Julia Wolf

Arithmetic Structure in Sets of Integers

Synopsis

This dissertation deals with four problems concerning arithmetic structures in dense sets of integers. In Chapter 1 we give an exposition of the state-of-the-art technique due to Pintz, Steiger and Szemerédi which yields the best known upper bound on the density of sets whose difference set is square-free. Inspired by the well-known fact that Fourier analysis is not sufficient to detect progressions of length 4 or more, we determine in Chapter 2 a necessary and sufficient condition on a system of linear equations which guarantees the correct number of solutions in any uniform subset of $\mathbb{F}_p^n$. This joint work with Tim Gowers constitutes the core of this thesis and relies heavily on recent progress in so-called "quadratic Fourier analysis" pioneered by Gowers, Green and Tao. In particular, we use a structure theorem for bounded functions which provides a decomposition into a quadratically structured and a quadratically uniform part. We also present an alternative decomposition leading to improved bounds for the main result, and discuss the connections with recent results in ergodic theory. Chapter 3 deals with improved upper and lower bounds on the minimum number of monochromatic 4-term progressions in any two-colouring of $\mathbb{Z}_N$. Finally, in Chapter 4 we investigate the structure of the set of popular differences of a subset of $\mathbb{Z}_N$. More precisely, we establish that, given a subset of size linear in N, the set of its popular differences does not always contain the complete difference set of another large set.

Arithmetic Structure in Sets of Integers

Julia Wolf

Clare College

UNIVERSITY OF CAMBRIDGE

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy
December 2007

About This Thesis

The research described in this dissertation was performed in the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge between October 2003 and December 2007, and was supervised by Professor W.T. Gowers. During this period, the author enjoyed the hospitality of the Universitat Politècnica de Catalunya, Barcelona (September - December 2005), the Massachusetts Institute of Technology, Cambridge MA (January - June 2006) and the Institute for Advanced Study, Princeton NJ (September - December 2007). The author was supported by a grant from the Engineering and Physical Sciences Research Council and a Gates Cambridge Scholarship for the duration of her graduate studies. The visit to Barcelona was financed by the EU Marie Curie Research Training Network COMBSTRU.

I hereby confirm that this dissertation is the result of my own work, in the process of which I have benefited from many useful discussions with Tim Gowers, Ben Green and Tom Sanders. The work presented in this dissertation is original apart from the results of Chapter 1 and Sections 2.5, 3.4 and 3.5, which are, with some simplifications and modifications, of an expository nature. The main result of Chapter 2 was conceived in collaboration with Tim Gowers. These contributions are reiterated and carefully referenced at the appropriate points in the text.

No part of this dissertation has been submitted or will be submitted for a degree or other qualification at any other university.

Acknowledgements

My warmest thanks must go to Tim Gowers, whose love for beautiful mathematics has inspired me from the day I attended his first lecture in April 2000. I am also greatly indebted to Ben Green, whose unbounded enthusiasm for the subject and constant encouragement have contributed greatly to the completion of this thesis. From my very first day at Cambridge, Andrew Thomason challenged me at every step along the way and thus deserves a great deal of credit for my mathematical development.

For the rather adventurous journey that is a Ph.D., I could not have hoped for better company than that of Tom Sanders. I shall always be grateful to Dominic Vella and Arabella Schelpe, without whose support I would have given up at various points over the past eight years. And if it hadn't been for the presence of Anne-Sophie Kaloghiros and Pablo Suarez-Serrato, our Head of Department would no doubt have been justified in calling Pavilion E "the morgue".

I would also like to thank Clare College for providing a second home and a stimulating environment throughout my time at Cambridge, as well as Sally Lowe and Sue Goodbody for dealing with all administrative matters promptly and efficiently.

And last but not least my family, who never questioned my decisions, and have no doubt suffered a great deal because of my stubbornness.

This thesis is dedicated to my grandmother. Für Oma.

Synopsis

This dissertation deals with four problems concerning arithmetic structures in dense sets of integers. In Chapter 1 we give an exposition of the state-of-the-art technique due to Pintz, Steiger and Szemerédi which yields the best known upper bound on the density of sets whose difference set is square-free. Inspired by the well-known fact that Fourier analysis is not sufficient to detect progressions of length 4 or more, we determine in Chapter 2 a necessary and sufficient condition on a system of linear equations which guarantees the correct number of solutions in any uniform subset of $\mathbb{F}_p^n$. This joint work with Tim Gowers constitutes the core of this thesis and relies heavily on recent progress in so-called "quadratic Fourier analysis" pioneered by Gowers, Green and Tao. In particular, we use a structure theorem for bounded functions which provides a decomposition into a quadratically structured and a quadratically uniform part. We also present an alternative decomposition leading to improved bounds for the main result, and discuss the connections with recent results in ergodic theory. Chapter 3 deals with improved upper and lower bounds on the minimum number of monochromatic 4-term progressions in any two-colouring of $\mathbb{Z}_N$. Finally, in Chapter 4 we investigate the structure of the set of popular differences of a subset of $\mathbb{Z}_N$. More precisely, we establish that, given a subset of size linear in N, the set of its popular differences does not always contain the complete difference set of another large set.

Contents

Chapter 0

Introduction and Motivation

0.1 Structures and Patterns

There is little doubt that the natural numbers were the first type of numbers to be conceived by mankind, long before the more creative amongst us human beings came up with the concept of rational and irrational numbers, the complex numbers and the p-adics. In spite of this long history, many of the simplest questions we can ask about these most primitive objects of mathematics remain unresolved to this day.

Consider one of the patterns that is most readily described, a 3-term arithmetic progression written as a triple $x, x+d, x+2d$. Given a large number N, how large can a subset of the natural numbers 1 up to N be assuming it contains no non-trivial 3-term progressions? It seems intuitively obvious that the larger the set, the harder it ought to be to avoid a given arithmetic pattern.

Roth [Rot53] was the first to provide a meaningful upper bound on the size of a set without 3-term progressions, while Behrend [Beh46] gave an explicit construction of a progression-free set of rather large density. Behrend's example has not been surpassed in the sixty years since its initial publication, and despite several recent improvements on Roth's upper bound there remains a significant gap in our understanding of this problem.

An extension of Roth's Theorem, namely the statement that any sufficiently dense subset of the integers contains a k-term arithmetic progression for arbitrary fixed k, was proved by Szemerédi [Sze75] in 1975. Both the search for better bounds in Roth's Theorem and the quest for effective control over long progressions have sparked the

development of many powerful new techniques in discrete harmonic analysis, some of which find application in this dissertation.

Arithmetic combinatorics is the now commonly accepted term for the area of mathematics that deals with structural questions of precisely the kind we just described. Both the type of structure under consideration (arithmetic progressions, square differences, sum and difference sets) as well as the setting in which they occur (sets of integers, graphs and hypergraphs, functions defined on the hypercube) may vary, but the common theme throughout is what has often been termed a *dichotomy* between structure and randomness: either the object under consideration behaves in a random-like way, in which case it is possible to count the desired arithmetic patterns to a high degree of accuracy, or else the object was highly structured to start with, which improves our situation from the outset. An important consequence of this dichotomy is our ability to decompose *any* object (a set, graph or function) into a structured and a random-looking component. For an extraordinarily insightful introduction to this sphere of results the reader is referred to [Tao05].

This dissertation deals with four problems concerning a variety of different arithmetic structures in dense sets of integers. In Chapter 1 we give an exposition of the state-of-the-art technique due to Pintz, Steiger and Szemerédi which yields the best known upper bound on the density of sets whose difference set is square-free. Inspired by the well-known fact that Fourier analysis is not sufficient to detect progressions of length 4 or more, we determine in Chapter 2 a necessary and sufficient condition on a system of linear equations which guarantees the correct number of solutions in any uniform subset of $\mathbb{F}_p^n$. This joint work with Tim Gowers constitutes the core of this thesis and relies heavily on recent progress in so-called "quadratic Fourier analysis" pioneered by Gowers, Green and Tao. In particular, we use a structure theorem for bounded functions which provides a decomposition into a quadratically structured and a quadratically uniform part. We also present an alternative decomposition leading to improved bounds for the main result, and discuss the connections with recent results in ergodic theory. Chapter 3 deals with improved upper and lower bounds on the minimum number of monochromatic 4-term progressions in any two-colouring of $\mathbb{Z}_N$. Finally, in Chapter 4 we investigate the structure of the set of popular differences of a subset of $\mathbb{Z}_N$. More precisely, we establish that, given a subset of size linear in N, the set of its popular differences does not always contain the complete difference set of another large set.

The variability of these problems is reflected in the relatively wide range of techniques needed to attack them. The next section gives a brief overview of the main results and the methods we use, sets up the notation that will be used throughout and serves

as a guide to the remaining chapters.

0.2 Methods and Techniques

We begin by setting up the basics of discrete Fourier analysis, which in one form or another pervades every chapter of this thesis.

0.2.1 Fourier Analysis on Finite Abelian Groups

Many times a problem concerning arithmetic structures in sets of integers can be transferred to a finite Abelian group G, which is advantageous from the point of view of performing harmonic analysis. In particular, we shall be thinking of G as either $\mathbb{F}_p^n$ with p a small fixed prime and n tending to infinity, or the cyclic group $\mathbb{Z}_N$ for N a large prime.

For a character $\gamma \in \widehat{G}$, we define the *Fourier transform* $\widehat{f}(\gamma)$ of f at the frequency γ by the formula

$$\widehat{f}(\gamma) := \mathbb{E}_{x \in G} f(x)\gamma(-x).$$

We use the expectation operator $\mathbb{E}_{x \in G}$ to denote the sum over the elements of G divided by the cardinality of G. Recall that when $G = \mathbb{F}_p^n$ or $\mathbb{Z}_N$, the Pontryagin dual $\widehat{G}$ of G is isomorphic to G itself, and the characters $\gamma(x)$ take the form $\omega^{t \cdot x}$, where ω denotes a p^{th} or N^{th} root of unity, respectively.

Amongst the very basic useful properties of the discrete Fourier transform are the *Fourier inversion formula*

$$f(x) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma(x)$$

and *Parseval's Identity*

$$\mathbb{E}_{x \in G} |f(x)|^2 = \sum_{\gamma \in \widehat{G}} |\widehat{f}(\gamma)|^2.$$

Another indispensable tool is that of discrete *convolution*, which for two functions f and $g : G \to \mathbb{C}$ is defined as

$$f * g(x) := \mathbb{E}_{y \in G} f(y)g(x - y).$$

It is easy to verify straight from the definitions that the Fourier transform of the convolution of two functions equals the product of their individual Fourier transforms,

in other words,

$$\widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma).$$

We shall use the norms $\|f\|_s^s := \mathbb{E}_{x\in G}|f(x)|^s$ on physical and $\|\widehat{f}\|_s^s := \sum_{\gamma\in\widehat{G}}|\widehat{f}(\gamma)|^s$ on Fourier space. It should be clear from the context which one is being used. With this notation, *Hölder's Inequality* becomes

$$\|f_1 f_2 ... f_k\|_s \leq \|f_1\|_{s_1}\|f_2\|_{s_2}...\|f_k\|_{s_k}$$

whenever $s^{-1} = s_1^{-1} + s_2^{-1} + ... + s_k^{-1}$. The *Cauchy-Schwarz Inequality*, our most versatile weapon, takes the simple form

$$|\mathbb{E}_{x\in G}f(x)g(x)| \leq \|f\|_2\|g\|_2,$$

although we shall often use it on Fourier space with the appropriate normalisation.

A subset $A \subseteq G$ will be referred to as *uniform* if all non-trivial Fourier coefficients of its characteristic function are small. Small will usually mean $o(1)$, which is a quantity tending to 0 as the size of the group G tends to infinity.

It is then easy to see that if G has odd order, any uniform subset $A \subseteq G$ of size $|A| = \alpha|G|$ contains roughly the same number of 3-term progressions as a random subset of G, where the elements are chosen independently at random with probability α. Indeed, it follows by expanding the characteristic function $A(x)$ of the set A in terms of its Fourier coefficients that

$$\mathbb{E}_{x,d\in G}A(x)A(x+d)A(x+2d) = \sum_{\gamma\in\widehat{G}}\widehat{A}(\gamma)^2\widehat{A}(-2\gamma).$$

It is straightforward to compute that the trivial character γ_0 makes a contribution of α^3, and the remaining sum can be bounded, using the Cauchy-Schwarz Inequality on Fourier space, as

$$|\sum_{\gamma\neq\gamma_0}\widehat{A}(\gamma)^2\widehat{A}(-2\gamma)| \leq \sup_{\gamma\neq\gamma_0}|\widehat{A}(\gamma)|\sum_{\gamma\in\widehat{G}}|\widehat{A}(\gamma)|^2.$$

The final sum is bounded by α as a consequence of Parseval's Theorem, and thus for $\sup_{\gamma\neq\gamma_0}|\widehat{A}(t)|$ sufficiently small the contribution from the non-trivial Fourier modes is negligible. We conclude that A really does contain $\alpha^3|G|^2$ 3-term progressions.

Having carefully set out Fourier analysis on a finite Abelian group, we shall use it in Chapter 1 in the case $G = \mathbb{Z}_N$ to give an exposition of a paper by Pintz, Steiger and

Szemerédi [PSS88] improving the bound in Sárközy's Theorem, which states that any sufficiently dense subset of $\{1, \ldots, N\}$ contains two distinct elements whose difference is a perfect square.

Theorem 1.2. *Any subset $A \subseteq \{1, \ldots, N\}$ whose difference set is square-free has density*

$$\alpha \ll (\log N)^{-\frac{1}{4}\log\log\log\log N}.$$

The original paper [PSS88] is rather difficult to digest, but the main idea is one that deserves clarification, as it is an ingenious extension of the now classical energy increment argument used in the proof of Szemerédi's Theorem for progressions of length 4, which may turn out to have other applications.

Similar to the case of 3-term progressions discussed above, the starting point is an identity of the form

$$\mathbb{E}_{x,y\in\mathbb{Z}_N} A(x)A(y)S(x-y) = \sum_{t\in\mathbb{Z}_N} |\widehat{A}(t)|^2\widehat{S}(t),$$

where S denotes the characteristic function of the set of squares. For subsets $A \subseteq \{1, \ldots, N\}$ containing no square differences, the left-hand side is equal to zero. It is a well-known fact that the set of squares has small Fourier transform at frequencies $t \in \mathbb{Z}_N$ such that t/N is close to a rational with large denominator, and bounded Fourier transform otherwise. Indeed, this observation gave rise to the development of the *circle method* by Hardy and Littlewood in the 1920s.

0.2.2 Quadratic Fourier Analysis

It was first observed in [Gow98] (and in the context of ergodic theory, by Furstenberg and Weiss [FW96]) that ordinary Fourier analysis is not sufficient to count progressions of length 4 or longer. In particular, it was shown that there exist uniform sets which contain significantly more than the expected number of 4-term progressions. Gowers established that progressions of length $k+1$ are governed by the so-called *U^k-norms*, which he defined as follows.

Definition 2.2. *Let G be a 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*

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

where $C^{|\omega|}f = f$ *if* $\sum_i \omega_i$ *is even and* $\overline{f}$ *otherwise.*

It is not hard to see that $\|f\|_{U^2}^4 = \sum_t |\widehat{f}(t)|^4$, and therefore having small U^2-norm is equivalent to being uniform in the sense discussed above. If the characteristic function of a set has small U^3-norm, we say that the set is *quadratically uniform.* Gowers showed, using nothing more than the Cauchy-Schwarz Inequality, that if a set is *uniform of order* k in the sense that its characteristic function is small in U^{k+1}, then it contains roughly the expected number of progressions of length $k+2$.

Following their paper on long arithmetic progressions in the primes [GT04], Green and Tao set out to investigate the behaviour of general linear systems in [GT06a]. They established a notion of complexity of a linear system which we shall refer to as the *Cauchy-Schwarz complexity* and for whose precise definition we refer the reader to the introduction of Chapter 2. Roughly speaking, Cauchy-Schwarz complexity k describes precisely those linear systems for which the Cauchy-Schwarz Inequality allows us to reduce to an estimate of the U^{k+1}-norm of the characteristic function of the set. That is, Cauchy-Schwarz complexity k determines a sufficient condition for a system to be governed by the U^{k+1}-norm.

The starting point of my joint investigations with Tim Gowers, which culminated in the paper [GW07b], was the question of which types of linear systems require which degree of uniformity. In other words, is Cauchy-Schwarz complexity k also a necessary condition for the system to be governed by the U^{k+1}-norm? In particular, are there systems of Cauchy-Schwarz complexity 2 which are in fact governed by the U^2-norm, that is, ordinary Fourier analysis? The surprising answer is yes, and in fact, we can give a necessary and sufficient condition on a linear system of Cauchy-Schwarz complexity 2 which guarantees that it is governed by the U^2-norm.

In order to make this statement more precise, we make the following definition.

Definition 2.5. *Let* $\mathcal{L}$ *be a system of* m *distinct linear forms* $L_1, L_2, \ldots, L_m$ *in* d *variables. The* true complexity *of* $\mathcal{L}$ *is the smallest* k *with the following property. For every* $\epsilon > 0$ *there exists* $\delta > 0$ *such that if* G *is any finite Abelian group and* $f : G \to \mathbb{C}$ *is any function with* $\|f\|_\infty \leq 1$ *and* $\|f\|_{U^{k+1}} \leq \delta$*, then*

$$\left|\mathbb{E}_{x_1,\ldots,x_d \in G} \prod_{i=1}^{m} f(L_i(x_1, \ldots, x_d))\right| \leq \epsilon.$$

The main result of Chapter 2, and indeed this thesis, comes in two parts. The first one says that if the linear system $\mathcal{L}$ on $\mathbb{F}_p^n$ is such that the squares of its linear forms

are linearly dependent over $\mathbb{F}_p$, then we can find a uniform set A which contains significantly more than the expected number of solutions to $\mathcal{L}$.

Theorem 2.7. *Let $\mathcal{L} = (L_1, \dots, L_m)$ be a system of linear forms in d variables and suppose that the quadratic forms $L_i^T L_i$ are linearly dependent over $\mathbb{F}_p$. Then there exists $\epsilon > 0$ such that for every $\delta > 0$ there exists n and a set $A \subset \mathbb{F}_p^n$ with the following two properties.*

(i) A is δ-uniform of degree 1.

(ii) If $\mathbf{x} = (x_1, \dots, x_d)$ is chosen randomly from $(\mathbb{F}_p^n)^d$, then the probability that $L_i(\mathbf{x})$ is in A for every i is at least $\alpha^m + \epsilon$, where α is the density of A.

In other words, the true complexity of $\mathcal{L}$ is at least 2.

