COKERNELS OF RANDOM MATRICES SATISFY THE COHEN-LENSTRA HEURISTICS

KENNETH MAPLES

ABSTRACT. Let A be an $n \times n$ random matrix with iid entries taken from the p-adic integers $\mathbb{Z}_p$ or $\mathbb{Z}/N\mathbb{Z}$. Then under mild non-degeneracy conditions the cokernel of A has a universal probability distribution. In particular, the p-part of a random matrix over $\mathbb{Z}$ has cokernel distributed according to the Cohen-Lenstra measure,

$$\mathbb{P}(\operatorname{coker} A \cong G) = \frac{1}{|\operatorname{Aut} G|} \prod_{k=1}^{\infty} (1 - p^{-k}) + O(e^{-cn})$$

where the constants are absolute.

1. INTRODUCTION

The Cohen-Lenstra measure is a probability distribution on isomorphism classes of finite abelian p-groups, given by

$$\nu(G) := \frac{1}{|\operatorname{Aut} G|} \prod_{k=1}^{\infty} (1 - p^{-k}).$$

It appears in various guises as a candidate for a random abelian p-group. Its best known appearance is in the conjectural work of Cohen and Lenstra [2], where based on numerical evidence they conjectured that the p-part of the ideal class group of quadratic number fields are distributed according to ν. These conjectures were extended to other number fields, in particular by Malle [6], as long as the number field does not contain pth roots of unity.

It was observed in [3] that if $B : \mathbb{Z}_p^n \to \mathbb{Z}_p^n$ is a random matrix with iid entries chosen according to Haar measure, then the cokernel distribution of B converges to the Cohen-Lenstra measure as $n \to \infty$. In this article we show that this property holds for random matrices in a much larger class.

Let $\xi \in \mathbb{Z}_p$ be a random variable. We define the concentration constant $\alpha = \alpha(\xi)$ to be the largest value $0 < \alpha < 1$ such that we have the non-degeneracy condition

$$\sup_{t \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\xi = t \bmod p) \leq 1 - \alpha.$$

If $A \in M(n, \mathbb{Z}_p)$ is an iid random matrix whose entries have concentration constant α, then we say that A has concentration constant α as well.

We have the following universality principle for random matrices with iid entries in the p-adic integers.

Theorem 1.1. *Let $A \in M(n, \mathbb{Z}_p)$ be a discrete random matrix with concentration constant α. Then for all finite abelian p-groups G,*

$$\mathbb{P}(\operatorname{coker} A \cong G) = \frac{1}{|\operatorname{Aut} G|} \prod_{k=1}^{\infty} (1 - p^{-k}) + O(e^{-c\alpha n})$$

where the constants are absolute.

Date: 1/7/13.

1991 *Mathematics Subject Classification.* Primary 15A52; Secondary 15A33, 60C05.

The author was partially supported by a Graduate Research Fellowship from the National Science Foundation.

We can also control the probability distribution of the cokernel of matrices over other rings. Let N denote a positive integer with $\omega = \omega(N)$ distinct prime factors. Let $B : (\mathbb{Z}/N\mathbb{Z})^n \to (\mathbb{Z}/N\mathbb{Z})^n$ be chosen uniformly from all $\mathbb{Z}/N\mathbb{Z}$-module morphisms and let A be a random $n \times n$ matrix over $\mathbb{Z}/N\mathbb{Z}$ with iid entries distributed according to a random variable ξ. We define the concentration constant $\alpha(\xi)$ analogously to the p-adic case; namely, $0 < \alpha < 1$ is the largest number such that

$$\sup_{p|N} \sup_{t \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(x = t \bmod p) \leq 1 - \alpha.$$

Then we have the following universality result.

Theorem 1.2. *Let $A \in M(n, \mathbb{Z}/N\mathbb{Z})$ be a discrete random matrix with concentration constant α. Then for all finite $\mathbb{Z}/N\mathbb{Z}$-modules G,*

$$\mathbb{P}(\operatorname{coker} A \cong G) = \mathbb{P}(\operatorname{coker} B \cong G) + O_\omega(e^{-c\alpha n})$$

where the constants depend only on $\omega = \omega(N)$, the number of distinct prime factors of N.

As a corollary, we have the following generalization of Theorem 1.2 from [7] to control the rank of random matrices over $\mathbb{Z}/p\mathbb{Z}$ for primes p.

Corollary 1.3. *Let $A \in M(n, \mathbb{Z}/p\mathbb{Z})$ be a discrete random matrix with concentration constant α. Then we have*

$$\mathbb{P}(\operatorname{rank} A = n - k) = p^{-k^2} \frac{\prod_{\ell=k+1}^{\infty}(1 - p^{-\ell})}{\prod_{\ell=1}^{k}(1 - p^{-\ell})} + O(e^{-c\alpha n})$$

for all $0 \leq k \leq n$, where the constants are absolute.

For now let $R = \mathbb{Z}_p$ or $\mathbb{Z}/N\mathbb{Z}$ and let $X_1, ..., X_n$ denote the columns of A. The main idea behind Theorem 1.1 and 1.2 is to expose the columns one by one and compute the probability distribution of the successive quotients

$$R^n / \langle X_{k+1}, ..., X_n \rangle$$

where $\langle X_{k+1}, ..., X_n \rangle$ denotes their span as R-modules. Conditioning on the isomorphism class of the previous quotients, we can partition R^n into sets according to the resulting class of the quotient by X_k. It turns out that these sets can be written as the set-theoretic difference of R-submodules of R^n that are constructed in a natural way from the span $\langle X_{k+1}, ..., X_n \rangle$. Modulo some simple technical restrictions, it suffices to consider the probability that X_k lies in the random submodule $\phi(\langle X_{k+1}, ..., X_n \rangle)$, where ϕ is a natural deterministic function that takes R-submodules of R^n to R-submodules of R^n. We will call such $\phi(\langle X_{k+1}, ..., X_n \rangle)$ enlarged submodules.

The analysis of enlarged submodules relies on the swapping lemma of Tao and Vu [10] [11] which the author subsequently adapted to finite fields in [7]. This allows us to show that enlarged submodules N where $\mathbb{P}(X_k \in N)$ deviates significantly from $|N^\perp|^{-1}$ can be induced with much higher probability by a different probability distribution.

For the remaining submodules, we still have $\mathbb{P}(X_k \in N)$ deviating slightly from uniform. In this setting we generalize the inverse theorem of [7] to show that $N^\perp$ contains a "structured" vector. Since we can assume that this probability is within a constant factor of $|N^\perp|^{-1}$, it is possible to explicitly enumerate all such vectors.

The key property of A used in the proof is the martingale independence of its columns. It should be possible to prove a version of this theorem with much weaker conditions on the construction of A. However the independence of the entries is crucial to the argument and it is unclear how to avoid this obstruction with current technology.

Although we only work over $R = \mathbb{Z}_p$ and $\mathbb{Z}/N\mathbb{Z}$ in this article, the argument is amenable to much more general analysis. We do not attempt maximum generality to avoid unnecessary complexity for the most common applications. It would, however, be interesting to extend Theorem 1.1 to the ring of integers of finite extensions of $\mathbb{Q}_p$. These arguments suffice when we

assume that the probability distribution on the entries, taken modulo the maximal ideal $\mathfrak{m}$, are "non-degenerate" in the sense that they do not concentrate on an additive coset of $\mathbb{F}_{p^f} \cong R/\mathfrak{m}$. However, it is intuitively clear that it should suffice, at least for bounded exponents f, for the probability distribution to not concentrate on affine *subfields*; i.e. subsets of the form $\beta\mathbb{F}_{p^d} + \gamma$ for some $d \mid f$. This limitation will be considered in forthcoming work.

From Theorem 1.2 we can recover the torsion-free results of Tao and Vu [10] as well as the mod-p results of Charlap, Rees, and Robbins [1] and Kahn, Komłós [5]. As mentioned above, we would need control over rings with non-prime residue fields $R/\mathfrak{m}$ for a full generalization.

2. Notation

We will use standard asymptotic notation. For an index variable n and functions f, g of n, we write $f = O(g)$ to mean that there are absolute constants n_0 and C such that for all $n > n_0$, $f(n) \leq Cg(n)$. These constants may change from line to line. Similarly, the equation $f = g + O(h)$ is shorthand for $|f - g| = O(h)$. We use the notations $f \lesssim g$ and $g \gtrsim f$ to denote $f = O(g)$ when convenient.

It is convenient to employ probabilistic notation. If E is an event, we let $\mathbb{P}(E)$ denote its probability; if F is another event, then we can express the conditional probability of E on F as $\mathbb{P}(E \mid F)$ and their intersection as $\mathbb{P}(E \wedge F)$. For random variables X we let $\mathbb{E}X$ denote their expectation.

For V an R-submodule of R^n, we define

$$V[t] := \{v \in R^n \mid tv \in V\}$$

and by abuse of notation

$$V[\infty] := \{v \in R^n \mid tv \in V \text{ for some } t \in R\}.$$

We let (a) denote the ideal in R generated by a; we will also let (a) denote the principal ideal aR^n in the free module R^n. For positive integers m, n we will let $[m, n]$ denote their least common multiple.

