Contents

1. The Khinchin axioms for entropy 2
   1.1. Entropy axioms 2
   1.2. Properties of entropy 2
2. A special case of Sidorenko's conjecture 7
3. Bregman's theorem 9
4. Shearer's lemma and applications 12
5. The union-closed conjecture 18
6. Entropy in additive combinatorics 23
7. A proof of Marton's conjecture in $\mathbb{F}_2^n$ 28

1. The Khinchin axioms for entropy

Note all random variables we deal with will be discrete, unless otherwise stated. We use $\log = \log_2$.

1.1. Entropy axioms

Definition 1.1 The **entropy** of a discrete random variable X is a quantity $H(X)$ that takes real values and satisfies the **Khinchin axioms**: Normalisation, Invariance, Extendability, Maximality, Continuity and Additivity.

Axiom 1.2 (Normalisation) If X is uniform on $\{0,1\}$ (i.e. $X \sim \text{Bern}(1/2)$), then $H(X) = 1$.

Axiom 1.3 (Invariance) If X takes values in A, Y takes values in B, $f : A \to B$ is a bijection and $\mathbb{P}(X = a) = \mathbb{P}(Y = f(a))$ for all $a \in A$, then $H(Y) = H(X)$ (i.e. the entropy of X depends only on its distribution).

Axiom 1.4 (Extendability) If X takes values on a set A, B is disjoint from A, Y takes values in $A \sqcup B$, and for all $a \in A$, $\mathbb{P}(Y = a) = \mathbb{P}(X = a)$, then $H(Y) = H(X)$.

Axiom 1.5 (Maximality) If X takes values in a finite set A and Y is uniformly distributed in A, then $H(X) \leq H(Y)$.

Definition 1.6 The **total variance distance** between X and Y is

$$\sup_E |\mathbb{P}(X \in E) - \mathbb{P}(Y \in E)|.$$

Axiom 1.7 (Continuity) H depends continuously on X (with respect to total variation distance).

Definition 1.8 Let X and Y be random variables. The **conditional entropy** of X given Y is

$$H(X \mid Y) := \sum_y \mathbb{P}(Y = y) H(X \mid Y = y).$$

Axiom 1.9 (Additivity) $H(X, Y) := H((X, Y)) = H(Y) + H(X \mid Y)$.

1.2. Properties of entropy

Lemma 1.10 If X and Y are independent, then $H(X, Y) = H(X) + H(Y)$.

Proof (Hints). Straightforward. □

Proof. $H(X \mid Y) = \sum_y \mathbb{P}(Y = y) H(X \mid Y = y)$ Since X and Y are independent, the distribution of X is unaffected by knowing Y, so $H(X \mid Y = y) = H(X)$ for all y, which gives the result. (Note we have implicitly used Invariance here). □

Corollary 1.11 If $X_1, ..., X_n$ are independent, then

$$H(X_1, ..., X_n) = H(X_1) + \cdots + H(X_n).$$

Proof (Hints). Straightforward. □

Proof. By Lemma 1.10 and induction. □

Lemma 1.12 (Chain Rule) Let $X_1, ..., X_n$ be RVs. Then

$$H(X_1, ..., X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \cdots + H(X_n \mid X_1, ..., X_{n-1}).$$

Proof (Hints). Straightforward. □

Proof. The case $n = 2$ is Additivity. In general,

$$H(X_1, ..., X_n) = H(X_1, ..., X_{n-1}) + H(X_n \mid X_1, ..., X_{n-1}),$$

so the result follows by induction. □

Lemma 1.13 Let X and Y be RVs. If $Y = f(X)$, then $H(X, Y) = H(X)$. Also, $H(Z \mid X, Y) = H(Z \mid X)$.

Proof (Hints). Consider an appropriate bijection. □

Proof. The map $g : x \mapsto (x, f(x))$ is a bijection, and $(X, Y) = g(X)$, so the first statement follows from Invariance. Also,

$$\begin{aligned} H(Z \mid X, Y) &= H(Z, X, Y) - H(X, Y) \quad \text{by additivity} \\ &= H(Z, X) - H(X) \quad \text{by first part} \\ &= H(Z \mid X) \quad \text{by additivity} \end{aligned}$$

□

Lemma 1.14 If X takes only one value, then $H(X) = 0$.

Proof (Hints). Consider an independent copy of X. □

Proof. Let X' be an independent copy of X. (X, X') takes only one value, so $H(X, X') = H(X)$ by Invariance. But by independence, $H(X, X') = H(X) + H(X') = 2H(X)$, so $H(X) = 0$. □

Proposition 1.15 If X is uniformly distributed on a set of size 2^n, then $H(X) = n$.

Proof (Hints). Straightforward. □

Proof. Let $X_1, ..., X_n$ be independent RVs, uniformly distributed on $\{0, 1\}$. By Corollary 1.11 and Normalisation, $H(X_1, ..., X_n) = n$. So the result follows by Invariance. □

Proposition 1.16 If X is uniformly distributed on a set A of size n, then $H(X) = \log n$.

Proof (Hints). Straightforward. □

Proof. Let $r \in \mathbb{N}$ and let $X_1, ..., X_r$ be independent copies of X. Then $(X_1, ..., X_r)$ is uniform on A^r, and $H(X_1, ..., X_r) = rH(X)$. Now pick k such that $2^k \leq n^r \leq 2^{k+1}$. Then by Proposition 1.15, Invariance and Maximality, $k \leq rH(X) \leq k + 1$. So $\frac{k}{r} \leq \log n \leq \frac{k+1}{r}$ and $\frac{k}{r} \leq H(X) \leq \frac{k+1}{r}$ for all $r \in \mathbb{N}$. So $H(X) = \log n$, as claimed. □

Theorem 1.17 (Khinchin) If H satisfies the Khinchin axioms and X takes values in a finite set A, then

$$H(X) = \sum_{a \in A} p_a \log(1/p_a) = \mathbb{E}\left[\log \frac{1}{P_X(X)}\right],$$

where $p_a = \mathbb{P}(X = a)$.

Proof (Hints).

- Explain why it is enough to prove for when the p_a are rational.
- Pick $n \in \mathbb{N}$ such that $p_a = \frac{m_a}{n}$, $m_a \in \mathbb{N}_0$. Let Z be uniform on $[n]$. Let $\{E_a : a \in A\}$ be a partition of $[n]$ such that $X = a \Leftrightarrow Z \in E_a$.
- Consider $H(Z \mid X)$.

□

Proof. First we do the case where all $p_a \in \mathbb{Q}$. Pick $n \in \mathbb{N}$ such that $p_a = \frac{m_a}{n}$, $m_a \in \mathbb{N}_0$. Let Z be uniform on $[n]$. Let $\{E_a : a \in A\}$ be a partition of $[n]$ into sets with $|E_a| = m_a$. By Invariance, we may assume that $X = a \Leftrightarrow Z \in E_a$. Then

$$\begin{aligned}
\log n = H(Z) &= H(Z, X) = H(X) + H(Z \mid X) \\
&= H(X) + \sum_{a \in A} p_a H(Z \mid X = a) \\
&= H(X) + \sum_{a \in A} p_a \log m_a \\
&= H(X) + \sum_{a \in A} p_a (\log p_a + \log n) \\
&= H(X) + \sum_{a \in A} p_a \log p_a + \log n.
\end{aligned}$$

Hence $H(X) = -\sum_{a \in A} p_a \log p_a$.

The general result follows by Continuity. □

Corollary 1.18 Let X and Y be random variables. Then $0 \leq H(X)$ and $0 \leq H(X \mid Y)$.

Proof (Hints). Trivial. □

Proof. Immediate consequence of Khinchin. □

Corollary 1.19 If $Y = f(X)$, then $H(Y) \leq H(X)$.

Proof (Hints). Straightforward. □

Proof. $H(X) = H(X, Y) = H(Y) + H(X \mid Y)$. But $H(X \mid Y) \geq 0$. □

Proposition 1.20 (Subadditivity) Let X and Y be RVs. Then $H(X, Y) \leq H(X) + H(Y)$.

Proof (Hints).

- Let $p_{ab} = \mathbb{P}(X = a, Y = b)$. Explain why it is enough to show for the case when the p_{ab} are rational.
- Pick n such that $p_{ab} = m_{ab}/n$ with each $m_{ab} \in \mathbb{N}_0$. Partition $[n]$ into sets E_{ab} of size m_{ab}. Let Z be uniform on $[n]$.

- Show that if X (or Y) is uniform, then $H(X \mid Y) \leq H(X)$ and $H(X,Y) \leq H(X) + H(Y)$.
- Let $E_b = \cup_a E_{ab}$ for each b. We can assume $Y = b$ iff $Z \in E_b$. Now define an RV W as follows: if $Y = b$, then W is uniformly distributed in E_b. Use conditional independence of X and W given Y to conclude the result.

□

Proof. Note that for any two RVs X, Y,

$$\begin{aligned} & H(X,Y) \leq H(X) + H(Y) \\ \iff & H(X \mid Y) \leq H(X) \\ \iff & H(Y \mid X) \leq H(Y) \end{aligned}$$

by Additivity. Next, observe that $H(X \mid Y) \leq H(X)$ if X is uniform on a finite set, since $H(X \mid Y) = \sum_y \mathbb{P}(Y = y) H(X \mid Y = y) \leq \sum_y \mathbb{P}(Y = y) H(X) = H(X)$ by Maximality. By the above equivalence, we also have $H(X \mid Y) \leq H(X)$ if Y is uniform on a finite set. Now let $p_{ab} = \mathbb{P}(X = a, Y = b)$, and assume that all p_{ab} are rational. Pick n such that $p_{ab} = m_{ab}/n$ with each $m_{ab} \in \mathbb{N}_0$. Partition $[n]$ into sets E_{ab} of size m_{ab}. Let Z be uniform on $[n]$. WLOG (by Invariance), $(X,Y) = (a,b)$ iff $Z \in E_{ab}$.

Let $E_b = \cup_a E_{ab}$ for each b. So $Y = b$ iff $Z = E_b$. Now define an RV W as follows: if $Y = b$, then $W \in E_b$, but then W is uniformly distributed in E_b and independent of X (and Z). So W and X are conditionally independent given Y, and W is uniform on $[n]$. Then $H(X \mid Y) = H(X \mid Y, W) = H(X \mid W)$ by conditional independence and by Lemma 1.13 (since W determines Y). Since W is uniform, $H(X \mid W) \leq H(X)$.

The general result follows by Continuity. □

Corollary 1.21 $H(X) \geq 0$ for any X.

Proof (Hints). (Without using the formula) straightforward. □

Proof. (Without using the formula). By subadditivity, $H(X \mid X) \leq H(X)$. But $H(X \mid X) = 0$. □

Corollary 1.22 Let $X_1, ..., X_n$ be RVs. Then

$$H(X_1, ..., X_n) \leq H(X_1) + \cdots + H(X_n).$$

Proof (Hints). Trivial. □

Proof. Trivial by induction. □

Proposition 1.23 (Submodularity) Let X, Y, Z be RVs. Then

$$H(X \mid Y, Z) \leq H(X \mid Z).$$

Proof (Hints). Use that $H(X \mid Y, Z = z) \leq H(X \mid Z = z)$ (why?). □

Proof. $H(X \mid Y, Z) = \sum_z \mathbb{P}(Z = z) H(X \mid Y, Z = z) \leq \sum_z \mathbb{P}(Z = z) H(X \mid Z = z) = H(X \mid Z)$ by Subadditivity. □

Remark 1.24 Submodularity can be expressed in several equivalent ways. Expanding using Additivity gives

$$H(X,Y,Z) - H(Y,Z) \leq H(X,Z) - H(Z)$$

and

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

and

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

Lemma 1.25 Let X, Y, Z be RVs with $Z = f(Y)$. Then $H(X \mid Y) \leq H(X \mid Z)$.

Proof (Hints). Straightforward. □

Proof. We have

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

by Submodularity. □

Lemma 1.26 Let X, Y, Z be RVs with $Z = f(X) = g(Y)$. Then

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

Proof (Hints). Straightforward. □

Proof. By Submodularity, we have $H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z)$, which implies the result, since Z depends on X and Y. □

Lemma 1.27 Let X be an RV taking values in a finite set A and let Y be uniform on A. If $H(X) = H(Y)$, then X is uniform.

Proof (Hints). Use Jensen's inequality. □

Proof. Let $p_a = \mathbb{P}(X = a)$. Then

$$H(X) = \sum_{a \in A} p_a \log(1/p_a) = |A| \cdot \mathbb{E}_{a \in A} p_a \log\left(\frac{1}{p_a}\right).$$

The function $x \mapsto x \log(1/x)$ is concave on $[0,1]$. So by Jensen's inequality,

$$H(X) \leq |A| \cdot (\mathbb{E}_{a \in A} p_a) \cdot \log\left(\frac{1}{\mathbb{E}_{a \in A} p_a}\right) = \log|A| = H(Y),$$

with equality iff $a \mapsto p_a$ is constant, i.e. X is uniform. □

Corollary 1.28 If $H(X,Y) = H(X) + H(Y)$, then X and Y are independent.

