Entropy and additive combinatorics

by

Marcel K. Goh

15 February 2024

Abstract. These expository notes give a gentle introduction to the notion of entropy as it is used in additive combinatorics, with a view towards understanding the proof of the polynomial Freiman–Ruzsa conjecture by W. T. Gowers, B. Green, F. Manners, and T. Tao. Much of the first section has been transcribed from lectures given by W. T. Gowers.

1. The Khintchine–Shannon axioms

Let X be a discrete random variable. Its entropy $\mathbf{H}\{X\}$ is a real number (or ∞) that measures the "information content" of X. For example, if X is a constant random variable, then $\mathbf{H}\{X\}$ should be zero (we do not gain any information from knowing the value of X), and if X is uniformly distributed on $\{0,1\}^n$, then $\mathbf{H}\{X\}$ should be proportional to n, since X is determined by n bits of information. It satisfies the following axioms, which are sometimes called the Khinchine–Shannon axioms.

a) (*Invariance.*) If X takes values in A, Y takes values in B, $\phi : A \to B$ is a bijection, and $\mathbf{P}\{Y = \phi(a)\} = \mathbf{P}\{X = a\}$ for all $a \in A$, then $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$.

b) (*Extensibility.*) If X takes values in A and Y takes values in B for a set B such that $A \subseteq B$, and furthermore $\mathbf{P}\{Y = a\} = \mathbf{P}\{X = a\}$ for all $a \in A$, then $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$.

c) (*Continuity.*) The quantity $\mathbf{H}\{X\}$ depends continuously on the probabilities $\mathbf{P}\{X = a\}$.

d) (*Maximisation.*) Over all possible random variables X taking values in a finite set A, the quantity $\mathbf{H}\{X\}$ is maximised for the uniform distribution.

e) (*Additivity.*) For X taking values in A and Y taking values in B, we have the formula

$$\mathbf{H}\{X, Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y \mid X\},$$

where $\mathbf{H}\{X, Y\} = \mathbf{H}\big\{(X, Y)\big\}$ and

$$\mathbf{H}\{Y \mid X\} = \sum_{x \in A} \mathbf{P}\{X = x\}\, \mathbf{H}\{Y \mid X = x\}.$$

We shall take it on faith that there really exists a function on random variables satisfying these axioms. In fact, the axioms only define entropy up to a multiplicative constant, so we shall add the following axiom. (It is very possible you have met this function somewhere on your travels, but you will not find it anywhere in these notes.)

f) (*Normalisation.*) If X is uniformly distributed on $\{0,1\}$, then $\mathbf{H}\{X\} = 1$.

Notationally, we would expect that $\mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\}$ if X and Y are independent. This is the first proposition we will carefully prove using the axioms above.

Proposition 1.1. *Let X and Y be independent random variables. Then $\mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\}$ and consequently $\mathbf{H}\{X,Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y\}$.*

Proof. Suppose X takes values in a finite set A. Then for all $x \in A$, the distribution of Y and Y given that $X = x$ is the same, so

$$\mathbf{H}\{Y \mid X\} = \sum_{x\in A} \mathbf{P}\{X = x\}\,\mathbf{H}\{Y \mid X = x\} = \sum_{x\in A} \mathbf{P}\{X = x\}\,\mathbf{H}\{Y\} = \mathbf{H}\{Y\}.$$

The second version of the statement follows from the additivity axiom. ∎

We will sometimes use the notation X^n to denote the vector $(X_1, \dots, X_n)$ where the X_i are independent copies of the random variable X. We have the following three corollaries, which are each proved by induction. The second also requires the normalisation axiom, and the third is often known as the *chain rule.*

Corollary 1.2. *We have $\mathbf{H}\{X^n\} = n\,\mathbf{H}\{X\}$.* ∎

Corollary 1.3. *If X is uniformly distributed on a set of size 2^n, then*

$$\mathbf{H}\{X\} = n. \quad ∎$$

Corollary 1.4 (*Chain rule*). *Let $X_1, \dots, X_n$ be random variables. Then*

$$\mathbf{H}\{X_1, \dots, X_n\} = \mathbf{H}\{X_1\} + \mathbf{H}\{X_2 \mid X_1\} + \cdots + \mathbf{H}\{X_n \mid X_1, \dots, X_{n-1}\}. \quad ∎$$

Next we establish the intuitive fact that the entropy of a uniform random variable supported on a set A is at most the entropy of a uniform random variable supported on a superset B of A.

Proposition 1.5. *Let $A \subseteq B$ with B finite, let X be uniformly distributed on A, and let Y be uniformly distributed on B. Then $\mathbf{H}\{X\} \le \mathbf{H}\{Y\}$, with equality if and only if $A = B$.*

Proof. By the extensibility axiom, $\mathbf{H}\{X\}$ is not affected if we regard X as a function taking values in B. Then by the maximisation axiom, $\mathbf{H}\{X\} \le \mathbf{H}\{Y\}$, since Y is uniform on B.

If $A = B$, then it is clear that $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$, since X and Y are the same random variable.

On the other hand, say $|A| = m$ and $|B| = n$ with $m < n$. If $m = 1$, then by the previous proposition we have $\mathbf{H}\{X\} = 0$, and by normalisation and invariance, $\mathbf{H}\{Y\} = 1$. When $m \ge 2$, pick k such that $m^k \le n^{k-1}$, so that $|A^k| \le |B^{k-1}|$. Then by Corollary 1.2 and the inequality we showed in the first paragraph of this proof, we have

$$n\,\mathbf{H}\{X\} = \mathbf{H}\{X^n\} \le \mathbf{H}\{X^{n-1}\} = (n-1)\,\mathbf{H}\{Y\},$$

whence $\mathbf{H}\{X\} < \mathbf{H}\{Y\}$. ▮

The link between entropy and additive combinatorics is rather a fine one. It is based on the following observation. (In this and the rest of the notes, lg denotes the base-2 logarithm. This base can be changed by modifying the normalisation axiom.)

Proposition 1.6. *Let X be a uniform random variable on a finite set A. Then*

$$\mathbf{H}\{X\} = \lg|A|.$$

Proof. For any positive integer n we can let X^n denote a tuple of independent copies of X; Corollary 1.2 tells us $\mathbf{H}\{X^n\} = n\,\mathbf{H}\{X\}$. Let m be such that $2^m \le |A|^n \le 2^{m+1}$ so that

$$\frac{m}{n} \le \lg|A| \le \frac{(m+1)}{n}.$$

Let Y be uniform on a set of size 2^m, and let Z be uniform on a set of size 2^{m+1}. Then by Corollary 1.3 we have $\mathbf{H}\{Y\} = m$ and $\mathbf{H}\{Z\} = (m+1)$. Then by Proposition 1.5 we have

$$\frac{m}{n} \le \mathbf{H}\{X\} \le \frac{(m+1)}{n}.$$

In other words, $\mathbf{H}\{X\}$ satisfies the same bounds as $\lg|A|$. Taking n large, we can make these bounds arbitrarily tight, proving the claim. ▮

The maximisation axiom gives the following corollary.

