SINGULARITY OF RANDOM MATRICES OVER FINITE FIELDS

KENNETH MAPLES

Abstract. Let \(A\) be an \(n \times n\) random matrix with iid entries over a finite field of order \(q\). Suppose that the entries do not take values in any additive coset of the field with probability greater than \(1 - \alpha\) for some fixed \(0 < \alpha < 1\). We show that the singularity probability converges to the uniform limit with error bounded by \(O(e^{-c\alpha n})\), where the implied constant and \(c > 0\) are absolute. We also show that the determinant of \(A\) assumes each non-zero value with probability \(q^{-1} \prod_{k=2}^{\infty} (1 - q^{-k}) + O(e^{-c\alpha n})\), where the constants are absolute.

1. Introduction

Let \(q = p^j\) be a prime power and let \(\mathbb{F}_q\) be the finite field with \(q\) elements. Let \(\xi\) be a random variable that takes values in \(\mathbb{F}_q\) with probability distribution \(\mu\). We need \(\mu\) to be suitably non-degenerate.

We say that \(\mu\) is \(\alpha\)-dense for some \(0 < \alpha < 1\) if for every additive subgroup \(T \triangleleft \mathbb{F}_q\) and \(s \in \mathbb{F}_q\),
\[
P(\xi \in s + T) \leq 1 - \alpha.
\]

Let \(A\) be an \(n \times n\) random matrix whose entries are iid copies of \(\xi\). We are interested in computing the typical spectral properties of \(A\); in particular, we would like to know how often the matrix is singular. As it turns out, as long as \(A\) is “dense” in the sense that its entries take values from an \(\alpha\)-dense distribution, the singularity probability converges rapidly to the expected value.

Theorem 1.1 (Charlap, Rees, Robbins [2]). With \(A\) and \(\mathbb{F}_q\) as above, we have
\[
P(A \text{ is non-singular}) = \prod_{k=1}^{\infty} (1 - q^{-k}) + o_{q,\mu}(1).
\]

Note that
\[
\prod_{k=1}^{n} (1 - q^{-k}) = \frac{|\text{GL}_n(\mathbb{F}_q)|}{|\text{M}_n(\mathbb{F}_q)|}
\]
is the density of invertible matrices in the set of all \(n \times n\) matrices over \(\mathbb{F}_q\).

In this article, we improve the error to an exponential decay depending only on the \(\alpha\)-density of the distribution.

Theorem 1.2. Let \(\mathbb{F}_q\) with \(q = p^j\) and suppose \(A \in M(n, \mathbb{F}_q)\) is a random matrix with iid entries which take values from an \(\alpha\)-dense probability distribution. Then we have the estimate
\[
P(A \text{ is non-singular}) = \prod_{k=1}^{\infty} (1 - q^{-k}) + O(e^{-c\alpha n})
\]
where the implied constant and \(c > 0\) are absolute.

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

The author is supported by a Graduate Research Fellowship from the National Science Foundation.
The dependence on $\alpha$ is optimum: if we take $\mu$ such that $\Pr(\xi = 0) > 1 - e^{\log_2 n}$ for $c$ suitably small, then we expect $A$ to have a positive proportion of zero columns.

The techniques used to prove Theorem 1.2 also generalize to control the distribution of the determinant.

**Theorem 1.3.** For all non-zero $t \in \mathbb{F}_q$, we have the formula

$$
\Pr(\det A = t) = q^{-1} \prod_{k=2}^{\infty} (1 - q^{-k}) + O(e^{-cn}).
$$

where $c > 0$ and the implied constant are absolute.

We will consider more refined estimates for the singularity of random matrices in [10]. In particular we will prove estimates for the rank of non-degenerate matrices over a finite field and for matrices over other arithmetic rings.

The singularity of random matrices with continuous distributions over $\mathbb{C}$ is well-understood. If the last $n-1$ columns of an $n \times n$ matrix are linearly independent, they span a hypersurface in $\mathbb{C}^n$. Hypersurfaces have Lebesgue measure zero, so as long as the probability distribution of $X$ is absolutely continuous the matrix is almost surely non-singular.

In contrast, matrices whose entries that take values in a discrete distribution may be singular with strictly positive probability. For example, if the probability distribution $\mu$ is the Bernoulli distribution then the probability that the first two columns are linearly dependent is $2^{1-n}$. Komlós was the first to show in [7] that the singularity probability of an $n \times n$ Bernoulli matrix over the integers converges to zero. He subsequently generalized this proof to general distributions in [8].

It was significantly more difficult to improve this bound to an exponential rate. For Bernoulli matrices, the first proof was found by Kahn, Komlós, and Szemerédi, who found the bound $c^n$ for some $c < 1$ in [6]. Tao and Vu improved the bound to $(3/4)^n$ in [13] and [14], and it was further improved to $(1/\sqrt{2})^n$ in [1] by Bourgain, Vu, and Wood.

These estimates on the singularity probability are derived from progress on the Littlewood-Offord problem. The classical Littlewood-Offord problem asks, for a fixed $(a_1, \ldots, a_n) \in \mathbb{R}^n$, how many of the signed sums $\pm a_1 \pm \cdots \pm a_n$ can lie in a given interval? The problem is so called after Littlewood and Offord’s analysis of the real zeros of random polynomials in [9]. Erdős improved the inequality in [3] to show that for $|a_k| \geq 1$ the number of sums in the interval is bounded above by $O(n^{-1/2}2^n)$, and that this inequality is sharp for the all-ones vector. In general, Littlewood-Offord inequalities give a correspondence between structure in the coefficients of the vector $(a_1, \ldots, a_n)$ and bounds on the number of signed sums in a given interval.

In this paper we prove three Littlewood-Offord theorems for finite fields. We have replaced the random sums from the classical problem with inner products $w \cdot X$, where $w \in \mathbb{F}_q^n$ is a fixed vector of (mostly) non-zero residues and $X$ is a random vector with iid entries taken from an $\alpha$-dense probability distribution. The techniques we employ go back to Halász in [4], who showed that the probability $\Pr(X \cdot w = 0)$ can be controlled by finding additive structure in the level sets of the Fourier transform of $1_{X \cdot w = 0}$. We also crucially rely on arguments developed in [6] and [13], where it was discovered that $\Pr(X \cdot w = 0)$ can be controlled by constructing auxiliary random vectors with larger concentration probabilities. We will discuss this idea further in Section 2.3.