Proof (Hints). Go through the proof of Subadditivity and check when equality holds. □

Proof. We go through the proof of subadditivity and check when equality holds. Suppose that X is uniform on A. Then

$$H(X \mid Y) = \sum_y \mathbb{P}(Y = y) H(X \mid Y = y) \leq H(X),$$

with equality iff $X \mid Y = y$ is uniform on A for all y (by Lemma 1.27), which implies that X and Y are independent.

At the last stage of the proof, we said $H(X \mid Y) = H(X \mid Y, W) = H(X \mid W) \leq H(X)$, where W was uniform, i.e. $H(W \mid X) \leq H(W)$. So equality holds only if X and W are independent, which implies (since Y depends on W), that X and Y are independent.□

Definition 1.29 Let X and Y be RVs. The **mutual information**

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

Remark 1.30 Subadditivity is equivalent to the statement that $I(X : Y) \geq 0$, and Corollary 1.28 implies that $I(X : Y) = 0$ iff X and Y are independent.

Note that $H(X, Y) = H(X) + H(Y) - I(X : Y)$ (note the similarity to the inclusion-exclusion formula for two sets).

Definition 1.31 Let X, Y, Z be RVs. The **conditional mutual information** of X and Y given Z is

$$\begin{aligned} I(X : Y \mid Z) &:= \sum_z \mathbb{P}(Z = z) I(X \mid Z = z : Y \mid Z = z) \\ &= \sum_z \mathbb{P}(Z = z)(H(X \mid Z = z) + H(Y \mid Z = z) - H(X, Y \mid Z = z)) \\ &= H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z) \\ &= H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z). \end{aligned}$$

Submodularity is equivalent to the statement that $I(X : Y \mid Z) \geq 0$.

2. A special case of Sidorenko's conjecture

Definition 2.1 Let G be a bipartite graph with (finite) vertex sets X and Y and density α (defined to be $\frac{|E(G)|}{|X| \cdot |Y|}$). Let H be another (think of it as small) bipartite graph with vertex sets U and V. Now let $\varphi : U \to X$ and $\psi : V \to Y$. We say that (φ, ψ) is a **homomorphism** if $\varphi(x)\psi(y) \in E(G)$ for every edge $xy \in E(H)$.

Conjecture 2.2 (Sidorenko's Conjecture) For every G, H, for random $\varphi : U \to X$, $\psi : V \to Y$,

$$\mathbb{P}((\varphi, \psi) \text{ is a homomorphism}) \geq \alpha^{|E(H)|}.$$

Remark 2.3 Sidorenko's Conjecture is not hard to prove when H is the complete bipartite graph $K_{r,s}$ (the case $K_{2,2}$ can be proved using Cauchy-Schwarz: exercise).

Theorem 2.4 Sidorenko's Conjecture is true if H is a path of length 3.

Proof (Hints).

- Let (X_1, Y_1) be a random edge of G (with $X_1 \in X$, $Y_1 \in Y$). Now let X_2 be a random neighbour of Y_1 and Y_2 be a random neighbour of X_2. Explain why it suffices to prove that $H(X_1, Y_1, X_2, Y_2) \geq \log(\alpha^3 m^2 n^2)$.
- Find an equivalent way of choosing a uniformly random edge (X_1, Y_1) of G (in terms of vertices). Use this to reaosn that X_2Y_1 and X_2Y_2 are uniformly random in $E(G)$.
- Find the lower bound for $H(X_1, Y_1, X_2, Y_2)$ using the Chain Rule and Maximality.

□

Proof. We want to show that if G is a bipartite graph of density α with vertex sets X, Y of size m and n, and we choose $x_1, x_2 \in X$, $y_1, y_2 \in Y$ independently at random, then $\mathbb{P}(x_1y_1, y_1x_2, x_2y_2 \in E(G)) \geq \alpha^3$.

It would be enough to let P be a path of length 3 chosen uniformly at random and show that $H(P) \geq \log(\alpha^3 m^2 n^2)$ (by Proposition 1.16). Instead, we shall define a different RV taking values in the set of all paths of length 3 (including degenerate paths). To do this, let (X_1, Y_1) be a random edge of G (with $X_1 \in X$, $Y_1 \in Y$). Now let X_2 be a random neighbour of Y_1 and Y_2 be a random neighbour of X_2. It will be enough to prove that

$$H(X_1, Y_1, X_2, Y_2) \geq \log(\alpha^3 m^2 n^2).$$

We can choose X_1, Y_1 in three equivalent ways:

1. Pick an edge uniformly from all edges
2. Pick a vertex x with probability proportional to its degree $\deg(x)$, and then a random neighbour Y of x.
3. Same as above with x and y exchanged.

By the equivalence, it follows that $Y_1 = y$ with probability $\deg(y)/|E(G)|$, so X_2Y_1 is uniform in $E(G)$, so $X_2 = x'$ with probability $d(x')/|E(G)|$, so X_2Y_2 is uniform in $E(G)$.

Let U_A be the uniform distribution on A. Therefore, by the Chain Rule,

$$\begin{aligned}
H(X_1, Y_1, X_2, Y_2) &= H(X_1) + H(Y_1 \mid X_1) + H(X_2 \mid X_1, Y_1) + H(Y_2 \mid X_1, Y_1, X_2) \\
&= H(X_1) + H(Y_1 \mid X_1) + H(X_2 \mid Y_1) + H(Y_2 \mid X_2) \\
&= H(X_1) + H(X_1, Y_1) - H(X_1) + H(X_2, Y_1) - H(Y_1) + H(X_2, Y_2) - H(Y_2) \\
&= 3H\left(U_{E(G)}\right) - H(Y_1) - H(X_2) \\
&\geq 3H\left(U_{E(G)}\right) - H(U_Y) - H(U_X) \\
&= 3\log(\alpha mn) - \log n - \log m \\
&= \log(\alpha^3 m^2 n^2).
\end{aligned}$$

So we are done, by Maximality (since $H(P) \geq H(X_1, Y_1, X_2, Y_2)$). Alternative finish to the proof: let X', Y' be uniform in X, Y and independent of each other and X_1, Y_1, X_2, Y_2. Then by the above inequality and Corollary 1.11,

$$H(X_1, Y_1, X_2, Y_2, X', Y') = H(X_1, Y_1, X_2, Y_2) + H(U_X) + H(U_Y)$$
$$\geq 3H\big(U_{E(G)}\big).$$

So by Maximality, the number of paths of length 3 times $|X|$ times $|Y|$ is $\geq |E(G)|^3$.□

3. Bregman's theorem

Definition 3.1 Let A be an $n \times n$ matrix over $\mathbb{R}$. The **permanent** of A is

$$\mathrm{per}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i\sigma(i)},$$

i.e. "the determinant without the signs".

Proposition 3.2 Let G be a bipartite graph with vertex sets X, Y of size n. Given $(x, y) \in X \times Y$, let

$$A_{xy} = \begin{cases} 1 & \text{if } xy \in E(G) \\ 0 & \text{if } xy \notin E(G) \end{cases},$$

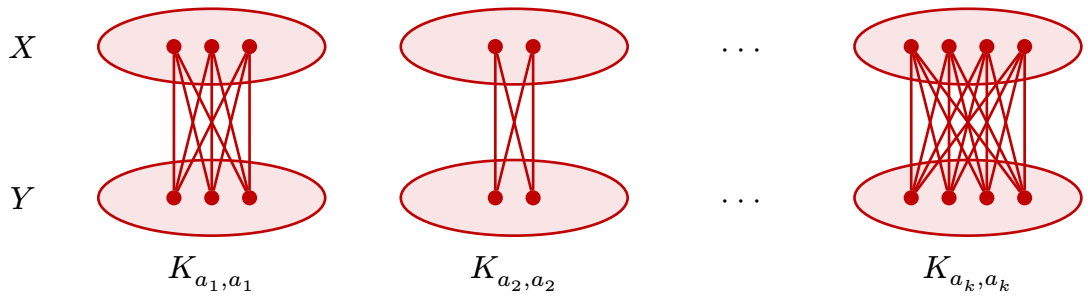
i.e. A is the bipartite adjacency matrix of G. Then $\mathrm{per}(A)$ is the number of perfect matchings in G. (Note that $\mathrm{per}(A)$ is well-defined as it is invariant under reordering of the vertices.)

Proof (Hints). Straightforward. □

Proof. Each (perfect) matching corresponds to a bijection $\sigma : X \to Y$ such that $x\sigma(x) \in E(G)$ for all $x \in X$. $\sigma \in S_n$ contributes 1 to the sum iff $x\sigma(x)$ is an edge of G for all $x \in X$ (i.e. iff σ corresponds to a perfect matching), and 0 otherwise. □

Bregman's theorem concerns how large $\mathrm{per}(A)$ can be if A is a $0,1$ matrix and the sum of the entries in the i-th row is d_i (i.e. if the degree of $x_i \in X$ is d_i).

Example 3.3 Let G be a disjoint union of K_{a_i,a_i}'s, $i = 1, ..., k$, with $a_1 + \cdots + a_k = n$. Then the number of perfect matchings in G is $\prod_{i=1}^{k} a_i!$.



Theorem 3.4 (Bregman) Let G be a bipartite graph with vertex sets X, Y of size n. Then the number of perfect matchings in G is at most

$$\prod_{x \in X} (\deg(x)!)^{1/\deg(x)}.$$

Proof (Hints).

- For an enumeration $x_1, ..., x_n$ of X and random matching (a bijection) σ, show that $H(\sigma) \leq \log \deg(x_1) + \mathbb{E}_\sigma \log \deg^\sigma_{x_1}(x_2) + \cdots + \mathbb{E}_\sigma \log \deg^\sigma_{x_1,...,x_{n-1}}(x_n)$ (find a suitable expression for $\deg^\sigma_{x_1,...,x_{i-1}}(x_i)$).
- Explain why

$$\deg^\sigma_{x_1,...,x_{i-1}}(x_i) = d(x_i) - |\{j : \sigma^{-1}(y_j) \text{ comes earlier than } x_i \text{ in } x_1, ..., x_n\}.$$

- Show that the average of $\log \deg^\sigma_{x_1,...,x_{i-1}}(x_i)$ is $\frac{1}{d(x)}(\log(d(x)!))$.

□

Proof (by Radhakrishnan). Each (perfect) matching corresponds to a bijection $\sigma : X \to Y$ such that $x\sigma(x) \in E(G)$ for all $x \in X$. Let σ be chosen uniformly from all such bijections. Then by the Chain Rule,

$$\begin{aligned} H(\sigma) &= H(\sigma(x_1), ..., \sigma(x_n)) \\ &= H(\sigma(x_1)) + H(\sigma(x_2) \mid \sigma(x_1)) + \cdots + H(\sigma(x_n) \mid \sigma(x_1), ..., \sigma(x_{n-1})), \end{aligned}$$

where $x_1, ..., x_n$ is some enumeration of X. We have $H(\sigma(x_1)) \leq \log \deg(x_1)$ by Maximality, and

$$H(\sigma(x_2) \mid \sigma(x_1)) \leq \mathbb{E}_\sigma \log \deg^\sigma_{x_1}(x_2),$$

where $\deg^\sigma_{x_1}(x_2) = |N(x_2) \setminus \{\sigma(x_1)\}|$, by the definition of conditional entropy and Maximality. In general,

$$H(\sigma(x_i) \mid \sigma(x_1), ..., \sigma(x_{i-1})) \leq \mathbb{E}_\sigma \log \deg^\sigma_{x_1,...,x_{i-1}}(x_i),$$

where $\deg^\sigma_{x_1,...,x_{i-1}}(x_i) = |N(x_i) \setminus \{\sigma(x_1), ..., \sigma(x_{i-1})\}|$.

