# PFR Blueprint

Terence Tao

April 18, 2024

# Chapter 1

# Applications of Jensen's inequality

In this chapter, $h$ denotes the function $h(x) := x \log \frac{1}{x}$ for $x \in [0, 1]$.

**Lemma 1.1** (Concavity)**.** *$h$ is strictly concave on $[0, 1]$.*

*Proof.* Check that $h'$ is strictly monotone decreasing. □

**Lemma 1.2** (Jensen)**.** *If $S$ is a finite set, and $\sum_{s \in S} w_s = 1$ for some non-negative $w_s$, and $p_s \in [0, 1]$ for all $s \in S$, then*

$$\sum_{s \in S} w_s h(p_s) \leq h(\sum_{s \in S} w_s p_s).$$

*Proof.* Apply Jensen and Lemma 1.1. □

**Lemma 1.3** (Converse Jensen)**.** *If equality holds in the above lemma, then $p_s = \sum_{s \in S} w_s h(p_s)$ whenever $w_s \neq 0$.*

*Proof.* Need some converse form of Jensen, not sure if it is already in Mathlib. May also wish to state it as an if and only if. □

# Chapter 2

# Shannon entropy inequalities

Random variables in this paper are measurable maps $X : \Omega \to S$ from a probability space $\Omega$ to a measurable space $S$, and called $S$-valued random variables. In many cases we will assume that singletons in $S$ are measurable. Often we will restrict further to the case when $S$ is finite with the discrete $\sigma$-algebra, which of course implies that $S$ has measurable singletons.

**Definition 2.1** (Entropy). *If $X$ is an $S$-valued random variable, the entropy $\mathbb{H}[X]$ of $X$ is defined*

$$\mathbb{H}[X] := \sum_{s \in S} \mathbb{P}[X = x] \log \frac{1}{\mathbb{P}[X = x]}$$

*with the convention that $0 \log \frac{1}{0} = 0$.*

**Lemma 2.2** (Entropy and relabeling).

*(i) If $X : \Omega \to S$ and $Y : \Omega \to T$ are random variables, and $Y = f(X)$ for some injection $f : S \to T$, then $\mathbb{H}[X] = \mathbb{H}[Y]$.*

*(ii) If $X : \Omega \to S$ and $Y : \Omega \to T$ are random variables, and $Y = f(X)$ and $X = g(Y)$ for some functions $f : S \to T$, $g : T \to S$, then $\mathbb{H}[X] = \mathbb{H}[Y]$.*

*Proof.* Expand out both entropies and rearrange. □

**Lemma 2.3** (Jensen bound). *If $X$ is an $S$-valued random variable, then $\mathbb{H}[X] \leq \log |S|$.*

*Proof.* This is a direct consequence of Lemma 1.2. □

**Definition 2.4** (Uniform distribution). *If $H$ is a subset of $S$, an $S$-random variable $X$ is said to be uniformly distributed on $H$ if $\mathbb{P}[X = s] = 1/|H|$ for $s \in X$ and $\mathbb{P}[X = s] = 0$ otherwise.*

**Lemma 2.5** (Uniform distributions exist). *Given a finite non-empty subset $H$ of a set $S$, there exists a random variable $X$ (on some probability space) that is uniformly distributed on $H$.*

*Proof.* Direct construction. □

**Lemma 2.6** (Entropy of uniform random variable). *If $X$ is $S$-valued random variable, then $\mathbb{H}[X] = \log |S|$ if and only if $X$ is uniformly distributed on $S$.*

*Proof.* Direct computation in one direction. Converse direction needs Lemma 1.3. □

**Lemma 2.7** (Entropy of uniform random variable, II)**.** *If $X$ is uniformly distributed on $H$, then, then $\mathbb{H}[X] = \log |H|$.*

*Proof.* Direct computation. □

**Lemma 2.8** (Bounded entropy implies concentration)**.** *If $X$ is an $S$-valued random variable, then there exists $s \in S$ such that $\mathbb{P}[X = s] \geq \exp(-\mathbb{H}[X])$.*

*Proof.* We have

$$\mathbb{H}[X] = \sum_{s \in S} \mathbb{P}[X = s] \log \frac{1}{\mathbb{P}[X = s]} \geq \min_{s \in S} \log \frac{1}{\mathbb{P}[X = s]}$$

and the claim follows. □

We use $X, Y$ to denote the pair $\omega \mapsto (X(\omega), Y(\omega))$.

**Lemma 2.9** (Commutativity and associativity of joint entropy)**.** *If $X : \Omega \to S$, $Y : \Omega \to T$, and $Z : \Omega \to U$ are random variables, then $\mathbb{H}[X, Y] = \mathbb{H}[Y, X]$ and $\mathbb{H}[X, (Y, Z)] = \mathbb{H}[(X, Y), Z]$.*

*Proof.* Set up an injection from $(X, Y)$ to $(Y, X)$ and use Lemma 2.2 for the first claim. Similarly for the second claim. □

**Definition 2.10** (Conditioned event)**.** *If $X : \Omega \to S$ is an $S$-valued random variable and $E$ is an event in $\Omega$, then the conditioned event $(X|E)$ is defined to be the same random variable as $X$, but now the ambient probability measure has been conditioned to $E$.*

Note: it may happen that $E$ has zero measure. In which case, the ambient probability measure should be replaced with a zero measure. (In our formalization we achieve this by working with arbitrary measures, and normalizing them to be probability measures if possible, else using the zero measure. Conditioning is also formalized using existing Mathlib definitions.)

**Definition 2.11** (Conditional entropy)**.** *If $X : \Omega \to S$ and $Y : \Omega \to T$ are random variables, the conditional entropy $\mathbb{H}[X|Y]$ is defined as*

$$\mathbb{H}[X|Y] := \sum_{y \in Y} \mathbb{P}[Y = y] \mathbb{H}[(X|Y = y)].$$

**Lemma 2.12** (Conditional entropy and relabeling)**.** *If $X : \Omega \to S$, $Y : \Omega \to T$, and $Z : \Omega \to U$ are random variables, and $Y = f(X, Z)$ almost surely for some map $f : S \times U \to T$ that is injective for each fixed $U$, then $\mathbb{H}[X|Z] = \mathbb{H}[Y|Z]$.*

*Similarly, if $g : T \to U$ is injective, then $\mathbb{H}[X|g(Y)] = \mathbb{H}[X|Y]$.*

*Proof.* For the first part, use Definition 2.11 and then Lemma 2.2. The second part is a direct computation. □

**Lemma 2.13** (Chain rule)**.** *If $X : \Omega \to S$ and $Y : \Omega \to T$ are random variables, then*

$$\mathbb{H}[X, Y] = \mathbb{H}[Y] + \mathbb{H}[X|Y].$$

*Proof.* Direct computation. □

**Lemma 2.14** (Conditional chain rule)**.** *If $X : \Omega \to S$, $Y : \Omega \to T$, $Z : \Omega \to U$ are random variables, then*

$$\mathbb{H}[X, Y|Z] = \mathbb{H}[Y|Z] + \mathbb{H}[X|Y, Z].$$

*Proof.* For each $z \in U$, we can apply Lemma 2.13 to the random variables $(X|Z = z)$ and $(Y|Z = z)$ to obtain

$$\mathbb{H}[(X|Z = z), (Y|Z = z)] = \mathbb{H}[Y|Z = z] + \mathbb{H}[(X|Z = z)|(Y|Z = z)].$$

Now multiply by $\mathbb{P}[Z = z]$ and sum. Some helper lemmas may be needed to get to the form above. This sort of "average over conditioning" argument to get conditional entropy inequalities from unconditional ones is commonly used in this paper. □

**Definition 2.15** (Mutual information)**.** *If $X : \Omega \to S$, $Y : \Omega \to T$ are random variables, then*

$$\mathbb{I}[X : Y] := \mathbb{H}[X] + \mathbb{H}[Y] - \mathbb{H}[X, Y].$$

**Lemma 2.16** (Alternative formulae for mutual information)**.** *With notation as above, we have*

$$\mathbb{I}[X : Y] = \mathbb{I}[Y : X]$$

$$\mathbb{I}[X : Y] = \mathbb{H}[X] - \mathbb{H}[X|Y]$$

$$\mathbb{I}[X : Y] = \mathbb{H}[Y] - \mathbb{H}[Y|X]$$

*Proof.* Immediate from Lemmas 2.9, 2.13. □

**Lemma 2.17** (Nonnegativity of mutual information)**.** *We have $\mathbb{I}[X : Y] \geq 0$.*

*Proof.* An application of Lemma 1.2 and Lemma 2.16. □

**Corollary 2.18** (Subadditivity)**.** *With notation as above, we have $\mathbb{H}[X, Y] \leq \mathbb{H}[X] + \mathbb{H}[Y]$.*

*Proof.* Use Lemma 2.17. □

**Corollary 2.19** (Conditioning reduces entropy)**.** *With notation as above, we have $\mathbb{H}[X|Y] \leq \mathbb{H}[X]$.*

*Proof.* Combine Lemma 2.17 with Lemma 2.16. □

**Corollary 2.20** (Submodularity)**.** *With three random variables $X, Y, Z$, one has $\mathbb{H}[X|Y, Z] \leq \mathbb{H}[X|Z]$.*

*Proof.* Apply the "averaging over conditioning" argument to Corollary 2.19. □

**Corollary 2.21** (Alternate form of submodularity)**.** *With three random variables $X, Y, Z$, one has*

$$\mathbb{H}[X, Y, Z] + \mathbb{H}[Z] \leq \mathbb{H}[X, Z] + \mathbb{H}[Y, Z].$$

*Proof.* Apply Corollary 2.20 and Lemma 2.13. □

**Definition 2.22** (Independent random variables)**.** *Two random variables $X : \Omega \to S$ and $Y : \Omega \to T$ are independent if the law of $(X, Y)$ is the product of the law of $X$ and the law of $Y$. Similarly for more than two variables.*

**Lemma 2.23** (Vanishing of mutual information)**.** *If $X, Y$ are random variables, then $\mathbb{I}[X : Y] = 0$ if and only if $X, Y$ are independent.*

*Proof.* An application of the equality case of Jensen's inequality, Lemma 1.3. □

**Corollary 2.24** (Additivity of entropy)**.** *If $X, Y$ are random variables, then $\mathbb{H}[X, Y] = \mathbb{H}[X] + \mathbb{H}[Y]$ if and only if $X, Y$ are independent.*

*Proof.* Direct from Lemma 2.23. □

**Definition 2.25** (Conditional mutual information)**.** *If $X, Y, Z$ are random variables, with $Z$ $U$-valued, then*

$$\mathbb{I}[X : Y|Z] := \sum_{z \in U} P[Z = z] \mathbb{I}[(X|Z = z) : (Y|Z = z)].$$

**Lemma 2.26** (Alternate formula for conditional mutual information)**.** *We have*

$$\mathbb{I}[X : Y|Z] := \mathbb{H}[X|Z] + \mathbb{H}[Y|Z] - \mathbb{H}[X, Y|Z]$$

*and*

$$\mathbb{I}[X : Y|Z] := \mathbb{H}[X|Z] - \mathbb{H}[X|Y, Z].$$

*Proof.* Routine computation. □

**Lemma 2.27** (Nonnegativity of conditional mutual information)**.** *If $X, Y, Z$ are random variables, then $\mathbb{I}[X : Y|Z] \geq 0$.*

*Proof.* Use Definition 2.25 and Lemma 2.20. □

**Definition 2.28** (Conditionally independent random variables)**.** *Two random variables $X : \Omega \to S$ and $Y : \Omega \to T$ are conditionally independent relative to another random variable $Z : \Omega \to U$ if $P[X = s \wedge Y = t|Z = u] = P[X = s|Z = u]P[Y = t|Z = u]$ for all $s \in S, t \in T, u \in U$. (We won't need conditional independence for more variables than this.)*

**Lemma 2.29** (Vanishing conditional mutual information)**.** *If $X, Y, Z$ are random variables, then $\mathbb{I}[X : Y|Z] = 0$ iff $X, Y$ are conditionally independent over $Z$.*

*Proof.* Immediate from Lemma 2.23 and Definition 2.28. □

**Corollary 2.30** (Entropy of conditionally independent variables)**.** *If $X, Y$ are conditionally independent over $Z$, then*

$$\mathbb{H}[X, Y, Z] = \mathbb{H}[X, Z] + \mathbb{H}[Y, Z] - \mathbb{H}[Z].$$

*Proof.* Immediate from Lemma 2.29 and Lemma 2.26. □

# Chapter 3

# Entropic Ruzsa calculus

In this section $G$ will be a finite additive group. (May eventually want to generalize to infinite $G$.)

**Lemma 3.1** (Negation preserves entropy). *If $X$ is $G$-valued, then $\mathbb{H}[-X] = \mathbb{H}[X]$.*

*Proof.* Immediate from Lemma 2.2. □

**Lemma 3.2** (Shearing preserves entropy). *If $X, Y$ are $G$-valued, then $\mathbb{H}[X \pm Y|Y] = \mathbb{H}[X|Y]$ and $\mathbb{H}[X \pm Y, Y] = \mathbb{H}[X, Y]$.*

*Proof.* Immediate from Lemma 2.12 and Lemma 2.13. □

**Lemma 3.3** (Lower bound of sumset). *If $X, Y$ are $G$-valued random variables on $\Omega$, we have*

$$\max(\mathbb{H}[X], \mathbb{H}[Y]) - \mathbb{I}[X : Y] \leq \mathbb{H}[X \pm Y].$$

*Proof.* By Lemma 2.19, 3.2, 2.16, 3.1 we have

$$\mathbb{H}[X \pm Y] \geq \mathbb{H}[X \pm Y|Y] = \mathbb{H}[X|Y] = \mathbb{H}[X] - \mathbb{I}[X : Y]$$

and similarly with the roles of $X, Y$ reversed, giving the claim. □

**Corollary 3.4** (Conditional lower bound on sumset). *If $X, Y$ are $G$-valued random variables on $\Omega$ and $Z$ is another random variable on $\Omega$ then*

$$\max(\mathbb{H}[X|Z], \mathbb{H}[Y|Z]) - \mathbb{I}[X : Y|Z] \leq \mathbb{H}[X \pm Y|Z],$$

*Proof.* This follows from Lemma 3.3 by conditioning to $Z = z$ and summing over $z$ (weighted by $\mathbb{P}[Z = z]$). □

**Corollary 3.5** (Independent lower bound on sumset). *If $X, Y$ are independent $G$-valued random variables, then*

$$\max(\mathbb{H}[X], \mathbb{H}[Y]) \leq \mathbb{H}[X \pm Y].$$

*Proof.* Combine Lemma 3.3 with Lemma 2.23. □

One random variable is said to be a copy of another if they have the same distribution.

