MASTER'S DEGREE THESIS

# Master of Science in Advanced Mathematics and Mathematical Engineering

**Title: Entropy methods for sumset inequalities**

**Author: Alberto Espuny Díaz**

**Advisor: Oriol Serra Albo**

**Department: Matemàtiques**

**Academic year: 2015-2016**



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master's Thesis

# Entropy methods for sumset inequalities

Alberto Espuny Díaz

Advisor: Oriol Serra Albo

Departament de Matemàtiques

Para Edu.

# Abstract

**Keywords:** Additive Combinatiorics, Shannon Entropy, Plünecke inequalities.

**MSC2010:** 05D40, 11B13, 94A17

In this thesis we present several analogies betweeen sumset inequalities and entropy inequalities. We offer an overview of the different results and techniques that have been developed during the last ten years, starting with a seminal paper by Ruzsa, and also studied by authors such as Bollobás, Madiman, or Tao. After an introduction to the tools from sumset theory and entropy theory, we present and prove many sumset inequalities and their entropy analogues, with a particular emphasis on Plünnecke-type results. Functional submodularity is used to prove many of these, as well as an analogue of the Balog-Szemerédi-Gowers theorem. Partition-determined functions are used to obtain many sumset inequalities analogous to some new entropic results. Their use is generalized to other contexts, such as that of projections or polynomial compound sets. Furthermore, we present a generalization of a tool introduced by Ruzsa by extending it to a much more general setting than that of sumsets. We show how it can be used to obtain many entropy inequalities in a direct and unified way, and we extend its use to more general compound sets. Finally, we show how this device may help in finding new expanders.

# Resum

**Paraules clau:** Combinatòria additiva, entropia de Shannon, desigualtats de Plünnecke.

**MSC2010:** 05D40, 11B13, 94A17

Aquesta tesi té com a objectiu presentar resultats recents que plantejen analogies entre desigualtats clàssiques entre cardinals de conjunts suma i desigualtats d'entropies de variables aleatòries. Es dona una vista panoràmica dels diferents resultats i tècniques que han aparegut a la literatura els darrers deu anys, motivats per un treball seminal de Ruzsa, que han estat desenvolupades per diversos autors que inclouen Bollobás, Madiman, Ruzsa i Tao. Després d'introduir les nocions bàsiques de combinatòria additiva i de teoria de la entropia, el treball presenta i prova una gran diversitat de teoremes sobre conjunts suma i els seus anàlegs entròpics, amb un èmfasi particular en les desigualtats de Plünecke. La submodularitat funcional es fa servir per provar molts d'aquests resultats, especialment la versió entròpica del teorema de Balog-Szemerédi-Gowers. La noció de funcions determinades per particions es fa servir per obtenir anàlegs entròpics d'altres desigualtats més generals i en particular s'estèn a tractar projeccions. Finalment el treball presenta una nova generalització d'una eina introduïda per Ruzsa que va molt més enllà dels problemes de conjunts suma. Provem com aquesta generalització permet reobtenir molts dels resultats exposats al treball i presentem una aplicació original a la obtenció de versions entròpiques d'expansors polinòmics, que podria proporcionar noves famílies d'aquests expansors.

# Contents

# Introduction

This work tries to put emphasis on an example of a phenomenon that occurs with increasing frequency in the study of mathematics: the intersection of different areas. As the title already suggests, there are two main areas from mathematics that will be studied: sumset theory (additive combinatorics) and entropy theory (information theory).

Additive combinatorics, also known as combinatorial number theory, is now a broad field in mathematics, developed mainly after an approach to the still open Goldbach's conjecture. A lot of intense activity has been carried out in this area in the most recent decades, and it is now very rich in results, tools, and interesting open problems. Among the different problems studied in additive combinatorics, sumset theory concerns itself with the study of the properties (such as size, density, structure) of iterated sums of sets. There are also inverse problems, that try to derive properties of the original sets once a property of their sumset is known. The techniques to tackle these problems come from many distinct areas of mathematics: one can find tools coming from elementary combinatorics, graph theory, number theory, ergodic theory, probability, harmonic analysis, convex geometry, incidence geometry, algebraic geometry, or information theory, among others.

Information theory is an area whose development in the last decades is much due to its applications in the digital era. Related to the study of transmission and reception of signals and information, the mathematical approach has provided many useful tools to other areas of mathematics. In particular, the notion of entropy (introduced by Shannon in order to "measure" the amount of information conveyed by the output of a random variable) has proved to have many applications in other areas; particularly, many results in combinatorics can be proved using simple entropic inequalities, and this has found applications in problems in extremal combinatorics (such as counting independent sets, subgraphs, or graph homomorphisms), algorithm theory, computer science, discrete geometry, or game theory.

There are many classical sumset inequalities that have been thoroughly studied throughout the last forty years. The seminal works of Gregory Freiman, Imre Ruzsa, Ben Green or Terence Tao have

left a wake of results with many applications in more sofisticated problems. An instance of these results is Plünnnecke's inequality, first proved in 1969, and the many Plünnecke-type inequalities that have been developed later. Similarly, many entropy inequalities have arisen from the study of information theory.

Starting with the seminal work by Ruzsa [25], it was shown that several sumset inequalities have an entropic analogue, providing a rich connection which allows for the use of tools from information theory in additive combinatorics, and inspiring a host of new entropic inequalities which have applications in information theory.

The main purpose of this Master's Thesis is to give a general and uniform overview of the many results appeared in this area in the last ten years. One of the motivations is to describe the entropy analogue of Plünnecke-type inequalities, which were thoroughly studied in the authors Bachelor's Thesis. The proofs of the main results which are included in the text, while inspired in the ones in the literature, have been rewritten in a hopefully simpler or clearer way. In this work, furthermore, we exploit the device introduced by Ruzsa by extending it to a more general context. The power of this device is illustrated by deriving most of the results in the literature in a simple and unified way, and by showing new applications in the area of polynomial expanders.

The basis for this study is presented in Chapter 1. The chapter is devoted to giving a brief introduction to some basic concepts of sumset theory and entropy theory. In particular, entropy is defined, and most of its basic properties, which will be used throughout the thesis, are presented and proved. It also serves as an introduction to some entropy theory definitions analogous to some of the classical sumset theory tools.

In Chapter 2 we strive to find entropic results which are analogous to many classical sumset inequalities. In particular, we deal with Ruzsa's triangle inequality, as well as with the Plünnecke-Ruzsa inequalities. Furthermore, we present and prove the very influencial Balog-Szemerédi-Gowers theorem, and give an entropic analogue.

In Chapter 3 we work in the opposite direction. Starting from the well known Han's inequality and Shearer's inequality, we present and prove some projection inequalities which are completely analogous to the entropic inequalities. The entropic inequalities are then generalized in several new ways, obtaining results proved in the last decade. All these entropic inequalities are then used to obtain several sumset inequalities.

Chapter 4 presents an extension of a device instroduced by Ruzsa in [25] which allows for a new unified approach to most of the results presented in previous chapters. The power and flexibility of this framework are also illustrated by giving new applications concerning polynomial expanders and their entropy analogues.

While the study of these inequalities is still recent and many new results may appear in the following years, we strive to give a wide overview of the tools and results that have been presented so far, in such a way that this thesis may serve as an introductory text for anyone interested in these exciting new developments.

**Acknowledgements:** The author would like to express his thanks to Professor Oriol Serra for his help in gathering information and discussing many details of this thesis. He would also like to thank Professor Madiman for several helpful remarks about the content of this work.

This investigation was partially funded by grant 2015 / COLAB / 00069 of the Spanish Ministerio de Educación, Cultura y Deporte.

# Chapter 1
# Preliminaries

The aim of this thesis is to present in a unified way the different analogies that appear between sumset inequalities and entropy inequalities. In order to do so, we shall present the results in both settings, and prove them when possible. Sometimes we will first prove the sumset inequality and then prove the entropy one, and then discuss the analogy between the results. Other times, the entropic results will be used to deduce the set-theoretic inequalities, so the order in which they are presented will be reversed. In both cases it is important to understand the theoretical basis in order to follow through the developments presented here, so let us begin by presenting this basis. It should be noted that some basic concepts about probability theory, such as the definition and some basic properties of random variables, are already assumed to be known by the reader. So are the basic operations of set theory, and some basic concepts from graph theory. For the rest, we will try to present every possible definition, so that this thesis is as self-contained as possible.

First of all, let us start with some basic notation. We shall use uppercase letters $A, B, C$ when talking about sets, and $X, Y, Z$ for random variables. We will write $|A|$ to denote the size of a set $A$. When talking about groups we will use $G, H, K$, whereas $\mathbf{H}$ shall denote entropy. When talking about graphs we shall use greek letters $\Gamma$, $\Delta$. We also denote as $[n]$ the set $\{1, 2, \dots n\}$.

In section 1.1 we present the definitions and ideas of the theory of set addition, while the basic definitions and properties of entropy are presented in section 1.2. Bear in mind, however, that along the thesis there will be many aspects from mathematics that will be used; usually, we will present and define more particular tools only when they are needed.

## 1.1. Introduction to the theory of set addition

The theory of set addition developed as a result of the study of Goldbach's conjecture. Although it did not result in any breakthrough in this aspect and was soon substituted by different approaches,

the interest in this theory lingered, and it has now become a field of intense research, with many interesting open problems. This theory has also recently been extended to the non-commutative case, but here we will deal mainly with additive sets, that is, sets in commutative additive groups $(G, +)$, to which we will refer as the ambient group.

**Definition 1.1.** Let $A$ and $B$ be two sets in a commutative group.

The *sumset* or *Minkowsky sum* of these two sets is

$$A + B = \{a + b : a \in A, b \in B\}.$$

A particular case of the sumsets occurs when adding a singleton to another set. In this case, what we have is a translation of the set, and we write

$$\{a\} + B = a + B.$$

The iterated $h$-fold sumset will be denoted as $hA$. It can be recursively defined as

$$hA = (h-1)A + A = A + A + \overset{h)}{\ldots} + A.$$

The inverse of a set $A$ is the set of the inverses of $A$, and can be denoted as

$$-A = \{-a : a \in A\}.$$

Then, one can easily define the *difference set* as

$$A - B = \{a + b : a \in A, b \in -B\} = \{a - b : a \in A, b \in B\}.$$

In general, we may write

$$kA - lB = \{a_1 + \ldots + a_k - b_1 - \ldots - b_l : a_i \in A, b_j \in B\}.$$

When dealing with different problems, we may use a more compact notation. For instance, assume that we have sets $A_1, \ldots, A_n$ in our additive ambient group. Then, for any set of indices $S \subseteq [n]$ we will write $A_S^+ = \sum_{i \in S} A_i$.

The main goal of the theory of set addition is to understand the properties of sets that are being operated with respect to each other. For example, many of the problems in additive combinatorics concern themselves with bounding the cardinalities of sumsets. Some trivial sharp bounds can be found, but this problem becomes interesting when considering stronger hypothesis about the sets that are being added. A clear example of this are the Plünnecke-Ruzsa inequalities (see Theorem 2.8, Theorem 2.9 and Theorem 2.10, or have a look at [8]), in which bounds for iterated

sum-and-difference sets are stated. Very often, the problems that concern themselves with bounding cardinalities of iterated sumsets do so by considering what is called the *doubling constant* of a set.

**Definition 1.2.** Given a finite set $A$ in a commutative group, its *doubling constant* is defined as the ratio

$$\sigma[A] = \frac{|A+A|}{|A|}.$$

Similarly, its *difference constant* is given by

$$\delta[A] = \frac{|A-A|}{|A|}.$$

When $\sigma[A]$ is "small" (or constant), these can be understood as a way to measure the "additive structure" of a set. These simple concepts can be somewhat generalized using the *Ruzsa distance*.

**Definition 1.3.** The *Ruzsa distance* between two sets $A$ and $B$ in a commutative group is given by

$$d_R(A,B) = \log \frac{|A-B|}{\sqrt{|A||B|}}.$$

The Ruzsa distance is not an actual distance as, in general, $d_R(A,A) > 0$. However, the symmetric and positive properties hold, as well as the triangle inequality, so thinking about this as a metric sometimes gives a good insight on the problems. Furthermore, observe that $d_R(A,-A) = \log \sigma[A]$ and $d_R(A,A) = \log \delta[A]$, which explains that this is somewhat a generalization of the doubling and difference constants. In this way, it can be understood as a measure of the amount of common additive structure between the two sets.

The kind of problems that try to find properties of a "higher order" sumset when some condition on the sets is imposed (as happens with the Plünnecke-Ruzsa inequalities) are called direct problems. A different type of problems adressed in additive combinatorics are the so called inverse problems. Often, the assumptions that are considered so that better bounds for iterated sumsets can be obtained are related to the structure of the sets that are being added. Inverse problems try to find converse results: if a bound for sumsets is known, can something be said about the structure of the sets? There are many very interesting and deep results in this area, and some entropy inverse theorems have been recently presented. However, throughout this thesis we will mainly concern ourselves with direct problems. Should the reader be interested, we recommend to check [31] and [16]. As for more information about additive combinatorics, we defer the reader to [32].

## 1.2. Introduction to Shannon entropy

Entropy and mutual information are widely used properties, whose study comes from information theory. They are defined as functionals of a probability distribution, and characterize the behaviour of random variables.

**Definition 1.4.** Given a discrete random variable $X$ with a probability mass function $p(x)$, its entropy is defined as

$$\mathbf{H}(X) = -\sum_x p(x) \log p(x).$$

It is interesting to notice that the above expression can be thought of as the expectation of a function of the random variable $X$.

In a way, the entropy is a measure of the uncertainty of a random variable, of the amount of information that its output conveys. It can also be thought of as the "measure of the surprise" of the outcome of a random variable. Entropy became usual with the development of information theory, and its use has been generalized in many different areas of mathematics. For instance, many results in combinatorics can be proved using an entropic approach. In this section, we give a brief overview of the basic properties of entropy, as well as mutual information, presenting only those that will be useful in the remainder of this work.

### 1.2.1. The entropy function

As was already established, the entropy of a random variable $X : \mathcal{X} \to G$ with probability mass function $p_X(x)$, where $G$ is a group, is defined as

$$\mathbf{H}(X) = \mathbb{E}\left[-\log p_X(x)\right],$$

where log denotes the natural logarithm (by convention) and $x \in X(\mathcal{X})$ (notice that, in fact, the space where $X$ takes values plays no role in the value of its entropy, which only depends on the distribution of probabilities). In most of the bibliography, the logarithms are taken in base 2; this, however, only results in a rescaling of the numerical results, which is something we pay no mind to throughout this thesis. If $X$ is a discrete random variable, we may expand this expression, which can be written as

$$\mathbf{H}(X) = -\sum_{x \in X(\mathcal{X})} p_X(x) \log p_X(x) = \sum_{x \in X(\mathcal{X})} p_X(x) \log \frac{1}{p_X(x)} = \sum_{x \in X(\mathcal{X})} F(p_X(x)),$$

where $F : \mathbb{R}^+ \to \mathbb{R}^+$ is the function $F(x) = x \log \frac{1}{x}$. As a convention, $F(0) = 0$ is taken. In such a way, adding terms with probability 0 does not change the entropy of the random variable.

This entropy function is crutial for the development of entropy theory, so here we concentrate on showing some of its most basic properties.

One of the first things one may notice is that $F$ is nonnegative in the range we are working on. Indeed, it is easy to check that $\log \frac{1}{x} > 0$ when $0 < x < 1$. This means that $\mathbf{H}(X) \geq 0$ for any random variable, and this is a key property that will be used once and again.

The first derivative of the entropy function is

$$F'(x) = \log \frac{1}{x} + x\frac{-1}{x^2}x = \log \frac{1}{x} - 1,$$

and the second derivative is

$$F''(x) = -\frac{1}{x^2}x = -\frac{1}{x},$$

which is negative for all positive values of $x$. Hence, $F$ is a concave function, and from this simple fact many properties can be obtained. For instance, we know that it has at most one maximum. This maximum is achieved when the first derivative is zero, so we have that

$$F'(x) = 0 \implies \log \frac{1}{x} - 1 = 0 \implies \log \frac{1}{x} = 1 \implies x = \frac{1}{e}.$$

We can then state that $F$ is increasing for $x < \frac{1}{e}$, decreasing for $x > \frac{1}{e}$, and we have the bound

$$F(x) \leq F\left(\frac{1}{e}\right) = \frac{1}{e}$$

(notice that if the entropy was defined in base 2, the maximum would be achieved for $x = \frac{1}{2}$, which is a widely known property). Considering the tangent lines to the function, we also have the bound

$$F(y) \leq F(x) + F'(x)(y - x) \qquad \forall\ x > 0,\, y \geq 0. \tag{1}$$

Similarly, because it is a concave function, we have the subadditivity property.

**Lemma 1.1 (Subadditivity).** *For any positive reals $x$ and $y$, the following inequality holds:*

$$F(x + y) \leq F(x) + F(y).$$

*Proof.* Remember that a function $f$ is said to be concave if $tf(a) + (1 - t)f(b) \leq f(ta + (1 - t)b)$ for all $a$, $b$ in the range of the function and for all $t \in [0, 1]$. As such, taking $b = 0$ we have that $F(ta) = F(ta + (1 - t)0) \geq tF(a) + (1 - t)F(0) = tF(a)$. Then,

$$\begin{aligned} F(x) + F(y) &= F\left((x + y)\frac{x}{x + y}\right) + F\left((x + y)\frac{y}{x + y}\right) \\ &\geq \frac{x}{x + y}F(x + y) + \frac{y}{x + y}F(x + y) = F(x + y). \end{aligned}$$

□

In particular, we have the triangle inequality

$$|F(a) - F(b)| \leq F(|a-b|) \tag{2}$$

for all $0 \leq a, b \leq \frac{1}{e}$. Indeed, assume without loss of generality that $a > b$. Then, taking $x = a - b$ and $y = b$ in Lemma 1.1, we have the desired result.

Finally, consider the identity

$$F(ax) = ax \log \frac{1}{a} + ax \log \frac{1}{x} = a \log \frac{1}{a} x \log \frac{1}{x} \left( \frac{1}{\log \frac{1}{a}} + \frac{1}{\log \frac{1}{x}} \right) = F(a)F(x) \left( \frac{1}{\log \frac{1}{a}} + \frac{1}{\log \frac{1}{x}} \right).$$

One can easily check that

$$\frac{1}{\log \frac{1}{a}} + \frac{1}{\log \frac{1}{x}} \leq 2$$

for all values of $a$ and $x$ such that $0 \leq a, x \leq \frac{1}{e}$, so from this we get the inequality

$$F(ax) \leq 2F(a)F(x).$$

**Lemma 1.2 (Jensen bound).** *Let $A$ be a finite set in a group, and let $X$ be an $A$-random variable. Then,*

$$\mathbf{H}(X) \leq \log |A|.$$

*Proof.* Simply check, using (1), that

$$\begin{aligned} \mathbf{H}(X) &= \sum_{x \in A} F(p_X(x)) \leq \sum_{x \in A} \left[ F\left(\frac{1}{|A|}\right) + F'\left(\frac{1}{|A|}\right)\left(p_X(x) - \frac{1}{|A|}\right)\right] \\ &= |A| \frac{1}{|A|} \log |A| + (\log |A| - 1) \sum_{x \in A} \left( p_X(x) - \frac{1}{|A|} \right) = \log |A|. \end{aligned}$$

□

### 1.2.2. Conditional entropy

Given a discrete random variable $X$, let $p_X$ denote its probability distribution, and consider an event $E$. We define the conditioned random variable $(X|E)$ by restricting the probability measure to $E$ and dividing by the measure of $E$ in order to normalize the result. The probability distribution can be written as

$$p_{(X|E)}(x) = \frac{\Pr(x \in X \wedge E)}{\Pr(E)}.$$

If $E$ is an event of the form $X \in A$ for some finite set $A$, then

$$p_{X|X\in A}(x) = \frac{\mathbb{1}_A(x)p_X(x)}{\sum_{y\in A} p_X(y)},$$

where $\mathbb{1}_A$ is the characteristic function of $A$. With this, one can define the entropy of a conditioned random variable in the same way as before, which reults in

$$\mathbf{H}(X|Y=y) = \sum_{x\in X(\mathcal{X})} \Pr(X=x|Y=y)\log\frac{1}{\Pr(X=x|Y=y)}.$$

Then, the conditional entropy is defined as

$$\mathbf{H}(X|Y) = \sum_{y\in\mathrm{range}(Y)} p_Y(y)\mathbf{H}(X|Y=y). \tag{3}$$

There is a simple formula relating the conditional entropy to normal entropies,

**Lemma 1.3.** *Given two discrete random variables $X$ and $Y$,*

$$\mathbf{H}(X|Y) = \mathbf{H}(X,Y) - \mathbf{H}(Y).$$

*Proof.* Indeed, we have that

$$\begin{aligned}
\mathbf{H}(X|Y) &= \sum_{y\in\mathrm{range}(Y)} p_Y(y)\mathbf{H}(X|Y=y) \\
&= \sum_{y\in\mathrm{range}(Y)} p_Y(y) \sum_{x\in\mathrm{range}(X)} \Pr(X=x|Y=y)\log\frac{1}{\Pr(X=x|Y=y)} \\
&= \sum_{y\in\mathrm{range}(Y)} \sum_{x\in\mathrm{range}(X)} \Pr(X=x,Y=y)\log\frac{\Pr(Y=y)}{\Pr(X=x,Y=y)} \\
&= \sum_{(x,y)\in\mathrm{range}((X,Y))} \Pr(X=x,Y=y)\log\frac{1}{\Pr(X=x,Y=y)} + \sum_{(x,y)\in\mathrm{range}((X,Y))} \Pr(X=x,Y=y)\log\Pr(Y=y) \\
&= \mathbf{H}(X,Y) - \sum_{(x,y)\in\mathrm{range}((X,Y))} \Pr(X=x,Y=y)\log\frac{1}{\Pr(Y=y)} \\
&= \mathbf{H}(X,Y) - \sum_{y\in\mathrm{range}(Y)} \log\frac{1}{\Pr(Y=y)} \sum_{x\in\mathrm{range}(X)} \Pr(X=x,Y=y) \\
&= \mathbf{H}(X,Y) - \sum_{y\in\mathrm{range}(Y)} \log\frac{1}{\Pr(Y=y)} \Pr(Y=y) = \mathbf{H}(X,Y) - \mathbf{H}(Y).
\end{aligned}$$

□

In particular, we have that

$$\mathbf{H}(X, Y|Y) = \mathbf{H}(X, Y, Y) - \mathbf{H}(Y) = \mathbf{H}(X, Y) - \mathbf{H}(Y) = \mathbf{H}(X|Y). \tag{4}$$

One can write a similar statement when considering three random variables.

**Lemma 1.4.** *Let $X$, $Y$ and $Z$ be three discrete random variables. Then,*

$$\mathbf{H}(X, Y|Z) = \mathbf{H}(X|Y, Z) + \mathbf{H}(Y|Z).$$

*Proof.* Indeed, applying Lemma 1.3 several times we can see that

$$\begin{aligned}\mathbf{H}(X, Y, Z) &= \mathbf{H}(X, Y|Z) + \mathbf{H}(Z),\\ \mathbf{H}(X, Y, Z) &= \mathbf{H}(X|Y, Z) + \mathbf{H}(Y, Z) = \mathbf{H}(X|Y, Z) + \mathbf{H}(Y|Z) + \mathbf{H}(Z).\end{aligned}$$

□

Applying this several times results in the *chain rule for entropy*.

**Lemma 1.5 (Chain rule for entropy).** *Let $X_1, \dots, X_n$ be $n$ discrete random variables. Then,*

$$\mathbf{H}(X_1, \dots, X_n) = \mathbf{H}(X_1) + \mathbf{H}(X_2|X_1) + \mathbf{H}(X_3|X_2, X_1) + \dots + \mathbf{H}(X_n|X_{n-1}, \dots, X_1).$$

As happened with the notation for sumsets, there is a compact notation when dealing with joint random variables. In this sense, if we have random variables $X_1, \dots, X_n$, for each subset of indices $S \subseteq [n]$ we will write $X_S = (X_i : i \in S)$. We take as a convention the notation $X_\varnothing$ to not be a random variable, in which case the entropy is, by definition, null. Using this notation, the chain rule can be rewritten as

$$\mathbf{H}(X_1, \dots, X_n) = \sum_{i=1}^{n} \mathbf{H}(X_i|X_{[i-1]}).$$

One can also use the total probability formula

$$p_X(x) = \sum_{y \in \text{range}(Y)} p_Y(y) p_{X|Y=y}(x)$$

to obtain the following result.

**Lemma 1.6 (Dropping condition).** *Let $X$ and $Y$ be two discrete random variables. Then,*

$$\mathbf{H}(X|Y) \leq \mathbf{H}(X).$$

*Proof.* Indeed, we have

$$\begin{aligned}\mathbf{H}(X) &= \sum_{x\in\text{range}(X)} F(p_X(x)) = \sum_{x\in\text{range}(X)} F\left(\sum_{y\in\text{range}(Y)} p_Y(y)p_{X|Y=y}(x)\right)\\ &\geq \sum_{x\in\text{range}(X)}\sum_{y\in\text{range}(Y)} p_Y(y)F(p_{X|Y=y}(x)) = \sum_{y\in\text{range}(Y)} p_Y(y)\mathbf{H}(X|Y=y) = \mathbf{H}(X|Y),\end{aligned}$$

where the inequality comes from using Jensen's inequality with the average weighted by the $p_Y(y)$, and the concavity of $F$, and the last equality corresponds to (3). □

Equality holds if, and only if, $p_{X|Y=y}(x) = p_X(x)$ for all $y \in$ range$(Y)$, $x \in$ range(X), that is, if, and only if, $(X|Y = y) \equiv X$ for all $y \in$ range$(Y)$. This is equivalent to saying that $X$ and $Y$ are independent. Combining this and Lemma 1.3, we have that

$$\mathbf{H}(X) \geq \mathbf{H}(X|Y) = \mathbf{H}(X,Y) - \mathbf{H}(Y) \implies \mathbf{H}(X,Y) \leq \mathbf{H}(X) + \mathbf{H}(Y), \tag{5}$$

with equality if, and only if, $X$ and $Y$ are independent random variables.

When considering the entropy of joint distributions, one can now prove the following lemma.

**Lemma 1.7 (Submodularity).** *Let $X = (X_1, \ldots, X_n)$ be the joint distribution of $n$ random variables. Then, $\mathbf{H}(X_S)$ is a submodular function of the set $S$, that is, given two sets $S, T \subseteq [n]$,*

$$\mathbf{H}(X_{S\cup T}) + \mathbf{H}(X_{S\cap T}) \leq \mathbf{H}(X_S) + \mathbf{H}(X_T).$$

*Proof.* By the dropping condition Lemma 1.6,

$$\mathbf{H}(X_{T\setminus S}|X_S) \leq \mathbf{H}(X_{T\setminus S}|X_{S\cap T}).$$

Using Lemma 1.3, we have that

$$\mathbf{H}(X_{T\cup S}) - \mathbf{H}(X_S) \leq \mathbf{H}(X_T) - \mathbf{H}(X_{S\cap T}),$$

and the result follows by reordering the terms. □

We say that a discrete random variable $Y$ is determined by another discrete random variable $X$ if there is a function $f$ : range$(X) \to$ range$(Y)$ such that $Y = f(X)$.

