An upper bound on ℓ_q norms of noisy functions

Alex Samorodnitsky

Abstract

Let T_ϵ, $0 \leq \epsilon \leq 1/2$, be the noise operator acting on functions on the boolean cube $\{0,1\}^n$. Let f be a nonnegative function on $\{0,1\}^n$ and let $q \geq 1$. We upper bound the ℓ_q norm of $T_\epsilon f$ by the average ℓ_q norm of conditional expectations of f, given sets of roughly $(1 - 2\epsilon)^{r(q)} \cdot n$ variables, where r is an explicitly defined function of q.

We describe some applications for error-correcting codes and for matroids. In particular, we derive an upper bound on the weight distribution of BEC-capacity achieving binary linear codes and their duals. This improves the known bounds on the linear-weight components of the weight distribution of constant rate binary Reed-Muller codes for all (constant) rates.

I. INTRODUCTION

This paper considers the well-known problem of quantifying the decrease in the ℓ_q norm of a function on the boolean cube when this function is acted on by the noise operator.

Given a noise parameter $0 \leq \epsilon \leq 1/2$, the noise operator T_ϵ acts on functions on the boolean cube as follows: for $f : \{0,1\}^n \to \mathbb{R}$, $T_\epsilon f$ at a point x is the expected value of f at y, where y is a random binary vector whose i^{th} coordinate is x_i with probability $1 - \epsilon$ and $1 - x_i$ with probability ϵ, independently for different coordinates. Namely, $(T_\epsilon f)(x) = \sum_{y \in \{0,1\}^n} \epsilon^{|y-x|} (1-\epsilon)^{n-|y-x|} f(y)$, where $|\cdot|$ denotes the Hamming distance. We will write f_ϵ for $T_\epsilon f$, for brevity.

Note that f_ϵ is a convex combination of shifted copies of f. Hence, the noise operator decreases norms. An effective way to quantify this decrease for ℓ_q norms is given by the hypercontractive inequality [3], [7], [2] (see also [10]):

$$\|f_\epsilon\|_q \leq \|f\|_{1+(q-1)(1-2\epsilon)^2}. \tag{1}$$

Entropy provides another example of a convex homogeneous functional on (nonnegative) functions on the boolean cube. For a nonnegative function f let the entropy of f be given by $Ent(f) = \mathbb{E} f \log_2 f - \mathbb{E} f \log_2 \mathbb{E} f$ (where the expectation is taken w.r.t. the uniform measure on $\{0,1\}^n$). The entropy of f is closely related to Shannon's entropy of the corresponding distribution $f/\Sigma f$ on $\{0,1\}^n$ (see e.g., the discussion in the introduction of [13]), and similarly the entropy of f_ϵ is related to Shannon's entropy of the output of a binary symmetric channel with error probability ϵ on input distributed according to $f/\Sigma f$. The decrease in entropy (or, correspondingly the increase in Shannon's entropy) after noise is quantified in the "Mrs. Gerber's Lemma" [15]: $Ent(f_\epsilon) \leq n \mathbb{E} f \cdot \phi \left(\frac{Ent(f)}{n \mathbb{E} f}, \epsilon \right)$, where ϕ is an explicitly given function on $[0,1] \times [0,1/2]$.

There is a well-known connection between the ℓ_q norms of a nonnegative function f and its entropy (see e.g., [6]): Assume, as we may by homogeneity, that $\mathbb{E} f = 1$. Then $Ent(f) = \lim_{q \to 1} \frac{1}{q-1} \log_2 \|f\|_q^q$. The quantity $Ent_q(f) = \frac{1}{q-1} \log_2 \|f\|_q^q$ is known as the qth Renyi entropy of f ([12]).

The starting point for this paper is an alternative way to quantify the decrease in entropy after noise given in [13]. To state this result, let us introduce some notation. For $0 \leq \lambda \leq 1$, let $T \sim \lambda$ denote a random subset

T of $[n]$ in which each element is chosen independently with probability λ. Let $\mathbb{E}(f|T)$ be the conditional[2] expectation of f with respect to T, that is a function on $\{0,1\}^n$ defined by $\mathbb{E}(f|T)(x) = \mathbb{E}_{y:y_{|T}=x_{|T}} f(y)$. Then ([13]):

$$Ent\,(f_\epsilon) \;\leq\; \mathbb{E}_{T\sim(1-2\epsilon)^2} Ent\big(\mathbb{E}(f|T)\big). \tag{2}$$

In light of the connection between entropy and ℓ_q norms, it is natural to ask whether (2) can be extended to an inequality between ℓ_q norms (equivalently, Renyi entropies). The main result of this paper is a positive answer to this question.

Theorem 1.1.: Let f be a nonnegative function on $\{0,1\}^n$. Then, for any $q > 1$ holds

$$\log \|f_\epsilon\|_q \;\leq\; \mathbb{E}_{T\sim\lambda} \log \|\mathbb{E}(f|T)\|_q, \tag{3}$$

with $\lambda = \lambda(q,\epsilon) = (1-2\epsilon)^r$, where the exponent $r = r(q)$ is given by the following expression:

$$r(q) \;=\; \begin{cases} \frac{1}{2\ln 2} \frac{2^{3-q}\left(2^{q-1}-1\right)}{q-1} & 1 < q \leq 2 \\ \frac{1}{2\ln 2}\frac{q}{q-1} & q \geq 2 \end{cases} \tag{4}$$

Note that (3) is in fact an extension of (2), since assuming $\mathbb{E}\,f = 1$, dividing both sides by $q - 1$ and taking the limit as $q \to 1$, recovers (2). The only thing to observe is that $\lim_{q\to 1} r(q) = 2$.

Remark 1.2.: Since conditional expectation of a function f is a convex combination of shifted copies of f, its norm is smaller than that of f, and hence (3) measures the decrease in ℓ_q-norm under noise, similarly to (1). In fact, the proof of (3) follows the approach of [7] to the proof of (1). In this approach, we view both sides of the corresponding inequality as functions of ϵ (for a fixed q) and compare the derivatives. Since noise operators form a semigroup it suffices to compare the derivatives at zero, and this is done via an appropriate logarithmic Sobolev inequality (see (7)). ∎

A. Applications

Let $C \subseteq \{0,1\}^n$ be a binary linear code. Let $r_C(\cdot)$ denote the rank function of the binary matroid defined by C. That is, $r_C(T)$ is the rank of the column submatrix of a generating matrix of C which contains columns indexed by T. Applying (3) to the scaled characteristic function of C, that is $f = \frac{2^n}{|C|} \cdot 1_C$, gives the following claim, connecting between the linear structure of C and its behavior under noise.