**Lemma 3.6** (Copy preserves entropy). *If $X'$ is a copy of $X$ then $\mathbb{H}[X'] = \mathbb{H}[X]$.*

*Proof.* Immediate from Definition 2.1. □

**Lemma 3.7** (Existence of independent copies). *Let $X_i : \Omega_i \to S_i$ be random variables for $i = 1, \dots, k$. Then if one gives $\prod_{i=1}^k S_i$ the product measure of the laws of $X_i$, the coordinate functions $(x_j)_{j=1}^k \mapsto x_i$ are jointly independent random variables which are copies of the $X_1, \dots, X_k$.*

*Proof.* Explicit computation. □

**Definition 3.8** (Ruzsa distance). *Let $X, Y$ be $G$-valued random variables (not necessarily on the same sample space). The* Ruzsa distance *$d[X; Y]$ between $X$ and $Y$ is defined to be*

$$d[X; Y] := \mathbb{H}[X' - Y'] - \mathbb{H}[X']/2 - \mathbb{H}[Y']/2$$

*where $X', Y'$ are (the canonical) independent copies of $X, Y$ from Lemma 3.7.*

**Lemma 3.9** (Distance from zero). *If $X$ is a $G$-valued random variable and $0$ is the random variable taking the value $0$ everywhere then*

$$d[X; 0] = \mathbb{H}(X)/2.$$

*Proof.* This is an immediate consequence of the definitions and $X - 0 \equiv X$ and $\mathbb{H}(0) = 0$. □

**Lemma 3.10** (Copy preserves Ruzsa distance). *If $X', Y'$ are copies of $X, Y$ respectively then $d[X'; Y'] = d[X; Y]$.*

*Proof.* Immediate from Definitions 3.8 and Lemma 3.6. □

**Lemma 3.11** (Ruzsa distance in independent case). *If $X, Y$ are independent $G$-random variables then*

$$d[X; Y] := \mathbb{H}[X - Y] - \mathbb{H}[X]/2 - \mathbb{H}[Y]/2.$$

*Proof.* Immediate from Definition 3.8 and Lemmas 2.2, 3.6. □

**Lemma 3.12** (Distance symmetric). *If $X, Y$ are $G$-valued random variables, then*

$$d[X; Y] = d[Y; X].$$

*Proof.* Immediate from Lemma 3.1 and Definition 3.8. □

**Lemma 3.13** (Distance controls entropy difference). *If $X, Y$ are $G$-valued random variables, then*

$$|\mathbb{H}[X] - H[Y]| \leq 2d[X; Y].$$

*Proof.* Immediate from Lemma 3.5 and Definition 3.8, and also Lemma 3.1. □

**Lemma 3.14** (Distance controls entropy growth). *If $X, Y$ are independent $G$-valued random variables, then*

$$\mathbb{H}[X - Y] - \mathbb{H}[X], \mathbb{H}[X - Y] - \mathbb{H}[Y] \leq 2d[X; Y].$$

*Proof.* Immediate from Lemma 3.5 and Definition 3.8, and also Lemma 3.1. □

**Lemma 3.15** (Distance nonnegative). *If $X, Y$ are $G$-valued random variables, then*

$$d[X; Y] \geq 0.$$

*Proof.* Immediate from Lemma 3.13. □

**Lemma 3.16** (Projection entropy and distance)**.** *If $G$ is an additive group and $X$ is a $G$-valued random variable and $H \leq G$ is a finite subgroup then, with $\pi : G \to G/H$ the natural homomorphism we have (where $U_H$ is uniform on $H$)*

$$\mathbb{H}(\pi(X)) \leq 2d[X; U_H].$$

*Proof.* WLOG, we make $X$, $U_H$ independent (Lemma 3.7). Now by Lemmas 2.20, 3.2, 2.3

$$\begin{aligned} \mathbb{H}(X - U_H|\pi(X)) &\geq \mathbb{H}(X - U_H|X) && = \mathbb{H}(U_H) \\ \mathbb{H}(X - U_H|\pi(X)) &\leq \log |H| && = \mathbb{H}(U_H) \end{aligned}$$

By Lemma 2.13

$$\mathbb{H}(X - U_H) = \mathbb{H}(\pi(X)) + \mathbb{H}(X - U_H|\pi(X)) = \mathbb{H}(\pi(X)) + \mathbb{H}(U_H)$$

and therefore

$$d[X; U_H] = \mathbb{H}(\pi(X)) + \frac{\mathbb{H}(U_H) - \mathbb{H}(X)}{2}.$$

Furthermore by Lemma 3.13

$$d[X; U_H] \geq \frac{|\mathbb{H}(X) - \mathbb{H}(U_H)|}{2}.$$

Adding these inequalities gives the result. □

**Lemma 3.17** (Improved Ruzsa triangle inequality)**.** *If $X, Y, Z$ are $G$-valued random variables on $\Omega$ with $(X, Y)$ independent of $Z$, then*

$$\mathbb{H}[X - Y] \leq \mathbb{H}[X - Z] + \mathbb{H}[Z - Y] - \mathbb{H}[Z] \tag{3.1}$$

This is an improvement over the usual Ruzsa triangle inequality because $X, Y$ are not assumed to be independent. However we will not utilize this improvement here.

*Proof.* Apply Corollary 2.21 to obtain

$$\mathbb{H}[X - Z, X - Y] + \mathbb{H}[Y, X - Y] \geq \mathbb{H}[X - Z, Y, X - Y] + \mathbb{H}[X - Y].$$

Using

$$\mathbb{H}[X - Z, X - Y] \leq \mathbb{H}[X - Z] + \mathbb{H}[Y - Z]$$

(from Lemma 2.2, Lemma 2.18),

$$\mathbb{H}[Y, X - Y] = \mathbb{H}[X, Y]$$

(from Lemma 2.2), and

$$\mathbb{H}[X - Z, Y, X - Y] = \mathbb{H}[X, Y, Z] = \mathbb{H}[X, Y] + \mathbb{H}[Z]$$

(from Lemma 2.2 and Lemma 2.24) and rearranging, we indeed obtain (3.1). □

**Lemma 3.18** (Ruzsa triangle inequality)**.** *If $X, Y, Z$ are $G$-valued random variables, then*

$$d[X; Y] \leq d[X; Z] + d[Z; Y].$$

*Proof.* By Lemma 3.10 and Lemmas 3.7, 3.11, it suffices to prove this inequality assuming that $X, Y, Z$ are defined on the same space and are independent. But then the claim follows from Lemma 3.17 and Definition 3.8. □

**Definition 3.19** (Conditioned Ruzsa distance). *If $(X, Z)$ and $(Y, W)$ are random variables (where $X$ and $Y$ are $G$-valued) we define*

$$d[X|Z;Y|W] := \sum_{z,w} \mathbb{P}[Z = z]\mathbb{P}[W = w]d[(X|Z = z); (Y|(W = w))].$$

*similarly*

$$d[X;Y|W] := \sum_{w} \mathbb{P}[W = w]d[X; (Y|(W = w))].$$

**Lemma 3.20** (Alternate form of distance). *The expression $d[X|Z;Y|W]$ is unchanged if $(X, Z)$ or $(Y, W)$ is replaced by a copy. Furthermore, if $(X, Z)$ and $(Y, W)$ are independent, then*

$$d[X|Z;Y|W] = \mathbb{H}[X - Y|Z, W] - \mathbb{H}[X|Z]/2 - \mathbb{H}[Y|W]/2$$

*and similarly*

$$d[X;Y|W] = \mathbb{H}[X - Y|W] - \mathbb{H}[X]/2 - \mathbb{H}[Y|W]/2.$$

*Proof.* Straightforward thanks to Lemma 3.6, Lemma 3.10, Lemma 3.11, Definition 3.19, Definition 2.11. □

**Lemma 3.21** (Kaimanovich-Vershik-Madiman inequality). *Suppose that $X, Y, Z$ are independent $G$-valued random variables. Then*

$$\mathbb{H}[X + Y + Z] - \mathbb{H}[X + Y] \leq \mathbb{H}[Y + Z] - \mathbb{H}[Y].$$

*Proof.* From Lemma 2.20 we have

$$\mathbb{H}[X, X + Y + Z] + \mathbb{H}[Z, X + Y + Z] \geq \mathbb{H}[X, Z, X + Y + Z] + \mathbb{H}[X + Y + Z].$$

However, using Lemmas 2.24, 2.2 repeatedly we have $\mathbb{H}[X, X + Y + Z] = \mathbb{H}[X, Y + Z] = \mathbb{H}[X]+\mathbb{H}[Y+Z]$, $\mathbb{H}[Z, X+Y+Z] = \mathbb{H}[Z, X+Y] = \mathbb{H}[Z]+\mathbb{H}[X+Y]$ and $\mathbb{H}[X, Z, X+Y+Z] = \mathbb{H}[X, Y, Z] = \mathbb{H}[X] + \mathbb{H}[Y] + \mathbb{H}[Z]$. The claim then follows from a calculation. □

**Lemma 3.22** (Existence of conditional independent trials). *For $X, Y$ random variables, there exist random variables $X_1, X_2, Y'$ on a common probability space with $(X_1, Y'), (X_2, Y')$ both having the distribution of $(X, Y)$, and $X_1, X_2$ conditionally independent over $Y'$ in the sense of Definition 2.28.*

*Proof.* Explicit construction. □

**Lemma 3.23** (Balog-Szemerédi-Gowers). *Let $A, B$ be $G$-valued random variables on $\Omega$, and set $Z := A + B$. Then*

$$\sum_{z} \mathbb{P}[Z = z]d[(A|Z = z); (B|Z = z)] \leq 3\mathbb{I}[A : B] + 2\mathbb{H}[Z] - \mathbb{H}[A] - \mathbb{H}[B]. \qquad (3.2)$$

*Proof.* Let $(A_1, B_1)$ and $(A_2, B_2)$ (and $Z'$, which by abuse of notation we call $Z$) be conditionally independent trials of $(A, B)$ relative to $Z$ as produced by Lemma 3.22, thus $(A_1, B_1)$ and $(A_2, B_2)$ are coupled through the random variable $A_1 + B_1 = A_2 + B_2$, which by abuse of notation we shall also call $Z$.

Observe from Lemma 3.11 that the left-hand side of (3.2) is

$$\mathbb{H}[A_1 - B_2|Z] - \mathbb{H}[A_1|Z]/2 - \mathbb{H}[B_2|Z]/2. \tag{3.3}$$

since, crucially, $(A_1|Z = z)$ and $(B_2|Z = z)$ are independent for all $z$.

Applying submodularity (Lemma 2.21) gives

$$\begin{aligned}\mathbb{H}[A_1 - B_2] &+ \mathbb{H}[A_1 - B_2, A_1, B_1] \\ &\leq \mathbb{H}[A_1 - B_2, A_1] + \mathbb{H}[A_1 - B_2, B_1].\end{aligned} \tag{3.4}$$

We estimate the second, third and fourth terms appearing here. First note that, by Lemma 2.30 and Lemma 2.2 (noting that the tuple $(A_1 - B_2, A_1, B_1)$ determines the tuple $(A_1, A_2, B_1, B_2)$ since $A_1 + B_1 = A_2 + B_2$)

$$\mathbb{H}[A_1 - B_2, A_1, B_1] = \mathbb{H}[A_1, B_1, A_2, B_2, Z] = 2\mathbb{H}[A, B] - \mathbb{H}[Z]. \tag{3.5}$$

Next observe that

$$\mathbb{H}[A_1 - B_2, A_1] = \mathbb{H}[A_1, B_2] \leq \mathbb{H}[A] + \mathbb{H}[B]. \tag{3.6}$$

Finally, we have

$$\mathbb{H}[A_1 - B_2, B_1] = \mathbb{H}[A_2 - B_1, B_1] = \mathbb{H}[A_2, B_1] \leq \mathbb{H}[A] + \mathbb{H}[B]. \tag{3.7}$$

Substituting (3.5), (3.6) and (3.7) into (3.4) yields

$$\mathbb{H}[A_1 - B_2] \leq 2\mathbb{I}[A : B] + \mathbb{H}[Z]$$

and so by Corollary 2.19

$$\mathbb{H}[A_1 - B_2|Z] \leq 2\mathbb{I}[A : B] + \mathbb{H}[Z].$$

Since

$$\begin{aligned}\mathbb{H}[A_1|Z] &= \mathbb{H}[A_1, A_1 + B_1] - \mathbb{H}[Z] \\ &= \mathbb{H}[A, B] - \mathbb{H}[Z] \\ &= \mathbb{H}[Z] - \mathbb{I}[A : B] - 2\mathbb{H}[Z] - \mathbb{H}[A] - \mathbb{H}[B]\end{aligned}$$

and similarly for $\mathbb{H}[B_2|Z]$, we see that (3.3) is bounded by $3\mathbb{I}[A : B] + 2\mathbb{H}[Z] - \mathbb{H}[A] - \mathbb{H}[B]$ as claimed. □

