An extension of the unified Brascamp-Lieb and the Entropy Power Inequality to finite Abelian groups

Chin Wa (Ken) Lau and Chandra Nair
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong (China)
Email:{kenlau,chandra}@ie.cuhk.edu.hk

Abstract

The doubling-followed-by-rotation trick to prove the extremality of Gaussian distributions has become a valuable tool in information theory. In particular, the above trick has been used to establish the Gaussian extremality of a family of inequalities that unifies the Entropy Power Inequality and the Brascamp-Lieb inequalities. Here, we develop a technique (similar to the one in the continuous case) to prove the extremality of Haar distributions for a similar family of inequalities in finite Abelian groups.

I. Introduction

A. Background

The Entropy Power Inequality (EPI) is a powerful tool that has found widespread applications in network information theory. It has been widely used to show the capacity region (for instance [1], [2]) in several multiuser information theory settings. Furthermore, various versions of this inequality have been formulated for discrete random variables. Shamai and Wyner, [3], established a discrete analog of EPI for binary-valued random variables. Harremoës and Vignaet, [4], discovered a discrete analog of EPI for a particular family of binomial random variables. Sharma, Das, and Muthukrishnan, [5] based on the work of [4], established another version of the discrete EPI. On the other hand, there have been several attempts to generalization Mrs. Gerber's Lemma by Wyner and Ziv [6]; for example, Jog and Anantharam have shown a generalization of Mrs. Gerber's Lemma for random variables on an Abelian group with order 2^n [7].

Definition 1 (Entropy Power)**.** Suppose X is an $\mathbb{R}^n$-valued random variable with a well-defined differential entropy $h(X)$. The entropy power of X is defined as

$$\mathcal{N}(X) = \frac{1}{2\pi e} e^{2h(X)/n}.$$

Theorem 1 (EPI [8], [9])**.** *Suppose X and Y are independent $\mathbb{R}^n$-valued random variables. Then*

$$\mathcal{N}(X) + \mathcal{N}(Y) \leq \mathcal{N}(X+Y),$$

and the equality holds if and only if X and Y are Gaussians with proportional covariance matrices.

An equivalent dimension-independent form of the EPI was formulated by Lieb [10].

Theorem 2. *Suppose X and Y are independent $\mathbb{R}^n$-valued random variables. For any $\lambda \in [0,1]$, we have*

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y) \geq 0,$$

where the equality holds if and only if X and Y are Gaussians with identical covariance matrices.

In other words, the functional defined by

$$f(X,Y) := h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y),$$

where X and Y are independent random variables, is minimized by Gaussians with identical covariance matrices.

Similarly, the Brascamp-Lieb inequality (BLI) [11] is a family of functional inequalities at the intersection of information and functional inequalities. Special cases of the BLI include Hölder's inequality, the Loomis-Whitney inequality, and sharp forms of Young's convolution inequalities [12]. One of the central results here is that the optimal constants can be computed by restricting to Gaussian distributions. Recently, in [13], the following theorem was proved that unified the family of Brascamp-Lieb inequalities and the Entropy-Power inequality.

A conference version of this result will be presented at the 2024 IEEE International Symposium on Information Theory.

Theorem 3 (Unified EPI and BLI, [13])**.** *Let* $(\mathbf{A},\mathbf{c},\mathbf{r},\mathbf{d})$ *be a BL-EPI datum. Define*

$$M_g := \sup_{Z\in\mathcal{P}_g(\mathbf{r})} \sum_{i=1}^{k} d_i h(Z_i) - \sum_{j=1}^{m} c_j h(A_j Z). \tag{1}$$

Then for any $X \in \mathcal{P}(\mathbf{r})$*, the following inequality holds:*

$$\sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X) \le M_g. \tag{2}$$

Remark 1. The readers are encouraged to look at [13] for a precise definition of BL-EPI datum and $\mathcal{P}(\mathbf{r})$. The main point is that $\mathcal{P}_g(\mathbf{r})$ restricts the distributions in $\mathcal{P}(\mathbf{r})$ to Gaussian distributions, and Z's are Gaussian random variables. For this paper, it suffices to note that $\{d_i\}$ and $\{c_j\}$ are positive constants, and $X_1, X_2, \dots, X_k$ are mutually independent random vectors. It is also worth noting that the same proof (of Gaussian extremality) goes through if one imposes covariance constraints, $\mathbb{E}[X_i X_i^T] \preceq K_i$, on the independent random vectors.