The complementary part says that if the system $\mathcal{L}$ has Cauchy-Schwarz complexity 2 and is *square-independent* in the sense that the squares of the linear forms defining $\mathcal{L}$ are linearly independent over $\mathbb{F}_p$, then any uniform set A contains approximately the expected number of solutions to $\mathcal{L}$.

Corollary 2.20. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let A be a c-uniform subset of $\mathbb{F}_p^n$ of density α. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, with Cauchy-Schwarz complexity at most 2. Let $\mathbf{x} = (x_1, \dots, x_d)$ be a random element of $(\mathbb{F}_p^n)^d$. Then the probability that $L_i(\mathbf{x}) \in A$ for every i differs from α^m by at most ϵ.*

More generally, we expect the following result to hold.

Conjecture 2.6. *The true complexity of a system of linear forms $\mathcal{L} = (L_1, \dots, L_m)$ is equal to the smallest k such that the functions L_i^{k+1} are linearly independent.*

Our main tool is what is known as a *structure theorem* for the U^3-norm. It allows us to decompose any bounded function into a quadratically structured and a quadratically uniform part, that is, we can write f as $f_1 + f_2$, where f_1 is a quadratically structured object and f_2 is small in U^3. Of course, there is a trade-off between the degree of structure we can achieve for f_1 and the degree of uniformity we can obtain for f_2. We shall be using the following version in the setting $\mathbb{F}_p^n$ due to Green and Tao [Gre05b].

Theorem 2.9. *Let p be a fixed prime, let $\delta > 0$ and suppose that $n > n_0(\delta)$ is sufficiently large. Given any function $f : \mathbb{F}_p^n \to [-1, 1]$, there exists a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ of complexity at most $d = d(\delta)$ together with a decomposition*

$$f = f_1 + f_2,$$

where

$$f_1 := \mathbb{E}(f|\mathcal{B}_2) \quad \textit{and} \quad \|f_2\|_{U^3} \leq \delta.$$

This structure theorem follows by iteration from the so-called U^3-*inverse theorem*, which was first given by Green and Tao in [GT05a] and has its roots in the work of Gowers [Gow98]. Gowers showed that if a function on $\mathbb{Z}_N$ has large U^3-norm, then it correlates with a quadratic phase along a long arithmetic progression. Observe that a U^2-inverse theorem is self-evident: It is easy to check that $\|f\|_{U^2} \leq \|f\|_\infty^{1/2}$, so that if a function has large U^2-norm, then it automatically correlates with a linear phase by definition of the Fourier transform. The analogous statement for U^3, on the other hand, is a deep result combining heavy combinatorial tools such as Freiman's Theorem and the Balog-Szemerédi-Gowers Theorem. Green and Tao added a symmetry argument to obtain correlation on a so-called *Bohr set*, which can be thought of as more "global" than a long arithmetic progression. (These last remarks become relevant only in the case $\mathbb{Z}_N$. In our model setting $\mathbb{F}_p^n$, correlation is always proved on a low-codimensional subspace.)

Our use of Theorem 2.9 (or rather, a variant thereof) results in tower-type bounds for Theorem 2.19. Using a decomposition into sums of quadratic phases rather than an ergodic-type factor approach, we present an alternative proof of Theorem 2.19 resulting in improved bounds in Section 4.1. More precisely, we derive the following dependence between the uniformity parameter c and the resulting error ϵ in the average over the linear system.

Theorem 2.21. *In Theorem 2.19, the uniformity parameter c can be taken to be a tower of exponentials of height $m+1$ in the error ϵ^{-1}.*

By exploiting a more precise version of the inverse theorem the authors have been able to improve this bound even further (to doubly exponential). This improvement together with a proof of Theorem 2.19 in the setting $\mathbb{Z}_N$, where one is forced to work with local quadratic phases (that is, phases defined on a Bohr set) from the outset, is contained in the forthcoming paper [GW07a]. An extension to the case $k = 2$ of Conjecture 2.6 for systems of Cauchy-Schwarz complexity 3, conditional on a conjectured U^4-inverse theorem, is in preparation.

Quadratic Fourier analysis also motivates some of the observations made in Chapter 3. In particular, we use (and slightly modify) a recent construction of Gowers [Gow06b] to exhibit a 2-colouring of $\mathbb{Z}_N$ which contains significantly fewer than the expected number of monochromatic 4-term progressions. This is the first non-trivial upper bound for the minimum number of monochromatic 4-term progressions we are aware of.

Theorem 3.2. *There exists a colouring of $\mathbb{Z}_N$ with N a prime containing fewer than $1/16(1-1/2025)N^2$ monochromatic 4-term progressions.*

By a careful counting argument we also improve the best known lower bound due to Cameron, Cilleruelo and Serra [CCS05].

Theorem 3.1. *Any 2-colouring of $\mathbb{Z}_N$ with N a prime contains at least $1/32N^2$ monochromatic four-term progresssions.*

These results together with a discussion of the corresponding colouring problem in graphs are due to appear in [Wol07].

0.2.3 Ergodic Theory

Motivated in part by a recent paper of Leibman [Lei07], which proves the main result of Chapter 2 in the ergodic theoretic setting, we have included a brief discussion of the connections between arithmetic combinatorics and ergodic theory in Section 2.5.

This is by no means the first time that the paths of these two seemingly unrelated areas of mathematics have crossed. Following on from classical examples such as the proof of Szemerédi's Theorem by Furstenberg [Fur77], there is a plethora of recent work that exemplifies the close connections between the two fields, notably by Green, Tao and Ziegler. While some results from ergodic theory find direct applications in number theory via the so-called *Furstenberg Correspondence Principle*, it is more often the case that in fact a similar phenomenon occurs in both contexts.

For example, the question of which degree of uniformity characterises the behaviour of a linear system corresponds to asking for the degree of the minimal characteristic factor of the associated multiple ergodic average. In particular, saying that ordinary Fourier analysis suffices to count solutions to a certain linear system corresponds to saying that the so-called *Kronecker factor* is characteristic for the ergodic average under consideration.

Although ergodic approaches suffer from the disadvantage that they do not give quantitative bounds and that they require an initial investment in acquiring the necessary jargon, the elegance of the subject often leads the way to an intuitive understanding of many structural questions that we are currently unable to answer quantitatively. For example, one point of envy is that ergodic theorists are able to deal with general polynomial (not just linear) systems of equations such as arithmetic progressions with square common difference, without any significant leap in conceptual difficulty

once the correct set-up is found. It would be of great interest to explore the parallels between the two areas in more detail with the aim of establishing similar results in arithmetic combinatorics.

0.2.4 Graphs and Hypergraphs

Because we shall touch upon the topic in Sections 2.3.4 and 3.4, we briefly mention the connection with the concept of uniformity in graphs and hypergraphs.

The concept of *quasirandomness* was introduced by Thomason in the 1980s and subsequently developed further through work by Chung, Graham and others. It turns out that there is a set of equivalent conditions such that if a graph satisfies one of them, it is guaranteed to contain the expected number of all small fixed subgraphs. Note how this stands in stark contrast to the world of subsets of the integers which we discussed at length in the preceding section: once we know a graph is quasirandom, we obtain all fixed substructures for free.

The result in graph theory which is analogous to the structure theorem we discussed above is the famous *Szemerédi Regularity Lemma*, which allows one to decompose any dense graph into a bounded number of components, the bipartite graph between any two of which behaves quasirandomly. This allows us to count almost any conceivable structure inside such graphs, and it is therefore not surprising that the result has found countless applications in graph theory.

The notion of quasirandomness was subsequently extended to hypergraphs by Chung and Graham [CG90]. A more sophisticated version of quasirandomness in hypergraphs, together with the corresponding regularity decomposition, was developed recently by Gowers [Gow06a] and independently by Rödl et al. [RS04],[NRS06].

0.2.5 Probabilistic Tools

Let G be a finite Abelian group of order N. Suppose that A is a subset of G of cardinality linear in N, and define the set of γ-*popular* differences of A to be

$$D_\gamma(A) := \{x \in G : A * -A(x) \geq \gamma\},$$

where we have written A for the indicator function of the subset A. In other words, $D_M(A)$ is the set of elements of G which can be written as a difference of elements of A in at least γN different ways. Because we are considering subsets of G of size linear in N, we shall take γ to be a small constant throughout the paper. Is it true

that $D_\gamma(A)$ always contains the complete difference set $A_0 - A_0$ for some large set A_0? Our aim in Chapter 4 is to show that this is not always so. More precisely, when $G = \mathbb{F}_2^n$ and $G = \mathbb{Z}_N$ with N a prime, we prove that there exists a set A of linear size such that any set A_0 whose difference set is contained in $D_\gamma(A)$ has density $o(1)$. Here $o(1)$ denotes a quantity tending to 0 as the order N of the group G tends to infinity.

Theorem 4.1. *Let $G = \mathbb{F}_2^n$ or $G = \mathbb{Z}_N$. Then there exists a set $A \subseteq G$ of size greater than $N/3$ with the property that any set A_0 whose difference set is contained in the set $D_\gamma(A)$ of γ-popular differences of A has density $o(1)$.*

Theorem 4.1 is not only an interesting result in its own right, but it rules out certain simpler approaches to counting sum-free sets in the spirit of Lev, Łuczak and Schoen [LŁS01] as well as Green and Ruzsa [GR05]. Theorem 4.1 is the main result of [Wol05].

The construction we use, namely a so-called *niveau set*, was originally introduced by Ruzsa [Ruz91] and has seen a number of interesting applications in arithmetic combinatorics to date. The main tool in determining many of its properties is a classical theorem from probability theory known as *Esseen's Inequality*, which allows us to compare two distribution functions provided we have enough information about the higher moments of the corresponding random variables.

Theorem 4.12. *Let F_1, F_2 be probability distribution functions with corresponding characteristic functions ϕ_1, ϕ_2. Assume F_1' exists and is pointwise bounded by a constant V. Then*

$$\sup_x |F_1(x) - F_2(x)| \ll \frac{V}{T} + \int_0^T \frac{|\phi_1(t) - \phi_2(t)|}{t} dt.$$

We also make use of *measure concentration* results in both the discrete cube $\mathbb{F}_2^n$ and the k-dimensional torus $\mathbb{T}^k$ in the form of Theorem 4.7 and Corollary 4.29, respectively.

Acknowledgements. The author is greatly indebted to Tim Gowers and Ben Green for suggesting many of the problems in this thesis, and for guiding her towards their solution through countless valuable insights. She has also benefited from discussions with Endre Szemerédi, Bryna Kra, Andrew Thomason and Geoffrey Grimmett, whose individual contributions are acknowledged at the appropriate points in the text.

Chapter 1

Sets Whose Difference Set is Square-Free

1.1 Introduction

The purpose of this chapter is to give an exposition of the best known bound on the density of sets whose difference set contains no squares, which was first derived by Pintz, Steiger and Szemerédi in [PSS88]. We show how their method can be brought in line with the modern view of the energy increment strategy employed in problems such as Szemerédi's Theorem on arithmetic progressions, and explore the extent to which the particularities of the method are specific to the set of squares.

Results about the types of arithmetic structures one is guaranteed to find inside dense sets of integers have been around since the 1950s when Roth [Rot53] first proved that any subset of the integers of positive upper density contains a 3-term arithmetic progression. Szemerédi [Sze75], and independently Furstenberg [Fur77], extended this result to longer progressions. Much of what drives arithmetic combinatorics these days is closely related to the search for better bounds for this problem.

Another type of structure mathematicians have always been fascinated by is that of perfect squares. Sárközy [Sár78a] proved the following beautiful theorem in 1978.

Theorem 1.1. *Any subset $A \subseteq [N]$ whose difference set is square-free has density*

$$\alpha \ll \frac{(\log\log N)^{2/3}}{(\log N)^{1/3}}.$$

Throughout this chapter, we shall take the symbol "$\ll$" to mean "is bounded above by a constant times", and write $[N]$ for the set $\{1, 2, \ldots, N\}$. We will be mainly

concerned with outlining the main steps leading to a proof of the best known bound for Theorem 1.1 by Pintz, Steiger and Szemerédi, which was first published in [PSS88] with a subsequent extension of the result to k^{th} powers in [BPPS94].

Theorem 1.2. *Any subset $A \subseteq [N]$ whose difference set is square-free has density*

$$\alpha \ll (\log N)^{-\frac{1}{4}\log\log\log\log N}.$$

This bound is quite extraordinary in the sense that it is by far superior to any bound known for the corresponding problem concerning arithmetic progressions. In particular, the best known bound for the existence of 3-term arithmetic progressions was very recently improved by Bourgain [Bou06] to

$$\alpha \ll (\log\log N)^2(\log N)^{-2/3},$$

and we refer the reader to Green and Tao [GT06b] for the currently best known bounds for progressions of length 4. For progressions of length $k \geq 5$, the best known bound is due to Gowers [Gow01] and of the form $(\log\log N)^{-c}$, where the constant c can be taken to be $2^{-2^{k+9}}$.

In fact, the bound in Theorem 1.2 is good enough to give us information about the existence of arithmetic structure in the prime numbers, which have asymptotic density $(\log N)^{-1}$. We cannot draw similar conclusions from the bounds on Roth's Theorem, although the existence of arithmetic progressions in the primes is now known by other methods [GT04].

In [Sár78a] Sárközy conjectures that $\alpha \ll N^{-1/2+\epsilon}$ for any positive ϵ. He also shows in Part II [Sár78b] of his impressive series of papers that $\alpha p^{1/2} > q(p)/2$ for all primes $p \equiv 1 \bmod 4$, where $q(p)$ is the least positive quadratic non-residue of p, so the conjecture would imply that $q(p) = O(p^{\epsilon})$ for all $p \equiv 1 \bmod 4$, which is believed to lie beyond the range of currently known techniques in analytic number theory.

Sárközy's conjecture should also be compared with the best known construction for this problem which is not surprisingly due to Ruzsa [Ruz84]. He constructs a subset of $[N]$ of density

$$\alpha \geq N^{-1/2(1-\log 7/\log 65)},$$

where the exponent is approximately equal to -0.266923.

The statement of Theorem 1.1 can, by very similar methods, be extended to polynomial structures other than the squares, more precisely, any polynomial that has an integer root. For example, it is true for $x^2 - 1$ (for a simple argument in the

spirit of [Gre02], see [Wol03]) but not x^2+1: since there are no squares congruent to 2 mod 3, the set of all multiples of 3 provides a counterexample. The general polynomial result is known as the *Bergelson-Leibmann Theorem*, and was first proved by ergodic theoretic methods [BL96]. Although these are extremely natural and elegant, no quantitative bounds can be obtained.

Let us also briefly mention that one can ask whether the set of differences of a dense set necessarily contains an element which is a prime minus 1. Again, the answer is yes and the interested reader is referred to [Sár78c]. Observe that this problem is of no interest for differences of the form $p-k$ with $k \neq 1$: If k is prime, the difference set always contains 0 which is of the form $p-k$. If k is composite, the set of all multiples of k is very dense and contains no differences of the form $p-k$. If $k=0$, we can take the multiples of any composite number to give us a dense counterexample. The methods of [PSS88] were recently applied to the shifted (by 1) primes by Lucier [Luc07], but the bounds are superseded by recent work of Ruzsa and Sanders [RS07]. We shall briefly discuss these matters in the final section.

Finally, let us remark that the corresponding problem for squares in sumsets was settled in [LOS82] by graph theoretic methods. In this case it is possible to find a set of density 11/32 whose sumset is square-free.

Let us first recall the comparatively simple iteration argument used by Green [Gre02] to tackle the question of square-free difference sets, which yields the bound $\alpha \ll (\log\log N)^{-1/11}$. It uses a standard *density increment* strategy: At the i^{th} step of this iteration argument, we have a set A_i of density α_i whose difference set is square-free. The latter property ensures the existence of a large Fourier coefficient, which in turn can be used to establish in a standard way that A_i has increased density on a long arithmetic progression with square common difference. After rescaling, we obtain a set A_{i+1} of increased density $\alpha_{i+1} \geq \alpha_i(1+\alpha_i^{12})$, whose difference set is again square-free. It is not difficult to see that if α were $\gg (\log\log N)^{-1/11}$ we could repeat this process until the density has increased beyond 1, which is clearly nonsense.

It has been shown in several instances that it can be more efficient to use a collection of large Fourier coefficients rather than a single one. This is what we shall refer to as the *energy increment* strategy, which originated in the work of Szemerédi [Sze90] in the late 80s and was also deployed around the same time by [HB87].

In order to obtain the radical improvement stated in Theorem 1.2, Pintz, Steiger and Szemerédi use such an energy increment argument, but in addition they employ a further iteration sitting inside the one just described, which aims to build up a very large collection of large Fourier coefficients. By the nature of the set of squares,

we should be able to locate these large Fourier coefficients near rationals with small denominator. Either we can increase the number of intervals supporting a large Fourier coefficient at each step of the iteration significantly, and we end up with large total L^2-mass (which gives a good bound on α by Parseval's Theorem), or we fail to do so at some point. Using combinatorial properties of the rational numbers with small denominators, the latter case implies a lower bound on the L^2-mass of Fourier coefficients near rationals with a specific (although unspecified) denominator, and as usual this allows us to pass down to a subprogression on which A has increased density.

Let us conclude the introduction by setting up our notation. Throughout the proof, we may assume that $\alpha \geq c\,(\log N)^{-c' \log\log\log\log N}$ for suitable constants c and c'. We shall use the letter A to denote the characteristic function of the set A, and for ease of notation we set $L = \log N, l = \log\log N$ and $\log_i N = \log\log ... \log N$, where the logarithm is always taken to base e. We also put $k = e^{2l}$ and $K = e^{l^2}$. Fourier analysis will be carried out on $\mathbb{Z}_N$ by defining the Fourier coefficient of a set $A \subseteq \mathbb{Z}_N$ at $t \in \widehat{\mathbb{Z}}_N$ via the formula

$$\widehat{A}(t) := \mathbb{E}_{x\in\mathbb{Z}_N} A(x)e(tx/N).$$

Also, write $I(a/q, \eta)$ for the interval of length η around a/q, and let

$$F_i(q,\eta) := \sum_{\substack{\frac{t}{N} \in \bigcup I(\frac{a}{q},\eta) \\ a\leq q,\ (a,q)=1}} |\widehat{A}_i(t)|^2,$$

that is, $F_i(q,\eta)$ is the sum of squares of Fourier coefficients near rationals with denominator q. *Parseval's Identity* takes the form

$$\sum_{t=1}^{N} |\widehat{A}(t)|^2 = \alpha,$$

which implies that $F_i(q,\eta)$ as defined above is bounded by α.

