Quadratic Goldreich-Levin Theorems

Madhur Tulsiani[*] Julia Wolf[†]

May 23, 2011

Abstract

Decomposition theorems in classical Fourier analysis enable us to express a bounded function in terms of few linear phases with large Fourier coefficients plus a part that is pseudorandom with respect to linear phases. The Goldreich-Levin algorithm [GL89] can be viewed as an algorithmic analogue of such a decomposition as it gives a way to efficiently find the linear phases associated with large Fourier coefficients.

In the study of "quadratic Fourier analysis", higher-degree analogues of such decompositions have been developed in which the pseudorandomness property is stronger but the structured part correspondingly weaker. For example, it has previously been shown that it is possible to express a bounded function as a sum of a few quadratic phases plus a part that is small in the U^3 norm, defined by Gowers for the purpose of counting arithmetic progressions of length 4. We give a polynomial time algorithm for computing such a decomposition.

A key part of the algorithm is a local self-correction procedure for Reed-Muller codes of order 2 (over $\mathbb{F}_2^n$) for a function at distance $1/2-\varepsilon$ from a codeword. Given a function $f : \mathbb{F}_2^n \to \{-1, 1\}$ at fractional Hamming distance $1/2 - \varepsilon$ from a quadratic phase (which is a codeword of Reed-Muller code of order 2), we give an algorithm that runs in time polynomial in n and finds a codeword at distance at most $1/2 - \eta$ for $\eta = \eta(\varepsilon)$. This is an algorithmic analogue of Samorodnitsky's result [Sam07], which gave a tester for the above problem. To our knowledge, it represents the first instance of a correction procedure for any class of codes, beyond the list-decoding radius.

In the process, we give algorithmic versions of results from additive combinatorics used in Samorodnitsky's proof and a refined version of the inverse theorem for the Gowers U^3 norm over $\mathbb{F}_2^n$.

[*] Princeton University and IAS, Princeton, NJ. Work supported by NSF grant CCF-0832797.

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

1 Introduction

Higher-order Fourier analysis, which has its roots in Gowers's proof of Szemerédi's Theorem [Gow98], has experienced a significant surge in the number of available tools as well as applications in recent years, including perhaps most notably Green and Tao's proof that there are arbitrarily long arithmetic progressions in the primes.

Across a range of mathematical disciplines, classical Fourier analysis is often applied in form of a *decomposition theorem*: one writes a bounded function f as

$$f = f_1 + f_2, \tag{1}$$

where f_1 is a structured part consisting of the frequencies with large amplitude, while f_2 consists of the remaining frequencies and resembles uniform, or random-looking, noise. Over $\mathbb{F}_2^n$, the Fourier basis consists of functions of the form $(-1)^{\langle \alpha, x\rangle}$ for $\alpha \in F_2^n$, which we shall refer to as *linear phase functions*. The part f_1 is then a (weighted) sum of a few linear phase functions.

From an algorithmic point of view, efficient techniques are available to compute the structured part f_1. The Goldreich-Levin [GL89] theorem gives an algorithm which computes, with high probability, the large Fourier coefficients of $f : \mathbb{F}_2^n \to \{-1, 1\}$ in time polynomial in n. One way of viewing this theorem is precisely as an algorithmic version of the decomposition theorem above, where f_1 is the part consisting of large Fourier coefficients of a function and f_2 is random-looking with respect to any test that can only detect large Fourier coefficients.

It was observed by Gowers (and previously by Furstenberg and Weiss in the context of ergodic theory) that the count of certain patterns is *not* almost invariant under the addition of a noise term f_2 as defined above, and thus a decomposition such as (1) is not sufficient in that context. In particular, for counting 4-term arithmetic progressions a more sensitive notion of uniformity is needed. This subtler notion of uniformity, called *quadratic uniformity*, is expressed in terms of the U^3 norm, which was introduced by Gowers in [Gow98] and which we shall define below.

In certain situations we may therefore wish to decompose the function f as above, but where the random-looking part is quadratically uniform, meaning $\|f_2\|_{U^3}$ is small. Naturally one needs to answer the question as to what replaces the *structured part*, which in (1) was defined by a small number of linear characters.

This question belongs to the realm of what is now called *quadratic Fourier analysis*. Its central building block, largely contained in Gowers's proof of Szemerédi's theorem but refined by Green and Tao [GT08] and Samorodnitsky [Sam07], is the so-called *inverse theorem for the U^3 norm*, which states, roughly 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 inverse theorem implies that the structured part f_1 has quadratic structure in the case where f_2 is small in U^3, and starting with [Gre07] a variety of such *quadratic decomposition theorems* have come into existence: in one formulation [GW10c], one can write f as

$$f = \sum_i \lambda_i (-1)^{q_i} + f_2 + h, \tag{2}$$

where the q_i are quadratic forms, the λ_i are real coefficients such that $\sum_i |\lambda_i|$ is bounded, $\|f_2\|_{U^3}$ is small and h is a small ℓ_1 error (that is negligible in all known applications.)

In analogy with the decomposition into Fourier characters, it is natural to think of the coefficients λ_i as the *quadratic Fourier coefficients* of f. As in the case of Fourier coefficients, there is a trade-off between the complexity of the structured part and the randomness of the uniform part. In the case of the quadratic decomposition above, the bound on the ℓ^1 norm of the coefficients λ_i depends inversely on the uniformity parameter $\|f_2\|_{U^3}$. However, unlike the decomposition into Fourier characters, the decomposition in terms of quadratic phases is not necessarily unique, as the quadratic phases do not form a basis for the space of functions on $\mathbb{F}_2^n$.

Quadratic decomposition theorems have found several number-theoretic applications, notably in a series of papers by Gowers and the second author [GW10c, GW10a, GW10b], as well as [Can10] and [HL11].

However, all decomposition theorems of this type proved so far have been of a rather abstract nature. In particular, work by Trevisan, Vadhan and the first author [TTV09] uses linear programming techniques and boosting, while Gowers and the second author [GW10c] gave a (non-constructive) existence proof using the Hahn-Banach theorem. The boosting proof is constructive in a very weak sense (see Section 3) but is quite far from giving an algorithm for computing the above decompositions. We give such an algorithm in this paper.

A computer science perspective. Algorithmic decomposition theorems, such as the weak regularity lemma of Frieze and Kannan [FK99] which decomposes a matrix as a small sum of cut matrices, have found numerous application in approximately solving constraint satisfaction problems. From the point of view of theoretical computer science, a very natural question to ask is if the simple description of a bounded function as a small list of quadratic phases can be computed efficiently. In this paper we give a probabilistic algorithm that performs this task, using a number of refinements of ingredients in the proof of the inverse theorem to make it more efficient, which will be detailed below.

Connections to Reed-Muller codes. A building block in proving the decomposition theorem is an algorithm for the following problem: given a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, which is at Hamming distance at most $1/2 - \varepsilon$ from an unknown quadratic phase $(-1)^q$, find (efficiently) a quadratic phase $(-1)^{q'}$ which is at distance at most $1/2 - \eta$ from f, for some $\eta = \eta(\varepsilon)$.

This naturally leads to a connection with Reed-Muller codes since for Reed-Muller codes of order 2, the codewords are precisely the (truth-tables of) quadratic phases.

Note that the list decoding radius of Reed-Muller codes of order 2 is 1/4 [GKZ08, Gop10], which means that if the distance were less than 1/4, we could find *all* such q, and there would only be poly(n) many of them. The distance here is greater than 1/4 and there might be exponentially many (in n) such functions q. However, the problem may still be tractable as we are required to find only *one* such q (which might be at a slightly larger distance than q').

The problem of *testing* if there is such a q was considered by Samorodnitsky [Sam07]. We show that in fact, the result can be turned into a *local self corrector* for Reed-Muller codes at distance $(1/2 - \varepsilon)$. We are not aware of any class of codes for which such a self-correcting procedure is known, beyond the list-decoding radius.

1.1 Overview of results and techniques

We state below the basic decomposition theorem for quadratic phases, which is obtained by combining Theorems 3.1 and 4.1 proved later. The theorem is stated in terms of the U^3 norm, defined formally in Section 2.

Theorem 1.1 *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 $O(n^4 \log n \cdot \text{poly}(1/\eta, \log(1/\delta)))$ which, given any function*

$g : X \to [-1, 1]$ as an oracle, outputs with probability at least $1 - \delta$ a decomposition into quadratic phases

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

satisfying $k \le 1/\eta^2$, $\|f\|_{U^3} \le \varepsilon$, $\|e\|_1 \le 1/2B$ and $|c_i| \le \eta$ for all i.

Note that in [GW10a] the authors had to work much harder to obtain a bound on the number of terms in the decomposition, rather than just the ℓ^1 norm of its coefficients. Our decomposition approach gives such a bound immediately and is equivalent from a quantitative point of view: we can bound the number of terms here by $1/\eta^2$, which is exponential in $1/\varepsilon$.

It is possible to further strengthen this theorem by combining the quadratic phases obtained into only $\mathrm{poly}(1/\varepsilon)$ *quadratic averages*. Roughly speaking, each quadratic average is a sum of few quadratic phases, which differ only in their linear part. We describe this in detail in Section 5.

The key component of the above decomposition theorem is the following self-correction procedure for Reed-Muller codes of order 2 (which are simply truth-tables of quadratic phase functions). The correlation between two functions f and g is defined as $\langle f, g\rangle = \mathbb{E}_{x\in\mathbb{F}_2^n}\left[f(x)g(x)\right]$.

Theorem 1.2 *Given $\varepsilon, \delta > 0$, there exists $\eta = \exp(-1/\varepsilon^C)$ and a randomized algorithm* `Find-Quadratic` *running in time $O(n^4 \log n \cdot \mathrm{poly}(1/\varepsilon, 1/\eta, \log(1/\delta)))$ 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 $\|f\|_{U^3} \ge \varepsilon$, then with probability at least $1 - \delta$ it finds a quadratic form q such that $\langle f, (-1)^q\rangle \ge \eta$.*
- *The probability that the algorithm outputs a quadratic form q with $\langle f, (-1)^q\rangle \le \eta/2$ is at most δ.*

We remark that all the results contained here can be extended to $\mathbb{F}_p^n$ for any constant p. We choose to present only the case of $\mathbb{F}_2^n$ for simplicity of notation.

Our results for computing the above decompositions comprise various components.

Constructive decomposition theorems. We prove the decomposition theorem using a procedure which, at every step, tests if a certain function has correlation at least $1/2 - \varepsilon$ with a quadratic phase. Given an algorithm to *find* such a quadratic phase, the procedure gives a way to combine them to obtain a decomposition.

Previous decomposition theorems have also used such procedures [FK99, TTV09]. However, they required that the quadratic phase found at each step have correlation $\eta = O(\varepsilon)$, if one exists with correlation ε. In particular, they require the fact that if we scale f to change its ℓ_∞ norm, the quantities η and ε would scale the same way (this would not be true if, say, $\eta = \varepsilon^2$).

We need and prove a general decomposition theorem, which works even as η degrades arbitrarily in $1/\varepsilon$. This requires a somewhat more sophisticated analysis and the introduction of a third error term for which we bound the ℓ_1 norm.

Algorithmic versions of theorems from additive combinatorics. Samorodnitsky's proof uses several results from additive combinatorics, which produce large sets in $\mathbb{F}_2^n$ with certain useful additive properties. The proof of the inverse theorem uses the description of these sets. However, in our setting, we do not have time to look at the entire set since they may be of size $\mathrm{poly}(\varepsilon) \cdot 2^n$, as in the case of the Balog-Szemerédi-Gowers theorem described later. We thus work by building efficient sampling procedures or procedures for efficiently deciding membership in such sets, which require new algorithmic proofs.

A subtlety arises when one tries to construct such a testing procedure. Since the procedure runs in polynomial time, it often works by sampling and estimating certain properties and the estimates may be erroneous. This leads to some noise in the decision of any such an algorithm, resulting a noisy version of the set (actually a distribution over sets). We get around this problem by proving a robust version of the Balog-Szemerédi-Gowers theorem, for which we can "sandwich" the output of such a procedure between two sets with desirable properties. This technique may be useful in other algorithmic applications.

Local inverse theorems and decompositions involving quadratic averages. Samorodnitsky's inverse theorem says that when a function f has U^3 norm ε, then one can find a quadratic phase q which has correlation η with f, for $\eta = \exp(-1/\varepsilon^C)$. A decomposition then requires $1/\eta^2$, that is exponentially many (in $1/\varepsilon$), terms.

A somewhat stronger result was implicit in the work of Green and Tao [GT08]. They showed that there exists a subspace of codimension $\mathrm{poly}(1/\varepsilon)$ and on all of whose cosets f correlates polynomially with a quadratic phase. Picking a particular coset and extending that quadratic phase to the whole space gives the previous theorem.

It turns out that the different quadratic phases on each coset in fact have the same quadratic part and differ only by a linear term. This was exploited in [GW10c] to obtain a decomposition involving only polynomially many quadratic objects, so-called *quadratic averages*, which are described in more detail in Section 5.

We remark that the results of Green and Tao [GT08] do not directly extend to the case of characteristic 2 since division by 2 is used at one crucial point in the argument. We combine their ideas with those of Samorodnitsky to give an algorithmic version of a decomposition theorem involving quadratic averages.

2 Preliminaries

Throughout the paper, we shall be using Latin letters such as x, y or z to denote elements of $\mathbb{F}_2^n$, while Greek letters α and β are used to denote members of the dual space $\widehat{\mathbb{F}_2^n} \cong \mathbb{F}_2^n$. We shall use δ as our error parameter, while $\varepsilon, \eta, \gamma$ and ρ are variously used to indicate correlation strength between a Boolean function f and a family of structured functions $\mathcal{Q}$. Throughout the manuscript N will denote the quantity 2^n. Constants C may change from line to line without further notice.

We shall be using the following standard probabilistic bounds without further mention.

Lemma 2.1 (Hoeffding bound for sampling [TV06]) *If $\mathbf{X}$ is a random variable with $|\mathbf{X}| \leq 1$ and $\hat{\mu}$ is the empirical average obtained from t samples, then*

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

A Hoeffding-type bound can also be obtained for polynomial functions of ±1-valued random variables.

Lemma 2.2 (Hoeffding bound for low-degree polynomials [O'D08]) *Suppose that* $\mathbf{F} = \mathbf{F}(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ *is a polynomial of degree* d *in random variables* $\mathbf{X}_1, \ldots, \mathbf{X}_N$ *taking value* ± 1*, then*

$$\mathbb{P}\left[|\mathbf{F} - \mathbb{E}[\mathbf{F}]| \; > \; \gamma\right] \; \leq \; \exp\left(-\Omega\left(d \cdot (\gamma/\sigma)^{2/d}\right)\right),$$

where $\sigma = \sqrt{\mathbb{E}[\mathbf{F}^2] - \mathbb{E}[\mathbf{F}]^2}$ *is the standard deviation of* $\mathbf{F}$*.*

We start off by stating two fundamental results in additive combinatorics which are often applied in sequence. For a set $A \subseteq \mathbb{F}_2^n$, we write $A + A$ for the set of elements $a + a'$ such that $a, a' \in A$. More generally, the k-fold *sumset*, denoted by kA, consists of all k-fold sums of elements of A.

First, the Balog-Szemerédi-Gowers theorem states that if a set has many additive quadruples, that is, elements a_1, a_2, a_3, a_4 such that $a_1 + a_2 = a_3 + a_4$, then a large subset of it must have small sumset.

Theorem 2.3 (Balog-Szemerédi-Gowers [Gow98]) *Let* $A \subseteq \mathbb{F}_2^n$ *contain at least* $|A|^3/K$ *additive quadruples. Then there exists a subset* $A' \subseteq A$ *of size* $|A'| \geq K^{-C}|A|$ *with the property that* $|A' + A'| \leq K^C|A'|$*.*

Freiman's theorem, first proved by Ruzsa in the context of $\mathbb{F}_2^n$, asserts that a set with small sumset is efficiently contained in a subspace.

Theorem 2.4 (Freiman-Ruzsa Theorem [Ruz99]) *Let* $A \subseteq \mathbb{F}_2^n$ *be such that* $|A + A| \leq K|A|$*. Then* A *is contained in a subspace of size at most* $2^{O(K^C)}|A|$*.*

We shall also require the notion of a *Freiman homomorphism.* We say the map l is a Freiman 2-homomorphism if $x + y = z + w$ implies $l(x) + l(y) = l(z) + l(w)$. More generally, a Freiman homomorphism of order k is a map l such that $x_1 + x_2 + \cdots + x_k = x'_1 + x'_2 + \cdots + x'_k$ implies that $l(x_1) + \cdots + l(x_k) = l(x'_1) + \cdots + l(x'_k)$. The order of the Freiman homomorphism measures the degree of linearity of l; in particular, a truly linear map is a Freiman homomorphism of all orders.

Next we recall the definition of the uniformity of U^k norms introduced by Gowers in [Gow98].

Definition 2.5 *Let* G *be any finite abelian group. For any positive integer* $k \geq 2$ *and any function* $f : G \to \mathbb{C}$*, define the* U^k-norm *by the formula*

$$\|f\|_{U^k}^{2^k} = \mathbb{E}_{x, h_1, \ldots, h_k \in G} \prod_{\omega \in \{0,1\}^k} C^{|\omega|} f(x + \omega \cdot h),$$

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

In the special case $k = 2$, a computation shows that

$$\|f\|_{U^2} = \|\widehat{f}\|_{l^4},$$

and hence any approach using the U^2 norm is essentially equivalent to using ordinary Fourier analysis. In the case $k = 3$, the U^3 norm counts the number of additive octuples "contained in" f, that is, we average over the product of f at all eight vertices of a 3-dimensional parallelepiped in G.

These uniformity norms satisfy a number of important properties: they are clearly nested

$$\|f\|_{U^2} \leq \|f\|_{U^3} \leq \|f\|_{U^4} \leq \dots$$

and can be defined inductively

$$\|f\|_{U^{k+1}}^{2^{k+1}} = \mathbb{E}_x \|f_x\|_{U^k}^{2^k},$$

where $k \geq 2$ and the function f_x stands for the assignment $f_x(y) = f(y)\overline{f(x+y)}$. Thinking of the function f as a complex exponential (a phase function), we can interpret the function f_x as a kind of *discrete derivative* of f.

It follows straight from a simple but admittedly ingenious sequence of applications of the Cauchy-Schwarz inequality that if the balanced function $1_A - \alpha$ of a set $A \subseteq G$ of density α has small U^k norm, then A contains the expected number of arithmetic progressions of length $k+1$, namely $\alpha^{k+1}|G|^2$. This fact makes the uniformity norms interesting for number-theoretic applications.

In computer science they have been used in the context of probabilistically checkable proofs (PCP) [ST06], communication complexity [VW07], as well as in the analysis of pseudo-random generators that fool low-degree polynomials [BV10].

