The Polynomial Freiman-Ruzsa Theorem

An Expository Note

Aparna Gupte
MIT

Seyoon Ragavan
MIT

September 17, 2026

Abstract

In this expository note, we motivate the polynomial Freiman-Ruzsa conjecture and describe the main ideas of its proof by Gowers, Green, Manners, and Tao [GGMT25]. Finally, we describe two applications of the PFR theorem to theoretical computer science, and end with the PBR conjecture.

Contents

1 Introduction

In these notes, we will be considering a subset $A \subseteq \mathbb{F}_2^n$. The restriction to characteristic 2 is primarily for simplicity.

Definition 1.1 (Sumsets). *For sets $A, B \subseteq \mathbb{F}_2^n$, we define $A + B = \{a + b : a \in A, b \in B\}$. We use mA to denote the set $\underbrace{A + A + \ldots + A}_{m \text{ times}}$.*

Remark 1.2. *mA should not be confused with the set $\{ma : a \in A\}$.*

Remark 1.3. *Recall that in last week's lecture (on Balog-Szemerédi-Gowers) we defined the energy $E(A, B) := \{a, a' \in A; b, b' \in B : a + b = a' + b'\}$, and saw that it is related to $|A + B|$ by $|E(A, B)| \cdot |A + B| \geq |A|^2 \cdot |B|^2$.*

Definition 1.4 (Doubling constant). *The doubling constant of a nonempty set A is defined as $K(A) := |A + A|/|A|$. We say that A has doubling constant at most K if we have $K(A) \leq K$.*

Remark 1.5 (Translation invariance). *Note that replacing A with $a + A$ for any $a \in \mathbb{F}_2^n$ preserves the sizes of both A and $A + A$, and hence also the doubling constant $K(A)$. This will enable us to assume without loss of generality that A contains 0 in certain proofs.*

This note is centred around the following question:

Motivating Question: What does a small doubling constant tell us about A?

We begin with the following straightforward observation:

Proposition 1.6. *We always have $|A + A| \geq |A|$. Equality holds if and only if A is a coset of a subgroup of $\mathbb{F}_2^n$.*

Proof. By Remark 1.5 it suffices to show the given inequality when $0 \in A$, and that under this assumption we have equality if and only if A is a subgroup. Indeed, we have $A + A \supseteq 0 + A = A \Rightarrow |A + A| \geq |A|$. Equality holds if and only if $A + A = A$ i.e., A is closed under addition and therefore a subgroup. □

So if $K = 1$, we know that A (up to shifts) must be a subgroup. This suggests answering our motivating question with a robust variant of this characterization: "even if $K > 1$, A is $\text{poly}(K)$-close to a subgroup." What does it mean to be close to a subgroup, and what characterisation should we be satisfied with? Our goal will be to find some property $\mathcal{P}_K$ that captures "closeness to a subgroup" such that:

$$|2A| \leq K|A| \Rightarrow A \text{ satisfies } \mathcal{P}_K; \text{ and} \tag{1}$$

$$A \text{ satisfies } \mathcal{P}_K \Rightarrow |2A| \leq \text{poly}(K) \cdot |A|. \tag{2}$$

Organisation. In Section 2, we present the weaker Freiman-Ruzsa theorem which considers $\mathcal{P}_K$ of the form "A can be covered by a small subgroup." We will see that this notion is insufficient, which motivates the PFR conjecture which we introduce in Section 3. In Sections 4-6, we sketch the proof by Gowers, Green, Manners, and Tao [GGMT25] of the PFR conjecture. Finally, we conclude in Sections 7 and 8 with applications of PFR to theoretical computer science and open questions.

Acknowledgements. This content is based on a presentation given by the authors for CS 2253 at Harvard in spring 2026. We thank Salil Vadhan, Jake Ruotolo, and all the class participants for discussions and comments that informed our presentation and these notes.

2 The Freiman-Ruzsa Theorem: Covering A with a Subgroup

The first natural idea for $\mathcal{P}_K$ is to try to cover A with as small a subgroup as possible. In other words, we might say A satisfies $\mathcal{P}_K$ exactly when the subgroup generated by A has size $\leq f(K, |A|)$ for some suitable function f. The Freiman-Ruzsa theorem nails down what exactly this function f can be:

Theorem 2.1 (Freiman-Ruzsa). *Assume A has doubling constant $\leq K$. Then the subgroup H generated by A has size at most $K2^{K^4}|A|$.*

In fact, this can be strengthened:

Theorem 2.2 ([Eve12]). *Assume A has doubling constant $\leq K$. Then the subgroup H generated by A has size at most $O(2^{2K}/K) \cdot |A|$.*

Note that this candidate for $\mathcal{P}_K$ satisfies (1), but not (2). Indeed, given this property, all we can infer is $|2A| \leq |H| \leq 2^{2K}|A|$, which is worse than the $\mathrm{poly}(K)$ loss we were hoping for. This suggests considering the property $\mathcal{P}_K$ to be "A is contained in a subgroup H of size at most $2^{2K}/K \cdot |A|$." We have shown that it satisfies (1). However, given this property all we can infer about A is $|2A| \leq |H| \leq 2^{2K}/K \cdot |A|$, which is worse than the $\mathrm{poly}(K)$ loss we were hoping for in (2). Unfortunately, Theorem 2.2 is tight up to constant factors, as indicated by the following example:

Example 2.3. *Assume $2K \leq n$, and let $A = \{e_1, \ldots, e_{2K}\}$, where e_i is the ith standard basis vector. Then $|A + A| = 1 + \binom{2K}{2} = 1 + K(2K-1) \leq 2K^2 = K|A|$. On the other hand, the subgroup generated by A has size $2^{2K} = \frac{2^{2K}}{2K}|A|$.*

Regardless, let us prove Theorem 2.1; it will serve as a warm-up and an opportunity to introduce a particularly nice tool, namely the Ruzsa covering. We start with some preliminary lemmas:

Lemma 2.4 (Plünnecke-Ruzsa [Plü69]). *If A has doubling constant at most K, then for any $m \geq 1$ we have $|mA| \leq K^m|A|$.*

Lemma 2.5 (Ruzsa covering lemma). *Let $A, B \subseteq \mathbb{F}_2^n$ be finite nonempty sets. Then A can be covered by $\leq |A+B|/|B|$ translates of $2B$.*

Proof. By Remark 1.5, let us assume without loss of generality that $0 \in B$. Let $\{a_1, \ldots, a_\ell\} \subseteq A$ be maximal such that the translates $\{a_i + B : i \in [\ell]\}$ are pairwise disjoint. In words, we are taking a maximal packing of $A + B$ with translates of B by elements of A. This packing cannot include too many translates, since we have

$$\bigcup_{i=1}^{\ell}(a_i + B) \subseteq A + B \Rightarrow \ell|B| \leq |A + B| \Rightarrow \ell \leq |A + B|/|B|.$$

Secondly, we claim that $A \subseteq \bigcup_{i=1}^{\ell}(a_i + 2B)$, which would imply the conclusion. Indeed, consider any $a \in A$. We have two cases:

- If $a \notin \bigcup_{i=1}^{\ell}(a_i + B)$, then we know by maximality of our packing that there exists $i \in [\ell]$ such that $a_i + B$ and $a + B$ intersect i.e., $a \in a_i + 2B$.
- If $a \in \bigcup_{i=1}^{\ell}(a_i + B)$, there must exist $i \in [\ell]$ such that $a \in a_i + B \subseteq a_i + 2B$ (noting that $0 \in B \Rightarrow B \subseteq 2B$).

This completes our proof. □

Remark 2.6. *A Ruzsa covering can be thought of as an ε-net over the space A, with the metric $\mathsf{dist}(u, v)$ being the minimal integer m such that $u - v \in mB$.*

Proof of Theorem 2.1. The idea is to show that there is a subgroup B of size $\leq 2^{K^4}$ such that $3A \subseteq 2A + B$. (Looking ahead, the reason for the constant 2 on the RHS is because this is what the Ruzsa covering lemma enables, and the reason for the 3 on the LHS is that we would like it to be bigger than the 2.) If we could show this, then we can plug this into itself repeatedly to show that A generates a small subgroup. Indeed for any $m \geq 4$, we have $mA = (m-3)A + 3A \subseteq (m-3)A + (2A + B) = (m-1)A + B$, so iterating this (and noting that $B + B = B$) implies that $mA \subseteq 2A + B$. So letting $H = \bigcup_{m=3}^{\infty}(mA)$ be the subgroup generated by A, we have $H \subseteq 2A + B$, which implies $|H| \leq |2A| \cdot |B| \leq K2^{K^4}|A|$, as desired.