**Lemma 3.24** (Upper bound on conditioned Ruzsa distance)**.** *Suppose that $(X, Z)$ and $(Y, W)$ are random variables, where $X, Y$ take values in an abelian group. Then*

$$d[X|Z; Y|W] \leq d[X; Y] + \tfrac{1}{2}\mathbb{I}[X : Z] + \tfrac{1}{2}\mathbb{I}[Y : W].$$

*In particular,*

$$d[X; Y|W] \leq d[X; Y] + \tfrac{1}{2}\mathbb{I}[Y : W].$$

*Proof.* Using Lemma 3.20 and Lemma 3.7, if $(X', Z'), (Y', W')$ are independent copies of the variables $(X, Z)$, $(Y, W)$, we have

$$\begin{aligned} d[X|Z;Y|W] &= \mathbb{H}[X' - Y'|Z', W'] - \tfrac{1}{2}\mathbb{H}[X'|Z'] - \tfrac{1}{2}H[Y'|W'] \\ &\leq \mathbb{H}[X' - Y'] - \tfrac{1}{2}\mathbb{H}[X'|Z'] - \tfrac{1}{2}H[Y'|W'] \\ &= d[X';Y'] + \tfrac{1}{2}\mathbb{I}[X' : Z'] + \tfrac{1}{2}\mathbb{I}[Y' : W']. \end{aligned}$$

Here, in the middle step we used Lemma 2.19, and in the last step we used Definition 3.8 and Definition 2.15. □

**Lemma 3.25** (Comparison of Ruzsa distances, I)**.** *Let $X, Y, Z$ be random variables taking values in some abelian group of characteristic* 2*, and with $Y, Z$ independent. Then we have*

$$\begin{aligned} d[X;Y+Z] - d[X;Y] &\leq \tfrac{1}{2}(\mathbb{H}[Y+Z] - \mathbb{H}[Y]) \\ &= \tfrac{1}{2}d[Y;Z] + \tfrac{1}{4}\mathbb{H}[Z] - \tfrac{1}{4}\mathbb{H}[Y]. \end{aligned} \tag{3.8}$$

*and*

$$\begin{aligned} d[X;Y|Y+Z] - d[X;Y] &\leq \tfrac{1}{2}(\mathbb{H}[Y+Z] - \mathbb{H}[Z]) \\ &= \tfrac{1}{2}d[Y;Z] + \tfrac{1}{4}\mathbb{H}[Y] - \tfrac{1}{4}\mathbb{H}[Z]. \end{aligned} \tag{3.9}$$

*Proof.* We first prove (3.8). We may assume (taking an independent copy, using Lemma 3.7 and Lemma 3.10, 3.11) that $X$ is independent of $Y, Z$. Then we have

$$\begin{aligned} &d[X;Y+Z] - d[X;Y] \\ &\qquad = \mathbb{H}[X+Y+Z] - \mathbb{H}[X+Y] - \tfrac{1}{2}\mathbb{H}[Y+Z] + \tfrac{1}{2}\mathbb{H}[Y]. \end{aligned}$$

Combining this with Lemma 3.21 gives the required bound. The second form of the result is immediate Lemma 3.11.

Turning to (3.9), we have from Definition 2.15 and Lemma 2.2

$$\begin{aligned} \mathbb{I}[Y : Y+Z] &= \mathbb{H}[Y] + \mathbb{H}[Y+Z] - \mathbb{H}[Y, Y+Z] \\ &= \mathbb{H}[Y] + \mathbb{H}[Y+Z] - \mathbb{H}[Y, Z] = \mathbb{H}[Y+Z] - \mathbb{H}[Z], \end{aligned}$$

and so (3.9) is a consequence of Lemma 3.24. Once again the second form of the result is immediate from Lemma 3.11. □

**Lemma 3.26** (Comparison of Ruzsa distances, II)**.** *Let $X, Y, Z, Z'$ be random variables taking values in some abelian group, and with $Y, Z, Z'$ independent. Then we have*

$$\begin{aligned} &d[X;Y+Z|Y+Z+Z'] - d[X;Y] \\ &\qquad \leq \tfrac{1}{2}(\mathbb{H}[Y+Z+Z'] + \mathbb{H}[Y+Z] - \mathbb{H}[Y] - \mathbb{H}[Z']). \end{aligned} \tag{3.10}$$

*Proof.* By Lemma 3.25 (with a change of variables) we have

$$d[X;Y+Z|Y+Z+Z'] - d[X;Y+Z] \leq \tfrac{1}{2}(\mathbb{H}[Y+Z+Z'] - \mathbb{H}[Z']).$$

Adding this to (3.8) gives the result. □

# Chapter 4

# The 100% version of PFR

**Definition 4.1** (Symmetry group)**.** *If $X$ is a $G$-valued random variable, then the symmetry group* $\mathrm{Sym}[X]$ *is the set of all $h \in G$ such that $X + h$ has the same distribution as $X$.*

**Lemma 4.2** (Symmetry group is a group)**.** *If $X$ is a $G$-valued random variable, then* $\mathrm{Sym}[X]$ *is a subgroup of $G$.*

*Proof.* Direct verification of the group axioms. □

**Lemma 4.3** (Zero Ruzsa distance implies large symmetry group)**.** *If $X$ is a $G$-valued random variable such that $d[X;X] = 0$, and $x, y \in G$ are such that $P[X = x], P[X = y] > 0$, then $x - y \in$* $\mathrm{Sym}[X]$*.*

*Proof.* Let $X_1, X_2$ be independent copies of $X$ (from Lemma 3.7). Let $A$ denote the range of $X$. From Lemma 3.11 and Lemma 3.10 we have

$$\mathbb{H}[X_1 - X_2] = \mathbb{H}[X_1].$$

Observe from Lemma 2.12 that

$$\mathbb{H}[X_1 - X_2 | X_2] = \mathbb{H}[X_1 | X_2] = \mathbb{H}[X_1]$$

and hence by Lemma 2.16

$$\mathbb{I}[X_1 - X_2 : X_1] = 0.$$

By Corollary 2.23, $X_1 - X_2$ and $X_1$ are therefore independent, thus the law of $(X_1 - X_2 | X_1 = x)$ does not depend on $x \in A$. The claim follows. □

**Lemma 4.4** (Translate is uniform on symmetry group)**.** *If $X$ is a $G$-valued random variable with $d[X;X] = 0$, and $x_0$ is a point with $P[X = x_0] > 0$, then $X - x_0$ is uniformly distributed on* $\mathrm{Sym}[X]$*.*

*Proof.* The law of $X - x_0$ is invariant under $\mathrm{Sym}[X]$, non-zero at the origin, and supported on $\mathrm{Sym}[X]$, giving the claim. □

**Lemma 4.5** (Symmetric 100% inverse theorem)**.** *Suppose that $X$ is a $G$-valued random variable such that $d[X;X] = 0$. Then there exists a subgroup $H \leq G$ such that $d[X;U_H] = 0$.*

*Proof.* Take $H$ to be the symmetry group of $X$, which is a group by Lemma 4.2. From Lemma 4.4, $X - x_0$ is uniform on $H$, and $d[X; X - x_0] = d[X;X] \leq 0$, and the claim follows. □

**Corollary 4.6** (General 100% inverse theorem). *Suppose that $X_1, X_2$ are $G$-valued random variables such that $d[X_1; X_2] = 0$. Then there exists a subgroup $H \leq G$ such that $d[X_1; U_H] = d[X_2; U_H] = 0$.*

*Proof.* Using Lemma 3.18 and Lemma 3.15 we have $d[X_1; X_1] = 0$, hence by Lemma 4.5 $d[X_1; U_H] = 0$ for some subgroup $H$. By Lemma 3.18 and Lemma 3.15 again we also have $d[X_2; U_H]$ as required. □

# Chapter 5

# The Fibring lemma

**Proposition 5.1** (General fibring identity)**.** *Let $\pi : H \to H'$ be a homomorphism additive groups, and let $Z_1, Z_2$ be $H$-valued random variables. Then we have*

$$d[Z_1; Z_2] \geq d[\pi(Z_1); \pi(Z_2)] + d[Z_1|\pi(Z_1); Z_2|\pi(Z_2)].$$

*Moreover, if $Z_1, Z_2$ are taken to be independent, then the difference between the two sides is*

$$I(Z_1 - Z_2 : (\pi(Z_1), \pi(Z_2))|\pi(Z_1 - Z_2)).$$

*Proof.* Let $Z_1, Z_2$ be independent throughout (this is possible by Lemma 3.10 and Lemma 3.7). By Lemma 3.20, We have

$$\begin{aligned}
&d[Z_1|\pi(Z_1); Z_2|\pi(Z_2)] \\
&= \mathbb{H}[Z_1 - Z_2|\pi(Z_1), \pi(Z_2)] - \tfrac{1}{2}\mathbb{H}[Z_1|\pi(Z_1)] - \tfrac{1}{2}\mathbb{H}[Z_2|\pi(Z_2)] \\
&\leq \mathbb{H}[Z_1 - Z_2|\pi(Z_1 + Z_2)] - \tfrac{1}{2}\mathbb{H}[Z_1|\pi(Z_1)] - \tfrac{1}{2}H[Z_2|\pi(Z_2)] \\
&= d[Z_1; Z_2] - d[\pi(Z_1); \pi(Z_2)].
\end{aligned}$$

In the middle step, we used Lemma 2.20, and in the last step we used the fact that

$$\mathbb{H}[Z_1 - Z_2|\pi(Z_1 - Z_2)] = \mathbb{H}[Z_1 - Z_2] - \mathbb{H}[\pi(Z_1 - Z_2)]$$

(thanks to Lemma 2.13 and Lemma 2.2) and that

$$\mathbb{H}[Z_i|\pi(Z_i)] = \mathbb{H}[Z_i] - \mathbb{H}[\pi(Z_i)]$$

(since $Z_i$ determines $\pi(Z_i)$). This gives the claimed inequality. The difference between the two sides is precisely

$$\mathbb{H}[Z_1 - Z_2|\pi(Z_1 - Z_2)] - \mathbb{H}[Z_1 - Z_2|\pi(Z_1), \pi(Z_2)].$$

To rewrite this in terms of (conditional) mutual information, we use the identity

$$\mathbb{H}[A|B] - \mathbb{H}[A|B, C] = \mathbb{I}[A : C|B],$$

(which follows Lemma 2.26) taking $A := Z_1 - Z_2$, $B := \pi(Z_1 - Z_2)$ and $C := (\pi(Z_1), \pi(Z_2))$, and noting that in this case $\mathbb{H}[A|B, C] = \mathbb{H}[A|C]$ since $C$ uniquely determines $B$ (this may require another helper lemma about entropy). This completes the proof. □

**Corollary 5.2.** *If $\pi : G \to H$ is a homomorphism of additive groups and $X, Y$ are $G$-valued random variables then*

$$d[X;Y] \geq d[\pi(X);\pi(Y)].$$

*Proof.* By Proposition 5.1 and the nonnegativity of conditional Ruzsa distance (from Lemma 3.15) we have

$$d[X;Y] \geq d[\pi(X);\pi(Y)] + d[X \mid \pi(X); Y \mid \pi(Y)].$$

The inequality follows from $d[X \mid \pi(X); Y \mid \pi(Y)] \geq 0$ (Lemma 3.15). □

**Corollary 5.3** (Specific fibring identity). *Let $Y_1, Y_2, Y_3$ and $Y_4$ be independent $G$-valued random variables. Then*

$$\begin{aligned} &d[Y_1 + Y_3; Y_2 + Y_4] + d[Y_1|Y_1 + Y_3; Y_2|Y_2 + Y_4] \\ &\quad + \mathbb{I}[Y_1 + Y_2 : Y_2 + Y_4|Y_1 + Y_2 + Y_3 + Y_4] = d[Y_1;Y_2] + d[Y_3;Y_4]. \end{aligned}$$

*Proof.* We apply Proposition 5.1 with $H := G \times G$, $H' := G$, $\pi$ the addition homomorphism $\pi(x, y) := x + y$, and with the random variables $Z_1 := (Y_1, Y_3)$ and $Z_2 := (Y_2, Y_4)$. Then by independence (Lemma 2.24)

$$d[Z_1; Z_2] = d[Y_1; Y_2] + d[Y_3; Y_4]$$

while by definition

$$d[\pi(Z_1); \pi(Z_2)] = d[Y_1 + Y_3; Y_2 + Y_4].$$

Furthermore,

$$d[Z_1|\pi(Z_1); Z_2|\pi(Z_2)] = d[Y_1|Y_1 + Y_3; Y_2|Y_2 + Y_4],$$

since $Z_1 = (Y_1, Y_3)$ and $Y_1$ are linked by an invertible affine transformation once $\pi(Z_1) = Y_1 + Y_3$ is fixed, and similarly for $Z_2$ and $Y_2$. (This has to do with Lemma 2.12) Finally, we have

$$\begin{aligned} &\mathbb{I}[Z_1 + Z_2 : (\pi(Z_1), \pi(Z_2)) \,|\, \pi(Z_1) + \pi(Z_2)] \\ &\quad = \mathbb{I}[(Y_1 + Y_2, Y_3 + Y_4) : (Y_1 + Y_3, Y_2 + Y_4) \,|\, Y_1 + Y_2 + Y_3 + Y_4] \\ &\quad = \mathbb{I}[Y_1 + Y_2 : Y_2 + Y_4 \,|\, Y_1 + Y_2 + Y_3 + Y_4] \end{aligned}$$

where in the last line we used the fact that $(Y_1 + Y_2, Y_1 + Y_2 + Y_3 + Y_4)$ uniquely determine $Y_3 + Y_4$ and similarly $(Y_2 + Y_4, Y_1 + Y_2 + Y_3 + Y_4)$ uniquely determine $Y_1 + Y_3$. (This requires another helper lemma about entropy.) □

# Chapter 6

# Entropy version of PFR

**Definition 6.1.** $\eta := 1/9$.

Throughout this chapter, $G = \mathbb{F}_2^n$, and $X_1^0, X_2^0$ are $G$-valued random variables.

