Random Vectors of Bounded Weight and Their Linear Dependencies

Nathan Linial * Dror Weitz †

December 16, 2000

Abstract

Let μ be a probability distribution on a vector space V. When m vectors $u_1, \ldots, u_m$ are drawn from μ, how likely are they to be linearly dependent? How is the dimension of their linear span distributed? Such questions have been addressed in a number of papers (e.g. [1],[2],[3],[6],[7]). Our work is motivated by problems in coding theory, and we address these problems in the following context: Here $V = F_q^n$, the n-dimensional vector space over the field of order q and the distribution μ is uniform over the set of vectors with Hamming weight $\leq w$. Let $M_{m\times n}$ be a random matrix whose rows $u_1, \ldots, u_m$ are sampled independently from μ. We investigate two associated random variables: (i) The rank of such a random matrix M, (ii) The cardinality of kernel(M). Finally, we consider the distribution of random sums of such randomly chosen vectors $u_1, \ldots, u_m$.

Of particular interest to us is to find the least Hamming weight w where the restriction on the vectors' weights hardly matters. Namely, where the answers become nearly identical with the case $w = n$, in which vectors are selected uniformly from the entire space.

*Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel. Work supported in part by grants from the Binational Israel-US Science Foundation and the Israel Academy of Science.

†Institute of Computer Science, Hebrew University, Jerusalem 91904, Israel.

1 Introduction

Our notation is rather standard: $F_q = GF(q)$ is the field of order q. The collection of $m \times n$ matrices over F_q is denoted by $M_{m,n,q}$, and the same set endowed with a uniform distribution is the probability space $\Omega_{m,n,q}$. The *rank* of matrices is thus viewed as an integer-valued random variable on $\Omega_{m,n,q}$. The distribution of the rank is known and easy to calculate. Namely, the probability that a matrix A drawn from $\Omega_{m,n,q}$ has rank r is exactly:

$$\frac{1}{q^{(n-r)(m-r)}} \prod_{i=0}^{r-1} \frac{(1-q^{i-n})(1-q^{i-m})}{1-q^{i-r}} \tag{1}$$

A sketchy proof for this standard fact is provided in the Appendix. In contrast, similar problems for real matrices are much more difficult. See Kahn, Komlós and Szemerédi [6] and the references therein.

Motivated by problems from Coding Theory, we study similar questions for matrices whose rows satisfy certain bounds on their Hamming weights. (The *Hamming weight* of a vector $v \in F_q^n$, denoted $|v|$ is the number of non-zero coordinates in v.) Let:

$$W = W(w,n,q) = \left\{ v \in F_q^n \mid |v| \leq w \right\}$$

be the set of n-dimensional vectors over F_q with Hamming weight bounded by w.

The set of those $m \times n$ matrices whose rows belong to $W(w,n,q)$ is called $M_{w,m,n,q}$, and this set endowed with uniform distribution, is the probability space $\Omega_{w,m,n,q}$. Clearly, $\Omega_{n,m,n,q} = \Omega_{m,n,q}$, and we ask how large w should be for the rank to be distributed over $\Omega_{w,m,n,q}$ in essentially the same way that it is distributed in $\Omega_{m,n,q}$ (i.e., very close to the expression given in Formula (1)). We do not answer this question in full, but give evidence in support of the conjecture that this happens iff $w \geq \ln n + \omega(1)$. Notice that $w = \ln n - \omega(1)$ is certainly too small. A standard coupon collector argument shows that when $m = n$, the (square) matrix almost surely has an all zeros column, and thus, its rank is less than n almost surely. This contradicts the fact that the probability of full rank is bounded away from 0 the matrix is selected uniformly at random with no restrictions on the weight (this fact is an easy consequence of Formula 1).

A closely related random variable that we investigate is the cardinality of the kernel of a matrix in $\Omega_{w,m,n,q}$. If the matrix has rank r, then there are q^{m-r} vectors in the kernel. It is shown that already for $w = \ln n + \omega(1)$, the expectation of this random variable, $\mathrm{E}(q^{m-r})$, is close to the expectation in the weight unbounded case where matrices are chosen from $\Omega_{m,n,q}$:

Theorem 1.1 *Let $\Omega = \Omega_{w,m,n,q}$ be the above probability space of $m \times n$ matrices. Consider the rank $r(\cdot)$ as a random variable on Ω. Then the expected cardinality of the kernel satisfies:*

$$1 + \frac{q^m - 1}{q^n} \;\leq\; \mathrm{E}(q^{m-r})$$

with equality when $w = n$. Moreover, if $w \geq \ln n + \omega(1)$, then for every m,

$$\mathrm{E}(q^{m-r}) \;\leq\; (1 + o(1))(1 + \frac{q^m - 1}{q^n})$$

as $n \to \infty$.