In many applications, being small in the U^k norm is a desirable property for a function to have. What can we say if this is not the case? It is not too difficult to verify that $\|f\|_{U^k} = 1$ if and only if f is a polynomial phase function of degree $k-1$, i.e. a function of the form $\omega^{p(x)}$ where p is a polynomial of degree $k-1$ and ω is an appropriate root of unity. But does every function with large U^k norm look like a polynomial phase function of degree $k-1$?

It turns out that any function with large U^k norm correlates, at the very least locally, with a polynomial phase function of degree $k-1$. This is known as the inverse theorem for the U^k norm, proved by Green and Tao [GT08] for $k = 3$ and $p > 2$ and Samorodnitsky [Sam07] for $k = 3$ and $p = 2$, and Bergelson, Tao and Ziegler [BTZ10, TZ10] for $k > 3$. We shall restrict our attention to the case $k = 3$ in this paper, which we can state as follows.

Theorem 2.6 (Global Inverse Theorem for U^3 [GT08], [Sam07]) *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \varepsilon$. Then there exists a a quadratic form q and a vector b such that*

$$|\mathbb{E}_x f(x)\omega^{q(x)+b\cdot x}| \geq \exp(-O(\varepsilon^{-C}))$$

In Section 5 we shall discuss various refinements of the inverse theorem, including correlations with so-called *quadratic averages*. These refinements allow us to obtain polynomial instead of exponential correlation with some quadratically structured object.

We discuss further potential improvements and extensions of the arguments presented in this paper in Section 6.

First of all, however, we shall turn to the problem of constructively obtaining a decomposition assuming that one has an efficient correlation testing procedure, which is done in Section 3.

3 From decompositions to correlation testing

In this section we reduce from the problem of finding a decomposition for given function to the problem of finding a single quadratic phase or average that correlates well with the function.

We state the basic decomposition result in somewhat greater generality as we believe it may be of independent interest. We will consider a real-valued function g on a finite domain X (which shall be $\mathbb{F}_2^n$ in the rest of the paper). We shall decompose the function g in terms of members from an arbitrary class $\mathcal{Q}$ of functions $\overline{q} : X \to [-1, 1]$. $\mathcal{Q}$ may later be taken to be the class of quadratic phases or quadratic averages. We will assume $\mathcal{Q}$ to be closed under negation of the functions i.e., $\overline{q} \in \mathcal{Q} \Rightarrow -\overline{q} \in \mathcal{Q}$. Finally, we shall consider a semi-norm $\|\cdot\|_S$ defined for functions on X, such that if $\|f\|_S$ is large for $f : X \to \mathbb{R}$ then f has large correlation with some function in $\mathcal{Q}$. The obvious choice for $\|\cdot\|_S$ is $\|f\|_S = \max_{\overline{q} \in \mathcal{Q}} |\langle f, \overline{q} \rangle|$, as is the case in many known decomposition results and the general result in [TTV09]. However, we will be able to obtain a stronger algorithmic guarantee by taking $\|\cdot\|_S$ to be the U^3 norm.

Theorem 3.1 *Let $\mathcal{Q}$ be a class of functions as above and 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 $\|f\|_S \geq \varepsilon$, outputs, with probability at least $1 - \delta$, a function $\overline{q} \in \mathcal{Q}$ such that $\langle f, \overline{q} \rangle \geq \eta$ for some $\eta = \eta(\varepsilon, B)$. Then there exists an algorithm which, given any function $g : X \to [-1, 1]$, outputs with probability at least $1 - \delta/\eta^2$ a decomposition*

$$g \;=\; c_1\overline{q}_1 + \ldots + c_k\overline{q}_k + e + f$$

satisfying $k \leq 1/\eta^2$, $\|f\|_S \leq \varepsilon$ and $\|e\|_1 \leq 1/2B$. Also, the algorithm makes at most k calls to A.

We prove the decomposition theorem building on an argument from [TTV09], which in turn generalizes an argument of [FK99]. Both the arguments in [TTV09, FK99] work well if for a function $f : X \to R$ satisfying $\max_{\overline{q} \in \mathcal{Q}} | \langle f, \overline{q} \rangle | \geq \varepsilon$, one can efficiently find a $\overline{q} \in \mathcal{Q}$ with $\langle f, \overline{q} \rangle \geq \eta = \Omega(\varepsilon)$. It is important there that $\eta = \Omega(\varepsilon)$, or at least that the guarantee is independent of how f is scaled.

Both proofs give an algorithm which, at each step t, checks if there exists $\overline{q}_t \in \mathcal{Q}$ which has good correlation with a given function f_t, and the decomposition is obtained by adding the functions $\overline{q}_t$ obtained at different steps. In both cases, the ℓ_∞ norm of the functions f_t changes as the algorithm proceeds.

Suppose $\varepsilon' = o(\varepsilon)$ and we only had the scale-dependent guarantee that for functions $f : X \to [-1, 1]$ with $\|f\|_S \geq \varepsilon$, we can efficiently find a $\overline{q} \in \mathcal{Q}$ such that $\langle f, \overline{q} \rangle \geq \varepsilon^2$ (say). Then at step t of the algorithm if we have $\|f_t\|_\infty = M$ (say), then $\|f_t\|_S \geq \varepsilon$ will imply $\|f/M\|_S \geq \varepsilon/M$ and one can only get a $\overline{q}_t$ satisfying $\langle f_t, \overline{q}_t \rangle \geq M \cdot (\varepsilon/M)^2 = \varepsilon^2/M$. Thus, the correlation of the functions $\overline{q}_t$ we can obtain degrades as the $\|f_t\|_\infty$ increases. This turns out to be insufficient to bound the number of steps required by these algorithms and hence the number of terms in the decomposition.

When testing correlations with quadratic phases using $\|\cdot\|_S$ as the U^3 norm, the correlation η obtained for $f : \mathbb{F}_2^n \to [-1, 1]$ has very bad dependence on ε and hence we run into the above problem. To get around it, we truncate the functions f_t used by the algorithm so that we have a uniform bound on their ℓ_∞ norms. However, this truncation introduces an extra term in the decomposition, for which we bound the ℓ_1 norm. Controlling the ℓ_1 norm of this term requires a somewhat more sophisticated analysis than in [FK99]. An analysis based on a similar potential function was also employed in [TTV09] (though not for the purpose of controlling the ℓ_1 norm).

We note that a third term with bounded ℓ_1 norm also appears in the (non-constructive) decompositions obtained in [GW10a].

Proof of Theorem 3.1: We will assume all calls to the algorithm A correctly return a q as above or declare $\|f\|_S < \varepsilon$ as the case may be. The probability of any error in the calls to A is at most $k\delta$.

We build the decomposition by the following simple procedure.

```
- Define functions $f_1 = h_1 = g$. Set $t = 1$.
- While $\|f_t\|_S \geq \varepsilon$
    – Let $\overline{q}_t$ be the output of $A$ when called with the function $f_t$.
    – $h_{t+1} := h_t - \eta\overline{q}_t$.
    – $f_{t+1} := \texttt{Truncate}_{[-B,B]}\left(h_{t+1}\right) = \max\{-B, \min\{B, h_{t+1}\}\}$
    – $t := t + 1$
```

If the algorithm runs for k steps, the decomposition it outputs is

$$g = \sum_{t=1}^{k} \eta \cdot \overline{q}_t \ + \ (h_k - f_k) \ + \ f_k$$

where we take $f = f_k$ and $e = h_k - f_k$. By construction, we have that $\|f_k\|_S \leq \varepsilon$. It remains to show that $k \leq 1/\eta^2$ and $\|h_k - f_k\|_1 \leq 1/2B$.

To analyze $\|h_t - f_t\|$, we will define an additional function $\Delta_t \overset{\text{def}}{=} f_t \cdot (h_t - f_t)$. Note that $\Delta_t(x) \geq 0$ for every x, since f_t is simply a truncation of h_t and hence $f_t = B$ when $h_t > f_t$ and $-B$ when $h_t < f_t$. This gives

$$\|\Delta_t\|_1 \ = \ \mathbb{E}\left[\Delta_t\right] \ = \ \mathbb{E}\left[f_t \cdot (f_t - h_t)\right] \ = \ \mathbb{E}\left[B \cdot |h_t - f_t|\right] \ = \ B \cdot \|h_t - f_t\|_1 \, .$$

We will in fact bound the ℓ_1 norm of Δ_k to obtain the required bound on $\|h_k - f_k\|_1$. The following lemma states the bounds we need at every step.

Lemma 3.2 *For every input x and every $t \leq k - 1$*

$$f_t^2(x) - f_{t+1}^2(x) + 2\Delta_t(x) - 2\Delta_{t+1}(x) + \eta^2 \ \geq \ 2\eta \cdot \overline{q}_t(x) f_t(x).$$

We first show how the above lemma suffices to prove the theorem. Taking expectations on both sides of the inequality gives, for all $t \leq k - 1$,

$$\|f_t\|_2^2 - \|f_{t+1}\|_2^2 + 2\,\|\Delta_t\|_1 - 2\,\|\Delta_{t+1}\|_1 + \eta^2 \ \geq \ 2\eta \cdot \langle \overline{q}_t, f_t \rangle \ \geq \ 2\eta^2.$$

Summing over all $t \leq k - 1$ gives

$$\|f_1\|_2^2 - \|f_k\|_2^2 + 2\,\|\Delta_1\|_1 - 2\,\|\Delta_k\|_1 \ \geq \ k \cdot \eta^2 \ \Longrightarrow \ k \cdot \eta^2 + \|f_k\|_2^2 + 2\,\|\Delta_k\|_1 \ \leq \ 1$$

since $\|f_1\|_2^2 = \|g\|_2^2 \leq 1$ and $\Delta_1 = 0$. However, this gives $k \leq 1/\eta^2$ and $\|\Delta_k\|_1 \leq 1/2$, which in turn implies $\|h_k - f_k\|_1 \leq 1/2B$, completing the proof of Theorem 3.1. ■

We now return to the proof of Lemma 3.2.

Proof of Lemma 3.2: We shall fix an input x and consider all functions only at x. We start by bringing the RHS into the desired form and collecting terms.

$$\begin{aligned} 2\eta\overline{q}_t \cdot f_t \ &= \ 2(h_t - h_{t+1}) \cdot f_t \\ &= \ 2(h_t - f_t) \cdot f_t - 2(h_{t+1} - f_{t+1}) \cdot f_{t+1} + 2f_t^2 - 2f_{t+1}^2 - 2h_{t+1} \cdot f_t + 2h_{t+1} \cdot f_{t+1} \\ &= \ 2\Delta_t - 2\Delta_{t+1} + f_t^2 - f_{t+1}^2 + \left(f_t^2 - f_{t+1}^2 - 2h_{t+1}(f_t - f_{t+1})\right) \end{aligned}$$

It remains to show that $f_t^2 - f_{t+1}^2 - 2h_{t+1}(f_t - f_{t+1}) = (f_t - f_{t+1})(f_t + f_{t+1} - 2h_{t+1}) \leq \eta^2$. We first note that if $|f_{t+1}| < B$, then $h_{t+1} = f_{t+1}$ and the expression becomes $(f_t - f_{t+1})^2$, which is at most η^2. Also, if $|f_t| = |f_{t+1}| = B$, then f_t and f_{t+1} must be equal (as f_t only changes in steps of η) and the expression is 0.

Finally, in the case when $|f_t| < B$ and $|f_{t+1}| = B$, we must have that $|f_t - h_{t+1}| = |h_t - h_{t+1}| \leq \eta$. We can then bound the expression as

$$(f_t - f_{t+1})(f_t + f_{t+1} - 2h_{t+1}) \ \leq \ \left(\frac{(f_t - f_{t+1}) + (f_t + f_{t+1} - 2h_{t+1})}{2}\right)^2 \ = \ (f_t - h_{t+1})^2 \ \leq \ \eta^2,$$

which proves the lemma. ∎

We next show that in the case when $\|\cdot\|_S$ is the U^3 norm and $\mathcal{Q}$ contains at most $\exp(o(2^n))$ functions, it is sufficient to test the correlations only for Boolean functions $f : \mathbb{F}_2^n \to \{-1, 1\}$. This can be done by simply scaling a function taking values in $[-B, B]$ to $[-1, 1]$ and then randomly rounding the value independently at each input to ±1 with appropriate probability.

Lemma 3.3 *Let $\varepsilon, \delta > 0$. Let A be an algorithm, which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1, 1\}$ satisfying $\|f\|_{U^3} \geq \varepsilon$, outputs, with probability at least $1 - \delta$, a function $\overline{q} \in \mathcal{Q}$ such that $\langle f, \overline{q} \rangle \geq \eta$ for some $\eta = \eta(\varepsilon)$. In addition, assume that the running time of A is* $\mathrm{poly}(n, 1/\eta, \log(1/\delta))$.

Then there exists an algorithm A' which, given oracle access to a function $f : \mathbb{F}_2^n \to [-B, B]$ satisfying $\|f\|_{U^3} \geq \varepsilon$, outputs, with probability at least $1 - 2\delta$, an element $\overline{q} \in \mathcal{Q}$ satisfying $\langle f, \overline{q} \rangle \geq \eta'$ for $\eta' = \eta'(\varepsilon, B)$. Moreover, the running time of A' is $\mathrm{poly}(n, 1/\eta', \log(1/\delta))$.

Proof: Consider a random Boolean function $\tilde{f} : \mathbb{F}_2^n \to \{-1, 1\}$ such that $\tilde{f}(x)$ is 1 with probability $(1 + f(x)/B)/2$ and -1 otherwise. A' simply calls A with the function $\tilde{f}$ and parameters $\varepsilon/2B, \delta$. This means that whenever A queries the value of the function at x, A' generates it independently of all other points by looking at $f(x)$. It then outputs the $\overline{q}$ given by A.

If $\|\tilde{f}\|_{U^3} \geq \varepsilon/2B$, then A outputs a $\overline{q}$ satisfying $\langle \tilde{f}, \overline{q} \rangle \geq \eta(\varepsilon/2B)$. If for the same q we also have $\langle f, \overline{q} \rangle \geq B \cdot \eta(\varepsilon/2B)/2 = \eta'(\varepsilon, B)$, then the output of A' is as desired. However, $\|\tilde{f}\|_{U^3}$ is a polynomial of degree 8 and the correlation with any $\overline{q}$ is a linear polynomial in the 2^n random variables $\{\tilde{f}(x)\}_{x \in \mathbb{F}_2^n}$. Thus, by Lemma 2.2, the probability that $\|\tilde{f}\|_{U^3} < \|f\|_{U^3}/B - \varepsilon/2B$, or $\langle \tilde{f}, \overline{q} \rangle \geq \langle f, \overline{q} \rangle / B - \eta(\varepsilon/2B)/2$ for any $\overline{q} \in \mathcal{Q}$, is at most $\exp(-\Omega_{\varepsilon,B}(-|\mathcal{Q}| \cdot 2^n)) \leq \delta$. ∎

Thus, to compute the required decomposition into quadratic phases, one only needs to give an algorithm for finding a phase $\overline{q} = (-1)^q$ satisfying $\langle f, (-1)^q \rangle \geq \eta$ when $f : \mathbb{F}_2^n \to \{-1, 1\}$ is a *Boolean* function satisfying $\|f\|_{U^3} \geq \varepsilon$.

4 Finding correlated quadratic phases over $\mathbb{F}_2^n$

In this section, we show how to obtain an algorithm for finding a quadratic phase which has good correlation with a given function Boolean $f : \mathbb{F}_2^n \to \{-1, 1\}$ (if one exists). For an f satisfying $\|f\|_{U^3} \geq \varepsilon$, we want to find a quadratic form q such that $\langle f, (-1)^q \rangle \geq \eta(\varepsilon)$. The following theorem provides such a guarantee.

Theorem 4.1 *Given $\varepsilon, \delta > 0$, there exists $\eta = \exp(-1/\varepsilon^C)$ and a randomized algorithm* `Find-Quadratic` *running in time $O(n^4 \log n \cdot \mathrm{poly}(1/\varepsilon, 1/\eta, \log(1/\delta)))$ which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, either outputs a quadratic phase $(-1)^{q(x)}$ 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.*

The fact that $\|f\|_{U^3} \geq \varepsilon$ implies the *existence* of a quadratic phase $(-1)^q$ with $\langle f, (-1)^q \rangle \geq \eta$ was proven by Samorodnitsky [Sam07]. We give an algorithmic version of his proof, starting with the proofs of the results from additive combinatorics contained therein.

Note that $\|f\|_{U^3}^8$ is simply the expected value of the product $\prod_{\omega \in \{0,1\}^3} f(x + \omega \cdot h)$ for random $x, h_1, h_2, h_3 \in \mathbb{F}_2^n$. Hence, Lemma 2.1 implies that $\|f\|_{U^3}$ can be easily estimated by sampling sufficiently many values of x, h_1, h_2, h_3 and taking the average of the products for the samples.

Corollary 4.2 *By making $O((1/\gamma^2) \cdot \log(1/\delta))$ queries to f, one can obtain an estimate $\hat{U}$ such that*

$$\mathbb{P}\left[\left| \|f\|_{U^3} - \hat{U} \right| > \gamma \right] \leq \delta.$$

The main algorithm begins by checking if $\hat{U} \geq 3\varepsilon/4$ and rejects if this is not the case. If $\hat{U} \geq 3\varepsilon/4$, then the above claim implies that $\|f\|_{U^3} \geq \varepsilon/2$ with high probability. So our algorithm will actually return a q with correlation $\eta(\varepsilon')$ with $\varepsilon' = \varepsilon/2$. We shall ignore this and just use ε in the sequel for the sake of readability.

4.1 Picking large Fourier coefficients in derivatives

The first step of the proof in [Sam07] is to find a choice function $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ which is "somewhat linear". The choice function is used to pick a Fourier coefficient for the derivative f_y. The intuition is that if f were indeed a quadratic phase of the form $(-1)^{\langle x, Mx \rangle}$, then

$$f_y(x) = f(x)f(x+y) = (-1)^{\langle x, (M+M^T)y \rangle} \cdot (-1)^{\langle y, My \rangle} \; .$$

Thus, the largest Fourier coefficient (with absolute value 1) would be $\hat{f}_y((M+M^T)y)$. Hence, there is a function $\varphi(y) \stackrel{\text{def}}{=} (M+M^T)y$, which is given by multiplying y by a *symmetric matrix* $M + M^T$, which selects a large Fourier coefficient for f_y. The proof attempts to construct such a symmetric matrix for any f with $\|f\|_{U^3} \geq \varepsilon$.

Expanding the U^3 norm and using Hölder's inequality gives the following lemma.

