Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 157 (2011)

An additive combinatorics approach relating rank to communication complexity

Eli Ben-Sasson* Shachar Lovett† Noga Ron-Zewi ‡

May 24, 2012

Abstract

For a $\{0,1\}$-valued matrix M let $\mathrm{CC}(M)$ denote the deterministic communication complexity of the boolean function associated with M. It is well-known since the work of Mehlhorn and Schmidt [STOC 1982] that $\mathrm{CC}(M)$ is bounded from above by $\mathrm{rank}(M)$ and from below by $\log \mathrm{rank}(M)$ where $\mathrm{rank}(M)$ denotes the rank of M over the field of real numbers. Determining where in this range lies the true worst-case value of $\mathrm{CC}(M)$ is a fundamental open problem in communication complexity. The state of the art is

$$\log^{1.631} \mathrm{rank}(M) \leq \mathrm{CC}(M) \leq 0.415\ \mathrm{rank}(M),$$

the lower bound is by Kushilevitz [unpublished, 1995] and the upper bound is due to Kotlov [Journal of Graph Theory, 1996]. Lovász and Saks [FOCS 1988] conjecture that $\mathrm{CC}(M)$ is closer to the lower bound, i.e., $\mathrm{CC}(M) \leq \log^c(\mathrm{rank}(M))$ for some absolute constant c — this is the famous "log-rank conjecture" — but so far there has been no evidence to support it, even giving a slightly non-trivial ($o(\mathrm{rank}(M))$) upper bound on the communication complexity.

Our main result is that, assuming the Polynomial Freiman-Ruzsa (PFR) conjecture in additive combinatorics, there exists a universal constant c such that

$$\mathrm{CC}(M) \leq c \cdot \mathrm{rank}(M)/\log \mathrm{rank}(M).$$

Although our bound is stated using the rank of M over the reals, our proof goes by studying the problem over the finite field of size 2, and there we bring to bear a number of new tools from additive combinatorics which we hope will facilitate further progress on this perplexing question.

In more detail, our proof is based on the study of the "approximate duality conjecture" which was suggested by Ben-Sasson and Zewi [STOC 2011] and studied there in connection to the PFR conjecture. First we improve the bounds on approximate duality assuming the PFR conjecture. Then we use the approximate duality conjecture (with improved bounds) to get our upper bound on the communication complexity of low-rank matrices.

*Department of Computer Science, Technion, Haifa, Israel and Microsoft Research New-England, Cambridge, MA. eli@cs.technion.ac.il. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 240258.

†School of Mathematics, Institute for Advanced Study, Princeton, NJ. slovett@math.ias.edu. Supported by NSF grant DMS-0835373.

‡Department of Computer Science, Technion, Haifa. nogaz@cs.technion.ac.il, Research supported by the Israel Ministry of Science and Technology.

ISSN 1433-8092

1 Introduction

This paper presents a new connection between communication complexity and additive combinatorics, showing that a well-known conjecture from additive combinatorics known as the *Polynomial Freiman-Ruzsa Conjecture* (PFR, in short), implies better upper bounds than currently known on the deterministic communication complexity of a boolean function in terms of the rank of its associated matrix. More precisely, our results show that the PFR Conjecture implies that every boolean function has communication complexity $O(\text{rank}(M)/\log \text{rank}(M))$ where $\text{rank}(M)$ is the rank, over the reals, of the associated matrix. We view this result as interesting not only due to its being the first sublinear bound (and the first advance on this problem since 1997) but also because of its suggestion of a new connection between the two vibrant, yet seemingly unrelated, fields of communication complexity and additive combinatorics.

Our analysis relies on the study of *approximate duality*, a concept closely related to the PFR Conjecture, which was introduced in [BZ11]. Our main technical contribution improves the bounds on approximate duality, assuming the PFR Conjecture, and it does so with simpler proof than in [BZ11]. We view this contribution as being of independent interest because of the growing number of applications of the "approximate duality method" to theoretical computer science. These include so-far the construction of bipartite Ramsey graphs and two-source extractors [BZ11], communication complexity (this work), and the subsequent lower bounds for matching vector locally decodable codes [BDL12].

1.1 On communication complexity and matrix rank

In the two-party communication complexity model two parties — Alice and Bob — wish to compute a function $f : X \times Y \to \{0,1\}$ on inputs x and y where x is known only to Alice and y is known only to Bob. In order to compute the function f they must exchange bits of information between each other according to some (deterministic) protocol. The (deterministic) communication complexity of a protocol is the maximum total number of bits sent between the two parties, where the maximum is taken over all pairs of inputs x, y. We henceforth omit the adjective "deterministic" from our discourse because our results deal only with the deterministic model. The communication complexity of the function f, denoted by $\text{CC}(f)$, is the minimum communication complexity of a protocol for f.

For many applications it is convenient to associate the function $f : X \times Y \to \{0,1\}$ with the matrix $M \in \{0,1\}^{X \times Y}$ whose (x,y) entry equals $f(x,y)$. For a $\{0,1\}$-valued matrix M, let $\text{CC}(M)$ denote the communication complexity of the boolean function associated with M. Let $\text{rank}(M)$ denote the rank of M over the reals. We will occasionally consider the rank of M over the two-element field $\mathbb{F}_2$ and will denote this by $\text{rank}_{\mathbb{F}_2}(M)$.

It is well-known since the work of Mehlhorn and Schmidt [MS82] that

$$\log \text{rank}(M) \leq \text{CC}(M) \leq \text{rank}(M) \tag{1}$$

and it is a fundamental question to find out what is the true worst-case dependency of $\text{CC}(M)$ on the rank. The famous log-rank conjecture due to Lovász and Saks [LS88] postulates that the true answer is closer to the lower bound of (1).

Conjecture 1.1 (Log-rank)**.** *For every $\{0,1\}$-valued matrix M $\text{CC}(M) = \log^{O(1)} \text{rank}(M)$.*

Lovász and Saks also point out that the above conjecture has several other interesting equivalent formulations. One of them, due to Nuffelen [Nuf76] and Fajtlowicz [Faj88], is the following:

Conjecture 1.2. *For every graph G, $\chi(\overline{G}) \leq \log^{O(1)} \operatorname{rank}(G)$, where $\chi(\overline{G})$ is the chromatic number of the complement of G, and $\operatorname{rank}(G)$ is the rank of the adjacency matrix of G over the reals.*

Though considerable effort has been made since 1982 in an attempt to narrow the gap between lower and upper bounds in (1), the state of the art is not far from where it was 30 years ago and currently stands at

$$\Omega(\log^{1.63\ldots} \operatorname{rank}(M)) = \Omega(\log^{\log_3 6} \operatorname{rank}(M)) \leq \mathrm{CC}(M) \leq \log(4/3)\operatorname{rank}(M) = (0.41\ldots)\operatorname{rank}(M) \tag{2}$$

The upper bound is due to Kotlov [Kot97] and improves on the previous best bound of $\mathrm{CC}(M) \leq \operatorname{rank}(M)/2$ by Kotlov and Lovász [KL96]. The lower bound is due to Kushilevitz (unpublished, cf. [NW95]) and improves on a previous bound of $\Omega(\log^{\log_2 3} \operatorname{rank}(M)) = \Omega(\log^{1.58\ldots} \operatorname{rank}(M))$ due to Nisan and Wigderson [NW95].

Our main result is stated next. It assumes a well-known conjecture from additive combinatorics — the Polynomial Freiman-Ruzsa (PFR) conjecture — discussed in the next section.

Theorem 1.3 (Main)**.** *Assuming the PFR Conjecture 1.5, for every $\{0,1\}$-valued matrix M*

$$\mathrm{CC}(M) = O(\operatorname{rank}(M)/\log \operatorname{rank}(M)).$$

1.2 Additive combinatorics and the Polynomial Freiman-Ruzsa conjecture

Quoting the (current) Wikipedia definition, additive combinatorics studies "combinatorial estimates associated with the arithmetic operations of addition and subtraction". As such, it deals with a variety of problems that aim to 'quantify' the amount of additive structure in subsets of additive groups. One such a problem is that which is addressed by the Polynomial Freiman-Ruzsa conjecture (we shall encounter a different problem in additive combinatorics when we get to "approximate duality" later on).

