Entropy and set cardinality inequalities for partition-determined functions

Mokshay Madiman[*] Adam W. Marcus[†] Prasad Tetali[‡]

Abstract

A new notion of partition-determined functions is introduced, and several basic new inequalities are presented for the entropy of such functions of independent random variables, as well as for cardinalities of compound sets obtained using these functions. Here a compound set means a set obtained by varying each argument of a function of several variables over a set associated with that argument, where all the sets are subsets of an appropriate algebraic structure so that the function is well defined. The compound set inequalities imply, in turn, several inequalities for sumsets, providing for instance partial progress towards a conjecture of Ruzsa (2007) for sumsets in nonabelian groups.

Keywords: Sumsets, additive combinatorics, entropy inequalities, cardinality inequalities.

1 Introduction

It is well known in certain circles that there appears to exist an informal parallelism between entropy inequalities on the one hand, and set cardinality inequalities on the other. In this paper, we clarify some aspects of this parallelism, while presenting new inequalities for both entropy and set cardinalities.

A natural connection between entropy and set cardinalities arises from the fact that the entropy of the uniform distribution on a finite set of size m is just $\log m$, and this is the maximum entropy of any distribution supported on the same set. Consequently, inequalities for entropy lead to inequalities for cardinalities of sets. For instance, by choosing (X, Y) to be uniformly distributed on the set $A \subset B \times C$, the classical inequality $H(X,Y) \leq H(X) + H(Y)$ implies $\log |A| \leq \log |B| + \log |C|$ or $|A| \leq |B| \cdot |C|$.

For the joint entropy, there is an elaborate history of entropy inequalities starting with the chain rule of Shannon, whose major developments include works of Han, Shearer, Fujishige, Yeung, Matúš, and others. The classical part of this work, involving so-called Shannon-type inequalities that use the submodularity of the joint entropy, was synthesized and generalized in [14, 15], where an array of lower as well as upper bounds were given, for the joint entropy of a collection of random variables generalizing inequalities of Han [10], Fujishige [7] and Shearer [4]. For the history of the non-classical part, involving so-called non-Shannon inequalities, one may consult for instance, Matúš [16] and references therein.

[*] Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA. Email: mokshay.madiman@yale.edu

[†] Department of Mathematics, Yale University, PO Box 208283, New Haven, CT 06520, USA. Email: adam.marcus@yale.edu

[‡] School of Mathematics and School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA. Email: tetali@math.gatech.edu

Entropies of sums, even in the setting of independent summands, are not as well understood as joint entropies. For continuous random variables, the so-called entropy power inequalities provide important lower bounds on entropy of sums, see, e.g., [13]. For discrete random variables, an unpublished work of Tao and Vu [21] gives some upper bounds on entropy of sums, and discusses the analogy between entropy of sums of random variables (instead of joint entropy) and sumset cardinality inequalities (instead of set projection inequalities). In this paper, we develop this analogy in directions not considered in [21], and prove more general inequalities both for entropies and sumsets.

The property of sumsets that allows us to apply ideas from entropy is that, for a fixed element a, the sum $a+b$ depends only on b (no further knowledge of how a and b are related is needed). We will formalize this idea into what we will call a *partition-determined* function. This means that the function (more precisely, an extension of it) can be defined on a subset of its arguments, in such a way that the original function can be recovered by looking at the functional value on the subset and on its complementary subset. The projection and sum functions are merely the simplest example of such partition-determined functions.

Definition 1. Let $X_1, X_2, \ldots, X_k$ be finite sets. Any nonempty subset $s \subset [k]$ corresponds to a different product space $\mathbf{X}_s = \prod_{i \in s} X_i$. For sets $s \subseteq t \subseteq [k]$, we define the projection function $\pi_s : \mathbf{X}_t \to \mathbf{X}_s$ in the natural way: for $x \in \mathbf{X}_t$, let $\pi_s(x) = (x_{i_1}, \ldots, x_{i_{|s|}})$ where $i_j \in s$. (To avoid ambiguity, one may assume that the indices i_j are labeled in increasing order, i.e., $i_{j+1} > i_j$.) When the meaning is clear, we will write $\pi_i(x)$ for $\pi_{\{i\}}(x)$.

We will use $Q(X_1, X_2, \ldots, X_k)$ to denote the space that is a disjoint union of each of the spaces $\mathbf{X}_s$, for nonempty $s \subseteq [k]$. Formally,

$$Q(X_1, X_2, \ldots, X_k) = \bigcup_{\phi \neq s \subseteq [k]} \left\{ (x_{i_1}, \ldots, x_{i_{|s|}}) : x_i \in X_i, s = \{i_1, \ldots, i_{|s|}\} \right\}$$

Let Y be any space and $f : Q(X_1, \ldots, X_k) \to Y$ be any function. Then, for a nonempty set $s \subset [k]$, we define $f_s : \mathbf{X}_s \to Y$ to be the restriction of f to only those inputs that came from $\mathbf{X}_s$. We will abuse notation by writing, for nonempty $s \subseteq t$ and $x \in t$, $f_s(x)$ to mean $f_s(\pi_s(x))$, and when the domain is clear, we will merely write $f(x)$.

Let s be a subset of $[k]$ and let $\overline{s}$ denote $[k] \setminus s$. We will say that a function f defined on $Q(X_1, X_2, \ldots, X_k)$ is *partition-determined with respect to* s if for all $x, y \in X_{[k]}$, we have that $f(x) = f(y)$ whenever both $f_s(x) = f_s(y)$ and $f_{\overline{s}}(x) = f_{\overline{s}}(y)$. Extending this idea to collections of subsets $\mathcal{C}$, we will say that f is *partition-determined with respect to* $\mathcal{C}$ if f is partition-determined with respect to s for all $s \in \mathcal{C}$. Finally, in the case that f is partition-determined with respects to all subsets $[k]$, we will simply write that f is *partition-determined.*

The definition above is intended to capture the property of sumsets that was mentioned earlier. For a function f to be partition-determined with respect to a single set $s \subset [k]$, it must be that $f_s(x)$ and $f_{\overline{s}}(x)$ uniquely determine the value of $f(x)$. Then being partition-determined with respect to a collection $\mathcal{C}$ is nothing more than being partition-determined with respect to all $s \in \mathcal{C}$.

Simple examples relevant for consideration include Cartesian products of sets and linear combinations of sets (and so, in particular, sumsets). Both these classes of examples are partition-determined with respect to $\mathcal{C}$ for any $\mathcal{C}$.

Example 1. *Let V be a vector space over the reals with basis vectors $\{v_1, \ldots, v_k\}$. Let $X_1, \ldots, X_k \subseteq \mathbb{R}$ and define $f : Q(X_1, \ldots, X_k) \to V$ such that $f_s(x) = \sum_{i \in s} \pi_i(x) v_i$. Then f is partition-determined with respect to $\mathcal{C}$ for all collections $\mathcal{C}$ of subsets of $[k]$.*

Proof. Let $x \in X_t$ for some $t \subseteq [k]$ and let $s \in \mathcal{C}$ where $\mathcal{C}$ is a collection of subsets of $[k]$. Then

$$f(x) \;=\; \sum_{i \in t} \pi_i(x) v_i \;=\; \sum_{i \in (s \cap t)} \pi_i(x) v_i \;+\; \sum_{i \in (\overline{s} \cap t)} \pi_i(x) v_i \;=\; f_s(x) + f_{\overline{s}}(x).$$

Thus knowing $f_s(x)$ and $f_{\overline{s}}(x)$ uniquely determines $f(x)$. Since this is true for any $s \in \mathcal{C}$, f is partition-determined with respect to $\mathcal{C}$. □

Example 2. *Let $(\mathcal{G}, +)$ be an Abelian group and $X_1, \ldots, X_k \subseteq \mathcal{G}$ and let $c_1, \ldots, c_k \in \mathbb{Z}$. Define $f : Q(X_1, \ldots, X_k) \to \mathcal{G}$ such that $f_s(x) = \sum_{i \in s} c_i \pi_i(x)$. Then f is partition-determined with respect to $\mathcal{C}$ for all collections $\mathcal{C}$ of subsets of $[k]$.*

Proof. The proof is identical to Example 1, only replacing v_i with c_i. □

Equipped with the notion of partition-determined functions, we prove an array of inequalities for both entropy and set cardinality. For instance, we have the following results for sums as corollaries of general statements for partition-determined functions.