The key new advance to study matrices over finite fields is an inverse theorem for random sums $w \cdot X$ which are almost uniformly distributed, but differ from the uniform distribution by an exponentially small quantity. In this setting we show that the coefficients of $w$ must lie in a small subset $R \subseteq \mathbb{F}_q^n$. Because the sums are almost uniformly distributed, we can enumerate all such vectors.
The requirement that \( \mu \) be an \( \alpha \)-dense probability distribution can be weakened. As discussed in [2], let \( \theta \in \mathbb{F}_8 \) be a primitive element and consider the uniform distribution \( \mu \) on \( \{0, 1, \theta, 1 + \theta\} \). With the methods from this article it is easy to show that we recover the expected singularity probability \( \prod_{k=1}^{\infty} (1 - 8^{-k}) + O(e^{-cn}) \) while \( \mu \) is supported on an additive subgroup of \( \mathbb{F}_2^3 \). This was first done by Kahn and Komlós in [5] where they showed that it suffices to assume that \( \mu \) does not concentrate on affine subfields; i.e. subsets of the form \( \beta \mathbb{F}_q^d + \gamma \) for \( d \mid f \) and \( \beta, \gamma \in \mathbb{F}_q \).

With this weaker condition we can construct examples that do not have an exponentially small error term. For example, let \( f \) be a large prime, \( \theta \in \mathbb{F}_p^d \) a primitive element and \( \mu \) uniformly distributed on \( \{0, 1, \theta, 1 + \theta\} \). After expanding the determinant we see that \( \det A \) can only take values in the additive subgroup \( \langle 1, \theta, \theta^2, ..., \theta^m \rangle \). This appears to be the only obstruction, and such probability distributions will be considered in a forthcoming paper.

In Section 2 we prove Theorem 1.2 for dense distributions over general finite fields. In Section 3 we prove the Littlewood-Offord theorems we need to establish Theorem 1.2. Finally, we prove Theorem 1.3 in Section 4.

2. Proof of Theorem 1.2

Let \( X_1, ..., X_n \) denote the columns of \( A \). For convenience we will let \( X \in \mathbb{F}_q^n \) denote an independent random vector with iid entries distributed according to \( \mu \); thus each \( X_k \) is an iid copy of \( X \) for \( 1 \leq \ell \leq n \).

We expose each column \( X_k \) in turn, from \( X_n \) to \( X_1 \), and check whether it lies in the span of the previously exposed columns. Let \( W_k := \langle X_{k+1}, ..., X_n \rangle \) denote the span of the final \( n - k \) columns of \( A \). By conditional expectation,

\[
\mathbb{P}(A \text{ is non-singular}) = \prod_{k=1}^{n} \mathbb{P}(X_k \notin W_k \mid \text{codim} \, W_k = k)
\]

Suppose that \( V \) is a deterministic subspace of \( \mathbb{F}_q^n \) of codimension \( k \). If \( X_k \) were chosen uniformly from \( \mathbb{F}_q^n \) then we would have \( \mathbb{P}(X_k \notin V) = 1 - q^{-k} \) and the theorem would follow. In fact, we can show that for sufficiently small \( k \) this equality holds with exponentially small error.

**Proposition 2.1.** There is an absolute constant \( \eta > 0 \) such that, for all \( 1 \leq k \leq \eta n \), we have the estimate

\[
\mathbb{P}(X_k \in W_k \mid \text{codim} \, W_k = k) = q^{-k} + O(e^{-cn})
\]

where the implied constant and \( c > 0 \) are absolute.

For columns \( X_k \) with \( k > \eta n \) we have \( q^{-k} = O(e^{-cn}) \), so it suffices to show that

\[
\mathbb{P}(X_k \in W_k \mid \text{codim} \, W_k = k) = O(e^{-cn}).
\]

This is guaranteed by the following lemma, first recorded in [11].

**Lemma 2.2** (Odlyzko). For any fixed subspace \( V \) of \( \mathbb{F}_q^n \) and random vector \( X \in \mathbb{F}_q^n \) that is \( \alpha \)-dense, we have the bound

\[
\mathbb{P}(X \in V) \leq (1 - \alpha)^{\text{codim} \, V}.
\]

**Proof of Lemma 2.2.** Let \( k \) denote the codimension of \( V \). We can find \( n - k \) coordinates \( \tau \subseteq [n] \) such that \( V \) is a graph over \( \tau \). If we condition on the coordinates of \( X \) in \( \tau \), then there is a unique choice for the remaining coordinates \( [n] \setminus \tau \) for \( X \in V \). Since \( \mu \) is \( \alpha \)-dense, 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. \( \square \)
We will now prove Proposition 2.1. It is convenient to distinguish four kinds of subspaces that \( W_k \) can represent as \( X_{k+1},...,X_n \) vary. Fix absolute constants \( \delta, d, \) and \( D; \) for intuition we can take \( \delta = 1/100, \) \( d = 1/100, \) and \( D = 10, \) but we do not compute exact values.

Let \( V \) be a fixed codimension \( k \) subspace of \( \mathbb{F}_q^m. \) Then we say that \( V \) is sparse, unsaturated, semi-saturated, or saturated as follows.

**Sparse:** There is a non-zero \( w \perp V \) with \( |\text{supp } w| \leq \delta n. \) We can directly count these subspaces.

**Unsaturated:** \( V \) is not sparse and we have the estimate

\[
\max(e^{-\delta n}, Dq^{-k}) < |\mathbb{P}(X \in V) - q^{-k}|.
\]

We adapt the swapping method from [13] and construct a random vector \( Y \) such that

\[
\mathbb{P}(X \in V) \leq (\frac{1}{2} + \frac{1}{D} + o(1))\mathbb{P}(Y \in V).
\]

**Semi-saturated:** \( V \) is not sparse and we have the estimates

\[
e^{-\delta n} < |\mathbb{P}(X_k \in V) - q^{-k}| \leq Dq^{-k}.
\]

In this range the swapping method does not yield a useful gain; however, we can enumerate semi-saturated \( V \) by finding a structured \( w \perp V. \) Note that for \( q \) sufficiently large there are no semi-saturated spaces.

**Saturated:** \( V \) is not sparse and we have the estimate

\[
|\mathbb{P}(X_k \in V) - q^{-k}| \leq e^{-\delta n}.
\]

Proposition 2.1 will follow if we can show that \( W_k \) represents a saturated subspace with probability \( 1 - O(e^{-\alpha n}) \) with absolute constants. It therefore suffices to show that \( W_k \) is sparse, semi-saturated, or unsaturated with probability \( O(e^{-\alpha n}). \)

### 2.1. Sparse subspaces

We adapt the counting method from [6]. If \( W_k \) is sparse, then we can find a non-zero \( w \perp W_k \) with \( |\text{supp } w| \leq \delta n. \) By the union bound,

\[
\mathbb{P}(W_k \text{ is sparse}) \leq \sum_{\sigma \subseteq [n] \atop 1 \leq |\sigma| \leq \delta n} \mathbb{P}(W_k \perp w \text{ for some } w \text{ with } \text{supp } w = \sigma).
\]

Fix \( \sigma. \) It suffices to bound

\[
Q_\sigma := \mathbb{P}(W_k \perp w \text{ for some } w \text{ with } \text{supp } w = \sigma) \leq O(e^{-\alpha n})
\]

with the implied constant and \( c > 0 \) depending on \( \delta. \) We will choose \( \delta \) in the proofs for unsaturated and semi-saturated subspaces.

