons of Combinatorics* (Springer, Berlin, Heidelberg, 2008), 257–280.

https://doi.org/10.1017/fms.2019.21 Published online by Cambridge University Press