SURJECTIVITY OF NEAR SQUARE RANDOM MATRICES

HOI H. NGUYEN AND ELLIOT PAQUETTE

Abstract. We show that a nearly square iid random integral matrix is surjective over the integral lattice with very high probability. This answers a question by Koplewitz [7]. Our result extends to sparse matrices as well as to matrices of dependent entries.

1. Introduction

In this note we study random rectangular matrices $M_{n\times(n+u)} = (M_{ij})$ of size $n \times (n+u)$, where $n \to \infty$ and $u \geq 0$, and the entries M_{ij} are i.i.d copies of a random variable ξ taking integral values and such that for any prime p

$$\max_{x\in\mathbf{Z}/p\mathbf{Z}} \mathbf{P}(\xi = x) \leq 1 - \alpha_n, \tag{1}$$

where $\alpha_n > 0$ is a parameter allowed to depend on n. Such distributions are called α_n-*balanced*.

For random square matrices of random Bernoulli entries taking values 0 and 1 with probability 1/2, the problem to estimate the probability p_n of $M_{n\times n}$ being singular has attracted quite a lot of attention. In the early 60's Komlós [5] showed $p_n = O(n^{-1/2})$. This bound was significantly improved by Kahn, Komlós, and Szemerédi in the 90's to $p_n \leq 0.999^n$. About ten years ago, Tao and Vu [14] improved the bound to $p_n \leq (3/4 + o(1))^n$. We also refer the reader to [6] by Rudelson and Vershynin for implicit bounds of type e^{-cn}. The most recent record is due to and Bourgain, Vu and Wood [2], who show:

Theorem 1.1.

$$p_n \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n.$$

These results imply that with very high probability the linear map $M_{n\times n}$ is injective over $\mathbf{Z}^n$ (and hence the lattice $M_{n\times n}(\mathbf{Z}^n)$ has full rank in $\mathbf{Z}^n$.) Another fundamental question of interest is the surjectivity onto $\mathbf{Z}^n$, more specifically:

Is it true that with high probability $M_{n\times n}$ is also surjective over $\mathbf{Z}^n$ (in other words, the quotient group $\mathbf{Z}^n/M_{n\times n}(\mathbf{Z}^n)$ is trivial)?

Unfortunately, the answer to this question turns out to be negative: with high probability M is never surjective over $\mathbf{Z}^n$. To explain this at the heuristic level, assume that the vector $\mathbf{e}_1 = (1, 0, \dots, 0)$ is in the image space $M_{n\times n}(\mathbf{Z}^n)$, then (assuming that $M_{n\times n}$ is non-singular)

$$\mathbf{x} = M_{n\times n}^{-1}(\mathbf{e}_1) = ((M_{n\times n}^{-1})_{11}, \dots, (M_{n\times n}^{-1})_{1n})^T \in \mathbf{Z}^n.$$

However, we have $(M_{n\times n}^{-1})_{1i} = \frac{\det(M^{1i})}{\det(M_{n\times n})}$, where M^{1i} is the matrix obtained from $M_{n\times n}$ by removing the first row and the i-th column. By the co-factor expansion

$$\det(M_{n\times n}) = \sum_{i=1}^{n} (-1)^{i-1} M_{1i} \det(M^{1i}). \tag{2}$$

But as the M_{1i} are independent from the submatrices $M^{1i}, 1 \leq i \leq n$, it is highly unlikely that the random sum $|\sum_{i=1}^{n}(-1)^{i-1}M_{1i}\det(M^{1i})|$ becomes smaller than all $|\det(M^{1i})|$ so that the components $\frac{\det(M^{1i})}{\det(M_{n\times n})}$ of $\mathbf{x}$ are all integral.

Having seen that $M_{n\times n} : \mathbf{Z}^n \to \mathbf{Z}^n$ is unlikely to be surjective, it is natural to think of rectangular matrices $M_{n\times(n+u)} : \mathbf{Z}^{n+u} \to \mathbf{Z}^n$ which might have better chance to be surjective. In fact, in the past several years there have been exciting developments (see for instance [8, 10, 17, 18]) in the study of $M_{n\times(n+u)}(\mathbf{Z}^{n+u})$ for

The first author is supported by research grant DMS-1600782.

various ensembles of $M_{n\times(n+u)}$. For instance a special version of a recent result by Wood [18, Corollary 3.4] shows:

Theorem 1.2. *Let $u \geq 0$ be a fixed integer. Let $M_{n\times(n+u)}$ be a random matrix with entries being iid copies of an α_n-balanced random variable of fixed $\alpha_n > 0$. Let P be a finite set of primes, then*

$$\lim_{n\to\infty} \mathbf{P}\Big(\mathbf{Cok}(M_{n\times(n+u)})_P \simeq \{id\}\Big) = \prod_{p\in P}\prod_{k=1}^{\infty}(1-p^{-k-u}),$$

where $G_P = \prod_{p\in P} G_p$ is the product of p-Sylow subgroups of G and the cokernel $\mathbf{Cok}(M_{n\times m})$ is the quotient group $\mathbf{Z}^n/M_{n\times m}(\mathbf{Z}^m)$.

We remark that P is fixed and $n \to \infty$ in this result. However as P increases, the probability on the right hand side of the limit becomes arbitrarily small. Hence it follows that

$$\limsup_{n\to\infty} \mathbf{P}\Big(\mathbf{Cok}(M_{n\times n}) \simeq \{id\}\Big) \leq \inf_P \lim_{n\to\infty} \mathbf{P}\Big(\mathbf{Cok}(M_{n\times n})_P \simeq \{id\}\Big) = 0$$

which officially answers our question above.
In the opposite direction, it has been conjectured by Koplewitz [7, 8] that

Conjecture 1.3. *Let the matrix entries be iid copies of an α_n-balanced random variable of fixed $\alpha_n > 0$. Then for any fixed constant $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\Big(\mathbf{Cok}(M_{n\times\lfloor(1+\varepsilon)n\rfloor}) \simeq \{id\}\Big) = 1.$$

Also, with $u \to \infty$ together with n

$$\lim_{n\to\infty} \mathbf{P}\Big(\mathbf{Cok}(M_{n\times(n+u)}) \simeq \{id\}\Big) = 1.$$

To support these conjectures, Koplewitz himself showed in [7, Theorem 1] (see also [8, Theorem 30]) that $\mathbf{P}\Big(M_{n\times\lfloor(2+\varepsilon)n\rfloor} \simeq \{id\}\Big) \geq 1 - e^{-c_\varepsilon n}$. In the same paper he also confirmed Conjecture 1.3 for random matrices of entries distributed according to the Haar measure over the profinite completion $\widehat{\mathbf{Z}}$ of $\mathbf{Z}$.
In this note we confirm the first conjecture. In fact we are able to extend the result to very sparse matrices. More specifically, we can assume ξ to take integer values as in (1) with

$$\alpha_n \geq \frac{C_0 \log n}{n} \tag{3}$$

for a sufficiently large constant C_0.

Theorem 1.4 (Main result)**.** *Let ξ be as in* (3)*. Assume furthermore that ξ is bounded with probability one. Then for every $A, \varepsilon_0 > 0$, there exist $B = B(A, C_0, \varepsilon_0)$ and an absolute constant c such that*

$$\mathbf{P}\Big(\mathbf{Cok}(M_{n\times(n+\lfloor B(\frac{\log n}{\alpha_n}\log(\frac{\log n}{\alpha_n})+\log n)\rfloor)}) \simeq \{id\}\Big) \geq 1 - O(n^{-A} + e^{-c\alpha_n n}).$$

In particular, if α_n is fixed then

$$\mathbf{P}\Big(\mathbf{Cok}(M_{n\times\lfloor n+\log^{1+o(1)} n\rfloor}) \simeq \{id\}\Big) \geq 1 - O(n^{-\omega(1)}); \tag{4}$$

as well as if $\alpha_n \geq \frac{\log^{O(1)} n}{n}$ then

$$\mathbf{P}\Big(\mathbf{Cok}(M_{n\times\lfloor(1+o(1))n\rfloor}) \simeq \{id\}\Big) \geq 1 - O(n^{-\omega(1)}). \tag{5}$$

Note that a balanced assumption on ξ is necessary as the results no longer hold for instance if we work with the Bernoulli ± 1 ensemble; in this case the matrix cannot be surjective modulo 2 for even n. Note also by considering the $\{0, 1\}$ ensemble with $\alpha_n = \mathbf{P}(\xi = 1)$, we can see that roughly the stated number of additional columns is necessary up to multiplicative constants, just by considering rows that are identically 0.
We will also discuss an extension to a family of matrices of dependent entries, see Section 3. Our method is short and direct. We will first prove a slightly weaker version (Theorem 2.5) by relying on a totally elementary lemma by Odlyzko (Lemma 2.3). We then refine the method by using a more involved result by Maples from [9] (Theorem 2.9). However, as [9] appears to be slightly incomplete and contains several

(minor) errors, we will take this opportunity to recast Maples' proof toward our sparsest settings. Along the way, we show that this approach also yields a completely new singularity bound for sparse integral matrices.

Theorem 1.5. *There exists an absolute constant $c > 0$ such that as long as the entries of $M_{n\times n}$ are iid copies of ξ distributed as in* (3) *(which is not necessarily bounded) then*

$$p_n \leq e^{-c\alpha_n n}.$$

We notice that the recent paper [1] by Basak and Rudelson addressed the singularity (and in more general the least singular value) for a general family of sparse matrices. Unfortunately, Theorem 1.5 does not seem to follow from [1] because we have no restriction on the spectral norm of $M_{n\times n}$.

2. Proof of Theorem 1.4

We assume that

$$\mathbf{P}(|\xi| \leq K_0) = 1,$$

for some positive constant K_0. This assumption is only for Theorem 1.4, but not for Theorem 1.5.
A natural approach is to show that the equation system

$$M_{n\times m}\mathbf{x} = \mathbf{e}_i,$$

has solutions $\mathbf{x} \in \mathbf{Z}^m$, for any standard unit vector $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0), 1 \leq i \leq n$. However such an approach does not look simple as we would have to prove cancellation of extremely large numbers involving determinants of minors (see also the discussion around (2)). Instead, we will prove surjectivity by reducing our matrices over finite fields via the following result.

Lemma 2.1. [7, Lemma 5] *Let $m \geq n$. A matrix $M_{n\times m} : \mathbf{Z}^m \to \mathbf{Z}^n$ is surjective if and only if the modulo matrix $M_{n\times m}/p : \mathbf{F}_p^m \to \mathbf{F}_p^n$ is surjective for every prime p. Here $M_{n\times m}/p$ is the matrix over $\mathbf{F}_p$ given by $(M_{n\times m}/p)_{ij} = M_{ij}(\mod p)$.*

Proof. (of Lemma 2.1) Assume that $M_{n\times m}/p : \mathbf{F}_p^m \to \mathbf{F}_p^n$ is surjective, then $M_{n\times m}/p$ contains a submatrix $M'_{n\times n} = M'_{n\times n}(p)$ (depending on p) of size $n\times n$ such that $\det(M'_{n\times n}/p) \neq 0(\mod p)$. Thus $\det(M'_{n\times n}) \neq 0$. This implies that the columns of $M'_{n\times n}$ generate $\mathbf{Q}^n$, and hence $M'_{n\times n}(\mathbf{Z}^n)$ is a full-rank integer lattice. In particular, the lattice co–volume $d = |\mathbf{Z}^n/M_{n\times m}(\mathbf{Z}^m)|$ (which is independent of p) is finite and d divides $\det(M'_{n\times n})$ for all p. Now assume that $d \neq 1$. Then as $\det(M'_{n\times n}) \neq 0(\mod p)$, d is not divisible by p. But this holds for all prime p, a contradiction. □

By this lemma, for our problem we need to show that $M_{n\times m}/p : \mathbf{F}_p^m \to \mathbf{F}_p^n$ is surjective (or equivalently, $M_{n\times m}/p$ has rank n in $\mathbf{F}_p^n$) for *every* prime p. This does not seem to be an easier task, but in what follows we show that there is a way to restrict the treatment to a set of a only a few primes.
Our first ingredient is the following simple bound (see also [**?**, Lemma 3.9]).