Finally, let us briefly outline the structure of the remainder of this chapter. Section 1.2 is devoted to describing the (by now pretty standard) energy increment iteration, which already gives some improvement over previously known bounds. Sections 1.3 and 3.2 contain the details of the inner iteration, while in Section 1.5 we will be concerned with working out bounds. After that we will be in a position to discuss the limitations of the method in Section 1.6. An appendix is included for readers who are not familiar with the intricacies of traditional circle method estimates for

the squares, although we do take some prior exposure to Fourier analysis for granted.

1.2 The Outer Iteration

At the step i^{th} of the *outer iteration* we are given a set $A_i \subseteq [N_i]$ of density α_i whose difference set is square-free. From now on we fix i, and dropping the index we write $A = A_i$, $\alpha = \alpha_i$, $N = N_i$ and $F(q,\eta) = F_i(q,\eta)$. Because we shall be working in $\mathbb{Z}_N$ and do not want to count square-difference that only exist modulo N, we set $N_1 := N/2$ and $\sigma^{-1} := \sqrt{N_1}$ and consider differences between the sets A and $A \cap [N_1]$. This is permissible because without loss of generality we may assume that A has density at least $\alpha/2$ on $[N_1]$. However, for convenience we shall not explicitly make the distinction between A and $A \cap [N_1]$ in this exposition.

We let the function S be defined by $S(x) = 2\sqrt{x/N_1}T(x)$, where T denotes the characteristic function of the set of squares less than N_1. Working with a weighted version S of the squares as defined above makes them uniformly distributed on $[N_1]$, a process which does not harm the validity of (1.1) but improves the major arc estimates for $\widehat{S}(t)$ significantly. This is discussed in more detail together with all Fourier estimates for S in Appendix A. Note that this strategy corresponds to replacing the characteristic function with the von-Mangoldt function in the corresponding problem for the primes, which is a completely standard procedure in analytic number theory.

Following the lines of the usual argument in the proof of Roth's Theorem, we can now regard A and T as subsets of $\mathbb{Z}_N$ and write

$$\mathbb{E}_{x\in\mathbb{Z}_N} A * -A(x)S(x) = 0. \tag{1.1}$$

Taking the Fourier transform and subtracting the trivial mode implies that

$$\sum_{t\neq 0} |\widehat{A}(t)|^2 |\widehat{S}(t)| \gg \alpha^2\sigma. \tag{1.2}$$

We shall see that Equation (1.2) implies that $\widehat{A}(t)$ takes rather large values rather frequently. By Hölder's Inequality, we can neglect those values of t for which $|\widehat{A}(t)|$ or $|\widehat{A}(t)|$ takes values at most α/K provided that $\alpha \gg K^{-2/5}$. Indeed, we have

$$\sum_{\text{these } t} |\widehat{A}(t)|^2 |\widehat{S}(t)| \ll \sup_{\text{these } t} |\widehat{A}(t)|^{1/3} \left(\sum_{t=1}^{N} |\widehat{A}(t)|^2\right)^{5/6} \left(\sum_{t=1}^{N} |\widehat{S}(t)|^6\right)^{1/6}.$$

The l^6-estimate for $\widehat{S}(t)$, which we have postponed to Lemma A.5, implies that this expression is bounded above by a small constant times $\alpha^2\sigma$. By a similar argument we can also neglect those values of t for which $|\widehat{S}(t)|$ is small. This is the case whenever t belongs to a set that is traditionally known as the *minor arcs*. It consists of those values of t for which t/N is close to a rational with large denominator, where "large" is determined by a parameter R defined in Appendix A. Indeed, for t/N close to rationals with denominator greater than R, Lemma A.4 implies that

$$\sum_{\text{these } t} |\widehat{A}(t)|^2|\widehat{S}(t)| \ll \frac{\alpha\sigma}{\sqrt{K/L}}.$$

This quantity is negligible provided that $\alpha \gg (K/L)^{-1/2}$. It follows that we need only consider those t for which $t/N \in I(a/q, (qQ)^{-1})$ for $q \leq R$, that is, the values of t on the *major arcs*.

Next we want to perform dyadic averaging over the remaining ranges of parameters to obtain a set of intervals on which A has large energy. For this purpose, we define for $1 \leq b \leq r \leq R$ with $(b, r) = 1$, the *A-special major arcs* as

$$\tau(b, r) = \left\{ t \neq 0 : \frac{t}{N} \in I\left(\frac{b}{r}, \frac{1}{rQ}\right), |\widehat{A}(t)| \geq \frac{\alpha}{K} \right\},$$

where $Q = N/K$ throughout. It turns out that we can bound the l^1-Fourier mass of the squares on the set $\tau(b, r)$ because we can usefully estimate the Fourier coefficients of S near rationals with small denominator.

Lemma 1.3. *Let $1 \leq b \leq r \leq R$ with $(b, r) = 1$. Then we have*

$$\sum_{t \in \tau(b,r)} |\widehat{S}(t)| \ll \frac{l^3\sigma}{\sqrt{r}}$$

with $\tau(b, r)$ defined as above.

Proof. This is another instance where we have to delve into the exponential sum estimates in the appendix. More precisely, we use Lemma A.1 and Lemma A.2 to obtain

$$\begin{aligned} N \sum_{t \in \tau(b,r)} |\widehat{S}(t)| &\ll \sum_{t \in \tau(b,r)} \left(\frac{\sqrt{\log r}}{\sqrt{r}} |F_S(t/N - b/r)| + \sqrt{r \log r}(1 + |t/N - b/r|N) \right) \\ &\ll \frac{\sqrt{\log r}}{\sqrt{r}} \sigma^{-1} \log K + \sqrt{r \log r} K^2, \end{aligned}$$

where by our choice of K the first term is bounded by $l^3\sigma/\sqrt{r}$ and the second term is clearly negligible. □

It follows easily from Lemma 1.3 and the preceding discussion that

$$\alpha^2\sigma \ll \sum_{r\leq R}\sum_{\substack{b\leq r\\(b,r)=1}}\sum_{t\in\tau(b,r)}|\widehat{A}(t)|^2|\widehat{S}(t)| \ll \sum_{r\leq R}\sum_{\substack{b\leq r\\(b,r)=1}}\sup_{t\in\tau(b,r)}|\widehat{A}(t)|^2\frac{l^3\sigma}{\sqrt{r}}.$$

Next we shall partition the set of relevant fractions b/r into sets of the form

$$\mathcal{L}_{X,V} = \left\{\frac{b}{r} : X < r \leq 2X,\ \frac{\alpha}{V} < \sup_{t\in\tau(b,r)}|\widehat{A}(t)| \leq \frac{2\alpha}{V}\right\}$$

for integers $X \leq R, V \leq K$. There are $\log R \log K$ of these sets. Hence there exist parameters $X \leq R, V \leq K$ such that

$$\frac{\alpha^2\sigma}{\log R\log K} \ll |\mathcal{L}_{X,V}|\frac{\alpha^2}{V^2}\frac{l^3\sigma}{\sqrt{X}},$$

which in turn immediately implies that

$$|\mathcal{L}_{X,V}| \gg \frac{V^2\sqrt{X}}{l^3\log R\log K}.$$

By definition, we also have the upper bound $|\mathcal{L}_{X,V}| \leq \alpha^{-2}XV^2\sup_{X<r\leq 2X}F(r,(rQ)^{-1})$, and it follows easily from Parseval's Identity that $|\mathcal{L}_{X,V}| \leq \alpha^{-1}V^2$. Putting everything together, we obtain a lower bound on the energy of A concentrated on Fourier modes near rationals with denominator of magnitude around X of the form

$$\sup_{X<r\leq 2X}F\left(r,\frac{1}{rQ}\right) \gg \frac{\alpha^3}{(l^3\log R\log K)^2}. \tag{1.3}$$

By our choice of the parameters R and K, we will always have $\log R = O(l^2) = \log K$, so that the denominator in (1.3) is always $O(l^{14})$. This bound will be useful in conjunction with the following standard lemma, which says that we can obtain a density increment of size about $F(q,(qQ)^{-1})$ on a progression of common difference q^2 and length at least $Q/(qL)$.

Lemma 1.4. *Let $q > 1, N' = \lfloor(\eta q^2 L)^{-1}\rfloor$, and let $A \subset [N]$ have density α. Then we can find a set $A' \subset [N']$ of density*

$$\alpha' \geq \alpha + \frac{F(q,\eta)}{8\alpha},$$

with the additional property that if $A - A$ was square-free, so is $A' - A'$.

Proof. We shall show that under the assumption that A has large Fourier mass near rationals with denominator q, A has large intersection with some translate of an arithmetic progression of common difference q^2 which is not too short. Let this progression be $P = \{q^2k : 1 \leq k \leq |P|\}$ with $|P| = N'$, and consider

$$J := \sum_{t=1}^{N} |\widehat{A * P}(t)|^2 = \mathbb{E}_x |A * P(x)|^2 = \mathbb{E}_x \left(\frac{|A \cap (P + x)|}{N} \right)^2, \tag{1.4}$$

which is the quantity we are trying to find a lower bound for. Now if $t/N \in I(a/q, \eta)$, then $q^2kt/N = aqk + O(\eta q^2|P|)$, so that $e(q^2kt/N) = 1 + O(L^{-1})$ and hence

$$|\widehat{P}(t)| = \frac{1}{N} \left| \sum_{k=1}^{|P|} e(q^2kt/N) \right| = \frac{|P|}{N} \left(1 + O\left(L^{-1}\right)\right).$$

It follows from this estimate and Equation (1.4) that

$$J = \sum_{t=1}^{N} |\widehat{A}(t)|^2 |\widehat{P}(t)|^2 \geq \alpha^2 \left(\frac{|P|}{N} \right)^2 \left(1 + O\left(L^{-1}\right)\right) \left(1 + \frac{F(q,\eta)}{\alpha^2}\right).$$

We therefore find that there exists an x such that

$$|A \cap (P + x)| \geq |P| \left(\alpha + \frac{F(q,\eta)}{8\alpha} \right),$$

and the lemma follows as stated. □

In the proof above we deliberately glossed over the fact that we need to ensure that P is not just a progression modulo N. This is easily achieved by discarding those translates that would split into two progression upon unravelling $\mathbb{Z}_N$, a procedure which results in a minor and ultimately insignificant loss in the density increment.

The argument so far shows that we can get a density increase of $F/8\alpha$ with $F \gg \alpha^3 l^{-14}$ at each step, and the length of the progression to which we scale after d steps is at least $N/(KRL)^d = \Omega(N/L^{cld})$, which means we can iterate $d \ll L/l^2$ times. This in turn gives rise to the condition $L/l^2 \ll \alpha^{-1} \log \alpha^{-1} l^{14}$, which results in a bound on the density of A of the form

$$\alpha \ll \frac{l^{17}}{L}. \tag{1.5}$$

For the benefit of readers familiar with the paper [PSS88], we point out that the

iteration argument presented in this section was originally phrased as a maximal counterexample. However, we believe that our presentation helps to align the part of the argument we have discussed so far with what follows. Using a further iteration, which we are about to describe in more detail, we shall be able to raise the exponent of the denominator in the bound (1.5) from 1 to a function of N tending very slowly to infinity.

1.3 The Inner Iteration

At the m^{th} step of what we from now on call the *inner iteration*, we inherit a set of large Fourier coefficients near rationals with denominator bounded by X_m, which we shall denote by

$$\mathcal{P}^{(m)}_{X_m,V_m} = \left\{ u : \frac{u}{N} \in I\left(\frac{a}{q}, \frac{m}{Q}\right),\ 1 \le a \le q \le X_m,\ (a,q) = 1,\ |\widehat{A}(u)| \ge \frac{\alpha}{V_m} \right\}.$$

Here X_m and V_m are the parameters maximizing the expression $|\mathcal{P}_{X,V}|V^{-2}$. Since trivially $\sup_{1\le X,1\le V} |\mathcal{P}_{X,V}|V^{-2} \ge 1$, we may assume that $V_m \le X_m$. Let $\mathcal{R}^{(m)}_{X_m,V_m}$ be the corresponding set of centres of intervals a/q.

For fixed $u \in \mathcal{P}^{(m)}$, write $A_u(x) = e(ux/N)A(x)$. We now consider the expression

$$\mathbb{E}_{x\in\mathbb{Z}_N} A * -A_u(x)S(x),$$

which is again zero under the assumption that $A-A$ is square-free. Observe that this is the point where we make definite use of that fact that $A - A$ contains no squares, as opposed to relatively few. It follows that for fixed $u \in \mathcal{P}^{(m)}$, we have

$$\sum_{t\neq 0} |\widehat{A}(t)\widehat{A}(u+t)\widehat{S}(t)| \gg \frac{\alpha^2\sigma}{V_m}.$$

Just as before, by a simple use of Hölder's Inequality we can neglect values of t for which one of $\widehat{A}(t), \widehat{A}(u+t)$ or $\widehat{S}(t)$ is small in modulus. Indeed, if t is such that $|\widehat{A}(t)|$ or $|\widehat{A}(u+t)|$ is at most α/K, then the contribution from these t is bounded by

$$\sup_{\text{these } t} |\widehat{A}(t)|^{1/3} \left(\sum_{t=1}^{N} |\widehat{A}(t)|^2\right)^{5/6} \left(\sum_{t=1}^{N} |\widehat{S}(t)|^6\right)^{1/6} \ll \frac{\alpha^2\sigma}{\alpha^{5/6}K^{1/3}},$$

which is negligible compared with $\alpha^2\sigma V_m^{-1}$ provided that $\alpha \gg (X_m K^{-1/3})^{6/5}$. On the other hand, minor and major arc estimates for $\widehat{S}(t)$ imply that for t to be taken into

account, t/N needs to be close to a rational with small denominator $r < X_{m+1}/X_m$. For if t were near a rational with denominator between $X_{m+1}/X_m = X_m^3 X_1$ and K, which corresponds to the *fairly major arcs*, Lemma A.3 yields

$$\sum_{\text{these } t} |\widehat{A}(t)\widehat{A}(u+t)\widehat{S}(t)| \leq \sup_{X_{m+1}/X_m < r \leq K} \frac{\alpha\sigma}{r^{1/3}} \leq \frac{\alpha\sigma}{X_m X_1^{1/3}},$$

which is bounded above by $\alpha^2\sigma V_m^{-1}$ provided that $\alpha \gg X_1^{-1/3}$. Similarly, for t on the *minor arcs*, that is, for denominators r satisfying $K \leq r \leq Q$, we have by Lemma A.4 that

$$\sum_{\text{these } t} |\widehat{A}(t)\widehat{A}(u+t)\widehat{S}(t)| \leq \frac{\alpha\sigma}{\sqrt{K/L}},$$

which is bounded above by $\alpha^2\sigma V_m^{-1}$ provided that $\alpha \gg X_m(K/L)^{-1/2}$.

We again perform dyadic averaging over the remaining ranges of parameters in order to obtain a set of intervals which supports a large proportion of the total energy of A. To this end, for $u \in \mathcal{P}^{(m)}$, $1 \leq b \leq r \leq Q$ and $(b,r) = 1$, we define the *A-special major arcs with respect to u* as

$$\tau(b,r,u) = \left\{ t \neq 0 : \frac{t}{N} \in I\left(\frac{b}{r}, \frac{1}{rQ}\right), |\widehat{A}(t)| \geq \frac{\alpha}{K}, |\widehat{A}(u+t)| \geq \frac{\alpha}{K} \right\}.$$

With this definition it follows by averaging that for each $u \in \mathcal{P}^{(m)}$,

$$\begin{aligned} \frac{\alpha^2\sigma}{V_m} &\ll \sum_{r \leq \frac{X_{m+1}}{X_m}} \sum_{\substack{b \leq r \\ (b,r)=1}} \sum_{t \in \tau(b,r,u)} |\widehat{A}(t)\widehat{A}(u+t)\widehat{S}(t)| \\ &\ll \sum_{r \leq \frac{X_{m+1}}{X_m}} \sum_{\substack{b \leq r \\ (b,r)=1}} \sup_{t \in \tau(b,r,u)} |\widehat{A}(t)| \sup_{t \in \tau(b,r,u)} |\widehat{A}(u+t)| \sum_{t \in \tau(b,r,u)} |\widehat{S}(t)| \end{aligned}$$

But as before, Lemma 1.3 implies that $\sum_{t \in \tau(b,r,u)} |\widehat{S}(t)| \ll l^3 \sigma r^{-1/2}$. Hence for each $u \in \mathcal{P}^{(m)}$, we can choose integers V_u, W_u and X_u satisfying $1 \leq V_u \leq K, 1 \leq W_u \leq K, 1 \leq X_u \leq X_{m+1}/X_m$ such that the set $\mathcal{L}_u$ given by

$$\left\{ \frac{b}{r} : X_u < r \leq 2X_u,\ \frac{\alpha}{V_u} < \sup_{t \in \tau(b,r,u)} |\widehat{A}(t)| \leq \frac{2\alpha}{V_u},\ \frac{\alpha}{W_u} < \sup_{t \in \tau(b,r,u)} |\widehat{A}(u+t)| \leq \frac{2\alpha}{W_u} \right\}$$

has cardinality at least

$$\frac{V_u W_u \sqrt{X_u}}{l^3 (\log K)^2 V_m \log(X_{m+1}/X_m)}.$$

When splitting the sum into dyadic ranges, the number of choices for V_u, W_u, X_u is bounded above by $(\log K)^2 \log(X_{m+1}/X_m)$. Hence we can make the same choice of V_u, W_u, X_u for at least $|\mathcal{P}^{(m)}|/(\log K)^2 \log(X_{m+1}/X_m)$ different $u \in \mathcal{P}^{(m)}$. Let us denote the set of such u by $\widetilde{\mathcal{P}}^{(m)}$, using parameters $\widetilde{V}, \widetilde{W}, \widetilde{X}$.