Corollary 1.7. *Let X be a random variable supported on a finite set A. Then*

$$\mathbf{H}\{X\} \le \lg|A|. \quad ▮$$

Hence the entropy $\mathbf{H}\{X\}$ is at most the exponential of the size of its support. As we will see, simply replacing (logarithms of) cardinalities with entropies, we at useful "entropic analogues" of combinatorial statements. But first, we need more lemmas.

If Y is a random variable such that $Y = f(X)$ for some random variable Y and some function f, then we say that Y is *determined by* X or X *determines* Y. We want to show that $\mathbf{H}\{Y\} \le \mathbf{H}\{X\}$, which reflects the idea that we get more information from X than from Y. This, rather annoyingly, seems to require a couple of steps.

Lemma 1.8. *If* $Y = f(X)$ *then* $\mathbf{H}\{X\} = \mathbf{H}\{Y\} + \mathbf{H}\{X \mid Y\}$.

Proof. There is a a bijection between values x taken by X and values $\big(x, f(x)\big)$ taken by (X, Y), so we have

$$\mathbf{H}\{X\} = \mathbf{H}\{X, Y\} = \mathbf{H}\{Y\} + \mathbf{H}\{X \mid Y\}$$

by invariance and additivity. ▮

We are now done if we can show that entropy is nonnegative. This is a corollary of the following lemma. The following proof is a modification of one due to S. Eberhard.

Proposition 1.9. *Let X be a discrete random variable supported on a set A and let*

$$a^* = \operatorname*{arg\,max}_{a \in A} \mathbf{P}\{X = a\}.$$

Then

$$\mathbf{P}\{X = a^*\} \geq 2^{-\mathbf{H}\{X\}}.$$

Proof. First we will work in the case where there exists n such that $\mathbf{P}\{X = a\}$ is a multiple of $1/n$ for all $a \in A$. Let Y be uniformly distributed on $[n]$ and let $\{E_a\}_{a \in A}$ be a partition of $[n]$ such that $|E_a| = n\,\mathbf{P}\{X = a\}$ for all $a \in A$, and let $Z = a$ if $Y \in E_a$. This definition makes Z and X identically distributed, so $\mathbf{H}\{Z\} = \mathbf{H}\{X\}$ by the invariance axiom, and it suffices to prove $\mathbf{H}\{Z\} \geq 0$.

For every $a \in A$, the conditional entropy $\mathbf{H}\{Y \mid Z = a\}$ is uniformly distributed on a set of size $|E_a|$. From our choice of a^* we have $|E_{a^*}| \geq |E_a|$ for all $a \in A$. Hence by Proposition 1.6, we have

$$\mathbf{H}\{Y \mid Z\} = \sum_{a \in A} \mathbf{P}\{X = a\}\,\mathbf{H}\{Y \mid X = a\} \sum_{a \in A} \mathbf{P}\{X = a\} \log |E_a| \leq \log |E_{a^*}|.$$

Since Z is determined by Y, we have $\mathbf{H}\{Y\} = \mathbf{H}\{Z\} + \mathbf{H}\{Y \mid Z\}$ by the previous lemma. So by another invocation of Proposition 1.6, we have

$$\begin{aligned}
\mathbf{H}\{Z\} &= \mathbf{H}\{Y\} - \mathbf{H}\{Y \mid Z\} \\
&\geq \log n - \log |E_{a^*}| \\
&\geq \log\left(\frac{n}{|E_{a^*}|}\right) \\
&= \log\left(\frac{1}{\mathbf{P}\{X = a^*\}}\right),
\end{aligned}$$

and hence $2^{-\mathbf{H}\{X\}} = 2^{-\mathbf{H}\{Z\}} \leq \mathbf{P}\{X = a^*\}$.

The general case follows from the continuity axiom. ▮

This proof came dangerously close to deriving the formula for entropy, but we will not need any such formula, so we will refrain from mentioning it. From the fact that $\mathbf{P}\{X = a^*\} \leq 1$ we can immediately conclude that entropy is nonnegative.

Corollary 1.10. *Let X be a discrete random variable taking values in a finite set A. Then* $\mathbf{H}\{X\} \geq 0$. ▮

This observation completes the proof that a random variable has a smaller entropy than one by which it is determined.

Corollary 1.11. *If* $Y = f(X)$ *then* $\mathbf{H}\{X\} \geq \mathbf{H}\{Y\}$. ▮

Next we show that a random variable has zero entropy if and only if it is constant. This reflects the idea that the variables from which we get no information are those which take the same value no matter what.

Proposition 1.12. *Let X be a discrete random variable. Then* $\mathbf{H}\{X\} = 0$ *if and only if it takes exactly one value.*

Proof. First suppose that X takes only one value. Let a be the value of X such that $\mathbf{P}\{X = a\} = 1$. Then (X, X) equals (a, a) with probability 1 as well, so $\mathbf{H}\{X\} = \mathbf{H}\{X, X\}$ by the invariance axiom. But it can easily be checked that X and (X, X) are independent (we have

$$\mathbf{P}\bigl\{X = a, (X, X) = (a, a)\bigr\} = \mathbf{P}\{X = a\}\,\mathbf{P}\bigl\{(X, X) = (a, a)\bigr\}$$

for instance), so $\mathbf{H}\{X, X\} = 2\,\mathbf{H}\{X\}$. Thus we conclude that $\mathbf{H}\{X\} = 0$.

Now suppose that X takes more than one value; let A be the set of a such that $\mathbf{P}\{X = a\} > 0$ and let $\alpha = \max_{a\in A} \mathbf{P}\{X = a\}$. For all n let X^n denote the tuple of n independent copies of X; the maximum probability of any particular value (in A^n) that X^n takes is α^n. But $\alpha < 1$ since X takes more than one value, so for any $\epsilon > 0$ we can find n such that $\alpha^n < \epsilon$. This means that we can partition A^n into two disjoint sets E and F such that $\mathbf{P}\{X^n \in E\}$ and $\mathbf{P}\{X^n \in F\}$ are both in the range $[1/2 - \epsilon, 1/2 + \epsilon]$.

Let Y be the random variable taking the value 0 if $X^n \in E$ and 1 if $X^n \in F$. Then by Corollary 1.2, $\mathbf{H}\{X^n\} = n\,\mathbf{H}\{X\}$, and since X^n determines Y,

$$\mathbf{H}\{X^n\} = \mathbf{H}\{Y\} + \mathbf{H}\{X^n \mid Y\} \geq \mathbf{H}\{Y\}.$$

But $\mathbf{H}\{Y\} > 0$ for ϵ small enough, the normalisation and continuity axioms. So $\mathbf{H}\{X\} \geq \mathbf{H}\{Y\}/n > 0$ as well. ▮

Mutual information. For random variables X and Y, the *mutual information* $\mathbf{I}\{X : Y\}$ is defined by the equivalent formulas

$$\begin{aligned}\mathbf{I}\{X : Y\} &= \mathbf{H}\{X\} + \mathbf{H}\{Y\} - \mathbf{H}\{X, Y\}\\ &= \mathbf{H}\{X\} - \mathbf{H}\{X \mid Y\}\\ &= \mathbf{H}\{Y\} - \mathbf{H}\{Y \mid X\}.\end{aligned}$$

It measures, roughly speaking, how much information one can get from one variable by looking at the other one. From the formula it is clear that $\mathbf{I}\{X : Y\} = 0$ if X and Y are independent. In general, we still have the inequality $\mathbf{I}\{X : Y\} \geq 0$, which a corollary of the following lemma, which expresses the intuitive fact that conditioning cannot increase the entropy of a random variable. The proof, which is due to C. West, uses an argument similar to the one we used to prove Lemma 1.9.