**Definition 6.2** ($\tau$ functional)**.** *If $X_1, X_2$ are two $G$-valued random variables, then*

$$\tau[X_1; X_2] := d[X_1; X_2] + \eta d[X_1^0; X_1] + \eta d[X_2^0; X_2].$$

**Lemma 6.3** ($\tau$ depends only on distribution)**.** *If $X_1', X_2'$ are copies of $X_1, X_2$, then $\tau[X_1'; X_2'] = \tau[X_1; X_2]$.*

*Proof.* Immediate from Lemma 3.6. □

**Definition 6.4** ($\tau$-minimizer)**.** *A pair of $G$-valued random variables $X_1, X_2$ are said to be a $\tau$-minimizer if one has*

$$\tau[X_1; X_2] \leq \tau[X_1'; X_2']$$

*for all $G$-valued random variables $X_1', X_2'$.*

**Proposition 6.5** ($\tau$ has minimum)**.** *A pair $X_1, X_2$ of $\tau$-minimizers exist.*

*Proof.* By Lemma 6.3, $\tau$ only depends on the probability distributions of $X_1, X_2$. This ranges over a compact space, and $\tau$ is continuous. So $\tau$ has a minimum. □

## 6.1 Basic facts about minimizers

In this section we assume that $X_1, X_2$ are $\tau$-minimizers. We also write $k := d[X_1; X_2]$.

**Lemma 6.6** (Distance lower bound)**.** *For any $G$-valued random variables $X_1', X_2'$, one has*

$$d[X_1'; X_2'] \geq k - \eta(d[X_1^0; X_1'] - d[X_1^0; X_1]) - \eta(d[X_2^0; X_2'] - d[X_2^0; X_2]).$$

*Proof.* Immediate from Definition 6.2 and Definition 6.5. □

**Lemma 6.7** (Conditional distance lower bound)**.** *For any $G$-valued random variables $X_1', X_2'$ and random variables $Z, W$, one has*

$$d[X_1'|Z; X_2'|W] \geq k - \eta(d[X_1^0; X_1'|Z] - d[X_1^0; X_1]) - \eta(d[X_2^0; X_2'|W] - d[X_2^0; X_2]).$$

*Proof.* Apply Lemma 6.6 to conditioned random variables and then average. □

## 6.2 First estimate

We continue the assumptions from the preceding section.

Let $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ be independent random variables, with $X_1, \tilde{X}_1$ copies of $X_1$ and $X_2, \tilde{X}_2$ copies of $X_2$. (This is possible thanks to Lemma 3.7.)

We also define the quantity

$$I_1 := I[X_1 + X_2 : \tilde{X}_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2].$$

**Lemma 6.8** (Fibring identity for first estimate)**.** *We have*

$$\begin{aligned} &d[X_1 + \tilde{X}_2; X_2 + \tilde{X}_1] + d[X_1 | X_1 + \tilde{X}_2; X_2 | X_2 + \tilde{X}_1] \\ &\quad + \mathbb{I}[X_1 + X_2 : \tilde{X}_1 + X_2 \,|\, X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2] = 2k. \end{aligned}$$

*Proof.* Immediate from Corollary 5.3. □

**Lemma 6.9** (Lower bound on distances)**.** *We have*

$$\begin{aligned} d[X_1 + \tilde{X}_2; X_2 + \tilde{X}_1] \geq k &- \eta(d[X^0_1; X_1 + \tilde{X}_2] - d[X^0_1; X_1]) \\ &- \eta(d[X^0_2; X_2 + \tilde{X}_1] - d[X^0_2; X_2]) \end{aligned}$$

*Proof.* Immediate from Lemma 6.6. □

**Lemma 6.10** (Lower bound on conditional distances)**.** *We have*

$$\begin{aligned} &d[X_1 | X_1 + \tilde{X}_2; X_2 | X_2 + \tilde{X}_1] \\ &\qquad \geq k - \eta(d[X^0_1; X_1 | X_1 + \tilde{X}_2] - d[X^0_1; X_1]) \\ &\qquad\qquad - \eta(d[X^0_2; X_2 | X_2 + \tilde{X}_1] - d[X^0_2; X_2]). \end{aligned}$$

*Proof.* Immediate from Lemma 6.7. □

**Lemma 6.11** (Upper bound on distance differences)**.** *We have*

$$\begin{aligned} d[X^0_1; X_1 + \tilde{X}_2] - d[X^0_1; X_1] &\leq \tfrac{1}{2}k + \tfrac{1}{4}\mathbb{H}[X_2] - \tfrac{1}{4}\mathbb{H}[X_1] \\ d[X^0_2; X_2 + \tilde{X}_1] - d[X^0_2; X_2] &\leq \tfrac{1}{2}k + \tfrac{1}{4}\mathbb{H}[X_1] - \tfrac{1}{4}\mathbb{H}[X_2], \\ d[X^0_1; X_1 | X_1 + \tilde{X}_2] - d[X^0_1; X_1] &\leq \tfrac{1}{2}k + \tfrac{1}{4}\mathbb{H}[X_1] - \tfrac{1}{4}\mathbb{H}[X_2] \\ d[X^0_2; X_2 | X_2 + \tilde{X}_1] - d[X^0_2; X_2] &\leq \tfrac{1}{2}k + \tfrac{1}{4}\mathbb{H}[X_2] - \tfrac{1}{4}\mathbb{H}[X_1]. \end{aligned}$$

*Proof.* Immediate from Lemma 3.25 (and recalling that $k$ is defined to be $d[X_1; X_2]$). □

**Lemma 6.12** (First estimate)**.** *We have $I_1 \leq 2\eta k$.*

*Proof.* Take a suitable linear combination of Lemma 6.8, Lemma 6.9, Lemma 6.10, and Lemma 6.11. □

One can also extract the following useful inequality from the proof of the above lemma.

**Lemma 6.13** (Entropy bound on quadruple sum)**.** *With the same notation, we have*

$$\mathbb{H}[X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2] \leq \tfrac{1}{2}\mathbb{H}[X_1] + \tfrac{1}{2}\mathbb{H}[X_2] + (2 + \eta)k - I_1. \tag{6.1}$$

*Proof.* Subtracting Lemma 6.10 from Lemma 6.8, and combining the resulting inequality with Lemma 6.11 gives the bound

$$d[X_1 + \tilde{X}_2; X_2 + \tilde{X}_1] \leq (1 + \eta)k - I_1,$$

and the claim follows from Lemma 3.11 and the definition of $k$. □

## 6.3 Second estimate

We continue the assumptions from the preceding section. We introduce the quantity

$$I_2 := \mathbb{I}[X_1 + X_2 : X_1 + \tilde X_1 | X_1 + X_2 + \tilde X_1 + \tilde X_2].$$

**Lemma 6.14** (Distance between sums)**.** *We have*

$$d[X_1 + \tilde X_1; X_2 + \tilde X_2] \geq k - \frac{\eta}{2}(d[X_1; X_1] + d[X_2; X_2]).$$

*Proof.* From Lemma 6.6 one has

$$\begin{aligned} d[X_1 + \tilde X_1; X_2 + \tilde X_2] \geq k &- \eta(d[X^0_1; X_1] - d[X^0_1; X_1 + \tilde X_1]) \\ &- \eta(d[X^0_2; X_2] - d[X^0_2; X_2 + \tilde X_2]). \end{aligned}$$

Now Lemma 3.25 gives

$$d[X^0_1; X_1 + \tilde X_1] - d[X^0_1; X_1] \leq \tfrac{1}{2} d[X_1; X_1]$$

and

$$d[X^0_2; X_2 + \tilde X_2] - d[X^0_2; X_2] \leq \tfrac{1}{2} d[X_2; X_2],$$

and the claim follows. □

**Lemma 6.15.** *We have*

$$d[X_1; X_1] + d[X_2; X_2] \leq 2k + \frac{2(2\eta k - I_1)}{1 - \eta}.$$

*Proof.* We may use Lemma 3.11 to expand

$$\begin{aligned} & d[X_1 + \tilde X_1; X_2 + \tilde X_2] \\ &= \mathbb{H}[X_1 + \tilde X_1 + X_2 + \tilde X_2] - \tfrac{1}{2}\mathbb{H}[X_1 + \tilde X_1] - \tfrac{1}{2}\mathbb{H}[X_2 + \tilde X_2] \\ &= \mathbb{H}[X_1 + \tilde X_1 + X_2 + \tilde X_2] - \tfrac{1}{2}\mathbb{H}[X_1] - \tfrac{1}{2}\mathbb{H}[X_2] \\ &\qquad - \tfrac{1}{2}\left(d[X_1; X_1] + d[X_2; X_2]\right), \end{aligned}$$

and hence by Lemma 6.13

$$d[X_1 + \tilde X_1; X_2 + \tilde X_2] \leq (2 + \eta)k - \tfrac{1}{2}\left(d[X_1; X_1] + d[X_2; X_2]\right) - I_1.$$

Combining this bound with Lemma 6.14 we obtain the result. □

**Lemma 6.16** (Second estimate)**.** *We have*

$$I_2 \leq 2\eta k + \frac{2\eta(2\eta k - I_1)}{1 - \eta}.$$

*Proof.* We apply Corollary 5.3, but now with the choice

$$(Y_1, Y_2, Y_3, Y_4) := (X_2, X_1, \tilde X_2, \tilde X_1).$$

Now Corollary 5.3 can be rewritten as

$$\begin{aligned} &d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] + d[X_1|X_1 + \tilde{X}_1; X_2|X_2 + \tilde{X}_2] \\ &\quad + \mathbb{I}[X_1 + X_2 : X_1 + \tilde{X}_1 \,|\, X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2] = 2k, \end{aligned}$$

recalling once again that $k := d[X_1; X_2]$. From Lemma 6.7 one has

$$\begin{aligned} d[X_1|X_1 + \tilde{X}_1; X_2|X_2 + \tilde{X}_2] \geq k &- \eta(d[X_1^0; X_1] - d[X_1^0; X_1|X_1 + \tilde{X}_1]) \\ &- \eta(d[X_2^0; X_2] - d[X_2^0; X_2|X_2 + \tilde{X}_2]). \end{aligned}$$

while from Lemma 3.25 we have

$$d[X_1^0; X_1|X_1 + \tilde{X}_1] - d[X_1^0; X_1] \leq \tfrac{1}{2} d[X_1; X_1],$$

and

$$d[X_2^0; X_2|X_2 + \tilde{X}_2] - d[X_2^0; X_2] \leq \tfrac{1}{2} d[X_1; X_2].$$

Combining all these inequalities with Lemma 6.14, we have

$$\mathbb{I}[X_1 + X_2 : X_1 + \tilde{X}_1|X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2] \leq \eta(d[X_1; X_1] + d[X_2; X_2]). \tag{6.2}$$

Together with Lemma 6.15, this gives the conclusion. □

## 6.4 Endgame

Let $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ be as before, and introduce the random variables

$$U := X_1 + X_2, \qquad V := \tilde{X}_1 + X_2, \qquad W := X_1 + \tilde{X}_1$$

and

$$S := X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2.$$

**Lemma 6.17** (Symmetry identity)**.** *We have*

$$I(U : W|S) = I(V : W|S).$$

*Proof.* This should follow from Lemma 3.6, Lemma 2.26, and Lemma 2.13. □

**Lemma 6.18** (Bound on conditional mutual informations)**.** *We have*

$$I(U : V \,|\, S) + I(V : W \,|\, S) + I(W : U \,|\, S) \leq 6\eta k - \frac{1 - 5\eta}{1 - \eta}(2\eta k - I_1).$$

*Proof.* From the definitions of $I_1, I_2$ and Lemma 6.17, we see that

$$I_1 = I(U : V \,|\, S), \qquad I_2 = I(W : U \,|\, S), \qquad I_2 = I(V : W \,|\, S).$$

Applying Lemma 6.12 and Lemma 6.16 we have the inequalities

$$I_2 \leq 2\eta k + \frac{2\eta(2\eta k - I_1)}{1 - \eta}.$$

We conclude that

$$I_1 + I_2 + I_2 \leq I_1 + 4\eta k + \frac{4\eta(2\eta k - I_1)}{1 - \eta}$$

and the claim follows from some calculation. □

**Lemma 6.19** (Bound on distance increments)**.** *We have*

$$\sum_{i=1}^{2} \sum_{A \in \{U,V,W\}} \left(d[X_i^0; A|S] - d[X_i^0; X_i]\right)$$
$$\leq (6 - 3\eta)k + 3(2\eta k - I_1).$$

*Proof.* By Lemma 3.26 (taking $X = X_1^0$, $Y = X_1$, $Z = X_2$ and $Z' = \tilde{X}_1 + \tilde{X}_2$, so that $Y + Z = U$ and $Y + Z + Z' = S$) we have, noting that $\mathbb{H}[Y + Z] = \mathbb{H}[Z']$,

$$d[X_1^0; U|S] - d[X_1^0; X_1] \leq \tfrac{1}{2}(\mathbb{H}[S] - \mathbb{H}[X_1]).$$