Observe that for each $u \in \widetilde{\mathcal{P}}^{(m)}$, we have found an element $w \in \tau(b, r, u)$ with the property that $|\widehat{A}(u+w)| \geq \alpha/\widetilde{W}$. We would like to count the number of distinct $u+w$ in order to determine whether we can achieve a significant increase in total L^2-mass by adding all points of the form $u+w$ to the set of u where we had already located a large Fourier coefficient. For the sake of clarity, the technical details of this counting argument as well as the rough explanation for why we should expect it to work have been postponed until the next section. Writing $F^{(m)} := \sup_{\widetilde{X} < r \leq 2\widetilde{X}} F(r, (rQ)^{-1})$ and $\tau := \sup_{q \leq X_m} \tau(q)$, we find by Lemma 1.5 below that there are at least

$$\frac{\alpha^2}{F^{(m)}} \frac{|\widetilde{\mathcal{P}}^{(m)}_{X_m, V_m}|}{\tau^4 \widetilde{X} \log \widetilde{X} \widetilde{V}^2} \inf |\mathcal{L}_u|^2$$

different $u+w$ with the property that $(u+w)/N \in I(c/s, (m+1)/Q)$ and $\alpha/\widetilde{W} < |\widehat{A}(u+w)| \leq 2\alpha/\widetilde{W}$, a quantity bounded below by

$$\frac{\alpha^2 \widetilde{W}^2}{F^{(m)}} \frac{|\widetilde{\mathcal{P}}^{(m)}_{X_m, V_m}|}{V_m^2} \left(\tau^4 (\log K)^{4+2} (\log(X_{m+1}/X_m))^{2+1+1} (l^3)^2\right)^{-1}.$$

This allows us to define the set $\mathcal{P}^{(m+1)}_{X_{m+1}, V_{m+1}}$, where we choose parameters $V_{m+1} := \widetilde{W}$ and $X_{m+1} := X_m^4 X_1$. (We briefly remark that before passing to the next step of the iteration, we may need to reset them so they correspond to the maximum of the expression $|\mathcal{P}_{X,V}| V^{-2}$). Thus we have just shown that

$$\frac{|\mathcal{P}^{(m+1)}_{X_{m+1}, V_{m+1}}|}{V_{m+1}^2} \geq \frac{|\mathcal{P}^{(m)}_{X_m, V_m}|}{V_m^2} \frac{\alpha^2 \eta}{F^{(m)}},$$

where we have set the parameter η to be

$$\eta := (\tau^4 (\log K)^6 (\log(X_{m+1}/X_m))^4 l^6)^{-1}.$$

When choosing our main parameters X_1 and M in Section 1.5 we shall ensure that $\eta = \Omega(L^{-1/2})$ always. Now we are faced with two possible cases.

- Suppose $\alpha^2 \eta / F^{(m)} \geq L^{1/2}$ for all $m \leq M$, then by Parseval we have $\alpha \leq L^{-M/2}$, and we will have completed the proof without leaving the inner iteration, simply by building up a collection of Fourier coefficients with large total Fourier energy.

- Otherwise, there exists $m \leq M$ such that $\alpha^2\eta/F^{(m)} \leq L^{1/2}$, i.e. $F^{(m)} \geq \alpha^2\eta/L^{1/2}$. This lower bound on $F^{(m)}$ enables us to pass down to a subprogression on which A has increased density.

Note that the density increase in the second case is significantly greater than the one we obtained in Section 1.2.

1.4 Combinatorics of Rational Numbers

This section is dedicated to an explanation of why one would expect to be able to locate a significant number of large Fourier modes $u + w$, where u was already large, under the assumption that both u/N and w/N lie near rationals with small denominator. Looking back on our work in the preceding section and recalling that that u/N is assumed to lie in the interval $I(a/q, m/Q)$, we had established that for all $u \in \widetilde{\mathcal{P}}$ there is a large set $\mathcal{L}_u$ defined by

$$\left\{ \frac{b}{r} : X_u < r \leq 2X_u,\ \frac{\alpha}{V_u} < \sup_{t\in\tau(b,r,u)} |\widehat{A}(t)| \leq \frac{2\alpha}{V_u},\ \frac{\alpha}{W_u} < \sup_{t\in\tau(b,r,u)} |\widehat{A}(u+t)| \leq \frac{2\alpha}{W_u} \right\}.$$

Since $X_m \leq X_1^{4^m}$, the intervals $I(\frac{a}{q}, \frac{m}{Q})$ are disjoint whenever $m \leq Q/X_1^{4^m}$ (which is yet another condition we have to satisfy when choosing our parameters in Section 1.5), so that counting the number of distinct $u + w$ is equivalent to counting the number of distinct $a/q + b/r$.

In lowest terms, $\frac{a}{q} + \frac{b}{r}$ can be expressed as a gigantic fraction of the form

$$\frac{(ar' + bq')/f}{(r'q'd)/f},$$

where $d = (q, r), q = dq', r = dr'$ and $f = (ar' + bq', d)$. We immediately note that $(q', r') = 1$ and $(f, q') = (f, r') = 1$.

For fixed a/q we associate a pair $\{d, f\}$ with every $b/r \in \mathcal{L}_{a/q} = \mathcal{L}_u$, where u is the unique element in $\widetilde{\mathcal{P}}$ associated with a/q. For each a/q, there exists a pair $\{d, f\}$ associated with lots of $b/r \in \mathcal{L}_{a/q}$, say all $b/r \in \widetilde{\mathcal{L}}_{a/q}$. By averaging, we find that $|\widetilde{\mathcal{L}}_{a/q}| \geq \tau(q)^{-2}|\mathcal{L}_{a/q}|$. Similarly, for each q, there exists $\{d, f\}$ associated with lots of a/q, say all a/q with $a \in \widetilde{A}(q)$. Again, by averaging, we must have $|\widetilde{A}(q)| \geq \tau(q)^{-2}|A(q)|$, while $\sum_{q\leq X_m} |A(q)| = \widetilde{\mathcal{P}}$.

Now fix c/s, and count the number of solutions to the equation

$$\frac{c}{s} = \frac{a}{q} + \frac{b}{r} \tag{1.6}$$

with $a/q \in \widetilde{\mathcal{Q}} = \{a/q : q \leq \widetilde{X}, a \in \widetilde{A}(q)\}$ and $b/rin\widetilde{\mathcal{L}}_{a/q}$. We write $s = q'r'e$ and then choose f, which immediately determines d, q, r. It is clear that $a \mod q'$ is determined by $ar' + bq' = cf$. Denote the number of distinct $a \mod f$ by $r(q)$. By the Chinese Remainder Theorem, we deduce that there are $r(q)q/q'f$ choices for a, which in turn automatically determines b. We conclude that the number of solutions to (1.6) is at most $\sum_{q=q'r'e} \sum_{f \leq d \leq r \leq \widetilde{X}} r(efq')d/f$, so we have an upper bound on the number of solutions provided we have an upper bound for $r(q)$.

Fix q, and the associated popular pair $\{d, f\}$. The crucial observation is that $\widetilde{\mathcal{L}}_{a_1/q}$ and $\widetilde{\mathcal{L}}_{a_2/q}$ are disjoint if $a_1 \not\equiv a_2 \mod f$. Then

$$r(q) \inf |\widetilde{\mathcal{L}}_{a/q}| \leq |\bigcup_{a \in \widetilde{A}(q)} \widetilde{\mathcal{L}}_{a/q}| \leq \left|\left\{\frac{b}{r} : \frac{b}{r} \in \cup \mathcal{L}_{a/q}\right\}\right| \leq \sum_{r \leq R, d|r} \left|\left\{b : \frac{b}{r} \in \cup \mathcal{L}_{a/q}\right\}\right| \leq \frac{\widetilde{X}}{d} B_r,$$

where B_r is the number of distinct numerators b such that $b/r \in \cup \mathcal{L}_{a/q}$, so that the number of solutions to (1.6) is bounded above by $\widetilde{X} \log \widetilde{X} B_r / \inf |\widetilde{\mathcal{L}}_{a/q}|$. It follows immediately that the number of distinct $a/q + b/r$ with $a/q \in \widetilde{\mathcal{R}}^{(m)}$ and $b/r \in \mathcal{L}_{a/q}$ is bounded below by

$$\frac{\sum_{q \leq \widetilde{X}} \sum_{a \in \widetilde{A}(q)} |\widetilde{\mathcal{L}}_{a/q}|}{\#\text{ solutions to (1.6)}} \gg \sum_{q \leq X_m} |\widetilde{A}(q)| \inf |\widetilde{\mathcal{L}}_{a/q}| \left(\frac{\widetilde{X} \log \widetilde{X} B_r}{\inf |\widetilde{\mathcal{L}}_{a/q}|}\right)^{-1} \gg \frac{\inf |\widetilde{\mathcal{L}}_{a/q}|^2 |\widetilde{\mathcal{R}}^{(m)}|}{\tau^4 \widetilde{X} \log \widetilde{X} B_r}.$$

But B_r and the L^2-mass of Fourier coefficients near rationals with denominator r are by sheer definition related via the inequality $F^{(m)}(r, (rQ)^{-1}) \geq \alpha B_r / \widetilde{V}^2$. Thus we have proved the following statement about the additive behaviour of rational numbers with small denominators.

Lemma 1.5. *Let $\widetilde{\mathcal{R}}^{(m)}$ be the set of centres of intervals corresponding to $\widetilde{\mathcal{P}}^{(m)}$, with parameters $\widetilde{V}, \widetilde{W}, \widetilde{X}$ as specified in the preceding section. For $u \in \widetilde{\mathcal{P}}$, let the set $\mathcal{L}_u$ be defined by*

$$\left\{\frac{b}{r} : X_u < r \leq 2X_u,\ \frac{\alpha}{V_u} < \sup_{t \in \tau(b,r,u)} |\widehat{A}(t)| \leq \frac{2\alpha}{V_u},\ \frac{\alpha}{W_u} < \sup_{t \in \tau(b,r,u)} |\widehat{A}(u+t)| \leq \frac{2\alpha}{W_u}\right\}.$$

Then the number of distinct $a/q + b/r$ with $a/q \in \widetilde{\mathcal{R}}^{(m)}$ and $b/r \in \mathcal{L}_{a/q}$ is at least

$$\frac{|\widetilde{\mathcal{R}}^{(m)}|}{\widetilde{V}^2}\frac{\alpha^2}{F^{(m)}}\frac{\inf|\mathcal{L}_{a/q}|^2}{\tau^4\widetilde{X}\log\widetilde{X}},$$

where $F^{(m)} := \sup_{\widetilde{X}<r\le 2\widetilde{X}} F^{(m)}(r,(rQ)^{-1})$ and $\tau := \sup_{q\le X_m}\tau(q)$.

Let us summarize what this section has achieved: We were trying to assess whether we could increase the L^2-mass of the large Fourier coefficients, and for this purpose we counted how many of them there are. That is, we counted the number of distinct new intervals with centres $a/q + b/r$. The obvious way of doing this is to divide the number of all relevant fractions of the form $a/q + b/r$, that is $\sum_{\text{appropriate } a,q}|\mathcal{L}_{a/q}|$, by the number of solutions to $c/s = a/q + b/r$ with $a/q \in \mathcal{R}^{(m)}, b/r \in \mathcal{L}_{a/q}$. The inequality $F^{(m)}(r,(rQ)^{-1}) \ge \alpha B_r/\widetilde{V}^2$ immediately gives us the desired connection between the L^2-mass near denominator r and the number of distinct numerators b such that $b/r \in \cup\mathcal{L}_{a/q}$, B_r. The upshot is that either we have lots of these for some r, that is, $B := \sup_r B_r$ is large, in which case we have (by definition) large L^2-mass near a specific denominator and we can scale. If not, that is if B is small, then by the above counting argument we obtain lots of new intervals so that the total L^2 mass increases significantly.

1.5 Working Out Bounds

To make the combinatorial counting arguments in the preceding section work, we need $\eta = \Omega(L^{-1/2})$ as remarked above, that is, we require that

$$\tau := \sup_{q\le X_m}\tau(q) \ll L^c \quad\text{and}\quad \log\frac{X_{m+1}}{X_m} \le l^2 \tag{1.7}$$

for some small constant c. It is a well-known number-theoretic fact that

$$\log\tau(X_m) \ll \frac{\log X_m}{\log\log X_m} \ll \frac{4^m\log X_1}{m+\log\log X_1} \le cl,$$

and we therefore choose

$$M := \frac{1}{2}\log_4 N \text{ as well as } X_1 := L^{(\log_3 N)^{1/4}}$$

in order to satisfy both conditions in (1.7). We can then check that the remaining conditions are satisfied. It was necessary to have

$$\alpha \gg X_1^{-1/3}, \quad \alpha \gg \frac{X_m}{\sqrt{K/L}} \quad \text{and} \quad \alpha \gg \left(\frac{X_m}{K^{1/3}}\right)^{6/5}$$

to ensure that we could neglect the contributions from the fairly major and the minor arcs, and $m \leq Q/X_1^{4^m}$ to force the intervals $I(a/q, m/Q)$ to be disjoint. We have made no attempts to optimize the constants involved here.

1.6 Remarks

The method which we have discussed was extended to cover the case of kth powers in [BPPS94]. Only minor modifications to the argument are necessary, and these occur almost exclusively through the Hardy-Littlewood type estimates in the appendix.

It should also be clear that similar progress can be made for polynomial differences such as $x^2 - 1$. Very recently, Lucier [Luc07] applied the method to the shifted primes to obtain a bound of

$$\left(\frac{(\log_3 N)^4}{\log\log N}\right)^{\log_5 N}$$

on the density of the set which avoids the set of all $p = 1$, p a prime. However, it should be noted that the currently best-known bound for this problem obtained in [RS07] is of the form

$$\exp(-c\sqrt[4]{\log N})$$

and does not use this technique. Indeed, at least assuming GRH it is relatively straightforward to obtain a density increment of size a constant times α in the case of primes, which cannot be improved by the technique described in this chapter. (For comparison, the straightforward density increase in the case of squares is of size α^2, and can be improved to α using combinatorics of rationals.)

Given the fact that the application to the primes is slightly bogus, it would be very interesting to find a genuinely new and useful application of this method.

Acknowledgements. The author would like to thank Endre Szemerédi for helpful comments and discussions.

Chapter 2

The True Complexity of a Linear System

2.1 Introduction

This chapter includes joint work with Tim Gowers. Section 2.3 has been submitted as [GW07b]. Section 4.1 is a precursor to the forthcoming paper [GW07a].

In this chapter we look for conditions that are sufficient to guarantee that a subset A of a finite Abelian group G contains the "expected" number of linear configurations of a given type. The simplest non-trivial result of this kind is the well-known fact that if G has odd order, A has density α and all Fourier coefficients of the characteristic function of A are significantly smaller than α (except the one at zero, which equals α), then A contains approximately $\alpha^3|G|^2$ triples of the form $(a, a+d, a+2d)$. This is "expected" in the sense that a random set A of density α has approximately $\alpha^3|G|^2$ such triples with very high probability.

More generally, it was shown in [Gow01] (in the case $G = \mathbb{Z}_N$ for N prime, but the proof generalizes) that a set A of density α has about $\alpha^k|G|^2$ arithmetic progressions of length k if the characteristic function of A is almost as small as it can be, given its density, in a norm that is now called the U^{k-1}-norm. Green and Tao [GT06a] have found the most general statement that follows from the technique used to prove this result, introducing a notion that they call the *complexity* of a system of linear forms. They prove that if A has almost minimal U^{k+1}-norm then it has the expected number of linear configurations of a given type, provided that the associated complexity is at most k. The main result of this chapter is that the converse is not true: in particular there are certain systems of complexity 2 that are controlled by the U^2-norm, whereas the result of Green and Tao requires the stronger hypothesis of U^3-control.

We say that a system of m linear forms $L_1, \dots, L_m$ in d variables has *true complexity* k if k is the smallest positive integer such that, for any set A of density α and almost minimal U^{k+1}-norm, the number of d-tuples $(x_1, \dots, x_d)$ such that $L_i(x_1, \dots, x_d) \in A$ for every i is approximately $\alpha^m |G|^d$. We conjecture that the true complexity k is the smallest positive integer s for which the functions $L_1^{s+1}, \dots, L_m^{s+1}$ are linearly independent.

Using the "quadratic Fourier analysis" of Green and Tao we prove this conjecture in Section 2.3 in the case where the complexity of the system (in Green and Tao's sense) is 2, $s = 1$ and G is the group $\mathbb{F}_p^n$ for some fixed odd prime p. Section 4.1 is devoted to obtaining improved bounds for this problem. Finally, a closely related result in ergodic theory was recently proved independently by Leibman [Lei07]. We discuss the connections between his result and ours in Section 2.5.

Let us now turn to a more detailed description of the problem. Suppose A is a subset of a finite Abelian group G and let $\alpha = |A|/|G|$ be the *density* of A. We say that A is *uniform* if it has one of several equivalent properties, each of which says in its own way that A "behaves like a random set". For example, writing A for the characteristic function of the set A, we can define the convolution $A * A$ by the formula

$$A * A(x) = \mathbb{E}_{y+z=x} A(y)A(z),$$

where the expectation is with respect to the uniform distribution over all pairs $(y, z) \in G^2$ such that $y + z = x$; one of the properties in question is that the variance of $A * A$ should be small. As we have already remarked above, if this is the case and G has odd order, then it is easy to show that A contains approximately $\alpha^3 |G|^2$ triples of the form $(x, x + d, x + 2d)$. Indeed, these triples are the solutions (x, y, z) of the equation $x + z = 2y$, and

$$\mathbb{E}_{x+z=2y} A(x)A(y)A(z) = \mathbb{E}_y A * A(2y)A(y).$$

The mean of the function $A * A$ is α^2, so if the variance is sufficiently small, then the right-hand side is approximately $\alpha^2 \mathbb{E}_y A(y) = \alpha^3$. This is a probabilistic way of saying that the number of solutions of $x + z = 2y$ inside A is approximately $\alpha^3 |G|^2$, which is what we would expect if A was a random set with elements chosen independently with probability α.

An easy generalization of the above argument shows that, given any linear equation in G of the form

$$c_1 x_1 + c_2 x_2 + \cdots + c_m x_m = 0,$$

for suitable fixed coefficients $c_1, c_2, ..., c_m$, the number of solutions in A is approxi-

mately $\alpha^m|G|^{m-1}$. Roughly speaking, you can choose $x_3, \dots, x_m$ in A however you like, and if A is sufficiently uniform then the number of ways of choosing x_1 and x_2 to lie in A and satisfy the equation will almost always be roughly $\alpha^2|G|$. By "suitable" we mean that there are certain divisibility problems that must be avoided. For example, if G is the group $\mathbb{F}_2^n$, $x+z=2y$ and x belongs to A, then z belongs to A for the trivial reason that it equals x. Throughout this chapter we shall consider groups of the form $\mathbb{F}_p^n$ for some prime p and assume that p is large enough for such problems not to arise.

When $k \geq 4$, uniformity of a set A does not guarantee that A contains approximately $\alpha^k|G|^2$ arithmetic progressions of length k. For instance, there are examples of uniform subsets of $\mathbb{Z}_N$ that contain significantly more, or even significantly fewer than, the expected number of four-term progressions [Gow06b]. It was established in [Gow98] that the appropriate measure for dealing with progressions of length 4 is a property known as *quadratic uniformity*: sets which are sufficiently quadratically uniform contain roughly the correct number of four-term progressions. We shall give precise definitions of higher-degree uniformity in the next section, but for now let us simply state the result, proved in [Gow01] in the case $G = \mathbb{Z}_N$, that if A is uniform of degree $k-2$, then A contains approximately $\alpha^k|G|^2$ arithmetic progressions of length k. Moreover, if A is uniform of degree j for some $j < k-2$, then it does *not* follow that A must contain approximately $\alpha^k|G|^2$ arithmetic progressions of length k.