So $H(\sigma) \leq \mathbb{E}_\sigma \left[\log d(x_1) + \log d^\sigma_{x_1}(x_2) + ... + \log d^\sigma_{x_1,...,x_{n-1}}(x_n)\right]$ and so

$$\begin{aligned} H(\sigma) &\leq \mathbb{E}_{\text{orderings}} \mathbb{E}_\sigma \left[\log d(x_1) + \log d^\sigma_{x_1}(x_2) + ... + \log d^\sigma_{x_1,...,x_{n-1}}(x_n)\right] \\ &= \mathbb{E}_\sigma \mathbb{E}_{\text{orderings}} \left[\log d(x_1) + \log d^\sigma_{x_1}(x_2) + ... + \log d^\sigma_{x_1,...,x_{n-1}}(x_n)\right] \end{aligned}$$

Key idea: we now regard $x_1, ..., x_n$ as a *random* enumeration of X and take the average. For each $x \in X$, define the **contribution** of x to be $\log\left(d^\sigma_{x_1,...,x_{i-1}}(x_i)\right)$, where $x_i = x$. We shall now fix σ and $x \in X$. Let the neighbours of x be $y_1, ..., y_k$. Then one of the y_j will be $\sigma(x)$, say y_h. Then $d^\sigma_{x_1,...,x_{i-1}}(x_i)$ (given that $x_i = x$) is

$$d(x) - \left|\{j : \sigma^{-1}(y_j) \text{ comes earlier than } x = \sigma^{-1}(y_h)\}\right|.$$

All positions of $\sigma^{-1}(y_h)$ are equally likely, so the average contribution of x is

$$\begin{aligned} \mathbb{E}_{\text{orderings}} \log d^\sigma_{x_1,...,x_{i-1}}(x_i) &= \frac{1}{d(x)}(\log d(x) + \log(d(x) - 1) + \cdots + \log(1)) \\ &= \frac{1}{d(x)} \log d(x)!. \end{aligned}$$

By linearity of expectation,

$$H(\sigma) \leq \sum_{x \in X} \frac{1}{d(x)} \log(d(x)!)$$

So the number of matchings is at most $\prod_{x \in X} (d(x)!)^{1/d(x)}$. □

Definition 3.5 Let G be a graph with $2n$ vertices. A **1-factor** in G is a collection of n disjoint edges.

Theorem 3.6 (Kahn-Lovasz) Let G be a graph with $2n$ vertices. Then the number of 1-factors in G is at most

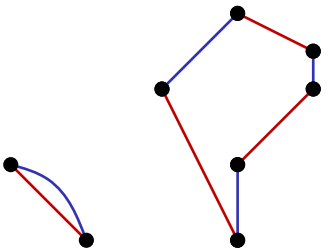
$$\prod_{x \in V(G)} (d(x)!)^{1/2d(x)}.$$

Proof (Hints).

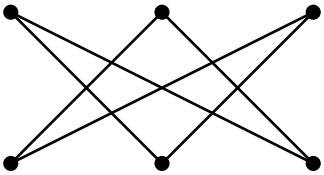
- Let M be the set of 1-factors of G and let (M_1, M_2) be a uniformly random element of $M \times M$.
- Given a cover of G by M_1 and M_2, find an expression for the number of pairs (M_1', M_2') that give rise to it, in terms of the number of even cycles.
- Let G_2 be the bipartite graph with two vertex sets V_1, V_2, which are both copies of $V(G)$. Join $x \in V_1$ to $y \in V_2$ iff $xy \in E(G)$.
- Explain why each perfect matching of G_2 gives a cover of $V(G)$ by isolated vertices, edges and cycles, and find an expression for the number of such perfect matchings that give rise to it.

□

Proof (by Alon, Friedman). Let M be the set of 1-factors of G and let (M_1, M_2) be a uniformly random element of $M \times M$. For each M_1, M_2, the union $M_1 \cup M_2$ is a collection of disjoint edges and even cycles that covers all the vertices of G.



Call such a union a **cover of G by edges and even cycles**. If we are given such a cover, then the number of pairs (M_1, M_2) that give rise to it is 2^k, where k is the number of even cycles. Now let's build a bipartite graph G_2 out of G. G_2 has two vertex sets V_1, V_2, which are both copies of $V(G)$. Join $x \in V_1$ to $y \in V_2$ iff $xy \in E(G)$.



G_2 if G is the triangle graph

By Bregman, the number of perfect matchings in G_2 is at most $\prod_{x\in V(G)} (d(x)!)^{1/d(x)}$. Each matching gives a permutation σ of $V(G)$ such that $x\sigma(x) \in E(G)$ for all $x \in V(G)$. Each such σ has a cycle decomposition, and each cycle gives a cycle in G. So σ gives a cover of $V(G)$ by isolated vertices, edges and cycles (not necessarily all even). Given such a cover with k cycles, each cycle can be directed in two ways, so the number of σ that give rise to it is $= 2^k$. So there is an injection from $M \times M$ to the set of matchings of G_2, since every cover by edges and and even cycles is a cover by vertices, edges and cycles. So $|M|^2 \leq \prod_{x\in V(G)} (d(x)!)^{1/d(x)}$. □

4. Shearer's lemma and applications

Notation 4.1 Given a random variable $X = (X_1, ..., X_n)$ and $A \subseteq [n]$, $A = \{a_1 < ... < a_k\}$, write X_A for the random variable $\left(X_{a_1}, ..., X_{a_k}\right)$.

Lemma 4.2 (Shearer) Let $X = (X_1, ..., X_n)$ be an RV and let $\mathcal{A}$ be a family of subsets of $[n]$ such that every $i \in [n]$ belongs to at least r of the sets $A \in \mathcal{A}$. Then

$$H(X_1, ..., X_n) \leq \frac{1}{r} \sum_{A\in\mathcal{A}} H(X_A).$$

Proof (Hints). For each $a \in [n]$, write $X_{<a}$ for $(X_1, ..., X_{a-1})$. Show that $H(X_A) \geq \sum_{a\in A} H(X_a \mid X_{<a})$. □

Proof. For each $a \in [n]$, write $X_{<a}$ for $(X_1, ..., X_{a-1})$. For each $A \in \mathcal{A}$, $A = \{a_1 < \cdots < a_k\}$, by the Chain Rule and Submodularity,

$$\begin{aligned} H(X_A) &= H\left(X_{a_1}\right) + H\left(X_{a_2} \mid X_{a_1}\right) + \cdots + H\left(X_{a_k} \mid X_{a_1}, ..., X_{a_{k-1}}\right) \\ &\geq H\left(X_{a_1} \mid X_{<a_1}\right) + H\left(X_{a_2} \mid X_{<a_2}\right) + \cdots + H\left(X_{a_k} \mid X_{<a_k}\right) \\ &= \sum_{a\in A} H(X_a \mid X_{<a}). \end{aligned}$$

Therefore, $\sum_{A\in\mathcal{A}} H(X_A) \geq r \sum_{a=1}^{n} H(X_a \mid X_{<a}) = rH(X)$. □

Example 4.3 $H(X_1, X_2, X_3) \leq \frac{1}{2}(H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3))$.

Lemma 4.4 Let $X = (X_1, ..., X_n)$ be an RV and let $A \subseteq [n]$ be a randomly chosen subset of $[n]$, according to some probability distribution. Suppose that for each $i \in [n]$, $\mathbb{P}(i \in A) \geq \mu$. Then

$$H(X) \leq \mu^{-1} \cdot \mathbb{E}_A[H(X_A)].$$

Proof (Hints). Very similar to proof of Shearer. □

Proof. As in Shearer,

$$H(X_A) \geq \sum_{a\in A} H(X_a \mid X_{<a}).$$

So

$$\begin{aligned}
\mathbb{E}_A[H(X_A)] &\geq \mathbb{E}_A\left[\sum_{a\in A} H(X_a \mid X_{<a})\right] \\
&= \mathbb{E}_A\left[\sum_{k=1}^{n} H(X_k \mid X_{<k})\mathbb{1}_{\{k\in A\}}\right] \\
&= \sum_{k=1}^{n} H(X_k \mid X_{<k})\mathbb{E}\left[\mathbb{1}_{\{k\in A\}}\right] \\
&= \sum_{k=1}^{n} H(X_k \mid X_{<k})\mathbb{P}(k \in A) \\
&\geq \mu \cdot H(X).
\end{aligned}$$

□

Definition 4.5 Let $E \subseteq \mathbb{Z}^n$ and let $A \subseteq [n]$. Then we write $P_A E$, if $A = \{a_1, ..., a_k\}$, for the set of $u \in \mathbb{Z}^A$ such that there exists $v \in \mathbb{Z}^{[n]\setminus A}$ such that $[u, v] \in E$, where $[u, v]$ is u suitably intertwined with v.

Corollary 4.6 Let $E \subseteq \mathbb{Z}^n$ and let $\mathcal{A}$ be a family of subsets of $[n]$ such that every $i \in [n]$ is contained in at least r sets in $\mathcal{A}$. Then

$$|E| \leq \prod_{A\in\mathcal{A}} |P_A E|^{1/r}.$$

Proof (Hints). Straightforward. □

Proof. Let X be a uniformly random element of E. Then by Shearer,

$$\log|E| = H(X) \leq \frac{1}{r} \cdot \sum_{A\in\mathcal{A}} H(X_A).$$

But X_A takes values in $P_A E$, so $H(X_A) \leq \log|P_A E|$ by Maximality. Hence,

$$\log|E| \leq \frac{1}{r} \sum_{A\in\mathcal{A}} |P_A E|.$$

□

Corollary 4.7 (Discrete Loomis-Whitney Theorem) If $\mathcal{A} = \{[n] \setminus \{i\} : i = 1, ..., n\}$, we get

$$|E| \leq \prod_{i=1}^{n} \left|P_{[n]\setminus\{i\}} E\right|^{1/(n-1)}.$$

Theorem 4.8 Let G be a graph with m edges. Then G has at most $\frac{1}{6}(2m)^{3/2}$ triangles.

Remark 4.9 If $m = \binom{n}{2}$, then this bound is fairly sharp.

Proof (Hints). Consider a uniformly random triangle with an ordering on the vertices, and use Shearer. □

Proof. Let (X_1, X_2, X_3) be a random triple of vertices such that X_1X_2, X_1X_3 and X_2X_3 are all edges (so pick a random triangle with an ordering of the vertices). Let t be the number of triangles in G. By Shearer,

$$\log(6t) = H(X_1, X_2, X_3) \leq \frac{1}{2}(H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3)).$$

Each (X_i, X_j) (for $i \neq j$) is supported in the set of edges of G, given a direction, so $H(X_i, X_j) \leq \log(2m)$ by Maximality. □

Definition 4.10 Let V be a set of size n and let $\mathcal{G}$ be a set of graphs, all with vertex set V. Then $\mathcal{G}$ is **Δ-intersecting** (triangle-intersecting) if $G_1 \cap G_2$ contains a triangle for all $G_1, G_2 \in \mathcal{G}$.

Theorem 4.11 If $|V| = n$, then a Δ-intersecting family of graphs with vertex set V has size at most $2^{\binom{n}{2}-2}$.

Proof (Hints).

- Let $\mathcal{G}$ be a Δ-intersecting family. View $G \in \mathcal{G}$ as a characteristic function from $V^{(2)}$ to $\{0,1\}$. Let $X = (X_e : e \in V^{(2)})$ be chosen uniformly at random from $\mathcal{G}$.
- For each $R \subseteq V$, let $G_R = K_R \cup K_{V \setminus R}$ and define $\mathcal{G}_R = \{G \cap G_R : G \in \mathcal{G}\}$. Explain why $\mathcal{G}_R$ is an intersecting family, use this to give upper bound on $|\mathcal{G}_R|$.
- What is the probability that an edge e is in a random G_R? By considering X_{G_R} taking values in the above family, conclude.

□

Proof. Let $\mathcal{G}$ be a Δ-intersecting family and let X be chosen uniformly at random from $\mathcal{G}$. We write $V^{(2)}$ for the set of (unordered) pairs of elements of V. We think of any $G \in \mathcal{G}$ as a characteristic function from $V^{(2)}$ to $\{0,1\}$. So $X = (X_e : e \in V^{(2)})$, $X_e \in \{0,1\}$ (where we fix an ordering of $V^{(2)}$). For each $R \subseteq V$, let G_R be the graph $K_R \cup K_{V \setminus R}$. For each R, we shall look at the projection X_{G_R}, which we can think of as taking values in the set $\{G \cap G_R : G \in \mathcal{G}\} =: \mathcal{G}_R$.

Note that if $G_1, G_2 \in \mathcal{G}$, $R \subseteq [n]$, then $G_1 \cap G_2 \cap G_R \neq \emptyset$, since $G_1 \cap G_2$ contains a triangle, which must intersect G_R by the pigeonhole principle (the triangle contains 3 vertices, one of which is contained in one of the two components of G_R). Thus, $\mathcal{G}_R$ is an intersecting family, so has size at most $2^{|E(G_R)|-1}$. By Lemma 4.4,

$$H(X) \leq 2 \cdot \mathbb{E}_R\left[H\left(X_{G_R}\right)\right] \leq 2 \cdot \mathbb{E}_R[|E(G_R)| - 1] = 2 \cdot \left(\frac{1}{2}\binom{n}{2} - 1\right) = \binom{n}{2} - 2,$$

since each e belongs to G_R with probability $1/2$ (and so $\mathbb{E}_R[|E(G_R)|] = \frac{1}{2}\binom{n}{2}$). □

Definition 4.12 Let G be a graph and let $A \subseteq V(G)$. The **edge-boundary** ∂A of A is the set of edges xy such that $x \in A$, $y \notin A$. If $G = \mathbb{Z}^n$ or $\{0,1\}^n$ and $i \in [n]$, the ***i*-th boundary** $\partial_i A$ is the set of edges $xy \in \partial A$ such that $x - y = \pm e_i$, i.e. $\partial_i A$ consists of edges in the edge-boundary in direction i.