Illustrative Entropy Result: Let $Z_1, \ldots, Z_n$ be independent discrete random variables taking values in the Abelian group $(\mathcal{G}, +)$, and let $\mathcal{C}$ be an r–regular hypergraph on $[n]$. Then

$$H(Z_1 + \cdots + Z_n) \leq \frac{1}{r} \sum_{s \in \mathcal{C}} H\biggl(\sum_{i \in s} Z_i\biggr).$$

Illustrative Set Cardinality Result: Let $A, B_1, B_2, \ldots, B_n \subset \mathcal{G}$ be finite subsets of the Abelian group $(\mathcal{G}, +)$. If $\mathcal{C}$ is an r-regular hypergraph on $[n]$, then for any $D \subseteq B_1 + \ldots + B_n$,

$$|A + D|^{|\mathcal{C}|} \leq |D|^{|\mathcal{C}|-r} \prod_{s \in \mathcal{C}} \biggl| A + \sum_{i \in s} B_i \biggr|.$$

This set inequality (and others for sums, projections etc.) is obtained as a corollary of inequalities for cardinalities of compound sets. Here a compound set means a set obtained by varying each argument of a function of several variables over a set associated with that argument, where all the sets are subsets of an appropriate algebraic structure so that the function is well defined. In other words, for subsets $X_1, \ldots, X_k$ of some ambient space $\mathcal{X}$, we will always use the notation

$$f(X_1, \ldots, X_k) = \{f(x_1, \ldots, x_k) : x_1 \in X_1, \ldots, x_k \in X_k\}.$$

When the ambient space is a group, the only operation available is the sum, and all compound sets are sumsets. When the ambient space is a ring, one may consider compound sets built from polynomials. For particular ambient spaces, such as Euclidean space, the class of functions available is extremely broad and therefore so is the class of compound sets that can be considered.

This paper is organized as follows. Section 2.1 presents preliminaries on mutual information, and a key lemma on entropies of partition-determined functions. Section 2.2 presents a rather general new submodularity property of the entropy for strongly partition-determined functions (defined there) of independent random variables. Surprisingly, this result is entirely elementary and relies on the classical properties of joint entropy. Section 2.3 uses this result to demonstrate new, general upper bounds on entropies of partition-determined functions. Section 2.4 considers applications to particular partition-determined functions such as sums.

Section 3.1 applies joint entropy inequalities to obtain a basic but powerful result about compound sets. Section 3.2 discusses the relation between this general result and the entropy inequalities for partition-determined functions obtained in Section 2. The remaining subsections of Section 3 discuss various consequences of the basic result for compound set cardinalities in Section 3.1. Section 3.3 gives a first easy application of the compound set result to obtaining well known cardinality inequalities for projections of sets.

Section 3.4 gives a second, and important, application to sumsets in Abelian groups. In this, we build on recent work by Gyarmati, Matolcsi and Ruzsa [8], who noted that Han-type inequalities can be applied to sumsets in much the same way that they can be applied to characteristic functions of sets of random variables (the usual situation). While it is not true in general that sumsets satisfy a log-submodular relation in an obvious way, it is natural to ask whether they permit a weaker property, by way of fractional subadditivity. It is classical (and recently reviewed in [15]) that fractional subadditivity is weaker than log-submodularity and more general than Han's inequalities. Here, by extending an idea embedded in [8], and making further use of entropy, we show a general fractional subadditivity property for sumsets that implies some of the results and conjectures in [8] as easy corollaries.

Section 3.5 applies the compound set inequalities to obtain results for sumsets in non-Abelian groups, motivated by a conjecture of Ruzsa [19]. In particular, we make partial progress towards resolving Ruzsa's conjecture. In Section 3.6, we present a novel application of our basic result to obtaining cardinality bounds for compound sets in rings that are more general than sumsets (such as the compound sets $g(A, B)$ in a ring obtained from $g(a, b) = a^2 + b^2$ or $g(a, b) = a^2 - b^2$).

It should be noted that multiple papers on this topic have appeared in the literature recently. In particular, two other parties have published work on topics similar to this one, and it seems that each work provides unique and independent contributions to the overall goal of applying entropy to sumsets. In the process, however, some of the corollaries that are obtained from the results in this paper have been shown by Balister and Bollobás [1] using independent techniques. Their primary result uses the ideas of Gyarmati, Matolcsi, and Ruzsa [8] and Madiman and Tetali [15] to develop a hierarchy of entropy and sumset inequalities, which intersects the collection of inequalities obtained in this paper (in which some of the corollaries lie). In yet a different direction, Ruzsa [20] explores more deeply the relationship between sums and sets, and was able to deduce some of the same corollaries using Han's inequalities. A second paper of Gyarmati, Matolcsi, and Ruzsa [9] also independently proves a subset of the corollaries derived here.

We will attempt point out the intersections with [1, 20, 9] at the appropriate places in the individual sections. We wish to emphasize, however, that apart from the novelty of our specific results, the approach we adopt using partition-determined functions provides a general and powerful framework for studying such inequalities, and in that sense goes beyond the aforementioned papers such as [1] and [9] that prove similar inequalities for the specific context of sumsets.

2 Entropy inequalities

2.1 Partition-determined functions and mutual information

As usual, we denote by $[n]$ the index set $\{1, 2, \ldots, n\}$. Let $Z_{[n]} = (Z_1, Z_2, \ldots, Z_n)$ be a collection of random variables, and assume each Z_i takes values in some finite set X_i. If the probability distribution of $Z_{[n]}$ has joint probability mass function $p(z_{[n]}) = \mathbb{P}(Z_{[n]} = z_{[n]})$, where $z_{[n]} = (z_1, \ldots, z_n)$, then the *entropy* of $Z_{[n]}$ is defined by

$$H(Z_{[n]}) = \sum_{z_{[n]} \in X_1 \times \ldots \times X_n} -p(z_{[n]}) \log p(z_{[n]}).$$

Recall that the *conditional entropy* of Z given Y, denoted $H(Z \mid Y)$, is defined by taking the mean using the distribution of Y of the entropy of the conditional distribution of Z given $Y = y$. The standard fact that $H(Z, Y) = H(Z) + H(Y|Z)$ has come to be known as Shannon's chain rule for entropy. For any function f, it is easy to see that $H(f(Z)) \leq H(Z)$, with equality if and only if f is a bijection.

The *mutual information* between two jointly distributed random variables Z and Y is defined by

$$I(Z; Y) = H(Z) - H(Z \mid Y),$$

and is a measure of the dependence between Z and Y. In particular, $I(Z; Y) = 0$ if and only if Z and Y are independent.

Analogous to conditional entropy, one may also define the *conditional mutual information* $I(Z; Y \mid X)$, which quantifies how much more information about Y one can glean from the pair (X, Z) as compared to simply X. More precisely, we can define

$$\begin{aligned} I(Z; Y \mid X) &= I(X, Z; Y) - I(X; Y) && (1)\\ &= H(X, Z) - H(X, Z \mid Y) - [H(X) - H(X \mid Y)] && (2)\\ &= H(X, Z) - H(X, Z, Y) - H(X) + H(X, Y), && (3) \end{aligned}$$

where the alternate form in the last display was obtained by adding and subtracting $H(Y)$ in the last step, and using Shannon's chain rule for entropy. The following lemma gives a simple and classical property of mutual information.

Lemma 2.1. *The mutual information cannot increase when one looks at functions of the random variables (the "data processing inequality"):*

$$I(f(Z); Y) \leq I(Z; Y).$$

A proof can be found in elementary texts on information theory such as Cover and Thomas [5]. The following strengthened notion of partition-determined functions will turn out to be useful:

Definition 2. We say that $f : Q(X_1, \ldots, X_n) \to Y$ is *strongly partition-determined* if for any disjoint sets s and t, the values of $f_{s \cup t}$ and f_t (together) completely determine the value of f_s.

Observe that the converse is always true; that is, given f_t and f_s, $f_{s \cup t}$ whenever f is partition-determined. Also, both running examples, namely projections and sums, are strongly partition-determined functions. For brevity, we simply write $H(f_s)$ for $H(f_s(Z))$.

Lemma 2.2. *Suppose X_i are finite sets, and $f : Q(X_1, \ldots, X_n) \rightarrow V$ is a partition-determined function. Let $Z_1, \ldots, Z_n$ be random variables, with Z_i taking values in X_i. Then, for disjoint sets $s, t \subset [n]$,*

$$I(f_{s \cup t}; f_t) \geq H(f_{s \cup t}) - H(f_s). \tag{4}$$

If, furthermore, f is strongly partition-determined and $Z_1, \ldots, Z_n$ are independent, then

$$I(f_{s \cup t}; f_t) = H(f_{s \cup t}) - H(f_s). \tag{5}$$

Proof. Since conditioning reduces entropy,

$$H(f_{s \cup t}) - H(f_s) \overset{(a)}{\leq} H(f_{s \cup t}) - H(f_s | f_t).$$

But since f is partition-determined, $f_{s \cup t} = \phi(f_s, f_t)$ for some function ϕ, and hence

$$H(f_s | f_t) = H(f_s, f_t | f_t) \overset{(b)}{\geq} H(f_{s \cup t} | f_t).$$

Thus

$$H(f_{s \cup t}) - H(f_s) \leq H(f_{s \cup t}) - H(f_{s \cup t} | f_t) = I(f_{s \cup t}; f_t).$$

This yields the first part of the lemma. For the second part, note that independence of $Z_1, \ldots, Z_n$ guarantees equality in (a), while f being strongly partition-determined guarantees equality in (b). □

2.2 A basic result for entropy

The inequality below, while a simple consequence of the above elementary facts, is rather powerful.