Proposition 1.13. *Let X and Y be discrete random variables. Then*

$$\mathbf{H}\{X \mid Y\} \leq \mathbf{H}\{X\}.$$

Proof. Let A be the support of X and B be the support of Y. First we consider the case that X is uniform on A (so A is finite). Then by the definition of conditional entropy,

$$\mathbf{H}\{X \mid Y\} = \sum_{b \in B} \mathbf{P}\{Y = b\}\, \mathbf{H}\{X \mid Y = b\}.$$

But for each b, the random variable $(X \,|\, Y = b)$ takes values in A, so its entropy is bounded above by $\mathbf{H}\{X\}$ by the maximisation axiom. Hence $\mathbf{H}\{X|Y\} \leq \mathbf{H}\{X\}$.

Next, suppose that A and B are both finite and suppose further that $\mathbf{P}\{Y = b\}$ is rational for all b. Then there is an integer n and integers $\{m_b\}_{b \in B}$ such that $\mathbf{P}\{Y = b\} = m_b/n$ for all $b \in B$. Now partition $[n]$ into sets $\{E_b\}_{b \in B}$, where $|E_b| = m_b$ for all $b \in B$. We define a random variable Z by sampling uniformly at random from E_b if $Y = b$, and doing so independently of $(X \mid Y = b)$. The result is a random variable Z that is uniform on $[n]$ and which is independent of $X \mid Y$. Furthermore, since Z determines Y, we have $\mathbf{H}\{Z\} = \mathbf{H}\{Y, Z\}$ by the invariance axiom. Hence

$$\begin{aligned}
\mathbf{H}\{X \mid Y\} &= \mathbf{H}\{X \mid Y, Z\} \\
&= \mathbf{H}\{X, Y, Z\} - \mathbf{H}\{Y, Z\} \\
&= \mathbf{H}\{X, Z\} - \mathbf{H}\{Z\} \\
&= \mathbf{H}\{X \mid Z\} \\
&\leq \mathbf{H}\{X\},
\end{aligned}$$

where the inequality on the last line follows from the previous paragraph.

The general case follows from the continuity axiom and the fact that any discrete random variable can be approximated by discrete random variables with finite support and rational probabilities. ∎

By the additivity axiom, the previous proposition is equivalent to

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

and we shall use this to prove the following submodularity inequality.

Proposition 1.14 (*Submodularity*). *Suppose X, Y, Z, and W are random variables such that (Z, W) determines X, Z determines Y, and W also determines Y. Then*

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

Proof. The hypotheses give the three inequalities

$$\mathbf{H}\{X\} \leq \mathbf{H}\{Z, W\}, \quad \mathbf{H}\{Y\} \leq \mathbf{H}\{Z\}, \quad \text{and} \quad \mathbf{H}\{Y\} \leq \mathbf{H}\{W\}.$$

From this we see that

$$2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} \le 2\,\mathbf{H}\{Z, W\} + \mathbf{H}\{Z\} + \mathbf{H}\{Y\}.$$

But then since conditioning does not increase entropy, we have

$$2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} \le 2\,\mathbf{H}\{Z\} + 2\,\mathbf{H}\{Y\},$$

whence dividing both sides by 2 completes the proof. ∎

The submodularity inequality is often stated in terms of a triple of random variables in terms of the *conditional mutual information*, which is defined by

$$\mathbf{I}\{X : Y \mid Z\} = \sum_{z\in C} p_Z(z)\,\mathbf{I}\bigl\{(X \mid Z = z) : (Y \mid Z = z)\bigr\}.$$

Proposition 1.15. *Let X, Y, and Z be discrete random variables. Then*

$$\mathbf{H}\{X, Y, Z\} + \mathbf{H}\{Z\} \le \mathbf{H}\{X, Z\} + \mathbf{H}\{Y, Z\},$$

which is equivalent to

$$\mathbf{I}\{X : Y \mid Z\} \ge 0.$$

Proof. It is clear from the additivity axiom that both sides in the first inequality are equal to $\mathbf{H}\{X\} + \mathbf{H}\{Y\} + 2\,\mathbf{H}\{Z\}$ if and only if X and Y are independent conditional on Z.

To prove the first inequality, note that (X, Y, Z) is jointly determined by (X, Z) and (Y, Z), and Z is determined by both (X, Z) and (Y, Z) separately, then apply the previous proposition. Now by the definition of conditional mutual information,

$$\begin{aligned}\mathbf{I}\{X : Y \mid Z\} &= \sum_{z\in C} p_Z(z)\,\mathbf{I}\bigl\{(X \mid Z = z) : (Y \mid Z = z)\bigr\}\\ &= \mathbf{H}\{X \mid Z\} + \mathbf{H}\{Y \mid Z\} - \mathbf{H}\{X, Y \mid Z\}\\ &= \mathbf{H}\{X, Z\} - 2\,\mathbf{H}\{Z\} + \mathbf{H}\{Y, Z\} - \mathbf{H}\{X, Y, Z\} + \mathbf{H}\{Z\}\\ &= \mathbf{H}\{X, Z\} + \mathbf{H}\{Y, Z\} - \mathbf{H}\{X, Y, Z\} - \mathbf{H}\{Z\},\end{aligned}$$

and this proves that the first statement is equivalent to the second. ∎

2. Group-valued random variables

Now we will examine the case where the random variables in question take values in an abelian group G, meaning we can take sums $X + Y$ and differences $X - Y$ of them. Note that if we condition on Y, then the values taken $X + Y$ are in bijection with values taken by X. This leads to the following proposition.

Proposition 2.1. *Let X and Y be random variables each taking finitely many values in an abelian group G. We have*

$$\max\big(\mathbf{H}\{X\}, \mathbf{H}\{Y\}\big) - \mathbf{I}\{X : Y\} \le \mathbf{H}\{X \pm Y\}.$$

Furthermore, for any random variable Z, we have the conditional version

$$\max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\} \le \mathbf{H}\{X \pm Y \mid Z\}$$

of the same statement.