For $A \subseteq \mathbb{F}_2^n$, let $A + A$ denote the sum-set of A

$$A + A := \{a + a' \mid a, a' \in A\}$$

where addition is over $\mathbb{F}_2$. It is easy to see that $|A + A| = |A|$ if and only if A is an affine subspace of $\mathbb{F}_2^n$. The question addressed by the Freiman-Ruzsa Theorem is whether the ratio of $|A + A|$ to $|A|$ also 'approximates' the closeness of A to being a subspace, or in other words, whether the fact that $A + A$ is small with respect to the size of A also implies that $\operatorname{span}(A)$ is small with respect to the size of A. The Freimnan-Ruzsa Theorem [Ruz99] says that this is indeed the case.

Theorem 1.4 (Freiman-Ruzsa Theorem [Ruz99])**.** *If $A \subseteq \mathbb{F}_2^n$ has $|A+A| \leq K|A|$, then $|\operatorname{span}(A)| \leq K^2 2^{K^4}|A|$.*

The above theorem was improved in a series of works [GR06, San08, GT09], culminating in the recent work [EZ11] which proved an upper bound on the ratio $\frac{|\operatorname{span}(A)|}{|A|}$ of the form $2^{2k}/(2k)$. This bound can be seen to be tight (up to a multiplicative factor of 2) by letting $A = \{u_1, u_2, \ldots, u_t\}$, where $u_1, u_2, \ldots, u_t \in \mathbb{F}_2^n$ are linearly independent vectors. Then in this case we have $|A+A| \approx \frac{t}{2}|A|$, while $|\operatorname{span}(A)| = 2^t$.

This example also shows that the ratio $\frac{|\operatorname{span}(A)|}{|A|}$ must depend exponentially on K. However, it does not rule out the existence of a large subset $A' \subseteq A$ for which the ratio $\frac{|\operatorname{span}(A')|}{|A'|}$ is just polynomial in K, and this is exactly what is suggested by the PFR Conjecture:

Conjecture 1.5 (Polynomial Freiman-Ruzsa (PFR))**.** *There exists an absolute constant r, such that if $A \subset \mathbb{F}_2^n$ has $|A+A| \leq K|A|$, then there exists a subset $A' \subseteq A$ of size at least $K^{-r}|A|$ such that $|\mathrm{span}\,(A')| \leq |A|$.*

Note that the above conjecture implies that $|\mathrm{span}\,(A')| \leq |A| \leq K^r|A'|$. The PFR conjecture has many other interesting equivalent formulations, see the survey of Green [Gre05] for some of them. It is conjectured to hold for subsets of general groups as well and not only for subsets of the group $\mathbb{F}_2^n$ but we will be interested only in the latter case. Significant progress on this conjecture has been achieved recently by Sanders [San10], using new techniques developed by Croot and Sisask [CS10]. Sanders proved an upper bound on the ratio $\frac{|\mathrm{span}(A')|}{|A'|}$ which is quasi-polynomial in K:

Theorem 1.6 (Quasi-polynomial Freiman-Ruzsa Theorem (QFR) [San10])**.** *Let $A \subset \mathbb{F}_2^n$ be a set such that $|A+A| \leq K|A|$. Then there exists a subset $A' \subseteq A$ of size at least $K^{-O(\log^3 K)}|A|$ such that $|\mathrm{span}\,(A')| \leq |A|$.*

We end this section by mentioning several other recent applications of the PFR Conjecture to theoretical computer science. The first application, due to Samorodnitsky [Sam07], is to the area of low-degree testing, with further results by Lovett [Lov10] and Green and Tao [GT10]. The second application is to the construction of two-source extractors due to Ben-Sasson and Zewi [BZ11]. The latter paper also introduced the notion of approximate duality which plays a central role in our proof method as well. The approximate duality method has recently found another application to proving lower bounds on locally decodable matching vector codes in the subsequent work by Bhowmick, Dvir and Lovett [BDL12]. In the next section we describe the approximate duality conjecture and our new contributions to its study.

1.3 Approximate duality

Our main technical contribution (Lemma 1.11) is improving the bounds on approximate duality, assuming the PFR conjecture. The new bound lies at the heart of our proof of the Main Theorem 1.3. We believe that Lemma 1.11 and its proof are of independent interest since they improve and simplify the proof of [BZ11], and have already found new interesting applications to the study of locally decodable codes [BDL12].

For $A, B \subseteq \mathbb{F}_2^n$, we define the *duality measure* of A, B in (3) as an estimate of how 'close' this pair is to being dual

$$D(A,B) := \left| \mathbb{E}_{a \in A, b \in B} \left[(-1)^{\langle a,b \rangle_2} \right] \right|, \tag{3}$$

where $\langle a,b \rangle_2$ denotes the binary inner-product of a, b over $\mathbb{F}_2$, defined by $\langle a,b \rangle_2 = \sum_{i=1}^n a_i \cdot b_i$ where all arithmetic operations are in $\mathbb{F}_2$.

Remark 1.7. The duality measure can be alternatively defined as the discrepancy of the inner product function on the rectangle $A \times B$ (up to a normalization factor of $\frac{2^n}{|A||B|}$). Nevertheless we chose to use the term 'duality measure' instead of 'discrepancy' because of the algebraic context in which we use it, as explained below.

It can be verified that if $D(A,B) = 1$ then A is contained in an affine shift of $B^\perp$ which is the space dual to the linear $\mathbb{F}_2$-span of B. The question is what can be said about the structure of A, B when $D(A,B)$ is sufficiently large, but strictly smaller than 1. The following theorem from [BZ11] says that if the duality measure is a constant very close to 1 (though strictly smaller than 1) then there exist relatively large subsets $A' \subseteq A$, $B' \subseteq B$, such that $D(A', B') = 1$.

Theorem 1.8 (Approximate duality for nearly-dual sets, [BZ11]). *For every $\delta > 0$ there exists a constant $\epsilon > 0$ that depends only on δ, such that if $A, B \subseteq \mathbb{F}_2^n$ satisfy $D(A,B) \geq 1-\epsilon$, then there exist subsets $A' \subseteq A, |A'| \geq \frac{1}{4}|A|$ and $B' \subseteq B, |B'| \geq 2^{-\delta n}|B|$, such that $D(A',B') = 1$.*

It is conjectured that a similar result holds also when the duality measure is relatively small, and in particular when it tends to zero as n goes to infinity. Furthermore, the following theorem from [BZ11] gives support to this conjecture, by showing that such bounds indeed follow from the PFR conjecture.

Theorem 1.9 (Approximate duality assuming PFR, exponential loss [BZ11]). *Assuming the PFR Conjecture 1.5, for every constant $\delta > 0$ there exists a constant $\zeta > 0$, depending only on δ, such that if $A, B \subseteq \mathbb{F}_2^n$ satisfy $D(A,B) \geq 2^{-\zeta n}$, then there exist subsets $A' \subseteq A$, $|A'| \geq 2^{-\delta n}|A|$ and $B' \subseteq B$, $|B'| \geq 2^{-\delta n}|B|$ such that $D(A',B') = 1$.*

Remark 1.10. The above theorem is stated a bit differently in [BZ11], namely the constant ζ there depends on another constant $\alpha > 0$ and there is an additional requirement that the sets A, B would both be of size at least $2^{\alpha n}$. Note however that this requirement is redundant since without loss of generality we may assume that A, B are both of size at least $2^{\delta n}$ since the theorem holds trivially otherwise.

Our main technical contribution is the following generalization of the above theorem.

Lemma 1.11 (Main technical lemma). *Assuming the PFR Conjecture 1.5 there exists a universal integer r such that the following holds. Suppose that $A, B \subseteq \{0,1\}^n$ satisfy $D(A,B) \geq \epsilon$. Then for every $K \geq 1$ and $t = n/\log K$, there exist subsets A', B' of A, B respectively such that $D(A',B') = 1$, and*

$$|A'| \geq \left(\left(\frac{(\epsilon/2)^{2^t}}{nK}\right)(4n)^{-t}\right)^r |A|, \qquad |B'| \geq \left(\left(\frac{(\epsilon/2)^{2^t}}{nK}\right)2^{-t}\right)^r |B|. \tag{4}$$

The proof of the above lemma appears in Section 2. To see that it is indeed a generalization of Theorem 1.9 set $K = 2^{\delta n/(3r)}$, $t = 3r/\delta$, $\zeta = \delta/(3r \cdot 2^t) = \delta/(3r \cdot 2^{3r/\delta})$, $\epsilon = 2^{-\zeta n}$, and note that in this case the above lemma assures the existence of $|A'| \geq 2^{-\delta n}|A|$, $|B'| \geq 2^{-\delta n}|B|$ such that $D(A',B') = 1$.