**Lemma 1.8.** *Let $X$ and $Y$ be two discrete random variables such that $Y$ is determined by $X$. Then,*

$$\mathbf{H}(Y) \leq \mathbf{H}(X).$$

*Proof.* Let $Y = f(X)$. We know that $\mathbf{H}(X, f(X)) = \mathbf{H}(X|f(X)) + \mathbf{H}(f(X)) = \mathbf{H}(f(X)|X) + \mathbf{H}(X)$. However, $\mathbf{H}(f(X)|X) = 0$ since $f(X)$ is determined by $X$ $\left( p_{Y|X=x}(y) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases} \right)$. The result follows because entropy is nonnegative. □

An important particular case of this comes when considering the random variable $(X, Y)$, which determines both $X$ and $Y$. Therefore,

$$\max\{\mathbf{H}(X), \mathbf{H}(Y)\} \leq \mathbf{H}(X, Y).$$

Then, using Lemma 1.3 and Lemma 1.6 we have that

$$\mathbf{H}(X) - \mathbf{H}(Y) \leq \mathbf{H}(X, Y) - \mathbf{H}(Y) = \mathbf{H}(X|Y) \leq \mathbf{H}(X).$$

Furthermore, if $X$ determines $Y$ then $X$ and $(X, Y)$ determine each other, and therefore $\mathbf{H}(X, Y) = \mathbf{H}(X)$. In this particular case we have that $\mathbf{H}(X|Y) = \mathbf{H}(X) - \mathbf{H}(Y)$ and $\mathbf{H}(Y|X) = 0$.

Now we can state a lemma which will come in useful when obtaining entropy analogues of set inequalities. It is a particuar form of the submodularity that is present in entropy.

**Lemma 1.9 (Functional submodularity inequality).** *Let $X_0$, $X_1$, $X_2$ and $X_{12}$ be discrete random variables such that $X_1$ and $X_2$ each determine $X_0$ and $(X_1, X_2)$ determines $X_{12}$. Then,*

$$\mathbf{H}(X_{12}) + \mathbf{H}(X_0) \leq \mathbf{H}(X_1) + \mathbf{H}(X_2).$$

*Proof.* As $X_{12}$ is determined by $(X_1, X_2)$, we also have that $X_{12}|X_0 = x$ is determined by $(X_1|X_0 = x, X_2|X_0 = x)$ for each possible value of $x \in \text{range}(X_0)$. By Lemma 1.8 and (5), this means that

$$\mathbf{H}(X_{12}|X_0 = x) \leq \mathbf{H}(X_1|X_0 = x, X_2|X_0 = x) \leq \mathbf{H}(X_1|X_0 = x) + \mathbf{H}(X_2|X_0 = x) \quad \forall\, x \in \text{range}(X_0).$$

Applying the definition of conditional entropy (3) we have that

$$\mathbf{H}(X_{12}|X_0) \leq \mathbf{H}(X_1|X_0) + \mathbf{H}(X_2|X_0)$$

and, by Lemma 1.3 and Lemma 1.6,

$$\mathbf{H}(X_{12}|X_0) = \mathbf{H}(X_{12}, X_0) - \mathbf{H}(X_0) \geq \mathbf{H}(X_{12}) - \mathbf{H}(X_0).$$

As $X_1$ and $X_2$ each determine $X_0$, we have that $\mathbf{H}(X_1|X_0) = \mathbf{H}(X_1) - \mathbf{H}(X_0)$ and $\mathbf{H}(X_2|X_0) = \mathbf{H}(X_2) - \mathbf{H}(X_0)$. Putting all of this together we have that

$$\mathbf{H}(X_{12}) - \mathbf{H}(X_0) \leq \mathbf{H}(X_{12}|X_0) \leq \mathbf{H}(X_1) + \mathbf{H}(X_2) - 2\mathbf{H}(X_0),$$

and the result follows by reordering the terms. □

### 1.2.3. Mutual information

The mutual information between two random variables measures the amount of information one of them contains about the other. Conversely, one can think of it as the reduction in the uncertainty of one due to the knowledge of the other. In this sense, it can be understood as a measure of their dependence. When given two random variables $X$ and $Y$, their mutual information is defined as

$$\mathbf{I}(X;Y) = \mathbf{H}(X) - \mathbf{H}(X|Y). \tag{6}$$

In fact, the definition is symmetric: by Lemma 1.3 we have that

$$\mathbf{I}(X;Y) = \mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{H}(X) - \mathbf{H}(X,Y) + \mathbf{H}(Y) = \mathbf{H}(Y) - \mathbf{H}(Y|X) = \mathbf{I}(Y;X).$$

The mutual information between two random variables is nonnegative, as a direct consequence of the dropping condition Lemma 1.6. In particular, one has that $\mathbf{I}(X;Y) = 0$ if, and only if, $X$ and $Y$ are independent. Also, notice that $\mathbf{I}(X;X) = \mathbf{H}(X)$, so the entropy is just the mutual information of a random variable and itself. In this sense, mutual information is just a generalization of the concept of entropy.

As happens with entropy, one can also define conditional mutual information in an analogous way. One has that

$$\mathbf{I}(X;Y|Z) = \mathbf{I}((X,Z);Y) - \mathbf{I}(Z;Y) \tag{7}$$

By applying (6) and Lemma 1.3 we have

$$\mathbf{I}(X;Y|Z) = \mathbf{H}(X,Z) - \mathbf{H}(X,Z|Y) - \mathbf{H}(Z) + \mathbf{H}(Z|Y) = \mathbf{H}(X,Z) + \mathbf{H}(Y,Z) - \mathbf{H}(X,Y,Z) - \mathbf{H}(Z).$$

It is interesting to note that this quantity is nonnegative as a simple consequence of Lemma 1.7 (submodularity).

One of the classical properties of mutual information, in a similar spirit to Lemma 1.8, is known as the data processing inequality.

**Lemma 1.10 (Data processing inequality).** *The mutual information cannot increase when looking at functions of the random variables. That is,*

$$\mathbf{I}(f(X);Y) \leq \mathbf{I}(X;Y).$$

*Proof.* By the symmetry of mutual information, we may write

$$\begin{aligned}\mathbf{I}((X,f(X));Y) &= \mathbf{I}(X;Y|f(X)) + \mathbf{I}(f(X);Y)\\ &= \mathbf{I}(f(X);Y|X) + \mathbf{I}(X;Y).\end{aligned}$$

Now observe that $Y$ and $f(X)$ are conditionally independent given $X$. Indeed, one has that

$$\begin{aligned}\Pr(Y=y, f(X)=z|X=x) &= \frac{\Pr(Y=y, X=x, f(X)=z)}{\Pr(X=x)} \\ &= \frac{\Pr(Y=y, X=x)\Pr(f(X)=z|X=x)}{\Pr(X=x)} \\ &= \Pr(Y=y|X=x)\Pr(f(X)=z|X=x),\end{aligned}$$

where we have used twice the definition of conditional probabilities, and the second equality comes from the fact that $f(X)$ is completely determined by $X$, so $f(X) = z$ can only occur if $z = f(x)$. With this, $\mathbf{I}(f(X); Y|X) = 0$ and we can write

$$\mathbf{I}(X;Y) = \mathbf{I}(X;Y|f(X)) + \mathbf{I}(f(X);Y) \geq \mathbf{I}(f(X);Y)$$

by the nonnegativity of mutual information. □

For further information about entropy and mutual information, we refer the reader to the vast bibliography about this topic. For instante, [7] or [3] provide detailed proofs and more careful explanations.

### 1.2.4. Analogues of set addition theory

In set theory, we work with the cardinalities of sets. In the entropy setting, we must first find an object with which we can find a correspondence, and this object is a random variable. Instead of size, we will consider its entropy. We shall see that this approach allows us to obtain many results in this setting that are perfectly analogous to those in the addition set setting. Usually, one thinks about random variables defined over the reals; however, in order to obtain analogies to general set cardinality inequalities, we must be able to define random variables over general sets.

**Definition 1.5.** Given any arbitrary set $A$, we will say that $X$ is an $A$-random variable if it takes values in a finite subset $\{x \in A : \Pr(X = x) \neq 0\} \subseteq A$.

Now, one can see there is a certain analogy between the size of sumsets and the entropy of $G$-random variables. Although they are clearly different, many ideas from one setting can be applied to the other. For example, one may define doubling constants and the Ruzsa distance.

**Definition 1.6.** Let $G$ be an additive group, and let $X$ be a $G$-random variable. Then, its *doubling constant* is defined as

$$\sigma[X] = e^{\mathbf{H}(X_1+X_2)-\mathbf{H}(X)},$$

where $X_1$ and $X_2$ are independent copies of $X$.

Observe that $\sigma[X] \geq 1$. This is a consequence of the fact that, if $X$ and $Y$ are independent random variables, $\mathbf{H}(X+Y) \geq \max\{\mathbf{H}(X), \mathbf{H}(Y)\}$, which will be proved in Lemma 2.2. In this sense, we have a first analogy, as one can trivially see that $\sigma[A] \geq 1$.

Furthermore, one can prove that, if $X$ is uniformly distributed over a set $A$, then $\sigma[X] \leq \sigma[A]$. Indeed, let $X$ be a uniform $A$-random variable, and let $X_1$ and $X_2$ be two independent copies of $X$. Observe that $\operatorname{range}(X_1 + X_2) \subseteq \operatorname{range}(X_1) + \operatorname{range}(X_2)$ holds for any random variables (not necessarily uniform). Then we have

$$\begin{aligned}\sigma[X] = e^{\mathbf{H}(X_1+X_2)-\mathbf{H}(X)} = \frac{e^{\mathbf{H}(X_1+X_2)}}{e^{\mathbf{H}(X)}} &\leq \frac{|\operatorname{range}(X_1+X_2)|}{|\operatorname{range}(X)|} \\ &\leq \frac{|\operatorname{range}(X)+\operatorname{range}(X)|}{|\operatorname{range}(X)|} = \frac{|A+A|}{|A|} = \sigma[A].\end{aligned}$$

where the first inequality holds because of Lemma 1.2 and the fact that $e^{\mathbf{H}(X)} = |\operatorname{range}(X)| = |A|$, since $X$ is uniform.

We thus have that the doubling constant of uniform random variables ranges between the same values as that of sets. However, it is important to note that it can be significantly smaller; one can construct many examples where this fact is observed.

One can also similarly define the Ruzsa distance for random variables.

**Definition 1.7.** Given two $G$-random variables $X$ and $Y$, where $G$ is an additive group, their *Ruzsa distance* is defined as

$$d_R(X,Y) = \mathbf{H}(X'-Y') - \frac{1}{2}(\mathbf{H}(X)+\mathbf{H}(Y)),$$

where $X'$ and $Y'$ are independent copies of $X$ and $Y$.

Again, this is not a metric, but it is nonnegative, symmetric, and a triangle inequality holds. Furthermore, we can observe that $d_R(X,-X) = \log \sigma[X]$, as happens with the Ruzsa distance defined over sets. As above, we can also prove that the entropy Ruzsa distance is upper bounded by the set Ruzsa distance when the random variables are uniform. Indeed, let $X$ be a uniform $A$-random variable, and let $Y$ be a uniform $B$-random variable. Then,

$$\begin{aligned}e^{d_R(X,Y)} = \frac{e^{\mathbf{H}(X'-Y')}}{e^{\frac{1}{2}(\mathbf{H}(X)+\mathbf{H}(Y))}} &= \frac{e^{\mathbf{H}(X'-Y')}}{\sqrt{e^{\mathbf{H}(X)}e^{\mathbf{H}(Y)}}} \\ &\leq \frac{|\operatorname{range}(X'-Y')|}{\sqrt{|\operatorname{range}(X)||\operatorname{range}(Y)|}} \leq \frac{|\operatorname{range}(X)-\operatorname{range}(Y)|}{\sqrt{|\operatorname{range}(X)||\operatorname{range}(Y)|}} = \frac{|A-B|}{\sqrt{|A||B|}} = e^{d_R(A,B)}.\end{aligned}$$

Again, the inequality comes from Lemma 1.2.

# Chapter 2
# Entropy analogues of sumset inequalities

This chapter's main goal is to show how one can prove many analogues of sumset inequalities in the context of random variables, by using entropy instead of sizes. There is not a standard way to do so, although very often there is a certain resemblance between the proofs of the sumset inequalities and their entropic versions. In this chapter we present several sumset inequalities, together with their proofs, and then present and prove their analogous versions in the entropy setting. Most of these entropy analogues were proved by Tao in [31]. In many cases, the entropy analogues to sumset inequalities can be used to prove the original sumset ones by choosing uniform distributions. This shows that the entropy counterparts are usually stronger and more general statements. This is one of the main values of the entropic approach.

In section 2.1 we present some of the traditional sumset results, as well as their entropy counterparts. The study of the very influential Balog-Szemerédi-Gowers theorem and its entropy analogue is presented in section 2.2

## 2.1. Entropy analogues of some basic sumset inequalities

We start by presenting some of the most basic sumset inequalities, and finding their corresponding equivalent statement in the entropy setting. First, one may consider the trivial bounds, which come from simple counting of the total number of possible sums and from the fact that adding a set to another cannot decrease the size of any of them.

**Lemma 2.1.** *Let $A, B \subseteq G$ be two finite sets, where $G$ is a group. Then,*

$$\max\{|A|, |B|\} \leq |A + B| \leq |A||B|.$$

This trivial bound can be found in the setting of random variables and entropies.

**Lemma 2.2.** *Let $X$ and $Y$ be two G-random variables. Then,*

$$\mathbf{H}(X+Y) \leq \mathbf{H}(X) + \mathbf{H}(Y).$$

*Furthermore, if $X$ and $Y$ are independent, then*

$$\max\{\mathbf{H}(X), \mathbf{H}(Y)\} \leq \mathbf{H}(X+Y).$$

*Proof.* The first part can be proved simply by observing that $(X, Y)$ determines $X + Y$. As a consequence, by Lemma 1.8, we know that $\mathbf{H}(X+Y) \leq \mathbf{H}(X, Y)$, and by (5), $\mathbf{H}(X, Y) \leq \mathbf{H}(X) + \mathbf{H}(Y)$.

The proof of the second result is somewhat different. By Lemma 1.6 we know that conditioning does not increase the entropy, so $\mathbf{H}(X+Y|Y) \leq \mathbf{H}(X+Y)$. Then, observe that $\Pr(X+Y=z|Y=y) = \Pr(X=z-y|Y=y)$ for all $z \in \text{range}(X+Y|Y=y)$, so their distributions are the same and $\mathbf{H}(X+Y|Y) = \mathbf{H}(X|Y)$. Finally, using independence in Lemma 1.6 we have that $\mathbf{H}(X|Y) = \mathbf{H}(X)$. The same can be done conditioning to $X$, which finishes the proof. □

It is very easy to see that the entropic inequalities in Lemma 2.2 imply the sumset inequalities in Lemma 2.1. For the first one, for each element $c \in A+B$ consider a unique pair of representatives $(a, b)_c$ such that $a + b = c$ (one may take whichever; for instance, the lexicographically minimal one), and let $Z = (X, Y)$ be a random variable that takes each value $(a, b) \in A \times B$ with probability $\frac{1}{|A+B|}$ if it is one of the representatives, and 0 otherwise. In such a way, $X + Y$ is uniformly distributed over $A + B$. This way of constructing uniform distributions will be used many times throughout this thesis. By Lemma 2.2,

$$\log|A+B| = \mathbf{H}(X+Y) \leq \mathbf{H}(X) + \mathbf{H}(Y) \leq \log|A| + \log|B|,$$

where the last inequality comes from Lemma 1.2 and the fact that the range of $X$ is contained in $A$, and similarly for $Y$. For the second inequality, let $X$ be a uniform random variable ranging over $A$, and $Y$ a uniform random variable ranging over $B$, such that they are independent. By Lemma 2.2,

$$\log\max\{|A|, |B|\} = \max\{\mathbf{H}(X), \mathbf{H}(Y)\} \leq \mathbf{H}(X+Y) \leq \log|A+B|.$$

The trivial bounds follow by exponentiating.

In this way, many sumset inequalities can be proved starting from entropy inequalities. This will be a recurrent technique throughout this thesis. The idea for this technique is the fact that, given a set $A$, the random variable that best describes the set is the uniform one. Indeed, if $|A| = n$ and we take a random variable $X$ defined with $\Pr(X = a) = \dfrac{1}{n}$ for all $a \in A$, we have that

$$\mathbf{H}(X) = \sum_{a \in A} \frac{1}{n} \log n = \log n = \log|A|$$

or, equivalently, $|A| = e^{\mathbf{H}(X)}$. Hence, one may establish this correspondence between sets and random variables. Note, however, that there is not such a straightforward way to go from general random variables to sets; we will refer to this issue in Chapter 3.

We now turn our attention towards Ruzsa's triangle inequality. First shown by Ruzsa in [24] while working on commutative groups, it has proved to be a very useful tool to find bounds for sumsets in both the commutative and noncommutative settings. Throughout this thesis, we will work mainly in the commutative setting; however, it is interesting to see how this result works in the more general case.

**Theorem 2.3 (Ruzsa's triangle inequality).** *Let $A$, $B$ and $C$ be finite non-empty sets in a (not necessarily commutative) group. Then,*

$$|A||B-C| \leq |B-A||A-C|.$$

*Proof.* The idea of the proof is to find an injection between $A \times (B-C)$ and $(B-A) \times (A-C)$. Since the sizes of these sets are $|A||B-C|$ and $|B-A||A-C|$, respectively, finding such an injection immediately yields the result.

Consider the following map:

$$\begin{aligned} \varphi : A \times (B-C) &\longrightarrow (B-A) \times (A-C) \\ (a, b-c) &\longmapsto (b-a, a-c) \end{aligned}$$

We would like to see that this is an injection. First, observe that an element $b-c \in B-C$ may come from different elements $b_1, b_2 \in B$ and $c_1, c_2 \in C$ such that $b_1 - c_1 = b_2 - c_2$. Hence, we must first fix a representation in $B, C$ for each element of $B-C$. We do so by defining an injection

$$f : B-C \longrightarrow B \times C$$

such that $f(x)_1 - f(x)_2 = x \ \ \forall\, x \in B-C$, where $f(x)_1$ denotes the first coordinate of $f(x)$, and $f(x)_2$ denotes the second. Such an injection exists because $|B-C| \leq |B||C|$ (this is the trivial bound, Lemma 2.1). For example, if we give the elements of $B$ some order $b_1, b_2, \ldots, b_k$, we could map $x$ to a pair $(b_i, c_j)$ such that the index $i$ is minimum.

Now, assume that $\varphi(a,x) = \varphi(a', x')$ for some $a, a' \in A$ and $x, x' \in B-C$. Then,

$$\begin{cases} f(x)_1 - a &= f(x')_1 - a', \\ a - f(x)_2 &= a' - f(x')_2. \end{cases}$$

Adding these two equalities, we get that

$$f(x)_1 - f(x)_2 = f(x')_1 - f(x')_2,$$

and since $f$ is an injection by definition, this means that $x = x'$. Substituting this in the former system of equations yields $a = a'$, so $\varphi$ is an injection. □

Observe that the need for commutativity is avoided by the special order in which the operations are made. There are many similar statements that only hold in the commutative case.

Ruzsa's triangle inequality gave rise to the idea of a *distance* in the theory of set addition. This is the so called *Ruzsa distance*, introduced in Chapter 1. Using this distance, Ruzsa's triangle inequality can be written as

$$d_R(B, C) \leq d_R(B, A) + d_R(A, C).$$

The analogue of this distance in the entropy setting was also introduced, and we use it in order to state the following theorem.

**Theorem 2.4.** *Let $X$, $Y$ and $Z$ be three independent $G$-random variables, where $G$ is any (not necessarily commutative) group. Then,*

$$d_R(Y, Z) \leq d_R(Y, X) + d_R(X, Z).$$

*Proof.* By applying the definition of the Ruzsa distance, we find that the statement holds if, and only if,

$$\mathbf{H}(Y - Z) - \frac{1}{2}\left(\mathbf{H}(Y) + \mathbf{H}(Z)\right) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z) - \frac{1}{2}\left(\mathbf{H}(Y) + \mathbf{H}(X) + \mathbf{H}(X) + \mathbf{H}(Z)\right),$$

that is, if

$$\mathbf{H}(X) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z).$$

Now observe that both $(Y - X, X - Z)$ and $(Y, Z)$ determine $Y - Z$, and that they jointly determine $(X, Y, Z)$. Hence, we may apply Lemma 1.9 to obtain

$$\mathbf{H}(X, Y, Z) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X, X - Z) + \mathbf{H}(Y, Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z) + \mathbf{H}(Y, Z),$$

where the second inequality comes from (5). Finally, by applying (5) and taking into account the independence of the random variables we have

$$\mathbf{H}(X) + \mathbf{H}(Y) + \mathbf{H}(Z) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z) + \mathbf{H}(Y) + \mathbf{H}(Z),$$

and the result follows. □

This is a very interesting result, but it depends strongly on the independence of the random variables. This condition can be slightly weakened, as was proved by Ruzsa in [25].

**Theorem 2.5.** *Let $X$, $Y$ and $Z$ be three $G$-random variables such that $X$ is independent of $(Y, Z)$, where $G$ is any (not necessarily commutative) group. Then,*

$$d_R(Y, Z) \leq d_R(Y, X) + d_R(X, Z).$$

*Proof.* Again, it is enough to prove that

$$\mathbf{H}(X) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z).$$

First, consider the special case in which for all $u \in \text{range}(Y - Z)$ there is only one pair $(y, z) \in \text{range}(Y, Z)$ such that $u = y - z$. In this case, $y$ and $z$ are functions of $u$ (as there is only one pair, we have a bijection), say $y = f(u)$, $z = g(u)$. Because of the independence between $X$ and $(Y, Z)$ we have that

$$\mathbf{H}(X, Y - Z) = \mathbf{H}(X) + \mathbf{H}(Y - Z).$$

On the other hand, since

$$Y - Z = (Y - X) + (X - Z)$$

and

$$X = (X - Z) + Z = (X - Z) + g(Y - Z) = (X - Z) + g((Y - X) + (X - Z)),$$

we may now use Lemma 1.8 to obtain that

$$\mathbf{H}(X, Y - Z) \leq \mathbf{H}(Y - X, X - Z).$$

The result follows since, by (5),

$$\mathbf{H}(Y - X, X - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z).$$

Now consider the more general case in which we may have collisions, but where the variables assume only finitely many values. Take a possible value $u \in \text{range}(Y - Z)$ such that $u = y_1 - z_1 = y_2 - z_2$ for different $y_1, y_2 \in \text{range}(Y)$, $z_1, z_2 \in \text{range}(Z)$ and $\Pr(Y = y_i, Z = z_i) > 0$ for $i \in \{1, 2\}$. Let $\Pr(Y = y_1, Z = z_1) = p_1$ and $\Pr(Y = y_2, Z = z_2) = p_2$, and set $p = p_1 + p_2$. Consider the family of variables $Y_t, Z_t$ such that $\Pr(Y_t = y_1, Z_t = z_1) = tp$ and $\Pr(Y_t = y_2, Z_t = z_2) = (1 - t)p$ for each $t \in [0, 1]$, and otherwise having the same distribution as $(Y, Z)$. Then, $Y_t - Z_t$ has the same distribution as $Y - Z$, so the left hand side of the inequality remains constant when substituted by these variables.

The right hand side, however, is a concave function of $t$ (because of the concavity of the entropy function), so its minimum must be achieved either at $t = 0$ or $t = 1$, which corresponds to having only one of the two different representations of $u$ that we were considering. Hence, substituting the initial variables by this minimal pair $(Y_t, Z_t)$ results in a sharper inequality. One can repeat this process as long as there are collisions, which at each step reduces the cardinality of $\text{range}(Y, Z)$ by one; as the variables only assumed finitely many values, this process must end. In the end, this reduces our study to the first special case, which we already proved.

Finally, the general case follows by a routine limiting argument. □

Again, it is now easy to prove the sumset inequality starting from the entropic one. Indeed, let $X$ be a uniform $G$-random variable with range $A$, and take $(Y, Z)$ uniformly distributed over unique representatives of each value of $B - C$, so that $Y - Z$ is uniform. Then,

$$\log|A| + \log|B - C| = \mathbf{H}(X) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z) \leq \log|B - A| + \log|A - C|.$$

Now we turn our attention to a different result, of which we want to find an entropy analogue. It is given by Tao and Vu in [32] (see Cor. 2.12). We refer to this reference for the proof of the result.

**Theorem 2.6.** *Let $A$ be a set in an additive group $G$. Then,*

$$d_R(A, -B) \leq 3\, d_R(A, B).$$

Again, one can prove an entropy theorem analogous to this result. First, however, let us introduce the concept of conditionally independent trials, which will be useful to prove this and several other results.

**Definition 2.1.** Given two random variables $X$ and $Y$, not necessarily independent, we can produce two *conditionally independent trials $X_1, X_2$ of $X$ relative to $Y$*, defined by declaring $(X_1|Y = y)$ and $(X_2|Y = y)$ to be independent trials of $(X|Y = y)$ for each $y \in \text{range}(Y)$.

In particular, this means that $X$, $X_1$ and $X_2$ have the same distribution, and that $X_1$ and $X_2$ are conditionally independent relative to $Y$. Then, by (5),

$$\mathbf{H}(X_1, X_2|Y) = \mathbf{H}(X_1|Y) + \mathbf{H}(X_2|Y) = 2\mathbf{H}(X|Y). \tag{8}$$

Thus, applying Lemma 1.3 results in

$$\mathbf{H}(X_1, X_2, Y) = \mathbf{H}(X_1, X_2|Y) + \mathbf{H}(Y) = 2\mathbf{H}(X, Y) - \mathbf{H}(Y). \tag{9}$$

We can now prove the entropy analogue of Theorem 2.6.

**Theorem 2.7.** *Let $X$ and $Y$ be two independent $G$-random variables, where $G$ is an additive group. Then,*

$$d_R(X, -Y) \leq 3\, d_R(X, Y).$$

*Proof.* By the definition of the Ruzsa distance, we must prove that

$$\mathbf{H}(X - Y) \leq 3\mathbf{H}(X + Y) - \mathbf{H}(X) - \mathbf{H}(Y).$$

Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be conditionally independent trials of $(X, Y)$ relative to $X - Y$. We have that $(X, Y)$ determines $X - Y$, so we conclude that $X_1 - Y_1 = X_2 - Y_2$. Let $(X_3, Y_3)$ be another

independent trial of $(X, Y)$. Then,

$$X_3 + Y_3 = (X_3 - Y_2) - (X_1 - Y_3) + X_2 + Y_1,$$

and we have that $(X_3 - Y_2, X_1 - Y_3, X_2, Y_1)$ and $(X_3, Y_3)$ each determine $X_3 + Y_3$ and together determine $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$. By Lemma 1.9 we have that

$$\mathbf{H}(X_1, X_2, X_3, Y_1, Y_2, Y_3) + \mathbf{H}(X_3 + Y_3) \leq \mathbf{H}(X_3, Y_3) + \mathbf{H}(X_3 - Y_2, X_1 - Y_3, X_2, Y_1). \tag{10}$$

Now we can rewrite or bound each of the previous terms. In each of the bounds we consider the fact that the independent trials and the original random variable are identically distributed. By the independence of $X$ and $Y$ we have that $\mathbf{H}(X_3, Y_3) = \mathbf{H}(X) + \mathbf{H}(Y)$. By the standard entropy inequality (5), $\mathbf{H}(X_3 - Y_2, X_1 - Y_3, X_2, Y_1) \leq 2\mathbf{H}(X - Y) + \mathbf{H}(X) + \mathbf{H}(Y)$. Finally, by the independence hypothesis and (9),

$$\begin{aligned}\mathbf{H}(X_1, X_2, X_3, Y_1, Y_2, Y_3) &= \mathbf{H}(X_1, X_2, Y_1, Y_2) + \mathbf{H}(X_3, Y_3) \\ &= 2\mathbf{H}(X, Y) - \mathbf{H}(X - Y) + \mathbf{H}(X) + \mathbf{H}(Y) = 3\mathbf{H}(X) + 3\mathbf{H}(Y) - \mathbf{H}(X - Y),\end{aligned}$$

as

$$\begin{aligned}\mathbf{H}(X_1, X_2, Y_1, Y_2) &= \mathbf{H}(X_1, X_2, Y_1, Y_2, X - Y) \\ &= 2\mathbf{H}(X, Y, X - Y) - \mathbf{H}(X - Y) = 2\mathbf{H}(X, Y) - \mathbf{H}(X - Y)\end{aligned}$$

because $X - Y$ is determined by $(X, Y)$. Substituting these into (10) yields

$$\mathbf{H}(X + Y) \leq 3\mathbf{H}(X - Y) - \mathbf{H}(X) - \mathbf{H}(Y),$$

as we wanted to see. □

Again, one can obtain a sumset inequality from the entropy one in Theorem 2.7. Indeed, assume that $X$ and $Y$ are distributed over $A$ and $B$, respectively, in such a way that $X - Y$ is uniform (again, this can be done by taking unique representatives in $A \times B$ of each value in $A - B$). Then, by the nonnegativity of the entropy, we trivially have that

$$\log |A - B| = \mathbf{H}(X - Y) \leq 3\mathbf{H}(X + Y) - \mathbf{H}(X) - \mathbf{H}(Y) \leq 3\mathbf{H}(X + Y) \leq 3 \log |A + B|,$$

so $|A - B| \leq |A + B|^3$. We can obtain a better result by using Lemma 2.2, whence we obtain

$$\log |A - B| = \mathbf{H}(X - Y) \leq 3\mathbf{H}(X + Y) - \mathbf{H}(X) - \mathbf{H}(Y) \leq 2\mathbf{H}(X + Y) \leq 2 \log |A + B|,$$

from where the bound is $|A - B| \leq |A + B|^2$. We must emphasize that this inequality, although interesting by itself, is not as strong as the one in Theorem 2.6.