Further applications of Lemma 3.26 give

$$\begin{aligned} d[X_2^0; U|S] - d[X_2^0; X_2] &\leq \tfrac{1}{2}(\mathbb{H}[S] - \mathbb{H}[X_2]) \\ d[X_1^0; V|S] - d[X_1^0; X_1] &\leq \tfrac{1}{2}(\mathbb{H}[S] - \mathbb{H}[X_1]) \\ d[X_2^0; V|S] - d[X_2^0; X_2] &\leq \tfrac{1}{2}(\mathbb{H}[S] - \mathbb{H}[X_2]) \end{aligned}$$

and

$$d[X_1^0; W|S] - d[X_1^0; X_1] \leq \tfrac{1}{2}(\mathbb{H}[S] + \mathbb{H}[W] - \mathbb{H}[X_1] - \mathbb{H}[W']),$$

where $W' := X_2 + \tilde{X}_2$. To treat $d[X_2^0; W|S]$, first note that this equals $d[X_2^0; W'|S]$, since for a fixed choice $s$ of $S$ we have $W' = W + s$ (here we need some helper lemma about Ruzsa distance). Now we may apply Lemma 3.26 to obtain

$$d[X_2^0; W'|S] - d[X_2^0; X_2] \leq \tfrac{1}{2}(\mathbb{H}[S] + \mathbb{H}[W'] - \mathbb{H}[X_2] - \mathbb{H}[W]).$$

Summing these six estimates and using Lemma 6.13, we conclude that

$$\sum_{i=1}^{2} \sum_{A \in \{U,V,W\}} \left(d[X_i^0; A|S] - d[X_i^0; X_i]\right)$$
$$\begin{aligned} &\leq 3\mathbb{H}[S] - \tfrac{3}{2}\mathbb{H}[X_1] - \tfrac{3}{2}\mathbb{H}[X_2] \\ &\leq (6 - 3\eta)k + 3(2\eta k - I_1) \end{aligned}$$

as required. □

**Lemma 6.20** (Key identity)**.** *We have $U + V + W = 0$.*

*Proof.* Obvious because we are in characteristic two. □

For the next two lemmas, let $(T_1, T_2, T_3)$ be a $G^3$-valued random variable such that $T_1 + T_2 + T_3 = 0$ holds identically. Set

$$\delta := \sum_{1 \leq i < j \leq 3} \mathbb{I}[T_i; T_j]. \tag{6.3}$$

**Lemma 6.21** (Constructing good variables, I)**.** *One has*

$$\begin{aligned} k \leq \delta + \eta(d[X_1^0; T_1] - d[X_1^0; X_1]) + \eta(d[X_2^0; T_2] - d[X_2^0; X_2]) \\ + \tfrac{1}{2}\eta\mathbb{I}[T_1 : T_3] + \tfrac{1}{2}\eta\mathbb{I}[T_2 : T_3]. \end{aligned}$$

(Note: in the paper, this lemma was phrased in a more intuitive formulation that is basically the contrapositive of the one here. Similarly for the next two lemmas.)

*Proof.* We apply Lemma 3.23 with $(A, B) = (T_1, T_2)$ there. Since $T_1 + T_2 = T_3$, the conclusion is that

$$\sum_{t_3} \mathbb{P}[T_3 = t_3] d[(T_1|T_3 = t_3); (T_2|T_3 = t_3)] \\ \leq 3\mathbb{I}[T_1 : T_2] + 2\mathbb{H}[T_3] - \mathbb{H}[T_1] - \mathbb{H}[T_2]. \tag{6.4}$$

The right-hand side in (6.4) can be rearranged as

$$\begin{aligned} &2(\mathbb{H}[T_1] + \mathbb{H}[T_2] + \mathbb{H}[T_3]) - 3\mathbb{H}[T_1, T_2] \\ &= 2(\mathbb{H}[T_1] + \mathbb{H}[T_2] + \mathbb{H}[T_3]) - \mathbb{H}[T_1, T_2] - \mathbb{H}[T_2, T_3] - \mathbb{H}[T_1, T_3] = \delta, \end{aligned}$$

using the fact (from Lemma 2.2) that all three terms $\mathbb{H}[T_i, T_j]$ are equal to $\mathbb{H}[T_1, T_2, T_3]$ and hence to each other. We also have

$$\begin{aligned} &\sum_{t_3} P[T_3 = t_3](d[X_1^0; (T_1|T_3 = t_3)] - d[X_1^0; X_1]) \\ &= d[X_1^0; T_1|T_3] - d[X_1^0; X_1] \leq d[X_1^0; T_1] - d[X_1^0; X_1] + \tfrac{1}{2}\mathbb{I}[T_1 : T_3] \end{aligned}$$

by Lemma 3.24, and similarly

$$\begin{aligned} &\sum_{t_3} \mathbb{P}[T_3 = t_3](d[X_2^0; (T_2|T_3 = t_3)] - d[X_2^0; X_2]) \\ &\qquad \leq d[X_2^0; T_2] - d[X_2^0; X_2] + \tfrac{1}{2}\mathbb{I}[T_2 : T_3]. \end{aligned}$$

Putting the above observations together, we have

$$\begin{aligned} \sum_{t_3} \mathbb{P}[T_3 = t_3]\psi[(T_1|T_3 = t_3); (T_2|T_3 = t_3)] &\leq \delta + \eta(d[X_1^0; T_1] - d[X_1^0; X_1]) \\ &+\eta(d[X_2^0; T_2] - d[X_2^0; X_2]) + \tfrac{1}{2}\eta\mathbb{I}[T_1 : T_3] + \tfrac{1}{2}\eta\mathbb{I}[T_2 : T_3] \end{aligned}$$

where we introduce the notation

$$\psi[Y_1; Y_2] := d[Y_1; Y_2] + \eta(d[X_1^0; Y_1] - d[X_1^0; X_1]) + \eta(d[X_2^0; Y_2] - d[X_2^0; X_2]).$$

On the other hand, from Lemma 6.6 we have $k \leq \psi[Y_1; Y_2]$, and the claim follows. □

**Lemma 6.22** (Constructing good variables, II)**.** *One has*

$$k \leq \delta + \frac{\eta}{3}\Bigg(\delta + \sum_{i=1}^{2}\sum_{j=1}^{3}(d[X_i^0; T_j] - d[X_i^0; X_i])\Bigg).$$

*Proof.* Average Lemma 6.21 over all six permutations of $T_1, T_2, T_3$. □

**Theorem 6.23** ($\tau$-decrement)**.** *Let $X_1, X_2$ be tau-minimizers. Then $d[X_1; X_2] = 0$.*

*Proof.* Set $k := d[X_1; X_2]$. Applying Lemma 6.22 with any random variables $(T_1, T_2, T_3)$ such that $T_1 + T_2 + T_3 = 0$ holds identically, we deduce that

$$k \le \delta + \frac{\eta}{3}\Big(\delta + \sum_{i=1}^{2}\sum_{j=1}^{3}(d[X_1^0; T_j] - d[X_i^0; X_i])\Big).$$

Note that $\delta$ is still defined by (6.3) and thus depends on $T_1, T_2, T_3$. In particular we may apply this for

$$T_1 = (U|S = s), \qquad T_2 = (V|S = s), \qquad T_3 = (W|S = s)$$

for $s$ in the range of $S$ (which is a valid choice by Lemma 6.20) and then average over $s$ with weights $p_S(s)$, to obtain

$$k \le \tilde{\delta} + \frac{\eta}{3}\Big(\tilde{\delta} + \sum_{i=1}^{2}\sum_{A\in\{U,V,W\}}(d[X_i^0; A|S] - d[X_i^0; X_i])\Big),$$

where

$$\tilde{\delta} := \mathbb{I}[U : V|S] + \mathbb{I}[V : W|S] + \mathbb{I}[W : U|S].$$

Putting this together with Lemma 6.18 and Lemma 6.19, we conclude that

$$\begin{aligned} k &\le \Big(1 + \frac{\eta}{3}\Big)\Big(6\eta k - \frac{1-5\eta}{1-\eta}(2\eta k - I_1)\Big) + \frac{\eta}{3}\Big((6-3\eta)k + 3(2\eta k - I_1)\Big) \\ &= (8\eta + \eta^2)k - \Big(\frac{1-5\eta}{1-\eta}\Big(1 + \frac{\eta}{3}\Big) - \eta\Big)(2\eta k - I_1) \\ &\le (8\eta + \eta^2)k \end{aligned}$$

since the quantity $2\eta k - I_1$ is non-negative (by Lemma 6.12), and its coefficient in the above expression is non-positive provided that $\eta(2\eta + 17) \le 3$, which is certainly the case with Definition 6.1. Moreover, from Definition 6.1 we have $8\eta + \eta^2 < 1$. It follows that $k = 0$, as desired. □

## 6.5 Conclusion

**Theorem 6.24** (Entropy version of PFR)**.** *Let $G = \mathbb{F}_2^n$, and suppose that $X_1^0, X_2^0$ are $G$-valued random variables. Then there is some subgroup $H \le G$ such that*

$$d[X_1^0; U_H] + d[X_2^0; U_H] \le 11 d[X_1^0; X_2^0],$$

*where $U_H$ is uniformly distributed on $H$. Furthermore, both $d[X_1^0; U_H]$ and $d[X_2^0; U_H]$ are at most $6d[X_1^0; X_2^0]$.*

*Proof.* Let $X_1, X_2$ be the $\tau$-minimizer from Lemma 6.5. From Theorem 6.23, $d[X_1; X_2] = 0$. From Corollary 4.6, $d[X_1; U_H] = d[X_2; U_H] = 0$. Also from $\tau$-minimization we have $\tau[X_1; X_2] \le \tau[X_2^0; X_1^0]$. Using this and the Ruzsa triangle inequality we can conclude. □

Note: a "stretch goal" for this project would be to obtain a 'decidable' analogue of this result (see the remark at the end of Section 2 for some related discussion).

# Chapter 7

# Proof of PFR

**Lemma 7.1** (Ruzsa covering lemma)**.** *If $A, B$ are finite non-empty subsets of a group $G$, then $A$ can be covered by at most $|A+B|/|B|$ translates of $B-B$.*

*Proof.* Cover $A$ greedily by disjoint translates of $B$. □

**Lemma 7.2.** *If $A \subset \mathbf{F}_2^n$ is non-empty and $|A+A| \leq K|A|$, then $A$ can be covered by at most $K^{13/2}|A|^{1/2}/|H|^{1/2}$ translates of a subspace $H$ of $\mathbf{F}_2^n$ with*

$$|H|/|A| \in [K^{-11}, K^{11}]. \tag{7.1}$$

*Proof.* Let $U_A$ be the uniform distribution on $A$ (which exists by Lemma 2.5), thus $\mathbb{H}[U_A] = \log |A|$ by Lemma 2.7. By Lemma 2.3 and the fact that $U_A + U_A$ is supported on $A + A$, $\mathbb{H}[U_A+U_A] \leq \log |A+A|$. By Definition 3.8, the doubling condition $|A+A| \leq K|A|$ therefore gives

$$d[U_A; U_A] \leq \log K.$$

By Theorem 6.24, we may thus find a subspace $H$ of $\mathbb{F}_2^n$ such that

$$d[U_A; U_H] \leq \tfrac{1}{2}C' \log K \tag{7.2}$$

with $C' = 11$. By Lemma 3.13 we conclude that

$$|\log |H| - \log |A|| \leq C' \log K,$$

proving (7.1). From Definition 3.8, (7.2) is equivalent to

$$\mathbb{H}[U_A - U_H] \leq \log(|A|^{1/2}|H|^{1/2}) + \tfrac{1}{2}C' \log K.$$

By Lemma 2.8 we conclude the existence of a point $x_0 \in \mathbb{F}_p^n$ such that

$$p_{U_A - U_H}(x_0) \geq |A|^{-1/2}|H|^{-1/2}K^{-C'/2},$$

or equivalently

$$|A \cap (H + x_0)| \geq K^{-C'/2}|A|^{1/2}|H|^{1/2}.$$

Applying Lemma 7.1, we may thus cover $A$ by at most

$$\frac{|A + (A \cap (H + x_0))|}{|A \cap (H + x_0)|} \leq \frac{K|A|}{K^{-C'/2}|A|^{1/2}|H|^{1/2}} = K^{C'/2+1}\frac{|A|^{1/2}}{|H|^{1/2}}$$

translates of

$$(A \cap (H + x_0)) - (A \cap (H + x_0)) \subseteq H.$$

This proves the claim. □

**Theorem 7.3** (PFR)**.** *If $A \subset \mathbf{F}_2^n$ is non-empty and $|A + A| \leq K|A|$, then $A$ can be covered by most $2K^{12}$ translates of a subspace $H$ of $\mathbf{F}_2^n$ with $|H| \leq |A|$.*

*Proof.* Let $H$ be given by Lemma 7.2. If $|H| \leq |A|$ then we are already done thanks to (7.1). If $|H| > |A|$ then we can cover $H$ by at most $2|H|/|A|$ translates of a subspace $H'$ of $H$ with $|H'| \leq |A|$. We can thus cover $A$ by at most

$$2K^{13/2}\frac{|H|^{1/2}}{|A|^{1/2}}$$

translates of $H'$, and the claim again follows from (7.1). □

**Corollary 7.4** (PFR in infinite groups)**.** *If $G$ is an abelian 2-torsion group, $A \subset G$ is non-empty finite, and $|A + A| \leq K|A|$, then $A$ can be covered by most $2K^{12}$ translates of a finite group $H$ of $G$ with $|H| \leq |A|$.*

*Proof.* Apply Theorem 7.3 to the group generated by $A$, which is isomorphic to $\mathbb{F}_2^n$ for some $n$. □

# Chapter 8

# Improving the exponents

The arguments here are due to Jyun-Jie Liao.

