Weight distribution of random linear codes and Krawchouk polynomials

Alex Samorodnitsky

Abstract

For $0 < \lambda < 1$ and $n \to \infty$ pick uniformly at random λn vectors in $\{0,1\}^n$ and let C be the orthogonal complement of their span. Given $0 < \gamma < \frac{1}{2}$ with $0 < \lambda < h(\gamma)$, let X be the random variable that counts the number of words in C of Hamming weight $i = \gamma n$ (where i is assumed to be an even integer). Linial and Mosheiff [3] determined the asymptotics of the moments of X of all orders $o\left(\frac{n}{\log n}\right)$. In this paper we extend their estimates up to moments of linear order. Our key observation is that the behavior of the suitably normalized k^{th} moment of X is essentially determined by the k^{th} norm of the Krawchouk polynomial K_i.

1 Introduction

This paper follows up on the recent result [3] by Linial and Mosheiff. In [3] the authors study the distribution of the number of codewords of a given weight in a random linear code of a given rate, providing tight estimates on the suitably normalized moments of this distribution, up to moments of order $o\left(\frac{n}{\log n}\right)$. In this paper we extend the estimates in [3] up to moments of linear order.

We refer to [3] for the introduction and initial discussion of the problem and for the description of the wider context of asymptotic problems in coding theory. Here we pass directly to technical definitions. In the discussion below we try to adhere to the notation of [3], where possible.

Definitions and notation.Let $h(t) = t\log_2\left(\frac{1}{t}\right) + (1-t)\log_2\left(\frac{1}{1-t}\right)$ be the binary entropy function. Given $0 < \gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$, assume that an integer n is such that γn and λn are integers. Furthermore, assume that $i = \gamma n$ is even. Let L be the set of all vectors of weight i in $\{0,1\}^n$. Let M be a random binary matrix with λn rows and n columns. Let C be a (random) linear code with parity check matrix M. Set $X = |L \cap C|$. Set $p = 2^{-\lambda n}$.

As observed in [3], it is well-known (and easy to see) that $\mathbb{E}\, X = |L| \cdot \mathbb{E}\,|C| = \binom{n}{i} \cdot p$ and $\mathrm{Var}(X) = \binom{n}{i} \cdot p(1-p) \approx \binom{n}{i} \cdot p$. Note that here and below we use $\approx$, $\gtrsim$, and $\lesssim$ notation to describe equalities and inequalities which hold up to lower order terms. We also write $\gg$ to indicate that the left hand side of an inequality is exponentially larger than its right hand side.

The following theorem from [3] describes higher central moments of X:

Theorem 1.1: *Assume* $2 \leq k \leq o\left(\frac{n}{\log n}\right)$. *Then*

$$\frac{(X-\mathbb{E}\,X)^k}{\mathrm{Var}(X)^{\frac{k}{2}}} \approx \begin{cases} o(1) & \text{if } k \text{ is odd and} < k_0 \\ k!! & \text{if } k \text{ is even and} < k_0 \\ 2^{\left(F(k,\gamma)-\frac{k}{2}h(\gamma)-\left(\frac{k}{2}-1\right)\lambda\right)\cdot n} & \text{if } k \geq k_0 \end{cases}$$

Here $F(k,\gamma)$ is a certain bivariate function defined in [3] and k_0 is the minimal value of k for which $F(k,\gamma)-\frac{k}{2}h(\gamma)-\left(\frac{k}{2}-1\right)\lambda > 0$ (it can be seen that this inequality holds for all $k \geq k_0$).

[3] discusses the question of extending the estimates above to higher values of k. In this paper we suggest a different approach to the problem, proving the following result.

Theorem 1.2: *Let* $\epsilon = \min\{\lambda, h(\gamma)-\lambda\}$. *There exist constants* c_ϵ *and* C_ϵ *such that for any* $C_\epsilon \leq k \leq c_\epsilon n$ *holds*

$$\frac{(X-\mathbb{E}\,X)^k}{\mathrm{Var}(X)^{\frac{k}{2}}} \approx 2^{\left(F(k,\gamma)-\frac{k}{2}h(\gamma)-\left(\frac{k}{2}-1\right)\lambda\right)\cdot n}.$$

Discussion. As observed in [3] higher central moments of $X = |C \cap L|$ are larger than those for a general random code C in which each word in chosen from $\{0,1\}^n$ uniformly and independently with probability p. In that case X is a binomial random variable, $X \sim B\left(\binom{n}{i}, p\right)$.

Roughly speaking, the k^{th} moment of X counts the number of k-tuples in L contained in C. It is easy to see that the advantage given by linearity comes from linearly dependent k-tuples. Our main observation is that, at least intuitively, for k much smaller than n a "typical" dependent k-tuple is a k-circuit, that is a linear dependence of the form $x_1+...+x_k=0$, which does not contain any smaller dependencies. In particular, the rank of the set $\{x_1, x_2, ..., x_k\}$ is $k-1$. The number of such dependencies is very close (this will be made more precise below) to $\mathbb{E}\left(\sum_{x\in L} w_x\right)^k = \mathbb{E}\,K_i^k$, where w_x is the Walsh-Fourier character corresponding to $x \in \{0,1\}^n$, and K_i is the i^{th} Krawchouk polynomial (all notions will be defined in Section 2 below). The expectation is taken over the uniform distribution on $\{0,1\}^n$, Given the assumption that i is an even integer, we have $\mathbb{E}\,K_i^k \approx \mathbb{E}\,|K_i|^k = \|K_i\|_k^k$ (see Section 2 for this as well). As observed in [3] the probability that a subset S of $\{0,1\}^n$ is contained in C is $p^{r(S)}$, where $r(S)$ is the rank of S. Hence, if this intuition is correct, the expected number of dependent k-tuples in C would be close to $p^{k-1}\|K_i\|_k^k$, and we would expect the advantage given by linearity to be visible iff