Finally, one can also prove a Plünnecke-Ruzsa-type inequality in the entropic setting. The original sumset theoretic inequalities have a very rich history. The first result was proved by Plünnecke [21], and can be stated as follows.

**Theorem 2.8 (Plünnecke's inequality).** *Let $j$, $h$ be two non-negative integers such that $j < h$, and let $A$ and $B$ be sets in a commutative group. Assume that $|A + jB| \leq \alpha|A|$. Then, there exists a non-empty set $A' \subseteq A$ such that*

$$|A' + hB| \leq \alpha^{\frac{h}{j}}|A'|.$$

In particular, this means that

$$|hB| \leq \alpha^{\frac{h}{j}}|A|. \tag{11}$$

This result was generalised using Ruzsa's triangle inequality. The new result, presented by Ruzsa [22, 23], states the following.

**Theorem 2.9 (Plünnecke-Ruzsa inequality).** *Let $A$ and $B$ be finite sets in a commutative group, and $j$ be a positive integer. Assume that $|A + jB| \leq \alpha|A|$. Then, for any nonnegative integers $k$ and $l$ such that $j \leq \min\{k, l\}$, we have that*

$$|kB - lB| \leq \alpha^{\frac{k+l}{j}}|A|.$$

Later on, further generalizations were presented by Gyarmati, Matolcsi, and Ruzsa [11], arriving to the following very general result.

**Theorem 2.10.** *Let $j$ and $h$ be two positive integers such that $j < h$. Let $A, B_1, \ldots, B_h$ be finite sets in a commutative group. For any $I \subseteq [h]$, let $B_I^+ = \sum_{i \in I} B_i$. For each $B_I^+$, let $\alpha_I$ be a rational number such that $|A + B_I^+| \leq \alpha_I|A|$. Assume that $\alpha_J$ is known for any $J \subseteq [h]$ such that $|J| = j$, and write*

$$\beta = \left( \prod_{J \subseteq [h] : |J| = j} \alpha_J \right)^{\frac{(j-1)!(h-j)!}{(h-1)!}}.$$

*Then, there exists a non-empty set $A' \subseteq A$ such that*

$$|A' + B_{[h]}^+| \leq \beta|A'|.$$

First of all, we are interested in the particular case where $j = 1$ and $B = A$ of Theorem 2.9, which is enough for many applications. In such a case, using the notation for doubling constants introduced in Chapter 1 we have that

$$|hA| \leq \sigma[A]^h|A|$$

and

$$|kA - lA| \leq \sigma[A]^{k+l}|A|.$$

The entropic inequality analogue to this one was first proved by Tao in [31]. In his paper, he claimed that the inequality loses a constant factor with respect to the Plünnecke-Ruzsa inequalities due to the fact that the graph theoretic proof of Plünnecke's inequality has not been adapted to the entropic setting; however, his arguments can be easily used to show the tight constant that corresponds to the sumset inequalities. For instance, the tight constant appears in a paper of Madiman and Kontoyiannis [16], and were probably first published by Madiman in [18]. For a thorough overview of sumset Plünnecke-type inequalities and their proofs, the reader is encouraged to check [8]. Here we shall simply prove the entropic version, by refining Tao's argument in a new way.

**Theorem 2.11.** *Let $X$ be a $G$-random variable, where $G$ is an additive group, and let $X_1, \ldots, X_k, X'_1, \ldots, X'_l$ be independent copies of $X$ for some positive integers $k$ and $l$. Then,*

$$\mathbf{H}(X_1 + \ldots + X_k - X'_1 - \ldots - X'_l) \leq \mathbf{H}(X) + (k+l)\log\sigma[X].$$

*Proof.* Let $X$ and $Y$ be two independent $G$-random variables, and let $(X_1, Y_1), \ldots, (X_k, Y_k)$ be independent trials of $(X, Y)$. Set $S_i = X_i + Y_i$ for each $i \in \{1, \ldots, k\}$. Observe that we may write $S_1 + \ldots + S_k = (Y_1 + X_2) + (Y_2 + X_3) + \ldots + (Y_{k-1} + X_k) + (Y_k + X_1)$. In particular, both $(X_1, Y_1, S_2, S_3, \ldots, S_k)$ and $(Y_1 + X_2, \ldots, Y_{k-1} + X_k, Y_k + X_1)$ determine $S_1 + \ldots + S_k$, while they jointly determine $(X_1, \ldots, X_k, Y_1, \ldots, Y_k)$. By Lemma 1.9 we have that

$$\begin{aligned}\mathbf{H}(X_1, \ldots, X_k, Y_1, \ldots, Y_k) + &\mathbf{H}(S_1 + \ldots + S_k) \\ &\leq \mathbf{H}(X_1, Y_1, S_2, S_3, \ldots, S_k) + \mathbf{H}(Y_1 + X_2, \ldots, Y_{k-1} + X_k, Y_k + X_1).\end{aligned}$$

By independence,

$$\mathbf{H}(X_1, \ldots, X_k, Y_1, \ldots, Y_k) = k(\mathbf{H}(X) + \mathbf{H}(Y))$$

and

$$\mathbf{H}(X_1, Y_1, S_2, S_3, \ldots, S_k) = \mathbf{H}(X) + \mathbf{H}(Y) + (k-1)\mathbf{H}(X+Y),$$

and by the standard inequality (5),

$$\mathbf{H}(Y_1 + X_2, \ldots, Y_{k-1} + X_k, Y_k + X_1) \leq k\mathbf{H}(X+Y).$$

Substituting these above yields

$$\mathbf{H}(S_1 + \ldots + S_k) \leq (2k-1)\mathbf{H}(X+Y) - (k-1)(\mathbf{H}(X) + \mathbf{H}(Y)).$$

Now let each $Y_i$ be one more independent copy of $X$. Using the definition of the doubling constant, we have that

$$\mathbf{H}(X_1 + \ldots + X_{2k}) \leq (2k-1)\mathbf{H}(X_1 + X_2) - (2k-2)\mathbf{H}(X) = \mathbf{H}(X) + (2k-1)\log\sigma[X].$$

It is clear that this equation holds when adding an even number $n = 2k$ of independent and identically distributed random variables. For the case when we add an odd number $n = 2k+1$ of independent identically distributed random variables $X_1, \ldots, X_{2k+1}$ we proceed in a similar way. Define $S_i = X_{2i-1} + X_{2i}$ and $S'_i = X_{2i} + X_{2i+1}$, for $i \in [k]$. It is clear that $S_1 + \ldots + S_k + X_{2k+1} = X_1 + S'_1 + \ldots + S'_k$. Hence, we have that both $(S_1, \ldots, S_k, X_{2k+1})$ and $(X_1, S'_1, \ldots, S'_k)$ determine $X_1 + \ldots + X_{2k+1}$, and it is easy to check that, combined, they determine $(X_1, \ldots, X_{2k+1})$. By Lemma 1.9,

$$\mathbf{H}(X_1, \ldots, X_{2k+1}) + \mathbf{H}(X_1 + \ldots + X_{2k+1}) \leq \mathbf{H}(S_1, \ldots, S_k, X_{2k+1}) + \mathbf{H}(X_1, S'_1 \ldots, S'_k).$$

By independence,

$$\mathbf{H}(X_1, \ldots, X_{2k+1}) = (2k+1)\mathbf{H}(X),$$

and by (5),

$$\mathbf{H}(S_1, \ldots, S_k, X_{2k+1}) \leq \mathbf{H}(X) + k\mathbf{H}(X + X'),$$
$$\mathbf{H}(X_1, S'_1, \ldots, S'_k) \leq \mathbf{H}(X) + k\mathbf{H}(X + X'),$$

where $X'$ is an independent copy of $X$. Substituting these above yields

$$\begin{aligned}\mathbf{H}(X_1 + \ldots + X_{2k+1}) &\leq \mathbf{H}(S_1, \ldots, S_k, X_{2k+1}) + \mathbf{H}(X_1, S'_1 \ldots, S'_k) - \mathbf{H}(X_1, \ldots, X_{2k+1}) \\ &\leq 2\mathbf{H}(X) + 2k\mathbf{H}(X + X') - (2k+1)\mathbf{H}(X) \\ &= \mathbf{H}(X) + 2k(\mathbf{H}(X + X') - \mathbf{H}(X)) = \mathbf{H}(X) + 2k\log\sigma[X].\end{aligned}$$

Putting the results for odd and even number of summands together, when considering sums of independent identically distributed random variables we may write

$$\mathbf{H}(X_1 + \ldots + X_n) \leq \mathbf{H}(X) + (n-1)\log\sigma[X]. \tag{12}$$

Finally, in order to obtain the inequality from the statement, apply Theorem 2.4 taking $X = -X''$, $X''$ being another independent copy of $X$, $Y = X_1 + \ldots + X_k$ and $Z = X'_1 + \ldots + X'_l$. As $\mathbf{H}(X) =$

$\mathbf{H}(-X)$, this yields

$$\begin{aligned}\mathbf{H}(X_1+\ldots+X_k-X_1'-\ldots-X_l') &\\ \leq \mathbf{H}(X''+X_1+\ldots+X_k)+\mathbf{H}(X''+X_1'+\ldots+X_l')-\mathbf{H}(X'') &\\ \leq \mathbf{H}(X)+k\log\sigma[X]+\mathbf{H}(X)+l\log\sigma[X]-\mathbf{H}(X) &\\ = \mathbf{H}(X)+(k+l)\log\sigma[X], &\end{aligned}$$

where the last inequality comes from applying (12). □

It is interesting to note that (12) is a perfect analogue of Plünnecke's inequality up to the constant factor (one may argue that this result is even better, as the exponent is reduced to $n-1$, but for the applications, this difference does not matter).

One may now try to obtain sumset inequalities from these results. However, it soon becomes aparent that it is not such an easy task as it was before. Several problems arise: first, we want to have several copies of the same random variable (that is, several independent and identically distributed random variables) such that their sum is uniformly distributed over the set $kA-lA$, which cannot be done. Secondly, we face the problem of upper bounding the entropic doubling constant by its additive counterpart, which, again, is not clear.

## 2.2. The Balog-Szemerédi-Gowers theorem

The Balog-Szemerédi-Gowers theorem was a big breakthrough in the theory of set addition. One may roughly think as follows. The traditional results in additive combinatorics are concerned with small sumsets, for in that case one can derive structural properties on the summands. If the sumset is small, then there must be many collisions in the sums. The Balog-Szemerédi-Gowers theorem states a partial converse of this idea: if we know that there are a lot of collisions, then the sumset must be small. This is actually not true; one can construct examples with many collisions where the sumset is big. However, what Balog and Szemerédi first proved is that one can find "big" subsets such that their sumset is "small".

So far we have been dealing with complete subsets, but very often in applications one does not have complete control over a sumset, and only knows a partial collection of these sums. This idea can be represented through partial sumsets.

**Definition 2.2.** Let $A$ and $B$ be two sets in an additive group $G$, and let $\Gamma$ be a subset of $A\times B$ (that is, a subgraph of the complete bipartite graph with independent sets $A$ and $B$). We define the

partial sumset of $A$ and $B$ along $\Gamma$ as

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

The partial difference set is defined in the same way.

In general, it is not easy to work on this setting, as the usual estimates we have been presenting do not work anymore. The Balog-Szemerédi-Gowers theorem allows to go from this setting to the complete sumset setting, by refining the sets that are being added.

The way to think about the "many collisions" is to consider that the partial sumset restricted to a "big" graph is small. One may build sets such that this is true and the complete sumset is very large, but the Balog-Szemerédi-Gowers theorem states that, by refining the sets that are being added, one can also control the size of the complete sumset.

We would like to prove an entropy analogue of this important result. But first, let us state and prove the Balog-Szemerédi-Gowers theorem carefully. The first version of the theorem was presented by Balog and Szemerédi in [2], roughly stating that, if there is a large number of quadruples $(a, a', b, b') \in A^2 \times B^2$ such that $a + a' = b + b'$, then there are large subsets $A' \subseteq A$ and $B' \subseteq B$ such that $A' + B'$ is small. The bounds they gave were greatly improved by Gowers [9, 10], who reduced them to a polynomial and gave a simpler proof. A more recent proof is due to Sudakov, Szemerédi and Vu [28], and the proof we shall present here, based on this work, can be found in [32]. An even more recent proof is due to Schoen [26], who provided a further improvement in the bounds. His new bounds are not presented here.

**Theorem 2.12 (Balog-Szemerédi-Gowers).** *Let $A$ and $B$ be additive sets in an ambient group $G$, and let $\Gamma \subseteq A \times B$ be such that $|\Gamma| \geq \frac{|A||B|}{K}$ and $|A \overset{\Gamma}{+} B| \leq K' \sqrt{|A||B|}$ for some $K \geq 1$ and $K' > 0$. Then, there exist sets $A' \subseteq A$ and $B' \subseteq B$ such that*

$$|A'| \geq \frac{|A|}{4\sqrt{2}K},$$

$$|B'| \geq \frac{|B|}{4K},$$

*and*

$$|A' + B'| \leq 2^{12} K^4 (K')^3 \sqrt{|A||B|}.$$

This can be understood as a statement about dense bipartite graphs. In order to prove it, one must first show two results about paths of small length in these dense bipartite graphs.

**Lemma 2.13.** *Let $\Gamma(A, B, E)$ be a bipartite graph such that $|E| \geq \frac{|A||B|}{K}$ for some $K \geq 1$. Then, for any $0 < \varepsilon < 1$ there exists a set $A' \subseteq A$ such that $|A'| \geq \frac{|A|}{\sqrt{2}K}$ and at least a $(1-\varepsilon)$ proportion of the pairs of vertices $(a, a') \in A' \times A'$ are connected by at least $\frac{\varepsilon}{2K^2}|B|$ paths of length 2 in $\Gamma$.*

*Proof.* We may assume that $|E| = \frac{|A||B|}{K}$ (if this is not the case, decrease the value of $K$). Then, by double counting, we may know the number of edges by counting incoming edges either to $A$ or to $B$,

$$|E| = \sum_{b \in B} |N(b)| = \sum_{a \in A} |N(a)|,$$

and we can find the average degree in each of the independent sets in this way. Writing this in the language of expectations, we have that

$$\mathbb{E}_{b \in B}\left[\frac{|N(b)|}{|A|}\right] = \mathbb{E}_{a \in A}\left[\frac{|N(a)|}{|B|}\right] = \frac{|E|}{|A||B|} = \frac{1}{K}.$$

Now consider the common neighbourhood of a pair of vertices in $A$. We have that

$$\sum_{a,a' \in A} |N(a) \cap N(a')| = \sum_{b \in B} |N(b)|^2$$

by double counting, as the left hand side of the equality counts each vertex $b \in B$ as many times as there are pairs $(a, a')$ such that $b \in N(a) \cap N(a')$, and this corresponds to all pairs $(a, a')$ such that $a, a' \in N(b)$. The number of such pairs is $|N(b)|^2$ (note that we consider the possibility of $(a, a)$ being a pair and the two possible orderings for each $a \neq a'$ at both sides of the equality). Consequently,

$$\mathbb{E}_{a,a' \in A}\left[\frac{|N(a) \cap N(a')|}{|B|}\right] = \mathbb{E}_{b \in B}\left[\frac{|N(b)|^2}{|A|^2}\right] \geq \mathbb{E}_{b \in B}\left[\frac{|N(b)|}{|A|}\right]^2 = \frac{1}{K^2}, \tag{13}$$

where the inequality comes from the Cauchy-Schwarz inequality.

Let $\Omega$ be the set of pairs of vertices of $A$ such that there are less than $\frac{\varepsilon}{2K^2}$ paths of length two connecting them, that is,

$$\Omega = \left\{(a, a') \in A \times A : |N(a) \cap N(a')| < \frac{\varepsilon}{2K^2}|B|\right\}.$$

Note that here we are considering ordered pairs and also pairs of the form $(a, a)$, that is, we consider walks of length two instead of paths. Let $\mathbb{1}_\omega$ denote the indicator function of the event $\omega$. Then, we clearly have that

$$\mathbb{E}_{a,a' \in A}\left[\mathbb{1}_{(a,a') \in \Omega}\frac{|N(a) \cap N(a')|}{|B|}\right] < \frac{|\Omega|}{|A|^2}\frac{\varepsilon|B|}{2K^2|B|} \leq \frac{\varepsilon}{2K^2},$$

so from this and (13)

$$\mathbb{E}_{a,a'\in A}\left[\left(1-\frac{1}{\varepsilon}\mathbb{1}_{(a,a')\in\Omega}\right)\frac{|N(a)\cap N(a')|}{|B|}\right]>\frac{1}{K^2}-\frac{1}{2K^2}=\frac{1}{2K^2}. \tag{14}$$

Again, in a similar way as before, the left hand side of this inequality can be written as

$$\mathbb{E}_{b\in B}\left[\frac{1}{|A|}\sum_{a,a'\in N(b)}\left(1-\frac{1}{\varepsilon}\mathbb{1}_{(a,a')\in\Omega}\right)\right],$$

so by the pigeonhole principle there must exist an element $b\in B$ such that

$$\frac{1}{|A|}\sum_{a,a'\in N(b)}\left(1-\frac{1}{\varepsilon}\mathbb{1}_{(a,a')\in\Omega}\right)>\frac{1}{2K^2}.$$

In particular, we have that

$$\frac{|N(b)|^2}{|A|^2}\geq\frac{1}{|A|}\sum_{a,a'\in N(b)}\left(1-\frac{1}{\varepsilon}\mathbb{1}_{(a,a')\in\Omega}\right),$$

so this means that $|N(b)|>\dfrac{|A|}{\sqrt{2}K}$. We also have that

$$\{(a,a')\in N(b):(a,a')\in\Omega\}\leq\varepsilon|N(b)|^2$$

(otherwise, we would have a negative quantity in (14), reaching a contradiction), so by taking the complement, the number of pairs of neighbours of $b$ that are connected by at least $\dfrac{\varepsilon}{2K^2}|B|$ paths of length two is at least $(1-\varepsilon)|N(b)|^2$. Hence, taking $A'=N(b)$ gives the desired result. □

**Lemma 2.14.** *Let $\Gamma(A,B,E)$ be a bipartite graph such that $|E|\geq\dfrac{|A||B|}{K}$ for some $K\geq 1$. Then, there exist sets $A'\subseteq A$ and $B'\subseteq B$ such that $|A'|\geq\dfrac{|A|}{4\sqrt{2}K}$, $|B'|\geq\dfrac{|B|}{4K}$, and every pair $(a,b)\in A'\times B'$ is connected by at least $\dfrac{|A||B|}{2^{12}K^4}$ paths of length three.*

*Proof.* Let $\tilde{A}\subseteq A$ be the set of vertices in $A$ that have degree at least $\dfrac{|B|}{2K}$ (this set is nonempty since, by hypothesis, the average degree of the vertices of $A$ is at least $\dfrac{|B|}{K}$). Let $\tilde{\Gamma}(\tilde{A},B,\tilde{E})$ be the subgraph of $\Gamma$ induced by $\tilde{A}$. Note that in this transformation we delete at most $\dfrac{|A||B|}{2K}$ edges (assuming all vertices were deleted, which cannot happen!), so $|\tilde{E}|\geq\dfrac{|A||B|}{2K}$ holds for the resulting set of edges. Write $|A|=L|\tilde{A}|$ for some $L\geq 1$.

Now, take $K' = \frac{2K}{L}$ and $\varepsilon = \frac{1}{16K}$ and apply Lemma 2.13 to $\tilde{\Gamma}$. This yields a subset $\tilde{A}' \subseteq \tilde{A}$ of size $|\tilde{A}'| \geq \frac{|\tilde{A}|}{\sqrt{2}\frac{2K}{L}} = \frac{|A|}{2\sqrt{2}K}$ and such that $\left(1 - \frac{1}{16K}\right)$ of the pairs $(a, a') \in \tilde{A}' \times \tilde{A}'$ are connected by at least $\frac{L^2|B|}{128K^3}$ paths of length two.

Let us call a pair $(a, a') \in \tilde{A}' \times \tilde{A}'$ "bad" if $a$ and $a'$ are not connected by at least $\frac{L^2|B|}{8K^2}$ paths of length two. Observe that there are at most $\frac{1}{16K}|\tilde{A}'|^2$ such bad pairs. Let $A' \subseteq \tilde{A}'$ be the set of all $a \in \tilde{A}'$ such that at most $\frac{1}{8K}|\tilde{A}'|$ pairs $(a, a')$ are bad. Then, $|\tilde{A}' \setminus A'| \leq \frac{|\tilde{A}'|}{2}$. Indeed, define a bipartite graph $\Delta(A_1, A_2, E^*)$, where $A_1$ and $A_2$ are two copies of $\tilde{A}'$ and $(a, a') \in E^*$ if, and only if, $(a, a')$ is a bad pair. As we said before, $|E^*| \leq \frac{1}{16K}|\tilde{A}'|$. Assume that $|\tilde{A}' \setminus A'| > \frac{|\tilde{A}'|}{2}$; then, the sum of the degrees in $A_1$, which equals the number of edges, is greater than $\frac{|\tilde{A}'|}{2}\frac{|\tilde{A}'|}{8K} = \frac{|\tilde{A}'|^2}{16K}$, thus reaching a contradiction. Hence, we must have that

$$|A'| \geq \frac{|\tilde{A}'|}{2} \geq \frac{|A|}{4\sqrt{2}K}.$$

Now we strive to define the corresponding $B'$. Since every element in $\tilde{A}$ (and in particular in $\tilde{A}'$) has degree at least $\frac{|B|}{2K}$, we have that

$$\sum_{b \in B} |\{a \in \tilde{A}' : (a, b) \in E\}| = \left|\{(a, b) \in E : a \in \tilde{A}'\}\right| \geq |\tilde{A}'|\frac{|B|}{2K}.$$

Let

$$B' = \left\{b \in B : |\{a \in \tilde{A}' : (a, b) \in E\}| \geq \frac{|\tilde{A}'|}{4K}\right\},$$

that is, the set of vertices of $B$ that have degree at least $\frac{|\tilde{A}'|}{4K}$. Then we have

$$|\tilde{A}'||B'| \geq \sum_{b \in B'} |\{a \in \tilde{A}' : (a, b) \in E\}| \geq |\tilde{A}'|\frac{|B|}{2K} - |B|\frac{|\tilde{A}'|}{4K} = \frac{|\tilde{A}'||B|}{4K},$$

where the difference comes from assuming $B'$ is empty as an upper bound, and therefore $|B'| \leq \frac{|\tilde{A}'||B|}{4K}$.

Let $a \in A'$ and $b \in B'$ be any vertices. By construction of $B'$, $b$ is adjacent to at least $\frac{|\tilde{A}'|}{4K}$ elements $a'$ of $\tilde{A}'$. By construction of $A'$, at most $\frac{|\tilde{A}'|}{8K}$ pairs $(a, a')$ are bad, so more than $\frac{7|\tilde{A}'|}{8K}$ are good.

Thus, there are at least $\frac{|\tilde{A}'|}{8K} \geq \frac{|A|}{16\sqrt{2}K}$ vertices $a'$ which are simultaneously adjacent to $b$ and connected to $a$ by at least $\frac{L^2|B|}{8K^2}$ paths of length two. This means that, overall, there are at least $\frac{|A|}{16\sqrt{2}K}\frac{L^2|B|}{8K^2} \geq \frac{|A||B|}{2^8K^3}$ paths of length three connecting $a$ and $b$. This is even better than what was stated. □

*Proof of Theorem 2.12.* First, we want to define the bipartite graph over $A$ and $B$. These two sets may not be disjoint, but we can make them so by replacing $G$ by $G \times \mathbb{Z}$, $A$ by $A \times \{0\}$, and $B$ by $B \times \{1\}$. Now we can view $\Gamma$ as a bipartite graph defined on the sets $A$ and $B$. A direct application of Lemma 2.14 gives subsets $A' \subseteq A$ and $B' \subseteq B$ of the right cardinalities, and such that every pair $a \in A'$, $b \in B'$ is connected by at least $\frac{|A||B|}{2^{12}K^4}$ paths of length three. We can rewrite this as

$$\left|\left\{(a',b') \in A' \times B' : (a,b'),(b',a'),(a',b) \in \Gamma\right\}\right| \geq \frac{|A||B|}{2^{12}K^4},$$

for any pair $(a,b) \in A' \times B'$. We obviously have that $a+b = (a+b') - (a'+b') + (a'+b)$, so by taking $x = a+b'$, $y = a'+b'$ and $z = a'+b$ we may write that, for any pair $(a,b) \in A' \times B'$,

$$\left|\left\{(x,y,z) \in \left(A \overset{\Gamma}{+} B\right)^3 : x-y+z = a+b\right\}\right| \geq \frac{|A||B|}{2^{12}K^4}.$$

Since the overall number of triples is $|A \overset{\Gamma}{+} B|^3 \leq (K')^3|A|^{\frac{3}{2}}|B|^{\frac{3}{2}}$, we conclude that the total number of possible values for $a+b$ is bounded as

$$|A'+B'| \leq \frac{(K')^3|A|^{\frac{3}{2}}|B|^{\frac{3}{2}}}{\frac{|A||B|}{2^{12}K^4}} = 2^{12}K^4(K')^3\sqrt{|A||B|},$$

as we wanted to see. □

Now we would like to prove an entropy analogue of the Balog-Szemerédi-Gowers theorem. One needs to find analogues of the different elements in the statement; particularly, we need an analogue of partial sumsets, as well as an analogue of the refinements of the sets through which the result is obtained. The second corresponds to conditioning the random variables, while the first can be identified with making variables weakly dependent, in a certain sense. The statement of the theorem is as follows.