However, the main significance of Lemma 1.11 is that it allows one to tradeoff the loss in the sizes of A' and B' with the value of ϵ for a wider range of parameters. More specifically it allows one to achieve a loss in the sizes of A' and B' which is only sub-exponential in n by requiring ϵ be a bit larger. In particular, the following corollary of Lemma 1.11 will enable us to prove the new upper bound of $O(\mathrm{rank}(M)/\log \mathrm{rank}(M))$ on the communication complexity of $\{0,1\}$-valued matrices assuming the PFR conjecture.

Corollary 1.12 (Approximate duality assuming PFR, sub-exponential loss). *Suppose that $A, B \subseteq \mathbb{F}_2^n$ satisfy $D(A,B) \geq 2^{-\sqrt{n}}$. Then assuming the PFR conjecture 1.5, there exist subsets A', B' of A, B respectively such that $D(A',B') = 1$, and $|A'| \geq 2^{-cn/\log n}|A|$, $|B'| \geq 2^{-cn/\log n}|B|$ for some absolute constant c.*

Proof of Corollary 1.12. Follows from Lemma 1.11 by setting $K = 2^{4n/\log n}$, $t = \frac{\log n}{4}$, $\epsilon = 2^{-\sqrt{n}}$. □

Note that in Corollary 1.12 the ratios $|A'|/|A|$, $|B'|/|B|$ are bounded from below by $2^{-cn/\log n}$, whereas in Theorem 1.9 we only get a smaller bound of the form $2^{-\delta n}$ for some constant $\delta > 0$. However, this improvement comes with a requirement that the duality measure $D(A,B)$ is larger — in the above corollary we require that it is at least $2^{-\sqrt{n}}$ while in Theorem 1.9 we only require

it to be at least $2^{-\zeta n} \ll 2^{-\sqrt{n}}$. We note that the bound $D(A,B) \geq 2^{-\sqrt{n}}$ can be replaced by $D(A,B) \geq \exp(-n^{1-\epsilon})$ for any $\epsilon > 0$ at the price of a larger constant $c = c(\epsilon)$.

Remark 1.13 (Exponential loss necessary). Generally speaking, the bound on $\min\left\{\frac{|A'|}{|A|}, \frac{|B'|}{|B|}\right\}$ — which in Corollary 1.12 above is $2^{-cn/\log n}$ — cannot be improved beyond $2^{-O(\sqrt{n})}$ even if we assume $D(A,B) > 0.99$. To see this take $A = B = \binom{n}{c'\sqrt{n}}$ to be the set of all $\{0,1\}$-vectors with exactly $c'\sqrt{n}$ ones, where c' is a sufficiently small positive constant that guarantees $D(A,B) \geq 0.99$. It can be verified that if $A' \subset A, B' \subset B$ satisfy $D(A',B') = 1$ then the smaller set of A', B' is of size $2^{-\Omega(\sqrt{n})} \cdot |A|$.

We stress that a benefit of the proof of Lemma 1.11 is that it simplifies the original proof of Theorem 1.9 in [BZ11]. Indeed, we believe that the presentation of the proof that appears in this paper is clearer and less involved than that in [BZ11]. Our proof method also allows us to deduce new equivalence between approximate duality and the PFR conjecture in the exponential range that was not previously known. We elaborate on this equivalence in Section 4.

1.4 Proof overview

First we show how our Main Theorem 1.3 is deduced from the improved bounds on approximate duality in Corollary 1.12. Then we give an overview of the proof of Lemma 1.11 itself.

From approximate duality to communication complexity upper bounds. We follow the approach of Nisan and Wigderson from [NW95]. Let the *size* of a matrix M be the number of entries in it and if M is $\{0,1\}$-valued let $\delta(M)$ denote its *(normalized) discrepancy*, defined as the absolute value of the difference between the fraction of zero-entries and one-entries in M. Informally, discrepancy measures how "unbalanced" is M, with $\delta(M) = 1$ when M is *monochromatic* — all entries have the same value — and $\delta(M) = 0$ when M is completely balanced.

Returning to the work of [NW95], they observed that to prove the log-rank conjecture it suffices to show that a $\{0,1\}$-valued matrix M of rank r always contains a monochromatic sub-matrix of size $|M|/\mathrm{qpoly}(r)$ where $\mathrm{qpoly}(r) = r^{\log^{O(1)} r}$ means quasi-polynomial in r. Additionally, they used spectral techniques (i.e., arguing about the eigenvectors and eigenvalues of M) to show that any $\{0,1\}$-valued matrix M of rank r contains a relatively large submatrix M' — of size at least $|M|/r^{3/2}$ — that is somewhat biased — its discrepancy is at least $1/r^{3/2}$. We show, using tools from additive combinatorics, that M' in fact contains a pretty large monochromatic submatrix (though not large enough to deduce the log-rank conjecture).

To this end we start by working over the two-element field $\mathbb{F}_2$. This seems a bit counter-intuitive because the log-rank conjecture is false over $\mathbb{F}_2$. The canonical counterexample is the inner product function $IP(x,y) = \langle x,y\rangle_2$ — It is well-known (see e.g. [KN97][Chapters 1.3., 2.5.]) that $\mathrm{rank}_{\mathbb{F}_2}(M_{IP}) = n$ while $\mathrm{CC}(IP) = n$. However, rather than studying M over $\mathbb{F}_2$ we focus on the biased submatrix M' and things change dramatically. (As a sanity-check notice that M_{IP} does not contain large biased submatrices and this does not contradict the work of [NW95] because the rank of M_{IP} over the reals is $2^n - 1$.)

Thus, our starting point is a large submatrix M' that has large discrepancy. It is well-known that $\mathrm{rank}_{\mathbb{F}_2}(M') \leq \mathrm{rank}(M') \leq r$ and that this implies M' can be written as $M' = A^\top \cdot B$ where A, B are matrices whose columns are vectors in $\mathbb{F}_2^r$. Viewing each of A, B as the set of its columns, we have in hand two sets that have a large duality measure as defined in (3), namely, $D(A,B) = \delta(M') \geq 1/r^{3/2}$. This is the setting in which we apply Corollary 1.12 and deduce that A, B contain relatively large subsets A', B' with $D(A',B') = 1$. One can now verify that the submatrix of M'

whose rows and columns are indexed by A', B' respectively is indeed monochromatic, as needed. We point out that to get our bounds we need to be able to find monochromatic submatrices of M' even when M' is both small and skewed (i.e., has many more columns than rows or vice versa). Fortunately, Corollary 1.12 is robust enough to use in such settings.

Improved bounds on approximate duality assuming PFR. We briefly sketch the proof of our Main Technical Lemma 1.11. We use the *spectrum* of a set as defined in [TV06, Chapter 4]:

Definition 1.14 (Spectrum)**.** For a set $B \subseteq \mathbb{F}_2^n$ and $\alpha \in [0,1]$ let the *α-spectrum* of B be the set

$$\mathrm{Spec}_\alpha(B) := \{x \in \mathbb{F}_2^n \mid \; |\, \mathbb{E}_{b \in B}\left[(-1)^{\langle x,b\rangle_2}\right] \,| \geq \alpha\}. \tag{5}$$

Notice that $A \subseteq \mathrm{Spec}_\epsilon(B)$ implies $D(A,B) \geq \epsilon$ (cf. (3)). In the other direction, Markov's inequality can be used to deduce that $D(A,B) \geq \epsilon$ implies the existence of $A' \subseteq A$ of relatively large size — $|A'| \geq \frac{\epsilon}{2}|A|$ — such that $A' \subseteq \mathrm{Spec}_{\epsilon/2}(B)$. To prove our lemma we start with $A_1 = A'$ and establish a sequence of sets

$$A_2 \subseteq A_1 + A_1, \quad A_3 \subseteq A_2 + A_2, \ldots$$

such that $A_i \subseteq \mathrm{Spec}_{\epsilon_i}(B)$ for all i. This holds by construction for A_1 with $\epsilon_1 = \epsilon/2$, and we show that it is maintained throughout the sequence for increasingly smaller values of ϵ_i (we shall use $\epsilon_i = \epsilon_{i-1}^2$).