Proposition 1.3.: For any $0 \leq \lambda \leq 1$ holds

$$\lambda n - \mathbb{E}_{T\sim\lambda} r_C(T) \;\geq\; \max_{1<q<\infty} \frac{1}{q-1} \cdot \log_2 \mathbb{E}\,f_{\epsilon(q)}^q, \quad \text{for} \quad \epsilon(q) = \frac{1-\lambda^{1/r(q)}}{2}. \tag{5}$$

Let us discuss the bound given by (5). First, it seems useful to extend it to the limiting values of q as well. Let $F(\lambda, q) = \frac{1}{q-1} \cdot \log_2 \mathbb{E}\,f_{\epsilon(q)}^q$, for $1 < q < \infty$. Set $F(\lambda, 1) = \lim_{q\to 1} F(\lambda, q) = Ent\left(f_{\frac{1-\sqrt{\lambda}}{2}}\right)$ and $F(\lambda, \infty) = \lim_{q\to\infty} F(\lambda, q) = \log_2 \|f_{\frac{1-\lambda^2\ln 2}{2}}\|_\infty$. Then we get $\lambda n - \mathbb{E}_{T\sim\lambda} r_C(T) \geq \max_{1\leq q\leq\infty} F(\lambda, q)$. For any λ, the three values of $F(\lambda, q)$ which seem to be easier to understand are $F(\lambda, 1)$, $F(\lambda, \infty)$ and $F(\lambda, 2) = \log_2 \mathbb{E}\,f_{\frac{1-\lambda^{\ln 2}}{2}}^2$. While simple examples show that $F(\lambda, 1)$ can be both smaller and greater than $F(\lambda, 2)$, intriguingly we always have $F(\lambda, 2) = F(\lambda, \infty)$, and moreover this quantity is easily expressible in terms of the weight distribution of the code C or of the dual code $C^\perp$.

Lemma 1.4: Let $(a_0, ..., a_n)$ be the weight distribution of C. That is $a_k = \Big| \{x \in C, |x| = k\} \Big|$, for $k = 0, ..., n$. Similarly, let $(b_0, ..., b_n)$ be the weight distribution of $C^\perp$. Then, writing θ for $\lambda^{2\ln 2}$, we have

$$F(\lambda, 2) \;=\; F(\lambda, \infty) \;=\; \log_2 \sum_{i=0}^{n} b_i \theta^i \;=\; \log_2\left(\frac{1}{|C|} \cdot \sum_{k=0}^{n} a_k (1-\theta)^k (1+\theta)^{n-k}\right).$$

We proceed to discuss some applications of (5).

1) Codes achieving capacity for the binary erasure channel: This is a family of codes for which we can effectively upper bound the LHS of (5), for λ close to the rate of the code. Recall that the rate of a linear code $R(C)$ is defined as $\frac{1}{n}\log_2(|C|) = \frac{\dim(C)}{n}$. A code $C \subseteq \{0,1\}^n$ (more precisely a family of codes indexed by n) achieves capacity in the block-error rate sense for the BEC (binary erasure channel) if there exist two functions δ_1, δ_2 which go to zero with n such that for $\lambda = R(C) + \delta_1$ the probability of decoding error given by $p_e = \Pr_{T \sim \lambda}\{r_C(T) < R(C) \cdot n\}$ is upperbounded by δ_2. This immediately implies (writing R for $R(C)$) that

$$\lambda n - \mathop{\mathbb{E}}_{T \sim \lambda} r_C(T) \ \leq \ \lambda n - (1 - \delta_2) \cdot Rn \ = \ \delta_1 n + \delta_2 Rn \ = \ o(n).$$

Hence we have the following corollary of (5).

Corollary 1.5.: Let C be a BEC capacity achieving code of rate R. Then there exists $\lambda = R + o(1)$ such that for all $1 \leq q \leq \infty$ holds $F(\lambda, q) \leq o(n)$.

The inequality $Ent\left(f_{\frac{1-\sqrt{\lambda}}{2}}\right) = F(\lambda, 1) \leq o(n)$ has been interpreted ([11], using (2)) as indicating that C is 'well-dispersed' in $\{0,1\}^n$ in the sense that the output of the binary symmetric channel with error probability $\epsilon = \frac{1-\sqrt{\lambda}}{2} \approx \frac{1-\sqrt{R}}{2}$ whose input is a random codeword from C is very close in the Kullback-Leibler distance to the uniform distribution. We observe that Corollary 1.5 can be interpreted as providing increasingly stronger measures of proximity of the channel output to uniform, as noise increases beyond $\frac{1-\sqrt{R}}{2}$.

The inequality $F(\lambda, 2) \leq o(n)$ together with Lemma 1.4 provide the following bound on the components of the weight distribution of a code which achieves BEC capacity, or a code whose dual achieves BEC capacity.

Proposition 1.6: Let C a linear code of rate R. Let $(a_0, ..., a_n)$ be the weight distribution of C. For $0 \leq k \leq n$, let $k^* = \min\{k, n-k\}$. Let $\theta = R^{2 \ln 2}$.

- If $C^\perp$ achieves BEC capacity, then for all $0 \leq k \leq n$ holds

$$a_k \ \leq \ 2^{o(n)} \cdot \left(\frac{1}{1-R}\right)^{(2 \ln 2) \cdot k^*}.$$

- If C achieves BEC capacity, then for all $0 \leq k \leq n$ holds

$$a_k \ \leq \ 2^{o(n)} \cdot \begin{cases} \frac{|C|}{(1-\theta)^{k^*}(1+\theta)^{n-k^*}} & 0 \leq k^* \leq \frac{1-\theta}{2} \cdot n \\ \frac{\binom{n}{k^*} \cdot |C|}{2^n} & \text{otherwise} \end{cases}$$

In particular, since the dual of a Reed-Muller code is a Reed-Muller code, and since Reed-Muller codes achieve BEC capacity [9], both bounds in Proposition 1.6 hold for Reed-Muller codes. The bounds in the literature [1], [8], [14] seem to focus mostly on Reed-Muller codes of rates close to 0 or 1, or on weights which increase sublinearly in n, but do extend to all weights and to all rates. Comparing with these bounds, it seems that Proposition 1.6 improves the bounds on $\{a_k\}$, for k growing linearly in n, for constant-rate Reed-Muller codes of all rates $0 < R < 1$. (See Section III-1 for a quick comparison between the bounds.)

Another implication of Proposition 1.6 which seems to be worth drawing attention to is that a linear code C of rate R that achieves BEC capacity has $a_k \leq 2^{o(n)} \cdot \frac{|C|\binom{n}{k}}{2^n}$, for $k \in \frac{1 \pm R^{2 \ln 2}}{2} \cdot n$. In other words, in this range of weights, the weight distribution of C is upper-bounded, up to a relatively small error, by the weight distribution of a typical random code of the same rate.

2) Bounds on weight distribution of linear codes: The following (more general) claim is an immediate corollary of the inequality $\lambda n - \mathbb{E}_{T \sim \lambda} r_C(T) \geq F(\lambda, 2)$ and of Lemma 1.4.