**Definition 8.1** (New definition of $\eta$)**.** *$\eta$ is a real parameter with $\eta > 0$.*

Previously in Definition 6.1 we had set $\eta = 1/9$. To implement this chapter, one should refactor the previous arguments so that $\eta$ is now free to be a positive number, though the specific hypothesis $\eta = 1/9$ would now need to be added to Theorem 6.23.

Let $X_1^0, X_2^0$ be $G$-valued random variables, and let $X_1, X_2$ be $\tau$-minimizers as defined in Definition 6.4.

For the next two lemmas, let $(T_1, T_2, T_3)$ be a $G^3$-valued random variable such that $T_1 + T_2 + T_3 = 0$ holds identically. Let $\delta$ be the quantity in (6.3).

We have the following variant of Lemma 6.21:

**Lemma 8.2** (Constructing good variables, I')**.** *One has*

$$k \leq \delta + \eta(d[X_1^0; T_1|T_3] - d[X_1^0; X_1]) + \eta(d[X_2^0; T_2|T_3] - d[X_2^0; X_2]).$$

*Proof.* We apply Lemma 3.23 with $(A, B) = (T_1, T_2)$ there. Since $T_1 + T_2 = T_3$, the conclusion is that

$$\begin{aligned} &\sum_{t_3} \mathbb{P}[T_3 = t_3] d[(T_1|T_3 = t_3); (T_2|T_3 = t_3)] \\ &\qquad \leq 3\mathbb{I}[T_1 : T_2] + 2\mathbb{H}[T_3] - \mathbb{H}[T_1] - \mathbb{H}[T_2]. \end{aligned} \tag{8.1}$$

The right-hand side in (8.1) can be rearranged as

$$\begin{aligned} &2(\mathbb{H}[T_1] + \mathbb{H}[T_2] + \mathbb{H}[T_3]) - 3\mathbb{H}[T_1, T_2] \\ &= 2(\mathbb{H}[T_1] + \mathbb{H}[T_2] + \mathbb{H}[T_3]) - \mathbb{H}[T_1, T_2] - \mathbb{H}[T_2, T_3] - \mathbb{H}[T_1, T_3] = \delta, \end{aligned}$$

using the fact (from Lemma 2.2) that all three terms $\mathbb{H}[T_i, T_j]$ are equal to $\mathbb{H}[T_1, T_2, T_3]$ and hence to each other. We also have

$$\begin{aligned} &\sum_{t_3} P[T_3 = t_3](d[X_1^0; (T_1|T_3 = t_3)] - d[X_1^0; X_1]) \\ &\quad = d[X_1^0; T_1|T_3] - d[X_1^0; X_1] \end{aligned}$$

and similarly

$$\sum_{t_3} \mathbb{P}[T_3 = t_3](d[X_2^0; (T_2|T_3 = t_3)] - d[X_2^0; X_2]) \\ \leq d[X_2^0; T_2|T_3] - d[X_2^0; X_2].$$

Putting the above observations together, we have

$$\sum_{t_3} \mathbb{P}[T_3 = t_3]\psi[(T_1|T_3 = t_3); (T_2|T_3 = t_3)] \leq \delta + \eta(d[X_1^0; T_1|T_3] - d[X_1^0; X_1]) \\ +\eta(d[X_2^0; T_2|T_3] - d[X_2^0; X_2])$$

where we introduce the notation

$$\psi[Y_1; Y_2] := d[Y_1; Y_2] + \eta(d[X_1^0; Y_1] - d[X_1^0; X_1]) + \eta(d[X_2^0; Y_2] - d[X_2^0; X_2]).$$

On the other hand, from Lemma 6.6 we have $k \leq \psi[Y_1; Y_2]$, and the claim follows. □

(One could in fact refactor Lemma 6.21 to follow from Lemma 8.2 and Lemma 3.24).

**Lemma 8.3** (Constructing good variables, II'). *One has*

$$k \leq \delta + \frac{\eta}{6} \sum_{i=1}^{2} \sum_{1 \leq j,l \leq 3; j \neq l} (d[X_i^0; T_j|T_l] - d[X_i^0; X_i])$$

*Proof.* Average Lemma 8.2 over all six permutations of $T_1, T_2, T_3$. □

Now let $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ be independent copies of $X_1, X_2, X_1, X_2$, and set

$$U := X_1 + X_2, \qquad V := \tilde{X}_1 + X_2, \qquad W := X_1 + \tilde{X}_1$$

and

$$S := X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2$$

and introduce the quantities

$$k = d[X_1; X_2]$$

and

$$I_1 = I(U : V \,|\, S).$$

**Lemma 8.4** (Constructing good variables, III'). *One has*

$$k \leq I(U : V \,|\, S) + I(V : W \,|\, S) + I(W : U \,|\, S) + \frac{\eta}{6} \sum_{i=1}^{2} \sum_{A,B \in \{U,V,W\}: A \neq B} (d[X_i^0; A|B, S] - d[X_i^0; X_i]).$$

*Proof.* For each $s$ in the range of $S$, apply Lemma 8.3 with $T_1, T_2, T_3$ equal to $(U|S = s)$, $(V|S = s)$, $(W|S = s)$ respectively (which works thanks to Lemma 6.20), multiply by $\mathbb{P}[S = s]$, and sum in $s$ to conclude. □

To control the expressions in the right-hand side of this lemma we need a general entropy inequality.

**Lemma 8.5** (General inequality)**.** *Let $X_1, X_2, X_3, X_4$ be independent $G$-valued random variables, and let $Y$ be another $G$-valued random variable. Set $S := X_1 + X_2 + X_3 + X_4$. Then*

$$\begin{aligned}
&d[Y; X_1 + X_2 | X_1 + X_3, S] - d[Y; X_1] \\
&\quad \leq \tfrac{1}{4}(d[X_1; X_2] + 2d[X_1; X_3] + d[X_2; X_4]) \\
&\qquad\quad + \tfrac{1}{4}(d[X_1 | X_1 + X_3; X_2 | X_2 + X_4] - d[X_3 | X_3 + X_4; X_1 | X_1 + X_2]) \\
&\qquad\quad + \tfrac{1}{8}(\mathbb{H}[X_1 + X_2] - \mathbb{H}[X_3 + X_4] + \mathbb{H}[X_2] - \mathbb{H}[X_3] \\
&\qquad\qquad + \mathbb{H}[X_2 | X_2 + X_4] - \mathbb{H}[X_1 | X_1 + X_3]).
\end{aligned}$$

*Proof.* On the one hand, by Lemma 3.24 and two applications of Lemma 3.25 we have

$$\begin{aligned}
&d[Y; X_1 + X_2 | X_1 + X_3, S] \\
&\quad \leq d[Y; X_1 + X_2 | S] + \tfrac{1}{2}\mathbb{I}[X_1 + X_2 : X_1 + X_3 | S] \\
&\quad \leq d[Y; X_1 + X_2] \\
&\qquad + \tfrac{1}{2}(d[X_1 + X_2; X_3 + X_4] + \mathbb{I}[X_1 + X_2 : X_1 + X_3 | S]) \\
&\qquad + \tfrac{1}{4}(\mathbb{H}[X_1 + X_2] - \mathbb{H}[X_3 + X_4]) \\
&\quad \leq d[Y; X_1] \\
&\qquad + \tfrac{1}{2}(d[X_1; X_2] + d[X_1 + X_2; X_3 + X_4] + \mathbb{I}[X_1 + X_2 : X_1 + X_3 | S]) \\
&\qquad + \tfrac{1}{4}(\mathbb{H}[X_1 + X_2] - \mathbb{H}[X_3 + X_4] + \mathbb{H}[X_2] - \mathbb{H}[X_1]).
\end{aligned}$$

From Corollary 5.3 (with $Y_1, Y_2, Y_3, Y_4$ set equal to $X_3, X_1, X_4, X_2$ respectively) one has

$$\begin{aligned}
&d[X_3 + X_4; X_1 + X_2] + d[X_3 | X_3 + X_4; X_1 | X_1 + X_2] \\
&+ \mathbb{I}[X_3 + X_1 : X_1 + X_2 | S] = d[X_3; X_1] + d[X_4; X_2].
\end{aligned}$$

Rearranging the mutual information and Ruzsa distances slightly, we conclude that

$$\begin{aligned}
&d[Y; X_1 + X_2 | X_1 + X_3, S] \\
&\quad \leq d[Y; X_1] \\
&\qquad + \tfrac{1}{2}(d[X_1; X_2] + d[X_1; X_3] + d[X_2; X_4] - d[X_3 | X_3 + X_4; X_1 | X_1 + X_2]) \\
&\qquad + \tfrac{1}{4}(\mathbb{H}[X_1 + X_2] - \mathbb{H}[X_3 + X_4] + \mathbb{H}[X_2] - \mathbb{H}[X_1]).
\end{aligned}$$

On the other hand, $(X_1 + X_2 | X_1 + X_3, S)$ has an identical distribution to the independent sum of $(X_1 | X_1 + X_3)$ and $(X_2 | X_2 + X_4)$. We may therefore apply Lemma 3.25 to conditioned variables $(X_1 | X_1 + X_3 = s)$ and $(X_2 | X_2 + X_4 = t)$ and average in $s, t$ to obtain the alternative bound

$$\begin{aligned}
&d[Y; X_1 + X_2 | X_1 + X_3, S] \\
&\quad \leq d[Y; X_1 | X_1 + X_3] + \tfrac{1}{2}d[X_1 | X_1 + X_3; X_2 | X_2 + X_4] \\
&\qquad + \tfrac{1}{4}(\mathbb{H}[X_2 | X_2 + X_4] - \mathbb{H}[X_1 | X_1 + X_3]) \\
&\quad \leq d[Y; X_1] \\
&\qquad + \tfrac{1}{2}(d[X_1; X_3] + d[X_1 | X_1 + X_3; X_2 | X_2 + X_4]) \\
&\qquad + \tfrac{1}{4}(\mathbb{H}[X_2 | X_2 + X_4] - \mathbb{H}[X_1 | X_1 + X_3] + \mathbb{H}[X_1] - \mathbb{H}[X_3]).
\end{aligned}$$

If one takes the arithmetic mean of these two bounds and simplifies using Corollary 5.3, one obtains the claim. □

Returning to our specific situation, we now have

**Lemma 8.6** (Bound on distance differences)**.** *We have*

$$\sum_{i=1}^{2} \sum_{A,B\in\{U,V,W\}:A\neq B} d[X_i^0;A|B,S] - d[X_i^0;X_i]$$
$$\leq 12k + \frac{4(2\eta k - I_1)}{1-\eta}.$$

*Proof.* If we apply Lemma 8.5 with $X_1 := X_1$, $Y := X_1^0$ and $(X_2,X_3,X_4)$ equal to the 3! permutations of $(X_2,\tilde X_1,\tilde X_2)$, and sums (using the symmetry $\mathbb{H}[X|X+Y] = \mathbb{H}[Y|X+Y]$, which follows from Lemma 2.12), we can bound

$$\sum_{A,B\in\{U,V,W\}:A\neq B} d[X_1^0;A|B,S] - d[X_1^0;X_1]$$

by

$$\begin{aligned}
&\tfrac14(6d[X_1;X_2] + 6d[X_1;\tilde X_2]\\
&\quad + 6d[X_1;\tilde X_1] + 2d[\tilde X_1;\tilde X_2] + 2d[\tilde X_1;X_2] + 2d[X_2;\tilde X_2])\\
&\quad + \tfrac18(2\mathbb{H}[X_1+X_2] + 2\mathbb{H}[X_1+\tilde X_1] + 2\mathbb{H}[X_1+\tilde X_2]\\
&\quad - 2\mathbb{H}[\tilde X_1+X_2] - 2\mathbb{H}[X_2+\tilde X_2] - 2\mathbb{H}[\tilde X_1+\tilde X_2])\\
&\qquad + \tfrac14(\mathbb{H}[X_2|X_2+\tilde X_2] + \mathbb{H}[\tilde X_1|\tilde X_1+\tilde X_2] + \mathbb{H}[\tilde X_1|X_1+\tilde X_2]\\
&\qquad\quad - \mathbb{H}[X_1|X_1+\tilde X_1] - \mathbb{H}[X_1|X_1+X_2] - \mathbb{H}[X_1|X_1+\tilde X_2]),
\end{aligned}$$

which simplifies to

$$\begin{aligned}
&\tfrac14(16k + 6d[X_1;X_1] + 2d[X_2;X_2])\\
&\quad + \tfrac14(H[X_1+\tilde X_1] - H[X_2+\tilde X_2] + d[X_2|X_2+\tilde X_2] - d[X_1|X_1+\tilde X_1]).
\end{aligned}$$

A symmetric argument also bounds

$$\sum_{A,B\in\{U,V,W\}:A\neq B} d[X_2^0;A|B,S] - d[X_2^0;X_2]$$

by

$$\begin{aligned}
&\tfrac14(16k + 6d[X_2;X_2] + 2d[X_1;X_1])\\
&\quad + \tfrac14(H[X_2+\tilde X_2] - H[X_1+\tilde X_1] + d[X_1|X_1+\tilde X_1] - d[X_2|X_2+\tilde X_2]).
\end{aligned}$$

On the other hand, from Lemma 6.15 one has

$$d[X_1;X_1] + d[X_2;X_2] \leq 2k + \frac{2(2\eta k - I_1)}{1-\eta}.$$

Summing the previous three estimates, we obtain the claim. □

**Theorem 8.7** (Improved $\tau$-decrement)**.** *Suppose* $0 < \eta < 1/8$*. Let* $X_1, X_2$ *be tau-minimizers. Then* $d[X_1;X_2] = 0$.