$$p^{k-1}\|K_i\|_k^k \gg \mathrm{Var}(X)^{\frac{k}{2}} = \left(\binom{n}{i}\cdot p\right)^{\frac{k}{2}} \iff p^{\frac{k}{2}-1}\|K_i\|_k^k \gg \left(\binom{n}{i}\right)^{\frac{k}{2}},$$

in which case we would expect to have (assuming that the advantage provided by linearity is greater than the corresponding binomial moment arising from independent choices)

$$\frac{(X-\mathbb{E}\,X)^k}{\mathrm{Var}(X)^{\frac{k}{2}}} \approx \frac{p^{\frac{k}{2}-1}\|K_i\|_k^k}{\left(\binom{n}{i}\right)^{\frac{k}{2}}}.$$

We would like to understand the exponential behavior of these estimates. Recall that $p = 2^{-\lambda n}$. We have $|L| = \binom{n}{i} \approx 2^{h(\gamma)n}$ (by the known estimates on binomial coefficients, see e.g. Theorem 1.4.5. in [4]). We also have (see Section 2) that $\frac{1}{n}\log_2 \|K_i\|_k^k \approx \psi(k,\gamma) + \frac{k}{2}\cdot h(x)$, where $\psi(k,\gamma)$ is a certain bivariate function defined in [2]. Using of all this, and passing to exponents, the constraint $p^{k-1}\|K_i\|_k^k \gg \mathrm{Var}(X)^{\frac{k}{2}}$ becomes

$$\psi(k,\gamma) > \left(\frac{k}{2} - 1\right)\lambda,$$

in which case we expect to have

$$\frac{(X - \mathbb{E}\,X)^k}{\mathrm{Var}(X)^{\frac{k}{2}}} \;\approx\; 2^{\left(\psi(k,\gamma) - \left(\frac{k}{2}-1\right)\lambda\right)\cdot n}.$$

It turns out that these are precisely the results proved in the third claim of Theorem 1.1 and in Theorem 1.2. In fact, the two pertinent bivariate functions defined in [3] and in [2] are the same, up to a multiple of the entropy function:

$$F(k,\gamma) \;=\; \psi(k,\gamma) + \frac{k}{2}\cdot h(\gamma).$$

To see that, cf. the definition of $F(k,\gamma)$ and Definition 10 in [3] with Lemma 2.5 and Proposition 2.6 in [2].

The remainder of the paper is devoted to proving Theorem 1.2. By the preceding discussion, this theorem is an immediate corollary of the two following claims.

Proposition 1.3: *Assume that $k \le \lambda n - 1$. Then*

$$\mathop{\mathbb{E}}_{C}(X - \mathbb{E}\,X)^k \;\ge\; \frac{1}{2}\cdot p^{k-1}\|K_i\|_k^k.$$

and

Proposition 1.4: *Let $\epsilon = \min\{\lambda, h(\gamma) - \lambda\}$. There exist constants c_ϵ and C_ϵ such that for any $C_\epsilon \le k \le c_\epsilon n$ holds*

$$\mathop{\mathbb{E}}_{C}(X - \mathbb{E}\,X)^k \;\lesssim\; p^{k-1}\cdot 2^{\left(\psi(k,\gamma) + \frac{k}{2}h(\gamma)\right)\cdot n}.$$

These two claims are proved in Section 3. Prior to that, some relevant notions and background are provided in Section 2.

2 Preliminaries

We collect some relevant facts on Fourier analysis on the boolean cube (see e.g., [7]).

For $x \in \{0,1\}^n$, define the Walsh-Fourier character w_x on $\{0,1\}^n$ by setting $w_x(y) = (-1)^{\sum_i x_i y_i}$, for all $y \in \{0,1\}^n$. The *weight* of the character w_x is the Hamming weight $|x|$ of x (that is the number of 1-coordinates in x). The characters $\{w_x\}_{x\in\{0,1\}^n}$ form an orthonormal basis in the space of real-valued functions on $\{0,1\}^n$, under the inner product $\langle f,g\rangle = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)g(x)$. The expansion $f = \sum_{x\in\{0,1\}^n} \widehat{f}(x)w_x$ defines the Fourier transform $\widehat{f}$ of f. We also have the Parseval identity, $\|f\|_2^2 = \sum_{x\in\{0,1\}^n} \widehat{f}^2(x)$. Here is one additional simple fact that we will require. Let C be a linear subspace of $\{0,1\}^n$. Then $\widehat{1_C} = \frac{|C|}{2^n}\cdot 1_{C^\perp}$.