Corollary 1.7.: Let C be a binary linear code and let $(b_0, ..., b_n)$ be the distance distribution of the dual code $C^\perp$. Then for any $0 \leq \lambda \leq 1$ and for any $0 \leq i \leq n$ holds

$$b_i \ \leq \ \lambda^{-(2 \ln 2) \cdot i} \cdot 2^{\lambda n - \mathbb{E}_{T \sim \lambda} r_C(T)}. \tag{6}$$

This claim gives upper bounds on the components of the weight distribution of $C^\perp$, provided we can upper bound $\lambda n - \mathbb{E}_{T \sim \lambda} r_C(T)$.

3) Ranks of random subsets in a binary matroid: We give a different way to write the inequality $\lambda n - \frac{4}{}$ $\mathbb{E}_{T \sim \lambda} r_C(T) \geq F(\lambda, 2)$, stating it as a lemma since it requires a (simple) argument.

Lemma 1.8:: Let $r_C(\cdot)$ be the rank function of the binary matroid on $\{1, ..., n\}$ defined by a generating matrix of a linear code C of length n. Let $0 \leq p \leq 1$ and let $t = p^{\frac{1}{2 \ln 2}}$. Then

$$\log_2 \mathop{\mathbb{E}}_{S \sim p} \left(2^{|S| - r_C(S)} \right) \quad \leq \quad \mathop{\mathbb{E}}_{T \sim t} \left(|T| - r_C(T) \right).$$

Let the dimension of C be k. Recall that the Tutte polynomial [5] of the matroid defined by C is $T_C(x, y) = \sum_{S \subseteq \{1, ..., n\}} (x - 1)^{k - r_C(S)} (y - 1)^{|S| - r_C(S)}$. It is easy to see that an alternative way to write the inequality of Lemma 1.8 is

$$\log_2 \left[p^k (1 - p)^{n-k} T_C \left(\frac{1}{p}, \frac{1+p}{1-p} \right) \right] \quad \leq \quad t^{k+1} (1 - t)^{n-k-1} \left(\frac{d}{dy} T_C \right) \left(\frac{1}{t}, \frac{1}{1-t} \right).$$

As an immediate implication of Lemma 1.8 we get the following tail bound.

Corollary 1.9:: For any $\Delta \geq 0$ holds

$$Pr_{S \sim p} \left\{ |S| - r_C(S) \geq \mathop{\mathbb{E}}_{T \sim t} \left(|T| - r_C(T) \right) + \Delta \right\} \quad \leq \quad 2^{-\Delta}.$$

Remark 1.10:: Consider a different approach to obtain tail bounds for the function $|S| - r_C(S)$. Let $f(S) = |S| - r_C(S)$. Then $|f(x) - f(y)| \leq \|x - y\|$, for any $x, y \in \{0, 1\}^n$ (here $\|x - y\|$ stands for the Hamming distance between x and y) and hence, by the 'bounded differences' inequality [4], for any $t \geq 0$ holds $Pr_{S \sim p} \{ f(S) \geq \mathbb{E}_{T \sim p} f(T) + t \} \leq e^{-2t^2/n}$.

Let $\mu : p \mapsto \mathbb{E}_{S \sim p} f(S)$. By the Margulis-Russo formula [10], $\mu' = \frac{1}{p} \mathbb{E}_{S \sim p} \sum_{i \in S} (f(S) - f(S \setminus \{i\}))$. Hence, since f is monotone increasing, $\mu' \geq 0$ and μ is increasing. Moreover, since f is supermodular (by submodularity of rank), it is easy to see that μ is convex. In particular, since $\mu(0) = 0$, we have $\mu(t) \geq \frac{t}{p} \mu(p)$ for $t \geq p$. Taking everything into account, we have $Pr_{S \sim p} \{ f(S) \geq \mathbb{E}_{T \sim t} f(T) + \Delta \} \leq e^{-\frac{2(\mu(t) - \mu(p) + \Delta)^2}{n}} \leq e^{-\frac{2((t-p)\mu(p) + p\Delta)^2}{p^2 n}}$.

Note that this does not seem to recover the bound of Corollary 1.9 when $\mu(t)$ is small, as in the case of BEC capacity achieving codes. ■

To conclude, we consider the special case of graphic matroids. For a graph $G = (V, E)$ with n edges, let M be the matroid on $\{1, ..., n\}$ whose independent sets are forests in G. This is a binary matroid, and Lemma 1.8 specializes as follows.

Corollary 1.11:: Let $G = (V, E)$ be a graph. For a subset of edges $S \subseteq E$, let $c(S)$ denote the number of the connected components in the subgraph (V, S). Then

$$\log_2 \mathop{\mathbb{E}}_{S \sim p} \left(2^{|S| + c(S)} \right) \quad \leq \quad t|E| + \mathop{\mathbb{E}}_{T \sim t} c(T).$$

This paper is organized as follows. We prove Theorem 1.1 in Section II. All the remaining proofs are in Section III.

II. PROOF OF THEOREM 1.1

We start with a log-Sobolev-type inequality. Recall that the Dirichlet form $\mathcal{E}(f, g)$ for functions f and g on the boolean cube is defined by $\mathcal{E}(f, g) = \mathbb{E}_x \sum_{y \sim x} \left(f(x) - f(y) \right) \left(g(x) - g(y) \right)$. Here $y \sim x$ means that x and y differ in precisely one coordinate.

Theorem 2.1.: Let f be a nonnegative function on $\{0,1\}^n$. Then for any $q \geq 1$ holds

$$\mathcal{E}\left(f^{q-1}, f\right) \geq 4r(q) \, \mathbb{E} \, f^q \cdot \left(n \cdot \ln ||f||_q \; - \sum_{|T|=n-1} \ln ||\, \mathbb{E}(f|T)||_q \right), \tag{7}$$

where $r(q)$ is given in (4).

Moreover for $1 < q < 2$ equality is attained if and only if f is a constant function, and for $q \geq 2$ if and only if f satisfies the following condition: If x and y differ in precisely one coordinate and if $f(x) \neq f(y)$, then f vanishes at either x or y.[1]

Corollary 2.2.: Let $\delta > 0$ be arbitrarily small. For any nonnegative non-constant function f on $\{0,1\}^n$ and for any $q > 1$ holds

$$\mathcal{E}\left(f^{q-1}, f\right) > 4\left(r(q) - \delta\right) \mathbb{E} \, f^q \cdot \left(n \cdot \ln ||f||_q \; - \sum_{|T|=n-1} \ln ||\, \mathbb{E}(f|T)||_q \right).$$

Theorem 2.1 is proved below. For now we assume that it holds and proceed with the proof of Theorem 1.1. We assume, as we may, that the logarithms on both sides of (3) are the natural logarithms. Observe that the claim of the theorem is immediate for constant functions, since both sides of (3) are $\ln(\mathbb{E} \, f)$. So we may and will assume that f is not constant (note that this implies f_ϵ is non-constant for all $0 \leq \epsilon < 1/2$).

Fix $q > 1$. Let $\delta > 0$. We will prove (3) for $\lambda(q, \epsilon) = (1 - 2\epsilon)^{r(q)-\delta}$. Taking δ to zero will then imply (3) for $\lambda = (1 - 2\epsilon)^{r(q)}$ as well.