Lemma 4.3 (Corollary 6.6 [Sam07]) *Suppose that $f : \mathbb{F}_2^n \to \{-1, 1\}$ is such that $\|f\|_{U^3} \geq \varepsilon$. Then*

$$\mathop{\mathbb{E}}_{x,y}\left[\sum_{\alpha,\beta} \hat{f_x}^2(\alpha) \cdot \hat{f_y}^2(\beta) \cdot \widehat{f_{x+y}}^2(\alpha+\beta) \right] \;\geq\; \varepsilon^{16}.$$

Choosing a random function $\varphi(x) = \alpha$ with probability $\hat{f_x}^2(\alpha)$ satisfies

$$\mathop{\mathbb{P}}_{x,y}\left[\varphi(x)+\varphi(y)=\varphi(x+y)\right] \;=\; \sum_{\alpha,\beta} \hat{f_x}^2(\alpha)\cdot\hat{f_y}^2(\beta)\cdot\widehat{f_{x+y}}^2(\alpha+\beta).$$

Thus, when $\|f\|_{U^3} \geq \varepsilon$, the above lemma gives that

$$\mathop{\mathbb{P}}_{\varphi,x,y}\left[\varphi(x)+\varphi(y)=\varphi(x+y)\right] \;=\; \mathop{\mathbb{E}}_{x,y}\left[\sum_{\alpha,\beta} \hat{f_x}^2(\alpha)\cdot\hat{f_y}^2(\beta)\cdot\widehat{f_{x+y}}^2(\alpha+\beta)\right] \;\geq\; \varepsilon^{16}.$$

The proof in [Sam07] works with a random function φ as described above. We define a slightly different random function φ, since we need its value at any input x to be samplable in time polynomial in n. Thus, we will only sample α for which the corresponding Fourier coefficients are sufficiently large. In particular, we need an algorithmic version of the decomposition of a function into linear phases, which follows from the Goldreich-Levin theorem.

Theorem 4.4 (Goldreich-Levin [GL89]) *Let $\gamma, \delta > 0$. There is a randomized algorithm* `Linear-Decomposition`*, which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1,1\}$, runs in time $O(n^2 \log n \cdot \mathrm{poly}(1/\gamma, \log(1/\delta)))$ and outputs a decomposition*

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

with the following guarantee:

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

Remark 4.5 *Note that the above is a slightly non-standard version of the Goldreich-Levin theorem. The usual one makes $O(n\log n\cdot\mathrm{poly}(1/\gamma, \log(1/\delta)))$ queries to f (where each query takes $O(n)$ time to write down) and guarantees that for any specific α such that $|\hat{f}(\alpha)| \geq \gamma$, there exists an i with $\alpha_i = \alpha$, with probability at least $1-\delta$. By repeating the algorithm $O(\log(1/\gamma))$ times, we can take a union bound over all α as in the last property guaranteed by the above theorem.*

It follows that in order to sample $\varphi(x)$, instead of sampling from all Fourier coefficients of f_x, we only sample from the large Fourier coefficients using the above decomposition. We shall denote the quantity $\varepsilon^{16}/4$ that appears below by ρ.

Lemma 4.6 *There exists a distribution over functions $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that $\varphi(x)$ is independently chosen for each $x \in \mathbb{F}_2^n$, and is samplable in time $O(n^3 \log n \cdot \mathrm{poly}(1/\varepsilon))$ given oracle access to f. Moreover, if $\|f\|_{U^3} \geq \varepsilon$, then we have*

$$\mathop{\mathbb{P}}_{\varphi}\left[\mathop{\mathbb{P}}_{x,y}\left[\varphi(x)+\varphi(y)=\varphi(x+y)\right] \geq \varepsilon^{16}/4\right] \geq \varepsilon^{16}/4.$$

Proof: We sample $\varphi(x)$ at each input x as follows. We run `Linear-Decomposition` for f_x with $\gamma = \delta = \varepsilon^{16}/18$ and sample $\varphi(x)$ to be α_i with probability c_i^2. If $\sum c_i^2 < 1$, we answer arbitrarily with the remaining probability. By Theorem 4.4, with probability at least $1 - 2\gamma$ over the run of `Linear-Decomposition`, each $\alpha \in \mathbb{F}_2^n$ with $|\hat{f}_x(\alpha)| \geq \gamma$ is sampled with probability at least $(\hat{f}_x(\alpha) - \gamma/2)^2 \geq \hat{f}_x{}^2(\alpha) - \gamma$. Let $[z]_0$ denote $\max\{0, z\}$. We have

$$\begin{aligned}\mathop{\mathbb{P}}_{\varphi,x,y}[\varphi(x)+\varphi(y)=\varphi(x+y)] &\geq \mathop{\mathbb{E}}_{x,y}\left[\sum_{\alpha,\beta}(1-2\gamma)^3\left[\hat{f}_x{}^2(\alpha)-\gamma\right]_0\left[\hat{f}_y{}^2(\beta)-\gamma\right]_0\left[\widehat{f_{x+y}}^2(\alpha+\beta)-\gamma\right]_0\right]\\ &\geq \varepsilon^{16} - 9\gamma,\end{aligned}$$

which by our choice of parameters is at least $\varepsilon^{16}/2$. This immediately implies that $\mathbb{P}_\varphi\left[\mathbb{P}_{x,y}[\varphi(x)+\varphi(y)=\varphi(x+y)] \geq \varepsilon^{16}/4\right] \geq \varepsilon^{16}/4$. ■

Thus, with probability $\rho = \varepsilon^{16}/4$ one gets a good φ which is somewhat linear. This φ is then used to recover an appropriate quadratic phase. We will actually delay sampling the function on all points and only query $\varphi(x)$ when needed in the construction of the quadratic phase (which we show can be done by querying φ on polynomially many points). Consequently, the construction procedures that follow will only work with a small probability, i.e. when we are actually working with a good φ. However, we can test the quadratic phase we obtain in the end and repeat the entire process if the phase does not correlate well with f. Also, note that we store the $(x, \varphi(x))$ already sampled in a data structure and re-use them if and when the same x is queried again.

4.2 Applying the Balog-Szemerédi-Gowers theorem

The next step of the proof uses φ to obtain a linear choice function Dx for some matrix D. This step uses certain results from additive combinatorics, for which we develop algorithmic versions below. In particular, it applies the Balog-Szemerédi-Gowers (BSG) theorem to the set

$$A_\varphi \overset{\text{def}}{=} \left\{(x,\varphi(x)) \;:\; |\hat{f}_x(\varphi(x))| \geq \gamma\right\},$$

where we will choose $\gamma = O(\varepsilon^{16})$ as in Lemma 4.6.

For any set $A \in \{0,1\}^n$ that is somewhat linear, the Balog-Szemerédi-Gowers theorem allows us to find a subset $A' \subseteq A$ which is large and does not grow too much when added to itself. We state the following version from [BS94], which is particularly suited to our application.

Theorem 4.7 (Balog-Szemerédi-Gowers Theorem [BS94]) *Let $A \subseteq \mathbb{F}_2^n$ be such that $\mathbb{P}_{a_1,a_2 \in A}[a_1 + a_2 \in A] \geq \rho$. Then there exists $A' \subseteq A$, $|A|' \geq \rho|A|$ such that $|A' + A'| \leq (2/\rho)^8|A|$.*

We are interested in finding the set A'_φ which results from applying the above theorem to the set A_φ. However, since the set A'_φ is of exponential size, we do not have time to write down the entire set (even if we can find it). Instead, we will need an efficient algorithm for testing membership in the set. To get the required algorithmic version, we follow the proof by Sudakov, Szemerédi and Vu [SSV05] and the presentation by Viola [Vio07].

In this proof one actually constructs a graph on the set A_φ and then selects a subset of the neighborhood of a random vertex as A'_φ, after removing certain problematic vertices. It can be deduced that the set A'_φ can be found in time polynomial in the size of the graph. However, as discussed above, this is still *exponential* in n and hence inadequate for our purposes. Below, we develop a test to check if a certain element $(x, \varphi(x))$ is in A'_φ.

We first define a (random) graph on the vertex set [1] $\{(x, \varphi(x)) \mid x \in \mathbb{F}_2^n\}$ and edge set E_γ for $\gamma > 0$, defined as

$$E_\gamma \overset{\text{def}}{=} \left\{ (x, \varphi(x)), (y, \varphi(y)) \;\middle|\; \begin{array}{c} \varphi(x) + \varphi(y) = \varphi(x+y) \\ \text{and} \\ |\hat{f}_x(\varphi(x))|, |\hat{f}_y(\varphi(y))|, |\widehat{f_{x+y}}(\varphi(x+y))| \geq \gamma \end{array} \right\}.$$

Lemma 4.6 implies that over the choice of φ, with probability at least $\rho = \varepsilon^{16}/4$, the graph defined with $\gamma = \varepsilon^{16}/18$, has density at least ρ. However, if a φ is good for a certain value of γ, then it is also good for all values $\gamma' \leq \gamma$ (as the density of the graph can only increase). For the remaining argument, we will assume that we have sampled φ completely and that it is good. We will later choose $\gamma \in [\varepsilon^{16}/180, \varepsilon^{16}/18]$.

Since we will be examining the properties of certain neighborhoods in this graph, we first write a procedure to test if two vertices in the graph have an edge between them.

```
Edge-Test (u,v,$\gamma$)
    - Let $u = (x, \varphi(x))$ and $v = (y, \varphi(y))$.
    - Estimate $|\hat{f}_x(\varphi(x))|, |\hat{f}_y(\varphi(y))|$ and $|\widehat{f_{x+y}}(\varphi(x+y))|$ using $t$ samples for each.
    - Answer 1 if $\varphi(x)+\varphi(y) = \varphi(x+y)$ and all estimates are at least $\gamma$, and 0 otherwise.
```

Unfortunately, since we are only estimating the Fourier coefficients, we will only be able to test if two vertices have an edge between them with a slight error in the threshold γ, and with high probability. Thus, if the estimate is at least γ, we can only say that with high probability, the Fourier coefficient must be at least $\gamma - \gamma'$ for a small error γ'. This leads to the following guarantee on `Edge-Test`.

Claim 4.8 *Given $\gamma', \delta > 0$, the output of* `Edge-Test` *(u, v, γ) with $t = O(1/\gamma'^2 \cdot \log(1/\delta))$ queries, satisfies the following guarantee with probability at least $1 - \delta$.*

- `Edge-Test`$(u, v, \gamma) = 1 \implies (u, v) \in E_{\gamma - \gamma'}$.
- `Edge-Test`$(u, v, \gamma) = 0 \implies (u, v) \notin E_{\gamma + \gamma'}$.

Proof: The claim follows immediately from Lemma 2.1 and the definitions of $E_{\gamma-\gamma'}$, $E_{\gamma+\gamma'}$. ■

The approximate nature of the above test introduces a subtle issue. Note that the outputs 1 and 0 of the test correspond to the presence or absence of edges in *different graphs* with edge sets $E_{\gamma-\gamma'}$ and $E_{\gamma+\gamma'}$. The edge sets of the two graphs are related as $E_{\gamma+\gamma'} \subseteq E_{\gamma-\gamma'}$. But the proof of Theorem 4.7 uses somewhat more complicated subsets of vertices, which are defined using both upper and lower bounds on the sizes of certain neighborhoods. Since the upper and lower bounds estimated using the above test will hold for slightly different graphs, we need to be careful in analyzing any algorithm that uses `Edge-Test` as a primitive.

[1] Since φ is random, the vertex set of the graph as defined is random. However, since φ is a function, the vertex set is isomorphic to $\mathbb{F}_2^n$ and one may think of the graph as being defined on a fixed set of vertices with edges chosen according to a random process.

We now return to the argument as presented in [SSV05]. It considers the neighborhood of a random vertex u and removes vertices that have too few neighbors in common with other vertices in the graph. Let the size of the vertex set be $N = 2^n$. For a vertex u, we define the following sets:

$$\begin{aligned}
N(u) &\overset{\text{def}}{=} \{v \ : \ (u,v) \in E_\gamma\} \\
S(u) &\overset{\text{def}}{=} \left\{ v \in N(u) \ : \ \mathop{\mathbb{P}}_{v_1}\left[v_1 \in N(u) \text{ and } |N(v) \cap N(v_1)| \leq \rho^3 N\right] \geq \rho^2 \right\} \\
&= \left\{ v \in N(u) \ : \ \mathop{\mathbb{P}}_{v_1}\left[v_1 \in N(u) \text{ and } \mathop{\mathbb{P}}_{v_2}\left[v_2 \in N(v) \cap N(v_1)\right] \leq \rho^3 \right] > \rho^2 \right\} \\
T(u) &\overset{\text{def}}{=} N(u) \setminus S(u) \\
&= \left\{ v \in N(u) \ : \ \mathop{\mathbb{P}}_{v_1}\left[v_1 \in N(u) \text{ and } \mathop{\mathbb{P}}_{v_2}\left[v_2 \in N(v) \cap N(v_1)\right] \leq \rho^3 \right] \leq \rho^2 \right\}
\end{aligned}$$

It is shown in [SSV05] (see also [Vio07]) that if the graph has density ρ, then picking $A'_\varphi = T(u)$ for a random vertex u is a good choice[2].

Lemma 4.9 *Let the graph with edge set E_γ have density at least ρ and let $A'_\varphi = T(u)$ for a random vertex u. Then, with probability at least $\rho/2$ over the choice of u, the set A'_φ satisfies*

$$\left|A'_\varphi\right| \geq \rho N \quad \textit{and} \quad \left|A'_\varphi + A'_\varphi\right| \leq (2/\rho)^8 N.$$

We now translate the condition for membership in the set $T(u)$ into an algorithm. Note that we perform different edge tests with different thresholds, the values of which will be chosen later.

```
BSG-Test $(u, v, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2)$          (Approximate test to check if $v \in T(u)$)

    - Let $u = (x, \varphi(x))$ and $v = (y, \varphi(y))$.
    - Sample $(z_1, \varphi(z_1)), \ldots, (z_r, \varphi(z_r))$.
    - For each $i \in [r]$, sample $(w_1^{(i)}, \varphi(w_1^{(i)})), \ldots, (w_s^{(i)}, \varphi(w_s^{(i)}))$.
    - If Edge-Test (u,v,$\gamma_1$) = 0, then output 0.
    - For $i \in [r], j \in [s]$, let
          $X_i = $ Edge-Test $((x, \varphi(x)), (z_i, \varphi(z_i)), \gamma_2)$
          $Y_{ij} = $ Edge-Test $\left((y, \varphi(y)), \left(w_j^{(i)}, \varphi\left(w_j^{(i)}\right)\right), \gamma_3\right)$
          $Z_{ij} = $ Edge-Test $\left((z_i, \varphi(z_i)), \left(w_j^{(i)}, \varphi\left(w_j^{(i)}\right)\right), \gamma_3\right)$
    - For each $i$, take $B_i = 1$ if $\frac{1}{s}\sum_j Y_{ij} \cdot Z_{ij} \leq \rho_1$ and 0 otherwise.
    - Answer 1 if $\frac{1}{r}\sum_i X_i \cdot B_i \leq \rho_2$ and 0 otherwise.
```

[2] Note that here we are choosing A'_φ to be the neighborhood of *any* vertex in the graph, instead of vertices in A_φ. However, this is not a problem since the only vertices with non-empty neighborhoods are the ones in A_φ.

Choice of parameters for `BSG-Test`: We shall choose the parameters for the above test as follows. Recall that $\rho = \varepsilon^{16}/4$. We take $\rho_1 = 21\rho^3/20$ and $\rho_2 = 19\rho^2/20$. Given an error parameter δ, we take r and s to be $\mathrm{poly}(1/\rho, \log(1/\delta))$, so that with probability at least $1-\delta$, the error in the last two estimates is at most $\rho^3/100$. Also, by using $\mathrm{poly}(1/\rho, \log(1/\delta))$ samples in each call to `Edge-Test`, we can assume that the error in all estimates used by `Edge-Test` is at most $\rho^3/100$.

To choose $\gamma_1, \gamma_2, \gamma_3$, we divide the interval $[\varepsilon^{16}/180, \varepsilon^{16}/18]$ into $4/\rho^2$ consecutive sub-intervals of size $\rho^3/20$ each. We then randomly choose a sub-interval and choose positive parameters γ, μ so that $\gamma - \mu$ and $\gamma + \mu$ are endpoints of this interval. We set $\gamma_1 = \gamma_3 = \gamma + \mu/2$ and $\gamma_2 = \gamma - \mu/2$.

To analyze `BSG-Test`, we "sandwich" the elements on which it answers 1 between a large set and a set with small doubling.

Lemma 4.10 *Let $\delta > 0$ and parameters ρ_1, ρ_2, r, s be chosen as above. Then for every $u = (x, \varphi(x))$ and every choice of $\gamma_1, \gamma_2, \gamma_3$ as above, there exist two sets $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$, such that the output of* `BSG-Test` *satisfies the following with probability at least $1-\delta$.*

- `BSG-Test`$(u, v, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2) = 1 \quad \Longrightarrow \quad v \in A_\varphi^{(2)}(u)$.
- `BSG-Test`$(u, v, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2) = 0 \quad \Longrightarrow \quad v \notin A_\varphi^{(1)}(u)$.

Moreover, with probability $\rho^3/24$ over the choice of u and $\gamma_1, \gamma_2, \gamma_3$, we have

$$|A_\varphi^{(1)}(u)| \geq (\rho/6) \cdot N \quad \text{and} \quad |A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)| \leq (2/\rho)^8 \cdot N.$$

Proof: To deal with the approximate nature of `Edge-Test`, we define the following sets:

$$N_\gamma(u) \overset{\text{def}}{=} \{v \,:\, (u,v) \in E_\gamma\}$$

$$T(u, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2) \overset{\text{def}}{=} \left\{ v \in N_{\gamma_1}(u) \,:\, \underset{v_1}{\mathbb{P}}\left[v_1 \in N_{\gamma_2}(u) \;\&\; \underset{v_2}{\mathbb{P}}\left[v_2 \in N_{\gamma_3}(v) \cap N_{\gamma_3}(v_1)\right] \leq \rho_1 \right] \leq \rho_2 \right\}$$

Going through the definitions and recalling that $E_\gamma \subseteq E_{\gamma - \gamma'}$ for $\gamma' > 0$, it can be checked that the sets $T(u, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2)$ are monotone in the various parameters. In particular, for $\gamma_1', \gamma_2', \gamma_3', \rho_1', \rho_2' > 0$

$$T(u, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2) \;\subseteq\; T(u, \gamma_1 - \gamma_1', \gamma_2 + \gamma_2', \gamma_3 - \gamma_3', \rho_1 - \rho_1', \rho_2 + \rho_2').$$

Recall that we have $\gamma_1 = \gamma_3 = \gamma + \mu/2$ and $\gamma_2 = \gamma - \mu/2$, where $[\gamma - \mu, \gamma + \mu]$ is a sub-interval of $[\varepsilon^{16}/180, \varepsilon^{16}/18]$ of length $\rho^3/20$.