Krawchouk polynomials. For $0 \le i \le n$, let F_i be the sum of all Walsh-Fourier characters of weight i, that is $F_i = \sum_{|x|=i} w_x$. It is easy to see that $F_i(x)$ depends only on the Hamming weight $|x|$ of x, and it can be viewed as a univariate function on the integer points $0,...,n$, given by the restriction to $\{0,...,n\}$ of the univariate polynomial $K_i = \sum_{k=0}^{i}(-1)^k\binom{x}{k}\binom{n-x}{i-k}$ of degree i. That is, $F_i(x) = K_i(|x|)$. The polynomial K_i is the i^{th} *Krawchouk polynomial* (see e.g., [6]). Abusing notation, we will also call F_i the i^{th} Krawchouk polynomial, and write K_i for F_i when the context is clear. Here are two simple properties of the Krawchouk polynomials which we will need: $K_i(0) = \|K_i\|_2^2 = \binom{n}{i}$.

Norms. It is easy to see from the definition of the Walsh-Fourier characters that for any integer k and for any subset S of $\{0,1\}^n$ holds $\mathbb{E}\left(\sum_{x\in S} w_x\right)^k = \Big|\{(x_1,...,x_k)\in S^k,\ x_1+...+x_k=0\}\Big|$ (the expectation is taken w.r.t. the uniform probability measure on $\{0,1\}^n$) . In particular, if L is the set of all vectors of weight i in $\{0,1\}^n$ and k is an even integer, we have

$$\|K_i\|_k^k \ = \ \mathbb{E}\left(\sum_{x\in L} w_x\right)^k \ = \ \Big|\{(x_1,...,x_k)\in L^k,\ x_1+...+x_k=0\}\Big|.$$

Let now $k>2$ be an odd integer. As observed e.g., in Section 2.2 in [2], the k^{th} norm of K_i is essentially attained outside the root region of K_i. For an even i, K_i is positive outside its root region and hence we also have

$$\|K_i\|_k^k \ = \ \mathbb{E}\,|K_i|^k \ \approx \ \mathbb{E}\,K_i^k \ = \ \Big|\{(x_1,...,x_k)\in L^k,\ x_1+...+x_k=0\}\Big|.$$

As observed e.g., in Section 2.2 in [2], for all $0 \le i \le n/2$ and $k \ge 2$ holds, writing $\gamma = i/n$:

- $\frac{\|K_i\|_k^k}{(\|K_i\|_2)^k} \ \le \ 2^{\psi\left(k,\frac{i}{n}\right)\cdot n} = 2^{\psi(k,\gamma)\cdot n}$, where ψ is a certain bivariate function defined in [2].
- Moreover, $\frac{\|K_i\|_k^k}{(\|K_i\|_2)^k} \ \approx \ 2^{\psi(k,\gamma)\cdot n}$.

As pointed out in the introduction, $F(k,\gamma) = \psi(k,\gamma) + \frac{k}{2}\cdot h(\gamma)$. Since $\|K_i\|_2^2 = \binom{n}{i} \approx 2^{h(\gamma)\cdot n}$, we have

$$\|K_i\|_k^k \ \approx \ 2^{\left(\psi(k,\gamma)+\frac{k}{2}h(\gamma)\right)\cdot n} \ = \ 2^{F(k,\gamma)\cdot n}.$$

Finally, we need some properties of the function ψ shown in [3]. We collect them in the following lemma, for convenience (writing them in terms of ψ). We remark that the second and the third properties listed in the lemma were also shown independently in [2].

Lemma 2.1:

1. $\psi(k,\gamma) \geq \frac{k}{2} \cdot H(\gamma) - 1$.
2. *For a fixed* γ *the function* $\psi(k,\gamma)$ *is convex in* k.
3. *Since, furthermore,* $\psi(2,\gamma) = 0$, *this implies that* $\frac{\psi(k,\gamma)}{k-2}$ *is an increasing function of* k *for* $k > 2$.

Remark 2.2: It seems worthwhile to sketch a way to derive these facts from the approximate identity $\|K_i\|_k^k \approx 2^{\left(\psi(k,\gamma)+\frac{k}{2}h(\gamma)\right)\cdot n}$. Observe first that $\|K_i\|_k^k \geq \frac{1}{2^n} \cdot K(0)^k = \frac{1}{2^n} \cdot \binom{n}{i}^k$; and second that for any function f on $\{0,1\}^n$, the function $\alpha \rightarrow \log \|f\|_{\frac{1}{\alpha}}$ is convex on $(0,1]$ (this is a consequence of Hölder's inequality, see e.g., [1], Theorems 196 and 197).

∎

3 Proof of Theorem 1.2

3.1 Proof of Proposition 1.3

We will show this proposition to be an immediate corollary of the lower bound given in the following claim. (The upper bound in this claim we be used in the next section as a step towards the proof of Proposition 1.4).

Proposition 3.1:

Assume that $k \leq \lambda n - 1$. *Then*

$$\frac{1}{2} \cdot \sum_{S=(u_1 \ldots u_k) \in L^k} p^{r(S)} \;\leq\; \mathop{\mathbb{E}}_{C} (X - \mathbb{E}\, X)^k \;\leq\; 2 \cdot \sum_{S=(u_1 \ldots u_k) \in L^k} p^{r(S)},$$

where the summation is over all sequences $(u_1 \ldots u_k)$ *of vectors which contain no coloops (vector which is not contained in the span of all the rest).*

Given the lower bound in Proposition 3.1, Proposition 1.3 can be derived as follows. Note that a sequence of vectors which sums to 0 necessarily contains no coloops. Hence we have (see Section 2)

$$p^{k-1} \cdot \|K_i\|_k^k \;=\; p^{k-1} \cdot \sum_{S=(u_1 \ldots u_k) \in L^k, u_1 + \ldots + u_k = 0} 1 \;\leq$$

$$\sum_{S=(u_1...u_k)\in L^k, u_1+...+u_k=0} p^{r(S)} \;\leq\; \sum_{S=(u_1...u_k)\in L^k} p^{r(S)},$$

where the last summation is over all sequences $(u_1...u_k)$ of vectors which contain no coloops.