The proof proceeds by induction on n. The claim clearly holds for $n = 0$. Let $n > 0$. We assume that the claim holds for all dimensions smaller than n, and show that it holds for n as well. For fixed q and δ both sides of (3) are functions of f and ϵ. Let $\mathcal{F}(f, \epsilon)$ denote the LHS and $\mathcal{G}(f, \epsilon)$ the RHS. Clearly $\mathcal{F}(f, 0) = \mathcal{G}(f, 0) = \ln ||f||_q$. We will argue that if $\mathcal{F}(f, \epsilon) = \mathcal{G}(f, \epsilon)$ for some $0 \leq \epsilon < 1/2$ then necessarily $\mathcal{F}'(f, \epsilon) < \mathcal{G}'(f, \epsilon)$ (here and below the derivatives are taken w.r.t. ϵ). It is easy to see that this implies $\mathcal{F}(f, \epsilon) \leq \mathcal{G}(f, \epsilon)$, for all $0 \leq \epsilon \leq 1/2$, which is precisely what we need to show.

Due to the fact that the noise operators form a semigroup under composition: $T_\rho \circ T_\epsilon = T_{\epsilon + \rho - 2\epsilon\rho}$, it will suffice to compare the derivatives at zero (see (8) and the argument preceding it). This comparison is done in the following lemma.

Lemma 2.3.: Let f be a nonnegative non-constant function on $\{0,1\}^n$. Then

$$\mathcal{F}'(f, 0) \; < \; \mathcal{G}'(f, 0).$$

Proof: As shown in [7], we have $\mathcal{F}'(f, 0) = -\frac{\mathcal{E}\left(f^{q-1}, f\right)}{2 \, \mathbb{E} \, f^q}$. We claim that

$$\mathcal{G}'(f, 0) = -2(r(q) - \delta) \cdot \left(n \ln ||f||_q - \sum_{|T|=n-1} \ln ||\, \mathbb{E}(f|T)||_q \right),$$

which will conclude the proof of the lemma by Corollary 2.2. In fact, note that $\lambda(q, 0) = 1$ and that $\lambda'(q, 0) = -2(r(q) - \delta)$. Hence we have:

$$\mathcal{G}'(f, 0) = \frac{d}{d\epsilon}_{|\epsilon = 0} \mathop{\mathbb{E}}_{T \sim \lambda} \ln ||\, \mathbb{E}(f|T)||_q = \frac{d}{d\epsilon}_{|\epsilon = 0} \left(\lambda^n \cdot \ln ||f||_q + \lambda^{n-1}(1 - \lambda) \cdot \sum_{|T|=n-1} \ln ||\, \mathbb{E}(f|T)||_q \right) =$$

---

[1]Conditions of equality in (7) are stated for completeness. They are not used in the proof of Theorem 1.1.

$$-2(r(q) - \delta) \cdot \left(n \cdot \ln \|f\|_q \; - \sum_{|T|=n-1} \ln \| \mathbb{E}(f|T) \|_q \right),$$

completing the proof.

∎

We can now conclude the proof of Theorem 1.1. Assume that $\mathcal{F}(f, \epsilon) = \mathcal{G}(f, \epsilon)$, for some $0 \le \epsilon < 1/2$. Consider the functions $\mathcal{F}(f_\epsilon, \rho)$ and $\mathcal{G}(f_\epsilon, \rho)$, for $0 \le \rho \le 1/2$. We have

$$\mathcal{F}(f_\epsilon, \rho) \;=\; \ln \| (f_\epsilon)_\rho \|_q \;=\; \ln \| f_{\epsilon + (1-2\epsilon)\cdot\rho} \|_q \;=\; \mathcal{F}\big(f, \epsilon + (1-2\epsilon) \cdot \rho\big),$$

and

$$\mathcal{G}(f_\epsilon, \rho) \;=\; \mathop{\mathbb{E}}_{R \sim \lambda(q,\rho)} \ln \| \mathbb{E}(f_\epsilon | R) \|_q \;\le\; \mathop{\mathbb{E}}_{R \sim \lambda(q,\rho)} \; \mathop{\mathbb{E}}_{T \subseteq R, T \sim \lambda(q,\epsilon)} \ln \| \mathbb{E}(f|T) \|_q \;=$$

$$\mathop{\mathbb{E}}_{T \sim \lambda(q,\rho)\cdot\lambda(q,\epsilon)} \ln \| \mathbb{E}(f|T) \|_q \;=\; \mathop{\mathbb{E}}_{T \sim \lambda(q,\epsilon + (1-2\epsilon)\cdot\rho)} \ln \| \mathbb{E}(f|T) \|_q \;=\; \mathcal{G}\big(f, \epsilon + (1-2\epsilon) \cdot \rho\big).$$

The inequality follows from the induction hypothesis for $R \subset [n]$ and from the assumption $\mathcal{F}(f, \epsilon) = \mathcal{G}(f, \epsilon)$ for $R = [n]$.

Hence, using by Lemma 2.3 in the first inequality and the fact that $\mathcal{G}(f_\epsilon, 0) = \ln \|f_\epsilon\|_q = \mathcal{F}(f, \epsilon) = \mathcal{G}(f, \epsilon)$ in the second inequality, we have

$$\mathcal{F}'(f, \epsilon) \;=\; \frac{1}{1 - 2\epsilon} \cdot \mathcal{F}'(f_\epsilon, 0) \;<\; \frac{1}{1 - 2\epsilon} \cdot \mathcal{G}'(f_\epsilon, 0) \;\le\; \mathcal{G}'(f, \epsilon), \tag{8}$$

concluding the proof.

∎

A. Proof of Theorem 2.1

We start with the base case $n = 1$. This case is dealt with in the following claim.

Proposition 2.4.: Let g be a nonnegative function on a 2-point space with $\mathbb{E}\, g = 1$. Then

$$\mathcal{E}\big(g^{q-1}, g\big) \;\ge\; 4r(q)\, \mathbb{E}\, g^q \ln \|g\|_q, \tag{9}$$

where $r(q)$ is given in (4).

Moreover for $1 < q < 2$ equality is attained if and only if g is a constant function, and for $q \ge 2$ if and only if g is a constant function, or a mutiple of a characteristic function of a point.

Proposition 2.4 is proved below. Here we assume its validity and proceed with the proof of Theorem 2.1. First, we restate the 2-point inequality (9) in an equivalent form, with two modifications: replacing $\|g\|_q$ with $\mathbb{E}\, g^q$ and extending the inequality by homogeneity to general nonnegative functions.