Moving our problem from the field of real numbers to the two-element field $\mathbb{F}_2$ now pays off. Each A_i is of size at most 2^n so there must be an index $i \leq n/\log K$ for which $|A_{i+1}| \leq K|A_i|$, let t be the minimal such index. We use the PFR conjecture together with the Balog–Szemerédi–Gowers Theorem 2.1 from additive combinatorics to show that our assumption that $|A_{t+1}| \leq K|A_t|$ implies that a large subset A_t'' of A_t has small span (over $\mathbb{F}_2$).

We now have in hand a set A_t'' which is a relatively large fraction of its span and additionally satisfies $D(A_t'', B) \geq \epsilon_t$ because by construction $A_t'' \subseteq \mathrm{Spec}_{\epsilon_t}(B)$. We use an approximate duality claim from [BZ11] (Lemma 2.2) which applies when one of the sets is a large fraction of its span (in our case the set which is a large fraction of its span is A_t''). This claim says that A_t'' and B each contain relatively large subsets A_t', B_t' satisfying $D(A_t', B_t') = 1$. Finally, recalling A_t' is a (carefully chosen) subset of $A_{t-1} + A_{t-1}$, we argue that A_{t-1} contains a relatively large subset A_{t-1}' that is "dual" to a large subset B_{t-1}' of B_t', where by "dual" we mean $D(A_{t-1}', B_{t-1}') = 1$ (in other words A_{t-1}' is contained in an affine shift of the space dual to span $\left(B_{t-1}'\right)$). We continue in this manner to find pairs of "dual" subsets for $t-2, t-3, \ldots, 1$ at which point we have found a pair of "dual" subsets of A, B that have relatively large size, thereby completing the proof.

1.5 Discussion and directions for future research

The new connection between additive combinatorics and communication complexity seems to us worthy of further study. In particular, the exciting recent advances in additive combinatorics [San10, CS10, EZ11] use a rich palette of tools that may yield further insights into problems in communication complexity. We end this section by briefly pointing out a few directions we find interesting.

Improved unconditional bounds on communication complexity Given the recent QFR result of [San10] (Theorem 1.6) which comes very close to proving the PFR conjecture, it is interesting to see if it implies any unconditional improvement on communication complexity of low-rank

matrices. Looking at our proof of Lemma 1.11, we apply the PFR conjecture to a subset A'_t of A_t which satisfies $|A'_t + A'_t| \leq K'|A'_t|$ for $K' \approx K/\epsilon^{2^t}$. For $\epsilon < \frac{1}{2}$ this gives a non-trivial bound only if $t = O(\log n)$. Since t could be as large as $n/\log K$ we are forced to choose $K = 2^{\Omega(n/\log n)}$ which implies in turn $K' = 2^{\Omega(n/\log n)}$. Thus, Sander's QFR Theorem 1.6 does not yield any non-trivial bounds in our case. However, for purposes of improving the unconditional upper bound of Kotlov (cf. 2) say, to $CC(M) \leq \mathrm{rank}(M)/4$, it suffices to improve the loss in the size of A in Theorem 1.6 from $K^{-O(\log^3 K)}$ to $K^{-c \log K}$ for a sufficiently small constant c.

Improved conditional bounds The bounds on approximate duality in Corllary 1.12 can possibly be significantly improved. For all we know, the exponential loss of $2^{-O(\sqrt{n})}$ shown in Remark 1.13 may be tight, and this would lead to an improved version of Corollary 1.12 in which the sizes of $|A'|, |B'|$ are a $2^{-O(\sqrt{n})}$ fraction of A and B respectively, instead of the $2^{-O(n/\log n)}$ loss we currently have. Such a result would translate directly to an upper bound on communication complexity of the form $\mathrm{CC}(M) \leq O(\sqrt{\mathrm{rank}(M)})$. In order to make further progress one might want to also consider working over finite fields that are larger than 2, or over the reals. As a first step in this direction, one may wish to investigate whether there are interesting approximate duality statements over such fields.

Does the log-rank conjecture imply the PFR conjecture? Alternatively, does it have any other non-trivial consequences in additive combinatorics? We believe the answer to this question is positive and make a step in this direction by showing an equivalence between approximate duality and PFR statements in the exponential range, namely, when the losses in the sizes of sets in both approximate duality and PFR is exponential in n (See Section 4 for an exact statement and details of the proof.)

1.6 Paper organization.

The next section contains the proof of the Main Technical Lemma 1.11. The proof of Main Theorem 1.3 given Corollary 1.12 appears in Section 3. Finally, in Section 4 we prove a new equivalence between approximate duality and the PFR conjecture in the exponential range.

2 Improved bounds on approximate duality assuming PFR

In this section we prove our Main Technical Lemma 1.11. We start of some additive combinatorics preliminaries.

Additive combinatorics preliminaries In what follows all arithmetic operations are taken over $\mathbb{F}_2$. For the proof of Lemma 1.11 we need two other theorems from additive combinatorics. The first is the well-known Balog–Szemerédi–Gowers Theorem of [BS94, Gow98].

Theorem 2.1 (Balog–Szemerédi–Gowers)**.** *There exist fixed polynomials $f(x,y)$, $g(x,y)$ such that the following holds for every subset A of an abelian additive group. If A satisfies $\Pr_{a,a' \in A}[a + a' \in S] \geq 1/K$ for $|S| \leq C|A|$, then one can find a subset $A' \subseteq A$ such that $|A'| \geq |A|/f(K,C)$, and $|A' + A'| \leq g(K,C)|A|$.*

The second is a lemma from [BZ11] which can be seen as an approximate duality statement which applies when one of the sets has small span:

Lemma 2.2 (Approximate-duality for sets with small span, [BZ11]). *If $D(A,B) \geq \epsilon$, then there exist subsets $A' \subseteq A, B' \subseteq B$, $|A'| \geq \frac{\epsilon}{4}|A|$, $|B'| \geq \frac{\epsilon^2}{4}\frac{|A|}{|\mathrm{span}(A)|}|B|$, such that $D(A',B') = 1$. If $A \subseteq \mathrm{Spec}_\epsilon(B)$ then we have $|A'| \geq |A|/2$ and $|B'| \geq \epsilon^2\frac{|A|}{|\mathrm{span}(A)|}|B|$ in the statement above.*

Recall the definition of the spectrum given in (5):

$$\mathrm{Spec}_\alpha(B) := \{x \in \mathbb{F}_2^n \mid \; |\, \mathbb{E}_{b\in B}\left[(-1)^{\langle x,b\rangle_2}\right] | \geq \alpha\}.$$

Finally, for $S \subset \mathbb{F}_2^n$ and $x \in \mathbb{F}_2^n$ let $\mathrm{rep}_S(x)$ be the number of different representations of x as an element of the form $s + s'$ where $s, s' \in S$. $\mathrm{rep}_S(x)$ can also be written, up to a normalization factor, as $1_S * 1_S(x)$ where 1_S is the indicating function of the set S and $*$ denotes convolution.

Proof overview We construct a decreasing sequence of constants

$$\epsilon_1 = \epsilon/2, \epsilon_2 = \epsilon_1^2/2, \epsilon_3 = \epsilon_2^2/2, \ldots$$

and a sequence of sets

$$A_1 := A \cap \mathrm{Spec}_{\epsilon_1}(B), \quad A_2 \subseteq (A_1 + A_1) \cap \mathrm{Spec}_{\epsilon_2}(B), \quad A_3 \subseteq (A_2 + A_2) \cap \mathrm{Spec}_{\epsilon_3}(B), \ldots$$

Since each of the sets in the sequence is of size at most 2^n there must be an index $i \leq n/\log K$ for which

$$|A_{i+1}| \leq K|A_i| \tag{6}$$

and let t be the minimal such index. The PFR Conjecture 1.5 together with the Balog–Szemerédi–Gowers Theorem 2.1 will be used to deduce from (6) that a large subset A_t'' of A_t has small span. Applying Lemma 2.2 to the sets A_t'' and B implies the existence of large subsets $A_t' \subseteq A_t$ and $B_t' \subseteq B$ such that $D(A_t', B_t') = 1$. Finally we argue inductively for $i = t-1, t-2, \ldots, 1$ that there exist large subsets $A_i' \subseteq A_i$ and $B_i' \subseteq B$ such that $D(A_i', B_i') = 1$. The desired conclusion will follow from the $i = 1$ case. To be able to "pull back" and construct a pair of large sets A_{i-1}', B_{i-1}' from the pair A_i', B_i' we make sure every element in A_i is the sum of roughly the same number of pairs in $A_{i-1} \times A_{i-1}$.