**Theorem 2.15.** *Let $G$ be an additive group, and let $X$ and $Y$ be two $G$-random variables which are weakly dependent, in the sense that $\mathbf{H}(X,Y) \geq \mathbf{H}(X) + \mathbf{H}(Y) - \log K$ for some $K \geq 1$. Suppose also that $\mathbf{H}(X+Y) \leq \frac{1}{2}(\mathbf{H}(X) + \mathbf{H}(Y)) + \log K'$, for some $K' > 0$. If we let $(X_1,Y_1)$ and $(X_2,Y_2)$ be conditionally independent trials of $(X,Y)$ conditioning on $Y$ and let $(X_1,X_2,Y)$ and $(X_1,Y')$ be conditionally*

*independent trials of* $(X_1, X_2, Y)$ *and* $(X_1, Y)$ *conditioning on* $X_1$*, then* $X_2$ *and* $Y'$ *are conditionally independent relative to* $(X_1, Y)$ *with*

$$\mathbf{H}(X_2|X_1, Y) \geq \mathbf{H}(X) - \log K, \tag{15}$$

$$\mathbf{H}(Y'|X_1, Y) \geq \mathbf{H}(Y) - \log K \tag{16}$$

*and*

$$\mathbf{H}(X_2 + Y'|X_1, Y) \leq \frac{1}{2}(\mathbf{H}(X) + \mathbf{H}(Y)) + 4\log K + 3\log K'. \tag{17}$$

Notice the analogies between this and Theorem 2.12. The entropies of the conditioned random variables are exactly analogous to the sizes of the refined sets, and the entropy of the sum is analogous to the size of the sum of the refinements. Furthermore, we can also find an analogy with the paths of length three. Indeed, the conditionally independent trials $(X_1, Y_1)$ and $(X_2, Y_2)$ conditioning on $Y$ can be though of as paths of length two (we obtain two values of $X$ related to a unique value of $Y$), and the conditionally independent trials $(X_1, X_2, Y)$ and $(X_1, Y')$ conditioning to $X_1$ extend the path on one of its sides. Saying that $X_2$ and $Y'$ are conditionally independent relative to $(X_1, Y)$ can somehow be interpreted as saying that they are independent, provided they are adjacent to the endpoint of an edge defined by $(X_1, Y)$. Overall, one can think that the random variables $(Y', X_1, Y, X_2)$ are drawn from the space of all paths of length three. The reader is encouraged to consider this analogy while studying the following proof.

In order to prove the theorem, we first present a lemma that we shall need.

**Lemma 2.16.** *Consider the same setting and random variables as in Theorem 2.15. Then,*

$$\mathbf{H}(X_1 - X_2|Y) \leq \mathbf{H}(X) + 2\log K + 2\log K'.$$

*Proof.* Let $(X_1, X_2, Y)$ and $(X_1, X_2, Y')$ be two conditionally independent trials of $(X_1, X_2, Y)$ relative to $(X_1, X_2)$. Observe that $(X_1, X_2, Y)$ and $(X_1 + Y', X_2 + Y', Y)$ both determine $(X_1 - X_2, Y)$, and they jointly determine $(X_1, X_2, Y, Y')$. Hence, we may apply Lemma 1.9 to obtain

$$\mathbf{H}(X_1, X_2, Y, Y') + \mathbf{H}(X_1 - X_2, Y) \leq \mathbf{H}(X_1, X_2, Y) + \mathbf{H}(X_1 + Y', X_2 + Y', Y).$$

By conditional independence (9), trivial joint entropy inequalities (5), and the definition of conditional entropy (Lemma 1.3), we have that

$$\begin{aligned}\mathbf{H}(X_1, X_2, Y, Y') &= 2\mathbf{H}(X_1, X_2, Y) - \mathbf{H}(X_1, X_2) = 4\mathbf{H}(X, Y) - 2\mathbf{H}(Y) - \mathbf{H}(X_1, X_2)\\ &\geq 4\mathbf{H}(X, Y) - 2\mathbf{H}(Y) - 2\mathbf{H}(X),\\ \mathbf{H}(X_1 - X_2, Y) &= \mathbf{H}(X_1 - X_2|Y) + \mathbf{H}(Y),\\ \mathbf{H}(X_1, X_2, Y) &= 2\mathbf{H}(X, Y) - \mathbf{H}(Y),\\ \mathbf{H}(X_1 + Y', X_2 + Y', Y) &\leq 2\mathbf{H}(X + Y) + \mathbf{H}(Y).\end{aligned}$$

Substituting these above and rearranging the terms yields

$$\mathbf{H}(X_1 - X_2|Y) \leq 2\mathbf{H}(X + Y) + \mathbf{H}(Y) + 2\mathbf{H}(X) - 2\mathbf{H}(X, Y) \leq \mathbf{H}(X) + 2\log K + 2\log K',$$

where the last inequality comes from the statement hypothesis. □

*Proof of Theorem 2.15.* By construction, $Y'$ and $(X_2, Y)$ are conditionally independent relative to $X_1$, so $X_2$ and $Y'$ are conditionally independent relative to $(X_1, Y)$. Also, since $X_1$ is conditionally independent from $X_2$ relative to $Y$, we have

$$\mathbf{H}(X_2|X_1, Y) = \mathbf{H}(X_2|Y) = \mathbf{H}(X|Y) = \mathbf{H}(X, Y) - \mathbf{H}(Y) \geq \mathbf{H}(X) + \mathbf{H}(Y) - \log K,$$

so (15) holds. On the other hand, since $Y$ and $Y'$ are conditionally independent relative to $X_1$,

$$\mathbf{H}(Y'|X_1, Y) = \mathbf{H}(Y'|X_1) = \mathbf{H}(Y|X) = \mathbf{H}(X, Y) - \mathbf{H}(X) \geq \mathbf{H}(Y) + \mathbf{H}(X) - \log K,$$

so (16) holds too. Finally, in order to prove (17), observe that $(X_2, Y', Y)$ and $(X_1 - X_2, X_1 + Y', Y)$ both determine $(X_2 + Y', Y)$ and jointly determine $(X_1, X_2, Y, Y')$. A direct application of Lemma 1.9 results in

$$\mathbf{H}(X_1, X_2, Y, Y') + \mathbf{H}(X_2 + Y', Y) \leq \mathbf{H}(X_2, Y', Y) + \mathbf{H}(X_1 - X_2, X_1 + Y', Y).$$

By conditional independence (9), trivial joint entropy inequalities (5), and the definition of conditional entropy (Lemma 1.3), we then have that

$$\begin{aligned}\mathbf{H}(X_1, X_2, Y, Y') &= \mathbf{H}(X_1, X_2, Y) + \mathbf{H}(X_1, Y') - \mathbf{H}(X_1),\\ \mathbf{H}(X_2 + Y', Y) &= \mathbf{H}(X_2 + Y'|Y) + \mathbf{H}(Y)\\ \mathbf{H}(X_2, Y', Y) &\leq \mathbf{H}(X_2, Y) + \mathbf{H}(Y') = \mathbf{H}(X, Y) + \mathbf{H}(Y),\end{aligned}$$

and

$$\begin{aligned}\mathbf{H}(X_1 - X_2, X_1 + Y', Y) &\leq \mathbf{H}(X_1 - X_2|Y) + \mathbf{H}(Y) + \mathbf{H}(X_1 + Y')\\ &\leq \mathbf{H}(X_1 - X_2|Y) + \mathbf{H}(Y) + \mathbf{H}(X + Y).\end{aligned}$$

By substituting we obtain that

$$\mathbf{H}(X_2 + Y'|Y) \leq \mathbf{H}(X_1 - X_2|Y) + 2\mathbf{H}(Y) + \mathbf{H}(X) + \mathbf{H}(X + Y) - 2\mathbf{H}(X, Y).$$

The result follows by applying Lemma 2.16 and the conditions from the statement. □

# Chapter 3
# Entropy, projections and sumsets

In a different direction than what was presented in Chapter 2, many people also noticed that there exists a certain parallelism between entropy inequalities and projection inequalities (meaning inequalities using the sizes of sets and their projections). This has been studied by several renowned researchers, such as Ruzsa [25], Balister and Bollobás [1], or Madiman, Marcus and Tetali [19]. Furthermore, Gyarmati, Matolcsi and Ruzsa [12] realized that one can use projection inequalities to obtain sumset inequalities which are analogous to their projection counterparts. Their idea has been further studied and developed, and we shall present here all the different results related to these analogies.

In order to better understand the analogies, and the very powerfull results that have been recently achieved, we believe it is better to start by presenting the original results, and slowly build up to the more general ones. This allows us to present the main ideas more clearly, and we hope that this also helps the reader to obtain a clear understanding of the techniques that are being used. In any case, we shall not prove every result separately, as we shall see that many of them can be deduced from the more general form.

We devote Section 3.1 to present the theorems and analogies in a very particular case, corresponding to the first historical results in each of the settings. The more general case, also well studied for a long time, is presented in Section 3.2, while Section 3.3 describes some more recent generalizations. In Section 3.4 we present a new approach, which provides a unified way to prove all the previous results and also obtain some new ones.

The diagram at the end of this chapter illustrates the relationship among the successive generalizations presented in the chapter.

## 3.1. The first results

In 1949, Loomis and Whitney [17] first proved an inequality bounding the volume of an $n$-dimensional body in terms of the volume of its $(n-1)$-dimensional projections. Their result extends to the discrete setting, where it may be written as follows.

**Theorem 3.1 (Loomis, Whitney).** *Let $n \geq 2$, let $B_1, \ldots, B_n$ be arbitrary finite sets, and let $A \subseteq B_1 \times \ldots \times B_n$ be a subset of their Cartesian product. Let $A_i = \pi_i(A) \subseteq B_1 \times \ldots \times B_{i-1} \times B_{i+1} \times \ldots \times B_n$ be the projection of $A$ to the coordinate hyperplanes, where $\pi_i(a) = \pi_i(a_1, \ldots, a_n) = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. Then,*

$$|A|^{n-1} \leq \prod_{i=1}^{n} |A_i|.$$

Almost thirty years later, Han [13] proved an exact analogue of the Loomis and Whitney inequality for the entropy of a family of random variables. With the properties of entropy introduced in Chapter 1 it is now easy to give a proof of this result.

**Theorem 3.2 (Han's inequality).** *Let $X_1, \ldots, X_n$ be $n$ discrete random variables. Then,*

$$\mathbf{H}(X_1, \ldots, X_n) \leq \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

*Proof.* By the definition of conditional entropy given in Lemma 1.3 and by Lemma 1.6, we may write

$$\begin{aligned} \mathbf{H}(X_1, \ldots, X_n) &= \mathbf{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + \mathbf{H}(X_i | X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\ &\leq \mathbf{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + \mathbf{H}(X_i | X_1, \ldots, X_{i-1}) \end{aligned}$$

for all $1 \leq i \leq n$. If these $n$ inequalities are added, we obtain

$$n\mathbf{H}(X_1, \ldots, X_n) = \sum_{i=1}^{n} \left[ \mathbf{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + \mathbf{H}(X_i | X_1, \ldots, X_{i-1}) \right]$$

(where the case for $i = 1$ of $\mathbf{H}(X_i | X_1, \ldots, X_{i-1})$ is understood as $\mathbf{H}(X_1)$), and by Lemma 1.5 the second summands add up to $\mathbf{H}(X_1, \ldots, X_n)$. Isolating this term yields the desired result. □

Notice that we do not make any assumptions about the independence of the random variables.

The proof of Theorem 3.1 is not difficult. The authors of [17] reduce the general problem to a combinatorial one, dividing the body into cubes, and then use induction on $n$. In fact, the approach using cubes is enough in the discrete case, as we may assume that we are dealing with sets in the $n$-dimensional lattice, where considering cubes of side 1 centered in the points of the lattice directly

gives a body in the $n$-dimensional real space with the same volume as the number of points of the discrete structure. However, the proof we shall present here is much simpler, based on the entropic method.

*Proof of Theorem 3.1.* Define independent random variables $X_1, \ldots, X_n$, $X_i$ ranging over the projection of $A$ to its $i$-th coordinate, in such a way that each of them is uniform. Clearly, the joint distribution $X = (X_1, \ldots, X_n)$ is uniformly distributed over $A$. Then, by Han's inequality and Lemma 1.2 we have that

$$\log |A| = \mathbf{H}(X) \leq \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \leq \frac{1}{n-1} \sum_{i=1}^{n} \log |A_i|,$$

whence the result follows by exponentiating. □

Observe that in this way we have obtained a completely elementary proof of a result in projection theory using the properties of entropy, and the key tool for this is the fact that a uniform random variable can somehow "represent" a set. This reminds us of the approach we presented in Chapter 2. However, in this setting Ruzsa [25] managed to prove the converse: starting from the Loomis and Whitney inequality, one can prove Han's inequality. In order to do this we must find a way to represent a random variable with a set. If the random variable $X$ were uniform, we could use the same approach as before: the set $A = \text{range}(X)$ reflects the distribution of $X$. But since Han's inequality holds for general random variables, we must find a way to build sets that reflect the distribution of the random variables.

In order to establish such a way to associate a set to a random variable $X$, consider the following. First, assume that $X$ assumes only finitely many values (say, $\text{range}(X) = \mathcal{X}$, $|\mathcal{X}| = m$), and that its distribution is given by $\Pr(X = x_i) = p_i$ for each $x_i \in \mathcal{X}$, with $\sum_{i=1}^{m} p_i = 1$. Assume, furthermore, that all the $p_i$ are rational. Then, we build a set $A \subseteq \mathcal{X}^k$ in the cartesian product of the range of $X$, for a certain $k$, as follows. For each $(y_1, \ldots, y_k) \in \mathcal{X}^k$, we include it in $A$ if among the $k$ coordinates exactly $p_i k$ of them are equal to $x_i$, for all $i$. For this to be possible, we need to take a $k$ such that $p_i k$ is an integer for all $i$, but this can be achieved because there are only finitely many values, and each of them is rational. In this way, if we take an element of $A$ and pick one of its coordinates uniformly at random, the probability that it is $x_i$ is $p_i$, so we recover the random variable. And $A$ is the set of all elements of $\mathcal{X}^k$ for which this can be done.

Using Stirling's formula and the construction for $A$ presented above, we have that

$$|A| = \binom{k}{p_1 k, p_2 k, \dots, p_m k} = \frac{k!}{(p_1 k)! \dots (p_m k)!} \sim \frac{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k}{\sqrt{2^m \pi^m k^m \prod_{i=1}^m p_i} \prod_{i=1}^m \left(\frac{p_i k}{e}\right)^{p_i k}} \tag{18}$$

$$= \frac{(2\pi k)^{\frac{1}{2}}}{(2\pi k)^{\frac{m}{2}}} \frac{1}{\left(\prod_{i=1}^m p_i\right)^{\frac{1}{2}} \left(\prod_{i=1^m} p_i^{p_i}\right)^k} = \exp\log\left(\frac{(2\pi k)^{\frac{1}{2}}}{(2\pi k)^{\frac{m}{2}}} \left(\prod_{i=1}^m p_i\right)^{-\frac{1}{2}} \left(\prod_{i=1^m} p_i^{p_i}\right)^{-k}\right)$$

$$= \exp\left(\frac{1-m}{2}\log(2\pi k) - \frac{1}{2}\sum_{i=1}^m \log p_i - k\sum_{i=1} p_i \log p_i\right) = e^{k\mathbf{H}(X)+O(\log k)},$$

where the last equality is obtained by observing that the random variable $X$ is fixed, so the only variable is $k$, and the definition of entropy.

With this construction, one can prove Han's inequality starting from the Loomis and Whitney inequality.

**Theorem 3.3 (Ruzsa).** *Theorem 3.1 and Theorem 3.2 are equivalent.*

*Proof.* We already proved the fact that the Loomis and Whitney inequality can be derived from Han's inequality, as this is the way in which we proved the Loomis and Whitney inequality. Now let us see the converse.

Let $X = (X_1, \dots, X_n)$. First, assume that $X$ assumes only finitely many values, each of them with a rational probability. Let $B_i$ be the range of $X_i$, for $i \in [n]$, and write $\mathcal{X} = B_1 \times \ldots \times B_n$ for the range of $X$. Now build a set $A \subseteq \mathcal{X}^k$ as described above. By (18) we know that $\log|A| \sim k\mathbf{H}(X)$.

Observe that there is a natural correspondence $\mathcal{X}^k = (B_1 \times \ldots \times B_n)^k \cong B_1^k \times \ldots \times B_n^k$. Then, the projections $\pi_i$ of $\mathcal{X}$ defined in Theorem 3.1 naturally induce the projections $\pi_i^k : \mathcal{X}^k \to B_1^k \times \ldots \times B_{i-1}^k \times B_{i+1}^k \times \ldots \times B_n^k$. Now we have two "operations" (functionals), the projection and the way in which we construct a set, and we may build the following diagram with them.

$$\begin{array}{ccc} X & \xrightarrow{\text{make set}} & A \\ \downarrow{\scriptstyle \pi_i} & & \downarrow{\scriptstyle \pi_i^k} \\ \pi_i(X) & \xrightarrow{\text{make set}} & A_i \end{array}$$

Observe that this diagram is commutative. Indeed, we have that $A$ is built by taking all those elements of $\mathcal{X}^k$ such that $x \in \mathcal{X}$ appears exactly $p_x k$ times in its vectorial expression, where $p_x = \Pr(X = x)$. Then, the projection onto one of the coordinate hyperplanes deletes one of the coordinates of the elements in the range of $X$. One may think of each element in $A$ as a matrix with $n$ columns, corresponding to the ranges of the $n$ random variables, and $k$ rows, in such a way that in the columns, the elements $x = (x_1, \dots, x_n)$ appear exactly $p_x k$ times. Then, the projection $\pi_i$ consists in deleting the $i$-th column in each of the elements in $A$, and each element $\pi_i(x) \in B_1^k \times \ldots \times B_{i-1}^k \times B_{i+1}^k \times \ldots \times B_n^k$ now appears as many times as it appeared for each of the possible values of the $i$-th coordinate, that is, it appears $k \sum_{x_i \in B_i} \Pr(X = (x_1, \dots, x_n))$ times.

On the other hand, the projection $\pi_i$ of $X$ deletes the $i$-th random variable in the joint distribution. Then, $\Pr(\pi_i(X) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)) = \sum_{x_i \in B_i} \Pr(X = (x_1, \dots, x_n))$, so the same value of $k$ for which $p_x k$ is an integer for all $x \in \mathcal{X}$ also works for their projections. Building now the set corresponding to $\pi_i(X)$ gives matrices with $n-1$ columns, each corresponding to each $B_j$ for $j \in [n] \setminus \{i\}$, and with the same number of rows as before, in such a way that each element appears $k \Pr(\pi_i(X) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n))$. But this is exactly the same as above (except maybe for a reordering of the rows).

As a consequence, we have that $\log |A_i| = k\mathbf{H}(\pi_i(X)) + O(\log k)$ by (18). Now, by Theorem 3.1 we have that

$$\begin{aligned}
|A|^{n-1} \leq \prod_{i=1}^{n} |A_i| &\implies (n-1) \log |A| \leq \sum_{i=1}^{n} \log |A_i| \\
&\implies (n-1) \left[k\mathbf{H}(X) + O(\log k)\right] \leq \sum_{i=1}^{n} \left[k\mathbf{H}(\pi_i(X)) + O(\log k)\right] \\
&\implies (n-1)k\mathbf{H}(X) \leq k \sum_{i=1}^{n} \mathbf{H}(\pi_i(X)) + O(\log k) \\
&\implies (n-1)\mathbf{H}(X) \leq \sum_{i=1}^{n} \mathbf{H}(\pi_i(X)) + O\left(\frac{\log k}{k}\right).
\end{aligned}$$

Now, if we let $k$ go to infinity (when building the set $A$, we may take the smallest $k$ such that $p_x k$ is integer for all $x$, and then take multiples of this $k$), the asymptotic term goes to zero. Since the other two terms do not depend on $k$, this becomes Han's inequality.

A standard limiting argument extends the above argument to the general case of real probabilities. □

With this, we have that Han's inequality and the Loomis and Whitney theorem are not only analogous in their formulation, but actually equivalent. Ruzsa took this idea even further, and managed

to prove an equivalence theorem between sizes of linear functions of sets restricted to graphs, and entropy of linear functions of random variables. First, let us write a more general version of Definition 2.2

**Definition 3.1.** Let $A$ and $B$ be two finite sets in an additive group $G$, and let $\Gamma$ be a bipartite graph $\Gamma(A, B)$. Given any function $f$ on two variables, we say that the *partial function of $A$ and $B$ along* $\Gamma$ is

$$f(\Gamma) = \{f(a, b) : a \in A, b \in B, a \sim_\Gamma b\}.$$

Using this notation, Ruzsa's equivalence theorem can be stated as follows.

**Theorem 3.4 (Ruzsa).** *Let $l, l_1, \dots, l_n$ be linear functions in two variables with integer coefficients. Let $\alpha_1, \dots, \alpha_n$ be positive real numbers. Then, the following are equivalent:*

*(i) For every pair $(A, B)$ of finite sets in a torsionfree commutative group and a bipartite graph $\Gamma(A, B)$,*

$$|l(\Gamma)| \leq \prod_{i=1}^{n} |l_i(\Gamma)|^{\alpha_i}.$$

*(ii) (i) holds when $A$ and $B$ are sets of integers.*

*(iii) For every pair of (not necessarily independent) random variables $X$ and $Y$ taking values in a torsionfree commutative group such that the entropy of each $l_i(X, Y)$ is finite, the entropy of $l(X, Y)$ satisfies*

$$\mathbf{H}(l(X, Y)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(l_i(X, Y)).$$

*(iv) (iii) holds for integer-valued random variables.*

One must note that linear functions in two variables restricted to a bipartite graph are the same as linear projections. Indeed, the set of edges of the graph can be written as $E = \{(a, b) : a \in A, b \in B, a \sim_\Gamma b\} \subseteq A \times B$, and then $f(\Gamma) = f(E)$. With this fact in mind, we can rewrite Theorem 3.4 in terms of subsets of Cartesian products of groups.

**Theorem 3.5 (Ruzsa).** *Let $l, l_1, \dots, l_n$ be linear functions in two variables with integer coefficients. Let $\alpha_1, \dots, \alpha_n$ be positive real numbers. Then, the following are equivalent:*

*(i) For every torsionfree commutative group $G$, for any finite set $A \subseteq G \times G$ we have that*

$$|l(A)| \leq \prod_{i=1}^{n} |l_i(A)|^{\alpha_i}.$$

*(ii) (i) holds when taking $G = \mathbb{Z}$.*

*(iii) For every pair of (not necessarily independent) random variables $X$ and $Y$ taking values in a torsionfree commutative group such that the entropy of each $l_i(X,Y)$ is finite, the entropy of $l(X,Y)$ satisfies*

$$\mathbf{H}(l(X,Y)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(l_i(X,Y)).$$

*(iv) (iii) holds for integer-valued random variables.*

*Proof.* We trivially have that (i) implies (ii) and (iii) implies (iv), as they are particular cases of the more general statements (i) and (iii), respectively. Now let us prove a few other implications.

First we prove that (ii) implies (i). To do so, replace the group $G$ by the group $H$ generated by all the coordinates of $A$. We observe that $H$ is a subgroup of $G$, and that it is finitely generated. It is known that finitely generated torsionfree groups are isomorphic to $\mathbb{Z}^d$, for some integer $d$. Now take a linear map $\mathbb{Z}^d \to \mathbb{Z}$ that separates the images enough; for example, one may take $(x_1, \ldots, x_d) \mapsto x_1 + kx_2 + \ldots + k^{d-1}x_d$, for a large enough integer $k$. With the composition of the isomorphism and this linear function, we map $A$ onto $A' \subseteq \mathbb{Z} \times \mathbb{Z}$. Finally, we can make $k$ large enough so that $|l(A')| = |l(A)|$ and $|l_i(A')| = |l_i(A)|$ for all $i$. Then,

$$|l(A)| = |l(A')| \leq \prod_{i=1}^{n} |l_i(A')|^{\alpha_i} = \prod_{i=1}^{n} |l_i(A)|^{\alpha_i}.$$

Next, we show that (iv) implies (ii). For each $z \in l(A)$ choose a pair $(x,y) \in A$ such that $l(x,y) = z$ (that is, a representative of the preimage of $z$ by $l$). Define a random variable $(X,Y)$ such that $\Pr(X = x, Y = y) = \dfrac{1}{|l(A)|}$ for pairs $(x,y)$ as defined above, and $\Pr(X = x, Y = y) = 0$ everywhere else. With this, we have that $l(X,Y)$ is distributed uniformly over $l(A)$, and using the properties of entropy we have

$$\log |l(A)| = \mathbf{H}(l(X,Y)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(l_i(X,Y)) \leq \sum_{i=1}^{n} \alpha_i \log |l_i(A)| = \log \prod_{i=1}^{n} |l_i(A)|^{\alpha_i},$$

where the last inequality comes from the fact that $l_i(\mathrm{range}(X), \mathrm{range}(Y)) \subseteq l_i(A)$, and the result follows by exponentiating.

Finally, we show that (i) implies (iii). For this, first assume that the random variable $(X,Y)$ assumes only finitely many values, and that each of these occurs with rational probability. Now build a set $A \subseteq (G \times G)^k$ as described before, so $\log |A| = k\mathbf{H}(X,Y) + O(\log k)$. We have a natural bijection $(G \times G)^k \cong G^k \times G^k$. Let $A'$ be the image of $A$ through this map. Then we can build the following diagram (in which we can consider $l$ to be represented among the $l_i$).

$$\begin{array}{ccc} (X,Y) & \xrightarrow{\text{make set}} & A' \\ \Big\downarrow{\scriptstyle l_i} & & \Big\downarrow{\scriptstyle l_i} \\ l_i(X,Y) & \xrightarrow{\text{make set}} & l_i(A') \end{array}$$

As in the proof of Theorem 3.3, the diagram is commutative. This is proved in the same way as before; we omit the details here. As a consequence, $\log|l_i(A')| = k\mathbf{H}(l_i(X,Y)) + O(\log k)$ by (18). Using (i) we have

$$\begin{aligned} |l(A)| \leq \prod_{i=1}^{n} |l_i(A)|^{\alpha_i} &\implies \log|l(A)| \leq \sum_{i=1}^{n} \alpha_i \, |l_i(A)| \\ &\implies k\mathbf{H}(l(X,Y)) + O(\log k) \leq k\sum_{i=1}^{n} \alpha_i \mathbf{H}(l_i(X,Y)) + O(\log k) \\ &\implies \mathbf{H}(l(X,Y)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(l_i(X,Y)) + O\left(\frac{\log k}{k}\right), \end{aligned}$$

which gives the desired result when letting $k$ tend to infinity. As before, the passage to the general case is a routine limiting argument. □

Let us briefly present an application of this equivalence theorem. The following result, due to Katz and Tao [15], is a well-known inequality relating the size of a sumset and a difference set along a graph $\Gamma$.

**Theorem 3.6.** *Let $A$ and $B$ be two sets in a commutative torsionfree group, and let $\Gamma$ be a bipartite graph on them. Then,*

$$|A \overset{\Gamma}{-} B| \leq (|A||B|)^{\frac{2}{3}} |A \overset{\Gamma}{+} B|^{\frac{1}{2}}.$$

Although we do not present its proof, we can easily obtain its entropy analogue, as presented by Ruzsa.

**Theorem 3.7.** *Let $X$ and $Y$ be two (not necessarily independent) discrete random variables with values in a torsionfree commutative group, and such that their entropies and the entropy of their sum are finite. Then, the entropy of $X - Y$ is also finite and satisfies*

$$\mathbf{H}(X-Y) \leq \frac{2}{3}(\mathbf{H}(X) + \mathbf{H}(Y)) + \frac{1}{2}\mathbf{H}(X+Y).$$

*Proof.* The result follows directly from Theorem 3.6 by applying Theorem 3.4. □

Observe that the graph $\Gamma$ along which the set operations are calculated corresponds to the non-independence of the random variables in this entropic result.

Recently, Gyarmati, Matolcsi and Ruzsa [12] proved an analogue of Han's inequality and the Loomis and Whitney theorem in the theory of set addition. They did so by combining the projection inequality of Loomis and Whitney with a lexicographic ordering. We present this result here.

**Theorem 3.8 (Gyarmati, Matolcsi, Ruzsa).** *Let $B_1, \ldots, B_n$ be finite nonempty sets in a commutative group. Let $S = B_1 + \ldots + B_n$ and $S_i = B_1 + \ldots + B_{i-1} + B_{i+1} + \ldots + B_n$. Then,*

$$|S|^{n-1} \leq \prod_{i=1}^{n} |S_i|.$$

*Proof.* First, list the elements of the sets in some order, say, $B_1 = \{b_{11}, b_{12}, \ldots, b_{1t_1}\}, \ldots, B_n = \{b_{n1}, b_{n2}, \ldots, b_{nt_n}\}$. For each element $s \in S$ consider the decomposition $s = a_{1i_1} + a_{2i_2} + \ldots + a_{ni_n}$ such that the sequence of second indices of the summands $(i_1, i_2, \ldots, i_n)$ is minimal in lexicographical order. Now, define a function $\varphi : S \to B_1 \times \ldots \times B_n$ by $\varphi(s) = (a_{1i_1}, a_{2i_2}, \ldots, a_{ni_n})$. This function is well defined, and its image is a set of "representatives" of each possible sum. Say that the image of $S$ by $\varphi$ is $A \subseteq B_1 \times \ldots \times B_n$; we obviously have that $|S| = |A|$. Now define $A_i = \pi_i(A)$ as in the statement of Theorem 3.1. Its application yields

$$|A|^{n-1} \leq \prod_{i=1}^{n} |A_i|,$$

so now it is enough to see that $|A_j| \leq |S_j|$.