Proof. Since conditioning does not increase entropy, we have

$$\mathbf{H}\{X \pm Y\} \ge \mathbf{H}\{X \pm Y \mid Y\},$$

and since the probabilities $\mathbf{P}\{X + Y = z \mid Y = y\} = \mathbf{P}\{X = z - y \mid Y = y\}$ for all $z \in G$, by invariance we have

$$\mathbf{H}\{X \pm Y\} \ge \mathbf{H}\{X \mid Y\} = \mathbf{H}\{X\} - \mathbf{I}\{X : Y\}.$$

Repeating the same argument but exchanging the roles of X and Y, we get

$$\mathbf{H}\{X \pm Y\} \ge \mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\} - \mathbf{I}\{X : Y\},$$

so

$$\mathbf{H}\{X \pm Y\} \ge \max\big(\mathbf{H}\{X\}, \mathbf{H}\{Y\}\big) - \mathbf{I}\{X : Y\}.$$

Now let Z be any random variable with finite support.

$$\begin{aligned}
\mathbf{H}\{X \pm Y \mid Z\} &= \sum_{z \in G} \mathbf{P}\{Z = z\}\, \mathbf{H}\{X \pm Y \mid Z = z\} \\
&\ge \Big(\max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\}\Big) \sum_{z \in G} \mathbf{P}\{Z = z\} \\
&= \max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\},
\end{aligned}$$

which completes the proof. ▮

Corollary 2.2. *If X and Y are independent, then*

$$\max\big(\mathbf{H}\{X\}, \mathbf{H}\{Y\}\big) \le \mathbf{H}\{X \pm Y\}.$$

Proof. The mutual information $\mathbf{I}\{X : Y\}$ is zero whenever X and Y are independent. ▮

Entropic Ruzsa distance. In additive combinatorics, whenever we have two finite subsets A and B of the same abelian group, we can compute the Ruzsa distance

$$d(A, B) = \lg \frac{|A - B|}{\sqrt{|A| \cdot |B|}}$$

between them. (This satisfies all the axioms of a metric except the one requiring $d(A, A) = 0$ for all sets A.)

The entropic analogue of the Ruzsa distance is defined as follows. For finitely supported random variables X and Y taking values in the same abelian group, we let X' and Y' be independent copies of X and Y, respectively, and define the *entropic Ruzsa distance* by

$$\mathbf{d}\{X, Y\} = \mathbf{H}\{X' - Y'\} - \frac{\mathbf{H}\{X'\}}{2} - \frac{\mathbf{H}\{Y'\}}{2}.$$

This definition only depends on the individual distributions of X and Y and does not require them to have the same sample space. Once again, we don't necessarily have $\mathbf{d}\{X, X\} = 0$, but we do have the triangle inequality, which shall now prove.

Proposition 2.3. *Let X, Y, and Z be random variables with finite support in the same abelian group. Then*

$$\mathbf{d}\{X, Z\} \le \mathbf{d}\{X, Y\} + \mathbf{d}\{Y, Z\},$$

which is equivalent to

$$\mathbf{H}\{X' - Z'\} \le \mathbf{H}\{X' - Y'\} + \mathbf{H}\{Y' - Z'\} - \mathbf{H}\{Y'\}$$

for X', Y', and Z' independent and distributed as X, Y, and Z, respectively.

Proof. That the two statements are equivalent is easily obtained by expanding the definition of entropic Ruzsa distance and cancelling some terms. So without loss of generality, we may assume that that X, Y, and Z are independent and just prove the second statement.

By submodularity, we have $\mathbf{I}\bigl\{(X - Y : Z) \mid X - Z\bigr\} \ge 0$, so

$$\begin{aligned} 0 &\le \mathbf{I}\bigl\{(X - Y : Z) \mid X - Z\bigr\} \\ &\le \mathbf{H}\{X - Y \mid X - Z\} + \mathbf{H}\{Z \mid X - Z\} - \mathbf{H}\{X - Y, Z \mid X - Z\} \\ &\le \mathbf{H}\{X - Y, X - Z\} + \mathbf{H}\{Z, X - Z\} - \mathbf{H}\{X - Y, Z, X - Z\} - \mathbf{H}\{X - Z\}. \end{aligned} \tag{1}$$

Now, since the values $(x - y, x - z)$ taken by $(X - Y, X - Z)$ are in bijection with values $(x - z, y - z)$ taken by $(X - Z, Y - Z)$ via the map $(v, w) \mapsto (w, w - v)$, by the invariance axiom we have

$$\mathbf{H}\{X - Y, X - Z\} = \mathbf{H}\{X - Z, Y - Z\},$$

and

$$\mathbf{H}\{X - Y, X - Z\} \le \mathbf{H}\{X - Y\} + \mathbf{H}\{Y - Z\}$$

follows by submodularity. Similar invocations of the invariance axiom give

$$\mathbf{H}\{Z, X - Z\} = \mathbf{H}\{X, Z\}$$

and

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

where in the latter statement the second equality follows from the fact that (X, Y) and Z are independent. Substituting these three inequalities into (1), we have

$$0 \leq \mathbf{H}\{X-Y\} + \mathbf{H}\{Y-Z\} + \mathbf{H}\{X, Z\} - \mathbf{H}\{X, Z\} + \mathbf{H}\{Y\} - \mathbf{H}\{X-Z\},$$

whence

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

which completes the proof. ▮

We also define a conditional version of the entropic Ruzsa distance. If X and Y are G-valued random variables with finite support and Z and W are any random variables with finite supports A and B respectively, then we define

$$\mathbf{d}\{X \mid Z; Y \mid W\} = \sum_{z \in A} \sum_{w \in B} \mathbf{P}\{Z = z\} \, \mathbf{P}\{W = w\} \, \mathbf{d}\big\{(X \mid Z = z); (Y \mid W = w)\big\}.$$

If (X', Z') and (Y', W') are independent copies of (X, Z) and (Y, W) respectively, then this distance is also given by the formula

$$\mathbf{d}\{X \mid Z; Y \mid W\} = \mathbf{H}\{X' - Y' \mid Z', W'\} - \frac{\mathbf{H}\{X' \mid Z'\}}{2} - \frac{\mathbf{H}\{Y' \mid W'\}}{2}.$$

The sum-difference inequality. The Ruzsa triangle inequality bounds the size of a difference set by passing through a different subset. There is another inequality that relates the size of a sumset with the size of a difference set. If A and B are nonempty finite subsets of an abelian group G, then

$$|A+B| \leq \frac{|A-B|^3}{|A||B|}.$$

If we replace cardinalities by exponentials of entropies, then we obtain the statement of the following proposition.

Proposition 2.4. *Let X and Y be independent random variables taking values in the same abelian group. Then*

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

Before we proceed to the proof, we establish the following definition, which will be needed later in these notes as well. Let X and Y be random variables (not necessarily independent). We say that X_1 and Y_1 are *conditionally independent trials of X and Y relative to Z* if for all z in the range of Z, the random variables distributed as $(X_1 \mid Z = z)$ and $(Y_1 \mid Z = z)$ are independent, $(X_1 \mid Z = z)$ has the same distribution as $(X \mid Z = z)$, and similarly for Y_1 and Y. In particular, if $X = Y$ and X_1 and X_2 are conditionally independent trials of X relative to Z, we have

$$\mathbf{H}\{X_1, X_2 \mid Z\} = \mathbf{H}\{X_1 \mid Z\} + \mathbf{H}\{X_2 \mid Z\} = 2\,\mathbf{H}\{X \mid Z\},$$

by additivity and independence. From this we obtain

$$\mathbf{H}\{X_1, X_2, Z\} = 2\,\mathbf{H}\{X \mid Z\} + \mathbf{H}\{Z\} = 2\,\mathbf{H}\{X, Z\} - \mathbf{H}\{Z\}. \tag{2}$$

It is also important to observe that (X_1, Z) and (X_2, Z) both have the same distributions as (X, Z).