The discrepancy between k and $k-2$ seems slightly unnatural until one reformulates the statement in terms of solutions of equations. We can define an arithmetic progression of length k either as a k-tuple of the form $(x, x+d, \dots, x+(k-1)d)$ or as a solution $(x_1, x_2, \dots, x_k)$ to the system of $k-2$ equations $x_i - 2x_{i+1} + x_{i+2} = 0$, $i = 1, 2, \dots, k-2$. In all the examples we have so far discussed, we need uniformity of degree precisely k in order to guarantee approximately the expected number of solutions of a system of k equations. It is tempting to ask whether this is true in general.

However, a moment's reflection shows that it is not. For example, the system of equations $x_1 - 2x_2 + x_3 = 0$, $x_4 - 2x_5 + x_6 = 0$ has about $\alpha^6|G|^4$ solutions in a uniform set, since the two equations are completely independent. This shows that a sensible conjecture must take account of how the equations interact with each other.

A more interesting example is the system that consists of the $\binom{m}{3}$ equations $x_{ij}+x_{jk} = x_{ik}$ in the $\binom{m}{2}$ unknowns x_{ij}, $1 \leq i < j \leq m$. These equations are not all independent, but one can of course choose an independent subsystem of them. It is not hard to see that there is a bijection between solutions of this system of equations where every

x_{ij} belongs to A and m-tuples $(x_1, \dots, x_m)$ such that $x_j - x_i \in A$ whenever $i < j$. Now one can form a bipartite graph with two vertex sets equal to G by joining x to y if and only if $y - x \in A$. It is well-known that if A is uniform, then this bipartite graph is quasirandom. The statement that every $x_j - x_i$ belongs to A can be reformulated to say that $(x_1, \dots, x_m)$ form a clique in an m-partite graph that is built out of quasirandom pieces derived from A. A "counting lemma" from the theory of quasirandom graphs then implies easily that the number of such cliques is approximately $\alpha^{\binom{m}{2}}|G|^m$. So uniformity of degree 1 is sufficient to guarantee that there are about the expected number of solutions to this fairly complicated system of equations.

In their recent work on configurations in the primes, Green and Tao [GT06a] analysed the arguments used to prove the above results, which are fairly simple and based on repeated use of the Cauchy-Schwarz inequality. They isolated the property that a system of equations, or equivalently a system of linear forms, must have in order for degree-k uniformity to be sufficient for these arguments to work, and called this property *complexity*. Since in this chapter we shall have more than one notion of complexity, we shall sometimes call their notion *Cauchy-Schwarz complexity*, or *CS-complexity* for short.

Definition 2.1. *Let $\mathcal{L} = (L_1, ..., L_m)$ be a system of m linear forms in d variables. For $1 \leq i \leq m$ and $s \geq 0$, we say that $\mathcal{L}$ is* s-complex *at i if one can partition the $m-1$ forms $\{L_j : j \neq i\}$ into $s+1$ classes such that L_i does not lie in the linear span of any of these classes. The* Cauchy-Schwarz complexity *(or CS-complexity) of $\mathcal{L}$ is defined to be the least s for which the system is s-complex at i for all $1 \leq i \leq m$, or ∞ if no such s exists.*

To get a feel for this definition, let us calculate the complexity of the system $\mathcal{L}$ of k linear forms $x, x + y, \dots, x + (k-1)y$. Any two distinct forms $x + iy$ and $x + jy$ in $\mathcal{L}$ contain x and y in their linear span. Therefore, whichever form L we take from $\mathcal{L}$, if we wish to partition the others into classes that do not contain L in their linear span, then we must take these classes to be singletons. Since we are partitioning $k-1$ forms, this tells us that the minimal s is $k-2$. So $\mathcal{L}$ has complexity $k-2$.

Next, let us briefly look at the system $\mathcal{L}$ of $\binom{m}{2}$ forms $x_i - x_j$ $(1 \leq i < j \leq m)$ that we discussed above. If L is the form $x_i - x_j$ then no other form $L' \in \mathcal{L}$ involves both x_i and x_j, so we can partition $\mathcal{L} \setminus \{L\}$ into the forms that involve x_i (which therefore do not involve x_j) and the forms that do not involve x_i. Since neither class includes L in its linear span, the complexity of $\mathcal{L}$ is at most 1. When $m \geq 3$ it can also be shown to be at least 1.

It follows from Green and Tao's result that if A is sufficiently uniform and $\mathcal{L} = (L_1, ..., L_m)$ has complexity at most 1, then A contains approximately the expected number of m-tuples of the form $(L_1(x_1, \dots, x_d), \dots, L_m(x_1, \dots x_d))$. (If the forms are defined over $\mathbb{Z}^d$, then this number is $\alpha^m |G|^d$.)

Notice that this statement adequately explains all the cases we have so far looked at in which uniformity implies the correct number of solutions. It is thus quite natural to conjecture that Green and Tao's result is tight. That is, one might guess that if the complexity $\mathcal{L}$ is greater than 1 then there exist sets A that do not have the correct number of images of $\mathcal{L}$.

But is this correct? Let us look at what is known in the other direction, by discussing briefly the simplest example that shows that uniform sets in $\mathbb{Z}_N$ do not have to contain the correct number of arithmetic progressions of length 4. (Here we are taking N to be some large prime.) Roughly speaking, one takes A to be the set of all x such that x^2 mod N is small. Then one makes use of the identity

$$x^2 - 3(x+d)^2 + 3(x+2d)^2 - (x+3d)^2 = 0$$

to prove that if x, $x+d$ and $x+2d$ all lie in A, then $x+3d$ is rather likely to lie in A as well, because $(x+3d)^2$ is a small linear combination of small elements of $\mathbb{Z}_N$. This means that A has "too many" progressions of length 4. (Later, we shall generalize this example and make it more precise.)

The above argument uses the fact that the squares of the linear forms x, $x+d$, $x+2d$ and $x+3d$ are linearly dependent. Later, we shall show that if $\mathcal{L}$ is *any* system of linear forms whose squares are linearly dependent, then essentially the same example works for $\mathcal{L}$. This gives us a sort of "upper bound" for the set of systems $\mathcal{L}$ that have approximately the right number of images in any uniform set: because of the above example, we know that the squares of the forms in any such system $\mathcal{L}$ must be linearly independent.

And now we arrive at the observation that motivated this project: the "upper bound" just described does not coincide with the "lower bound" of Green and Tao. That is, there are systems of linear forms of complexity greater than 1 with squares that are linearly independent. One of the simplest examples is the system $(x, y, z, x+y+z, x+2y-z, x+2z-y)$. Another, which is translation-invariant (in the sense that if you add a constant to everything in the configuration, you obtain another configuration of the same type), is $(x, x+y, x+z, x+y+z, x+y-z, x+z-y)$. Both these examples have complexity 2, but it is not hard to produce examples with arbitrarily high complexity.

In the light of such examples, we are faced with an obvious question: which systems of linear forms have roughly the expected number of images in any sufficiently uniform set? We conjecture that the correct answer is given by the "upper bound"—that is, that square independence is not just necessary but also sufficient. When the group G is $\mathbb{F}_p^n$ for a fixed prime p, we prove this conjecture for systems of complexity 2. This includes the two examples above, and shows that having Cauchy-Schwarz complexity at most 1 is not a necessary condition, even if it is a natural sufficient one.

However, the proof is much deeper for systems of complexity 2. Although the statement of our result is completely linear, we use "quadratic Fourier analysis", recently developed by Green and Tao [GT05a], to prove it, and it seems that we are forced to do so. Thus, it appears that Cauchy-Schwarz complexity captures the systems for which an easy argument exists, while square independence captures the systems for which the result is true.

Very recently, and independently, Leibman [Lei07] described a similar phenomenon in the ergodic-theoretic context. In Section 2.5 of this chapter we shall briefly outline how his results relate to ours.

So far, we have concentrated on uniform sets. However, in the next section we shall define higher-degree uniformity and formulate a more complete conjecture, which generalizes the above discussion in a straightforward way. Green and Tao proved that a system of Cauchy-Schwarz complexity k has approximately the correct number of images in a set A if A is sufficiently uniform of degree $k+1$. Once again, it seems that this is not the whole story, and that the following stronger statement should be true: a linear system $\mathcal{L} = (L_1, \dots, L_m)$ has the right number of images in any set A that is sufficiently uniform of degree k if and only if the functions L_i^{k+1} are linearly independent. The reason we have not proved this is that the natural generalization of our existing argument would have to use an as yet undeveloped general "polynomial Fourier analysis", which is known only in the quadratic case. However, it is easy to see how our arguments would generalize if such techniques were available, which is compelling evidence that our conjecture (which we will state formally in a moment) is true.

2.2 Uniformity Norms and True Complexity

As promised, let us now give a precise definition of higher-degree uniformity. We begin by defining a sequence of norms, known as *uniformity norms*.

Definition 2.2. *Let G be a 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*

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

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

These norms were first defined in [Gow01] (in the case where G is the group $\mathbb{Z}_N$). Of particular interest in this chapter will be the U^2-norm and the U^3-norm. The former can be described in many different ways. The definition above expresses it as the fourth root of the average of

$$f(x)\overline{f(x+h)f(x+h')}f(x+h+h')$$

over all triples (x, h, h'). It is not hard to show that this average is equal to $\|f * f\|_2^2$, and also to $\|\widehat{f}\|_4^4$. (These identities depend on appropriate normalizations—we follow the most commonly used convention of taking averages in physical space and sums in frequency space.)

We shall call a function f *c-uniform* if $\|f\|_{U^2} \leq c$ and *c-quadratically uniform* if $\|f\|_{U^3} \leq c$. We shall often speak more loosely and describe a function as uniform if it is c-uniform for some small c, and similarly for higher-degree uniformity. We remark here that if $j \leq k$ then $\|f\|_{U^j} \leq \|f\|_{U^k}$, so c-uniformity of degree k implies c-uniformity of all lower degrees.

If A is a subset of an Abelian group G and the density of A is α, then we say that A is uniform of degree k if it is close in the U^k-norm to the constant function α. More precisely, we define the *balanced function* $f(x) = A(x) - \alpha$ and say that A is *c-uniform of degree k* if $\|f\|_{U^k} \leq c$.

The following theorem is essentially Theorem 3.2 in [Gow01]. (More precisely, in that paper the theorem was proved for the group $\mathbb{Z}_N$, but the proof is the same.)

Theorem 2.3. *Let $k \geq 2$ and let G be a finite Abelian group such that there are no non-trivial solutions to the equation $jx = 0$ for any $1 \leq j < k$. Let $c > 0$ and let $f_1, f_2, \dots, f_k$ be functions from G to $\mathbb{C}$ such that $\|f_i\|_\infty \leq 1$ for every i. Then*

$$\left|\mathbb{E}_{x,y \in G} f_1(x) f_2(x+y) \dots f_k(x+(k-1)y)\right| \leq \|f_k\|_{U^{k-1}}.$$

It follows easily from this result that if A is a set of density α and A is c-uniform for sufficiently small c, then A contains approximately $\alpha^k |G|^2$ arithmetic progressions

of length k. Very briefly, the reason for this is that we are trying to show that the average

$$\mathbb{E}_{x,y}A(x)A(x+y)\dots A(x+(k-1)y)$$

is close to α^k. Now this average is equal to

$$\mathbb{E}_{x,y}A(x)A(x+y)\dots f(x+(k-1)y)+\alpha\mathbb{E}_{x,y}A(x)A(x+y)\dots A(x+(k-2)y).$$

The first of these terms is at most c, by Theorem 2.3, and the second can be handled inductively. The bound we obtain in this way is $c(1+\alpha+\dots+\alpha^{k-1})\leq kc$.

We can now state formally Green and Tao's generalization in terms of CS-complexity in the case where G is the group $\mathbb{Z}_N$, which is implicit in [GT06a].

Theorem 2.4. *Let N be a prime, let $f_1,\dots,f_m$ be functions from $\mathbb{Z}_N$ to $[-1,1]$, and let $\mathcal{L}$ be a linear system of CS-complexity k consisting of m forms in d variables. Then, provided $N\geq k$,*

$$\left|\mathbb{E}_{x_1,\dots,x_d\in\mathbb{Z}_N}\prod_{i=1}^m f(L_i(x_1,\dots,x_d))\right|\leq\min_i\|f_i\|_{U^{k+1}}.$$

Just as in the case of arithmetic progressions, it follows easily that if A is a subset of G of density α, then the probability, given a random element $(x_1,\dots,x_d)\in G^d$, that all the m images $L_i(x_1,\dots,x_d)$ lie in A is approximately α^m. (The inductive argument depends on the obvious fact that if $\mathcal{L}$ has complexity at most k then so does any subsystem of $\mathcal{L}$.)

Green and Tao proved the above theorem because they were investigating which linear configurations can be found in the primes. For that purpose, they in fact needed a more sophisticated "relative" version of the statement. Since the proof of the version we need here is simpler (partly because we are discussing systems of complexity at most 2, but much more because we do not need a relative version), we give it for the convenience of the reader. This is another result where the proof is essentially the same for all Abelian groups, give or take questions of small torsion. Since we need it in the case $G=\mathbb{F}_p^n$, we shall just prove it for this group. The reader should bear in mind that for this group, one should understand linear independence of a system of forms as independence over $\mathbb{F}_p$ when one is defining complexity (and also square-independence).

The first step of Green and Tao's proof was to put an arbitrary linear system into a convenient form for proofs. Given a linear form L in d variables $x_1,\dots,x_d$, let us define the *support* of L to be the set of j such that L depends on x_j. That is,

if $L(x_1, \dots, x_d) = \lambda_1 x_1 + \cdots + \lambda_d x_d$ then the support of L is $\{i : \lambda_i \neq 0\}$. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a system of linear forms and let the support of L_i be σ_i for each i. Then $\mathcal{L}$ is said to be in *s-normal form* if it is possible to find subsets $\tau_i \subset \sigma_i$ for each i with the following two properties.

(i) Each τ_i has cardinality at most $s + 1$.

(ii) If $i \neq j$ then τ_i is not a subset of σ_j.

If a linear system $\mathcal{L}$ is in s-normal form, then it has complexity at most s. Indeed, if τ_i has r elements $\{i_1, \dots, i_r\}$, then one can partition the remaining forms into r sets $\mathcal{L}_1, \dots, \mathcal{L}_r$ in such a way that no form in $\mathcal{L}_h$ uses the variable x_{i_h}. Since L_i *does* use the variable x_{i_h} it is not in the linear span of $\mathcal{L}_h$.

The converse of this statement is false, but Green and Tao prove that every linear system of complexity s can be "extended" to one that is in s-normal form. This part of the proof is the same in both contexts, so we do not reproduce it. All we need to know here is that if we prove Theorem 2.4 for systems in normal form then we have it for general systems.

Just to illustrate this, consider the obvious system associated with arithmetic progressions of length 4, namely $(x, x + y, x + 2y, x + 3y)$. This is not in 2-normal form, because the support of the first form is contained in the supports of the other three. However, the system $(-3x - 2y - z, -2x - y + w, -x + z + 2w, y + 2z + 3w)$ *is* in 2-normal form (since the supports have size 3 and are distinct) and its images are also uniformly distributed over all arithmetic progressions of length 4 (if we include degenerate ones).

Now let us prove Theorem 2.4 when $k = 2$. Without loss of generality we may assume that $\mathcal{L}$ is in 2-normal form at 1, and that it is the only form using all three variables $x_1 = x, x_2 = y$ and $x_3 = z$. We use the shorthand $h(x, y, z) = f(L_1(x_1, x_2, ..., x_d))$, and denote by $b(x, y)$ any general bounded function in two variables x and y. It is then possible to rewrite

$$\mathbb{E}_{x_1,\dots,x_d \in \mathbb{F}_p^n} \prod_{i=1}^{m} f(L_i(x_1, ..., x_d))$$

as

$$\mathbb{E}_{x_4,x_5,\dots,x_d} \mathbb{E}_{x,y,z} h(x, y, z) b(x, y) b(y, z) b(x, z).$$

Here, the functions h and b depend on the variables $x_4, \dots, x_d$ but we are suppressing this dependence in the notation.

Estimating the expectation over (x, y, z) is a well-known argument from the theory of quasirandom hypergraphs. (See for instance Theorem 4.1 in [Gow06a].) First, we apply Cauchy-Schwarz and use the boundedness of b to obtain an upper bound of

$$(\mathbb{E}_{x,y}(\mathbb{E}_z h(x,y,z)b(x,z)b(y,z))^2)^{1/2}.$$

Expanding out the square and rearranging yields

$$(\mathbb{E}_{y,z,z'}b(y,z)b(y,z')\mathbb{E}_x h(x,y,z)h(x,y,z')b(x,z)b(x,z'))^{1/2},$$

and by a second application of Cauchy-Schwarz we obtain an upper bound of

$$(\mathbb{E}_{y,z,z'}(\mathbb{E}_x h(x,y,z)h(x,y,z')b(x,z)b(x,z'))^2)^{1/4}.$$

A second round of interchanging summation followed by a third application of Cauchy-Schwarz gives us an upper bound of

$$(\mathbb{E}_{x,x',z,z'}(\mathbb{E}_y h(x,y,z)h(x,y,z')h(x',y,z)h(x',y,z'))^2)^{1/8}.$$

This expression equals the "octahedral norm" of the function $h(x, y, z)$—a hypergraph analogue of the U^3-norm. Because for fixed $x_4, \ldots, x_d$, h depends only on the linear expression $L_1(x, y, z)$, a simple change of variables can be used to show that it is in fact equal to $\|f\|_{U^3}$.

Now all that remains is to take the expectation over the remaining variables and the proof is complete. It is also not hard to generalize to arbitrary k, but this we leave as an exercise to the reader.

Now, as we stated earlier, Theorem 2.4 does not settle the question of which systems are controlled by which degrees of uniformity. Accordingly, we make the following definition.

Definition 2.5. *Let $\mathcal{L}$ be a system of m distinct linear forms $L_1, L_2, \ldots, L_m$ in d variables. The* true complexity *of $\mathcal{L}$ is the smallest k with the following property. For every $\epsilon > 0$ there exists $\delta > 0$ such that if G is any finite Abelian group and $f : G \to \mathbb{C}$ is any function with $\|f\|_\infty \leq 1$ and $\|f\|_{U^{k+1}} \leq \delta$, then*

$$\left|\mathbb{E}_{x_1,\ldots,x_d \in G} \prod_{i=1}^{m} f(L_i(x_1, \ldots, x_d))\right| \leq \epsilon.$$

The main conjecture of this chapter is now simple to state precisely.