If we have such a perpendicular vector \( w \) we can write the matrix equation

\[
w^t [X_{\ell+1} \cdots X_n] = 0.
\]

Restricting the product to indices in \( \sigma \) and denoting this reduction by \( \tilde{\cdot}, \)

\[
\tilde{w}^t [\tilde{X}_{\ell+1} \cdots \tilde{X}_n] = 0.
\]

The matrix of reduced columns has size \( |\sigma| \times (n-\ell) \) and has rank less than \( |\sigma|, \) so we conclude that the dimension of the column space is at most \( |\sigma| - 1. \) There are at most \( \binom{n-\ell}{|\sigma|-1} \) possible choices for a set \( \tau \) of spanning columns; we do not require that they be linearly independent. Regardless of the choice of \( \tau, \) the remaining columns must be perpendicular to \( \tilde{w}. \) Collecting these bounds, we find

\[
Q_\sigma \leq \sum_{\tau \subseteq [n] \atop |\tau| = |\sigma| - 1} \sup_{\text{supp } w = \sigma} \mathbb{P}(X_t \perp w \text{ for all } t \notin \tau \mid \text{codim } W_k = k)
\]
We expect linearly independent vectors to be less likely to lie in a given subspace than average. The next proposition verifies that intuition.

**Proposition 2.3.** Let $Z_1, ..., Z_r$ be non-trivial iid random vectors in $\mathbb{F}_q^n$. Then we have the bound

$$
P(Z_1, \ldots, Z_r \in V \mid Z_1, \ldots, Z_r \text{ are linearly independent}) \leq P(Z \in V)^r.
$$

**Proof.** Expanding the left hand side with conditional expectation,

$$
\prod_{j=1}^r P(Z_j \in V \mid Z_1, \ldots, Z_{j-1} \in V \text{ and } Z_1, \ldots, Z_j \text{ are linearly independent})
$$

Let $U := \langle Z_1, \ldots, Z_{j-1} \rangle$ denote the span of the exposed vectors. It suffices to show that

$$
P(Z \in V \setminus U) P(Z \notin U) \leq P(Z \in V).
$$

In fact,

$$
P(Z \in V \setminus U) = P(Z \in U) P(Z \in V \setminus U) + P(Z \notin U) P(Z \in V \setminus U)
$$

$$
\leq P(Z \in U) P(Z \notin U) + P(Z \in V \setminus U) P(X \notin U)
$$

$$
= (P(Z \in U) + P(Z \in V \setminus U)) P(Z \notin U)
$$

and the proposition follows. \hfill \Box

Combining terms we get

$$
P(W_k \text{ is sparse}) \leq \sum_{\sigma \subseteq [n], 1 \leq |\sigma| \leq \delta n} \left( \frac{n - \ell}{|\sigma| - 1} \right) \sup_{\text{supp } w = \sigma} P(X \perp w)^{n - |\sigma| + 1}
$$

It remains to bound $P(X \perp w)$ for $w$ with support $\sigma$. For this task we can use the following Littlewood-Offord theorem.

**Lemma 2.4** (Littlewood-Offord). Let $X \in \mathbb{F}_q^n$ be a random vector with iid entries taken from an $\alpha$-dense probability distribution $\mu$. Suppose $w \in \mathbb{F}_q^n$ has at least $m$ non-zero coefficients. Then we have the estimate

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

for all $r \in \mathbb{F}_q$, where the implied constant is absolute.

We only require the estimate for $r = 0$. We will prove Lemma 2.4 in Section 3.

If we combine this with the trivial inequality $P(X \perp w) \leq 1 - \alpha$ for small $|\sigma|$, we deduce

$$
P(W \text{ is sparse}) \leq O(e^{-c \alpha n})
$$

with absolute constants for $\delta$ sufficiently small.

2.2. **Semi-Saturated subspaces.** Let $V$ be a semi-saturated subspace of codimension $k$. We first claim that we can find a non-zero $\xi \perp V$ that is structured in the following sense.

**Proposition 2.5.** For all $\beta > 0$ there is a value of $d$ in the definition of semi-saturated and a subset

$$
R \subseteq \mathbb{F}_q^n, \quad |R| \leq \beta^n q^n
$$

such that every semi-saturated $V$ is perpendicular to a non-zero $\xi \in R$. 
We will prove Proposition 2.5 in Section 3.

We have therefore found a $\xi \in V^\perp$ that is “structured” in that it lies in an exponentially small subset of $\mathbb{F}_q^n$. It turns out that this is enough to attain the desired estimate on $\mathbb{P}(W_k$ is semi-saturated). In fact, we estimate

$$\mathbb{P}(W_k \text{ is semi-saturated} \mid \text{codim } W_k = k) \leq \sum_{\text{codim } V = k \atop V \text{ is semi-saturated}} \mathbb{P}(W_k = V \mid \text{codim } W = k).$$

Using Proposition 2.3, we can bound

$$\mathbb{P}(W_k = V \mid \text{codim } W_k = k) \leq \mathbb{P}(X \in V)^{n-k} \leq D^{n-k}q^{-k(n-k)}$$

where the last inequality is from the definition of semi-saturated $V$. It now suffices to count the number of semi-saturated subspaces.

The subspace $V$ is completely determined by its annihilator $V^\perp$. We therefore count the number of possible annihilators that meet $R$. We can choose $k$ generators $v_1, \ldots, v_k$ for $V^\perp$ and force $v_1 \in R$; we then divide by the number of ways we could generate the same subspace with different choices for $v_2, \ldots, v_k$. This gives the upper bound

$$\#\{\text{semi-saturated } V\} \lesssim \beta^n q^n (q^n)^{k-1} \leq \beta^n q^{n(k^2+1)}.$$

Collecting terms we find

$$\mathbb{P}(W_k \text{ is semi-saturated}) \lesssim D^{n-k} \beta^n q^k$$

If there are any semi-saturated subspaces, we must have the inequality $e^{-d \alpha n} \leq Dq^{-k}$. With fixed $D$ we can choose $\beta$ and therefore an upper bound for $d$ such that the right hand side converges to zero at an exponential rate.