The sequence of sets Let $\epsilon_1 := \epsilon/2$, $A_1 := A \cap \mathrm{Spec}_{\epsilon_1}(B)$. Assuming A_{i-1}, ϵ_{i-1} have been defined set $\epsilon_i = \epsilon_{i-1}^2/2$ and let $j_i \in \{0, \ldots, n-1\}$ be an integer index which maximizes the size of

$$\left\{(a,a') \in A_{i-1} \mid a + a' \in \mathrm{Spec}_{\epsilon_i}(B) \text{ and } 2^{j_i} \leq \mathrm{rep}_{A_{i-1}}(a+a') \leq 2^{j_i+1}\right\}. \tag{7}$$

and set

$$A_i := \{a + a' : a, a' \in A_{i-1}, a + a' \in \mathrm{Spec}_{\epsilon_i}(B) \text{ and } 2^{j_i} \leq \mathrm{rep}_{A_{i-1}}(a+a') \leq 2^{j_i+1}\}. \tag{8}$$

Claim 2.3. *For $i = 1$ we have $|A_1| \geq (\epsilon/2)|A|$. For $i > 1$ we have*

$$\Pr_{a,a' \in A_{i-1}}[a + a' \in A_i] \geq \epsilon_i/n \tag{9}$$

and additionally

$$|A_i| \geq \frac{\epsilon_i}{2^{j_i+1}n}|A_{i-1}|^2. \tag{10}$$

Proof. The case of $i = 1$ follows directly from Markov's inequality. For larger i we argue that

$$\Pr_{a,a' \in A_{i-1}}[a + a' \in \mathrm{Spec}_{\epsilon_i}(B)] \geq \epsilon_i.$$

To see this use Cauchy-Schwarz to get

$$\mathbb{E}_{a,a' \in A_{i-1}} |\mathbb{E}_{b \in B} (-1)^{\langle a+a', b \rangle}| = E_{b \in B} (\mathbb{E}_{a \in A_{i-1}} [(-1)^{\langle a,b \rangle}])^2 \geq (\mathbb{E}_{a \in A_{i-1}, b \in B} [(-1)^{\langle a,b \rangle}])^2 = \epsilon_{i-1}^2$$

and apply Markov's inequality to deduce that an ϵ_i-fraction of $(a, a') \in A_{i-1} \times A_{i-1}$ sum to an element of $\mathrm{Spec}_{\epsilon_i}(B)$. Selecting j_i to maximize (7) yields inequality (9). Since every element $x \in A_i$ can be represented as $x = a + a'$ with $a, a' \in A_{i-1}$ in at most 2^{j_i+1} different ways we deduce (10) from (9) and complete the proof. □

The inductive claim Let t be the minimal index such that $|A_{t+1}| \leq K|A_t|$ and note that $t \leq n/\log K$ because all sets A_i are contained in $\mathbb{F}_2^n$. We shall prove the following claim by backward induction.

Claim 2.4 (Inductive claim)**.** *For $i = t, t-1, \ldots, 1$ there exist subsets*

$$A_i' \subseteq A_i, \quad B_i' \subseteq B$$

such that $D(A_i', B_i') = 1$ and A_i', B_i' are not too small:

$$|A_i'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-i)}\left(\prod_{\ell=i}^{t} \epsilon_{\ell+1}\right)|A_i|, \qquad |B_i'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)2^{-(t-i)}|B|$$

We split the proof of the claim to two parts. The base case (Proposition 2.5) is proved using the tools from additive combinatorics listed in the beginning of this section. The inductive step is proved in Proposition 2.6 using a graph construction. Before proving Claim 2.4 we show how it implies Lemma 1.11.

Proof of Main Technical Lemma 1.11. Set $i = 1$ in Claim 2.4 above. Recall that $\epsilon_{i+1} = \epsilon_i^2/2$ for all i, so

$$\epsilon_{\ell+1} = \epsilon^{2^\ell}/2^{2^\ell - 1} \geq (\epsilon/2)^{2^\ell}.$$

Thus we have $\epsilon_{t+1} \geq (\epsilon/2)^{2^t}$ and $\prod_{\ell=1}^{t} \epsilon_{\ell+1} \geq (\epsilon/2)^{2^{t+1}}$. This gives the bounds on A', B' stated in (4). □

Proposition 2.5 (Base case of Claim 2.4 ($i = t$))**.** *There exist subsets $A_t' \subseteq A_t$, $B_t' \subseteq B_t$ such that $D(A_t', B_t') = 1$ and A_t', B_t' are not too small:*

$$|A_t'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|A_t|, \qquad |B_t'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|B|.$$

Proof. By assumption $|A_{t+1}| \leq K|A_t|$ and $Pr_{a,a' \in A_t}[a + a' \in A_{t+1}] \geq \epsilon_{t+1}/n$ by (10). Hence we can apply the Balog–Szemerédi–Gowers Theorem (Theorem 2.1) to the set A_t to obtain a subset $\tilde{A}_t \subseteq A_t$ such that

$$|\tilde{A}_t| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|A_t|,$$

and

$$|\tilde{A}_t + \tilde{A}_t| \leq \mathrm{poly}\left(\frac{nK}{\epsilon_{t+1}}\right)|A_t| = \mathrm{poly}\left(\frac{nK}{\epsilon_{t+1}}\right)|\tilde{A}_t|.$$

Now we can apply the PFR Conjecture 1.5 to the set $\tilde{A}_t$ which gives a subset $A_t'' \subseteq \tilde{A}_t$ such that

$$|A_t''| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|\tilde{A}_t| = \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|A_t|,$$

and

$$|\mathrm{span}\,(A_t'')\,| \leq |\tilde{A}_t| = \mathrm{poly}\left(\frac{nK}{\epsilon_{t+1}}\right)|A_t''|.$$

Recall that $A_t'' \subseteq \mathrm{Spec}_{\epsilon_t}(B)$, and in particular $D(A_t'', B) \geq \epsilon_t$. Applying Lemma 2.2 to the sets A_t'' and B we conclude that there exist subsets $A_t' \subseteq A_t''$, $B' \subseteq B$ such that $D(A_t', B') = 1$, and which satisfy $|A_t'| \geq \frac{1}{2}|A_t''|$ and

$$|B_t'| \geq \epsilon_t^2 \frac{|A_t''|}{|\mathrm{span}\,(A_t'')\,|}|B| = \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)|B|.$$

This completes the proof of the base case. □

Proposition 2.6 (Inductive step of Claim 2.4)**.** *For every $i = t-1, \ldots, 1$ there exist subsets $A_i' \subseteq A_i$, $B_i' \subseteq B$ such that $D(A_i', B_i') = 1$ and A_i', B_i' are not too small:*

$$|A_i'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-i)}\left(\prod_{\ell=i}^{t}\epsilon_{\ell+1}\right)|A_i|, \qquad |B_i'| \geq \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)2^{-(t-i)}|B|.$$

Proof. Suppose that the claim is true for i and argue it holds for index $i-1$. Let $G = (A_{i-1}, E)$ be the graph whose vertices are the elements in A_{i-1}, and (a, a') is an edge if $a + a' \in A_i'$. We bound the number of edges in this graph from below. Recall from (8) that every $a \in A_i'$ (where $A_i' \subseteq A_i$) satisfies $2^{j_i} \leq \mathrm{rep}_{A_{i-1}}(a) \leq 2^{j_i+1}$. Using this we get

$$\begin{aligned}
|E| &\geq 2^{j_i} \cdot |A_i'| && (\mathrm{rep}_{A_{i-1}}(x) \geq 2^{j_i} \text{ for all } x \in A_i')\\
&\geq 2^{j_i} \cdot \mathrm{poly}\left(\tfrac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-i)}\left(\textstyle\prod_{\ell=i}^{t}\epsilon_{\ell+1}\right)|A_i| && (\text{induction hypothesis})\\
&\geq 2^{j_i} \cdot \mathrm{poly}\left(\tfrac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-i)}\left(\textstyle\prod_{\ell=i}^{t}\epsilon_{\ell+1}\right)\frac{\epsilon_i}{2^{j_i+1}n}|A_{i-1}|^2 && (\text{by (10)})\\
&= 2 \cdot \mathrm{poly}\left(\tfrac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-(i-1))}\left(\textstyle\prod_{\ell=i-1}^{t}\epsilon_{\ell+1}\right)|A_{i-1}|^2.
\end{aligned}$$

Let $M := \mathrm{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-(i-1))}\left(\prod_{\ell=i-1}^{t}\epsilon_{\ell+1}\right)$. Since our graph has at least $2M|A_{i-1}|^2$ edges and $|A_{i-1}|$ vertices, it has a connected component with at least $2M|A_{i-1}|$ vertices and denote by A_{i-1}'' the set of vertices in it.