Theorem 2.3. [Submodularity for Strongly Partition-determined Functions] *Suppose X_i are finite sets, and $f : Q(X_1, \ldots, X_n) \rightarrow V$ is a strongly partition-determined function. Let $Z_1, \ldots, Z_n$ be independent random variables, with Z_i taking values in X_i. Then*

$$H(f_{s \cup t}) + H(f_{s \cap t}) \leq H(f_s) + H(f_t) \tag{6}$$

for any nonempty subsets s and t of $[k]$.

Proof. First note that it suffices to prove the result for $n = 3$ (since we can consider collections of random variables to be a single random variables under the joint distribution). For this case, we have

$$\begin{aligned} & H(f_{\{1,2\}}) + H(f_{\{2,3\}}) - H(f_{\{1,2,3\}}) - H(f_{\{2\}}) \\ & = H(f_{\{1,2\}}) - H(f_{\{2\}}) - \big[H(f_{\{1,2,3\}}) - H(f_{\{2,3\}})\big] \\ & = I(f_{\{1,2\}}; f_{\{1\}}) - I(f_{\{1,2,3\}}; f_{\{1\}}), \end{aligned}$$

using (5) from Lemma 2.2. Thus we simply need to show that

$$I(f_{\{1,2\}}; f_{\{1\}}) \geq I(f_{\{1,2,3\}}; f_{\{1\}}).$$

Now

$$\begin{aligned} I(f_{\{1,2,3\}}; f_{\{1\}}) &\overset{(a)}{\leq} I(f_{\{1,2\}}, f_{\{3\}}, f_{\{1\}}) \\ &\overset{(b)}{=} I(f_{\{1,2\}}; f_{\{1\}}) + I(f_{\{3\}}; f_{\{1\}}|f_{\{1,2\}}) \\ &\overset{(c)}{=} I(f_{\{1,2\}}; f_{\{1\}}) \end{aligned}$$

where (a) follows from the data processing inequality, (b) follows from (1) and (c) follows from the hypothesis of independence; so the proof is complete. □

Remark 1. Observe that we can allow the consideration of empty sets in Theorem 2.3 (and below) if we set $H(f_\phi) = 0$. Indeed, this is natural because $f_\phi(x) = f(\pi_\phi(x))$ does not have any actual arguments by definition, and hence must be the constant function, for which the entropy is of course 0.

A consequence of the submodularity of entropy of sums is an entropy inequality that obeys the partial order constructed using compressions, as introduced by Bollobás and Leader [2]. Following Balister and Bollobás [1], we introduce some notation. Let $M(n,m)$ be the family of multisets (allowing repetition) of non-empty subsets of $[n]$ with a total of m elements. Consider a given multiset $\mathcal{C} = \{s_1, \ldots, s_l\} \in M(n,m)$. The idea is to consider an operation that takes two sets in $\mathcal{C}$ and replaces them by their union and intersection; however, note that (i) if s_i and s_j are nested (i.e., either $s_i \subset s_j$ or $s_j \subset s_i$), then replacing (s_i, s_j) by $(s_i \cap s_j, s_i \cup s_j)$ does not change $\mathcal{C}$, and (ii) if $s_i \cap s_j = \phi$, the null set may enter the collection, which would be undesirable. Thus, take any pair of non-nested sets $\{s_i, s_j\} \subset \mathcal{C}$ and let $\mathcal{C}' = \mathcal{C}(ij)$ be obtained from $\mathcal{C}$ by replacing s_i and s_j by $s_i \cap s_j$ and $s_i \cup s_j$, keeping only $s_i \cup s_j$ if $s_i \cap s_j = \phi$. $\mathcal{C}'$ is called an *elementary compression* of $\mathcal{C}$. The result of a sequence of elementary compressions is called a *compression*.

Define a partial order on $M(n,m)$ by setting $\mathcal{A} > \mathcal{B}$ if $\mathcal{B}$ is a compression of $\mathcal{A}$. To check that this is indeed a partial order, one needs to rule out the possibility of cycles, which can be done by noting that

$$\sum_{s \in \mathcal{A}} |s|^2 < \sum_{s \in \mathcal{A}'} |s|^2;$$

if $\mathcal{A}'$ is an elementary compression of $\mathcal{A}$.

Theorem 2.4. *Suppose X_i are finite sets, and $f : Q(X_1, \ldots, X_n) \to V$ is a strongly partition-determined function. Let $Z_1, \ldots, Z_n$ be independent random variables, with Z_i taking values in X_i. Let $\mathcal{A}$ and $\mathcal{B}$ be finite multisets of subsets of $[n]$, with $\mathcal{A} > \mathcal{B}$. Writing $f_s = f_s(Z_1, \ldots, Z_n)$,*

$$\sum_{s \in \mathcal{A}} H(f_s) \geq \sum_{t \in \mathcal{B}} H(f_t).$$

Proof. The proof follows exactly the same reasoning as given by Balister and Bollobás [1] for the special case of f being the identity function (or f_s being the projection function). When $\mathcal{B}$ is an *elementary* compression of $\mathcal{A}$, the statement is immediate from the submodularity of $H(f_s)$ proved in Theorem 2.3, and transitivity of the partial order gives the full statement. □

Note that for every multiset $\mathcal{A} \in M(n,m)$ there is a unique minimal multiset $\mathcal{A}^\#$ dominated by $\mathcal{A}$ consisting of the sets $s_j^\# = \{i \in [n] : i$ lies in at least j of the sets $s \in \mathcal{A}\}$. Thus a particularly nice instance of Theorem 2.4 is for the special case of $\mathcal{B} = \mathcal{A}^\#$.

2.3 Upper bounds for entropy of a partition-determined function

Let $\mathcal{C}$ be a collection of subsets of $[n]$. For any index i in $[n]$, define the *degree* of i in $\mathcal{C}$ as $r(i) = |\{t \in \mathcal{C} : i \in t\}|$. A function $\alpha : \mathcal{C} \to \mathbb{R}_+$, is called a *fractional covering*, if for each $i \in [n]$, we have $\sum_{s\in\mathcal{C}:i\in s} \alpha_s \geq 1$. If α satisfies the equalities $\sum_{s\in\mathcal{C}:i\in s} \alpha_s = 1$ for each $i \in [n]$, it is called a *fractional partition*. If the degree of every index i in $\mathcal{C}$ is exactly r, $\mathcal{C}$ is called an r-regular hypergraph, and $\alpha_s = 1/r$ for every $s \in \mathcal{C}$ constitutes a fractional partition using $\mathcal{C}$.

Theorem 2.5. [Upper Bound for Entropy of Strongly Partition-determined Function] *Let $Z_1, \dots, Z_n$ be independent random variables, each taking values in a finite set. Then, writing $f_s = f_s(Z_1, \dots, Z_n)$, we have*

$$H(f_{[n]}) \leq \sum_{s\in\mathcal{C}} \alpha_s H\big(f_s\big),$$

for any fractional covering α using any collection $\mathcal{C}$ of subsets of $[n]$.

Proof. Define the set function $g(s) = H(f_s)$, and note that $g(\phi) = 0$ is the appropriate convention (see Remark 1). Theorem 2.3 says that g is a submodular function. Now the corollary follows from the general fact that a submodular function g with $g(\phi) = 0$ is "fractionally subadditive" (see, e.g., [15]). □

For any collection $\mathcal{C}$ of subsets, [15] introduced the degree covering, given by

$$\alpha_s = \frac{1}{r_-(s)},$$

where $r_-(s) = \min_{i\in s} r(i)$. Specializing Theorem 2.5 to this particular fractional covering, we obtain

$$H(f_{[n]}) \leq \sum_{s\in\mathcal{C}} \frac{1}{r_-(s)} H\big(f_s\big).$$

A simple example is the case of the collection $\mathcal{C}_m$, consisting of all subsets of $[n]$ with m elements, for which the degree of each index with respect to $\mathcal{C}_m$ is $\binom{n-1}{m-1}$.

2.4 Special cases

The first (and best studied) special case of interest is when f is the identity mapping, so that f_s is the projection onto the subset s of coordinates. In this case, Theorems 2.3, 2.4 and 2.5 reduce to the following fact. For any set $s \subset [n]$, let Z_s stand for the random variable $(Z_i : i \in s)$, with the indices taken in their increasing order. If $Z_1, \dots, Z_n$ are independent discrete random variables taking values in the group $\mathcal{G}$, then

$$H(Z_1, \dots, Z_n) \leq \sum_{s\in\mathcal{C}} \alpha_s H\big(Z_s\big),$$

for any fractional covering α using any collection $\mathcal{C}$ of subsets of $[n]$. In the context of independent random variables, however, this fact is not particularly enlightening (although true, and with equality for fractional partitions). For more on its validity in the general dependent case, see [15].