2.3. Unsaturated subspaces. In [13] it was observed that there is a random vector $Y$ such that if $X \in \mathbb{R}^n$ is a Bernoulli random vector and $V$ is a non-sparse hyperplane, then we can bound

$$\mathbb{P}(X \in V) \leq (\frac{1}{2} + o(1))\mathbb{P}(Y \in V).$$

Ignoring difficulties with independence, this suggests the inequality

$$\mathbb{P}(X_{k+1}, \ldots, X_n \text{ span } V) \leq c^n \mathbb{P}(Y_{k+1}, \ldots, Y_n \text{ span } V)$$

for some $1/2 < c < 1$. Summing over non-saturated subspaces $V$ and using the trivial bound

$$\sum_{V \text{ unsaturated} \atop \text{codim } V = k} \mathbb{P}(Y_{k+1}, \ldots, Y_n \text{ span } V) \leq 1$$

would complete the argument.

Over the finite field $\mathbb{F}_q$ we cannot quite get the above inequality, but rather an inequality of the form

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

This reflects our intuition that Fourier analysis over $\mathbb{F}_q$ controls errors from univormity rather than absolute probabilities. If we want to use this inequality to get an exponential strength gain, then we must require $\mathbb{P}(X \in V) - q^{-k} > Dq^{-k}$ for some $D > 0$. It turns out that this is enough for the argument to work.

Let $\nu$ denote a probability distribution to be chosen later. Suppose $\nu$ is $\beta$-dense for some $\beta > 0$; we will later show that $\beta = \alpha/8$. Let $Y_1, \ldots, Y_r \in \mathbb{F}_q^n$ be iid random vectors with iid entries taken from $\nu$ and let $Z_1, \ldots, Z_s \in \mathbb{F}_q^n$ be iid copies of $X$. Here $r, s$ are parameters to be chosen later.
We will need control over $\mathbb{P}(X \in V)$ in the sequel. We therefore make the following definition, first given in [13].

**Definition 2.6.** Let $V$ be a deterministic subspace in $\mathbb{F}_q^n$. We say that $V$ has combinatorial codimension $d_\pm \in \mathbb{Z}^+ / n$ and write $d_\pm(V) = d_\pm$ if

$$(1 - \alpha)^{d_\pm} \leq \mathbb{P}(X \in V) < (1 - \alpha)^{d_\pm - 1/n}$$

Note that the combinatorial codimension of a subspace depends on the choice of $\alpha$ and $\mu$. There are $O(n^2)$ possible combinatorial codimensions, so it suffices to control each separately.

In this section, we will assume that $X_{k+1}, \ldots, X_n$ are conditioned to be linearly independent.

Fix an unsaturated subspace $V$ with codimension $k$ and combinatorial codimension $d_\pm$. Let $B_V$ denote the event

$$B_V := \text{"$Y_1, \ldots, Y_r, Z_1, \ldots, Z_s$ are linearly independent in $V$."}$$

By probabilistic independence we can write

$$\mathbb{P}(W_k = V) = \frac{\mathbb{P}(B_V \land W_k = V)}{\mathbb{P}(B_V)}.$$ 

If $X_{k+1}, \ldots, X_n$ span $V$, we can find $n - k - r - s$ columns that complete $Y_1, \ldots, Y_r, Z_1, \ldots, Z_s$ to a basis for $V$. The remaining vectors must also lie in $V$. We therefore define the event

$$C_V := \text{"$X_{k+r+s+1}, \ldots, X_n, Y_1, \ldots, Y_r, Z_1, \ldots, Z_s$ span $W$."}$$

so after relabeling the columns of $A$,

$$\mathbb{P}(B_V \land W = V) \leq \left(\frac{n - k}{r + s}\right) \mathbb{P}(X_{k+1}, \ldots, X_{k+r+s} \in V) \mathbb{P}(C_V)$$

By Proposition 2.3, recalling that our vectors $X_{k+1}, \ldots, X_n$ are conditioned to be linearly independent,

$$\mathbb{P}(X_{k+1}, \ldots, X_{k+r+s} \in V) \leq \mathbb{P}(X \in V)^{r+s}.$$ 

Next we consider $\mathbb{P}(B_V)$. We can write by conditional expectation

$$\mathbb{P}(B_V) = \mathbb{P}(B_V \mid Y_1, \ldots, Y_r, Z_1, \ldots, Z_s \in V) \mathbb{P}(Y \in V)^r \mathbb{P}(Z \in V)^s.$$ 

We need to control the probability that the vectors $Y_1, \ldots, Y_r, Z_1, \ldots, Z_s$ are linearly independent. It turns out that Odlyzko’s lemma is strong enough for what we need, as long as $r$ and $s$ are not too large and the combinatorial codimension is not too small.

**Proposition 2.7.** Let $Y_1, \ldots, Y_r$ be iid vectors taken from a $\beta$-dense probability distribution $\nu$ and let $Z_1, \ldots, Z_s$ be iid vectors taken from an $\alpha$-dense probability distribution $\mu$. Then if $V$ has combinatorial codimension $d_\pm \leq O_{\alpha, \beta}(n)$ we have

$$\mathbb{P}(B_V \mid Y_1, \ldots, Y_r, Z_1, \ldots, Z_s \in V) \geq \frac{1}{2}.$$ 

**Proof.** Define the events

$$F_V := \text{"$Y_1, \ldots, Y_r, Z_1, \ldots, Z_s \in V$."}$$

and, for convenience,

$$F_V(i) := \text{"$F_V \land Y_1, \ldots, Y_{i-1}$ are linearly independent"}$$

$$\bar{F}_V(j) := \text{"$F_V \land Y_1, \ldots, Y_r, Z_1, \ldots, Z_{j-1}$ are linearly independent."}$$
Expanding the probability with conditional expectation,
\[
P(B_V | F_V) = \prod_{i=1}^{r} P(Y_i \notin \langle Y_1, \ldots, Y_{i-1} \rangle | F_V(i))
\times \prod_{j=1}^{s} P(Z_j \notin \langle Y_1, \ldots, Y_r, Z_1, \ldots, Z_{j-1} \rangle | \tilde{F}_V(j)).
\]

With Lemma 2.2,
\[
P(Y_i \notin \langle Y_1, \ldots, Y_{i-1} \rangle | F_V(i)) \geq 1 - (1 - \beta)^{n-i+1}(1 - \alpha)^{-d_{\pm}}
\]
and
\[
P(Z_j \notin \langle Y_1, \ldots, Y_r, Z_1, \ldots, Z_{j-1} \rangle | \tilde{F}_V(j)) \geq 1 - (1 - \alpha)^{n-r-j+1}(1 - \alpha)^{-d_{\pm}}
\]
We therefore have the lower bound
\[
P(B_V | F_V) \geq 1 - (1 - \beta)^{n-i+1}(1 - \alpha)^{-d_{\pm}} - (1 - \alpha)^{n-r-j+1}(1 - \alpha)^{-d_{\pm}} \geq 1/2
\]
as long as \(d_{\pm}\) is sufficiently small and \(r, s\) are sufficiently small. \(\square\)

Collecting estimates, we have
\[
P(W = V) \lesssim \left(\frac{n-k}{r+s}\right)^r \frac{P(X \in V)^r}{P(Y \in V)^r} P(C_V)
\]
We are now ready to state the key lemma to compare the random vectors \(X\) and \(Y\).