Theorem 1.1 implies, for sufficiently large w, that a random matrix is very likely to have full rank, or at least nearly full rank. This is expressed by the two results below. Recall that $f = \omega(g)$ means that $\frac{f(n)}{g(n)}$ tends to infinity with n. When $|n-m|$ grows to infinity (with n), full rank is almost sure:

Corollary 1.2 *If $w > \ln n + \omega(1)$ and if $|n - m| \geq \omega(1)$ then almost every matrix in $\Omega_{w,m,n,q}$ has full rank (i.e. $r = \min\{m, n\}$).*

In any case, even when $|n - m|$ is bounded, *almost* full rank is almost certain:

Corollary 1.3 *If $w > \ln n + \omega(1)$ and if $r' \leq \min\{m, n\} - \omega(1)$ then $\Pr(r \leq r') = o(1)$.*

Notice, in comparison, (and this easily follows from Formula 1) that if row weights are not restricted, then almost sure full rank is attained iff $|n - m| = \omega(1)$. Furthermore, if $|n - m|$ is bounded, $\Pr(r \leq r') \to 0$ (as n grows) iff $r' \leq \min\{m, n\} - \omega(1)$. In words, rank $\leq r'$ is vanishingly rare only when we are far from the case of full rank. (The Appendix contains the derivations of these facts about the unrestricted case). Thus, corollaries 1.2 and 1.3 imply this: Those values of the rank which, without restrictions on the weight, are (asymptotically) either extremely likely or extremely unlikely, exhibit the same qualitative behavior, provided that $w \geq \ln n + \omega(1)$.

This still leaves out those ranks that occur with probabilities which are bounded away from zero and one. Namely, when the differences among n, m, and r are all bounded. We do not know if here, too, the probabilities are asymptotically the same with and without restrictions on the weights. What we do know, however, is that restricting the weight, even to $\ln n + \omega(1)$, has almost (asymptotically) no effect on the expected size of the kernel of the matrix for any values of n and m (Theorem 1.1). It seems reasonable to suspect that the same bound on weights would lead to a distribution of the rank that agrees with the unrestricted situation. This statement receives some support from computer simulations that we have carried out.

Similar problems have been considered (e.g.[1], [3], and [7]) with a different probability distribution on random matrices. There, each matrix entry a_{ij} is independently drawn from some probability distribution P_{ij} on the elements of the field F_q, that is uniform on the non-zero elements of the field.

In [7], Kovalneko shows that for the binary case ($q = 2$), as long as $P_{ij}(1)$ is bounded away from zero and one, the distribution of the rank converges as $n \to \infty$ to the same distribution with $P_{ij}(1) = \frac{1}{2}$ (this is the same probability space as our $\Omega_{m,n,q}$). Cooper [3] improves this result and shows that for $m = n$ (the square matrix), the probability of a full rank matrix converges to the same probability with $P_{ij}(1) = \frac{1}{2}$ as long as $P_{ij}(0) \leq 1 - \frac{\ln n + \Delta}{n}$ where $\Delta = \Delta(n)$ grows to infinity slowly enough. We also mention Balakin [1], who shows, for general q, that under the same restriction on P_{ij} as the last one and when $|m - n| \geq \omega(1)$, the chosen matrix has full rank almost surely.

Remark 1.4 *Notice that the results mentioned here do not translate automatically into our random model. It would be reasonable to presume that our results for $\Omega_{w,m,n,q}$ should more-or-less coincide with those for $P_{ij}(0) = 1 - \frac{w}{n}$. However, the probability spaces of matrices generated by the two models are different, even asymptotically. For example, if in our model, $w = o(n)$ then the weight of the first row is exactly w with probability $1 - o(1)$. Clearly no such cocentration takes place at the P_{ij} model.*

Calkin [2] has considered a random model similar to ours. In his paper the m vectors are

chosen independently and uniformly among those with weight exactly w (whereas we choose from the vectors with weight *at most* w). One of the questions he addresses is this: How large can m be (as a function of w) for these vectors to still be almost surely independent (i.e., the matrix should have full rank)? In order to answer that question, Calkin calculates the expected size of the kernel. His answer implies that for $w = \omega(1)$, as long as $\frac{m}{n}$ is bounded away from 1, $E(q^{m-r}) \to 1$ as $n \to \infty$, and therefore, under these conditions, these vectors are almost surely independent. Our answer, in contrast, applies for *any* unbounded difference between n and m. This, however, makes it necessary for us to require a larger w (logarithmic in n). We still do not know what the exact conclusions are for $\log n \gg w \gg 1$. Perhaps $n - m > \max\{\omega(1), ne^{-w}\}$ is enough in general (this seems plausible and is consistent with what is known).

The proof of Theorem 1.1 is based on an analysis of the distribution of *sums* of vectors with bounded weights. Let $w_1, ..., w_m \leq n$ be nonnegative integers. We let $\mu = \mu_{w_1,...,w_m}$ be the probability distribution of the vector $v_1 \oplus ... \oplus v_m$ where each v_i is chosen independently and uniformly from $W(w_i, n, q)$. Specifically, we estimate how much $\mu(\vec{0})$, the probability of the zero vector deviates from uniform. We show:

Theorem 1.5 *Let $w_1, ..., w_m$ be nonnegative integers such that $\sum w_i \geq (\frac{q-1}{q})n \ln n + \omega(n)$. Then*

$$q^{-n} \leq \mu_{w_1,...,w_m}(\vec{0}) \leq (1 + o(1))q^{-n}$$

as $n \to \infty$.

The proof proceeds by first reducing the problem to the case where $w_i = 1$ for every i. When $q = 2$, this translates to a problem concerning random walks on the cube. To deal with general values of q, we need to adapt several known results (e.g. [4],[5],[9]) about this random walk.