The above inequality was proved using the doubling and rotation idea ([14]), a technique that differs from the previous proof methods of the EPI. Our main result (Theorem 6) is a discrete analog (in finite Abelian groups) of Theorem 3. Further, we demonstrate that the proof technique in [13] can be essentially mimicked (modulo some differences in the technical arguments) in this setting. The proof technique in [13] can be summarized (pushing some technical conditions under the carpet) as follows: Gaussian optimality was deduced by demonstrating the sub-additivity of an entropic functional and by further using the proof of the sub-additivity to show that rotated forms of the optimizers are independent. This implied that the optimizers must be Gaussian by applying the Darmois-Skitovich theorem.

Theorem 4 (Darmois-Skitovich theorem [15], [16])**.** *Let* $X_1, \dots, X_n$ *be independent random variables. Let* $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ *be non-zero constants for each coordinate. If the linear statistics* $L_1 = \sum_{i=1}^n \alpha_i X_i$ *and* $L_2 = \sum_{i=1}^n \beta_i X_i$ *are independent, then all random variables* $X_1, \dots, X_n$ *are Gaussians.*

The following finite Abelian group analog of the above result was discovered by Feldman [17].

Theorem 5 (Feldman [17])**.** *Let* $\mathbb{G}$ *be a finite Abelian group, and* X_1, X_2 *be independent random variables with values in* $\mathbb{G}$*. Let* $\alpha_1, \alpha_2, \beta_1, \beta_2$ *be automorphisms of the group* $\mathbb{G}$*. Then if the linear statistics* $L_1 = \alpha_1(X_1)+\alpha_2(X_2)$ *and* $L_2 = \beta_1(X_1)+\beta_2(X_2)$ *are independent, then* X_1 *and* X_2 *are shifts of a Haar distribution of some subgroup* $\mathbb{H}$ *of* $\mathbb{G}$*, or equivalently,* X_1 *and* X_2 *are uniform distributions on a coset of some subgroup* $\mathbb{H}$ *of the group* $\mathbb{G}$*.*

Therefore, it is natural to guess that Gaussians can be replaced by uniform distributions on a coset of some subgroup (or shifts of Haar distributions) when working in finite Abelian groups. However, while this intuition is correct, we show a way to overcome some technical hassles (different from the continuous case) in our proof. Furthermore, just like the rotation trick in the continuous case, we believe this argument can find several other applications to establish the optimality of Haar distributions.

Notation: We use $(\mathbb{G},+)$ or $\mathbb{G}$ to denote a finite Abelian group. We use $|A|$ to denote the cardinality of a finite set A and $\mathrm{supp}(p_X)$ to denote the support of p_X. For an additive group $\mathbb{G}$ and random variable X taking value in $\mathbb{G}$, we define the following: $cX := X + \cdots + X$ (c times) for $c \in \mathbb{Z}_+$. If $c = 0$, then cX is the identity element; and when $c \in \mathbb{Z}_-$, then cX is the additive inverse of $|c|X$.

II. Main

Theorem 6. *Let* $X_1, \dots, X_n$ *be independent random variables taking values in subgroups* $\mathbb{H}_1, \dots, \mathbb{H}_n$ *of a finite Abelian group* $\mathbb{G}$*, respectively. Let* $a_1, \dots, a_n$*, and* $b_1, \dots, b_\ell$ *be positive constants, and* $c_{i,j}^{(1)}, \dots, c_{i,j}^{(m_j)}$ *be integers. Then, the following optimization problem*

$$\max_{\prod_{i=1}^n p_{X_i}} \sum_{i=1}^n a_i H(X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \dots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i\right),$$

has an optimizer $(X_1^*, \dots, X_n^*)$*, where* X_i^* *is uniformly distributed on a coset of a subgroup* $\mathbb{K}_i \subseteq \mathbb{H}_i$*.*

Remark 2. The following points are worth noting:

1) One can relax the assumption on the sign of a_i. Note that, if any $a_k \le 0$, it is optimal to set the corresponding X_k to be a constant random variable. To see this, observe that

$$H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \dots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i,\right) \ge H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \dots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | X_k\right).$$

2) Unlike the continuous case, where Lieb's formulation of EPI was known, the extremality of the uniform distribution on a coset of some subgroup for $a_1 H(X_1) + a_2 H(X_2) - H(X_1 + X_2)$ was not known. There have been conjectures (and some results), [7], of a similar flavor.

We establish the following lemma before providing proof of Theorem 6. This is the analogous result to the Darmois-Skitovich theorem needed in our proof. We are unable to use Theorem 5 directly in our arguments. This lemma should also be of independent interest.