We will use the number theorist's exponential function $e(t) := \exp(2\pi it)$ and $e_p(t) := \exp(2\pi it/p)$.

If A is a set we will let $\#A$ and $|A|$ both denote its cardinality.

3. The column exposure process

Consider the set of $n \times n$ matrices over the finite field $\mathbb{F}_p$. It is well-known that the proportion of invertible matrices in this set can be calculated from the sequence of column vectors $X_1, \ldots, X_n$ and the associated sequence of subspaces

$$W_\ell := \langle X_{\ell+1}, \ldots, X_n \rangle \subseteq \mathbb{F}_p^n$$

for $\ell = 0, \ldots, n$; we have chosen to consider subspaces starting with X_n for the most convenient normalization. In fact, we see that the entire matrix is invertible if and only if $\operatorname{codim} W_\ell = n - \dim W_\ell = \ell$ for each ℓ. If we let A denote a random $n \times n$ matrix over $\mathbb{F}_p$ chosen uniformly, so that the columns $X_1, \ldots, X_n$ are independent random vectors taken uniformly from $\mathbb{F}_p^n$, then we compute

$$\begin{aligned}\mathbb{P}(\operatorname{codim} W_{\ell-1} = \ell - 1 \mid \operatorname{codim} W_\ell = \ell) &= 1 - \mathbb{P}(X_\ell \in W_\ell \mid \operatorname{codim} W_\ell = \ell) \\ &= 1 - p^{-\ell}\end{aligned}$$

so, in particular,

$$\mathbb{P}(A \text{ is invertible}) = \prod_{\ell=1}^{n} (1 - p^{-\ell}).$$

We can generalize this computation to compute the cokernel of a random $n \times n$ matrix A over the ring R, given by

$$\operatorname{coker} A := R^n / \langle X_1, ..., X_n \rangle.$$

We consider the sequence of submodules

$$W_\ell := \langle X_{\ell+1}, ..., X_n \rangle$$

as $\ell = 0, \ldots, n$. We have the natural quotient map

$$\phi_\ell : R^n/W_\ell \longrightarrow R^n/W_{\ell-1}$$

with kernel equal to $\langle X_\ell \rangle$, the span of X_ℓ in R_n/W_ℓ.

In this section we show that the isomorphism class of these quotients can be computed by testing the membership of the column vectors X_ℓ in various submodules constructed in a natural way from W_ℓ. We define an enlarged submodule to be a submodule of the form

$$\bigcap_{(a,b)\in\mathcal{F}} (a) + W_\ell[b]$$

where $\mathcal{F}$ is a finite set of pairs in $R \times R$. Since enlarged submodules are functorial in W_ℓ, we will typically denote the enlarged submodule by $\phi(W_\ell)$.

The following universality result forms the heart of the argument.

Proposition 3.1. *Let $\xi \in R$ be a random variable with concentration constant α. Let $k \leq \eta n$. Then for every enlarged submodule $\phi(W_k)$, conditioning on the isomorphism class*

$$R^n/\phi(W_k) \cong V^\perp$$

for some R-submodule V, we have

$$|\mathbb{P}(X_k \in \phi(W_k)) - |V^\perp|^{-1}| \leq O(e^{-c\alpha n})$$

Here $0 < \eta < 1$, $c > 0$ and the implied constant are absolute.

The bulk of this chapter concerns the proof of Proposition 3.1. However it remains to show that it suffices to consider enlarged submodules.

The partial quotients contain information about the syzygies of $X_{\ell+1}, ..., X_n$ over quotients of R. We first observe that for $R = \mathbb{Z}_p$ we can extract the free part of $\mathbb{Z}_p^n/W_\ell$ almost surely.

Proposition 3.2. *For $R = \mathbb{Z}_p$ we have $\mathbb{P}(X_\ell \in W_\ell[\infty]) = O(e^{-c\alpha n})$.*

Proof. For all $L > 0$, by Proposition 3.1 we have

$$\mathbb{P}(X_\ell \in W_\ell[p^L]) = p^{-L} + O(e^{-c\alpha n}).$$

Now send $L \to \infty$ and note that $W_\ell[\infty] = \bigcap_{L=1}^\infty W_\ell[p^L]$. □

Proposition 3.3. *For $R = \mathbb{Z}/N\mathbb{Z}$ we can factor*

$$R^n/W_\ell = \oplus_{p|N} R_{(p)}^n/(W_\ell)_{(p)}$$

where $R_{(p)}$, $(W_\ell)_{(p)}$ denote the p-th part of R, W_ℓ respectively.

Proof. This is immediate from the Chinese remainder theorem. □

Proposition 3.4. *With probability $1 - O(e^{-c\alpha n})$ there are isomorphisms*

$$R^n/W_\ell \cong R^\ell \oplus T_\ell$$

where T_ℓ is a uniquely defined finite R-module.

Proof. **Case $R = \mathbb{Z}_p$.** We induct downward on ℓ starting with $\ell = n$. For $\ell = n$ the result is trivial with $T_n = 0$. If the result is shown for ℓ, then it suffices to find a finite $\mathbb{Z}_p$-module $T_{\ell-1}$ such that $\mathbb{Z}_p^{\ell-1} \oplus T_{\ell-1} \cong \mathbb{Z}_p^\ell \oplus T_\ell/\langle X_\ell \rangle$. Let $v_1, ..., v_\ell$ be some basis for $\mathbb{Z}_p^\ell$ and write

$$X_\ell = a_1 v_1 + \cdots + a_\ell v_\ell + t$$

with $t \in T_\ell$; this decomposition is unique. By Proposition 3.2 we know that with probability $1 - O(e^{-c\alpha n})$ at least one of the coefficients $a_1, \ldots, a_\ell$ is non-zero. Suppose without loss of generality that the power of p dividing a_ℓ is the smallest among $a_1, \ldots, a_\ell$. Consider the short exact sequence

$$0 \longrightarrow \mathbb{Z}_p^{\ell-1} \xrightarrow{\iota} \mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell\rangle \xrightarrow{\pi} \mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell, v_1, ..., v_{\ell-1}\rangle \longrightarrow 0$$

where ι is the inclusion map on $v_1, ..., v_{\ell-1}$ and π is the indicated quotient map. It is easy to see that $\mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell, v_1, ..., v_{\ell-1}\rangle$ is finite: each element of T_ℓ is torsion, while the generators of $\mathbb{Z}_p^{\ell}$ either map to 0 or are torsion. Let

$$T_{\ell-1} := \mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell, v_1, ..., v_{\ell-1}\rangle.$$

By the splitting lemma, it suffices to define a map of $\mathbb{Z}_p$-modules $\psi : \mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell\rangle \to \mathbb{Z}_p^{\ell-1}$ such that $\psi \circ \iota$ is the identity on $\mathbb{Z}_p^{\ell-1}$.

Indeed, for $w \in \mathbb{Z}_p^{\ell} \oplus T_\ell/\langle X_\ell\rangle$ choose a representative

$$w = b_1 v_1 + \cdots + b_\ell v_\ell + t'$$

where $t' \in T_\ell$. Now $b_\ell = p^r\beta$ for some $r \geq 0$ and $\beta \in \mathbb{Z}_p^\times$; similarly let $a_\ell = p^s\alpha$ for some $s \geq 0$ and $\alpha \in \mathbb{Z}_p^\times$. If $r \geq s$ then $b_\ell/a_\ell \in \mathbb{Z}_p$ and so there is a unique representative

$$w - \frac{b_\ell}{a_\ell} X_\ell \in \langle v_1, \ldots, v_{\ell-1}\rangle \oplus T_\ell.$$

We then let $\psi(w)$ be the projection onto $\langle v_1, ..., v_{\ell-1}\rangle$ of this representative. If $r < s$, then $p^{s-r}b_\ell/a_\ell \in \mathbb{Z}_p$ and so there is a unique representative of $p^{s-r}w$ given by

$$p^{s-r}w - \frac{p^{s-r}b_\ell}{a_\ell} X_\ell \in \langle v_1, \ldots, v_{\ell-1}\rangle \oplus T_\ell.$$

Furthermore, the coefficients of $v_1, \ldots, v_{\ell-1}$ of this representative must all be divisible by p^{s-r} (this is because a_ℓ generates the largest ideal), so we let $\psi(w)$ be $p^{-(s-r)}$ times the projection onto $\langle v_1, ..., v_{\ell-1}\rangle$ of this representative. It is clear that the resulting map is linear and commutes with multiplication by elements of $\mathbb{Z}_p$, as required.