A perhaps more interesting second case is when the sets X_i are finite subsets of an ambient abelian group $(\mathcal{G}, +)$, and f is just the sum function.

Corollary 2.6. [Entropy of Sums in Abelian Groups] *Let $Z_1, \dots, Z_n$ be independent discrete random variables taking values in the abelian group $\mathcal{G}$, and let*

$$Z_s^+ = \sum_{i \in s} Z_i.$$

Then:

1. *The set function $f(s) = H(Z_s^+)$ is submodular.*

2. *If $\mathcal{A} > \mathcal{B}$ then*

$$\sum_{s \in \mathcal{A}} H(Z_s^+) \geq \sum_{t \in \mathcal{B}} H(Z_t^+).$$

3. *For any fractional covering α using any collection $\mathcal{C}$ of subsets of $[n]$,*

$$H(Z_1 + \cdots + Z_n) \leq \sum_{s \in \mathcal{C}} \alpha_s H\big(Z_s^+\big).$$

Since the sum function in an abelian group is strongly partition-determined, the three statements in Corollary 2.6 follow from Theorems 2.3, 2.4 and 2.5 respectively. Parts of Corollary 2.6 were presented in [11]. Let us note, in particular, that the first part of Corollary 2.6 resolves affirmatively a strengthened form of "Entropy Conjecture 3" in the recent paper of Ruzsa [20].

Note that while the simplest applications of Corollary 2.6 are to discrete abelian groups, it continues to hold for arbitrary abelian groups $\mathcal{G}$ provided the random variables Z_i are supported on discrete subsets of $\mathcal{G}$. In fact, a modified version of Corollary 2.6 holds even beyond the discrete support setting. Specifically, one can work with $\mathcal{G}$-valued random variables on any abelian locally compact topological group $\mathcal{G}$, whose distribution is absolutely continuous with respect to the Haar measure on $\mathcal{G}$. Such continuous versions, and their application to matrix analysis, are explored in [12]. For instance, [12] shows that the following modified inequality holds even in the continuous setting where the entropy can be negative: if $H(X) \geq 0$, then

$$H\bigg(X + \sum_{i \in [n]} X_i\bigg) \quad \leq \textstyle\sum_{s \in \mathcal{C}} \alpha_s H\bigg(X + \sum_{i \in s} X_i\bigg).$$

This is identical to Corollary 2.6 except that the random variable X is an additional summand in every sum. Some of the sumset inequalities we will discuss below have a similar flavor, in that a set appears in every sum.

3 Set cardinality inequalities

3.1 A basic result for reduced sets

The proofs in this section extend the arguments of Gyarmati, Matolcsi, and Ruzsa [8] (more precisely, an idea in Theorem 1.2 of their paper) while making further use of entropy. Given a collection of spaces X_i, their idea was to endow each X_i with an arbitrary linear order. Then for any set $s \subseteq [k]$ and elements $A, B \in X_s$, they define $A <_{lex} B$ if the vector representation of A comes before that of B in lexicographical order.

In the statements and proofs that follow, we will work in the following setting: Let $X_1, X_2, \ldots, X_k$ be finite sets, and f be a function defined on $Q(X_1, X_2, \ldots, X_k)$. Now let $Y \subseteq f(\mathbf{X}_{[k]})$ be a given set that we wish to bound in size. For each $y \in Y$, we define $r(y)$ to be the smallest element of $f^{-1}(y) \subseteq \mathbf{X}_{[k]}$ in lexicographical order and set $R = \{r(y) : y \in Y\}$. Thus each $y \in Y$ has a unique "representative preimage" $r(y)$, and $|Y| = |R|$.

The main observation is the following somewhat surprising lemma, which requires nothing more than the elementary properties of entropy discussed above; this is also the essence (in addition to Han's Inequality) of the proof of Theorem 1.2 in [8].

Lemma 3.1. *Let Z be a random variable that chooses uniformly from the elements of R. If f is partition-determined with respect to s, then*

$$H\big(Z_s \mid f(Z_s)\big) = 0.$$

Proof. Since Z_s takes values in $\pi_s(R)$, it suffices to show that the restriction of f to the domain $\pi_s(R)$ is a one-to-one function. Assume, for the sake of contradiction, that there are two elements $a \neq b \in X_s$ such that $f(a) = f(b)$ and both $Pr(Z_s = a)$ and $Pr(Z_s = b)$ are non-zero. Thus there must be "preimages" of a, b with respect to π_s; that is, elements $A, B \in R$ such that $\pi_s(A) = a$ and $\pi_s(B) = b$. Furthermore, $A \neq B$ since $a \neq b$.

Without loss of generality, let $A <_{lex} B$, and define the vectors $A', B' \in X_{[k]}$ by switching the values of a and b in A and B; that is

$$A'_i = \begin{cases} B_i \ (= b_i) & \text{for } i \in s \\ A_i & \text{for } i \notin s \end{cases} \qquad \text{and} \qquad B'_i = \begin{cases} A_i \ (= a_i) & \text{for } i \in s \\ B_i & \text{for } i \notin s \end{cases}.$$

See Figure 1 for an example.

$$\begin{array}{lcllcl} a & = & (a_2, a_4, a_5) & b & = & (b_2, b_4, b_5) \\ A & = & (A_1, a_2, A_3, a_4, a_5) & B & = & (B_1, b_2, B_3, b_4, b_5) \\ A' & = & (A_1, b_2, A_3, b_4, b_5) & B' & = & (B_1, a_2, B_3, a_4, a_5) \end{array}$$

Figure 1: Example of a, b, A, B, A', B' with $k = 5$ and $s = \{2, 4, 5\}$.

Note that since $f(a) = f(b)$ and f is partition-determined with respect to s, it follows that $f(A) = f(A')$ and $f(B) = f(B')$. Due to the fact that A and B were chosen to be the representatives of $f(A)$ and $f(B)$ respectively, one must have $A <_{lex} A'$ and $B <_{lex} B'$. On the other hand A and A' agree on all $i \notin s$, so it must be that $a <_{lex} b$. The same argument for B and B', however, implies that $b <_{lex} a$, a contradiction. □

The following result now becomes basic when considering such constructions.

Theorem 3.2. *Suppose f is partition-determined with respect to $\mathcal{C}$, and α is a fractional covering of $[k]$ using $\mathcal{C}$. For any $Y \subseteq f(X_{[k]})$, we have that*

$$|Y| \leq \prod_{s \in \mathcal{C}} \left| f_s \left(f_{[k]}^{-1}(Y) \right) \right|^{\alpha_s}.$$

Proof. Let Z be a random variable distributed uniformly on the elements of R and let $Z_i = \pi_i(Z)$ for all $i \in [k]$. Then by the usual fractional subadditivity of entropy, we have

$$\log\big(|R|\big) = H(Z) = H(Z_1, \ldots, Z_k) \leq \sum_{s \in \mathcal{C}} \alpha_s H(Z_s) \tag{7}$$

where $Z_s = \pi_s(Z)$. By the chain rule for entropy,

$$H\big(Z_s \mid f(Z_s)\big) + H\big(f(Z_s)\big) = H\big(Z_s, f(Z_s)\big) = H\big(f(Z_s) \mid Z_s\big) + H(Z_s) \tag{8}$$

for each $s \in \mathcal{C}$. Here, $H\big(f(Z_s)|Z_s\big) = 0$ since f is completely determined by its input. On the other hand, $H(Z_s|f(Z_s)) = 0$ by Lemma 3.1, so Equation 8 reduces to $H(Z_s) = H\big(f(Z_s)\big)$.

Plugging this into Equation (7) yields:

$$\log\big(|R|\big) \leq \sum_{s \in \mathcal{C}} \alpha_s H\big(f(Z_s)\big) \leq \sum_{s \in \mathcal{C}} \alpha_s \log\big(\,|f_s\,(R)|\,\big) \leq \sum_{s \in \mathcal{C}} \alpha_s \log\Big(\big|f_s\big(f_{[k]}^{-1}(Y)\big)\big|\Big)$$

where the last inequality is due to the fact that $R \subseteq f_{[k]}^{-1}(Y)$. Since $|Y| = |R|$, our claimed result is true. $\square$

Considering the full compound set rather than a subset of it yields the following corollary.