Lemma 1. *Let X_A and X_B be two independent random variables taking values in some finite Abelian group $\mathbb{H}$. Let $S = \mathrm{supp}(p_{X_B})$ and $\mathbb{D}$ denote the subgroup generated by pairwise differences of the elements of S. For $X_A + X_B$ to be independent of X_B, it is necessary and sufficient that $\mathrm{P}(X_A = h_1) = \mathrm{P}(X_A = h_2)$ whenever h_1, h_2 belong to the same coset of $\mathbb{D}$ (in other words, p_{X_A} is uniformly distributed conditioned on it taking values in a given coset of $\mathbb{D}$). Consequently $|\mathrm{supp}(p_{X_A})| = k|\mathbb{D}| \geq k|\mathrm{supp}(p_{X_B})|$ for some $k \in \mathbb{N}$ satisfying $1 \leq k \leq \frac{|\mathbb{H}|}{|\mathbb{D}|}$, and $k = 1$ only if X_A is uniformly distributed on a coset of $\mathbb{D}$.*

Proof. First, assume that X_A is uniform on the cosets of $\mathbb{D}$. Let T be a transversal for $\mathbb{D}$ in $\mathbb{H}$. Therefore, any element $h \in \mathbb{H}$ can be uniquely represented as $h = t + d$, for some $t \in T$ and $d \in \mathbb{D}$. If X_A is uniform on the cosets of $\mathbb{D}$, then $\mathrm{P}(X_A = h) = \mathrm{P}(X_A = t + d) = \frac{1}{|\mathbb{D}|}\mathrm{P}_T(T = t)$ for some arbitrary distribution, P_T, on the transversal. If X_A and X_B are independent, note that $\mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\mathrm{P}(X_B = b)$.

On the other hand $\mathrm{P}(X_A + X_B = h + b) = \sum_{\hat{b} \in S} \mathrm{P}(X_A = h + b - \hat{b})\mathrm{P}(X_B = \hat{b})$. Since $b - \hat{b} \in D$, $h + b - \hat{b}$ belongs to the same coset as h. Therefore, for all $\hat{b}$, we have $\mathrm{P}(X_A = h + b - \hat{b}) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)$. Consequently, $\mathrm{P}(X_A + X_B = h + b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\sum_{\hat{b} \in S}\mathrm{P}(X_B = \hat{b}) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)$. Therefore $\mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\mathrm{P}(X_B = b) = \mathrm{P}(X_A + X_B = h + b)\mathrm{P}(X_B = b)$. This implies that $X_A + X_B$ is independent of X_B, as desired.

Conversely, let us assume that X_A and X_B are independent, and additionally, $X_A + X_B$ is also independent of X_B. Therefore $\mathrm{P}(X_A + X_B = h + b)\mathrm{P}(X_B = b) = \mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b)$. This implies that for all $b \in S$, we have $\mathrm{P}(X_A = h) = \mathrm{P}(X_A + X_B = h + b) = \sum_{\hat{b} \in S}\mathrm{P}(X_A = h + b - \hat{b})\mathrm{P}(X_B = \hat{b})$. Rewriting h as $h - b$, we see that $\mathrm{P}(X_A = h - b) = \sum_{\hat{b} \in S}\mathrm{P}(X_A = h - \hat{b})\mathrm{P}(X_B = \hat{b})$. Since the right-hand-side does not depend on b, we obtain that $\mathrm{P}(X_A = h - b_1) = \mathrm{P}(X_A = h - b_2)$, for all $b_1, b_2 \in S$ and $h \in \mathbb{H}$. Replacing $h - b_1$ by h, we note that $\mathrm{P}(X_A = h) = \mathrm{P}(X_A = h + b_1 - b_2)$. The pairwise differences $b_i - b_j$ generate $\mathbb{D}$, and from above p_{X_A} is invariant under a shift by a pairwise difference, it follows that p_{X_A} is invariant under a shift by an element in $\mathbb{D}$. In other words, X_A is uniform on the cosets of $\mathbb{D}$.

Finally note that $|\mathrm{supp}(p_{X_A})| = |\mathrm{supp}(p_T)||\mathbb{D}|$, and $|\mathrm{supp}(p_{X_A})| = |\mathbb{D}|$ only if T is a constant random variable, implying that X_A is uniform on a coset of $\mathbb{D}$. We also have that $|\mathbb{D}| \geq |\mathrm{supp}(p_{X_B})|$, since $b \mapsto b - b_0$ is an injection from $\mathrm{supp}(p_{X_B})$ to $\mathbb{D}$, where b_0 is an arbitrary fixed element from $\mathrm{supp}(p_{X_B})$. □

Remark 3. The proof is similar to that in [18, Section 5]. In [18], X_A and X_B are assumed to be identically distributed.