Choose an arbitrary element a in A_{i-1}''. Partition B_i' into two sets $B_{i,0}'$ and $B_{i,1}'$ such that all elements in $B_{i,0}'$ have inner product 0 with a, and all elements in $B_{i,1}'$ have inner product 1 with a. Let B_{i-1}' be the larger of $B_{i,0}', B_{i,1}'$, and note that $|B_{i-1}'| \geq |B_i'|/2$. Recall that our assumption was that $D(A_i', B_i') = 1$. Abusing notation, let $\langle A_i', B_i'\rangle_2$ denote the value of $\langle a', b'\rangle_2$ for some $a' \in A_i', B_i'$ (the choice of a', b' does not matter because $D(A_i', B_i') = 1$). Next we consider two cases — the case where $\langle A_i', B_i'\rangle_2 = 0$, and the case where $\langle A_i', B_i'\rangle_2 = 1$.

In the first case we have that for every $a, a' \in A_{i-1}''$ which are neighbors in the graph, $a + a' \in A_i'$, and therefore $\langle a + a', b\rangle_2 = 0$ for every $b \in B_{i-1}'$. This implies in turn that $\langle a, b\rangle_2 = \langle a', b\rangle_2$ for all elements $a, a' \in A_{i-1}''$ which are neighbors in the graph, $b \in B_{i-1}'$. Since A_{i-1}'' induces a

connected component, and due to our choice of B'_{i-1}, this implies that $D(A''_{i-1}, B'_{i-1}) = 1$ so we set $A'_{i-1} = A''_{i-1}$.

In the second case we have that $\langle a + a', b\rangle_2 = 1$ for every $a, a' \in A''_{i-1}$ which are neighbors in the graph, $b \in B'_{i-1}$. In particular this implies that $\langle a, b\rangle_2 = \langle a', b\rangle_2 + 1$ for every elements $a, a' \in A''_{i-1}$ which are neighbors in the graph, $b \in B'_{i-1}$. This means that A''_{i-1} can be partitioned into two sets $A'_{i-1,0}, A'_{i-1,1}$, where the first one contains all elements in A''_{i-1} that have inner product 0 with all elements in B'_{i-1}, while the second set contains all elements in A'_{i-1} that have inner product 1 with all elements in B'_{i-1}. We set A'_{i-1} to be the larger of these two sets and get $D(A'_{i-1}, B'_{i-1}) = 1$ and $|A'_{i-1}| \geq M|A_{i-1}|$.

Concluding, in both cases we obtained subsets A'_{i-1}, B'_{i-1} of A_{i-1}, B respectively, such that $D(A'_{i-1}, B'_{i-1}) = 1$ and A'_{i-1}, B'_{i-1} are not too small:

$$|A'_{i-1}| \geq \text{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)(4n)^{-(t-(i-1))}\left(\prod_{\ell=i-1}^{t}\epsilon_{\ell+1}\right)|A_{i-1}|,$$

and

$$|B'_{i-1}| \geq \frac{1}{2}|B'_i| \geq \frac{1}{2}\text{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)2^{-(t-i)}|B| = \text{poly}\left(\frac{\epsilon_{t+1}}{nK}\right)2^{-(t-(i-1))}|B|.$$

This concludes the proof of the inductive claim.

□

3 From approximate duality to communication complexity upper bounds

In this section we prove our main theorem, Theorem 1.3 given Corollary 1.12. The proof of the main technical lemma is deferred to Section 2.

We start by repeating the necessary definitions. For a $\{0,1\}$-valued matrix M, let $\text{CC}(M)$ denote the communication complexity of the boolean function associated with M. Let $\text{rank}(M)$ and $\text{rank}_{\mathbb{F}_2}(M)$ denote the rank of M over the reals and over $\mathbb{F}_2$, respectively. We denote by $|M|$ the total number of entries in M, and by $|M_0|$ and $|M_1|$ the number of zero and non-zero entries of M, respectively. We say that M is *monochromatic* if either $|M| = |M_0|$ or $|M| = |M_1|$. Finally, we define the *discrepancy* $\delta(M)$ of M to be the ratio $\frac{||M_0|-|M_1||}{|M|}$.

Recall the statements of Theorem 1.3 and Corollary 1.12.

Main Theorem 1.3 (restated). *Assuming the PFR conjecture (Conjecture 1.5), for every $\{0,1\}$-valued matrix M,*

$$\text{CC}(M) = O(\text{rank}(M)/\log \text{rank}(M)).$$

Corollary 1.12 (restated). *Suppose that $A, B \subseteq \mathbb{F}_2^n$ satisfy $D(A,B) \geq 2^{-\sqrt{n}}$. Then assuming the PFR conjecture, there exist subsets A', B' of A, B respectively such that $D(A', B') = 1$, and $|A'| \geq 2^{-cn/\log n}|A|$, $|B'| \geq 2^{-cn/\log n}|B|$ for some absolute constant c.*

We first prove that the above corollary is equivalent to the following one:

Lemma 3.1 (Main technical lemma, equivalent matrix form)**.** *Let M be a $\{0,1\}$-valued matrix with no identical rows or columns, of rank at most r over $\mathbb{F}_2$, and of discrepancy at least $2^{-\sqrt{r}}$. Then assuming the PFR conjecture (Conjecture 1.5), there exists a monochromatic submatrix M' of M of size at least $2^{-cr/\log r}|M|$ for some absolute constant c.*

Proof. We prove only the Corollary 1.12 $\Rightarrow$ Lemma 3.1 implication. The proof of the converse implication is similar. Denote the number of rows and columns of M by k, ℓ respectively. It is well known that the rank of M over a field $\mathbb{F}$ equals r if and only if M can be written as the sum of r rank one matrices over the field $\mathbb{F}$. Since $\text{rank}_{\mathbb{F}_2}(M) \leq r$ this implies in turn that there exist subsets $A, B \subseteq \mathbb{F}_2^r$, $A = \{a_1, a_2, \ldots, a_k\}$, $B = \{b_1, b_2, \ldots, b_\ell\}$ such that $M_{i,j} = \langle a_i, b_j \rangle_2$ for all $1 \leq i \leq k, 1 \leq j \leq \ell$. Since M has no identical rows or columns we know that $|A| = k$, $|B| = \ell$. Note that $D(A, B) = \delta(M) \geq 2^{-\sqrt{r}}$.

Corollary 1.12 now implies the existence of subsets $A' \subseteq A$, $B' \subseteq B$, $|A'| \geq 2^{-cr/\log r}|A|$, $|B'| \geq 2^{-cr/\log r}|B|$, such that $D(A', B') = 1$. Let M' be the submatrix of M whose rows and columns correspond to the indices in A' and B' respectively. The fact that $D(A', B') = 1$ implies that $M_{i,j} = \langle a_i, b_j \rangle_2 \equiv \text{const}$ for all $a_i \in A'$, $b_j \in B'$. Therefore M' is a monochromatic submatrix of M of which satisfies

$$|M'| = |A'||B'| \geq 2^{-2cr/\log r}|A||B| = 2^{-2cr/\log r}|M|,$$

as required. □

In order to prove Theorem 1.3 we follow the high-level approach of Nisan and Wigderson [NW95] which was explained in the previous section. They showed that in order to prove the log-rank conjecture it suffices to prove that every $\{0,1\}$-valued matrix of low rank has a large monochromatic submatrix. We start with the following lemma.