Lemma 2.2 (quadratic estimate)**.** *Let p be a prime. Let $0 < \varepsilon_n < 1$ be a given parameter that might depend on n. Let $M_{n\times n}$ be a matrix of size $n \times n$ whose entries are iid copies of a random variable ξ from* (3)*. Then the probability that $M_{n\times n}$ has rank at most $(1-\varepsilon_n)n$ in $\mathbf{F}_p^n$ is smaller than $e^{-\alpha_n\varepsilon_n^2 n^2+n}$.*

To prove this result we rely on a useful result by Odlyzko [4].

Lemma 2.3. *Let H be a subspace of dimension $1 \leq d \leq n$ in $\mathbf{F}_p^n$. Then if X is a random vector whose entries are iid copies of a random variable ξ from* (3)*, then*

$$\mathbf{P}(X \in H) \leq (1-\alpha_n)^{n-d}.$$

We insert a proof of this well-known result here for completion.

Proof of Lemma 2.3. Let $\{H_1, H_2, \dots, H_d\}$ be a basis for H. By permuting coordinates, we may assume without loss of generality that the restrictions $\tilde{H}_1, \tilde{H}_2, \dots, \tilde{H}_d$ of these vectors to the first d coordinates

are again linearly independent. Consider the event $X = (\xi_1, \dots, \xi_d, \xi_{d+1}, \dots, \xi_n)^t \in H$. From the linear independence of $\left\{\tilde{H}_i\right\}_1^d$, there are unique $c_1, c_2, \dots, c_d \in \mathbf{F}_p$ so that

$$(\xi_1, \dots, \xi_d)^t = \sum_{i=1}^{d} c_i \tilde{H}_i.$$

Hence conditioning on $(\xi_1, \dots, \xi_d)$, if $X = (\xi_1, \dots, \xi_d, \xi_{d+1}, \dots, \xi_n)^t \in H$ then

$$X = \sum_{i=1}^{d} c_i H_i.$$

In particular, each value $\xi_{d+1}, \xi_{d+2}, \dots, \xi_n$ is determined. However the probability of each of these events is at most $1 - \alpha_n$, and so by independence the event $X \in H$ holds with probability at most $(1 - \alpha_n)^{n-d}$. □

Now we turn to the quadratic estimate.

Proof of Lemma 2.2. Let $d = \lfloor (1 - \varepsilon_n)n \rfloor$. Assume that the columns $X_{i_1}, \dots, X_{i_d}$ span the column space of $M_{n\times n}$. For now assume that $\{i_1, \dots, i_d\} = \{1, \dots, d\}$. Let H be the subspace spanned by $X_1, \dots, X_d$. We are considering the event $\mathcal{E}_{1,\dots,d}$ that $X_i \in H, d+1 \le i \le n$. By Lemma 2.3, for any $i \ge d+1$,

$$\mathbf{P}(X_i \in H) \le (1 - \alpha_n)^{n-d} \le e^{-\alpha_n \varepsilon_n n}.$$

Applying this bound for $d+1 \le i \le n$ and using independence we obtain

$$\mathbf{P}_{X_i, d+1 \le i \le n}\Big(\mathcal{E}_{1,\dots,d} | X_1, \dots, X_d\Big) = \mathbf{P}_{X_i, d+1 \le i \le n}\Big(X_{d+1}, \dots, X_n \in H | H\Big) \le e^{-\alpha_n \varepsilon_n^2 n^2}.$$

Taking the union bound over at most 2^n choices of $\{i_1, \dots, i_d\}$ we conclude the proof. □

2.4. **A simpler result.** To get the main idea, in this subsection we show:

Theorem 2.5. *For every $A > 0$, there exists sufficiently large B such that*

$$\mathbf{P}\Big(\mathbf{Cok}(M_{n\times(n+u)}) \simeq \{id\}\Big) \ge 1 - n^{-A},$$

where

$$u = \left\lfloor B \frac{\log^2 n}{\alpha_N} + \sqrt{\frac{n \log n}{\alpha_n}} \right\rfloor.$$

Note that u does not drop below $\sqrt{n \log n}$.
In what follows we prove Theorem 2.5. The same argument will also be used to deal with matrices of dependent entries. Let $\mathcal{P}_n$ be the set of primes up to $(K_0 n)^{n/2}$

$$\mathcal{P}_n := \Big\{p \text{ prime }, p \le (K_0 n)^{n/2}\Big\}. \tag{6}$$

By taking the union bound, Lemma 2.2 then implies:

Corollary 2.6. *Let $\mathcal{E}$ be the event that the matrix $M_{n\times n}$ has rank at least $(1 - \varepsilon_n)n$ in $\mathbf{F}_p^n$ for all $p \in \mathcal{P}_n$. Then*

$$\mathbf{P}(\mathcal{E}) \ge 1 - e^{-\alpha_n \varepsilon_n^2 n^2 + n \log n + n + n \log K_0/2}.$$

Set

$$\varepsilon_n := \sqrt{\frac{3 \log n}{\alpha_n n}}.$$

With this value of $\varepsilon_n < 1$, Corollary 2.6 implies

$$\mathbf{P}(\mathcal{E}) \ge 1 - e^{-n \log n}.$$

Lemma 2.7 (surjectivity for small primes)**.** *For any A there is a B sufficiently large so that with probability at least $1 - n^{-A}$, the random matrix $M_{n\times(n+u)}$ is surjective over $\mathbf{F}_p^n$ for all $p \in \mathcal{P}_n$ simultaneously, and $u = u(B)$ is as in Theorem 2.5.*

Proof. (of Lemma 2.7) It suffices to show that with high probability $M_{n\times(n+u)}$ has full rank in $\mathbf{F}_p^n$ (which would then imply surjectivity in $\mathbf{F}_p^n$).
We consider the submatrix $M_{n\times n}$, the restriction of $M_{n\times(n+u)}$ to the first n columns. Let $\mathcal{E}$ be the event defined in Corollary 2.6, i.e. that $M_{n\times n}$ has rank at least $(1-\varepsilon_n)n$ over $\mathbf{F}_p^n$ for all $p\in\mathcal{P}_n$. We thus have

$$\mathbf{P}_{M_{n\times n}}(\mathcal{E})\geq 1-e^{-n\log n}.$$

Consider also the event $\mathcal{E}_{\neq 0}$ that $\det(M_{n\times n})\neq 0$, where by Theorem 1.5

$$\mathbf{P}(\mathcal{E}_{\neq 0})\geq 1-n^{-A},$$

provided that $\frac{C_0\log n}{n}\leq\alpha$ for large C_0.
Now we condition on $M_{n\times n}$ satisfying $\mathcal{E}$ and $\mathcal{E}_{\neq 0}$ and show that with high probability (with respect to the last u columns) that $M_{n\times(n+u)}$ is surjective over all $\mathbf{F}_p^n, p\in\mathcal{P}_n$.
Let $\mathcal{P}^*=\mathcal{P}^*(M_{n\times n})$ be the collection of prime divisors of $\det(M_{n\times n})$. Because $|\det(M_{n\times n})|\leq(K_0n)^{n/2}$ by the Hadamard bound, the random set $\mathcal{P}^*$ has small size, say

$$|\mathcal{P}^*|\leq n^2.$$

Case 1. When $p\in\mathcal{P}_n$ but $p\notin\mathcal{P}^*$, then $M_{n\times n}$ has full rank in $\mathbf{F}_p^n$, and so does $M_{n\times(n+u)}$.
Case 2. Consider $p\in\mathcal{P}^*$, we estimate the probability of the event $\mathcal{E}_p$ that $M_{n\times(n+u)}/p$ has full rank.
Let $H_0\subset\mathbf{F}_p^n$ be the column subspace of $M_{n\times n}$, for which by assumption

$$d_0:=n-\dim(H_0)\leq n-(1-\varepsilon_n)n\leq\sqrt{\frac{3n\log n}{\alpha_n}}.$$

We next expose the remaning u vectors in groups. For $1\leq i\leq d_0$, at step i we will add k_i column vectors $X_{n+\sum_{l=1}^{i-1}k_l+j}, 1\leq j\leq k_i$ to the set of already exposed column vectors $X_1,\dots,X_{n+\sum_{l=1}^{i-1}k_l}$, where

$$k_i:=\left\lceil\frac{B\log n}{\alpha d_{i-1}}\right\rceil,$$

and where d_{i-1} is the codimension of the subspace H_{i-1} generated by $\langle X_1,\dots,X_{n+\sum_{l=1}^{i-1}k_l}\rangle$. Notice that in this exposing process the choice of k_i depends on d_{i-1}, a decreasing sequence throughout the process.
Next let $\mathcal{F}_i$ be the event that $\dim(H_i)\geq\dim(H_{i-1})+1$. In other words, $\mathcal{F}_i$ is the event that after adding the vectors of group i we have a strict decrease in the co-rank,

$$d_i\leq d_{i-1}-1.$$

Assuming that $\dim(H_{i-1})<n$, then by Lemma 2.3, and by independence of the column vectors,

$$\begin{aligned}\mathbf{P}\Big(\mathcal{F}_i|\wedge_{j=0}^{i-1}\mathcal{F}_j\wedge\mathcal{E}\wedge\mathcal{E}_{\neq 0},\dim(H_{i-1})<n\Big)&\geq 1-((1-\alpha_n)^{\operatorname{codim}(H_{i-1})})^{k_i}\\&\geq 1-((1-\alpha_n)^{d_{i-1}})^{k_i}\\&\geq 1-n^{-B}.\end{aligned}$$

By Bayes' rule, with probability at least $(1-n^{-B})^{d_0}\geq 1-n^{-B+1}$, after adding $\sum_i k_i\leq\sum_i\frac{B\log n}{\alpha_n d_i}+1\leq\frac{B\log n}{\alpha_n}\log d_0+d_0$ columns, the matrix $M_{n\times(n+\frac{B\log n}{\alpha_n}\log d_0+d_0)}$ has full rank in $\mathbf{F}_p^n$.
Taking union bound over all primes $p\in\mathcal{P}^*$, we obtain that with probability at least $1-n^{-B+3}$ the obtained matrix has full rank in $\mathbf{F}_p^n$ for all $p\in\mathcal{P}^*$.
By **Case 1.** and **Case 2.**, we have seen that with $M_{n\times n}$ satisfying $\mathcal{E}$ and $\mathcal{E}_{\neq 0}$, the matrix $M_{n\times(n+u)}$ is surjective simultaneously over $\mathbf{F}_p^n$ for all $p\in\mathcal{P}_n$ with the desired probability. The proof is then complete after unfolding the conditioning on $M_{n\times n}$ (using Corollary 2.6). □

Proof. (of Theorem 2.5) We condition on the event $\mathcal{E}_{\neq 0}$. Note that with probability one $|\det(M_{n\times n})|\leq(K_0n)^{n/2}$. This shows that with prime $p>(K_0n)^{n/2}$, $\det(M_{n\times n})\neq 0(\mod p)$. Hence on $\mathcal{E}_{\neq 0}$ the matrix $M_{n\times n}$ is surjective over $\mathbf{F}_p^n$ for all $p\geq(K_0n)^{n/2}$.
Furthermore, Lemma 2.7 implies that with probability at least $1-n^{-A}$, for all $p\in\mathcal{P}_n$ the random matrix $M_{n\times(n+u)}$ is surjective over $\mathbf{F}_p^n$. Hence altogether our matrix $M_{n\times(n+u)}$ is surjective in $\mathbf{Z}^n$ by Lemma 2.1. □

2.8. **Proof of Theorem 1.4.** Now we turn to our main theorem, where the proof is similar but instead of Lemma 2.3 we will be using the following result by Maples (see either [9, Theorem 1.2] or [10, Corollary 1.3].)