**Lemma 2.8 (Swapping).** There is a \(\beta\)-dense probability distribution \(\nu\) on \(\mathbb{F}_q^n\) with \(\beta = \alpha/8\) such that, if \(Y \in \mathbb{F}_q^n\) is a random vector with iid coefficients distributed according to \(\nu\), then
\[
|P(X \in V) - q^{-1}| \leq \left(\frac{1}{2} + o(1)\right) |P(Y \in V) - q^{-1}|.
\]

If \(V\) is unsaturated, then as an immediately corollary we have
\[
P(X \in V) \leq \left(\frac{1}{2} + \frac{1}{D} + o(1)\right) P(Y \in V)
\]
We will prove this lemma in Section 3. With this estimate, we can sum over all subspaces of codimension \(k\) and combinatorial codimension \(e\). Since a set of vectors can span at most one subspace, the events \(C_V\) for \(V\) varying are disjoint and we can conclude
\[
\sum_{V: \text{codim} V = k \atop d_{\pm}(V) = d_{\pm}} P(W = V) \lesssim \left(\frac{n-k}{r+s}\right)^r 2^{-r} = O(e^{-cn}).
\]
Here we picked \(r = \delta_1 n, s = n - k - r - \delta_2 n\). \(\square\)

3. **Littlewood-Offord Theorems**

We now come to the heart of the argument: proving the three Littlewood-Offord type lemmas used in the preceding section.

We briefly review some theory from additive combinatorics. For more discussion, see [12].

The following cosine inequality is elementary.

**Lemma 3.1.** For all positive integers \(k\) and for any \(\beta_1, \ldots, \beta_k \in \mathbb{R}\) we have the inequality
\[
\cos(\beta_1 + \cdots + \beta_k) \geq k \sum_{\ell=1}^{k} \cos \beta_\ell - k^2 + 1.
\]
Proof. We can assume that $-\pi/2 \leq \beta_\ell \leq \pi/2$ for all $\ell$, as otherwise the inequality is trivial. On this interval $\cos$ is concave, so we have the inequality

$$k^{-1} \sum_{\ell=1}^{k} \cos \beta_\ell \leq \cos \left( \frac{\beta_1 + \cdots + \beta_k}{k} \right).$$

It suffices to show that

$$\cos(\beta/k) \leq k^{-2} \cos \beta + 1 - k^{-2}$$

for all $\beta \in \mathbb{R}$, but this is immediate from the power series. \qed

Let $\mu$ be a probability measure on the finite field $\mathbb{F}_q$. We need estimates on the Fourier transform

$$\hat{\mu}(\psi) := \sum_{t \in \mathbb{F}_q} \mu(t) \psi(t).$$

Recall that $\mathbb{F}_q \cong \mathbb{F}_q$ via the isomorphism that sends $t \in \mathbb{F}_q$ to the character $x \mapsto e_p(\Tr(tx))$, where $\Tr : F_{p^\ell} \to F_p$ is the field trace. We define the additive spectrum $\text{Spec}_{1-\epsilon} \mu$ to be the set

$$\text{Spec}_{1-\epsilon} \mu := \{ \psi \in \hat{\mathbb{F}}_q \mid |\hat{\mu}(\psi)| \geq 1 - \epsilon \}.$$

For $\epsilon$ small, we can find additive structure in $\text{Spec}_{1-\epsilon} \mu$. The next lemma makes this explicit; see Lemma 4.37 in [12].

**Lemma 3.2.** For $\epsilon_1, \ldots, \epsilon_k < 1$ we have the sum-set inclusion

$$\text{Spec}_{1-\epsilon_1} \mu + \cdots + \text{Spec}_{1-\epsilon_k} \mu \subseteq \text{Spec}_{1-k(\epsilon_1+\cdots+\epsilon_k)} \mu.$$

**Proof.** Let $\psi_\ell \in \text{Spec}_{1-\epsilon_\ell} \mu$ for each $\ell$. We write $\psi_\ell(t) = e(\Tr(s_\ell t)/p)$ for appropriate $s_\ell$. We can find $\theta_\ell \in \mathbb{R}/\mathbb{Z}$ so that

$$\Re \sum_{t \in \mathbb{F}_q} \mu(t)e(\Tr(s_\ell t)/p + \theta_\ell) \geq 1 - \epsilon_\ell.$$

Summing, we derive

$$\Re \sum_{t \in \mathbb{F}_q} \mu(t) \left( ke(\Tr(s_1 t)/p + \theta_1) + \cdots + ke(\Tr(s_k t)/p + \theta_k) - k^2 + 1 \right) \geq 1 - k(\epsilon_1 + \cdots + \epsilon_k).$$

The result now follows from Lemma 3.1. \qed

A subset $A \subseteq Z$ of an abelian group induces a symmetry subgroup of $Z$ given by

$$\text{Sym} A := \{ h \in Z \mid h + A = A \}.$$

Clearly $A$ can be decomposed into the union of cosets of $\text{Sym} A$.

We need to bound sumsets from below. For $q = p$ the estimate we need is the Cauchy-Davenport inequality: any $A, B \subseteq \mathbb{F}_p$ satisfy $|A + B| \geq \min(|A| + |B| - 1, p)$. The next lemma generalizes the Cauchy-Davenport inequality to non-cyclic groups; see Theorem 5.5 in [12] for a proof.

**Lemma 3.3 (Knese’s Theorem).** Let $A, B \subseteq Z$ be finite subsets of an abelian group $Z$. We have the lower bound

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

Since $\text{Sym}(A_1 + \cdots + A_k)$ is increasing in $k$, we get the following iterated version.

**Corollary 3.4.** Let $A_1, \ldots, A_k \subseteq Z$ be finite subsets of an abelian group $Z$. We have the lower bound

$$|A_1 + \cdots + A_k| + (k-1)|\text{Sym}(A_1 + \cdots + A_k)| \geq |A_1| + \cdots + |A_k|.$$
3.1. The Classical Littlewood-Offord Estimate. We start by bounding the concentration probability \( \mathbb{P}(X \cdot w = r) \) for arbitrary \( r \in \mathbb{F}_q \), \( X \in \mathbb{F}_q^n \) a random vector with iid entries taken from an \( n \)-dense probability measure \( \mu \), and \( w \in \mathbb{F}_q^n \) a vector with at least \( m \) non-zero entries.