Corollary 3.3. *Suppose f is partition-determined with respect to $\mathcal{C}$, and α is a fractional covering of $[k]$ using $\mathcal{C}$. Then*

$$|f(X_{[k]})| \leq \prod_{s \in \mathcal{C}} |f(X_s)|^{\alpha_s}\,.$$

The only potential problem in obtaining Corollary 3.3 from Theorem 3.2 is that $f^{-1} \circ f(X_{[k]})$ could in general be a superset of $X_{[k]}$, if the X_i are finite subsets of an ambient space G. However this is not a problem, since throughout we only require f to be defined on $X_{[k]}$ (and hence can work with the range of the inverse function being thought of as restricted to $X_{[k]}$).

Let us also mention here that Corollary 3.3 is not a consequence of an underlying submodularity property; in particular, one can generally find examples for which

$$\big|f_{s\cup t}\left(f_{[k]}^{-1}(Y)\right)\big| \cdot \big|f_{s\cap t}\left(f_{[k]}^{-1}(Y)\right)\big| > \big|f_s\left(f_{[k]}^{-1}(Y)\right)\big| \cdot \big|f_t\left(f_{[k]}^{-1}(Y)\right)\big|.$$

Such counterexamples in the cases of f corresponding to projections and sums are discussed in Sections 3.3 and 3.4 respectively. This phenomenon appears to be in stark contrast to the corresponding entropy statement (Theorem 2.3), which asserts for strongly partition-determined functions such as sums that such a submodularity holds for entropy. The discrepancy is explained by the fact that one needs to invoke a set of representatives to pass from entropy to set cardinality.

3.2 Remarks on the connection between the entropy and set inequalities

Let $Z_1, \ldots, Z_n$ be random variables supported by the sets $X_1, \ldots, X_n$ respectively, and note that for any subset $s \subset [n]$, the random variable $f_s(Z)$ is supported on the compound set $f_s(X)$. Thus the left hand side of Theorem 2.5 has the bound

$$H(f_s(Z)) \leq \log|f_s(X)|, \tag{9}$$

while its right hand side has the bound

$$\sum_{s\in\mathcal{C}} \alpha_s H\big(f_s(Z)\big) \le \sum_{s\in\mathcal{C}} \alpha_s \log \big|f_s(X)\big|.$$

Corollary 3.3 says that these bounds themselves are ordered. This would be implied by Theorem 2.5 if we could find a product distribution on $(X_1,\ldots,X_n)$ that made their sum uniformly distributed on its range, since then (9) would simply hold with equality.

Interestingly, while it is in general not possible to find such product distributions, it is always possible to find a *joint* distribution (with dependence) that makes the sum uniformly distributed on its range.

Lemma 3.4. *There exists a joint distribution for Z on X which makes $f_{[k]}(Z)$ uniformly distributed on $f_{[k]}(X)$.*

Proof. Let $|X_j| = m_j, j\in[k]$. Clearly $f_{[k]}^{-1}$ partitions $f_{[k]}^{-1}(Y)$ into l subsets, where $l=|Y|$. For each $y\in Y$, want

$$\sum_{(x_1,\ldots,x_k)\in f_{[k]}^{-1}(y)} p(x_1,\ldots,x_k) = P(Z_{[k]}=y) = \frac{1}{l}.$$

Thus there are l linear equations (one for each possible value of y) in the $\prod_{j\in[k]} m_j$ variables. (It is not necessary to write down a separate equation for all the probabilities summing to 1 since that is implied by this linear system.) Observe that since $l \le |f(X_{[k]})| \le |X_{[k]}| = \prod_{j\in[k]} m_j$, there are always more variables than equations, and that the linear independence of the equations ensures consistency and hence existence of solutions to the system. In fact, the following is an explicit positive solution: Let the element of the partition corresponding to y have size l_y; for each $(x_1,\ldots,x_k)\in f_{[k]}^{-1}(y)$, set

$$p(x_1,\ldots,x_k) = \frac{1}{l l_y}.$$

Thus one can find a probability distribution on $X_{[k]}$ that makes $f(X_{[k]})$ uniformly distributed on Y. □

If one insists on a product distribution, one gets $l-1$ nonlinear (polynomial) equations in $\sum_{j\in[k]}(m_j-1)$ variables. In the case of sumsets, the Cauchy-Davenport theorem implies that the number of equations is at least the number of variables. It is intriguing that both the simplest upper bound as well as the simplest lower bound for sumset cardinalities appear in counting equations needed for the existence of general and product distributions respectively, leading to the uniform distribution on the sumset. Even when the number of equations exactly equals the number of variables, however, the nonlinearity implies that real solutions may fail to exist. For instance, consider the sets $X_1 = X_2 = \{0,1\}$. If Z_1 and Z_2 take the value 1 with probabilities p and q respectively, the conditions

$$pq = \frac{1}{3}, \quad (1-p)(1-q) = \frac{1}{3}$$

can be solved only if p and q are allowed to be complex.

In summary, we have the following observation. The hoped-for simple method of proof of Corollary 3.3 from Theorem 2.5 as outlined above fails because of the independence requirement of Theorem 2.5. The method of proof of Corollary 3.3 in Section 3.1 using the uniform distribution on a set of representatives is precisely designed to address this problem.

It is pertinent to note that the inequality of Theorem 2.5 does in fact hold for one particular choice of non-product distribution, namely, for the random variable Z that is uniformly distributed on the set of representatives. Indeed, by Lemma 3.1, the submodularity of $H(Z_s)$ implies that $H(f_s(Z))$ is submodular.

3.3 Corollaries on set projections

There are some well-known inequalities for the cardinalities of set projections that come directly out of Theorem 3.2. We mention this way of proving them only for completeness, and to illustrate the unifying nature of Theorem 3.2.

Corollary 3.5. *Let α be a fractional covering using the hypergraph $\mathcal{C}$ on $[k]$. Let $X_1, \dots, X_k$ be arbitrary finite sets, and $Y \subset \mathbf{X}_{[k]}$. Then*

$$|Y| \leq \prod_{s \in \mathcal{C}} |\pi_s(Y)|^{\alpha_s}.$$

Proof. Apply Theorem 3.2 to the function $f_s = \pi_s$, which is obviously partition-determined with respect to any collection $\mathcal{C}$. □

For instance, Lemma 3.1 of Gyarmati, Matolcsi and Ruzsa [8] follows from Corollary 3.5. As suggested there, these inequalities can be directly derived either from Shearer-type entropy inequalities (such as those of [15]) or from the so-called 'Box Theorem' of Bollobás and Thomason [3].

It is pertinent to note that the corresponding (stronger) submodularity inequality

$$|\pi_{s \cup t}(Y)| \cdot |\pi_{s \cap t}(Y)| \leq |\pi_s(Y)| \cdot |\pi_t(Y)|$$

does not hold. For a counterexample, consider $X_i = \{0, 1\}$ for $i = 1, 2, 3$, and let $Y = \{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,0,1)\}$. Then $|\pi_{\{1,2\}}(Y)| = 3$, $|\pi_{\{2,3\}}(Y)| = 3$, while $|\pi_{\{1,2,3\}}(Y)| = 5$ and $|\pi_{\{2\}}(Y)| = 2$. (This counterexample is implicit in [1].)

3.4 Corollaries on Abelian sumsets

The following corollaries are a slight variant on Theorem 3.2 in that they do not directly define R; rather, they pick a subset D of the image of one of the subspaces, then lift the preimage of D up to the top space.

Theorem 3.6. *Let $A, B_1, B_2, \dots, B_k \subset \mathcal{G}$, where $(\mathcal{G}, +)$ is an Abelian group under the operation $+$. For any $s \subset [k]$, define*

$$B_s^+ = \sum_{i \in s} B_i,$$

which is well defined by commutativity of addition. Let α be any fractional partition on $[k]$ using the collection $\mathcal{C}$ of subsets of $[k]$. Then, for any $D \subseteq B_{[k]}^+$,

$$|A + D|^c \leq |D|^{c-1} \prod_{s \in \mathcal{C}} |A + B_s^+|^{\alpha_s},$$

where $c = \sum_{s \in \mathcal{C}} \alpha_s$.