Theorem 2.9. *Let p be any prime. Assume that the entries of $M_{n\times n}$ are iid copies of ξ from* (1) *with α from* (3)*. Then for all $k \leq \eta n$ with a sufficiently small absolute constant η we have*

$$\mathbf{P}(\mathrm{rank}(M_{n\times n}/p) = n-k) = O\Big(n^k(p^{-k^2} + e^{-c\alpha n})\Big). \tag{7}$$

In fact [10, Corollary 1.3] says much more, that the bound is precisely

$$p^{-k^2}\prod_{i=1}^{k}(1-p^{-i})^{-1}\prod_{i=k+1}^{\infty}(1-p^{-i}) + O(e^{-c\alpha n}). \tag{8}$$

However, we will not need this later result (given that it has not been formally verified, especially for the sparse case). Note that (8), in its limit form ($n \to \infty$), is a simple consequence of the aforementioned paper [18, Corollary 3.5] by Wood. Back to [9], as this paper has some mistakes (for instance [9, Proposition 2.3] is incorrect, see the appendix for further discussion), for transparency we will recast an almost complete proof of Theorem 2.9 in the appendix. Theorem 2.9 and Theorem 1.5 will then follow as a byproduct.
Let $\mathcal{P}'_n$ be the set of primes up to $e^{c\alpha_n n/2}$, where c is the sufficiently small constant from Theorem 2.9, i.e.

$$\mathcal{P}'_n := \Big\{p \text{ prime }, p \leq e^{c\alpha_n n/2}\Big\}.$$

Note that $\mathcal{P}'_n \subset \mathcal{P}_n$ (defined in (6)). By applying (7) with $k = C_1 \log n$ for a sufficiently large constant C_1 to each $p \in \mathcal{P}'_n$ and taking the union bound

$$\sum_{p\in\mathcal{P}'_n} n^k(p^{-k^2} + O(e^{-c\alpha_n n})) = n^{-\omega(1)}.$$

Corollary 2.10. *Let $\mathcal{E}'$ be the event that the matrix $M_{n\times n}$ has rank at least $n - C_1 \log n$ in $\mathbf{F}_p^n$ for all $p \in \mathcal{P}'_n$. Then*

$$\mathbf{P}(\mathcal{E}') \geq 1 - n^{-\omega(1)}.$$

We next prove an analog of Lemma 2.7. Set

$$u = \left\lfloor B \cdot \left(\frac{\log n}{\alpha_n}\log\frac{\log n}{\alpha_n} + \log n\right)\right\rfloor,$$

for a sufficiently large constant B.

Lemma 2.11. *With probability at least $1 - n^{-A}$, the random matrix $M_{n\times(n+u)}$ is surjective over $\mathbf{F}_p^n$ for all $p \in \mathcal{P}_n$ simultaneously.*

Proof of Lemma 2.11. Again, if suffices to show that with high probability$M_{n\times(n+u)}/p$ has full rank in each $\mathbf{F}_p^n$.
We consider the submatrix $M_{n\times n}$, the restriction of $M_{n\times(n+u)}$ to the first n columns. Let $\mathcal{E}'$ be the event implied by Corollary 2.10 that this matrix $M_{n\times n}$ has rank at least $n - C_1 \log n$ over $\mathbf{F}_p^n$ for all $p \in \mathcal{P}'_n$. We thus have

$$\mathbf{P}_{M_{n\times n}}(\mathcal{E}') \geq 1 - n^{-\omega(1)}.$$

Consider also the event $\mathcal{E}_{\neq 0}$ that $\det(M_{n\times n}) \neq 0$ from Theorem 1.5. Conditioning on $M_{n\times n}$ satisfying $\mathcal{E}'$ and $\mathcal{E}_{\neq 0}$, we will show that with high probability (with respect to the last u columns) that $M_{n\times(n+u)}$ is surjective over all $\mathbf{F}_p^n, p \in \mathcal{P}_n$.
To do this, similarly to the proof of Lemma 2.7, let $\mathcal{P}^* = \mathcal{P}^*(M_{n\times n})$ be the collection of prime divisors of $\det(M_{n\times n})$, then clearly the random set $\mathcal{P}^*$ has size at most n^2.
Case 1. When $p \in \mathcal{P}_n$ but $p \notin \mathcal{P}^*$, then $M_{n\times n}$ has full rank in $\mathbf{F}_p^n$, and so does $M_{n\times(n+u)}$.
Case 2. Consider $p \in \mathcal{P}^*$, we estimate the probability of the event $\mathcal{E}_p$ that $M_{n\times(n+u)}$ has full rank over $\mathbf{F}_p^n$. For this, first note that under $\mathcal{E}'$, if $p \in \mathcal{P}'_n$ (that is $p \leq e^{c\alpha_n n/2}$) then the corank of $M_{n\times n}$ over $\mathbf{F}_p^n$ is at most $C_1 \log n$. Now if $e^{c\alpha_n n/2} < p \leq (K_0 n)^{n/2}$, as $|\det(M_{n\times n})| \leq (K_0 n)^{n/2}$, the corank of $M_{n\times n}$ over $\mathbf{F}_p^n$

for these large p must be at most $\frac{\log(K_0 n)^{n/2}}{\log p} \leq (c\alpha_n)^{-1} \log(K_0 n)$. So in either case the corank is at most $(c\alpha_n)^{-1} \log n + C_1 \log n$.

Let $H_0 \subset \mathbf{F}_p^n$ be the column subspace of $M_{n\times n}$, for which by assumption

$$d_0 := n - \dim(H_0) \leq (c\alpha_n)^{-1} \log n + C_1 \log n.$$

Similarly to the proof of Theorem 2.5, for $1 \leq i \leq d_0$, we will add $k_i = \lceil \frac{B \log n}{\alpha_n d_{i-1}} \rceil$ column vectors $X_{n+\sum_{l=1}^{i-1} k_l + j}, 1 \leq j \leq k_i$ to the set of already exposed column vectors $X_1, \dots, X_{n+\sum_{l=1}^{i-1} k_l}$, where d_{i-1} is the codimension of the subspace H_{i-1} generated by $\langle X_1, \dots, X_{n+\sum_{l=1}^{i-1} k_l} \rangle$.

Let $\mathcal{F}_i$ be the event that $\dim(H_i) \geq \dim(H_{i-1}) + 1$. By Lemma 2.3, and by independence of the column vectors,

$$\mathbf{P}\Big(\mathcal{F}_i | \wedge_{j=0}^{i-1} \mathcal{F}_j \wedge \mathcal{E}' \wedge \mathcal{E}_{\neq 0}, \dim(H_{i-1}) < n\Big) \geq 1 - ((1-\alpha)^{\mathrm{codim}(H_{i-1})})^{k_i} \geq 1 - n^{-B}.$$

By Bayes' rule, with probability at least $(1-n^{-B})^{d_0} \geq 1 - n^{-B+1}$, after adding $\sum_i k_i = O(\frac{\log n}{\alpha_n} \log d_0 + d_0) := u$ columns, the matrix $M_{n\times(n+u)}$ has full rank in $\mathbf{F}_p^n$. (It is possible to improve the total number of extra columns by a more careful analysis of the d_i but we will not do so here for simplicity.)

Taking the union bound over all primes $p \in \mathcal{P}^*$, we obtain that with probability at least $1 - n^{-B+3}$ the matrix $M_{n\times(n+u)}$ has full rank in $\mathbf{F}_p^n$ for all $p \in \mathcal{P}^*$.

We have seen that with $M_{n\times n}$ satisfying $\mathcal{E}$ and $\mathcal{E}_{\neq 0}$, the matrix $M_{n\times(n+u)}$ is surjective simultaneously over $\mathbf{F}_p^n$ for all $p \in \mathcal{P}_n$ with the desired probability. The proof is then complete after unfolding the conditioning on $M_{n\times n}$, knowing that these events hold with very high probability. □

Finally, for Theorem 1.4, conditioning on the event $\mathcal{E}_{\neq 0}$, with prime $p > (K_0 n)^{n/2}$ we have $\det(M_{n\times n}) \neq 0(\mod p)$, and hence on $\mathcal{E}_{\neq 0}$ the matrix $M_{n\times n}$ is surjective over $\mathbf{F}_p^n$ for all $p \geq (K_0 n)^{n/2}$.

On the other hand, Lemma 2.11 implies that with probability at least $1 - n^{-A}$, for all $p \in \mathcal{P}_n$ the random matrix $M_{n\times(n+u)}$ is surjective over $\mathbf{F}_p^n$.

3. Some remarks

We have studied random matrices of independent entries. It is natural to consider Conjecture 1.3 for other families of matrices of dependent entries. Here we discuss one such model.

Let $M_{n\times n}$ be a random symmetric matrix, where for simplicity we assume that the entries $(M_{n\times n})_{ij}, 1 \leq i \leq j \leq n$ are iid copies of a bounded random variable ξ from (1) with fixed α. It follows from [12, 16] that for this model the singularity probability can be bounded by

$$p_n = n^{-\omega(1)}. \tag{9}$$

Heuristically, arguing similarly to (2) (where we expose both columns and rows at the same time to obtain a quadratic variant of (2)), we can show that with high probability the matrix $M_{n\times n}$ is not surjective over $\mathbf{Z}^n$. Actually an analog of Theorem 1.2 has been established in [10] for this model [1], which confirms the above heuristic. However, we will show that by adding a couple of few more (say) independent rows, the matrix becomes surjective.

Theorem 3.1. *Let $M_{n\times(n+u)}$ be a random matrix where its restriction $M_{n\times n}$ to the first n columns is a symmetric matrix as above, and the last u columns are independent with entries being iid copies of ξ. Then for any $A > 0$, there exists B such that for $u = \lfloor B\sqrt{n \log n} \rfloor$*

$$\mathbf{P}\Big(\mathbf{Cok}(M_{n\times(n+u)}) \simeq \{id\}\Big) \geq 1 - n^{-A}.$$

To justify this result, we establish the following analog of Lemma 2.2.

Lemma 3.2 (quadratic estimate)**.** *Let p be a prime. Let $0 < \varepsilon_n < 1$ be a given parameter that might depend on n. Let $M_{n\times n}$ be a symmetric matrix where $(M_{n\times n})_{ij}, 1 \leq i \leq j \leq n$ are iid copies of a bounded random variable ξ from (1) with fixed α_n. Then the probability that $M_{n\times n}$ has rank at most $(1-\varepsilon_n)n$ in $\mathbf{F}_p^n$ is smaller than $e^{-\alpha_n \varepsilon_n^2 n^2/2+n}$.*

[1] To be more precise, M. M. Wood studied the Laplacian, but her result also covers the non-normalized ensemble.

Proof of Lemma 3.2. Let $d = \lfloor (1-\varepsilon_n)n \rfloor$. Assume that the columns $X_{i_1}, \dots, X_{i_d}$ spans the column space of $M_{n\times n}$. For now assume that $\{i_1, \dots, i_d\} = \{1, \dots, d\}$. Let H be the span of $X_1, \dots, X_d$. We are considering the event $\mathcal{E}_{1,\dots,d}$ that $X_i \in H, d+1 \le i \le n$. Now as X_i is dependent on H, we cannot estimate the probability of $X_i \in H$ directly by Odlyzko's bound. However, we can get rid of the dependence by deleting the corresponding common entries as below.
For $1 \le j \le n-d$ set

$$I_{d+j} := \{1, \dots, d+j\}.$$

For any $X \in \mathbf{F}_p^n$ and $J \subset [n]$ we denote $X|_J$ by the restriction of X over the components indexed by J. For convenience we also denote $H|_J$ by the subspace generated by $X_1|_J, \dots, X_d|_J$. Assume that $X_{d+1}, \dots, X_n \in H$, then the following holds

- $X_{d+2}|_{I_{d+1}} \in H_{I_{d+1}}$, and more generally $X_{d+j+1}|_{I_{d+j}} \in H|_{I_{d+j}}, 1 \le j \le n-d-1$;
- the vector $X_{d+j+1}|_{I_{d+j}}$ is independent of $H|_{I_{d+j}}$;
- the vectors $X_{d+j+1}|_{I_{d+j}}, 1 \le j \le n-d-1$ are mutually independent.