Theorem 4.13 (Edge-isoperimetric Inequality in $\mathbb{Z}^n$) Let $A \subseteq \mathbb{Z}^n$ be a finite set. Then

$$|\partial A| \geq 2n \cdot |A|^{(n-1)/n}.$$

Proof (Hints). Use Discrete Loomis-Whitney Theorem, the AM-GM inequality, and a suitable lower bound on $|\partial_i A|$. □

Proof. By the Discrete Loomis-Whitney Theorem,

$$\begin{aligned}
|A| &\leq \prod_{i=1}^{n} \left|P_{[n]\setminus\{i\}} A\right|^{1/(n-1)} \\
&= \left(\prod_{i=1}^{n} \left|P_{[n]\setminus\{i\}} A\right|^{1/n}\right)^{n/(n-1)} \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n} \left|P_{[n]\setminus\{i\}} A\right|\right)^{n/(n-1)} \qquad \text{by AM-GM inequality}
\end{aligned}$$

But $|\partial_i A| \geq 2\left|P_{[n]\setminus\{i\}} A\right|$ since each fibre contributes at least 2. So

$$\begin{aligned}
|A| &\leq \left(\frac{1}{2n}\sum_{i=1}^{n} |\partial_i A|\right)^{n/(n-1)} \\
&= \left(\frac{1}{2n}|\partial A|\right)^{n/(n-1)}
\end{aligned}$$

□

Theorem 4.14 (Edge-isoperimetric Inequality in the Cube) Let $A \subseteq \{0,1\}^n$ (where we take usual graph on $\{0,1\}^n$). Then

$$|\partial A| \geq |A|(n - \log|A|).$$

Proof (Hints).

- Let $X = (X_1, ..., X_n)$ be a uniformly random element of A. Write $X_{\setminus i} = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$.
- Use Shearer to show that $\sum_{i=1}^{n} H\left(X_i \mid X_{\setminus i}\right) \leq H(X)$.
- What are the possible values of $\left|P^{-1}_{[n]\setminus\{i\}}(u)\right|$, and what is $H\left(X_i \mid X_{\setminus i} = u\right)$ in each case? How many u satisfy $\left|P^{-1}_{[n]\setminus\{i\}}(u)\right| = 1$? Use this to deduce an expression for $H\left(X_i \mid X_{\setminus i}\right)$.

□

Proof. Let X be a uniformly random element of A and write $X = (X_1, ..., X_n)$. Write $X_{\setminus i}$ for $(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$. By Shearer,

$$\begin{aligned}
H(X) &\leq \frac{1}{n-1}\sum_{i=1}^{n} H\left(X_{\setminus i}\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(H(X) - H\left(X_i \mid X_{\setminus i}\right)\right).
\end{aligned}$$

Hence, $\sum_{i=1}^n H\big(X_i \mid X_{\setminus i}\big) \le H(X)$. But

$$H\big(X_i \mid X_{\setminus i} = u\big) = \begin{cases} 1 \text{ if } \left|P^{-1}_{[n]\setminus\{i\}}(u)\right| = 2 \\ 0 \text{ if } \left|P^{-1}_{[n]\setminus\{i\}}(u)\right| = 1 \end{cases}$$

(Note that we always have $\left|P^{-1}_{[n]\setminus\{i\}}(u)\right| \in \{0,1,2\}$). The number of points of the second kind is $|\partial_i A|$. So

$$\begin{aligned} H\big(X_i \mid X_{\setminus i}\big) &= \sum_u \mathbb{P}\big(X_{\setminus i} = u\big) H\big(X_i \mid X_{\setminus i = u}\big) \\ &= \sum_{u \notin \partial_i A} \mathbb{P}\big(X_{\setminus i} = u\big) \\ &= 1 - \sum_{u \in \partial_i A} \mathbb{P}\big(X_{\setminus i} = u\big) \\ &= 1 - \frac{|\partial_i A|}{|A|}. \end{aligned}$$

So

$$\begin{aligned} H(X) &\ge \sum_{i=1}^n \left(1 - \frac{|\partial_i A|}{|A|}\right) \\ &= n - \frac{|\partial A|}{|A|}. \end{aligned}$$

Also, $H(X) = \log|A|$. So we are done. □

Definition 4.15 Let $\mathcal{A}$ be a family of sets of size d. The **lower shadow** of $\mathcal{A}$ is

$$\partial\mathcal{A} = \{B : |B| = d-1, \exists A \in \mathcal{A} \text{ s.t. } B \subseteq A\}.$$

Theorem 4.16 (Kruskal-Katona) If $|\mathcal{A}| = \binom{t}{d} = \frac{t(t-1)\cdots(t-d+1)}{d!}$ for some real number t, then

$$|\partial\mathcal{A}| \ge \binom{t}{d-1}.$$

Proof (Hints).

- Let $X = (X_1, ..., X_d)$ be a random ordering of the elements of a uniformly random $A \in \mathcal{A}$. Give an expression for $H(X)$.
- Explain why it is enough to show $H(X_1, ..., X_{d-1}) \ge \log\big((d-1)!\binom{t}{d-1}\big)$.
- Let $T \sim \text{Bern}(p)$ be independent of $X_1, ..., X_{k-1}$, and given $X_1, ..., X_{k-1}$, let

$$X^* = \begin{cases} X_{k+1} \text{ if } T = 0 \\ X_k \quad \text{if } T = 1 \end{cases}.$$

- Use that X_k and X^* have the same distribution (why?) to show that $H(X_k \mid X_{<k}) \ge H(X^*, T \mid X_{\le k}) = h(p) + pH(X_{k+1} \mid X_{\le k})$.
- Find the maximum of the lower bound to show that $H(X_k \mid X_{<k}) \ge \log\big(2^{H(X_{k+1} \mid X_{\le k})} + 1\big)$.

- Using the chain rule and the fact that $\log\big(d!\binom{t}{d}\big)$ is an increasing function in t for $t \geq d$, show that $r + d - 1 \leq t$ where $r = 2^{H(X_d \mid X_{<d})}$, and use this to conclude the desired bound on $H(X_{<d})$.

□

Proof. Let $X = (X_1, ..., X_d)$ be a random ordering of the elements of a uniformly random $A \in \mathcal{A}$. Then $H(X) = \log(d!|A|) = \log\big(d!\binom{t}{d}\big)$. Note that $(X_1, ..., X_{d-1})$ is an ordering of the elements of some $B \in \partial_i A$, so

$$H(X_1, ..., X_{d-1}) \leq \log((d-1)!|\partial_i A|)$$

So it's enough to show $H(X_1, ..., X_{d-1}) \geq \log\big((d-1)!\binom{t}{d-1}\big)$. Also, $H(X) = H(X_1, ..., X_{d-1}) + H(X_d \mid X_1, ..., X_{d-1})$ and $H(X) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_d \mid X_1, ..., X_{d-1})$. We would like an upper bound for $H(X_d \mid X_{<d})$. Our strategy will be to obtain a lower bound for $H(X_k \mid X_{<k})$ in terms of $H(X_{k+1} \mid X_{<k+1})$. We shall prove that $2^{H(X_k \mid X_{<k})} \geq 2^{H(X_{k+1} \mid X_{<k+1})} + 1$ for all k.

Let T be chosen independently of X. Let $T \sim \text{Bern}(1-p)$ ($T = 0$ with probability p, p is to be chosen later). Given $X_1, ..., X_{k-1}$, let

$$X^* = \begin{cases} X_{k+1} & \text{if } T = 0 \\ X_k & \text{if } T = 1 \end{cases}$$

Note that X_k and X_{k+1} have the same distribution (given $X_1, ..., X_{k-1}$), so X^* does as well. Then

$$\begin{aligned}
H(X_k \mid X_{<k}) &= H(X^* \mid X_{<k}) \text{ since } X_k \sim X^* \\
&\geq H(X^* \mid X_{\leq k}) \quad \text{by Submodularity} \\
&= H(X^*, T \mid X_{\leq k}) \quad \text{since } X_{\leq k} \text{ and } X^* \text{ determine } T \text{ (since } X_{k+1} \neq X_k) \\
&= H(T \mid X_{\leq k}) + H(X^* \mid T, X_{\leq k}) \quad \text{by Additivity} \\
&= H(T) + pH(X^* \mid X_{\leq k}, T = 0) + (1-p)H(X^* \mid X_{\leq k}, T = 1) \\
&= H(T) + pH(X_{k+1} \mid X_{\leq k}) + (1-p)H(X_k \mid X_{\leq k}) \\
&= h(p) + ps.
\end{aligned}$$

where $s = H(X_{k+1} \mid X_{\leq k})$ and $h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$. This is maximised when $p = \frac{2^s}{2^s+1}$. Then we get

$$\frac{2^s}{2^s+1}(\log(2^s+1) - \log(2^s)) + \frac{1}{2^s+1}(\log(2^s+1)) + \frac{s2^s}{2^s+1} = \log(2^s+1).$$

This proves the claim.

Let $r = 2^{H(X_d \mid X_{<d})}$. Then by the claim,

$$\begin{aligned}
H(X) &= H(X_1) + \cdots + H(X_d \mid X_{<d}) \\
&\geq \log(r + d - 1) + \cdots + \log(r)
\end{aligned}$$

$$= \log\left(\frac{(r+d-1)!}{(r-1)!}\right) = \log\left(d!\binom{r+d-1}{d}\right).$$

Since $H(X) = \log\big(d!\binom{t}{d}\big)$ is an increasing function (for $t \geq d$), it follows that $r+d-1 \leq t$, i.e. $r \leq t+1-d$. It follows that

$$\begin{aligned} H(X_{<d}) &= \log\left(d!\binom{t}{d}\right) - \log r \\ &\geq \log\left(d!\frac{t!}{d!(t-d)!(t+1-d)}\right) \\ &= \log\left((d-1)!\binom{t}{d-1}\right). \end{aligned}$$

□

5. The union-closed conjecture

Definition 5.1 Let $\mathcal{A}$ be a finite family of sets. $\mathcal{A}$ is **union-closed** if $A \cup B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$.

Conjecture 5.2 (Union-closed Conjecture) If $\mathcal{A}$ is a non-empty union-closed family, then there exists x that belongs to at least $\frac{1}{2}|\mathcal{A}|$ sets in $\mathcal{A}$.

Theorem 5.3 (Gilmer) There exists a constant $c > 0$ such that if $\mathcal{A}$ is any union-closed family, then there exists x that belongs to at least $c|\mathcal{A}|$ of the sets in $\mathcal{A}$.

Example 5.4 Let $\mathcal{A} = [n]^{(pn)} \cup [n]^{(\geq(2p-p^2-o(1))n)}$. Then with high probability, if A, B are random elements of $[n]^{(pn)}$, then $|A \cup B| \geq (2p-p^2-o(1))n$ (since the intersect is likely of size at most $p^2 n$). If $1-(2p-p^2-o(1)) = p$, then almost all of $\mathcal{A}$ is contained in $[n]^{(pn)}$. The solutions of p occur roughly when $1-3p+p^2 = 0$, which has solutions $p = \frac{1}{2}\big(3 \pm \sqrt{5}\big)$.

If we want to prove Gilmer, it is natural to let A, B be independent uniformly random elements of $\mathcal{A}$ and to consider $H(A \cup B)$. Since $\mathcal{A}$ is union-closed, $A \cup B \in \mathcal{A}$, so $H(A \cup B) \leq \log|\mathcal{A}|$. Now we would like to get a lower bound for $H(A \cup B)$ assuming that no x belongs to more than $p|\mathcal{A}|$ sets in $\mathcal{A}$.

Lemma 5.5 Suppose $c > 0$ is such that $h(xy) \geq c(xh(y) + yh(x))$ for every $x, y \in [0,1]$. Let $\mathcal{A}$ be a family of sets such that every element (of $\cup\,\mathcal{A}$) belongs to fewer than $p|\mathcal{A}|$ members of $\mathcal{A}$. Let A, B be independent uniformly random members of $\mathcal{A}$. Then

$$H(A \cup B) > c(1-p)(H(A) + H(B)).$$

Proof (Hints).

- Think of A, B as characteristic functions. Write $A_{<k}$ for $(A_1, ..., A_{k-1})$.
- Explain why it is enough to prove that $H((A \cup B)_k \mid A_{<k}, B_{<k}) > c(1-p)\big(H(A_k \mid A_{<k}) + H\big(B_k \mid H_{B_{<k}}\big)\big)$ for all k.
- For each $u, v \in \{0,1\}^{k-1}$, write $p(u) = \mathbb{P}(A_k = 0 \mid A_{<k} = u)$ and $q(v) = \mathbb{P}(B_k = 0 \mid B_{<k} = v)$. Find a (simple) expression for $H((A \cup B)_k \mid A_{<k} = u, B_{<k} = v)$.