We define the sets $A_\varphi^{(1)}(u)$ and $A_\varphi^{(2)}(u)$ as below.

$$A_\varphi^{(1)}(u) \;\overset{\text{def}}{=}\; T(u, \gamma + \mu, \gamma - \mu, \gamma + \mu, 11\rho^3/10, 9\rho^2/10)$$

$$A_\varphi^{(2)}(u) \;\overset{\text{def}}{=}\; T(u, \gamma, \gamma, \gamma, \rho^3, \rho^2)$$

By the monotonicity property noted above, we have that $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$. Also, by the choice of parameters r, s and the number of samples in `Edge-Test`, we know that with probability $1-\delta$, the error in all estimates used in `BSG-Test` is at most $\rho^3/100$. Hence, we get that with probability at least $1-\delta$, if `BSG-Test` answers 1, then the input is in $A_\varphi^{(2)}$ and if `BSG-Test` answers 0, then it is not in $A_\varphi^{(1)}$. It remains to prove the bounds on the size and doubling of these sets.

By our choice of parameters, $A_\varphi^{(2)}(u)$ is the same set as the one defined in Sudakov et al. [SSV05]. They show that if u is such that $|A_\varphi^{(2)}(u)| \geq 3 \cdot (\rho/2)^2 N$, then $|A_\varphi^{(2)}(u) + A_\varphi^{(2)}(u)| \leq (2/\rho)^8 \cdot N$ (see Lemma 3.2 in [Vio07] for a simplified proof of the version mentioned here). To show the lower bound on the size of $A_\varphi^{(2)}(u)$, we will show that in fact with probability at least $\rho^3/24$ over the choice of u and $\gamma_1, \gamma_2, \gamma_3$, we will have $|A_\varphi^{(1)}(u)| \geq (\rho/6) \cdot N$. Since $A_\varphi^{(1)}(u) \subseteq A_\varphi^{(2)}(u)$, this suffices for the proof.

We consider a slight modification of the argument of [SSV05], showing an upper bound on the expected size of the set $S'(u)$ defined as

$$\begin{aligned} S'(u) &\stackrel{\text{def}}{=} N_{\gamma+\mu}(u) \setminus T(u, \gamma+\mu, \gamma-\mu, \gamma+\mu, 11\rho^3/10, 9\rho^2/10) \\ &= \left\{ v \in N_{\gamma+\mu}(u) \ : \ \mathop{\mathbb{P}}_{v_1}\left[v_1 \in N_{\gamma-\mu}(u) \ \& \ \mathop{\mathbb{P}}_{v_2}\left[v_2 \in N_{\gamma+\mu}(v) \cap N_{\gamma+\mu}(v_1)\right] \leq 11\rho^3/10 \right] \geq 9\rho^2/10 \right\}. \end{aligned}$$

We know from Lemma 4.6 that since $\gamma + \mu \leq \varepsilon^{16}/18$, the quantity $\mathbb{E}_u\left[|N_{\gamma+\mu}(u)|\right]$, which is the average degree of the graph, is at least ρN (assuming that we are working with a good function φ). Combining this with an upper bound on $\mathbb{E}_u\left[|S'(u)|\right]$ will give the required lower bound on the size of $A_\varphi^{(1)}(u) = T(u, \gamma+\mu, \gamma-\mu, \gamma+\mu, 11\rho^3/10, 9\rho^2/10)$.

We call a pair (v, v_1) *bad* if $|N_{\gamma+\mu}(v) \cap N_{\gamma+\mu}(v)| \leq 11\rho^3 N/10$. We need the following bound.

Claim 4.11 *There exists a choice for the sub-interval* $[\gamma-\mu, \gamma+\mu]$ *of length* $\rho^3/20$ *in* $[\varepsilon^{16}/180, \varepsilon^{16}/18]$ *such that*

$$\mathop{\mathbb{E}}_u\left[\#\left\{\textit{bad pairs } (v, v_1) \ : \ v \in N_{\gamma+\mu}(u) \ \& \ v_1 \in N_{\gamma-\mu}(u)\right\}\right] \ \leq \ 3\rho^3 N^2/5$$

We first prove Lemma 4.10 assuming the claim. From the definition of $S'(u)$,

$$\#\{\text{bad pairs } (v, v_1) \ : \ v \in N_{\gamma+\mu}(u) \ \& \ v_1 \in N_{\gamma-\mu}(u)\} \ \geq \ |S'(u)| \cdot (9\rho^2 N/10).$$

Claim 4.11 gives $\mathbb{E}_u\left[|S'(u)|\right] \leq (3\rho^3 N^2/5)/(9\rho^2 N/10) = (2\rho/3)N$, for at least one choice of the interval $[\gamma-\mu, \gamma+\mu]$. Since there are $4/\rho^2$ choices for the sub-interval, this happens with probability at least $\rho^2/4$.

For this choice of γ and μ (and hence of $\gamma_1, \gamma_2, \gamma_3$), we also have $\mathbb{E}_u\left[|N_{\gamma+\mu}(u)|\right] \geq \rho N$. Since $S'(u) = N_{\gamma+\mu}(u) \setminus A_\varphi^{(1)}$, we get that $\mathbb{E}_u\left[|A_\varphi^{(1)}|\right] \geq \rho N - (2\rho/3)N = (\rho/3)N$. Hence, with probability at least $\rho/6$ over the choice of u, $|A_\varphi^{(1)}| \geq (\rho/6)N$. Thus, we obtain the desired outcome with probability at least $\rho^3/24$ over the choice of u and $\gamma_1, \gamma_2, \gamma_3$. ■

Proof of Claim 4.11: We begin by observing that the expected number of bad pairs (v, v_1) such that $v \in N_{\gamma+\mu}(u)$ & $v_1 \in N_{\gamma-\mu}(u)$ is equal to

$$\begin{aligned} &\mathbb{E}_u\left[\#\left\{\text{bad pairs } (v, v_1) \ : \ v \in N_{\gamma+\mu}(u) \ \& \ v_1 \in N_{\gamma+\mu}(u)\right\}\right] \\ &+ \mathbb{E}_u\left[\#\left\{\text{bad pairs } (v, v_1) \ : \ v \in N_{\gamma+\mu}(u) \ \& \ v_1 \in N_{\gamma-\mu}(u) \setminus N_{\gamma+\mu}(u)\right\}\right]. \end{aligned}$$

Note that for each of the $\binom{N}{2}$ choices for v, v_1, if they form a bad pair, then each u is in $N_{\gamma+\mu}(v) \cap N_{\gamma+\mu}(v_1)$ with probability at most $11\rho^3/10$. Hence, the first term is at most $(11\rho^3/20)N^2$. Also, the second term is at most

$$N \cdot \mathop{\mathbb{E}}_u\left[|N_{\gamma-\mu}(u) \setminus N_{\gamma+\mu}(u)|\right] \ = \ N \cdot \left(\mathop{\mathbb{E}}_u\left[|N_{\gamma-\mu}(u)|\right] - \mathop{\mathbb{E}}_u\left[|N_{\gamma+\mu}(u)|\right]\right)$$

We know that $\mathbb{E}_u\left[|N_\gamma(u)|\right]$ is monotonically decreasing in γ. Since it is at most N for $\gamma = \varepsilon^{16}/180$, there is at least one interval of size $\rho^3/20$ in $[\varepsilon^{16}/180, \varepsilon^{16}/18]$, where the change is at most $\rho^3 N/20$. Taking $\gamma+\mu$ and $\gamma-\mu$ to be the endpoints of this interval finishes the proof. ■

4.3 Obtaining a linear choice function

Using the subset given by the Balog-Szemerédi-Gowers theorem, one can use the somewhat linear choice function φ to find an *linear transformation* $x \mapsto Tx$ which also selects large Fourier coefficients in derivatives. In particular, it satisfies $\mathbb{E}_x\left[\hat{f_x}^2(Tx)\right] \geq \eta$ for some $\eta = \eta(\varepsilon)$. This map T can then be used to find an appropriate quadratic phase.

In this subsection, we give an algorithm for finding such a transformation, using the procedure `BSG-Test` developed above. In the lemma below, we assume as before that φ is a good function satisfying the guarantee in Lemma 4.6. We also assume that we have chosen a good vertex u and parameters $\gamma_1, \gamma_2, \gamma_3$ satisfying the guarantee in Lemma 4.10.

Lemma 4.12 *Let φ be as above and $\delta > 0$. Then there exists an $\eta = \exp(-1/\varepsilon^C)$ and an algorithm which makes $O(n^2 \log n \cdot \mathrm{poly}(1/\eta, \log(1/\delta)))$ calls to* `BSG-Test` *and uses additional running time $O(n^3)$ to output a linear map T or the symbol $\perp$. If* `BSG-Test` *is defined using a good u and parameters $\gamma_1, \gamma_2, \gamma_3$ as above, then with probability at least $1 - \delta$ the algorithm outputs a map T satisfying $\mathbb{E}_x\left[\hat{f_x}^2(Tx)\right] \geq \eta$.*

Proof: Let $t = 4n^2 + \log(10/\delta)$. We proceed by first sampling $K = 100t/\rho$ elements $(x, \varphi(x))$ and running `BSG-Test` $(u, \cdot)$ on each of them with parameters as in Lemma 4.10 and $\delta' = \delta/(5K)$. We retain only the points $(x, \varphi(x))$ on which `BSG-Test` outputs 1. Since $\delta' = \delta/(5K)$, `BSG-Test` does not satisfy the guarantee of Lemma 4.10 on some query with probability at most $\delta/5$. We assume this does not happen for any of the points we sampled.

If `BSG-Test` outputs 1 on fewer than t of the queries, we stop and output $\perp$. The following claim shows that the probability of this happening is at most $\delta/5$. In fact, the claim shows that with probability $1 - \delta/5$ there must be at least t samples from $A_\varphi^{(1)}$ itself, on which we assumed that `BSG-Test` outputs 1.

Claim 4.13 *With probability at least $1 - \delta/5$, the sampled points contain at least t samples from $A_\varphi^{(1)}$.*

Proof: Since $|A_\varphi^{(1)}| \geq \rho N/6$, the expected number of samples from $A_\varphi^{(1)}$ is at least $\rho K/6$. By a Hoeffding bound, the probability that this number is less than t is at most $\exp(-\Omega(\rho K)) \leq \delta/5$ if $\rho K = \Omega(\log(1/\delta))$. ■

Note that conditioned on being in $A_\varphi^{(1)}$, the sampled points are in fact *uniformly* distributed in $A_\varphi^{(1)}$. We show that then they must span a subspace of large dimension, and that their span must cover at least half of $A_\varphi^{(1)}$.

Claim 4.14 *Let $z_1, \ldots, z_t \in A_\varphi^{(1)}$ be uniformly sampled points. Then for $t \geq 4n^2 + O(\log(1/\delta))$ it is true with probability $1 - \delta/5$ that*

- $| < z_1, \ldots, z_t > \cap A_\varphi^{(1)}| \geq (1/2)|A_\varphi^{(1)}|$
- $\dim(< z_1, \ldots, z_t >) \geq n - \log(12/\rho)$.

Proof: For the first part, we consider the span $< z_1, \ldots, z_t >$, which is a subspace of $\mathbb{F}_2^n$. The probability that it has small intersection with $A_\varphi^{(1)}$ is

$$\sum_{|S \cap A_\varphi^{(1)}| \leq |A_\varphi^{(1)}|/2} \mathbb{P}\left[z_1, \ldots, z_t \in S\right] \cdot \mathbb{P}\left[< z_1, \ldots, z_t > \;=\; S \mid z_1, \ldots, z_t \in S\right],$$

where the sum is taken over all subspaces S of $\mathbb{F}_2^n$. Since $|S \cap A_\varphi^{(1)}| \leq |A_\varphi^{(1)}|/2$, we have that $\mathbb{P}\left[z_1, \ldots, z_t \in S\right] \leq (1/2)^t$. Thus, the required probability bounded above by

$$\sum_{|S \cap A_\varphi^{(1)}| \leq |A_\varphi^{(1)}|/2} (1/2)^t \cdot 1 \;\leq\; 2^{-t} O(2^{4n^2}).$$

The last bound uses the fact that the number of subspaces of $\mathbb{F}_2^{2n}$ is $O(2^{4n^2})$. Thus, for $t = 4n^2 + \log(10/\delta)$, the probability is at most $\delta/10$.

We now bound the probability that the sampled points $z_1, \ldots, z_t$ span a subspace of dimension at most $n - k$. The probability that a random of $A_\varphi^{(1)}$ lies in a *specific* subspace of dimension $n - k$ is at most $(2^{-k}/(\rho/6))$. Hence, the probability that all t points lie in any subspace of dimension $n - k$ is bounded above by

$$\left(\frac{2^{-k}}{\rho/6}\right)^t \cdot \#\{\text{subspaces of dim } n-k\} \;\leq\; \left(\frac{2^{-k}}{\rho/6}\right)^t \cdot 2^{n(n-k)}.$$

For $t \geq n^2 + O(\log(1/\delta))$ and $k = \log(12/\rho)$, this probability is at most $\delta/10$. Hence the dimension of the span of the sampled vectors is at least $n - \log(12/\rho)$ with high probability. ■

Next, we *upper bound* the dimension of the span of the retained points (on which `BSG-Test` answered 1). By the assumed correctness of `BSG-Test`, we get that all the points must lie inside $A_\varphi^{(2)}$. Applying the Freiman-Ruzsa Theorem (Theorem 2.4), it follows that

$$|< A_\varphi^{(2)} >| \;\leq\; \exp(1/\rho^C)N.$$

The above implies that all the points are inside a space of dimension at most $n + \log(1/\nu)$, where we have written $\nu = \exp(-1/\rho^C)$. From here, we can proceed in a similar fashion to [Sam07].

Let V denote the span of the retained points and let $v_1, \ldots, v_r$ be a basis for V. We can add vectors to complete it to $v_1, \ldots, v_s$ so that the projection onto the first n coordinates has full rank. Let $V' =< v_1, \ldots, v_s >$. We can also assume, by a change of basis, that for $i \leq n$ we have the coordinate vectors $v_i = (e_i, u_i)$. This can all be implemented by performing Gaussian elimination, which takes time $O(n^3)$.

Consider the $2n \times s$ matrix with $v_1, \ldots, v_s$ as columns. By the previous discussion, this matrix is of the form

$$P = \begin{pmatrix} I & 0 \\ T & U \end{pmatrix},$$

where I is the $n \times n$ identity matrix, and T and U are $n \times n$ and $n \times (s - n)$ matrices, respectively. By Claim 4.14, we know that v' contains $|A_\varphi^{(1)}|/2 \geq (\rho/12)N$ vectors of the form $(x, \varphi(x))^T$. For each such vector, there exists a $w \in \mathbb{F}_2^s$ such that $P \cdot w = (x, \varphi(x))^T$. Because of the form of P, we must have that $w = (x, z)$ for $z \in \mathbb{F}_2^{s-n}$. Thus, we get that for each vector $(x, \varphi(x))$, we in fact have $\varphi(x) = Tx + Uz$ for some $z \in \mathbb{F}_2^{s-n}$.

Therefore, for at least one $z_0 \in \mathbb{F}_2^{s-n}$ and $y_0 = Uz_0$ we find that

$$\mathop{\mathbb{P}}_{x \in \mathbb{F}_2^n} [\varphi(x) = Tx + y_0] \ \geq \ (\rho/12) \cdot 2^{-(s-n)}.$$

We next upper bound $s - n$. Note that $s \leq r + k$ since by Claim 4.14, V had dimension at least $n - k$ for $k = \log(12/\rho)$. Also, we know that $r \leq n + \log(1/\nu)$ by the bound on $| < A_\varphi^{(2)} > |$, implying that $s \leq n + \log(12/\rho) + \log(1/\nu)$. We conclude that $2^{-(s-n)} \geq (\rho/12)\nu$.

Moreover, for each element of the form $(x, \varphi(x)) \in A_\varphi^{(1)}$, we know that $|\hat{f}_x(\varphi(x))| \geq \gamma \geq \varepsilon^{16}/180$. This implies that

$$\mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} \left[\hat{f_x}^2(Tx + y_0) \right] \geq \gamma^2 \cdot (\rho/12) \cdot (\rho\nu/12).$$

Samorodnitsky shows that we can in fact take y_0 to be 0. In fact, he shows the following general claim.

Claim 4.15 (Consequence of Lemma 6.10 [Sam07]) *For any matrix T and $y \in \mathbb{F}_2^n$, $\mathbb{E}_{x \in \mathbb{F}_2^n} \left[\hat{f_x}^2(Tx + y) \right] \leq \mathbb{E}_{x \in \mathbb{F}_2^n} \left[\hat{f_x}^2(Tx) \right]$.*

Thus, we simply output the matrix T constructed as above. For $\eta = \gamma^2 \rho^2 \nu/144$, it satisfies $\mathbb{E}_{x \in \mathbb{F}_2^n} \left[\hat{f_x}^2(Tx) \right] \geq \eta$. Finally, we calculate the probability that the algorithm outputs $\perp$ or outputs a T not satisfying this guarantee. This can happen only when the guarantee on `BSG-Test` is not satisfied for one of the sampled points, or when the guarantees in Claims 4.13 and 4.14 are not satisfied. Since each of these happen with probability at most $\delta/5$, the probability of error is at most $3\delta/5 < \delta$. ■

4.4 Finding a quadratic phase function

Once we have identified the linear map T above, the remaining argument is identical to the one in [Sam07].

Equipped with T, one can find a symmetric matrix B with zero diagonal that satisfies a slightly weaker guarantee. This step is usually referred to as the *symmetry argument*, and we shall encounter a modification of it in Section 5. The only algorithmic steps used in the process are Gaussian elimination and finding a basis for a subspace, which can both be done in time $O(n^3)$.

Lemma 4.16 (Proof of Theorem 2.3 [Sam07]) *Let T be as above. Then in time $O(n^3)$ one can find a symmetric matrix B with zero diagonal such that $\mathbb{E}_{x \in \mathbb{F}_2^n} \left[\hat{f_x}^2(Bx) \right] \geq \eta^2$.*

Now that we have correlation of the derivative f_x of the function with a truly linear map, it remains to "integrate" this relationship to obtain that f itself correlates with a quadratic map. Following Green and Tao, we shall henceforth refer to this part of the argument as the *integration step.*

Having obtained B above, we can find a matrix M such that $M + M^T = B$. We take the quadratic part of the phase function to be $h(x) = (-1)^{\langle x, Mx \rangle}$. The following claim helps establish the linear part.

Lemma 4.17 (Corollary 6.4 [Sam07]) *Let B and h be as above. Then there exists $\alpha \in \mathbb{F}_2^n$ such that $|\widehat{fh}(\alpha)| \geq \eta^2$.*

An appropriate α can be found using the algorithm `Linear-Decomposition` with parameter $\gamma' = \eta^2$ (by picking any element from the list it outputs). We take $q(x) = \langle x, Mx\rangle + \langle \alpha, x\rangle + c$ where $(-1)^c$ is the sign of the coefficient for $(-1)^{\langle \alpha, x\rangle}$ in the linear decomposition. The running time of this step is $O(n^3 \log n \cdot \mathrm{poly}(1/\eta, \log(1/\delta)))$, where δ is the probability of error we want to allow for this invocation of `Linear-Decomposition`.

