Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 135 (2012)

Sampling-based proofs of almost-periodicity results and algorithmic applications

Eli Ben-Sasson [*] Noga Ron-Zewi [†] Madhur Tulsiani[‡] Julia Wolf[§]

April 3, 2013

Abstract

We give new and simple combinatorial proofs of *almost-periodicity results* for sumsets of sets with small doubling in the spirit of Croot and Sisask [CS10], whose almost-periodicity lemma has had far-reaching implications in additive combinatorics. We provide an alternative point of view which relies only on Chernoff's bound for sampling, and avoids the need for L^p-norm estimates used in the original proof of Croot and Sisask.

We demonstrate the usefulness of our new approach by showing that one can easily deduce from it two significant recent results proved using Croot and Sisask almost-periodicity – the *quasipolynomial Bogolyubov-Ruzsa lemma* due to Sanders [San10] and a result on large subspaces contained in sumsets of dense sets due to Croot, Laba and Sisask [CŁS11].

We then turn to algorithmic applications, and show that our approach allows for almost-periodicity proofs to be converted in a natural way to probabilistic algorithms that decide membership in almost-periodic sumsets of dense subsets of $\mathbb{F}_2^n$. Exploiting this, we give a new algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma.

Together with the results by the last two authors [TW11], this implies an algorithmic version of the *quadratic Goldreich-Levin theorem* in which the number of terms in the quadratic Fourier decomposition of a given function, as well as the running time of the algorithm, are quasipolynomial in the error parameter ε. The algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma also implies an improvement in running time and performance of the self-corrector for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$ in [TW11].

[*] Department of Computer Science, Technion, Haifa, Israel and MIT, 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.

[†] Department of Computer Science, Technion, Haifa. nogaz@cs.technion.ac.il, Research supported by a scholarship from the Israel Ministry of Science and Technology and by the US-Israel Binational Science Foundation. Part of the research was conducted while the author was a visiting researcher at Microsoft Research New England, Cambridge, MA. The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

[‡] Toyota Technological Institute at Chicago.

[§] Centre de Mathématiques Laurent Schwartz, École Polytechnique, 91128 Palaiseau, France.

ISSN 1433-8092

Contents

1 Introduction

When Croot and Sisask introduced “A probabilistic technique for finding almost-periods of convolutions” in 2009 [CS10], it created quite a splash in the additive combinatorics community. Roughly speaking, their main result says that if $A \subseteq \mathbb{F}_2^n$ is a set whose sumset $A + A = \{a + a' : a, a' \in A\}$ is small, then there exists a dense set X such that the convolution $\mathbb{1}_A * \mathbb{1}_A(\cdot)$ of the indicator function of A with itself, and its translates $\mathbb{1}_A * \mathbb{1}_A(\cdot + x)$ for $x \in X$, are almost indistinguishable in the L^2 and higher L^p norms. This set X may then be referred to as the set of “almost periods”.

The two main combinatorial applications of the above technique were the proof of the quasipolynomial Bogolyubov-Ruzsa lemma due to Sanders [San10] and a result saying that sumsets of dense sets contain large subspaces due to Croot, Łaba and Sisask [CŁS11]. Both applications made crucial use of the L^p-norm estimates of Croot and Sisask, where p was taken to be very large (a function of the density α of the set $A \subseteq \mathbb{F}_2^n$ under investigation, such as $\log \alpha^{-1}$).

Our main result is a simple combinatorial proof of almost-periodicity results in the spirit of Croot and Sisask that proceeds entirely without recourse to L^p-norms, instead only relying on the Chernoff bound for sampling. This is in contrast to Croot and Sisask's approach which obtained L^p-norm estimates using a simple sampling technique combined with tailbounds for a multinomial distribution, which Sanders replaced by the Marcinkiewicz-Zygmund inequality. It is our hope that this proof will appeal to a larger part of the theoretical computer science community than the currently existing ones, thereby increasing the likelihood of further novel applications of the almost-periodicity techniques.

We prove our almost-periodicity results in Section 4. We stress that our almost-periodicity approach works for arbitrary abelian groups, but for simplicity we state our results only over $\mathbb{F}_2^n$. We illustrate the use of our new approach by presenting simplified combinatorial proofs of known additive combinatorics results as well as new algorithmic applications. Let us describe these in more detail.

1.1 Combinatorial applications

In Section 5 we show that both the quasipolynomial Bogolyubov-Ruzsa lemma [San10] and the results of Croot, Laba and Sisask on large subspaces contained in dense sets [CŁS11] can be easily deduced from our almost-periodicity approach.

The quasipolynomial Bogolyubov-Ruzsa lemma. In its original form, the Bogolyubov-Ruzsa lemma states that if $A \subseteq \mathbb{F}_2^n$ is a set of density α, then $4A := A + A + A + A$ contains a subspace of codimension at most $2\alpha^{-2}$. One of the first applications Croot and Sisask gave of their new technique was a *quasipolynomial* Bogolyubov-Ruzsa lemma, which asserted that $4A$ contains iterated sumsets, of density at least $2^{-O(\log^4(1/\alpha))}$ inside $\mathbb{F}_2^n$. It was quickly recognized by Sanders [San10] that the latter result could be boot-strapped, using a little Fourier analysis, to a quasipolynomial version of the Bogolyubov-Ruzsa lemma in which the codimension of the subspace that is found within $4A$ is at most $O(\log^4(\alpha^{-1}))$.

This result has important implications for the bounds in Freiman's theorem, which describes the structure of sets with small sumsets [Ruz99], and to the inverse theorem for the Gowers U^3 norm [Sam07, GT08]. It is also a crucial ingredient in Sanders's groundbreaking upper bound of $C(\log\log N)^5 N/\log N$ for the size of a subset of $\{1, \ldots, N\}$ not containing any 3-term arithmetic progressions [San11b]. An improvement to $O(\log(1/\alpha))$ of the bound on the codimension of the subspace which is contained in $4A$ implies the *polynomial Freiman-Ruzsa conjecture* in additive combinatorics, which has found several applications to complexity theory so far [BZ11, BLR12, BDL13]. See the survey of Green [Gre05] for more information on the polynomial Freiman-Ruzsa conjecture and its combinatorial applications.

Sumsets of dense sets contain large subspaces. Our second combinatorial application concerns the problem of finding large subspaces within sumsets of a dense set. Green [Gre05] had shown that if $A \subseteq \mathbb{F}_2^n$ has density α, then $A + A$ contains a subspace of dimension $\Omega(\alpha^2 n)$. This was improved by Sanders who proved in [San11a] using a Fourier-iteration lemma that this subspace must be of dimension at least $\Omega(\alpha n)$. Croot, Łaba and Sisask [CŁS11], who addressed the more general problem in the integers, asking for long arithmetic progressions in sumsets of dense sets, remarked that a slightly worse bound of the form $\Omega\big(\frac{\alpha}{\log^3(1/\alpha)}n\big)$ follows implicitly from their techniques.

In Section 5.2 we show that the finite-field analogue of the Croot-Łaba-Sisask bound also follows from our almost-periodicity approach. This requires a more careful analysis of our sampling technique, which we shall give in detail in the appendix.

1.2 Algorithmic applications

An advantage of our sampling-based almost-periodicity proofs is that they can be turned into algorithms that decide membership in almost-periodic sumsets of dense subsets of $\mathbb{F}_2^n$ in a rather natural way. In particular, using our new techniques we present in Section 6.1 an algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma. This is an algorithm which runs in time polynomial in n and quasipolynomial in $1/\alpha$, which, given an oracle access to an indicator function of a set A of density at least α inside $\mathbb{F}_2^n$, finds, with high probability, a basis to a subspace $V \subseteq 4A$ of codimension $O(\log^4(1/\alpha))$.

The main problem we encounter when converting our almost-periodicity proofs into such an algorithm is that our combinatorial proofs produce large subsets of $\mathbb{F}_2^n$ of size exponential in n. Since we are interested in an algorithm which runs in time polynomial in n, we cannot afford to store or compute with these sets explicitly. Instead, we develop probabilistic procedures to efficiently test membership in such sets. We elaborate on these procedures in Section 2.

Combined with the results of [TW11], the algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma implies an improvement in the running time and performance guarantee of a self-correction procedure for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$, given in [TW11]. This in turn leads to an improved quadratic Goldreich-Levin theorem in which the running time of the algorithm, as well as the number of terms in the quadratic Fourier decomposition of a given function, are quasipolynomial in the error parameter ε. We elaborate on these applications below.

One major difficulty with obtaining the above applications from the algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma, encountered already in [TW11], is that the individual subroutines in these applications, which correspond to algorithmic versions of theorems in additive combinatorics, are probabilistic in nature. Since they are applied in sequence, this means that the input for the next subroutine comes with a certain amount of noise, and it is therefore necessary to prove robust algorithmic versions of the theorems from additive combinatorics. This applies in particular to the quasipolynomial Bogolyubov-Ruzsa lemma, of which we prove such a robust version in Section 6.1.

An improved self-corrector for the Reed-Muller code of order 2. A central ingredient in the quadratic Goldreich-Levin theorem of [TW11] is a self-correction procedure for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$. More precisely, the authors present a procedure which runs in time polynomial in n and exponential in $1/\varepsilon$, which given a function $f : \mathbb{F}_2^n \to \{-1, 1\}$ of distance at least $1/2 - \varepsilon$ from a quadratic phase $(-1)^q$ (which is a codeword of the Reed-Muller code of order 2), finds a quadratic phase $(-1)^{q'}$ which has distance at most $1/2 - \eta(\varepsilon)$ from f for $\eta(\varepsilon) = \exp(-1/\varepsilon)$.

This self-correction procedure is essentially an algorithmic version of the proof of the inverse theorem for the U^3 norm [Sam07, GT08], which states that if a bounded function f has large U^3 norm, then it correlates with a quadratic phase. As stated above, the Bogolyubov-Ruzsa lemma is crucial in the proof of the inverse theorem, and hence plugging our new algorithmic proof of the quasipolynomial Bogolyubov-Ruzsa lemma into the self-correction procedure of [TW11] we improve the running time of the procedure, as well as the parameter η, to depend only quasipolynomially on $1/\varepsilon$. We elaborate on this in Section 6.2.

We remark that the list decoding radius of the Reed-Muller code of order 2 is $1/4$ [GKZ08, Gop10], and hence at distance $1/2 - \varepsilon$ one cannot expect to find all codewords of distance $1/2 - \varepsilon$ from a given codeword. Instead our self-correction procedure (as well as that of [TW11]) returns only a single codeword that correlates with the original codeword.

An improved quadratic Goldreich-Levin theorem. As mentioned above, the self-correction procedure for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$ plays a substantial role in the work on quadratic decomposition theorems by the last two authors. The aim of such theorems is to decompose any bounded function $f : \mathbb{F}_2^n \to \mathbb{C}$ as a sum $g + h$, where g is quadratically uniform, in the sense that the Gowers U^3 norm $\|g\|_{U^3}$ is small, and h is quadratically structured, in the sense that it is a bounded sum of quadratic phases. These types of decompositions constitute a higher-order analogue of classical Fourier decompositions, and they have found several number-theoretic applications [Can10, GW10b, GW12, GW10a, HL11]. Such decomposition theorems had previously been obtained in an abstract and non-constructive way (either using a form of the Hahn-Banach theorem [GW12], or a so-called energy increment approach [Gre07]).

From a computer science perspective, it is a natural question to ask whether such a decomposition could be computed efficiently. In [TW11], the authors gave a probabilistic algorithm that, given any function $f : \mathbb{F}_2^n \to \mathbb{C}$, would with high probability compute, in time polynomial in n and exponential in $1/\varepsilon$, a quadratic decomposition for that function with a specified U^3 error ε, in which the number of quadratic terms is exponential in $1/\varepsilon$. This essentially amounts to computing a "quadratic Fourier decomposition" for f, and was therefore termed a *quadratic Goldreich-Levin theorem* in analogy with the well-known linear case [GL89].

The quadratic Goldreich-Levin algorithm consists of two parts: a deterministic part which is able to construct the quadratically structured part of f under the assumption that we have an algorithm which provides some quadratic phase function that f correlates with (if there is no such phase function, we just set $g = f$). The algorithm for finding a quadratic phase function, which constitutes the second part of the overall algorithm, is basically the self-correction procedure for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$ described above. Using our improved self-correction procedure we improve the running time, as well as the number of terms that are obtained in the final quadratic decomposition, to depend only quasipolynomially in the uniformity parameter ε. More details are given in Section 6.3.

Acknowledgements. The last two authors would like to thank Tom Sanders for numerous helpful remarks and discussions. The first two authors would like to thank Shachar Lovett for sharing with them his view of Sanders's proof of the quasipolynomial Bogolyubov-Ruzsa lemma (cf. [Lov12]).

2 Techniques

2.1 Sampling-based proofs of almost-periodicity results

The following is the precise statement of the original almost-periodicity lemma of Croot and Sisask (Proposition 1.3 in [CS10]). Since it is valid for general abelian groups G, it is written in multiplicative notation.

Proposition 2.1 (Croot-Sisask Lemma, L^p local version) *Let $\varepsilon > 0$ and let $m \geq 1$ be an integer. Let G be an abelian group and let $A, B \subseteq G$ be finite subsets such that $|B \cdot A| \leq K|B|$. Then there is a set $X \subseteq A$ of size $|X| \geq |A|/(2K)^{50m/\varepsilon}$ such that for each $x \in XX^{-1}$,*

$$\|\mathbb{1}_A * \mathbb{1}_B(yx) - \mathbb{1}_A * \mathbb{1}_B(y)\|_{2m}^{2m} \leq \max\{\varepsilon^m |AB||B|^m, \|\mathbb{1}_A * \mathbb{1}_B\|_m^m\}\varepsilon^m |B|^m.$$

It is known that L^p bounds and Chernoff's inequality are, in a certain sense, equivalent. Specifically, a random variable X obeys a Chernoff-type tail bound of the form

$$\mathbb{P}[|X| \geq t\|X\|_2] \leq C \exp(-\Omega(t^2))$$

if and only if its L^p-norm satisfies

$$\|X\|_p \leq C\sqrt{p}\|X\|_2$$

for all $p \in [2, \infty)$, the latter representing a Khinchine-type inequality (from which Marcinkiewicz-Zygmund can be derived). For a proof of this statement we refer the reader to the excellent lecture notes by Sanders [San12].

Thus it is natural to ask whether one could formulate an 'L^p-norm free' almost-periodicity statement that suffices for applications and whether such a statement could be proven without recourse to L^p-norms. In Section 4.1 we answer this question in the affirmative by proving Proposition 2.2 below. We stress again that this proposition holds over any abelian group, but for simplicity we state it only for the special case of $\mathbb{F}_2^n$.

To state Proposition 2.2 we start with fixing some notation. For a subset $A \subseteq \mathbb{F}_2^n$ we let $\mathbb{1}_A$ denote the indicator function of A and μ_A denote the function $\mathbb{1}_A \cdot (2^n/|A|)$. For two real numbers α, β we write $\alpha \approx_\varepsilon \beta$ to denote $|\alpha - \beta| \leq \varepsilon$ and if $|\alpha - \beta| > \varepsilon$ we write $\alpha \not\approx_\varepsilon \beta$. Given subsets $A, B \subset \mathbb{F}_2^n$, define the *measure of additive containment* $\rho_{A\to B} : \mathbb{F}_2^n \to [0,1]$ by

$$\rho_{A\to B}(y) := \mathbb{P}_{a\in A}[y + a \in B] = \frac{|(y+A) \cap B|}{|A|} = \mu_A * \mathbb{1}_B(y), \tag{1}$$

for each $y \in \mathbb{F}_2^n$. Notice that $\rho_{A\to B}(y) = 1$ when $y + A \subseteq B$ and $\rho_{A\to B}(y) = 0$ when $(y+A) \cap B = \emptyset$.