Let g be a nonnegative function on a 2-point space. Then

$$\mathcal{E}\big(g^{q-1}, g\big) \;\ge\; \frac{4r(q)}{q} \mathbb{E}\, g^q \ln \frac{\mathbb{E}\, g^q}{(\mathbb{E}\, g)^q}.$$

We can now extend this inequality to general n by the following standard argument. Let f be a nonnegative function on $\{0, 1\}^n$. Then, denoting by e_i the ith unit vector, we have

$$\mathcal{E}\big(f^{q-1}, f\big) \;=\; \sum_{i=1}^n \mathop{\mathbb{E}}_x \big(f^{q-1}(x) - f^{q-1}(x + e_i)\big)\big(f(x) - f(x + e_i)\big).$$

For x and i, let $f_{x,i}$ denote the restriction of f to the 2-point space $(x, x+e_i)$. Then the RHS is $\sum_{i=1}^{n} \frac{1}{2^{n-1}} \sum_{x:x_i=0}^{7} \mathcal{E}\left(f_{x,i}^{q-1}, f_{x,i}\right)$.
By the 2-point inequality, this is at least

$$\frac{4r(q)}{q} \cdot \sum_{i=1}^{n} \frac{1}{2^{n-1}} \sum_{x:x_i=0} \mathbb{E}\, f_{x.i}^{q} \ln \frac{\mathbb{E}\, f_{x.i}^{q}}{(\mathbb{E}\, f_{x,i})^{q}} \quad = \quad \frac{4r(q)}{q} \cdot \sum_{i=1}^{n} \frac{1}{2^{n-1}} \sum_{x:x_i=0} \mathbb{E}\, f_{x.i}^{q}(-\ln)\frac{(\mathbb{E}\, f_{x,i})^{q}}{\mathbb{E}\, f_{x.i}^{q}}$$

Let $\theta_{x,i} = \frac{\mathbb{E}\, f_{x,i}^{q}}{\sum_{y:y_i=0} \mathbb{E}\, f_{y,i}^{q}}$. Then $\sum_{x:x_i=0} \theta_{x,i} = 1$. Note also that $\sum_{y:y_i=0} \mathbb{E}\, f_{y,i}^{q} = 2^{n-1} \cdot \mathbb{E}\, f^{q}$. By convexity
of $(-\ln)$ the RHS above is at least

$$\frac{4r(q)}{q} \cdot \mathbb{E}_x\, f^{q} \cdot \sum_{i=1}^{n}(-\ln)\left(\sum_{x:x_i=0} \theta_{x,i} \frac{(\mathbb{E}\, f_{x,i})^{q}}{\mathbb{E}\, f_{x,i}^{q}}\right) \quad = \quad \frac{4r(q)}{q} \cdot \mathbb{E}_x\, f^{q} \cdot \sum_{i=1}^{n}(-\ln)\frac{\sum_{x:x_i=0}(\mathbb{E}\, f_{x,i})^{q}}{2^{n-1} \cdot \mathbb{E}_x\, f^{q}}$$

$$= \frac{4r(q)}{q} \cdot \mathbb{E}_x\, f^{q} \cdot \sum_{i=1}^{n} \ln \frac{\mathbb{E}\, f^{q}}{\mathbb{E}\left(\mathbb{E}\left(f|[n] \setminus \{i\}\right)^{q}\right)} \quad = \quad 4r(q)\, \mathbb{E}_x\, f^{q} \cdot \left(n \cdot \ln \|f\|_q - \sum_{|T|=n-1} \ln \|\mathbb{E}(f|T)\|_q\right),$$

proving (7).

It remains to consider the cases of equality. Tracking back the conditions for equality in the proof above, it is easy to see that equality holds for f if and only if it holds for all one-dimensional restrictions $f_{x,i}$. For $1 < q < 2$ this means that all these restrictions are constant functions, implying f is constant. For $q \geq 2$ this means these restrictions are either constant, or multiples of a characteristic function of a point. Alternatively, if x and y differ in precisely one coordinate and if $f(x) \neq f(y)$, then f vanishes at either x or y. ∎

B. Proof of Proposition 2.4

There are two cases to consider, $1 < q < 2$ and $q \geq 2$. These cases are dealt with in the following subsections, starting with the somewhat easier case $q \geq 2$.

1) The case $q \geq 2$: We will show that for a nonnegative function g on a 2-point space with $\mathbb{E}\, g = 1$ holds

$$\mathcal{E}\left(g^{q-1}, g\right) \geq \frac{2}{\ln 2}\frac{1}{q-1} \cdot \mathbb{E}\, g^{q} \ln \mathbb{E}\, g^{q}. \tag{10}$$

Observe that equality holds in two cases: if g is a constant-1 function, a multiple of a characteristic function of a point. We will show that these are the only two cases for which equality holds.

Assume, w.l.o.g., that $g(0) \leq g(1)$. Let $0 \leq x \leq 1$, and let $g(0) = x$ and $g(1) = 2 - x$. Then (10) becomes:

$$\left((2-x)^{q-1} - x^{q-1}\right) \cdot \left((2-x) - x\right) \quad \geq \quad \frac{2}{\ln 2} \cdot \frac{1}{q-1} \cdot \frac{x^{q} + (2-x)^{q}}{2} \ln\left(\frac{x^{q} + (2-x)^{q}}{2}\right)$$

Setting $y = x(2-x)^{q-1} + x^{q-1}(2-x)$, $z = x^{q} + (2-x)^{q}$, and rearranging, this is the same as $1 - \frac{y}{z} \geq \frac{1}{q-1} \cdot \log_2 \frac{z}{2}$.
Note that since $x = 0$ is one of the cases for which equality holds, we may assume that $x > 0$. Now, let $t = \frac{2-x}{x}$.
Then $t \geq 1$, $y = 2^{q} \cdot \frac{t^{q-1}+t}{(t+1)^{q}}$, and $z = 2^{q} \cdot \frac{t^{q}+1}{(t+1)^{q}}$. Rearranging and simpifying, (10) becomes

$$\log_2\left(\left(\frac{(t+1)^{q}}{t^{q}+1}\right)^{\frac{1}{q-1}}\right) \geq \frac{t^{q-1}+t}{t^{q}+1}, \quad \text{for} \quad t \geq 1,\ q \geq 2.$$

Let $s = \left(\frac{(t+1)^{q}}{t^{q}+1}\right)^{\frac{1}{q-1}}$. Clearly $s > 1$. By convexity of the function $a \to a^{q}$, we have $s \leq 2$. By concavity of logarithm $\log_2(s) \geq s - 1$. (We remark that the only case in which equality holds in this inequality is if

$s = 2$, that is $t = 1$, and this implies g is constant.) Hence, it suffices to prove $\left(\frac{(t+1)^q}{t^q+1}\right)^{\frac{1}{q-1}} \geq \frac{(t+1)\left(t^{q-1}+1\right)^8}{t^q+1}$.

Simplifying and rearranging, this is the same as $(t+1) \cdot (t^q+1)^{q-2} \geq \left(t^{q-1}+1\right)^{q-1}$.