Now as $H|_{I_{d+j}}$ has rank at most d in $\mathbf{F}_p^{d+j}$, by Lemma 2.3 we have

$$\mathbf{P}(X_{d+j+1}|_{I_{d+j}} \in H|_{I_{d+j}}) \le (1-\alpha_n)^j.$$

Applying this bound for $1 \le j \le n-d-1$ and using the independence of $X_{d+j+1}|_{I_{d+j}}$, we obtain

$$\mathbf{P}_{X_i, d+1\le i\le n}\Big(\mathcal{E}_{1,\dots,d}|X_1, \dots, X_d\Big) \le \prod_{j=1}^{n-d-1} (1-\alpha_n)^j \le e^{-\alpha_n \varepsilon_n^2 n^2/2}.$$

Taking union bound over at most 2^n choices of $\{i_1, \dots, i_d\}$ we conclude the proof. □

We can now complete the proof of Theorem 3.1 verbatim as in the proof of Theorem 2.5 with fixed $\alpha_n L$. Indeed, Corollary 2.6 follows from Lemma 3.2, and Lemma 2.7 can be justified similarly (conditioning on (9)) because the last u columns are mutually independent, and are independent from $M_{n\times n}$.

Acknowledgement. The authors thank Kyle Luh for helpful comments.

Appendix A. The corank estimate: proof proof of Theorem 2.9

We will work in a more general setting. Let $q = p^f$ be a prime power and $\mathbf{F}_q$ be the finite field with q elements. We say that a probability distribution μ in $\mathbf{F}_q$ is α_n-*balanced* (for some $0 < \alpha_n < 1$) if for every additive subgroup T in $\mathbf{F}_q$ and $s \in \mathbf{F}_q$

$$\mu(s+T) \le 1-\alpha_n.$$

In the general finite field setting, we will assume

$$\alpha_n \ge n^{-1/2+\varepsilon} \text{ for any } \varepsilon > 0\ . \tag{10}$$

In the more specific setting when $q = p$ (which is the setting of Theorem 2.9), as there is no non-trivial additive subgroup in $\mathbf{F}_p$, we will assume

$$\max_{x\in\mathbf{F}_p} \mu(x) = 1-\alpha_n$$

where

$$\alpha_n \ge \frac{C_0 \log n}{n}, \text{ for a sufficiently large constant } C_0. \tag{11}$$

In what follows $M_{n\times n}$ is a random matrix where the entries are independent and identically distributed according to an α_n-balanced μ either from (10) or (11), and $n \to \infty$. Notice that in either case, we do not assume the support of μ to be bounded. Recall that $X_1, \dots, X_n$ are the columns of $M_{n\times n}$ and W_{n-k} is the subspace $\langle X_1, \dots, X_{n-k}\rangle$ generated by $X_1, \dots, X_{n-k}$. Our first goal is to reprove the following variant of [9, Proposition 2.1] and [11, Proposition 2.1.1].

Theorem A.1. *Assume that μ is distributed according to either* (10) *or* (11) *depending on q. Then there exist positive constants c, η such that the following holds for $1 \leq k \leq \eta n$: there exists an event $\mathcal{E}_{n-k}$ on the σ-algebra generated by $X_1, \dots, X_{n-k}$ of probability at least $1 - e^{-c\alpha_n n}$ such that for any $k \leq k_0 \leq \eta n$*

$$\mathbf{P}_{X_{n-k+1}}\Big(X_{n-k+1} \in W_{n-k} \big| \mathcal{E}_{n-k} \wedge \mathrm{codim}(W_{n-k}) = k_0\Big) = q^{-k_0} + O(e^{-c\alpha_n n}).$$

Notice that there are some modifications of this result compared to the original statement by Maples in [9, Proposition 2.1] or [11, Proposition 2.1.1] that

(i) the statement also holds when the codimension of W_{n-k} is not necessarily k;

(ii) the statement also holds for sparse settings such as (10) and (11).

Note that (i) is not new as it also appeared in a subsequent (unpublished) preprint by Maples ([10, Proposition 3.1]). We then deduce Theorem 2.9 restated here for finite field.

Corollary A.2. *Assume that μ is distributed according to either* (10) *or* (11)*. Assume that $k \leq \eta n$ for some sufficiently small η, then*

$$\mathbf{P}\Big(\mathrm{rank}(M_{n\times n}) = n - k\Big) = O\Big(n^k (q^{-k^2} + e^{-c\alpha n})\Big). \tag{12}$$

It seems plausible to get rid of the factor n^k here (especially for fixed α) but we do not attempt to do so, as the reader can check that any improvement along this line has little affect on the bounds in Theorem 1.4.

Proof. (of Corollary A.2) The event $\mathrm{rank}(M_{n\times n}) = n-k$ implies that there exist k column vectors $X_{i_1}, \dots, X_{i_k}$ which belong to the subspace of dimension $n-k$ generated by the remaining column vectors $X_i, i \neq i_1, \dots, i_k$. With a loss of a factor of n^k in probability, we can assume that $\{i_1, \dots, i_k\} = \{n-k+1, \dots, n\}$. We then use Theorem A.1 to show

$$\begin{aligned}
&\mathbf{P}\Big(X_{n-k+1}, \dots, X_n \in W_{n-k} \wedge \mathrm{codim}(W_{n-k}) = k\Big)\\
=&\mathbf{P}\Big(X_{n-k+1}, \dots, X_n \in W_{n-k} \wedge \mathcal{E}_{n-k} \wedge \mathrm{codim}(W_{n-k}) = k\Big) + O(e^{-c\alpha_n n})\\
\leq&\mathbf{P}\Big(X_{n-k+1}, \dots, X_n \in W_{n-k} | \mathcal{E}_{n-k} \wedge \mathrm{codim}(W_{n-k}) = k\Big) + O(e^{-c\alpha_n n})\\
\leq&\Big(q^{-k} + O(e^{-c\alpha_n n})\Big)^k + O(e^{-c\alpha_n n}) = O(q^{-k^2} + e^{-c\alpha_n n}).
\end{aligned}$$

□

Taking $k = 1$ and $q = p \to \infty$ in Corollary A.2 we then obtain Theorem 1.5.

Corollary A.3. *Assume that the entries of $M_{n\times n}$ are iid copies of a discrete random variable ξ taking integer values such that*

$$\max_{x \in \mathbf{Z}} \mathbf{P}(\xi = x) \leq 1 - \frac{C_0 \log n}{n}, \text{ for a sufficiently large constant } C_0.$$

Then the matrix $M_{n\times n}$ is non-singular with probability at least $1 - e^{-c\alpha_n n}$.

Proof. (of Corollary A.3) Choose a prime p to be large such that $p \geq \max\{n^{2n}, |\det(M_{n\times n})|\}$ and $\mathbf{supp}(\xi) \subset (-p, p)$. It then suffices to show that $M_{n\times n}/p$ has full rank with probability at least $1 - e^{-c\alpha_n n}$. To this end, and by Corollary A.6 (to be discussed in the sequel), it suffices to bound the probability that $M_{n\times n}/p$ has corank k between 1 and ηn, but then the statement follows from corollary A.2 (stated for μ distributed according to (11)) by taking union bound. □

Note that the trick to pass the singularity problem over $\mathbf{Z}$ to over $\mathbf{F}_p$, and let $p \to \infty$, is not new. See for instance [2, 14] and the references therein.

Finally, we will also show the following more general variant of Theorem A.1 for rectangular matrices $M_{n\times(n+u)}$ where $W_k = \langle X_{k+1}, \dots, X_{n+u}\rangle$.

Theorem A.4. *Assume that μ is distributed according to either* (10) *or* (11) *depending on q. Then there exist positive constants c, η such that the following holds. Let $0 \le u \le \eta n$ be given. Then for $k \le \eta n$, there exist an event $\mathcal{E}_{n+u-k}$ on the σ-algebra generated by $X_1, \dots, X_{n+u-k}$ of probability at least $1 - e^{-c\alpha_n n}$ such that for any $(k-u)^+ \le k_0 \le \eta n$*

$$\mathbf{P}_{X_{n+u-k+1}}\Big(X_{n+u-k+1} \in W_{n+u-k} \big| \mathcal{E}_{n+u-k} \wedge \operatorname{codim}(W_{n+u-k}) = k_0\Big) = q^{-k_0} + O(e^{-c\alpha_n n}).$$

The rest of the appendix is mainly dedicated to verify Theorem A.1. The proof of Theorem A.4 will be deduced shortly. As already mentioned, our approach mainly follows [9].

Part I: proof of Theorem A.1: In what follows the constants η, β, δ, d are sufficiently small but fixed (see for instance (25) for a choice of d), and α_n is allowed to depend on n as from (10) or (11) for sufficiently large constant C_0. The only place we have to treat (10) and (11) separately is in the proof of Proposition A.8 below.
We first note that Odlyzko's lemma in fact holds in any finite field.

Lemma A.5. *For a deterministic subspace V of $\mathbf{F}_q^n$ and a random vector X of iid entries from an α_n-balanced distribution*

$$\mathbf{P}(X \in V) \le (1 - \alpha_n)^{\operatorname{codim}(V)}.$$

Corollary A.6. *Let $X_1, \dots, X_{n-k}$ be the columns of $M_{n\times n}$. Then the probability that $X_1, \dots, X_{n-k}$ are linearly independent in $\mathbf{F}_q^n$ is at least $1 - n(1-\alpha_n)^k$.*

Proof. (of Corollary A.6) Let $0 \le i \le n-k-1$ be smallest such that $X_{i+1} \in \operatorname{Span}(X_1, \dots, X_i)$. By Lemma 2.3, this event is bounded by $(1-\alpha_n)^{n-i}$. Summing over $0 \le i \le n-k-1$, the probability under consideration is bounded by $n(1-\alpha_n)^k$. □

A.7. **Sparse subspace.** Let $0 < \delta, \eta$ be small constants (independently from α_n). Given a vector space $H \subset \mathbf{F}_q^n$, we say that H is *δ-sparse* if there is a non-zero vector w with $|\mathbf{supp}(w)| \le \delta n$ (i.e. w is *δ-sparse*) such that $w \perp H$, where $\mathbf{supp}(w)$ is the set of non-zero coordinates of w.