Proposition 2.2 (Almost-periodicity of sumsets) *If $A \subset \mathbb{F}_2^n$ satisfies $|2A| \leq K|A|$, then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*
2. *The size of X is at least $|A|/(2K^{t-1})$.*
3. *For all $x \in X$ and for all subsets $S \subseteq \mathbb{F}_2^n$,*

$$\mathbb{P}_{y\in S}[\rho_{A\to B}(y) \approx_{2\varepsilon} \rho_{A\to B}(y + x)] \geq 1 - 8\frac{|A+B|}{|S|} \cdot \exp\left(-2\varepsilon^2 t\right). \tag{2}$$

Our proof differs from the original proof of Croot and Sisask in that it disposes of L^p-norms and tail bounds for a multinomial distribution (or the Marcinkiewicz-Zygmund inequality), and replaces them with sampling arguments relying on the Chernoff-Hoeffding bound.

We now sketch the proof. To obtain X we replace $\rho_{A\to B}$ by an estimator function computed by taking a sequence of t independent random samples distributed uniformly over A. Denoting the sample sequence by $\mathbf{a} = (a_1, \dots, a_t)$, we estimate $\rho_{A\to B}(y)$ by the fraction of $a_i \in \mathbf{a}$ satisfying $y + a_i \in B$. Denote the estimator function corresponding to $\mathbf{a}$ by $\hat{\rho}_{\mathbf{a}}$. Fixing y, the Chernoff-Hoeffding bound says that the probability that $\hat{\rho}_{\mathbf{a}}(y)$ differs from $\rho_{A\to B}(y)$ by more than ε, i.e., the probability of the event "$\rho_{A\to B}(y) \not\approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)$" when $\mathbf{a} = (a_1, \dots, a_t)$ is distributed uniformly over A^t, is at most $\exp\left(-\Omega\left(\varepsilon^2 t\right)\right)$.

The key observation in the construction of the set X is that there are many pairs of good estimator sequences $\mathbf{a} = (a_1, \dots, a_t), \hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_t)$ for which there exists a "special" element $x \in \mathbb{F}_2^n$ such that $\hat{\mathbf{a}} = x + \mathbf{a}$, where $x + \mathbf{a} := (x + a_1, \ldots, x + a_t)$. Such x are called "special" for the following reason. We say y is "good" if both of the following conditions hold,

$$\rho_{A \to B}(y) \approx_\varepsilon \hat{\rho}_{\hat{\mathbf{a}}}(y) \quad \text{and} \quad \rho_{A \to B}(y + x) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y + x). \tag{3}$$

Now if $\hat{\mathbf{a}} = x + \mathbf{a}$, then we have

$$\hat{\rho}_{\hat{\mathbf{a}}}(y) = \hat{\rho}_{\hat{\mathbf{a}}}(y + x + x) = \hat{\rho}_{\hat{\mathbf{a}}+x}(y + x) = \hat{\rho}_{\mathbf{a}}(y + x),$$

and combining this with (3) implies that for "good" y we have $\rho_{A \to B}(y) \approx_{2\varepsilon} \rho_{A \to B}(y + x)$. Thus, to prove the proposition we only need to bound from below the number of "special" elements x, which is done based on the assumption that A has small doubling.

For applications, one would like a version of almost-periodicity in which the set X is replaced with a subspace. It was observed by Sanders [San10] that such a version could be deduced from Proposition 2.1 above using Fourier analysis arguments. In Section 4.2 we use Sanders's approach to deduce such a version also from our Proposition 2.2. In Section 5 we show how known combinatorial results can easily be deduced from this version. We now turn to describing the techniques used in our algorithmic version of almost-periodicity results.

2.2 Algorithmic versions of almost-periodicity results

With a view to algorithmic applications, we prove an algorithmic version of our almost-periodicity results in Section 6.1. As previously noted, the main difficulty in obtaining such an algorithmic version is that the combinatorial proofs of these results use the description of large subsets of $\mathbb{F}_2^n$, whose size is exponential in n. Since we are interested in an algorithm which runs in time polynomial in n we do not have time to describe and inspect these sets as a whole. Instead, we use random sampling methods to decide efficiently membership in such sets.

For instance, one of the first issues we need to deal with is that for our algorithmic applications we need to compute the measure $\rho_{A \to 2A}(y) = \mathbb{P}_{a \in A}[a + y \in 2A]$. In order to compute this measure algorithmically one has to have access to the indicator function for the sumset $2A$, whereas we only have oracle access to the indicator function of A. In order to deal with this, we observe that an element $y \in \mathbb{F}_2^n$ satisfies $a + y \in 2A$ if and only if $\mathbb{1}_A * \mathbb{1}_A(a + y) > 0$, and that the latter quantity can be estimated using sampling. In order to handle possible error in the estimation of $\mathbb{1}_A * \mathbb{1}_A(a + y) > 0$ we need to make some further modifications in the combinatorial proof.

Another core issue that we need to deal with is how to efficiently find the set X of almost periods. In the combinatorial proof the existence of the set X is shown in a non-constructive way using the pigeonhole principle. In order to find the set X in a constructive manner we prove that for a random string $\hat{\mathbf{a}} = (a_1, \ldots, a_t) \in (\mathbb{F}_2^n)^t$ many translates $\hat{\mathbf{a}} + x := (a_1 + x, \ldots, a_t + x)$ of $\hat{\mathbf{a}}$, as well as $\hat{\mathbf{a}}$ itself, will be good estimator sequences. In particular, one can take the set X to be the set of all $x \in \mathbb{F}_2^n$ such that $\hat{\mathbf{a}} + x$ is a good estimator sequence. For this one needs an efficient procedure for testing whether a given sequence is a good estimator sequence, and we obtain such a procedure using the above-mentioned idea for estimating $\rho_{A \to 2A}(y)$.

Finally, in our combinatorial proof we show that X can be approximated by a subspace V using Fourier analysis. It follows that in order to find the subspace V, it suffices to inspect the large Fourier coefficients of $\mathbb{1}_X$. This can be done efficiently using the (standard) Goldreich-Levin theorem (Theorem 6.7).

3 Preliminaries

In this section we fix our notation and collect some results that we shall use throughout the paper. Fundamental to our approach will be the following Chernoff-type tail bound for sampling [TV06].

Lemma 3.1 (Hoeffding bound for sampling) *If $\mathbf{X}$ is a random variable with $|\mathbf{X}| \leq 1$ and $\hat{\mu}$ is the empirical average obtained from t samples, then*

$$\mathbb{P}\left[|\mathbb{E}\left[\mathbf{X}\right] - \hat{\mu}| \; > \; \gamma\right] \; \leq \; 2\exp(-2\gamma^2 t).$$

Throughout the paper we shall make use of the discrete Fourier transform, which we define as follows. For $f : \mathbb{F}_2^n \to \mathbb{C}$, let

$$\widehat{f}(t) = \mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{x\cdot t}$$

for any $t \in \widehat{\mathbb{F}_2^n} = \mathbb{F}_2^n$, where $\mathbb{E}_{x\in\mathbb{F}_2^n}$ simply stands for the normalized sum $2^{-n}\sum_{x\in\mathbb{F}_2^n}$ and $x \cdot t = \sum_{i=1}^n x_i t_i$ for a pair of vectors $x = (x_1, \ldots, x_n), t = (t_1, \ldots, t_n) \in \mathbb{F}_2^n$. The inversion formula states that

$$f(x) = \sum_{t\in\mathbb{F}_2^n} \widehat{f}(t)(-1)^{x\cdot t}$$

for all $x \in \mathbb{F}_2^n$, and Parseval's identity takes the form

$$\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)\overline{g(x)} = \sum_{t\in\mathbb{F}_2^n} \widehat{f}(t)\overline{\widehat{g}(t)},$$

for any two functions $f, g : \mathbb{F}_2^n \to \mathbb{C}$. Finally, the convolution of two such functions is defined by

$$f * g(x) = \mathbb{E}_{y\in\mathbb{F}_2^n} f(y)g(x-y),$$

and the fact that the Fourier transform diagonlizes the convolution operator is expressed via the idenity

$$\widehat{f * g}(t) = \widehat{f}(t)\widehat{g}(t),$$

which holds for all $t \in \mathbb{F}_2^n$.

The set of large Fourier coefficients determines the value of a function to a significant extent, and for many arguments it is important to be able to estimate its size and determine its structure. For a function $f : \mathbb{F}_2^n \to \mathbb{C}$, let

$$\mathrm{Spec}_\rho(f) = \{t \in \mathbb{F}_2^n : |\widehat{f}(t)| \geq \rho\|f\|_1\}. \tag{4}$$

For a subset $A \subseteq \mathbb{F}_2^n$ we let $\mathbb{1}_A$ denote the indicator function of A and μ_A denote the function $\mathbb{1}_A \cdot (2^n/|A|)$ so that $\mathbb{E}_{x\in\mathbb{F}_2^n}[\mu_A(x)] = 1$. In the special case where $f = \mathbb{1}_A$ for a subset $A \subseteq \mathbb{F}_2^n$ of density α, Parseval's identity tells us that $|\mathrm{Spec}_\rho(\mathbb{1}_A)| \leq \rho^{-2} \cdot \alpha^{-1}$. A more precise result is known: Chang's theorem [Cha02] states that $\mathrm{Spec}_\rho(\mathbb{1}_A)$ is in fact contained in a subspace of dimension at most $C\rho^{-2}\log\alpha^{-1}$.

Theorem 3.2 (Chang's theorem) *Let $\rho \in (0, 1]$ and $A \subseteq \mathbb{F}_2^n$. Then there is a subspace V of $\mathbb{F}_2^n$ such that $\mathrm{Spec}_\rho(\mathbb{1}_A) \subseteq V$ and*

$$\dim(V) \leq 8\frac{\log(2^n/|A|)}{\rho^2}.$$

For an elegant recent proof of this result using entropy, see Impagliazzo et al. [IMR12].

Finally, for two real numbers α, β we write $\alpha \approx_\varepsilon \beta$ to denote $|\alpha - \beta| \leq \varepsilon$ and if $|\alpha - \beta| > \varepsilon$ we write $\alpha \not\approx_\varepsilon \beta$. All logarithms in this paper are taken to base 2.

4 Sampling-based proofs of almost-periodicity results

4.1 Croot-Sisask almost-periodicity

We start with the proof of Proposition 2.2 which we restate below for convenience.

Proposition 2.2 (restated). *If $A \subset \mathbb{F}_2^n$ satisfies $|2A| \leq K|A|$, then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*

2. *The size of X is at least $|A|/(2K^{t-1})$.*

3. *For all $x \in X$ and for all subsets $S \subseteq \mathbb{F}_2^n$,*

$$\mathop{\mathbb{P}}_{y \in S}\left[\rho_{A \to B}(y) \approx_{2\varepsilon} \rho_{A \to B}(y+x)\right] \geq 1 - 8\frac{|A+B|}{|S|} \cdot \exp\left(-2\varepsilon^2 t\right).$$

Proof: Recall the definition of $\rho_{A \to B}(y)$ given in (1). To simplify notation let $\rho(y) := \rho_{A \to B}(y)$, and for a sequence $\mathbf{a} = (a_1, \ldots, a_t)$ of length t, define the $\mathbf{a}$-*estimator* of ρ to be the function $\hat{\rho}_{\mathbf{a}} : \mathbb{F}_2^n \to [0,1]$ defined for $y \in \mathbb{F}_2^n$ by

$$\hat{\rho}_{\mathbf{a}}(y) := \frac{|\{y + a_i \in B \mid i = 1, \ldots, t\}|}{t}.$$

We say that $\mathbf{a}$ is an ε-*good estimator* for y if $\rho(y) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)$.

Fix $y \in \mathbb{F}_2^n$. Our first step towards constructing X is to show that most sample-sequences from A are ε-good for y, provided that t, the sample size, is large enough with respect to $1/\varepsilon$. Let Y_i be the indicator random variable for the event "$y + a_i \in B$" when a_i is chosen uniformly at random from A. Then $\hat{\rho}_{\mathbf{a}}(y) = \frac{1}{t}\sum_{i=1}^t Y_i$ is the average of t i.i.d. indicator random variables each having mean $\rho(y)$, so the Chernoff-Hoeffding bound (Lemma 3.1) implies that for each $y \in \mathbb{F}_2^n$,

$$\mathop{\mathbb{P}}_{\mathbf{a} \in A^t}\left[\rho(y) \not\approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)\right] \leq 2\exp\left(-2\varepsilon^2 t\right). \tag{5}$$

Now we proceed to show that most $\mathbf{a} \in A^t$ are ε-good estimators for most y. Let $Z_{\mathbf{a}}$ be the random variable measuring the fraction of $y \in A + B$ for which $\mathbf{a}$ is an ε-good estimator, that is,

$$Z_{\mathbf{a}} := \mathop{\mathbb{P}}_{y \in A+B}\left[\rho(y) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)\right].$$

Setting $\delta = 2\exp\left(-2\varepsilon^2 t\right)$, we conclude from (5) via linearity of expectation that

$$\mathbb{E}_{\mathbf{a} \in A^t}\left[Z_{\mathbf{a}}\right] \geq 1 - \delta.$$

Markov's inequality now shows that at least half of the sequences $\mathbf{a} \in A^t$ are ε-good estimators for all but a (2δ)-fraction of $y \in A + B$, in which case we say that $\mathbf{a}$ is an $(\varepsilon, 2\delta)$-*good estimator* for ρ. Denote by $\mathbf{G}[\varepsilon, 2\delta] \subset A^t$ the set of these sequences,

$$\mathbf{G}[\varepsilon, 2\delta] = \left\{\mathbf{a} \in A^t \mid \mathop{\mathbb{P}}_{y \in A+B}\left[\rho(y) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)\right] \geq 1 - 2\delta\right\}. \tag{6}$$

To obtain X we partition $\mathbf{G}[\varepsilon, 2\delta]$ as follows. Define a mapping $\varphi : A^t \mapsto \{0\} \times (2A)^{t-1}$ by shifting a sequence $\mathbf{a} = (a_1, \ldots, a_t)$ by its first element a_1,

$$\varphi(\mathbf{a}) = \mathbf{a} + a_1 := (a_1 + a_1, a_1 + a_2, \ldots, a_1 + a_t) \tag{7}$$

Then φ maps the set $\mathbf{G}[\varepsilon, 2\delta]$, which has size at least $|A|^t/2$, into a set of size $|2A|^{t-1} \leq (K|A|)^{t-1}$ so by the pigeonhole principle, there is a subset $\mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}} \subset \mathbf{G}[\varepsilon, 2\delta]$ that is mapped to the same element $\mathbf{b} = (0, b_2, \ldots, b_t)$. In addition, this subset is pretty large,

$$|\mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}| \geq \frac{|A|^t}{2K^{t-1}|A|^{t-1}} = \frac{|A|}{2K^{t-1}}. \tag{8}$$

Finally, fix an arbitrary $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_t) \in \mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}$ and set

$$X = \{\hat{a}_1 + a_1 \mid (a_1, \ldots, a_t) \in \mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}\}.$$

To complete our proof we show that X has the three properties listed in the statement of the lemma.

1. By definition, $X \subseteq \hat{a}_1 + A$.

2. The mapping $\mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}} \mapsto X$ given by $(a_1, \ldots, a_t) \mapsto \hat{a}_1 + a_1$ is invertible, because both $\hat{a}_1$ and $\mathbf{b}$ are fixed. Hence $|X| = |\mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}|$ and the size of X is bounded from below using (8).

3. Suppose $x = \hat{a}_1 + a_1$, where a_1 is the first element of an $(\varepsilon, 2\delta)$-good estimator $\mathbf{a} = (a_1, \ldots, a_t) \in \mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}$. The key observation is that $\hat{\mathbf{a}} + x = \mathbf{a}$. Indeed, the definition of $\mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}$ implies $\varphi(\hat{\mathbf{a}}) = \varphi(\mathbf{a})$, so using (7) we have

   $$\hat{a}_1 + \hat{a}_i \;=\; a_1 + a_i, \quad i = 1, \ldots, t,$$

   which, rearranging, comes out to

   $$a_i = x + \hat{a}_i, \quad i = 1, \ldots, t.$$

   In other words, $\hat{\mathbf{a}} + x = \mathbf{a}$ as claimed.