Assume there are two elements $z, z' \in A_j$, with coordinates $z = (a_{1i_1}, \ldots, a_{j-1i_{j-1}}, a_{j+1i_{j+1}}, \ldots, a_{ni_n})$ and $z' = (a_{1i'_1}, \ldots, a_{j-1i'_{j-1}}, a_{j+1i'_{j+1}}, \ldots, a_{ni'_n})$, such that $z \neq z'$ and

$$a_{1i_1} + \ldots + a_{j-1i_{j-1}} + a_{j+1i_{j+1}} + \ldots + a_{ni_n} = a_{1i'_1} + \ldots + a_{j-1i'_{j-1}} + a_{j+1i'_{j+1}} + \ldots + a_{ni'_n}.$$

We may assume without loss of generality that $(i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_n) < (i'_1, \ldots, i'_{j-1}, i'_{j+1}, \ldots, i'_n)$ in lexicographical order. Now, since $z' \in A_j$, there must exist elements $u \in S$ and $d \in B_j$ such that $u = a_{1i'_1} + \ldots + a_{j-1i'_{j-1}} + d + a_{j+1i'_{j+1}} + \ldots + a_{ni'_n}$ and $\varphi(u) = (a_{1i'_1}, \ldots, a_{j-1i'_{j-1}}, d, a_{j+1i'_{j+1}}, \ldots, a_{ni'_n}) \in A$. But then we also have that $u = a_{1i_1} + \ldots + a_{j-1i_{j-1}} + d + a_{j+1i_{j+1}} + \ldots + a_{ni_n}$ and, by the hypothesis about the lexicographical order, $(a_{1i_1}, \ldots, a_{j-1i_{j-1}}, d, a_{j+1i_{j+1}}, \ldots, a_{ni_n}) < (a_{1i'_1}, \ldots, a_{j-1i'_{j-1}}, d, a_{j+1i'_{j+1}}, \ldots, a_{ni'_n})$, so $\varphi(u) \neq (a_{1i'_1}, \ldots, a_{j-1i'_{j-1}}, d, a_{j+1i'_{j+1}}, \ldots, a_{ni'_n})$, and we reach a contradiction. With this, the sum of the coordinates of each element in $A_j$ is distinct, but since these coordinates belong to $B_1, \ldots, B_{j-1}, B_{j+1}, \ldots, B_n$, respectively, we have that $|A_j| \leq |B_1 + \ldots + B_{j-1} + B_{j+1} + \ldots + B_n|$ and the proof is complete. □

Notice that Theorem 3.8 is a particular case of Theorem 2.10. Indeed, consider the case when $j = h - 1$ in the statement of Theorem 2.10. In such a case,

$$|B^+_{[h]}| \leq |A' + B^+_{[h]}| \leq \left( \prod_{i=1}^{n} \frac{|A' + B^+_{[h]\setminus\{i\}}|}{|A'|} \right)^{\frac{1}{h-1}} |A'| \leq \left( \prod_{i=1}^{n} |B^+_{[h]\setminus\{i\}}| \right)^{\frac{1}{h-1}}.$$

All the inequalities follow either from the statement of the theorem or from trivial estimates. We recover Theorem 3.8 by taking the $(h-1)$-th power, so this is a Plünnecke-type inequality. However, it is not clear that one can recover this particular case of Theorem 2.10 from Theorem 3.8.

Now one might want to try to prove one of the previous Theorem 3.1 or Theorem 3.2 starting from here, and in such a way obtain an equivalence among all three. However, this has not been achieved so far, and it seems likely that is cannot be done. The reason for this is the fact that, during the last proof, we have only considered subsets of the projections, sorted in lexicographical order. In this way we have reduced the cardinalities of the sets, so it seems unlikely that one can work the other way around.

One may also wonder whether there is an analogue of this theorem in the entropy setting (there is Han's inequality as an analogue, but we refer to an analogue using entropies of sums of random variables). If there was, it may be stated as follows: given $n$ independent discrete random variables $X_1, \ldots, X_n$, we have

$$(n-1)\mathbf{H}(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} \mathbf{H}(X_1 + \ldots + X_{i-1} + X_{i+1} + \ldots + X_n).$$

This problem will be adressed in section 3.4.

## 3.2. Shearer's inequality

From here on, we shall use the more compact notations introduced in Chapter 1 for sumsets and random variables. For projections, we introduce a similar notation, corresponding to the one of random variables: given a set $A$ in the Cartesian product of $n$ spaces, for any subset of indices $S \subseteq [n]$ we write $A_S$ to denote the projection of $A$ onto the coordinates indexed by $S$. In particular, $A = A_{[n]}$. In general, we have that writing $S$ as a subscript will stand for the projection onto the coordinates indexed by $S$, while if this subscript is accompanied by a '$+$' sign as a superscript, then we refer to the sum of the coordinates indexed by $S$. Using this notation, we can write the main theorems from section 3.1 as follows:

- *Theorem 3.1*: $|A_{[n]}|^{n-1} \leq \prod_{i=1}^{n} |A_{[n]\setminus\{i\}}|$.

- *Theorem 3.2*: $(n-1)H(X_{[n]}) \leq \sum_{i=1}^{n} \mathbf{H}(X_{[n]\setminus\{i\}})$.
- *Theorem 3.8*: $|A^+_{[n]}|^{n-1} \leq \prod_{i=1}^{n} |A^+_{[n]\setminus\{i\}}|$.

Generalizations of these results came by using what are known as covers.

**Definition 3.2.** A *k-cover* is a multiset $\mathcal{S}$ of subsets of $[n]$ (this may be noted as $\mathcal{S} \subseteq 2^{[n]}$) such that each element $i \in [n]$ appears in at least $k$ of the members of $\mathcal{S}$. We say that a $k$-cover is *uniform* if every element $i \in [n]$ appears in exactly $k$ members of $\mathcal{S}$. When we have a 1-cover, we will simply call it a cover.

For example, the multiset $\{\{i\} : i \in [n]\}$ is a trivial cover of $[n]$, and with it one has the trivial bounds

$$|A_{[n]}| \leq \prod_{i=1}^{n} |A_i|$$

for projections, or

$$\mathbf{H}(X_{[n]}) \leq \sum_{i=1}^{n} \mathbf{H}(X_i)$$

for entropies. A different example is that of the multiset $\{[n] \setminus \{i\} : i \in [n]\}$, which gives an $(n-1)$-cover. This cover corresponds to the statements of the theorems from section 3.1. It is appropriate then to wonder if similar inequalities can be obtained for different covers of $[n]$.

The first generalization of Han's inequality came soon after the original result, when Shearer [6] proved the following.

**Theorem 3.9 (Shearer).** *Let $X = (X_1, \ldots, X_n)$ be the joint distribution of n discrete random variables. If $\mathcal{S}$ is a uniform k-cover of $[n]$, then*

$$k\mathbf{H}(X) \leq \sum_{S \in \mathcal{S}} \mathbf{H}(X_S).$$

*Proof.* Take a member $S \in \mathcal{S}$, and say that $S = \{i_1, \ldots, i_t\}$, ordered in such a way that $i_1 < i_2 < \ldots < i_t$. We can use the chain rule for entropy (Lemma 1.5) to write

$$\mathbf{H}(X_S) = \mathbf{H}(X_{i_1}, \ldots, X_{i_t}) = \sum_{j=1}^{t} \mathbf{H}(X_{i_j} | X_{i_1}, \ldots, X_{i_{j-1}}) \geq \sum_{j=1}^{t} \mathbf{H}(X_{i_j} | X_1, X_2, \ldots, X_{i_j - 1}),$$

where the inequality comes from the dropping condition (Lemma 1.6). Now, summing over all $S \in \mathcal{S}$ we have that

$$\sum_{S \in \mathcal{S}} \mathbf{H}(X_S) \geq k \sum_{i=1}^{n} \mathbf{H}(X_i | X_1, \ldots, X_{i-1}) = k\mathbf{H}(X_1, \ldots, X_n),$$

where the factor $k$ in the equality appears because $\mathcal{S}$ is a uniform $k$-cover, that is, each element $i \in [n]$ appears exactly $k$ times, and the inequality comes from a second use of the chain rule. □

The proof is, indeed, very similar to that of Han's inequality, so we observe that it is a very direct generalization. Notice that, in fact, it is enough to have a $k$-cover, not necessarily uniform. This is a consecuence of the fact that $\mathbf{H}(X_S)$ is monotone increasing on $S$, so if we make any of the members of $\mathcal{S}$ bigger we are only incresing the right hand side of the inequality.

The generalization of the projection inequality of Loomis and Whitney is known as the uniform cover inequality and, as happened before, can be proved directly using its entropy analogue.

**Theorem 3.10 (Uniform cover inequality).** *Let $n \geq 2$, let $B_1, \ldots, B_n$ be arbitrary finite sets, and let $A \subseteq B_1 \times \ldots \times B_n$ be a subset of their Cartesian product. If $\mathcal{S}$ is a uniform $k$-cover of $[n]$, then*

$$|A|^k \leq \prod_{S \in \mathcal{S}} |A_S|.$$

*Proof.* Define a random variable $X = (X_1, \ldots, X_n)$ with a uniform distribution over $A$. By Shearer's inequality,

$$k \log |A| = k\mathbf{H}(X) \leq \sum_{S \in \mathcal{S}} \mathbf{H}(X_S) \leq \sum_{S \in \mathcal{S}} \log |A_S|,$$

whence the statement follows by exponentiating. □

As in the case of Shearer's inequality, notice that it is enough to have a $k$-cover. Indeed, given any uniform $k$-cover $\mathcal{S}$, by adding more elements to its members one simply increases the right hand side of the inequality. The reason why this result is known as the uniform cover inequality comes from its original formulation for bodies, where uniformity is needed because the size of a projection with a set of indices $S$ may be smaller than one, and by duplicating this set one can make the right hand side product smaller than any positive number.

Using the same approach as Ruzsa, Balister and Bollobás [1] proved that the uniform cover inequality and Shearer's inequality are in fact equivalent.

**Theorem 3.11.** *Theorem 3.9 and Theorem 3.10 are equivalent.*

*Proof.* The fact that Shearer's inequality implies the uniform cover inequality is trivial, as this is how we showed the second statement. We now prove the converse.

Let $X = (X_1, \ldots, X_n)$ be the joint distribution of $n$ random variables, assuming $X_i$ takes values in $B_i$. First, assume that the random variable takes only finitely many values, each of them with a rational probability. Take a value of $k$ such that $kp_x \in \mathbb{N}$ for all $x \in \text{range}(X)$, and build the set $A \subseteq \mathcal{X}^k$ as described in the previous section, so that $\log |A| = k\mathbf{H}(X) + O(\log k)$ by (18).

Now, given any set of indices $S \subseteq [n]$, we may define the projection $\pi_S$ as the projection onto the coordinates indexed by $S$. For $A \subseteq (B_1 \times \ldots \times B_n)^k \cong B_1^k \times \ldots \times B_n^k$ we consider the projection $\pi_S^k$ analogous to that of the proof of Theorem 3.3. In such a way, we have the diagram

$$\begin{array}{ccc} X & \xrightarrow{\text{make set}} & A \\ \Big\downarrow{\scriptstyle \pi_S} & & \Big\downarrow{\scriptstyle \pi_S^k} \\ X_S & \xrightarrow{\text{make set}} & A_S \end{array}$$

which, again, is commutative, as we already saw. Then, $\log |A_S| = k\mathbf{H}(X_S) + O(\log k)$ for all $S \subseteq [n]$. By the uniform cover inequality, given a $k$-cover $\mathcal{S}$ of $[n]$ we have that

$$\begin{aligned} |A|^k \leq \prod_{S\in\mathcal{S}} |A_S| &\implies k \log |A| \leq k \sum_{S\in\mathcal{S}} \log |A_S| \\ &\implies k\mathbf{H}(X) + O(\log k) \leq k \sum_{S\in\mathcal{S}} \mathbf{H}(X_S) + O(\log k) \\ &\implies \mathbf{H}(X) \leq \sum_{S\in\mathcal{S}} \mathbf{H}(X_S) + O\left(\frac{\log k}{k}\right), \end{aligned}$$

whence the result follows by letting $k$ tend to infinity.

The passage to the general case follows from a routine limiting argument. □

Finally, we turn to the sumset analogue of Shearer's inequality and the uniform cover inequality. As happened with the other results in this section, the generalization follows in a very natural way, and using the same type of arguments.

**Theorem 3.12.** *Let $B_1, \ldots, B_n$ be finite nonempty sets in an additive group. Given a uniform $k$-cover $\mathcal{S}$ of $[n]$, we have*

$$|B_{[n]}^+|^k \leq \prod_{S\in\mathcal{S}} |B_S^+|.$$

*Proof.* Again, we first give an order to the elements of the sets, and we embed $B_{[n]}^+$ into $B = B_1 \times \ldots \times B_n$ by mapping $s \in B_{[n]}^+$ to the minimal element (in lexicographical order) of $B$ whose coordinates add up to $s$. Let this embedding be noted as $\varphi$, and call $A = \varphi(B_{[n]}^+) \subseteq B_1 \times \ldots \times B_n$. In more generality, define maps $\varphi_S : B_S^+ \to B_S$ for each $S \subseteq [n]$ (again, by taking each element to the lexicographically minimal whose sum of coordinates gives the element). Note that $|B_{[n]}^+| = |A|$ and, in general, $|B_S^+| = |\varphi_S(B_S^+)|$. Now let $\mathcal{S}$ be a uniform $k$-cover of $[n]$ and apply Theorem 3.10 to this set $A$, which results in

$$|A|^k \leq \prod_{S\in\mathcal{S}}^{n} |A_S|.$$

Finally, we want to see that $|A_S| \leq |B_S^+|$.

It is enough to see that $A_S \subseteq \varphi_S(B_S^+)$, as $|B_S^+| = |\varphi_S(B_S^+)|$. Indeed, assume $S = \{i_1, \ldots, i_{|S|}\}$ and take $z = (z_1, \ldots, z_n) \in A$, so that it is the lexicographically minimal element of $B$ whose coordinates sum up to some $s \in B_{[n]}^+$. Then, $(z_{i_1}, \ldots, z_{i_{|S|}})$ must also be lexicographically minimal among those elements of $B_S$ whose coordinates have the same sum as this one. Indeed, assume there was a lexicographically smaller element $(z'_{i_1}, \ldots, z'_{i_{|S|}})$ with the same sum. Then, taking $z'_i = z_i$ for all $i \notin S$, we would have that $\sum_{i=1}^n z_i = \sum_{i=1}^n z'_i$ but $(z'_1, \ldots, z'_n) < (z_1, \ldots, z_n)$, which contradicts the fact that $(z_1, \ldots, z_n) \in A$. □

As we did with Theorem 3.8, we may now want to compare this result to Theorem 2.10. We can use the same arguments as before to obtain a weaker bound from Theorem 2.10. If we are given the multiset of all sets of size $k$, then

$$|B_{[n]}^+| \leq |A' + B_{[n]}^+| \leq \left( \prod_{S \in \mathcal{S}} \frac{|A' + B_S^+|}{|A'|} \right)^{\frac{1}{\binom{n-1}{k-1}}} |A'| \leq \left( \prod_{S \in \mathcal{S}} |B_S^+| \right)^{\frac{1}{\binom{n-1}{k-1}}}$$
$$\implies |B_{[n]}^+|^{\binom{n-1}{k-1}} \leq \prod_{S \in \mathcal{S}} |B_S^+|.$$

But the multiset of all sets of size $k$ results in a uniform $\binom{n-1}{k-1}$-cover, so Theorem 3.12 gives the same bound as above. One must note that the bound is slightly weaker than that of Theorem 2.10 but, on the other hand, it extends the result to general uniform $k$-covers. It is reasonable to argue, then, that this entropic result is more general than the former.

As happened with Shearer's inequality and the uniform cover lemma, uniformity is not needed in Theorem 3.12, as each of the sets has size at least 1. Also, this again gives rise to the following question: given $n$ independent random variables $X_1, \ldots, X_n$ and a (uniform) $k$-cover $\mathcal{S}$ of $[n]$, does

$$k\mathbf{H}(X_{[n]}^+) \leq \sum_{S \in \mathcal{S}} \mathbf{H}(X_S^+)$$

hold? This, again, will be answered in section 3.4.

## 3.3. Compressions and fractional covers

### 3.3.1. Madiman and Tetali's work

Recently, Madiman and Tetali [20] strengthened Shearer's inequality by introducing conditional entropies. Furthermore, they were able to use their approach to obtain a lower bound as well. Their result may be stated as follows.

**Theorem 3.13 (Madiman, Tetali).** *Let $X = (X_1, \dots, X_n)$ be a sequence of n random variables such that $\mathbf{H}(X)$ is finite, and let $\mathcal{S}$ be a uniform k-cover of $[n]$. For each $S \subseteq [n]$ with minimal element $a \geq 1$ and maximal element $b \leq n$, define $S_* = \{1, \dots, a-1\}$ and $S^* = \{i \notin S : 1 \leq i \leq b-1\}$. Then,*

$$\sum_{S \in \mathcal{S}} \mathbf{H}(X_S | X_{S^*}) \leq k\mathbf{H}(X) \leq \sum_{S \in \mathcal{S}} \mathbf{H}(X_S | X_{S_*}).$$

In fact, what they proved is a stronger result, from which Theorem 3.13 follows as a simple corollary. In order to prove this result, we must first provide a generalization of the definition of covers given in section 3.2.

**Definition 3.3.** Consider a multiset $\mathcal{S}$ of subsets of $[n]$. A function $\alpha : \mathcal{S} \to \mathbb{R}_+$ such that $\alpha(S) = \alpha_S$ is called a *fractional cover* if for each $i \in [n]$ we have that

$$\sum_{S \in \mathcal{S} : i \in S} \alpha_S \geq 1.$$

Similarly, we say that a function $\beta : \mathcal{S} \to \mathbb{R}_+$ with $\beta(S) = \beta_S$ is called a *fractional packing* if for each $i \in [n]$ we have that

$$\sum_{S \in \mathcal{S} : i \in S} \alpha_S \leq 1.$$

Finally, a function $\gamma : \mathcal{S} \to \mathbb{R}_+$ with $\gamma(S) = \gamma_S$ is called a *fractional partition* if for each $i \in [n]$,

$$\sum_{S \in \mathcal{S} : i \in S} \alpha_S = 1.$$

In particular, observe that if a collection of subsets of $[n]$, $\mathcal{S}$, is a cover, then the trivial function taking $\alpha_S = 1$ for all $S \in \mathcal{S}$ is a fractional cover. Furthermore, if $\mathcal{S}$ is a uniform $k$-cover of $[n]$, then taking $\alpha_S = \frac{1}{k}$ for every $S \in \mathcal{S}$ results in a fractional partition of $[n]$ using $\mathcal{S}$.

With this, we may already prove the more general result which follows.

**Theorem 3.14.** *Let $X = (X_1, \dots, X_n)$ be a sequence of n random variables such that $\mathbf{H}(X)$ is finite, and consider a collection $\mathcal{S}$ of subsets of $[n]$. For each $S \subseteq [n]$ with minimal element $a \geq 1$ and maximal element $b \leq n$, define $S_* = \{1, \dots, a-1\}$ and $S^* = \{i \notin S : 1 \leq i \leq b-1\}$. Given any fractional covering $\alpha$,*

*fractional packing $\beta$, and fractional partition $\gamma$, the following two statements hold:*

$$\sum_{S\in\mathcal{S}} \beta_S \mathbf{H}(X_S|X_{S^*}) \leq \mathbf{H}(X_{[n]}) \leq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_S|X_{S_*}),$$
$$\sum_{S\in\mathcal{S}} \gamma_S \mathbf{H}(X_S|X_{S^*}) \leq \mathbf{H}(X_{[n]}) \leq \sum_{S\in\mathcal{S}} \gamma_S \mathbf{H}(X_S|X_{S_*}).$$

*Proof.* We first prove the second statement, and see that the proof of the first is done in the same way. Consider the chain rule given in Lemma 1.5. If we consider a further conditioning in both sides of the equality, it still holds (we may think of this as applying the chain rule to a random variable which is already conditioned). Hence, we may write

$$\mathbf{H}(X_S|X_{S_*}) = \sum_{j\in S} \mathbf{H}(X_j|X_{S\cap[j-1]}, X_{S_*}).$$

With this,

$$\begin{aligned}
\sum_{S\in\mathcal{S}} \gamma_S \mathbf{H}(X_S|X_{S_*}) &= \sum_{S\in\mathcal{S}} \gamma_S \sum_{j\in S} \mathbf{H}(X_j|X_{S\cap[j-1]}, X_{S_*}) \\
&\geq \sum_{S\in\mathcal{S}} \gamma_S \sum_{j\in S} \mathbf{H}(X_j|X_{[j-1]}) \\
&= \sum_{j=1}^{n} \mathbf{H}(X_j|X_{[j-1]}) \sum_{S\in\mathcal{S}} \gamma_S \mathbb{1}_{j\in S} \\
&= \sum_{j=1}^{n} \mathbf{H}(X_j|X_{[j-1]}) = \mathbf{H}(X_{[n]}),
\end{aligned}$$

where the inequality comes from the dropping condition Lemma 1.6, the next equality comes from interchanging the sums, the next from the definition of a fractional packing, and the last, from the chain rule (Lemma 1.5). Thus finishes the proof of the upper bound. Notice that, if we had a fractional covering $\alpha$, the second to last equality would be an inequality, and that would be the proof for the upper bound in the first expression.

We now turn our attention to the lower bound. It is proved in a similar way. As before, from the chain rule we may write

$$\mathbf{H}(X_S|X_{S^*}) = \sum_{j\in S} \mathbf{H}(X_j|X_{S\cap[j-1]}, X_{S^*}),$$

and using this,

$$\begin{aligned}\sum_{S\in\mathcal{S}}\gamma_S\mathbf{H}(X_S|X_{S^*}) &= \sum_{S\in\mathcal{S}}\gamma_S\sum_{j\in S}\mathbf{H}(X_j|X_{S\cap[j-1]},X_{S^*})\\ &\leq \sum_{S\in\mathcal{S}}\gamma_S\sum_{j\in S}\mathbf{H}(X_j|X_{[j-1]})\\ &= \sum_{j=1}^{n}\mathbf{H}(X_j|X_{[j-1]})\sum_{S\in\mathcal{S}}\gamma_S\mathbb{1}_{j\in S}\\ &= \sum_{j=1}^{n}\mathbf{H}(X_j|X_{[j-1]}) = \mathbf{H}(X_{[n]}),\end{aligned}$$

where each of the inequalities and equalities follow in the same way as before, completing the proof of the lower bound. In the case of a fractional packing, we again have that one more inequality appears in the chain of inequalities. □

In fact, Madiman and Tetali worked in a more general setting than this. Instead of working with entropy, they work with a function $f$, under the assumption that $f(\varnothing) = 0$ and that $f$ is submodular. With these two conditions, one can prove that many of the basic properties of $\mathbf{H}$, such as the chain rule or the dropping condition, actually hold in a more general setting, so one can obtain results for any function for which the conditions hold. As the topic of this thesis is the application of entropy, we have only showed this particular case.

Observe that, indeed, Theorem 3.13 follows from Theorem 3.14 by considering that the uniform $k$-cover gives a fractional partition when taking $\gamma_S = \frac{1}{k}$ for all $S \in \mathcal{S}$.

### 3.3.2. Balister and Bollobás's work

Balister and Bollobás [1] obtained a further generalization of Theorem 3.13, and their approach turned out to be rather simple: one simply needs to define a certain poset, and use basic properties of entropy. This poset is given by the definition of compressions of multisets.

**Definition 3.4.** Let $\mathcal{M}_{n,m}$ be the family of multisets of nonempty subsets of $[n]$ with a total of $m$ elements (the sum of the number of elements in each set in each multiset of the family is $m$). Given a multiset $\mathcal{S} = \{S_1,\dots,S_l\} \in \mathcal{M}_{n,m}$ with non-nested sets $S_i$ and $S_j$ (meaning that neither $S_i \subseteq S_j$ nor $S_j \subseteq S_i$), let $\mathcal{S}' = \mathcal{S}_{(ij)}$ be obtained by replacing $S_i$ and $S_j$ by $S_i \cup S_j$ and $S_i \cap S_j$, keeping only $S_i \cup S_j$ when $S_i \cap S_j = \varnothing$. We then say that $\mathcal{S}'$ is an *elementary compression* of $\mathcal{S}$. The result of a sequence of elementary compressions is called a *compression*.

Note that the need for sets $S_i$ and $S_j$ in the definition to be non-nested comes from the fact that we want $\mathcal{S}$ to be different from $\mathcal{S}'$. If one set was nested in the other, then their union and intersection would remain the same as the original sets.

Now let us define a partial order on $\mathcal{M}_{n,m}$ by setting $\mathcal{S} > \mathcal{T}$ when $\mathcal{T}$ is a compression of $\mathcal{S}$. The fact that this defines a partial order follows from the fact that, if $\mathcal{S}'$ is an elementary compression of $\mathcal{S}$, then

$$\sum_{S\in\mathcal{S}} |S|^2 < \sum_{S\in\mathcal{S}'} |S|^2.$$

One can notice that for every multiset $\mathcal{S} \in \mathcal{M}_{n,m}$ there exists a unique minimal multiset $\mathcal{S}^\sharp$ dominated by $\mathcal{S}$. This minimal multiset consists of the sets $S_j^\sharp = \{i \in [n] : i \text{ lies in at least } j \text{ of the sets } S \in \mathcal{S}\}$. As it turns out, $\mathcal{S}^\sharp$ is the unique multiset that is totally ordered by inclusion and has the same multiset union as $\mathcal{S}$. We encourage the reader to check these simple facts.

**Example 3.1.** Consider $\mathcal{S} = \{\{1,2\},\{1,3,4\},\{2,4\}\} \in \mathcal{M}_{4,7}$. An elementary compression with the first two sets results in $\mathcal{S}' = \{\{1,2,3,4\},\{1\},\{2,4\}\}$. A second elementary compression gives $\mathcal{S}'' = \{\{1,2,3,4\},\{1,2,4\}\}$. No more elementary compressions are possible, so $\mathcal{S}^\sharp = \mathcal{S}''$.

With only this, we may already prove a very strong generalization of Shearer's inequality.

**Theorem 3.15 (Balister, Bollobás).** *Let $X = (X_1, \dots, X_n)$ be a sequence of random variables with finite entropy, and let $\mathcal{S}$ and $\mathcal{T}$ be finite multisets of subsets of $[n]$. If $\mathcal{S} > \mathcal{T}$, then*

$$\sum_{S\in\mathcal{S}} \mathbf{H}(X_S) \geq \sum_{T\in\mathcal{T}} \mathbf{H}(X_T).$$

*Proof.* It is enough to check that the inequality holds for an elementary compression. Assume that $\mathcal{S} = \{S_1, \dots, S_l\}$ and $\mathcal{T} = \mathcal{S}_{(ij)}$ for some $i \neq j$. Then, the statement is equivalent to

$$\mathbf{H}(X_{S_i}) + \mathbf{H}(X_{S_j}) \geq \mathbf{H}(X_{S_i\cup S_j}) + \mathbf{H}(X_{S_i\cap S_j}),$$

which follows by submodularity (Lemma 1.7). □

With this, we have a very simple proof of a very powerful result. From here, one may obtain many others.