- Expand $H((A \cup B)_k \mid A_{<k}, B_{<k})$, give an upper bound, then simplify it (hint: law of total probability).

□

Proof. Think of A, B as characteristic functions. Write $A_{<k}$ for $(A_1, ..., A_{k-1})$. By the Chain Rule, it is enough to prove for every k that

$$H((A \cup B)_k \mid (A \cup B)_{<k}) > c(1-p)\Big(H(A_k \mid A_{<k}) + H\Big(B_k \mid H_{B_{<k}}\Big)\Big).$$

By Lemma 1.25,

$$H((A \cup B)_k \mid (A \cup B)_{<k}) \geq H((A \cup B)_k \mid A_{<k}, B_{<k})$$

For each $u, v \in \{0,1\}^{k-1}$, write $p(u) = \mathbb{P}(A_k = 0 \mid A_{<k} = u)$ and $q(v) = \mathbb{P}(B_k = 0 \mid B_{<k} = v)$. Then, since A and B are independent,

$$\begin{aligned}
&H((A \cup B)_k \mid A_{<k} = u, B_{<k} = v) \\
&= -\sum_{i=0}^{1} \mathbb{P}((A \cup B)_k = i \mid A_{<k} = u, B_{<k} = v) \log \mathbb{P}((A \cup B)_k = i \mid A_{<k} = u, B_{<k} = v) \\
&= h(p(u)q(v)).
\end{aligned}$$

which by hypothesis is at least $c(p(u)h(q(v)) + q(v)h(p(u)))$. So

$$\begin{aligned}
H((A \cup B)_k \mid (A \cup B)_{<k}) &\geq c \sum_{u,v} \mathbb{P}(A_{<k} = u)\mathbb{P}(B_{<k} = v)(p(u)h(q(v)) + q(v)h(p(u))) \\
&= c \cdot \sum_u \mathbb{P}(A_{<k} = u)p(u) \cdot \sum_v \mathbb{P}(B_{<k} = v)h(q(v)) \\
&+ c \cdot \sum_u \mathbb{P}_{A_{<k}=u} h(p(u)) \cdot \sum_v \mathbb{P}(B_{<k} = v)q(v)
\end{aligned}$$

But by law of total probability,

$$\sum_u \mathbb{P}(A_{<k} = u)\mathbb{P}(A_k = 0 \mid A_{<k} = u) = \mathbb{P}(A_k = 0),$$

and

$$\sum_v \mathbb{P}(B_{<k} = v)h(q(v)) = \sum_v \mathbb{P}(B_{<k} = v)H(B_k \mid B_{<k} = v) = H(B_k \mid B_{<k})$$

Similarly for the other term, so the RHS of the inequality equals

$$c(\mathbb{P}(A_k = 0)H(B_k \mid B_{<k}) + \mathbb{P}(B_k = 0)H(A_k \mid A_{<k})),$$

which by hypothesis (since $\mathbb{P}(A_k = 0) = \mathbb{P}(B_k = 0) > 1 - p$) is greater than

$$c(1-p)(H(A_k \mid A_{<k}) + H(B_k \mid B_{<k}))$$

as required. □

Corollary 5.6 Let $\mathcal{A}$, p and c be as in Lemma 5.5. If $\mathcal{A}$ is union-closed, then we must have $p \geq 1 - 1/2c$.

Proof (Hints). Straightforward. □

Proof. Let A and B be independent uniformly random elements of $\mathcal{A}$. Since $\mathcal{A}$ is union-closed, $A \cup B \in \mathcal{A}$, so $H(A \cup B) \leq \log|\mathcal{A}|$. Also, $H(A) = H(B) = \log|\mathcal{A}|$. Hence, by Lemma 5.5, $2c(1-p) \leq 1$. □

Corollary 5.6 gives a non-trivial bound as long as $c > 1/2$. We shall obtain $1/\left(\sqrt{5}-1\right)$.

We start by proving the diagonal case, i.e. where $x = y$.

Lemma 5.7 (Boppana) For every $x \in [0,1]$,

$$h(x^2) \geq \varphi \cdot x \cdot h(x),$$

where $\varphi = \frac{1}{2}\left(\sqrt{5}+1\right)$.

Proof (Hints).

- Let $\psi = 1/\varphi$. Show that equality holds when $x = \psi, 0, 1$.
- Let $f(x) = h(x^2) - \varphi \cdot x \cdot h(x)$. Show that $f'''(x) = 0$ iff $-\varphi x^3 - 4x^2 + 3\varphi x - 4 + 2\varphi = 0$. (Advice: use natural logs and find expressions for $h'(x)$, $h''(x)$ and $h'''(x)$ first).
- Explain why f''' has at most two roots in $(0,1)$ and so f has at most five roots in $[0,1]$.
- Show that f has a double root at 0 and at ψ.
- Explain why f must have constant sign on $[0,1]$, and by considering small x, show that there is x with $f(x) > 0$.

□

Proof. Write $\psi = 1/\varphi = \frac{1}{2}\left(\sqrt{5}-1\right)$. Then $\psi^2 = 1 - \psi$. So $h(\psi^2) = h(1-\psi) = h(\psi)$ and $\varphi\psi = 1$, so $h(\psi^2) = \varphi \cdot \psi \cdot h(\psi)$. So equality holds when $x = \psi$, and also when $x = 0, 1$.

Toolkit: $\ln(2) \cdot h(x) = -x \ln x - (1-x)\ln(1-x)$. Then

$$\ln(2) \cdot h'(x) = -\ln x - 1 + \ln(1-x) + 1 = \ln(1-x) - \ln(x)$$

and

$$\ln(2) \cdot h''(x) = -\frac{1}{x} - \frac{1}{1-x} = -\frac{1}{x(1-x)}$$

and

$$\ln(2) \cdot h'''(x) = \frac{1}{x^2} - \frac{1}{(1-x)^2} = \frac{1-2x}{x^2(1-x)^2}.$$

Let $f(x) = h(x^2) - \varphi \cdot x \cdot h(x)$. Then

$$f'(x) = 2xh'(x^2) - \varphi h(x) - \varphi x h'(x)$$

$$
\begin{aligned}
f''(x) &= 2h'(x^2) + 4x^2h''(x^2) - 2\varphi h'(x) - \varphi x h''(x) \\
f'''(x) &= 4xh''(x^2) + 8xh''(x^2) + 8x^3h'''(x^2) - 3\varphi h''(x) - \varphi x h'''(x) \\
&= 12xh''(x^2) + 8x^3h'''(x^2) - 3\varphi h''(x) - \varphi x h'''(x)
\end{aligned}
$$

So

$$
\begin{aligned}
\ln(2) f'''(x) &= \frac{-12x}{x^2(1-x^2)} + \frac{8x^3(1-2x^2)}{x^4(1-x^2)^2} + \frac{3\varphi}{x(1-x)} - \frac{\varphi x(1-2x)}{x^2(1-x)^2} \\
&= \frac{-12}{x(1-x^2)} + \frac{8(1-2x^2)}{x(1-x^2)^2} + \frac{3\varphi}{x(1-x)} - \frac{\varphi(1-2x)}{x(1-x)^2} \\
&= \frac{-12(1-x^2) + 8(1-2x^2) + 3\varphi(1-x)(1+x)^2 - \varphi(1-2x)(1+x)^2}{x(1-x)^2(1+x)^2}
\end{aligned}
$$

which is zero iff

$$
\begin{aligned}
&-12 + 12x + 8 - 16x^2 + 3\varphi(1 + x - x^2 - x^3) - \varphi(1 - 3x^2 - 2x^3) \\
&= -\varphi x^3 - 4x^2 + 3\varphi x - 4 + 2\varphi = 0.
\end{aligned}
$$

So the numerator of $f'''(x)$ is a cubic with negative leading coefficient and constant term, so it has a negative root, so it has at most two roots in $(0,1)$. It follows (by Rolle's theorem) that f has at most five roots in $[0,1]$, up to multiplicity. But

$$
f'(x) = 2x(\log(1-x^2) - \log(x^2)) + \varphi(x\log x + (1-x)\log(1-x)) - \varphi x(\log(1-x) - \log x)
$$

So $f'(0) = 0$, so f has a double root at 0. Now

$$
\begin{aligned}
f'(\psi) &= 2\psi(\log\psi - 2\log\psi) + \varphi(\psi\log\psi + 2(1-\psi)\log\psi) - (2\log\psi - \log\psi) \\
&= -2\psi\log\psi + \log\psi + 2\varphi\log\psi - 2\log\psi \\
&= 2\log\psi(-\psi + \varphi - 1) \\
&= 2\varphi\log\psi(-\psi^2 - 1 - \psi) = 0
\end{aligned}
$$

So there is a double root at ψ. Also, $f(1) = 0$. So f is either non-negative on all of $[0,1]$ or non-positive on all of $[0,1]$. If x is small,

$$
\begin{aligned}
f(x) &= x^2\log\frac{1}{x^2} + (1-x^2)\log\frac{1}{1-x^2} - \varphi x\left(x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x}\right) \\
&= 2x^2\log\frac{1}{x} - \varphi x^2\log\frac{1}{x} + O(x^2).
\end{aligned}
$$

So, because $2 > \varphi$, there exists x such that $f(x) > 0$. □

Lemma 5.8 The function $f(x,y) = \frac{h(xy)}{xh(y)+yh(x)}$ is minimised on $(0,1)^2$ at a point where $x = y$.

Proof (Hints).

- Show that we can extend f continuously to the boundary by setting $f(x,y) = 1$ whenever x or y is 0 or 1 (for the case when x or y tend to 0 separately, consider an

expansion for xy small, and for the case when x and y tend to 1, consider when one of x or y is 1).

- Pick any point in $(0,1)^2$ to show that f is minimised somewhere in that region.
- Let (x^*, y^*) be a minimum with $f(x^*, y^*) = \alpha$. Let $g(x) = h(x)/x$.
- By considering the expression $g(xy) - \alpha(g(x) + g(y))$ and partial derivatives, show that $x^* g'(x^*) = y^* g'(y^*)$.
- Show that $xg'(x)$ is an injection by considering its derivative.

□

Proof. We can extend f continuously to the boundary by setting $f(x, y) = 1$ whenever x or y is 0 or 1. To see this, note first that it is easy if neither x nor y is 0. If either x or y is small then $h(xy) = -xy(\log x + \log y) + O(xy)$, and

$$\begin{aligned} xh(y) + yh(x) &= -x(y \log y + O(y)) - y(x \log x + O(x)) \\ &= h(xy) \quad \text{up to } O(xy) \end{aligned}$$

So it tends to 1 again.

We can check that $f(1/2, 1/2) < 1$, so f is minimised somewhere in $(0,1)^2$. Let (x^*, y^*) be a minimum with $f(x^*, y^*) = \alpha$. For convenience, let $g(x) = h(x)/x$ and note that $f(x, y) = \frac{g(xy)}{g(x)+g(y)}$. Also, $g(xy) - \alpha(g(x) + g(y)) \geq 0$ with equality at (x^*, y^*). So the partial derivatives of the LHS are both 0 at (x^*, y^*):

$$\begin{aligned} y^* g'(x^* y^*) - \alpha g'(x^*) &= 0 \\ x^* g'(x^* y^*) - \alpha g'(y^*) &= 0. \end{aligned}$$

So $x^* g'(x^*) = y^* g'(y^*)$. So it is enough to prove that $xg'(x)$ is an injection. We have

$$g'(x) = \frac{h'(x)}{x} - \frac{h(x)}{x^2}$$

so

$$\begin{aligned} xg'(x) &= h'(x) - \frac{h(x)}{x} \\ &= \log(1-x) - \log x + \frac{x \log x + (1-x)\log(1-x)}{x} \\ &= \frac{\log(1-x)}{x}. \end{aligned}$$

Differentiating gives

$$-\frac{1}{x(1-x)} - \frac{\log(1-x)}{x^2} = \frac{-x - (1-x)\log(1-x)}{x^2(1-x)}$$

The numerator differentiates to $-1 + 1 + \log(1-x)$ which is negative. Also, it equals 0 at 0, so it has a constant sign. Thus, $xg'(x)$ is indeed an injection. □

Combining this with Boppana we get that

$$h(xy) \geq \frac{\varphi}{2}(xh(y) + yh(x))$$

This allows us to take $p = 1 - \frac{1}{\varphi} = \frac{3-\sqrt{5}}{2}$.

6. Entropy in additive combinatorics

We shall need two "simple" results from additive combinatorics due to Imre Ruzsa.