Proposition A.8 (random subspaces are not sparse)**.** *Let $0 < \varepsilon_0 < 1/2$ be any fixed constants. Then for any $0 \le \delta, \eta$ such that $\delta + \eta \le \varepsilon_0$, and with α_n from* (10) *or* (11) *the following holds for $0 \le k < \eta n$: with probability at least $1 - e^{-c\alpha_n n}$ with respect to $X_1, \dots, X_{n-k}$ the random subspace W_{n-k} is not δ-sparse. Here $c = c(\varepsilon_0)$ is an absolute constant.*

Proof. (of Proposition A.8) For $\sigma \subset [n]$ with $1 \le t = |\sigma| \le \delta n$, let $\mathcal{E}_\sigma$ be the event that W_{n-k} is orthogonal to a vector w with $\mathbf{supp}(w) = \sigma$, but is not orthogonal to any vector w' with $|\mathbf{supp}(w')| \le t-1$. Note that in this case the σ-restricted vector $w|_\sigma$ is orthogonal to the σ-restricted column vectors $X_1|_\sigma, \dots, X_{n-k}|_\sigma$. The dimension of the annihilator of W_{n-k} in $\mathbf{F}_q^\sigma$ and the dimension of $\operatorname{Span}(X_1|_\sigma, \dots, X_{n-k}|_\sigma)$ sum to t. If the annihilator were more than 1 dimensional, there would necessarily exist a nonzero linear combination of annihilators with support strictly contained in σ. Hence it follows that the column vectors $X_1|_\sigma, \dots, X_{n-k}|_\sigma$ span a subspace of dimension $t-1$, and there are $t-1$ linearly independent column vectors $X_{i_1}|_\sigma, \dots, X_{i_{t-1}}|_\sigma$ in $\mathbf{F}_q^\sigma$.
We therefore define the event $\mathcal{E}_{\sigma, i_1, \dots, i_{t-1}}$ to be that

(1) $X_{i_1}|_\sigma, \dots, X_{i_{t-1}}|_\sigma$ are linearly independent,
(2) $X_i|_\sigma \in \operatorname{Span}(X_{i_1}|_\sigma, \dots, X_{i_{t-1}}|_\sigma)$ for all $1 \le i \le n-k$,
(3) the annihilator of $\operatorname{Span}(X_{i_1}|_\sigma, \dots, X_{i_{t-1}}|_\sigma)$ in $\mathbf{F}_q^\sigma$ contains no nonzero vectors with support begin a proper set of σ,

and observe that $\mathcal{E}_\sigma$ is a union over all such $\mathcal{E}_{\sigma, i_1, \dots, i_{t-1}}$.
Case 1. If $144\alpha_n^{-1} \le t \le \delta n$, then by Theorem A.21 (whose proof is given in **Part II**),

$$\mathbf{P}(\mathcal{E}_{\sigma, i_1, \dots, i_{t-1}}) \le \left(\frac{1}{q} + \frac{2}{\sqrt{\alpha_n t}}\right)^{n-k-t+1} \le \left(\frac{1}{q} + \frac{2}{\sqrt{\alpha_n t}}\right)^{(1-\varepsilon_0)n} \le \left(\frac{2}{3}\right)^{n/2}.$$

Thus

$$\mathbf{P}(\mathcal{E}_\sigma) \le \sum_{i_1, \dots, i_{t-1}} \mathbf{P}(\mathcal{E}_{\sigma, i_1, \dots, i_{t-1}}) \le \binom{n-k}{t-1}\left(\frac{2}{3}\right)^{n/2}.$$

So

$$\sum_{\sigma,|\sigma|\geq 144\alpha_n^{-1}} \mathbf{P}(\mathcal{E}_\sigma) \leq \sum_{144\alpha_n^{-1}\leq t\leq \delta n} \binom{n}{t}\binom{n-k}{t-1}\left(\frac{2}{3}\right)^{n/2} \leq \left(\frac{2}{3}\right)^{n/4}.$$

provided that ε_0 (and hence δ) are sufficiently small.

Case 2. If $1 \leq t \leq 144\alpha_n^{-1}$. We use the simple bound

$$\mathbf{P}\Big(X_i|_\sigma \in \mathrm{Span}(X_{i_1}|_\sigma, \dots, X_{i_{t-1}}|_\sigma)\Big) \leq 1-\alpha_n.$$

Hence

$$\mathbf{P}(\mathcal{E}_\sigma) \leq \sum_{i_1,\dots,i_{t-1}} \mathbf{P}(\mathcal{E}_{\sigma,i_1,\dots,i_{t-1}}) \leq \binom{n-k}{t-1}(1-\alpha_n)^{(1-\delta-\eta)n} \leq \binom{n-k}{t-1}(1-\alpha_n)^{(1-\varepsilon_0)n}.$$

Consequently,

$$\sum_{\sigma,|\sigma|\leq 144\alpha_n^{-1}} \mathbf{P}(\mathcal{E}_\sigma) \leq \sum_{t\leq 144\alpha_n^{-1}} \binom{n}{t}\binom{n-k}{t-1}(1-\alpha_n)^{(1-\varepsilon_0)n}.$$

Subcase 2.1. Assume that α_n is from (10). Then as $\alpha_n \geq n^{-1/2+\varepsilon}$, the above can be easily bounded by

$$\sum_{\sigma,|\sigma|\leq 144\alpha_n^{-1}} \mathbf{P}(\mathcal{E}_\sigma) \leq (1-\alpha_n)^{\varepsilon_0 n/2}.$$

Subcase 2.2. Assume that α_n is from (11),

$$\frac{C_0 \log n}{n} \leq \alpha_n = 1 - \max_x \mu(x) \leq n^{-1/2+\varepsilon}.$$

We will rely on the following observation, which is a simple variant of Lemma [1, Lemma 3.2].

Claim A.9. *The following holds with probability at least $1-e^{-c\alpha_n n}$ with respect to $X_1,\dots,X_{n-k}$. For any $1\leq t\leq 144\alpha_n^{-1}$, and any $\sigma \in \binom{[n]}{t}$, there are at least two columns X_i, X_{i+1} whose restriction $(X_{i+1}-X_i)|\sigma$ has exactly one non-zero entry.*

Suppose that $w \in \mathbf{F}_q^n$ has support σ of size $t = |\sigma| = |\mathbf{supp}(w)| \leq 144\alpha_n^{-1}$. Then, conditioning on the event in the lemma, there is some $1 \leq i \leq n-k-1$ so that $(X_{i+1}-X_i)|_\sigma$ has exactly one nonzero entry, and hence $w|\sigma$ is not orthogonal to it. Hence, it cannot be simultaneously orthogonal to all $X_i, 1\leq i\leq n-k$. Thus, it suffices to prove the claim.

Proof of Claim A.9. For each $i \in \{1,3,\dots,2\lfloor (n-k-1)/2\rfloor+1\}$, consider the vectors $Y_i = X_{i+1}-X_i$. The entries of this vector are iid copies of the symmetrized random variable $\psi = \xi - \xi'$, where ξ', ξ are independent and have distribution μ. With $1-\alpha_n' = \mathbf{P}(\psi = 0)$, then we have

$$\alpha_n \leq \alpha_n' \leq 2\alpha_n$$

as this can be seen by

$$(1-\alpha_n)^2 \leq \max_x \mathbf{P}(\xi=x)^2 \leq \sum_x \mathbf{P}(\xi=x)^2 = \mathbf{P}(\psi=0) \leq \max_x \mathbf{P}(\xi=x) = 1-\alpha_n.$$

Now let p_σ be the probability that all $Y_i|_\sigma, i\in\{1,3,\dots,2\lfloor(n-k-1)/2\rfloor+1\}$ fail to have exactly one non-zero entry, then by independence of the columns and of the entries

$$p_\sigma = (1-t\alpha_n'(1-\alpha_n')^{t-1})^{(n-k)/2} \leq (1-t\alpha_n' e^{-t\alpha_n'})^{n-k} \leq e^{-nt\alpha_n' e^{-t\alpha_n'}/2}.$$

Notice that as $1\leq t\leq 8^{3/\varepsilon_0}\alpha_n^{-1}$, $e^{-t\alpha_n'}/2 \geq c$ for some positive constant c, and hence

$$e^{-nt\alpha_n' e^{-t\alpha_n'}/2} \leq (e^{-cn\alpha_n'})^t \leq n^{-cC_0 t/2} e^{-cn\alpha_n/2}.$$

Thus

$$\sum_{1\leq t\leq 144\alpha_n^{-1}} \sum_{\sigma\in\binom{[n]}{t}} p_\sigma \leq \sum_{1\leq t\leq 144\alpha_n^{-1}} \binom{n}{t} e^{-(n-k)t\alpha_n e^{-t\alpha_n}} \leq \sum_{1\leq t\leq 144\alpha_n^{-1}} (n^t n^{-cC_0 t/2}) e^{-cn\alpha_n/2} < e^{-cn\alpha_n/2},$$

provided that C_0 is sufficiently large. □

Our proof of Proposition A.8 is then complete by combining the cases considered above. □

Remark A.10. *Our treatment here is quite different from* [9, Section 2.1] *as there is no need to use* [9, Proposition 2.3] *(which states that "Let $Z_1, \dots, Z_r$ be non-trivial iid random vectors in $\mathbf{F}_q^n$, then $\mathbf{P}(Z_1, \dots, Z_r \in V | Z_1, \dots, Z_r$ are linearly independent) $\leq \mathbf{P}(Z \in V)^r$." An elementary counterexample to this proposition is that Z_i are chosen uniformly at random from $\{a_1, 2a_1, 3a_1, a_2, a_3\}$, where $a_1 \notin V$ and a_2, a_3 are linearly independent in V: in this case the LHS probability bound is larger than the RHS* [2]*. Also, Maples did not provide any treatment for the sparse case (Case 2) for Proposition A.8 in* [9] *or* [11]*, which we have added. See also the remark after Lemma A.12 below.*

To conclude, given constants η, δ and the parameter α_n from (10) or (11), let $\mathcal{E}_{n-k,dense}$ denote the event considered in Proposition A.8,

$$\mathbf{P}(\mathcal{E}_{n-k,dense}) \geq 1 - e^{-c\alpha_n n}. \tag{13}$$

We next turn to another type of subspace.

A.11. **Semi-saturated subspace.** Given $0 < \alpha_n, \delta, d < 1$. We call a subspace V of co-dimension k_0 *semi-saturated* (or *semi-sat* for short), where $k_0 \leq \eta n$, if V is not δ-sparse and

$$e^{-d\alpha_n n} < |\mathbf{P}(X \in V) - \frac{1}{q^{k_0}}| \leq \frac{16}{q^{k_0}}. \tag{14}$$

Here we assume

$$e^{-d\alpha_n n} < \frac{16}{q^{k_0}}.$$

If this condition is not satisfied (such as when q is sufficiently large), then the semi-saturated case can be omitted.

Lemma A.12. [9, Proposition 2.5] *For all $\beta > 0$ and $\delta > 0$ there exists $0 < d = d(\beta, \delta) < 1$ in the definition of semi-saturation and a deterministic set $\mathcal{R} \subset \mathbf{F}_q^n$ of non δ-sparse vectors and of size $|\mathcal{R}| \leq (2\beta^\delta)^n q^n$ such that every semi-saturated V is orthogonal to a vector $R \in \mathcal{R}$. In fact the conclusion holds for any subspace V satisfying the LHS of* (14)*.*

In short, semi-saturated subspaces are necessarily orthogonal to one of a small number of non-sparse vectors in $\mathbf{F}_q^n$. A proof of this result will be given in **Part II** where we emphasize that α_n can be as small as (11), in contrast to the proof of [9, Proposition 2.5] where α_n was treated as a constant.
Let $\mathcal{F}_{n-k,k_0,semi-sat}$ be the event that $\mathrm{codim}(W_{n-k}) = k_0$ and W_{n-k} is semi-saturated.