Finally, our desired claim follows by applying Lemma 2.5 to the sets $3A$ and A. This implies that $3A$ can be covered by $\leq |3A+A|/|A| = |4A|/|A| \leq K^4$ (by Lemma 2.4) translates of $2A$. Labelling these translates $b_1+2A, \ldots, b_\ell+2A$ and letting B be the subgroup generated by $b_1, \ldots, b_\ell$ implies the claim. □

3 The PFR Conjecture: Covering A with Few *Cosets* of a Subgroup

In the last section, we learned that simply looking at the subgroup generated by A is too coarse. However, a finer characterisation in terms of covering A with *multiple cosets* of a subgroup works better, and this is the heart of Katalin Marton's polynomial Freiman-Ruzsa conjecture:

Conjecture 3.1 (Marton's polynomial Freiman-Ruzsa (PFR) conjecture)**.** *If A has doubling constant $\leq K$, then there exists a subgroup H of $\mathbb{F}_2^n$ such that (i) $|H| \leq |A|$; and (ii) A can be covered by at most $\mathrm{poly}(K)$ translates of H.*

The property $\mathcal{P}_K$ which this conjecture suggests is therefore "A can be covered by at most $\mathrm{poly}(K)$ translates of a subgroup of size $\leq |A|$". Now we do have the reverse implication (2)! If $A \subseteq (x_1 + H) \cup \ldots \cup (x_T + H)$, then we have $A + A \subseteq \bigcup_{i,j \in [T]} (x_i + x_j + H) \Rightarrow |A + A| \leq T^2 |H| \leq T^2 |A|$. So if $T = \mathrm{poly}(K)$ then the loss T^2 is also $\mathrm{poly}(K)$. Of course, we need to show the PFR conjecture to show (1). The progress on this was marked by two breakthroughs:

- In 2012, Sanders [San12] proved a quasipolynomial version of the Freiman-Ruzsa conjecture, with $2^{O(\log^4 K)}$ translates in place of the polynomial covering bound.
- In 2023, Gowers, Green, Manners, and Tao [GGMT25] proved the PFR conjecture!

Theorem 3.2 ([GGMT25])**.** *Conjecture 3.1 holds. Additionally, the alternate formulations stated in Conjectures 3.3, 4.5, and 5.4 all hold.*

We also state an alternate formulation of the PFR conjecture that is often easier to prove. Later, we will see two additional variants (Conjectures 4.5 and 5.4) of the PFR conjecture. We compare and contrast these four variants at the end of Section 5.2.

Conjecture 3.3 (An alternate formulation of PFR)**.** *If A has doubling constant $\leq K$, then there exists a subgroup H of $\mathbb{F}_2^n$ such that (i) $|H| \leq \mathrm{poly}(K) \cdot |A|$; and (ii) $|A \cap H| \geq |A| / \mathrm{poly}(K)$.*

Note that the conclusion of this variant of the PFR conjecture does not give a satisfactory property $\mathcal{P}_K$; it does not imply that A has a small doubling constant because it does not control the "rogue elements" of A that fall outside H. Regardless, these two conjectures turn out to be equivalent because the restriction on $A + A$ allows us to control these rogue elements:

Lemma 3.4. *Conjectures 3.1 and 3.3 are equivalent.*

Proof. Showing that Conjecture 3.1 implies Conjecture 3.3. One of the translates $x+H$ must be such that $|A \cap (x+H)| \geq |A| / \mathrm{poly}(K)$. So we can take H' to be the subgroup generated by H and x. It has size at most $|H'| \leq 2|H| \leq 2|A|$, as desired.

Showing that Conjecture 3.3 implies Conjecture 3.1. Ruzsa covering strikes again! Assume Conjecture 3.3 and let A be such that $|A + A| \leq K|A|$. Then we know that there exists a subgroup H of $\mathbb{F}_2^n$ such that $|H| \leq 2K^C |A|$ and $|A \cap H| \geq |A|/(2K^C)$, for some constant C.

By Lemma 2.5, it follows that A can be covered by $\leq |A + (A \cap H)|/|A \cap H|$ translates of $2(A \cap H) \subseteq H$. (The reason that $A \cap H$ is useful is that its containment in A allows us to use the bound on $|A + A|$, and its containment in H allows us to pass from translates of $2(A \cap H)$ to translates of H.) Now note that:

$$\frac{|A + (A \cap H)|}{|A \cap H|} \leq \frac{|A + A|}{|A \cap H|} \leq K \frac{|A|}{|A \cap H|} \leq 2K^{C+1}.$$

So A can be covered by $\mathrm{poly}(K)$ translates of a subgroup H of size $\leq \mathrm{poly}(K) \cdot |A|$. If $|H| \leq |A|$, we are already done. Otherwise, we can take H' to be a subgroup of H with $|H'| \in [|A|/2, |A|]$ and note that H can be covered by $|H|/|H'| \leq \mathrm{poly}(K)$ translates of H'. □

4 A Convenient Language: The Ruzsa Distance

We have been thinking about the doubling constant as a way of quantifying how far a set A is from being closed under addition, i.e., how far A is from being a (coset of a) subgroup. One could hope that there is some underlying metric/distance, and indeed, this is almost true.

Definition 4.1 (Ruzsa distance)**.** *For nonempty $A, B \subseteq \mathbb{F}_2^n$, the Ruzsa distance is defined as*

$$d_{\mathsf{R}}(A, B) := \log |A + B| - \frac{1}{2} \log |A| - \frac{1}{2} \log |B|.$$

We write $d_{\mathsf{R}}(A) := d_{\mathsf{R}}(A, A)$ to denote the Ruzsa distance of a set A from itself.

The Ruzsa distance is not a true distance—as we will see below, the distance from a set A to itself is 0 if and only if it is itself a (coset of a) subgroup. So for most sets A, $d_{\mathsf{R}}(A, A) \neq 0$. However, like true metrics, it is non-negative, symmetric, and satisfies the triangle inequality.

The utility of the Ruzsa distance is because it has some of the nice properties of distances such as the triangle inequality, but also serves as a progress measure for us because $d_{\mathsf{R}}(A, A)$ captures the proximity of A to a subgroup in some way. Indeed, $d_{\mathsf{R}}(A, A)$ is exactly $\log K(A)$, and $d_{\mathsf{R}}(A, B)$ can be thought of as capturing whether A and B are close to subgroups as well as how these subgroups are aligned:

Proposition 4.2 (Compare with Proposition 1.6)**.** *We have $d_{\mathsf{R}}(A, B) \geq 0$ always, with equality if and only if A, B are cosets of the same subgroup.*

Proof. By Remark 1.5, assume WLOG that $0 \in A, B$. Now the inequality follows from noting that $|A + B| \geq \max(|A|, |B|) \geq \sqrt{|A| \cdot |B|}$. For equality to hold, the second inequality forces $|A| = |B|$ and the first inequality then forces $A = A + B = B$, implying the conclusion. □

Lemma 4.3. *Both of the following hold:*

- *Symmetry: $d_{\mathsf{R}}(A, B) = d_{\mathsf{R}}(B, A)$; and*
- *Triangle inequality: $d_{\mathsf{R}}(A, B) \leq d_{\mathsf{R}}(A, C) + d_{\mathsf{R}}(C, B)$.*

Proof. Symmetry is evident. For the triangle inequality: the idea in one equation is that we can write $a + b$ as $(a + c) + (b + c)$. The desired inequality is $|A + B| \cdot |C| \leq |A + C| \cdot |B + C|$. To prove this, we will exhibit an injective function $f : (A + B) \times C \to (A + C) \times (B + C)$. For any $s \in A + B$, let $a_s \in A, b_s \in B$ be canonical such that $a_s + b_s = s$. Then we will take $f(s, c) = (a_s + c, b_s + c)$. It is injective because $(a_s + c, b_s + c)$ defines $a_s + c + b_s + c = a_s + b_s = s$, which then defines a_s and therefore $(a_s + c) - a_s = c$. □

Remark 4.4 (Some equality cases of the triangle inequality)**.** *There does not seem to be a simple characterisation of when equality holds. Here are some concrete examples of equality cases that can get messy:*

- *Let $A \subseteq C \subseteq B$ be such that B, C are subgroups (but A can be arbitrary).*
  Then $A + B = B + C = B$ and $A + C = C$. So:

$$d_{\mathsf{R}}(A, C) = \log |A + C| - \frac{1}{2} \log |A| - \frac{1}{2} \log |C| = \frac{1}{2}(\log |C| - \log |A|)$$
$$d_{\mathsf{R}}(A, B) = \log |A + B| - \frac{1}{2} \log |A| - \frac{1}{2} \log |B| = \frac{1}{2}(\log |B| - \log |A|)$$
$$d_{\mathsf{R}}(B, C) = \log |B + C| - \frac{1}{2} \log |B| - \frac{1}{2} \log |C| = \frac{1}{2}(\log |B| - \log |C|),$$

  and it is apparent that equality holds.

- *Now suppose A, B, C are all subgroups. In this case, it turns out that equality holds if and only if both (i) $A \cap B \subseteq C$; and (ii) $(B \cap C) + (A \cap C) = C$.*

It will be intuitively useful to rephrase Marton's conjecture in the language of the Ruzsa distance.

Conjecture 4.5 (Ruzsa distance PFR conjecture)**.** *There exists a constant $C > 0$ such that the following holds: for any set $A \subseteq \mathbb{F}_2^n$ such that $d_{\mathsf{R}}(A) \leq \log K$, there is a subgroup $H \leq \mathbb{F}_2^n$ such that $d_{\mathsf{R}}(A, H) \leq C \log K$.*

Lemma 4.6. *Conjectures 3.1 and 4.5 are equivalent.*

Proof. Since $d_{\mathsf{R}}(A) = \log K(A)$, the two conjectures have equivalent preconditions.