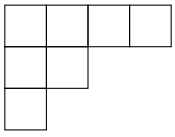
Case $R = \mathbb{Z}/N\mathbb{Z}$. By the chinese remainder theorem it suffices to prove the result for $N = p^r$ for some $r \geq 1$. But this follows from the previous case. □

We now have a sequence of finite R-modules $T_n, \ldots, T_0$ with $T_n = 0$ and $T_0 = \operatorname{coker} A$. For any matrix A (with probability $1 - O(e^{-c\alpha n})$) this sequence is uniquely defined. Therefore, for $H_n, \ldots, H_0$ any sequence of finite R-modules with $H_n = 0$ and $H_0 = \operatorname{coker} A$ we can write

$$\mathbb{P}(\operatorname{coker} A \cong G) = \sum_{H_{n-1},\ldots,H_1} \mathbb{P}(\operatorname{coker} A \cong G \text{ and } T_j \cong H_j \text{ for all } j) + O(e^{-c\alpha n}).$$

Next we will consider what sequences $H_n, \ldots, H_0$ can arise as the torsion parts of the partial quotients R^n/W_ℓ.

3.1. **Classifying partial cokernels.** We can classify finite R-modules in terms of Young diagrams. Recall that a **Young diagram** (or **diagram**) is a graphical display of a partition, where to the set $[m]$ and the partition $j_1 \geq j_2 \geq \cdots \geq j_t > 0$ with $j_1 + \cdots + j_t = m$ we have the corresponding Young diagram with t rows, where in the kth row there are j_k boxes. For example, the partition of [7] into $4 \geq 2 \geq 1$ corresponds to the Young diagram



For $R = \mathbb{Z}_p$, any finite $\mathbb{Z}_p$-module T is a finite abelian p-group; by the fundamental theorem of finite abelian groups, we know that T is isomorphic to

$$T \cong (\mathbb{Z}/p^{j_1}\mathbb{Z}) \oplus (\mathbb{Z}/p^{j_2}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p^{j_t}\mathbb{Z})$$

with $j_1 \geq j_2 \geq \cdots \geq j_t > 0$, and this representation is unique. We therefore have a one-to-one correspondence between partitions $j_1 \geq j_2 \geq \cdots \geq j_t > 0$ of $[\log_p M]$ and finite $\mathbb{Z}_p$-modules of cardinality M.

For $R = \mathbb{Z}/N\mathbb{Z}$, if we factor $R = \oplus_{p^k \| N} \mathbb{Z}/p^k\mathbb{Z}$ then we have a correspondence between finite R-modules T and tuples of Young diagrams, one for each prime dividing N, such that there are at most k columns in the Young diagram for p if $p^k \| N$. We will call the diagram of T corresponding to the prime p the **p-diagram** of T.

Recall that the matrix A induces a sequence $T_n, \ldots, T_0$ of finite R-modules with probability $1 - O(e^{-c\alpha n})$. It turns out this sequence is strictly growing in the sense that $T_\ell \hookrightarrow T_{\ell-1}$.

Proposition 3.5. *With probability $1 - O(e^{-c\alpha n})$ there is an injection $T_\ell \hookrightarrow T_{\ell-1}$.*

Proof. We see that $\phi_\ell : R^n/W_\ell \to R^n/W_{\ell-1}$ commutes with the isomorphisms from Proposition 3.4 to give a map

$$\phi_\ell : R^\ell \oplus T_\ell \to R^{\ell-1} \oplus T_{\ell-1}$$

with kernel equal to $\langle X_\ell \rangle$, the span of the image of X_ℓ in $R^\ell \oplus T_\ell$.

If $R = \mathbb{Z}_p$ then by Proposition 3.2 we have $X_\ell \notin T_\ell$ with probability $1 - O(e^{-c\alpha n})$. In particular $\langle X_\ell \rangle \cap T_\ell = \{0\}$, so ϕ_ℓ is injective when restricted to T_ℓ. Since all elements of T_ℓ have finite order their images must have finite order as well, and in particular must be contained in $T_{\ell-1}$. Thus we have $\phi_\ell : T_\ell \hookrightarrow T_{\ell-1}$ as required.

If $R = \mathbb{Z}/N\mathbb{Z}$, it suffices to prove that $(T_\ell)_{(p)} \hookrightarrow (T_{\ell-1})_{(p)}$. This is obvious after changing coordinates by the classification theorem for finite abelian p-groups. □

For any finite R-module T, we define the j-rank of T at p to be

$$r_{j,(p)}(T) := \operatorname{rank}_{\mathbb{F}_p}(p^{j-1}T/p^jT).$$

From the correspondence we see that $r_{j,(p)}$ is equal to the length of the jth column of the p-diagram of T. It turns out that there is considerable rigidity in the change of $r_{j,(p)}(T_\ell)$ as ℓ varies.

Proposition 3.6. *For all $j \geq 1$, $1 \leq \ell \leq n$, and p we have*

$$r_{j,(p)}(T_\ell) \leq r_{j,(p)}(T_{\ell-1}) \leq r_{j,(p)}(T_\ell) + 1.$$

Furthermore, $r_{j,(p)}(T_{\ell-1}) = r_{j,(p)}(T_\ell) + 1$ if and only if

$$\langle X_\ell \rangle \cap p^{j-1}(R^\ell \oplus T_\ell) \bmod p^j = \{0\}$$

Proof. Let $\phi_\ell : R^\ell \oplus T_\ell \to R^{\ell-1} \oplus T_{\ell-1}$ be the map from the proof of Proposition 3.5. We can restrict ϕ_ℓ to the submodule $p^{j-1}(R^\ell \oplus T_\ell)$ to derive a map $\phi'_\ell : p^{j-1}(R^\ell \oplus T_\ell) \to p^{j-1}(R^{\ell-1} \oplus T_{\ell-1})$. We can then quotient by p^j and note that $p^j(R^\ell \oplus T_\ell)$ maps to zero to derive a map

$$\widetilde{\phi} : p^{j-1}(R^\ell \oplus T_\ell)/p^j(R^\ell \oplus T_\ell) \to p^{j-1}(R^{\ell-1} \oplus T_{\ell-1})/p^j(R^{\ell-1} \oplus T_{\ell-1})$$

This is a map of finite dimensional vector spaces over $\mathbb{F}_p$. The quotients factor over the sums so we can write

$$\widetilde{\phi} : \mathbb{F}_p^\ell \oplus (p^{j-1}T_\ell/p^jT_\ell) \to \mathbb{F}_p^{\ell-1} \oplus (p^{j-1}T_{\ell-1}/p^jT_{\ell-1})$$

$\widetilde{\phi}$ is the composition of two quotient maps so it is surjective. Therefore, by the definition of the j-rank, we have

$$\ell + r_j(T_\ell) \geq \ell - 1 + r_j(T_{\ell-1}).$$

The other inequality follows immediately from Proposition 3.5.

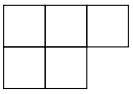
For the second statement, note that $\widetilde{\phi}$ is an isomorphism if and only if the kernel of ϕ restricted to $p^{j-1}(R^\ell \oplus T_\ell)$ is contained in $p^j(R^\ell \oplus T_\ell)$. □

This last proposition characterizes sequences $H_n, \dots, H_0 = \operatorname{coker} A$ which arise from sequences $T_n, \dots, T_0 = \operatorname{coker} A$. We first recall some standard notation on Young tableaux.

On a Young diagram λ, we define a **numbering** of λ from $[n]$ to be an assignment, for each box in the diagram, of an integer from $[n]$. We then say that λ is the **shape** of the numbering. We define a **semi-standard tableau** of shape λ to be a numbering which is weakly decreasing along rows and strictly decreasing along columns; for example,

4	3	3	1
2	2		
1			

is a tableau of shape $\lambda = 4 \geq 2 \geq 1$. We will write H/λ to indicate that H is a Young tableau of shape λ. Note that we have swapped the usual convention of increasing rows and columns for later notational convenience. Finally, to a tableau H/λ we define the **sub-diagram of H above ℓ** to be the diagram consisting of those boxes of H whose labels j satisfy $j \geq \ell$. For example, the sub-diagram of the tableau above at 2 is



We will denote the sub-diagram of a tableau H at ℓ by H_ℓ.

Young tableaux are a convenient notation for sequences of partial cokernels, as the following corollary illustrates.

Corollary 3.7. *Let A be an $n \times n$ matrix over R and let λ denote the diagram of* $\operatorname{coker} A$. *Then there is a one-to-one correspondence between sequences $T_n, \dots, T_0$ with $T_n = \{0\}$ and $T_0 = \operatorname{coker} A$ which can occur as the partial cokernels of A and tableaux H/λ from $[n]$, such that T_ℓ has diagram equal to H_ℓ.*

Proof. Fix the sequence $T_n, \dots, T_0$. By Proposition 3.5 we see that the diagrams λ_ℓ associated to T_ℓ are nested, so define H/λ to be the numbering of λ given by the time the box appears in the nested sequence. Clearly the rows of λ are weakly increasing. Proposition 3.6 shows that the columns increase by at most one each step, so there cannot be two adjacent equal numbers in a vertical configuration. The converse is obvious. □

3.2. **Enlarged submodules.** We now see that we can expand

$$\mathbb{P}(\operatorname{coker} A \cong G) = \sum_{H/\lambda \text{ from } [n]} \mathbb{P}(T_\ell \cong H_\ell \text{ for all } \ell) + O(e^{-c\alpha n}).$$

Each term in the sum can be expanded into a product by conditional expectation; namely,

$$\mathbb{P}(T_\ell \cong H_\ell \text{ for all } \ell) = \prod_{\ell=0}^{n} \mathbb{P}(T_\ell \cong H_\ell \mid T_j \cong H_j \text{ for all } j > \ell).$$

Each term in the product is not yet in a form that we can estimate. In particular, we would like to represent the set of X_ℓ such that $T_{\ell-1}$ has the desired isomorphism class as the set-theoretic difference of certain submodules of R^n.