   Recalling that $\hat{\mathbf{a}}$ is an $(\varepsilon, 2\delta)$-good estimator, we know that for all but a (2δ)-fraction of $y \in A + B$,

   $$\rho(y) \approx_\varepsilon \hat{\rho}_{\hat{\mathbf{a}}}(y) \tag{9}$$

   and (9) also holds for all $y \notin A + B$ since in this case $\hat{\rho}_{\hat{\mathbf{a}}}(y) = \rho(y) = 0$. Hence we have that (9) holds for all but a $(2\delta|A + B|/|S|)$-fraction of $y \in S$.

   Similarly, since $\mathbf{a}$ is also an $(\varepsilon, 2\delta)$-good estimator, we have that

   $$\rho(y + x) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y + x) \tag{10}$$

   for all but a $(2\delta|A + B|/|x + S|)$-fraction of $y \in S$. Using a union bound and the fact that $|S + x| = |S|$, we find that for all but a $(4\delta|A + B|/|S|)$-fraction of $y \in S$ both (9) and (10) hold. For such y we conclude that $\rho(y) \approx_{2\varepsilon} \rho(y + x)$ using the triangle inequality and the fact that $\hat{\rho}_{\hat{\mathbf{a}}}(y) = \hat{\rho}_{\hat{\mathbf{a}}+x}(y + x) = \hat{\rho}_{\mathbf{a}}(y + x)$.

This completes the proof of the proposition. ■

By an inductive application of Proposition 2.2 (using the triangle inequality and the union bound), one can prove the following iterated version.

Corollary 4.1 (Almost-periodicity of sumsets, iterated) *If $A \subset \mathbb{F}_2^n$ satisfies $|2A| \leq K|A|$, then for every integer t and set $B \subseteq \mathbb{F}_2^n$, there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*
2. *The size of X is at least $|A|/(2K^{t-1})$.*
3. *For all $x_1, \ldots, x_\ell \in X$ and for all subsets $S \subseteq \mathbb{F}_2^n$,*

$$\mathop{\mathbb{P}}_{y \in S}\left[\rho_{A\to B}(y) \approx_{2\varepsilon\ell} \rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] \geq 1 - 8\ell\frac{|A+B|}{|S|} \cdot \exp\left(-2\varepsilon^2 t\right). \tag{11}$$

Proof: The proof is by induction on ℓ. Proposition 2.2 establishes the case $\ell = 1$. For the induction step, suppose that the lemma holds for some integer $\ell \geq 1$ with a set $X \subseteq \mathbb{F}_2^n$. We shall show that the same set X satisfies the above requirements for $\ell + 1$.

Let $\delta := 8(|A+B|/|S|) \cdot \exp\left(-2\varepsilon^2 t\right)$. By the induction hypothesis, for at least a $(1-\ell\delta)$-fraction of $y \in S$, it is true that $\rho_{A\to B}(y) \approx_{2\varepsilon\ell} \rho_{A\to B}(y+x_1+\ldots+x_\ell)$. The case $\ell = 1$ implies that for at least a $(1-\delta)$-fraction of $y \in S$, it is true that $\rho_{A\to B}(y + x_1 + \ldots + x_\ell) \approx_{2\varepsilon} \rho_{A\to B}(y + x_1 + \ldots + x_{\ell+1})$. Thus by a union bound we have that for at least a $(1 - (\ell+1)\delta)$-fraction of $y \in S$ we have both $\rho_{A\to B}(y) \approx_{2\varepsilon\ell} \rho_{A\to B}(y + x_1 + \ldots + x_\ell)$ and $\rho_{A\to B}(y + x_1 + \ldots + x_\ell) \approx_{2\varepsilon} \rho_{A\to B}(y + x_1 + \ldots + x_{\ell+1})$. The proof is completed by noting that by the triangle inequality, for each such y, we also have $\rho_{A\to B}(y) \approx_{2\varepsilon(\ell+1)} \rho_{A\to B}(y + x_1 + \ldots + x_{\ell+1})$. ■

4.2 Almost-periodicity over a subspace

For applications one would like a version of Proposition 2.2 in which the set X of periods is in fact a subspace. It was observed by Sanders [San10] that one can use iterated almost-periodicity statements such as Corollary 4.1, combined with some Fourier analysis, to obtain such a subspace. Here we use Sanders's argument to deduce the following statement from Corollary 4.1.

Corollary 4.2 (Almost-periodicity of sumsets over a subspace) *If $A \subset \mathbb{F}_2^n$ is a subset of density α, then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a subspace V of codimension* $\mathrm{codim}(V) \leq 32\log(2/\alpha^t)$ *with the following property.*

For every $v \in V$, for all subsets $S \subseteq \mathbb{F}_2^n$ and for every $\varepsilon, \eta > 0$ and integer ℓ,

$$\mathop{\mathbb{P}}_{y \in S}\left[\rho_{A\to B}(y) \approx_{\varepsilon'} \rho_{A\to B}(y + v)\right] \geq 1 - 16\frac{\ell}{\eta}\frac{|A+B|}{|S|} \cdot \exp\left(-2\varepsilon^2 t\right), \tag{12}$$

where $\varepsilon' = 4\varepsilon\ell + 2\eta + 2^{-\ell}\sqrt{|B|/|A|}$.

As we shall see in Section 5, the proof of the quasipolynomial Bogolyubov-Ruzsa lemma (Theorem 5.1) follows easily from the above corollary, and the result of Croot, Laba and Sisask on the existence of subspaces in sumsets of dense sets (Theorem 5.3) follows easily from a refinement of the above corollary (which we will give as Corollary 5.4 below). Note that for the proof of Corollary 4.2 we need the stronger assumption that A has density at least α in $\mathbb{F}_2^n$, instead of the doubling hypothesis $|2A| \leq K|A|$.

The idea of the proof of Corollary 4.2 is the following. Let X be the subset guaranteed by Corollary 4.1 for $K = 1/\alpha$, and define the subspace V as $V = \mathrm{Spec}_{1/2}(X)^{\perp}$ (see Section 3 for the definition of Spec_{ρ}). The intuition is that if X were a subspace then $\mathrm{Spec}_{1/2}(X) = V^{\perp}$, and hence $V = X$. Thus V serves as an "approximate subspace" for X. Since A is dense in $\mathbb{F}_2^n$, by Corollary 4.1 we also have that X is dense in $\mathbb{F}_2^n$ and hence Chang's theorem (Theorem 3.2) implies that the subspace V is also dense in $\mathbb{F}_2^n$ (this is the only place where we need the stronger assumption on the density of A).

In order to show that (12) holds we first show, using Corollary 4.1, a simple averaging argument and the triangle inequality, that for most $y \in S$,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] \approx_{2\varepsilon\ell+\eta} \rho_{A\to B}(y)\,. \tag{13}$$

Similarly, for all $v \in V$ and for most $y \in S$,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + v + x_1 + \ldots + x_\ell)\right] \approx_{2\varepsilon\ell+\eta} \rho_{A\to B}(y+v)\,. \tag{14}$$

We then use Fourier analysis, following Sanders's argument closely, to show that for *all* $y \in \mathbb{F}_2^n$,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] \approx_{2^{-\ell}\sqrt{|B|/|A|}} \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + v + x_1 + \ldots + x_\ell)\right]\,, \tag{15}$$

where v is again an arbitrary element of V. The final conclusion follows from (13), (14) and (15) using the union bound and the triangle inequality. We start by establishing (13) and (14).

Lemma 4.3 *Let $\varepsilon, \delta > 0$, and let $A, B, X, S \subseteq \mathbb{F}_2^n$ be such that for all $x_1, \ldots, x_\ell \in X$,*

$$\mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to B}(y) \approx_{\varepsilon} \rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] \geq 1 - \delta\,.$$

Then for every $\eta > 0$ we have that

$$\mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to B}(y) \approx_{\varepsilon+\eta} \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right]\right] \geq 1 - \delta/\eta\,.$$

Proof: From Markov's inequality it follows that for at least a $(1-\delta/\eta)$-fraction of $y \in S$, the relation

$$\rho_{A\to B}(y) \approx_{\varepsilon} \rho_{A\to B}(y + x_1 + \ldots + x_\ell)$$

holds for at least a $(1-\eta)$-fraction of ℓ-tuples $(x_1, \ldots, x_\ell) \in X^\ell$. Thus for at least a $(1-\delta/\eta)$-fraction of $y \in S$, we have that

$$\left|\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] - \rho_{A\to B}(y)\right| \leq \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\left|\rho_{A\to B}(y + x_1 + \ldots + x_\ell) - \rho_{A\to B}(y)\right|\right]$$

which is seen to be bounded above by $(1-\eta)\cdot\varepsilon + \eta \cdot 1 \leq \varepsilon + \eta$. ■

The next lemma establishes (15).

Lemma 4.4 *Let $X \subseteq \mathbb{F}_2^n$, and let $V \subseteq \mathrm{Spec}_{1/2}(X)^{\perp}$. Then for all $y \in \mathbb{F}_2^n$ and $v \in V$,*

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + x_1 + \ldots + x_\ell)\right] \approx_{\varepsilon''} \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X}\left[\rho_{A\to B}(y + v + x_1 + \ldots + x_\ell)\right]\,, \tag{16}$$

where $\varepsilon'' = 2^{-\ell}\sqrt{|B|/|A|}$.

Proof: We can write the difference between the two sides of (16) using the convolution operator as

$$(\mu_X)^{*\ell} * \mu_A * \mathbb{1}_B(y) - (\mu_X)^{*\ell} * \mu_A * \mathbb{1}_B(y+v),$$

which in terms of the Fourier basis equals

$$\sum_{t \in \mathbb{F}_2^n} \hat{\mu}_A(t) \cdot (\hat{\mu}_X(t))^\ell \cdot \widehat{\mathbb{1}_B}(t) \cdot \left((-1)^{y \cdot t} - (-1)^{(y+v) \cdot t} \right) .$$

This expression in turn is bounded in absolute value by

$$\sum_{t \in \mathbb{F}_2^n} |\hat{\mu}_A(t)| \cdot |\hat{\mu}_X(t)|^\ell \cdot \left| \widehat{\mathbb{1}_B}(t) \right| \cdot \left| (-1)^{y \cdot t} \right| \cdot \left| 1 - (-1)^{v \cdot t} \right| = \sum_{t \in \mathbb{F}_2^n} |\hat{\mu}_A(t)| \cdot |\hat{\mu}_X(t)|^\ell \cdot \left| \widehat{\mathbb{1}_B}(t) \right| \cdot \left| 1 - (-1)^{v \cdot t} \right| .$$

By definition of V as the orthogonal complement of $\mathrm{Spec}_{1/2}(X)$, the right-hand side can be bounded as

$$\sum_{t \notin V^\perp} |\hat{\mu}_A(t)| \cdot |\hat{\mu}_X(t)|^\ell \cdot \left| \widehat{\mathbb{1}_B}(t) \right| \cdot \left| 1 - (-1)^{v \cdot t} \right| \leq 2^{-\ell} \sum_{t \notin V^\perp} |\hat{\mu}_A(t)| \cdot \left| \widehat{\mathbb{1}_B}(t) \right| .$$

By the Cauchy-Schwarz inequality and Parseval's indentity, this is bounded above by

$$2^{-\ell} \sqrt{\sum_{t \notin V^\perp} (\hat{\mu}_A(t))^2} \sqrt{\sum_{t \notin V^\perp} (\hat{\mathbf{1}}_B(t))^2} \leq 2^{-\ell} \sqrt{\mathbb{E}_{y \in \mathbb{F}_2^n} (\mu_A(y))^2} \sqrt{\mathbb{E}_{y \in \mathbb{F}_2^n} (\mathbf{1}_B(y))^2} = 2^{-\ell} \sqrt{|B|/|A|} .$$

■

We are now ready for the proof of Corollary 4.2.

Proof of Corollary 4.2: Let X be the set guaranteed by Corollary 4.1 for $K = 1/\alpha$, and let $V = \mathrm{Spec}_{1/2}(X)^\perp$. First, note that Property 2 of Corollary 4.1 implies that $|X| \geq |A|/(2(1/\alpha)^{t-1}) \geq \alpha^t \cdot 2^{n-1}$. It now follows from Chang's theorem (Theorem 3.2) that

$$\dim(\mathrm{Spec}_{1/2}(X)) = \mathrm{codim}(V) \leq 8 \frac{\log(2/\alpha^t)}{(1/2)^2} = 32 \log(2/\alpha^t).$$

It remains to show that (12) holds. Let $\delta := 8\ell(|A+B|/|S|) \cdot \exp\left(-2\varepsilon^2 t\right)$. From Corollary 4.1 and Lemma 4.3 we have that

$$\rho_{A \to B}(y) \approx_{2\varepsilon\ell+\eta} \mathbb{E}_{x_1,\ldots,x_\ell \in X}[\rho_{A \to B}(y + x_1 + \ldots + x_\ell)] \tag{17}$$

for at least a $(1 - \delta/\eta)$-fraction of $y \in S$, and similarly that for all $v \in V$,

$$\rho_{A \to B}(y+v) \approx_{2\varepsilon\ell+\eta} \mathbb{E}_{x_1,\ldots,x_\ell \in X}[\rho_{A \to B}(y + v + x_1 + \ldots + x_\ell)] \tag{18}$$

for at least a $(1 - \delta/\eta)$-fraction of $y \in S$. Moreover, Lemma 4.4 implies that for every $y \in S$ and $v \in V$,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X} [\rho_{A \to B}(y + x_1 + \ldots + x_\ell)] \approx_{2^{-\ell}\sqrt{|B/A|}} \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in X} [\rho_{A \to B}(y + v + x_1 + \ldots + x_\ell)] . \tag{19}$$

Applying the union bound and the triangle inequality to (17), (18) and (19), we conclude that

$$\rho_{A \to B}(y) \approx_{\varepsilon'} \rho_{A \to B}(y+v)$$

for $\varepsilon' = 4\varepsilon\ell + 2\eta + 2^{-\ell}\sqrt{|B|/|A|}$ for at least a $(1 - 2\delta/\eta)$-fraction of $y \in S$, which is the desired conclusion. ■

5 Combinatorial applications

5.1 The quasipolynomial Bogolyubov-Ruzsa lemma

In the context of $\mathbb{F}_2^n$, the traditional Bogolyubov-Ruzsa lemma states that if a set A has density at least α in its ambient group, then its fourfold sumset $A+A+A+A$ contains a subspace of codimension at most $2\alpha^{-2}$. It is easily proved using a few lines of Fourier analysis: the orthogonal complement of the subspace is given by the frequencies at which the indicator function of A has relatively large Fourier coefficients.

The bound on the codimension of V was improved to $O(\log^4(\alpha^{-1}))$ by Sanders [San10]. This improvement has far-reaching quantitative implications for other problems, in particular to the bounds in Freiman's theorem [Ruz99], the U^3 inverse theorem [Sam07, GT08] and the bound in Roth's theorem [San11b]. We now deduce the quasipolynomial Bogloyubov-Ruzsa lemma of Sanders from Corollary 4.2. In Section 6.1 we give an algorithmic version of the proof, which allows us to explicitly find a basis for $V^{\perp}$.