Proof: (Of Proposition 3.1)

We will use the following notation. For $u \in \{0,1\}^n$, let Y_u be the indicator of the event u is in C. Let $Z_u = Y_u - p$. Then $X = \sum_{u\in L} Y_u$ and $X - \mathbb{E}\, X = \sum_{u\in L} Z_u$. We have

$$\mathop{\mathbb{E}}_{C}(X - \mathbb{E}\, X)^k \;=\; \mathop{\mathbb{E}}_{C}\left(\sum_{u\in L} Z_u\right)^k \;=\; \sum_{(u_1...u_k)\in L^k} \mathop{\mathbb{E}}_{C}\prod_{r=1}^{k} Z_{u_r}.$$

So, the claim of the proposition will follow from the next lemma.

Lemma 3.2: *Assume that $k \leq \lambda n - 1$. Let $S = (u_1...u_k)$ be a sequence of vectors in $\{0,1\}^n$. Then there are two cases.*

- *These vectors contain no coloops. In this case*

$$\frac{1}{2}p^{r(S)} \;\leq\; \mathop{\mathbb{E}}_{C}\prod_{r=1}^{k} Z_{u_r} \;\leq\; 2p^{r(S)}.$$

- *These vectors contain a coloop. Then*

$$\mathop{\mathbb{E}}_{C}\prod_{r=1}^{k} Z_{u_r} \;=\; 0.$$

Proof: (Of Lemma 3.4)

Let $g = g_S = \prod_{r=1}^{k} Z_{u_r}$. This is a function on subspaces which depends only on the number of vectors from S contained in the given subspace. We need to estimate $\mathbb{E}\, g$. It will be convenient for us to transform the setting in the following manner. Let t be the number of distinct vectors in S. Arranging them in some order, and writing the statistics of their appearance in S, converts S into a t-tuple of integers $(s_1...s_t)$ summing to k. A subspace corresponds to a vector $x \in \{0,1\}^t$ whose 1-coordinates indicate which distinct vectors from S this subspace contains. The function g then transforms into a function on $\{0,1\}^t$ defined by $g(x) = (1-p)^{\langle S,x\rangle}(-p)^{k-\langle S,x\rangle}$. We have a measure $\mathcal{L}$ on $\{0,1\}^t$ induced by the measure on subspaces, which is determined by the following property: for all $R \subseteq [t]$ holds $\mathcal{L}\{x : x_j = 1\ \forall j \in R\} = p^{r(R)}$, where $r(R)$ stands for $r\left(\{u_i\}_{i\in R}\right)$. With all this, we have

$$\mathop{\mathbb{E}}_{C}\prod_{r=1}^{k} Z_{u_r} \;=\; \mathop{\mathbb{E}}_{x\sim\mathcal{L}} g(x).$$

Let f be a function on $\{0,1\}^t$ defined for $R \subseteq [t]$ by $f(R) = p^{r(R)}$. Thinking of $\mathcal{L}$ as a function on $\{0,1\}^t$, we have $f(R) = \sum_{x:\, R\subseteq x} \mathcal{L}(x)$. Hence, by the Möbius inversion formula for the

boolean lattice (see e.g., [5]) we have $\mathcal{L} = \mu f$, where μ is the Möbius function of the boolean lattice: $\mu(x,y) = (-1)^{|y|-|x|}$ if x is a subset of y, and 0 otherwise.

Writing $\langle \cdot, \cdot \rangle$ for the inner product of functions on $\{0,1\}^t$ endowed with the counting measure, we have

$$\mathop{\mathbb{E}}_{z\sim\mathcal{L}} g(z) \;=\; \langle \mathcal{L}, g\rangle \;=\; \langle \mu f, g\rangle = \left\langle f, \mu^t g\right\rangle .$$

Next, we compute $\mu^t g$. We have

$$\left(\mu^t g\right)(x) \;=\; \sum_z \mu^t(x,z) g(z) \;=\; \sum_{z\subseteq x} (-1)^{|x|-|z|}(1-p)^{\langle S,z\rangle}(-p)^{k-\langle S,z\rangle} \;=$$

$$(-1)^{k+|x|}p^k \cdot \sum_{z\subseteq x}(-1)^{\langle S,z\rangle+|z|}\left(\frac{1-p}{p}\right)^{\langle S,z\rangle} .$$

Using the fact that $\sum_{z\subseteq x}\prod_{i\in z}\alpha_i = \prod_{i\in x}(1+\alpha_i)$, the last expression is

$$(-1)^{k+|x|}p^k \prod_{i\in x}\left(1-\left(\frac{p-1}{p}\right)\right)^{s_i} .$$

Hence

$$\left\langle f, \mu^t g\right\rangle \;=\; (-1)^k p^k \sum_{x\in\{0,1\}^t}(-1)^{|x|}p^{r(x)}\prod_{i\in x}\left(1-\left(\frac{p-1}{p}\right)\right)^{s_i} \;=$$

$$(-1)^k p^k \sum_{x\in\{0,1\}^t}(-1)^{|x|}p^{r(x)-\langle S,x\rangle}\prod_{i\in x}\left(p^{s_i}-(p-1)^{s_i}\right).$$

Now there are two cases to consider. Let σ_x denote the summand corresponding to x in the last expression. Assume first that S contains a coloop (w.l.o.g. vector number t). We claim that in this case for all x holds $\sigma_x = -\sigma_{x\oplus e_t}$ and hence the total sum is 0. In fact, this follows by inspection, since (clearly) $s_t = 1$ and $r(x\cup t) = r(x\setminus t)+1$.