Showing that Conjecture 3.1 implies Conjecture 4.5. Let H be a subgroup such that $|H| \leq |A|$ and A is covered by at most K^C translates of H. Then $A + H$ is also covered by these same translates, so we have $|A + H| \leq K^C|H|$. Consequently, we have:

$$d_{\mathsf{R}}(A, H) = \log \frac{|A + H|}{\sqrt{|A| \cdot |H|}} \leq \log \frac{K^C \sqrt{|H|}}{\sqrt{|A|}} \leq C \log K,$$

as desired.

Showing that Conjecture 4.5 implies Conjecture 3.1. Let H be a subgroup such that $d_{\mathsf{R}}(A, H) \leq C \log K$. Then we have $|A + H| \leq K^C \sqrt{|A| \cdot |H|}$. Firstly, since $|A + H| \geq |H|$ it follows that $|H| \leq K^{2C}|A|$. Similarly, we have $|A| \leq K^{2C}|H|$. Secondly, note that $A + H$ is invariant under shifts by elements of H, so it is exactly a union of translates of H. The number of translates is therefore $|A + H|/|H| \leq K^C \sqrt{|A|/|H|} \leq K^{2C}$. These translates together form $A + H$ which covers A. Thus H almost satisfies our desiderata, except we have $|H| \leq K^{2C}|A|$ instead of $|H| \leq |A|$. This can be handled as in the end of the proof of Lemma 3.4. □

5 Proof Idea

In this section, we give the main proof idea of [GGMT25]. In Section 5.1, we lay out the high-level plan, and describe the obstacle this plan runs into. In Section 5.2 we introduce the key notion that gets around this issue. Section 6 describes the entire proof sketch.

5.1 Proof Strategy: Incremental Progress

Now that we have rephrased the PFR conjecture in terms of the Ruzsa distance in Conjecture 4.5, a natural proof strategy would be to enumerate a sequence $(A_n)_n$ of sets starting from $A_0 = A$, such that A_{n+1} is not too far (in Ruzsa distance) from A_n, yet A_{n+1} improves on the doubling constant of A_n. Once the doubling constant sinks below a certain threshold, we know we have reached a subgroup $A_N = H$. Then, we can hope to bound $d_{\mathsf{R}}(A, H)$ via the Ruzsa triangle inequality. Our task now becomes:

Given $A \subseteq \mathbb{F}_2^n$, find $A' \subseteq \mathbb{F}_2^n$ such that $d_{\mathsf{R}}(A') \leq (1 - \eta) \cdot d_{\mathsf{R}}(A)$ and $d_{\mathsf{R}}(A, A') \leq C \cdot d_{\mathsf{R}}(A)$,
for some universal constants $C > 0$ and $0 < \eta < 1$.

Taking the sumset of a random dense subset of a subgroup. A natural idea would be to consider $A' = A + A$. If $K = \frac{|2A|}{|A|}$, by Lemma 2.4, we know that $A + A$ is not too far from A

$$d_{\mathsf{R}}(A, A + A) = \log \frac{|3A|}{\sqrt{|A| \cdot |2A|}} \leq \log \frac{K^3 \cdot |A|}{\sqrt{K} \cdot |A|} = \frac{5}{2} d_{\mathsf{R}}(A).$$

And indeed, for many instances of A, taking the sumset substantially improves the doubling constant. For example, form a random subset $A \subseteq H$ of a subgroup $H \leq \mathbb{F}_2^n$ by including each element independently with probability $1/\ell$. When $|H| \gg \ell^2$, we typically have $|A| \approx |H|/\ell$. For each nonzero $v \in H$, there are $|H|/2$ disjoint pairs $\{h, h+v\}$, so $\Pr[v \notin 2A] = (1 - 1/\ell^2)^{|H|/2} \leq \exp\big(-|H|/(2\ell^2)\big)$. The expected fraction of missing nonzero elements is therefore small. Thus $A' = A + A \subseteq H$ typically occupies most of H, and

$$K(A + A) = \frac{|4A|}{|2A|} \leq \frac{|H|}{\Omega(|H|)} \leq O(1).$$

For large ℓ, this is a substantial improvement on the typical doubling constant of A:

$$K(A) = \frac{|2A|}{|A|} \approx \frac{|H|}{|H|/\ell} = \ell,$$

which grows with ℓ, while $K(A+A)$ stays bounded.

While the sumset operation makes progress when A is a dense subset of a subgroup, it actually makes things a lot worse when A is structured as a (sparse) union of ℓ unevenly-spaced cosets of a subgroup

$$A = \bigcup_{i \in [\ell]} H + v_i.$$

Assuming that the v_i's are randomly spaced, we expect $|A+A| = \left|\bigcup_{i,j\in[\ell]} H + v_i + v_j\right| \approx \ell^2|H|$, so taking the sumset degrades the doubling constant by a factor of ℓ

$$K(A) = \frac{|2A|}{|A|} \approx \frac{\ell^2|H|}{\ell|H|} = \ell, \text{ and}$$
$$K(A+A) = \frac{|4A|}{|2A|} \approx \frac{\ell^4|H|}{\ell^2|H|} = \ell^2.$$

Picking out a coset from a random sparse collection of cosets. For such $A = \bigcup_{i\in[\ell]} H + v_i$, however, there is a natural candidate for A', namely, H (or a coset of H)! Taking $A' = H$ drops the doubling constant to 1, and $A + H = A$, so H is not too far from A:

$$d_{\mathsf{R}}(A, H) = \log \frac{|A+H|}{\sqrt{|A|\cdot|H|}} = \log \frac{|A|}{\sqrt{|A|\cdot|H|}} = \frac{1}{2}\log \ell,$$

so in the sparse-coset example above, where $K \approx \ell$, we have $d_{\mathsf{R}}(A, H) = O(\log K)$ for large ℓ.

It turns out that taking $A' = A \cap (A + v)$ for a random $v \leftarrow A + A$ almost achieves this operation. Assume that the cosets $H + v_i$ are distinct and that their sums $H + v_i + v_j$ for $i < j$ are all distinct. For $v \in (A+A) \setminus H$, relabel the representatives so that $v = h + v_1 + v_2$ for some $h \in H$. Then $A + v = \bigcup_{i\in[\ell]} H + v_i + v_1 + v_2$, and $A' = A \cap (A+v) = (H+v_1) \cup (H+v_2)$. This is a coset of the subgroup generated by H and $v_1 + v_2$, so $K(A') = 1$. If $v \in H$, then $A' = A$; this occurs with probability $1/(1 + \binom{\ell}{2})$ when v is uniform on $A + A$.

Maybe at least one of these operations makes progress? We have seen two "orthogonal" ways in which A can have large doubling constant:

1. A is a random dense subset of a subgroup, and
2. A is a sparse random collection of cosets of a subgroup.

In the first case, taking the sumset $A' = A + A$ allowed us to make progress, and in the second case, trying to pick out a few of the cosets $A' = A \cap (A + v)$ for random $v \leftarrow A + A$ allowed us to make progress. One could hope that for *any* set A, at least one of these operations makes progress. To justify this hope, consider a set A that is a combination of the above two cases: let A be a random subset of $\bigcup_{i\in[\ell_2]} H + v_i$, with density $\frac{1}{\ell_1}$, and assume that the cosets and their pairwise sums satisfy the distinctness conditions above. Then, $|A| = \frac{\ell_2}{\ell_1}|H|$, $|2A| \approx \ell_2^2|H|$, and

$$K(A) = \frac{|2A|}{|A|} \approx \ell_1 \cdot \ell_2.$$

Taking the sumset gives doubling constant

$$K(A+A) = \frac{|4A|}{|2A|} \approx \frac{\ell_2^4|H|}{\ell_2^2|H|} = \ell_2^2,$$

For a typical $v \leftarrow A + A$ outside H, the intersection $A \cap (A + v)$ has density approximately $1/\ell_1^2$ in each of two cosets. In the regime where $|H| \gg \ell_1^4$, its sumset occupies a constant fraction of each of the two resulting cosets, so

$$|A \cap (A + v)| \approx \frac{1}{\ell_1^2}|H|,$$
$$K(A \cap (A + v)) = \frac{|(A \cap (A + v)) + (A \cap (A + v))|}{|A \cap (A + v)|} \approx \frac{|H|}{\frac{1}{\ell_1^2}|H|} = \ell_1^2.$$

That is, for random $v \leftarrow A + A$, we expect the original doubling constant $K(A)$ to be the geometric mean of the two resulting doubling constants, implying that at least one of the operations must be making progress.

