A maximal-coupling information inequality of sums on finite subsets of Abelian groups

Ken Lau, Chandra Nair, and Zhaobang Zhu

Abstract

We develop a Han(Shearer)-Inequality type maximal-coupling information inequality of sums on finite subsets of Abelian groups. This is motivated by the exploration of connections between inequalities in additive combinatorics and in information theory. The proof technique employed may be of potential interest in further developing these connections.

Index Terms

Information inequality, Abelian groups, Sumset inequalities, Additive combinatorics.

I. Introduction

In additive combinatorics, inequalities primarily refer to bounds on the sizes of sumsets and related combinatorial structures. These inequalities help understand how the size of the sums of sets relates to the sizes of the original sets and reveal structural properties of the sets involved. Ruzsa's work [1] established a deep connection between sumset inequalities in additive combinatorics and inequalities involving the entropy of discrete random variables. This connection, explored by Ruzsa, Madiman, Marcus, Tetali, Tao, and others, allows for translating combinatorial results into information-theoretic statements and vice versa.

In the words of Ruzsa in [1]: *We do not have a general "theory of translation" between these languages. We illustrate, by examples, the following three scenarios*:

1) *The entropy analog of a cardinality inequality can be formulated, and its equivalence can be proved.*
2) *Their equivalence cannot be established, but the proof of one can be adapted to prove the other.*
3) *There is a cardinality inequality, and the entropy analog as a conjecture is formulated.*

Substantial progress has been made since then towards a "theory of translation". This paper proves another information inequality that aims to provide novel insights into some missing parts of this theory.

Notation and Definitions: We use $(\mathbb{G},+)$ to denote an Abelian group, $(\mathbb{T},+)$ to denote a finitely generated torsion-free Abelian group, which means if $\exists t\in\mathbb{T}, s.t \sum_{i=1}^{n} t = 0_{\mathbb{T}}$ for some $n\in\mathbb{N}-\{0\}$, then $t=0_{\mathbb{T}}$.

For marginal distributions $p_{X_1}, p_{X_2}, ..., p_{X_n}$, let $\prod(p_{X_1}, p_{X_2}, ..., p_{X_n})$ denote the set of joint distributions whose marginals are $p_{X_1}, p_{X_2}, ..., p_{X_n}$ (we will also refer to this as the set of *couplings*). We use $\mathrm{supp}(p_X)$ to denote the support of the distribution p_X.

Let A, B be finite subsets of some Abelian group. Define $A+B \equiv \{a+b, a\in A, b\in B\}$, to be the "sumset of A, B". Let $[1:n]$ denote the set $\{1,2,\ldots,n\}$).

A. Ruzsa's equivalence theorem on sumset inequalities

In [1], Ruzsa made the first significant step to establish an equivalence relationship between sumset inequalities and entropic inequalities. Rather than focusing on conventional sumsets, his equivalence theorem addresses inequalities involving G-restricted sumsets, defined as follows:

Definition 1. (G-restricted sumset [1]) Suppose G is a subset of $A\times B$, where A, B are finite subsets of $(\mathbb{G},+)$. We denote the G-restricted sumset and difference set of A and B as $A \overset{G}{+} B$ and $A \overset{G}{-} B$, where

$$A \overset{G}{+} B = \{a+b : a\in A, b\in B, (a,b)\in G\},$$
$$A \overset{G}{-} B = \{a-b : a\in A, b\in B, (a,b)\in G\}.$$

Using a typicality-based argument, Ruzsa established the following equivalence theorem:

Theorem 1 (Ruzsa equivalence theorem, Equivalence Theorem 2 of [1])**.** *Let $f, g_1, \ldots, g_k$ be linear functions in two variables with integer coefficients, and let $\alpha_1, \ldots, \alpha_k$ be positive real numbers. The following statements are equivalent:*

Ken Lau is with Huawei Research (e-mail: lau.chin.wa1@huawei.com).
Chandra Nair is with CUHK (e-mail: cnair@ie.cuhk.edu.hk).
Zhaobang Zhu is with CUHK (e-mail: zz124@ie.cuhk.edu.hk).

1) For every finite $A \subseteq \mathbb{T} \times \mathbb{T}$ we have

$$|f(A)| \leq \prod |g_i(A)|^{\alpha_i},$$

where $|f(A)|$ denotes the cardinality of the image $f(A)$.