*Proof.* From Lemma 8.4, Lemma 8.6, and Lemma 6.18 one has

$$k \leq 8\eta k - \frac{(1-5\eta-\frac{4}{6}\eta)(2\eta k - I_1)}{(1-\eta)}.$$

For any $\eta < 1/8$, we see from Lemma 6.12 that the expression $\frac{(1-5\eta-\frac{4}{6}\eta)(2\eta k-I_1)}{(1-\eta)}$ is nonnegative, and hence $k = 0$ as required. □

**Theorem 8.8** (Limiting improved $\tau$-decrement)**.** *For $\eta = 1/8$, there exist tau-minimizers $X_1, X_2$ satisfying $d[X_1; X_2] = 0$.*

*Proof.* For each $\eta < 1/8$, consider minimizers $X_1^\eta$ and $X_2^\eta$ from Lemma 6.5. By Theorem 8.7, they satisfy $d[X_1^\eta; X_2^\eta] = 0$. By compactness of the space of probability measures on $G$, one may extract a converging subsequence of the distributions of $X_1^\eta$ and $X_2^\eta$ as $\eta \to 1/8$. By continuity of all the involved quantities, the limit is a pair of tau-minimizers for 1/8 satisfying additionally $d[X_1; X_2] = 0$. □

**Theorem 8.9** (Improved entropy version of PFR)**.** *Let $G = \mathbb{F}_2^n$, and suppose that $X_1^0, X_2^0$ are $G$-valued random variables. Then there is some subgroup $H \leq G$ such that*

$$d[X_1^0; U_H] + d[X_2^0; U_H] \leq 10 d[X_1^0; X_2^0],$$

*where $U_H$ is uniformly distributed on $H$. Furthermore, both $d[X_1^0; U_H]$ and $d[X_2^0; U_H]$ are at most $6d[X_1^0; X_2^0]$.*

*Proof.* Let $X_1, X_2$ be the good $\tau$-minimizer from Theorem 8.8. By construction, $d[X_1; X_2] = 0$. From Corollary 4.6, $d[X_1; U_H] = d[X_2; U_H] = 0$. Also from $\tau$-minimization we have $\tau[X_1; X_2] \leq \tau[X_2^0; X_1^0]$. Using this and the Ruzsa triangle inequality we can conclude. □

One can then replace Lemma 7.2 with

**Lemma 8.10.** *If $A \subset \mathbf{F}_2^n$ is non-empty and $|A + A| \leq K|A|$, then $A$ can be covered by at most $K^6 |A|^{1/2} / |H|^{1/2}$ translates of a subspace $H$ of $\mathbf{F}_2^n$ with*

$$|H|/|A| \in [K^{-10}, K^{10}].$$

*Proof.* By repeating the proof of Lemma 7.2 and using Theorem 8.9 one can obtain the claim with 13/2 replaced with 6 and 11 replaced by 10. □

This implies the following improved version of Theorem 7.3:

**Theorem 8.11** (Improved PFR)**.** *If $A \subset \mathbf{F}_2^n$ is non-empty and $|A+A| \leq K|A|$, then $A$ can be covered by most $2K^{11}$ translates of a subspace $H$ of $\mathbf{F}_2^n$ with $|H| \leq |A|$.*

*Proof.* By repeating the proof of Theorem 7.3 and using Lemma 8.10 one can obtain the claim with 11 replaced by 10. □

Of course, by replacing Theorem 7.3 with Theorem 8.11 we may also improve constants in downstream theorems in a straightforward manner.

# Chapter 9

# Homomorphism version of PFR

In this section, $G, G'$ are finite abelian 2-groups.

**Lemma 9.1** (Hahn-Banach type theorem)**.** *Let $H_0$ be a subgroup of $G$. Then every homomorphism $\phi : H_0 \to G'$ can be extended to a homomorphism $\tilde{\phi} : G \to G'$.*

*Proof.* By induction it suffices to treat the case where $H_0$ has index 2 in $G$, but then the extension can be constructed by hand. □

**Lemma 9.2** (Goursat type theorem)**.** *Let $H$ be a subgroup of $G \times G'$. Then there exists a subgroup $H_0$ of $G$, a subgroup $H_1$ of $G'$, and a homomorphism $\phi : G \to G'$ such that*

$$H := \{(x, \phi(x) + y) : x \in H_0, y \in H_1\}.$$

*In particular, $|H| = |H_0||H_1|$.*

*Proof.* We can take $H_0$ to be the projection of $H$ to $G$, and $H_1$ to be the slice $H_1 := \{y : (0, y) \in H\}$. One can construct $\phi$ on $H_0$ one generator at a time by the greedy algorithm, and then extend to $G$ by Lemma 9.1. The cardinality bound is clear from direct counting. □

**Theorem 9.3** (Homomorphism form of PFR)**.** *Let $f : G \to G'$ be a function, and let $S$ denote the set*

$$S := \{f(x + y) - f(x) - f(y) : x, y \in G\}.$$

*Then there exists a homomorphism $\phi : G \to G'$ such that*

$$|\{f(x) - \phi(x) : x \in G\}| \leq |S|^{12}.$$

*Proof.* Consider the graph $A \subset G \times G'$ defined by

$$A := \{(x, f(x)) : x \in G\}.$$

Clearly, $|A| = |G|$. By hypothesis, we have

$$A + A \subset \{(x, f(x) + s) : x \in G, s \in S\}$$

and hence $|A + A| \leq |S||A|$. Applying Lemma 8.10, we may find a subspace $H \subset G \times G'$ such that $|H|/|A| \in [K^{-10}, K^{10}]$ and $A$ is covered by $c + H$ with $|c| \leq |S|^6|A|^{1/2}/|H|^{1/2}$. If we let $H_0, H_1$ be as in Lemma 9.2, this implies on taking projections that $G$ is covered by at most $|c|$ translates of $H_0$. This implies that

$$|c||H_0| \geq |G|;$$

since $|H_0||H_1| = |H|$, we conclude that

$$|H_1| \leq |c||H|/|G| = |c||H|/|A|.$$

By hypothesis, $A$ is covered by at most $|c|$ translates of $H$, and hence by at most $|c||H_1|$ translates of $\{(x, \phi(x)) : x \in G\}$. As $\phi$ is a homomorphism, each such translate can be written in the form $\{(x, \phi(x) + d) : x \in G\}$ for some $d \in G'$. Since

$$|c||H_1| \leq |c|^2 \frac{|H|}{|A|} \leq \left(|S|^6 \frac{|A|^{1/2}}{|H|^{1/2}}\right)^2 \frac{|H|}{|A|} = |S|^{12},$$

the result follows. □

# Chapter 10

# Approximate homomorphism version of PFR

**Definition 10.1** (Additive energy). *If $G$ is a group, and $A$ is a finite subset of $G$, the* additive energy *$E(A)$ of $A$ is the number of quadruples $(a_1, a_2, a_3, a_4) \in A^4$ such that $a_1 + a_2 = a_3 + a_4$.*

**Lemma 10.2** (Cauchy–Schwarz bound). *If $G$ is a group, $A, B$ are finite subsets of $G$, then*

$$E(A) \geq \frac{|\{(a, a') \in A \times A : a + a' \in B\}|^2}{|B|}.$$

*Proof.* If $B$ is empty then the claim is trivial (with the Lean convention 0/0), so without loss of generality $B$ is non-empty. We can rewrite

$$|\{(a, a') \in A \times A : a + a' \in B\}| = \sum_{b \in B} r(b)$$

where $r : G \to \mathbb{N}$ is the counting function

$$r(b) := |\{(a, a') \in A \times A : a + a' = b\}|.$$

From double counting we have

$$\sum_{b \in G} r(b)^2 = E(A).$$

The claim now follows from the Cauchy–Schwarz inequality

$$(\sum_{b \in B} r(b))^2 \leq |B| \sum_{b \in B} r(b)^2.$$

□

**Lemma 10.3** (Balog–Szemerédi–Gowers lemma). *Let $G$ be an abelian group, and let $A$ be a finite non-empty set with $E(A) \geq |A|^3/K$ for some $K \geq 1$. Then there is a subset $A'$ of $A$ with $|A'| \geq |A|/(C_1 K^{C_2})$ and $|A' - A'| \leq C_3 K^{C_4} |A|$, where (provisionally)*

$$C_1 = 2^4, C_2 = 1, C_3 = 2^{10}, C_4 = 5.$$

*Proof.* See `https://terrytao.files.wordpress.com/2024/01/simplebsg.pdf` □

**Theorem 10.4** (Approximate homomorphism form of PFR)**.** *Let $G, G'$ be finite abelian 2-groups. Let $f : G \to G'$ be a function, and suppose that there are at least $|G|^2/K$ pairs $(x, y) \in G^2$ such that*

$$f(x + y) = f(x) + f(y).$$

*Then there exists a homomorphism $\phi : G \to G'$ and a constant $c \in G'$ such that $f(x) = \phi(x) + c$ for at least $|G|/(2^{172} * K^{146})$ values of $x \in G$.*

*Proof.* Consider the graph $A \subset G \times G'$ defined by

$$A := \{(x, f(x)) : x \in G\}.$$

Clearly, $|A| = |G|$. By hypothesis, we have $a + a' \in A$ for at least $|A|^2/K$ pairs $(a, a') \in A^2$. By Lemma 10.2, this implies that $E(A) \geq |A|^3/K^2$. Applying Lemma 10.3, we conclude that there exists a subset $A' \subset A$ with $|A'| \geq |A|/C_1K^{2C_2}$ and $|A' + A'| \leq C_1C_3K^{2(C_2+C_4)}|A'|$. Applying Lemma 8.10, we may find a subspace $H \subset G \times G'$ such that $|H|/|A'| \in [L^{-10}, L^{10}$ and a subset $c$ of cardinality at most $L^6|A'|^{1/2}/|H|^{1/2}$ such that $A' \subseteq c + H$, where $L = C_1C_3K^{2(C_2+C_4)}$. If we let $H_0, H_1$ be as in Lemma 9.2, this implies on taking projections the projection of $A'$ to $G$ is covered by at most $|c|$ translates of $H_0$. This implies that

$$|c||H_0| \geq |A'|;$$

since $|H_0||H_1| = |H|$, we conclude that

$$|H_1| \leq |c||H|/|A'|.$$

By hypothesis, $A'$ is covered by at most $|c|$ translates of $H$, and hence by at most $|c||H_1|$ translates of $\{(x, \phi(x)) : x \in G\}$. As $\phi$ is a homomorphism, each such translate can be written in the form $\{(x, \phi(x) + c) : x \in G\}$ for some $c \in G'$. The number of translates is bounded by

$$|c|^2 \frac{|H|}{|A'|} \leq \left(L^6 \frac{|A'|^{1/2}}{|H|^{1/2}}\right)^2 \frac{|H|}{|A'|} = L^{12}.$$

By the pigeonhole principle, one of these translates must then contain at least $|A'|/L^{12} \geq |G|/(C_1C_3K^{2(C_2+C_4)})^{12}(C_1K^{2C_2})$ elements of $A'$ (and hence of $A$), and the claim follows. □

# Chapter 11

# Weak PFR over the integers

**Lemma 11.1.** *If $G$ is torsion-free and $X, Y$ are $G$-valued random variables then $d[X; 2Y] \leq 5d[X; Y]$.*

*Proof.* Let $Y_1, Y_2$ be independent copies of $Y$ (also independent of $X$). Since $G$ is torsion-free we know $X, Y_1 - Y_2, X - 2Y_1$ uniquely determine $X, Y_1, Y_2$ and so

$$\mathbb{H}(X, Y_1, Y_2, X - 2Y_1) = \mathbb{H}(X, Y_1, Y_2) = \mathbb{H}(X) + 2\mathbb{H}(Y).$$

Similarly

$$\mathbb{H}(X, X - 2Y_1) = \mathbb{H}(X) + \mathbb{H}(2Y_1) = \mathbb{H}(X) + \mathbb{H}(Y).$$

Furthermore

$$\mathbb{H}(Y_1 - Y_2, X - 2Y_1) = \mathbb{H}(Y_1 - Y_2, X - Y_1 - Y_2) \leq \mathbb{H}(Y_1 - Y_2) + \mathbb{H}(X - Y_1 - Y_2).$$

By submodularity (Corollary 2.21)

$$\mathbb{H}(X, Y_1, Y_2, X - 2Y_1) + \mathbb{H}(X - 2Y_1) \leq \mathbb{H}(X, X - 2Y_1) + \mathbb{H}(Y_1 - Y_2, X - 2Y_1).$$

Combining these inequalities

$$\mathbb{H}(X - 2Y_1) \leq \mathbb{H}(Y_1 - Y_2) + \mathbb{H}(X - Y_1 - Y_2) - \mathbb{H}(Y).$$

Similarly we have

$$\mathbb{H}(Y_1, Y_2, X - Y_1 - Y_2) = \mathbb{H}(X) + 2\mathbb{H}(Y),$$
$$\mathbb{H}(Y_1, X - Y_1 - Y_2) = \mathbb{H}(Y) + \mathbb{H}(X - Y_2),$$

and

$$\mathbb{H}(Y_2, X - Y_1 - Y_2) = \mathbb{H}(Y) + \mathbb{H}(X - Y_1)$$

and by submodularity (Corollary 2.21) again

$$\mathbb{H}(Y_1, Y_2, X - Y_1 - Y_2) + \mathbb{H}(X - Y_1 - Y_2) \leq \mathbb{H}(Y_1, X - Y_1 - Y_2) + \mathbb{H}(Y_2, X - Y_1 - Y_2).$$

Combining these inequalities (and recalling the definition of Ruzsa distance) gives

$$\mathbb{H}(X - Y_1 - Y_2) \leq \mathbb{H}(X - Y_1) + \mathbb{H}(X - Y_2) - \mathbb{H}(X) = 2d[X; Y] + \mathbb{H}(Y).$$