Consider the two-parameter family of events

$$G_{a,b} := \{X_\ell \in p^a R^n + W_\ell[p^b]\}$$

with $a \geq 1$ and $b \geq 0$. We have the obvious inclusions

$$G_{a+1,b} \subseteq G_{a,b} \subseteq G_{a,b+1}$$

so we can define events

$$E_{j,0} := G_{j,0}$$

and

$$E_{j,t} := G_{j-t,t} \setminus G_{j-t,t-1}.$$

for $1 \leq j$ and $1 \leq t \leq j-1$. The events $E_{j,t}$ as t varies can be used to test the j-rank of the p-primary part of T as follows.

Proposition 3.8. *The events $E_{j,0}, \ldots, E_{j,j-1}$ are disjoint, and the equality*

$$r_{j,(p)}(T_{\ell-1}) = r_{j,(p)}(T_\ell) + 1$$

holds if and only if $E_{j,0} \cup \cdots \cup E_{j,j-1}$ holds.

Proof. We have $E_{j,t} \subseteq G_{j-t,t} \subseteq G_{g-t-s,t+s-1}$ for all $s \geq 1$, but $E_{j,t+s} = G_{g-t-s,t+s} \setminus G_{g-t-s,t+s-1}$, so the family is pairwise disjoint.

First we observe that the condition $r_{j,(p)}(T_{\ell-1}) = r_{j,(p)}(T_\ell)+1$ is equivalent to $\langle X_\ell \rangle \cap p^{j-1}(R^\ell \oplus T_\ell) \bmod p^j = \{0\}$ by Proposition 3.6, so it suffices to show that $E_{j,0} \cup \cdots \cup E_{j,j-1}$ is equivalent to this event.

Suppose that $\langle X_\ell \rangle \cap p^{j-1}(R^\ell \oplus T_\ell) \bmod p^j = \{0\}$. Let $t \geq 0$ be the minimum exponent such that $p^t X_\ell \in p^j R^n + W_\ell$, or equivalently, $X_\ell \in p^{j-t} + W_\ell[p^t]$. If $t = 0$ then $E_{j,0}$ holds; otherwise, $t > 0$ and thus $p^{t-1} X_\ell \notin p^j R^n + W_\ell$. But this implies that $p^{t-1} X_\ell \notin p^{j-1} R^n + W_\ell$, or equivalently $X_\ell \notin p^{j-t} R^n + W_\ell[p^{t-1}]$. Therefore $E_{j,t} = G_{j-t,t} \setminus G_{j-t,t-1}$ holds. It remains to show that $t < j$, but if $t = j$ we would have shown that $p^{t-1} X_\ell \notin p^{j-1} R^n + W_\ell$ which is absurd.

Conversely, suppose $E_{j,t}$ holds. If $t = 0$ then $X_\ell \in p^j R^n + W_\ell$ so $\langle X_\ell \rangle = \{0\} \bmod p^j R^n + W_\ell$. If $t > 0$ then $p^{t-1} X_\ell \notin p^{j-1} R^n + W_\ell$ which implies that $\langle X_\ell \rangle \cap p^{j-1} R^n + W_\ell \subseteq \langle p^t X_\ell \rangle$. However, $p^t X_\ell \in p^j + W_\ell$ so $\langle X_\ell \rangle \cap p^{j-1} R^n + W_\ell = \{0\} \bmod p^j$. □

We have therefore reduced the problem of computing the isomorphism class of $T_{\ell-1}$ conditioned on W_ℓ to testing the membership of

$$X_\ell \in p^a R^n + W_\ell[p^b]$$

for various a, b, and p.

Precisely, we have the following. Fix ℓ and let $S \subseteq \mathbb{Z}^+$ denote those indices where $r_k(T_{\ell-1}) = r_k(T_\ell) + 1$. By Proposition 3.2, S is finite with probability $1 - O(e^{-c\alpha n})$. We define random variables $V_{a,b}$ for $a \geq 1$ and $b \geq 0$ by

$$V_{a,b} := \begin{cases} G_{a,0} - G_{a+1,0}, & b = 0 \\ G_{a,b} - G_{a+1,b} - G_{a,b-1} + G_{a+1,b} \cap G_{a,b-1}, & b > 0 \end{cases}.$$

We can then compute the following.

Proposition 3.9. *Fix $\ell \geq 1$ and tableau H/λ. Let $S \subseteq \mathbb{Z}^+$ denote as above the set of columns of H which contain the number ℓ. Then we have the equality of events*

$$\{T_\ell \cong H_\ell\} = \prod_{k \in S} \sum_{t=0}^{k-1} V_{k-t,t}$$

Proof. By Proposition 3.8, $k \in S$ if and only if $E_{k,0} \cup \cdots \cup E_{k,k-1}$ holds.

First suppose $S = \{k\}$ for some k. In this case we want to compute the event that some $G_{k-t,t}$ holds for some $0 \leq t \leq k-1$ while no other $G_{a,b}$ holds. By the nesting property of $G_{a,b}$ and the inclusion-exclusion principle this is equal to $V_{k-t,t}$. The case for larger S follows by induction and the inclusion-exclusion principle. □

Note that if the product in Proposition 3.9 is expanded over the sum, then some of the terms will cancel identically.

The various submodules $(p^j) + W_\ell[p^k] \subseteq R^n_{(p)}$ for $j \geq 1$ and $k \geq 0$ form a grid of inclusions,

$$\begin{array}{ccc} (p^{j+1}) + W_\ell[p^k] & \hookrightarrow & (p^{j+1}) + W_\ell[p^{k+1}] \\ \downarrow & & \downarrow \\ (p^j) + W_\ell[p^k] & \hookrightarrow & (p^j) + W_\ell[p^{k+1}] \end{array}$$

We abbreviate

$$G_{j,k} := \text{“}X_\ell \in (p^j) + W_\ell[p^k]\text{”}$$

and observe that

$$E_{j,0} = G_{j,0}$$

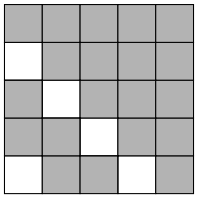
and

$$E_{j,t} = G_{j-t,t} \setminus G_{j-t,t-1}.$$

Let $\eta : \mathbb{Z}^+ \to \{0,1\}$ be the function such that

$$r_j(H_{\ell-1}) = r_j(H_\ell) + \eta(j).$$

We can associate a diagram to η by considering a grid with indices $j \geq 1$, $k \geq 0$ such that (j,k) is shaded if and only if $\eta(j+k) = 0$. For example, if $\operatorname{supp}\eta = \{1,4\}$ we have the diagram



To check that $T_{\ell-1} \cong H_{\ell-1}$ conditioned on $T_\ell \cong H_\ell$, we observe the following properties

Proposition 3.10. (1) *At most one of the events $E_{j,t}$ can hold for j fixed.*
(2) *If none of the events $E_{j,0}, ..., E_{j,t}$ hold, then no $E_{j',s}$ holds for $j' \geq j$ and $s \leq t$.*

Proof. (1) Since $E_{j,t} = G_{j,t} \setminus G_{j,t-1}$, by the inclusion $G_{j,s} \subseteq G_{j,t-1}$ for $s < t$ the conclusion follows.
(2) This follows from $E_{j',s} \subseteq G_{j',s} \subseteq G_{j,t} = \cup_{k=0}^{t} G_{j,k}$. □

We now show that these two conditions imply that there is a unique minimal enlarged submodule which contains the set which induces H_ℓ.

Algorithm 1. The minimal enlarged submodule $\phi(W_\ell)$ which contains the set which induces H_ℓ corresponds to the value of $\mathcal{F}$ when the following algorithm terminates.

(1) Initialize $\mathcal{F} = \emptyset$, $j = 1$, $i = 0$.
(2) Find the minimal $t \geq i$ such that (j,t) is clear in the diagram. If none, then STOP.
(3) Set $\mathcal{F}' := \mathcal{F} \cup \{(j,t)\}$ and relabel $\mathcal{F}$ to $\mathcal{F}'$.
(4) Increment j.
(5) Jump to 2.

We finally observe that for each $(j,t) \in \mathcal{F}$ with $t > 0$, it suffices to subtract the probability for the enlarged submodule $\mathcal{F} \cup \{j, t-1\}$; the correct value is then found by the inclusion-exclusion principle.

3.3. **Entropy bounds.** For $\ell > \delta n$ we observe that any non-trivial column would require that $X_\ell \in (p) + W[\infty]$ for some prime p. However, by Odlyzko's lemma, first recorded in [8], this occurs with probability $O(e^{-c\alpha n})$ so these columns can be ignored.