Theorem 5.1 (Quasipolynomial Bogloyubov-Ruzsa Lemma) *Let $A \subseteq \mathbb{F}_2^n$ be a subset of density α. Then there exists a subspace V of $\mathbb{F}_2^n$ satisfying $V \subseteq 4A$ and*

$$\operatorname{codim}(V) = O(\log^4(\alpha^{-1})).$$

Proof of Theorem 5.1: Applying Corollary 4.2 with $B = 2A$, $S = A$, $\ell = \log(30^2/\alpha)/2$, $\eta = 1/60$, $\varepsilon = 1/(120\ell)$ and $t = O(\log^3(1/\alpha))$, we conclude the existence of a subspace V of $\operatorname{codim}(V) = O(\log^4(1/\alpha))$ which has the property that for all $v \in V$,

$$\mathop{\mathbb{P}}_{a\in A}\left[\rho_{A\to 2A}(a) \approx_{\varepsilon'} \rho_{A\to 2A}(a+v)\right] \geq 1 - 16\frac{\ell}{\eta}\frac{|3A|}{|A|}\cdot \exp\left(-2\varepsilon^2 t\right),$$

where $\varepsilon' = 4\varepsilon\ell + 2\eta + 2^{-\ell}\sqrt{|2A|/|A|} \leq 1/30 + 1/30 + (\sqrt{\alpha}/30)\cdot\sqrt{1/\alpha} \leq 1/10$.

Since $\rho_{A\to 2A}(a) = 1$ for all $a \in A$, this implies that

$$\mathop{\mathbb{P}}_{a\in A}\left[\rho_{A\to 2A}(a+v) \geq 0.9\right] \geq 1 - 16\frac{\ell}{\eta}\frac{|3A|}{|A|}\cdot \exp\left(-2\varepsilon^2 t\right) \geq 0.9,$$

where the last inequality is a result of our choice of parameters.

Recalling the definition of $\rho_{A\to B}$ in (1), the inequality above implies that for all $v \in V$,

$$\mathop{\mathbb{P}}_{a,a'\in A}[a + a' + v \in 2A] \geq 0.9^2 = 0.81.$$

By averaging, there therefore exists a pair $a, a' \in A$ such that $\mathbb{P}_{v\in V}[a + a' + v \in 2A] \geq 0.81$, or equivalently $|V \cap (a + a' + 2A)\ | \geq 0.81|V|$. But it is easy to see that if $|V \cap B| > \frac{1}{2}|V|$ for some subset $B \subseteq \mathbb{F}_2^n$, then $V \subseteq 2B$ (since every element $v \in V$ has precisely $|V|$ different representations as $v = v_1 + v_2$ where $v_1, v_2 \in V$). We conclude that $V \subseteq 2(a + a' + A + A) \subseteq 4A$, which finishes the proof. ■

5.2 Sumsets of dense sets contain large subspaces

Inspired by the question of whether dense subsets of $\{1, \ldots, N\}$ contain long arithmetic progressions, which has received extensive coverage in the literature [Bou90, Gre02, San08], Ben Green asked an analogous question in the finite field setting and obtained the following result [Gre05].

Theorem 5.2 (Green's theorem on subspaces in sumsets) *Let $A \subseteq \mathbb{F}_2^n$ be a subset of density α. Then $A + A$ contains a subspace V of $\mathbb{F}_2^n$ of dimension*

$$\dim(V) = \Omega(\alpha^2 n).$$

In [San11a] Sanders showed, using a Fourier-based density-increment strategy, that one can in fact take the subspace V to have dimension $\dim(V) = \Omega(\alpha n)$. Croot, Łaba and Sisask [CŁS11] remark that a bound of the form $\Omega\big(\frac{\alpha}{\log^3(1/\alpha)} n\big)$ could be obtained via a finite-field analogue of their methods. Our main theorem in this section is the following, replicating the result from [CŁS11].

Theorem 5.3 (Sumsets of dense sets contain large subspaces) *Let $A \subseteq \mathbb{F}_2^n$ be a subset of density α. Then $A + A$ contains an affine subspace V of $\mathbb{F}_2^n$ of dimension*

$$\dim(V) = \Omega\left(\frac{\alpha}{\log^3(1/\alpha)} n\right).$$

For the proof of the above theorem we shall need a refined version of the almost-periodicity results from Section 4. In particular, we shall need the following refined version of Corollary 4.2.

Corollary 5.4 (Refined almost-periodicity of sumsets over a subspace) *If $A \subset \mathbb{F}_2^n$ is a subset of density α, then for every integer t and set $B \subseteq \mathbb{F}_2^n$, there exists a subspace V of codimension $\operatorname{codim}(V) \leq 32\log(2/\alpha^t)$ with the following property.*

For every $v \in V$, for all subsets $S \subseteq \mathbb{F}_2^n$ and for every $\varepsilon, \eta > 0$ and integer ℓ,

$$\mathop{\mathbb{P}}_{y \in S}\left[\rho_{A \to B}(y) - \rho_{A \to B}(y + v) \leq \varepsilon'\right] \geq 1 - 16\frac{\ell}{\eta}\frac{|A + B|}{|S|} \cdot \exp\left(-\varepsilon^2 t/4\right), \tag{20}$$

where $\varepsilon' = 4\varepsilon\ell\sqrt{\rho_{A \to B}(y)} + 2\eta + 2^{-\ell}\sqrt{|B|/|A|}$.

The main difference between the above corollary and Corollary 4.2 lies in the term $\sqrt{\rho_{A \to B}(y)}$ which appears in the expression for ε' in the above corollary. This term makes ε' smaller which in turn makes the above corollary stronger. For the sake of simplicity, we only consider in the above corollary one-sided bounds of the form $\rho_{A \to B}(y) - \rho_{A \to B}(y + v) \leq \varepsilon'$ instead of two-sided bounds of the form $\rho_{A \to B}(y) \approx_{\varepsilon'} \rho_{A \to B}(y + v)$. This will suffice for the proof of Theorem 5.3.

The proof of Corollary 5.4 is similar to the proof of Corollary 4.2, and the main difference is that in the proof of Corollary 5.4 we perform a more detailed analysis of the distribution $\mathbb{1}_B(a + y)$ when a is distributed uniformly over A and y is a fixed point in $\mathbb{F}_2^n$ and use information on the *variance* of this distribution. More specifically, in the proof Corollary 5.4, instead of using the regular Hoeffding bound for sampling (Lemma 3.1), we use the following well-known refinement involving the variance [TV06].

Lemma 5.5 (Refined Hoeffding bound for sampling) *If* $\mathbf{X}$ *is a random variable satisfying* $|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \leq \mathbf{1}$ *and* $\hat{\mu}$ *is the empirical average obtained from* t *samples, then*

$$\mathbb{P}[|\mathbb{E}[\mathbf{X}] - \hat{\mu}| \ > \ \gamma] \ \leq \ 2\exp\left(-\frac{\gamma^2 t}{4\sigma^2(X)}\right)$$

provided that $\gamma < 2\sigma^2$.

For completeness, we include the proof of Corollary 5.4 in full in Appendix A. The rest of this section is devoted to the proof of Theorem 5.3 assuming that Corollary 5.4 is true.

The idea of the proof of Theorem 5.3 is as follows. Applying Corollary 5.4 with $B = A$ implies the existence of a relatively large subspace V such that for every $v \in V$, for a large fraction of $y \in \mathbb{F}_2^n$, it holds that $\rho_{A\to A}(y+v) > 0$. Our goal will be to show that an affine shift of V is contained in $2A$, or equivalently to show the existence of an affine shift $y \in \mathbb{F}_2^n$ such that $\rho_{A\to A}(y+v) > 0$ for all $v \in V$. Suppose that we have chosen the parameters in Corollary 5.4 in such a way that for every $v \in V$, at least a $(1-\delta)$-fraction of $y \in \mathbb{F}_2^n$ satisfy that $\rho_{A\to A}(y+v) > 0$. Then the union bound implies that at least a $(1 - |V|\delta)$-fraction of $y \in \mathbb{F}_2^n$ satisfy the condition $\rho_{A\to A}(y+v) > 0$ for all $v \in V$. Thus in order to guarantee the existence of the desired affine shift y, it suffices to choose the parameters in Corollary 5.4 in such a way that $|V|\delta < 1$.

Note that we wouldn't have gained anything from considering the variance in the proof of the quasipolynomial Bogolyubov-Ruzsa lemma (Theorem 5.1) since there Corollary 4.2 is applied to elements y for which $\rho_{A\to B}(y)$ is very large (between 0.9 and 1), and we have no better handle on the variance. In contrast, here the typical element to which we apply Corollary 5.4 satisfies $\rho_{A\to A}(y) = \alpha$, so that the variance is small as well.

For the proof of Theorem 5.3 we shall need the following simple lemma.

Lemma 5.6 *Let* $f(t) = t^2 - bt - c$ *for* $b > 0$, $c \geq 0$, *and suppose that* $0 \leq t' \leq t''$ *are such that* $f(t') > 0$. *Then* $f(t') \leq f(t'')$.

Proof: The fact that $b > 0$, $c \geq 0$ implies that $f(t)$ has a root $t_1 \leq 0$ and another root $t_2 > 0$. Thus we have that $f(t)$ is negative in the interval $(0, t_2)$ and is positive in the interval (t_2, ∞). The fact that $t' \geq 0$ and $f(t') > 0$ thus implies that $t' > t_2$. The lemma follows by noting that f is monotonically increasing in the interval (t_2, ∞). ■

Proof of Theorem 5.3: Apply Corollary 5.4 with $B = A$, $\eta = \alpha/24$, $\ell = \log(12/\alpha)$, $\varepsilon = \sqrt{2\alpha}/(48\ell)$, t to be determined later on and

$$S = \{y \in \mathbb{F}_2^n \mid \rho_{A\to A}(y) \geq \alpha/2\}.$$

Noting that

$$\mathbb{E}_{y\in\mathbb{F}_2^n}[\rho_{A\to A}(y)] = \mathop{\mathbb{P}}_{y\in\mathbb{F}_2^n, a\in A}[a + y \in A] = \mathbb{E}_{a\in A}\left[\mathop{\mathbb{P}}_{y\in\mathbb{F}_2^n}[a+y\in A]\right] = \alpha,$$

Markov's inequality implies that $|S| \geq (\alpha/2)\cdot 2^n$.

With this choice of parameters Corollary 5.4 implies that for every $v \in V$,

$$\mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to A}(y+v) \geq \rho_{A\to A}(y) - \alpha/6 - \frac{\sqrt{2\alpha\cdot\rho_{A\to A}(y)}}{12}\right] \geq 1 - 16\frac{\ell}{\eta}\cdot\frac{|2A|}{|S|}\cdot\exp(-\varepsilon^2 t/4)$$

Let $\delta := 16(\ell/\eta) \cdot (|2A|/|S|) \cdot \exp(-\varepsilon^2 t/4)$. Since $\rho_{A\to A}(y) \geq \alpha/2$ for every $y \in S$, the inequality above implies that

$$\begin{aligned}
\mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to A}(y+v) \geq \alpha/4\right] &= \mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to A}(y+v) \geq \alpha/2 - \alpha/6 - \frac{\sqrt{2\alpha\cdot\alpha/2}}{12}\right] \\
&\geq \mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to A}(y+v) \geq \rho_{A\to A}(y) - \alpha/6 - \frac{\sqrt{2\alpha\cdot\rho_{A\to A}(y)}}{12}\right] \\
&\geq 1-\delta
\end{aligned}$$

where the first inequality follows by applying Lemma 5.6 with $f(t) = t^2 - \left(\sqrt{2\alpha}/12\right)t - \alpha/6$, $t' = \sqrt{\alpha/2}$, $t'' = \sqrt{\rho_{A\to A}(y)}$, and noting that our assumptions imply that $0 \leq t' \leq t''$ and that $f(t') = \alpha/4 > 0$.

A union bound then implies that

$$\mathop{\mathbb{P}}_{y\in S}\left[\rho_{A\to A}(y+v) \geq \alpha/4 \;\forall v \in V\right] \geq 1 - |V|\cdot\delta.$$

To conclude the proof we shall show that for sufficiently small integer t one can guarantee that $|V|\delta < 1$. This in turn will imply the existence of an affine shift $y \in S$ such that $\rho_{A\to A}(y+v) > 0$ for every $v \in V$, and consequently $y + V \subseteq 2A$. Our choice of parameters implies that

$$\begin{aligned}
|V|\delta &= \left(\frac{\alpha^t}{2}\right)^{32} \cdot 2^n \cdot 16 \cdot \frac{\ell}{\eta} \cdot \frac{|2A|}{|S|} \cdot \exp(-\varepsilon^2 t/4) \\
&\leq \exp\left(-t\left(32\log(1/\alpha) + \frac{2\alpha}{4\cdot 48^2\cdot\log^2(12/\alpha)}\right) + \left(2\log(1/\alpha) + n + \log\log(12/\alpha)\right)\right)
\end{aligned}$$

Thus, $|V|\delta < 1$ is guaranteed by letting

$$t = \frac{2\log(1/\alpha) + n + \log\log(12/\alpha)}{32\log(1/\alpha) + \frac{2\alpha}{4\cdot 48^2\cdot\log^2(12/\alpha)}} = \frac{n + O(\log(1/\alpha))}{32\log(1/\alpha) + \Omega(\alpha/\log^2(1/\alpha))}.$$

But for such a choice of t we have that

$$\dim(V) = n - 32\log(1/\alpha)t - 32 = \Omega\left(\frac{\alpha}{\log^3(1/\alpha)}n\right).$$

■

6 Algorithmic applications

6.1 Algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma

Here we develop a robust algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma. In other words, we give an efficient (probabilistic) algorithm for finding a basis for the orthogonal complement of the subspace V in Theorem 5.1.

Theorem 6.1 (Algorithmic Quasipolynomial Bogolyubov-Ruzsa Lemma) *There exists a randomized algorithm* `Quasipolynomial-Bogolyubov` *with input parameters $\gamma' \geq \alpha > 0$ which,* *given oracle access to a function $h : \mathbb{F}_2^n \to \{0,1\}$ with $\mathbb{E}h \geq \alpha$, outputs a subspace $V \subseteq \mathbb{F}_2^n$ of codimension at most $O(\log^4(1/\alpha))$ (by giving a basis for $V^\perp$) such that with probability at least $1-\gamma'$, we have $h * h * h * h(v) > 0$ for each $v \in V$. The algorithm runs in time $2^{O(\log^4(1/\alpha))} \cdot \mathrm{polylog}(1/\gamma') \cdot n^3 \log^2 n$.*

Note that if the function h equals the indicator function of a subset $A \subseteq \mathbb{F}_2^n$, then the condition $h * h * h * h(v) > 0$ implies that $v \in 4A$, and if this condition is satisfied for all $v \in V$, then $V \subseteq 4A$. While it will be convenient to think of the set $A = \{x \in \mathbb{F}_2^n \mid h(x) = 1\}$ in the proof, in the applications described in Sections 6.2 and 6.3 we will actually apply the theorem to the output of a randomized algorithm, and hence we shall need a robust version that works for any function $h : \mathbb{F}_2^n \to \{0,1\}$ for which $\mathbb{E}h \geq \alpha$. We also assume for convenience that $\mathbb{E}h$ is *exactly* α.

In the combinatorial proof we considered the measure $\rho_{A\to 2A}(y) = \mathbb{P}_{a\in A}[y + a \in 2A]$, and the subspace V was defined in terms of a set X which was described using this measure. However, now this measure is difficult to compute since it might not be possible to test membership in $2A$ simply using oracle access to h, which is the indicator function of A. We give a robust version of the combinatorial proof by noting that $y + a \in 2A$ is equivalent to saying that $h * h(a + y) > 0$. But since we do not have noise-free access to $h * h$, we cannot test this function directly. Instead, we test if $h * h(a + y) \geq \zeta\alpha^2$ for some $\zeta > 0$. For this purpose, we define the set

$$Z_\zeta \;:=\; \left\{x \in \mathbb{F}_2^n \mid h * h(x) \geq \zeta \cdot \alpha^2\right\} .$$

The following procedure tests membership in Z_ζ by estimating $h * h$ using few samples.

```
Z-Test $(x)$

    - Estimate $h * h(x) = \mathbb{E}_{y\in\mathbb{F}_2^n} h(y) \cdot h(x - y)$ using $r$ samples of elements $y \in \mathbb{F}_2^n$.

    - Answer 1 if the estimate is at least $\zeta\alpha^2$ and 0 otherwise.
```

However, since we are *estimating* the value of $h * h$, we only have the following kind of guarantee.