Proof of Lemma 2.4. Let \( \xi_1, \ldots, \xi_n \) denote the entries of \( X \). We can decompose the concentration probability into its Fourier transform,

\[
\mathbb{P}(X \cdot w = r) = q^{-1} + q^{-1} \sum_{t \in \mathbb{F}_q \setminus \{0\}} e_p(\text{Tr}(-rt)) \prod_{\ell=1}^n \mathbb{E} e_p(\text{Tr}(\xi_\ell w_\ell t)).
\]

By the triangle inequality,

\[
|\mathbb{P}(X \cdot w = r) - q^{-1}| \leq q^{-1} \sum_{t \in \mathbb{F}_q \setminus \{0\}} \prod_{\ell=1}^n |\mathbb{E} e_p(\text{Tr}(\xi_\ell w_\ell t))|
\]

Note that \( \mathbb{E} e_p(\text{Tr}(\xi_\ell w_\ell t)) = \hat{\mu}(w_\ell t) \).

We define \( \psi(t) := 1 - |\hat{\mu}(t)|^2 \) so that, with the inequality \( |\theta| \leq \exp(-\frac{1}{2}(1 - \theta^2)) \), we have

\[
|\mathbb{P}(X \cdot w = r) - q^{-1}| \leq q^{-1} \sum_{t \in \mathbb{F}_q \setminus \{0\}} \exp \left( -\frac{1}{2} \sum_{\ell=1}^n \psi(w_\ell t) \right)
\]

Put \( f(t) := \sum \psi(w_\ell t) \). We can decompose the sum into level sets,

\[
|\mathbb{P}(X \cdot w = r) - q^{-1}| \leq \frac{1}{2} \int_0^\infty q^{-1}|\{t \neq 0 \mid f(t) \leq v\}| e^{-v/2} dv.
\]

Let \( T(v) := \{ t \mid f(t) \leq v \} \) and \( T'(v) := T(v) \setminus \{0\} \).

We claim the following sum-set inequality; see [4] for the torsion-free case.

**Proposition 3.5.** For any \( v > 0 \), we have the inclusion

\[
T(v) + \cdots + T(v) \subseteq T(k^2 v)
\]

where there are \( k \) terms in the sum.

**Proof.** We first observe that for any \( \beta_1, \ldots, \beta_k \in \mathbb{F}_q \), we have the inequality

\[
\psi(\beta_1 + \cdots + \beta_k) \leq k(\psi(\beta_1) + \cdots + \psi(\beta_k)).
\]

In fact, we can rewrite this equation as

\[
1 - \sum_{a, b \in \mathbb{F}_q} \mu(a)\mu(-b) \cos(\frac{2\pi}{p} \text{Tr}((a + b)(\beta_1 + \cdots + \beta_k)))
\]

\[
\leq k^2 - k \sum_{j=1}^k \sum_{a, b \in \mathbb{F}_q} \mu(a)\mu(-b) \cos(\frac{2\pi}{p} \text{Tr}((a + b)\beta_j))
\]

which follows from Lemma 3.1.

Suppose \( t_1, \ldots, t_k \) satisfy \( f(t_k) \leq v \). Then we have

\[
f(t_1 + \cdots + t_k) = \sum_{t_1}^{n} \psi(w_\ell t_1 + \cdots + w_\ell t_k) \leq k \sum_{j=1}^k \sum_{\ell=1}^n \psi(w_\ell t_j) \leq k^2 v
\]

as required. \( \square \)
By Corollary 3.4 we deduce
\[ k |T(v)| \leq |T(k^2 v)| + (k - 1) |\text{Sym}(T(v) + \cdots + T(v))|. \]
This inequality is effective as long as \(|\text{Sym}(T(v) + \cdots + T(v))| = 1\). If \(\text{Sym}(T(v) + \cdots + T(v)) \neq \{0\}\), then because \(T(v) + \cdots + T(v) \subseteq T(k^2 v)\) we can find a non-trivial additive subgroup \(H < \mathbb{F}_q\) contained in the set \(T(k^2 v)\). It therefore suffices to choose \(k\) such that \(T(k^2 v)\) contains no non-trivial additive subgroups.

Fix \(H\); we will find a \(t \in H\) where \(f\) is large. Averaging \(f\) over the subgroup,
\[
|H|^{-1} \sum_{t \in H} f(t) = \sum_{\ell=1}^n |H|^{-1} \sum_{t \in H} \psi(w_\ell t) = \sum_{\ell=1}^n |H|^{-1} \sum_{t \in H} (1 - |\hat{\mu}(w_\ell t)|^2).
\]
By the inverse Fourier transform and the \(\alpha\)-density of \(\mu\),
\[
|H|^{-1} \sum_{t \in H} |\hat{\mu}(w_\ell t)|^2 = \sum_{\xi, \zeta \in \mathbb{F}_q} \mu(\xi)\mu(\zeta) 1_{H^\perp}(w_\ell(\xi - \zeta)) \leq 1 - \alpha
\]
Since at least \(m\) of the coefficients \(w_\ell\) are non-zero,
\[
|H|^{-1} \sum_{t \in H} f(t) \geq \alpha m.
\]
By the pigeonhole principle, there must be a \(t \in H\) with \(f(t) \geq \alpha m\).

We therefore conclude that
\[
|T'(v)| \lesssim \sqrt{\frac{|v|}{\alpha m}} |T'(\alpha m)|
\]
for all \(v \leq \alpha m\). Inserting this inequality into the level set estimate gives the bound
\[
|\mathbb{P}(X \cdot w \equiv r) - q^{-1}| \lesssim \frac{1}{\sqrt{\alpha m}} \int_0^\infty \sqrt{v} e^{-v} dv + e^{-\alpha m/2}
\]
as required. \(\square\)

3.2. The Inverse Theorem. We can find our structured perpendicular vector \(\xi\) with the pigeonhole principle.

Proof of Proposition 2.5. Let \(k := \text{codim} V\). We take Fourier transforms to find
\[
|\mathbb{P}(X \in V) - q^{-k}| \leq q^{-k} \sum_{\xi \in V^\perp \setminus \{0\}} \prod_{\ell=1}^n \left| \hat{\mu}(\xi_\ell) \right|
\]
By the pigeonhole principle, we can bound this above by
\[
\prod_{\ell=1}^n \left| \hat{\mu}(\xi_\ell) \right|
\]
for some fixed \(\xi \in V^\perp \setminus \{0\}\). Since \(V\) is not sparse, \(|\text{supp} \xi| \geq \delta n\).

Because \(V\) is semi-saturated, we get the lower bound
\[
e^{-dn} \leq \prod_{\ell=1}^n \left| \hat{\mu}(\xi_\ell) \right|.
\]
With the estimate \(|\theta| \leq \exp(-\frac{1}{2}(1 - \theta^2))\) we can take logarithms to find
\[
\sum_{\ell=1}^n 1 - |\hat{\mu}(\xi_\ell)|^2 \leq dn.
\]
Let \( \epsilon = 5d \). We can choose \( \sigma \subseteq [n] \) with \( |\sigma| \geq 0.9n \) such that
\[
\xi_{\ell} \in \text{Spec}_{1-\epsilon} \mu
\]
for all \( \ell \in \sigma \).