A. Proof of Theorem 6

The first step in proving the optimality of the uniform distribution of a coset of some subgroup is to identify a sub-additive functional. To this end, given an n-tuple of distributions $(p_{X_1}, \ldots, p_{X_n})$, such that X_i has support on $\mathbb{H}_i$, let us define:

$$F(X_1, \ldots, X_n) := \sup_{\substack{p_{U|X_1,\ldots,X_n}: \\ p_{X_1,\ldots,X_n|U} \\ = \prod_{i=1}^n p_{X_i|U}}} \sum_{i=1}^n a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right).$$

Observe that the maximum value of $F(X_1, \ldots, X_n)$ is the same as the value of the optimization problem in Theorem 6, as the average is always less than the maximum (the other direction is immediate by taking $X_1, \ldots, X_n$ to be mutually independent and U to be a constant).

Remark 4. This is essentially the same function as the one employed in [13].

Now consider an n-tuple of distributions $(p_{X_1,\hat{X}_1}, \ldots, p_{X_n,\hat{X}_n})$, such that $(X_i, \hat{X}_i)$ has support on $\mathbb{H}_i \times \mathbb{H}_i$, let us define (ignoring the abuse of notation):

$$F((X_1, \hat{X}_1), \ldots, (X_n, \hat{X}_n)) := \sup_{\substack{p_{U|(X_1,\hat{X}_1),\ldots,(X_n,\hat{X}_n)}: \\ p_{(X_1,\hat{X}_1),\ldots,(X_n,\hat{X}_n)|U} = \\ \prod_{i=1}^n p_{(X_i,\hat{X}_i)|U}}} \sum_{i=1}^n a_i H(X_i, \hat{X}_i|U)$$
$$- \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i, \sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i | U\right).$$

Observe that

$$
\begin{aligned}
&\sum_{i=1}^n a_i H(X_i,\hat{X}_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i, \sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i | U\right)\\
&= \sum_{i=1}^n a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right)\\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_i|U, X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i | U, \sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i\right)\\
&\overset{(a)}{=} \sum_{i=1}^n a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right)\\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_i|U, \mathbf{X}) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i | U, \sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i\right)\\
&\overset{(b)}{\leq} \sum_{i=1}^n a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right)\\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_i|U, \mathbf{X}) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i | U, \mathbf{X}\right)\\
&\overset{(c)}{\leq} F(X_1, \ldots, X_n) + F(\hat{X}_1, \ldots, \hat{X}_n).
\end{aligned}
$$

In the above $\mathbf{X} = (X_1, \ldots, X_n)$. Equality (a) follows, as conditioned on U, $\{(X_i, \hat{X}_i)\}$ are mutually independent and equality (b) follows from data-processing inequality as

$$
(U, \sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i) \to (U, \mathbf{X}) \to (U, \sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i)
$$

is Markov. Finally inequality (c) follows since conditioned on U, the random variables $\{X_i\}$ are mutually independent, and conditioned on $(U, \mathbf{X})$, the random variables $\{\hat{X}_i\}$ are mutually independent.

In the next part of the proof, we will argue that certain linear forms of the maximizer are independent. To this end, consider the two maximization problems listed below:

$$
\max_{\prod_{i=1}^n p_{X_i}} \sum_{i=1}^n a_i H(X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i\right), \qquad \text{(Optimization problem 1)}
$$

$$
\max_{\prod_{i=1}^n p_{\hat{X}_i}} \sum_{i=1}^n a_i H(\hat{X}_i) - \sum_{i=1}^n \epsilon H(\hat{X}_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_i\right). \qquad \text{(Optimization problem 2)}
$$

In the above two problems, the random variables X_i and $\hat{X}_i$ are assumed to take values in the subgroup $\mathbb{H}_i$. Let $(X_1^*, \ldots, X_n^*)$ and $(\hat{X}_{1,\epsilon}^*, \ldots, \hat{X}_{n,\epsilon}^*)$ be maximizers of the two optimization problems respectively and V, V_ϵ be the maximum value attained by the two optimization problems. Further, let us assume that among all possible maximizers of the first problem, $(X_1^*, \ldots, X_n^*)$ minimizes the function $\prod_{i=1}^n (1 + |\mathrm{supp}(p_{X_i})|)$.