2) For every pair X, Y of (not necessarily independent) random variables with values in $(\mathbb{T}, +)$ such that the entropy of each $g_i(X, Y)$ is finite, the entropy of $f(X, Y)$ is also finite and it satisfies

$$H(f(X, Y)) \leq \sum \alpha_i H(g_i(X, Y)).$$

Remark 1. However, the applicability of Theorem 1 is limited for two reasons: (i) most sumset inequalities in additive combinatorics are not established or valid in the graph-restricted form, and (ii) they do not require the underlying group structure to be a finitely generated torsion-free group. Consequently, to further extend the equivalence relationships between entropic and sumset inequalities, new equivalence theorems were needed.

In [2], a Ruzsa-type theorem was established for regular sumset inequalities. This addresses the first shortcoming mentioned in Remark 1.

Theorem 2. *(Generalized Ruzsa-type equivalence theorem), [3]*

Let $(\mathbb{T}, +)$ be a finitely generated torsion-free Abelian group. Let $f_1, \ldots, f_k$ and $g_1, \ldots, g_\ell$ be linear functions on $\mathbb{T}^n$ with integer coefficients, and let $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_\ell$ be positive real numbers. For the linear function f_i, let $S_i \subseteq [1 : n]$ denote the index set of non-zero coefficients. Similarly, for g_i let $T_i \subseteq [1 : n]$ denote the corresponding index set of non-zero coefficients. (So, effectively, f_i and g_i are linear functions on $\mathbb{T}_{S_i}$ and $\mathbb{T}_{T_i}$ respectively). Further, let us assume that S_i is a pairwise disjoint collection of sets. The following statements are equivalent:

a) For any $A_1, A_2, \ldots, A_n$ that are finite subsets of $\mathbb{T}$, we have

$$\prod_{i=1}^{k} |f_i(A_{S_i})|^{\alpha_i} \leq \prod_{i=1}^{\ell} |g_i(A_{T_i})|^{\beta_i},$$

where $A_S = \otimes_{i \in S} A_i$.

b) For any $m \in \mathbb{N}$, and for any $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n$ that are finite subsets of $\mathbb{T}^m$, we have

$$\prod_{i=1}^{k} |\hat{f}_i(\hat{A}_{S_i})|^{\alpha_i} \leq \prod_{i=1}^{\ell} |\hat{g}_i(\hat{A}_{T_i})|^{\beta_i},$$

where $\hat{A}_S = \otimes_{i \in S} \hat{A}_i$, and $\hat{f}_i$ (and $\hat{g}_i$) are the natural coordinate-wise extensions of f_i (and g_i) respectively, mapping points in

$$\underbrace{\mathbb{T}^m \times \mathbb{T}^m \times \cdots \times \mathbb{T}^m}_{n \text{ times}} \mapsto \mathbb{T}^m.$$

c) For every sequence of random variables $(X_1, \ldots, X_n)$, with fixed marginals p_{X_i} and having finite support in $\mathbb{T}$, we have

$$\sum_{i=1}^{k} \alpha_i \max_{\Pi(X_{S_i})} H(f_i(X_{S_i})) \leq \sum_{i=1}^{\ell} \beta_i \max_{\Pi(X_{T_i})} H(g_i(X_{T_i})),$$

where $\Pi(X_S)$ is collection of joint distributions p_{X_S} that are consistent with the marginals $p_{X_i}, i \in S$.

A key concept that is widely used in additive-combinatorics is the Ruzsa-distance between sets.

Definition 2 (Ruzsa distance)**.** Given sets A, B (finite subsets of some Abelian group),

$$d_R(A, B) := \log \frac{|A - B|}{|A|^{\frac{1}{2}} |B|^{\frac{1}{2}}}.$$

Remark 2. While $d_R(A, A)$ is not zero in general, we have $d_R(A, B) = d_R(B, A) \geq 0$, and $d_R(A, C) \leq d_R(A, B) + d_R(B, C)$, the latter being called the Ruzsa triangle inequality.

The following entropic-Ruzsa distances were defined in the literature.