Claim 6.2 *Given $\gamma_1 > 0$, the output of* Z-Test(x) *with $r = O((1/(\zeta^2\alpha^4)) \cdot \log(1/\gamma_1))$ queries satisfies the following guarantee with probability at least $1 - \gamma_1$.*

- Z-Test$(x) = 1 \implies x \in Z_{\zeta/2}$.
- Z-Test$(x) = 0 \implies x \notin Z_{3\zeta/2}$.

Proof: This follows immediately from the Hoeffding bound (Lemma 3.1). ■

Let Z denote the (random) set containing all elements for which Z-Test$(x) = 1$, that is $Z = \{x \in \mathbb{F}_2^n \mid \texttt{Z-Test}(x) = 1\}$. Then the measure

$$\rho(y) \;:=\; \rho_{A\to Z}(y) \;=\; \mathop{\mathbb{P}}_{a\in A}[a + y \in Z] \;=\; \frac{1}{\alpha} \cdot h * \mathbb{1}_Z(y) .$$

can be efficiently estimated by sampling.

The main ingredient in the proof of Theorem 6.1 is an algorithmic version of Corollary 4.1, that is, a procedure to test for membership in the set X which satisfies the iterated almost-periodicity condition in Corollary 4.1. We will present such an algorithmic version for the special case in which $B = Z$ and $S = \mathbb{F}_2^n$.

Lemma 6.3 (Iterated almost-periodicity of sumsets, algorithmic version) *Let $\gamma_2 > 0$ and let h be the indicator function of a subset $A \subseteq \mathbb{F}_2^n$ of density α. Then for any integers t, ℓ and $\varepsilon > 0$ there exists a randomized procedure* `X-Test` *with outputs in $\{0,1\}$ which makes at most $O\big((1/\alpha)^{O(t)} \cdot \exp(O(\varepsilon^2 t)) \cdot \log^4(\ell/\gamma_2) \cdot (1/\varepsilon^2)\big)$ calls to* `Z-Test` *and has the following properties.*

- *With probability at least $1 - \gamma_2$,*
  $\mathbb{P}_{x \in \mathbb{F}_2^n}[\texttt{X-Test}(x) = 1] \;\geq\; \alpha^{2t}/4.$
- *For all $x_1, \ldots, x_\ell \in \mathbb{F}_2^n$, we have with probability at least $1 - \gamma_2$,*

$$\forall i \in [\ell] \; \texttt{X-Test}(x_i) = 1 \quad \Longrightarrow$$

$$\mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n}\left[\rho_{A \to Z}(y) \approx_{4\varepsilon\ell} \rho_{A \to Z}(y + x_1 + \ldots + x_\ell)\right] \geq 1 - 4\ell \exp\big(-\Omega(\varepsilon^2 t)\big)\,.$$

For the proof we proceed as follows. First, as in the proof of Proposition 2.2, define $\mathbf{G}[\varepsilon, \delta]$ to be the set of sequences $\mathbf{a} \in (\mathbb{F}_2^n)^t$ which can be used to estimate ρ well (for our new definition of ρ). For $\mathbf{a} = (a_1, \ldots, a_t) \in A^t$, define

$$\begin{aligned} \hat{\rho}_{\mathbf{a}}(y) &:= \frac{|\{y + a_i \in Z \mid i = 1, \ldots, t\}|}{t}. \\ \mathbf{G}[\varepsilon, \delta] &:= \left\{ \mathbf{a} \in A^t \mid \mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n}\left[\rho(y) \approx_\varepsilon \hat{\rho}_{\mathbf{a}}(y)\right] \geq 1 - \delta \right\}. \end{aligned}$$

Note that in the definition of $\mathbf{G}[\varepsilon, \delta]$ above, the probability is taken over all elements $y \in \mathbb{F}_2^n$, and not only over the elements $y \in A + Z$ as in the proof of proposition 2.2 (cf., (6)). The reason is that it will be easier for us to test membership in $\mathbf{G}[\varepsilon, \delta]$ when the probability is taken over all elements $y \in \mathbb{F}_2^n$. As above, we will only be able to test membership in $\mathbf{G}[\varepsilon, \delta]$ approximately, using the following randomized procedure.

```
G-Test ($\mathbf{a} = (a_1, \ldots, a_t)$)
    - Check if $h(a_1) = \ldots = h(a_t) = 1$. If not output 0.
    - Pick $r$ independent samples $y_1, \ldots, y_r \in \mathbb{F}_2^n$.
    - For each $y_i$, estimate $\rho(y_i)$ using $r'$ independent samples. Also compute $\hat{\rho}_{\mathbf{a}}(y_i)$ for
      each $y_i$.
    - If $|\{y_i \mid |\rho(y_i) - \hat{\rho}_{\mathbf{a}}(y_i)| \geq \varepsilon\}| > \delta r$ then output 0, else output 1.
```

We prove the following guarantee for the above test.

Claim 6.4 *Given $\gamma_3 > 0$, the output of* `G-Test`$(\mathbf{a})$ *with $r = O((1/\delta^2) \cdot \log(1/\gamma_3))$ and $r' = O((1/\varepsilon^2) \cdot \log(r/\gamma_3))$ queries, satisfies the following guarantee with probability at least $1 - \gamma_3$.*

- `G-Test`$(\mathbf{a}) = 1 \implies \mathbf{a} \in \mathbf{G}[2\varepsilon, 2\delta]$.
- `G-Test`$(\mathbf{a}) = 0 \implies \mathbf{a} \notin \mathbf{G}[\varepsilon/2, \delta/2]$.

Proof: Again, this is a direct consequence of the Hoeffding bound (Lemma 3.1). ■

Note that the definition of the procedure `G-Test` actually depends on the parameters ε, δ and the error parameter γ_3, for choosing the appropriate values of r and r'. However, we choose to hide this dependence for the sake of readability. From now on let $G(\mathbf{a})$ denote the output of `G-Test` on the input $\mathbf{a}$.

The following claim is the key step in the proof of Lemma 6.3. It gives an efficient method for constructing the set X which was found in the combinatorial proof in a non-constructive manner using the pigeonhole principle. In order to construct the set X we will now show that a random sequence $\hat{\mathbf{a}} \in (\mathbb{F}_2^n)^t$ will satisfy with high probability that $G(\hat{\mathbf{a}}) = 1$ and $G(\hat{\mathbf{a}} + x) = 1$ for large number of elements $x \in \mathbb{F}_2^n$. This will enable us to define the set X as the set of all elements $x \in \mathbb{F}_2^n$ such that $G(\hat{\mathbf{a}} + x) = 1$. Membership in X can be then easily tested using the procedure `G-Test` above.

Claim 6.5 *Given $\gamma_4 > 0$, there exists an algorithm which makes $O((1/\alpha^{6t}) \cdot \log^2(1/\gamma_4))$ calls to* `G-Test` *with error parameter $\gamma_3 < 0.04$ and $\delta = \exp(-\Omega(\varepsilon^2 t))$, and finds an $\hat{\mathbf{a}} \in (\mathbb{F}_2^n)^t$ such that with probability at least $1 - \gamma_4$, we have $G(\hat{\mathbf{a}}) = 1$ and $\mathbb{E}_{x\in\mathbb{F}_2^n}[G(\hat{\mathbf{a}} + x)] \geq \alpha^{2t}/4$.*

Proof: As in the proof of Proposition 2.2, the Hoeffding bound gives that at least $0.99\left|A^t\right|$ sequences $\mathbf{a} \in A^t$ satisfy that $\mathbb{P}_{y\in A+Z}\left[\rho(y) \approx_{\varepsilon/2} \hat{\rho}_{\mathbf{a}}(y)\right] \geq 1-\delta/2$ for $\delta = \exp(-\Omega(\varepsilon^2 t))$. Noting that $\rho(y) = \hat{\rho}_{\mathbf{a}}(y) = 0$ for every $y \notin A + Z$, this implies in turn that $\mathbb{P}_{y\in\mathbb{F}_2^n}\left[\rho(y) \approx_{\varepsilon/2} \hat{\rho}_{\mathbf{a}}(y)\right] \geq 1 - \delta/2$. Consequently, $|\mathbf{G}[\varepsilon/2, \delta/2]| \geq 0.99\left|A^t\right|$ for $\delta = \exp(-\Omega(\varepsilon^2 t))$. Since A has density α in $\mathbb{F}_2^n$ we have for $\gamma_3 < 0.04$ that $\mathbb{E}_{\mathbf{a}\in(\mathbb{F}_2^n)^t}[G(\mathbf{a})] \geq (1 - \gamma_3) \cdot (0.99\alpha^t) \geq 0.95\alpha^t$. Using convexity gives

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{a}\in(\mathbb{F}_2^n)^t, x\in\mathbb{F}_2^n}[G(\mathbf{a}) \cdot G(\mathbf{a} + x)] &= \mathop{\mathbb{E}}_{\mathbf{a}\in(\mathbb{F}_2^n)^t, x, x'\in\mathbb{F}_2^n}\left[G(\mathbf{a} + x) \cdot G(\mathbf{a} + x')\right] \\
&= \mathop{\mathbb{E}}_{\mathbf{a}\in(\mathbb{F}_2^n)^t}\left[\left(\mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}[G(\mathbf{a} + x)]\right)^2\right] \\
&\geq \left(\mathop{\mathbb{E}}_{\mathbf{a}\in(\mathbb{F}_2^n)^t, x\in\mathbb{F}_2^n}[G(\mathbf{a} + x)]\right)^2 \\
&\geq (0.95 \cdot \alpha^t)^2 \;\geq\; 0.9 \cdot \alpha^{2t}\,.
\end{aligned}$$

Hence, by Markov's inequality

$$\mathop{\mathbb{P}}_{\mathbf{a}\in(\mathbb{F}_2^n)^t}\left[G(\mathbf{a}) \cdot \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}[G(\mathbf{a} + x)] \geq \alpha^{2t}/2\right] \geq \alpha^{2t}/4\,.$$

The algorithm then simply tries random sequences $\mathbf{a}$ until it finds one for which $G(\mathbf{a}) = 1$. For such an $\mathbf{a}$, it estimates $\mathbb{E}_{x\in\mathbb{F}_2^n}[G(\mathbf{a} + x)]$ using $O((1/\alpha^{4t}) \cdot \log(1/\gamma_4))$ samples. Using the Hoeffding bound once more, we have that with probability at least $1-\gamma_4/2$, the estimate is accurate to within an additive $\alpha^{2t}/8$. The algorithm stops and outputs an $\hat{\mathbf{a}}$ for which $G(\hat{\mathbf{a}}) = 1$ and the estimate computed by the algorithm is at least $3\alpha^{2t}/8$. By the above, it finds such an $\hat{\mathbf{a}}$ with probability at least $1-\gamma_4/2$ in at most $O((1/\alpha^{2t}) \cdot \log(1/\gamma_4))$ attempts, and $\mathbb{E}_{x\in\mathbb{F}_2^n}[G(\hat{\mathbf{a}} + x)] \geq \alpha^{2t}/4$ for such an $\hat{\mathbf{a}}$. If not, it simply outputs a random $\hat{\mathbf{a}}$. ■

We are now ready for the proof of Lemma 6.3.

Proof of Lemma 6.3: We apply Claim 6.5 with parameters to be specified later on to find $\hat{\mathbf{a}}$, and define X to be the set

$$X \;:=\; \{x \in \mathbb{F}_2^n \mid G(\hat{\mathbf{a}} + x) = 1\}\;.$$

Note that $G(\hat{\mathbf{a}}) = 1$ and $|X| \geq (\alpha^{2t}/4) \cdot 2^n$ with probability at least $1 - \gamma_4$. Also, membership in X can be tested efficiently. We simply define the procedure `X-Test` as

$$\texttt{X-Test}(x) = \texttt{G-Test}(\hat{\mathbf{a}} + x),$$

where `G-Test` is applied with input parameters ε and δ.

We now prove that this set X suffices for our purposes. We will prove using induction that for $x_1, \ldots, x_\ell$ satisfying $\texttt{X-Test}(x_i) = 1 \ \forall i \in [\ell]$, we have with probability at least $1 - \gamma_2$ that

$$\mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n} \left[\rho(y) \approx_{4\varepsilon\ell} \rho(y + x_1 + \ldots + x_\ell)\right] \geq 1 - 4\ell \exp\left(-\Omega(\varepsilon^2 t)\right).$$

By Claims 6.4 and 6.5 we have that $\hat{\mathbf{a}}, \hat{\mathbf{a}} + x_1, \ldots, \ldots, \hat{\mathbf{a}} + x_\ell \in \mathbf{G}[2\varepsilon, 2\delta]$ with probability at least $1 - (\ell+1)\gamma_3 - \gamma_4$. We will prove, by induction on r, that whenever $\hat{\mathbf{a}}, x_1, \ldots, x_\ell$ satisfy this condition, then for all $r = 1, \ldots, \ell$ we have

$$\mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n} \left[\rho(y) \approx_{4\varepsilon r} \rho(y + x_1 + \ldots + x_r)\right] \geq 1 - 4r\delta$$

for $\delta = \exp(-\Omega(\varepsilon^2 t))$.

For the base case, note that the fact that $\hat{\mathbf{a}} \in G[2\varepsilon, 2\delta]$ implies that $\rho(y) \approx_{2\varepsilon} \hat{\rho}_{\hat{\mathbf{a}}}(y)$ for all but a (2δ)-fraction of $y \in \mathbb{F}_2^n$. Similarly, since $\hat{\mathbf{a}} + x_1 \in G[2\varepsilon, 2\delta]$, we have that $\rho(y + x_1) \approx_{2\varepsilon} \hat{\rho}_{\hat{\mathbf{a}}+x_1}(y + x_1)$ for all but a (2δ)-fraction of $y \in \mathbb{F}_2^n$. This implies that for all but a (4δ)-fraction of $y \in \mathbb{F}_2^n$ both $\rho(y) \approx_{2\varepsilon} \hat{\rho}_{\hat{\mathbf{a}}}(y)$ and $\rho(y + x_1) \approx_{2\varepsilon} \hat{\rho}_{\hat{\mathbf{a}}+x_1}(y + x_1)$ hold. For such y we conclude that $\rho(y) \approx_{4\varepsilon} \rho(y + x_1)$ using the triangle inequality and the fact that $\hat{\rho}_{\hat{\mathbf{a}}}(y) = \hat{\rho}_{\hat{\mathbf{a}}+x_1}(y + x_1)$.

Next assume by induction that for $x_1, \ldots, x_r$ we have that $\rho(y) \approx_{4\varepsilon r} \rho(y + x_1 + \ldots + x_r)$ for at least a $(1 - 4r\delta)$-fraction of $y \in \mathbb{F}_2^n$. The base case implies that for at least a $(1 - 4\delta)$-fraction of $y \in \mathbb{F}_2^n$, it is true that $\rho(y + x_1 + \ldots + x_r) \approx_{4\varepsilon} \rho(y + x_1 + \ldots + x_{r+1})$. Thus by a union bound we have that for at least a $(1 - 4(r + 1)\delta)$-fraction of $y \in \mathbb{F}_2^n$ we have both $\rho(y) \approx_{4\varepsilon r} \rho(y + x_1 + \ldots + x_r)$ and $\rho(y + x_1 + \ldots + x_r) \approx_{4\varepsilon} \rho(y + x_1 + \ldots + x_{r+1})$. The proof is completed by noting that by the triangle inequality, for each such y, we also have $\rho(y) \approx_{4(r+1)\varepsilon} \rho(y + x_1 + \ldots + x_{r+1})$.