Note that of all the steps involved in finding a quadratic phase, finding the *linear* part of the phase is the only step for which running time depends exponentially on ε (since $\eta = \exp(-1/\varepsilon^{\Omega(1)})$). The running time of all other steps depends polynomially on $1/\varepsilon$.

4.5 Putting things together

We are now ready to finish the proof of Theorem 4.1.

Proof of Theorem 4.1: For the procedure `Find-Quadratic` the function $\varphi(x)$ will be sampled using Lemma 4.6 as required. We start with a random $u = (x, \varphi(x))$ and a random choice for the parameters $\gamma_1, \gamma_2, \gamma_3$ as described in the analysis of `BSG-Test`. We run the algorithm in Lemma 4.12 using `BSG-Test` with the above parameters and with error parameter $1/2$.

If the algorithm outputs a quadratic form $q(x)$, we estimate $|\langle f, (-1)^q\rangle|$ using $O((1/\eta^4) \cdot \log^2(\rho/\delta))$ samples. If the estimate is less than $\eta^2/2$, or if the algorithm stopped with output $\perp$ we discard q and repeat the entire process. For a M to be chosen later, if we do not find a quadratic phase in M attempts, we stop and output $\perp$.

With probability $\rho/2$, all samples of $\varphi(x)$ (sampled with error $1/n^5$) correspond to a good function φ. Conditioned on this, we have a good choice of u and $\gamma_1, \gamma_2, \gamma_3$ for `BSG-Test` with probability $\rho^3/24$. Conditioned on both the above, the algorithm in Lemma 4.12 finds a good transformation with probability $1/2$. Thus, for $M = O((1/\rho^4) \cdot \log(1/\delta))$, the algorithm stops in M attempts with probability at least $1 - \delta/2$. By choice of the number of samples above, the probability that we estimate $|\langle f, (-1)^q\rangle|$ incorrectly at any step is at most $\delta/2M$. Thus, with probability at least $1-\delta$, we output a good quadratic phase.

One call to the algorithm in Lemma 4.12 requires $O(n^2)$ calls to `BSG-Test`, which in turn requires $\mathrm{poly}(1/\varepsilon)$ calls to `Linear-Decomposition`, each taking time $O(n^2 \log n)$. This dominates the running time of the algorithm, which is $O(n^4 \log n \cdot \mathrm{poly}(1/\varepsilon, 1/\eta, \log(1/\delta)))$. ■

5 A refinement of the inverse theorem

In this section we shall work with a number of refinements of the inverse theorem as stated in Theorem 2.6. For the purposes of the preliminary discussion we shall think of p being any prime, and later specialize to the case $p = 2$.

It was observed (but not exploited) by Green and Tao [GT08] that a slightly stronger form of the inverse theorem holds. If V is a subspace of $\mathbb{F}_p^n$ and $y \in \mathbb{F}_p^n$, then one can define a seminorm $\|.\|_{u^3(y+V)}$ on functions from $\mathbb{F}_p^n$ to $\mathbb{C}$ by setting

$$\|f\|_{u^3(y+V)} = \sup_q |\mathbb{E}_{x\in y+V} f(x)\omega^{-q(x)}|,$$

where the supremum is taken over all quadratic forms q on $y+V$ and ω denotes a pth root of unity. This semi-norm measures the correlation over a coset of the subspace V. We shall be interested in the co-dimension of the subspace, which we shall denote by $\mathsf{cod}\, V$. With this notation, the inverse theorem in [GT08] can be stated as follows.

Theorem 5.1 (Local Inverse Theorem for U^3 [GT08]) *Let $p > 2$, and let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \varepsilon$. Then there exists a subspace V of $\mathbb{F}_p^n$ such that $\mathsf{cod}\, V \leq \varepsilon^{-C}$ and*

$$\mathbb{E}_{y \in V^*} \|f\|_{u^3(y+V)} \geq \varepsilon^C.$$

Here we have denoted the set of coset representatives of V by V^*, so that $V \oplus V^* = \mathbb{F}_2^n$. Actually, the theorem as usually stated involves an averages over the whole of $\mathbb{F}_p^n$ as opposed to just V^*, but the result can be obtained with this modification without difficulty by averaging over coset representatives throughout the proof.

One can deduce the usual inverse theorem from this version without too much effort: by an averaging argument, there must exist y such that f correlates well on $y + V$ with some quadratic phase function ω^q; this function can be extended to a function on the whole of $\mathbb{F}_p^n$ in many different ways, and a further averaging argument yields the usual bounds. However, extending the quadratic phase results in an exponential loss in correlation. (See, for example, Proposition 3.2 in [GT08].)

It turns out that, as Green and Tao remark, an even more precise theorem holds. The result as stated tells us that for each y we can find a local quadratic phase function ω^{q_y} defined on $y + V$ such that the average of $|\mathbb{E}_{x \in y+V} f(x) \omega^{q_y(x)}|$ is at least ε^C. However, it is actually possible to do this in such a way that the quadratic parts of the quadratic phase functions q_y are the same. More precisely, it can be done in such a way that each $q_y(x)$ has the form $q(x-y) + l_y(x-y)$ for a single quadratic function $q : V \to \mathbb{F}_p$ (that is independent of y) and some Freiman 2-homomorphisms $l_y : V \to \mathbb{F}_p$.

This *parallel correlation* was heavily exploited by Gowers and the second author [GW10a, GW10b] in a series of papers on what they called the *true complexity* of a system of linear equations, leading to radically improved bounds compared with the original approach in [GW10c], which was based on an ergodic-style decomposition theorem due to Green and Tao [Gre07].

For $p = 2$, the equivalent of Theorem 5.1 follows directly neither from Green and Tao's nor Samorodnitsky's approach but instead requires a merging of the two. The Green-Tao approach is not directly applicable since the so-called *symmetry argument* in that paper uses division by 2, while Samorodnitsky's approach loses the local information after an application of Freiman's theorem. Section 5 is dedicated to showing how to obtain this *local correlation*[3] in the case where the characteristic is equal to 2. We shall therefore restrict our attention to this case for the remainder of the discussion, bearing in mind that it applies almost verbatim to general p.

In order to be able to refer to the parallel correlation property more concisely, we shall use the concept of *quadratic averages* introduced in [GW10a]. As explained above, for each coset $y + V, y \in V^*$, we can specify a quadratic phase $q_y(x) = q(x-y) + l_y(x-y)$. We extend the definition of q_y to all $y \in \mathbb{F}_p^n$ by setting them equal to $q_{\hat{y}}$ where $\hat{y} \in V^*$ is such that $y \in \hat{y} + V$. Now we can define a quadratic average via the formula

$$Q(x) = \mathbb{E}_{y \in x - V} (-1)^{q_y(x)}.$$

[3] The term "local correlation" may be slightly confusing. It is often used to refer to the fact that in $\mathbb{Z}/N\mathbb{Z}$, no global quadratic correlation with a quadratic phase can be guaranteed. Indeed, such a phase function must be restricted to a Bohr set, or the correlation assumed to only take place on a long arithmetic progression, as in Gowers's original work. However, in $\mathbb{F}_p^n$, the setting we are working in here, there should be no ambiguity.

Notice that the q_y are the same whenever the y lie in the same coset of V. So in fact, since all the q_ys occurring here are such that $y \in x + V$, they are all identical. Thus the value of the quadratic average only depends on the coset of V that x lies in. More precisely, we can write

$$Q(x) = \sum_{y \in V^*} 1_{y+V}(x)(-1)^{q_y(x)}.$$

This tells us that at most $|V^*|$ many linear phases are needed to specify the quadratic average.

Combining the Green-Tao approach with Samorodnitsky's symmetry argument in characteristic 2, we shall obtain an algorithmic version of the analogue of the Local Inverse Theorem (Theorem 5.1) for $p = 2$. In order to use this result in our decomposition algorithm Theorem 3.1, we in fact state it as an algorithm for finding a *quadratic average* $Q(x) = \sum_{y \in V^*} 1_{y+V}(x)(-1)^{q_y(x)}$, which has correlation poly(ε) with the given function. Using this, Theorem 3.1 will then yield a decomposition into poly($1/\varepsilon$) quadratic averages.

Following [GW10c], we shall call the codimension of V the *complexity* of the quadratic average. We will find quadratic averages with complexity poly($1/\varepsilon$). Note that while this means that the description of a quadratic average is still of size $\exp(1/\varepsilon)$, the different quadratic forms appearing in a quadratic average only differ in the linear part.

Theorem 5.2 *Given $\varepsilon, \delta > 0$ and $n \in \mathbb{N}$, there exist $K, C = O(1)$ and a randomized algorithm* `Find-QuadraticAverage` *running in time $O(n^4 \log^2 n \cdot \exp(1/\varepsilon^K) \cdot \log(1/\delta))$, which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, either outputs a quadratic average $Q(x)$ of complexity $O(\varepsilon^{-C})$, or the symbol $\perp$. The algorithm satisfies the following guarantee:*

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

We briefly outline the key modifications in the proof that allow us to obtain this result. Recall that in the previous section we only obtained correlation $\eta = \exp(1/\varepsilon^C)$ because we applied the Freiman-Ruzsa theorem to the set $A_\varphi^{(2)}$: we were only able to assert that $| < A_\varphi^{(2)} > | \leq \exp(1/\varepsilon^C)|A_\varphi^{(2)}|$. Because we had correlation poly(ε) over $A_\varphi^{(2)}$, we obtained correlation $\exp(-1/\varepsilon^C)$ with the linear function we defined on $< A_\varphi^{(2)} >$.

They key difference in the new argument, which borrows heavily from Green and Tao [GT08], is that instead of looking for a subspace *containing* $A_\varphi^{(2)}$, which we previously used to find a linear function, we will look for a subspace *inside* $4A_\varphi^{(2)}$. Given the properties of $A_\varphi^{(2)}$, we will be able to find such a subspace by an application of Bogolyubov's lemma (described in more detail below), with the property that the co-dimension of the subspace is poly($1/\varepsilon$). We will also find a quadratic form such that *restricted to inputs from this subspace*, it has correlation poly($1/\varepsilon$) with the function f. We shall then show (Lemma 5.18) how to extend this quadratic form to all the cosets of the subspace, by adding a *different linear form for each coset* so that the correlation of the resulting quadratic average is still poly($1/\varepsilon$).

We begin by developing algorithmic version of some of the new ingredients in the proof.

5.1 An algorithmic version of Bogolyubov's lemma

We follow Green and Tao in using a form of Bogolyubov's lemma, which has become a standard tool in arithmetic combinatorics. Bogolyubov's lemma as it is usually stated allows one to find a large subspace inside the 4-fold sumset of any given set of large size. We briefly remind the reader of the relationship between sumsets and convolutions, which is used in the proof of the lemma.

For functions $h_1, h_2 : \mathbb{F}_2^n \to \mathbb{R}$, we define their convolution as $h_1 * h_2(x) \overset{\text{def}}{=} \mathbb{E}_y\left[h_1(y)h_2(x-y)\right]$. The Fourier transform diagonalizes the convolution operator, that is, $\widehat{h_1 * h_2}(\alpha) = \widehat{h_1}(\alpha)\widehat{h_2}(\alpha)$ for any two functions h_1, h_2 and any $\alpha \in \mathbb{F}_2^n$, which is easy to verify from the definition. Also, if 1_A is the indicator function for a set $A \subseteq \mathbb{F}_2^n$, then

$$1_A * 1_A(x) \;=\; \mathop{\mathbb{E}}_y\left[1_A(y)\cdot 1_A(x-y)\right] \;=\; \left|\{(y_1,y_2)\;:\;y_1,y_2 \in A \text{ and } y_1+y_2 = x\}\right|/2^n.$$

In particular, $1_A * 1_A$ is supported only on $A + A$ and gives the number of representations of x as the sum of two elements in A. In general, the k-fold convolution is supported on the k-fold sumset.

The proof of Bogolyubov's lemma constructs an explicit subspace by looking at the large Fourier coefficients (using the Goldreich-Levin theorem) and shows that the 4-fold convolution is positive on this subspace. Since we will actually apply this lemma not to a subset but to the output of a randomized algorithm, we state it for an arbitrary function h and its convolution.

We will output a subspace $V \subseteq \mathbb{F}_2^n$ by specifying a basis for the space $V^\perp \overset{\text{def}}{=} \{x : x^T y = 0 \;\; \forall y \in V\}$. Since $\left(V^\perp\right)^\perp = V$, this will also give us a way of checking if $x \in V$: we simply test if $x^T y = 0$ for all basis vectors y of $V^\perp$.

Lemma 5.3 (Bogolyubov's Lemma) *There exists a randomized algorithm* `Bogolyubov` *with parameters ρ and δ which, given oracle access to a function $h : \mathbb{F}_2^n \to \{0,1\}$ with $\mathbb{E}h \geq \rho$, outputs a subspace $V \leqslant \mathbb{F}_2^n$ (by giving a basis for $V^\perp$) of codimension at most $O(\rho^{-3})$ such that with probability at least $1-\delta$, we have $h*h*h*h(x) > \rho^4/2$ for all $x \in V$. The algorithm runs in time $n^2 \log n \cdot \mathrm{poly}(1/\rho, \log(1/\delta))$.*

Proof: We shall use the Goldreich-Levin algorithm `Linear-Decomposition` for the function h with parameter $\gamma = \rho^{3/2}/4$ and error δ to produce a list $K = \{\alpha_1, \ldots, \alpha_k\}$ of length $k = O(\gamma^{-2}) = O(\rho^{-3})$. We take V to be the subspace $\{x \in \mathbb{F}_2^n \;:\; \langle \alpha, x\rangle = 0 \;\; \forall \alpha \in K\}$ and output $\langle K \rangle$. Clearly $\mathsf{cod}(V) \leq |K|$. We next consider the convolution

$$h*h*h*h(x) = \sum_\alpha |\widehat{h}(\alpha)|^4(-1)^{\langle \alpha,x\rangle} = \sum_{\alpha\in K} |\widehat{h}(\alpha)|^4(-1)^{\langle \alpha,x\rangle} + \sum_{\alpha\notin K} |\widehat{h}(\alpha)|^4(-1)^{\langle \alpha,x\rangle}.$$

If $x \in V$, then

$$\sum_{\alpha\in K} |\widehat{h}(\alpha)|^4(-1)^{\langle \alpha,x\rangle} + \sum_{\alpha\notin K} |\widehat{h}(\alpha)|^4(-1)^{\langle \alpha,x\rangle} \geq |\widehat{h}(0)|^4 - \sup_{\alpha\notin K} |\widehat{h}(\alpha)|^2 \cdot \rho$$

The final part of the guarantee in Theorem 4.4 states that the probability of a Fourier coefficient being larger than γ and not being on our list K is at most δ. We conclude that with probability at least $1-\delta$, the expression $h*h*h*h(x)$ is bounded below, for all $x \in V$, by

$$\rho^4 - \rho \cdot \rho^3/2 \geq \rho^4/2,$$

and thus strictly positive. ■

We will, in fact, need a further twist of the above lemma. The function h to which will apply Lemma 5.3 will be defined by the output of a randomized algorithm. Thus, h can be thought of as a random variable, where we choose the value $h(x)$ on each input x by running the randomized algorithm. As in the case of `BSG-Test`, we will have the guarantee that there exist two sets $A^{(1)} \subseteq A^{(2)}$ and $\delta' > 0$ such that *for each input* x, with probability $1 - \delta'$ (over the choice of $h(x)$) we have $1_{A^{(1)}}(x) \leq h(x) \leq 1_{A^{(2)}}(x)$. We will want to use this to conclude that *for the entire subspace* V given by the algorithm `Bogolyubov`, $V \subseteq 4A^{(2)}$.

To argue this, it will be useful to consider the function h' defined as $h' \stackrel{\text{def}}{=} \min\{1_{A^{(2)}}, \max\{h, 1_{A^{(1)}}\}\}$. By definition, we always have that $1_{A^{(1)}}(x) \leq h'(x) \leq 1_{A^{(2)}}(x)$. Also, if for each x, we have with probability $1 - \delta'$ $1_{A^{(1)}}(x) \leq h(x) \leq 1_{A^{(2)}}(x)$, this means that for each x, $\mathbb{P}[h(x) \neq h'(x)] \leq \delta'$. The following claim gives the desired conclusion for the subspace given by the algorithm `Bogolyubov`.

Claim 5.4 *Let h be a random function such that for $\delta' > 0$ and for sets $A^{(1)} \subseteq A^{(2)} \subseteq \mathbb{F}_2^n$, we have that for every x with probability at least $1 - \delta'$, $1_{A_{(1)}}(x) \leq h(x) \leq 1_{A^{(2)}}(x)$. Also, let $\mathbb{E}1_{A^{(1)}} \geq \rho$. Let $h' = \min\{1_{A^{(2)}}, \max\{h, 1_{A^{(1)}}\}\}$ Let V be the subspace returned by the algorithm* `Bogolyubov` *when run with oracle access to h and error parameter δ. Then with probability at least $1 - \delta - \delta' \cdot n^2 \log n \cdot \mathrm{poly}(1/\rho, \log(1/\delta))$, we have that for all $x \in V$, $1_{A^{(2)}} * 1_{A^{(2)}} * 1_{A^{(2)}} * 1_{A^{(2)}}(x) \geq h' * h' * h' * h'(x) > \rho^4/2$. In particular, with above probability, $V \subseteq 4A^{(2)}$.*

Proof: Consider the behavior of the algorithm `Bogolyubov` when run with oracle access to h' instead of h. Since it is always true that $h' \leq 1_{A^{(2)}}$ and $\mathbb{E}[h'] \geq \mathbb{E}[1_{A^{(1)}}] \geq \rho$, the algorithm outputs, with probability $1 - \delta$, a subspace V such that for every $x \in V$, $1_{A^{(2)}} * 1_{A^{(2)}} * 1_{A^{(2)}} * 1_{A^{(2)}}(x) \geq h' * h' * h' * h'(x) > \rho^4/2$. Thus, with probability $1 - \delta$, it outputs a subspace V such that $V \subseteq 4A^{(2)}$.

Finally, we observe that the probability that the algorithm outputs different subspaces when run with oracle access to h and h' is small. The probability of having different outputs is at most the probability that h and h' differ on any of inputs queried by the algorithm `Bogolyubov`. Since it runs in time $n^2 \log n \cdot \mathrm{poly}(1/\rho, \log(1/\delta))$, this probability is at most $\delta' \cdot n^2 \log n \cdot \mathrm{poly}(1/\rho, \log(1/\delta))$. Thus, even when run with oracle access to h, with probability at least $1 - \delta - \delta' \cdot n^2 \log n \cdot \mathrm{poly}(1/\rho, \log(1/\delta))$, the algorithm `Bogolyubov` outputs a subspace $V \subseteq 4A^{(2)}$. ■

Next we require a version of Plünnecke's inequality in order to deal with the size of iterated sumsets. For a proof we refer the interested reader to [TV06], or the recent short and elegant proof by Petridis [Pet11].