Lemma 3.11 (Odlyzko)**.** *For any fixed subspace V of $\mathbb{F}_q^n$ and random vector $X \in \mathbb{F}_q^n$ with concentration constant α, we have the bound*

$$\mathbb{P}(X \in V) \leq (1-\alpha)^{\operatorname{codim} V}.$$

Proof of Lemma 3.11. Let k denote the codimension of V. We can find $n-k$ coordinates $\tau \subseteq [n]$ such that V is a graph over τ. If we condition on the coordinates of X in τ, then there is a unique choice for the remaining coordinates $[n] \setminus \tau$ for $X \in V$. Since μ has concentration constant α, the probability that each entry of X assumes the required value is bounded by $1-\alpha$, and the result follows from the independence of the entries. □

For G sufficiently small, i.e. $\log_p|G| = O(1)$ for all p, Theorem 1.1 and Theorem 1.2 follow immediately from Theorem 3.1 and the decomposition into Young tableaux. However, for G sufficiently large there are a super-exponential number of tableaux H/λ and we require a finer analysis.

Define the weight of a diagram λ to be the sum

$$w(\lambda) = \sum_{(i,j)\in\lambda} \frac{j(j+1)}{2}$$

Note that $w(\lambda)$ is the sum of the entries of the minimum semi-standard tableau on λ. Then there are two possibilities.

Proposition 3.12. *If G is an R-module such that $w(\lambda) > \epsilon n$, then Theorem 1.1 holds.*

Proof. Any semi-standard tableau H on λ is dominated by the event that $|T_\ell|$ is increased by $p^{\mu(k)}$, where $\mu(k)$ is the number of copies of k in H. This is bounded from above by the enlarged submodule $(p^{k\mu(k)} + W[\infty]$. These enlarged submodules form a nested sequence in $\mu(k)$, so it suffices to consider those μ where $\mu(k) \lesssim k^{-1}(\log p)^{-1}\eta n =: \nu(k)$ for η sufficiently small. Then in the product we can factor out the error,

$$\prod_{\ell=1}^{n}(p^{\ell\mu(\ell)} + O(e^{-c\alpha n})) = (1 + O(e^{-c\alpha n}))\prod_{\ell=1}^{n} p^{\ell\mu(\ell)}.$$

Finally we note that the number of possible labelings controlled by the envelope $\nu(k)$ is exponential with constant controlled by η, so Theorem 1.1 follows with η sufficiently small. □

If $w(\lambda) < \epsilon n$, then we can derive a bound on the number of possible diagrams of shape λ.

Proposition 3.13. *If $w(\lambda) < \epsilon n$ then there are $O(e^{-c\sqrt{\epsilon n}})$ semi-standard tableaux with letters from $[n]$ on λ.*

Proof. Let $d_\lambda(n)$ denote the number of semi-standard tableaux with letters from $[n]$ on λ. Recall from [4] that

$$d_\lambda(n) = \prod_{(i,j)\in\lambda} \frac{n+i-j}{r_i(\lambda) + c_j(\lambda) - j - i + 1}$$

where $r_i(\lambda)$ is the length of the ith row of λ and $c_j(\lambda)$ is the length of the jth column. We can bound

$$\begin{aligned}\prod_{(i,j)\in\lambda} (r_i(\lambda) + c_j(\lambda) - j - i + 1) &\geq \prod_{j=1}^{r_1(\lambda)} (c_j(\lambda)!) \\ &= \exp(\sum_{j=1}^{r_1} c_j(\log c_j - 1))O(\exp(c\sqrt{n})).\end{aligned}$$

Similarly,

$$\prod_{(i,j)}(n+i-j) \lesssim \exp(\sum_{j=1}^{r_1} c_j \log n).$$

Therefore

$$d_\lambda(n) \lesssim \exp(\sum_{j=1}^{r_1} c_j(\log n - \log c_j + 1)).$$

Now

$$\sum_{j=1}^{r_1} c_j(\log n - \log c_j + 1) \lesssim \sum_{t=0}^{\log n} \sum_{j:e^t \le c_j < e^{t+1}} e^t(\log n - t + 1)$$
$$\lesssim \sum_{t=0}^{\log n} e^t(\log n - t + 1)\#\{j \mid e^t \le c_j < e^{t+1}\}.$$

Note that

$$e^t\#\{j \mid e^t \le c_j \le e^{t+1}\}^2 \le \sum_{i=1}^{c_1} r_i^2 \lesssim \epsilon n$$

so that

$$\#\{j \mid e^t \le c_j \le e^{t+1}\} \le e^{-t/2}\sqrt{\epsilon n}$$

and thus

$$d_\lambda(n) \lesssim \exp(\sum_{t=0}^{\log n} e^{t/2}(\log n - t + 1)\sqrt{\epsilon n}) \lesssim \exp(cn\sqrt{\epsilon})$$

as required. □

We are now ready to prove Theorem 1.1 and Theorem 1.2.

Proof. Let λ be the Young diagram for G. If $w(\lambda) > \epsilon n$ then we are done. Otherwise, we have by the calculation above

$$\begin{aligned}\mathbb{P}(\operatorname{coker} A \cong G) &= \sum_{H/\lambda} \mathbb{P}(\cdots) \\ &= \sum_{H/\lambda} \prod_{\ell=0}^{n-1} (\mathbb{P}(B \cdots) + O(e^{-c\alpha n})) \\ &= \sum_{H/\lambda} \prod_{\ell=0}^{n-1} \mathbb{P}(B \cdots) + O(e^{-c\alpha n}) \\ &= \mathbb{P}(\operatorname{coker} B \cong G) + O(e^{-c\alpha n})\end{aligned}$$

where the third inequality is by Proposition 3.13. □

We have therefore reduced the probability distribution of the cokernel to computing the probability that X_ℓ lies in the expanded span of the previous columns. The remainder of this chapter describes how to prove Proposition 3.1.

4. Universality for enlarged submodules

We will distinguish between four possible types of submodules N that $\phi(W_{\ell+1})$ can represent. Let δ, d, and D be constants to be chosen later. Intuitively, δ and d will be of order $1/100$ while D will be about 10.

sparse: There is a non-zero $w \perp N$ such that $|\operatorname{supp} w| \le \delta n$; here $0 < \delta < 1$ is an absolute constant. We will show that sparse submodules are represented with exponentially small probability by direct counting in Section 6.

unsaturated: N is not sparse and

$$\max(e^{-d\alpha n}, \frac{D}{|N^\perp|}) \le \left|\mathbb{P}(X \in N) - \frac{1}{|N^\perp|}\right|.$$

Here D is an absolute constant. We will use a generalized version of the swapping argument from [10], [11] to show that these appear with exponentially small probability in Section 7.

semi-saturated: N neither sparse nor unsaturated, and we have the inequality

$$e^{-d\alpha n} \leq \left| \mathbb{P}(X \in N) - \frac{1}{|N^\perp|} \right| < \frac{D}{|N^\perp|}$$

Note that the set of semi-saturated N may be empty if $|N^\perp|$ is sufficiently large. We will count the semi-saturated submodules by finding a structured $w \perp N$ and then directly counting to show that they are represented with exponentially small probability. This is done in Section 9.

saturated: N is neither sparse, unsaturated, nor semi-saturated. In particular, these submodules satisfy the universality property.

Proposition 3.1 follows if we can show that $\phi(W_\ell)$ represents a saturated subspace with probability $1 - O(e^{-c\alpha n})$, where the constants are absolute.

5. Analytic and Combinatorial Tools

Before we finish the proof of Proposition 3.1, we must recall some theory from analysis and additive combinatorics. For more information see [9] and [7].

Let μ be a measure on a compact commutative topological ring. For any $0 < \epsilon < 1$ we define the spectrum $\operatorname{Spec}_{1-\epsilon} \mu$ to be the set of Fourier coefficients with magnitude at least $1 - \epsilon$; i.e.

$$\operatorname{Spec}_{1-\epsilon} \mu := \{\psi \in \widehat{R} \mid |\mu(\psi)| \geq 1 - \epsilon\}$$

The importance of Spec is that it is closed under a bounded number of set additions, as long as ϵ is enlarged sufficiently. Recall that for sets A, B in an additive group Z, we define $A + B$ to be the sumset $\{a + b \mid a \in A, b \in B\}$.

Lemma 5.1. *For all $\epsilon > 0$ and k a positive integer, we have*

$$\operatorname{Spec}_{1-\epsilon} \mu + \cdots + \operatorname{Spec}_{1-\epsilon} \mu \subseteq \operatorname{Spec}_{1-k^2\epsilon} \mu$$

where there are k summands on the left.

Recall that $\operatorname{Sym}(A)$ denotes the largest subgroup of Z such that A is the union of cosets of $\operatorname{Sym}(A)$.

Lemma 5.2 (Kneser)**.** *If $A, B \subseteq Z$ are additive sets in an ambient, finite abelian group Z, then we have the bound*

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

Iterating Kneser's inequality we have the following corollary.