It is immediate that $V_\epsilon \to V$ and $\epsilon \to 0$ (as the difference between the objective functions at any point is bounded by $\epsilon\left(\sum_{i=1}^n \log|\mathbb{H}_i|\right)$. Furthermore, by the compactness of the probability simplex and continuity of the function, we know that there is a sequence of maximizers $(\hat{X}_{1,\epsilon_m}^*, \ldots, \hat{X}_{n,\epsilon_m}^*)$ that converge to a maximizer of the first optimization problem.

Finally, we define

$$
F_\epsilon(X_1, \ldots, X_n) := \sup_{\substack{p_{U|X_1,\ldots,X_n}:\\ p_{X_1,\ldots,X_n|U} = \prod_{i=1}^n p_{X_i|U}}} \sum_{i=1}^n a_i H(X_i|U) - \sum_{i=1}^n \epsilon H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right).
$$

We have $F_\epsilon(X_1, \ldots, X_n) \leq V_\epsilon$.

Observe that by taking independent copies of the maximizers $(X_1^*, \ldots, X_n^*)$ and $(\hat{X}_{1,\epsilon}^*, \ldots, \hat{X}_{n,\epsilon}^*)$, we obtain

$$
\begin{aligned}
&V + V_\epsilon\\
&= \sum_{i=1}^n a_i H(X_i^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i^*\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{n} a_i H(\hat{X}_{i,\epsilon}^*) - \sum_{i=1}^{n} \epsilon H(\hat{X}_{i,\epsilon}^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) \\
&\overset{(a)}{=} \sum_{i=1}^{n} a_i H(X_i^*, \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^{n} \epsilon H(\hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i^*, \sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) \\
&\overset{(b)}{=} \sum_{i=1}^{n} a_i H(X_i^* + \hat{X}_{i,\epsilon}^*, \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^{n} \epsilon H(\hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*), \sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) \\
&= \sum_{i=1}^{n} a_i H(X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\quad + \sum_{i=1}^{n} a_i H(\hat{X}_{i,\epsilon}^* | X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^* | \sum_{i=1}^{n} c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\quad - \sum_{i=1}^{n} \epsilon H(\hat{X}_{i,\epsilon}^* | X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^{n} \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} a_i H(X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\quad + \sum_{i=1}^{n} a_i H(\hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*\right) \\
&\quad - \sum_{i=1}^{n} \epsilon H(\hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*) - \sum_{i=1}^{n} \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\overset{(d)}{\leq} F(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*) + F_\epsilon(\hat{X}_{1,\epsilon}^*, \ldots, \hat{X}_{n,\epsilon}^*) - \sum_{i=1}^{n} \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\overset{(e)}{\leq} V + V_\epsilon - \sum_{i=1}^{n} \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*).
\end{aligned}
$$

Here $\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*$ stands for the vector $(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*)$. In the above, equality (a) follows from the independence of $\mathbf{X}^*$ and $\hat{\mathbf{X}}_\epsilon^*$ and equality (b) follows from $H(X_1, X_2) = H(X_1 + X_2, X_2)$. Equality (c) follows from data-processing and the independence of the components of $(\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*)$, and (d) follows from the definition of F and F_ϵ as elaborated next. Note that $(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*)$ satisfies the support constraints and is a valid input for the function F (with U taken to be a constant). Now take $U = \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*$ and use independence of the components of $(\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*)$ to justify that this choice is a valid extension $p_{U|\hat{\mathbf{X}}_\epsilon^*}$ in the definition of F_ϵ. Finally, we note that the maximum of F and F_ϵ are V and V_ϵ to justify the inequality (e).

For $\epsilon > 0$, note that the above manipulations imply that $I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) = 0$ using the non-negativity of mutual information, or in other words, that $X_i^* + \hat{X}_{i,\epsilon}^*$ is independent of $\hat{X}_{i,\epsilon}^*$. Since X_i^* was independent of $\hat{X}_{i,\epsilon}^*$ by construction, note that we can apply Lemma 1 to deduce that the distribution of X_i^* is uniform on the cosets of $\mathbb{D}_{i,\epsilon}$. Here $\mathbb{D}_{i,\epsilon}$ is the subgroup of $\mathbb{H}_i$ generated by the pairwise differences of the support of $\hat{X}_{i,\epsilon}^*$. Further $|\text{supp}(p_{X_i^*})| = k_{i,\epsilon}|\mathbb{D}_{i,\epsilon}|$ for some $k_{i,\epsilon} \in \mathbb{N}$ satisfying $1 \leq k_{i,\epsilon} \leq \frac{|\mathbb{H}_i|}{|\mathbb{D}_{i,\epsilon}|}$.