Proof. Set $X_i = B_i$ for $i \in [k]$, and $X_{k+1} = A$. Let $\mathbf{X}_{[k+1]}$ be the Cartesian product set. Clearly, the collection of functions $f_s(x) = \sum_{i \in s} x_i$, for $s \subset [k+1]$, is partition-determined with respect to any collection $\mathcal{C}'$ of subsets of $[k+1]$. Since $D \subseteq f_{[k]}(X)$, set $Q = f_{[k]}^{-1}(D)$, then $Q \subset \mathbf{X}_{[k]}$. Note that $Y = D + A$ can be written as

$$Y = \{f(b_1, \ldots, b_k, a) : (b_1, \ldots, b_k) \in Q, a \in A\}.$$

Now choose

$$\mathcal{C}' = \{[k]\} \cup \left\{s' : s' = s \cup \{k+1\}, s \in \mathcal{C}\right\}.$$

For each set $s \in \mathcal{C}$, let

$$\gamma_{s \cup \{k+1\}} = \frac{\alpha_s}{\sum_{t \in \mathcal{C}} \alpha_t}$$

and let

$$\gamma_{[k]} = 1 - \frac{1}{\sum_{t \in \mathcal{C}} \alpha_t}.$$

Note that γ is a fractional covering (partition, in fact) for $[k+1]$ using $\mathcal{C}'$; for the index $k+1$, one has

$$\sum_{s \in \mathcal{C}} \gamma_{s \cup \{k+1\}} = 1$$

and for each $j \in [k]$, one has

$$\gamma_{[k]} + \sum_{s \in \mathcal{C} : s \ni j} \gamma_{s \cup \{k+1\}} = 1.$$

By Theorem 3.2, we have that

$$|Y| \leq \prod_{s' \in \mathcal{C}'} \left| f_{s'}\left(f_{[k+1]}^{-1}(Y)\right)\right|^{\gamma_{s'}}.$$

Using the fact that $f_{[k+1]}^{-1}(Y) = Q \times A$, it is easy to see that $f_{[k]}\left(f_{[k+1]}^{-1}(Y)\right) = D$, and that if $s' = s \cup \{k+1\}$ for $s \subset [k]$,

$$f_{s'}\left(f^{-1}(Y)\right) \subseteq B_s^+ + A.$$

The result follows. □

By applying Theorem 3.6 to an r-regular hypergraph $\mathcal{C}$, for which $\alpha_s = \frac{1}{r}$ gives a fractional partition, we obtain the following corollary (the Illustrative Set Result in the Introduction).

Corollary 3.7. *Under the assumptions of Theorem 3.6, if* $\mathcal{C}$ *is an* r*-regular hypergraph, then for any* $D \subseteq B_{[k]}^+$*,*

$$|A + D|^{|\mathcal{C}|} \leq |D|^{|\mathcal{C}| - r} \prod_{s \in \mathcal{C}} |A + B_s^+|.$$

Note that when $\mathcal{C} = \mathcal{C}_1$ is the collection of singleton sets, Corollary 3.7 reduces to Theorem 1.5 of Gyarmati, Matolcsi, and Ruzsa [8], namely

$$|A + D|^k \leq |D|^{k-1} \prod_{i=1}^{k} |A + B_i|.$$

When $\mathcal{C} = \mathcal{C}_{k-1}$ is the collection of leave-one-out sets, Corollary 3.7 resolves a conjecture stated in [8], namely if $\overline{B_i} = B_1 + \cdots + B_{i-1} + B_{i+1} + \cdots + B_k$ for $i = 1, \ldots, k$,

$$|A + D|^k \leq |D| \prod_{i=1}^{k} |A + \overline{B_i}|.$$

Clearly, various other choices of $\mathcal{C}$ yield other similar corollaries. We mention these only because they offer direct generalizations of Theorem 1.5 in Gyarmati, Matolcsi, and Ruzsa [8] and are independent of the results in Balister and Bollobás [1].

Incidentally, Ruzsa has observed that sumset cardinality is not log submodular. To see this, note that log submodularity of sumset cardinality would imply that $|kA|$ is a log concave function of k, which is not the case. If $|A| = n$ and $|2A| = m$, then $|3A|$ can be anywhere between cm and $C \min(m^{3/2}, m^3/n^2)$.

3.5 Corollaries on non-Abelian sumsets

One particular generalization that has been explored recently is that to non-Abelian groups. In particular, given $X_1, \ldots, X_k$ subsets of a non-Abelian group $(\mathcal{G}, \circ)$, can we find similar bounds on $|X_1 \circ \ldots \circ X_k|$ as we did when the underlying group was Abelian? Unfortunately, the non-Abelian addition function is no longer partition-determined with respect to any collection of subsets; however, it is partition-determined with respect to some collections of subsets, and so with a little added work we can still use Lemma 3.1 to find bounds. To see that, in fact, the same bounds cannot hold, consider the following example.

Example 3. *Let $\mathcal{G} = \{e, R, R^2, F, RF, R^2F\}$ be the dihedral group on 6 elements, $S = \{e, F\}, T = \{R\}$, and $U = \{e, F\}$. Then it is* not *the case that $|S \circ T \circ U|^2 \leq |S \circ T||T \circ U||S \circ U|$.*

Proof. On one hand, we have that $S \circ T = \{R, FR\}, S \circ U = \{e, F\}$, and $T \circ U = \{R, RF\}$, and so $|S \circ T||T \circ U||S \circ U| = 8$. On the other hand, $S \circ T \circ U = \{R, FR, RF, R^2\}$ and so $|S \circ T \circ U|^2 = 16$. □

The underlying reason that non-Abelian groups cannot be bounded in such a way is that, as in the example above, $|S \circ U|$ need not have any relation to $|S \circ T \circ U|$. So any bound will need to find some way to link the two. To do so, we will need to use stronger inequalities than the ones mentioned in the previous sections. We will derive them using the simplied proof of Shearer's Lemma due to Llewellyn and Radhakrishnan (see [18]), and which has been used to find extended Shearer-type bounds (the most general of these appearing in [6, 15]). However, the lemma below seems to be disjoint from results mentioned previously in the literature (and is a considerable strengthening of the similar Han inequality). The general idea is that the entropy of a collection of random variables can be compared to the sum of the pairs of random variables conditioned on all of the random variables falling in between the pair.

Lemma 3.8. *Let $Z = Z_1, Z_2, \ldots, Z_k$ be random variables, and define*

$$Z_{(i,j)} = \{Z_t : i < t < j\}$$

for all $1 \leq i < j \leq k$. Then, for $k \geq 2$,

$$(k-1)H(Z_1, Z_2, \ldots, Z_k) \leq \sum_{i=1}^{k} \sum_{j>i}^{k} H(Z_i, Z_j \mid Z_{(i,j)})$$

Proof. We will prove this by induction on k. The base case (when $k = 2$) is trivial, so assume the hypothesis to be true for $k-1$ random variables and consider a collection of k random variables. The general idea will be to peel off all of the pairs that contain the random variable Z_k and then appeal to the induction hypothesis for the other $\binom{k-1}{2}$ pairs. We write $H(Z)$ a total of $k-1$ times, in different forms:

$$\begin{array}{llll}
H(Z) = & & H(Z_1, Z_k \mid Z_2, \ldots, Z_{k-1}) & + H(Z_2, \ldots, Z_{k-1}) \\
H(Z) = & H(Z_1 \mid Z_2, \ldots, Z_{k-1}, Z_k) & + H(Z_2, Z_k \mid Z_3, \ldots, Z_{k-1}) & + H(Z_3, \ldots, Z_{k-1}) \\
H(Z) = & H(Z_1, Z_2 \mid Z_3, \ldots, Z_{k-1}, Z_k) & + H(Z_3, Z_k \mid Z_4, \ldots, Z_{k-1}) & + H(Z_4, \ldots, Z_{k-1}) \\
\vdots & \vdots & \vdots & \vdots \\
H(Z) = & H(Z_1, \ldots, Z_{k-3} \mid Z_{k-2}, Z_{k-1}, Z_k) & + H(Z_{k-2}, Z_k \mid Z_{k-1}) & + H(Z_{k-1}) \\
H(Z) = & H(Z_1, \ldots, Z_{k-3}, Z_{k-2} \mid Z_{k-1}, Z_k) & + H(Z_{k-1}, Z_k) &
\end{array}$$

Note that the middle terms are all of the type $H(Z_i, Z_k \mid Z_{(i,k)})$ (in particular, $H(Z_{k-1}, Z_k) = H(Z_{k-1}, Z_k \mid Z_{(k-1,k)})$ since $Z_{(k-1,k)}$ is empty). Furthermore, removing a random variable from the conditioning can only increase the entropy, so if we remove Z_k from all of the leftmost terms, we get

$$\begin{array}{lllllll}
H(Z) & \leq & & & H(Z_1, Z_k \mid Z_{(1,k)}) & + & H(Z_2, \ldots, Z_{k-1}) \\
H(Z) & \leq & H(Z_1 \mid Z_2, \ldots, Z_{k-1}) & + & H(Z_2, Z_k \mid Z_{(2,k)}) & + & H(Z_3, \ldots, Z_{k-1}) \\
H(Z) & \leq & H(Z_1, Z_2 \mid Z_3, \ldots, Z_{k-1}) & + & H(Z_3, Z_k \mid Z_{(3,k)}) & + & H(Z_4, \ldots, Z_{k-1}) \\
\vdots & & \vdots & & \vdots & & \vdots \\
H(Z) & \leq & H(Z_1, \ldots, Z_{k-3} \mid Z_{k-2}, Z_{k-1}) & + & H(Z_{k-2}, Z_k \mid Z_{(k-2,k)}) & + & H(Z_{k-1}) \\
H(Z) & \leq & H(Z_1, \ldots, Z_{k-3}, Z_{k-2} \mid Z_{k-1}) & + & H(Z_{k-1}, Z_k \mid Z_{(k-1,k)}) & &
\end{array}$$

And now if we sum all of these inequalities, the entries in the leftmost column can be combined with the entries in the rightmost column that is one level higher to get

$$(k-1)H(Z) \leq \sum_{i=1}^{k-1} H(Z_i, Z_k \mid Z_{(i,k)}) + (k-2)H(Z_1, Z_2, \ldots, Z_{k-1})$$

and the rest of the inequality follows from the induction hypothesis. □

Using Lemma 3.8, we can give a collection of inequalities for non-Abelian groups.