2 The distribution of vector sums

2.1 Preliminaries

2.1.1 Harmonic Analysis on the q-Cube

The q-cube in the title is simply a vector space over the field F_q, the familiar case being $q = 2$. Henceforth, we refer to the q-cube simply as cube. If μ_{w_i} is the uniform distribution on $W(w_i, n, q)$, then $\mu_{w_1,\ldots,w_m}$ is the convolution of the distributions $\mu_{w_1}, \mu_{w_2}, \ldots$. We need to make some preliminary remarks on convolutions in general.

If f and g are real functions on F_q^n, then their convolution $f * g$, is the function:

$$f * g(x) = \sum_{y \in F_q^n} f(x \ominus y) g(y) \tag{2}$$

Sums and differences of vectors in F_q^n are denoted by $\oplus$ and $\ominus$.

Remark 2.1 *We purposely suppress a normalization factor in this definition. As mentioned below, this will make our notation more convenient.*

A function f on the cube is *symmetric* if $f(x)$ depends only on $|x|$, the Hamming weight of x. We say that a symmetric function is nonincreasing if it is nonincreasing in $|x|$.

Theorem 2.2 *The class of symmetric nonnegative nonincreasing real functions on the cube is closed under convolution.*

Proof: A symmetric nonnegative nonincreasing function on the cube, can be uniquely expressed as:

$$f = \sum_{i=1}^{n} c_i 1_{W(i,n,q)}$$

where $1_{W(i,n,q)}$ is the characteristic function of $W(i, n, q)$ and the coefficients c_i are nonnegative. The convolution of two such functions is, therefore,

$$f * g \;=\; (\sum_{i=1}^{n} c_i 1_{W(i,n,q)}) * (\sum_{i=1}^{n} d_i 1_{W(i,n,q)}) \;= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i d_j \; (1_{W(i,n,q)} * 1_{W(j,n,q)}).$$

Our claim clearly follows if we can show that the (symmetric nonnegative) function

$$h_{i,j} \equiv 1_{W(i,n,q)} * 1_{W(j,n,q)}$$

is nonincreasing. To this end, it suffices (by induction) to consider x and y whose Hamming weight differ by one, say $|x| = |y| + 1$, and show that $h_{i,j}(y) \geq h_{i,j}(x)$. Since $h_{i,j}$ is symmetric, it is enough to consider the case when $x \ominus y = e_1$ with $x_1 = 1$ and $y_1 = 0$. Since

$$h_{i,j}(x) \;=\; |(x \ominus W(i,n,q)) \cap W(j,n,q)|$$

we need to show that

$$|(x \ominus W(i,n,q)) \cap W(j,n,q)| \leq |(y \ominus W(i,n,q)) \cap W(j,n,q)|.$$

In order to do so, we define a bijection $\phi : y \ominus W(i,n,q) \longrightarrow x \ominus W(i,n,q)$ (these two sets have the same cardinality, of course) such that $|\phi(v)| \geq |v|$. Let $H = \{u | |u| = i \;\; and \;\; u_1 = 0\} \subset W(i,n,q)$ and let $L = W(i,n,q) \setminus H$. Let $v \in y \ominus W(i,n,q) = y \ominus u$, then define ϕ via:

$$\phi(v) = \left\{ \begin{array}{ll} v \oplus e_1 & u \in H \\ v & u \in L \end{array} \right\} \tag{3}$$

That ϕ is bijective follows, since $x \ominus W(i,n,q)$ is the disjoint union of $y \ominus L$ and $y \oplus e_1 \ominus H$ (notice that $y \ominus L = x \ominus L$). The requirement $|\phi(v)| \geq |v|$, has to be checked only at the first coordinate and for $u \in H$. Then, however, $v_1 = y_1 - u_1 = 0$, whereas $\phi(v)_1 = 1$. $\square$

Consider the Fourier Transform:

$$\widehat{f}(x) = \sum_{y \in F_q^n} \omega_q^{<x,y>} f(y) \tag{4}$$

where ω_q is the primitive unit root of order q.

It is well known that $\widehat{f * g} = \widehat{f}\widehat{g}$. It is here that the absence of normalization factors in our definitions of the transform and convolutions rather convenient.

Lemma 2.3

$$f_1 * ... * f_m(x) = q^{-n}[\prod_{i=1}^{m} (\sum_{y \in F_q^n} f_i(y)) \;\; + \;\; \sum_{y \neq \vec{0}} \omega_q^{<x,y>} \prod_{i=1}^{m} \widehat{f}_i(y) \;]$$

Proof: Start with

$$g \equiv \widehat{(f_1 * ... * f_m)} = \prod_{i=1}^{m} \widehat{f_i}.$$

Reapplying the transform, we can write:

$$f_1 * ... * f_m(x) = q^{-n}\widehat{g}(x) = q^{-n} \sum_{y \in F_q^n} \omega_q^{<x,y>} g(y) =$$

$$= q^{-n} \left[g(\vec{0}) \ + \ \sum_{y \neq \vec{0}} \omega_q^{<x,y>} g(y) \right] =$$

$$= q^{-n} \left[\prod_{i=1}^{m} \widehat{f_i}(\vec{0}) \ + \ \sum_{y \neq \vec{0}} \omega_q^{<x,y>} \prod_{i=1}^{m} \widehat{f_i}(y) \right] =$$

$$= q^{-n} \left[\prod_{i=1}^{m} \left(\sum_{y \in F_q^n} f_i(y) \right) \ + \ \sum_{y \neq \vec{0}} \omega_q^{<x,y>} \prod_{i=1}^{m} \widehat{f_i}(y) \right]. \quad \square$$