Let h be a function on the two-point space with values t and 1. Then the inequality above is the same as $\|h\|_1 \cdot \|h\|_q^{q(q-2)} \geq \|h\|_{q-1}^{(q-1)^2}$. This can be shown as follows, using Hölder's inequality in the third step,

$$\|h\|_{q-1}^{q-1} \;=\; \mathbb{E}\, h^{q-1} \;=\; \left\langle h^{1/(q-1)}, h^{q(q-2)/(q-1)}\right\rangle \;\leq$$

$$\|h^{1/(q-1)}\|_{q-1} \cdot \|h^{q(q-2)/(q-1)}\|_{(q-1)/(q-2)} \;=\; \|h\|_1^{1/(q-1)} \cdot \|h\|_q^{q(q-2)/(q-1)}.$$

This completes the proof of (10). Tracing back the conditions for equality in (10), it is easy to see that it holds only in the two cases mentioned above. ∎

2) The case $1 < q < 2$: We will show that

$$\mathcal{E}\left(g^{q-1}, g\right) \;\geq\; \frac{2}{\ln 2} \cdot r(q) \cdot \mathbb{E}\, g^q \ln \mathbb{E}\, g^q, \quad r(q) = \frac{2^{3-q}\left(2^{q-1}-1\right)}{q(q-1)}, \tag{11}$$

and that equality holds iff g is a constant-1 function.

Let us first deal with the case $g(0) = 0$. In this case, the inequality reduces to $q \geq 2^{3-q}\left(2^{q-1}-1\right) = 4 - 2^{3-q}$, and it is easy to verify that this is a strict inequality for $1 < q < 2$.

Proceeding as in the proof of the first case, using the same notation, and writing r for $r(q)$, the inequality (11) is equivalent to $1 - \frac{y}{z} \geq r \cdot \log_2 \frac{z}{2}$. Substituting $t = \frac{2-x}{x}$ (note that we may assume $x = g(0) > 0$) and rearranging, this is the same as

$$r \cdot \log_2\left(\frac{(t+1)^q}{t^q+1}\right) \;\geq\; \frac{t + t^{q-1}}{t^q+1} + r \cdot (q-1) - 1.$$

Since $1 < \frac{(t+1)^q}{t^q+1} \leq 2^{q-1}$, concavity of the logarithm implies $\log_2\left(\frac{(t+1)^q}{t^q+1}\right) \geq \frac{q-1}{2^{q-1}-1} \cdot \left(\frac{(t+1)^q}{t^q+1} - 1\right)$. (Equality holds only if $\frac{(t+1)^q}{t^q+1} = 2^{q-1}$, that is $t = 1$, which means that g is constant.) So, it would suffice to show that $r \cdot \frac{q-1}{2^{q-1}-1}\left(\frac{(t+1)^q}{t^q+1} - 1\right) \geq \frac{t+t^{q-1}}{t^q+1} + r \cdot (q-1) - 1$, which is the same as $\frac{r(q-1)}{2^{q-1}-1} \frac{(t+1)^q}{t^q+1} \geq \frac{t+t^{q-1}}{t^q+1} + \frac{r(q-1)2^{q-1} - 2^{q-1}+1}{2^{q-1}-1}$.

Multiplying both sides by $\left(2^{q-1}-1\right)\left(t^q+1\right)$ gives

$$r(q-1)(t+1)^q \;\geq\; \left(2^{q-1}-1\right)\left(t+t^{q-1}\right) + \left(r(q-1)2^{q-1} - 2^{q-1}+1\right)\left(t^q+1\right).$$

Writing the above as $F(t) \geq G(t) + H(t)$, it is easy to see that $F(1) = G(1) + H(1)$ and that $F'(1) = G'(1) + H'(1)$. So it suffices to show $F''(t) \geq G''(t) + H''(t)$, which amounts to (after some simplification and rearranging):

$$rq(q-1)(t+1)^{q-2} + 2^{q-1}t^{q-3} \cdot \left((2-q) + \left(q - rq(q-1)\right)\cdot t\right) \;\geq\; (2-q)t^{q-3} + qt^{q-2},$$

or

$$rq(q-1)\left(\frac{t+1}{t}\right)^{q-3}(t+1) + 2^{q-1} \cdot \left((2-q) + \left(q - rq(q-1)\right)\cdot t\right) \;\geq\; (2-q) + qt.$$

Substituting $y = \frac{t+1}{t}$ and rearranging, this is the same as proving for $1 < y \leq 2$ that

$$rq(q-1)y^{q-2} + (2-q)\left(2^{q-1}-1\right)y \;\geq\; \left(rq(q-1) - (2q-2)\right)2^{q-1} + (2q-2).$$

Denote the LHS by $h(y)$. Then h' is proportional to $\left(2^{q-1}-1\right) - rq(q-1)y^{q-3}$. Substituting $r = r(q) = \frac{2^{3-q}\left(2^{q-1}-1\right)}{q(q-1)}$, we see that h' is proportional to $1 - \left(\frac{2}{y}\right)^{3-q}$. Hence $h' < 0$ for all $y < 2$, and it suffices to verify the last inequality for $y = 2$. Substituting $y = 2$ we get an identity in q.

This completes the proof of (11). Tracing back the conditions for equality, it is easy to see that it holds only if g is a constant function. ∎

III. REMAINING PROOFS

Proof of Proposition 1.3: We will need some facts from Fourier analysis on the boolean cube [10]. First, recall that for a function f on $\{0,1\}^n$ and for a subset $T \subseteq [n]$ holds $\mathbb{E}(f|T) = \sum_{R \subseteq T} \widehat{f}(R) w_R$, where $\{w_R\}$ are the Walsh-Fourier characters. Recall also that if $f = \frac{2^n}{|C|} \cdot 1_C$, where C is a linear code, then $\widehat{f} = 1_{C^\perp}$.

We also recall a fact from linear algebra. For $T \subseteq [n]$, let $C_T^\perp$ be the subspace $\{R \subseteq T, R \in C^\perp\}$. It is well known (and easy to see) that $\dim C_T^\perp = |T| - r_C(T)$, where $r_C(\cdot)$ is the dimension of the subset of columns indexed by T in a generating matrix of C. Hence, for $x \in \{0,1\}^n$:

$$\mathbb{E}(f|T)(x) = \sum_{R \in C_T^\perp} w_R(x) = \begin{cases} |C_T^\perp| = 2^{|T| - r_C(T)} & \text{if} \quad x \in \left(C_T^\perp\right)^\perp \\ 0 & \text{otherwise} \end{cases}$$

And hence, $\|\mathbb{E}(f|T)\|_q^q = \mathbb{E}(\mathbb{E}(f|T)^q) = 2^{(q-1)(|T| - r_C(T))}$. Using this in (3) gives, for any $q > 1$ that

$$\log_2 \mathbb{E} f_\epsilon^q \leq (q-1) \mathbb{E}_{T \sim \lambda}(|T| - r_C(T)) = (q-1) \cdot \left(\lambda n - \mathbb{E}_{T \sim \lambda} r_C(T)\right),$$

where $\lambda = (1 - 2\epsilon)^{r(q)}$. Rearranging, we get the claim of the proposition. ∎