Lemma 3.2 (Existence of large monochromatic submatrix assuming PFR)**.** *Assuming the PFR conjecture, every $\{0,1\}$-valued matrix M with no identical rows or columns has a monochromatic submatrix of size at least $2^{-O(\text{rank}(M)/\log \text{rank}(M))}|M|$.*

In order to prove the above lemma we use Lemma 3.1, together with the following theorem from [NW95], which says that every $\{0,1\}$-valued matrix M contains a submatrix of high discrepancy:

Theorem 3.3 (Existence of submatrix with high discrepancy [NW95])**.** *Every $\{0,1\}$-valued matrix M has a submatrix M' of size at least $(\text{rank}(M))^{-3/2}|M|$ and with $\delta(M') \geq (\text{rank}(M))^{-3/2}$.*

Proof of Lemma 3.2. Let $r = \text{rank}(M)$. Theorem 3.3 implies the existence of a submatrix M' of M with $|M'| \geq (\text{rank}(M))^{-3/2}|M|$, and $\delta(M') \geq r^{-3/2} \gg 2^{-\sqrt{r}}$. Note also that

$$\text{rank}_{\mathbb{F}_2}(M') \leq \text{rank}(M') \leq \text{rank}(M) = r.$$

Lemma 3.1 then implies the existence of a monochromatic submatrix M'' of M' of size at least $2^{-cr/\log r}|M'|$ for some absolute constant c. So we have that M'' is a monochromatic submatrix of M which satisfies

$$|M''| \geq 2^{-cr/\log r}|M'| \geq 2^{-cr/\log r} r^{-3/2}|M| = 2^{-O(r/\log r)}|M|$$

□

Proof of Theorem 1.3. Let M be a $\{0,1\}$-valued matrix. We will construct a deterministic protocol for M with communication complexity $O(\text{rank}(M)/\log \text{rank}(M))$. We may assume w.l.o.g that M has no repeated rows or columns, otherwise we can eliminate the repeated row or column and the protocol we construct for the "compressed" matrix (with no repeated rows/columns) will also be a protocol for M.

We follow the high level approach of the proof of Theorem 2 from [NW95]. We will show a protocol with $2^{O(r/\log r)}$ leaves. This will suffice since it is well-known that a protocol with t leaves has communication complexity at most $O(\log t)$ (cf. [KN97, Chapter 2, Lemma 2.8]).

Now we describe the protocol. Let Q be the largest monochromatic submatrix of M. Then Q induces a natural partition of M into 4 submatrices Q, R, S, T with R sharing the rows of Q and S sharing the columns of Q.

$$M = \begin{pmatrix} Q & R \\ S & T \end{pmatrix}$$

Let U_1 be a subset of the rows of $(Q|R)$ whose restriction to the columns of R span the rows of R. Similarly, let U_2 be a subset of the rows of $(S|T)$ whose restriction to the columns of S span the rows of S. Note that if Q is the all zeros matrix then the rows of U_1 are independent of the rows of U_2. Otherwise, if Q is the all ones matrix then the rows of U_1 are independent of all the rows of U_2 except possibly for the vector in U_2 whose restriction to the columns of S is the all ones vector (if such vector exists). Thus since Q is monochromatic we have that $\mathrm{rank}(R) + \mathrm{rank}(S) = |U_1| + |U_2| \leq \mathrm{rank}(M) + 1$.

If $\mathrm{rank}(R) \leq \mathrm{rank}(S)$ then the row player sends a bit saying if his input belongs to the rows of Q or not. The players continue recursively with a protocol for the submatrix $(Q|R)$ or the submatrix $(S|T)$ according to the bit sent. If $\mathrm{rank}(R) \geq \mathrm{rank}(S)$ the roles of the row and column players are switched.

Suppose without loss of generality that $\mathrm{rank}(R) \leq \mathrm{rank}(S)$. Then after sending one bit we continue with either the matrix $(Q|R)$ which is of rank at most $\mathrm{rank}(M)/2$ or with the matrix $(S|T)$ which — thanks to Lemma 3.2 — is of size at most $(1-\delta)|M|$ for $\delta \geq 2^{-cr/\log r}$.

Let $L(m, r)$ denote the number of leaves in the protocol starting with a matrix of area at most m and rank at most r. Then we get the following recurrence relation:

$$L(m,r) \leq \begin{cases} L(m, r/2) + L(m(1-\delta), r) & r > 1 \\ 1 & r = 1 \end{cases}$$

It remains to show that in the above recursion $L(m, r) = 2^{O(r/\log r)}$. Applying the recurrence iteratively $1/\delta$ times to the right-most summand we get

$$L(m,r) \leq \delta^{-1} L(m, r/2) + L(m(1-\delta)^{1/\delta}, r) \leq 2^{cr/\log(r)} L(m, r/2) + L(m/2, r).$$

Set $A(m, r) := 2^{-2cr/\log r} L(m, r)$. Then we have $A(m, r) \leq A(m, r/2) + A(m/2, r)$ which together with $A(1, r), A(m, 1) \leq 1$ imply $A(m, r) \leq \binom{\log m + \log r}{\log r}$ since we may apply the recursion iteratively at most $\log r$ times to the left term and $\log m$ times to the right term before we reach $A(1, r)$ or $A(m, 1)$. This in turn implies $A(m, r) \leq \binom{\log m + \log r}{\log r} \leq r^{O(\log r)}$ due to the fact that $r \leq m \leq 2^{2r}$, since we may assume there are no identical rows or columns in the matrix M.

Concluding, we have $L(m, r) \leq 2^{2cr/\log r + O(\log^2 r)}$, which implies in turn $\mathrm{CC}(M) = O(r/\log r)$ as claimed. □

4 Equivalence between approximate duality and the PFR conjecture in the exponential range

In this section we show a new equivalence between approximate duality and the PFR conjecture in the exponential range which follows from the proof of our main technical Lemma 1.11. Before we elaborate on this we discuss the previously known relations between approximate duality and the PFR conjecture. Recall first that Theorem 1.9 (which was proven in [BZ11]) shows that the following version of approximate duality is implied by the PFR conjecture.

Conjecture 4.1 (Approximate duality conjecture, exponential loss)**.** *For every constant $\delta > 0$ there exists a constant $\zeta > 0$, depending only on δ, such that if $A, B \subseteq \mathbb{F}_2^n$ satisfy $D(A,B) \geq 2^{-\zeta n}$, then there exist subsets $A' \subseteq A$, $|A'| \geq 2^{-\delta n}|A|$ and $B' \subseteq B$, $|B'| \geq 2^{-\delta n}|B|$ such that $D(A',B') = 1$.*

As to the converse direction, it was shown in [BZ11] that the above conjecture implies the following weakening of the PFR conjecture.

Theorem 4.2 ([BZ11])**.** *Assuming Conjecture 4.1, for every $1 > \alpha > 0$ and $1 > \delta > 0$, there exists an integer r which depends only on α and δ, such that if $A \subset \mathbb{F}_2^n$ has $2^{\alpha n} \leq |A| \leq 2^{(1-\alpha)n}$ and $|A+A| \leq K|A|$, then there exists a subset A' of A of size at least $2^{-\delta n}K^{-r}|A|$ such that $|\mathrm{span}\,(A')| \leq |A|$.*

Note that the above conjecture differs from the standard PFR conjecture in two ways. First, in the above conjecture the set A must be of high density, and the exponent r depends on the density of the set. Second, the loss in the size of A is multiplied by an exponential factor. An interesting problem raised by the work of [BZ11] was whether one could find an approximate duality type conjecture which is equivalent to another PFR type conjecture. In what follows we give an example of a pair of such conjectures, by showing that the above Conjecture 4.1 is equivalent to the following weakening of the PFR conjecture.

Conjecture 4.3 (PFR conjecture, exponential range)**.** *For every constant δ' there exists a constant ζ', depending only on δ', such that if $A \subseteq \mathbb{F}_2^n$ has $|A+A| \leq 2^{\zeta' n}|A|$, then there exists a subset $A' \subseteq A$ of size at least $2^{-\delta' n}|A|$ such that $|\mathrm{span}\,(A')| \leq |A|$.*

Note that the PFR conjecture (Conjecture 1.5) implies the above conjecture with $\zeta' = \delta'/r$ for some universal integer r. The above conjecture is in fact weaker than the PFR conjecture since we allow ζ' to be an arbitrary function of δ'. Our main result in this section is that the above Conjectures 4.1 and 4.3 are equivalent.