2.1.2 Krawtchouk Polynomials

The Krawtchouk polynomials $K_k^{(n,q)}$ are defined as follows:

$$K_k^{(n,q)}(x) = \sum_{j=0}^{k} (-1)^j \binom{x}{j} \binom{n-x}{k-j} (q-1)^{k-j} \tag{5}$$

It is known that K_k is the transform of the characteristic function of the k-th level of the cube (i.e., those vectors of Hamming weight k). Thus, if f is the characteristic function of $W(w,n,q)$, then

$$\widehat{f}(v) = \sum_{k=0}^{w} K_k^{(n,q)}(|v|) \tag{6}$$

We need the following standard fact on these polynomials (e.g. Equation (1.2.15) in [8]):

$$\sum_{k=0}^{w} K_k^{(n,q)}(x) = K_w^{(n-1,q)}(x-1) \tag{7}$$

We also recall the first Krawtchouk Polynomial:

$$K_1^{(n,q)}(x) = (q-1)n - qx \tag{8}$$

2.1.3 Domination

All vectors in this section are real and have real *nonnegative* entries. A vector $v = (v_1, v_2, \ldots)$ is called nonincreasing, if $v_1 \geq v_2 \geq \ldots$. The vector $v = (a_1, ..., a_k)$ is said to *dominate* $u = (b_1, ..., b_k)$ (denoted $v \succeq u$) iff

$$\sum_{j=1}^{i} a_j \;\geq\; \sum_{j=1}^{i} b_j$$

for every $1 \leq i \leq k-1$, and

$$\sum_{j=1}^{k} a_j \;=\; \sum_{j=1}^{k} b_j.$$

We make the following easy observations:

Lemma 2.4 *Let $v = (v_1, v_2, \ldots)$ be a nonincreasing real vector with exactly k positive coordinates. Define the vector u via: $u_1 = u_2 = \ldots = u_k = \frac{1}{k}\sum v_j$, and $u_{k+1} = u_{k+2} = \ldots = 0$. Then $v \succeq u$.*

The inner product of vectors u, v is denoted $< u, v >$.

Lemma 2.5 *Let u, v, w be real vectors of the same dimension. If v is nonnegative and nonincreasing, and $u \succeq w$, then $< v, u > \geq < v, w >$.*

2.2 Proof of Theorem 1.5

First we prove:

Lemma 2.6 *For any m and any $w_1, ..., w_m$, the (symmetric nonnegative) function $\mu_{w_1,...,w_m}$ is nonincreasing (and therefore $\mu_{w_1,...,w_m}(\vec{0}) \geq q^{-n}$).*

Proof: The proof is by induction on m and follows easily from Theorem 2.2. □

Note that the claim of Theorem 1.5 depends only on the sum $\sum w_i$ and not the specific distribution. What we show is that for a given sum, the worst case occurs when $w_i = 1$ for every i. We show this by considering what happens when a weight w_i is replaced by w_i repeats of the weight 1. Namely:

Lemma 2.7 *For any m integers $w_1, ..., w_m$ and any $1 \leq i \leq m$:*

$$\mu_{w_1,...,w_m}(\vec{0}) \leq \mu_{w_1,..,w_{i-1},1,1,..,1,w_{i+1},..,w_m}(\vec{0})$$

(There are w_i 1's in the r.h.s. expression.)

Proof: It will be useful to express

$$\mu_{w_1,..,w_m} = \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} * \mu_{w_i}$$

and

$$\mu_{w_1,..,w_{i-1},1,1,..,1,w_{i+1},..,w_m} = \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} * (\mu_1^{*w_i}).$$

The intuition underlying our proof is that μ_{w_i} is the uniform probability distribution on $W(w_i, n, q)$. On the other hand, $\mu_1^{*w_i}$ is some nonincreasing distribution over the same set. Thus, adding a random vector sampled from μ_{w_i} gets us further away, on average, than one sampled from $\mu_1^{*w_i}$.

We need to evaluate the relevant measures at $\vec{0}$. Note that for any two symmetric real functions on the cube $f * g(\vec{0}) =< f, g >$. This is because $x \oplus y = \vec{0}$ implies that $|x| = |y|$, and therefore, by symmetry, $g(x) = g(y)$. We therefore have:

$$\mu_{w_1,..,w_m}(\vec{0}) = < \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} \ , \ \mu_{w_i} >$$

and

$$\mu_{w_1,..,w_{i-1},1,1,..,1,w_{i+1},..,w_m}(\vec{0}) = < \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} \ , \ \mu_1^{*w_i} >$$

The notion of domination is easily extended to the realm of symmetric functions on the cube. Coordinates are points of the cube, and they are arranged in increasing order of Hamming weight. The internal ordering of points with equal weights is immaterial, when dealing with symmetric functions. We plan to deduce from Lemma 2.4 that $\mu_1^{*w_i} \succeq \mu_{w_i}$. Once this is shown, Lemma 2.5 implies

$$\mu_{w_1,..,w_{i-1},1,1,..,1,w_{i+1},..,w_m}(\vec{0}) \geq \mu_{w_1,..,w_m}(\vec{0}) \tag{9}$$

Let us verify the assumptions in Lemma 2.4. Indeed:

1) If $|v| > w_i$, then $\mu_{w_i}(v) = \mu_1^{*w_i}(v) = 0$

2) If $|v| \leq w_i$, then $\mu_{w_i}(v) = \frac{1}{|W(w_i,n,q)|}$ (independent of the vector v)