Lemma 5.5 (Plünnecke's Inequality) *Let $B \subseteq \mathbb{F}_2^n$ be such that $|B + B| \leq K|B|$ for some $K > 1$. Then for any positive integer k, we have $|kB| \leq K^k|B|$.*

5.2 Finding a good model set

Again, as in Section 4 we may assume that φ is a good function satisfying the guarantee in Lemma 4.6. Recall that $A_\varphi = \{(x, \varphi(x)) : x \in A\}$, where A was defined to be $A = \{x : |\widehat{f_x}(\varphi(x))| \geq \gamma\}$. We will use the routine `BSG-Test` described in Section 4. We assume we have chosen a good vertex u and parameters $\gamma_1, \gamma_2, \gamma_3$ satisfying the guarantee in Lemma 4.10 for `BSG-Test`.

We will need to restrict the sets $A_\varphi^{(1)}$ and $A_\varphi^{(2)}$ given by Lemma 4.10 a bit more before we can apply Bogolyubov's lemma to find an appropriate subspace. Because the subspace sits inside the sumset $4A_\varphi^{(2)}$, an element of the subspace is of the form $(x_1 + x_2 + x_3 + x_4, \varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4))$.

However, unlike tuples of the form $(x, \varphi(x))$, the second half of the tuple $(\varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4))$ may not uniquely depend on the first $(x_1 + x_2 + x_3 + x_4)$.

Since we will require this uniqueness property from our subspace, we restrict our sets to get new sets $A'^{(1)}_\varphi \subseteq A'^{(2)}_\varphi$. These restrictions will satisfy the following property: for all tuples x_1, x_2, x_3, x_4 and x'_1, x'_2, x'_3, x'_4 satisfying $x_1 + x_2 + x_3 + x_4 = x'_1 + x'_2 + x'_3 + x'_4$, we also have $\varphi(x_1) + \varphi(x_2) + \varphi(x_3) + \varphi(x_4) = \varphi(x'_1) + \varphi(x'_2) + \varphi(x'_3) + \varphi(x_4)'$. In other words, φ is a Freiman 4-homomorphism on the first n coordinates of $A'^{(2)}_\varphi$. We will, in fact, need to ensure that it is a Freiman 8-homomorphism in order to obtain a truly linear map.

We shall obtain these restrictions by intersecting the original sets with a subspace, which will be defined using a random linear map $\Gamma : \mathbb{F}_2^n \to \mathbb{F}_2^m$ and a random element $c \in \mathbb{F}_2^m$ (for $m = O(\log(1/\varepsilon))$). This step is often called *finding a good model*, and appears (in non-algorithmic form) as Lemma 6.2 in [GT08]. We shall apply the restriction $\Gamma(\varphi(x)) = c$ to the elements $v = (x, \varphi(x))$ on which `BSG-Test` outputs 1. Since we assume we have already chosen good parameters $u, \rho_1, \rho_2, \gamma_1, \gamma_2, \gamma_3$ for the routine `BSG-Test`, we hide these parameters in the description of the procedure below.

```
Model-Test (v, Γ, c)
    - Let $v = (y, \varphi(y))$.
    - Answer 1 if BSG-Test returns 1 on $v$ and $\Gamma(\varphi(y)) = c$, and 0 otherwise.
```

We shall first show that there exist *good* choices of Γ and c for our purposes. Let $A^{(2)}_\varphi$ be the set provided by Lemma 4.10 for a good choice of parameters. Let $B \subseteq \mathbb{F}_2^n \setminus \{0\}$ be the set of all t such that $(0, t) \in 16A^{(2)}_\varphi$.

Claim 5.6 *Let $\theta' = \varepsilon^{2448}/2^{487}$. The set B has size at most θ'^{-1}.*

Proof: Write $(0, B)$ for the set of all $(0, b), b \in B$. Since $A^{(2)}_\varphi$ is of the form $(x, \varphi(x))$ for some function φ, we have $|A^{(2)}_\varphi + (0, B)| = |A^{(2)}_\varphi||B|$, but at the same time $A^{(2)}_\varphi + (0, B) \subseteq 17A^{(2)}_\varphi$. By Lemma 5.5 we have $|17A^{(2)}_\varphi| \leq (3(2/\rho)^9)^{17}|A^{(2)}_\varphi| \leq (2^{181}/\rho^{153})|A^{(2)}_\varphi|$ since $A^{(2)}_\varphi$ has small sumset, and therefore $|B| \leq 2^{181}/\rho^{153} = \theta'^{-1}$, since $\rho = \varepsilon^{16}/4$. ■

Claim 5.7 *Let $m = 2\lceil \log_2 \theta'^{-1} \rceil$. Then with probability at least 1/2 a random linear map $\Gamma : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is non-zero on all of B.*

Proof: Let $\Gamma : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a randomly chosen linear transformation. Let E_t be the event that $\Gamma(t) = 0$. Clearly $\mathbb{P}(E_t) \leq 2^{-m}$ for each $t \in B$, and thus the probability that Γ is non-zero on all of B is $\mathbb{P}(\cap_t(E_t^C)) = \mathbb{P}((\cup_t E_t)^C) = 1 - \mathbb{P}(\cup_t E_t) \geq 1 - \sum_t \mathbb{P}(E_t) \geq 1 - |B|2^{-m} \geq 1/2$ by choice of m. So with probability at least $1/2$ we have a map Γ that is non-zero on B. ■

Claim 5.8 *Let $\theta = \theta'^2\rho/12$, where θ' is the constant obtained in Claim 5.6, that is, we set $\theta = \varepsilon^{4912}/(3 \cdot 2^{977})$. Fix a map Γ as in Claim 5.7. Then with probability at least θ a randomly chosen element $c \in \mathbb{F}_2^m$ is such that the set*

$$A'^{(1)}_\varphi \stackrel{\text{def}}{=} \{(x, \varphi(x)) \in A^{(1)}_\varphi : \Gamma(\varphi(x)) = c\}$$

has size at least θN.

Proof: The expected size of this set is at least $|A_\varphi^{(1)}|/2^m \geq (\rho N/6)/(\theta'^{-2}) \geq (\theta'^2\rho/6)N$, so with probability θ we can get it to be of size at least θN. ■

We shall of course also define

$$A'^{(2)}_\varphi \overset{\text{def}}{=} \{(x, \varphi(x)) \in A_\varphi^{(2)} : \Gamma(\varphi(x)) = c\},$$

and since $A_\varphi^{(1)} \subseteq A_\varphi^{(2)}$, we have a similar containment for the new subsets, immediately giving a similar lower bound on the size of $A'^{(2)}_\varphi$.

We summarize the above claims in the following refinement of Lemma 4.10.

Lemma 5.9 *Let the calls to* `BSG-Test` *in* `Model-Test` *be with a good choice of parameters* $u, \rho_1, \rho_2, \gamma_1, \gamma_2, \gamma_3$ *and with error parameter* $\delta > 0$*. Then, there exist two sets* $A'^{(1)}_\varphi \subseteq A'^{(2)}_\varphi$*, the output of* `Model-Test` *on input* $v = (y, \varphi(y))$ *satisfies the following with probability* $1 - \delta$*.*

- `Model-Test`$(v, \Gamma, c) = 1 \quad \Longrightarrow \quad v \in A'^{(2)}_\varphi$.
- `Model-Test`$(v, \Gamma, c) = 0 \quad \Longrightarrow \quad v \notin A'^{(1)}_\varphi$.

Moreover, with probability $\theta/2$ *over the choice of* Γ *and c , we have*

$$|A'^{(1)}_\varphi| \geq \theta N \quad \text{and} \quad \varphi \text{ is a Freiman 8-homomorphism on } A^{(2)},$$

where we denote the projection of $A'^{(2)}_\varphi$ *onto the first n coordinates by* $A^{(2)}$*.*

Proof: If `Model-Test` outputs 1, then $v = (y, \varphi(y)) \in A_\varphi^{(2)}$ with probability $1-\delta$ and $\Gamma(\varphi(y)) = c$, so $v \in A'^{(2)}_\varphi$. Similarly, if `Model-Test` outputs 0 then either `BSG-Test` gave 0 or $\Gamma(\varphi(y)) \neq c$, so in any case $v \notin A'^{(1)}_\varphi$.

By Claims 5.8 and 5.7, with probability at least $\theta/2$ over the choice of Γ and c, $|A'^{(1)}_\varphi| \geq \theta N$ and Γ is non-zero on all of B. It remains to verify that φ is a Freiman 8-homomorphism on $A^{(2)}$ in this case.

For any $(0, t) \in 16A'^{(2)}_\varphi$, we have $t \neq 0 \Rightarrow t \in B$ by definition. Also $\Gamma(t) = 16c = 0$ by linearity of Γ. Since Γ is non-zero on all of B, we must have $t = 0$. We also have $16A'^{(2)}_\varphi = 8A'^{(2)}_\varphi + 8A'^{(2)}_\varphi$, and so if we take $(0, t) = (x_1 + \cdots + x_8 + x'_1 + \ldots x'_8, \varphi(x_1) + \cdots + \varphi(x_8) + \varphi(x'_1) + \ldots \varphi(x'_8))$, we have that $x_1 + \cdots + x_8 + x'_1 + \ldots x'_8 = 0$ implies $\varphi(x_1) + \cdots + \varphi(x_8) + \varphi(x'_1) + \ldots \varphi(x'_8) = 0$, making φ a Freiman 8-homomorphism on $A^{(2)}$. ■

5.3 Obtaining a linear choice function on a subspace

As before, we now identify a linear transform (actually, an affine transform) that selects large Fourier coefficients in derivatives. However, as opposed to Section 4 where we defined a linear transform on the whole of $\mathbb{F}_2^n$, here we will just define it on a *coset a subspace* V such that $\mathsf{cod}(V) = \mathrm{poly}(1/\varepsilon)$. In particular, we will prove the following local version of Lemma 4.12.

Lemma 5.10 *Let* φ *be as above and let the parameters for* `BSG-Test` *and* `Model-Test` *be so that they satisfy the guarantees of lemmas 4.10 and 5.9. Let* $\delta > 0$ *and* ε *be as above. Then there exists* *an algorithm running in time $O(n^4 \log^2 n \cdot \exp(1/\varepsilon^K) \cdot \log^2(1/\delta))$ which outputs with probability at least $1-\delta$ a subspace V of codimension at most ε^{-C} as well as a linear linear map $x \mapsto Tx$ and $c_1, c_2 \in \mathbb{F}_2^n$ satisfying $\mathbb{E}_{x \in V+c_1}\left[\widehat{f_x}^2(Tx+Tc_1+c_2)\right] \geq \varepsilon^C$.*

Throughout the argument that follows, we shall assume that we have already chosen good parameters for `BSG-Test` and `Model-Test` so that the conclusions of Lemmas 4.10 and 5.9 hold. We also assume we have access to a good function φ as given by Lemma 4.6.

To find the subspace V we will apply Bogolyubov's lemma to the set identified by the procedure `Model-Test`. We shall look at the second half of the tuples in this subspace (coordinates $n+1$ to $2n$) to find a linear choice function.

Let $h : \mathbb{F}_2^n \to \{0,1\}$ be the (random) function defined by $h(y) = 1$ if $\texttt{Model-Test}(u, (y, \varphi(y)), \Gamma, c) = 1$ and 0 otherwise. The error parameter δ' for `Model-Test` is taken to be δ/n^3. We shall apply the algorithm `Bogolyubov` from Lemma 5.3 with queries to h and with error parameter $\delta_1 = \delta/20$.

Note that the function h is defined on points in $\mathbb{F}_2^n$. Let $A^{(1)}$ and $A^{(2)}$ denote projection on the first n coordinates of the sets $A'^{(1)}_\varphi$ and $A'^{(2)}_\varphi$ given by Lemma 5.9.

Since the last n coordinates are a function (namely φ) of the first n coordinates, we also have $|A'^{(1)}_\varphi| \geq \theta N$, for θ a function of ε as defined in Claim 5.8. Also, with probability $1-\delta'$ for each input x, the inequality $1_{A^{(1)}}(x) \leq h(x) \leq 1_{A^{(2)}}(x)$ holds.

By Claim 5.4, we obtain a subspace V_0 of codimension θ^{-3} such that with probability at least $1 - \delta_1 - \delta' \cdot n^2 \log n \cdot \mathrm{poly}(1/\theta, \log(1/\delta_1)) > 1 - \delta/10$, we have $V_0 \subseteq 4A^{(2)}$. Thus, each element $x \in V_0$ can we written as $x_1+x_2+x_3+x_4$ for $x_1, x_2, x_3, x_4 \in A^{(2)}$. We next show that the set

$$Z_0 \overset{\text{def}}{=} \left\{ (x_1+x_2+x_3+x_4, \varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)) \;\middle|\; \begin{array}{c} x_1+x_2+x_3+x_4 \in V_0, \\ x_1, x_2, x_3, x_4 \in A^{(2)} \end{array} \right\}$$

is also a subspace of $\mathbb{F}_2^{2n}$. Observe that the value of $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)$ is uniquely determined by $x_1+x_2+x_3+x_4$.

Claim 5.11 *There exists a linear map $\zeta : V_0 \to \mathbb{F}_2^n$ satisfying for any $x_1, x_2, x_3, x_4 \in A^{(2)}$ such that $x_1+x_2+x_3+x_4 \in V_0$, we have $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(x_1+x_2+x_3+x_4)$. Thus, the set Z_0 can be written as $Z_0 = \{(x, \zeta(x)) : x \in V_0\}$ and is a subspace of $\mathbb{F}_2^n$.*

Proof: We first show that the value of $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4)$ is uniquely determined by $x_1+x_2+x_3+x_4$. By Lemma 5.9, we know that φ is a Freiman 8-homomorphism on $A^{(2)}$ and hence it is also a Freiman 4-homomorphism. In particular, if for $x_1, x_2, x_3, x_4 \in A^{(2)}$ and $x'_1, x'_2, x'_3, x'_4 \in A^{(2)}$, we have that $x_1+x_2+x_3+x_4 = x'_1+x'_2+x'_3+x'_4$, then it also holds that $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \varphi(x'_1)+\varphi(x'_2)+\varphi(x'_3)+\varphi(x'_4)$. Thus, we can write the set Z_0 as $\{(x, \zeta(x)) : x \in V_0\}$, where ζ if some function on V. We next show that ζ must be a linear function.

We first show that $\zeta(0) = 0$. Since $0 \in V_0$, we must have elements $x_1, x_2, x_3, x_4 \in A^{(2)}$ with the property that $x_1+x_2+x_3+x_4 = 0$, in other words, $x_1+x_2 = x_3+x_4$. But since φ is also a Freiman 2-homomorphism, we get that $\varphi(x_1)+\varphi(x_2) = \varphi(x_3)+\varphi(x_4)$, which implies that $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(0) = 0$.

Since φ is a Freiman 8-homomorphism on $A^{(2)}$ and $V_0 \subseteq 4A^{(2)}$, it follows that ζ is a Freiman 2-homomorphism on V_0. Since V_0 is closed under addition, for $x, y \in V_0$ we can write $x+y = 0+(x+y)$ with all four summands in V_0. Since ζ is 2-homomorphic, we get that $\zeta(x)+\zeta(y) = \zeta(0)+\zeta(x+y) = \zeta(x+y)$. ■

We would like to use the linear map ζ to obtain the choice function on a coset of the space V_0. However, the problem is that we do not *know* the function ζ. We get around this obstacle by generating random tuples $(x_1+x_2+x_3+x_4, \varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4))$ such that $x_1+x_2+x_3+x_4$ and each $x_i \in A^{(2)}$. We show that for sufficiently many samples, the sampled points span a large subspace V of V_0. Since $\varphi(x_1)+\varphi(x_2)+\varphi(x_3)+\varphi(x_4) = \zeta(x_1+x_2+x_3+x_4)$ on V_0, we will be able to obtain the desired linear map on the subspace V.

We sample a point as follows. For the j^{th} sample, we generate four pairs $(x_1^j, \varphi(x_1^j)), \ldots, (x_4^j, \varphi(x_4^j))$. We accept the sample if all four pairs are accepted by `Model-Test` and if $x_1^j+x_2^j+x_3^j+x_4^j \in V$. If a sample is accepted, we store the point $y^j = x_1^j+x_2^j+x_3^j+x_4^j$ and $\zeta(y^j) = \varphi(x_1^j)+\varphi(x_2^j)+\varphi(x_3^j)+\varphi(x_4^j)$. Note that membership in V_0 can be tested efficiently since we know the basis for $V_0^{\perp}$. We first estimate the probability that a point $(y, \zeta(y))$ for $y \in V_0$ is accepted by the above test. This also gives a bound on the number of samples to be tried so that at least $t = O(n^2)$ samples are accepted.

Claim 5.12 *For a $y \in V_0$, the probability that a sample is accepted by the above procedure and the stored pair is equal to $(y, \zeta(y))$ is at least $\theta^4/4N$. Moreover, for some sufficiently large constant C, the probability that out of $C\exp(1/\theta^3)\cdot(1/\theta^4)\cdot t\cdot\log(10/\delta)$ samples fewer than t are accepted is at most $\delta/10$.*

Proof: Since the function $h(x)=1$ exactly when `Model-Test` accepts $(x,\varphi(x))$, the probability that a sample $(x_1,\varphi(x_1)),\ldots,(x_4,\varphi(x_4))$ is accepted and that $x_1+x_2+x_3+x_4=y$, is equal to

$$\mathbb{P}\left[\bigwedge_{i=1}^{4}(h(x_i)=1)\wedge(x_1+x_2+x_3+x_4=y)\right] \;=\; (1/N)\cdot \mathop{\mathbb{E}}_{h,x_1+x_2+x_3+x_4=y}[h(x_1)h(x_2)h(x_3)h(x_4)]$$

As in Claim 5.4, we define the function $h' = \max\{1_{A^{(1)}}, \min\{h, 1_{A^{(2)}}\}\}$. As before, we have that for each x, $\mathbb{P}[h(x)\neq h'(x)] \leq \delta'$, and that $h' * h' * h' * h'(x) > \theta^4/2$ for each $x \in V_0$. We can now estimate the above expectation as

$$\begin{aligned}
&\mathop{\mathbb{E}}_{h,x_1+x_2+x_3+x=y}[h(x_1)h(x_2)h(x_3)h(x_4)]\\
\geq\;& \mathop{\mathbb{P}}_{h,x_1+x_2+x_3+x_4=y}\left[\wedge_{i=1}^{4}(h(x_i)=h'(x_i))\right]\cdot \mathop{\mathbb{E}}_{h,x_1,x_2,x_3}\left[h'(x_1)h'(x_2)h'(x_3)h'(y+x_1+x_2+x_3)\right]\\
\geq\;& (1-4\delta')\cdot h' * h' * h' * h'(y)\\
\geq\;& (1-4\delta')\cdot(\theta^4/2) \;\geq\; \theta^4/4.
\end{aligned}$$

The last inequality exploited the fact that $h' * h' * h' * h'(y) \geq \theta^4/2$ for $y \in V_0$.