To get the required bounds we choose $\gamma_3 = \gamma_2/(2(\ell + 1))$ and $\gamma_4 = \gamma_2/2$. For this choice of parameters, the procedure in Claim 6.5 makes $O((1/\alpha^{6t}) \cdot \log^2(1/\gamma_2))$ calls to `G-Test`. Also, the procedure `G-Test` makes $O\big(\log^2(\ell/\gamma_2) \cdot \exp(O(\varepsilon^2 t)) \cdot (1/\varepsilon)^2\big)$ calls to `Z-Test`. This gives a total number of $O\big((1/\alpha)^{O(t)} \cdot \exp(O(\varepsilon^2 t)) \cdot \log^4(\ell/\gamma_2) \cdot (1/\varepsilon^2)\big)$ queries to `Z-Test`. ■

Next we prove the following algorithmic analogue of Corollary 4.2 which replaces the set X of almost-periods with a subspace V.

Lemma 6.6 (Almost-periodicity of sumsets over a subspace, algorithmic version) *Let $\gamma' > 0$ and let h be the indicator function of a subset $A \subseteq \mathbb{F}_2^n$ of density α. Then for every integers t, ℓ and for every $\eta, \varepsilon > 0$ there exists a randomized algorithm* `V-Test` *which outputs a subspace $V \subseteq \mathbb{F}_2^n$ of codimension at most $O(\log(1/\alpha^{2t}))$ (by giving a basis for $V^\perp$) such that with probability at least $1 - \gamma'$, we have that for all $v \in V$,*

$$\mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n} \left[\rho_{A\to Z}(y) \approx_{\varepsilon'} \rho_{A\to Z}(y + v)\right] \geq 0.99 - 24\frac{\ell}{\gamma'\eta} \exp\left(-\Omega(\varepsilon^2 t)\right).$$

where $\varepsilon' = 8\varepsilon\ell + 2\eta + 2^{-\ell}\sqrt{|Z|/|A|}$. The algorithm makes at most $n^2 \log n \cdot \mathrm{poly}((1/\alpha)^t \cdot log(1/\gamma'))$ queries to `X-Test` *with error parameter $\gamma_2 < \gamma'\eta/600$.*

As was the case in the combinatorial proof, we define V to be the subspace orthogonal to the large Fourier coefficients of $\mathbb{1}_X$, where X is the set of elements for which the procedure `X-Test` from Lemma 6.3 outputs 1. For this we need a procedure which determines the large Fourier coefficients of $\mathbb{1}_X$ with reasonable accuracy. This procedure is given by the Goldreich-Levin theorem below.

Theorem 6.7 (Goldreich-Levin Theorem, [GL89]) *Let $\nu, \delta > 0$. There is a randomized algorithm which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, runs in time $O(n^2 \log n \cdot \mathrm{poly}(1/\nu, \log(1/\delta)))$ and outputs a decomposition*

$$f = \sum_{i=1}^{k} c_i \cdot (-1)^{\langle \alpha_i, x\rangle} + g$$

with the following guarantee.

- $k = O(1/\nu^2)$.
- $\mathbb{P}\left[\exists i \ |c_i - \hat{f}(\alpha_i)| > \nu/2\right] \leq \delta$.
- $\mathbb{P}\left[\forall \alpha \text{ such that } |\hat{f}(\alpha)| \geq \nu, \quad \exists i \ \alpha_i = \alpha\right] \geq 1 - \delta$.

For the proof of Lemma 6.6 we shall also need the following algorithmic analogue of Lemma 4.3.

Lemma 6.8 *Let $\varepsilon, \delta > 0$, let `X-Test` be a randomized algorithm with outputs in $\{0, 1\}$ and let X denote the set of all elements in $\mathbb{F}_2^n$ for which `X-Test` outputs 1. Suppose that `X-Test` satisfies that for all $x_1, \ldots, x_\ell \in X$, with probability at least $1 - \gamma_2$ it holds that*

$$\underset{y \in \mathbb{F}_2^n}{\mathbb{P}} \left[\rho_{A \to Z}(y) \approx_\varepsilon \rho_{A\to Z}(y + x_1 + \ldots + x_\ell)\right] \geq 1 - \delta\,.$$

Then for every $\eta, \gamma > 0$ we have that with probability at least $1 - \gamma$,

$$\underset{y \in \mathbb{F}_2^n}{\mathbb{P}} \left[\rho_{A\to Z}(y) \approx_{\varepsilon+\eta} \underset{x_1,\ldots,x_\ell \in X}{\mathbb{E}} \left[\rho_{A\to Z}(y + x_1 + \ldots + x_\ell)\right]\right] \geq 1 - (\delta + \gamma_2)/(\gamma\eta)\,.$$

Proof: From Markov's inequality it follows that with probability at least $1 - \gamma$, for at least a $(1 - (\delta + \gamma_2)/(\gamma\eta))$-fraction of $y \in \mathbb{F}_2^n$, the relation

$$\rho_{A\to Z}(y) \approx_\varepsilon \rho_{A\to Z}(y + x_1 + \ldots + x_\ell)$$

holds for at least a $(1-\eta)$-fraction of ℓ-tuples $(x_1, \ldots, x_\ell) \in X^\ell$. Thus for at least a $(1-(\delta+\gamma_2)/(\gamma\eta))$-fraction of $y \in \mathbb{F}_2^n$, we have that

$$\left|\underset{x_1,\ldots,x_\ell \in X}{\mathbb{E}} \left[\rho_{A\to Z}(y + x_1 + \ldots + x_\ell)\right] - \rho_{A\to Z}(y)\right| \leq \underset{x_1,\ldots,x_\ell \in X}{\mathbb{E}} \left[\left|\rho_{A\to Z}(y + x_1 + \ldots + x_\ell) - \rho_{A\to Z}(y)\right|\right]$$

which is seen to be bounded above by $(1 - \eta) \cdot \varepsilon + \eta \cdot 1 \leq \varepsilon + \eta$. ■

Proof of Lemma 6.6: Let X be the set of elements for which the procedure X-Test from Lemma 6.3 outputs 1. Define the subspace V_0 as

$$V_0 \;:=\; \left(\mathrm{Spec}_{1/2}(X)\right)^{\perp} = \left\{\xi \in \mathbb{F}_2^n \mid \left|\widehat{\mathbb{1}_X}(\xi)\right| \geq \widehat{\mathbb{1}_X}(0)/2\right\}^{\perp}.$$

To find (an approximation to) V_0, we first estimate $\widehat{\mathbb{1}_X}(0) = \mathbb{E}\left[\mathbb{1}_X\right] = \mathbb{P}_{x\in\mathbb{F}_2^n}\left[\texttt{X-Test}(x) = 1\right]$ using $O((1/\alpha^{4t})\cdot\log(1/\gamma'))$ samples so that with probability at least $1-\gamma'/6$, the error is at most $\alpha^{2t}/8$. By Lemma 6.3, with probability at least $1-\gamma_2$, the quantity $\mathbb{P}_{x\in\mathbb{F}_2^n}\left[\texttt{X-Test}(x) = 1\right]$ is at least $\alpha^{2t}/4$. Taking $\gamma_2 < \gamma'/6$, we get that with probability at least $1-\gamma'/3$, the estimate is at least $\alpha^{2t}/8$. Call this estimate μ_0.

Next, the Goldreich-Levin theorem (Theorem 6.7) enables us to determine the large Fourier coefficients of $\mathbb{1}_X$ with reasonable accuracy. We run Theorem 6.7 with error parameter $\delta = \gamma'/6$ and an oracle access to the procedure X-Test, to find all characters with Fourier coefficients larger than $\nu = \mu_0$ in absolute value, up to an additive accuracy of $\nu/2 = \mu_0/2$. Let K be the list of characters given by the algorithm. We take

$$V = \left\{\xi \in \mathbb{F}_2^n \mid \xi \in K\right\}^{\perp}.$$

Now with probability at least $1-2\gamma'/3$, the trivial coefficient $\widehat{\mathbb{1}_X}(0)$ is at least $\alpha^{2t}/4$, K contains all ξ such that $|\widehat{\mathbb{1}_X}(\xi)| \geq \widehat{\mathbb{1}_X}(0)/2$ and $|\widehat{\mathbb{1}_X}(\xi)| \geq \widehat{\mathbb{1}_X}(0)/4$ for all $\xi \in K$, so $\mathrm{Spec}_{1/2}(X) \subseteq K \subseteq \mathrm{Spec}_{1/32}(X)$. By Chang's theorem (Theorem 3.2) and our choice of parameters, the codimension of V is then at most $O(\log(1/\alpha^{2t}))$.

Let $\delta := 4\ell\exp\left(-\Omega(\varepsilon^2 t)\right)$. From Lemma 6.3 and Lemma 6.8 we have that with probability at least $1-\gamma'/3$,

$$\rho_{A\to Z}(y) \approx_{4\varepsilon\ell+\eta} \mathbb{E}_{x_1,\ldots,x_\ell\in X}\left[\rho_{A\to Z}(y+x_1+\ldots+x_\ell)\right] \tag{21}$$

for at least a $(1-3(\gamma_2+\delta)/(\gamma'\eta))$-fraction of $y\in\mathbb{F}_2^n$, and also that for all $v\in V$,

$$\rho_{A\to Z}(y+v) \approx_{4\varepsilon\ell+\eta} \mathbb{E}_{x_1,\ldots,x_\ell\in X}\left[\rho_{A\to Z}(y+v+x_1+\ldots+x_\ell)\right] \tag{22}$$

for at least a $(1-3(\gamma_2+\delta)/(\gamma'\eta))$-fraction of $y\in\mathbb{F}_2^n$. Moreover, Lemma 4.4 implies that for every $y\in\mathbb{F}_2^n$ and $v\in V$,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell\in X}\left[\rho_{A\to Z}(y+x_1+\ldots+x_\ell)\right] \approx_{2^{-\ell}\sqrt{|Z/A|}} \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell\in X}\left[\rho_{A\to Z}(y+v+x_1+\ldots+x_\ell)\right]. \tag{23}$$

Applying the union bound and the triangle inequality to (21), (22) and (23), we conclude that

$$\rho_{A\to Z}(y) \approx_{\varepsilon'} \rho_{A\to Z}(y+v)$$

for $\varepsilon' = 8\varepsilon\ell + 2\eta + 2^{-\ell}\sqrt{|Z|/|A|}$ for at least a $(1-6(\gamma_2+\delta)/(\gamma'\eta))$-fraction of $y\in\mathbb{F}_2^n$, which by our choice of $\gamma_2 < \gamma'\eta/600$ gives the desired conclusion.

The running time is dominated by the $O(n^2\log n\cdot\mathrm{poly}((1/\alpha)^t, \log(1/\gamma')))$ calls made by the Goldreich-Levin algorithm to the procedure X-Test.

■

We now proceed to the proof of our main Theorem 6.1.

Proof of Theorem 6.1: Applying Lemma 6.6 with $\ell = \log(60^2/\alpha)/2$, $\varepsilon = 1/(480\ell)$, $\eta = 1/120$ and $t = O(\log^3(1/\alpha) + \log(1/\gamma')) = O(\log^3(1/\alpha))$ (provided that $\gamma' \geq \alpha$), where the basic procedure `Z-Test` is run with $\zeta = 1/300$ and $\gamma_1 = \min\left\{3\zeta\alpha^2/2, \gamma' \cdot \left(n^2 \log n \cdot \text{poly}\log(1/\gamma') \cdot 2^{O(\log^4(1/\alpha))}\right)^{-1}\right\}$, we obtain with probability at least $1-\gamma'$ a subspace V of $\text{codim}(V) \leq O(\log^4(1/\alpha))$ which has the property that for all $v \in V$,

$$\mathop{\mathbb{P}}_{y \in \mathbb{F}_2^n}\left[\rho_{A\to Z}(y) \approx_{1/20} \rho_{A\to Z}(y+v)\right] \geq 1 - 0.1 \cdot \alpha.$$

Since A has density α in $\mathbb{F}_2^n$ this implies in turn that for all $v \in V$,

$$\mathop{\mathbb{P}}_{a \in A}\left[\rho_{A\to Z}(a) \approx_{1/20} \rho_{A\to Z}(a+v)\right] \geq 0.9.$$

Next we show that, with high probability, the value of $\rho_{A\to Z}(a)$ is close to one for most elements $a \in A$.

$$\begin{aligned}
\mathop{\mathbb{E}}_{a\in A}\left[\rho_{A\to Z}(a)\right] &= \mathbb{E}_{a,a'\in A}[\mathbb{1}_Z(a+a')] \\
&= \frac{1}{\alpha^2} \cdot \langle h*h, \mathbb{1}_Z\rangle \\
&= \frac{1}{\alpha^2} \cdot \langle h*h, 1\rangle - \frac{1}{\alpha^2} \cdot \langle h*h, (1-\mathbb{1}_Z)\rangle \\
&= 1 - \frac{1}{\alpha^2} \cdot \langle h*h, (1-\mathbb{1}_Z)\rangle
\end{aligned}$$

From Claim 6.2, with probability at least $1-\gamma_1$, $h*h$ is at most $3\zeta\alpha^2/2$ when $1-\mathbb{1}_Z = 1$, and consequently the inner product in the second term is at most $3\zeta\alpha^2/2 + \gamma_1$. This gives $\mathbb{E}_{a\in A}\left[\rho_{A\to Z}(a)\right] \geq 1 - 3\zeta/2 - (\gamma_1/\alpha^2)$. Hence we have that $\rho_{A\to Z}(a) \geq 1 - \sqrt{3\zeta/2 + (\gamma_1/\alpha^2)}$ for at least a $(1 - \sqrt{3\zeta/2 + (\gamma_1/\alpha^2)})$-fraction of $a \in A$, which by the choice of $\gamma_1 \leq 3\zeta\alpha^2/2$ and $\zeta = 1/300$ implies that for all $v \in V$,

$$\mathop{\mathbb{P}}_{a\in A}\left[\rho_{A\to Z}(a+v) \geq 0.9\right] \geq 0.9.$$

Recalling the definition of $\rho_{A\to Z}$, this inequality implies that for all $v \in V$,

$$\mathop{\mathbb{P}}_{a,a'\in A}[a+a'+v \in Z] \geq 0.9^2 = 0.81.$$

By averaging, there therefore exists a pair $a, a' \in A$ such that $\mathbb{P}_{v\in V}[a+a'+v \in Z] \geq 0.81$, or equivalently $|V \cap (a+a'+Z)\ | \geq 0.81|V|$. But it is easy to see that if $|V \cap B| > \frac{1}{2}|V|$ for some subset $B \subseteq \mathbb{F}_2^n$, then $V \subseteq 2B$ (since every element $v \in V$ has precisely $|V|$ different representations as $v = v_1 + v_2$ where $v_1, v_2 \in V$). We conclude that $V \subseteq 2(a+a'+Z+Z) = 2Z$. This implies in turn that $h*h*h*h(v) > 0$ for all $v \in V$, provided that γ_1 is sufficiently small such that with probability at least $1-\gamma'$, the procedure `Z-Test` does not err on any of the elements queried during the execution of the algorithm.

It remains to analyze the running time. The procedure `V-Test` makes $O(n^2 \log n \cdot \text{poly}((1/\alpha)^t, \log(1/\gamma')))$ calls to the procedure `X-Test` with error parameter $\gamma_2 < \gamma'\eta/600$

for $\eta = 1/120$. The procedure `X-Test` makes in turn $O\big((1/\alpha)^{O(t)} \cdot \exp(O(\varepsilon^2 t)) \cdot \log^4(\ell/\gamma_2) \cdot (1/\varepsilon^2)\big)$ queries to `Z-Test` with $\zeta = 1/300$ and error parameter $\gamma_1 = \min\left\{3\zeta\alpha^2/2, \gamma' \cdot \left(n^2 \log n \cdot \operatorname{poly}\log(1/\gamma') \cdot 2^{O(\log^4(1/\alpha))}\right)^{-1}\right\}$. This choice of γ_1 guarantees that with probability at least $1-\gamma'$, the procedure `Z-Test` does not err on any of the elements queried the execution of the algorithm.