3) $\mu_1^{*w_i}$ is a symmetric nonnegative nonincreasing function

4) $|\mu_{w_i}| = |\mu_1^{*w_i}| = 1$ (probability measures)

By Lemma 2.4 $\mu_1^{*w_i} \succeq \mu_{w_i}$ holds, and since $\mu_{w_1,..,w_{i-1},w_{i+1},..,w_m}$ is a symmetric nonnegative nonincreasing function by Lemma 2.5:

$$< \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} \ , \ \mu_1^{*w_i} > \ \geq \ < \mu_{w_1,..,w_{i-1},w_{i+1},..,w_m} \ , \ \mu_{w_i} > .$$

Inequality 9 follows. $\square$

Repeated application of this lemma yields:

Corollary 2.8 *For any m and any nonnegative integers $w_1, ..., w_m$ whose sum is w:*

$$\mu_{w_1,...,w_m}(\vec{0}) \ \leq \ \mu_1^{*w}(\vec{0})$$

Now we need to extend a known fact about the rate of convergence of a random walk on the cube from $q = 2$ to general q. We use Fourier analysis to calculate the probability of the zero vector as done e.g., in [4] and [5], but we do it for general q.

Theorem 2.9 *If $m \geq \frac{(q-1)n+1}{q}(\ln n + b)$, then $\mu_1^{*m}(\vec{0}) \leq q^{-n}e^{(q-1)e^{-b}}$ for any b.*

Proof: Let

$$f = \mu_1 = \frac{\chi_1}{(q-1)n+1}$$

where χ_1 is the characteristic function of $W(1, n, q)$. According to Lemma 2.3:

$$f^{*m}(\vec{0}) = q^{-n}[\ (\sum_{u \in F_q^n} f(u))^m \ + \sum_{u \neq \vec{0}} (\widehat{f}(u))^m \ ]$$

Recall that

$$\widehat{f}(u) = \frac{1}{(q-1)n+1} \sum_{k=0}^{1} K_k^{(n,q)}(|u|) = \frac{1}{(q-1)n+1} K_1^{(n-1,q)}(|u|-1),$$

by Identity (7) from the preliminaries. The other term is easy to evaluate, since f is a probability measure, i.e., $\sum_u f(u) = 1$. Consequently,

$$f^{*m}(\vec{0}) = q^{-n}(1 + \sum_{u \neq \vec{0}} [\frac{K_1^{(n-1,q)}(|u|-1)|}{(q-1)n+1}]^m) \ =$$

$$= \ q^{-n}(1 + \sum_{x=1}^{n} (q-1)^x \binom{n}{x} [\frac{K_1^{(n-1,q)}(x-1)}{(q-1)n+1}]^m)$$

Now use the expression for K_1 (Equation (8) from the preliminaries) to conclude:

$$f^{*m}(\vec{0}) = q^{-n}\sum_{x=0}^{n}(q-1)^x\binom{n}{x}[1-\frac{qx}{(q-1)n+1}]^m.$$

We use the inequality $(1-t)^m \leq e^{-mt}$ that holds for every real t and every odd integer m. Therefore, for odd m:

$$\begin{aligned} q^n\mu_1^{*m}(\vec{0}) &\leq \sum_{x=0}^{n}(q-1)^x\binom{n}{x}e^{-\frac{qmx}{(q-1)n+1}} = [1+(q-1)e^{-\frac{qm}{(q-1)n+1}}]^n \leq \qquad (10)\\ &\leq [1+(q-1)e^{-(\ln n+b)}]^n \leq e^{(q-1)e^{-b}} \end{aligned}$$

Since $\mu_1^{*m}(\vec{0})$ is a monotone nonincreasing function of m this bound applies to even m as well. □ Notice that if b tends to infinity with n, then $e^{(q-1)e^{-b}} = 1+o(1)$, and we get Theorem 1.5 by applying Corollary 2.8 to the last result.

3 The Expected Cardinality of the Kernel

In this section we discuss the distribution of the size of the kernel of the random matrix and its relation to the distribution of the rank of the matrix. We show that already for $w = \ln n + \omega(1)$ the expected size of the kernel is essentially the same as in the case where no bounds are placed on the weights (Theorem 1.1). We use this result to bound the expected value of the rank of the random matrix and in turn the probability of full and small rank matrices (Corollaries 1.2 and 1.3).

If an $m \times n$ matrix A over F_q has rank r, then its left kernel $\{x \mid xA = \vec{0}\}$ has cardinality q^{m-r}. Consider the size of the left kernel as a random variable over $\Omega_{w,m,n,q}$. We denote its expectation by E_w^m. It is not hard to verify that the same random variable over $\Omega_{m,n,q}$ (no bound on weights) has expectation $\mathrm{E}^m = 1 + \frac{q^m-1}{q^n}$. We now prove Theorem 1.1 which states that as long as $w \geq \ln n + \omega(1)$:

$$\mathrm{E}^m \leq \mathrm{E}_w^m \leq (1+o(1))\mathrm{E}^m$$

Notice that this also yields the same relationship for the expected size of the right kernels by multiplying by q^{n-m}.