Theorem 3.9. *Let $X_1, X_2, \ldots, X_k$ be subsets of a non-Abelian group, and define*

$$A(i,j) = \max\{|X_i \circ x_{i+1} \circ \ldots \circ x_{j-1} \circ X_j| : x_{i+1} \in X_{i+1}, \ldots, x_{j-1} \in X_{j-1}\}$$

for all $1 \leq i < j \leq k$. Then, for $k \geq 2$,

$$|X_1 \circ X_2 \circ \ldots \circ X_k|^{k-1} \leq \prod_{1 \leq i < j \leq k} A(i,j)$$

Proof. We define Z as before. By Lemma 3.8, we have that

$$(k-1)H(Z_1, Z_2, \ldots, Z_k) \leq \sum_{i=1}^{k} \sum_{j>i}^{k} H(Z_i, Z_j \mid Z_{(i,j)})$$

so it suffices to show that $H(Z_i, Z_j \mid Z_{(i,j)}) \leq \log A(i,j)$. Note that in the Abelian case, we would have been pleased enough with $H(Z_i, Z_j)$ but in the non-Abelian case, the sumset f that we used before is not (necessarily) partition-determined with respect to the partition $\{\{i,j\}, \overline{\{i,j\}}\}$. In light of this, for each subset $s \subset [k]$, we define the set function g as

$$g_s(x_1, \ldots, x_k) = \begin{cases} x_{i_1} \circ \ldots \circ x_{i_{|s|}} & \text{if } s \text{ forms a consecutive interval in } [k] \\ (y_1, y_2, \ldots, y_k) & \text{where } y_i = x_i \text{ for } i \in s \text{ and } y_i = 0 \text{ for } i \notin s, \text{ otherwise} \end{cases}$$

It is easy to check that this g is partition-determined with respect to any partition. And while we have seemingly increased the sizes of a number of the sets that we wish to bound, the conditioning allows us to overcome this obstacle.

$$H(Z_i, Z_j \mid Z_{(i,j)}) = H(Z_{[i,j]}) - H(Z_{(i,j)}) = H(g_{[i,j]}(Z)) - H(Z_{(i,j)}) = H(g_{[i,j]}(Z) \mid Z_{(i,j)})$$

by Lemma 3.1. Furthermore, we have

$$\begin{aligned} H(g_{[i,j]}(Z) \mid Z_{(i,j)}) &= \sum_{z_{(i,j)} \in Z_{(i,j)}} H(g_{[i,j]}(Z) \mid Z_{(i,j)} = z_{(i,j)}) Pr(Z_{(i,j)} = z_{(i,j)}) \\ &\leq \max_{z_{(i,j)} \in Z_{(i,j)}} H(g_{[i,j]}(Z) \mid Z_{(i,j)} = z_{(i,j)}) \\ &\leq \log \max\{|X_i \circ x_{i+1} \circ \ldots \circ x_{j-1} \circ X_j| : x_{i+1} \in X_{i+1}, \ldots, x_{j-1} \in X_{j-1}\} \\ &= \log A(i,j) \end{aligned}$$

since, assuming that $Z_{(i,j)} = z_{(i,j)}$, the range of $g_{[i,j]}$ is at most $|X_i \circ z_{i+1} \circ \ldots \circ z_{j-1} \circ X_j|$. □

Two interesting properties of our methods are revealed in Theorem 3.9. The first is that g does not have to (necessarily) be the obvious compound function — in fact, the normal sumset function is not partition-determined in the non-Abelian setting, and our choice of g was able to overcome this. The second is that the size of the sets $g_{\overline{s}}$ only play a factor in the bound if $\overline{s}$ happens to be equal to some other t which is chosen in a different partition. Thus if we are careful, we can inflate some of the sets in order to make g partition-determined without changing the bound. The following corollary, which inspired Theorem 3.9, was originally proved by Ruzsa [19].

Corollary 3.10. *Let S, T, U be subsets of a non-Abelian group. Then*

$$|S \circ T \circ U|^2 \leq \max_{t \in T} |S \circ T||T \circ U||S \circ t \circ U|.$$

Curiously, the method seems to break down in other cases. In particular, it is unknown whether Theorem 3.9 remains true in the asymmetric case (using the same covering set of pairs, but having different weights α_s). In addition, different covering sets are also still unexplored. For example, the following problem posed by Ruzsa [19] remains open.

Problem 1. *Let S, T, U, V be subsets of a non-Abelian group. Is it true that*

$$|S \circ T \circ U \circ V|^3 \leq \max_{t,u} |S \circ T \circ U||S \circ T \circ u \circ V||S \circ t \circ U \circ V||T \circ U \circ V|?$$

Observe that the corresponding entropy inequality is *not* true; indeed, if one chooses $Z_2 = Z_3$ and $Z_1 = Z_4 = 0$, then

$$\begin{aligned} H(Z_2) &= H(Z_1, Z_2, Z_3, Z_4) \\ &> \frac{1}{3}\big[H(Z_1, Z_2, Z_3) + H(Z_2, Z_3, Z_4) + H(Z_1, Z_3, Z_4 \mid Z_2) + H(Z_1, Z_2, Z_4 \mid Z_3)\big] \\ &= \frac{2}{3}H(Z_2). \end{aligned}$$

3.6 Corollaries on polynomial compound sets

Although all of the examples and applications of partition-determined set functions thus far have been associated with sumsets, this does not need to be the case. Indeed, one can consider arbitrary "compound sets" obtained by plugging sets in as arguments of any function involving well-defined operations on the ambient space. In other words, for subsets $X_1, \ldots, X_k$ of some ambient space $\mathcal{X}$, we will always use the notation

$$f(X_1, \ldots, X_k) = \{f(x_1, \ldots, x_k) : x_1 \in X_1, \ldots, x_k \in X_k\}.$$

When the ambient space is a group, the only operation available is the sum, and all compound sets are sumsets. When the ambient space is a ring, one may consider compound sets built from polynomials. For particular ambient spaces, such as Euclidean space, the class of functions available is extremely broad and therefore so is the class of compound sets that can be considered.

In this paper, we only illustrate the possible uses of our Theorem 3.2 in the context of a ring. The result below works for possibly non-commutative, non-unital rings. Note that while one must take care to maintain the order of terms within any finite monomial over a non-commutative ring, one can still define polynomials (a sum of finite monomials) and corresponding polynomial functions although they are no longer themselves commutative. For our purposes, monomials in 2 indeterminates $\mathbf{x}, \mathbf{y}$ over a non-commutative ring could include, for instance, $a\mathbf{x}^b c\mathbf{y}^d e\mathbf{x}^f g$. We allow all such monomials with finitely many terms; all that matters is that when two such monomials appear with opposite signs in a polynomial, they cancel to the additive identity 0 because of the commutativity of addition.

Given sets $A_1, \ldots, A_k$, and a function g, we will write $g(A_1, \ldots, A_k)$ to denote the set $\{g(a_1, \ldots, a_k) : a_i \in A_i\}$. While it is perhaps overly pedantic to draw specific attention to notation, it is nonetheless necessary here, due to the proliferation of similar notation in the additive combinatorics community (see [22]). In the usual notation for additive combinatorics, $AB + BA$ is used to refer to $\{ab + b'a' : a, a' \in A, b, b' \in B\}$. However, when thinking of the compound set obtained by applying $g(a, b) = ab + ba$ to sets A and B, we need to consider the set $A \cdot B \oplus B \cdot A = \{ab + ba : a \in A, b \in B\}$. The symbols $\cdot$ and $\oplus$ are used to denote *bound* multiplication and addition operations, i.e., when they appear in an expression, repeated appearances of a set symbol in that expression mean that the *same* element of the set is substituted for each such appearance in computing the various instances of the expression. Our results could also be used to obtain results for the usual setting by using (for example) $A \cdot B \oplus C \cdot D$ where A, D (likewise B, C) are a posteriori taken to be the same set, but they will tend to be weaker than the bounds that exploit the algebra of the underlying ring.