Proof or Lemma 1.4: Recall that for a function f on $\{0,1\}^n$ and for $0 \leq \epsilon \leq 1/2$ holds $\widehat{f_\epsilon}(R) = \widehat{f}(R) \cdot (1 - 2\epsilon)^{|R|}$. Let $(b_0, ..., b_n)$ be the distance distribution of $C^\perp$. Then, since $\epsilon(2) = \frac{1 - \lambda^{\ln 2}}{2}$, we have

$$2^{F(\lambda, 2)} = \sum_{R \in \{0,1\}^n} \widehat{f}^2(R) \cdot (1 - 2\epsilon)^{2|R|} = \sum_{i=0}^n b_i \lambda^{(2 \ln 2) \cdot i} = \sum_{i=0}^n b_i \theta^i.$$

On the other hand, it is easy to see that for $f = \frac{2^n}{|C|} \cdot 1_C$ and for any $0 \leq \epsilon \leq 1/2$ holds $\|f_\epsilon\|_\infty = \|\sum_{R \in \{0,1\}^n} \widehat{f}(R)(1 - 2\epsilon)^{|R|} w_R\|_\infty = \sum_{R \in \{0,1\}^n} \widehat{f}(R)(1 - 2\epsilon)^{|R|} = f_\epsilon(0)$ (since $\widehat{f}$ is nonnegative). Hence, recalling $\epsilon(\infty) = \frac{1 - \lambda^{2 \ln 2}}{2}$,

$$2^{F(\lambda, \infty)} = \|f_\epsilon\|_\infty = f_\epsilon(0) = \sum_{R \in \{0,1\}^n} \widehat{f}(R) \cdot (1 - 2\epsilon)^{|R|} = \sum_{i=0}^n b_i \lambda^{(2 \ln 2) \cdot i} = \sum_{i=0}^n b_i \theta^i.$$

Next, recall that for the function $g(x) = (1 - \theta)^{|x|}(1 + \theta)^{n - |x|}$ on $\{0,1\}^n$ holds $\widehat{g}(R) = \theta^{|R|}$. Hence, by Parseval's identity,

$$\sum_{i=0}^n b_i \theta^i = \sum_{R \in \{0,1\}^n} \widehat{f}(R)\widehat{g}(R) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x) = \frac{1}{|C|} \cdot \sum_{k=0}^n a_k (1 - \theta)^k (1 + \theta)^{n-k}.$$

∎

Proof or Proposition 1.6: Assume that $C^\perp$ achieves BEC capacity. Note that the rate of $C^\perp$ is $1 - R$ and that the dual of $C^\perp$ is C. By Corollary 1.5 and Lemma 1.4, there exists $\lambda = 1 - R + o(1)$ such that $\sum_{k=0}^n a_k \lambda^{(2 \ln 2) \cdot k} = 2^{o(n)}$. Hence, for all $0 \leq k \leq n/2$ holds

$$a_k \leq 2^{o(n)} \cdot \left(\frac{1}{\lambda}\right)^{(2 \ln 2) \cdot k} \leq 2^{o(n)} \cdot \left(\frac{1}{1 - R}\right)^{(2 \ln 2) \cdot k^*}.$$

Next, we claim that for any $0 \leq \alpha \leq 1$ holds $\sum_{k=0}^n a_k \alpha^{n-k} \leq \sum_{k=0}^n a_k \alpha^k$. To see this, recall that the Fourier transform on $\{0,1\}^n$ is its own inverse, up to normalization, and hence (see the proof of Lemma 1.4) we have

for any $\alpha \in \mathbb{R}$ that $\widehat{\alpha^{|x|}}(R) = \frac{1}{2^n}(1-\alpha)^{|R|}(1+\alpha)^{n-|R|}$. Let $f = \frac{2^n}{|C|} \cdot 1_C$. Then, for $0 < \alpha \le 1$, [2] we have, by Parseval's identity,

$$\sum_{k=0}^{n} a_k \alpha^{n-k} = \frac{|C|}{2^n} \cdot \alpha^n \sum_{x \in \{0,1\}^n} f(x)\left(\frac{1}{\alpha}\right)^{|x|} = \frac{|C|}{2^n} \cdot \sum_{R \in \{0,1\}^n} (-1)^{|R|}\widehat{f}(R)(1-\alpha)^{|R|}(1+\alpha)^{n-|R|} \le$$

$$\frac{|C|}{2^n} \cdot \sum_{R \in \{0,1\}^n} \widehat{f}(R)(1-\alpha)^{|R|}(1+\alpha)^{n-|R|} = \frac{|C|}{2^n} \cdot \sum_{x \in \{0,1\}^n} f(x)\alpha^{|x|} = \sum_{k=0}^{n} a_k \alpha^k.$$

For the inequality in the third step, note that $\widehat{f} = 1_{C^\perp}$ is nonnegative.

In particular, setting $\alpha = \lambda^{2\ln 2}$, we have for all $n/2 < k \le n$ that $a_k \le 2^{o(n)} \cdot \left(\frac{1}{\lambda}\right)^{(2\ln 2)\cdot(n-k)} \le 2^{o(n)} \cdot \left(\frac{1}{1-R}\right)^{(2\ln 2)\cdot k^*}$. This completes the proof of the first claim of the proposition.

Assume now that C achieves BEC capacity. Let $(b_0, ..., b_n)$ be the weight distribution of $C^\perp$. By Corollary 1.5 and Lemma 1.4, for $\theta = R^{2\ln 2}$ holds $\sum_{i=0}^{n} b_i \theta^i = 2^{o(n)}$. Clearly this also holds if we replace θ with any $0 \le \alpha \le \theta$. Using Parseval's identity as in the proof of Lemma 1.4, we have that $\sum_{k=0}^{n} a_k(1-\alpha)^k(1+\alpha)^{n-k} \le 2^{o(n)} \cdot |C|$, for any $0 \le \alpha \le \theta$.

Hence, for all $0 \le k \le n/2$ holds $a_k \le 2^{o(n)} \cdot \frac{|C|}{\max_{0 \le \alpha \le \theta}\{(1-\alpha)^k(1+\alpha)^{n-k}\}}$. Let $H(x) = x\log_2\left(\frac{1}{x}\right) + (1-x)\log_2\left(\frac{1}{1-x}\right)$ be the binary entropy function. It is easy to see that

$$\max_{0 \le \alpha \le \theta}\left\{(1-\alpha)^k(1+\alpha)^{n-k}\right\} = \begin{cases} (1-\theta)^k(1+\theta)^{n-k} & \text{if} \quad 0 \le k \le \frac{1-\theta}{2}\cdot n \\ 2^{(1-H(k/n))\cdot n} & \text{if} \quad \frac{1-\theta}{2}\cdot n \le k \le n/2 \end{cases}.$$

Recalling that, by Stirling's formula, for any $0 \le k \le n$ holds $2^{H(k/n)\cdot n} \le O\left(\sqrt{n}\right)\cdot\binom{n}{k}$, this proves the second claim of the proposition for $0 \le k \le n/2$.