Conjecture 2.6. *The true complexity of a system of linear forms $\mathcal{L} = (L_1, \dots, L_m)$ is equal to the smallest k such that the functions L_i^{k+1} are linearly independent.*

In the next section, we shall prove this conjecture in the simplest case that is not covered by the result of Green and Tao, namely the case when $k = 1$ and $\mathcal{L}$ has CS-complexity 2. All other cases would require a more advanced form of polynomial Fourier analysis than the quadratic Fourier analysis that is so far known, but we shall explain why it will almost certainly be possible to generalize our argument once such a theory is developed.

2.3 True Complexity for Vector Spaces over Finite Fields

We shall now follow the course that is strongly advocated by Green [Gre05a] and restrict attention to the case where G is the group $\mathbb{F}_p^n$, where p is a fixed prime and n tends to infinity. The reason for this is that it makes many arguments technically simpler than they are for groups with large torsion such as $\mathbb{Z}_N$. In particular, one can avoid the technicalities associated with Bohr sets. These arguments can then almost always be converted into more complicated arguments for $\mathbb{Z}_N$. (In the forthcoming paper [GW07a], we give a different proof for the case $\mathbb{F}_p^n$ and carry out the conversion process. That proof is harder than the proof here but gives significantly better bounds and is easier to convert.)

We begin this section with the easier half of our argument, showing that if $\mathcal{L}$ is a system of linear forms $(L_1, \dots, L_m)$ and if there is a linear dependence between the squares of these forms, then the true complexity of $\mathcal{L}$ is greater than 1. This part can be proved almost as easily for $\mathbb{Z}_N$, but we shall not do so here.

2.3.1 Square-Independence is Necessary

Let us start by briefly clarifying what we mean by square-independence of a linear system $\mathcal{L} = (L_1, \dots, L_m)$. When the group G is $\mathbb{Z}_N$, then all we mean is that the functions L_i^2 are linearly independent, but when it is $\mathbb{F}_p^n$, then this definition does not make sense any more. Instead, we ask for the quadratic forms $L_i^T L_i$ to be linearly independent. If $L_i(x_1, \dots, x_d) = \sum_r \gamma_r^{(i)} x_r$, then $L_i^T L_i(x_1, \dots, x_d) = \sum_r \sum_s \gamma_r^{(i)} \gamma_s^{(i)} x_r x_s$. Therefore, what we are interested in is linear independence of the matrices $\Gamma_{rs}^{(i)} = \gamma_r^{(i)} \gamma_s^{(i)}$ over $\mathbb{F}_p$. (Note that in the case of $\mathbb{Z}_N$, this is equivalent to independence of the functions L_i^2.)

Theorem 2.7. *Let $\mathcal{L} = (L_1, \dots, L_m)$ be a system of linear forms in d variables and suppose that the quadratic forms $L_i^T L_i$ are linearly dependent over $\mathbb{F}_p$. Then there exists $\epsilon > 0$ such that for every $\delta > 0$ there exists n and a set $A \subset \mathbb{F}_p^n$ with the following two properties.*

(i) A is δ-uniform of degree 1.

(ii) If $\mathbf{x} = (x_1, \dots, x_d)$ is chosen randomly from $(\mathbb{F}_p^n)^d$, then the probability that $L_i(\mathbf{x})$ is in A for every i is at least $\alpha^m + \epsilon$, where α is the density of A.

In other words, the true complexity of $\mathcal{L}$ is at least 2.

For the proof we require the following standard lemma, which says that certain Gauss sums are small. A proof can be found in [Gre05b], for example.

Lemma 2.8. *Suppose that $q : \mathbb{F}_p^n \to \mathbb{F}_p$ is a quadratic form of rank r. That is, suppose that $q(x) = x^T M x + b^T x$ for some matrix M of rank r and some vector $b \in \mathbb{F}_p^n$. Then*

$$|\mathbb{E}_{x \in \mathbb{F}_p^n} \omega^{q(x)}| \leq p^{-r/2},$$

with equality if $b = 0$. In particular,

$$|\mathbb{E}_{x \in \mathbb{F}_p^n} \omega^{\eta x^T x}| \leq p^{-n/2}$$

for any non-zero $\eta \in \mathbb{F}_p$.

Proof of Theorem 2.7. Let A be the set $\{x \in \mathbb{F}_p^n : x^T x = 0\}$. Then the characteristic function of A can be written as

$$A(x) = \mathbb{E}_u \omega^{u x^T x},$$

where $\omega = \exp(2\pi i/p)$ and the expectation is taken over $\mathbb{F}_p$. Let us now take any square-*independent* system $\mathcal{L} = (L_1, \dots, L_m)$ of linear forms in $\mathbf{x} = (x_1, \dots, x_d)$ and estimate the expectation $\mathbb{E}_{\mathbf{x}} \prod_i A(L_i(\mathbf{x}))$.

Using the formula for $A(x)$, we can rewrite this expectation as

$$\mathbb{E}_{\mathbf{x} \in (\mathbb{F}_p^n)^d} \mathbb{E}_{u_1, \dots, u_m \in \mathbb{F}_p} \omega^{\sum_i u_i L_i(\mathbf{x})^T L_i(\mathbf{x})}.$$

We can break this up into p^m expectations over $\mathbf{x}$, one for each choice of $u_1, \dots, u_m$.

If the u_i are all zero, then the expectation over $\mathbf{x}$ is just the expectation of the constant function 1, so it is 1. Otherwise, since the quadratic forms $L_i^T L_i$ are linearly independent, the sum $\sum_i u_i L_i(\mathbf{x})^T L_i(\mathbf{x})$ is a non-zero quadratic form $q(x) = \sum_{i,j} \gamma_{ij} x_i^T x_j$.

Without loss of generality, there exists j such that $\gamma_{1j} \neq 0$. If in addition $\gamma_{11} = 0$, then for every choice of $x_2, \dots, x_d$ we can write $q(x)$ in the form $r^T x_1 + z$, where $r = \sum_j \gamma_{1j} x_j$ and z depends on $x_2, \dots, x_d$ only. This is a non-constant linear function of x_1 except when $\sum_j \gamma_{1j} x_j = 0$. Since not every γ_{1j} is zero, this happens with probability p^{-n}. Therefore, $|\mathbb{E}_{\mathbf{x}} \omega^{q(\mathbf{x})}| \leq p^{-n}$ in this case. If $\gamma_{11} \neq 0$, then this same function has the form $\gamma_{11} x_1^T x_1 + r^T x_1$ for some element $r \in \mathbb{F}_p^n$ (which depends on $x_2, \dots, x_d$). In this case, Lemma 2.29 implies that the expectation is at most $p^{-n/2}$.

Since the probability that $u_1 = \cdots = u_m = 0$ is p^{-m}, this shows that

$$\left| \mathbb{E}_{\mathbf{x} \in (\mathbb{F}_p^n)^d} \prod_i A(L_i(\mathbf{x})) - p^{-m} \right| \leq p^{-n/2}.$$

Applying this result in the case where $\mathcal{L}$ consists of the single form x, we see that the density of A differs from p^{-1} by at most $p^{-n/2}$. Therefore, we have shown that for this particular set A, square-independence of $\mathcal{L}$ guarantees approximately the "correct" probability that every $L_i(\mathbf{x})$ lies in A.

This may seem like the opposite of what we were trying to prove, but in fact we have almost finished, for the following simple reason. If we now take $\mathcal{L}$ to be an arbitrary system $(L_1, \dots, L_m)$ of linear forms, then we can choose from it a maximal square-independent subsystem. Without loss of generality this subsystem is $(L_1, \dots, L_l)$. Then all the quadratic forms $L_i^T L_i$ with $i > l$ are linear combinations of $L_1^T L_1, \dots, L_l^T L_l$, so a sufficient condition for every $L_i^T L_i(\mathbf{x})$ to be zero is that it is zero for every $i \leq l$. But this we know happens with probability approximately p^{-l} by what we have just proved. Therefore, if $\mathcal{L}$ is not square-independent, then A^m contains "too many" m-tuples of the form $(L_1(\mathbf{x}), \dots, L_m(\mathbf{x}))$. □

2.3.2 A Review of Quadratic Fourier Analysis

We shall now turn our attention to the main result of this chapter, which states that if $\mathcal{L}$ has CS-complexity at most 2 and is square-independent, then the true complexity of $\mathcal{L}$ is at most 1. We begin with a quick review of quadratic Fourier analysis for functions defined on $\mathbb{F}_p^n$. Our aim in this review is to give precise statements of the results that we use in our proof. The reader who is prepared to use quadratic Fourier analysis as a black box should then find that this chapter is self-contained.

So far in our discussion of uniformity, we have made no mention of Fourier analysis at all. However, at least for the U^2-norm, there is a close connection. Let f be a complex-valued function defined on a finite Abelian group G. If γ is a character on G, the

Fourier coefficient $\widehat{f}(\gamma)$ is defined to be $\mathbb{E}_x f(x)\gamma(x)$. The resulting Fourier transform satisfies the convolution identity $\widehat{f * g} = \widehat{f}\,\widehat{g}$, Parseval's identity $\|\widehat{f}\|_2 = \|f\|_2$ and the inversion formula $f(x) = \sum_\gamma \widehat{f}(\gamma)\gamma(-x)$. (The second and third identities depend on the correct choice of normalization: $\|f\|_2^2$ is defined to be $\mathbb{E}_x|f(x)|^2$, whereas $\|\widehat{f}\|_2^2$ is defined to be $\sum_\gamma |\widehat{f}(\gamma)|^2$. That is, as mentioned earlier, we take averages in G and sums in $\widehat{G}$.) It follows that $\|f\|_{U^2}^4 = \|\widehat{f}\|_4^4$, since both are equal to $\|f * f\|_2^2$.

It is often useful to split a function f up into a "structured" part and a uniform part. One way of doing this is to let K be the set of all characters γ for which $|\widehat{f}(\gamma)|$ is larger than some δ and to write $f = f_1 + f_2$, where $f_1 = \sum_{\gamma\in K} \widehat{f}(\gamma)\gamma(-x)$ and $f_2 = \sum_{\gamma\notin K} \widehat{f}(\gamma)\gamma(-x)$. If $\|f\|_\infty \leq 1$, (as it is in many applications), then Parseval's identity implies that $|K| \leq \delta^{-2}$, and can also be used to show that $\|f_2\|_{U^2} \leq \delta^{1/2}$. That is, K is not too large, and f_2 is $\delta^{1/2}$-uniform.

When G is the group $\mathbb{F}_p^n$, the characters all have the form $x \mapsto \omega^{r^T x}$. Notice that this character is constant on all sets of the form $\{x : r^T x = u\}$, and that these sets partition $\mathbb{F}_p^n$ into p affine subspaces of codimension 1. Therefore, one can partition $\mathbb{F}_p^n$ into at most $p^{|K|}$ affine subspaces of codimension $|K|$ such that f_1 is constant on each of them. This is the sense in which f_1 is "highly structured".

The basic aim of quadratic Fourier analysis is to carry out a similar decomposition for the U^3-norm. That is, given a function f, we would like to write f as a sum $f_1 + f_2$, where f_1 is "structured" and f_2 is *quadratically* uniform. Now this is a stronger (in fact, much stronger) property to demand of f_2, so we are forced to accept a weaker notion of structure for f_1.

Obtaining any sort of structure at all is significantly harder than it is for the U^2-norm, and results in this direction are much more recent. The first steps were taken in [Gow98] and [Gow01] for the group $\mathbb{Z}_N$ in order to give an analytic proof of Szemerédi's theorem. The structure of that proof was as follows: Theorem 2.3 can be used to show that if a set A is sufficiently uniform of degree $k-2$, then it must contain an arithmetic progression of length k. Then an argument that is fairly easy when $k = 3$ but much harder when $k \geq 4$ can be used to show that if A is *not* c-uniform of degree k, then it must have "local correlation" with a function of the form $\omega^{\phi(x)}$, where $\omega = \exp 2\pi i/N$ and ϕ is a polynomial of degree d. "Local" in this context means that one can partition $\mathbb{Z}_N$ into arithmetic progressions of size N^η (for some η that depends on c and k only) on a large proportion of which one can find such a correlation.

This was strong enough to prove Szemerédi's theorem, but for several other applications the highly local nature of the correlation is too weak. However, in the quadratic

case, this problem has been remedied by Green and Tao [GT05a]. In this case, the obstacle to "globalizing" the argument is that a certain globally-defined bilinear form that occurs in the proof of [Gow01] is not symmetric, and thus does not allow one to define a corresponding globally-defined quadratic form. (In the context of $\mathbb{Z}_N$, "global" means something like "defined on a proportional-sized Bohr set". For $\mathbb{F}_p^n$ one can take it to mean "defined everywhere".) Green and Tao discovered an ingenious "symmetry argument" that allows one to replace the bilinear form by one that *is* symmetric, and this allowed them to prove a quadratic structure theorem for functions with large U^3-norm that is closely analogous to the linear structure theorem that follows from conventional Fourier analysis.

An excellent exposition of this structure theorem when the group G is a vector space over a finite field can be found in [Gre05b]. This contains proofs of all the background results that we state here.

Recall that in the linear case, we called f_1 "structured" because it was constant on affine subspaces of low codimension. For quadratic Fourier analysis, we need a quadratic analogue of the notion of a decomposition of $\mathbb{F}_p^n$ into parallel affine subspaces of codimension d_1. In order to define such a decomposition, one can take a surjective linear map $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ and for each $a \in \mathbb{F}_p^{d_1}$ one can set V_a to equal $\Gamma_1^{-1}(\{a\})$. If we want to make this idea quadratic, we should replace the linear map Γ_1 by a "quadratic map" Γ_2, which we do in a natural way as follows. We say that a function $\Gamma_2 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_2}$ is *quadratic* if it is of the form $x \mapsto (q_1(x), \dots, q_{d_2}(x))$, where $q_1, \dots, q_{d_2}$ are quadratic forms on $\mathbb{F}_p^n$. Then, for each $b \in \mathbb{F}_p^{d_2}$ we define W_b to be $\{x \in \mathbb{F}_p^n : \Gamma_2(x) = b\}$.

In [GT05b], Green and Tao define $\mathcal{B}_1$ to be the algebra generated by the sets V_a and $\mathcal{B}_2$ for the finer algebra generated by the sets $V_a \cap W_b$. They call $\mathcal{B}_1$ a *linear factor of complexity* d_1 and $(\mathcal{B}_1, \mathcal{B}_2)$ a *quadratic factor of complexity* (d_1, d_2). This is to draw out a close analogy with the "characteristic factors" that occur in ergodic theory.

These definitions give us a suitable notion of a "quadratically structured" function—it is a function f_1 for which we can find a linear map $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ and a quadratic map $\Gamma_2 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_2}$ such that d_1 and d_2 are not too large and f_1 is constant on the sets $V_a \cap W_b$ defined above. This is equivalent to saying that f_1 is measurable with respect to the algebra $\mathcal{B}_2$, and also to saying that $f_1(x)$ depends on $(\Gamma_1(x), \Gamma_2(x))$ only.

The quadratic structure theorem of Green and Tao implies that a bounded function f defined on $\mathbb{F}_p^n$ can be written as a sum $f_1 + f_2$, where f_1 is quadratically structured in the above sense, and $\|f_2\|_{U^3}$ is small. In [GT05b] the result is stated explicitly for

$p = 5$, but this is merely because of the emphasis placed on 4-term progressions. The proof is not affected by the choice of p (as long as it stays fixed).

In the statement below, which is taken from [GT05b], we write $\mathbb{E}(f|\mathcal{B}_2)$ for the conditional expectation, or averaging projection, of f. That is, if $X = V_a \cap W_b$ is an atom of $\mathcal{B}_2$ and $x \in X$, then $\mathbb{E}(f|\mathcal{B}_2)(x)$ is the average of f over X. Since the function $\mathbb{E}(f|\mathcal{B}_2)$ is constant on the sets $V_a \cap W_b$, it is quadratically structured in the sense that interests us.

Theorem 2.9. *Let p be a fixed prime, let $\delta > 0$ and suppose that $n > n_0(\delta)$ is sufficiently large. Given any function $f : \mathbb{F}_p^n \to [-1, 1]$, there exists a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ of complexity at most $((4\delta^{-1})^{3C_0+1}, (4\delta^{-1})^{2C_0+1})$ together with a decomposition*

$$f = f_1 + f_2,$$

where

$$f_1 := \mathbb{E}(f|\mathcal{B}_2) \quad \textit{and} \quad \|f_2\|_{U^3} \leq \delta.$$

The absolute constant C_0 can be taken to be 2^{16}.

As it stands, the above theorem is not quite suitable for applications, because technical problems arise if one has to deal with quadratic forms of low rank. (Notice that so far we have said nothing about the quadratic forms q_i—not even that they are distinct.) Let $\Gamma_2 = (q_1, \dots, q_k)$ be a quadratic map and for each i let β_i be the symmetric bilinear form corresponding to q_i: that is, $\beta_i(x, y) = (q_i(x + y) - q_i(x) - q_i(y))/2$. We shall say that Γ_2 *is of rank at least r* if the bilinear form $\sum_i \lambda_i \beta_i$ has rank at least r whenever $\lambda_1, \dots, \lambda_{d_2}$ are elements of $\mathbb{F}_p$ that are not all zero. If Γ_2 is used in combination with some linear map Γ_1 to define a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$, then we shall also say that this quadratic factor has rank at least r.

Just to clarify this definition, let us prove a simple lemma that will be used later.

Lemma 2.10. *Let β be a symmetric bilinear form of rank r on $\mathbb{F}_p^n$ and let W be a subspace of $\mathbb{F}_p^n$ of codimension d_1. Then the rank of the restriction of β to W is at least $r - 2d_1$.*

Proof. Let $V = \mathbb{F}_p^n$. For every subspace W of V, let us write $W^\perp$ for the subspace

$$\{v \in V : \beta(v, w) = 0 \text{ for every } w \in W\}.$$

The rank of β is just the codimension of $V^\perp$, and equals r by hypothesis. Now let W have codimension d_1, and let Y be a complement for W, which will therefore have

dimension d_1. Then $V^\perp = W^\perp \cap Y^\perp$ and $Y^\perp$ has dimension at least $n - d_1$. It follows easily that

$$r = \text{codim} V^\perp \leq \text{codim} W^\perp + \text{codim} Y^\perp \leq \text{codim} W^\perp + d_1,$$

which implies that the codimension of $W^\perp$ is at least $r - d_1$. Hence the codimension of $W^\perp$ inside W is at least $r - 2d_1$. □

We are now in a position to state the version of the structure theorem that we shall be using. It can be read out of (but is not explicitly stated in) [Gre05b] and [GT05b].