Definition 3 (Independent Entropic Ruzsa distance, [4])**.** Given two distributions p_X, p_Y having support on some finite subsets of an Abelian group $\mathbb{G}$,

$$d_{ER}(X, Y) := H(X - Y) - \frac{1}{2} H(X) - \frac{1}{2} H(Y).$$

Here, the independent coupling, $(X, Y) \sim p_X p_Y$ is assumed.

Based on Theorem 2, one would venture that a more natural entropic-distance between two distributions

Definition 4 (Coupling Entropic Ruzsa distance, [2])**.** Given two distributions p_X, p_Y having support on some finite subsets of an Abelian group $\mathbb{G}$,

$$d_{CR}(X,Y) := \max_{\Pi(p_X,p_Y)} H(X-Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y).$$

Here $\Pi(p_X, p_Y)$ denotes the set of all couplings consistent with the marginals p_X and p_Y.

B. Background

We begin with the following definition:

Definition 5 (Fractional Partition)**.** A function $\gamma : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is a *fractional partition* of $[1:n]$ w.r.t. $\mathcal{F}$ if

$$\forall i \in [n], \quad \sum_{\substack{S \subseteq [1:n] \\ i \in S}} \gamma(S) = 1.$$

Remark 3. Note that this definition allows the weights of some of the subsets to be zero.

The following theorem serves as a primary motivation for our result.

Theorem 3 (Theorem 1.2, [5])**.** *Let $T_1, T_2, \ldots, T_n$ be finite subsets of an underlying Abelian group $(\mathbb{G}, +)$. Let γ denote a fractional partition of $[1:n]$. The following inequality holds:*

$$|T_1 + T_2 + \cdots + T_n| \leq \prod_{S \subseteq [1:n]} |\sum_{i \in S} T_i|^{\gamma(S)}.$$

If the underlying group is a finitely generated torsion-free Abelian group, then a formally equivalent information inequality can be inferred from an application of the equivalence theorem established in [3] to Theorem 3. However, the authors were concerned by the absence of a self-contained proof of the aforementioned inequality based solely on standard tools from information theory. In this work, we provide such a proof and moreover, our techniques allow us to generalize the result to finite subsets of Abelian groups. This addresses the second concern in Remark 1 and possibly points at a way to develop the theory of equivalence further.

Another result that is related to our main result is called Han's inequality or Shearer's lemma.

Theorem 4 (Han's inequality, [6])**.** *Let $X_1, X_2, \ldots, X_n$ be a collection of random variables with finite support. Let γ denote a fractional partition of $[1:n]$. The following information inequality holds:*

$$\max_{\Pi_{[1:n]}} H(X_1, X_2, \cdots, X_n) \leq \sum_{S \subseteq [1:n]} \gamma(S) H(\{X_i\}_{i \in S}).$$

II. Main

The main result that we establish in this paper is the following information inequality.

Theorem 5. *Let $X_1, X_2, \ldots, X_n$ be a collection of random variables supported on finite subsets of an ambient Abelian group $(\mathbb{G}, +)$. Further, let p_{X_i} denote the probability mass function of the random variable X_i. Given $S \subseteq [1:n]$, let Π_S denote the set of couplings (or joint distributions) of $p_{\boldsymbol{X}_S}$ consistent with the given marginals. Let γ denote a fractional partition of $[1:n]$. The following information inequality holds:*

$$\max_{\Pi_{[1:n]}} H(X_1 + X_2 + \cdots + X_n) \leq \sum_{S \subseteq [1:n]} \gamma(S) \max_{\Pi_S} H(\sum_{i \in S} X_i).$$

Remark 4. The following points are worth noting:

1) Instead of the maximum coupling employed in Theorem 5, if one takes the independent coupling, then the corresponding inequality is known in the literature. The interested readers can refer to [7] or [8]. The use of independence coupling usually leads to analogous inequalities, and tools like data-processing can be used. However, the situation for maximal-coupling is usually much less understood.
2) Note that Theorem 5 implies Theorem 3. To see this, take a distribution of $(X_1, ..., X_n)$ such that it is uniformly supported on a set that contains only a unique pre-image of each element in the support of $(X_1 + \cdots + X_n)$. Hence, clearly, the left hand side evaluates to the logarithm of the cardinality of the sumset. Since, for finite sets, the entropy is upper bounded by the logarithm of the cardinality of the support, Theorem 3 is immediate. Also, note that such an implication to Theorem 3 is absent if you take the independent coupling.