Theorem 3.11. *Let R be a (possibly non-commutative and non-unital) ring. Suppose $f : R^m \to R$ can be extended to a function $\bar{f}$ on Q that is partition-determined with respect to $\mathcal{C}$. For each $i \in [m]$, let $g_i \in R[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ be a polynomial in n indeterminates with coefficients in R, let $g_i : R^n \to R$ be the corresponding polynomial function, and let $g : R^n \to R^m$ be the function whose i-th component is g_i. Let $F : R^n \to R$ be the polynomial function associated with the polynomial $F \in R[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ obtained by reducing the expression*

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_m), \quad \textit{with} \quad \mathbf{y}_i = g_i(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

in the $\mathbf{x}$-indeterminates, i.e., removing some pairs of monomials that are additive inverses of each other after substitution and expansion of terms. Then, for any collection $X_1, \ldots, X_n$ of finite subsets of R, and for any fractional covering α with respect to the hypergraph $\mathcal{C}$ on $[m]$,

$$|F(X_1, \ldots, X_n)| \leq \prod_{s \in \mathcal{C}} |\bar{f} \circ \pi_s \circ g(X_1, \ldots, X_n)|^{\alpha_s}.$$

Proof. Observe that when $\mathbf{y}_i = g_i(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are substituted in $f(\mathbf{y}_1, \ldots, \mathbf{y}_m)$ and the resulting expression is expanded as a sum of monomials in the $\mathbf{x}$-indeterminates, there is a polynomial $C \in R[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ such that, *before cancellation* of any monomials,

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_m) = F(\mathbf{x}_1, \ldots, \mathbf{x}_n) + C(\mathbf{x}_1, \ldots, \mathbf{x}_n) - C(\mathbf{x}_1, \ldots, \mathbf{x}_n),$$

by the definition of F. Consequently,

$$f(Y_1, \ldots, Y_m) = F(X_1, \ldots, X_n) \oplus C(X_1, \ldots, X_n) \ominus C(X_1, \ldots, X_n).$$

Since the set $C \ominus C$ is just the singleton $\{0\}$ containing the additive identity (because these are bound operations),

$$\begin{aligned} |F(X_1, \ldots, X_n)| &= |f(Y_1, \ldots, Y_m)| \\ &\leq \prod_{s \in \mathcal{C}} |\bar{f}(Y_s)|^{\alpha_s} \\ &= \prod_{s \in \mathcal{C}} |\bar{f} \circ \pi_s \circ g(X_1, \ldots, X_n)|^{\alpha_s}, \end{aligned}$$

where we used Corollary 3.3 for the inequality. □

Example 4. *Let A, B be subsets of a ring R. Then*

$$|A^2 \oplus B^2| \leq |(A \oplus B)^2| \cdot |A \cdot B \oplus B \cdot A|,$$

where the squares are understood to be bound squares.

Remark: When R is the ring of real numbers $\mathbb{R}$, the inequality in Example 4 implies that for any finite rectangular grid of points in the plane (corresponding to the Cartesian product $A \times B$), the number of distinct circles centered at the origin passing through grid points is bounded by the product of the number of distinct lines of slope -1 passing through grid points and the number of distinct hyperbolae with the axes as their asymptotes that pass through the grid points.

Proof. Let $f(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{y}_1^2 - \mathbf{y}_2$. With $\mathbf{y}_1 = \mathbf{x}_1 + \mathbf{x}_2$ and $\mathbf{y}_2 = \mathbf{x}_1\mathbf{x}_2 + \mathbf{x}_2\mathbf{x}_1$, one finds $F(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 + \mathbf{x}_2)^2 - \mathbf{x}_1\mathbf{x}_2 - \mathbf{x}_2\mathbf{x}_1 = \mathbf{x}_1^2 + \mathbf{x}_2^2$. Since $\bar{f}$ is partition-determined with respect to $\big\{\{1\}, \{2\}\big\}$ if we set $\bar{f}(\mathbf{y}_1) = \mathbf{y}_1^2$ and $\bar{f}(\mathbf{y}_2) = \mathbf{y}_2$, one can apply Theorem 3.11. □

A large class of further examples is provided by a general rule — whenever F can be factorized, one can use the product function for f (which is always partition-determined with respect to any $\mathcal{C}$), and obtain bounds for the cardinality of the polynomial set in terms of the cardinalities of the "factor sets".

Corollary 3.12. *Let R be a commutative ring. Suppose $F \in R[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ (written without any redundant monomials) has a factorization of the form*

$$F(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{j \in [m]} g_j(\mathbf{x}_1, \ldots, \mathbf{x}_n),$$

where each g_j is also a polynomial (typically not involving some of the $\mathbf{x}$-indeterminates). Then for any collection $X_1, \ldots, X_n$ of finite subsets of R, and for any fractional covering α with respect to the hypergraph $\mathcal{C}$ on $[m]$,

$$|F(X_1, \ldots, X_n)| \leq \prod_{s \in \mathcal{C}} \left| \prod_{j \in s} g_j(X_1, \ldots, X_n) \right|^{\alpha_s}.$$

Proof. Simply note that $F(\mathbf{x}_1, \ldots, \mathbf{x}_n) = f(\mathbf{y}_1, \ldots, \mathbf{y}_m)$, with the correspondence $\mathbf{y}_j = g_j(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $f(\mathbf{y}_1, \ldots, \mathbf{y}_m) = \prod_{j \in [m]} \mathbf{y}_j$. Now f has an obvious extension to Q, whose restriction to Y_s is given by $\bar{f}((\mathbf{y}_j : j \in s)) = \prod_{j \in s} \mathbf{y}_j$. An application of Theorem 3.11 completes the proof. □

Acknowledgments

We thank Imre Ruzsa for sharing the preprint [8], for helpful discussions, and for informing us of the independent and recent work of Balister–Bollobás. We also thank Béla Bollobás for promptly sending us the preprint [1], which contains results of independent interest.

References

[1] P. Balister and B. Bollobás, "Projections, entropy, and sumsets," *Preprint*, October 2007.

[2] B. Bollobás and I. Leader, "Compressions and isoperimetric inequalities," *J. Combinatorial Theory Ser. A*, vol. 56, no. 1, pp. 47–62, 1991.

[3] B. Bollobás and A. Thomason, "Projections of bodies and hereditary properties of hypergraphs," *Bull. London Math. Soc.*, vol. 27, no. 5, pp. 417–424, 1995.

[4] F. Chung, R. Graham, P. Frankl, and J. Shearer, "Some intersection theorems for ordered sets and graphs," *J. Combinatorial Theory, Ser. A*, vol. 43, pp. 23–37, 1986.

[5] T. Cover and J. Thomas, *Elements of Information Theory.* New York: J. Wiley, 1991.

[6] E. Friedgut. Hypergraphs, entropy, and inequalities. *The American Mathematical Monthly*, 111(9):749–760, November 2004.

[7] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Information and Control*, vol. 39, pp. 55–72, 1978.

[8] K. Gyarmati, M. Matolcsi, and I. Ruzsa, "A superadditivity and submultiplicativity property for cardinalities of sumsets," *Combinatorica*, to appear.

[9] K. Gyarmati, M. Matolcsi, and I. Z. Ruzsa, "Plünnecke's inequality for different summands," *Preprint,* `arXiv:0810:1488v1`, 2008.

[10] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Information and Control*, vol. 36, no. 2, pp. 133–156, 1978.

[11] M. Madiman. On the entropy of sums. In *Proc. IEEE Inform. Theory Workshop.* Porto, Portugal, 2008.

[12] M. Madiman, "Determinant and trace inequalities for sums of positive-definite matrices," *Preprint*, 2008.

[13] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2317–2329, July 2007.

[14] M. Madiman and P. Tetali. Sandwich bounds for joint entropy. *Proceedings of the IEEE International Symposium on Information Theory, Nice*, June 2007.

[15] M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," *Submitted*, 2007.

[16] F. Matús, "Two constructions on limits of entropy functions," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 320–330, 2007.

[17] M. Nathanson, *Additive Number Theory: Inverse Problems and Geometry of Sumsets*, ser. Graduate Texts in Mathematics. Springer, 1996, no. 165.

[18] J. Radhakrishnan. Entropy and counting. In *Computational Mathematics, Modelling and Algorithms* (ed. J. C. Misra), Narosa, 2003.
See: `http://www.tcs.tifr.res.in/~jaikumar/Papers/EntropyAndCounting.pdf`

[19] I. Z. Ruzsa, "Cardinality questions about sumsets," in *Additive combinatorics*, ser. CRM Proc. Lecture Notes. Providence, RI: Amer. Math. Soc., 2007, vol. 43, pp. 195–205.

[20] I. Z. Ruzsa, "Sumsets and entropy," Random Struct. Algorithms, 34, no. 1, 2009, pp. 1–10.

[21] T. Tao, and V. Vu, “Entropy methods,” *Unpublished note*, 2006.

[22] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Studies in Advanced Mathematics 105, CUP 2006.