Corollary 5.3. *Let $A_1, ..., A_k$ and B be additive sets in an ambient finite abelian group Z. Suppose we have the sumset inclusion*

$$A_1 + \cdots + A_k \subseteq B.$$

Suppose further that B contains no additive cosets of Z. Then we have

$$|B| + (k - 1) \geq |A_1| + \cdots + |A_k|$$

6. Sparse

We would like to control the probability that $\phi(W_{\ell+1})$ is sparse. By definition,

$$\mathbb{P}(\phi(W_{\ell+1}) \text{ is sparse}) = \mathbb{P}(\phi(W_{\ell+1}) \perp w \text{ non-zero with } |\operatorname{supp} w| \leq \delta n)$$

In particular, if we construct the $n \times (n - \ell)$ rectangular matrix

$$B := \begin{bmatrix} X_{\ell+1} & \cdots & X_n \end{bmatrix}$$

then $w^t B = 0$ for some non-zero w with $|\operatorname{supp} w| \leq \delta n$.

The coefficients of w are contained in a finite R-submodule of $\widehat{R}$, since elements of $\widehat{R}$ have finite order. We see that there is a maximal ideal $\mathfrak{m}$ such that $w \bmod \mathfrak{m}$ is not constant zero.

Reducing modulo this ideal, and possibly shrinking $\operatorname{supp} w$, we conclude that $w^t B = 0$ modulo $\mathfrak{m}$.

We can now finish the argument as in [7]. In particular, we can restrict B to those rows corresponding to non-zero entries of w. This gives a $|\sigma| \times n - \ell$ matrix that is not of full rank. Regardless of the choice of spanning columns, the remaining columns must lie in their span and in particular be perpendicular to w. We recall the "classical" Littlewood-Offord theorem for finite fields from [7].

Lemma 6.1. *Let $X \in \mathbb{F}_q^n$ be a random vector with iid entries taken from a probability distribution μ with concentration constant α. Suppose $w \in \mathbb{F}_q^n$ has at least m non-zero coefficients. Then we have the estimate*

$$\left|\mathbb{P}(X \cdot w = r) - q^{-1}\right| \lesssim \frac{1}{\sqrt{\alpha m}}$$

for all $r \in \mathbb{F}_q$.

Combining these estimates, we see that

$$\mathbb{P}(\phi(W_\ell) \text{ is sparse}) \lesssim \sum_{k=1}^{\delta n} \binom{n}{k}\binom{n-\ell}{k-1} \min(1-\alpha, q^{-1} + \frac{1}{\sqrt{\alpha k}})^{n-\ell-k+1}$$

where q is the smallest residue field for a maximal ideal of R. If we choose δ sufficiently small than this quantity is bounded by $O(e^{-c\alpha n})$.

7. Unsaturated

The singularity bounds of Tao and Vu in [10], [11] are based on swapping the columns of the matrix A with new columns drawn from a more singular probability distribution. Informally, if we have random vectors X and Y such that

$$\mathbb{P}(X \in V) \leq c\mathbb{P}(X \in V)$$

for some $0 < c < 1$ and all V in some class of vector spaces, then we can conclude that

$$\mathbb{P}(X_2, ..., X_n \text{ span } V) \leq c^{n-1}\mathbb{P}(Y_2, ..., Y_n \text{ span } V)$$

modulo some difficulties with linear independence. If we would like to show that the columns of A span such V with small probability, it therefore suffices to use much worse bounds (such as the trivial bound) for the "swapped in" vectors Y.

In [7] the author showed that this argument can also be used in the finite field setting as long as the vector spaces are not close to saturation for the random vectors. Explicitly, it was required that

$$|\mathbb{P}(X \in V) - |V^\perp|^{-1}| \geq D|V^\perp|^{-1}$$

for some D (about 10).

In our more general setting, the actual swapping lemma is a straightforward generalization of the swapping lemma from [7]. We will prove the following swapping lemma in 8.

Lemma 7.1. *There exists a probability distribution $\nu \in R$ with concentration constant $\alpha/8$ with the following property. Suppose $Y \in R^n$ is a random vector with iid entries taken from ν. Then for every submodule $N \lhd R^n$ that is not sparse, we have the inequality*

$$|\mathbb{P}(X \in N) - |N^\perp|^{-1}| \leq \left(\frac{1}{2} + o(1)\right)|\mathbb{P}(Y \in N) - |N^\perp|^{-1}|.$$

We need a new notion for linear independence. In addition to avoiding linear dependencies, it is important that our vectors not introduce additional cokernel. We therefore say that $Y_1, ..., Y_r$ are *simply independent in N* if $Y_1, ..., Y_r \in N$ and $R^n/\langle Y_1, ..., Y_r\rangle \cong R^{n-r}$. We will abbreviate *simply independent* by *s.a.*.

As in [10] and [7], we require a dyadic decomposition on the magnitude of $\mathbb{P}(X \in N)$. We recall that the *combinatorial codimension* of N is the unique fraction $d_\pm \in n^{-1}Z^+$ such that

$$(1-\alpha)^{d_\pm/n+1/n} \leq \mathbb{P}(X \in N) \leq (1-\alpha)^{d_\pm/n}.$$

Since we have the trivial bounds $2^{-n} \leq \mathbb{P}(X \in N) \leq 1$ we see that there are at most $O(\alpha n^2)$ possible combinatorial codimensions, so that it suffices to estimate

$$\mathbb{P}(\phi(W_{\ell+1}) \text{ unsaturated with } d_\pm(W) = d_\pm) \leq O(e^{-cn})$$

for all $d_\pm$.

Let $V^\perp$ denote the isomorphism class of $\phi(W_{\ell+1})^\perp$. Let N be an unsaturated submodules with $N^\perp \cong V^\perp$ of combinatorial codimension $d_\pm$. Let $Y_1, ..., Y_r$ be iid copies of Y given by Lemma 7.1 and let $X'_1, ..., X'_s$ be iid copies of X. Then we have

$$\mathbb{P}(\phi(W_{\ell+1}) = N) = \frac{\mathbb{P}(\phi(W_{\ell+1}) = N \wedge Y_1, ..., Y_r, X'_1, ..., X'_s \text{ are s.i. in } N)}{\mathbb{P}(Y_1, ..., Y_r, X_1, ..., X'_s \text{ are s.i. in } N)}.$$

Because we have assumed that $R^n/\langle Y_1, ..., Y_r, X'_1, ..., X'_s\rangle \cong R^{n-r-s}$, we can replace $r+s$ of the column vectors $X_{\ell+1}, ..., X_n$ with $Y_1, ..., Y_r, X'_1, ..., X'_s$ with the following proposition.

Proposition 7.2. *Suppose $\phi(W_{\ell+1}) = N$ and $Z_1, ..., Z_j$ are simply independent in N. Then there is a subset $\sigma \subseteq [\ell+1, n]$ with $|\sigma| = r$ such that*

$$\phi(\langle \{X_k\}_{k\notin\sigma}, Z_1, ..., Z_j\rangle) = N.$$

Furthermore, we can choose σ so that if $k \in \sigma$, we have $\ker H_{k+1} \longrightarrow H_k \cong R$.

Proof. It suffices to prove this proposition for $r = 1$, since the simple independence of $Z_1, ..., Z_j$ prevents a subsequent choice of X_k from colliding with previously chosen vectors.

From the definition of ϕ, we have

$$t(Z - v) = a_{\ell+1}X_{\ell+1} + \cdots + a_nX_n$$

for some $v \in IR^n$ and coefficients $a_k \in R$. We can find k such that $(a_k) = (t)$ as ideals in R. In particular, we have $a_k = tu$ for some $u \in R^\times$, and we therefore write

$$tX_k = u^{-1}t(Z - v) - u^{-1}a_{\ell+1}X_{\ell+1} - \cdots - u^{-1}a_nX_n.$$

Now, any $y \in N$ satisfies

$$t(y - v') = b_{\ell+1}X_{\ell+1} + \cdots + a_nX_n$$

but this equation can be rewritten to replace X_k with Z. □

Now it is convenient to abbreviate events. We define the events

$$\begin{aligned} D_n &:= Y_1, ..., Y_r, X'_1, ..., X'_s \in N \\ E_N &:= Y_1, ..., Y_r, X'_1, ..., X'_s \text{ s.i. in } N \\ F_{N,\sigma} &:= \phi(\langle \{X_k\}_{k\notin\sigma}, Y_1, ..., Y_r, X'_1, ..., X'_s\rangle) = N \end{aligned}$$

By Proposition 7.2, we have

$$\mathbb{P}(\phi(W_{\ell+1}) = N) \leq \sum_{\substack{\sigma\subseteq[\ell+1,n]\\|\sigma|=r+s}} \frac{\mathbb{P}(\{X_k\}_{k\in\sigma} \in N)}{\mathbb{P}(E_N)}\mathbb{P}(F_{N,\sigma})$$

First we consider the ratio

$$\frac{\mathbb{P}(\{X_k\}_{k\in\sigma} \in N)}{\mathbb{P}(E_N)}$$

We can expand the denominator with conditional expectation,

$$\begin{aligned} \mathbb{P}(E_N) &= \mathbb{P}(D_N)\mathbb{P}(E_N \mid D_N) \\ &= \mathbb{P}(Y \in N)^r\mathbb{P}(X \in N)^s\mathbb{P}(E_N \mid D_N) \end{aligned}$$

For the numerator we need the following independence lemma.