Proposition A.13. *Let $\beta, \delta > 0$ be parameters such that $\beta^\delta < 17^{-2}/2$. With $d = d(\beta, \delta)$ from Lemma A.12 we have*

$$\mathbf{P}(\mathcal{F}_{n-k,k_0,semi-sat}) \leq e^{-n}.$$

In particularly, with $\mathcal{E}_{n-k,semi-sat}$ being the event $\wedge_{k \leq k_0 \leq \eta n} \overline{\mathcal{F}}_{n-k,k_0,semi-sat}$ in the σ-algebra generated by $X_{k+1}, \dots, X_n$, then

$$\mathbf{P}(\mathcal{E}_{n-k,semi-sat}) \geq 1 - e^{-n/2}. \tag{15}$$

Proof of Proposition A.13. We have

$$\mathbf{P}(\mathcal{F}_{n-k,k_0,semi-sat}) = \sum_{V semi-sat} \mathbf{P}(W_{n-k} = V) \leq \sum_{V semi-sat} \mathbf{P}(X_1, \dots, X_{n-k} \in V).$$

Now for each fixed V that is semi-saturated of co-dimension $k_0 \geq k$, by definition $\mathbf{P}(X \in V) \leq 17 q^{-k_0}$. So

$$\mathbf{P}(X_1, \dots, X_{n-k} \in V) \leq 17^{n-k} q^{-k_0(n-k)}.$$

We next use Lemma A.12 to count the number $N_{semi-sat}$ of semi-saturated subspaces V. Each V is determined by its annihilator $V^\perp$ (of cardinality q^{k_0}). To count $V^\perp$, we first choose a vector $v \in \mathcal{R}$, and then another $(k_0 - 1)$ dimensional subspace that is linearly independent of v. The number of ways to complete

[2] We thank M. M. Wood for pointing out the mistake, as well as for supplying a counterexample.

this space equals the number of ways to pick a (k_0-1)–dimensional space from $\mathbf{F}_q^{n-1}$. The number of such subspaces is given by the well–known (see [13, Proposition 1.3.18]) exact formula

$$\prod_{j=0}^{k_0-2}\frac{q^{n-1-j}-1}{q^{k_0-1-j}-1}\leq Cq^{(k_0-1)(n-k_0)},$$

for some absolute constant $C>0$. Therefore

$$N_{semi-sat}\leq C(2\beta^\delta)^n q^{nk_0-k_0^2+k_0}.$$

Therefore

$$\begin{aligned}\mathbf{P}(\mathcal{F}_{k,k_0,semi-sat})&\leq\sum_{V\,semi-sat}\mathbf{P}(X_1,\dots,X_{n-k}\in V)=O\Big((2\beta^\delta)^n q^{nk_0-k_0^2+k_0}17^{n-k}q^{-k_0(n-k)}\Big)\\&=O\Big(17^{n-k}(2\beta^\delta)^n q^{k_0}q^{k_0(k-k_0)}\Big)=O\Big(17^{n-k}(2\beta^\delta)^n q^{k_0}\Big).\end{aligned}$$

Now recall that $e^{-d\alpha_n n}\leq 16q^{-k_0}$, and so

$$\mathbf{P}(\mathcal{F}_{n-k,k_0,semi-sat})=O(17^{n-k}(2\beta^\delta)^n q^{k_0})=O(17^{n+1-k}(2\beta^\delta)^n e^{d\alpha_n n}).$$

We then choose β so that $2\beta^\delta<17^{-2}$ and with $d<1$ we have $\mathbf{P}(\mathcal{F}_{n-k,k_0,semi-sat})\leq e^{-n}$. □

A.14. **Unsaturated subspace.** Recall that $k\leq\eta n$ for sufficiently small η. Let V be a subspace of codimension $k_0\geq k$ in $\mathbf{F}_q^n$. We say that V is *unsaturated* if V is not δ-sparse and

$$\max(e^{-d\alpha_n n},16q^{-k_0})<|\mathbf{P}(X\in V)-q^{-k_0}|.$$

In particularly this implies that

$$\mathbf{P}(X\in V)\geq\max\{17q^{-k_0},\frac{16}{17}e^{-d\alpha_n n}\}.$$

The following is from [9, Lemma 2.8].

Lemma A.15. *There is an α_n'-balanced probability distribution ν on $\mathbf{F}_q$ with $\alpha_n'=\alpha_n/64$ such that if $Y\in\mathbf{F}_q^n$ is a random vector with iid coefficients distributed according to ν, then for any unsaturated proper subspace V*

$$|\mathbf{P}(X\in V)-\frac{1}{q^{k_0}}|\leq(\frac{1}{2}+o(1))|\mathbf{P}(Y\in V)-\frac{1}{q^{k_0}}|.$$

A proof of this lemma will be given in **Part II**. By definition, if V is unsaturated then

$$\mathbf{P}(Y\in V)\geq(2-o(1))(\mathbf{P}(X\in V)-\frac{1}{q^{k_0}})+\frac{1}{q^{k_0}}>\frac{3}{2}\mathbf{P}(X\in V).$$

Definition A.16. Let V be a subspace in $\mathbf{F}_q^n$. Let $d_{comb}\in\{1/n,\dots,n^2/n\}$. We say that V has combinatorial codimension d_{comb} if

$$(1-\alpha_n)^{d_{comb}}\leq\mathbf{P}(X\in V)\leq(1-\alpha_n)^{d_{comb}-1/n}.$$

Now as we are in the unsaturated case, $\mathbf{P}(X\in V)\geq\frac{16}{17}e^{-d\alpha_n n}$, and so

$$d_{comb}\leq 2dn.$$

In what follows we will fix d_{comb} from the above range, noting that d is sufficiently small, and there are only $O(n^2)$ choices of d_{comb}.
Let be fixed any $0<\delta_1<\delta_2<1/3$ such that

$$16(\delta_2-\delta_1)(1+\log\frac{1}{\delta_2-\delta_1})<\delta_1. \tag{16}$$

Set

$$r:=\lfloor\delta_1 n\rfloor\text{ and }s:=n-k-\lfloor\delta_2 n\rfloor.$$

Let $Y_1,\dots,Y_r$ be random vectors with entries distributed by ν obtained by Lemma A.15, and let $Z_1,\dots,Z_s$ and $X_{r+s+1},\dots,X_{n-k}$ be random vectors with entries distributed by μ.

Proposition A.17. *Let $W = \mathrm{Span}\{X_1, \dots, X_{n-k}\}$. We have*

$$\mathbf{P}\Big(r+s \le \dim(W) \le n-k, W \textit{ unsaturated}\Big) \le (3/2)^{-r/2}\binom{n-k}{r+s} \le (3/2)^{-\delta_1 n/4}.$$

The second inequality follows directly from (16) and the standard bound $\binom{n}{k} \le \left(\frac{en}{n-k}\right)^{n-k}$. Notice that we do not require $\{X_i\}$ to be linearly independent. In other words, let $\mathcal{E}_{n-k,unsat}$ denote the complement of the event above in the σ-algebra generated by $X_1, \dots, X_{n-k}$, then

$$\mathbf{P}(\mathcal{E}_{n-k,unsat}) \ge 1 - (3/2)^{-\delta_1 n/4}. \tag{17}$$

To prove Proposition A.17 we show the following:

Theorem A.18. *Let V be any subspace of dimension between $r+s$ and $n-k$ and having $d_{comb} \le 2dn$. Then we have*

$$\mathbf{P}\Big(\mathrm{Span}\{X_1, \dots, X_{n-k}\} = V\Big) \le (3/2)^{-r/2}\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\Big\{\{Y_i\}_1^r, \{Z_i\}_1^s, \{X_i\}_{r+s+1}^{n-k}\Big\} = V\Big).$$

To conclude Proposition A.17 we then just use

$$\sum_{\substack{V \le \mathbf{F}_q^n \\ \mathrm{codim}(V) \ge k}} \mathbf{P}\Big(\mathrm{Span}\Big\{\{Y_i\}_1^r, \{Z_i\}_1^s, \{X_i\}_{r+s+1}^{n-k}\Big\} = V\Big) = 1. \tag{18}$$

Proof of Theorem A.18. First of all, by independence between X_i, Y_j, Z_l,

$$\begin{aligned}
&\mathbf{P}\Big(\mathrm{Span}\{X_1, \dots, X_{n-k}\} = V\Big) \times \mathbf{P}\Big(Y_1, \dots, Y_r, Z_1, \dots, Z_s \text{ linearly independent in } V\Big) \\
=&\mathbf{P}\Big(\mathrm{Span}\{X_1, \dots, X_{n-k}\} = V \wedge Y_1, \dots, Y_r, Z_1, \dots, Z_s \text{ linearly independent in } V\Big).
\end{aligned} \tag{19}$$

We next estimate $\mathbf{P}(Y_1, \dots, Y_r, Z_1, \dots, Z_s \text{ linearly independent in } V)$. By conditioning,

$$\begin{aligned}
&\mathbf{P}\Big(Z_1, \dots, Z_s, Y_1, \dots, Y_r \text{ linearly independent in } V\Big) \\
=&\mathbf{P}\Big(Y_r \in V\Big)\mathbf{P}\Big(Y_{r-1} \in V, Y_{r-1} \notin \langle Y_r\rangle | Y_r \in V\Big) \cdots \mathbf{P}\Big(Y_1 \in V, Y_1 \notin \langle Y_2, \dots, Y_r\rangle | Y_2, \dots, Y_r \text{ lin. in } V\Big) \\
&\times \cdots \times \mathbf{P}\Big(Z_1 \in V, Z_1 \notin \langle Z_2, \dots, Z_r, Y_1, \dots, Y_r\rangle | Z_2, \dots, Z_s, Y_1, \dots, Y_r \text{ lin. in } V\Big).
\end{aligned}$$

We first estimate the terms involving Y_i. By Lemma 2.3

$$\begin{aligned}
&\mathbf{P}\Big(Y_i \in V, Y_i \notin \langle Y_{i+1}, \dots, Y_r\rangle | Y_{i+1}, \dots Y_r \text{ lin. in } V\Big) \ge \mathbf{P}(Y_i \in V) - (1-\alpha')^{n-(r-i)} \\
\ge&\frac{3}{2}\mathbf{P}(X_i \in V) - (1-\alpha_n')^{n-(r-i)} \ge \frac{3}{2}(1-\alpha_n)^{d_{comb}} - (1-\alpha_n')^{n-(r-i)} \\
\ge&\frac{3}{2}(1-\alpha_n)^{d_{comb}}(1-(1-\alpha_n)^{n/256-d_{comb}}),
\end{aligned}$$

where we used that $\alpha_n' = \alpha_n/64$ and $n - r \ge (1-\delta_1)n \ge n/2$.
Similarly, the terms involving Z_i can be estimated as

$$\begin{aligned}
&\mathbf{P}\Big(Z_i \in V, Z_i \notin \langle Z_{i+1}, \dots, Z_s, Y_1, \dots, Y_r\rangle | Z_{i+1}, \dots, Z_s, Y_1, \dots, Y_r\Big) \\
\ge&\mathbf{P}(Z_i \in V) - (1-\alpha_n')^{n-(r+s-i)} \\
\ge&(1-\alpha_n)^{d_{comb}} - (1-\alpha_n')^{n-(r+s-i)} \ge (1-\alpha_n)^{d_{comb}} - (1-\alpha_n)^{n/256},
\end{aligned}$$

where we used that $r+s = n-k-(\lfloor\delta_2 n\rfloor - \lfloor\delta_1 n\rfloor) \ge n/2$. So

$$\begin{aligned}
\mathbf{P}\Big(Y_1, \dots, Y_r, Z_1, \dots, Z_s \text{ linearly independent in } V\Big) &\ge (3/2)^r(1-\alpha_n)^{(r+s)d_{comb}}\Big(1-(1-\alpha_n)^{n/256-d_{comb}}\Big)^{r+s} \\
&\ge (3/2)^{r-1}(1-\alpha_n)^{(r+s)d_{comb}},
\end{aligned} \tag{20}$$

where we used $d_{comb} \le 2dn$ and d is sufficiently small.

Now we estimate the probability $\mathbf{P}(\mathrm{Span}\{X_1,\dots,X_{n-k}\}=V\wedge Y_1,\dots,Y_r,Z_1,\dots,Z_s$ linearly independent in $V)$ in (19). Since $Y_1,\dots,Y_r,Z_1,\dots,Z_s$ are linearly independent in V and $\mathrm{Span}\{X_1,\dots,X_{n-k}\}=V$, there exists $n-k-r-s$ vectors $X_{i_1},\dots,X_{i_{n-k-r-s}}$ which together with $Y_1,\dots,Y_r,Z_1,\dots,Z_s$ are a basis for V. With a loss of a factor $\binom{n-k}{r+s}$ in probability, we can assume that $\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V$, and the remaining vectors $X_1,\dots,X_{r+s}$ belong to V. Thus,

$$\begin{aligned}
&\mathbf{P}\Big(\mathrm{Span}\{X_1,\dots,X_{n-k}\}=V\wedge Y_1,\dots,Y_r,Z_1,\dots,Z_s \text{ linearly independent in } V\Big)\\
\le&\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V\wedge X_1,\dots,X_{r+s}\in V\Big)\\
\le&\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V\Big)\mathbf{P}(X_1,\dots,X_{r+s}\in V)\\
\le&\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V\Big)(1-\alpha_n)^{(r+s)(d_{comb}-1/n)}. \qquad (21)
\end{aligned}$$

Putting (19), (20) and (21) together,

$$\begin{aligned}
&\mathbf{P}\Big(\mathrm{Span}\{X_1,\dots,X_{n-k}\}=V\Big)\\
=&\frac{\mathbf{P}\Big(\mathrm{Span}\{X_1,\dots,X_{n-k}\}=V\wedge Y_1,\dots,Y_r,Z_1,\dots,Z_s \text{ linearly independent in } V\Big)}{\mathbf{P}\Big(Y_1,\dots,Y_r,Z_1,\dots,Z_s \text{ linearly independent in } V\Big)}\\
\le&(3/2)^{-r+1}(1-\alpha_n)^{-(r+s)d_{comb}}\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V\Big)(1-\alpha_n)^{(r+s)(d_{comb}-1/n)}\\
\le&(3/2)^{-r/2}\binom{n-k}{r+s}\mathbf{P}\Big(\mathrm{Span}\left\{\{Y_i\}_1^r,\{Z_i\}_1^s,\{X_i\}_{r+s+1}^{n-k}\right\}=V\Big).
\end{aligned}$$

□

We remark that our proof above follows [14, Section 4]. The treatment of [9] is similar but the author oversimplified the process by relying on the aforementioned incorrect result [9, Proposition 2.3]. We now conclude the main result.