Theorem 2.11. *Let p be a fixed prime, let $\delta > 0$, let $r : \mathbb{N} \to \mathbb{N}$ an arbitrary function (which may depend on δ) and suppose that $n > n_0(r, \delta)$ is sufficiently large. Then given any function $f : \mathbb{F}_p^n \to [-1, 1]$, there exists $d_0 = d_0(r, \delta)$ and a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ of rank at least $r(d_1 + d_2)$ and complexity at most $(d_1, d_2), d_1, d_2 \leq d_0$, together with a decomposition*

$$f = f_1 + f_2 + f_3,$$

where

$$f_1 := \mathbb{E}(f|\mathcal{B}_2), \quad \|f_2\|_2 \leq \delta \quad \textit{and} \quad \|f_3\|_{U^3} \leq \delta.$$

Note that $\mathbb{E}f_1 = \mathbb{E}f$. In particular $\mathbb{E}f_1 = 0$ whenever f is the balanced function of a subset of $\mathbb{F}_p^n$. It can be shown that f_1 is uniform whenever f is uniform: roughly speaking, the reason for this is that $\mathbb{E}(f|\mathcal{B}_1)$ is approximately zero and the atoms of $\mathcal{B}_2$ are uniform subsets of the atoms of $\mathcal{B}_1$. However, we shall not need this fact.

We shall apply Theorem 2.11 when r is the function $d \mapsto 2md + C$ for a constant C. Unfortunately, ensuring that factors have high rank is an expensive process: even for this modest function the argument involves an iteration that increases d_0 exponentially at every step. For this reason we have stated the theorem in a qualitative way. A quantitative version would involve a tower-type bound.

2.3.3 Square-Independence is Sufficient

We now have the tools we need to show that square-independence coupled with CS-complexity 2 is sufficient to guarantee the correct number of solutions in uniform sets. The basic idea of the proof is as follows. Given a set $A \subset \mathbb{F}_p^n$ of density α, we first replace it by its balanced function $f(x) = A(x) - \alpha$. Given a square-independent

linear system $\mathcal{L}$ of complexity at most 2, our aim is to show, assuming that $\|f\|_{U^2}$ is sufficiently small, that

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d}\prod_{i=1}^m f(L_i(\mathbf{x}))$$

is also small. (Once we have done that, it will be straightforward to show that the same average, except with A replacing f, is close to α^m.) In order to carry out this estimate, we first apply the structure theorem to decompose f as $f_1+f_2+f_3$, where f_1 is quadratically structured, f_2 is small in L_2 and f_3 is quadratically uniform. This then allows us to decompose the product into a sum of 3^m products, one for each way of choosing f_1, f_2 or f_3 from each of the m brackets. If we ever choose f_2, then the Cauchy-Schwarz inequality implies that the corresponding term is small, and if we ever choose f_3 then a similar conclusion follows from Theorem 2.4. Thus, the most important part of the proof is to use the linear uniformity and quadratic structure of f_1 to prove that the product

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d}\prod_{i=1}^m f_1(L_i(\mathbf{x}))$$

is small. This involves a calculation that generalizes the one we used to prove Theorem 2.7. The main step is the following lemma, where we do the calculation in the case where the linear factor $\mathcal{B}_1$ is trivial.

To understand its significance, let us briefly think about what happens when we map a 4-term progression to $(\mathbb{F}_p^{d_2})^4$ using the quadratic map Γ_2. Because of the relation between the squares of the forms defining the 4-term progression, we find that there is roughly the expected number of progression in the pre-image $(\Gamma_2^{-1}(b^{(1)}),\Gamma_2^{-1}(b^{(2)}),\Gamma_2^{-1}(b^{(3)}),\Gamma_2^{-1}(b^{(4)}))\subseteq(\mathbb{F}_p^n)^4$ whenever the $b^{(i)}\in\mathbb{F}_p^{d_2}$ satisfy $b^{(1)}-3b^{(2)}+3b^{(3)}-b^{(4)}=0$, and precisely no progressions otherwise. For a general square-independent linear system, it turns out that the pre-images are roughly uniformly distributed independent of any relations between the $b^{(i)}$s.

Lemma 2.12. *Let $\mathcal{L}=(L_1,\dots,L_m)$ be a square-independent system of linear forms and let $\Gamma_2=(q_1,\dots,q_{d_2})$ be a quadratic map from $\mathbb{F}_p^n$ to $\mathbb{F}_p^{d_2}$ of rank at least r. Let $\phi_1,\dots,\phi_m$ be linear maps from $(\mathbb{F}_p^n)^d$ to $\mathbb{F}_p^{d_2}$ and let $b_1,\dots,b_m$ be elements of $\mathbb{F}_p^{d_2}$. Let $\mathbf{x}=(x_1,\dots,x_d)$ be a randomly chosen element of $(\mathbb{F}_p^n)^d$. Then the probability that $\Gamma_2(L_i(\mathbf{x}))=\phi_i(\mathbf{x})+b_i$ for every i differs from p^{-md_2} by at most $p^{-r/2}$.*

Proof. Let Λ be the set of all $m\times d_2$ matrices $\lambda=(\lambda_{ij})$ over $\mathbb{F}_p$ and let us write $\phi_i=(\phi_{i1},\dots,\phi_{id_2})$ and $b_i=(b_{i1},\dots,b_{id_2})$ for each i. The probability we are interested

in is the probability that $q_j(L_i(\mathbf{x})) = \phi_{ij}(\mathbf{x}) + b_{ij}$ for every $i \leq m$ and every $j \leq d_2$. This equals

$$\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\lambda\in\Lambda}\prod_{i=1}^{m}\prod_{j=1}^{d_2}\omega^{\lambda_{ij}(q_j(L_i(\mathbf{x}))-\phi_{ij}(\mathbf{x})-b_{ij})},$$

since if $q_j(L_i(\mathbf{x})) = \phi_{ij}(\mathbf{x}) + b_{ij}$ for every i and j, then the expectation over λ is 1, and otherwise if we choose i and j such that $q_j(L_i(\mathbf{x})) \neq \phi_{ij}(\mathbf{x}) + b_{ij}$ and consider the expectation over λ_{ij} while all other entries of λ are fixed, then we see that the expectation over λ is zero.

We can rewrite the above expectation as

$$\mathbb{E}_{\lambda\in\Lambda}\mathbb{E}_{\mathbf{x}}\omega^{\sum_{i,j}\lambda_{ij}(q_j(L_i(\mathbf{x}))-\phi_{ij}(\mathbf{x})-b_{ij})}.$$

If $\lambda = 0$, then obviously the expectation over $\mathbf{x}$ is 1. This happens with probability p^{-md_2}. Otherwise, for each i let us say that the coefficients of L_i are $c_{i1}, \dots, c_{id}$. That is, let $L_i(\mathbf{x}) = \sum_{u=1}^{d} c_{iu}x_u$. Then

$$q_j(L_i(\mathbf{x})) = \sum_{u,v} c_{iu}c_{iv}\beta_j(x_u, x_v),$$

where β_j is the bilinear form associated with q_j. Choose some j such that λ_{ij} is non-zero for at least one i. Then the square-independence of the linear forms L_i implies that there exist u and v such that $\sum_i \lambda_{ij}c_{iu}c_{iv}$ is not zero.

Fix such a j, u and v and do it in such a way that $u = v$, if this is possible. We shall now consider the expectation as x_u and x_v vary with every other x_w fixed. Notice first that

$$\sum_{i,j}\lambda_{ij}q_j(L_i(\mathbf{x})) = \sum_{i,j}\sum_{t,w}\lambda_{ij}c_{it}c_{iw}\beta_j(x_t, x_w).$$

Let us write β_{tw} for the bilinear form $\sum_{i,j}\lambda_{ij}c_{it}c_{iw}\beta_j$, so that this becomes $\sum_{t,w}\beta_{tw}(x_t, x_w)$. Let us also write $\phi(\mathbf{x})$ for $\sum_{ij}\lambda_{ij}\phi_{ij}(\mathbf{x})$ and let $\phi_1, \dots, \phi_d$ be linear maps from $\mathbb{F}_p^n$ to $\mathbb{F}_p$ such that $\phi(\mathbf{x}) = \sum_t \phi_t(x_t)$ for every $\mathbf{x}$. Then

$$\sum_{i,j}\lambda_{ij}(q_j(L_i(\mathbf{x})) - \phi_{ij}(\mathbf{x})) = \sum_{t,w}\beta_{tw}(x_t, x_w) - \sum_t \phi_t(x_t).$$

Notice that if we cannot get u to equal v, then $\sum_i \lambda_{ij}c_{iu}^2 = 0$ for every u and every j, which implies that $\beta_{uu} = 0$. Notice also that the assumption that Γ_2 has rank at least r and the fact that $\sum_i \lambda_{ij}c_{iu}c_{iw} \neq 0$ for at least one j imply that β_{uv} has rank at least r.

If we fix every x_t except for x_u and x_v, then $\sum_{t,w}\beta_{tw}(x_t,x_w)-\sum_t\phi_t(x_t)$ is a function of x_u and x_v of the form

$$\beta_{uv}(x_u,x_v)+\psi_u(x_u)+\psi_v(x_v),$$

where ψ_u and ψ_v are linear functionals on $\mathbb{F}_p^n$ (that depend on the other x_t).

Now let us estimate the expectation

$$\mathbb{E}_{x_u,x_v}\omega^{\sum_{i,j}\lambda_{ij}(q_j(L_i(\mathbf{x}))-\phi_{ij}(\mathbf{x})-b_{ij})},$$

where we have fixed every x_t apart from x_u and x_v. Letting $b=\sum\lambda_{ij}b_{ij}$ and using the calculations we have just made, we can write this in the form

$$\mathbb{E}_{x_u,x_v}\omega^{\beta_{uv}(x_u,x_v)+\psi_u(x_u)+\psi_v(x_v)-b}.$$

If $u=v$, then the expectation is just over x_u and the exponent has the form $q(x_u)+w^Tu-b$ for some quadratic form q of rank at least r. Therefore, by Lemma 2.29, the expectation is at most $p^{-r/2}$. If $u\neq v$ (and therefore every $\mathbf{b}_{uu}$ is zero) then for each x_v the exponent is linear in u. This means that either the expectation over x_u is zero or the function $\beta_{uv}(x_u,x_v)+\psi_u(x_u)$ is constant. If the latter is true when $x_v=y$ and when $x_v=z$, then $\beta_{uv}(x_u,y-z)$ is also constant, and therefore identically zero. Since β_{uv} has rank at least r, $y-z$ must lie in a subspace of codimension at least r. Therefore, the set of x_v such that $\beta_{uv}(x_u,x_v)+\psi_u(x_u)$ is constant is an affine subspace of $\mathbb{F}_p^n$ of codimension at least r, which implies that the probability (for a randomly chosen x_v) that the expectation (over x_u) is non-zero is at most p^{-r}. When the expectation is non-zero, it has modulus 1.

In either case, we find that, for any non-zero $\lambda\in\Lambda$, the expectation over $\mathbf{x}$ is at most $p^{-r/2}$, and this completes the proof of the lemma. □

We now want to take into account Γ_1 as well as Γ_2. This turns out to be a short deduction from the previous result. First let us do a simple piece of linear algebra.

Lemma 2.13. *Let $\mathcal{L}=(L_1,\dots,L_m)$ be a collection of linear forms in d variables, and suppose that the linear span of $L_1,\dots,L_m$ has dimension d'. Let $\Gamma_1:\mathbb{F}_p^n\to\mathbb{F}_p^{d_1}$ be a surjective linear map and let $\phi:(\mathbb{F}_p^n)^d\to(\mathbb{F}_p^{d_1})^m$ be defined by the formula*

$$\phi:\mathbf{x}\mapsto(\Gamma_1(L_1(\mathbf{x})),\dots,\Gamma_1(L_m(\mathbf{x}))).$$

Then the image of ϕ is the subspace Z of $(\mathbb{F}_p^{d_1})^m$ that consists of all sequences

$(a_1, \dots, a_m)$ such that $\sum_i \mu_i a_i = 0$ whenever $\sum_i \mu_i L_i = 0$. The dimension of Z is $d'd_1$.

Proof. Since the m forms L_i span a space of dimension d', the set of sequences $\mu = (\mu_1, \dots, \mu_m)$ such that $\sum_i \mu_i L_i = 0$ is a linear subspace W of $\mathbb{F}_p^m$ of dimension $m - d'$. Therefore, the condition that $\sum_i \mu_i a_i = 0$ for every sequence $\mu \in W$ restricts $(a_1, \dots, a_m)$ to a subspace of $(\mathbb{F}_p^{d_1})^m$ of codimension $d_1(m - d')$. (An easy way to see this is to write $a_i = (a_{i1}, \dots, a_{id_1})$ and note that for each j the sequence $(a_{1j}, \dots, a_{mj})$ is restricted to a subspace of codimension $m - d'$.) Therefore, the dimension of Z is $d'd_1$, as claimed.

Now let us show that Z is the image of ϕ. Since ϕ is linear, Z certainly contains the image of ϕ, so it will be enough to prove that the rank of ϕ is $d'd_1$.

Abusing notation, let us write $\Gamma_1(\mathbf{x})$ for the sequence $(\Gamma_1 x_1, \dots, \Gamma_1 x_d)$, which belongs to $(\mathbb{F}_p^{d_1})^d$. Then $\phi(\mathbf{x})$ can be rewritten as $(L_1(\Gamma_1(\mathbf{x})), \dots, L_m(\Gamma_1(\mathbf{x})))$. Since Γ_1 is a surjection, it is also a surjection when considered as a map on $(\mathbb{F}_p^n)^d$. Therefore, the rank of ϕ is the rank of the map $\psi : (\mathbb{F}_p^{d_1})^d \to (\mathbb{F}_p^{d_1})^m$ defined by

$$\psi : \mathbf{y} \mapsto (L_1(\mathbf{y}), \dots, L_m(\mathbf{y})).$$

Since the L_i span a space of dimension d', the nullity of this map is $d_1(d - d')$, so its rank is $d_1 d'$. Therefore, the image of ϕ is indeed Z. □

Lemma 2.14. *Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, and suppose that the linear span of $L_1, \dots, L_m$ has dimension d'. Let $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ be a surjective linear map and let $\Gamma_2 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_2}$ be a quadratic map of rank at least r. Let $a_1, \dots, a_m$ be elements of $\mathbb{F}_p^{d_1}$ and let $b_1, \dots, b_m$ be elements of $\mathbb{F}_p^{d_2}$, and let ϕ and Z be as defined in the previous lemma. Then the probability, if $\mathbf{x}$ is chosen randomly from $(\mathbb{F}_p^n)^d$, that $\Gamma_1(L_i(\mathbf{x})) = a_i$ and $\Gamma_2(L_i(\mathbf{x})) = b_i$ for every $i \le m$ is zero if $(a_1, \dots, a_m) \in Z$, and otherwise it differs from $p^{-d_1 d' - d_2 m}$ by at most $p^{d_1 - d' d_1 - r/2}$.*

Proof. If $\mathbf{a} = (a_1, \dots, a_m) \notin Z$, then there exists $\mu \in \mathbb{F}_p^m$ such that $\sum_i \mu_i a_i \neq 0$ and $\sum_i \mu_i L_i(\mathbf{x}) = 0$ for every $\mathbf{x}$. Since Γ_1 is linear, it follows that there is no $\mathbf{x}$ such that $\Gamma_1(L_i(\mathbf{x})) = a_i$ for every i.

Otherwise, by Lemma 2.13, $\mathbf{a}$ lies in the image of ϕ, which has rank $d'd_1$, so $\phi^{-1}(\{\mathbf{a}\})$ is an affine subspace of $(\mathbb{F}_p^n)^d$ of codimension $d'd_1$. Therefore, the probability that $\phi(\mathbf{x}) = \mathbf{a}$ is $p^{-d'd_1}$. Now let us use Lemma 2.12 to estimate the probability, conditional on this, that $\Gamma_2(L_i(\mathbf{x})) = b_i$ for every $i \le m$.

In the proof of Lemma 2.13, we observed that $\phi(\mathbf{x})$ depends on $\Gamma_1(\mathbf{x})$ only, so we shall estimate the required probability, given the value of $\Gamma_1(\mathbf{x})$. (Recall that this is notation for $(\Gamma_1 x_1, \dots, \Gamma_1 x_d)$.) In order to specify the set on which we are conditioning, let V be the kernel of Γ_1 (considered as a map defined on $\mathbb{F}_p^n$), and given a sequence $(w_1, \dots, w_d) \in (\mathbb{F}_p^n)^d$, let us estimate the required probability, given that $x_u \in V + w_u$ for every u.

Let us write $x_u = y_u + w_u$. Thus, we are estimating the probability that $\Gamma_2(L_i(\mathbf{y} + \mathbf{w})) = b_i$ for every $i \leq m$. But for each i we can write $\Gamma_2(L_i(\mathbf{y} + \mathbf{w}))$ as $\Gamma_2(L_i(\mathbf{y})) + \phi_i(\mathbf{y}) + b_i'$ for some linear function $\phi_i : V^d \to \mathbb{F}_p^{d_2}$ and some vector $b_i' \in \mathbb{F}_p^{d_2}$.

Because Γ_2 has rank at least r and the codimension of V in $\mathbb{F}_p^n$ is d_1, Lemma 2.10 implies that the rank of the restriction of Γ_2 to V is at least $r - 2d_1$. Therefore, by Lemma 2.12, the probability that $\Gamma_2(L_i(\mathbf{y})) = -\phi_i(\mathbf{y}) + b_i - b_i'$ for every i differs from p^{-md_2} by at most $p^{d_1 - r/2}$.

Since this is true for all choices of $\mathbf{w}$, we have the same estimate if we condition on the event that $\phi(\mathbf{x}) = \mathbf{a}$ for some fixed $\mathbf{a} \in Z$. Therefore, the probability that $\Gamma_1(L_i(\mathbf{x})) = a_i$ and $\Gamma_2(L_i(\mathbf{x})) = b_i$ for every i differs from $p^{-d'd_1 - md_2}$ by at most $p^{d_1 - d'd_1 - r/2}$, as claimed. □

Next, we observe that Lemma 2.14 implies that all the atoms of $\mathcal{B}_2$ have approximately the same size.

Corollary 2.15. *Let Γ_1 and Γ_2 be as above and let x be a randomly chosen element of $\mathbb{F}_p^n$. Then for every $a \in \mathbb{F}_p^{d_1}$ and every $b \in \mathbb{F}_p^{d_2}$, the probability that $\Gamma_1(x) = a$ and $\Gamma_2(x) = b$ differs from $p^{-d_1 - d_2}$ by at most $p^{-r/2}$.*

Proof. Let us apply Lemma 2.14 in the case where $\mathcal{L}$ consists of the single one-variable linear form $L(x) = x$. This has linear rank 1 and is square-independent, so when we apply the lemma we have $d' = m = 1$. If we let $a_1 = a$ and $b_1 = b$, then the conclusion of the lemma tells us precisely what is claimed. □

The next two lemmas are simple technical facts about projections on to linear factors. The first one tells us that if g is any function that is uniform and constant on the atoms of a linear factor, then it has small L_2-norm. The second tells us that projecting on to a linear factor decreases the U^2-norm.