Assume now that S does not contain coloops. Since by assumption $p = 2^{-\lambda n} \le 2^{-k-1}$ and since $s_i \ge 1$ for all $i\in[t]$, we have that, up to a $1+o(1)$ multiplicative factor, $\prod_{i\in x}\left(p^{s_i}-(p-1)^{s_i}\right) \approx \prod_{i\in x}(-1)^{s_i+1} = (-1)^{\langle S,x\rangle+|x|}$. So, $\sigma_x \approx (-1)^k p^k (-1)^{\langle S,x\rangle}p^{r(x)-\langle S,x\rangle}$. In particular, $\sigma_{[t]} \approx p^{r(S)}$. We claim that this is the dominant term in $\sum_{x\in\{0,1\}^t}\sigma_x$. Specifically, we claim that $\sigma_{[t]} \ge 2\cdot\sum_{x\ne[t]}|\sigma_x|$. Note that showing this will complete the proof of the lemma.

Since the number of summands is smaller than $\frac{1}{2p}$, it suffices to show that for all $x\ne[t]$ holds $r(x)-\langle S,x\rangle > r(S)-k$. In fact,

$$(r(x)-\langle S,x\rangle)-(r(S)-k) \;=\; \sum_{i\notin x} s_i - (r(S)-r(x)).$$

Since $r(S)-r(x)\le t-|x|$, this can only be 0 iff $s_i = 1$ for all $i\notin x$ and $r(S)-r(x) = |t|-|x|$. But this means that all vectors in x^c are coloops, contradicting the assumptions.

∎ (Lemma 3.4).

This concludes the proof of Proposition 3.1 and of Proposition 1.3.

3.2 Proof of Proposition 1.4

Assume $k \leq \lambda n - 1$. From the upper bound of Proposition 3.1, we have

$$\mathop{\mathbb{E}}_{C} (X - \mathbb{E}\, X)^k \;\leq\; 2 \cdot \sum_{S=(u_1 \ldots u_k) \in L^k} p^{r(S)},$$

where the summation is over all sequences of vectors which contain no coloops. Note that the rank of any such sequence is smaller than k. For $1 \leq r \leq k-1$ let $N(r)$ denote the number of such sequences S with $r(S) = r$. Then the RHS in the last inequality is $2 \cdot \sum_{r=1}^{k} p^r N(r)$.Observe that $N(k-1)$ is the number of k-circuits in L. Hence (see Section 2) $p^{k-1} N(k-1) \leq p^{k-1} \cdot \mathbb{E}\, \|K_i\|_k^k \leq p^{k-1} \cdot 2^{\left(\psi(k,\gamma) + \frac{k}{2} h(\gamma)\right) \cdot n}$. Therefore, to prove Proposition 1.4 it will suffice to show the following claim.

Proposition 3.3: *Let $\epsilon = \min\{\lambda, H(\gamma) - \lambda\}$. There exist constants c_ϵ and C_ϵ such that for any $C_\epsilon \leq k \leq c_\epsilon n$ and for any $1 \leq r \leq k-2$ holds*

$$N(r) \;\lesssim\; p^{k-r-1} \cdot 2^{\left(\psi(k,\gamma) + \frac{k}{2} h(\gamma)\right) \cdot n}.$$

In the remainder of this section we prove Proposition 3.3. Let $S = (u_1 \ldots u_k)$ be a sequence of k vectors in $\{0,1\}^n \setminus \{0\}$, with $r(S) = r$. Let $B = \{b_1, \ldots, b_r\}$ be a basis of S, that is a subset of r linearly independent vectors in S (which span all vectors in S). We build on a simple observation, which we state in the following lemma.

Lemma 3.4: *If S contains no coloops, then for some $1 \leq v \leq \min\{r, k-r\}$ there are additional v vectors $x_1, \ldots, x_v$ in S with the following property: For $1 \leq d \leq v$, let $x_d = \sum_{i \in S_d} b_i$. Then $S_1, \ldots, S_v$ are non-empty subsets of $[r]$, for all $1 \leq d \leq v$ holds $S_d \not\subseteq \cup_{i=1}^{d-1} S_i$, and $\cup_{i=1}^{v} S_i = [r]$.*

Proof: Note that each element of B participates in at least one linear dependence among the elements of S. Let x_1 be a vector in $S \setminus B$ whose representation as a linear combination of elements of B contains b_1. If $S_1 \neq [r]$, let $2 \leq j \leq r$ be the minimal index in $[r] \setminus S_1$. Let x_2 be a vector in $S \setminus B$ whose representation as a linear combination of elements of B contains b_j, etc. ∎

We proceed with a technical lemma. We write K for $K_i = K_{\gamma n}$ for notational convenience from now on.