Definition 6.1 Let G be an abelian group and let $A, B \subseteq G$. The **sumset** $A + B$ of A and B is the set

$$\{x + y : x \in A, y \in B\}$$

and the **difference set** $A - B$ is the set

$$\{x - y : x \in A, y \in B\}.$$

Write $2A$ for $A + A$, $3A$ for $A + A + A$, etc.

Definition 6.2 The **Ruzsa distance** $d(A, B)$ is

$$\frac{|A - B|}{|A|^{1/2} \cdot |B|^{1/2}}.$$

Lemma 6.3 (Ruzsa Triangle Inequality) $d(A, C) \leq d(A, B) \cdot d(B, C)$.

Proof (Hints). Expand the stated inequality and consider an appropriate injection. □

Proof. This is equivalent to the statement

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

For each $x \in A - C$, pick $a(x) \in A$, $c(x) \in C$ such that $x = a(x) - c(x)$. Define the map

$$\begin{aligned} \varphi : (A - C) \times B &\to (A - B) \times (B - C), \\ (x, b) &\mapsto (a(x) - b, b - c(x)). \end{aligned}$$

Adding the coordinates of $\varphi(x, b)$ gives x, so we can calculate $a(x)$ and $c(x)$ from $\varphi(x, b)$, and hence b. So φ is an injection. □

Lemma 6.4 (Ruzsa Covering Lemma) Let G be an abelian group and let $A, B \subseteq G$ be finite. Then A can be covered by at most $|A + B|/|B|$ translates of $B - B$.

Proof (Hints). Consider a maximal subset $\{x_1, ..., x_k\} \subseteq A$ such that the $x_i + B$ are disjoint, show $k \leq |A + B|/|B|$. □

Proof. Let $\{x_1, ..., x_k\}$ be a maximal subset of A such that the sets $x_i + B$ are disjoint. Then for all, $a \in A$, there exists i such that $(a + B) \cap (x_i + B) \neq \emptyset$, i.e. $a \in (x_i + (B - B))$. So A can be covered by k translates of $B - B$. But since the $x_i + B$ are disjoint,

$$|B|k = |\{x_1, ..., x_k\} + B| \leq |A + B|.$$

□

Let X, Y be discrete random variables taking values in an abelian group. What is $X+Y$ when X and Y are independent? For each z, $\mathbb{P}(X+Y=z) = \sum_{x+y=z} \mathbb{P}(X=x)\mathbb{P}(Y=y)$. Writing p_x and q_y for $\mathbb{P}(X=x)$ and $\mathbb{P}(Y=y)$, this gives

$$\sum_{x+y=z} p_x p_y = (p * q)(z)$$

where $p(x) = p_x$, $q(y) = q_y$. So sums of independent random variables correspond to convolutions.

Definition 6.5 Let G be an abelian group and let X, Y be G-valued random variables. The **(entropic) Ruzsa distance** between X and Y is

$$\begin{aligned} d(X;Y) &= H(X'-Y') - \frac{1}{2}H(X) - \frac{1}{2}H(Y) \\ &= H(X'-Y') - \frac{1}{2}H(X') - \frac{1}{2}H(Y'). \end{aligned}$$

where X', Y' are independent copies of X, Y.

Lemma 6.6 If A, B are finite subsets of G and X, Y are uniform on A, B respectively, then

$$d(X;Y) \leq \log d(A,B).$$

Proof (Hints). Straightforward. □

Proof. WLOG X, Y are independent. Then

$$\begin{aligned} d(X,Y) &= H(X-Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y) \\ &\leq \log|A-B| - \frac{1}{2}\log|A| - \frac{1}{2}\log|B| = \log d(A,B). \end{aligned}$$

□

Lemma 6.7 Let X, Y be G-valued random variables. Then

$$H(X-Y) \geq \max\{H(X), H(Y)\} - I(X:Y).$$

Proof (Hints). Use that $H(X-Y) \geq H(X-Y \mid Y)$ and $H(X-Y) \geq H(X-Y \mid X)$. □

Proof. We have

$$\begin{aligned} H(X-Y) &\geq H(X-Y \mid Y) \text{ by Subadditivity} \\ &= H(X-Y, Y) - H(Y) \\ &= H(X,Y) - H(Y) \text{ by Invariance} \\ &= H(X) + H(Y) - H(Y) - I(X:Y) \\ &= H(X) - I(X:Y). \end{aligned}$$

We use Invariance with the bijection $(x, y) \mapsto (x - y, y)$. By symmetry, we also have $H(X - Y) \geq H(Y) - I(X : Y)$. □

Corollary 6.8 If X, Y are G-valued RVs, then $d(X; Y) \geq 0$.

Proof (Hints). Straightforward. □

Proof. WLOG X and Y are independent. Then $I(X : Y) = 0$, so $H(X - Y) \geq \max\{H(X), H(Y)\} \geq \frac{1}{2}(H(X) + H(Y))$. □

Lemma 6.9 If X, Y are G-valued RVs, then $d(X; Y) = 0$ iff there is some (finite) subgroup H of G such that X and Y are uniform on cosets of H.

Proof (Hints).

- $\Longleftarrow$: straightforward.
- $\Longrightarrow$: assume WLOG that X and Y are independent. By considering entropy, explain why $X - Y$ and Y are independent.
- Deduce that for X supported on A and Y supported on B, for all $z \in A - B$ and $y_1, y_2 \in B$, $\mathbb{P}(X = y_1 + z) = \mathbb{P}(X = y_2 + z)$.
- Show by contradiction that thus $\mathbb{P}(X = x)$ is non-zero on $z + B$, and so that $z + B \subseteq A$.
- Deduce that $A = B + z$ for all $z \in A - B$, and so that $A - x$ is constant over $x \in A$.
- Deduce that $A - A$ is a subgroup.

□

Proof. $\Longleftarrow$: If X, Y are uniform on $x + H, y + H$ then $X' - Y'$ is uniform on $(x - y) + H$, so $H(X' - Y') = H(X) = H(Y)$.

$\Longrightarrow$: WLOG X and Y are independent. We have $H(X - Y) = \frac{1}{2}(H(X) + H(Y))$. So equality must hold throughout the proof of Lemma 6.7 and Corollary 6.8, thus $H(X - Y \mid Y) = H(X - Y)$. Therefore, $X - Y$ and Y are independent. So for every $z \in A - B$ and $y_1, y_2 \in B$,

$$\mathbb{P}(X - Y = z \mid Y = y_1) = \mathbb{P}(X - Y = z \mid Y = y_2),$$

where $A = \{x : \mathbb{P}(X = x) \neq 0\}$ and $B = \{y : \mathbb{P}(Y = y) \neq 0\}$. We can write this as

$$\mathbb{P}(X = y_1 + z) = \mathbb{P}(X = y_2 + z)$$

So $\mathbb{P}(X = x)$ is constant on $z + B$. In particular, $z + B \subseteq A$ ($\mathbb{P}(X = x)$ must be non-zero on $z + B$, as otherwise $(z + B) \cap A = \emptyset$, i.e. $z \notin A - B$). By the same argument, $A - z \subseteq B$. So $A = B + z$ for all $z \in A - B$. So for every $x \in A$ and $y \in B$, $A = B + x - y$, so $A - x = B - y$. Hence, $A - x$ is the same for every $x \in A$. Therefore, $A - x = \cup_{x \in A} (A - x) = A - A$ for all $x \in A$. It follows that

$$A - A + A - A = (A - A) - (A - A) = A - x - (A - x) = A - A.$$

So $A - x = A - A$ is a subgroup, and so A is a coset of $A - A$. $B = A + x$, so B is also a coset of $A - A$. Also, as stated above, X is uniform on $z + B = A$ and Y is uniform on $A - z = B$. □

Lemma 6.10 (Entropic Ruzsa Triangle Inequality) Let X, Y, Z be G-valued random variables. Then $d(X;Z) \leq d(X;Y) + d(Y;Z)$.

Proof (Hints). Simplify the desired inequality and use Lemma 1.26 (where $X - Z$ depends on two different (pairs of) random variables). □

Proof. We must show (assuming WLOG that X, Y, Z are independent) that

$$\begin{aligned} &H(X-Z) - \frac{1}{2}H(X) - \frac{1}{2}H(Z) \\ &\leq H(X-Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y) + H(Y-Z) - \frac{1}{2}H(Y) - \frac{1}{2}H(Z), \end{aligned}$$

i.e. that $H(X-Z) + H(Y) \leq H(X-Y) + H(Y-Z)$. Since $X - Z$ depends on $(X - Y, Y - Z)$ and on (X, Z), by Lemma 1.26,

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

i.e. $H(X, Y, Z) + H(X-Z) \leq H(X, Z) + H(X-Y, Y-Z)$. By independence and Subadditivity, we get $H(X-Z) + H(Y) \leq H(X-Y) + H(Y-Z)$. □

Lemma 6.11 (Submodularity for Sums) If X, Y, Z are independent G-valued RVs, then

$$H(X+Y+Z) + H(Z) \leq H(X+Z) + H(Y+Z).$$

Proof (Hints). Use Lemma 1.26. □

Proof. $X + Y + Z$ is a function of $(X + Z, Y)$ and of $(X, Y + Z)$. Therefore, by Lemma 1.26,

$$H(X+Z, Y, X, Y+Z) + H(X+Y+Z) \leq H(X+Z, Y) + H(X, Y+Z),$$

thus $H(X, Y, Z) + H(X+Y+Z) \leq H(X+Z) + H(Y) + H(X) + H(Y+Z)$. By independence and cancelling equal terms, we get the desired inequality. □

Lemma 6.12 Let G be an abelian group and let X be a G-valued random variable. Then $d(X; -X) \leq 2d(X; X)$.

Proof (Hints). Consider independent copies X_1, X_2, X_3 of X, use Lemma 6.7 and Submodularity for Sums. □

Proof. Let X_1, X_2, X_3 be independent copies of X. Then by Lemma 6.7,

$$\begin{aligned} d(X;-X) &= H(X_1 + X_2) - \frac{1}{2}H(X_1) - \frac{1}{2}H(X_2) \\ &\leq H(X_1 + X_2 - X_3) - H(X) \\ &\leq H(X_1 - X_3) + H(X_2 - X_3) - H(X_3) - H(X) \\ &= 2d(X;X) \end{aligned}$$

by Submodularity for Sums and since X_1, X_2, X_3 are all copies of X. □

Corollary 6.13 Let X and Y be G-valued random variables. Then $d(X; -Y) \leq 5d(X; Y)$.

Proof (Hints). Straightforward. □

Proof. By the Entropic Ruzsa Triangle Inequality,

$$\begin{aligned} d(X;-Y) &\le d(X;Y) + d(Y;-Y) \\ &\le d(X;Y) + 2d(Y;Y) \\ &\le d(X;Y) + 2(d(Y;X) + d(X;Y)) = 5d(X;Y). \end{aligned}$$

□

Definition 6.14 Let X, Y, U, V be G-valued random variables. The **conditional distance** is

$$d(X \mid U; Y \mid V) = \sum_{u,v} \mathbb{P}(U=u)\mathbb{P}(V=v) d(X \mid U=u; Y \mid V=v).$$

Definition 6.15 Let X, Y, U be G-valued random variables. The **simultaneous conditional distance** of X to Y given U is

$$d(X; Y \parallel U) := \sum_{u} \mathbb{P}(U=u) d(X \mid U=u; Y \mid U=u).$$

Definition 6.16 We say that X', Y' are **conditionally independent trials** of X, Y given U if X' is distributed like X, Y' like Y, and for each u, $X' \mid U=u$ is distributed like $X \mid U=u$, $Y' \mid U=u$ is distributed like $Y \mid U=u$, and $X' \mid U=u$ and $Y' \mid U=u$ are independent.

In that case, $d(X; Y \parallel U) = H(X' - Y' \mid U) - \frac{1}{2}H(X' \mid U) - \frac{1}{2}H(Y' \mid U)$.

Lemma 6.17 (Entropic BSG Theorem) Let A, B be G-valued RVs. Then

$$d(A; B \parallel A+B) \le 3I(A:B) + 2H(A+B) - H(A) - H(B).$$

Proof (Hints).

- Let A', B' be conditionally independent trials of A, B given $A+B$.
- Show that $H(A' \mid A+B) = H(A) + H(B) - I(A:B) - H(A+B)$.
- Let (A_1, B_1) and (A_2, B_2) be conditionally independent trials of (A, B) given $A+B$.
- Explain why $H(A_1 - B_2) \le H(A_1 - B_2, A_1) + H(A_1 - B_2, B_1) - H(A_1 - B_2, A_1, B_1)$.
- Use that $A_1 + B_1 = A_2 + B_2$ to bound each of the first two terms on the RHS of the above, and rewrite the $H(A_1 - B_2, A_1, B_1)$ term, using the conditional independence of (A_1, B_1) and (A_2, B_2), to conclude the result.