Proof of Theorem A.1. Let $\mathcal{E}_{n-k,dense},\mathcal{E}_{n-k,semi-sat},\mathcal{E}_{n-k,unsat}$ be the events introduced in (13), (15), (17). By definition, on these events, if $\mathrm{codim}(W_{n-k})=k_0$ then

$$|\mathbf{P}(X\in W_{n-k})-\frac{1}{q^{k_0}}|\le e^{-d\alpha_n n}.$$

□

Next we give a proof of Theorem A.4. The statement is clearly equivalent to Theorem A.1 if $k>u$. In what follows we assume $k\le u$.

Proof of Theorem A.4. . By Proposition A.8, with probability at least $1-e^{-c\alpha_n n}$ the subspace $\langle X_1,\dots,X_{n-k}\rangle$ is not δ-sparse, and hence $\langle X_1,\dots,X_{n+u-k}\rangle$ is also not δ-sparse on this event.
For the semi-saturated subspace, the conclusion of Proposition A.13 continues to hold using Lemma A.12. Indeed, with the same choice of parameters, by the proof of Proposition of A.13

$$\begin{aligned}
\mathbf{P}(\mathcal{F}_{n+u-k,k_0,semi-sat})=\sum_{V semi-sat}\mathbf{P}(W_{n+u-k}=V)&\le\sum_{V semi-sat}\mathbf{P}(X_1,\dots,X_{n+u-k}\in V)\\
&\le\sum_{V semi-sat}\mathbf{P}(X_1,\dots,X_{n-k}\in V)\le e^{-n}.
\end{aligned}$$

Finally, for unsaturated subspaces, by the same method we can show the following analog of Proposition A.17 with the same parameters

$$\mathbf{P}\Big(\mathrm{Span}\{X_1,\dots,X_{n+u-k}\}\text{ unsaturated and of dim. between } r+s \text{ and } n\Big)\le(3/2)^{-r/2}\binom{n+u-k}{r+s}. \qquad (22)$$

Indeed, to justify this result we just use the same swapping method of Theorem A.18; the only difference is that there are $\binom{n+u-k}{r+s}$ ways to choose the $X_1, \dots, X_{r+s}$ in (21). The bound (22) is again smaller than $(3/2)^{-\delta_1 n/4}$ if $u \le \eta n$ with η sufficiently small compared to δ_1. □

Part II. Proofs of the lemmas: Here we sketch the proof of Lemma A.12, Lemma A.15, and Theorem A.21 by following [9].

Recall that $\mathrm{tr} : \mathbf{F}_q \to \mathbf{F}_p$ is the field trace, which gives rise to the isomorphism between $\mathbf{F}_q$ and $\widehat{\mathbf{F}}_q$ by $x \to e_p(\mathrm{tr}(tx)) = \exp(2\pi i \mathrm{tr}(tx)/p)$ (and so we will identify $\widehat{\mathbf{F}}_q$ with $\mathbf{F}_q$). Let μ be a probability measure on $\mathbf{F}_q$. The Fourier transform of μ is

$$\widehat{\mu}(x) = \sum_{t \in \mathbf{F}_q} \mu(t) e_p(\mathrm{tr}(xt)) = \mathbf{E} e_p(\mathrm{tr}(x\xi)), \text{ where } \xi \text{ has distribution } \mu.$$

We define the additive spectrum by

$$\mathbf{Spec}_{1-\varepsilon}\mu := \{x \in \mathbf{F}_q, |\widehat{\mu}(x)| \ge 1 - \varepsilon\}.$$

We will be using the following two important results.

Theorem A.19 (Kneser)**.** [15, Theorem 5.5] *Let $A, B \subset Z$ be finite subsets of an abelian group Z. Then*

$$|A + B| + |\operatorname{Sym}(A + B)| \ge |A| + |B|,$$

where $\operatorname{Sym}(X) = \{h \in Z : h + X = X\}$.

Note that $\operatorname{Sym}(X)$ is a symmetric additive subgroup of Z. As $\operatorname{Sym}(A_1 + \cdots + A_k)$ is increasing in k, iterating the result we obtain

Corollary A.20.

$$|A_1 + \cdots + A_k| + (k-1)|\operatorname{Sym}(A_1 + \cdots + A_k)| \ge |A_1| + \cdots + |A_k|.$$

To start with, we prove the following version of the classical Erdős-Littlewood-Offord result in finite field, which was used in the proof of Proposition A.8.

Theorem A.21. [9, Theorem 2.4] *Let $X \in \mathbf{F}_q^n$ be a random vector with iid entries taken from an α-balanced distribution. Suppose $w \in \mathbf{F}_q^n$ has at least m non-zero coefficients. Then we have*

$$|\mathbf{P}(X \cdot w = r) - \frac{1}{q}| \le \frac{2}{\sqrt{\alpha m}}.$$

Proof of Theorem A.21. Let $\xi_1, \dots, \xi_n$ denote the entries of X, and $w_1, \dots, w_n$ denote the components of w. By $\mathbf{F}_p$–linearity of the field trace, we can write

$$1_{X \cdot w = r} = \frac{1}{q} \sum_{t \in \mathbf{F}_q} e_p\Big(\mathrm{tr}(\sum_l \xi_l w_l)\Big) e_p\Big(-\mathrm{tr}(rt)\Big).$$

By independence of the $\{\xi_i\}_1^n$ we can therefore write

$$\mathbf{P}(X \cdot w = r) = \mathbf{E}(1_{X \cdot w = r}) = \frac{1}{q} \sum_{t \in \mathbf{F}_q} \prod_{l=1}^n \mathbf{E} e_p(\mathrm{tr}(\xi_l w_l t)) e_p(-\mathrm{tr}(rt)).$$

By the triangle inequality

$$|\mathbf{P}(X \cdot w = r) - \frac{1}{q}| \le \frac{1}{q} \sum_{t \in \mathbf{F}_q, t \ne 0} \prod_{l=1}^n |\mathbf{E} e_p(\mathrm{tr}(\xi_l w_l t))| = \frac{1}{q} \sum_{t \in \mathbf{F}_q, t \ne 0} \prod_{l=1}^n |\widehat{\mu}(w_l t)|.$$

Define $\psi(t) = 1 - |\widehat{\mu}(t)|^2$. Using $|x| \le \exp(-(1-x^2)/2)$ for $|x| \le 1$,

$$|\mathbf{P}(X \cdot w = r) - \frac{1}{q}| \le \frac{1}{q} \sum_{t \in \mathbf{F}_q, t \ne 0} \exp\Big(-\frac{1}{2} \sum_{l=1}^n \psi(w_l t)\Big).$$

Set $f(t) = \sum_l \psi(w_l t)$, then

$$|\mathbf{P}(X \cdot w = r) - \frac{1}{q}| \leq \frac{1}{2} \int_0^\infty \frac{1}{q} |\{t \neq 0, f(t) \leq v\}| e^{-v/2} dv = \frac{1}{2} \int_0^\infty \frac{1}{q} |T'(v)| e^{-v/2} dv \tag{23}$$

where the level sets are defined as

$$T(v) := \{t, f(t) \leq v\} \text{ and } T'(v) = T(v) \backslash \{0\}.$$

Claim A.22. *For any $v > 0$ we have*

$$kT(v) = T(v) + \cdots + T(v) \subset T(k^2 v).$$

Proof. It suffices to show that for any $\beta_1, \ldots, \beta_k \in \mathbf{F}_q$

$$\psi(\beta_1 + \cdots + \beta_k) \leq k(\psi(\beta_1) + \cdots + \psi(\beta_k)).$$

Indeed, by definition of ψ, after squaring out the above is equivalent to

$$1 - \sum_{t_1, t_2 \in \mathbf{F}_q} \mu(t_1)\mu(-t_2) \cos\Big(\frac{2\pi}{p} \mathrm{tr}((t_1+t_2)(\beta_1 + \cdots + \beta_k))\Big) \leq k^2 - k \sum_i \sum_{t_1, t_2 \in \mathbf{F}_q} \mu(t_1)\mu(-t_2) \cos\Big(\frac{2\pi}{p} \mathrm{tr}((t_1+t_2)\beta_i)\Big).$$

Hence it suffices to observe that for all real numbers $(\beta_i)_1^k$, $\cos(\beta_1 + \cdots + \beta_k) \geq k \sum \cos \beta_i - k^2 + 1$, which we justify now. If for even a single β_i, $\cos(\beta_i) \leq 0$, then the largest value that can be attained on the right hand side is $1 - k$, from which it follows the equality holds. Hence by periodicity, we may assume that all these $\beta_i \in (-\pi/2, \pi/2)$. The function

$$(\beta_i)_1^k \mapsto \cos(\beta_1 + \cdots + \beta_k) - k \sum \cos(\beta_i)$$

is smooth and its only local minimum in the domain considered can be checked to occur at 0, at which value equality is attained.

□

By Corollary A.20

$$kT(v) \leq |T(k^2 v)| + (k-1) \operatorname{Sym}(T(v) + \cdots + T(v)). \tag{24}$$

We next claim that if $k^2 v < \alpha_n m$, that is $k < \sqrt{\frac{\alpha_n m}{v}}$, then $T(k^2 v)$ contains no-nontrivial additive subgroup H of $\mathbf{F}_q$, and so $|\operatorname{Sym}(T(v) + \cdots + T(v))| = 1$. Indeed, fix a subgroup H, then

$$|H|^{-1} \sum_{t \in H} f(t) = \sum_{l=1}^n |H|^{-1} \sum_{t \in H} \psi(w_l t) = \sum_{l=1}^n |H|^{-1} \sum_{t \in H} (1 - |\widehat{\mu}(w_l t)|^2).$$

By the Fourier inversion formula, and by the α-balanced assumption on the distribution μ

$$|H|^{-1} \sum_{t \in H} (1 - |\widehat{\mu}(w_l t)|^2) = \sum_{s_1, s_2 \in \mathbf{F}_q} \mu(s_1)\mu(s_2) 1_{H^\perp}(w_l(s_1 - s_2)) \leq 1 - \alpha_n.$$

Since at least m of the choices w_l are non-zero

$$|H|^{-1} \sum_t f(t) \geq \alpha_n m.$$

Thus there exists $t \in H$ such that $f(t) \geq \alpha_n m$, and so $H \not\subset T(k^2 v)$. So by (24) we have

$$|T'(v)| \leq \sqrt{\frac{v}{\alpha_n m}} |T'(\alpha_n m)| \leq \sqrt{\frac{v}{\alpha_n m}} q, \text{ for all } v \leq \alpha_n m.$$

Substitute back to (23) we obtain

$$|\mathbf{P}(X \cdot w = r) - \frac{1}{q}| \leq \frac{1}{2} \frac{1}{\sqrt{\alpha_n m}} \int_0^\infty \sqrt{v} e^{-v/2} dv + e^{-\alpha_n m/2}.$$

□

Now we prove the other lemmas that were used in the treatment of semi-saturated and un-saturated subspaces.