To deal with the complementary case $n/2 < k \le n$, observe that $\sum_{k=0}^{n} a_k(1-\alpha)^k(1+\alpha)^{n-k} = (1+\alpha)^n \cdot \sum_{k=0}^{n} a_k\left(\frac{1-\alpha}{1+\alpha}\right)^k \ge (1+\alpha)^n \cdot \sum_{k=0}^{n} a_k\left(\frac{1-\alpha}{1+\alpha}\right)^{n-k} = \sum_{k=0}^{n} a_k(1-\alpha)^{n-k}(1+\alpha)^k$. ∎

Proof or Lemma 1.8: Let $\epsilon = \frac{1-\sqrt{p}}{2}$. Using (3) with $q = 2$ we have, for $f = \frac{2^n}{|C|} \cdot 1_C$, that

$$\mathbb{E}f_\epsilon^2 = \sum_{S \in \{0,1\}^n} \widehat{f}^2(S)(1-2\epsilon)^{2|S|} = \sum_{S \in \{0,1\}^n} \widehat{f}^2(S)p^{|S|} = \mathop{\mathbb{E}}_{S \sim p}\mathbb{E}\left(\mathbb{E}^2(f|S)\right) = \mathop{\mathbb{E}}_{S \sim p} 2^{|S|-r_C(S)}.$$

Hence the claim of the lemma follows from the inequality $\lambda n - \mathbb{E}_{T \sim \lambda} r_C(T) \ge F(\lambda, 2)$ with $\lambda = (1-2\epsilon)^{r(2)} = (1-2\epsilon)^{\frac{1}{\ln 2}} = p^{\frac{1}{2\ln 2}}$. ∎

Proof or Corollary 1.11: For $S \subseteq E$, the matroid rank of S is given by $r(S) = \sum_{i=1}^{c(S)}(|C_i|-1) = |V|-c(S)$, where $\{C_i\}$ are the connected components of (V, S). Hence, $|S| - r(S) = |S| + c(S) - |V|$. Substituting this into the claim of Lemma 1.8 gives the claim of this corollary. ∎

1) Comparing the bounds on the weight distribution of Reed-Muller codes: As far as we know, the best known bounds in the literature are due to [14] (Theorem 1.3). They state, in the notation of Proposition 1.6, that for the Reed-Muller code of rate R and for k linear in n we have

$$a_k \le 2^{CRk^*\left(2\log_2\left(\frac{n}{k^*}\right)+3\right)}, \tag{12}$$

where C is an absolute constant, $C \ge 30$. We proceed to compare (12) with the bounds in Proposition 1.6. We ignore the $2^{o(n)}$-factor in these bounds, since it is negligible for $0 < R < 1$ and k linear in n.

---

[2] Note that for $\alpha = 0$ the claim $\sum_{k=0}^{n} a_k\alpha^{n-k} \le \sum_{k=0}^{n} a_k\alpha^k$ is trivial, since $a_0 = 1 \ge a_n$.

Replacing the term $\log_2\left(\frac{n}{k^*}\right)$ in (12) by 1 (which is a smaller quantity, since $k^* \leq n/2$), it is easy to see [11] that (12) is larger than the first bound of Proposition 1.6 unless $R \geq 1 - \delta$, for a very small δ (specifically, $\delta < 2^{-100}$). Assume that $R = 1 - \epsilon$, where ϵ is very small. Consider the second bound of Proposition 1.6. If $k^* \geq \frac{1 - R^{2\ln 2}}{2} \cdot n$, then this bound is at most $\binom{n}{k^*}$, and it is easy to see that it is smaller than (12). For $0 < k^* \leq \frac{1 - R^{2\ln 2}}{2} \cdot n$, the term $\log_2\left(\frac{n}{k^*}\right)$ in (12) behaves like $\log_2\left(\frac{1}{\epsilon}\right)$, and hence (12) behaves like $\left(\frac{1}{\epsilon}\right)^{C_1 k^*}$, for $C_1 \geq 60$, and this is larger than the first bound in Proposition 1.6, which is $\left(\frac{1}{\epsilon}\right)^{(2\ln 2)\cdot k^*}$. ∎

ACKNOWLEDGMENTS

REFERENCES

[1] E. Abbe, A. Shpilka, and A. Wigderson, *Reed-Muller Codes for Random Erasures and Errors*, IEEE Trans. Information Theory 61(10): 5229-5252, (2015).
[2] W. Beckner, *Inequalities in Fourier Analysis*, Annals of Math., 102(1975), pp. 159-182.
[3] A. Bonami, *Etude des coefficients Fourier des fonctions de Lp(G)*, Annales de lInstitut Fourier, 20(2) (1970), 335-402.
[4] S. Boucheron, G. Lugosi, and P. Massart, **Concentration inequalities**, Oxford University Press, 2013.
[5] T. H. Brylawski and J. G. Oxley, *The Tutte polynomial and its applications*, in Matroid Applications, Encyclopedia Math. Appl. 40, N. White, ed., Cambridge University Press, 1992.
[6] T. Cover and J. Thomas, **Elements of Information Theory**, Wiley 2006.
[7] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. of Math., 97, 1975, pp. 1061-1083.
[8] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. L. Urbanke, *Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels*, ISIT 2016.
[9] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Sasoglu and R. L. Urbanke, *Reed-Muller Codes Achieve Capacity on Erasure Channels*, IEEE Trans. Information Theory, 63(7), 4298-4316 (2017).
[10] R. O'Donnell, **Analysis of Boolean functions**, Cambridge University Press, 2014.
[11] O. Ordentlich, Y. Polyanskiy, and R. Urbanke, personal communication, 2016.
[12] A. Renyi, *On the measures of entropy and information*, in Proc. 4th Berkeley Symp. on Math. Statist. Probability, (Berkeley CA 1961), vol. 1, 547-561.
[13] A. Samorodnitsky, *On the entropy of a noisy function*, IEEE Trans. Information Theory 62(10): 5446-5464 (2016).
[14] A Shplika and O. Sberlo, *Weight Distribution of Reed-Muller Codes with Linear Degrees*, manuscript 2018. See also O. Sberlo, MSc thesis, WIS 2018.
[15] A. D. Wyner and J. Ziv, *A theorem on the entropy of certain binary sequences and applications: Part I*, IEEE Trans. Inform. Theory, vol. 19, no. 6, pp. 769-772, 1973.

Alex Samorodnitsky received the B.A. degree in computer science and mathematics (1987), the M.Sc. degree in mathematics (1990), and the Ph.D. degree in mathematics (1998), all from the Hebrew University of Jerusalem.

He is now a Professor of Computer Science at the School of Engineering and Computer Science in the Hebrew University, which he joined in 2001 as an Assistant Professor. His research interests are in theoretical computer science, coding theory, and combinatorics.