We could hope that any set A with large doubling constant has approximately such structure, leading us to conjecture that

$$K(A)^2 \overset{?}{\geq} K(A + A) \cdot K(A \cap (A + v)) \tag{3}$$

holds for all A. In the language of Ruzsa distance,

$$2 \cdot d_{\mathsf{R}}(A) \overset{?}{\geq} d_{\mathsf{R}}(A + A) + d_{\mathsf{R}}(A \cap (A + v)). \tag{4}$$

This choice of two operations is not entirely ad hoc—considering the homomorphism $\pi : A \times A \to A + A$ where $\pi(a, b) = a + b$, we notice that for some $v \in A + A$, its fiber is

$$\begin{aligned}\pi^{-1}(v) &= \{(a, b) \in A \times A : a + b = v\} \\ &= \{(a, v - a) : a, v - a \in A\} \\ &\cong A \cap (A + v).\end{aligned}$$

It is no coincidence that our two candidate operations for obtaining A' are the image and fibers of some homomorphism. Imagine for a moment that our set is actually a group $A = H$. Then for $\pi : H \times H \to H + H$, the first isomorphism theorem tells us that $H \times H \cong \operatorname{im}\pi \times \ker\pi$, so the doubling constants on both sides have to equal:

$$\begin{aligned}K(H \times H) = K(H)^2 = K(\operatorname{im}\pi) \cdot K(\ker\pi), \text{ or} \\ 2 \cdot d_R(H) = d_R(\operatorname{im}\pi) + d_R(\ker\pi)\end{aligned} \tag{5}$$

Here $\operatorname{im}\pi = H + H$, and $\ker\pi = \pi^{-1}(0)$ is exactly the fiber of 0. But of course Equation (5) is vacuous—the doubling constant of a group is always 1.

In our setting, where A is a set, $A \times A$ does not nicely decompose into a product of the image and the kernel, since the different fibers can have vastly different structure in general—it decomposes into what is called a *skew* product

$$A \times A = \bigsqcup_{v \in A+A} \{v\} \times (A \cap (A + v)).$$

These preimage sets $A \cap (A + v)$ could have wildly differing sizes, structures and doubling constants, so it is not clear that we can argue that at least one of these operations will decrease the doubling constant. To try to make this work, we need some way of averaging over all these fibers. For example, we could conjecture

$$2 \cdot d_{\mathsf{R}}(A) \overset{?}{\geq} d_{\mathsf{R}}(A + A) + \mathop{\mathbb{E}}_{v \leftarrow A+A}\big[d_{\mathsf{R}}(A \cap (A + v))\big].$$

Unfortunately, it turns out that this is actually false: there exist sets A such that $d_{\mathsf{R}}(A + A) > 2 \cdot d_{\mathsf{R}}(A)$.

The beautiful insight of [GGMT25] is that using the Ruzsa distance/doubling constant as a measure of progress is too pessimistic. This is because Ruzsa distance treats all elements in the sumset $A + A$ equally, but in reality some elements might appear more frequently than others as the sum of pairs $(a, b) \in A \times A$. Taking this entire *histogram* of elements in $A + A$ into account, $A + A$ might already be concentrated on a set with small doubling constant, but the Ruzsa distance $d_{\mathsf{R}}(A)$ is too coarse to capture this progress. As a workaround, Gowers et al. replace sets with random variables, and the Ruzsa distance with its entropic cousin, which *does* capture progress in the form of histograms concentrated on groups.

5.2 A Finer Measure of Progress: The Entropic Ruzsa Distance

Definition 5.1. *For **independent** random variables X, Y supported on $\mathbb{F}_2^n$, define the **entropic** Ruzsa distance as:*

$$d_{\mathsf{ER}}(X,Y) := H(X+Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y),$$

where $H(\cdot)$ is the Shannon entropy.[1] For arbitrary X, Y, we compute this distance using independent copies, so it depends only on their distributions. As in Definition 4.1, we write $d_{\mathsf{ER}}(X) := d_{\mathsf{ER}}(X_1, X_2)$, where X_1, X_2 are independent copies of X, to denote the entropic Ruzsa distance of X from itself.

The entropic Ruzsa distance looks very similar to the Ruzsa distance, and shares many of its properties. Like the Ruzsa distance, the entropic Ruzsa distance is also almost a metric—it satisfies non-negativity, symmetry and the triangle inequality, and reaches 0 only with uniform random variables over cosets of the same group (see Section B). If U_A, U_B are random variables that are uniform on sets $A, B \subseteq \mathbb{F}_2^n$ respectively, it might look like we exactly recover the Ruzsa distance between A, B. However, this is not true! While the sumset $A+B$ treats all elements $c \in A+B$ equally, the sum $U_A + U_B$ of the corresponding uniform random variables captures the entire histogram of how often an element $c \in A+B$ appears as a sum of $(a,b) \in A \times B$. This difference is what makes the entropic Ruzsa distance a better measure of progress.

Lemma 5.2 (Comparison with Ruzsa distance)**.** *For all sets $A, B \subseteq \mathbb{F}_2^n$, we have $d_{\mathsf{ER}}(U_A, U_B) \le d_{\mathsf{R}}(A,B)$.*

Even stronger, the entropic Ruzsa distance gives us the right analog of Equation (4)—the fibering identity. This identity is what makes the entropic Ruzsa distance so suitable in finishing the proof strategy in Section 5.1.

Lemma 5.3 (Fibering identity)**.** *Let Y be an $\mathbb{F}_2^n$-valued random variable, and let $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a homomorphism, where n and m may differ. Then,*

$$d_{\mathsf{ER}}(Y) = d_{\mathsf{ER}}\big(\pi(Y)\big) + d_{\mathsf{ER}}\big(Y|\pi(Y)\big) + \Delta,$$

where the conditional entropic Ruzsa distance $d_{\mathsf{ER}}(Y|\pi(Y))$ is defined as in Definition B.3, and the defect Δ is

$$\Delta = I\big[Y_1 + Y_2 : \big(\pi(Y_1), \pi(Y_2)\big) \mid \pi(Y_1) + \pi(Y_2)\big],$$

where we write Y_1, Y_2 for independent copies of the same random variable Y.

In particular, since $\Delta \ge 0$, applying the fibering identity to our setting with the addition map $\pi(X_1, X_2) = X_1 + X_2$, we get the entropic analog of Equation (4) for free. The fibering identity also matches the intuition of decomposing the product $A \times A$ into $A + A$ and the fibers of the addition map—note how in Lemma 5.3, $\pi(Y)$ is the image of π, and the conditional random variable $Y|\pi(Y)$ looks like an average over the fibers of π.

We defer proofs of Lemma 5.2 and Lemma 5.3, as well as other useful properties of the entropic Ruzsa distance to Section B.

The Entropic PFR Conjecture. Let us first rephrase Conjecture 4.5 in the language of *entropic* Ruzsa distance.

Conjecture 5.4 (Entropic Marton's conjecture)**.** *Let X be a $\mathbb{F}_2^n$-valued random variable with $d_{\mathsf{ER}}(X) \le \log K$. Then, there exists a uniform random variable U_H on a subgroup $H \subseteq \mathbb{F}_2^n$ such that $d_{\mathsf{ER}}(X, U_H) \le C \log K$ for some universal constant $C > 0$.*

Lemma 5.5. *Conjecture 5.4 implies Conjecture 3.3.*

Proof. Assume Conjecture 5.4 and let $A \subseteq \mathbb{F}_2^n$ be such that $|A+A| \le K|A|$. Then by Lemma 5.2, we have $d_{\mathsf{ER}}(U_A, U_A) \le d_{\mathsf{R}}(A,A) = \log|A+A| - \log|A| \le \log K$. So Conjecture 5.4 tells us that there is a subgroup H such that $d_{\mathsf{ER}}(U_A, U_H) \le C \log K \Leftrightarrow H(U_A + U_H) - \frac{1}{2}\log|A| - \frac{1}{2}\log|H| \le C \log K$.

[1] See Section A for the information theory background relevant to this note.

Consider the random variable A' (supported on the quotient group $\mathbb{F}_2^n/H$) induced by sampling $a \leftarrow A$ uniformly then taking the coset $a + H$. Then we have $H(U_A + U_H) = H(A') + \log|H|$, so we have $H(A') + \frac{1}{2}\log|H| - \frac{1}{2}\log|A| \leq C\log K$. The first implication of this is that $\frac{1}{2}\log|H| - \frac{1}{2}\log|A| \leq C\log K \Rightarrow |H| \leq K^{2C}|A|$. Secondly, by the lower bound in Proposition A.5, there must exist a coset $a + H$ such that:

$$\begin{aligned}
-\log\Pr[U_A \in a+H] \leq H(A') &\leq C\log K + \frac{1}{2}\log|A| - \frac{1}{2}\log|H| \\
\Leftrightarrow \log\Pr[U_A \in a+H] &\geq -C\log K - \frac{1}{2}\log|A| + \frac{1}{2}\log|H| \\
\Leftrightarrow \Pr[U_A \in a+H] &\geq \frac{\sqrt{|H|}}{K^C\sqrt{|A|}} \\
\Leftrightarrow |A \cap (a+H)| &\geq \frac{\sqrt{|A|\cdot|H|}}{K^C}.
\end{aligned}$$

The LHS is bounded above by $|H|$ so this tells us that $|H| \geq |A|/K^{2C}$. Therefore we have:

$$|A \cap (a+H)| \geq \frac{\sqrt{|A|\cdot|H|}}{K^C} \geq \frac{|A|}{K^{2C}}.$$

So if we define H' to be the subgroup generated by H and a, then we have $|H'| \leq 2|H| \leq 2K^{2C}|A|$ and $|A \cap H'| \geq |A|/K^{2C}$, implying the conclusion of Conjecture 3.3 as desired. □

Comparing PFR Variants. At this point, we have seen four variants of the PFR conjecture. We have shown that they are all equivalent, but it is worth pausing to reflect on the relative strength of their conclusions. In the following, we refer to Conjectures 3.1 and 4.5 as the "strong variants" and Conjectures 3.3 and 5.4 as the "weak variants".