**Example 3.2.** Consider the same multiset from Example 3.1. If we have four random variables $X_1$, $X_2$, $X_3$ and $X_4$, applying Theorem 3.15 to this multiset and its compressions yields

$$\begin{aligned}\mathbf{H}(X_1,X_2) + \mathbf{H}(X_1,X_3,X_4) + \mathbf{H}(X_2,X_4) &\geq \mathbf{H}(X_1,X_2,X_3,X_4) + \mathbf{H}(X_1) + \mathbf{H}(X_2,X_4)\\ &\geq \mathbf{H}(X_1,X_2,X_3,X_4) + \mathbf{H}(X_1,X_2,X_4).\end{aligned}$$

One can also easily derive Shearer's inequality from this result. Indeed, consider a uniform $k$-cover $\mathcal{S}$ of $[n]$. Then, it is easy to check that $\mathcal{S}^\sharp$ is composed of $k$ copies of $[n]$, so a direct application of the theorem yields

$$k\mathbf{H}(X) \leq \sum_{S\in\mathcal{S}} \mathbf{H}(X_S).$$

Even more, we can easily derive Theorem 3.13.

*Proof of Theorem 3.13.* By Lemma 1.3 we know that $\mathbf{H}(X_S|X_T) = \mathbf{H}(X_{S\cup T}) - \mathbf{H}(X_T)$. Using this, we can rewrite each of the inequalities in the statement. The upper bound becomes

$$k\mathbf{H}(X) + \sum_{S\in\mathcal{S}} \mathbf{H}(X_{S_*}) \leq \sum_{S\in\mathcal{S}} \mathbf{H}(X_{S\cup S_*}).$$

Now we observe that $\mathcal{C}_1 = k\{[n]\} \cup \{S_* : S \in \mathcal{S}\}$ is a totally ordered multiset (by inclusion). Furthermore, since $\mathcal{S}$ is a uniform $k$-cover of $[n]$, it is clear that it has the same multiset union as $\mathcal{C}_2 = \{S \cup S_* : S \in \mathcal{S}\}$, so $\mathcal{C}_1 = \mathcal{C}_2^\sharp$, and the inequality holds by Theorem 3.15.

Similarly, the lower bound can be rewritten as

$$k\mathbf{H}(X) + \sum_{S\in\mathcal{S}} \mathbf{H}(X_{S^*}) \geq \sum_{S\in\mathcal{S}} \mathbf{H}(X_{S\cup S^*}).$$

Here, we have that $\mathcal{C}_3 = \{S \cup S^* : S \in \mathcal{S}\}$ is a totally ordered multiset. Indeed, if we denote $b_S = \max\{i : i \in S\}$, we have that $\mathcal{C}_3 = \{[b_S] : S \in \mathcal{S}\}$. Furthermore, it has the same multiset union as $\mathcal{C}_4 = k\{[n]\} \cup \{S^* : S \in \mathcal{S}\}$, so $\mathcal{C}_3 = \mathcal{C}_4^\sharp$, and the inequality follows by Theorem 3.15. □

It is important to note, however, that this general result in entropy theory does no longer have an analogue with projections. If it did, it would be stated along the lines *Let $B_1, \ldots, B_n$ be arbitrary finite sets, and let $A \subseteq B_1 \times \ldots \times B_n$ be a subset of their Cartesian product. Let $\mathcal{S}$ and $\mathcal{T}$ be finite multisets of subsets of $[n]$. If $\mathcal{S} > \mathcal{T}$, then*

$$\prod_{S\in\mathcal{S}} |A_S| \geq \prod_{T\in\mathcal{T}} |A_T|.$$

However, there are simple counterexamples for this. For instance, let $A = \{(0,0,0), (1,0,0), (0,1,0), (1,1,0), (0,0,1)\} \subseteq \mathbb{Z}^3$. It is clear that $|A| = 5$, $|A_1| = |A_2| = |A_3| = 2$, $|A_{\{1,2\}}| = 4$, $|A_{\{1,3\}}| = |A_{\{2,3\}}| = 3$. Then we have that $10 = |A||A_3| > |A_{1,3}||A_{2,3}| = 9$, but $\{\{1,2,3\}, \{3\}\} < \{\{2,3\}, \{1,3\}\}$, which contradicts the statement.

Along with Theorem 3.12, Balister and Bollobás also presented a generalization of the sumset theorem, in which they take subsets of the sumsets. Their result is as follows.

**Theorem 3.16.** *Let $A, B_1, \ldots, B_n$ be finite sets in an abelian group. Consider a set $D \subseteq B^+_{[n]}$. Then,*

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

*Proof.* Call $B_{n+1} = A$, and for each $S \subseteq [n+1]$ define the map $\varphi_S$ as in the proof of Theorem 3.12 (set of lexicographically minimal representatives of $B_S^+$). Let $A' = \varphi_{[n+1]}(D + B_{n+1})$. In the same way as in the proof of Theorem 3.12, one can check that $|A'_{[n]}| \leq |D|$ and $|A'_{\{i,n+1\}}| \leq |B_i + B_{n+1}|$. Now, consider a uniform $n$-cover given by taking the pairs $\{1, n+1\}, \{2, n+1\}, \ldots, \{n, n+1\}$, and $n-1$ copies of the set $[n]$. Applying Theorem 3.10 yields the desired inequality. □

It is clear that the covers that are considered in this case are much more restrictive than those of the more general theorem. It is natural to wonder whether this can be extended to more general covers. We shall answer this question in section 3.4.

## 3.4. Partition-determined functions

Recently, Madiman, Marcus and Tetali published a paper [19] in which they present a unifying theory for most of the results presented in this chapter. Furthermore, their approach allows them to obtain several very general results in different contexts as simple corollaries. For instance, the very general Theorem 3.15 of Balister and Bollobás will turn out to be a direct corollary of these new results, and we will see that their proofs are also elementary. We now start with all the due definitions in order to present this new theory; several of them correspond to notation that we have already been using, and are simply restated to remind the reader.

### 3.4.1. Definitions and examples

Let $A_1, \ldots, A_n$ be finite sets. We may define $A = A_{[n]} = A_1 \times \ldots \times A_n$. For any $S \subseteq [n]$ we write $A_S = \prod_{i \in S} A_i = \pi_S(A)$ for the projection of $A$ onto the coordinates indexed by $S$. This same projection $\pi_S : A_T \to A_S$ can be defined for every $S \subseteq T \subseteq [n]$ in the natural way: if $S = \{i_1, \ldots, i_{|S|}\}$ (assume the indices are ordered increasingly), for each $a \in A_T$ let $\pi_S(a) = (a_{i_1}, \ldots, a_{i_{|S|}})$.

Write $Q(A_1, A_2, \ldots, A_n)$ to denote the space that is the disjoint union of the spaces $A_S$ corresponding to non-empty $S \subseteq [n]$. This can be written as

$$Q(A_1, A_2, \ldots, A_n) = \bigsqcup_{\varnothing \neq S \subseteq [n]} A_S.$$

Let $B$ be any space and consider a function $f : Q(A_1, \ldots, A_n) \to B$. For any non-empty $S \subseteq [n]$, we define $f_S : A_S \to B$ as the restriction of $f$ to the inputs that come from $A_S$. When given $S \subseteq T \subseteq [n]$ and $a \in T$, we abuse notation and write $f_S(a)$ to mean $f_S(\pi_S(a))$. The reader may think of this as $Q(A_1, \ldots, A_n) = A_{[n]}$ since, for most applications, we will consider functions that are defined

in the same way in each of the projections $A_S$ of the bigger space $A$. However, it is interesting to write the results in the more general way.

**Definition 3.5.** Consider $S \subseteq [n]$, and let $\overline{S} = [n] \setminus S$. We say that a function $f$ defined over $Q(A_1, \dots, A_n)$ is *partition-determined with respect to* $S$ if for every $x, y \in Q(A_1, \dots, A_n)$ we have that $f(x) = f(y)$ whenever $f_S(x) = f_S(y)$ and $f_{\overline{S}}(x) = f_{\overline{S}}(y)$. Informally, one may say that $f$ is partition-determined with respect to $S$ if $f_S(x)$ and $f_{\overline{S}}(x)$ uniquely determine $f(x)$. If we have a collection of subsets $\mathcal{S} \subseteq 2^{[n]}$, we say that $f$ is *partition-determined with respect to* $\mathcal{S}$ if it is partition-determined with respect to every $S \in \mathcal{S}$. Finally, we say that $f$ is *partition-determined* if it is partition-determined with respect to $2^{[n]}$.

**Remark 3.1.** It is reasonable to wonder what happens when defining $f_\varnothing$, which corresponds to taking $S = [n]$ and considering its complement. It is natural to take the projection, considering the full notation, to write $f_\varnothing(x) = f(\pi_\varnothing(x))$, which does not have any actual arguments. Hence, it should be a constant function. Later we will see that this is indeed useful when considering entropies, in which case we will want this function to bring no information.

Consider the following example.

**Example 3.3.** Let $A_1 = A_2 = \ldots = A_n = \{0, 1, \dots, 9\}$ (so we can think of $A_{[n]}$ as numbers with $n$ digits), and consider $f : Q(A_1, \dots, A_n) \to \mathbb{Z}$ to compute the sum of the digits. It is clear that, if we have the sum of a certain set of digits, and the sum of all the others, the overall sum is computed as the sum of these two. Hence, the sum of the digits is a partition-determined function.

There are many more examples of partition-determined functions. For instance, we present two which we have been using so far: projections and sumsets.

**Lemma 3.17.** *Let $V$ be a vector space over the reals with basis vectors $\{v_1, \dots, v_n\}$. Let $A_1, \dots, A_n \subseteq \mathbb{R}$ be sets of real numbers, and define a function $f : Q(A_1, \dots, A_n) \to V$ by*

$$f_S(a) = \sum_{i \in S} \pi_i(a) v_i$$

*for every $S \subseteq [n]$. Then, $f$ is partition-determined.*

*Proof.* Consider a subset $T \subseteq [n]$ and take $a \in A_T$. Consider a collection of subsets $\mathcal{S} \subseteq 2^{[n]}$. For any $S \in \mathcal{S}$ we have that

$$f(a) = \sum_{i \in T} \pi_i(a) v_i = \sum_{i \in S \cap T} \pi_i(a) v_i + \sum_{i \in \overline{S} \cap T} \pi_i(a) v_i = f_S(a) + f_{\overline{S}}(a),$$

so knowing $f_S(a)$ and $f_{\overline{S}}(a)$ uniquely determines $f(a)$. This can be done for any collection $\mathcal{S}$. □

**Lemma 3.18.** *Let $(G,+)$ be an additive group, and let $A_1,\dots,A_n \subseteq G$ be finite. Let $c_1,\dots,c_n \in \mathbb{Z}$. Define a function $f : Q(A_1,\dots,A_n) \to G$ by*

$$f_S(a) = \sum_{i\in S} c_i \pi_i(a)$$

*for every $S \subseteq [n]$. Then, $f$ is partition-determined.*

*Proof.* It works in the same way as before. Consider a subset $T \subseteq [n]$ and take $a \in A_T$, and consider a collection of subsets $\mathcal{S} \subseteq 2^{[n]}$. For any $S \in \mathcal{S}$ we have that

$$f(a) = \sum_{i\in T} c_i \pi_i(a) = \sum_{i\in S\cap T} c_i \pi_i(a) + \sum_{i\in \overline{S}\cap T} c_i \pi_i(a) = f_S(a) + f_{\overline{S}}(a),$$

so knowing $f_S(a)$ and $f_{\overline{S}}(a)$ uniquely determines $f(a)$. □

The following is an example of a function that is not partition-determined.

**Example 3.4.** As in Example 3.3, consider the sets $A_1 = A_2 = \ldots = A_n = \{0,1,\dots,9\}$. Now, the function $f$ computes the number of distinct digits of any element in $Q(A_1,\dots,A_n)$. This function is clearly not partition-determined. For instance, assume $n = 4$, $S = \{1,2\}$, and $a = (1,0,1,2) \in A_{[n]}$. It is clear that $f_S(a) = 2$ and $f_{\overline{S}}(a) = 2$ while $f(a) = 3$. However, for $a' = (1,0,1,0)$ and $a'' = (1,0,2,3)$ we have the same values for $f_S$ and $f_{\overline{S}}$, while $f(a') = 2$ and $f(a'') = 4$.

Partition determined functions can have an even stronger property.

**Definition 3.6.** We say that $f : Q(A_1,\dots,A_n) \to B$ is *strongly partition-determined* if it is partition-determined and for any $a \in A_{[n]}$ and disjoint sets $S,T \subseteq [n]$ one has that $f_{S\cup T}(a)$ and $f_T(a)$ completely determine $f_S(a)$.

For example, the function defined in Example 3.3 is strongly partition-determined. So are the functions defined in Lemma 3.17 and Lemma 3.18.

Madiman, Marcus and Tetali were able to prove several entropic results for these partition-determined functions. From them, they obtain several results for sums, projections, and others, as corollaries of the general results they obtain, bounding the cardinalities of compound sets.

**Definition 3.7.** A *compound set* is a set of the form $\{f(a_1,\dots,a_n) : a_1 \in A_1,\dots,a_n \in A_n\}$, where the sets $A_1,\dots,A_n$ are subsets of an appropriate algebraic structure so that $f$ is well defined. This compound set is denoted as $f(A_1,\dots,A_n)$.

Obviously, a similar notation can be used when considering random variables. The image in this case is not a set, but a new random variable, which we may write as $f(X_1,\dots,X_n)$. Similarly, we would have $f_S(X_1,\dots,X_n)$ to be a random variable for each $S \subseteq [n]$. For ease of notation, however, we will usually note this random variable simply as $f_S$.

### 3.4.2. General results

Once the definitions have been established, one can obtain several entropy inequalities in the very general setting of partition-determined functions. In particular, we shall prove three very strong results. We first start with a useful lemma.

**Lemma 3.19.** *Suppose that $A_1, \dots, A_n$ are finite sets, and consider a partition-determined function $f : Q(A_1, \dots, A_n) \to B$, where $B$ is any space. Let $X_1, \dots, X_n$ be random variables such that $X_i$ ranges over $A_i$. Then, for any disjoint sets $S, T \subseteq [n]$,*

$$\mathbf{I}(f_{S\cup T}; f_T) \geq \mathbf{H}(f_{S\cup T}) - \mathbf{H}(f_S)$$

*holds, with equality if $f$ is strongly partition-determined and the variables $X_1, \dots, X_n$ are independent.*

*Proof.* Since conditioning reduces entropy (Lemma 1.6), we have that

$$\mathbf{H}(f_{S\cup T}) - \mathbf{H}(f_S) \leq \mathbf{H}(f_{S\cup T}) - \mathbf{H}(f_S|f_T).$$

Since $f$ is partition-determined, we have that $f_{S\cup T} = \phi(f_S, f_T)$ for some function $\phi$. Then, by the data processing inequality (Lemma 1.10) and a trivial property of entropy we have that

$$\mathbf{H}(f_S|f_T) = \mathbf{H}(f_S, f_T|f_T) \geq \mathbf{H}(f_{S\cup T}|f_T).$$

Substituting this above readily yields the desired inequality, by the definition (6) of mutual information. For the case of equality, observe that we have equality in the first inequality when the random variables are independent, and equality in the second when the function is strongly partition-determined (in such a case, the function $\phi$ corresponds to a bijection, so no information is lost and the entropy remains the same). □

We can now prove a submodularity result, in the same spirit as Lemma 1.7, for partition-determined functions. This is a simple result that follows from the previous lemma and other basic properties of mutual information and entropy.

**Theorem 3.20 (Submodularity for strongly partition-determined functions).** *Let $A_1, \dots, A_n$ be finite sets, and take a strongly partition-determined function $f : Q(A_1, \dots, A_n) \to B$. Let $X_1, \dots, X_n$ be independent random variables, with $X_i$ taking values in $A_i$. Then, for any nonempty sets $S, T \subseteq [n]$,*

$$\mathbf{H}(f_{S\cup T}) + \mathbf{H}(f_{S\cap T}) \leq \mathbf{H}(f_S) + \mathbf{H}(f_T).$$

*Proof.* First of all, notice that it is enough to prove that the statement holds when $n = 3$. This is so because any other case can be written using only three random variables, considering each of

them as a joint distribution of several others. For this particular case we have that

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

where the last equality comes from Lemma 3.19. We want to see that this quantity is nonnegative. Indeed, we have that

$$\mathbf{I}(f_{\{1,2,3\}}; f_{\{2\}}) \leq \mathbf{I}(f_{\{1,2\}}, f_{\{3\}}; f_{\{2\}}) = \mathbf{I}(f_{\{1,2\}}; f_{\{2\}}) + \mathbf{I}(f_{\{3\}}; f_{\{2\}} | f_{\{1,2\}}) = \mathbf{I}(f_{\{1,2\}}; f_{\{2\}}), \tag{19}$$

where the inequality comes from the fact that $f$ is partition determined and the data processing inequality, the first equality corresponds to (7), and the second comes as a consequence of the independence of the random variables. The result follows by reordering the terms. □

**Remark 3.2.** Here we realise that we must define $\mathbf{H}(f_\varnothing)$ for the case when $S$ and $T$ are disjoint. Corresponding to what we said in Remark 3.1, we have that this entropy is zero, which is what we needed to enssure that the above statement makes sense.

Notice that Lemma 1.7 corresponds to the case when $f$ is the identity function (the projection to the coordinates indexed by $S$ and $T$).

The next result is stated in a very similar way to Theorem 3.15. In fact, it is its straightforward generalization for strongly partition-determined functions, once we have been able to show their submodularity. It is based in the same partial order defined using compressions over multisets of subsets of $[n]$.

**Theorem 3.21.** *Let $X_1, \dots, X_n$ be a sequence of independent random variables taking values in $A_1, \dots, A_n$, respectively, where the $A_i$ are finite sets, and let $f : Q(A_1, \dots, A_n) \to B$ be a strongly partition-determined function. Let $\mathcal{S}$ and $\mathcal{T}$ be finite multisets of subsets of $[n]$. If $\mathcal{S} > \mathcal{T}$ with the partial order defined by compressions, then*

$$\sum_{S \in \mathcal{S}} \mathbf{H}(f_S) \geq \sum_{T \in \mathcal{T}} \mathbf{H}(f_T).$$

*Proof.* The proof follows that of Theorem 3.15. First, we observe that it is enough to prove it for an elementary compression, because of the transitivity of the partial order. But in the case of an elementary compression we have that $\mathcal{T} = \mathcal{S}_{(ij)}$ for some $i \neq j$, and since most terms cancel out, we simply have to prove that

$$\mathbf{H}(f_{S_i}) + \mathbf{H}(f_{S_j}) \geq \mathbf{H}(f_{S_i \cup S_j}) + \mathbf{H}(f_{S_i \cap S_j}),$$

which follows by submodularity (Theorem 3.20). □

Observe that Theorem 3.15 is just a particular case of this new theorem, the case in which $f$ is taken to be the identity (or projection) function. Many other results can be obtained by considering different functions. We refer to this later on.

In order to prove one final general result, we use again the fractional covers introduced in section 3.3. With them, we can prove the following, which can be though of as a different generalization of Shearer's inequality. It is the generalization of Theorem 3.14 to the case when we use entropies of partition-determined functions.

**Theorem 3.22.** *Let $A_1, \dots, A_n$ be finite sets, and let $f : Q(A_1, \dots, A_n) \to B$ be a strongly partition-determined function. Consider independent random variables $X_1, \dots, X_n$ such that $X_i$ ranges over $A_i$. Then, for any fractional covering $\alpha$ using any collection $\mathcal{S}$ of subsets of $[n]$,*

$$\mathbf{H}(f_{[n]}) \leq \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(f_S).$$

*Proof.* The proof closely follows that of Theorem 3.14. As was stated in section 3.3, this result actually holds for any submodular function $g$ such that $g(\varnothing) = 0$. In particular, we may define $g(S) = \mathbf{H}(f_S)$, which means that $g(\varnothing) = 0$ by the convention we have taken. Furthermore, define $g(S|T) = g(S \cup T) - g(T)$, and call this $g$ conditional on $T$ (assuming $S$ and $T$ are disjoint). By the chain rule (we encourage the reader to prove the chain rule for general submodular functions such that $g(\varnothing) = 0$, which can be proved by induction), we may write

$$g(S) = \sum_{i \in S} g(\{i\}|S \cap [i-1]),$$

and using this,

$$\begin{aligned}
\sum_{S \in \mathcal{S}} \alpha_S g(S) &= \sum_{S \in \mathcal{S}} \alpha_S \sum_{i \in S} g(\{i\}|S \cap [i-1]) \\
&\geq \sum_{S \in \mathcal{S}} \alpha_S \sum_{i \in S} g(\{i\}|[i-1]) \\
&= \sum_{i=1}^{n} g(\{i\}|[i-1]) \sum_{S \in \mathcal{S}} \alpha_S \mathbb{1}_{i \in S} \\
&\geq \sum_{i=1}^{n} g(\{i\}|[i-1]) = g([n]),
\end{aligned}$$

where the first inequality comes from the dropping condition, the first equality is an interchange of the sums, the second inequality is a result of the definition of a fractional covering, and the last equality is a second use of the chain rule (Lemma 1.5). □

### 3.4.3. Corollaries

The three theorems from the previous subsection are the key elements for the remainder of this chapter. We shall see how we can obtain some very general results in very different settings. We begin with entropy corollaries: the first example is obtained when taking $f$ to be the identity function. In this case we obtain many of the results that have already been proved throughout this thesis: Theorem 3.20 becomes Lemma 1.7, Theorem 3.21 becomes Theorem 3.15, and Theorem 3.22 turns into a part of Theorem 3.14. We have already studied all these results, but with this we have presented a unified way to prove them.

Now let us consider a different function. In this case, let us assume that $A_1, \dots, A_n$ are sets in an abelian group $(G, +)$, and let $f$ be the sum function (which is strongly partition-determined, as we proved in Lemma 3.18).

**Corollary 3.23.** *Let $X_1, \dots, X_n$ be independent random variables taking values in a commutative group $(G, +)$. Then,*

*(i) The set function $\mathbf{H}(X_S^+)$ is submodular, that is, for any sets $S, T \subseteq [n]$,*

$$\mathbf{H}(X_{S\cup T}^+) + \mathbf{H}(X_{S\cap T}^+) \leq \mathbf{H}(X_S^+) + \mathbf{H}(X_T^+).$$

*(ii) Let $\mathcal{S}$ and $\mathcal{T}$ be two collections of subsets of $[n]$. If $\mathcal{S} > \mathcal{T}$, then*

$$\sum_{S\in\mathcal{S}} \mathbf{H}(X_S^+) \geq \sum_{T\in\mathcal{T}} \mathbf{H}(X_T^+).$$

*(iii) For any fractional covering $\alpha$ using any collection $\mathcal{S}$ of subsets of $[n]$,*

$$\mathbf{H}(X_{[n]}^+) \leq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_S^+).$$

*Proof.* Substitute $f$ for the sum of the variables in Theorem 3.20, Theorem 3.21, and Theorem 3.22, respectively. □

Some of these results had already been stated before. For instance, the submodularity of the entropy of sums of random variables was already discussed by Madiman in [18]. He also mentioned (ii) and stated and proved (iii) of the previous corollary. The approach we have showed has the advantage that it gives the same inequalities for a very wide class of functions, and not only for sums.

Notice, in particular, that if $\mathcal{S}$ is a uniform $k$-cover of $[n]$ and we let $\alpha_S = \frac{1}{k}$ for all $S \in \mathcal{S}$, then item (iii) positively answers the questions stated at the end of section 3.1 and section 3.2. Furthermore, item (i) serves to give an entropy analogue of Ruzsa's twin to the triangle inequality. This result states the following.

**Proposition 3.24.** *Let $A$, $B$ and $C$ be finite non-empty sets in a (not necessarily commutative) group. Then,*

$$|A||B+C| \leq |B+A||A+C|.$$

Its entropy analogue, in the abelian setting, can be stated as follows.

**Proposition 3.25.** *Let $X$, $Y$ and $Z$ be independent random variables with values in any commutative group. Then,*

$$\mathbf{H}(X) + \mathbf{H}(Y+Z) \leq \mathbf{H}(X+Y) + \mathbf{H}(X+Z).$$

*Proof.* Consider item (i) in Corollary 3.23. By submodularity, we have that $\mathbf{H}(X) + \mathbf{H}(X+Y+Z) \leq \mathbf{H}(X+Y) + \mathbf{H}(X+Z)$, and it is clear that $\mathbf{H}(X+Y+Z) \geq \mathbf{H}(Y+Z)$ by Lemma 2.2. □

**Remark 3.3.** In the previous proof, we could have used item (ii) instead. The proof would be equally simple and direct.

One can now use this to prove a slightly stronger version of Theorem 2.11.

**Theorem 3.26.** *Let $X$ and $Y$ be two $G$-random variables, where $G$ is an additive group, and let $\mathbf{H}(X+Y) - \mathbf{H}(X) = \beta$. For any positive integers $k, l$, let $Y_1, \dots, Y_k, Y'_1, \dots, Y'_l$ be independent copies of $Y$. Then,*

$$\mathbf{H}(Y_1 + \ldots + Y_k - Y'_1 - \ldots - Y'_l) \leq \mathbf{H}(X) + (k+l)\beta.$$

*Proof.* Observe that, by item (i) of Corollary 3.23, given any three $G$-random variables we have that

$$\mathbf{H}(X) + \mathbf{H}(X+Y+Z) \leq \mathbf{H}(X+Y) + \mathbf{H}(X+Z). \tag{20}$$

First we are going to prove by induction that

$$\mathbf{H}(X + Y_1 + \ldots + Y_n) \leq \mathbf{H}(X) + n\beta. \tag{21}$$

for any $n \geq 1$. The base case $n = 1$ holds with equality by assumption, so now let us assume that the statement holds for some $n \geq 1$, and let us prove that it also holds for $n+1$. Simply, let $Z = Y_1 + \ldots + Y_n$. Then, by (20),

$$\begin{aligned}\mathbf{H}(X + Y_1 + \ldots + Y_n + Y_{n+1}) &\leq \mathbf{H}(X + Y_1 + \ldots + Y_n) + \mathbf{H}(X + Y_{n+1}) - \mathbf{H}(X) \\ &\leq \mathbf{H}(X) + n\beta + \mathbf{H}(X) + \beta - \mathbf{H}(X) = \mathbf{H}(X) + (n+1)\beta.\end{aligned}$$

Now, as we did in the proof of Theorem 2.11, apply Theorem 2.4, taking $X = -X'$, $X'$ being an independent copy of $X$, $Y = Y_1 + \ldots + Y_k$ and $Z = Y_1' + \ldots + Y_l'$. As $\mathbf{H}(X) = \mathbf{H}(-X)$, this yields

$$\begin{aligned}\mathbf{H}(Y_1 + \ldots + Y_k - Y_1' - \ldots - Y_l') & \\ &\leq \mathbf{H}(X' + Y_1 + \ldots + Y_k) + \mathbf{H}(X' + Y_1' + \ldots + Y_l') - \mathbf{H}(X') \\ &\leq \mathbf{H}(X) + k\log\sigma[X] + \mathbf{H}(X) + l\log\sigma[X] - \mathbf{H}(X) \\ &= \mathbf{H}(X) + (k+l)\log\sigma[X],\end{aligned}$$

where the last inequality comes from applying (21). □

Here, $\beta$ somewhat substitutes the role of the Ruzsa distance. It is the entropy analogue of the value $\alpha$ from the statement of Theorem 2.9. Observe that we have obtained an exact analogue of this theorem in the case $j = 1$. Observe, furthermore, that (21) is the analogue of Theorem 2.8 in the case $j = 1$. And the proof we have shown here is much simpler than that from chapter 2.

From Corollary 3.23, one can also obtain entropy analogues of Plünnecke-Ruzsa-type inequalities, in a very general fashion. These Plünnecke-Ruzsa-type inequalities will be presented later on.