As argued earlier, we have a sequence of optimizers $\hat{\mathbf{X}}_{\epsilon_m}^*$ such that as $\epsilon_m \downarrow 0$ and $\hat{\mathbf{X}}_{\epsilon_m}^*$ converges to a maximizer, say $\tilde{\mathbf{X}}^*$, of the problem with $\epsilon = 0$. Now, we have for any $\epsilon > 0$,

$$
\begin{aligned}
\prod_{i=1}^{n}(1 + k_{i,\epsilon}|\mathbb{D}_{i,\epsilon}|) = \prod_{i=1}^{n}(1 + |\text{supp}(p_{X_i^*})|) &\overset{(a)}{\leq} \prod_{i=1}^{n}(1 + |\text{supp}(p_{\tilde{X}_i^*})|) \\
&\overset{(b)}{\leq} \limsup_{m\to\infty} \prod_{i=1}^{n}(1 + |\text{supp}(p_{\hat{X}_{i,\epsilon_m}^*})|) \overset{(c)}{\leq} \limsup_{m\to\infty} \prod_{i=1}^{n}(1 + |\mathbb{D}_{i,\epsilon_m}|).
\end{aligned}
$$

Here (a) follows from the minimality of the choice of X^*, i.e. $\mathbf{X}^*$ minimizes $\prod_{i=1}^n (1+|\mathrm{supp}(p_{X_i})|)$ among all the maximizers of the optimization problem. Inequality (b) follows from the observation that as $\hat{X}^*_{i,\epsilon_m}$ converges (weakly) to $\tilde{X}^*_i$, $\mathrm{supp}(p_{\tilde{X}^*_i}) \subseteq \mathrm{supp}(p_{\hat{X}^*_{i,\epsilon_m}})$ for sufficiently large m. Finally $|\mathbb{D}_{i,\epsilon_m}| \geq |\mathrm{supp}(p_{\hat{X}^*_{i,\epsilon_m}})|$, as argued earlier, since $b \mapsto b - b_0$ is an injection from $\mathrm{supp}(p_{\hat{X}^*_{i,\epsilon_m}})$ to $\mathbb{D}$, where b_0 is an arbitrary fixed element from $\mathrm{supp}(p_{\hat{X}^*_{i,\epsilon_m}})$. Therefore for some large, enough m, have $k_{i,\epsilon_m} = 1$ for all i, where $1 \leq i \leq m$. Therefore, again invoking Lemma 1, we see that X^*_i is uniformly distributed on some coset of a subgroup $\mathbb{D}_{i,\epsilon_m} \subseteq \mathbb{H}_i$. This completes the proof of Theorem 6.

Remark 5. Note that the above argument also establishes some properties of the maximizers of the optimization problem. Suppose $\mathbf{X}^*_a$ is another maximizer such that $\prod_{i=1}^n (1 + |\mathrm{supp}(p_{X^*_{a,i}})|) > \prod_{i=1}^n (1 + |\mathrm{supp}(p_{X^*_i})|)$. Then, the above argument implies that one cannot have a sequence of maximizers of the perturbed problem that converges to $\mathbf{X}^*_a$.

III. Future and Related work

An extension to torsion-free groups is certainly interesting. Along these lines, Tao proposed a conjecture [18] on the discrete analog of EPI as below

Conjecture 1. *Suppose $X_1, \ldots X_{n+1}$ are identically distributed and independent random variables on some torsion-free group $\mathbb{T}$. Then, for any $\epsilon > 0$, as long as $H(X)$ is sufficiently large (depending on n, ϵ), we have*

$$H(X_1 + \cdots + X_{n+1}) \geq H(X_1 + \ldots X_n) + \frac{1}{2} \log \frac{n+1}{n} - \epsilon.$$

There is a recent proof of this conjecture, under the assumption that the distribution of X is log-concave, by Gavalakis [19].

Optimality of uniform distribution in a discrete setting for an information functional has recently been established by Gowers, Green, Manners, and Tao [20]. There, they consider the following functional on a finite Abelian group, $\mathbb{G}$, with characteristic 2.

Definition 2 (Polynomial Freiman-Ruzsa Functional)**.** [20, Equation 2.1] For any random variables X^0, Y^0 with support contained inside $\mathbb{G}$, a finite Abelian group with characteristic 2, define the functional

$$\begin{aligned}\tau(X,Y) := &\left(H(X-Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y)\right) + \eta\left(H(X-X^0) - \frac{1}{2}H(X) - \frac{1}{2}H(X^0)\right)\\ &+ \eta\left(H(Y-Y^0) - \frac{1}{2}H(Y) - \frac{1}{2}H(Y^0)\right),\end{aligned}$$

where X, Y, X^0, Y^0 are mutually independent. Here X, Y also take values in $\mathbb{G}$.