Proof of Theorem 1.1: Since the left kernel is $\{x \mid xA = \vec{0}\}$,

$$\mathrm{E}_w^m = \sum_{x \in F_q^m} \Pr(xA = \vec{0})$$

For a given vector x of weight t, $\Pr(xA = \vec{0}) = \mu_w^{*t}(\vec{0})$. The zero vector is always in the kernel, whence:

$$\mathrm{E}_w^m = 1 + \sum_{t=1}^{m} (q-1)^t \binom{m}{t} \mu_w^{*t}(\vec{0}) \tag{11}$$

According to Lemma 2.6, $\mu_w^{*t}(\vec{0}) \geq q^{-n} = \mu_n^{*t}(\vec{0})$ whence $\mathrm{E}_w^m \geq \mathrm{E}^m$. We also wish to upper bound E_w^m by $(1+o(1))(1+q^{m-n})$. To this end, using (11), it is enough to show that

$$\sum_{t=1}^{m} (q-1)^t \binom{m}{t} \mu_w^{*t}(\vec{0}) \ \leq \ q^{m-n} + o(\max\{1, q^{m-n}\}) \tag{12}$$

In order to obtain this inequality, we use the upper bounds on $\mu_w^{*t}(\vec{0})$ that follow from Theorem 1.5 and some intermediate calculations from Theorem 2.9.

Notice that $wt \geq \frac{q-1}{q} n \ln n + \omega(n)$ for the vast majority of the terms in the sum on the l.h.s. of (12). Thus Theorem 1.5 yields an upper bound on $\mu_w^{*t}(\vec{0})$ for these terms. Namely, when $t \geq (\frac{q-1}{q} - \frac{1}{\ln n})n$ then $wt \geq (\frac{q-1}{q} - \frac{1}{\ln n})n(\ln n + \omega(1)) \geq \frac{q-1}{q} n \ln n + \omega(n)$. It follows that

$$\sum_{t=(\frac{q-1}{q} - \frac{1}{\ln n})n}^{m} (q-1)^t \binom{m}{t} \mu_w^{*t}(\vec{0}) \ \leq \ (1+o(1)) q^{-n} \sum_{t=0}^{m} (q-1)^t \binom{m}{t} \ = \ (1+o(1)) q^{m-n}$$

It remains to show that

$$\Delta(m) \ \equiv \ \sum_{t=1}^{(\frac{q-1}{q} - \frac{1}{\ln n})n} (q-1)^t \binom{m}{t} \mu_w^{*t}(\vec{0}) \ \leq \ o(\max\{1, q^{m-n}\})$$

One may expect that the hardest case to prove is when $m = n$. Indeed it suffices to consider this case as we now explain: For $m \leq n$ we wish to show that $\Delta(m) \leq o(1)$, but $\Delta(m)$ is an increasing function of m, so that $m = n$ is clearly the hardest case in this range. For $m \geq n$ we wish to show that $\Delta(m) \leq o(q^{m-n})$. Now $\Delta(m) \leq q\Delta(m-1)$ when $m \geq n$, and therefore, again, one should only consider the case $m = n$. To see this inequality, note that $\binom{m}{t} = \frac{m}{m-t}\binom{m-1}{t} \leq q\binom{m-1}{t}$ as long as $t \leq \frac{q-1}{q} m$, which holds throughout the entire range of summation.

It remains to show that $\Delta \equiv \Delta(n) \leq o(1)$. In order to establish an upper bound on Δ, we need an upper bound on $\mu_w^{*t}(\vec{0})$. Throughout the rest of the proof we use the following fact:

$$\mu_w^{*t}(\vec{0}) \leq \mu_1^{*wt}(\vec{0}) \leq q^{-n}[1+(q-1)e^{-\frac{qwt}{(q-1)n+1}}]^n.$$

The first inequality is an application of Corollary 2.8 and the second is inequality (10) from the proof of Lemma 2.9. We also define $\Lambda \equiv \frac{(q-1)n+1}{qw}$. Using the inequality $e^{-x} \leq 1 - \frac{x}{1+x}$ we obtain that

$$\mu_w^{*t}(\vec{0}) \leq q^{-n}[1+(q-1)e^{-\frac{t}{\Lambda}}]^n \leq q^{-n}[1+(q-1)(1-\frac{\frac{t}{\Lambda}}{1+\frac{t}{\Lambda}})]^n = [\frac{q-(q-1)\frac{\frac{t}{\Lambda}}{1+\frac{t}{\Lambda}}}{q}]^n \leq$$

$$\leq e^{-\frac{(q-1)nt}{q\Lambda(1+\frac{t}{\Lambda})}} = e^{-\frac{wt}{1+\frac{t}{\Lambda}}} \leq e^{-(1-\frac{t}{\Lambda})wt}$$

Before going into the detailed calculations concerning Δ, we note a few standard facts about binomial coefficients that we need:

1. $\binom{n}{t} \leq (\frac{ne}{t})^t$ [Stirling's approximation]
2. Let the function H_q be defined as follows: $H_q(x) \equiv x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x)$. Then $\sum_{t=0}^{k}(q-1)^t\binom{n}{t} \leq q^{H_q(\frac{k}{n})n}$ for any $0 \leq k \leq \frac{q-1}{q}n$. See (5.1.5) in [8] for a proof of this fact.
3. $H_q(x)$ increases from zero to one as x goes from zero to $\frac{q-1}{q}$.
4. $H_q(\frac{q-1}{q} - x) \leq 1 - \frac{2}{\ln q}x^2$ [This fact can be verified by comparing the first two derivatives of both functions].