The probability that a sample is accepted is equal to the probability that one selects a pair $(y,\zeta(y))$ for *some* $y \in V_0$. This is least $(|V_0|/N)\cdot(\theta^4/2) = \exp(-1/\theta^3)\cdot(\theta^4/2)$. The bound on the probability of accepting fewer than t samples is then given by a Hoeffding bound. ∎

Let $(y^1,\zeta(y^1)),\ldots,(y^t,\zeta(y^t))$ be t stored points corresponding to t samples accepted by the above procedure. The following claim analogous to Claim 4.14 shows that for $t=O(n^2)$, the projection on the first n coordinates of these points must span a large subspace of V_0.

Claim 5.13 *Let $(y^1,\zeta(y^1)),\ldots,(y^t,\zeta(y^t))$ be t points stored according to the above procedure. For $t = n^2+\log(10/\delta)$, the probability that $\mathsf{cod}(<y^1,\ldots,y^t>) \geq \mathsf{cod}(V_0)+\log(4/\theta^4)$ is at most $\delta/10$.*

Proof: Let $k = \mathsf{cod}(V_0) + 4\log(4/\theta)$ and let S be any subspace of codimension k. The probability that a sample $(x_1, \varphi(x_1)), \ldots, (x_4, \varphi(x_4))$ is accepted and has $x_1 + x_2 + x_3 + x_4 = y$ for a specific $y \in S$ is at most $1/N$. Thus, the probability that an accepted sample $(y^j, \zeta(y^j))$ has $y^j \in S$, conditioned on being accepted, is at most $(|S|/N)/((|V_0|/N) \cdot (\theta^4/2))$. Thus, the probability that all t stored points lie in *any* subspace of co-dimension k is at most

$$\left(\frac{|S|/N}{(|V_0|/N) \cdot (\theta^4/2)}\right)^t \cdot \#\{\text{suspaces of co-dimension } k\} \;=\; \left(\frac{\theta^4/4}{\theta^4/2}\right)^t \cdot 2^{n(n-k)} \;\le\; 2^{-t} \cdot 2^{n^2},$$

which is at most $\delta/10$ for $t = n^2 + \log(10/\delta)$. ■

Let $V =< y^1, \ldots, y^t >$. The above claim shows that with high probability, the codimension of V satisfies $\mathsf{cod}(V) = \exp(1/\theta^3)$. From the way the samples were generated, we also know $\zeta(y^1), \ldots, \zeta(y^t)$. Since ζ is a linear function by Claim 5.11, we can extend it to a linear transform $x \mapsto Tx$ such that $\forall x \in V$, $Tx = \zeta(x)$ (as in Section 4).

We now show that there is a coset of V on which Tx identifies large Fourier coefficients of the derivative f_x. We define the set $Z \overset{\text{def}}{=} \{(x, Tx) \,:\, x \in V\}$. We will find a coset of Z such that a significant fraction of points in this coset are of the form $(x, \varphi(x)) \in A'^{(2)}_\varphi$. Recall that a point $(x, \varphi(x))$ in $A'^{(2)}_\varphi$ satisfies $|\hat{f}_x(\varphi(x))| \ge \gamma = O(\varepsilon^{16})$. Thus, Tx will be a linear function selecting large Fourier coefficients for a significant fraction of points in this coset.

The following claim shows the existence of such a coset.

Claim 5.14 *The sets $Z + A'^{(1)}_\varphi$ and $Z + A'^{(2)}_\varphi$ both consist of at most $(1/\theta) \cdot (N/\,|Z|)$ cosets of Z. Hence, for some $c \in A'^{(1)}_\varphi$ we have $|(Z + c) \cap A'^{(2)}_\varphi| \ge |(Z + c) \cap A'^{(1)}_\varphi| \ge \theta^2 \cdot |Z|$.*

Proof: Since $Z \subseteq 4A'^{(2)}_\varphi$ and $A'^{(1)}_\varphi \subseteq A'^{(2)}_\varphi$, we have that

$$Z + A'^{(1)}_\varphi \;\subseteq\; Z + A'^{(2)}_\varphi \;\subseteq\; 5A'^{(2)}_\varphi \;\subseteq\; 5A^{(2)}_\varphi.$$

The last inclusion follows from the fact that $A'^{(2)}_\varphi$ was obtained by intersecting $A'^{(2)}_\varphi$ (given by Lemma 4.10) with a subspace.

We know from Lemma 4.10 that $|A^{(2)}_\varphi + A^{(2)}_\varphi| \le (2/\rho)^8 \cdot N \le (2/\rho)^8 \cdot (6/\rho) \cdot |A^{(2)}_\varphi|$. Lemma 5.5 (Plünnecke's inequality) then gives that $|5A^{(2)}_\varphi| \le (6/\rho)^{45} \cdot |A^{(2)}_\varphi| \le (1/\theta) \cdot |A^{(2)}_\varphi| \le (1/\theta) \cdot N$. Thus, $|Z + A'^{(2)}_\varphi| \le (1/\theta) \cdot N$ and it is the union of at most $(1/\theta) \cdot (N/|Z|)$ cosets.

Since $A'^{(1)}_\varphi \subseteq Z + A'^{(1)}_\varphi$, there must exist at least one coset $Z + c$ for $c \in A'^{(1)}_\varphi$, such that

$$\left|(Z + c) \cap A'^{(1)}_\varphi\right| \;\ge\; \frac{|A'^{(1)}_\varphi|}{(1/\theta) \cdot (N/|Z|)} \;\ge\; \theta^2 \cdot |Z|,$$

where the last inequality used the fact that $|A'^{(1)}_\varphi| \ge \theta N$, as guaranteed by Lemma 5.9. ■

We now show how to computationally identify this coset of Z. We will simply sample a sufficiently large number of points on which `Model-Test` answers 1. We will then divide the points into different cosets of Z and pick the coset with the most number of elements. The following claim shows that this procedure succeeds in finding the desired coset with high probability.

Claim 5.15 *Let $s = C \cdot (N/|Z|) \cdot (\log(1/\delta)/\theta^5) \leq C \cdot \exp(1/\theta^3) \cdot (\log(1/\delta)/\theta^5)$ for a sufficiently large constant C. There exists an algorithm which runs in time $O(n^3 \cdot s^2)$ and finds, with probability at least $1 - \delta/5$, a point $c \in A'^{(2)}_{\varphi}$ such that $|(Z + c) \cap A'^{(2)}_{\varphi}| \geq (\theta^2/2) \cdot |Z|$.*

Proof: We sample s independent elements of the form $(x, \varphi(x))$ and reject all the ones on which `Model-Test` outputs 0, where we run `Model-Test` with error parameter $\delta' = \delta/(10s)$. For some $r \leq s$, let $(x_1, \varphi(x_1)), \ldots, (x_r, \varphi(x_r))$ be the accepted elements.

For each $i, j \leq r$, we test if $(x_i, \varphi(x_i))$ and $(x_j, \varphi(x_j))$ lie in the same coset of Z, by checking if $(x_i - x_j, \varphi(x_i) - \varphi(x_j)) \in Z$. This takes time $O(n^3)$ for each i, j as we need to check if $(x_i - x_j, \varphi(x_i) - \varphi(x_j))$ can be expressed as a linear combination of the basis vectors for Z, which requires solving a system of linear equations.

Lying in the same coset is an equivalence relation, which divides the points $(x_1, \varphi(x_1)), \ldots, (x_r, \varphi(x_r))$ into equivalence classes. We pick the class with the maximum number of elements. Since $(0, 0) \in Z$, for any element $(x_i, \varphi(x_i))$ in this class, we can write the coset as $Z + (x_i, \varphi(x_i))$. We thus pick an arbitrary element of the form $(x_i, \varphi(x_i))$ in the largest class and output $c = (x_i, \varphi(x_i))$.

The running time of the above algorithm is $O(s^2 \cdot n^3)$. We need to argue that with probability at least $1 - \delta/5$, the coset $Z + c$ with the maximum number of samples satisfies $|(Z + c) \cap A'^{(2)}_{\varphi}| \geq (\theta^2/2) \cdot |Z|$.

With probability at least $1 - \delta' \cdot s = 1 - \delta/10$, `Model-Test` answers 1 on all elements in $A'^{(1)}_{\varphi}$ and 0 on all elements outside $A'^{(2)}_{\varphi}$. For any coset of the form $Z + c$, let $N(Z + c)$ be the number of samples that land in the coset. Conditioned on the correctness of `Model-Test`, we have that for any coset of the form $Z + c$,

$$s \cdot \frac{|(Z + c) \cap A'^{(1)}_{\varphi}|}{N} \leq \mathbb{E}\left[N(Z + c)\right] \leq s \cdot \frac{|(Z + c) \cap A'^{(2)}_{\varphi}|}{N},$$

which by definition of s implies that

$$C \cdot \frac{\log(1/\delta)}{\theta^5} \cdot \frac{|(Z + c) \cap A'^{(1)}_{\varphi}|}{|Z|} \leq \mathbb{E}\left[N(Z + c)\right] \leq C \cdot \frac{\log(1/\delta)}{\theta^5} \cdot \frac{|(Z + c) \cap A'^{(2)}_{\varphi}|}{|Z|}.$$

By a Hoeffding bound, the probability that $N(Z + c)$ deviates by an additive $(C/4) \cdot (\log(1/\delta)/\theta^3)$ from the expectation is at most $\delta \cdot \exp(-C'(1/\theta^3))$ for any fixed coset. Since the number of cosets is at most $(1/\theta) \cdot \exp(1/\theta^3)$ by Claim 5.14, the probability that on *any* coset $N(Z + c)$ deviates from the expectation by the above amount is at most $\delta \cdot \exp(-C'(1/\theta^3)) \cdot (1/\theta) \cdot \exp(1/\theta^3) < \delta/10$ for an appropriate value of C'.

By Claim 5.14, we know that there is a coset $Z + c$ with $|(Z + c) \cap A'^{(1)}_{\varphi}| \geq \theta^2|Z|$ and hence $\mathbb{E}\left[N(Z + c)\right] \geq C \cdot (\log(1/\delta)/\theta^3)$. By the above deviation bound, we should have that $N(Z + c) \geq (3C/4) \cdot (\log(1/\delta)/\theta^3)$ for this coset. Thus, the coset with the maximum number of samples, say $Z + c'$, will certainly also satisfy $N(Z + c') \geq (3C/4) \cdot (\log(1/\delta)/\theta^3)$. Again, by the deviation bound, it must be true that $\mathbb{E}\left[N(Z + c')\right] \geq (C/2) \cdot (\log(1/\delta)/\theta^3)$, and hence $|(Z + c) \cap A'^{(2)}_{\varphi}| \geq \theta^2|Z|/2$. ■

We can now combine the previous argument to prove Lemma 5.10.

Proof of Lemma 5.10: We follow the steps described above to find the subspace V_0, and subsequently the subspace V together with the transformation T. This immediately yields the subspace $Z = \{(x,Tx) : x \in V\}$. Claim 5.15 finds $c = (c_1,c_2) \in \mathbb{F}_2^n$ such that a fraction of at least $\theta^2/2$ of points $(y+c_1, Ty+c_2)$ in the coset $Z + (c_1,c_2)$ are of the form $(x,\varphi(x))$ for $(x,\varphi(x)) \in A'^{(2)}_{\varphi}$, and so $|\hat{f_x}^2(\varphi(x))| \geq \gamma = O(\varepsilon^{16})$. Since $(y, Ty+c_2) = (x+c_1, \varphi(x))$ for these points, we have $T(x+c_1)+c_2 = \varphi(x)$. This implies

$$\mathop{\mathbb{E}}_{x\in c_1+V}\left[\widehat{f_x}^2(Tx+Tc_1+c_2)\right] \ \geq\ (\theta^2/2)\cdot\gamma^2 \ \geq\ \varepsilon^C. \tag{3}$$

The errors in the application of Bogolyubov's lemma and in Claims 5.12, 5.13 and 5.15 add up to $\delta/2 < \delta$. The running time is dominated by the $C\exp(1/\theta^3)\cdot(1/\theta^4)\cdot t\cdot\log(10/\delta)$ calls to `Model-Test` in Claim 5.12 for $t = O(n^2)$. Since each call to `Model-Test` takes $O(n^2\log n\cdot\mathrm{poly}(1/\varepsilon)\cdot\log(\delta/n^3))$ time, the total running time is $O(n^4\log^2 n\cdot\exp(O(1/\theta^3))\cdot\log^2(1/\delta))$. ■

Fourier analysis over a subspace

To begin with we collect some basic facts about Fourier analysis over a subspace of $\mathbb{F}_2^n$, which will be required for the remaining part of the argument. Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be a function and let $W \subseteq \mathbb{F}_2^n$ be a subspace. We define the Fourier coefficients of f with respect to the subspace as the correlation with a linear phase over the subspace.

As in the case of Fourier analysis over $\mathbb{F}_2^n$, it is easy to verify that the functions $\{\chi_\alpha\}_{\alpha\in W}$ with $\chi_\alpha(x) \stackrel{\text{def}}{=} (-1)^{\langle\alpha,x\rangle}$ form an orthonormal basis for functions from W to $\mathbb{R}$ with respect to the inner product $\langle f_1, f_2\rangle_W \stackrel{\text{def}}{=} \mathbb{E}_{x\in W}\left[f_1(x)f_2(x)\right]$. Thus the dual group $\hat{W}$ of these basis functions is isomorphic to W. As in the case of $\mathbb{F}_2^n$, we have Parseval's identity saying that $\sum_{\alpha\in W}\langle f,\chi_\alpha\rangle_W^2 = \mathbb{E}_{x\in W}\left[f^2(x)\right]$.

It is easy to modify the proof of the Goldreich-Levin theorem so that it can be used to identify the linear functions χ_α for $\alpha \in W$ that have large correlation with a Boolean function f over a subspace W. We omit the details.

Theorem 5.16 (Goldreich-Levin theorem for a subspace) *Let $\gamma,\delta > 0$ and $W \subseteq \mathbb{F}_2^n$ be a given subspace. There is a randomized algorithm which, given oracle access to a function $f : \mathbb{F}_2^n \to \{-1,1\}$, runs in time $O(n^2\log n\cdot\mathrm{poly}(1/\gamma,\log(1/\delta)))$ and outputs a list $L = \{\alpha_1,\ldots,\alpha_k\}$ with each $\alpha_i \in W$ such that*

- $k = O(1/\gamma^2)$.
- $\mathbb{P}\left[\exists\alpha_i \in L\ |\langle f,\chi_{\alpha_i}\rangle_W| \leq \gamma/2\right] \leq \delta$.
- $\mathbb{P}\left[\exists\alpha \notin L\ |\langle f,\chi_{\alpha_i}\rangle_W| \geq \gamma\right] \leq \delta$.

5.4 Finding a quadratic phase on a subspace

In order to deduce the refined inverse theorem (Theorem 5.1) for $p = 2$, we need to redo the symmetry argument and integration phase with this local expression obtained in Lemma 5.10. The modifications to Samorodnitsky's approach are relatively minor but we give complete proofs nonetheless. One significant difference is that we will need to take Fourier transforms relative to subspaces.

We begin by obtaining a subspace $W \leqslant V$ on which the matrix T obtained in the previous step is symmetric, thereby providing the "local" analogue of Lemma 4.16.

Lemma 5.17 (Symmetry Argument) *Given a subspace V and a linear map T with the property that*

$$\mathbb{E}_{x\in c_1+V}\widehat{f_x}^2(Tx+z_c)\geq \varepsilon^C,$$

we can output a subspace $W \leqslant V$ of codimension at most $\log(\varepsilon^{-C})$ inside V together with a symmetric matrix B on W with zero diagonal such that

$$\mathbb{E}_{x\in c_1+W}\widehat{f_x}^2(Bx+z_c)\geq \varepsilon^C$$

in time $O(n^3)$.

Proof: We let $g(x)=(-1)^{\langle x,Tx+z_c\rangle}$ and $F(x)=\widehat{f_x}^2(Tx+z_c)$, and begin by noting that by Lemma 6.11 in [Sam07], we have that $g(x)=-1$ implies $F(x)=0$. Therefore we have

$$\varepsilon^C\leq \mathbb{E}_{x\in c_1+V}\widehat{f_x}^2(Tx+z_c)=\mathbb{E}_{x\in c_1+V}g(x)F(x)=\mathbb{E}_{x\in V}g^{c_1}(x)F^{c_1}(x),$$

we have written $h^y(x)$ for the shift $h(x+y)$. Taking the Fourier transform relative to the subspace V, we obtain

$$\varepsilon^C\leq (\sum_{\alpha\in\widehat{V}}\widehat{g^{c_1}}(\alpha)\widehat{F^{c_1}}(\alpha))^2,$$

and by the Cauchy-Schwarz inequality and Parseval's theorem this is bounded above by

$$\sum_{\alpha\in\widehat{V}}\widehat{g^{c_1}}(\alpha)^2\sum_{\alpha\in\widehat{V}}\widehat{F^{c_1}}(\alpha)^2\leq \mathbb{E}_{x\in V}g^{c_1}*_V g^{c_1}(x).$$

The latter (local) convolution can easily be computed:

$$g^{c_1}*_V g^{c_1}(x)=\mathbb{E}_{y\in V}(-1)^{\langle x+y+c_1,T(x+y)+c_2\rangle}(-1)^{\langle y+c_1,Ty+c_2\rangle}=g^{c_1}(x)(-1)^{\langle c_1,c_2\rangle}\mathbb{E}_{y\in V}(-1)^{\langle (T+T^T)x,y\rangle}.$$

The final expectation gives the indicator function of the subspace

$$W'=\{x\in V:\langle (T+T^T)x,y\rangle=0 \text{ for all } y\in V\},$$

that is, W' is a linear subspace on which T is symmetric. Note that W' is the space of solutions of a linear system of equations, a basis of which can be computed by Gaussian elimination in time $O(n^3)$.

We denote the map that takes x to Tx for $x\in W'$ by B. We have just shown that

$$|\mathbb{E}_{x\in V}1_{W'}(x)g^{c_1}(x)|\geq \varepsilon^C,$$

and in particular since g is bounded, we quickly observe that W' has density at least ε^C inside V. This means the codimension can have gone up by at most $\log(\varepsilon^{-C})$, which is negligible in the grand scheme of things.