It suffices to find an absolute \( \eta > 0 \) such that \( |\text{Spec}_{1-\eta} \mu| \leq \beta \eta \). We observe that there is a value \( \gamma > 0 \) such that \( \text{Spec}_{1-\gamma} \mu \) does not contain any non-trivial additive subgroups \( H < \mathbb{F}_q^n \). In fact, by Markov’s inequality and Fourier inversion,
\[
(1 - \gamma)^2 \# H \cap \text{Spec}_{1-\gamma} \mu \leq \sum_{t \in H} |\hat{\mu}(t)|^2 \leq |H|(1 - \alpha).
\]

We then choose \( \gamma = \alpha/2 \).

We can now use Corollary 3.4 to show that for any \( k \geq 1 \),
\[
|\text{Spec}_{1-\gamma} \mu \setminus \{0\}| \geq k |\text{Spec}_{1-2\gamma} \mu \setminus \{0\}|
\]
so we pick \( k = \beta^{-1} \) and let \( \delta = \beta^2 \gamma = \beta^2 \alpha/2 \). We then deduce that \( \beta = \sqrt{c/(5\alpha)} \). \( \square \)

### 3.3. The Swapping Lemma

Let \( \mu \) be an \( \alpha \)-dense probability distribution and \( V \) an unsaturated subspace of codimension \( k \). We want to find \( \nu \) depending only on \( \mu \) such that, if \( Y \) is a random vector with iid entries taken from \( \nu \), we have the inequality
\[
|\mathbb{P}(X \in V) - q^{-k}| \leq \left( \frac{1}{2} + o(1) \right) |\mathbb{P}(Y \in V) - q^{-k}|
\]
Let us postpone the definition of \( \nu \) and define functions \( f,g : \mathbb{F}_q \to \mathbb{R}^+ \) to be
\[
f(t) = \prod_{\ell=1}^{n} |\hat{\mu}(w_{\ell} t)| \quad g(t) = \prod_{\ell=1}^{n} \hat{\nu}(w_{\ell} t).
\]
The lemma would follow immediately if we had \( \hat{\nu} \geq 0 \) and we could establish
\[
\sum_{t \in V^\perp \setminus \{0\}} f(t) \leq \left( \frac{1}{2} + o(1) \right) \sum_{t \in V^\perp \setminus \{0\}} g(t).
\]
Let \( F(u) = \{ t \mid f(t) \geq u \} \) and \( G(u) = \{ t \mid g(t) \geq u \} \) denote level sets. We define \( \nu \) so that the level sets \( G(u) \) control the additive structure of \( F(u) \).

**Proposition 3.6.** There is a probability distribution \( \nu : \mathbb{F}_q \to [0,1] \) depending on \( \mu \) and \( \alpha \) with the following properties.

1. For all \( 0 < u < 1 \) we have the sumset inclusion \( F(u) + F(u) \subseteq G(u) \).
2. For all \( t \in V^\perp \), \( f(t) \leq g(t)^4 \).
3. \( \hat{\nu}(t) \geq 0 \) for all \( t \in \mathbb{F}_q \).
4. \( \nu \) is \( \beta \)-dense for \( \beta = \alpha/8 \).

We will prove Proposition 3.6 in a moment. First we will show how to use Proposition 3.6 to prove Lemma 2.8.

**Proof of Lemma 2.8.** Let \( \epsilon > 0 \) be determined later. We decompose the sum of \( f \) into the domains where \( f \leq \epsilon \) and \( f > \epsilon \),
\[
\sum_{t \in V^\perp \setminus \{0\}} f(t) \leq \sum_{t \in V^\perp \setminus \{0\}, f(t) \leq \epsilon} f(t) + \sum_{t \in V^\perp \setminus \{0\}, f(t) > \epsilon} f(t).
\]
Therefore if we can set \( \epsilon = o(1) \) as \( n \to \infty \) this part is complete.

We write the sum over the domain where \( f > \epsilon \) into level sets,

\[
\sum_{\substack{t \in V^\perp \setminus \{0\} \\mid f(t) > \epsilon}} f(t) = \int_\epsilon^\infty |F'(u)| \, du + \epsilon |F'(\epsilon)|
\]

Here we let \( F'(u) := F(u) \setminus \{0\} \) and similarly define \( G'(u) := G(u) \setminus \{0\} \).

From the sumset inequality \( F(u) + F(u) \subseteq G(u) \) and Kneser’s inequality,

\[
2|F(u)| \leq |\text{Sym}(F(u) + F(u))| + |G(u)|
\]

We would like to pick \( \epsilon = o(1) \) such that \( |\text{Sym}(F(u) + F(u))| = 1 \). Since \( F(u) + F(u) \subseteq G(u) \) and \( G(u) \) is increasing in \( u \), we require that every non-trivial additive subgroup \( H < V^\perp \) contain a non-zero element \( t \notin G(\epsilon) \).

Fix \( H < V^\perp \). We can clearly assume that \( H \cong \mathbb{Z}/p\mathbb{Z} \); pick \( w \in V^\perp \) that generates \( H \). Since \( V \) is unsaturated, we know that \( w \) contains at least \( \delta n \) non-zero entries.

Define the function

\[
h(t) := \sum_{\ell=1}^n 1 - \tilde{\nu}(t_\ell)^2.
\]

for \( t \in H \). Averaging \( h \) over \( H \), we can argue as in the proof of Lemma 2.4 to find

\[
|H|^{-1} \sum_{t \in H} h(t) \geq \beta \delta n.
\]

Note that we need \( \nu \) to be \( \beta \)-dense. By the pigeonhole principle we can find a (non-zero) \( t \in H \) with \( h(t) \geq \beta \delta n \). We then deduce that

\[
g(t) \leq \exp\left(-\frac{1}{2} h(t)\right) \leq \exp\left(-\frac{1}{2} \beta \delta n\right)
\]

so we set \( \epsilon = \exp\left(-\frac{1}{2} \beta \delta n\right) \). For every \( u \geq \epsilon \) we now have

\[
2|F'(u)| \leq |G'(u)|,
\]

so returning to our integral of level sets we find

\[
\int_{\epsilon}^\infty |F'(u)| \, du + \epsilon |F'(\epsilon)| \leq \frac{1}{2} \int_0^\infty |G'(u)| \, du.
\]

The lemma now follows. \( \square \)

**Proof of Proposition 3.6.** Let \( \gamma = 1/8 \) be a parameter and define

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

Clearly \( \nu \) is a probability measure if \( 0 < \gamma < 1 \). We also have \( \tilde{\nu} > 1 - 2\gamma \). Let \( \beta = \gamma \alpha \). It is easy to see that \( \nu \) is \( \beta \)-dense: for \( H < \mathbb{F}_q \) additive we have

\[
\nu(H) = 1 - \sum_{t \notin H} \gamma \mu \ast \mu^-(t) \leq 1 - \gamma \alpha
\]
and for any $x \notin H$ we have
\[
\nu(x + H) = \sum_{t \in x + H} \gamma \mu * \mu^-(t) \leq \gamma(1 - \alpha) \leq 1 - \gamma \alpha.
\]
as desired.