Proof. Let (X_1, Y_1) and (X_2, Y_2) be conditionally independent trials of (X, Y) relative to $X - Y$. Since (X, Y) determines $X - Y$, we have $X_1 - Y_1 == X - Y = X_2 - Y_2$. Let (X_3, Y_3) be another trial of (X, Y) independent of (X_1, X_2, Y_1, Y_2). Then

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

so $(X_3 - Y_2, X_1 - Y_3, X_2, Y_1)$ and (X_3, Y_3) each determine $X_3 + Y_3$. On the other hand, $(X_3 - Y_2, X_1 - Y_3, X_2, Y_1)$ and (X_3, Y_3) together determine the sextuple $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$, so by the submodularity inequality, we have

$$\begin{aligned}\mathbf{H}\{X_1, X_2, X_3, Y_1, Y_2,&Y_3\} + \mathbf{H}\{X_3 + Y_3\} \\ &\le \mathbf{H}\{X_3 - Y_2, X_1 - Y_3, X_2, Y_1\} + \mathbf{H}\{X_3, Y_3\}.\end{aligned} \tag{3}$$

We have

$$\mathbf{H}\{X_3, Y_3\} = \mathbf{H}\{X, Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y\}$$

by independence of X and Y, and since (X_3, Y_3) and (X_1, X_2, Y_1, Y_2) are independent, we have

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

where in the third line we applied (2). On the other hand,

$$\mathbf{H}\{X_3 + Y_3\} = \mathbf{H}\{X + Y\}$$

and

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

Substituting everything into (3) yields

$$\begin{aligned} 3\,\mathbf{H}\{X\} + 3\,\mathbf{H}\{Y\} - \mathbf{H}\{X-Y\} + \mathbf{H}\{X+Y\} \\ \leq 2\,\mathbf{H}\{X-Y\} + 2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\}, \end{aligned}$$

and the desired inequality follows upon rearrangement of terms. ▮

The entropic sum-difference inequality can also be stated in terms of entropic Ruzsa distances; independence is not necessary here because independent trials are baked into the definition of entropic Ruzsa distance.

Corollary 2.5. *Let X and Y be discrete random variables taking values in the same abelian group. Then*

$$\mathbf{d}\{X, -Y\} \leq 3\,\mathbf{d}\{X, Y\}.$$ ▮

3. The Plünnecke–Ruzsa inequality

In additive combinatorics, one of the most useful sumset inequalities is the following.

Theorem 3.1 (*Plünnecke–Ruzsa inequality*). *Let A and B be finite subsets of an abelian group and suppose that $|A+B| \leq K|A|$ for some constant K. Then for any integers $r, s \geq 0$, not both zero, we have $|rB - sB| \leq K^{r+s}|A|$.* ▮

In this section we will develop an entropic analogue of this statement, in which sets are replaced by random variables of finite support and cardinality is replaced with the exponential of entropy. First, a technical lemma.

Lemma 3.2. *Let X, Y, and Z be independent random variables taking values in a common abelian group. Then*

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

Proof. By submodularity, the quantity $\mathbf{I}\{X : Z \mid X+Y+Z\}$ is nonnegative, so we have

$$\begin{aligned} 0 &\leq \mathbf{I}\{X : Z \mid X+Y+Z\} \\ &= \mathbf{H}\{X, X+Y+Z\} + \mathbf{H}\{Z, X+Y+Z\} \\ &\qquad - \mathbf{H}\{X, Z, X+Y+Z\} - \mathbf{H}\{X+Y+Z\}. \end{aligned}$$

Since X, Y, and Z are independent, we have

$$\mathbf{H}\{X, X+Y+Z\} = \mathbf{H}\{X, Y+Z\} = \mathbf{H}\{X\} + \mathbf{H}\{Y+Z\},$$

where in the first equality we use invariance. By similar reasoning we have

$$\mathbf{H}\{Z, X+Y+Z\} = \mathbf{H}\{Z\} + \mathbf{H}\{X+Y\}$$

and

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

Plugging these three identities into the inequality above yields

$$\begin{aligned} 0 &\le \mathbf{H}\{X\} + \mathbf{H}\{Y+Z\} + \mathbf{H}\{Z\} + \mathbf{H}\{X+Y\} \\ &\qquad - \mathbf{H}\{X\} - \mathbf{H}\{Y\} - \mathbf{H}\{Z\} - \mathbf{H}\{X+Y+Z\} \\ &= \mathbf{H}\{Y+Z\} + \mathbf{H}\{X+Y\} - \mathbf{H}\{Z\} - \mathbf{H}\{X+Y+Z\}, \end{aligned}$$

whence the claim follows upon rearranging. ▮

From here we are not far from proving the entropic Plünnecke–Ruzsa inequality, a result of T. Tao.

Theorem 3.3. *Let $X, Y_1, \ldots, Y_m$ be independent random variables of finite entropy taking values in an abelian group G, such that*

$$\mathbf{H}\{X + Y_i\} \le \mathbf{H}\{X\} + \log K_i$$

for all $1 \le i \le m$ and some scalars $K_1, \ldots, K_m \ge 1$. Then

$$\mathbf{H}\{X + Y_1 + \cdots + Y_m\} \le \mathbf{H}\{X\} + \log(K_1 \cdots K_m).$$

Proof. We prove the claim by induction on m. If $m = 1$, then we are done by hypothesis. Now suppose that $\mathbf{H}\{X+Y_1+\cdots+Y_{m-1}\} \le \mathbf{H}\{X\}+\log(K_1 \cdots K_{m-1})$. Then by the previous lemma, the induction hypothesis, and the hypothesis on $\mathbf{H}\{X + Y_m\}$, we have

$$\begin{aligned} \mathbf{H}\{Y_1 + \cdots + Y_{m-1} + X + Y_m\} &\le \mathbf{H}\{Y_1 + \cdots + Y_{m-1} + X\} \\ &\qquad\qquad + \mathbf{H}\{X + Y_m\} - \mathbf{H}\{X\} \\ &\le \mathbf{H}\{X\} + \log(K_1 \cdots K_{m-1}) + \log K_m \\ &\le \mathbf{H}\{X\} + \log(K_1 \cdots K_m), \end{aligned}$$

which is what we sought to prove. ▮

We can make this look bit more like the version of the Plünnecke–Ruzsa inequality above by using the triangle inequality.

Corollary 3.4 (*Entropic Plunnecke–Ruzsa inequality*). *Let X and Y be random variables with $\mathbf{H}\{X+Y\} \le \mathbf{H}\{X\} + \log K$. Then for any $r, s \ge 0$ not both zero, we have*

$$\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \le \mathbf{H}\{X\} + (r+s)\log K,$$

where $Y_1, \ldots, Y_r, Z_1, \ldots, Z_s$ are independent copies of Y.