Lemma 3.5: *Let $1 \leq v \leq r$. Let $m = r + v$. Let $S_1, \ldots, S_v$ be non-empty subsets of $[r]$ with the following property: for all $1 \leq d \leq v$ holds $S_d \not\subseteq \cup_{i=1}^{d-1} S_i$, and $\cup_{i=1}^{v} S_i = [r]$.*

Let $\mathcal{A} \subseteq L^m$ contain the m-tuples $(x_1, \ldots, x_m)$ which satisfy $x_{r+d} = \sum_{i \in S_d} x_i$, $d = 1, \ldots, v$. Then there exist integers $a_1, \ldots, a_v \geq 2$ with $\sum_{i=1}^{v} a_i = m$ so that

$$|\mathcal{A}| \;\leq\; \prod_{i=1}^{v} \mathop{\mathbb{E}}_{y_i} |K(y_i)|^{a_i}.$$

In fact, the sequence $\{a_d\}$ is defined as follows: $a_d = |S_d \setminus \cup_{i=1}^{d-1} S_i| + 1$, $d = 1, \ldots, v$.

Proof:

Viewing an m-tuple of vectors $x_1, ..., x_m$ in $\{0,1\}^n$ as rows of an $m \times n$ binary matrix $\bar{x}$, let the columns of this matrix be denoted by $c_1(\bar{x}), ..., c_n(\bar{x})$. Let M be the following $r \times m$ binary matrix: The first r columns of M form an $r \times r$ identity matrix, and for $1 \leq d \leq v$, the column $r+d$ is the characteristic vector of the set S_d. Let $C \subseteq \{0,1\}^m$ be the linear code generated by the rows of M. The rows of M are linearly independent, and hence $\dim(C) = r$.

Observe that

$$\mathcal{A} \;=\; \left\{\bar{x} = (x_1, ..., x_m) \in L^m,\; c_j(\bar{x}) \in C,\; j = 1, ..., n\right\}$$

Let

$$\mathcal{B} \;=\; \left\{\bar{x} = (x_1, ..., x_m) \in (\{0,1\}^n)^m, c_j(\bar{x}) \in C,\; j = 1, ..., n\right\}.$$

Note that $\mathcal{A} = \mathcal{B} \cap L^m$.

Let

$$\mathcal{B}^* \;=\; \left\{\bar{y} = (y_1, ..., y_m) \in (\{0,1\}^n)^m,\; Mc_j(\bar{y}) = 0,\; j = 1, ..., n\right\}$$

We need an auxiliary lemma.

Lemma 3.6: *Let $y_1, ..., y_m \in \{0,1\}^n$. Then*

$$\sum_{(x_1,...,x_m)\in\mathcal{B}} \prod_{i=1}^{m} w_{x_i}(y_i) \;=\; \begin{cases} 2^{rn} & \text{if} \quad (y_1, ..., y_m) \in \mathcal{B}^* \\ 0 & \text{otherwise} \end{cases}$$

Proof: (of Lemma 3.6)

Identifying $(\{0,1\}^n)^m$ with $\{0,1\}^{mn}$, we view $\bar{x} = (x_1, ..., x_m), \bar{y} = (y_1, ..., y_m)$ as points in $\{0,1\}^{mn}$, and then $\prod_{i=1}^{m} w_{x_i}(y_i) = w_{\bar{y}}(\bar{x})$. Note that in this identification $\mathcal{B}$ becomes an (rn)-dimensional linear subspace of $\{0,1\}^{mn}$, and $\mathcal{B}^*$ becomes its dual, and we have (see Section 2)

$$\sum_{(x_1,...,x_m)\in\mathcal{B}} \prod_{i=1}^{m} w_{x_i}(y_i) \;=\; \sum_{\bar{x}\in\mathcal{B}} w_{\bar{y}}(\bar{x}) \;=\; 2^{mn} \cdot \widehat{1_{\mathcal{B}}}(\bar{y}) \;=$$

$$\begin{cases} 2^{rn} & \text{if} \quad \bar{y} \in \mathcal{B}^\perp \\ 0 & \text{otherwise} \end{cases} \;=\; \begin{cases} 2^{rn} & \text{if} \quad (y_1, ..., y_m) \in \mathcal{B}^* \\ 0 & \text{otherwise} \end{cases}$$

∎ (Lemma 3.6)

We continue with the proof of Lemma 3.5. Since $K = \sum_{z\in L} w_z$ (see Section 2), for any $x \in \{0,1\}^n$ holds $\mathbb{E}_{y\in\{0,1\}^n} w_x(y)K(y) = 1_{x\in L}$. Hence, we have, using Lemma 3.6 in the last step,

$$|\mathcal{A}| \;=\; \sum_{(x_1,...,x_m)\in\mathcal{B}} \prod_{i=1}^{m} \mathop{\mathbb{E}}_{y_i} w_{x_i}(y_i) K_i(y_i) \;=\; \sum_{(x_1,...,x_m)\in\mathcal{B}} \mathop{\mathbb{E}}_{y_1,...,y_m} \prod_{i=1}^{m} K(y_i) \prod_{i=1}^{m} w_{x_i}(y_i) \;=$$

$$\mathop{\mathbb{E}}_{y_1,...,y_m} \prod_{i=1}^{m} K(y_i) \sum_{(x_1,...,x_m)\in\mathcal{B}} \prod_{i=1}^{m} w_{x_i}(y_i) \;=\; 2^{-vn} \sum_{(y_1,...,y_m)\in\mathcal{B}^*} \prod_{i=1}^{m} K(y_i)$$