It follows that

$$\mathbb{H}(X-2Y_1) \leq \mathbb{H}(Y_1-Y_2) + 2d[X;Y]$$

and so (using $\mathbb{H}(2Y)=\mathbb{H}(Y)$)

$$\begin{aligned} d[X;2Y] &= \mathbb{H}(X-2Y_1) - \mathbb{H}(X)/2 - \mathbb{H}(2Y)/2 \\ &\leq \mathbb{H}(Y_1-Y_2) + 2d[X;Y] - \mathbb{H}(X)/2 - \mathbb{H}(Y)/2 \\ &= d[Y_1;Y_2] + \frac{\mathbb{H}(Y)-\mathbb{H}(X)}{2} + 2d[X;Y]. \end{aligned}$$

Finally note that by the triangle inequality (Lemma 3.18) we have

$$d[Y_1;Y_2] \leq d[Y_1;X] + d[X;Y_2] = 2d[X;Y].$$

The result follows from $(\mathbb{H}(Y)-\mathbb{H}(X))/2 \leq d[X;Y]$ (Lemma 3.13). □

**Lemma 11.2.** *If $G$ is a torsion-free group and $X, Y$ are $G$-valued random variables and $\phi : G \to \mathbb{F}_2^d$ is a homomorphism then*

$$\mathbb{H}(\phi(X)) \leq 10d[X;Y].$$

*Proof.* By Corollary 5.2 and Lemma 11.1 we have

$$d[\phi(X);\phi(2Y)] \leq d[X;2Y] \leq 5d[X;Y]$$

and $\phi(2Y) = 2\phi(Y) \equiv 0$ so the left-hand side is equal to $d[\phi(X);0] = \mathbb{H}(\phi(X))/2$ (using Lemma 3.9). □

**Lemma 11.3.** *Let $G = \mathbb{F}_2^n$ and $\alpha \in (0,1)$ and let $X, Y$ be $G$-valued random variables such that*

$$\mathbb{H}(X) + \mathbb{H}(Y) > \frac{20}{\alpha} d[X;Y].$$

*There is a non-trivial subgroup $H \leq G$ such that*

$$\log|H| < \frac{1+\alpha}{2}(\mathbb{H}(X)+\mathbb{H}(Y))$$

*and*

$$\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)) < \alpha(\mathbb{H}(X)+\mathbb{H}(Y))$$

*where $\psi : G \to G/H$ is the natural projection homomorphism.*

*Proof.* By Theorem 8.9 there exists a subgroup $H$ such that $d[X;U_H]+d[Y;U_H] \leq 10d[X;Y]$. Using Lemma 3.16 we deduce that $\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(X)) \leq 20d[X;Y]$. The second claim follows adding these inequalities and using the assumption on $\mathbb{H}(X)+\mathbb{H}(Y)$.

Furthermore we have by Lemma 3.13

$$\log|H| - \mathbb{H}(X) \leq 2d[X;U_H]$$

and similarly for $Y$ and thus

$$\begin{aligned} \log|H| &\leq \frac{\mathbb{H}(X)+\mathbb{H}(Y)}{2} + d[X;U_H] + d[Y;U_H] \leq \frac{\mathbb{H}(X)+\mathbb{H}(Y)}{2} + 10d[X;Y] \\ &< \frac{1+\alpha}{2}(\mathbb{H}(X)+\mathbb{H}(Y)). \end{aligned}$$

Finally note that if $H$ were trivial then $\psi(X) = X$ and $\psi(Y) = Y$ and hence $\mathbb{H}(X)+\mathbb{H}(Y) = 0$, which contradicts Lemma 3.15. □

**Lemma 11.4.** *If $G = \mathbb{F}_2^d$ and $\alpha \in (0,1)$ and $X, Y$ are $G$-valued random variables then there is a subgroup $H \leq \mathbb{F}_2^d$ such that*

$$\log|H| \leq \frac{1+\alpha}{2(1-\alpha)}(\mathbb{H}(X) + \mathbb{H}(Y))$$

*and if $\psi : G \to G/H$ is the natural projection then*

$$\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)) \leq \frac{20}{\alpha} d[\psi(X); \psi(Y)].$$

*Proof.* Let $H \leq \mathbb{F}_2^d$ be a maximal subgroup such that

$$\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)) > \frac{20}{\alpha} d[\psi(X); \psi(Y)]$$

and such that there exists $c \geq 0$ with

$$\log|H| \leq \frac{1+\alpha}{2(1-\alpha)}(1-c)(\mathbb{H}(X) + \mathbb{H}(Y))$$

and

$$\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)) \leq c(\mathbb{H}(X) + \mathbb{H}(Y)).$$

Note that this exists since $H = \{0\}$ is an example of such a subgroup or we are done with this choice of $H$.

We know that $G/H$ is a 2-elementary group and so by Lemma 11.3 there exists some non-trivial subgroup $H' \leq G/H$ such that

$$\log|H'| < \frac{1+\alpha}{2}(\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y))$$

and

$$\mathbb{H}(\psi' \circ \psi(X)) + \mathbb{H}(\psi' \circ \psi(Y)) < \alpha(\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)))$$

where $\psi' : G/H \to (G/H)/H'$. By group isomorphism theorems we know that there exists some $H''$ with $H \leq H'' \leq G$ such that $H' \cong H''/H$ and $\psi' \circ \psi(X) = \psi''(X)$ where $\psi'' : G \to G/H''$ is the projection homomorphism.

Since $H'$ is non-trivial we know that $H$ is a proper subgroup of $H''$. On the other hand we know that

$$\log|H''| = \log|H'| + \log|H| < \frac{1+\alpha}{2(1-\alpha)}(1-\alpha c)(\mathbb{H}(X) + \mathbb{H}(Y))$$

and

$$\mathbb{H}(\psi''(X)) + \mathbb{H}(\psi''(Y)) < \alpha(\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y))) \leq \alpha c(\mathbb{H}(X) + \mathbb{H}(Y)).$$

Therefore (using the maximality of $H$) it must be the first condition that fails, whence

$$\mathbb{H}(\psi''(X)) + \mathbb{H}(\psi''(Y)) \leq \frac{20}{\alpha} d[\psi''(X); \psi''(Y)].$$

□

We could use the previous lemma for any value of $\alpha \in (0,1)$, which would give a whole range of estimates in Theorem 11.10. For definiteness, we specialize only to $\alpha = 3/5$, which gives a constant 2 in the first bound below.

**Lemma 11.5.** *If $G = \mathbb{F}_2^d$ and $\alpha \in (0,1)$ and $X, Y$ are $G$-valued random variables then there is a subgroup $H \leq \mathbb{F}_2^d$ such that*

$$\log|H| \leq 2(\mathbb{H}(X) + \mathbb{H}(Y))$$

*and if $\psi : G \to G/H$ is the natural projection then*

$$\mathbb{H}(\psi(X)) + \mathbb{H}(\psi(Y)) \leq 34d[\psi(X); \psi(Y)].$$

*Proof.* Specialize Lemma 11.4 to $\alpha = 3/5$. In the second inequality, it gives a bound $100/3 < 34$. □

**Lemma 11.6.** *Let $\phi : G \to H$ be a homomorphism and $A, B \subseteq G$ be finite subsets. If $x, y \in H$ then let $A_x = A \cap \phi^{-1}(x)$ and $B_y = B \cap \phi^{-1}(y)$. There exist $x, y \in H$ such that $A_x, B_y$ are both non-empty and*

$$d[\phi(U_A); \phi(U_B)] \log \frac{|A||B|}{|A_x||B_y|} \leq (\mathbb{H}(\phi(U_A)) + \mathbb{H}(\phi(U_B)))(d(U_A, U_B) - d(U_{A_x}, U_{B_y})).$$

*Proof.* The random variables $(U_A \mid \phi(U_A) = x)$ and $(U_B \mid \phi(U_B) = y)$ are equal in distribution to $U_{A_x}$ and $U_{B_y}$ respectively (both are uniformly distributed over their respective fibres). It follows from Lemma 5.1 that

$$\begin{aligned}\sum_{x,y \in H} \frac{|A_x||B_y|}{|A||B|} d[U_{A_x}; U_{B_y}] &= d[U_A \mid \phi(U_A); U_B \mid \phi(U_B)] \\ &\leq d[U_A; U_B] - d[\phi(U_A); \phi(U_B)].\end{aligned}$$

Therefore with $M := \mathbb{H}(\phi(U_A)) + \mathbb{H}(\phi(U_B))$ we have

$$\left(\sum_{x,y \in H} \frac{|A_x||B_y|}{|A||B|} Md[U_{A_x}; U_{B_y}]\right) + Md[\phi(U_A); \phi(U_B)] \leq Md[U_A; U_B].$$

Since

$$M = \sum_{x,y \in H} \frac{|A_x||B_y|}{|A||B|} \log \frac{|A||B|}{|A_x||B_y|}$$

we have

$$\sum_{x,y \in H} \frac{|A_x||B_y|}{|A||B|} \left(Md[U_{A_x}; U_{B_y}] + d[\phi(U_A); \phi(U_B)] \log \frac{|A||B|}{|A_x||B_y|}\right) \leq Md[U_A; U_B].$$

It follows that there exists some $x, y \in H$ such that $|A_x|, |B_y| \neq 0$ and

$$Md[U_{A_x}; U_{B_y}] + d[\phi(U_A); \phi(U_B)] \log \frac{|A||B|}{|A_x||B_y|} \leq Md[U_A; U_B].$$

□

**Definition 11.7.** *If $A \subseteq \mathbb{Z}^d$ then by $\dim(A)$ we mean the dimension of the span of $A - A$ over the reals – equivalently, the smallest $d'$ such that $A$ lies in a coset of a subgroup isomorphic to $\mathbb{Z}^{d'}$.*

**Theorem 11.8.** *If $A, B \subseteq \mathbb{Z}^d$ are finite non-empty sets then there exist non-empty $A' \subseteq A$ and $B' \subseteq B$ such that*

$$\log \frac{|A||B|}{|A'||B'|} \leq 34d[U_A; U_B]$$

*such that* $\max(\dim A', \dim B') \leq \frac{40}{\log 2} d[U_A; U_B]$.

*Proof.* Without loss of generality we can assume that $A$ and $B$ are not both inside (possibly distinct) cosets of the same subgroup of $\mathbb{Z}^d$, or we just replace $\mathbb{Z}^d$ with that subgroup. We prove the result by induction on $|A| + |B|$.

Let $\phi : \mathbb{Z}^d \to \mathbb{F}_2^d$ be the natural mod-2 homomorphism. By Lemma 11.2

$$\max(\mathbb{H}(\phi(U_A)), \mathbb{H}(\phi(U_B))) \leq 10d[U_A; U_B].$$

We now apply Lemma 11.5, obtaining some subgroup $H \leq \mathbb{F}_2^d$ such that

$$\log|H| \leq 40d[U_A; U_B]$$

and

$$\mathbb{H}(\tilde{\phi}(U_A)) + \mathbb{H}(\tilde{\phi}(U_B)) \leq 34d[\tilde{\phi}(U_A); \tilde{\phi}(U_B)]$$

where $\tilde{\phi} : \mathbb{Z}^d \to \mathbb{F}_2^d/H$ is $\phi$ composed with the projection onto $\mathbb{F}_2^d/H$.

By Lemma 11.6 there exist $x, y \in \mathbb{F}_2^d/H$ such that, with $A_x = A \cap \tilde{\phi}^{-1}(x)$ and similarly for $B_y$,

$$\log \frac{|A||B|}{|A_x||B_y|} \leq 34(d[U_A; U_B] - d[U_{A_x}; U_{B_y}]).$$

Suppose first that $|A_x| + |B_y| = |A| + |B|$. This means that $\tilde{\phi}(A) = \{x\}$ and $\tilde{\phi}(B) = \{y\}$, and hence both $A$ and $B$ are in cosets of $\ker \tilde{\phi}$. Since by assumption $A, B$ are not in cosets of a proper subgroup of $\mathbb{Z}^d$ this means $\ker \tilde{\phi} = \mathbb{Z}^d$, and so (examining the definition of $\tilde{\phi}$) we must have $H = \mathbb{F}_2^d$. Then our bound on $\log|H|$ forces $d \leq \frac{40}{\log 2} d[U_A; U_B]$ and we are done with $A' = A$ and $B' = B$.

Otherwise,

$$|A_x| + |B_y| < |A| + |B|.$$

By induction we can find some $A' \subseteq A_x$ and $B' \subseteq B_y$ such that $\dim A', \dim B' \leq \frac{40}{\log 2} d[U_{A_x}; U_{B_y}] \leq \frac{40}{\log 2} d[U_A; U_B]$ and

$$\log \frac{|A_x||B_y|}{|A'||B'|} \leq 34d[U_{A_x}; U_{B_y}].$$

Adding these inequalities implies

$$\log \frac{|A||B|}{|A'||B'|} \leq 34d[U_A; U_B]$$

as required. □

**Theorem 11.9.** *If $A \subseteq \mathbb{Z}^d$ is a finite non-empty set with $d[U_A; U_A] \leq \log K$ then there exists a non-empty $A' \subseteq A$ such that*

$$|A'| \geq K^{-17}|A|$$

*and* $\dim A' \leq \frac{40}{\log 2} \log K$.

*Proof.* Immediate from Theorem 11.8 and rearranging. □

**Theorem 11.10.** *Let $A \subseteq \mathbb{Z}^d$ and $|A - A| \leq K|A|$. There exists $A' \subseteq A$ such that $|A'| \geq K^{-17}|A|$ and $\dim A' \leq \frac{40}{\log 2} \log K$.*

*Proof.* As in the beginning of Theorem 7.3 the doubling condition forces $d[U_A; U_A] \leq \log K$, and then we apply Theorem 11.9. □