□

Proof. We have

$$d(A, B \parallel A+B) = H(A' - B' \mid A+B) - \frac{1}{2}H(A' \mid A+B) - \frac{1}{2}H(B' \mid A+B),$$

where A', B' are conditionally independent trials of A, B given $A+B$. Now

$$\begin{aligned} H(A' \mid A+B) &= H(A \mid A+B) = H(A, A+B) - H(A+B) \\ &= H(A,B) - H(A+B) \\ &= H(A) + H(B) - I(A:B) - H(A+B). \end{aligned}$$

Similarly, $H(B' \mid A+B) = H(A) + H(B) - I(A:B) - H(A+B)$, so

$$\frac{1}{2}H(A' \mid A+B) + \frac{1}{2}H(B' \mid A+B)$$

is also the same. By Subadditivity, $H(A'-B' \mid A+B) \leq H(A'-B')$. Let (A_1, B_1) and (A_2, B_2) be conditionally independent trials of (A, B) given $A+B$ (here, A_1 plays the role of A', B_2 plays the role of B', and each comes with another RV since we know the value of $A+B$). Then $H(A'-B') = H(A_1 - B_2)$. By Submodularity,

$$H(A_1 - B_2) \leq H(A_1 - B_2, A_1) + H(A_1 - B_2, B_1) - H(A_1 - B_2, A_1, B_1)$$

Also,

$$H(A_1 - B_2, A_1) = H(A_1, B_2) \leq H(A_1) + H(B_2) = H(A) + H(B)$$

and since $A_1 + B_1 = A_2 + B_2$,

$$H(A_1 - B_2, B_1) = H(A_2 - B_1, B_1) = H(A_2, B_1) \leq H(A) + H(B).$$

Finally, since $A_1 + B_1 = A_2 + B_2 = A + B$,

$$\begin{aligned} H(A_1 - B_2, A_1, B_1) &= H(A_1, B_1, A_2, B_2) \\ &= H(A_1, B_1, A_2, B_2 \mid A+B) + H(A+B) \\ &= 2H(A, B \mid A+B) + H(A+B) \\ &= 2H(A,B) - H(A+B) \\ &= 2H(A) + 2H(B) - 2I(A:B) - H(A+B). \end{aligned}$$

where the third line is by conditional independence of (A_1, B_1) and (A_2, B_2). Adding or subtracting as appropriate all these terms gives the required inequality. □

7. A proof of Marton's conjecture in $\mathbb{F}_2^n$

We shall prove the following theorem.

Theorem 7.1 (Polynomial Freiman-Ruzsa) There is a polynomial p with the following property: if $n \in \mathbb{N}$ and $A \subseteq \mathbb{F}_2^n$ is such that $|A+A| \leq C|A|$, then there is a subspace $H \subseteq \mathbb{F}_2^n$ of size at most $|A|$ such that A is contained in the union of at most $p(C)$ translates of H. Equivalently, there exists $K \subseteq \mathbb{F}_2$, $|K| \leq p(C)$, such that $A \subseteq K + H$.

Remark 7.2 Polynomial Freiman-Ruzsa is also known as the **Green-Manners-Tao-Gowers theorem**.

In fact, we shall prove the following statement:

Theorem 7.3 (EPFR) Let $G = \mathbb{F}_2^n$. There is an absolute constant α with the following property:

Let X, Y be G-valued random variables. Then there exists a subgroup H of G such that

$$d(X;U_H) + d(U_H;Y) \leq \alpha d(X;Y),$$

where U_H is a random variable distributed uniformly on H.

Remark 7.4 "EPFR" stands for entropic polynomial Freiman-Ruzsa.

Lemma 7.5 Let X be a discrete random variable and write $p_x = \mathbb{P}(X = x)$. Then there exists x such that $p_x \geq 2^{-H(X)}$.

Proof (Hints). By contradiction. □

Proof. If not, then $H(X) = \sum_x p_x \log(1/p_x) > H(X) \sum_x p_x = H(X)$: contradiction. □

Proposition 7.6 EPFR implies Polynomial Freiman-Ruzsa.

Proof (Hints).

- Let $A \subseteq \mathbb{F}_2^n$ and $|A+A| \leq C|A|$. Let U_H be uniformly distributed on H, let X and Y be independent copies of U_A. Show that $d(X;U_H) \leq \frac{1}{2}\alpha \log C$.
- Deduce that there exists z such that

  $$\mathbb{P}(X + U_H = z) \geq |A|^{-1/2}|H|^{-1/2}C^{-\alpha/2}$$

  and find an expression for the LHS.
- Let $B = A \cap (z + H)$. Show that A can be covered by at most $\frac{|A+B|}{|B|}$ translates of H.
- Use that $B \subseteq A, z + H$ to show that

  $$\frac{|A+B|}{|B|} \leq C^{\alpha/2+1}\frac{|A|^{1/2}}{|H|^{1/2}} \leq C^{\alpha+1}.$$
- Consider the cases $|H| \leq |A|$ and $|H| > |A|$: if the latter, then consider a subgroup H' of H of size between $|A|/2$ and $|A|$ (why does this exist?).

□

Proof. Let $A \subseteq \mathbb{F}_2^n$ and $|A+A| \leq C|A|$. Let X and Y be independent copies of U_A. Then by EPFR, there exists a subgroup H such that $d(X;U_H) + d(U_H;X) \leq \alpha d(X;Y)$, so $d(X;U_H) \leq \frac{\alpha}{2} d(X;Y)$. But since we are in $\mathbb{F}_2^n$,

$$\begin{aligned} d(X;Y) &= H(U_A - U_A') - \frac{1}{2}H(U_A) - \frac{1}{2}H(U_A') = H(U_A + U_A') - H(U_A) \\ &\leq \log C|A| - \log|A| = \log C, \end{aligned}$$

by Maximality. So $d(X;U_H) \leq \frac{1}{2}\alpha \log C$, i.e.

$$\begin{aligned} H(X + U_H) &\leq \frac{1}{2}H(X) + \frac{1}{2}H(U_H) + \frac{1}{2}\alpha \log C \\ &= \frac{1}{2}\log|A| + \frac{1}{2}\log|H| + \frac{1}{2}\alpha \log C. \end{aligned}$$

Therefore by Lemma 7.5, there exists z such that

$$\mathbb{P}(X + U_H = z) \geq |A|^{-1/2}|H|^{-1/2}C^{-\alpha/2}.$$

But $\mathbb{P}(X + U_H = z) = \frac{|A \cap (z-H)|}{|A||H|} = \frac{|A \cap (z+H)|}{|A||H|}$. So there exists $z \in G$ such that

$$|A \cap (z + H)| \geq C^{-\alpha/2}|A|^{-1/2}|H|^{-1/2}.$$

Let $B = A \cap (z + H)$. Let $B = A \cap (z + H)$. By Ruzsa Covering Lemma, we can cover A by at most $\frac{|A+B|}{|B|}$ translates of $B - B = B + B$. But $B \subseteq z + H$ so $B + B \subseteq 2z + H + H = H$. So A can be covered by at most $\frac{|A+B|}{|B|}$ translates of H. But since $B \subseteq A$, $|A + B| \leq |A + A| \leq C|A|$. So

$$\frac{|A + B|}{|B|} \leq \frac{C|A|}{C^{-\alpha/2}|A|^{1/2}|H|^{1/2}} = C^{\alpha/2+1}\frac{|A|^{1/2}}{|H|^{1/2}}.$$

Since B is contained in $z + H$, $|H| \geq C^{-\alpha/2}|A|^{1/2}|H|^{1/2}$, which implies $|H| \geq C^{-\alpha}|A|$. So

$$C^{\alpha/2+1}\frac{|A|^{1/2}}{|H|^{1/2}} \leq C^{\alpha+1}.$$

If $|H| \leq |A|$, then we are done (with polynomial $p(x) = x^{\alpha+1}$). Otherwise, since $B \subseteq A$, $|A| \geq C^{-\alpha/2}|A|^{1/2}|H|^{1/2}$, which implies $|H| \leq C^{\alpha}|A|$. Pick a subgroup H' of H of size between $|A|/2$ and $|A|$. Then H is a union of $|H|/|H'| \leq 2C^{\alpha}$ translates of H', so A is a union of at most $2C^{2\alpha+1}$ translates of H'. □

Now we reduce further. We shall prove the following statement.

Theorem 7.7 (EPFR') There is an absolute constant $\eta > 0$ such that if X and Y are any two $\mathbb{F}_2^n$-valued RVs, with $d(X;Y) > 0$, then there exist $\mathbb{F}_2^n$-valued RVs U and V such that

$$\tau_{X,Y}(U;V) := d(U;V) + \eta(d(U;X) + d(V;Y)) < d(X;Y).$$

Proposition 7.8 EPFR' with constant η implies EPFR with constant $1/\eta$.

Proof (Hints).

- By compactness, we can find $\mathbb{F}_2^n$-valued RVs U, V such that $\tau_{X,Y}(U;V)$ is minimised.
- Assuming that $d(U;V) \neq 0$, use the Ruzsa Triangle Inequality to derive a contradiction.
- Conclude using Lemma 6.9.

□

Proof. By compactness, we can find $\mathbb{F}_2^n$-valued RVs U, V such that $\tau_{X,Y}(U;V)$ is minimised. If $d(U;V) \neq 0$, then by EPFR', there exist $\mathbb{F}_2^n$-valued RVs Z, W such that $\tau_{UV}(Z;W) < d(U;V)$. But then by the Ruzsa Triangle Inequality,

$$\begin{aligned}
\tau_{X,Y}(Z;W) &= d(Z;W) + \eta(d(Z;X) + d(W;Y)) \\
&\leq d(Z;W) + \eta(d(Z;U) + d(W;V)) + \eta(d(U;X) + d(V;Y)) \\
&< d(U;V) + \eta(d(U;X) + d(V;Y)) \\
&= \tau_{X,Y}(U;V),
\end{aligned}$$

which is a contradiction. It follows that $d(U;V) = 0$. So by Lemma 6.9, there exists H such that U and V are uniform on cosets of H, so

$$\eta(d(U;X) + d(V;Y)) = \eta(d(U_H;X) + d(U_H;Y)) < d(X;Y),$$

since $d(\cdot;\cdot)$ is invariant under constant shifts of either of its arguments. This gives EPFR with constant $1/\eta$. □

Notation 7.9 Write $\tau_{X,Y}(U \mid Z; V \mid W)$ for $\sum_{z,w} \mathbb{P}(Z=z)\mathbb{P}(W=w)\tau_{X,Y}(U \mid Z = z; V \mid W = w)$ and $\tau_{X,Y}(U;V \parallel Z)$ for $\sum_z \mathbb{P}(Z=z)\tau_{X,Y}(U \mid Z=z; V \mid Z=z)$.

Remark 7.10 If we can prove EPFR' for conditioned random variables, then by averaging, we get it for some pair of random variables (e.g. of the form $U \mid Z = z$ and $V \mid W = w$).

Lemma 7.11 (Fibring) Let G and H be abelian groups and let $\varphi : G \to H$ be a homomorphism. Let X, Y be G-valued random variables. Then

$$d(X;Y) = d(\varphi(X);\varphi(Y)) + d(X \mid \varphi(X); Y \mid \varphi(Y)) + I(X - Y : (\varphi(X), \varphi(Y)) \mid \varphi(X) - \varphi(Y)).$$

Proof (Hints).
- May assume WLOG that X and Y are independent.
- Use Lemma 1.13 and Additivity.

□

Proof. We may assume WLOG that X and Y are independent. We have

$$\begin{aligned} d(X;Y) &= H(X-Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y) \\ &= H(\varphi(X) - \varphi(Y)) + H(X - Y \mid \varphi(X) - \varphi(Y)) \\ &\quad -\frac{1}{2}H(\varphi(X)) - \frac{1}{2}H(X \mid \varphi(X)) - \frac{1}{2}H(\varphi(Y)) - \frac{1}{2}H(Y \mid \varphi(Y)) \\ &= d(\varphi(X);\varphi(Y)) + d(X \mid \varphi(X); Y \mid \varphi(Y)) \\ &\quad + H(X - Y \mid \varphi(X) - \varphi(Y)) - H(X - Y \mid \varphi(X), \varphi(Y)) \end{aligned}$$

But the last line equals

$$\begin{aligned} &H(X - Y \mid \varphi(X) - \varphi(Y)) - H(X - Y \mid \varphi(X), \varphi(Y), \varphi(X) - \varphi(Y)) \\ &= I(X - Y : (\varphi(X), \varphi(Y)) \mid \varphi(X) - \varphi(Y)). \end{aligned}$$

□

We shall be interested in the following special case.