Consider the last expression. Let M^* be the following generating matrix of $C^\perp$. It is a $v \times m$ binary matrix whose first r columns form a $v \times r$ matrix whose rows are the characteristic vectors of the sets $S_1, ..., S_v$. The remaining v columns form a $v \times v$ identity matrix. Let the subsets of $[v]$ corresponding to the first r columns of M^* be denoted by $T_1, ..., T_r$. Observe that

$$2^{-vn} \sum_{(y_1,...,y_m)\in\mathcal{B}^*} \prod_{i=1}^{m} K(y_i) \;=\; \mathop{\mathbb{E}}_{y_1,...,y_v\in\{0,1\}^n} \prod_{i=1}^{v} K(y_i) \prod_{j=1}^{r} K\left(\sum_{i\in T_j} y_i\right).$$

For $1 \le d \le v$, set $R_d = S_d \setminus \cup_{i=1}^{d-1} S_i$. Note that the sets $R_1 ... R_v$ are non-empty and they partition $[r]$. Note also that the sets $\{T_j\}_{j\in R_d}$ are subsets of $\{d, ..., v\}$ and they all contain d. We have

$$\mathop{\mathbb{E}}_{y_1,...,y_v} \prod_{i=1}^{v} K(y_i) \prod_{j=1}^{r} K\left(\sum_{i\in T_j} y_i\right) \;\le$$

$$\mathop{\mathbb{E}}_{y_1,...,y_v} \prod_{i=1}^{v} |K|(y_i) \prod_{j=1}^{r} |K|\left(\sum_{i\in T_j} y_i\right) \;=$$

$$\mathop{\mathbb{E}}_{y_v} |K|(y_v) \cdot \prod_{j\in R_v} |K|\left(\sum_{i\in T_j} y_i\right) \cdot \mathop{\mathbb{E}}_{y_{v-1}} |K|(y_{v-1}) \cdot \prod_{j\in R_{v-1}} |K|\left(\sum_{i\in T_j} y_i\right) \cdots \mathop{\mathbb{E}}_{y_1} |K|(y_1) \cdot \prod_{j\in R_1} |K|\left(\sum_{i\in T_j} y_i\right)$$

Consider the expectation $E_{y_d} |K|(y_d) \cdot \prod_{j\in R_d} |K|\left(\sum_{i\in T_j} y_i\right)$, for fixed $y_{d+1}, ..., y_v$. Note that this is the expectation of a product of $a_d = |R_d| + 1 = |S_d \setminus \cup_{i=1}^{d-1} S_i| + 1$ functions $f_1, ..., f_{a_d}$ of y_d, each of which is a rearrangement of $|K|$. By Hölder's inequality we have

$$\mathop{\mathbb{E}}_{y_d} f_1 \cdots f_{a_d} \;\le\; \prod_{j=1}^{a_d} \|f_j\|_{a_d} \;=\; \|K\|_{a_d}^{a_d} \;=\; \mathop{\mathbb{E}}_{y_d} |K(y_d)|^{a_d}.$$

Since this holds for all $1 \le d \le v$, we have

$$\mathop{\mathbb{E}}_{y_v} |K|(y_v) \cdot \prod_{j\in R_v} |K|\left(\sum_{i\in T_j} y_i\right) \cdot \mathop{\mathbb{E}}_{y_{v-1}} |K|(y_{v-1}) \cdot \prod_{j\in R_{v-1}} |K|\left(\sum_{i\in T_j} y_i\right) \cdots \mathop{\mathbb{E}}_{y_1} |K|(y_1) \cdot \prod_{j\in R_1} |K|\left(\sum_{i\in T_j} y_i\right) \le$$

$$\prod_{i=1}^{v} \mathop{\mathbb{E}}_{y_i} |K(y_i)|^{a_i},$$

completing the proof of the lemma.

∎ (Lemma 3.5)

We can now complete the proof of Proposition 3.3. Fix $1 \leq r \leq k-2$. Let $S = (u_1, ..., u_k)$ be a sequence of vectors in L of rank r and with no coloops. This sequence is determined fully if we provide the following information: A subset B of r vectors in S which is a basis of S, an integer $1 \leq v \leq \min\{r, k-r\}$, a subset V of $S \setminus B$ of cardinality v which satisfies the conditions of Lemma 3.4; and given that, we describe the remaining $k-r-v$ vectors in S.

There are $\binom{k}{r}$ ways to choose the location of the subset B, $\binom{k-r}{v} \cdot v!$ ways to describe the location and the ordering of vectors in V. Then we need to choose the sets $S_1, ..., S_v$ described in Lemma 3.4. The number of such sets is at most 2^{rv}. Provided this information, the number of possible $(r+v)$-tuples of vectors in $B \cup V$ is bounded, by Lemma 3.5, by $\prod_{i=1}^{v} \mathbb{E}_{y_i} |K(y_i)|^{a_i}$, where $a_1, ..., a_v$ are determined by $S_1, ..., S_v$. Given such an $(r+v)$-tuple of vectors in $B \cup V$, the remaining $k-r-v$ vectors in S can be chosen in at most $2^{r(k-r-v)}$ ways (since they belong to the span of B).