A. Previous Work

There is an elegant proof of Theorem 3 using information theory that can be found in [7]. The main difficulty of obtaining a proof of Theorem 5 from this argument lies in the fact that we are given fixed marginal distributions p_{X_i}, and we have to work within the space of couplings induced by these marginals. The ideas in this paper outline how to overcome these non-trivial obstacles using concepts from convex analysis.

Next, we revisit some of the key ideas and results in [7] as they are useful in our proof as well.

Definition 6 (Good Representative)**.** Let $S_1, .., S_n$ be finite subsets of some ambient Abelian group $\mathbb{G}$. A collection $A \subseteq S_1 \times \cdots \times S_n$ is called a *good representative* of the sumset $S_1 + S_2 + \cdots + S_n$ if the following holds: for every coordinate projection of A into $S_{i_1} \times \cdots \times S_{i_k}$, where $1 \leq k \leq n$ and $S = \{i_1, ..., i_k\} \subseteq [1:n]$, the mapping $\mathbf{a} = (a_{i_1}, .., a_{i_k}) \mapsto a_{i_1} + \cdots + a_{i_k}$ forms a bijection between the appropriate spaces of projections.

If $S_1, ..., S_n$ are all finite sets, it is natural to define a permutation for each support and further define a permutation in their Cartesian product. For a permutation σ in a finite set A, let $a_1 <_\sigma a_2$ denote a_1 before a_2 in permutation σ.

Definition 7 (Lexicographic Permutation)**.** Let π be a permutation of $[1:n]$, and let σ_i be a permutation of the elements of S_i, $1 \leq i \leq n$. The lexicographic permutation $\rho(\pi, \sigma_1, ..., \sigma_n)$ is a permutation of $S_1 \times \cdots \times S_n$ where $\forall s_1, s_2 \in S_1 \times \cdots \times S_n$, $s_1 <_\rho s_2$ if $\exists k \in [1:n]$, s.t $(s_1)_k <_{\sigma_k} (s_2)_k$ and $(s_1)_t = (s_2)_t$ for all $t \in \{i <_\pi k, i \in [1:n]\}$.

Lemma 6 (Lemma 3.1, [7])**.** *Use the notation in Definition 7. Let $s \in S_1 + \cdots + S_n$, and $a_s \in S_1 \times \cdots \times S_n$ denote the first element with respect to ρ of the set $\{\boldsymbol{a} \in S_1 \times \cdots \times S_n, \sum_{i=1}^n a_i = s\}$. The collection $A \equiv \{a_s : s \in S_1 + \cdots + S_n\}$ is a good representative, called the lexicographic representative.*

Proof. We reproduce the proof for completeness, and also to show how our arguments borrowed ideas from this proof. The proof proceeds by contradiction. Assume $\exists \boldsymbol{a}, \hat{\boldsymbol{a}}$, $T = \{i_1, \ldots, i_k\} \subset [1:n]$, such that

$$a_{i_1} + \cdots + a_{i_k} = \hat{a}_{i_1} + \cdots + \hat{a}_{i_k}$$

where $a_{i_l} \neq \hat{a}_{i_l}, 1 \leq l \leq k$. WLOG, assume $(a_{i_1}, ..., a_{i_k}) <_\rho (\hat{a}_{i_1}, ..., \hat{a}_{i_k})$. Consider a new $\boldsymbol{a}_1$ as:

$$\hat{\mathbf{a}}_{1j} = \begin{cases} \hat{\mathbf{a}}_j & j \notin T \\ \mathbf{a}_j & i \in T \end{cases} .$$

It is clear that the sum of the elements of $\hat{\mathbf{a}}$ and $\hat{\mathbf{a}}_1$ are the same, and that $\hat{\mathbf{a}}_1 <_\rho \hat{\mathbf{a}}$, which contradicts that $\hat{\mathbf{a}}$ is prior to any other element that has the same sum. □

The proof, in [7], proceeds by taking a uniform distribution on A. Let us call the induced random variables $(X_1, ..., X_n)$. Because A is a good-representative, we have $H(X_S | \sum_{i \in S} X_i) = 0$, for all $S \subseteq [1:n]$. Now from Han's inequality, for any fractional partition γ,

$$\begin{aligned} \log |S_1 + \cdots + S_n| &= \log |A| \\ &= H(X_1, \ldots, X_n) \\ &\leq \sum_{S \subseteq [1:n]} \gamma(S) H(X_S) \\ &= \sum_{S \subseteq [1:n]} \gamma(S) H(\sum_{i \in S} X_i) \\ &\leq \sum_{S \subseteq [1:n]} \gamma(S) \log |\sum_{i \in S} S_i|. \end{aligned}$$

Our first step in adapting this proof to establish Theorem 5 is to make the following definition.