It was shown in [20, Proposition 2.1] that all minimizers of $\tau(X,Y)$ must be uniform distributions on a coset of a subgroup for all X^0, Y^0 with support in $\mathbb{G}$, when $\eta \leq \frac{1}{9}$.

Let us consider a slight modification of the above functional and define for a pair of distributions p_X, p_Y

$$\begin{aligned}T(X,Y) := \min_{\substack{p_{U|XY}:\\ p_{XY|U} = p_{X|U}p_{Y|U}}} &H(X-Y|U) - \frac{1}{2}H(X|U) - \frac{1}{2}H(Y|U) + \eta\left(H(X-X^0|U) - \frac{1}{2}H(X|U) - \frac{1}{2}H(X^0|U)\right)\\ &+ \eta\left(H(Y-Y^0|U) - \frac{1}{2}H(Y|U) - \frac{1}{2}H(Y^0|U)\right),\end{aligned}$$

where the triple (U,X,Y), X^0, and Y^0 are mutually independent.

Define the two-letter form

$$\begin{aligned}T((X_a,X_b),(Y_a,Y_b)) := \min_{\substack{p_{U|X_a,X_bY_aY_b}:\\ p_{X_aX_bY_aY_b|U} = p_{X_aX_b|U}p_{Y_a,Y_b|U}}} &H(X_a-Y_a, X_b-Y_b|U) - \frac{1}{2}H(X_a,X_b|U) - \frac{1}{2}H(Y_a,Y_b|U)\\ &+ \eta\left(H(X_a-X^0_a, X_b-X^0_b|U) - \frac{1}{2}H(X_a,X_b|U) - \frac{1}{2}H(X^0_a,X^0_b|U)\right)\\ &+ \eta\left(H(Y_a-Y^0_a, Y_b-Y^0_b|U) - \frac{1}{2}H(Y_a,Y_b|U) - \frac{1}{2}H(Y^0_a,Y^0_b|U)\right),\end{aligned}$$

where the tuple $(U,(X_a,X_b),(Y_a,Y_b))$, X^0_a, X^0_b, Y^0_a, and Y^0_b are mutually independent.

Lemma 2. *For any $\eta \geq 0$, following super-additivity inequality holds:*

$$T((X_a,X_b),(Y_a,Y_b)) \geq T(X_a,Y_a) + T(X_b,Y_b)$$

Proof. Observe that the following holds:

$$H(X_a-Y_a, X_b-Y_b|U) - \frac{1}{2}H(X_a,X_b|U) - \frac{1}{2}H(Y_aY_b|U)$$

$$
\begin{aligned}
&= H(X_a - Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b - Y_b|U, X_a - Y_a) - \frac{1}{2}H(X_b|U, X_a) - \frac{1}{2}H(Y_b|U, Y_a) \\
&\overset{a}{=} H(X_a - Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b - Y_b|U, X_a - Y_a) - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) \\
&\quad - \frac{1}{2}H(Y_b|U, X_a, Y_a, X_a^0, Y_a^0) \\
&\geq H(X_a - Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b - Y_b|U, X_a, Y_a, X_a^0, Y_a^0)) - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) \\
&\quad - \frac{1}{2}H(Y_b|U, X_a, Y_a, X_a^0, Y_a^0).
\end{aligned}
$$

Here (a) follows from the independence and the Markov structure of the random variables.

In an identical fashion, we can also show that

$$
\begin{aligned}
&H(X_a - X_a^0, X_b - X_b^0|U) - \frac{1}{2}H(X_a, X_b|U) - \frac{1}{2}H(X_a^0, X_b^0|U) \\
&\geq H(X_a - X_a^0|U) - \frac{1}{2}H(X_a|U) - H(X_a^0|U) + H(X_b - X_b^0|U, X_a, Y_a, X_a^0, Y_a^0)) - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) \\
&\quad - \frac{1}{2}H(X_b^0|U, X_a, Y_a, X_a^0, Y_a^0),
\end{aligned}
$$

and

$$
\begin{aligned}
&H(Y_a - Y_a^0, Y_b - Y_b^0|U) - \frac{1}{2}H(Y_a, Y_b|U) - \frac{1}{2}H(Y_a^0, Y_b^0|U) \\
&\geq H(Y_a - Y_a^0|U) - \frac{1}{2}H(Y_a|U) - H(Y_a^0|U) + H(Y_b - Y_b^0|U, X_a, Y_a, X_a^0, Y_a^0)) - \frac{1}{2}H(Y_b|U, X_a, Y_a, X_a^0, Y_a^0) \\
&\quad - \frac{1}{2}H(Y_b^0|U, X_a, Y_a, X_a^0, Y_a^0).
\end{aligned}
$$