The Fourier transform of $\nu$ is given by
\[
\hat{\nu}(\xi) = 1 - \gamma + \gamma|\hat{\mu}(\xi)|^2.
\]
We would next like to show that $F(u) + F(u) \subseteq G(u)$ for all $0 < u < 1$. It suffices to show, for all $\theta, \psi \in \hat{F}_q$,
\[
|\hat{\mu}(\theta)\hat{\mu}(\psi)| \leq \hat{\nu}(\theta + \psi)^2.
\]
We will consider two cases.

1. Suppose $|\hat{\mu}(\theta)| < 1 - 4\gamma$ or $|\hat{\mu}(\psi)| < 1 - 4\gamma$. Then
\[
|\hat{\mu}(\theta)\hat{\mu}(\psi)| < 1 - 4\gamma < (1 - 2\gamma)^2 < \hat{\nu}(\theta + \psi).
\]

2. Now suppose that $|\hat{\mu}(\theta)|, |\hat{\mu}(\psi)| \geq 1 - 4\gamma$. Define $\theta_1 = 1 - |\hat{\mu}(\theta)|$ and $\theta_2 = 1 - |\hat{\mu}(\psi)|$. By Lemma 3.2, we know that $|\hat{\mu}(\theta + \psi)|^2 \geq 1 - 2(\theta_1 + \theta_2)$.

We have the inequality
\[
\hat{\nu}(\theta + \psi) = 1 - \gamma + \gamma|\hat{\mu}(\theta + \psi)|^2 \geq 1 - 4\gamma(\theta_1 + \theta_2)
\]
Since we have $\gamma = 1/8$, we conclude that
\[
\hat{\nu}(\theta + \psi)^2 \geq |\hat{\mu}(\theta)\hat{\mu}(\psi)|
\]
as required.

It remains to show that $|\hat{\mu}(\theta)| \leq \hat{\nu}(\theta)^4$ for all $\theta$. By the geometric-arithmetic mean inequality,
\[
(|\hat{\mu}(\theta)|^2 \cdot 1^7)^{1/8} \leq \frac{1}{8}(|\hat{\mu}(\theta)|^2 + 7) = \hat{\nu}(\theta)
\]
as required. \hfill \Box

4. Probability distribution of the determinant

We will now indicate how to modify the proof of Theorem 1.2 to prove Theorem 1.3.

Again let $X_1, \ldots, X_n$ denote the columns of $A$. We begin by revealing all but the first column of the matrix. If we abbreviate $W := \langle X_2, \ldots, X_n \rangle$ then we find
\[
P(\det M = t) = P(\det M = t | \text{codim } W = 1)P(\text{codim } W = 1)
\]
We now use Proposition 2.1 to control the last $n - 1$ vectors,
\[
P(\text{codim } W = 1) = \prod_{k=2}^{n} P(X_k \notin \langle X_{k+1}, \ldots, X_n \rangle | \text{codim } \langle X_{k+1}, \ldots, X_n \rangle = k)
\]
\[
= \prod_{k=2}^{\infty} (1 - q^{-k}) + O(e^{-cn}).
\]

Pick $w \perp W$ such that $\det A = X_1 \cdot w$; namely, $w$ is the first row of the adjugate of $A$.
We can classify the possible hyperplanes $V$ that $W$ can represent. These are similar to the definitions made in Section 2, but the definition of semi-saturated has been expanded.

**sparse:** We have $|\text{supp } w| \leq \delta n$. Note that this is well-defined independent of the choice of $w \perp V$. 


unsaturated: $V$ is not sparse and either
\[
\max(e^{-d\alpha n}, Dq^{-1}) \leq |\mathbb{P}(X \in V) - q^{-1}|
\]
or
\[
Dq^{-1} \leq \mathbb{P}(X \in V) \leq e^{-d\alpha n} \leq \mathbb{P}(X \cdot w = t)
\]
for some $t \in \mathbb{F}_q$.

semi-saturated: $V$ is not sparse,
\[
|\mathbb{P}(X \in V) - q^{-1}| < Dq^{-1}
\]
and there is a $t \in \mathbb{F}_q$ with
\[
e^{-d\alpha n} < |\mathbb{P}(X \cdot w = t) - q^{-1}|.
\]
We can control these with a modified the inverse theorem.

saturated: $V$ is not sparse and
\[
|\mathbb{P}(X \cdot w = t) - q^{-1}| \leq e^{-d\alpha n}
\]
for all $t \in \mathbb{F}_q$.

We will now show that $W$ represents sparse, semi-saturated, and unsaturated subspaces with probability $O(e^{-c\alpha n})$.

4.1. Sparse subspaces. The argument in Section 2.1 shows that these occur with probability $O(e^{-c\alpha n})$.

4.2. Unsaturated subspaces. Since $\mathbb{P}(X \cdot w = t) \leq \mathbb{P}(Y \in V)$, we see that regardless of which set of inequalities hold, we have
\[
Dq^{-1} \leq \mathbb{P}(X \in V)
\]
and
\[
e^{-d\alpha n} \leq \mathbb{P}(Y \in V).
\]
Therefore the argument from Section 2.3 applies, so that unsaturated subspaces appear with probability $O(e^{-c\alpha n})$.

4.3. Semi-saturated subspaces. For all $t \in \mathbb{F}_p$ we can calculate
\[
\mathbb{P}(X \cdot w = t) = q^{-1} \sum_{\xi \in \mathbb{Z}/(p)} e_p(-\text{Tr}(t\xi)) \prod_{\ell=1}^{n} \mathbb{E}_p(\text{Tr}(\psi_w t\xi))
\]
Rearranging and applying the triangle inequality,
\[
|\mathbb{P}(X \cdot w = t) - q^{-1}| \leq q^{-1} \sum_{\xi \in \mathbb{F}_p \setminus \{0\}} \prod_{\ell=1}^{n} |\cos(2\pi w t\xi)|
\]
The argument can now be completed as in Theorem 1.2.

5. Acknowledgments

The author thanks Terence Tao for guidance and helpful conversation.
References


Department of Mathematics, University of California, Los Angeles, CA 90024

E-mail address: maples@math.ucla.edu