Proof. By the entropic Ruzsa triangle inequality, we have

$$\begin{aligned} &\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \le \\ &\qquad \mathbf{H}\{Y_1 + \cdots + Y_r + X\} + \mathbf{H}\{-X - Z_1 - \cdots - Z_s\} - \mathbf{H}\{-X\}. \end{aligned}$$

The values of $-X$ are in bijection with values of X, and the values of $-X - Z_1 - \cdots - Z_s$ are in bijection with the values of $X + Z_1 + \cdots + Z_s$ (with the same probabilities in both cases), so by the invariance axiom, we have

$$\begin{aligned}&\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \leq \\ &\qquad \mathbf{H}\{X + Y_1 + \cdots + Y_r\} + \mathbf{H}\{X + Z_1 + \cdots + Z_s\} - \mathbf{H}\{X\},\end{aligned}$$

and we can apply the the previous theorem twice to get

$$\begin{aligned}\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} &\leq \mathbf{H}\{X\} + \log(K^r) + \log(K^s),\\ &= \mathbf{H}\{X\} + (r+s)\log K. \quad ∎\end{aligned}$$

4. The Balog–Szemerédi–Gowers theorem

If A and B are subsets of the same abelian group such that $|A+B|$ is much smaller than $|A| \cdot |B|$, then it stands to reason that there must be a lot of redundancy in $A + B$; that is, many elements of $A + B$ can be expressed as $a + b$ in lots of different ways. To capture this notion, we can define the *additive energy* between two sets A and B to be

$$E(A, B) = \left|\left\{(a, a', b, b') \in A \times A \times B \times B : a + b = a' + b'\right\}\right|.$$

We now define an entropic version of $E(A, B)$. Let X and Y be discrete random variables taking values in the same abelian group. Let (X_1, Y_1) and (X_2, Y_2) be conditionally independent trials of (X, Y) relative to $X + Y$. These $X_1 + Y_1 = X_2 + Y_2 = A + B$. The *entropic additive energy* between X and Y is

$$\mathbf{e}\{X, Y\} = \mathbf{H}\{X_1, Y_1, X_2, Y_2\}.$$

This definition makes clear the analogy between this value and the additive energy of sets, but by conditional independence and the fact that (X_1, Y_1, X_2, Y_2) determines $X_1 + Y_1 = X + Y$, we can rewrite

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

where in the second equality we applied (2). This formula is something we will use often, as it makes no direct mention of the variables (X_1, Y_1) and (X_2, Y_2).

Quantifying the idea that small sumset must imply large additive energy, we have the following proposition.

Proposition 4.1. *Let A and B be finite subsets of an abelian group. If $|A+B| \leq K|A|^{1/2}|B|^{1/2}$ for some constant K, then we have*

$$E(A, B) \geq \frac{1}{K}|A|^{3/2}|B|^{3/2}. \quad ∎$$

Somewhat surprisingly, if we convert these statements into their entropic analogues in the naïve way, as we've been doing, the implication goes the other way! However, we have a weak equivalence (with worse constants in one direction) under the further assumption that the random variables in question are not too dependent.

Proposition 4.2. *Let X and Y be discrete random variables taking values in the same abelian group. If*

$$\mathbf{e}\{X,Y\} \ge \frac{3}{2}\,\mathbf{H}\{X\} + \frac{3}{2}\,\mathbf{H}\{Y\} - \log K, \tag{4}$$

for some constant K, then

$$\mathbf{H}\{X+Y\} \le \frac{1}{2}\,\mathbf{H}\{X\} + \frac{1}{2}\,\mathbf{H}\{Y\} + \log K. \tag{5}$$

If one adds the further assumption that $\mathbf{H}\{X,Y\} \ge \mathbf{H}\{X\} + \mathbf{H}\{Y\} - C$, then (5) implies (4) with a worse constant, namely, we may only conclude

$$\mathbf{e}\{X,Y\} \ge \frac{3}{2}\,\mathbf{H}\{X\} + \frac{3}{2}\,\mathbf{H}\{Y\} - \log K - 2C. \tag{6}$$

In particular, if X and Y are independent, then we can recover (4) from (5).

Proof. Assuming the lower bound on the additive energy, we have

$$2\,\mathbf{H}\{X,Y\} - \mathbf{H}\{X+Y\} \ge \frac{3}{2}\,\mathbf{H}\{X\} + \frac{3}{2}\,\mathbf{H}\{Y\} - \log K,$$

so

$$\begin{aligned}
\mathbf{H}\{X+Y\} &\le 2\,\mathbf{H}\{X,Y\} - \frac{3}{2}\,\mathbf{H}\{X\} - \frac{3}{2}\,\mathbf{H}\{Y\} + \log K \\
&\le 2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} - \frac{3}{2}\,\mathbf{H}\{X\} - \frac{3}{2}\,\mathbf{H}\{Y\} + \log K \\
&= \frac{1}{2}\,\mathbf{H}\{X\} + \frac{1}{2}\,\mathbf{H}\{Y\} + \log K.
\end{aligned}$$

On the other hand, assuming this upper bound on $\mathbf{H}\{X+Y\}$, we have

$$\begin{aligned}
\mathbf{e}\{X,Y\} &= 2\,\mathbf{H}\{X,Y\} - \mathbf{H}\{X+Y\} \\
&\ge 2\,\mathbf{H}\{X,Y\} - \frac{1}{2}\,\mathbf{H}\{X\} - \frac{1}{2}\,\mathbf{H}\{Y\} - \log K,
\end{aligned}$$

and if $2\,\mathbf{H}\{X,Y\} \ge 2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} - 2C$, then (6) follows directly. ∎

The fact that the implication goes the "wrong" way may seem somewhat baffling at first. We will now get to the bottom of this. If A and B are subsets of a finite abelian group, we can let X and Y be the uniform distributions on A and B, respectively. We have been operating under the belief that $\mathbf{H}\{X+Y\}$ should correspond (up to taking powers or logarithms) to the size of $A+B$. But this is not true, since X and Y may be given a joint distribution that is not uniform on $A \times B$, even if its marginals are uniform on A and B.

For example, let A and B be subsets of G and consider any regular bipartite graph H on the vertex set $A \cup B$. Let (X,Y) be defined by sampling an edge from H uniformly at random, letting X be its endpoint in A and Y be its endpoint in B. Since the graph is regular, X is uniform on A and Y is uniform on B, but $X+Y$ can only take values $a+b$ where (a,b) is an edge of H. In other words, $X+Y$ samples from the *partial sumset*

$$A +_H B = \big\{a+b : (a,b) \in E(H)\big\},$$

where elements that are represented more times as the sum of edge endpoints are given a greater weight. The way to properly recover the ordinary sumset $A+B$ is to let H be all of $A \times B$, in which case X and Y are independent. The extra assumption we added in Proposition 4.2 is analogous to stipulating that $|H| \geq K|A| \cdot |B|$, so that the resulting X and Y are "nearly" independent.