Denote $U_a = U$, and observe that $p_{X_aY_a|U_a} = p_{X_a|U_a}p_{Y_a|U_a}$ and $(U_a, X_a, Y_a), X_a^0$, and Y_a^0 are mutually independent. Denote $U_b = (U, U, X_a, Y_a, X_a^0, Y_a^0)$, and observe that $p_{X_bY_b|U_b} = p_{X_b|U_b}p_{Y_b|U_b}$ and $(U_b, X_b, Y_b), X_a^0$, and Y_a^0 are mutually independent. Putting the above inequalities together, the requisite super-additivity follows. □

However, we cannot do the transformation $(X_a - X_a^0, X_b - X_b^0) \mapsto (X_a - X_a^0 + X_b - X_b^0, X_b - X_b^0)$ as this would replace X_a^0 by $X_a^0 + X_b^0$. This is not permitted as X_a^0 is a fixed distribution. However, one can place X_a, X_b, Y_A, Y_b at the minimizer by alternate linear forms and use the minimality to force an independence of some linear forms. For $\eta \leq \frac{1}{9}$ (and when the field has characteristic two), the authors can deduce such independence of the linear forms and prove the optimality of the uniform distribution.

Remark 6. The argument presented above is slightly different from what is presented by the authors, but the difference is mainly superficial.

Acknowledgement

This work was supported by the following grants from the University Grants Committee of the Hong Kong Special Administrative Region: GRF14210120, GRF14221822.

References

[1] P. F. Bergmans, "Coding theorem for broadcast channels with degraded components," *IEEE Trans. Info. Theory*, vol. IT-15, pp. 197–207, 3 March, 1973.

[2] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept 2006.

[3] S. Shamai and A. Wyner, "A binary analog to the entropy-power inequality," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1428–1430, 1990.

[4] P. Harremoës and C. VIGNAT, "An entropy power inequality for the binomial family," *JIPAM. Journal of Inequalities in Pure & Applied Mathematics [electronic only]*, vol. 4, 01 2003.

[5] N. Sharma, S. Das, and S. Muthukrishnan, "Entropy power inequality for a family of discrete random variables," in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 1945–1949.

[6] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov 1973.

[7] V. Jog and V. Anantharam, "The entropy power inequality and mrs. gerber's lemma for groups of order 2^n," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3773–3786, 2014.

[8] C. E. Shannon, "A mathermatical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October, 1948.

[9] A. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019995859903481

[10] E. H. Lieb, "Proof of an entropy conjecture of wehrl," *Comm. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978. [Online]. Available: https://projecteuclid.org:443/euclid.cmp/1103904300

[11] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances in Mathematics*, vol. 20, no. 2, pp. 151–173, 1976.

[12] J. Bennett, A. Carbery, M. Christ, and T. Tao, “The Brascamp-Lieb inequalities: Finiteness, structure and extremals,” *Geometric and Functional Analysis*, vol. 17, no. 5, pp. 1343–1415, 2008.

[13] V. Anantharam, V. Jog, and C. Nair, “Unifying the brascamp-lieb inequality and the entropy power inequality,” *IEEE Transactions on Information Theory*, vol. 68, no. 12, pp. 7665–7684, 2022.

[14] Y. Geng and C. Nair, “The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages,” *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2087–2104, April 2014.

[15] G. Darmois, “Analyse générale des liaisons stochastiques: etude particulière de l’analyse factorielle linéaire,” *Revue de l’Institut International de Statistique / Review of the International Statistical Institute*, vol. 21, no. 1/2, pp. pp. 2–8, 1953. [Online]. Available: http://www.jstor.org/stable/1401511

[16] V. P. Skitovitch, “On a property of the normal distribution,” *DAN SSSR*, vol. 89, pp. 217–219, 1953.

[17] G. Feldman, “More on the skitovich-darmois theorem for finite abelian groups,” *Theory of Probability and Its Applications*, vol. 45, 09 1999.

[18] T. Tao, “Sumset and inverse sumset theory for Shannon entropy,” *Combinatorics, Probability and Computing*, vol. 19, no. 4, pp. 603–639, 2010.

[19] L. Gavalakis, “Discrete generalised entropy power inequalities for log-concave random variables,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 42–47.

[20] W. Gowers, B. Green, F. Manners, and T. Tao, “On a conjecture of Marton,” *arXiv preprint arXiv:2311.05762*, 2023.