- Proving that a strong variant implies any variant, or that the weak variants imply each other, is essentially just a matter of shuffling expressions around.
- However, proving that a weak variant implies a strong variant (see Lemma 3.4) requires (a) additional use of the precondition that A has a small doubling constant; and (b) the Ruzsa covering lemma.

  Intuitively, this is because the two weak variants do not fully control A. In Conjecture 3.3, the conclusion does not say anything about the elements of A that fall outside the subgroup H; in Conjecture 5.4, we can infer that there is a quotient-valued random variable A' that records the coset of H containing a sample from X and has low entropy, but what we really need is that it has small support. In both cases, we get around this (in Lemma 3.4) by using the bound on $A + A$ (via Ruzsa covering) to establish full control over A.
- Finally, note that upper bounds on $|H|$ are explicitly specified by Conjectures 3.1 and 3.3 (the two variants that do not make reference to Ruzsa distances), but not by Conjectures 4.5 and 5.4. The reason for this is that the constraint on the Ruzsa distance of A from a subgroup implicitly forces H to be not too large.

6 Proof Sketch of the Entropic PFR Conjecture

In this section, we give a proof sketch of Conjecture 5.4, and therefore also Conjecture 3.1. First, we argue that it suffices to prove the following "distance decrement" statement.

Conjecture 6.1 (Distance decrement). *There exist universal constants $C > 0, 0 < \eta < 1$ such that for every $\mathbb{F}_2^n$-valued random variable X, there exists an $\mathbb{F}_2^n$-valued random variable X' such that*

$$d_{\mathsf{ER}}(X') \leq (1-\eta)\cdot d_{\mathsf{ER}}(X), \text{ and} \tag{6}$$
$$d_{\mathsf{ER}}(X, X') \leq C \cdot d_{\mathsf{ER}}(X). \tag{7}$$

Claim 6.2. *Conjecture 6.1 implies Conjecture 5.4.*

Proof. Suppose Conjecture 6.1 holds. Then we can take a sequence of such random variables $(X_n)_n$ where $X_0 = X$ and X_n for $n \geq 1$ is such that $d_{\mathsf{ER}}(X_n) \leq (1-\eta)^n \, d_{\mathsf{ER}}(X)$ and $d_{\mathsf{ER}}(X_n, X_{n-1}) \leq C(1-\eta)^{n-1} \, d_{\mathsf{ER}}(X)$. Then, by weak compactness and continuity of the entropic Ruzsa distance, there exists a limiting random variable X_∞ such that $d_{\mathsf{ER}}(X_\infty) = 0$ and $d_{\mathsf{ER}}(X_\infty, X) \leq \frac{C}{\eta} \, d_{\mathsf{ER}}(X)$. Then Proposition B.2 tells us that there must be a subgroup H of $\mathbb{F}_2^n$ such that $d_{\mathsf{ER}}(X_\infty, U_H) = 0$. Therefore, by the entropic Ruzsa distance triangle inequality,

$$d_{\mathsf{ER}}(X, U_H) \leq d_{\mathsf{ER}}(X, X_\infty) + d_{\mathsf{ER}}(X_\infty, U_H) \leq \frac{C}{\eta} \cdot d_{\mathsf{ER}}(X).$$

□

We now attempt to prove Conjecture 6.1.[2] For convenience, let us call a random variable X' nearby, if it satisfies Equation (7).[3]

Definition 6.3 (C-nearby (conditional) random variable with respect to X)**.** *Let $C > 0$ be a constant. For a $\mathbb{F}_2^n$-valued random variable X, a $\mathbb{F}_2^n$-valued random variable X' is C-nearby if $d_{\mathsf{ER}}(X, X') \leq C \cdot d_{\mathsf{ER}}(X)$. Similarly, if A, X' are $\mathbb{F}_2^n$-valued random variables, the conditional random variable $X'|A$ is C-nearby if $d_{\mathsf{ER}}(X'|A, X) \leq C \cdot d_{\mathsf{ER}}(X)$.*

If X_1, X_2 are two independent copies of X, by Claim C.3, we know that $X_1 + X_2$ and $X_1|X_1 + X_2$ are both $\frac{3}{2}$-nearby random variables (cf. the two operations in Section 5.1). Now, if either

$$\begin{aligned} d_{\mathsf{ER}}(X_1 + X_2) &\leq (1-\eta) \cdot d_{\mathsf{ER}}(X), \text{ or} \\ d_{\mathsf{ER}}(X_1|X_1 + X_2) &\leq (1-\eta) \cdot d_{\mathsf{ER}}(X), \end{aligned}$$

we are done.[4] Otherwise, we have that both

$$d_{\mathsf{ER}}(X_1 + X_2) > (1-\eta) \, d_{\mathsf{ER}}(X), \text{ and} \tag{8}$$

$$d_{\mathsf{ER}}(X_1|X_1 + X_2) > (1-\eta) \, d_{\mathsf{ER}}(X), \tag{9}$$

and we now have to look for a different nearby random variable X' with small enough $d_{\mathsf{ER}}(X')$. Luckily, such a random variable conveniently falls out from the fibering identity (Lemma 5.3). Specializing the general fibering identity to our map of interest $\pi(X_1, X_2) = X_1 + X_2$, we get that

$$\begin{aligned} 2 \cdot d_{\mathsf{ER}}(X) &= d_{\mathsf{ER}}((X_1, X_2)) \\ &= d_{\mathsf{ER}}(X_1 + X_2) + d_{\mathsf{ER}}(X_1|X_1 + X_2) + \Delta, \end{aligned}$$

where

$$\begin{aligned} \Delta &= I\big((X_1 + X_3, X_2 + X_4); (X_1 + X_2, X_3 + X_4) \big| X_1 + X_2 + X_3 + X_4\big) \\ &= I\big(X_1 + X_3; X_1 + X_2 \big| X_1 + X_2 + X_3 + X_4\big), \end{aligned}$$

[2] For reasons that we comment on in a later footnote, we will not be able to prove this distance decrement conjecture, but will prove a related distance decrement statement that suffices to prove Conjecture 5.4.

[3] In [Tao24], Tao refers to these as relevant random variables.

[4] Technically, we are not quite done, since $X_1|X_1 + X_2$ is a *conditional* random variable, not an ordinary random variable like Conjecture 6.1 asks for. In general, if $d_{\mathsf{ER}}(X|A) = \mathbb{E}_{a,a'}[d_{\mathsf{ER}}(X_a, X_{a'})]$ is small, all we can conclude is that there exist a, a' such that $d_{\mathsf{ER}}(X_a, X_{a'})$ is small. (To appease Conjecture 6.1, we would like to show that there exists a such that $d_{\mathsf{ER}}(X_a)$ is small.) This can be fixed by modifying Conjecture 6.1 to the following statement: there exist universal constants $C > 0$ and $0 < \eta < 1$ such that, for every pair of $\mathbb{F}_2^n$-valued random variables X, Y, there are $\mathbb{F}_2^n$-valued random variables X', Y' satisfying

$$\begin{aligned} d_{\mathsf{ER}}(X', Y') &\leq (1-\eta) d_{\mathsf{ER}}(X, Y), \\ d_{\mathsf{ER}}(X, X'), \; d_{\mathsf{ER}}(Y, Y') &\leq C d_{\mathsf{ER}}(X, Y). \end{aligned}$$

See [Tao24, Proposition 3].

where X_1, X_2, X_3, X_4 are four independent copies of X. In the case where Equation (8) and Equation (9) hold, the fibering identity means that $\Delta < 2\eta \cdot d_{\mathsf{ER}}(X)$. For the sake of readability, let us name the following sums of random variables:

$$\begin{aligned} U &:= X_1 + X_2, \\ V &:= X_1 + X_3, \\ W &:= X_2 + X_3, \\ S &:= X_1 + X_2 + X_3 + X_4. \end{aligned}$$

Since we are in characteristic 2, we have that $W = U + V$. Since U, V, W are identically distributed, even conditioned on S, we have that $H(U|S) = H(V|S) = H(W|S)$.

Suppose for a moment that $\Delta = 0$, so that U, V are independent conditioned on S. Defining $U_s := (U|S = s)$, $V_s := (V|S = s)$, $W_s := (W|S = s)$, conditional independence means that the sum of two independent copies $U_s + V_s$ is the same as the sum of the random variable $U + V$ conditioned on $S = s$. That is,

$$\Delta = 0 \quad \Longrightarrow \quad U_s + V_s = \Big((U+V)|S = s\Big) = W_s.$$

In this idealized $\Delta = 0$ setting, we claim that there exists some s such that U_s has small entropic doubling constant:

$$\begin{aligned} \mathop{\mathbb{E}}_s\big[d_{\mathsf{ER}}(U_s)\big] &= \mathop{\mathbb{E}}_s\big[d_{\mathsf{ER}}(U_s, V_s)\big] \\ &= \mathop{\mathbb{E}}_s\big[H(U_s + V_s) - \frac{1}{2}H(U_s) - \frac{1}{2}H(V_s)\big] \\ &= \mathop{\mathbb{E}}_s\big[H(W_s) - \frac{1}{2}H(U_s) - \frac{1}{2}H(V_s)\big] \\ &= 0, \end{aligned}$$

which is a massive decrement.