Proof of Lemma A.12. Note that V is not δ-sparse. Let $k_0 = \mathrm{codim}(V)$, let $\xi_1, \dots, \xi_n$ denote the entries of X, we have

$$\mathbf{P}(X \in V) = \mathbf{E} q^{-k_0} \sum_{t \in V^\perp} e_p\big(\mathrm{tr}(\sum_l \xi_l t_l)\big) = q^{-k_0} \sum_{t \in V^\perp} \prod_{l=1}^n \mathbf{E} e_p(\mathrm{tr}(\xi_l t_l)) = q^{-k_0} \sum_{t \in V^\perp} \prod_{l=1}^n \widehat{\mu}(t_l),$$

where $t_1, \dots, t_n$ denote the entries of t, and where we used the fact that $\sum_{t \in V^\perp} e_p(\mathrm{tr}(\sum_l \xi_l t_l)) = 0$ if and only if $X \notin V$. By the triangle inequality,

$$|\mathbf{P}(X \in V) - q^{-k_0}| \leq q^{-k_0} \sum_{t \in V^\perp, t \neq 0} \prod_{l=1}^n |\widehat{\mu}(t_l)|.$$

By the pigeonhole principle, for some $t \in V^\perp, t \neq 0$,

$$e^{-d\alpha_n n} \leq |\mathbf{P}(X \in V) - q^{-k_0}| \leq \prod_{l=1}^n |\widehat{\mu}(t_l)|.$$

Again, using $|x| \leq \exp(-\frac{1}{2}(1-x^2))$ for $|x| \leq 1$,

$$\sum_{l=1}^n 1 - |\widehat{\mu}(t_l)|^2 \leq 2d\alpha_n n.$$

By averaging, there exists an index set $\sigma \subset [n]$ with $|\sigma| \geq (1-\delta/2)n$ and $|\widehat{\mu}(t_l)| \geq 1 - 10d\delta^{-1}\alpha_n$ for $l \in \sigma$. In other words, for all $l \in \sigma$

$$t_l \in \mathbf{Spec}_{1-10d\delta^{-1}\alpha_n} \mu.$$

Claim A.23. *The set $\mathbf{Spec}_{1-\alpha_n/2}$ does not contain any non-trivial additive subgroup H of $\mathbf{F}_q$.*

Proof. By Fourier's inversion formula

$$(1-\alpha_n/2)^2 |H \cap \mathbf{Spec}_{1-\alpha_n/2}(\mu)| \leq \sum_{s \in H} |\widehat{\mu}(s)|^2 \leq |H|(1-\alpha_n),$$

where in the last estimate we used the fact that μ is α_n-balanced. □

Set $k := \lfloor \beta^{-1} \rfloor$ and choose d so that

$$d \leq k^{-2}\delta/5. \tag{25}$$

Then

$$\mathbf{Spec}_{1-10d\delta^{-1}\alpha_n} \subset \mathbf{Spec}_{1-2k^{-2}\alpha_n}(\mu) := A.$$

We next claim that

$$|A \backslash \{0\}| \leq \beta q. \tag{26}$$

Indeed, this is because by applying Cauchy-Schwarz

$$kA \subset \mathbf{Spec}_{1-2\alpha_n}(\mu).$$

Furthermore,

$$\mathrm{Sym}(kA) \subset (kA) - (kA) = 2kA \subset \mathbf{Spec}_{1-\alpha_n/2}(\mu).$$

By Claim A.23, $\mathrm{Sym}(kA)$ is trivially $\{0\}$. So by Corollary A.20

$$k|A| \leq |kA| + (k-1) \leq q + (k-1),$$

proving (26).
Finally, let $\mathcal{R}$ be the set of non-zero t in $\mathbf{F}_q^n$ which have at least $(1-\delta/2)n$ components t_l in $\mathbf{Spec}_{1-10d\delta^{-1}\alpha}$. By (26), as $t \perp V$ and V is not δ-sparse, at least $\delta/2$ of the components ξ_l are non-zero

$$|\mathcal{R}| \leq 2^n (\beta q)^{\delta n/2} q^{n - \delta n/2} \leq (2\beta^{\delta/2} q)^n.$$

□

To prove Lemma A.15, we use the following rather standard Halász–type construction from [9, Proposition 3.6] (see also [14, Lemma 7.1] or [3]).

Proposition A.24. *There is a probability distribution $\nu : \mathbf{F}_q \to [0,1]$ depending on μ and α such that the following properties hold with $t = (t_1, \dots, t_n)$ and*

$$f(t) := \prod_{l=1}^{n} |\widehat{\mu}(t_l)|, g(t) := \prod_{l=1}^{n} |\widehat{\nu}(t_l)|.$$

- *For all $0 < u < 1$ we have $4F(u) \subset G(u)$, where*
$$F(u) = \{t \in \mathbf{F}_q^n, |f(t)| > u\} \text{ and } G(u) = \{t \in \mathbf{F}_q^n, |g(t)| > u\}.$$
- *For all $t \in V^\perp$,*
$$f(t) \leq g^{16}(t).$$
- *For all t,*
$$\widehat{\nu}(t) \geq 0.$$
- *ν is α_n'-balanced with $\alpha_n' = \alpha_n/64$.*

Proof of Lemma A.15. Note that

$$\mathbf{P}(Y \in V) - q^{-k} = q^{-k} \sum_{t \in V^\perp, t \neq 0} \prod_{l=1}^{n} \widehat{\nu}(t_l) = q^{-k} \sum_{t \in V^\perp, t \neq 0} g(t).$$

Thus to show $|\mathbf{P}(X \in V) - q^{-k}| \leq (\frac{1}{2} + o(1))|\mathbf{P}(Y \in V) - q^{-k}|$, as $\widehat{\nu} \geq 0$ it suffices to show that

$$\sum_{t \in V^\perp, t \neq 0} f(t) \leq (\frac{1}{2} + o(1)) \sum_{t \in V^\perp, t \neq 0} g(t).$$

Let $\varepsilon > 0$ be a parameter to be sent to 0. We write

$$\sum_{t \in V^\perp, t \neq 0} f(t) = \sum_{t \in V^\perp, t \neq 0, f(t) < \varepsilon} f(t) + \sum_{t \in V^\perp, t \neq 0, f(t) \geq \varepsilon} f(t) := \sum_{<\varepsilon}(f) + \sum_{\geq \varepsilon}(f).$$

As $f(t) \leq g(t)^{16}$, we have

$$\sum_{<\varepsilon}(f) \leq \varepsilon^{15/16} \sum_{<\varepsilon}(g) < (\frac{1}{2} + o(1)) \sum_{<\varepsilon}(g).$$

We also write

$$\sum_{\geq \varepsilon}(f) = \int_{\varepsilon}^{\infty} |F'(u)| du + \varepsilon |F'(\varepsilon)|,$$

where

$$F'(u) = F(u) \backslash \{0\} = \{t \in V^\perp, t \neq 0, |f(t)| > u\}.$$

Claim A.25. *With $\varepsilon = \exp(-\frac{1}{2}\alpha_n' \delta n)$, the set $G(\varepsilon)$ does not contain any non-trivial additive subgroup $H \leq V^\perp$. In particularly, as $G(u)$ is decreasing, the same happens for any $u \geq \varepsilon$.*

Proof. Clearly we can assume $H \cong \mathbf{Z}/p\mathbf{Z}$. Assume that $w = (w_1, \dots, w_n) \in V^\perp$ that generates H. Since V is unsaturated (and hence not δ-sparse), w has at least δn non-zero components. Define $h(t) := \sum_{l=1}^{n} 1 - |\widehat{\nu}(t_l)|^2, t \in H$. We can also write $h(t) := \sum_{i=1}^{k} 1 - |\widehat{\nu}(t w_{l_i})|^2, t \in \mathbf{Z}/p\mathbf{Z}$, where w_{l_i} are non-zero. By Fourier's inversion formula, and as ν is α_n'-balanced, $\sum_{t \in \mathbf{Z}/p\mathbf{Z}} |\widehat{\nu}(t w_{l_i})|^2 \leq p(1 - \alpha_n')$. So

$$\sum_{t \in H} h(t) \geq p\alpha_n' k \geq p\alpha_n' \delta n.$$

By pigeon-hole principle, there exists $t \in H$ such that $h(t) \geq \alpha_n' \delta n$. On the other hand

$$g(t) = \prod_{l=1}^{n} |\widehat{\nu}(t_l)| \leq \exp(-\frac{1}{2} \sum_{l=1}^{n} 1 - |\widehat{\nu}(t_l)|^2) = \exp(-\frac{1}{2} h(t)) \leq \exp(-\frac{1}{2}\alpha_n' \delta n) = \varepsilon.$$

We thus have found an element $t \in H$ which lies outside $G(\varepsilon)$. □

Let $u \geq \varepsilon$. By Proposition A.24, $\mathrm{Sym}(2F(u)) \subset 4F(u) \subset G(u)$. Thus by Claim A.25, the additive subgroup $\mathrm{Sym}(2F(u))$ must be trivial, and so by Corollary A.20

$$2|F(u)| \leq |\mathrm{Sym}(F(u)+F(u))| + |F(u)+F(u)| \leq 1 + |G(u)|.$$

It thus follows that for all $u \geq \varepsilon$ we have $2|F'(u)| \leq |G'(u)|$. So

$$\int_{\varepsilon}^{\infty} |F'(u)|du + \varepsilon|F'(\varepsilon)| \leq \frac{1}{2}(\int_{\varepsilon}^{\infty} |G'(u)|du + \varepsilon|G'(\varepsilon)|),$$

completing the proof of Lemma A.15. □

References

[1] A. Basak and M. Rudelson, Invertibility of sparse non-Hermitian matrices. Advances in Mathematics, 310, 426483, 2017.

[2] J. Bourgain, V. Vu and P. M. Wood, On the singularity probability of discrete random matrices. Journal of Functional Analysis 258 (2010), no.2, 559-603.

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

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

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

[6] M. Rudelson and R. Vershynin, The Littlewood-Offord Problem and invertibility of random matrices. Advances in Mathematics 218 (2008), 600-633.

[7] S. Koplewitz, The corank of a rectangular random integer matrix, `arxiv.org/abs/1611.06441`.

[8] S. Koplewitz, Random Graphs, Sandpile Groups, and Surjectivity of Random Matrices. Thesis (Ph.D.)-Yale University. 2017. 58 pp. ISBN: 978-0355-01826-4.

[9] K. Maples, Singularity of Random Matrices over Finite Fields, `arxiv.org/abs/1012.2372`.

[10] K. Maples, Cokernels of random matrices satisfy the Cohen-Lenstra heuristics, `arxiv.org/abs/1301.1239`.

[11] K. Maples, Arithmetic Properties of Random Matrices, Thesis (Ph.D.)-University of California, Los Angeles. 2011. 77 pp. ISBN: 978-1267-18973-8.

[12] H. Nguyen, Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. Duke Mathematics Journal Vol. 161, 4 (2012), 545-586.

[13] R. Stanley, Enumerative Combinatorics, vol. 1. Cambridge University Press. 1997.

[14] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices. Journal of the A. M. S, 20 (2007), 603-673.

[15] T. Tao and V. Vu, Additive combinatorics, Volume 105, Cambridge University Press, Cambridge, 2006.

[16] R. Vershynin, Invertibility of symmetric random matrices, Random Structures & Algorithms, 44 (2014), no. 2, 135-182.

[17] M. M. Wood, The distribution of sandpile groups of random graphs, Journal of the A. M. S, 30 (2017), pp. 915-958.

[18] M. M. Wood, Random integral matrices and the Cohen-Lenstra Heuristics, `arxiv.org/abs/1504.04391`.

E-mail address: `nguyen.1261@math.osu.edu`

E-mail address: `paquette.30@math.osu.edu`

Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, OH 43210