Assuming the query to `Z-Test` can be answered in constant time and it takes $O(n)$ time to write down the input, the running time for `Z-Test` is $O((1/\alpha^4)\log(1/\gamma_1) \cdot n)$. For our choice of parameters, this implies a total running time of $2^{O(\log^4(1/\alpha))} \cdot \operatorname{polylog}(1/\gamma') \cdot n^3 \log^2 n$.

■

6.2 An improved self-corrector for the Reed-Muller code of order 2

A key component in the quadratic Goldreich-Levin algorithm of [TW11] was the following self-correction procedure for the Reed-Muller code of order 2 (whose codewords are simply truth tables of quadratic phase functions). It is essentially an algorithmic version of the U^3 inverse theorem of [Sam07, GT08] and it states, qualitatively speaking, that a function with large U^3 norm correlates with a *quadratic phase function*, by which we mean a function of the form $(-1)^q$ for a quadratic form $q : \mathbb{F}_2^n \to \mathbb{F}_2$. The U^3 norm was defined for by Gowers for the purpose of counting arithmetic progressions of length 4, and is defined for $g : \mathbb{F}_2^n \to \mathbb{C}$ by the formula

$$\|g\|_{U^3}^8 = \mathbb{E}_{x,h_1,h_2,h_3 \in G} \prod_{\omega \in \{0,1\}^3} C^{|\omega|} g(x + \omega \cdot h),$$

where $\omega \cdot h$ is shorthand for $\sum_i \omega_i h_i$, and $C^{|\omega|} g = g$ if $\sum_i \omega_i$ is even and $\overline{g}$ otherwise.

Theorem 6.9 *Given $\varepsilon, \delta > 0$, there exists $\eta = \exp(-1/\varepsilon^C)$ and a randomized algorithm* `Find-Quadratic` *running in time $O(n^4 \log n \cdot \operatorname{poly}(1/\varepsilon, 1/\eta, \log(1/\delta)))$ which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, either outputs a quadratic form $q(x)$ or $\perp$. The algorithm satisfies the following guarantee.*

- *If $\|f\|_{U^3} \geq \varepsilon$, then with probability at least $1-\delta$ it finds a quadratic form q such that $\langle f, (-1)^q \rangle \geq \eta$.*
- *The probability that the algorithm outputs a quadratic form q with $\langle f, (-1)^q \rangle \leq \eta/2$ is at most δ.*

It can be easily verified that if $\langle f, (-1)^q \rangle \geq \varepsilon$ for some quadratic form q then $\|f\|_{U^3} \geq \varepsilon$. Consequently, the above theorem implies a self-corrector for the Reed-Muller code of order 2 which given a function $f : \mathbb{F}_2^n \to \{-1, 1\}$ which satisfies $\langle f, (-1)^q \rangle \geq \varepsilon$ for some quadratic form q, finds a quadratic form q' which satisfies $\langle f, (-1)^{q'} \rangle \geq \eta$ for $\eta(\varepsilon) = \exp(-1/\varepsilon^C)$.

The proof of Theorem 6.9 follows that of the inverse theorem very closely, except that many of the results from additive combinatorics that are used in the process need to be replaced by new "sampling versions": since the subsets of $\mathbb{F}_2^n$ that appear in the proof are generally very dense, it is too expensive to even write them down (let alone perform operations on them) if one is aiming for an algorithm that runs in time polynomial in n.

A crucial ingredient in the proof of Theorem 6.9 was an algorithmic version of the Bogolyubov-Ruzsa lemma (Lemma 5.3 in [TW11]), which reads as follows.

Lemma 6.10 (Algorithmic Bogolyubov-Ruzsa Lemma) *There exists a randomized algorithm* `Bogolyubov` *with input parameters α and δ which, given oracle access to a function $h : \mathbb{F}_2^n \to \{0,1\}$ with $\mathbb{E}h \geq \alpha$, outputs a subspace $V \subseteq \mathbb{F}_2^n$ of codimension at most $O(\alpha^{-3})$ (by giving a basis for $V^{\perp}$) such that with probability at least $1-\delta$, we have $h * h * h * h(v) > 0$ for all $v \in V$. The algorithm runs in time $n^2 \log n \cdot \mathrm{poly}(1/\alpha, \log(1/\delta))$.*

Replacing the above algorithm with our new algorithmic version of the quasipolynomial Bogolyubov-Ruzsa lemma (Theorem 6.1) improves the dependency of η on ε in Theorem 6.9 to be $\eta = \exp(-\mathrm{polylog}(1/\varepsilon))$. This in turn reduces the dependence on ε in the number of terms in a quadratic decomposition of a function, as well as in the running time of the quadratic decomposition algorithm, to quasipolynomial, as explained in the next section.

6.3 An improved quadratic Goldreich-Levin theorem

Both in number theory and theoretical computer science, there are certain situations where we may wish to decompose a bounded function $f : \mathbb{F}_2^n \to \mathbb{C}$ as a sum $g + h$, where g is a "uniform" or "random-looking", and h is a somewhat "structured" part. Such situations include the counting of arithmetic progressions [Gre07], the analysis of Probabilistically Checkable Proofs (PCPs) [ST06] and the approximation of matrices and tensors [FK99].

In the case where one is looking for "linear uniformity" in the function g, for example when counting arithmetic progressions of length 3, such a decomposition is achieved by separating large and small Fourier coefficients (corresponding to "linearly structured" and "linearly uniform" parts, respectively). This task can be handled algorithmically by the Goldreich-Levin theorem (Theorem 6.7), which provides an algorithm that computes, with high probability, the large Fourier coefficients of $f : \mathbb{F}_2^n \to \{-1, 1\}$ in time polynomial in n.

However, these linear decompositions have been shown to not be sensitive enough to handle many other situations, such as the counting of arithmetic progressions of length 4. In the latter case, one instead needs the function g to be "quadratically uniform" in the sense of Gowers [Gow98]: we say that a function g is *quadratically uniform* if it is small in the U^3 norm.

A hint as to what might constitute the *quadratically structured* part of a decomposition in which g is quadratically uniform is given by the so-called inverse theorem for the U^3 norm, whose proof was largely contained in Gowers's proof of Szemerédi's theorem but brought to the point by Samorodnitsky [Sam07] (in the case of characteristic 2) and by Green and Tao [GT08].

The inverse theorem implies that the structured part h has quadratic structure in the case where g is small in U^3, and starting with [Gre07] a variety of such *quadratic decomposition theorems* have come into existence: in one formulation [GW12], one can write f as

$$f = \sum_i c_i(-1)^{q_i} + g + l, \tag{24}$$

where the q_i are quadratic forms, the c_i are real coefficients such that $\sum_i |c_i|$ is bounded, $\|g\|_{U^3}$ is small and l is a small ℓ_1 error (which is negligible in all known applications). Such a decomposition is not unique and clearly non-trivial since the quadratic phases $(-1)^q$, unlike linear exponentials, do not form an orthonormal basis. In analogy with the decomposition into Fourier characters, it is natural to think of the coefficients c_i as the *quadratic Fourier coefficients* of f.

An algorithmic version of a quadratic decomposition theorem was given by the last two authors in [TW11]. Prior to [TW11], all quadratic decomposition theorems proved had been of a rather abstract nature. In particular, the work by Trevisan, Vadhan and the third author [TTV09] used linear programming techniques and boosting, while Gowers and the last author [GW12] gave a (non-constructive) existence proof using the Hahn-Banach theorem. The main result of [TW11] then was the following.

Theorem 6.11 (Quadratic Goldreich-Levin theorem) *Let $\varepsilon, \delta > 0$, $n \in \mathbb{N}$ and $B > 1$. Then there exists $\eta = \exp(-(B/\varepsilon)^C)$ and a randomized algorithm running in time $O(n^4 \log n \cdot \mathrm{poly}(1/\eta, \log(1/\delta)))$ which, given any function $f : \mathbb{F}_2^n \to [-1, 1]$ as an oracle, outputs with probability at least $1-\delta$ a decomposition into quadratic phases*

$$f \;=\; c_1(-1)^{q_1} + \ldots + c_k(-1)^{q_k} + g + l$$

satisfying $k \leq 1/\eta^2$, $\|g\|_{U^3} \leq \varepsilon$, $\|l\|_1 \leq 1/(2B)$ and $|c_i| \leq \eta$ for all $i = 1, \ldots, k$.

The algorithm comprised two parts. The first was a (entirely deterministic) procedure for assembling the quadratic phases with which the function f correlates into an actual decomposition, if these quadratic phases can indeed be found.

Theorem 6.12 *Let $\varepsilon, \delta > 0$ and $B > 1$. Let A be an algorithm which, given oracle access to a function $f : X \to [-B, B]$ satisfying $\|f\|_{U^3} \geq \varepsilon$, outputs, with probability at least $1-\delta$, a quadratic phase $(-1)^q$ such that $\langle f, (-1)^q \rangle \geq \eta$ for some $\eta = \eta(\varepsilon, B)$. Then there exists an algorithm which, given any function $f : X \to [-1, 1]$, outputs with probability at least $1 - \delta/\eta^2$ a decomposition*

$$f \;=\; c_1(-1)^{q_1} + \ldots + c_k(-1)^{q_k} + g + l$$

satisfying $k \leq 1/\eta^2$, $\|g\|_{U^3} \leq \varepsilon$, $\|l\|_1 \leq 1/(2B)$ and $|c_i| \leq \eta$ for all $i = 1, \ldots, k$.
The algorithm makes at most k calls to A.

Theorem 6.12 is proved using a boosting argument, for which we refer the reader to [TW11]. The other key component in the quadratic Goldreich-Levin algorithm is the self-correction procedure for the Reed-Muller code of order 2 at distance $1/2 - \varepsilon$, described in previous section (Theorem 6.9). Inserting our improved self-correction procedure we obtain the following quasipolynomial version of the quadratic Goldreich-Levin Theorem.

Theorem 6.13 (Quasipolynomial Quadratic Goldreich-Levin) *Let $\varepsilon, \delta > 0$, $n \in \mathbb{N}$ and $B > 1$. Then there exists $\eta = \exp(-\mathrm{poly}(B, \log(1/\varepsilon)))$ and a randomized algorithm running in time $O(n^4 \log n \cdot \mathrm{poly}(1/\eta, \log(1/\delta)))$ which, given any function $f : \mathbb{F}_2^n \to [-1, 1]$ as an oracle, outputs with probability at least $1-\delta$ a decomposition into quadratic phases*

$$f \;=\; c_1(-1)^{q_1} + \ldots + c_k(-1)^{q_k} + g + l$$

satisfying $k \leq 1/\eta^2$, $\|g\|_{U^3} \leq \varepsilon$, $\|l\|_1 \leq 1/(2B)$ and $|c_i| \leq \eta$ for all $i = 1, \ldots, k$.

A further variant of Theorem 6.12 was proved in [TW11], in which the quadratic phases in the decomposition were replaced with slightly more complicated quadratic object, namely so-called *quadratic averages*, which were first introduced in [GW12] by Gowers and the last author. In this case the authors of [TW11] obtained a bound on the number of terms in the decomposition that was polynomial in ε^{-1} in time exponential in ε^{-1}, at the cost of the description size of each quadratic average being exponential in ε^{-1}. Inserting the Algorithmic Quasipolynomial Bogolyubov-Ruzsa Lemma (Theorem 6.1) in the work of Section 5 in [TW11], we obtain an algorithm which finds a decomposition into polynomially many quadratic averages in time quasipolynomial in ε^{-1}, where the description size of each average is now quasipolynomial in ε^{-1}. We leave the details to the interested reader.

References

[BDL13] Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett, *Lower bounds on vector matching codes*, STOC, ACM, 2013.

[BLR12] Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi, *An additive combinatorics approach relating rank to communication complexity*, FOCS, IEEE Computer Society, 2012, pp. 177–186.

[Bou90] Jean Bourgain, *On arithmetic progressions in sums of sets of integers*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 105–109.

[BZ11] Eli Ben-Sasson and Noga Zewi, *From affine to two-source extractors via approximate duality*, STOC (Lance Fortnow and Salil P. Vadhan, eds.), ACM, 2011, pp. 177–186.

[Can10] Pablo Candela, *On the structure of steps of three-term arithmetic progressions in a dense set of integers*, Bull. Lond. Math. Soc. **42** (2010), no. 1, 1–14.

[Cha02] Mei-Chu Chang, *A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.

[CŁS11] Ernie Croot, Izabella Łaba, and Olof Sisask, *Arithmetic progressions in sumsets and L^p-almost-periodicity*, http://arxiv.org/abs/1103.6000v1 (2011).

[CS10] Ernie Croot and Olof Sisask, *A probabilistic technique for finding almost-periods of convolutions*, Geom. Funct. Anal. **20** (2010), no. 6, 1367–1396.

[FK99] A. M. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, 175–220.

[GKZ08] Parikshit Gopalan, Adam R. Klivans, and David Zuckerman, *List-decoding reed-muller codes over small fields*, STOC (Cynthia Dwork, ed.), ACM, 2008, pp. 265–274.

[GL89] O. Goldreich and L. Levin, *A hard-core predicate for all one-way functions*, stoc, 1989, pp. 25–32.

[Gop10] Parikshit Gopalan, *A fourier-analytic approach to reed-muller decoding*, FOCS, IEEE Computer Society, 2010, pp. 685–694.

[Gow98] Timothy Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), no. 3, 529–551.

[Gre02] Ben Green, *Arithmetic progressions in sumsets*, Geom. Funct. Anal. **12** (2002), no. 3, 584–597.

[Gre05] ______, *Finite field models in additive combinatorics*, Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 1–27.

[Gre07] ______, *Montréal notes on quadratic Fourier analysis*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 69–102.

[GT08] Ben Green and Terence Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153.

[GW10a] Timothy Gowers and Julia Wolf, *Linear forms and quadratic uniformity for functions on* $\mathbb{Z}_N$, To appear, J. Anal. Math., arXiv:1002.2210 (2010).

[GW10b] ______, *The true complexity of a system of linear equations*, Proc. Lond. Math. Soc. (3) **100** (2010), no. 1, 155–176.

[GW12] ______, *Linear forms and quadratic uniformity for functions on* $\mathbb{F}_p^n$, Mathematika **57** (2012), no. 2, 215–237.

[HL11] Hamed Hatami and Shachar Lovett, *Higher-order Fourier analysis of* $\mathbb{F}_p^n$ *and the complexity of systems of linear forms*, To appear, Geom. Func. Anal. (2011).

[IMR12] Russell Impagliazzo, Cristopher Moore, and Alexander Russell, *An Entropic Proof of Chang's Inequality*, http://arxiv.org/abs/1205.0263v1 (2012).

[Lov12] Shachar Lovett, *An exposition of Sanders's quasi-polynomial Freiman-Ruzsa theorem*, Electronic Colloquium on Computational Complexity (ECCC) **19** (2012), 29.

[Ruz99] Imre Ruzsa, *An analog of Freiman's theorem in groups*, Astérisque (1999), no. 258, xv, 323–326, Structure theory of set addition.

[Sam07] Alex Samorodnitsky, *Low-degree tests at large distances*, stoc, 2007, pp. 506–515.

[San08] Tom Sanders, *Additive structures in sumsets*, Math. Proc. Cambridge Philos. Soc. **144** (2008), no. 2, 289–316.

[San10] ______, *On the Bogolyubov-Ruzsa lemma*, To appear, Anal. PDE (2010).

[San11a] ______, *Green's sumset problem at density one half*, Acta Arith. **146** (2011), no. 1, 91–101.

[San11b] ______, *On Roth's theorem on progressions*, Ann. of Math. (2) **174** (2011), no. 1, 619–636.