However, U, V are not conditionally independent in general. The Balog-Szemerédi-Gowers theorem (Lemma B.5) allows us to salvage a decrement that will be good enough:

$$\begin{aligned} \mathop{\mathbb{E}}_s \mathop{\mathbb{E}}_w\big[d_{\mathsf{ER}}(U_{s,w}, V_{s,w})\big] &\le \mathop{\mathbb{E}}_s\big[3\, I(U_s : V_s) + 2H(W_s) - H(U_s) - H(V_s)\big] \\ &= 3\, I(U : V|S) + 2H(W|S) - H(U|S) - H(V|S) \\ &= 3\, I(U : V|S) + 0 \\ &= 3\Delta < 6\eta\, d_{\mathsf{ER}}(X). \end{aligned}$$

Here the averages are over $s \leftarrow S$ and $w \leftarrow (W \mid S = s)$, and $U_{s,w} := (U \mid S = s, W = w)$ and $V_{s,w} := (V \mid S = s, W = w)$. Since $V_{s,w}$ is a translate of $U_{s,w}$, we have $d_{\mathsf{ER}}(U_{s,w}, V_{s,w}) = d_{\mathsf{ER}}(U_{s,w})$. By Claim C.3 and Lemma B.4, we can show that $U|(S, W)$ is nearby with respect to X. Taking $0 < \eta < 1/12$, we may apply Lemma C.1 with 6η in place of its parameter η and $\eta' = 1/2$. This gives some s, w for which $U_{s,w}$ is nearby and $d_{\mathsf{ER}}(U_{s,w}) \le \frac{1}{2} d_{\mathsf{ER}}(X)$.

Remark 6.4. *We conclude this section by observing that given a random variable X, one of the following three conditional random variables gives the required distance decrement:*

1. $X_1 + X_2$,
2. $X_1|X_1 + X_2$,
3. $X_1 + X_2|(X_2 + X_3, X_1 + X_2 + X_3 + X_4)$,

where the third conditional random variable (in the proof's notation, this is $U|(S, W)$) is obtained by composing the first two:

$$\begin{aligned} (X_2|X_2 + X_3) + (X_1|X_1 + X_4) &= X_1 + X_2|(X_2 + X_3, X_1 + X_4) \\ &= X_1 + X_2|(X_2 + X_3, X_1 + X_2 + X_3 + X_4). \end{aligned}$$

7 Applications of the PFR Conjecture

7.1 The Hadamard Extractor

Definition 7.1 (Two-source extractors). *Let U, V be sets of size 2^n. A function $E : U \times V \to \{0,1\}$ is an (n, k, ε)-two-source extractor if for any independent random variables X, Y supported on U, V such that $H_\infty(X), H_\infty(Y) \geq k$, we have $\Pr[E(X,Y) = 0], \Pr[E(X,Y) = 1] \in [1/2 - \varepsilon, 1/2 + \varepsilon]$.*

Definition 7.2 (Two-source dispersers). *E is an (n, k)-two-source disperser if for any independent random variables X, Y supported on U, V such that $H_\infty(X), H_\infty(Y) \geq k$, we have $\Pr[E(X,Y) = 0], \Pr[E(X,Y) = 1] \in (0, 1)$.*

Definition 7.3 (Hadamard extractor). *Now suppose U, V are subsets of $\mathbb{F}_2^m$ of size 2^n. The Hadamard extractor $E_{U,V}$ is defined by $E_{U,V}(x, y) = \langle x, y \rangle$.*

Note that the Hadamard extractor should really be thought of as a function from $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. The inputs are first converted into elements of U and V using some canonical encoding, then the inner product is taken. So the design choice in the Hadamard extractor lies in the choice of the sets U and V.

The following theorem informally states that the Hadamard construction can automatically be "upgraded" from being a disperser to being an extractor. We note that we have other constructions of extractors achieving better parameters; the reason to study the properties of the Hadamard extractor is for its simplicity.

Theorem 7.4. *Assume the PFR conjecture and assume U, V are such that the Hadamard extractor on U, V is an (n, k)-disperser. Then for any constant $\varepsilon > 0$ there is a constant $c := c(\varepsilon) > 0$ such that the Hadamard extractor is also an $(n, k + c(\varepsilon)m/\log m, \varepsilon)$-extractor.*

We sketch the ingredients of the proof here.

Conjecture 7.5 (γ-approximate duality conjecture). *Assume $A, B \subseteq \mathbb{F}_2^n$ are such that $\langle a, b \rangle$ is ε-biased for $a, b \leftarrow A \times B$. Then the conjecture states that there exist subsets $A' \subseteq A, B' \subseteq B$ such that $\langle a', b' \rangle$ is constant for all $a', b' \in A' \times B'$, with $|A|/|A'|, |B|/|B'| \leq 2^{c\gamma(n)}$, where $c = c(\varepsilon)$ is some constant.*

It was shown by Ben-Sasson, Lovett, and Ron-Zewi [BLR12] that the PFR conjecture implies the approximate duality conjecture with $\gamma(n) = n/\log n$. This can be directly used to prove Theorem 7.4 via the following argument:

- One can assume WLOG that the random variables X, Y are independent and uniform on some subsets of size $2^{k + c(\varepsilon)m/\log m}$, and assume for the sake of contradiction that on these sets the inner product is biased.
- We can then apply the AD conjecture to these subsets to find a large subgraph on which all inner products are 0 or all inner products are 1.
- Then by the condition that E is a disperser, this forces this large subgraph to have $< 2^k$ vertices on one of the sides, which gives a contradiction.

7.2 The BLR Linearity Test

Imagine that a large technology corporation releases an oracle for some function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, and claims that f is linear. Can we efficiently (i.e., with few oracle queries) validate their claim? The answer is evidently no: f could be linear on every input except one, and we would need to make exponentially many queries to find this exception.

So the tech corporation begrudgingly settles for the more modest claim that f is close to a linear function (i.e., there is some linear function that it agrees with on at least $1 - \varepsilon$ fraction of inputs). This can be validated with the celebrated Blum-Luby-Rubinfeld (BLR) test [BLR90]: sample random points $x, y \leftarrow \mathbb{F}_2^n$ and check whether $f(x + y) = f(x) + f(y)$. The soundness of this test admits a beautiful analysis via Fourier techniques:

Theorem 7.6. *If the BLR test accepts with probability $\geq \delta$, then there exists a linear function with which f agrees on $\geq \delta$ fraction of inputs.*

Proof Sketch. Recall the Fourier decomposition of f: for each vector $y \in \mathbb{F}_2^n$ we define $\chi_y : \mathbb{F}_2^n \to \mathbb{R}$ by $\chi_y(x) = (-1)^{\langle x,y\rangle}$ and we can write $F(x) := (-1)^{f(x)}$ as

$$F(x) = \sum_{y \in \mathbb{F}_2^n} \hat{F}(y)\chi_y(x).$$

Then the BLR test accepting with probability $\geq \delta$ is exactly equivalent to:

$$\mathop{\mathbb{E}}_{x,y \leftarrow \mathbb{F}_2^n}[F(x)F(y)F(x+y)] \geq 2\delta - 1. \tag{10}$$

A direct computation establishes that:

$$\mathop{\mathbb{E}}_{x,y \leftarrow \mathbb{F}_2^n}[F(x)F(y)F(x+y)] = \sum_{z \in \mathbb{F}_2^n} \hat{F}(z)^3 \leq \max_{z \in \mathbb{F}_2^n} \hat{F}(z),$$

so there exists some $z \in \mathbb{F}_2^n$ such that $\hat{F}(z) \geq 2\delta - 1$. This is equivalent to f agreeing with the linear function $\langle z, \cdot \rangle$ on at least δ fraction of inputs. □

An attractive feature of this analysis is that it gives a handle on the BLR test even in the low-soundness regime (δ close to 0). Now let us tweak the setting slightly: assume f is vector valued so that it now maps $\mathbb{F}_2^n \to \mathbb{F}_2^m$ for some $m > 1$. It turns out that we do not know how to adapt the Fourier soundness analysis to this setting, because (10) is relying heavily on the fact that f has scalar outputs.