Simply put, in the entropic setting the bound (4) is stronger than (5) because the entropy $\mathbf{H}\{X,Y\}$ of the joint distribution appears in the formula for $\mathbf{e}\{X,Y\}$, whereas (5) says nothing whatsoever about this joint distribution.

The converse to Proposition 4.1 does not hold in general; that is, large additive energy does not necessarily imply a small sumset. However, there does exist a partial converse, which says that if sets A and B have a large additive energy, then there are dense subsets $A' \subseteq A$ and $B' \subseteq B$ such that the sumset $|A+B|$ is small. This is the celebrated Balog–Szemerédi–Gowers theorem.

Theorem 4.3 (*Balog–Szemerédi–Gowers theorem*). *Let A be a finite subset of an abelian group with $E(A,B) \geq c|A|^{3/2}|B|^{3/2}$. Then there are subsets $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geq c'|A|$ and $|B'| \geq c''|B|$ such that*

$$|A' + B'| \leq C|A|^{1/2}|B|^{1/2},$$

where c', c'', and C depend only on c.

In the entropy setting, the operation on random variables that corresponds to taking subsets is to conditioning. (As a sanity check, recall that conditioning never increases entropy, just as taking subsets never increases cardinality.) The Balog–Szemerédi–Gowers theorem gives us subsets between which we can take a *bona fide* sumset, so its entropic analogue should return conditionings X' and Y' of X and Y relative to some random variable Z, such that

i) X' and Y' are conditionally independent relative to Z;
ii) the entropies $\mathbf{H}\{X' \mid Z\}$ and $\mathbf{H}\{Y' \mid Z\}$ are not too small compared to their unconditioned analogues; and
iii) $\mathbf{H}\{X' + Y' \mid Z\}$ is small.

In fact, the conditioning we shall perform is exactly the one used to define additive energy.

First, we need a lemma, which we state separately since it will also be used later in the proof of the polynomial Freiman–Ruzsa theorem.

Lemma 4.4. *Let X and Y be discrete random variables taking values in the same abelian group. Let (X_1, Y_1) and (X_2, Y_2) be conditionally independent trials of (X, Y) relative to $X + Y$. Then we have*

$$\max\bigl(\mathbf{H}\{X_1 - X_2\}, \mathbf{H}\{X_1 - Y_2\}\bigr) \le \mathbf{H}\{X + Y\} - \mathbf{I}\{X : Y\}.$$

The right-hand side of this expression can also be written $2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} - \mathbf{e}\{X, Y\}$.

Proof. First we perform the proof for $X_1 - Y_2$. Submodularity gives us

$$\mathbf{H}\{X_1, Y_1, X_1 - Y_2\} + \mathbf{H}\{X_1 - Y_2\} \le \mathbf{H}\{X_1, X_1 - Y_2\} + \mathbf{H}\{Y_1, X_1 - Y_2\}.$$

Since $X_1 + Y_1 = X + Y = X_2 + Y_2$, given $(X_1, Y_1, X_1 - Y_2)$ we can recover the values of X_2 and Y_2. So $(X_1, Y_1, X_1 - Y_2)$ and (X_1, Y_1, X_2, Y_2) determine each other and hence

$$\mathbf{H}\{X_1, Y_1, X_1 - Y_2\} = \mathbf{H}\{X_1, Y_1, X_2, Y_2\} = 2\,\mathbf{H}\{X, Y\} - \mathbf{H}\{X + Y\}.$$

On the other side of the inequality, we have

$$\mathbf{H}\{X_1, X_1 - Y_2\} = \mathbf{H}\{X_1, Y_2\} \le \mathbf{H}\{X\} + \mathbf{H}\{Y\},$$

and similarly

$$\mathbf{H}\{Y_1, X_1 - Y_2\} = \mathbf{H}\{Y_1, X_2 - Y_1\} = \mathbf{H}\{X_2, Y_1\} \le \mathbf{H}\{X\} + \mathbf{H}\{Y\}.$$

Therefore,

$$\begin{aligned}\mathbf{H}\{X_1 - Y_2\} &\le \mathbf{H}\{X + Y\} - 2\,\mathbf{H}\{X, Y\} + 2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\}\\ &= \mathbf{H}\{X + Y\} - \mathbf{I}\{X : Y\}.\end{aligned}$$

The same holds with the roles of X_2 and Y_2 exchanged. ∎

Theorem 4.5 (*Entropic Balog–Szemerédi–Gowers theorem*). *Let X and Y be discrete random variables taking values in the same abelian group, and suppose that*

$$\mathbf{e}\{X, Y\} \ge \frac{3}{2}\mathbf{H}\{X\} + \frac{3}{2}\mathbf{H}\{Y\} + \log K$$

for some constant K. Then letting (X_1, Y_1) and (X_2, Y_2) be conditionally independent trials of (X, Y) relative to $X + Y$, we have

$$\mathbf{H}\{X_1 \mid X + Y\} \ge \mathbf{H}\{X\} - 2\log K$$

and

$$\mathbf{H}\{Y_2 \mid X + Y\} \ge \mathbf{H}\{Y\} - 2\log K.$$

Furthermore, the variables X_1 and Y_2 are conditionally independent relative to $X+Y$, and we have

$$\mathbf{H}\{X_1+Y_2 \mid X+Y\} \leq \frac{1}{2}\,\mathbf{H}\{X\}+\frac{1}{2}\,\mathbf{H}\{Y\}+\log K.$$

Proof. Using the coupling $X+Y = X_1+Y_1 = X_2+Y_2$, we have

$$\begin{aligned}
\mathbf{H}\{X_1 \mid X+Y\} &= \mathbf{H}\{X_1, X_1+Y_1\} - \mathbf{H}\{X+Y\} \\
&= \mathbf{H}\{X,Y\} - \mathbf{H}\{X+Y\} \\
&= \mathbf{e}\{X,Y\} - \mathbf{H}\{X,Y\} \\
&\geq \frac{3}{2}\,\mathbf{H}\{X\}+\frac{3}{2}\,\mathbf{H}\{Y\} - \log K - \mathbf{H}\{X,Y\} \\
&\geq \frac{1}{2}\,\mathbf{H}\{X\}+\frac{1}{2}\,\mathbf{H}\{Y\} - \log K.
\end{aligned}$$

where in the fourth line we used the hypothesis on $\mathbf{e}\{X,Y\}$, and in the last line we observed that $\mathbf{H}\{X\}+\mathbf{H}\{Y\}-\mathbf{H}\{X,Y\} \geq 0$. The bound

$$\mathbf{H}\{Y_2, X+Y\} \geq \frac{1}{2}\,\mathbf{H}\{X\}+\frac{1}{2}\,\mathbf{H}\{Y\} - \log K$$

is shown in the exact same way; only the first step differs.