[San12] ______, *Lecture notes on applications of commutative harmonic analysis*, http://people.maths.ox.ac.uk/∼sanders (2012).

[ST06] Alex Samorodnitsky and Luca Trevisan, *Gowers uniformity, influence of variables, and PCPs*, STOC, 2006, pp. 11–20.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Boosting, regularity and efficiently simulating every high-entropy distribution*, Proceedings of the 24th IEEE Conference on Computational Complexity, 2009.

[TV06] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge University Press, 2006.

[TW11] Madhur Tulsiani and Julia Wolf, *Quadratic Goldreich-Levin theorems*, FOCS, 2011, pp. 619–628.

A Appendix: Proof of Corollary 5.4

In order to prove Corollary 5.4, we start with refined versions of Proposition 2.2 and Corollary 4.1, given as Proposition A.1 and Corollary A.2 below.

Proposition A.1 (Refined version of almost-periodicity of sumsets) *Let $A \subset \mathbb{F}_2^n$ be a subset satisfying $|2A| \leq K|A|$. Then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*

2. *The size of X is at least $|A|/(2K^{t-1})$.*

3. *For all $x \in X$ and for all subsets $S \subseteq \mathbb{F}_2^n$,*

$$\mathop{\mathbb{P}}_{y \in S}\left[\rho_{A \to B}(y) - \rho_{A \to B}(y+x) \leq 2\varepsilon\sqrt{\rho_{A \to B}(y)}\right] \geq 1 - 8\frac{|A+B|}{|S|} \cdot \exp\left(-\varepsilon^2 t/4\right). \tag{25}$$

Proof: The proof is very similar to the proof of Proposition 2.2 and we only point out the differences here. As in Proposition 2.2, let $\rho(y) := \rho_{A \to B}(y)$ and for a vector $\mathbf{a} = (a_1, \ldots, a_t) \in A^t$ let

$$\hat{\rho}_{\mathbf{a}}(y) = \frac{|\{y + a_i \in B \mid i = 1, \ldots, t\}|}{t}.$$

For the purpose of this proof, we say that $\mathbf{a}$ is an ε-*good estimator* for y if $\rho(y) \approx_{\varepsilon'} \hat{\rho}_{\mathbf{a}}(y)$ for $\varepsilon' = \varepsilon\sqrt{\rho(y)}$ (this is the main point in which this proof differs from the proof of Proposition 2.2). Fix $y \in \mathbb{F}_2^n$, and let Y_i be the indicator random variable for the event "$y + a_i \in B$" where a_i is chosen uniformly at random from A. Then $\hat{\rho}_{\mathbf{a}}(y) = \frac{1}{t}\sum_{i=1}^t Y_i$ is the average of t i.i.d. indicator random variables each having mean $\rho(y)$ and variance $\rho(y)(1-\rho(y)) \leq \rho(y)$, so the refined Chernoff-Hoeffding bound (Lemma 5.5) implies that for all $y \in \mathbb{F}_2^n$,

$$\mathop{\mathbb{P}}_{\mathbf{a} \in A^t}\left[\rho(y) \not\approx_{\varepsilon\sqrt{\rho(y)}} \hat{\rho}_{\mathbf{a}}(y)\right] \leq 2\exp\left(-\varepsilon^2 t/4\right). \tag{26}$$

Set $\delta := 2\exp\left(-\varepsilon^2 t/4\right)$. Similarly to the proof of Proposition 2.2, by an averaging argument we get that at least half of the sequences $\mathbf{a} \in A^t$ are ε-good estimators for all but a (2δ)-fraction of $y \in A + B$, in which case we say that $\mathbf{a}$ is $(\varepsilon, 2\delta)$-good estimator for ρ. From here we continue as in the proof of Proposition 2.2, letting $\mathbf{G}[\varepsilon, 2\delta]$ be the set of $(\varepsilon, 2\delta)$-good estimators, that is

$$\mathbf{G}[\varepsilon, 2\delta] = \left\{\mathbf{a} \in A^t \mid \mathop{\mathbb{P}}_{y \in A+B}\left[\rho(y) \approx_{\varepsilon\sqrt{\rho(y)}} \hat{\rho}_{\mathbf{a}}(y)\right] \geq 1 - 2\delta\right\},$$

and defining $G[\varepsilon, 2\delta]_b$, $\hat{\mathbf{a}}$ and X accordingly.

It can be easily verified that the first two properties listed in the statement are satisfied. Next we show that the third one is satisfied as well.

Suppose $x = \hat{a}_1 + a_1$, where a_1 is the first element of an $(\varepsilon, 2\delta)$-good estimator $\mathbf{a} = (a_1, \ldots, a_t) \in \mathbf{G}[\varepsilon, 2\delta]_{\mathbf{b}}$. Recalling that $\hat{\mathbf{a}}$ is an $(\varepsilon, 2\delta)$-good estimator, we know that for all but a $(2\delta|A+B|/|S|)$-fraction of $y \in S$,

$$\rho(y) \approx_{\varepsilon\sqrt{\rho(y)}} \hat{\rho}_{\hat{\mathbf{a}}}(y). \tag{27}$$

Similarly, we have that

$$\rho(y+x) \approx_{\varepsilon\sqrt{\rho(y+x)}} \hat{\rho}_{\mathbf{a}}(y+x) \tag{28}$$

for all but a $(2\delta|A+B|/|x+S|)$-fraction of $y \in S$. Using a union bound and the fact that $|S+x| = |S|$, for all but a $(4\delta|A+B|/|S|)$-fraction of $y \in S$ both (27) and (28) hold. For such y we conclude $\rho(y) \approx_{\varepsilon'} \rho(y+x)$ for $\varepsilon' = \varepsilon\sqrt{\rho(y)} + \varepsilon\sqrt{\rho(y+x)}$ using the triangle inequality and the fact that $\hat{\rho}_{\hat{\mathbf{a}}}(y) = \hat{\rho}_{\hat{\mathbf{a}}+x}(y+x) = \hat{\rho}_{\mathbf{a}}(y+x)$. The proof is completed by noting that $\rho(y) - \rho(y+x) \le 2\varepsilon\sqrt{\rho(y)}$ holds trivially if $\rho(y+x) \ge \rho(y)$, and hence without loss of generality we may assume that $\rho(y+x) \le \rho(y)$. This implies in turn that $\varepsilon' \le 2\varepsilon\sqrt{\rho(y)}$.

■

As before, by an inductive application of Proposition A.1 one can prove the following iterated version.

Corollary A.2 (Refined almost-periodicity of sumsets, iterated) *If $A \subset \mathbb{F}_2^n$ satisfies $|2A| \le K|A|$ then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*

2. *The size of X is at least $|A|/(2K^{t-1})$.*

3. *For all $x_1, \ldots, x_\ell \in X$ and for all subsets $S \subseteq \mathbb{F}_2^n$,*

$$\underset{y\in S}{\mathbb{P}}\left[\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_\ell) \le 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)}\right] \ge 1 - 8\ell\frac{|A+B|}{|S|}\cdot\exp\left(-\varepsilon^2 t/4\right). \tag{29}$$

Proof of Corollary A.2: Proposition A.1 establishes the case $\ell = 1$. For the induction step, suppose that the lemma holds for some integer $\ell \ge 1$ with a set X, and we shall prove that the lemma holds for $\ell+1$ with the same set X.

Let $\delta := 8(|A+B|/|S|)\cdot\exp\left(-\varepsilon^2 t/4\right)$, and fix $x_1, \ldots, x_{\ell+1} \in X$. By the induction hypothesis, for at least a $(1-\ell\delta)$-fraction of $y \in S$ it holds that

$$\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_\ell) \le 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)}. \tag{30}$$

The $\ell = 1$ case implies that for at least a $(1-\delta)$-fraction of $y \in S$ it holds that

$$\rho_{A\to B}(y+x_1+\ldots+x_\ell) - \rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) \le 2\varepsilon\sqrt{\rho_{A\to B}(y+x_1+\ldots+x_\ell)}. \tag{31}$$

Thus by union bound we have that at least a $(1-(\ell+1)\delta)$-fraction of $y \in S$ satisfy both (30) and (31). This implies in turn that for at least a $(1-(\ell+1)\delta)$-fraction of $y \in S$ it holds that

$$\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) \le 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)} + 2\varepsilon\sqrt{\rho_{A\to B}(y+x_1+\ldots+x_\ell)}. \tag{32}$$

If $\rho_{A\to B}(y+x_1+\ldots+x_\ell) \le \rho_{A\to B}(y)$, Equation (32) implies that

$$\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) \le 2\varepsilon(\ell+1)\sqrt{\rho_{A\to B}(y)}$$

and hence we are done.

Otherwise assume that $\rho_{A\to B}(y+x_1+\ldots+x_\ell) \geq \rho_{A\to B}(y)$. Without loss of generality we may also assume that $\rho_{A\to B}(y) - 2\varepsilon\sqrt{\rho_{A\to B}(y)} > 0$ since otherwise the fact that $\rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) \geq 0$ implies that

$$\begin{aligned}\rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) &\geq \rho_{A\to B}(y) - 2\varepsilon\sqrt{\rho_{A\to B}(y)} \\ &\geq \rho_{A\to B}(y) - 2\varepsilon(\ell+1)\sqrt{\rho_{A\to B}(y)}\end{aligned}$$

and hence we are done.

Equation (31) then implies that

$$\begin{aligned}\rho_{A\to B}(y+x_1+\ldots+x_{\ell+1}) &\geq \rho_{A\to B}(y+x_1+\ldots+x_\ell) - 2\varepsilon\sqrt{\rho_{A\to B}(y+x_1+\ldots+x_\ell)} \\ &\geq \rho_{A\to B}(y) - 2\varepsilon\sqrt{\rho_{A\to B}(y)} \\ &\geq \rho_{A\to B}(y) - 2\varepsilon(\ell+1)\sqrt{\rho_{A\to B}(y)},\end{aligned}$$

where the second inequality follows from Lemma 5.6 by letting $f(t) = t^2 - 2\varepsilon t$, $t' = \sqrt{\rho_{A\to B}(y)}$, $t'' = \sqrt{\rho_{A\to B}(y+x_1+\ldots+x_\ell)}$ and noting that our assumptions imply that $0 \leq t' \leq t''$ and $f(t') > 0$. ■

One final ingredient needed for the proof of Corollary 5.4 is the following refined version of Lemma 4.3.

Lemma A.3 *Let $\delta > 0$, and let $\varepsilon : (\mathbb{F}_2^n)^{\ell+1} \to [0,1]$ be an arbitrary function in $\ell+1$ variables. Let $A, B, X, S \subseteq \mathbb{F}_2^n$ be such that for all $x_1, \ldots, x_\ell \in X$,*

$$\mathbb{P}_{y\in S}\left[\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_\ell) \leq \varepsilon(y, x_1, \ldots, x_\ell)\right] \geq 1-\delta.$$

Then for every $\eta > 0$ we have

$$\mathbb{P}_{y\in S}\left[\rho_{A\to B}(y) - \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+x_1+\ldots+x_\ell)] \leq \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\varepsilon(y,x_1,\ldots,x_\ell)] + \eta\right] \geq 1-\delta/\eta.$$

Similarly, if for all $x_1, \ldots, x_\ell \in X$,

$$\mathbb{P}_{y\in S}\left[\rho_{A\to B}(y+x_1+\ldots+x_\ell) - \rho_{A\to B}(y) \leq \varepsilon(y, x_1, \ldots, x_\ell)\right] \geq 1-\delta,$$

then for every $\eta > 0$ we have

$$\mathbb{P}_{y\in S}\left[\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+x_1+\ldots+x_\ell)] - \rho_{A\to B}(y) \leq \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\varepsilon(y,x_1,\ldots,x_\ell)] + \eta\right] \geq 1-\delta/\eta.$$

Proof: We shall prove only the first part of the lemma, the second part being almost identical. It follows from Markov's inequality that for at least a $(1-\delta/\eta)$-fraction of $y \in S$, we have

$$\rho_{A\to B}(y) - \rho_{A\to B}(y+x_1+\ldots+x_\ell) \leq \varepsilon(y, x_1, \ldots, x_\ell)$$

for at least a $(1-\eta)$-fraction of ℓ-tuples $(x_1, \ldots, x_\ell) \in X^\ell$. Taking expectations, we find that for at least a $(1-\delta/\eta)$-fraction of $y \in S$,

$$\rho_{A\to B}(y) - \mathbb{E}_{x_1,\ldots,x_\ell\in X}\left[\rho_{A\to B}(y+x_1+\ldots+x_\ell)\right] \leq \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\varepsilon(y+x_1+\ldots+x_\ell)] + \eta.$$

■

We are now ready for the proof of Corollary 5.4.

Proof of Corollary 5.4: Again, let $V = \mathrm{Spec}_{1/2}(X)^{\perp}$. As before, Chang's theorem (Theorem 3.2) implies that $\mathrm{codim}(V) \le 32\log(2/\alpha^t)$.

Let $\delta := 8\ell(|A+B|/|S|)\cdot\exp\left(-\varepsilon^2 t/4\right)$. From Lemma A.2 and Lemma A.3 we have that

$$\rho_{A\to B}(y) - \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+x_1+\ldots+x_\ell)] \le 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)}+\eta \tag{33}$$

for at least a $(1-\delta/\eta)$-fraction of $y \in S$, and similarly that for all $v \in V$,

$$\begin{aligned}
& \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)] - \rho_{A\to B}(y+v) \\
\le\ & 2\varepsilon\ell\cdot\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\sqrt{\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)}]+\eta \\
\le\ & 2\varepsilon\ell\sqrt{\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)]}+\eta
\end{aligned} \tag{34}$$

for at least a $(1-\delta/\eta)$-fraction of $y \in S$, where the last inequality is due to convexity.

From Lemma 4.4 we have that for every $y \in S$ and $v \in V$ it holds that

$$\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+x_1+\ldots+x_\ell)] - \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)] \le 2^{-\ell}\sqrt{|B|/|A|}. \tag{35}$$

If $\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)] \le \rho_{A\to B}(y)$ then applying the union bound to (33), (34) and (35) we conclude that

$$\begin{aligned}
& \rho_{A\to B}(y) - \rho_{A\to B}(y+v) \\
\le\ & 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)} + 2\varepsilon\ell\sqrt{\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)]} + 2\eta + 2^{-\ell}\sqrt{|B|/|A|} \\
\le\ & 4\varepsilon\ell\sqrt{\rho_{A\to B}(y)} + 2\eta + 2^{-\ell}\sqrt{|B|/|A|}
\end{aligned}$$

for at least a $(1-2\delta/\eta)$-fraction of $y \in S$, thus arriving at the desired conclusion.

Otherwise, assume that $\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)] \ge \rho_{A\to B}(y)$. Without loss of generality we may also assume that $\rho_{A\to B}(y) - 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)} - \eta > 0$ since otherwise we have that

$$\rho_{A\to B}(y+v) \ge \rho_{A\to B}(y) - 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)} - \eta \ge \rho_{A\to B}(y) - \varepsilon'$$

and hence we are done. Inequality (34) then implies that $\rho_{A\to B}(y+v)$ is at least

$$\begin{aligned}
& \mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)] - 2\varepsilon\ell\sqrt{\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)]} - \eta \\
\ge\ & \rho_{A\to B}(y) - 2\varepsilon\ell\sqrt{\rho_{A\to B}(y)} - \eta \\
\ge\ & \rho_{A\to B}(y) - \varepsilon'
\end{aligned}$$

where the first inequality follows from Lemma 5.6 by letting $f(t) = t^2 - 2\varepsilon\ell t - \eta$, $t' = \sqrt{\rho_{A\to B}(y)}$, $t'' = \sqrt{\mathbb{E}_{x_1,\ldots,x_\ell\in X}[\rho_{A\to B}(y+v+x_1+\ldots+x_\ell)]}$ and noting that our assumptions imply that $0 \le t' \le t''$ and $f(t') > 0$. ■