Proposition 7.3. *Let $Z_1, ..., Z_j$ be independent random vectors in R^n. Then we have the bound*

$$\mathbb{P}(Z_1, \ldots, Z_j \in N \mid Z_1, \ldots, Z_j \text{ are s.i.}) \leq \prod_{\ell=1}^{j} \mathbb{P}(Z_\ell \in N).$$

Proof. The proof is the same as in [7] except for minor notational differences. Expanding the left hand side with conditional expectation,

$$\prod_{j=1}^{r} \mathbb{P}(Z_j \in N \mid Z_1, \ldots, Z_{j-1} \in N \text{ and } Z_1, ..., Z_r \text{ are s.i.})$$

Let U_j denote the R span of $Z_1, \ldots, Z_{j-1}$. We claim that

$$\frac{\mathbb{P}(Z \in N \setminus U_j)}{\mathbb{P}(Z \notin U_j)} \leq \mathbb{P}(Z \in N).$$

In fact,

$$\begin{aligned}
\mathbb{P}(Z \in N \setminus U_j) &= \mathbb{P}(Z \in U_j)\mathbb{P}(Z \in N \setminus U_j) + \mathbb{P}(Z \notin U_j)\mathbb{P}(Z \in N \setminus U_j) \\
&\leq \mathbb{P}(Z \in U_j)\mathbb{P}(Z \notin U_j) + \mathbb{P}(Z \in N \setminus U_j)\mathbb{P}(Z \notin U_j) \\
&= \mathbb{P}(Z \in N)\mathbb{P}(Z \notin U_j).
\end{aligned}$$

□

Since $\{X_k\}_{k\in\sigma}$ are simply independent in N by construction, we therefore conclude that

$$\mathbb{P}(\{X_k\}_{k\in\sigma} \in N) \leq \mathbb{P}(X \in N)^{r+s}.$$

We now apply Lemma 7.1. Since N is unsaturated,

$$\frac{\mathbb{P}(X \in N)}{\mathbb{P}(Y \in N)} \leq \left(\frac{1}{2} + \frac{1}{D} + o(1)\right).$$

Collecting terms, we see that so far we have shown that

$$\mathbb{P}(\phi(W_{\ell+1}) = N) \leq \sum_{\substack{\sigma \subseteq [\ell+1,n] \\ |\sigma| = r+s}} \left(\frac{1}{2} + \frac{1}{D} + o(1)\right)^r \frac{1}{\mathbb{P}(E_N \mid D_N)} \mathbb{P}(F_{N,\sigma})$$

Next we control $\mathbb{P}(E_N \mid D_N)$ from below. We can expand this probability into the product

$$\prod_{j=1}^{r} \mathbb{P}(Y_1, ..., Y_j \text{ s.c. in } N \mid Y_1, ..., Y_{j-1} \text{ s.c. in } N \text{ and } Y_1, ..., Y_r \in N)$$

along with analogous terms for X'. We can write each term of the product in the form

$$\prod_{\mathfrak{m}} (1 - \mathbb{P}(Y_j \in \langle Y_1, ..., Y_{j-1} \rangle + \mathfrak{m}R^n \mid Y_j \in N)).$$

However, we can bound this with Odlyzko's lemma and the condition on the combinatorial codimension of N.

To finish the argument, we now sum over all submodules N of prescribed isomorphism class. We then find that

$$\mathbb{P}(\phi(W_{\ell+1}) \text{ is unsat}) \lesssim \sum_{\substack{\sigma \subset [\ell+1,n] \\ |\sigma| = r+s}} (\frac{1}{2} + \frac{1}{D} + o(1))^r \sum_{N \text{ unsat}} \mathbb{P}(F_{N,\sigma}).$$

We see that the inner sum on the right hand side is always bounded by 1, since the collection of vectors can only induce a single submodule. We then choose r,s appropriately to bound the whole quantity by $O(e^{-c\alpha n})$. □

8. Swapping Lemma

We will now prove a generalization of the swapping lemma from [10], [11], [7].

Proof of Lemma 7.1. We will abbreviate $\gamma = 1/8$ for convenience. Let

$$\nu(t) := \begin{cases} \gamma\mu * \mu^-(t), & t \neq 0 \\ 1 - \sum_{s\neq 0} \nu(s), & t = 0. \end{cases}$$

Here $\mu^-(t) := \mu(-t)$. It is trivial to verify that ν is a probability distribution and $\widehat{\nu} > 1 - 2\gamma$. It also has concentration constant $\beta = \gamma\alpha = \alpha/8$.

The Fourier transform of ν is

$$\widehat{\nu} = 1 - \gamma + \gamma|\widehat{\mu}|^2.$$

It now suffices to verify the swapping inequality. We observe that

$$\mathbb{P}(X \in N) - |N^\perp|^{-1} = |N^\perp|^{-1} \sum_{\psi\in N^\perp\setminus\{0\}} \mathbb{E}\psi(X) = |N^\perp|^{-1} \sum_{\psi\in N^\perp\setminus\{0\}} \prod_{\ell=1}^{n} \widehat{\mu}(\psi_\ell)$$

and similar for $\mathbb{P}(Y \in N)$. We therefore define

$$f(\psi) := \prod_{\ell=1}^{n} |\widehat{\mu}(\psi_\ell)|$$

and

$$g(\psi) := \prod_{\ell=1}^{n} \widehat{\nu}(\psi_\ell)$$

so that it suffices to show that

$$\sum_{\psi\in N^\perp\setminus\{0\}} f(\psi) \leq \left(\frac{1}{2} + o(1)\right) \sum_{\psi\in N^\perp\setminus\{0\}} g(\psi).$$

We do this by level sets. For $u > 0$ define

$$F(u) := \{\psi \in N^\perp \mid f(\psi) \geq u\}$$

and

$$G(u) := \{\psi \in N^\perp \mid g(\psi) \geq u\}$$

and likewise let $F'(u) = F(u) \setminus \{0\}$ and $G'(u) = G(u) \setminus \{0\}$.

Let $\epsilon > 0$ be chosen later. We must split the sum for f into two parts: those frequencies ψ where $f(\psi) \leq \epsilon$ and those where $f(\psi) > \epsilon$.

We first claim that $f(\psi) \leq g(\psi)^4$. This follows from the pointwise estimate $|\widehat{\mu}(t)| \leq \widehat{\nu}(t)^4$, which we can prove with the arithmetic-geometric mean inequality:

$$(|\widehat{\mu}(t)|^2)^{1/8} \leq \frac{1}{8}(|\widehat{\mu}|^2 + 7) = \nu(t).$$

This inequality quickly controls those frequencies where f is small. In fact,

$$\sum_{\substack{\psi\in N^\perp\setminus\{0\} \\ f(\psi)\leq\epsilon}} f(\psi) \leq \epsilon^{3/4} \sum_{\substack{\psi\in N^\perp\setminus\{0\} \\ f(\psi)\leq\epsilon}} g(\psi)$$

so as long as $\epsilon \to 0$ as $n \to \infty$ this portion is done.

Now we will consider those frequencies where f is large. In this case, we will apply Kneser's ineuality to the sumset inclusion $F(u) + F(u) \subseteq G(u)$.

It suffices to show that $|\widehat{\mu}(t)\widehat{\mu}(s)| \leq \widehat{\nu}(t+s)^2$. As in [7] we consider two cases. If either $\widehat{\mu}(t) < 1 - 4\gamma$ or $\widehat{\mu}(t) < 1 - 4\gamma$ then the inequality is trivial from the lower bound for $\widehat{\nu}$. Otherwise,

we write $|\widehat{\mu}(t)| = 1-\theta_1$ and $|\widehat{\mu}(s)| = 1-\theta_2$. By Lemma 5.1 we have $|\widehat{\mu}(t+s)| \geq 1-2(\theta_1+\theta_2)$. But by the definition of ν we get $\widehat{\nu}(t+s) = 1-\gamma+\gamma|\widehat{\mu}(t+s)|^2 \geq |\widehat{\mu}(t)\widehat{\mu}(s)|$.

Kneser's theorem tells us that $2|F(u)| \leq |\operatorname{Sym} F(u) + F(u)| + |G(u)|$. We would like to show that $\operatorname{Sym}(F(u)+F(u)) = \{0\}$. It suffices to show that $G(u)$ does not contain any non-trivial subgroup, as this would in turn guarantee that $F(u)+F(u)$ does not either.

Let $H \lhd G(u)$ be minimal, so that $H \cong \mathbb{Z}/p\mathbb{Z}$ for some prime p. Choose $w \in N^\perp$ that generates H as a $\mathbb{Z}/p\mathbb{Z}$-module; since N is not sparse, we can assume that w contains at least δn non-zero entries.