Definition 8 (Good Representative Distribution)**.** Let $T_1, T_2, \ldots, T_n$ be finite subsets of an underlying Abelian group $(\mathbb{G}, +)$. Let $X_1, ..., X_n$ be random variables taking values in $T_1, ..., T_n$ respectively. We call $p_{X_1,...,X_n}$ to be a *good representative distribution* of $X_1 + \cdots + X_n$, if the $\mathrm{supp}(p_{X_1,...,X_n})$ is a good representative of the sumset $\mathrm{supp}(p_{X_1}) + \mathrm{supp}(p_{X_2}) + \cdots + \mathrm{supp}(p_{X_n})$.

The following lemma yields an alternate equivalent characterization of a good representative distribution.

Lemma 7. *Let $T_1, T_2, \ldots, T_n$ be finite subsets of an underlying Abelian group $(\mathbb{G}, +)$. Let $X_1, ..., X_n$ be random variables taking values in $T_1, ..., T_n$ respectively. Then $p_{X_1,..,X_n}$ is a good representative distribution if and only if $H(X_S | \sum_{i \in S} X_i) = 0, \ \forall S \subseteq [1:n]$.*

The proof is immediate from the observation that for finite-valued random variables, $H(X|f(X)) = 0$ if and only if there is a bijection between the $\mathrm{supp}(p_X)$ and $\mathrm{supp}(p_{f(X)})$.

We cannot proceed in a straightforward fashion because of the following non-trivial obstacles:

1) Given marginal p_{X_i}, if we take any joint distribution on $\Pi_{[1:n]}$ consistent with the marginals, there is no guarantee that the induced distribution on the collection of lexicographic representatives would have the same marginal law.
2) While there is a choice of permutations π and $\sigma_1, .., \sigma_n$ in defining the lexicographic permutation, it is not a priori clear that this freedom can be useful in the presence of having the need to preserve the marginal distributions.

To circumvent the above obstacles, as shown in the next section, we utilize ideas from convex analysis to establish the extremality of good representative distributions.

B. Extremality of good representative distributions

Let $S_1, .., S_n$ be finite subsets of an Abelian group. Let U be a random variable taking values in $S_1 + \cdots + S_n$. Let $\mathcal{M}(p_U)$ denote the set of joint distributions $p_{X_1,...,X_n}$, such that the following two conditions hold:

1) $\text{supp}(p_{X_i}) \subseteq S_i$, $1 \leq i \leq n$
2) the distribution of $X_1 + \cdots + X_n$ is p_U.

In other words, $\mathcal{M}(p_U)$ is the collection of all joint distributions that are supported within $S_1 \times S_2 \cdots \times S_n$, and whose distribution of the sum of the variables is fixed to be p_U.

It is immediate to see that $\mathcal{M}(p_U)$ is a convex set. Further, observe that the collection of good representative distributions is a closed set. This is immediate from the entropic characterization that $H(X_T|\sum_{i\in T} X_i) = 0,\ \forall T \subseteq [1:n]$, and the continuity of the entropy function.

Given $p_{X_1,...,X_n}$, let $\text{Proj}(p_{X_1,...,X_n}) = (p_{X_1}, .., p_{X_n})$ denote the projection of the joint distribution to its marginal distributions. Define the set $\hat{\mathcal{M}}(p_U)$ to be the projection of $\mathcal{M}(p_U)$ on to the set of marginal distributions $(p_{X_1}, p_{X_2}, ..., p_{X_n})$. We say that $p_{X_1,..,X_n}$ is a lift of $(p_{X_1}, p_{X_2}, ..., p_{X_n})$, if $\text{Proj}(p_{X_1,..,X_n}) = (p_{X_1}, p_{X_2}, ..., p_{X_n})$.