Let $M(r,v) = \max_{(a_1,...,a_v)} \prod_{i=1}^{v} \mathbb{E}_{y_i} |K(y_i)|^{a_i}$, where the maximum is taken over all possible integer v-tuples $a_1, ..., a_v$ which satisfy $a_d \geq 2$ for all $1 \leq d \leq v$ and $\sum_{d=1}^{v} a_d = r+v$. Altogether, we get

$$N(r) \;\leq\; \sum_{v=1}^{min\{r,k-r\}} M(r,v) \cdot \min\{r, k-r\} \cdot \binom{k}{r} \cdot \left(\binom{k-r}{v} \cdot v! \right) \cdot 2^{rv} \cdot 2^{r(k-r-v)} \;\leq$$

$$\max_v \left\{ M(r,v) \right\} \cdot n \cdot k^{\min\{r,k-r\}} \cdot (k-r)^{\min\{r,k-r\}} \cdot 2^{r(k-r)} \;\leq\; \max_v \left\{ M(r,v) \right\} \cdot n^{2\min\{r,k-r\}+1} \cdot 2^{r(k-r)}.$$

We need to show that, up to lower order terms, this is at most $p^{k-r-1} \cdot 2^{\left(\psi(k,\gamma) + \frac{k}{2} h(\gamma)\right) \cdot n}$. Passing to exponents, it suffices to show that for any $1 \leq v \leq \min\{r, k-r\}$

$$\frac{1}{n} \cdot \log_2(M(r,v)) + \frac{2\min\{r,k-r\}+1}{n} \cdot \log_2(n) + \frac{r(k-r)}{n} \;\leq\; \psi(k,\gamma) + \frac{k}{2} h(\gamma) - (k-r-1)\lambda.$$

We refer to Section 2 for facts about the function ψ used in the following argument.

First, we upperbound $M(r,v)$. Let $a_1, ..., a_v$ satisfy $a_d \geq 2$ for all $1 \leq d \leq v$ and $\sum_{d=1}^{v} a_d = r+v$. Then

$$\frac{1}{n} \log_2 \left(\prod_{i=1}^{v} \mathop{\mathbb{E}}_{y_i} |K(y_i)|^{a_i} \right) \;\leq\; \sum_{i=1}^{v} \left(\psi(a_i,\gamma) + \frac{a_i}{2} h(\gamma) \right) \;=\; \sum_{i=1}^{v} \psi(a_i,\gamma) + \frac{r+v}{2} h(\gamma).$$

Using the fact that $\frac{\psi(a,\gamma)}{a-2}$ increases in a for $a > 2$, we have that $\sum_{i=1}^{v} \psi(a_i,\gamma) \leq \frac{r-v}{k-2} \cdot \psi(k,\gamma)$. Hence $\frac{1}{n} \log_2(M(r,v)) \leq \frac{r-v}{k-2} \cdot \psi(k,\gamma) + \frac{r+v}{2} h(\gamma)$. Therefore, it suffices to show that

$$\frac{2\min\{r,k-r\}+1}{n} \cdot \log_2(n) + \frac{r(k-r)}{n} \;\leq\; \frac{k-r+v-2}{k-2} \cdot \psi(k,\gamma) + \frac{k-r-v}{2} \cdot h(\gamma) - (k-r-1)\lambda.$$

We now use the fact that $\psi(k,\gamma) \geq \frac{k}{2} \cdot h(\gamma) - 1 = \frac{k-2}{2} \cdot h(\gamma) - (1 - h(\gamma))$. Substituting in the above expression, we need to show that

$$\frac{2\min\{r,k-r\}+1}{n} \cdot \log_2(n) + \frac{r(k-r)}{n} + \frac{k-r+v-2}{k-2} \cdot (1 - h(\gamma)) \;\leq\; (k-r-1) \cdot (h(\gamma) - \lambda).$$

Recall that $h(\gamma) - \lambda \geq \epsilon$, for a positive absolute constant ϵ. It is easy to see that there exist constants c_ϵ and C_ϵ such that for sufficiently large n and $C_\epsilon \leq k \leq c_\epsilon n$ each of the three summands on the LHS is upperbounded by $\frac{k-r-1}{3} \cdot \epsilon$, completing the proof of the proposition.

∎

References

[1] G. H. Hardy, J. E. Littlewood, and G. Polya, **Inequalities**, Cambridge University Press, 1988.

[2] N. Kirshner and A. Samorodnitsky, *A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres*, IEEE Trans. Inform. Theory 67, 2021, pp. 3509-3541.

[3] N. Linial and J. Mosheiff, *On the weight distribution of random binary linear codes*, Random Struct. Algorithms 56(1): 5-36 (2020).

[4] J.H. van Lint, **Introduction to Coding Theory**, third edition, Graduate Texts in Mathematics, vol. 86, Springer-Verlag, Berlin, 1999.

[5] J.H. van Lint and R.M. Wilson, **A Course in Combinatorics**, Second edition, Cambridge University Press, Cambridge, 2001.

[6] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory, vol. 23, 1977, pp. 157-166.

[7] R. O'Donnel, **Analysis of Boolean functions**, Cambridge University Press, 2014.