**Theorem 3.27.** *Let $X_0, X_1, \ldots, X_n$ be independent discrete random variables taking values in an abelian group $G$, and let $\alpha$ be a fractional covering using the collection $\mathcal{S}$ of subsets of $[n]$. Let $c = \sum_{S\in\mathcal{S}} \alpha_S$. Then,*

$$c\mathbf{H}(X_0 + X_{[n]}^+) \leq (c-1)\mathbf{H}(X_{[n]}^+) + \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0 + X_S^+).$$

*Proof.* For convenience, write $X_0 = X_{n+1}$. Consider the collection of subsets of $[n+1]$ given by

$$\mathcal{S}' = \{[n]\} \cup \{S' = S \cup \{n+1\} : S \in \mathcal{S}\}.$$

For each set $S \in \mathcal{S}$ let $\gamma_{S\cup\{n+1\}} = \frac{\alpha_S}{c}$, and let $\gamma_{[n]} = 1 - \frac{1}{c}$, which is nonnegative, since $c \geq 1$. With this, $\gamma$ is a fractional covering for $[n+1]$ using $\mathcal{S}'$. Indeed, for the index $n+1$ one has

$$\sum_{S\in\mathcal{S}} \gamma_{S\cup\{n+1\}} = 1,$$

and for each $j \in [n]$, since $\alpha$ is a fractional covering,

$$\gamma_{[n]} + \sum_{S\in\mathcal{S}:j\in S} \gamma_{S\cup\{n+1\}} = 1 - \frac{1}{c} + \sum_{S\in\mathcal{S}:j\in S} \frac{\alpha_S}{c} \geq 1.$$

The statement follows by applying item (iii) of Corollary 3.23 to this fractional covering. □

**Theorem 3.28.** *Let $X_0, X_1, \dots, X_n$ be independent discrete random variables taking values in an abelian group $G$, and let $\alpha$ be a fractional covering using the collection $\mathcal{S}$ of subsets of $[n]$. Let $c = \sum_{S\in\mathcal{S}} \alpha_S$. Then,*

$$\mathbf{H}(X_0 + X^+_{[n]}) \leq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0 + X^+_S) - (c-1)\mathbf{H}(X_0).$$

*Proof.* By Lemma 3.19, we have that $\mathbf{H}(X_1 + X_2) = \mathbf{H}(X_1) + \mathbf{I}(X_1 + X_2; X_2)$, because $X_1$ and $X_2$ are independent and the sum is a strongly partition-determined function. By applying this recursively to independent random variables, we obtain a chain rule,

$$\mathbf{H}(X^+_{[n]}) = \mathbf{H}(X_1) + \sum_{i=2}^{n} \mathbf{I}(X^+_{[i]}; X_i). \tag{22}$$

In our particular case, for each $S \in \mathcal{S}$ we may write

$$\mathbf{H}(X_0 + X^+_S) = \mathbf{H}(X_0) + \sum_{i\in S} \mathbf{I}(X^+_{[i]\cap S}; X_i).$$

Hence,

$$\sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0 + X^+_S) = \sum_{S\in\mathcal{S}} \alpha_S \left[\mathbf{H}(X_0) + \sum_{i\in S}^{n} \mathbf{I}(X^+_{[i]\cap S}; X_i)\right] \geq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0) + \sum_{S\in\mathcal{S}} \alpha_S \sum_{i\in S} \mathbf{I}(X^+_{[i]}; X_i),$$

where the inequality comes from (19) in the particular case when $f$ is the sum. By an interchange of sums, the definition of fractional coverings and, again, (19), we have

$$\begin{aligned}
\sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0 + X^+_S) &\geq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0) + \sum_{i\in[n]} \mathbf{I}(X^+_{[i]}; X_i) \sum_{S\in\mathcal{S}_i : i\in S} \alpha_S \\
&\geq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0) + \sum_{i\in[n]} \mathbf{I}(X^+_{[i]}; X_i) \\
&\geq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0) + \sum_{i\in[n]} \mathbf{I}(X_0 + X^+_{[i]}; X_i).
\end{aligned}$$

Now, using again (22),

$$\sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_0 + X^+_S) \geq \left(\sum_{S\in\mathcal{S}} \alpha_S - 1\right) \mathbf{H}(X_0) + \mathbf{H}(X_0 + X^+_{[n]}).$$

The statement follows by rearranging the terms. □

These two results can be written in a way that looks closer to the usual way of writing Plünnecke-type inequalities. In order to do so, por each $S \in \mathcal{S}$ define a constant $\beta_S$ such that $\mathbf{H}(X_0 + X^+_S) = \mathbf{H}(X_0) + \beta_S$. Then, the inequalities from Theorem 3.27 and Theorem 3.28 can be written as

$$\mathbf{H}(X_0 + X^+_{[n]}) \leq \mathbf{H}(X_0) + \left(1 - \frac{1}{c}\right) \mathbf{H}(X^+_{[n]}) + \frac{1}{c} \sum_{S\in\mathcal{S}} \alpha_S \beta_S$$

and

$$\mathbf{H}(X_0 + X_{[n]}^+) \leq \mathbf{H}(X_0) + \sum_{S \in \mathcal{S}} \alpha_S \beta_S,$$

respectively. It is interesting to remark that neither of these bounds seems to be better than the other. When trying to compare them, one must compare $\mathbf{H}(X_{[n]}^+)$ to $\sum_{S \in \mathcal{S}} \alpha_S \beta_S$, which is not simple at all. One may write

$$\sum_{S \in \mathcal{S}} \alpha_S \beta_S = \sum_{S \in \mathcal{S}} \alpha_S \left[ \mathbf{H}(X_0 + X_S^+) - \mathbf{H}(X_0) \right] \leq \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(X_S^+).$$

This is also an upper bound for $\mathbf{H}(X_{[n]}^+)$ (item (iii) from Corollary 3.23), so we may write a common upper bound as a weaker result,

$$\mathbf{H}(X_0 + X_{[n]}^+) \leq \mathbf{H}(X_0) + \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(X_S^+).$$

In particular, take $\mathcal{S}$ to be the collection of singletons in Theorem 3.28, taking the fractional covering given by $\alpha_i = 1$ for all $i \in [n]$. This results in

$$\mathbf{H}(X_0 + X_{[n]}^+) \leq \mathbf{H}(X_0) + \sum_{i=1}^{n} (\mathbf{H}(X_0 + X_i) - \mathbf{H}(X_0)),$$

which is the analogue of Theorem 2.10, the case with different summands, when $j = 1$. In general, we may consider $\mathcal{S}$ as the multiset of all sets of size $k$, which results in a uniform $\binom{n-1}{k-1}$-cover. Taking fractional partitions given by $\alpha_S = \binom{n-1}{k-1}^{-1}$ for all $S \in \mathcal{S}$ results in

$$\mathbf{H}(X_0 + X_{[n]}^+) \leq \mathbf{H}(X_0) + \sum_{S \subseteq [n] : |S| = k} \frac{(k-1)!(n-k)!}{(n-1)!} (\mathbf{H}(X_0 + X_S^+) - \mathbf{H}(X_0)),$$

which is the general analogue of Theorem 2.10. Therefore, as Theorem 3.28 serves for any fractional covering, it is a generalised version of the entropic analogue of Theorem 2.10.

We now turn our attention towards corollaries related to set cardinalities. Most of the proofs now are an extension of the arguments used by Gyarmati, Matolcsi and Ruzsa to prove Theorem 3.8, defining arbitrary linear orders in the spaces, and considering the lexicographical order for elements in the Cartesian product of the spaces. We will consider the following: we are given $n$ finite discrete spaces, $A_1, \ldots, A_n$, and a function $f : Q(A_1, \ldots, A_n) \to \mathcal{B}$, where $\mathcal{B}$ is any space. Then, given a set $B \in f(A_{[n]})$ that we want to bound in size, we define $r(b)$ to be the smallest element of $f^{-1}(b)$ in lexicographical order, for each $b \in B$. Then, set $R = \{r(b) : b \in B\}$. In this way, each $b \in B$ has a unique representative preimage $r(b)$, and $|R| = |B|$.

We first present a simple lemma, which will be the key for most of the results we are going to present.

**Lemma 3.29.** *Let $X$ be a discrete random variable uniformly distributed over $R$. If $f$ is a partition-determined function with respect to $S \subseteq [n]$, then*

$$\mathbf{H}(X_S|f(X_S)) = 0.$$

*Proof.* $X_S$ takes values in $\pi_S(R)$, so it is enough to show that the restriction of $f$ to this domain is a one-to-one function (this would mean that $X_S$ can be retrieved from $f(X_S)$, so we already know all the information about it and there is no additional uncertainty). Assume that there are two elements $\alpha \neq \alpha'$ in $\pi_S(R)$ such that $f(\alpha) = f(\alpha')$ and $\Pr(X_S = \alpha) \neq 0$, $\Pr(X_S = \alpha') \neq 0$. Consider their preimages with respect to the projection $\pi_S$, that is, the elements $a, a' \in R$ such that $\pi_S(a) = \alpha$ and $\pi_S(a') = \alpha'$. It is clear that $a \neq a'$, as $\alpha \neq \alpha'$. Without loss of generality, assume that $a <_{\text{lex}} a'$, and define $a'' \in A_{[n]}$ by

$$a''_i = \begin{cases} a_i & \text{for } i \in S, \\ a'_i & \text{for } i \notin S. \end{cases}$$

Since $f(\alpha) = f(\alpha')$ and $f$ is partition determined with respect to $S$, we have that $f_S(a'') = f(\alpha) = f(\alpha') = f_S(a')$ (as $a_i = \alpha_i$ when $i \in S$) and $f_{\overline{S}}(a'') = f_{\overline{S}}(a')$, so $f(a'') = f(a')$. As $a'$ was chosen to be the representative of $f(a')$, by construction we must have $a' <_{\text{lex}} a''$. However, as $a <_{\text{lex}} a'$, it is clear that $a'' <_{\text{lex}} a'$, so we reach a contradiction. Hence, there cannot be two such elements, and the restriction of $f$ to $\pi_S(R)$ is a one-to-one function. □

We can now start proving several results.

**Theorem 3.30.** *Suppose that $f$ is a partition-determined function with respect to a collection $\mathcal{S}$ of subsets of $[n]$, and let $\alpha$ be a fractional covering of $[n]$ using $\mathcal{S}$. Then, for any set $B \subseteq f(A_{[n]})$ we have that*

$$|B| \leq \prod_{S \in \mathcal{S}} \left| f_S\left(f^{-1}_{[n]}(B)\right)\right|^{\alpha_S}.$$

*Proof.* For any set $B$ as in the statement, define its respective set of representatives $R$. Let $X$ be a random variable uniformly distributed over $R$, and define $X_i = \pi_i(X)$ for all $i \in [n]$. Then, we have

$$\log|R| = \mathbf{H}(X) \leq \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(X_S),$$

where the inequality comes from Theorem 3.22 in the case when $f$ is the identity. On the other hand, by the chain rule for entropy we have

$$\mathbf{H}(X_S|f(X_S)) + \mathbf{H}(f(X_S)) = \mathbf{H}(X_S, f(X_S)) = \mathbf{H}(f(X_S)|X_S) + \mathbf{H}(X_S)$$

for each $S \in \mathcal{S}$. We know that $\mathbf{H}(f(X_S)|X_S) = 0$ since $f(X_S)$ is completely determined by $X_S$, and, furthermore, $\mathbf{H}(X_S|f(X_S)) = 0$ because of Lemma 3.29, so the previous becomes

$$\mathbf{H}(f(X_S)) = \mathbf{H}(X_S).$$

With this,

$$\begin{aligned}\log|B| = \log|R| &\leq \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(X_S) = \sum_{S\in\mathcal{S}} \alpha_S \mathbf{H}(f(X_S)) \\ &\leq \sum_{S\in\mathcal{S}} \alpha_S \log|f_S(R)| \leq \sum_{S\in\mathcal{S}} \alpha_S \log|f_S(f_{[n]}^{-1}(B))|.\end{aligned}$$

The result follows by exponentiating. □

In particular, by considering the whole set we have the following corollary.

**Corollary 3.31.** *Suppose that $f$ is a partition-determined function with respect to $\mathcal{S}$, and that $\alpha$ is a fractional covering of $[n]$ using $\mathcal{S}$. Then,*

$$|f(A_{[n]})| \leq \prod_{S\in\mathcal{S}} |f(A_S)|^{\alpha_S}.$$

Using Theorem 3.30 we can now obtain results in many settings. Consider, for instance, that of projections, with which we have been dealing all along this chapter. In this particular case, we obtain the following.

**Corollary 3.32.** *Let $B_1, \ldots, B_n$ be arbitrary finite sets, and take some set $A \subseteq B_{[n]}$. Then, for any fractional covering $\alpha$ of $[n]$ using a collection $\mathcal{S}$ of subsets of $[n]$,*

$$|A| \leq \prod_{S\in\mathcal{S}} |\pi_S(A)|^{\alpha_S}.$$

*Proof.* Apply Theorem 3.30 taking $f_S = \pi_S$, which is a partition-determined function for every collection $\mathcal{S}$. □

Note that Theorem 3.1 and Theorem 3.10 are simple particular cases of this result, when $\mathcal{S}$ is a uniform $k$-cover of $[n]$. Thus, we have obtained a much more general result than we had so far.

When dealing with sumsets in abelian groups, we also obtain some very general results. From Corollary 3.31 we obtain the following.

**Corollary 3.33.** *Let $A_1, \ldots, A_n$ be finite sets in an abelian group $(G, +)$, and let $\alpha$ be a fractional covering of $[n]$ using a collection $\mathcal{S}$ of subsets of $[n]$. Then,*

$$|A^+_{[n]}| \leq \prod_{S\in\mathcal{S}} |A^+_S|^{\alpha_S}.$$

*Proof.* Substitute $f$ by the sum in Corollary 3.31; the sum is a partition determined function, as we showed in Lemma 3.18. □

Again, Theorem 3.8 and Theorem 3.12 are particular cases of this result, which is the sumset analogue of Corollary 3.32. In particular, since this result is more general than Theorem 3.12, one can also think of it as a more general, although slightly weaker, version of Theorem 2.10. A more general variant of the previous result can be obtained when considering subsets of the sumset.

**Theorem 3.34.** *Let $A, B_1, \dots, B_n$ be finite sets in an abelian group $(G, +)$. Let $\alpha$ be a fractional covering of $[n]$ using the collection $\mathcal{S}$ of subsets of $[n]$, and let $c = \sum_{S \in \mathcal{S}} \alpha_S$. Then, for any set $D \subseteq B^+_{[n]}$,*

$$|A + D|^c \leq |D|^{c-1} \prod_{S \in \mathcal{S}} |A + B^+_S|^{\alpha_S}.$$

*Proof.* The proof is analogous to that of Theorem 3.27. Call $B_{n+1} = A$, and set $B_{[k+1]}$ to be the Cartesian product of all the sets $B_i$, as usual. As we saw in Lemma 3.18, the function $f_S(b) = \sum_{i \in S} b_i$ is partition-determined with respect to any collection $\mathcal{S}'$ of subsets of $[n+1]$. As $D \subseteq B^+_{[n]}$, let us define $C = f^{-1}_{[n]}(D)$, so $C \subseteq B_{[n]}$, and let us write

$$E = D + A = \{f(b_1, \dots, b_n, a) : (b_1, \dots, b_n) \in C, a \in A\}.$$

Now, choose the collection of subsets $\mathcal{S}'$ defined by

$$\mathcal{S}' = \{[n]\} \cup \{S \cup \{n+1\} : S \in \mathcal{S}\},$$

and take the fractional covering $\gamma$ for $[n+1]$ given by $\gamma_{S \cup \{n+1\}} = \frac{\alpha_S}{c}$ for each $S \in \mathcal{S}$, and $\gamma_{[n]} = 1 - \frac{1}{c}$. Applying Theorem 3.30,

$$|E| \leq \prod_{S' \in \mathcal{S}'} \left| f_{S'} \left( f^{-1}_{[n+1]}(E) \right) \right|^{\gamma_{S'}}.$$

As $f^{-1}_{[n+1]}(E) = C \times A$, it is clear that $f_{[n]} \left( f^{-1}_{[n+1]}(E) \right) = D$ and $f_{S \cup \{n+1\}} \left( f^{-1}_{[n+1]}(E) \right) \subseteq A + B^+_S$ for any $S' \neq [n]$, so

$$|E| \leq |D|^{\gamma_{[n]}} \prod_{S \in \mathcal{S}} |A + B^+_S|^{\gamma_{S \cup \{n+1\}}}.$$

Finally, the result follows by substituting the values of each $\gamma_{S'}$ and taking the $c$-th power. □

From this, one can obtain many different corollaries by considering different coverings. For example, given a uniform $k$-covering with a collection $\mathcal{S}$ of subsets of $[n]$, and taking the fractional covering given by $\alpha_S = \frac{1}{k}$ for all $S \in \mathcal{S}$, we have that

$$|A + D|^{|\mathcal{S}|} \leq |D|^{|\mathcal{S}| - k} \prod_{S \in \mathcal{S}} |A + B^+_S|.$$

In the particular case where $\mathcal{S}$ is the uniform 1-cover given by the collection of all singletons, the inequality becomes

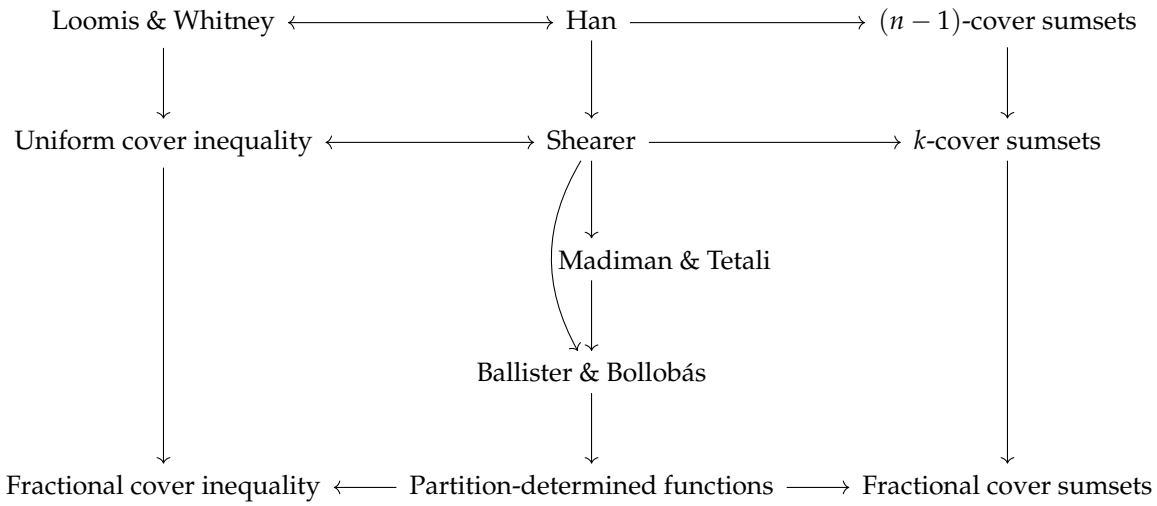
$$|A + D|^n \leq |D|^{n-1} \prod_{i=1}^{n} |A + B_i|.$$

so we are generalizing Theorem 3.16 of Balister and Bollobás for all possible fractional coverings.

Theorem 3.34 would be the sumset analogue of Theorem 3.27. However, we do not have an analogue for Theorem 3.28 so far, and its statement remains as an open conjecture.

Using Theorem 3.30, Madiman, Marcus and Tetali obtained corollaries for several other compound sets using partition-determined functions. In particular, they proved an extension to the non-commutative setting of some of the previous results, as well as some bounds for polynomial compound sets. In this thesis, we are mainly interested in sumsets in abelian groups, so should the reader be interested, we recommend to check [19]

The following diagram illustrates the relationships between the main statements presented in the chapter.



# Chapter 4
# The Ruzsa device

In this chapter we revisit the way in which Ruzsa contructed sets that capture the distribution of a random variable $X$ [25]. We first presented this at the beginning of Chapter 3. We shall show that his method can in fact be generalised and used in a much wider context, in such a way that it allows us to obtain many of the results presented in he previous chapters in a very general and unified way. In fact, what we shall see is that we can prove sumset inequalities starting from entropy inequalities, or viceversa. In this sense, we can trivially obtain sumset inequalities whose entropy counterpart is known and proved, and entropy inequalities whose sumset counterpart is known too. In order to illustrate the power and flexibility of the method, we introduce a new application related to expanders.

## 4.1. The Ruzsa device

Remember that, given a random variable $X$ defined over a set $\mathcal{X}$ that takes a finite number of values, each of them with rational probability, we construct a set $A \subseteq \mathcal{X}^k$ for a suitable $k$. The way in which we construct this set is by considering vectors of length $k$ whose coordinates are such that, if one of these coordinates is chosen uniformly at random, then we are choosing an element in $\mathcal{X}$ with the same probability as the random variable $X$ does. To be precise, assume $X$ takes values $\{x_1, \ldots, x_n\}$, each with probability $p_i = \frac{q_i}{r_i}$ for some $q_i, r_i \in \mathbb{N}$ (in such a way that $\sum_{i=1}^{n} p_i = 1$). Then, take $k = \text{lcm}(r_i : i \in [n])$, and consider all vectors of length $k$ with entries in $\mathcal{X}$ such that each element $x_i$ appears exactly $p_i k$ times. $A$ is the set of all these vectors.

Obviously, once this set has been constructed for a certain value of $k$ (for which $\text{lcm}(r_i : i \in [n])$ is the minimum possible value), then a set $A$ can be constructed for any multiple of $k$; simply put, each of the elements $x_i$ will appear as many times as the same multiple of the original number of

times it appeared before. Remember that, with such a construction, one can prove (18),

$$\log |A| = k\,\mathbf{H}(X) + O(\log k).$$

Ruzsa used this construction to prove the equivalence between Han's inequality and the Loomis and Whitney theorem. This same approach was later used by Ballister and Bollobás to prove the equivalence between Shearer's inequality and the uniform cover inequality in [1]. The main idea they used came from observing that one can build a set from a random variable and a different set from its projection onto a certain subspace, and that the resulting set in the second case is nothing more than the projection of the first one. That is, the following diagram is commutative.

$$\begin{array}{ccc} X & \xrightarrow{\text{make set}} & A \\ \downarrow{\scriptstyle \pi_i} & & \downarrow{\scriptstyle \pi_i^k} \\ \pi_i(X) & \xrightarrow{\text{make set}} & A_i \end{array}$$

Using this fact, one can separately compute the sizes of $A$ and its projections in terms of the entropy of the random variables, through (18). If a relationship between the sizes of the set and its projections is known, a relationship between the entropies of the variable and its projections follows (when letting $k$ tend to infinity). The converse is found by considering the uniform distribution on one of the sets, and using the most standard entropy inequalities (namely, Lemma 1.2).

Ruzsa took the idea behind these commutative diagrams a bit further. Instead of considering simple projections, he took linear functions defined over two variables, and again proved that making a set and applying linear functions commute. He used this to prove his equivalence theorem, Theorem 3.4. In this section, we try to go even further, and see that the diagram must always be commutative, no matter which function $f$ we consider.

Let $X$ be a random variable (any random variable, it may be the joint distribution of several others) that takes finitely many values in a set $\mathcal{X}$, each of them with rational probability. Consider the set $A$ that is built when capturing the information given by $X$ in the way that was described above. Let $f$ be any function defined over $\mathcal{X}$, $f : \mathcal{X} \to f(\mathcal{X})$. This function takes an element in $\mathcal{X}$ and transforms it in some way; hence, the same thing can be done for each of the entries of the elements of $A$. Let us denote $f^k$ to the function $f^k : \mathcal{X}^k \to f(\mathcal{X})^k$ such that $f^k(x_1', \ldots, x_k') = (f(x_1'), \ldots, f(x_k'))$. The diagram now looks like this.

$$\begin{array}{ccc} X & \xrightarrow{\text{make set}} & A \\ \Big\downarrow f & & \Big\downarrow f^k \\ f(X) & \xrightarrow{\text{make set}} & f^k(A) \end{array}$$

**Proposition 4.1.** *Let $X$ be a random variable taking values in a finite set $\mathcal{X}$ and let $f$ be a function defined over $\mathcal{X}$. Assume that the probability function of $X$ takes only rational values. Then, the following diagram*

$$\begin{array}{ccc} X & \xrightarrow{\textit{make set}} & A \\ \Big\downarrow f & & \Big\downarrow f^k \\ f(X) & \xrightarrow{\textit{make set}} & f^k(A) \end{array}$$

*is commutative.*

*Proof.* Assume that $X$ takes values $\{x_1, \ldots, x_n\}$, each with probability $p_i = \dfrac{q_i}{r_i}$ for some $q_i, r_i \in \mathbb{N}$, and construct the set $A$ as described above. Now consider $f(X)$. It is clear that the number of values it may take is bounded by $n$. Assume that it takes $m$ values, $f(X) \in \{y_1, \ldots, y_m\}$. The probability that each value $y_j$ is taken by $f(X)$ is the sum over all the preimages of $y_j$ of their probabilities; that is, if $f^{-1}(y_j) = \{x_{i_1}, \ldots, x_{i_l}\}$, then $\Pr(f(X) = y_j) = \sum_{h=1}^{l} p_{i_h}$. Each of these probabilities can be computed once the original distribution of $X$ is known, so the distribution of $f(X)$ is also known. So assume that $f(X)$ takes each value $y_j$ with probability $p'_j = \dfrac{q'_j}{r'_j}$ for some $q'_j, r'_j \in \mathbb{N}$ (it is clear that $p'_j \in \mathbb{Q}$ since it is the sum of finitely many rational values). In this case, one may have that $\mathrm{lcm}(\mathrm{r}'_\mathrm{j} : \mathrm{j} \in [\mathrm{m}]) < k$, but it will always be one of its divisors. This means, in particular, that we can construct a set $B \subseteq f(\mathcal{X})^k$, since $k$ is a good value, by making each of the $y_j$ appear exactly $p'_j k$ times in each vector. $B$ will be the set of all such vectors. We would like to see that $f^k(A) = B$.

Clearly, the image by $f^k$ of a vector $\mathbf{x} \in A$ is a vector in which every $y \in f(\mathcal{X})$ appears precisely $k \sum_{x \in f^{-1}(y)} \Pr(X = x)$ times, and thus $f^k(A) \subseteq B$. Reciprocally, let $\mathbf{y}$ be a vector in $B$. Each $y \in f(\mathcal{X})$ appears $k \sum_{x \in f^{-1}(y)} \Pr(X = x)$ times in $\mathbf{y}$. For each $y \in f(\mathcal{X})$ let $J_y \subseteq [k]$ be the set of coordinates of $\mathbf{y}$ which are equal to $y$, $\mathbf{y}_j = y$ for each $j \in J_y$, so that $|J_y| = k \sum_{x \in f^{-1}(y)} \Pr(X = x)$. There is a vector $\mathbf{x} \in A$ which has the elements in $f^{-1}(y)$ placed at the coordinates of $J_y$ for each $y \in f(\mathcal{X})$. For this vector we have $f^k(\mathbf{x}) = \mathbf{y}$. This shows that $B \subseteq f^k(A)$. □

Once we have that the diagram is commutative, we can prove the following result.