Now, taking the sum of both these bounds, we arrive at

$$\mathbf{H}\{X_1 \mid X+Y\} + \mathbf{H}\{Y_2 \mid X+Y\} \geq \mathbf{H}\{X\}+\mathbf{H}\{Y\} - 2\log K.$$

From this one deduces

$$\begin{aligned}
\mathbf{H}\{X_1 \mid X+Y\} &\geq \mathbf{H}\{X\}+\mathbf{H}\{Y_2\} - \mathbf{H}\{Y_2 \mid X+Y\} - 2\log K \\
&\geq \mathbf{H}\{X\} - 2\log K.
\end{aligned}$$

The corresponding lower bound on $\mathbf{H}\{Y_2 \mid X+Y\}$ is proved similarly.

It remains to prove the upper bound on $\mathbf{H}\{X_1+Y_2 \mid X+Y\}$. Note that $(X_1, Y_2, X+Y)$ and $(X_1-X_2, X+Y)$ jointly determine $(X_1, X_2, X+Y)$. Then given X_1-X_2 and $X+Y$ we can calculate

$$X_1+Y_2 = X_1-X_2+X_2+Y_2 = (X_1-X_2)+(X+Y),$$

so $(X_1, Y_2, X+Y)$ and $(X_1-X_2, X+Y)$ each separately determine $(X_1+Y_2, X+Y)$. Hence the submodularity inequality yields

$$\mathbf{H}\{X_1, X_2, X+Y\}+\mathbf{H}\{X_1+Y_2, X+Y\} \leq \mathbf{H}\{X_1, Y_2, X+Y\}+\mathbf{H}\{X_1-X_2, X+Y\}.$$

From $(X_1, X_2, X+Y)$ we can calculate $Y_1 = X+Y-X_1$ and $Y_2 = X+Y-X_2$, so this triple and the triple (X_1, X_2, Y_1, Y_2) determine each other. So the first term above is simply the additive energy between X and Y; that is

$$\mathbf{H}\{X_1, X_2, X+Y\} = \mathbf{H}\{X_1, X_2, Y_1, Y_2\} = 2\,\mathbf{H}\{X,Y\} - \mathbf{H}\{X+Y\}.$$

Now since X_1 and Y_2 are conditionally independent relative to $X+Y$, we have

$$\begin{aligned}\mathbf{H}\{X_1, Y_2, X+Y\} &= \mathbf{H}\{X_1, X+Y\} + \mathbf{H}\{Y_2, X+Y\} - \mathbf{H}\{X+Y\} \\ &= \mathbf{H}\{X, X+Y\} + \mathbf{H}\{Y, X+Y\} - \mathbf{H}\{X+Y\} \\ &= 2\,\mathbf{H}\{X,Y\} - \mathbf{H}\{X+Y\}\end{aligned}$$

For the last term above we split

$$\mathbf{H}\{X_1 - X_2, X+Y\} = \mathbf{H}\{X_1 - X_2 \mid X+Y\} - \mathbf{H}\{X+Y\}.$$

Putting everything together, we obtain

$$\begin{aligned}2\,\mathbf{H}\{X,Y\} - \mathbf{H}\{X{+}Y\} + \mathbf{H}\{X_1+Y_2, X+Y\}& \\ \le 2\,\mathbf{H}\{X,Y\} + \mathbf{H}\{X_1 - X_2 \mid X+Y\} - 2\,\mathbf{H}\{X+Y\},&\end{aligned}$$

so that

$$\mathbf{H}\{X_1 + Y_2 \mid X+Y\} \le \mathbf{H}\{X_1 - X_2 \mid X+Y\} - 2\,\mathbf{H}\{X+Y\} \le \mathbf{H}\{X_1 - X_2\}.$$

The previous lemma then gives

$$\mathbf{H}\{X_1 + Y_2 \mid X+Y\} \le 2\,\mathbf{H}\{X\} + 2\,\mathbf{H}\{Y\} - \mathbf{e}\{X,Y\},$$

and from our lower bound on $\mathbf{e}\{X,Y\}$, we conclude that

$$\mathbf{H}\{X_1 + Y_2 \mid X+Y\} \le \frac{1}{2}\,\mathbf{H}\{X\} + \frac{1}{2}\,\mathbf{H}\{Y\} + \log K,$$

which is what we wanted to show. ▮

5. The Freiman–Ruzsa theorem

Both Plünnecke's theorem and the Balog–Szemerédi–Gowers theorem have to do with the ratio $|A+A|/|A|$ of a finite set A. Let us now give this ratio a name. It is called the *doubling constant* of A. Plünnecke's theorem says that if the doubling constant is at most K, then the ratio $|rA - sA|/|A|$ is at most K^{r+s}. The Balog–Szemerédi–Gowers theorem, on the other hand, says that sets A with large additive energy contain large subsets A and A'' such that $|A' + A''|/|A|$ is bounded from above by some constant.

It is natural to ask whether there is some way to characterise the structure of sets A with small doubling constant (in some asymptotic sense). If the doubling constant is 1, then the structure of A is completely determined, as shown by the following proposition.

Proposition 5.1. *Let A be a finite subset of an abelian group G. If $|A+A| = |A|$, then A is a coset of a subgroup H.*

Proof. First we assume that A contains 0. Then

$$A = A + \{0\} \subseteq A + A,$$

but since $|A + A| = |A|$ this implies that $A + A = A$. Now let $H = \{h \in G : A + h = A$. The above observation shows that $A \subseteq H$. But if $h \in H$, then $A = A + h$ contains the element $0 + h = h$, so in fact $H = A$.

If $x, y \in H$, then $A+x = A$ and $A+y = A$, so adding y to both sides of the second identity we obtain $A - y = A$, and adding x to both sides of this, we get

$$A + x - y = A + x = A,$$

so $x - y \in H$. This along with the fact that $0 \in H$ shows that H is a subgroup of G.

Thus in the case where A contains 0, we see that A is actually a subgroup of G, and in the general case we may translate A without changing $|A + A|$, so A is the translate of a subgroup; that is, A is a coset of H. ∎

In the case that A is a union of not too many cosets, we also expect $A + A$ to be quite small, since each individual coset does not grow under addition (the only possible growth comes from additions between cosets).

It is believed that the converse holds; that is, if A has doubling constant bounded above by K, then A is contained in a union of cosets, where the number of cosets needed can be taken to be no more than polynomial in the doubling constant. This conjecture is due to K. Marton.

Conjecture 5.2. *Let $A \subseteq \left(\mathbf{Z}/r\mathbf{Z}\right)^n$ have $|A+A| \leq K|A|$ for some constant K. Then there is a subgroup H with $|H| \leq |A|$ such that A is contained in a union of K^C cosets of H, where C is a constant that can depend on r but not on n or K.*

In 2023, the first special case of this conjecture was proved by W. T. Gowers, B. Green, F. Manners, and T. Tao.

Theorem 5.3 (*Gowers–Green–Manners–Tao,* 2023). *There is a constant C such that the following holds. Let $A \subseteq \mathbf{F}_2^n$ have $|A+A| \leq K|A|$ for some constant K. Then there is a subgroup H with $|H| \leq |A|$ such that A is contained in a union of $2K^C$ cosets of H.*

The proof, which is almost entirely information-theoretic in nature, will be our main concern for the remainder of these notes.

References