Corollary 7.12 Let $G = \mathbb{F}_2^n$ and let X_1, X_2, X_3, X_4 be independent G-valued RVs. Then

$$\begin{aligned} d(X_1;X_3) + d(X_2;X_4) &= d((X_1,X_2);(X_3,X_4)) \\ &= d(X_1+X_2; X_3+X_4) + d(X_1 \mid X_1+X_2; X_3 \mid X_3+X_4) \\ &\quad + I(X_1+X_3, X_2+X_4 : X_1+X_2, X_3+X_4 \mid X_1+X_2+X_3+X_4). \end{aligned}$$

Proof (Hints). Straightforward. □

Proof. The first equality is easy to see. For the second, apply Fibring with $X = (X_1, X_2)$, $Y = (X_3, X_4)$ and $\varphi(x, y) = x + y$. □

Lemma 7.13 Let X_1, X_2, X_3, X_4 be independent $\mathbb{F}_2^n$-valued RVs and $W = X_1 + X_2 + X_3 + X_4$. Then

$$\begin{aligned}
&2d(X_1; X_3) + 2d(X_2; X_4) + d(X_1; X_4) + d(X_2; X_3) \\
\geq\ &2d(X_1 + X_2; X_3 + X_4) + d(X_1 + X_4; X_2 + X_3) \\
&+ 2d(X_1 \mid X_1 + X_2; X_3 \mid X_3 + X_4) + d(X_1 \mid X_1 + X_4; X_2 \mid X_2 + X_3) \\
&+ \frac{1}{3}d(X_1 + X_2; X_1 + X_3 \parallel X_2 + X_3, W) + \frac{1}{3}d(X_1 + X_2; X_1 + X_4 \parallel X_2 + X_4, W) \\
&+ \frac{1}{3}d(X_1 + X_4; X_1 + X_3 \parallel X_3 + X_4, W)
\end{aligned}$$

Proof (Hints). Use Corollary 7.12 and Entropic BSG Theorem on (X_1, X_2, X_3, X_4), (X_1, X_2, X_4, X_3) and (X_1, X_4, X_3, X_2), making heavy use of the observation that if i, j, k, l are some permutation of $1, 2, 3, 4$, then $H\big(X_i + X_j \mid W\big) = H(X_k + X_l \mid W)$. □

Proof. Recall that $d(X; Y \parallel X + Y) \leq 3I(X : Y) + 2H(X + Y) - H(X) - H(Y)$ by the Entropic BSG Theorem. Equivalently, $I(X : Y) \geq \frac{1}{3}(d(X; Y \parallel X + Y) + H(X) + H(Y) - 2H(X + Y))$. Applying this to the mutual information term in Corollary 7.12, we get that it is at least

$$\begin{aligned}
&\frac{1}{3}d(X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4 \parallel X_2 + X_3, W) + \frac{1}{3}H(X_1 + X_3, X_2 + X_4 \mid W) \\
&+\frac{1}{3}H(X_1 + X_2, X_3 + X_4 \mid W) - \frac{2}{3}H(X_2 + X_3, X_2 + X_3 \mid W).
\end{aligned}$$

which simplifies to

$$\begin{aligned}
&\frac{1}{3}d(X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4 \parallel X_2 + X_3, W) \\
&+ \frac{1}{3}H(X_1 + X_3 \mid W) + \frac{1}{3}H(X_1 + X_2 \mid W) - \frac{2}{3}H(X_2 + X_3 \mid W) \\
=\ &\frac{1}{3}d(X_1 + X_2; X_1 + X_3 \parallel X_2 + X_3, W) \\
&+ \frac{1}{3}H(X_1 + X_3 \mid W) + \frac{1}{3}H(X_1 + X_2 \mid W) - \frac{2}{3}H(X_2 + X_3 \mid W)
\end{aligned}$$

Now apply this inequality and Corollary 7.12 to (X_1, X_2, X_3, X_4), (X_1, X_2, X_4, X_3) and (X_1, X_4, X_3, X_2). We look at the first entropy terms in the above expression: we get

$$\begin{aligned}
&2H(X_1 + X_2 \mid W) + H(X_1 + X_4 \mid W) \\
&+H(X_1 + X_3 \mid W) + H(X_1 + X_4 \mid W) + H(X_1 + X_2 \mid W) \\
&-2H(X_2 + X_3 \mid W) - 2H(X_2 + X_4 \mid W) - 2H(X_1 + X_2 \mid W).
\end{aligned}$$

which is equal to 0, where we have made heavy use of the observation that if i, j, k, l are some permutation of $1, 2, 3, 4$, then $H\big(X_i + X_j \mid W\big) = H(X_k + X_l \mid W)$, which also allowed us e.g. to replace

$$d(X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4 \parallel X_2 + X_3, W)$$

by

$$d(X_1 + X_2; X_1 + X_3 \parallel X_2 + X_3, W).$$

□

Lemma 7.14 Let X_1, X_2 be independent copies of X and Y_1, Y_2 be independent copies of Y. Then

$$\begin{aligned} 6d(X;Y) \geq\ & 2d(X_1 + X_2; Y_1 + Y_2) + d(X_1 + Y_2; X_2 + Y_1) \\ & + 2d(X_1 \mid X_1 + X_2; Y_1 \mid Y_1 + Y_2) + d(X_1 \mid X_1 + Y_1; X_2 \mid X_2 + Y_2) \\ & + \frac{2}{3} d(X_1 + X_2; X_1 + Y_1 \parallel X_2 + Y_1, X_1 + Y_2) \\ & + \frac{1}{3} d(X_1 + Y_1; X_1 + Y_2 \parallel X_1 + X_2, Y_1 + Y_2). \end{aligned}$$

Proof (Hints). Straightforward. □

Proof. Apply Lemma 7.13 to (X_1, X_2, Y_1, Y_2) (all independent). We can swap Y_1 and Y_2 in the third last term since they are independent copies. □

Recall that we want U, V such that $\tau_{X,Y}(U, V) < d(X;Y)$. Lemma 7.14 gives us a collection of distances (some conditioned), at least one of which is at most $\frac{6}{7} d(X;Y)$. So it will be enough to show that for all of them, we get $d(U;X) + d(V;Y) \leq C d(X;Y)$ for some absolute constant C. Then we can take $\eta < 1/7C$.

Definition 7.15 We say that (U, V) is ***C*-relevant** to (X, Y) if $d(U;X) + d(V;Y) \leq C d(X;Y)$.

Lemma 7.16 (Y, X) is 2-relevant to (X, Y).

Proof (Hints). Straightforward. □

Proof. $d(Y;X) + d(X;Y) = 2d(X;Y)$. □

Lemma 7.17 Let U, V, X be independent $\mathbb{F}_2^n$-valued RVs. Then

$$d(U + V; X) \leq \frac{1}{2}(d(U;X) + d(V;X) + d(U;V)).$$

Proof (Hints). Write $-\frac{1}{2}H(U+V) = -H(U+V) + \frac{1}{2}H(U+V)$ and $H(U+V+X) = \frac{1}{2}H(U+V+X) + \frac{1}{2}H(U+V+X)$ and use Submodularity for Sums twice. □

Proof. We have

$$d(U + V; X) = H(U + V + X) - \frac{1}{2}H(U + V) - \frac{1}{2}H(X)$$

$$
\begin{aligned}
&= H(U+V+X) - H(U+V) + \frac{1}{2}H(U+V) - \frac{1}{2}H(X) \\
&\leq \frac{1}{2}H(U+X) - \frac{1}{2}H(U) + \frac{1}{2}H(V+X) - \frac{1}{2}H(V) + \frac{1}{2}H(U+V) - \frac{1}{2}H(X) \\
&= \frac{1}{2}(d(U;X) + d(V;X) + d(U;V))
\end{aligned}
$$

by Submodularity for Sums. □

Corollary 7.18 If (U, V) is C-relevant to (X, Y) and U_1, U_2, V_1, V_2 are copies of U, V, then $(U_1 + U_2, V_1 + V_2)$ is $2C$-relevant to (X, Y).

Proof (Hints). Use Lemma 7.17 and Entropic Ruzsa Triangle Inequality. □

Proof.

$$
\begin{aligned}
d(U_1+U_2;X) + d(V_1+V_2;Y) &\leq \frac{1}{2}(2d(U;X) + d(U;U) + 2d(V;Y) + d(V;V)) \\
&\leq 2(d(U;X) + d(V;Y)) \leq 2Cd(X;Y)
\end{aligned}
$$

by Lemma 7.17 and the Ruzsa Triangle Inequality. □

Corollary 7.19 Let X_1, X_2, Y_1, Y_2 be copies of X, Y. Then $(X_1 + X_2, Y_1 + Y_2)$ is 4-relevant to (Y, X).

Proof (Hints). Straightforward. □

Proof. By Lemma 7.16, (X, Y) is 2-relevant to (Y, X) so by Corollary 7.18, we are done. □

Corollary 7.20 If (U, V) is C-relevant to (X, Y), then $(U + V, U + V)$ is $(2C + 2)$-relevant to (X, Y).

Proof (Hints). Use Lemma 7.17 on $d(U + V; X)$, similarly for $d(U + V; Y)$. □

Proof. By Lemma 7.17 and the Ruzsa Triangle Inequality,

$$
\begin{aligned}
d(U+V;X) &\leq \frac{1}{2}(d(U;X) + d(V;X) + d(U;V)) \\
&\leq \frac{1}{2}(d(U;X) + d(V;Y) + d(X;Y) + d(U;X) + d(X;Y) + d(V;Y)) \\
&= d(U;X) + d(V;Y) + d(X;Y),
\end{aligned}
$$

and similarly for $d(U + V; Y)$. □

Lemma 7.21 Let U, V, X be independent $\mathbb{F}_2^n$-valued RVs. Then

$$
d(U \mid U+V;X) \leq \frac{1}{2}(d(U;X) + d(V;X) + d(U;V))
$$

Proof (Hints).

- Show that $d(U \mid U + V; X) \leq H(U + X) - \frac{1}{2}H(U) - \frac{1}{2}H(V) + \frac{1}{2}H(U + V) - \frac{1}{2}H(X)$.

- Show that $d(U \mid U+V; X) = d(V \mid U+V; X)$, and average the above inequality with a similar bound.

□

Proof. We have

$$\begin{aligned} d(U \mid U+V; X) &= H(U+X \mid U+V) - \frac{1}{2}H(U \mid U+V) - \frac{1}{2}H(X) \\ &= H(U+X \mid U+V) - \frac{1}{2}H(U, U+V) + \frac{1}{2}H(U+V) - \frac{1}{2}H(X) \\ &\leq H(U+X) - \frac{1}{2}H(U) - \frac{1}{2}H(V) + \frac{1}{2}H(U+V) - \frac{1}{2}H(X). \end{aligned}$$

But $d(U \mid U+V; X) = d(V \mid U+V; X)$ since $H(U+X, U+V) = H(X-V, U+V) = H(V+X, U+V)$, so also

$$d(U \mid U+V; X) \leq H(V+X) - \frac{1}{2}H(U) - \frac{1}{2}H(V) + \frac{1}{2}H(U+V) - \frac{1}{2}H(X).$$

Averaging the two inequalities gives the result (as earlier). □

Corollary 7.22 Let U, V be independent RVs and suppose that (U, V) is C-relevant to (X, Y). Let U_1, U_2, V_1, V_2 be independent copies of U, V. Then
1. $(U_1 \mid U_1+U_2, V_1 \mid V_1+V_2)$ is $2C$-relevant to (X, Y).
2. $(U_1 \mid U_1+V_1, U_2 \mid U_2+V_2)$ is $(2C+2)$-relevant to (X, Y).

Proof. Use Lemma 7.21. Then as soon as it is used, we are in exactly the situation we were in when bounding the relevance of (U_1+U_2, V_1+V_2) and (U_1+V_1, U_2+V_2) (with Corollary 7.18 and Corollary 7.20). □

It remains to tackle the last two terms in Lemma 7.14. For the fifth term, we need to bound

$$d(X_1+X_2 \mid X_2+Y_1, X_1+Y_2; X) + d(X_1+Y_1 \mid X_2+Y_1, X_1+Y_2; Y).$$

But the first term of this is at most (by Lemma 7.17)

$$\begin{aligned} &\frac{1}{2}d(X_1 \mid X_2+Y_1, X_1+Y_2; X) + \frac{1}{2}d(X_2 \mid X_2+Y_1, X_1+Y_2; X) + \frac{1}{2}d(X_1; X_2 \parallel X_2+Y_1, X_1+Y_2) \\ &\leq d(X_1 \mid X_1+Y_2; X) + d(X_2 \mid X_2+Y_1; X) = 2d(X \mid X+Y; X) \end{aligned}$$

by the Ruzsa Triangle Inequality and independence. Now we can use Lemma 7.21, and similarly for the other terms. In this way, we get that the fifth and sixth terms have relevances bounded above by λC for an absolute constant λ.