The following theorem is key to our proof:

Theorem 8. *The lifts of the extreme points of $\hat{\mathcal{M}}(p_U)$ are good representative distributions.*

Before we prove this theorem, we recall some preliminaries from convex analysis.

Definition 9. A point x in a convex set $K \subset \mathbb{R}^n$ is called an *extreme point* if x cannot be written as a nontrivial convex combination of two distinct points in K. Formally,

$$\begin{aligned} &x \text{ is an extreme point of } K \\ \iff\ &\text{if } x = ty + (1-t)z \text{ for } y, z \in K,\ t \in (0,1), \\ &\text{then } y = x \text{ and/or } z = x. \end{aligned}$$

Definition 10. A point x in a convex set $K \subset \mathbb{R}^n$ is called an *exposed point* if there exists a supporting hyperplane to K that touches K only at x. Formally,

$$\begin{aligned} &x \text{ is an exposed point of } K \\ \iff\ &\exists\, \phi \in (\mathbb{R}^n)^*,\ a \in \mathbb{R} \text{ such that } \phi(x) = a \\ &\text{and } \phi(y) < a \text{ for all } y \in K \setminus \{x\}. \end{aligned}$$

Theorem 9 (Straszewicz's Theorem, Theorem 18.6 [9])**.** *Every compact convex set in a finite-dimensional real vector space is the closed convex hull of its exposed points. Moreover, the set of exposed points is dense in the set of extreme points of the convex set.*

Proof of Theorem 8. Since $\mathcal{M}(p_U)$ is a compact set in finite dimensions, its projection $\hat{\mathcal{M}}(p_U)$ is compact. By Straszewicz's Theorem and the continuity of entropic functions (see Lemma 7 for an entropic characterization of good representative distribution), we can restrict ourselves to showing that the lifts of exposed points are good representative distributions.

Let $p^*_{X_1,...,X_n} \in \mathcal{M}(p_U)$ be any lift of an exposed point in $\hat{\mathcal{M}}(p_U)$. Let $(\lambda_{S_1}, \lambda_{S_2}, \cdots, \lambda_{S_n}) \in \mathbb{R}^{\sum_{i=1}^n |S_i|}$ denote the supporting hyperplane to $\hat{\mathcal{M}}(p_U)$ that intersects the set only at the exposed point. In other words, the optimization problem

$$\max_{p_{X_1,..,X_n}: p_{X_1,..,X_n} \in \mathcal{M}(p_U)} \sum_{i=1}^n p_{X_i}^T \lambda_{S_i}, \tag{1}$$

has a unique maximizer on its projection onto the marginal distributions.

Let $p^*_{X_1,...,X_n}$ be a maximizer of the optimization problem, (1). Assume that $p^*_{X_1,...,X_n}$ is not a good representative distribution. In other words, let $p^*_{X_1,...,X_n}$ have a support A, which is not a good representative of $\text{supp}(p_{X_1}) + \cdots + \text{supp}(p_{X_n})$. This implies that $\exists\ \mathbf{a}, \hat{\mathbf{a}} \in A$ such that their projection into some coordinates of $T = \{i_1, \ldots, i_k\}$ induce the same partial sum: i.e. $a_{i_1} + \cdots + a_{i_k} = \hat{a}_{i_1} + \cdots + \hat{a}_{i_k}$, where $a_{i_l} \neq \hat{a}_{i_l}, 1 \leq l \leq k$.

Assume w.l.o.g. $\sum_{j=1}^k (\lambda_{a_{i_j}} - \lambda_{\hat{a}_{i_j}}) \geq 0$. Let $\mathbf{a}_1$ be obtained from $\mathbf{a}$ by swapping $(a_{i_1}, .., a_{i_k})$ with $(\hat{a}_{i_1}, .., \hat{a}_{i_k})$, like what we did in Lemma 6. Denote $U = X_1 + \cdots + X_n$. Let u_a denote the sum of the elements of $\mathbf{a}, \mathbf{a}_1$.