On the other hand, there is an alternate analysis (in fact, the original one by [BLR90]) based on majority decoding that adapts to the vector-valued setting, but does not work in the low-soundness regime. The best result we have for vector-valued functions in the low-soundness regime actually uses the PFR conjecture, as shown by Samorodnitsky [Sam07]:

Theorem 7.7 (Samorodnitsky). *Assuming the PFR conjecture, if the BLR test accepts with probability $\geq \delta$, then there exists a linear function with which f agrees on $\geq \mathrm{poly}(\delta)$ fraction of inputs.*

Proof Sketch. Consider the set $A = \{(x, f(x)) : x \in \mathbb{F}_2^n\} \subseteq \mathbb{F}_2^{n+m}$. We have $|A| = 2^n$. Then the BLR test accepting with probability $\geq \delta$ is exactly equivalent to $\Pr_{a,b \leftarrow A}[a+b \in A] \geq \delta$. By the Balog-Szemerédi-Gowers theorem and Lemma 2.4, it can be shown that there is a set $A' \subseteq A$ of size $\geq \mathrm{poly}(\delta) \cdot 2^n$ such that $|A' + A'| \leq \mathrm{poly}(1/\delta) \cdot |A'|$. By Conjecture 3.3 (which can be shown to follow from Conjecture 3.1), this means there is a subgroup $H \subseteq \mathbb{F}_2^{n+m}$ with $|H| \leq \mathrm{poly}(1/\delta) \cdot |A'|$ such that $|H \cap A'| \geq \mathrm{poly}(\delta) \cdot |A'|$. Informally, H captures a union of linear maps from $\mathbb{F}_2^n \to \mathbb{F}_2^m$. One of these linear maps will agree with f on at least $\mathrm{poly}(\delta)$ fraction of inputs. □

8 The PBR conjecture

A strengthening of PFR, called the Polynomial Bogolyubov–Ruzsa conjecture is still open. The proof of the PFR conjecture shows that there exists a subspace H contained in tA where $t = O(\log\log K)$ such that $|H| \geq K^{-c}|A|$. The PBR conjecture says that $t = O(1)$ suffices. Curiously, the PBR conjecture would imply the PFR conjecture, and the partial result by Sanders [San12] towards PFR took this route; however, the proof by [GGMT25] does not so the below remains open.

Conjecture 8.1 (Polynomial Bogolyubov–Ruzsa). *There are absolute constants $t, c > 0$ such that if $A \subseteq \mathbb{F}_2^n$ satisfies*

$$|A + A| \leq K|A|,$$

then tA contains a subspace $H \leq \mathbb{F}_2^n$ with

$$|H| \geq K^{-c}|A|.$$

References

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In Harriet Ortiz, editor, *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*, pages 73–83. ACM, 1990. 13, 14

[BLR12] Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 177–186. IEEE Computer Society, 2012. 13

[Eve12] Chaim Even-Zohar. On sums of generating sets in $\mathbb{Z}_2^n$. *Combinatorics, Probability and Computing*, 21(6):916–941, 2012. 3

[GGMT25] William Timothy Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of Marton. *Annals of Mathematics*, 201(2):515–549, 2025. 1, 2, 4, 6, 8, 14

[Plü69] Helmut Plünnecke. Eigenschaften und Abschätzungen von Wirkungsfunktionen. 1969. 3

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In David S. Johnson and Uriel Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 506–515. ACM, 2007. 14

[San12] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, October 2012. 4, 14

[Tao24] Terence Tao. An abridged proof of Marton's conjecture, 2024. 11

A Background on Information Theory

Definition A.1 (Entropy). *For a random variable X supported on a finite set $\mathcal{X}$, define the entropy:*

$$H(X) := -\sum_{x \in \mathcal{X}} \Pr[X = x] \log \Pr[X = x].$$

We will make a slight abuse of notation and let $H(X, Y)$ denote the entropy of the joint distribution of X, Y.

Definition A.2 (Min-entropy). *The min-entropy $H_\infty(X)$ is defined to be $-\log \max_{x \in \mathcal{X}} \Pr[X = x]$.*

One can informally think of entropy as the "expected number of bits" required to write down a sample from X.

Definition A.3 (Conditional entropy). *For an event E, we can define the conditional entropy $H(X|E)$ to be the entropy of the conditional distribution of X given E. For a random variable Y, we define the conditional entropy $H(X|Y)$ by:*

$$H(X|Y) := \mathop{\mathbb{E}}_{y \sim Y}[H(X|Y = y)].$$

Definition A.4 ((Conditional) mutual information). *We define $I(X;Y) := H(X) + H(Y) - H(X,Y)$. We can similarly define the conditional mutual information $I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z)$.*

Informally, mutual information quantifies the following question: given a sample (x, y) from the joint distribution of X, Y, how many bits of information does y reveal about x (on average)? Formulated in this way, it may not even be obvious that it should be a symmetric quantity, but it is.

Proposition A.5 (Basic properties of entropy). *We have all of the following:*

- $H(X) \geq H_\infty(X)$.
- $H(X) \leq \log |\mathcal{X}|$, *with equality if and only if X is uniform on $\mathcal{X}$.*

- *Chain rule:* $H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$.
- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.
- $I(X;Y) \geq 0$.
- *For a deterministic function f defined on $\mathcal{X}$, we have $H(f(X)) \leq H(X)$. If f is injective, then we have $H(f(X)) = H(X)$.*

A.1 Useful Lemmas

Lemma A.6. *For independent random variables A, B, C, D in an abelian group, we have that*

$$H(A+B+C) - H(A+B) \leq H(B+C) - H(B),$$

and

$$H(A+B+C+D) - H(A+B+C) \leq H(A+B) + H(B+C) + H(C+D) - H(B) - H(C)$$

Proof. For independent random variables A, B, C, we have that

$$\begin{aligned} I(C; B+C) &= H(B+C) - H(B+C|C) = H(B+C) - H(B) \\ I(C; A+B+C) &= H(A+B+C) - H(A+B+C|C) = H(A+B+C) - H(A+B) \end{aligned}$$

Then, the first statement follows by the data-processing inequality. The second statement follows by applying the first inequality twice:

$$\begin{aligned} H(A+B+C+D) - H(A+B+C) \leq H(B+C+D) - H(B+C) \leq H(C+D) - H(C). \\ H(A+B+C+D) \leq H(A+B) + H(B+C) + H(C+D) - H(B) - H(C) \end{aligned}$$

□

B Properties of the Entropic Ruzsa Distance

Proposition B.1. *All of the following hold:*

- *Symmetry:* $d_{\mathsf{ER}}(X,Y) = d_{\mathsf{ER}}(Y,X)$.
- *Triangle inequality:* $d_{\mathsf{ER}}(X,Y) \leq d_{\mathsf{ER}}(X,Z) + d_{\mathsf{ER}}(Z,Y)$.

Proof. Symmetry is once again apparent. Finally, we turn to the triangle inequality. Our goal is to show that:

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

The idea when we proved the triangle inequality in Lemma 4.3 was to consider the function $(a+c)+(b+c) = a+b$. This function is many-to-one, and we captured this there by including c in the input of f. We will use a similar idea here. Define the function $f : (X+Z, Y+Z) \mapsto (X+Y, Y+Z)$. f is bijective, so by Proposition A.5, we have:

$$H(X+Z) + H(Y+Z) \geq H(X+Z, Y+Z) = H(X+Y, Y+Z). \tag{11}$$

Applying Proposition A.5 again implies:

$$H(X+Y, Y+Z) = H(X+Y) + H(Y+Z|X+Y) \geq H(X+Y) + H(Y+Z|X+Y, Y). \tag{12}$$

However, once we have conditioned on $X+Y$ and Y, $Y+Z$ is just distributed as a translate of Z. Thus $H(Y+Z|X+Y,Y) = H(Z)$. This together with chaining (11) and (12) implies the conclusion. □

Proposition B.2 (100% inverse theorem; compare with Propositions 1.6 and 4.2)**.** *We have $d_{\mathsf{ER}}(X, Y) \geq 0$ always, with equality if and only if X, Y are uniform on cosets of the same subgroup.*

Proof. Assume without loss of generality that $0 \in \mathsf{supp}(X) \cap \mathsf{supp}(Y)$. We have $H(X+Y) \geq H(X+Y|Y) = H(X)$ and similarly $H(X+Y) \geq H(Y)$. Therefore we have $H(X+Y) \geq \frac{1}{2}H(X) + \frac{1}{2}H(Y)$.

For equality to hold, we must have $H(X+Y) = H(X+Y|Y)$. In other words, for any $y \in \mathsf{supp}(Y)$, the distribution of $X + y$ is identically distributed to $X + 0 = X$. Letting B be the subgroup generated by $\mathsf{supp}(Y)$, it follows that $X + b$ is identically distributed to X for any $b \in B$. Therefore we have

$$H(X) = H(X + U_B) \geq H(X + U_B|X) = H(U_B) = \log|B| \geq \log|\mathsf{supp}(Y)| \geq H(Y).$$

We are assuming $H(X) = H(Y)$, so this forces $Y \equiv U_{\mathsf{supp}(Y)} \equiv U_B$. We can similarly show that there is a subgroup A of the same size as B such that $X \equiv U_A$. Now $X + b \equiv X$ for all $b \in B$ forces $B \subseteq A \Rightarrow A = B$, so we are done. □

Proof of Lemma 5.2. Note that $U_A + U_B$ is supported on $A + B$. Then we have:

$$\begin{aligned} d_{\mathsf{ER}}(U_A, U_B) &= H(U_A + U_B) - \frac{1}{2}H(U_A) - \frac{1}{2}H(U_B) \\ &= H(U_A + U_B) - \frac{1}{2}\log|A| - \frac{1}{2}\log|B| \text{ (Proposition A.5)} \\ &\leq \log|A + B| - \frac{1}{2}\log|A| - \frac{1}{2}\log|B| \text{ (Proposition A.5)} \\ &= d_{\mathsf{R}}(A, B). \end{aligned}$$