It remains to ensure that B has zero diagonal. Again this can be rectified in a small number of steps. Denote this diagonal by $v\in\mathbb{F}_2^n$. Let $W=W'\cap <v+z_c>^\perp$ if $\langle c_1,c_2\rangle=0$, otherwise intersect W' with the (unique) coset of $<v+z_c>^\perp$. Since $\langle x,Bx\rangle=\langle x,v\rangle$ over $\mathbb{F}_2$, we have that $\langle x+c_1,v+z_c\rangle=\langle x,Bx+z_c\rangle+\langle c_1,c_2\rangle$, and thus by Lemma 6.11 in [Sam07] if $x+c_1\in W'$ but $\notin W$, that is, $x+c_1\notin <v+z_c>^\perp$, then $\widehat{f_x}^2(Bx+z_c)=0$.

Hence we obtain

$$2\mathbb{E}_{x\in c_1+W}\widehat{f_x}^2(Bx+z_c)=\mathbb{E}_{x\in c_1+W'}\widehat{f_x}^2(Bx+z_c),$$

which yields the desired conclusion. ■

Finally, we need to perform the integration. The procedure is very similar to Lemma 4.17, but again we have to work relative to a subspace.

Lemma 5.18 (Integration Step) *Let $f : \mathbb{F}_2^n \to [-1,1]$. Let B be a symmetric $n \times n$ matrix with zero diagonal such that $\mathbb{E}_{x\in c_1+W}\widehat{f_x}^2(Bx+z_c) \geq \varepsilon^C$. Let $A \in \mathbb{F}_2^{n\times n}$ be a matrix such that $B = A + A^T$. Then there exist, for every $y \in \mathbb{F}_2^n$, a vector $r_y \in W$ such that*

$$\mathbb{E}_{y\in W^*}|\mathbb{E}_{x\in y+W} f(x)(-1)^{\langle x,Ax\rangle+\langle By,x\rangle+\langle r_y,x\rangle}| \geq \varepsilon^C.$$

Proof: Consider the quadratic phase $g(x) = (-1)^{\langle x,Ax\rangle}$ and the linear phase $l(z) = (-1)^{\langle z,z_c\rangle}$. (Note that this is where we require B to have zero diagonal.) We shall first prove that

$$\mathbb{E}_{x\in c_1+W}\widehat{f_x}^2(Bx+z_c) = \mathbb{E}_{x\in c_1+W}(\mathbb{E}_{y\in W^*}\langle f_x, g_x l\rangle_{y+W})^2 \leq \mathbb{E}_{y\in W^*}\sum_{\alpha\in\widehat{W}}\widehat{(fgl)^y}^2(\alpha)\widehat{(fg)^y}^2(\alpha),$$

where again we have written $h^y(x)$ for the shift $h(x+y)$ and the final Fourier transform is taken with respect to W. The equality follows from the fact that

$$\widehat{f_x}(Bx+z_c) = \mathbb{E}_y f_x(y)(-1)^{\langle y,Bx+z_c\rangle} = \mathbb{E}_{y\in W^*}\mathbb{E}_{z\in y+W} f_x(z)(-1)^{\langle z,Bx+z_c\rangle}$$

and so

$$(-1)^{\langle x,Ax\rangle}\widehat{f_x}(Bx+z_c) = \mathbb{E}_{y\in W^*}\mathbb{E}_{z\in y+W} f_x(z)(-1)^{\langle z+x,A(z+x)\rangle+\langle z,Az\rangle} l(z) = \mathbb{E}_{y\in W^*}\langle f_x, g_x l\rangle_{y+W},$$

where the inner product is taken over the translate $y+W$. For the inequality write

$$\mathbb{E}_{x\in c_1+W}(\mathbb{E}_{y\in W^*}\langle f_x, g_x l\rangle_{y+W})^2 \leq \mathbb{E}_{y\in W^*}\mathbb{E}_{x\in c_1+W}\langle f_x, g_x l\rangle_{y+W}^2,$$

which equals

$$\mathbb{E}_{y\in W^*}\mathbb{E}_{x\in c_1+W}(\mathbb{E}_{z\in y+W} fgl(z)fg(z+x))^2 = \mathbb{E}_{y\in W^*}\mathbb{E}_{x\in W}(\mathbb{E}_{z\in y+W} fgl(z)fg(z+x+c_1))^2,$$

which in turn can be reexpressed as

$$\mathbb{E}_{y\in W^*}\mathbb{E}_{x\in W}(\mathbb{E}_{z\in W}(fgl)^y(z)(fg)^y(z+x+c_1))^2 = \mathbb{E}_{y\in W^*}\mathbb{E}_{x\in W}((fgl)^y *_W (fg)^y)(x+c_1)^2.$$

Taking the Fourier transform with respect to W, it can be seen that the latter expression equals

$$\mathbb{E}_{y\in W^*}\sum_{\alpha\in\widehat{W}}\widehat{(fgl)^y}^2(\alpha)\widehat{(fg)^y}^2(\alpha),$$

completing the proof of the claim from the beginning. But since all functions involved are bounded,

$$\mathbb{E}_{y\in W^*}\sum_{\alpha\in\widehat{W}}\widehat{(fgl)^y}^2(\alpha)\widehat{(fg)^y}^2(\alpha) \leq \mathbb{E}_{y\in W^*}\sup_{\alpha\in\widehat{W}}|\widehat{(fg)^y}(\alpha)|.$$

Now for each $y \in W^*$, we fix a $\alpha_y \in \widehat{W}$ such that the supremum is attained. Then we have shown that

$$\varepsilon^C \leq \mathbb{E}_{y\in W^*}|\widehat{(fg)^y}(\alpha_y)| = \mathbb{E}_{y\in W^*}|\mathbb{E}_{x\in W} f(x+y)(-1)^{\langle x+y,A(x+y)\rangle+\langle\alpha_y,x\rangle}|,$$

which, after some rearranging of the phase, completes the proof. ■

5.5 Obtaining a quadratic average

Finally, we use the subspace W from Section 5.4 to obtain the required quadratic average.

Lemma 5.19 *Let $W \leqslant \mathbb{F}_2^n$ be a subspace with $\mathsf{cod}(V) \leq (1/\varepsilon^C)$. Let $A \in \mathbb{F}_2^{n\times n}$ and $B = A + A^T$ be such that there exist vectors $r_y \in W$ for each $y \in W^*$ satisfying*

$$\mathop{\mathbb{E}}_{y\in W^*}\left[\left|\mathop{\mathbb{E}}_{x\in y+W}\left[f(x)(-1)^{\langle x,Ax\rangle+\langle By,x\rangle+\langle r_y,x\rangle}\right]\right|\right] \geq \sigma.$$

Then for $\delta > 0$, one can find in time $n^2\log n \cdot |W^*| \cdot \mathrm{poly}(1/\sigma, \log(1/\delta))$ a quadratic average with a vector l_y and a constant c_y for each $y \in W^*$ satisfying

$$\mathop{\mathbb{E}}_{y\in W^*}\left[\mathop{\mathbb{E}}_{x\in y+W}\left[f(x)(-1)^{\langle x,Ax\rangle+\langle l_y,x\rangle+c_y}\right]\right] \geq \sigma^2/10.$$

Proof: Let $h_y(x) \stackrel{\text{def}}{=} f(x)(-1)^{\langle x,Ax\rangle+\langle x,By\rangle}$. By assumption we immediately find that

$$\mathop{\mathbb{E}}_{y\in W^*}\left[\left|\mathop{\mathbb{E}}_{x\in y+W}\left[h_y(x)(-1)^{\langle r_y,x\rangle}\right]\right|\right] \;=\; \mathop{\mathbb{E}}_{y\in W^*}\left[\left|\mathop{\mathbb{E}}_{x\in W}\left[h_y^y(x)(-1)^{\langle r_y,x\rangle}\right]\right|\right] \;\geq\; \sigma.$$

Here $h_y^y(x) = h_y(x+y)$ as before. Without loss of generality, we may assume that the vectors r_y maximize the above expression. Thus, we know that on average (over y), the functions h_y^y have a large Fourier coefficient (that is, significant correlation with some vector $r_y \in W$) over the subspace W. For every $y \in W^*$, we will use Theorem 5.16 to find this Fourier coefficient when it is indeed large. For those y for which the expression $\left|\mathbb{E}_{x\in W}\left[h_y^y(x)(-1)^{\langle r_y,x\rangle}\right]\right|$ is small for all $r_y \in W$, we will simply pick an arbitrary phase.

Let us describe this procedure in more detail. First, by an averaging argument we know that

$$\mathop{\mathbb{E}}_{y\in W^*}\left[\left|\mathop{\mathbb{E}}_{x\in W}\left[h_y^y(x)(-1)^{\langle r_y,x\rangle}\right]\right|\right] \;\geq\; \sigma \;\Rightarrow\; \mathop{\mathbb{P}}_{y\in W^*}\left[\left|\mathop{\mathbb{E}}_{x\in W}\left[h_y^y(x)(-1)^{\langle r_y,x\rangle}\right]\right| \geq \sigma/2\right] \;\geq\; \sigma/2.$$

Let $S \stackrel{\text{def}}{=} \{y \in W^* : \left|\mathbb{E}_{x\in W}\left[h_y^y(x)(-1)^{\langle r_y,x\rangle}\right]\right| \geq \sigma/2\}$. The above inequality shows that $|S| \geq (\sigma/2)\cdot W^*$. Now for each $y \in W^*$, we run the Goldreich-Levin algorithm for the subspace W from Theorem 5.16 with the function h_y^y, the parameter $\gamma = \sigma/2$ and error probability $\delta^2/2$.

For each $y \in S$ the algorithm finds, with probability $1-\delta^2$, an $r'_y \in W$ and a $c_y \in \mathbb{F}_2$ satisfying $\mathbb{E}_{x\in W}\left[h_y^y(x)(-1)^{\langle r'_y,x\rangle+c_y}\right] \;\geq\; \sigma/4$. Thus, with probability $1-\delta/2$, it finds such an r'_y for at least a $1-\delta$ fraction of $y \in S$. For $y \notin S$, that is for those y for which the algorithm fails to find a good linear phase, we choose an r'_y arbitrarily. If we can force the contribution of terms for $y \notin S$ to be non-negative, then we have that with probability $1-\delta/2$

$$\mathop{\mathbb{E}}_{y\in W^*}\left[1_S(y)\cdot \mathop{\mathbb{E}}_{x\in W}\left[h_y^y(x)(-1)^{\langle r'_y,x\rangle+c_y}\right]\right] \;\geq\; (1-\delta)\cdot(\sigma/2)\cdot(\sigma/8) \geq \sigma^2/9.$$

It remains to choose constants c_y for $y \notin S$ in such a way that their contribution to the average is non-negative. Consider the two potential assignments $c_y = 0\ \forall y \notin S$ and $c_y = 1\ \forall y \notin S$. Clearly the contribution of the terms for $y \notin S$ must be non-negative for at least one of the aforementioned assignments, in which case we obtain

$$\mathop{\mathbb{E}}_{y\in W^*}\left[\mathop{\mathbb{E}}_{x\in W}\left[h_y^y(x)(-1)^{\langle r'_y,x\rangle+c_y}\right]\right] \;\geq\; \sigma^2/9.$$

In order to determine which of the two assignments works, we can try both sets of signs and estimate the corresponding quadratic average using $O((1/\sigma^4) \cdot \log(1/\delta))$ samples, and choose the set of signs for which the estimate is larger. By Lemma 2.1, with probability at least $1 - \delta/2$, we select a set of values c_y such that

$$\mathop{\mathbb{E}}_{y \in W^*}\left[\mathop{\mathbb{E}}_{x \in y+W}\left[f(x)(-1)^{\langle x,Ax\rangle+\langle x,By\rangle+\langle x,r'_y\rangle+c_y}\right]\right] = \mathop{\mathbb{E}}_{y \in W^*}\left[\mathop{\mathbb{E}}_{x \in W}\left[h^y_y(x)(-1)^{\langle r'_y,x\rangle+c_y}\right]\right] \geq \sigma^2/10.$$

Choosing $l_y = By + r'_y$ then completes the proof. ■

5.6 Putting things together

We now give the proof of Theorem 5.2.

Proof of Theorem 5.2: For the procedure `Find-QuadraticAverage` the function $\varphi(x)$ will be sampled using Lemma 4.6 as required. We start with a random $u = (x, \varphi(x))$ and a random choice of the parameters $\gamma_1, \gamma_2, \gamma_3$ as described in the analysis of `BSG-Test`. We also choose the map Γ and the value c randomly for `Model-Test`. We run the algorithm in Lemma 5.10 using `BSG-Test` and `Model-Test` with the above parameters, and with error parameter 1/4.

Given a coset of the subspace V and the map T, we find a subspace $W \subseteq V$ and a symmetric matrix B with zero diagonal, using Lemma 5.17. We then use the algorithm in Lemma 5.19 to obtain the required quadratic average, with probability 1/4.

Given a quadratic average $Q(x)$, we estimate $|\langle f, Q\rangle|$ using $O((1/\sigma^4) \cdot \log^2(\theta/\delta))$ samples. If the estimate is less than $\sigma^2/20$, we discard Q and repeat the entire process. For a M to be chosen later, if we do not find a quadratic average in M attempts, we stop and output $\perp$.

With probability $\rho/2$, all samples of $\varphi(x)$ (sampled with error $1/n^5$) correspond to a good function φ. Conditioned on this, we have a good choice of u and $\gamma_1, \gamma_2, \gamma_3$ for `BSG-Test` with probability $\rho^3/24$. Also, we have a good choice of the map Γ and c for `Model-Test` with probability at least $\theta/2 = \varepsilon^{O(1)}$. Conditioned on the above, the algorithm in Lemma 5.10 finds a good transformation with probability 3/4 and thus the output of the algorithm in Lemma 5.19 is a good quadratic average with probability at least 1/2.

Thus, for $M = O((1/\rho^4) \cdot (1/\theta)\log(1/\delta))$, the algorithm stops in M attempts with probability at least $1 - \delta/2$. By choice of the number of samples above, the probability that we estimate $|\langle f, (-1)^q\rangle|$ incorrectly at any step is at most $\delta/2M$. Therefore we output a good quadratic average with probability at least $1 - \delta$.

The complexity of the quadratic average obtained, which is equal to the co-dimension of the space W, is at $O(1/\theta^3) = O(1/\varepsilon^C)$. The running time of each of the M steps is dominated by that of the algorithm in Lemma 5.10, which is $O(n^4 \log^2 n \cdot \exp(1/\varepsilon^K))$. We conclude that the total running time is $O(n^4 \log^2 n \cdot \exp(1/\varepsilon^K) \cdot \log(1/\delta))$. ■

6 Discussion

One way in which one might want extend the results in this paper is to consider the cyclic group of integers modulo of prime $\mathbb{Z}_N$. A (linear) Goldreich-Levin algorithm exists in this context [AGS03], and some quadratic decomposition theorems have been proven (see for example [GW10b]). However, strong quantitative results involving the U^3 norm require a significant amount of effort to even state.

For example, the role of the subspace relative to which the quadratic averages are defined will be played by so-called Bohr sets, which act as approximate subgroups in $\mathbb{Z}_N$. Moreover, it is no longer true that the inverse theorem can guarantee the existence of a globally defined quadratic phase with which the function correlates; instead, this correlation may be forced to be (and remain) local. Since there is an informal dictionary for translating analytic arguments from $\mathbb{F}_p^n$ to $\mathbb{Z}_N$, it seems plausible that many of our arguments could be extended to this setting, at the cost of adding a significant layer of (largely technical) complexity to the current presentation.

7 Acknowledgements

The authors would like to thank Tim Gowers, Swastik Kopparty, Tom Sanders and Luca Trevisan for helpful conversations.

References

[AGS03] Adi Akavia, Shafi Goldwasser, and Shmuel Safra, *Proving hard-core predicates using list decoding*, FOCS, 2003, pp. 146–.

[BS94] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica **14** (1994), 263–268, 10.1007/BF01212974.

[BTZ10] V. Bergelson, T. Tao, and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of* $\mathbb{F}^\omega$, Geom. Funct. Anal. **16** (2010), no. 6, 1539–1596.

[BV10] Andrej Bogdanov and Emanuele Viola, *Pseudorandom bits for polynomials*, SIAM J. Comput. **39** (2010), no. 6, 2464–2486.

[Can10] P. 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. MR 2586962 (2011a:11017)

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

[GKZ08] P. Gopalan, A.R. Klivans, and D. Zuckerman, *List-decoding Reed-Muller codes over small fields*, STOC, 2008, pp. 265–274.

[GL89] O. Goldreich and L. Levin, *A hard-core predicate for all one-way functions*, Proceedings of the 21st ACM Symposium on Theory of Computing, 1989, pp. 25–32.

[Gop10] P. Gopalan, *A Fourier-analytic approach to Reed-Muller decoding*, FOCS, 2010, pp. 685–694.

[Gow98] W.T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), no. 3, 529–551.

[Gre07] B.J. Green, *Montréal notes on quadratic Fourier analysis*, Additive combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 69–102. MR 2359469 (2008m:11047)

[GT08] B.J. Green and T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153. MR 2391635 (2009g:11012)

[GW10a] W.T. Gowers and J. Wolf, *Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$*, To appear, Mathematika. doi:10.1112/S0025579311001264, arXiv:1002.2209 (2010).

[GW10b] ———, *Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$*, To appear, J. Anal. Math., arXiv:1002.2210 (2010).

[GW10c] ———, *The true complexity of a system of linear equations*, Proc. Lond. Math. Soc. (3) **100** (2010), no. 1, 155–176. MR 2578471 (2011a:11019)

[HL11] Hamed Hatami and Shachar Lovett, *Correlation testing for affine invariant properties on $\mathbb{F}_p^n$ in the high error regime*, 2011.

[O'D08] R. O'Donnell, *Some topics in analysis of Boolean functions*, STOC, 2008, pp. 569–578.

[Pet11] G. Petridis, *Plünnecke's Inequality*, Preprint, arXiv:1101.2532 (2011).

[Ruz99] I.Z. Ruzsa, *An analog of Freiman's theorem in groups*, Astérisque (1999), no. 258, xv, 323–326, Structure theory of set addition. MR 1701207 (2000h:11111)

[Sam07] A. Samorodnitsky, *Low-degree tests at large distances*, Proceedings of the 39th ACM Symposium on Theory of Computing, 2007, pp. 506–515.

[SSV05] B. Sudakov, E. Szemerédi, and V.H. Vu, *On a question of Erdös and Moser*, Duke Mathematical Jounal **129** (2005), no. 1, 129–155.

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

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

[TV06] T. Tao and V. Vu, *Additive combinatorics*, Cambridge University Press, 2006.

[TZ10] T. Tao and T. Ziegler, *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, Analysis and PDE **3** (2010), 1–20.

[Vio07] E. Viola, *Selected results in additive combinatorics: An exposition (preliminary version)*, 2007.

[VW07] Emanuele Viola and Avi Wigderson, *Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols*, IEEE Conference on Computational Complexity, 2007.