Define, for $\epsilon > 0$ and small enough, a perturbed conditional distribution defined by

$$p^{\epsilon}_{x_1,\ldots,x_n|U} = \begin{cases} p^*(\mathbf{a}|u_a) - \epsilon & U = u_a, \mathbf{x} = \mathbf{a} \\ p^*(\mathbf{a}_1|u_a) + \epsilon & U = u_a, \mathbf{x} = \mathbf{a}_1 \\ p^*_{x_1,\ldots,x_n|U} & o.w. \end{cases}$$

Further define $p^{\epsilon}_{x_1,\ldots,x_n} = \sum_u p_U(u) p^{\epsilon}_{x_1,\ldots,x_n|U}$. Observe that $p^{\epsilon}_{x_1,\ldots,x_n} \in \mathcal{M}(p_U)$. Note that when $\epsilon = 0$, then $p^{\epsilon}_{x_1,\ldots,x_n} = p^*_{x_1,\ldots,x_n}$. For $\epsilon > 0$, note that the marginal of $p^{\epsilon}_{x_{i_\ell}} \neq p^*_{x_{i_\ell}}, 1 \leq \ell \leq k$. Therefore, for $\epsilon > 0$, the projections of $p^{\epsilon}_{x_1,\ldots,x_n}$ and $p^*_{x_1,\ldots,x_n}$ are distinct points of $\hat{\mathcal{M}}(p_U)$.

Define $f(p) = \sum_{i=1}^n p^T_{X_i} \lambda_{S_i}$. Observe that

$$\begin{aligned} &f(p^{\epsilon}) - f(p^*) \\ &= p_U(u_a)\epsilon \sum_{j=1}^{k} (\lambda_{\hat{a}_{i_j}} - \lambda_{a_{i_j}}) \geq 0. \end{aligned}$$

If the inequality above is strict, then $p^*_{X_1,\ldots,X_n}$ is not the maximizer of (1) (violating our assumption). If it is an equality above, then as projections of $p^{\epsilon}_{x_1,\ldots,x_n}$ and $p^*_{x_1,\ldots,x_n}$ are distinct points of $\hat{\mathcal{M}}(p_U)$, this contradicts that the supporting plane characterized by $(\lambda_{S_1}, \lambda_{S_2}, \cdots, \lambda_{S_n})$ touches $\hat{\mathcal{M}}(p_U)$ only at a single point. In either case, $p^*_{X_1,..,X_n}$ is not the lift of an exposed point of $\hat{\mathcal{M}}(p_U)$. □

A useful corollary of this result on extreme points is the following proposition.

Proposition 10. *Let $p_{X_1,\ldots,X_n}$ be a joint distribution where $supp(X_i) = S_i \subseteq (\mathbb{G}, +)$ and $|S_i| < \infty$. Let $U = \sum_{i=1}^n X_i$. Then there is a distribution $q_{V,X_1,..,X_n}$ such that $q_{X_1+\cdots+X_n} = p_U$, $q_{X_i} = p_{X_i}, 1 \leq i \leq n$, V is independent of U, and $\forall T \subseteq [1:n]$,*

$$H_q(X_T|V) = H_q(\sum_{i\in T} X_i|V). \tag{2}$$

Proof. Observe that $p_{X_1,\ldots,X_n} \in \mathcal{M}(p_U)$. Therefore, from Theorem 8, we can express its projection into $\hat{\mathcal{M}}(p_U)$ as a convex combination of a finite number of extreme points of $\hat{\mathcal{M}}(p_U)$, i.e.,

$$\mathrm{Proj}(p_{X_1,\ldots,X_n}) = \sum_{i=1}^{m} \alpha_i \,\mathrm{Proj}(q^{(i)}_{X_1,\ldots,X_n}),$$

where $q^{(i)}_{X_1,\ldots,X_n} \in \mathcal{M}(p_U)$, and $q^{(i)}_{X_1,\ldots,X_n}$ is a good representative distribution. Let V be a random variable taking values $\{1, 2, .., m\}$ and $\mathbb{P}(V = i) = \alpha_i$, $q_{X_1,\ldots,X_n|V=i} = q^{(i)}_{X_1,\ldots,X_n}$. Since $q^{(i)}_{X_1,\ldots,X_n} \in \mathcal{M}(p_U)$, we have V is independent of U. Further, since $q^{(i)}_{X_1,\ldots,X_n}$ is a good representative distribution, we have

$$H_q(X_T|\sum_{i\in T} X_i, V) = 0, \tag{3}$$

for all $T \subseteq [1:n]$, completing the proof. □

Remark 5. This proposition can be potentially very useful when converting sums to joint distributions, and then employing sub-modularity.