We show that $\Delta \leq o(1)$ by splitting the range of summation into four subranges s.t. $\Delta = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, and showing that each of the Δ_i is at most $o(1)$. The exact definition of the subranges will be given as we go along and prove the upper bound for each of them.

We start with Δ_1. Here, t goes from 1 to $\frac{\Lambda}{\ln n}$. Notice that $\frac{t}{\Lambda} \leq \frac{1}{\ln n}$ for all t in this range. Thus, in this range,

$$\mu_w^{*t}(\vec{0}) \leq e^{-(1-\frac{t}{\Lambda})wt} \leq e^{-(1-\frac{1}{\ln n})wt} \leq e^{-(\ln n+\omega(1))t}$$

Hence,

$$\Delta_1 \;=\; \sum_{t=1}^{\frac{\Lambda}{\ln n}}(q-1)^t\binom{n}{t}\mu_w^{*t} \;\leq\; \sum_{t=1}^{\infty}[(q-1)ne^{-(\ln n+\omega(1))}]^t = \sum_{t=1}^{\infty}[(q-1)e^{-\omega(1)}]^t \leq o(1)$$

We go on to Δ_2. Here, t goes from $\frac{\Lambda}{\ln n}$ to Λ. $\frac{t}{\Lambda} \leq 1$ for all t in this range, and thus,

$$\mu_w^{*t}(\vec{0}) \;\leq\; e^{-\frac{wt}{1+\frac{t}{\Lambda}}} \;\leq\; e^{-\frac{wt}{2}}$$

Hence,

$$\Delta_2 \;=\; \sum_{t=\frac{\Lambda}{\ln n}}^{\Lambda}(q-1)^t\binom{n}{t}\mu_w^{*t} \;\leq\; \sum_{t=\frac{\Lambda}{\ln n}}^{\Lambda}(\frac{(q-1)ne}{t})^t e^{-\frac{wt}{2}}$$

Where the last inequality is an application of Stirling's approximation. Notice that since $t \geq \frac{(q-1)n+1}{qw\ln n}$, in this range $\frac{(q-1)n}{t} \leq qw\ln n$, and thus

$$\Delta_2 \;\leq\; \sum_t[O(e^{-\frac{w}{2}}w\ln n)]^t \;\leq\; \sum_t[O(\frac{\ln^2 n}{\sqrt{n}})]^t \;\leq\; o(1)$$

Before we continue to Δ_3, we let γ be some constant s.t. $\frac{1+(q-1)e^{-1}}{q} < \gamma < 1$, and define δ as the unique solution to the equation $H_q(\delta) = (\log_q(\frac{\gamma q}{1+(q-1)e^{-1}}))$ in the range $\frac{q-1}{q} > \delta > 0$. (This is well-defined, since $H_q(x)$ increases from zero to one in the interval $[0, \frac{q-1}{q}]$).

In Δ_3, the range of t is from Λ to δn. Since $\frac{t}{\Lambda} \geq 1$ for all t in this range,

$$\mu_w^{*t}(\vec{0}) \;\leq\; [\frac{1+(q-1)e^{-\frac{t}{\Lambda}}}{q}]^n \;\leq\; [\frac{1+(q-1)e^{-1}}{q}]^n$$

Hence,

$$\Delta_3 \;=\; \sum_{t=\Lambda}^{\delta n}(q-1)^t\binom{n}{t}\mu_w^{*t} \;\leq\; [\frac{1+(q-1)e^{-1}}{q}]^n\sum_{t=0}^{\delta n}(q-1)^t\binom{n}{t} \;\leq$$

$$\leq\; [\frac{(1+(q-1)e^{-1})q^{H_q(\delta)}}{q}]^n \;=\; \gamma^n \;\leq\; o(1)$$

We go on to Δ_4. Here, t goes from δn to $(\frac{q-1}{q} - \frac{1}{\ln n})n$ (the end of the range for Δ). We let $\epsilon = \frac{\delta q}{q-1} > 0$. Since $\frac{t}{\Lambda} \geq \epsilon w$ for all t in this range,

$$\mu_w^{*t}(\vec{0}) \;\leq\; [\frac{1+(q-1)e^{-\frac{t}{\Lambda}}}{q}]^n \;\leq\; q^{-n}[1+(q-1)e^{-\epsilon w}]^n \;\leq\; q^{-n}e^{(q-1)n^{1-\epsilon}}$$

Hence,

$$\Delta_4 \;=\; \sum_{t=\delta n}^{(\frac{q-1}{q}-\frac{1}{\ln n})n} (q-1)^t \binom{n}{t} \mu_w^{*t} \;\le\; q^{-n} e^{(q-1)n^{1-\epsilon}} \sum_{t=0}^{(\frac{q-1}{q}-\frac{1}{\ln n})n} (q-1)^t \binom{n}{t} \;\le$$

$$\le\; q^{-n} e^{(q-1)n^{1-\epsilon}} q^{nH_q(\frac{q-1}{q}-\frac{1}{\ln n})}$$

We now use the fact that $H_q(\frac{q-1}{q}-x) \le 1 - \frac{2}{\ln q}x^2$ to deduce that $H_q(\frac{q-1}{q}-\frac{1}{\ln n}) \le 1 - \frac{2}{\ln q \ln^2 n}$, and thus,

$$\Delta_4 \;\le\; q^{-n} e^{(q-1)n^{1-\epsilon}} q^{n(1-\frac{2}{\ln q \ln^2 n})} \;=\; e^{(q-1)n^{1-\epsilon}-\frac{2n}{\ln^2 n}} \le o(1) \qquad \square$$

We now use this result to bound the expected value of the rank of the random matrix, and in turn, the probability of full, resp. small rank matrices.