□

B.1 The Fibering Identity

Proof of Lemma 5.3. Expanding the definition of entropic Ruzsa distance and using the entropy chain rule, it suffices to prove

$$\begin{aligned} H(Y_1 + Y_2) - H(Y) = H(\pi(Y_1) + \pi(Y_2)) - H(\pi(Y)) + H(Y_1 + Y_2 \mid (\pi(Y_1), \pi(Y_2))) - H(Y|\pi(Y)) \\ + I\big(Y_1 + Y_2; (\pi(Y_1), \pi(Y_2)) \mid \pi(Y_1) + \pi(Y_2)\big), \end{aligned}$$

or, equivalently,

$$\begin{aligned} H(Y_1 + Y_2) = H(\pi(Y_1) + \pi(Y_2)) + H(Y_1 + Y_2 \mid (\pi(Y_1), \pi(Y_2))) \\ + I\big(Y_1 + Y_2; (\pi(Y_1), \pi(Y_2)) \mid \pi(Y_1) + \pi(Y_2)\big). \end{aligned}$$

By the chain rule,

$$H(Y_1 + Y_2) = H(\pi(Y_1) + \pi(Y_2)) + H(Y_1 + Y_2 \mid \pi(Y_1) + \pi(Y_2)).$$

Since $\pi(Y_1) + \pi(Y_2)$ is determined both by $Y_1 + Y_2$ and by $(\pi(Y_1), \pi(Y_2))$, the definition of conditional mutual information gives

$$\begin{aligned} H(Y_1 + Y_2 \mid \pi(Y_1) + \pi(Y_2)) = H(Y_1 - Y_2 \mid \pi(Y_1), \pi(Y_2)) \\ + I\big[Y_1 - Y_2 : \pi(Y_1), \pi(Y_2) \mid \pi(Y_1) - \pi(Y_2)\big]. \end{aligned}$$

Substituting this into the chain rule equation above gives the desired identity. □

B.2 Conditional Entropic Ruzsa Distance

Definition B.3 (Conditional entropic Ruzsa distance)**.** *Let (X, A) and (Y, B) be random variables, and let X, Y be G-valued for some group G. Then, define*

$$d_{\mathsf{ER}}(X|A, Y|B) := \mathop{\mathbb{E}}_{a,b}\big[d_{\mathsf{ER}}(X_a, Y_b)\big],$$

where the expectation is taken over $a \leftarrow A$ and $b \leftarrow B$. In particular,

$$d_{\mathsf{ER}}(X|A, Y) := \mathop{\mathbb{E}}_{a}\big[d_{\mathsf{ER}}(X_a, Y)\big].$$

Lemma B.4 (Conditioning costs mutual information)**.** *Let (X, A) and (Y, B) be random variables, and let X, Y be G-valued for some group G. Then,*

$$d_{\mathsf{ER}}(X|A, Y|B) \le d_{\mathsf{ER}}(X, Y) + \frac{1}{2}I[X : A] + \frac{1}{2}I[Y : B].$$

In particular,

$$d_{\mathsf{ER}}(X|A, Y) \le d_{\mathsf{ER}}(X, Y) + \frac{1}{2}I[X : A].$$

B.3 Entropic Balog-Szemerédi-Gowers

Lemma B.5 (Entropic Balog-Szemerédi-Gowers)**.** *Let U, V be possibly dependent $\mathbb{F}_2^n$-valued random variables, and let $W := U + V$. For each w with $\Pr[W = w] > 0$, define $U_w := (U \mid W = w)$ and $V_w := (V \mid W = w)$. Then,*

$$\mathop{\mathbb{E}}_{w \leftarrow W}\big[d_{\mathsf{ER}}(U_w, V_w)\big] \le 3I[U : V] + 2H(W) - H(U) - H(V).$$

C Useful Lemmas

Lemma C.1. *Let $C > 0$ and $0 < \eta < 1$ be constants. Let $X, (U, S)$ be random variables such that X, U are G-valued for some group G. Suppose that*

$$\begin{aligned} d_{\mathsf{ER}}(X, U|S) &\le C\, d_{\mathsf{ER}}(X), \text{ and} \\ \mathop{\mathbb{E}}_{s}\big[d_{\mathsf{ER}}(U_s)\big] &\le \eta\, d_{\mathsf{ER}}(X). \end{aligned}$$

Then, for every $0 < \eta' < 1 - \eta$ there exists some s such that

$$\begin{aligned} d_{\mathsf{ER}}(X, U_s) &\le C'\, d_{\mathsf{ER}}(X), \text{ and} \\ d_{\mathsf{ER}}(U_s) &\le (1 - \eta')\, d_{\mathsf{ER}}(X), \end{aligned}$$

where $C' = \frac{1-\eta'}{1-\eta-\eta'}C$.

Proof. By the definition of conditional entropic Ruzsa distance, taking a suitable linear combination of the hypotheses shows that

$$\mathop{\mathbb{E}}_{s}\left[\frac{1 - \eta - \eta'}{C} d_{\mathsf{ER}}(X, U_s) + d_{\mathsf{ER}}(U_s)\right] \le (1 - \eta')\, d_{\mathsf{ER}}(X).$$

Therefore there must be some s for which

$$\frac{1 - \eta - \eta'}{C} d_{\mathsf{ER}}(X, U_s) + d_{\mathsf{ER}}(U_s) \le (1 - \eta')\, d_{\mathsf{ER}}(X)$$

holds. Then, the non-negativity of the entropic Ruzsa distance gives us the desired bounds. □

Lemma C.2. *Let $C > 0$ and $0 < \eta < 1$ be constants. Let $X, (U, S)$ be random variables such that X, U are G-valued for some group G. Suppose that*

$$d_{\mathsf{ER}}(X, U|S) \le C\, d_{\mathsf{ER}}(X)$$
$$d_{\mathsf{ER}}(U|S) \le (1-\eta)\, d_{\mathsf{ER}}(X).$$

Then, there exist constants $C' > 0$ and $0 < \eta' < 1$, and a G-valued random variable X' such that

$$d_{\mathsf{ER}}(X, X') \le C'\, d_{\mathsf{ER}}(X)$$
$$d_{\mathsf{ER}}(X') \le (1-\eta')\, d_{\mathsf{ER}}(X).$$

Partial proof. By the hypotheses, averaging independently over $s, s' \leftarrow S$ gives

$$\mathop{\mathbb{E}}_{s,s'}\Big[d_{\mathsf{ER}}(U_s, U_{s'}) + \frac{\eta}{4C}\big(d_{\mathsf{ER}}(X, U_s) + d_{\mathsf{ER}}(X, U_{s'})\big)\Big] \le \Big(1 - \frac{\eta}{2}\Big)\, d_{\mathsf{ER}}(X).$$

There exist s, s' such that

$$d_{\mathsf{ER}}(U_s, U_{s'}) + \frac{\eta}{4C}\big(d_{\mathsf{ER}}(X, U_s) + d_{\mathsf{ER}}(X, U_{s'})\big) \le \Big(1 - \frac{\eta}{2}\Big)\, d_{\mathsf{ER}}(X).$$

A possible candidate is $X' = U_s + U_{s'}$, with independent copies in the sum. The required self-distance and proximity bounds for this choice remain to be established. □

Claim C.3. *For a $\mathbb{F}_2^n$-valued random variable X, let X_1, X_2 be two independent copies of X. Then there exists a constant $C > 0$ such that $X_1 + X_2$ and $X_1|X_1 + X_2$ are C-nearby with respect to X.*

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

$$d_{\mathsf{ER}}(X) = H(X_1 + X_2) - H(X).$$

Bounding $d_{\mathsf{ER}}(X_1 + X_2, X_3)$,

$$\begin{aligned} d_{\mathsf{ER}}(X_1 + X_2, X_3) &= H(X_1 + X_2 + X_3) - \frac{1}{2}H(X_1 + X_2) - \frac{1}{2}H(X_3) \\ &\le \big(H(X_2 + X_3) + H(X_1 + X_2) - H(X_2)\big) - \frac{1}{2}H(X_1 + X_2) - \frac{1}{2}H(X_3) \quad \text{by Lemma A.6} \\ &= \frac{3}{2} \cdot (H(X_1 + X_2) - H(X)) \\ &= \frac{3}{2} d_{\mathsf{ER}}(X). \end{aligned}$$

Bounding $d_{\mathsf{ER}}(X_1|X_1 + X_2, X_3)$,

$$\begin{aligned} d_{\mathsf{ER}}(X_1|X_1 + X_2, X_3) &= H(X_1 + X_3|X_1 + X_2) - \frac{1}{2}H(X_1|X_1 + X_2) - \frac{1}{2}H(X_3) \\ &\le H(X_1 + X_3) - \frac{1}{2} \cdot \big(H(X_1, X_1 + X_2) - H(X_1 + X_2)\big) - \frac{1}{2}H(X_3) \\ &= \frac{3}{2} \cdot (H(X_1 + X_2) - H(X)) \\ &= \frac{3}{2} \cdot d_{\mathsf{ER}}(X). \end{aligned}$$

□