We are now ready to prove our main theorem.

Proof of Theorem 5. For p_{X^n} with marginals $p_{X_1}, ..., p_{X_n}$, consider the q_{X^n}, V induced by Proposition 10. Then $V \perp\!\!\!\perp U := X_1 + \cdots + X_n$, and $H_q(X_T|\sum_{i\in T} X_i, V) = 0$. The last can be rewritten as $H_q(X_T|V) = H(\sum_{i\in T} X_i|V)$. Therefore,

$$\begin{aligned} H_p(\sum_{i=1}^n X_i) &\overset{(a)}{=} H_q(\sum_{i=1}^n X_i) \\ &\overset{(a')}{=} H_q(\sum_{i=1}^n X_i|V) \\ &\overset{(b)}{=} H_q(X_1, ..., X_n|V) \\ &\overset{(c)}{\leq} \sum_{T\subseteq[1:n],} \alpha_T H_q(X_T|V) \\ &\overset{(d)}{=} \sum_{T\subseteq[1:n]} \alpha_T H_q(\sum_{i\in T} X_i|V) \end{aligned}$$

$$\leq \sum_{T\subseteq[1:n]} \alpha_T H_q(\sum_{i\in T} X_i)$$
$$\overset{(e)}{\leq} \sum_{S\subseteq[1:n]} \gamma(S) \max_{\Pi_S} H(\sum_{i\in S} X_i).$$

In the above (a) is a consequence of $p_U = q_U$, where $U = X_1 + \cdots + X_n$, and (a') is due to $V \perp\!\!\!\perp U$. The equalities (b) and (d) are a consequence of $q_{X_1,\ldots,X_n|V=v}$ being a good representative distribution. (c) is an instance of Shearer's lemma or Han's inequality about fractional coverings [6] for random vectors. Finally (e) follows as $q_{X_i} = p_{X_i}$ for all $1 \leq i \leq n$. □

III. Future directions and Conclusion

One immediate direction is to generalize the result to non-commutative group along the lines discussed by Ruzsa.

On another direction, motivated by yet another additive combinatorics inequality, the authors conjecture that the following inequality holds for every p_X, p_Y, p_Z, where X, Y, Z take values in some finite subsets of an ambient Abelian group.

$$H(X) + \max_{\Pi(P_Y,P_Z)} H(Y+Z)$$
$$\leq \max_{\Pi(P_X,P_Y)} H(X+Y) + \max_{\Pi(P_X,P_Z)} H(X+Z).$$

We developed a Han(Shearer)-Inequality type maximal-coupling information inequality of sums on finite subsets of Abelian groups. This was motivated by the exploration of connections between inequalities in additive combinatorics and in information theory. Unlike previous equivalences that involved Sanov's theorem, and hence the need for the underlying group to be finitely generated torsion-free group, this proof builds on ideas from convex analysis.

References

[1] I. Z. Ruzsa, "Sumsets and entropy," *Random Structures & Algorithms*, vol. 34, no. 1, pp. 1–10, 2009.

[2] C. W. Ken Lau and C. Nair, "Information inequalities via ideas from additive combinatorics," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 2452–2457.

[3] C. W. Lau and C. Nair, "Information inequalities via ideas from additive combinatorics," *IEEE Transactions on Information Theory*, vol. 71, no. 6, pp. 4055–4068, 2025.

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

[5] K. Gyarmati, M. Matolcsi, and I. Z. Ruzsa, "A superadditivity and submultiplicativity property for cardinalities of sumsets," *Combinatorica*, vol. 30, no. 2, pp. 163–174, 2010.

[6] M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2699–2713, June 2010.

[7] M. Madiman, A. Marcus, and P. Tetali, "Entropy and set cardinality inequalities for partition-determined functions," *Random Structures & Algorithms*, vol. 40, pp. 399–424, 2012.

[8] C. W. Lau, C. Nair, and D. Ng, "A mutual information inequality and some applications," *IEEE Trans. Inf. Theor.*, vol. 69, no. 10, p. 6210–6220, Oct. 2023.

[9] R. Rockafellar, *Convex Analysis*, ser. Convex Analysis. Princeton University Press, 1997.