**Theorem 4.2.** *Let $f, f_1, \dots, f_n$ be any functions defined over a set $\mathcal{X}$. Let $\alpha_1, \dots, \alpha_n$ be positive real numbers. Then, the following are equivalent:*

*(i) For any finite set $A \subseteq \mathcal{X}$ we have that*

$$|f(A)| \leq \prod_{i=1}^{n} |f_i(A)|^{\alpha_i}.$$

*(ii) For any finite set $A \subseteq \mathcal{X}$, for every random variable $X$ with support in $A$, the entropy of $f(X)$ satisfies*

$$\mathbf{H}(f(X)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(f_i(X)).$$

*Proof.* Showing that (ii) implies (i) is easy. One needs to define an adecuate random variable $X$. To do so, consider $f(A)$, and for each $b \in f(A)$ consider a unique representative of its preimage, $a^* \in f^{-1}(b)$. Let the set of these representatives be $A^*$, so that $f(A^*) = f(A)$. Define a random variable $X$ has having probability $\frac{1}{|f(A)|}$ of taking each value in $A^*$, and zero probability otherwise. In such a way, $f(X)$ is uniformly distributed over $f(A)$, so, by the properties of entropy,

$$\log |f(A)| = \mathbf{H}(f(X)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(f_i(X)) \leq \sum_{i=1}^{n} \alpha_i \log |f_i(A)|, \tag{23}$$

as it is clear that $f_i(A^*) \subseteq f_i(A)$ for all $i$.

To prove the converse, first assume that $X$ takes finitely many values, each of them with rational probability, and build sets $B \subseteq \mathcal{X}^k$ and $f(B), f_i(B) \subseteq f(\mathcal{X})^k$ for each $i \in [n]$ as in Proposition 4.1. For each of these, by (18), we know their asymptotic size. Using (i), we have that

$$\begin{aligned} k\,\mathbf{H}(f(X)) + O(\log k) &\leq \sum_{i=1}^{n} \alpha_i k\,\mathbf{H}(f_i(X)) + O(\alpha_i \log k) \\ \implies \mathbf{H}(f(X)) &\leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(f_i(X)) + O\left(\frac{\log k}{k}\right), \end{aligned}$$

and the result follows by letting $k$ tend to infinity. The passage to the general case of variables with real probabilities follows by a routine limiting argument. □

The reason that the numbers $\alpha_1, \dots, \alpha_n$ have to be positive is that the inequality in (23) is not guaranteed to hold otherwise. However, the second part of the proof also works when these values are negative. This allows us to write the following.

**Theorem 4.3.** *Let $f, f_1, \dots, f_n$ be any functions defined over a set $\mathcal{X}$, and let $\alpha_1, \dots, \alpha_n$ be any real numbers. If for any finite set $A \subseteq \mathcal{X}$ we have that*

$$|f(A)| \leq \prod_{i=1}^{n} |f_i(A)|^{\alpha_i},$$

*then*

$$\mathbf{H}(f(X)) \leq \sum_{i=1}^{n} \alpha_i \mathbf{H}(f_i(X))$$

*holds for every random variable $X$ with support in $A$.*

The fact that we have negative coefficients is what prevented us from proving the sumset versions of Theorem 2.7 and Theorem 2.11 directly from their entropy counterparts. We observe that this same problems extends to the general use of Ruzsa's device.

Now, Theorem 4.2 and Theorem 4.3 can be used as a black box to prove several of the entropy theorems we showed in Chapter 2. As they are entropy analogues of sumset inequalities, having their counterpart proved in the sumset theory is enough to ensure that they also hold in the entropy setting. For instance, the trivial inequalities given by Lemma 2.2 can be obtained by considering the projections and the sum as functions. Similarly, both Ruzsa's triangle inequality and its twin inequality can be obtained.

**Theorem 4.4.** *Let $X$, $Y$ and $Z$ be three independent $G$-random variables, where $G$ is any (not necessarily commutative) group. Then,*

$$d_R(Y, Z) \leq d_R(Y, X) + d_R(X, Z).$$

*Proof.* We must prove that

$$\mathbf{H}(X) + \mathbf{H}(Y - Z) \leq \mathbf{H}(Y - X) + \mathbf{H}(X - Z).$$

Let $A = \text{range}(X)$, $B = \text{range}(Y)$, and $C = \text{range}(Z)$, and let $D = A \times B \times C$. Define several functions over $D$ as follows: $f(x) = (\pi_2(x) - \pi_3(x))$, $f_1(x) = (\pi_2(x) - \pi_1(x))$, $f_2(x) = (\pi_1(x) - \pi_3(x))$, and $f_3(x) = \pi_1(x)$, where $\pi_i$ refers to the projection to the $i$-th coordinate as usual. By Theorem 2.3, $|f(D)| \leq |f_1(D)||f_2(D)||f_3(D)|^{-1}$, so the result follows by Theorem 4.3. □

Notice, in particular, that we have do not remove any assumption about the independence of the random variables, so this result is the same as Theorem 2.5. The twin to the triangle inequality is proved in the same way. We can also prove the analogue of Theorem 2.6 using this approach (notice that, in this case, it is actually enough to use Theorem 3.5, as all the sets we are considering are linear functions of two sets).

**Theorem 4.5.** *Let $X$ and $Y$ be two independent $G$-random variables, where $G$ is an additive group. Then,*

$$d_R(X,-Y) \leq 3\, d_R(X,Y).$$

*Proof.* We want to see that

$$\mathbf{H}(X-Y) \leq 3\mathbf{H}(X+Y) - \mathbf{H}(X) - \mathbf{H}(Y).$$

Let $A = \mathrm{range}(X)$ and $B = \mathrm{range}(Y)$, and let $C = A \times B$. For each $x \in C$ define the functions $f(x) = \pi_1(x) - \pi_2(x)$, $f_1(x) = \pi_1(x) + \pi_2(x)$, $f_2(x) = \pi_1(x)$ and $f_3(x) = \pi_2(x)$. By Theorem 2.6, $|f(C)| \leq |f_1(C)|^3 |f_2(C)|^{-1} |f_3(C)|^{-1}$, so Theorem 4.3 directly yields the result. □

Even more interesting, we can very easily prove and strengthen the entropic Plünnecke-Ruzsa-type inequalities. We may start with the Plünnecke-Ruzsa inequalities. We can prove the following generalization, in the entropic setting, to the general case (for different values of $j$) of the sumset inequalities.

**Theorem 4.6.** *Let $X$ and $Y$ be $G$-random variables ranging over finite sets in a commutative group, and let $j$ be a positive integer. Let $Y_1, \ldots, Y_k, Y'_1, \ldots, Y'_l$ be independent copies of $Y$ for some positive integers $k$ and $l$ such that $j \leq \min\{k,l\}$. Then,*

$$\mathbf{H}(Y_1 + \ldots + Y_k - Y'_1 - \ldots - Y'_l) \leq \mathbf{H}(X) + \frac{k+l}{j} \left( \mathbf{H}(X + Y_1 + \ldots + Y_j) - \mathbf{H}(X) \right).$$

*Proof.* Let $A = \mathrm{range}(X)$ and $B = \mathrm{range}(Y)$. Define $C = A \times B \times B \times \overset{k+l)}{\ldots} \times B$. For each $x \in C$ define the functions $f(x) = \pi_2(x) + \ldots + \pi_{k+1}(x) - \pi_{k+2}(x) - \ldots - \pi_{k+l+1}(x)$, $f_1(x) = \pi_1(x) + \pi_2(x) + \ldots + \pi_{j+1}(x)$ and $f_2(x) = \pi_1(x)$. By Theorem 2.9, $|f(C)| \leq |f_1(C)|^{\frac{k+l}{j}} |f_2(C)|^{1-\frac{k+l}{j}}$. Theorem 4.3 yields the desired result. □

Even more, we can easily obtain an analogue of Theorem 2.10.

**Theorem 4.7.** *Let $j$ and $h$ be two positive integers such that $j < h$. Let $X, Y_1, \ldots, Y_h$ be $G$-random variables, where $G$ is any commutative group. For any $I \subseteq [h]$, let $Y_I^+ = \sum_{i \in I} Y_i$. Then,*

$$\mathbf{H}(Y_{[h]}^+) \leq \mathbf{H}(X) + \frac{(j-1)!(h-j)!}{(h-1)!} \sum_{J \subseteq [h] : |J| = j} \Big( \mathbf{H}(X + Y_J^+) - \mathbf{H}(X) \Big).$$

*Proof.* Let $A = \mathrm{range}(X), B_1 = \mathrm{range}(Y_1), \ldots, B_h = \mathrm{range}(Y_h)$, and define $C = A \times B \times B \times \overset{h)}{\ldots} \times B$. Consider the functions $f(X) = \pi_2(x) + \ldots + \pi_{h+1}(x)$, $f_1(x) = \pi_1(x)$ and, for each $J \subseteq [h]$

such that $|J| = j$, $f_J(x) = \pi_1(x) + \sum_{i \in J} \pi_i(x)$. By Theorem 2.10, we have that

$$|f(C)| \le |A' + B^+_{[h]}| \le \left( \prod_{J \subseteq [h] : |J| = j} \frac{|A + B^+_J|}{|A|} \right)^{\frac{(j-1)!(h-j)!}{(h-1)!}} |A'| \le |f_1(C)| \prod_{J \subseteq [h] : |J| = j} \left( \frac{|f_J(C)|}{|f_1(C)|} \right)^{\frac{(j-1)!(h-j)!}{(h-1)!}}.$$

The claim follows by applying Theorem 4.3. □

Hence we have a very simple proof of a corollary of Theorem 3.28.

These are just a few examples of the many entropic results that can be achieved in this way. Similarly, we can obtain sumset results starting from entropy inequalities, provided there are no negative coefficients in the entropy inequalities, by using Theorem 4.2. For instance, from Theorem 3.22 we obtain the following.

**Theorem 4.8.** *Let $A_1, \dots, A_n$ be finite sets, and let $f : Q(A_1, \dots, A_n) \to B$ be a strongly partition-determined function. Then, for any fractional covering $\alpha$ using any collection $\mathcal{S}$ of subsets of $[n]$,*

$$|f_{[n]}(A_1, \dots, A_n)| \le \prod_{S \in \mathcal{S}} |f_S(A_1, \dots, A_n)|^{\alpha_S}.$$

*Proof.* Consider independent random variables $X_1, \dots, X_n$ ranging over $A_1, \dots, A_n$, respectively. By Theorem 3.22, we have that

$$\mathbf{H}(f_{[n]}) \le \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(f_S).$$

Now, since all the values $\alpha_S$ are nonnegative, we may apply apply Theorem 4.2. The statement follows swiftly. □

In the particular case of sums, we have the following.

**Corollary 4.9.** *Let $A_1, \dots, A_n$ be finite sets, and let $f : Q(A_1, \dots, A_n) \to B$ be a strongly partition-determined function. Then, for any fractional covering $\alpha$ using any collection $\mathcal{S}$ of subsets of $[n]$,*

$$|A^+_{[n]}| \le \prod_{S \in \mathcal{S}} |A^+_S|^{\alpha_S}.$$

*Proof.* The statement follows from Theorem 4.8, since the sum is a strongly partition-determined function. □

Similarly, from Theorem 3.27 we get the following.

**Theorem 4.10.** *Let $A_0, A_1, \ldots, A_n$ be finite sets in an abelian group $G$, and let $\alpha$ be a fractional covering using the collection $\mathcal{S}$ of subsets of $[n]$. Let $c = \sum_{S \in \mathcal{S}} \alpha_S$. Then,*

$$|A_0 + A_{[n]}^+|^c \leq |A_{[n]}^+|^{c-1} \prod_{S \in \mathcal{S}} |A_0 + A_S^+|^{\alpha_S}.$$

*Proof.* Consider any independent random variables $X_0, X_1, \ldots, X_n$ ranging over $A_0, A_1, \ldots, A_n$, respectively. By Theorem 3.27 we have

$$c\mathbf{H}(X_0 + X_{[n]}^+) \leq (c-1)\mathbf{H}(X_{[n]}^+) + \sum_{S \in \mathcal{S}} \alpha_S \mathbf{H}(X_0 + X_S^+).$$

Since all the coefficients are nonnegative, the result follows by applying Theorem 4.2. □

Notice that we cannot obtain a sumset analogue of Theorem 3.28 in a similar way as we did for Theorem 3.27 because there are negative coefficients involved. This is related to some observations about the difference in nature of Theorem 3.27 and Theorem 3.28 pointed out in [19]. We believe the existence of negative coefficients plays an important role in this sense.

## 4.2. Expanding functions

There is a very important open conjecture, by Erdős and Szemerédi, that states that if $A$ is a finite set of integers, then

$$\max\{|A + A|, |A \cdot A|\} \geq c|A|^{2-\delta}$$

for any positive $\delta$ and for some positive constant $c$. A lot of effort has been directed at solving this open problem, but the best bounds known today are still far from the conjectured $2 - \delta$. The wide interest directed at this problem derived in many other similar problems, such as the statement of similar ones in sets other than the integers. For instance, Bourgain, Katz and Tao [5] extended the problem to finite fields $\mathbb{F}_p$, for prime $p$, and were able to prove a non-trivial lower bound, namely, that given a set $A \subseteq \mathbb{F}_p$ such that $p^\alpha < |A| < p^{1-\alpha}$, then $\max\{|A + A|, |A \cdot A|\} \geq c(\alpha)|A|^{1+\varepsilon}$, for some $\varepsilon = \varepsilon(\alpha) > 0$.

In a different sense, this problem arose the question of whether other functions have a similar effect on the size of their compound sets, and some interest was directed to polynomials. The natural question is as follows: given a polynomial in two (or more) variables $f$, does its domain blow when applied on any sets? That is, is it true that, for any sets $A, B \subseteq \mathbb{F}_p$ of comparable size, $f(A, B)$ is ampler than the size of $A$?

We know that the answer is negative for $f(x,y) = x + y$ because we have $|f(A,A)| = 2|A| - 1$ when $A$ is an arithmetic progression, and we are looking for functions with $|f(A,A)| \geq |A|^{1+\varepsilon}$. To be more precise, the following definition of expander polynomials is given in [4].

**Definition 4.1.** For any prime $p$, let $f_p : \mathbb{F}_p^k \to \mathbb{F}_p$ be an arbitrary function in $k$ variables in $\mathbb{F}_p$. We say that the family of maps $\mathcal{F} = \{f_p : p \text{ is a prime}\}$ is an *expander* (in $k$ variables) if, for any $\alpha$, $0 < \alpha < 1$, and for any $L_1, L_2 \in \mathbb{R}_+$ there exist $\varepsilon = \varepsilon(\alpha) > 0$ and a positive constant $c = c(\mathcal{F}, L_1, L_2)$ not depending on $\alpha$ such that for any prime $p$ and any $k$-tuple $(A_1, \ldots, A_k)$ of subsets of $\mathbb{F}_p$ satisfying $L_1 p^\alpha \leq |A_i| \leq L_2 p^\alpha$, one has that

$$|f(A_1, \ldots, A_k)| \geq c p^{\alpha+\varepsilon}.$$

Informally, we may understand this definition as saying that a function $f(x_1, \ldots, x_k)$ is an expander if $\lim_{p\to\infty} \frac{|f(A_1,\ldots,A_k)|}{|A|} = \infty$ whenever $A_1, \ldots, A_k$ are of comparable sizes (meaning that they differ by a constant factor, as in the definition above). Very little is known about explicit constructions of expanders. In 2005, Bourgain [4] showed that the bivariant function $f(x,y) = x^2 + xy$ is an expander. In fact, what he proved was that, given any two sets $A$ and $B$ of comparable sizes such that $p^\alpha < |A| < p^{1-\alpha}$, $\frac{|f(A,B)|}{|A|} > p^\gamma$ for some positive but inexplicit $\gamma$. More recently, Hegyvári and Hennecart [14] proved that a more general class of functions is a class of expanders.

**Theorem 4.11 (Hegyvári, Hennecart).** *Let $k \geq 1$ be a positive integer, and let $f$ and $g$ be two polynomials with integer coefficients. Now, consider the function $F : \mathbb{Z}^2 \to \mathbb{Z}$ given by $F(x,y) = f(x) + x^k g(y)$. Assume, furthermore, that $f(x)$ and $x^k$ are affinely independent. Then, $F$ induces a family of expanders.*

The tools used by these authors come from incidence geometry. Other authors have used tools from spectral graph theory (see [27], by Solymosi) to obtain similar results about expanders. In any case, the examples of expanders are scarse, and building any new ones would be an interesting topic.

In this sense, we now consider Theorem 4.2. On the one hand, we can use it to prove entropy analogues of expanders. For instance, we can use the expander defined by Bourgain, $f(x,y) = x^2 + xy$. On the other hand, if we have a bound in the entropy setting, we may obtain bounds for cardinalities. And we may use this to try and find new expanders.

In the case of Bourgain's expander, remember that we have

$$\frac{|f(A,B)|}{|A|} > p^\gamma$$

for some positive $\gamma$, for any sets $A$ and $B$ of comparable size and such that $p^\alpha < |A| < p^{1-\alpha}$. Using basic properties of logarithms, we may write that

$$|f(A,B)| > p^\gamma |A| = |A|^{\gamma \log_{|A|} p} |A| = |A|^{1+\gamma \frac{\log p}{\log |A|}}.$$

A direct application of Theorem 4.2 tells us that, for any random variable $X$ defined over $\mathbb{F}_p^2$ with support in $A \times B$,

$$\mathbf{H}(f(X)) > \left(1 + \gamma \frac{\log p}{\log |A|}\right) \mathbf{H}(\pi_1(X)).$$

This means that

$$\mathbf{H}(f(X)) - \mathbf{H}(\pi_1(X)) > \gamma \frac{\log p}{\log |A|} \mathbf{H}(\pi_1(X)),$$

and this difference would tend to infinity as the entropy of the projection of the random variable does.

This motivates the following definition.

**Definition 4.2.** For any prime $p$, let $f_p : \mathbb{F}_p^k \to \mathbb{F}_p$ be an arbitrary function in $k$ variables in $\mathbb{F}_p$. We say that the family of maps $\mathcal{F} = \{f_p : p \text{ is a prime}\}$ is an *entropic expander* (in $k$ variables) if, for any $\alpha$, $0 < \alpha < 1$, and for any $L_1, L_2 \in \mathbb{R}_+$ there exist $\varepsilon = \varepsilon(\alpha) > 0$ and a positive constant $c = c(\mathcal{F}, L_1, L_2)$ not depending on $\alpha$ such that for any prime $p$ and any random variable defined over a $k$-tuple $(A_1, \ldots, A_k)$ of subsets of $\mathbb{F}_p$ satisfying $L_1 + \alpha \log p \leq \mathbf{H}(X_i) \leq L_2 + \alpha \log p$, one has that

$$\mathbf{H}(f(X)) \geq c + (\alpha + \varepsilon) \log p.$$

Now, if we were able to find a function which is an entropic expander we would automatically recover an expander, by using Theorem 4.2. The author has no knowledge of such an entropic inequality, but if it were found, it would lead to the contruction of new families of expanders.

# Conclusions

Throughout this work, we have been able to observe a great deal of the interaction between entropy theory and additive combinatorics. The importance of the entropic method in additive combinatorics has been made clear through its many varied and useful applications, especially in the way in which it simplifies many proofs or allows to extend previous ideas to some new results. Conversely, we have seen how many ideas from additive combinatorics can be applied in the entropic setting, and how this has allowed for a lot of information theoretic inequalities to be proven. Furthermore, we have seen that not only basic inequalities can be obtained in such a way: the analogue of the Balog-Szemerédi-Gowers theorem is a much more involved result, and many other important results might be obtained in similar ways.

The different techniques that have been showed are part of a still developing new theory. The functional submodularity method introduced by Tao is a very versatile one, and can be used in many settings, as long as one can find the right functions. The method presented by Madiman, Marcus and Tetali using partition-determind functions, with mutual information being a key ingredient, is equally powerful in many cases, and allows for the statement of theorems in settings other than that of sumset inequalities. Similarly, the Ruzsa device becomes a powerful and easy-to-use tool. Although still under study, it is clear that it can be used to obtain results in a very general way, and also in settings other than that of sumsets.

This work may be continued in several ways. On the one hand, Kontoyiannis and Madiman have shown [16] that mutual information techniques allow to prove similar entropy inequalities when considering differential entropies. Differential entropy is the generalization of entropy to continuous random variables, and the more traditional results do no longer hold in this setting (for one, differential entropy may be negative, and many of the entropy results we have are based on its nonnegativity; similarly, one can prove that differential entropy is not functionally submodular). A thorough study of differential entropy inequalities analogous to sumset inequalities is certainly interesting by itself, and more so if accompanied by the results discussed in this thesis.

In a different direction, one may consider inverse problems in additive combiantorics, and ask if there are any parallel results for entropy. And the answer is positive: Tao [31] studies several inverse properties of entropic inequalities, and proves a theorem analogous to the celebrated Freiman-Ruzsa theorem. A study of these results, although quite more advanced in technique than what we have been showing, would be a nice complement for this thesis.

Finally, one must consider a more thorough study of the Ruzsa device. Its usefulness has been clearly shown, and what we still lack is a clear application in more general settings. The problem of finding new polynomial expanders is an exciting one.

# References


[1] Balister, P. and Bollobás, B. "Projections, entropy and sumsets". In: *Combinatorica* 32.2 (2012), pp. 125–141. ISSN: 0209-9683. DOI: 10.1007/s00493-012-2453-1.

[2] Balog, A. and Szemerédi, E. "A statistical theorem of set addition". In: *Combinatorica* 14.3 (1994), pp. 263–268. ISSN: 0209-9683. DOI: 10.1007/BF01212974.

[3] Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities*. A nonasymptotic theory of independence. Oxford University Press, Oxford, 2013, pp. x+481. ISBN: 978-0-19-953525-5. DOI: 10.1093/acprof:oso/9780199535255.001.0001.

[4] Bourgain, J. "More on the sum-product phenomenon in prime fields and its applications". In: *Int. J. Number Theory* 1.1 (2005), pp. 1–32. ISSN: 1793-0421. DOI: 10.1142/S1793042105000108.

[5] Bourgain, J., Katz, N., and Tao, T. "A sum-product estimate in finite fields, and applications". In: *Geom. Funct. Anal.* 14.1 (2004), pp. 27–57. ISSN: 1016-443X. DOI: 10.1007/s00039-004-0451-1.

[6] Chung, F. R. K. et al. "Some intersection theorems for ordered sets and graphs". In: *J. Combin. Theory Ser. A* 43.1 (1986), pp. 23–37. ISSN: 0097-3165. DOI: 10.1016/0097-3165(86)90019-1.

[7] Cover, T. M. and Thomas, J. A. *Elements of information theory*. Second. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006, pp. xxiv+748. ISBN: 978-0-471-24195-9; 0-471-24195-4.

[8] Espuny Díaz, A. "Classical and modern approaches for Plünnecke-type inequalities". B.S. Thesis. Universitat Politècnica de Catalunya, 2015. URL: http://upcommons.upc.edu/handle/2117/77022.

[9] Gowers, W. T. "A new proof of Szemerédi's theorem for arithmetic progressions of length four". In: *Geom. Funct. Anal.* 8.3 (1998), pp. 529–551. ISSN: 1016-443X. DOI: 10.1007/s000390050065.

[10] Gowers, W. T. "A new proof of Szemerédi's theorem". In: *Geom. Funct. Anal.* 11.3 (2001), pp. 465–588. ISSN: 1016-443X. DOI: 10.1007/s00039-001-0332-9.

[11] Gyarmati, K., Matolcsi, M., and Ruzsa, I. Z. "Plünnecke's inequality for different summands". In: Bolyai Soc. Math. Stud. 19 (2008), pp. 309–320. DOI: 10.1007/978-3-540-85221-6_10.

[12] Gyarmati, K., Matolcsi, M., and Ruzsa, I. Z. "A superadditivity and submultiplicativity property for cardinalities of sumsets". In: *Combinatorica* 30.2 (2010), pp. 163–174. ISSN: 0209-9683. DOI: 10.1007/s00493-010-2413-6.

[13] Han, T. S. "Nonnegative entropy measures of multivariate symmetric correlations". In: *Information and Control* 36.2 (1978), pp. 133–156. ISSN: 0890-5401.

[14] Hegyvári, N. and Hennecart, F. "Explicit constructions of extractors and expanders". In: *Acta Arith.* 140.3 (2009), pp. 233–249. ISSN: 0065-1036. DOI: 10.4064/aa140-3-2.

[15] Katz, N. H. and Tao, T. "Bounds on arithmetic projections, and applications to the Kakeya conjecture". In: *Math. Res. Lett.* 6.5-6 (1999), pp. 625–630. ISSN: 1073-2780. DOI: 10.4310/MRL.1999.v6.n6.a3.

[16] Kontoyiannis, I. and Madiman, M. "Sumset and inverse sumset inequalities for differential entropy and mutual information". In: *IEEE Trans. Inform. Theory* 60.8 (2014), pp. 4503–4514. ISSN: 0018-9448. DOI: 10.1109/TIT.2014.2322861.

[17] Loomis, L. H. and Whitney, H. "An inequality related to the isoperimetric inequality". In: *Bull. Amer. Math. Soc* 55 (1949), pp. 961–962. ISSN: 0002-9904.

[18] Madiman, M. "On the entropy of sums". In: *Proc. IEEE Inform. Theory Workshop*. 2008, pp. 303–307.

[19] Madiman, M., Marcus, A. W., and Tetali, P. "Entropy and set cardinality inequalities for partition-determined functions". In: *Random Structures Algorithms* 40.4 (2012), pp. 399–424. ISSN: 1042-9832. DOI: 10.1002/rsa.20385.

[20] Madiman, M. and Tetali, P. "Information inequalities for joint distributions, with interpretations and applications". In: *IEEE Trans. Inform. Theory* 56.6 (2010), pp. 2699–2713. ISSN: 0018-9448. DOI: 10.1109/TIT.2010.2046253.

[21] Plünnecke, H. "Eine zahlentheoretische Anwendung der Graphentheorie". In: *J. Reine Angew. Math.* 243 (1970), pp. 171–183. ISSN: 0075-4102.

[22] Ruzsa, I. Z. "An application of graph theory to additive number theory". In: *Sci. Ser. A Math. Sci. (N.S.)* 3 (1989), pp. 97–109. ISSN: 0716-8446.

[23] Ruzsa, I. Z. "Addendum to: An application of graph theory to additive number theory". In: *Sci. Ser. A Math. Sci. (N.S.)* 4 (1990/1991), pp. 93–94.

[24] Ruzsa, I. Z. "Sums of finite sets". In: *Number theory (New York, 1991–1995)*. Springer, New York, 1996, pp. 281–293. DOI: 10.1007/978-1-4612-2418-1_21.

[25] Ruzsa, I. Z. "Sumsets and entropy". In: *Random Structures Algorithms* 34.1 (2009), pp. 1–10. ISSN: 1042-9832. DOI: 10.1002/rsa.20248.

[26] Schoen, T. "New bounds in Balog-Szemerédi-Gowers theorem". In: *Combinatorica* 35.6 (2015), pp. 695–701. ISSN: 0209-9683. DOI: 10.1007/s00493-014-3077-4.

[27] Solymosi, J. "Incidences and the spectra of graphs". In: *Combinatorial number theory and additive group theory*. Adv. Courses Math. CRM Barcelona. Birkhäuser Verlag, Basel, 2009, pp. 299–314. DOI: 10.1007/978-3-7643-8962-8_22.

[28] Sudakov, B., Szemerédi, E., and Vu, V. H. "On a question of Erdős and Moser". In: *Duke Math. J.* 129.1 (2005), pp. 129–155. ISSN: 0012-7094. DOI: 10.1215/S0012-7094-04-12915-X.

[29] Tao, T. *An entropy Plünnecke-Ruzsa inequality*. 2009. URL: https://terrytao.wordpress.com/2009/10/27/an-entropy-plunnecke-ruzsa-inequality/.

[30] Tao, T. *Sumset and inverse sumset theorems for Shannon entropy*. 2009. URL: https://terrytao.wordpress.com/2009/06/25/sumset-and-inverse-sumset-theorems-for-shannon-entropy/.

[31] Tao, T. "Sumset and inverse sumset theory for Shannon entropy". In: *Combin. Probab. Comput.* 19.4 (2010), pp. 603–639. ISSN: 0963-5483. DOI: 10.1017/S0963548309990642.

[32] Tao, T. and Vu, V. H. *Additive combinatorics*. Vol. 105. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006, pp. xviii+512. ISBN: 978-0-521-85386-6; 0-521-85386-9. DOI: 10.1017/CBO9780511755149.