Theorem 4.4. *Conjecture 4.1 is equivalent to conjecture 4.3.*

The fact that Conjecture 4.3 implies Conjecture 4.1 follows from our proof of the main technical Lemma 1.11. We have already noted in Section 1.3 that Lemma 1.11 implies that Conjecture 4.1 holds assuming the PFR conjecture (by setting $K = 2^{\delta n/(3r)}$, $t = 3r/\delta$, $\zeta = \delta/(3r\cdot 2^t) = \delta/(3r\cdot 2^{3r/\delta})$, $\epsilon = 2^{-\zeta n}$ in Lemma 1.11). Inspecting the proof of Lemma 1.11 it turns out that plugging the weaker Conjecture 4.3 instead of the PFR conjecture in the proof of Lemma 1.11 suffices for obtaining Conjecture 4.1.

In the remaining of the section we show that Conjecture 4.1 implies Conjecture 4.3. For the proof of this implication we follow the approach of [BZ11]. In particular we use the following lemma from [TV06] (appearing there as Lemma 4.38) which shows that a set having a small sum set must have large spectrum:

Lemma 4.5 (Small sumset forces large spectrum)**.** *Let A be a subset of a finite abelian group Z, and let $0 < \epsilon \leq 1$. Then we have the following lower bound on the sum set:*

$$|A-A| \geq \frac{|A||Z|}{|A||\mathrm{Spec}_\epsilon(A)| + |Z|\epsilon^2}$$

Note that in $\mathbb{F}_2^n$ we have that $A - A = A + A$.

Proof of (Conjecture 4.1 ⇒ Conjecture 4.3). The idea of the proof of is as follows. Suppose that A has a small sum set. Then Lemma 4.5 implies that A has large spectrum, denote the spectrum set by B. Assuming Conjecture 4.1, we have that A and B contain large subsets A', B' respectively which lie in affine shifts of dual subspaces. But this implies in turn that $\dim(A') \leq n - \dim(B')$, i.e. A' has a small span, and setting the parameters correctly we arrive at the desired result. Details follow.

Let ζ be the constant guaranteed by Conjecture 4.1 for the constant $\delta = \delta'/3$. Also, let $\zeta' = \min\{\delta'/6, \zeta\}$, and suppose that $|A + A| \leq 2^{\zeta' n}|A|$. In Lemma 4.5 set $\epsilon = 2^{-\zeta n}$. Then from the lemma and the assumption that $|A + A| \leq 2^{\zeta' n}|A|$ we have

$$2^{\zeta' n}|A| \geq |A - A| \geq \frac{|A|2^n}{|A||\mathrm{Spec}_\epsilon(A)| + 2^n\epsilon^2}$$

And rearranging we obtain

$$|\mathrm{Spec}_\epsilon(A)| \geq \frac{2^n(1 - 2^{\zeta' n}\epsilon^2)}{2^{\zeta' n}|A|} \geq \frac{2^n(1 - 2^{-\zeta' n})}{2^{\zeta' n}|A|} \geq \frac{2^n}{2^{2\zeta' n}|A|}$$

where the second inequality is due to our choice of $\epsilon = 2^{-\zeta n} \leq 2^{-\zeta' n}$.

Conjecture 4.1 then implies (noting that $\epsilon = 2^{-\zeta n}$) the existence of subsets $A' \subseteq A$, $B' \subseteq \mathrm{Spec}_\epsilon(A)$ which lie in affine shifts of dual spaces such that $|A'| \geq 2^{-(\delta'/3)n}|A|$, $|B'| \geq 2^{-(\delta'/3)n}|\mathrm{Spec}_\epsilon(A)|$.

But this implies in turn that $\dim(A') + \dim(B') \leq n$, and consequently

$$|\mathrm{span}\,(A')\,| \leq \frac{2^n}{|B'|} \leq \frac{2^{(\delta'/3)n} \cdot 2^n}{|\mathrm{Spec}_\epsilon(A)|} \leq 2^{(\delta'/3)n}2^{2\zeta' n}|A| \leq 2^{2\delta' n/3}|A|$$

where the last inequality is due to our choice of $\zeta' \leq \delta'/6$.

Concluding, we have that $|\mathrm{span}\,(A')\,| \leq 2^{2\delta' n/3}|A|$ where A' is a subset of A of size at least $2^{-(\delta'/3)n}|A|$. Write $\mathrm{span}\,(A')$ as a direct sum of subspaces L_1 and L_2, where L_2 is a subspace of size $2^{2\delta' n/3}$, and L_1 is a subspace of size at most $|A|$ which maximizes the size of $A'' = A' \cap L_1$. We have that A'' is a subset of A of size at least $2^{-\delta' n}|A|$ such that $|\mathrm{span}\,(A'')\,| \leq |L_1| \leq |A|$ which concludes the proof that Conjecture 4.1 ⇒ Conjecture 4.3.

□

Acknowledgements. We thank Nati Linial and Eyal Kushilevitz for drawing our attention, independently, to the similarities between the notions of discrepancy and approximate duality, which led us to consider the question addressed in this paper.

References

[BDL12] Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds for matching vector codes. 2012. Submitted.

[BS94] Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[BZ11] Eli Ben-Sasson and Noga Zewi. From affine to two-source extractors via approximate duality. In *the Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC'2011*, New York, 2011. ACM Press.

[CS10] Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions, September 14 2010. Comment: 29 pages, to appear in GAFA.

[EZ11] Chaim Even-Zohar. On sums of generating sets in $(Z_2)^n$, 2011.

[Faj88] Siemion Fajtlowicz. On conjectures of graffiti. *Discrete Mathematics*, 72(1-3):113–118, 1988.

[Gow98] William Timothy Gowers. A new proof of Szemerèdi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[GR06] Ben Green and Imre Z. Ruzsa. Sets with small sumset and rectification. *Bulletin of the London Mathematical Society*, 1(38):43–52, 2006.

[Gre05] Ben Green. Finite field models in additive combinatorics. In *London Mathematical Society Lecture Note Series*, volume 324. Cambridge University Press, 2005.

[GT09] Ben Green and Terence Tao. Freiman's theorem in finite fields via extremal set theory. *Combinatorics, Probability & Computing*, 18(3):335–355, 2009.

[GT10] Ben Green and Terence Tao. An equivalence between inverse sumset theorems and inverse conjectures for the u^3 norm. *Math. Proc. Cambridge Philos. Soc.*, 149(1):1–19, 2010.

[KL96] Andrew Kotlov and László Lovász. The rank and size of graphs. *JGT: Journal of Graph Theory*, 23(2):185–189, 1996.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, New York, 1997.

[Kot97] Andrew Kotlov. Rank and chromatic number of a graph. *JGT: Journal of Graph Theory*, 26(1):1–8, 1997.

[Lov10] Shachar Lovett. Equivalence of polynomial conjectures in additive combinatorics. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:10, 2010.

[LS88] László Lovász and Michael E. Saks. Lattices, Möbius functions and communication complexity. In *FOCS*, pages 81–90. IEEE, 1988.

[MS82] Kurt Mehlhorn and Erik M. Schmidt. Las vegas is better than determinism in vlsi and distributed computing (extended abstract). In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, STOC '82, pages 330–337, New York, NY, USA, 1982. ACM.

[Nuf76] C. Van Nuffelen. A bound for the chromatic number of a graph. *American Mathematical Monthly*, (83):265–266, 1976.

[NW95] Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.

[Ruz99] Imre Z. Ruzsa. An analog of Freiman's theorem in groups. *Astèrique*, 258:323–326, 1999.

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 506–515. ACM, 2007.

[San08] Tom Sanders. A note on freiman's theorem in vector spaces. *Combinatorics, Probability & Computing*, 17(2):297–305, 2008.

[San10] Tom Sanders. On the bogolyubov-ruzsa lemma, October 30 2010. Comment: 23 pp.

[TV06] Terence Tau and Van Vu. *Additive Combinatorics*. Cambridge University Press, Cambridge, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de