Corollary 3.1 *If* $w \ge \ln n + \omega(1)$ *then* $\mathrm{E}(\min\{m,n\} - r) \le \frac{q^{-|m-n|}}{\ln q} + o(1)$

Proof: According to Theorem 1.1 (and the fact that the same holds for the right kernel as well):

$$\mathrm{E}(q^{\min\{m,n\}-r}) \le (1+o(1))(1+q^{-|m-n|})$$

By Jensen inequality (q^x is a convex function of x):

$$q^{\mathrm{E}(\min\{m,n\}-r)} \;\le\; \mathrm{E}(q^{\min\{m,n\}-r}) \;\le\; (1+o(1))(1+q^{-|m-n|})$$

thus

$$\mathrm{E}(\min\{m,n\} - r) \;\le\; \frac{q^{-|m-n|}}{\ln q} + o(1). \qquad \square$$

Corollaries 1.2 and 1.3 follow immediately from Corollary 3.1 using the Markov inequality (notice that $\min\{m,n\} - r$ is a nonnegative integer valued random variable).

References

[1] G. V. Balakin, The distribution of the rank of random matrices over a finite field, *Theory Prob. App.*, **13**(4), 594-605 (1968).

[2] N. J. Calkin, Dependent sets of constant weight vectors in $GF(q)$, *Random Struct. Alg.*, **9**(1), 49-53 (1996).

[3] C. Cooper, On the rank of random matrices, *Random Struct. Alg.*, **16**(2), 209-232 (2000).

[4] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, CA, 1988.

[5] P. Diaconis, R.L. Graham, and J.A Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, *Random Struct. Alg.*, **1**(1), 51-72 (1990).

[6] J. Kahn, J. Komlós, and E. Szemerédi, On the probability that a random ±1-matrix is singular, *J. Am. Math. Soc.*, **8**(1), 233-240 (1995).

[7] I. N. Kovalenko, On the limit distribution of the number of solutions of a random system of linear equations in the class of boolean functions, *Theory Prob. App.*, **12**(1), 47-56 (1967).

[8] J. H. van Lint, *Introduction to Coding Theory*, Springer-Verlag, 1982.

[9] P. Matthews, Mixing rates for a random walk on the cube, *SIAM J. Alg. Disc. Meth.*, **8**(4), 746-752 (1987).

A Appendix

Here we provide some more details concerning the rank distribution with no bounds on weights. Some of these facts were mentioned in the introduction. We start with the explicit expression for this distribution - Formula (1).

$$\Pr(rank = r) \;=\; \frac{1}{q^{(n-r)(m-r)}} \prod_{i=0}^{r-1} \frac{(1-q^{i-n})(1-q^{i-m})}{1-q^{i-r}}$$

for every $0 \leq r \leq \min\{m,n\}$. Other values of r do not occur, of course. This is a standard fact and here is a sketch of a proof: Start with the case $r = m$. Here $\Pr(rank = r) = \prod_{i=0}^{r-1}(1-q^{i-n})$ expresses the fact that the rank equals to the number of row vectors iff each new randomly selected row vector does not belong to the linear span of the previously chosen vectors.

For general values of r, we observe that an $m \times n$ matrix A has rank r iff it can be expressed as

$A = BC$, where B is $m \times r$ and C is $r \times n$, and both have rank r. Moreover, this representation of A is unique up to selecting a nonsingular $r \times r$ matrix D, and expressing $A = B'C'$ where $B' = BD$ and $C' = D^{-1}C$. A proof of the formula now follows by direct counting.

In the introduction we made two claims about the tails of the above distribution. Namely, we claimed that:

1. A random $m \times n$ matrix has, almost surely, full rank (an n grows) iff $|n - m|$ is unbounded.

2. If $|n - m|$ is bounded, then $\Pr(rank \leq r) \to 0$ iff $r \leq \min\{m, n\} - \omega(1)$.

We first note that for any positive integers $r \leq n$, and $q \geq 2$:

$$(1 - q^{r-n-1})^2 \;\leq\; \prod_{i=0}^{r-1}(1 - q^{i-n}) \;\leq\; 1 - q^{r-n-1}$$

The upper bound is clear, and the lower bound can be easily derived by induction on r.

To prove the first claim, recall that $\Pr(rank = m) = \prod_{i=0}^{m-1}(1 - q^{i-n})$, and thus this probability equals to $1 - \Theta(q^{m-n})$. By symmetry, $\Pr(rank = \min\{m, n\}) = 1 - \Theta(q^{-|n-m|})$ and thus, the first claim.

For the second claim we observe that $\Pr(rank = r) = \Theta(q^{-(n-r)(m-r)})$. This simply means that there are some absolute positive constants A_1 and A_2, such that for any positive integers $r \leq \min\{m, n\}$, and $q \geq 2$:

$$A_1 \;\leq\; \prod_{i=0}^{r-1} \frac{(1 - q^{i-n})(1 - q^{i-m})}{1 - q^{i-r}} \;\leq\; A_2$$

(e.g. $A_1 = 1/16$ and $A_2 = 4$ can be established easily). We omit the easy derivation of these bounds.