Define the function

$$h(t) := \sum_{\ell=1}^{n} 1 - \widehat{\nu}(t_\ell)^2$$

Averaging h over H,

$$p^{-1}\sum_{t\in\mathbb{Z}/p\mathbb{Z}} h(tw) = n - p^{-1}\sum_{t\in\mathbb{Z}/p\mathbb{Z}}\sum_{\ell=1}^{n}\widehat{\nu}(tw_\ell)^2$$

By Plancherel's theorem and the fact that N is not saturated we see that this entire quantity is bounded below by $\gamma\alpha\delta n$. Therefore we simply require that $u \leq \gamma\alpha\delta n$, which gives us the bound

$$\sum_{\substack{\psi\in N^\perp\setminus\{0\}\\ f(\psi)>\epsilon}} f(\psi) \leq \frac{1}{2}\sum_{\substack{\psi\in N^\perp\setminus\{0\}\\ f(\psi)>\epsilon}} g(\psi)$$

as required. □

9. Semi-saturated

Let N be a semi-saturated submodule. By the inverse Fourier transform,

$$\begin{aligned} e^{-dn} &\leq \left|\mathbb{P}(X\in N) - \frac{1}{|N^\perp|}\right| \\ &\leq \frac{1}{|N^\perp|}\sum_{\xi\in N^\perp\setminus\{0\}}\prod_{\ell=1}^{n}|\widehat{\mu}(\xi_\ell)| \end{aligned}$$

In particular, we can find a $\zeta \in N^\perp \setminus \{0\}$ such that

$$\exp(-dn) \leq \exp(-\frac{1}{2}\sum_{\ell=1}^{n}\psi(\zeta_\ell)),$$

with $\psi(t) := 1-|\widehat{\mu}(t)|^2$; taking logarithms gives

$$\sum_{\ell=1}^{n}\psi(\zeta_\ell) \leq 2dn.$$

Let κ be a parameter to be chosen later. For all but κ of ξ_ℓ, we have

$$\psi(\zeta_\ell) \leq \frac{2dn}{\kappa}$$

or, equivalently, $\zeta_\ell \in \operatorname{Spec}_{1-\epsilon}\mu$ for $\epsilon = \frac{dn}{\kappa}$.

Proposition 9.1. *For all $\beta > 0$ there exists $\epsilon > 0$ such that for all orders T, $\operatorname{Spec}_{1-\epsilon}\mu$ contains at most βT elements of order T.*

Proof. We first notice that there is an $\eta > 0$ such that $\operatorname{Spec}_{1-\eta}\mu$ does not contain any non-trivial additive cosets $H+s$ with $H \lhd \widehat{R}$ and $s \in \widehat{R}$. In fact, for each such H and s we can apply Markov's inequality and use the α-concentration of μ,

$$(1-\eta)^2\#(H+s\cap\operatorname{Spec}_{1-\eta}\mu) \leq \sum_{h\in H}|\widehat{\mu}(h+s)|^2 \leq |H|(1-\alpha)$$

We therefore choose $\eta = \alpha/2$. With this value, we apply the iterated form of Kneser's inequality to find

$$k|\{t \in \widehat{R} \mid |t| = T\} \cap \operatorname{Spec}_{1-\epsilon} \mu \setminus \{0\}| \leq |\{t \in \widehat{R} \mid |t| = T\} \cap \operatorname{Spec}_{1-k^2\epsilon} \mu \setminus \{0\}|$$

so we pick $k = \beta^{-1}$ and $\epsilon = k^{-2}\eta$. □

We now choose ϵ, and therefore d, such that $\operatorname{Spec}_{1-\epsilon} \mu$ has small cardinality in every finite torsion submodule of $\widehat{R}$; namely βT for order T elements.

We can count directly the number of vectors ζ of given order with the above constraint. In fact, for order T we have

$$\#\{\text{structured } \zeta \text{ of order } T\} \leq \binom{n}{\kappa} T^{\kappa} (\beta T)^{n-\kappa} \lesssim \beta^n T^n$$

where in the last inequality we chose $\kappa = n/10$ and replaced β with a comparable value.

We can count the number of submodules N with prescribed perpendicular isomorphism class and perpendicular to a fixed vector.

Proposition 9.2. *Let $W^\perp$ be a fixed finite submodule of $\widehat{R}^n$ and let $\zeta \in \widehat{R}^n$ have finite order T. Then the number of submodules $N \lhd R^n$ with $N^\perp \cong W^\perp$ and $\zeta \in N^\perp$ is bounded by*

$$\#\{N \mid \zeta \in N^\perp\} \lesssim \frac{|W^\perp|^n}{|\operatorname{Aut} W^\perp| T^n} \#\{\zeta \in W^\perp \mid |\zeta| = T\}$$

Proof. We double count the number of pairs $\zeta \in N^\perp$,

$$\sum_{|\zeta|=T} \#\{N \mid \zeta \in N^\perp\} = \#\{(N^\perp, \zeta) \mid \zeta \in N^\perp, |\zeta| = T\} = \sum_N \#\{\zeta \in N^\perp \mid |\zeta| = T\}$$

We observe that $\#\{N \mid \zeta \in N^\perp\}$ is independent of ζ. In fact, for any two ζ_1, ζ_2 of order T we can find an automorphism $\widehat{R}^n \to \widehat{R}^n$ that maps ζ_1 to ζ_2; this induces a correspondence between submodules of the same isomorphism class. We conclude that

$$\#\{N \mid \zeta \in N^\perp\} = \frac{\#\{N \mid N^\perp \cong W^\perp\} \#\{\zeta \in N^\perp \mid |\zeta| = T\}}{\#\{\zeta \in \widehat{R}^n \mid |\zeta| = T\}}.$$

We have the trivial bound

$$\#\{N \mid N^\perp \cong W^\perp\} \leq \frac{|N^\perp|^n}{|\operatorname{Aut} N^\perp|},$$

so collecting terms we find

$$\#\{N \mid \zeta \in N^\perp\} \lesssim \frac{|W^\perp|^n}{|\operatorname{Aut} W^\perp| T^n} \#\{\zeta \in W^\perp \mid |\zeta| = T\}$$ □

Now we are ready to estimate the probability that W is semi-saturated. We count the number of semi-saturated spaces,

$$\begin{aligned}\mathbb{P}(W \text{ is semi-saturated}) &\leq \sum_{N \text{ semi-saturated}} \mathbb{P}(W = N) \\ &\leq \#\{N \text{ semi-saturated} \mid N^\perp \cong W^\perp\} \left(\frac{D}{|W^\perp|}\right)^{n-\ell}\end{aligned}$$

With the above bounds on the number of structured vectors and the number of submodules perpendicular to a given vector, we bound

$$\#\{N \mid \text{semi-sat and } \zeta \in N^\perp \text{ struct., order } T\} \lesssim \beta^n \frac{|W^\perp|^n}{|\operatorname{Aut} W^\perp|} \#\{\zeta \in W^\perp \mid |\zeta| = T\}$$

Summing over every possible order T,

$$\#\{N \mid \text{semi-sat and } \zeta \in N^\perp \text{ struct.}\} \lesssim \beta^n \frac{|W^\perp|^n}{|\operatorname{Aut} W^\perp|} |W^\perp|.$$

We now combine this with the bound on $\mathbb{P}(W = N)$ to find

$$\mathbb{P}(W \text{ semi-sat}) \lesssim \beta^n D^{n-\ell} \frac{|W^\perp|^\ell}{|\operatorname{Aut} W^\perp|} |W^\perp|.$$

By construction of W, its perpendicular module $W^\perp$ must contain at least ℓ terms of maximal order, so

$$|\operatorname{Aut} W^\perp| \geq |W^\perp|^\ell.$$

We also know that $|W^\perp| \leq De^{cn}$ since W is semi-saturated, so if we take β (and therefore d) sufficiently small then we find

$$\mathbb{P}(W \text{ is semi-saturated}) = O(e^{-cn})$$

as required. □

10. Acknowledgements

The author would like to thank Terence Tao for helpful criticism and guidance.

References

[1] L. S. Charlap, H. D. Rees, and D. P. Robbins. The asymptotic probability that a random biased matrix is invertible. *Discrete Math.*, 82(2):153–163, 1990.

[2] H. Cohen and H. W. Lenstra, Jr. Heuristics on class groups of number fields. In *Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983)*, volume 1068 of *Lecture Notes in Math.*, pages 33–62. Springer, Berlin, 1984.

[3] E. Friedman and L. C. Washington. On the distribution of divisor class groups of curves over a finite field. In *Théorie des nombres (Quebec, PQ, 1987)*, pages 227–239. de Gruyter, Berlin, 1989.

[4] W. Fulton. *Young tableaux*, volume 35 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1997. With applications to representation theory and geometry.

[5] J. Kahn and J. Komlós. Singularity probabilities for random matrices over finite fields. *Combin. Probab. Comput.*, 10(2):137–157, 2001.

[6] G. Malle. Cohen-Lenstra heuristic and roots of unity. *J. Number Theory*, 128(10):2823–2835, 2008.

[7] K. Maples. Singularity of random matrices over finite fields. preprint `arXiv:1012.2372 [math.CO]`.

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

[9] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

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

[11] T. Tao and V. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628 (electronic), 2007.

Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich
E-mail address: `kenneth.maples@math.uzh.ch`