Lemma 2.16. *Let G be a function from $\mathbb{F}_p^{d_1}$ to $[-1, 1]$, let $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ be a surjective linear map and let $g = G \circ \Gamma_1$. Then $\|g\|_2^4 \leq p^{d_1} \|g\|_{U^2}^4$.*

Proof. Since Γ_1 takes each value in $\mathbb{F}_p^{d_1}$ the same number of times, $\|g\|_{U^2} = \|G\|_{U^2}$. But

$$\|G\|_{U^2}^4 = \mathbb{E}_a(\mathbb{E}_b G(b)G(b+a))^2 \geq p^{-d_1}(\mathbb{E}_b G(b)^2)^2 = p^{-d_1}\|G\|_2^4,$$

which proves the result, since $\|g\|_2 = \|G\|_2$ as well. □

Lemma 2.17. *Let f be a function from $\mathbb{F}_p^n$ to $\mathbb{R}$, let $\mathcal{B}_1$ be a linear factor and let $g = \mathbb{E}(f|\mathcal{B}_1)$. Then $\|g\|_{U^2} \leq \|f\|_{U^2}$.*

Proof. On every atom of $\mathcal{B}_1$, g is constant and $f - g$ averages zero. Let Γ_1 be the linear map that defines $\mathcal{B}_1$ and, as we did earlier, for each $a \in \mathbb{F}_p^{d_1}$ let V_a stand for $\Gamma_1^{-1}(\{a\})$. Then

$$\|f\|_{U^2}^4 = \mathbb{E}_{a_1+a_2=a_3+a_4}\mathbb{E}_{\substack{x_1+x_2=x_3+x_4\\ \Gamma_1(x_i)=a_i}} f(x_1)f(x_2)f(x_3)f(x_4)\ .$$

Let us fix a choice of $a_1 + a_2 = a_3 + a_4$ and consider the inner expectation. Setting $g' = f - g$, this has the form

$$\mathbb{E}_{\substack{x_1+x_2=x_3+x_4\\ \Gamma_1(x_i)=a_i}} (\lambda_1 + g'(x_1))(\lambda_2 + g'(x_2))(\lambda_3 + g'(x_3))(\lambda_4 + g'(x_4))$$

This splits into sixteen parts. Each part that involves at least one λ_i and at least one $g'(x_i)$ is zero, because any three of the x_is can vary independently and g' averages zero on every atom of $\mathcal{B}_1$. This means that the expectation is

$$\lambda_1\lambda_2\lambda_3\lambda_4 + \mathbb{E}_{\substack{x_1+x_2=x_3+x_4\\ \Gamma_1(x_i)=a_i}} g'(x_1)g'(x_2)g'(x_3)g'(x_4)\ .$$

If we now take expectations over $a_1 + a_2 = a_3 + a_4$ we find that $\|f\|_{U^2}^4 = \|g\|_{U^2}^4 + \|f - g\|_{U^2}^4$. Notice that this is a general result about how the U^2-norm of a function is related to the U^2-norm of a projection on to a linear factor. □

Now we are ready to estimate the product we are interested in, for functions that are constant on the atoms of $\mathcal{B}_2$.

Lemma 2.18. *Let $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ be a linear function and $\Gamma_2 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_2}$ be a quadratic function. Let $(\mathcal{B}_1, \mathcal{B}_2)$ be the corresponding quadratic factor and suppose that this has rank at least r. Let $c > 0$ and let $f : \mathbb{F}_p^n \to [-1, 1]$ be a function with $\|f\|_{U^2} \leq c$ and let $f_1 = \mathbb{E}(f|\mathcal{B}_2)$. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms. Then*

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d}\prod_{i=1}^m f_1(L_i(\mathbf{x})) \leq 4^m c p^{d_1/4} + 2^{m+1}p^{m(d_1+d_2)-r/2}.$$

Proof. Let $g = \mathbb{E}(f_1|\mathcal{B}_1)$ and let $h = f_1 - g$. Then $\|g\|_1 \leq \|g\|_2 \leq p^{d_1/4}\|g\|_{U^2}$, by Cauchy-Schwarz and Lemma 2.16. By Lemma 2.17, $\|g\|_{U^2} \leq \|f\|_{U^2}$, which is at most c, by hypothesis. Therefore, $\|g\|_1 \leq cp^{d_1/4}$.

Since $f_1 = g + h$, we can split the product up into a sum of 2^m products, in each of which we replace $f_1(L_i(\mathbf{x}))$ by either $g(L_i(\mathbf{x}))$ or $h(L_i(\mathbf{x}))$. Since $\|g\|_1 \leq cp^{d_1/4}$ and $\|h\|_\infty \leq 2$, any product that involves at least one g has average at most $2^m cp^{d_1/4}$. It remains to estimate

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m h(L_i(\mathbf{x})).$$

Let Z be as defined in Lemma 2.13, and for each $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{b} = (b_1, \dots, b_m)$, let $P(\mathbf{a}, \mathbf{b})$ be the probability that $\Gamma_1(L_i(\mathbf{x})) = a_i$ and $\Gamma_2(L_i(\mathbf{x})) = b_i$ for every i. By Lemma 2.14, we can set $P(\mathbf{a}, \mathbf{b}) = p^{-d'd_1 - md_2} + \epsilon(\mathbf{a}, \mathbf{b})$, with $|\epsilon(\mathbf{a}, \mathbf{b})| \leq p^{d_1 - d'd_1 - r/2}$.

Now let H be defined by the formula $h(x) = H(\Gamma_1 x, \Gamma_2 x)$. Because h is constant on the atoms of $\mathcal{B}_2$, H is well-defined on the set of all elements of $\mathbb{F}_p^{d_1} \times \mathbb{F}_p^{d_2}$ of the form $(\Gamma_1 x, \Gamma_2 x)$. Since h takes values in $[-2, 2]$, so does H.

Next, we show that $\mathbb{E}_b H(a, b)$ is small for any fixed $a \in \mathbb{F}_p^{d_1}$, using the facts that h averages 0 on every cell of $\mathcal{B}_1$ and that it is constant on the cells of $\mathcal{B}_2$. Let us fix an a and write $P(b)$ for the probability that $\Gamma_2(x) = b$ given that $\Gamma_1(x) = a$—that is, for the density of $V_a \cap W_b$ inside V_a. Then

$$0 = \mathbb{E}_{x\in V_a} h(x) = \mathbb{E}_{x\in V_a} H(\Gamma_1 x, \Gamma_2 x) = \sum_b P(b) H(a, b).$$

By Corollary 2.15, we can write $P(b) = p^{-d_2} + \epsilon(b)$, with $|\epsilon(b)| \leq p^{d_1 - r/2}$ for every b. Therefore, the right-hand side differs from $\mathbb{E}_b H(a, b)$ by at most $2p^{d_1 + d_2 - r/2}$, which implies that $|\mathbb{E}_b H(a, b)| \leq 2p^{d_1 + d_2 - r/2}$.

Now

$$\mathbb{E}_{\mathbf{x}} \prod_{i=1}^m h(L_i(\mathbf{x})) = \mathbb{E}_{\mathbf{x}} \prod_{i=1}^m H(\Gamma_1(L_i(\mathbf{x})), \Gamma_2(L_i(\mathbf{x}))) = \sum_{\mathbf{a}\in Z} \sum_{\mathbf{b}} P(\mathbf{a}, \mathbf{b}) \prod_{i=1}^m H(a_i, b_i).$$

Let us split up this sum as

$$p^{-d'd_1 - md_2} \sum_{\mathbf{a}\in Z} \sum_{\mathbf{b}} \prod_{i=1}^m H(a_i, b_i) + \sum_{\mathbf{a}\in Z} \sum_{\mathbf{b}} \epsilon(\mathbf{a}, \mathbf{b}) \prod_{i=1}^m H(a_i, b_i).$$

The first term equals $\mathbb{E}_{a\in Z} \prod_{i=1}^m (\mathbb{E}_b H(a_i, b))$, which is at most $(2p^{d_1 + d_2 - r/2})^m$. The second is at most $p^{(d'd_1 + md_2)} 2^m p^{d_1 - d'd_1 - r/2} = 2^m p^{d_1 + md_2 - r/2}$. Therefore, the whole

sum is at most $2^{m+1}p^{m(d_1+d_2)-r/2}$. Together with our estimate for the terms that involved g, this proves the lemma. □

We have almost finished the proof of our main result.

Theorem 2.19. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let $f : \mathbb{F}_p^n \to [-1,1]$ be a c-uniform function. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, with Cauchy-Schwarz complexity at most 2. Then*

$$\left| \mathbb{E}_{\mathbf{x} \in (\mathbb{F}_p^n)^d} \prod_{i=1}^m f(L_i(\mathbf{x})) \right| \leq \epsilon.$$

Proof. Let $\delta > 0$ be a constant to be chosen later. Let C be such that $2^{m+1}p^{-C/2} \leq \epsilon/3$ and let r be the function $d \mapsto 2md + C$. Then according to the structure theorem (Theorem 2.11) there exists d_0, depending on r and δ only, and a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ of rank at least $2m(d_1+d_2)+C$ and complexity (d_1, d_2), with d_1 and d_2 both at most d_0, such that we can write $f = f_1 + f_2 + f_3$, with $f_1 = \mathbb{E}(f|\mathcal{B}_2)$, $\|f_2\|_2 \leq \delta$ and $\|f_3\|_{U^3} \leq \delta$.

Let us show that the sum does not change much if we replace $f(L_m(\mathbf{x}))$ by $f_1(L_m(\mathbf{x}))$. The difference is what we get if we replace $f(L_m(\mathbf{x}))$ by $f_2(L_m(\mathbf{x})) + f_3(L_m(\mathbf{x}))$. Now $\|f_2\|_1 \leq \|f_2\|_2$ and $\|f\|_\infty \leq 1$, so the contribution from the f_2 part is at most δ. As for the f_3 part, since $\|f_3\|_{U^3} \leq \delta$ and $\|f\|_\infty \leq 1$, Theorem 2.4 tells us that the contribution is at most δ. Therefore, the total difference is at most $\delta + \delta \leq 2\delta$.

Now let us replace f by f_1 in the penultimate bracket. The same argument works, since $\|f_1\|_\infty \leq 1$. Indeed, we can carry on with this process, replacing every single f by f_1, and the difference we make will be at most $2m\delta$.

We are left needing to show that the product with every f replaced by f_1 is small. This is what Lemma 2.18 tells us. It gives us an upper bound of $4^m cp^{d_1/4} + 2^{m+1}p^{m(d_1+d_2)-r/2}$, where for r we can take $2m(d_1+d_2)+C$. Therefore, the upper bound is $4^m cp^{d_0/4} + 2^{m+1}p^{-C/2}$, which, by our choice of C, is at most $4^m cp^{d_0/4} + \epsilon/3$.

To finish, let $\delta = \epsilon/6m$. This determines the value of d_0 and we can then set c to be $4^{-m}p^{-d_0/4}\epsilon/3$, which will be a function of ϵ only. □

Because of our use of Theorem 2.11, the bounds in the above result and in the corollary that we are about to draw from it are both very weak. However, we have been explicit about all the bounds apart from d_0, partly in order to make it clear how the parameters depend on each other and partly to demonstrate that our weak bound derives just from the weakness of d_0 in the structure theorem.

Corollary 2.20. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let A be a c-uniform subset of $\mathbb{F}_p^n$ of density α. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, with Cauchy-Schwarz complexity at most 2. Let $\mathbf{x} = (x_1, \dots, x_d)$ be a random element of $(\mathbb{F}_p^n)^d$. Then the probability that $L_i(\mathbf{x}) \in A$ for every i differs from α^m by at most ϵ.*

Proof. We shall choose as our c the c that is given by the previous theorem when ϵ is replaced by $\epsilon/2^m$. Our assumption is then that we can write $A = \alpha + f$ for a c-uniform function f. The probability we are interested in is

$$\mathbb{E}_{\mathbf{x} \in (\mathbb{F}_p^n)^d} \prod_{i=1}^{m} A(L_i(\mathbf{x})),$$

which we can split into 2^m parts, obtained by replacing each occurrence of A either by α or by f.

For each part that involves at least one occurrence of f, we have a power of α multiplied by a product over some subsystem of $\mathcal{L}$. This subsystem will also be square-independent and have CS-complexity at most 2. Moreover, the number of linear forms will have decreased. Therefore, the previous theorem and our choice of c tell us that the contribution it makes is at most $\epsilon/2^m$. Therefore, the contribution from all such parts is at most ϵ. The only remaining part is the one where every $A(L_i(\mathbf{x}))$ has been replaced by α, and that gives us the main term α^m. □

2.3.4 Remarks

First, we remark that Corollary 2.20 allows us to deduce rather straightforwardly a Szemerédi-type theorem for square-independent systems of CS-complexity 2 which have the additional property that they are *translation-invariant*. That is, one can show that any sufficiently dense subset of $\mathbb{F}_p^n$ contains a configuration of the given type.

Without the result of the preceding section, establishing that any sufficiently dense subset contains a solution to systems of this type would require a quadratic argument of the form used by Green and Tao to prove Szemerédi's Theorem for progressions of length 4 in finite fields [GT05b]. This would involve obtaining density increases on quadratic subvarieties of $\mathbb{F}_p^n$, which then need to be linearized in a carefully controlled manner. Although it is certainly possible to adapt their argument in this way, for purely qualitative purposes it is much simpler to use the result that configurations of this type are governed by the U^2-norm, which allows one to produce a density increase

on an affine subspace. The resulting argument is almost identical to the well-known argument for 3-term progressions [Mes95]. Translation invariance is needed because the subspace on which we find a density increment may be an affine and not a strictly linear one. (It is not hard to show that the result is false if the system is not translation invariant.)

There are several ways in which the results of Section 3 might be generalized. An obvious one is to prove comparable results for the group $\mathbb{Z}_N$. As we mentioned earlier, we have a different proof for $\mathbb{F}_p^n$ and this can be transferred to $\mathbb{Z}_N$ by "semi-standard" methods. (That is, the general approach is clear, but the details can be complicated and sometimes require more than merely technical thought.) The alternative proof for $\mathbb{F}_p^n$ gives a doubly exponential bound for the main result rather than the tower-type bound obtained here.

Possibly even more obvious is to try to extend the main result of this paper to a proof of Conjecture 2.6. This involves a generalization in two directions: to systems of CS-complexity greater than 2, and to systems with true complexity greater than 1. All further cases will require polynomial Fourier analysis for a degree that is greater than 2: the simplest is likely to be to show that a square-independent system with CS-complexity 3 has true complexity 1. In this case, we would use a decomposition into a structured part (a projection onto a cubic factor) and a uniform part (which would be small in U^4 and therefore negligible) and then, as before, concentrate on the structured part. Square-independence (which implies cube-indepence) would ensure that we could reduce to the linear part of the factor as before.

This state of affairs leaves us very confident that Conjecture 2.6 is true. Although cubic and higher-degree Fourier analysis have not yet been worked out, they do at least exist in local form for $\mathbb{Z}_N$: they were developed in [Gow01] to prove the general case of Szemerédi's theorem. It is therefore almost certain that global forms will eventually become available, both for $\mathbb{Z}_N$ and for $\mathbb{F}_p^n$. And then, given a statement analogous to Theorem 2.11, it is easy to see how to generalize the main steps of our proof. In particular, the Gauss-sum estimates on which we depend so heavily have higher-degree generalizations.

A completely different direction in which one might consider generalizing the above results is to hypergraphs. For example, very similar proofs to those of Theorems 2.3 and 2.4 can be used to prove so-called "counting lemmas" for quasirandom hypergraphs—lemmas that assume that a certain norm is small and deduce that the hypergraph contains approximately the expected number of small configurations of a given kind.

One can now ask whether, as with sets, weaker quasirandomness assumptions about a

hypergraph suffice to guarantee the right number of certain configurations, and if so, which ones. It turns out to be possible to give a complete answer to a fairly natural formulation of this question. Unfortunately, however, the proof is rather too easy to be interesting, so here we content ourselves with somewhat informal statements of results concerning the special case of 3-uniform hypergraphs. The proofs we leave as exercises for any reader who might be interested.

Recall that if X, Y and Z are finite sets and $f : X \times Y \times Z \to \mathbb{R}$, then the *octahedral norm* of f is the eighth root of

$$\mathbb{E}_{x(0),x(1)\in X}\mathbb{E}_{y(0),y(1)\in Y}\mathbb{E}_{z(0),z(1)\in Z}\prod_{\epsilon\in\{0,1\}^3} f(x(\epsilon_1), y(\epsilon_2), z(\epsilon_3)).$$

It is easy to verify that if $X = Y = Z = G$ for some Abelian group G and $f(x, y, z) = g(x + y + z)$ for some function g, then the octahedral norm of f is the same as the U^3-norm of g. Therefore, it is natural to consider the octahedral norm of functions defined on $X \times Y \times Z$ as the correct analogue of the U^3-norm of functions defined on Abelian groups.

An important fact about the octahedral norm is that f has small octahedral norm if and only if it has a small correlation with any function of the form $u(x, y)v(y, z)w(x, z)$. Another important fact, the so-called "counting lemma" for quasirandom hypergraphs, states the following. Let X be a finite set and let H be a 3-uniform hypergraph with vertex set X and density α. Suppose that H is quasirandom in the sense that the function $H(x, y, z)-\alpha$ has small octahedral norm (where $H(x, y, z) = 1$ if $\{x, y, z\} \in H$ and 0 otherwise). Then H has about the expected number of copies of any fixed small hypergraph. For instance, if you choose x, y, z and w randomly from X, then the probability that all of $\{x, y, z\}, \{x, y, w\}, \{x, z, w\}$ and $\{y, z, w\}$ belong to H is approximately α^4.

Now let us suppose that g is uniform but not necessarily quadratically uniform, and that we again define $f(x, y, z)$ to be $g(x+y+z)$. What can we say about f? It is not necessarily the case that f has small octahedral norm, or that it has low correlation with functions of the form $u(x, y)v(y, z)w(x, z)$. However, it is not hard to show that it has low correlation with any function of the form $a(x)b(y)c(z)$, a property that was referred to as *vertex uniformity* in [Gow06a].

One might therefore ask whether vertex uniformity was sufficient to guarantee the right number of copies of some small hypergraphs. However, well-known and easy examples shows that it does so only for hypergraphs such that no pair $\{x, y\}$ is contained in more than one hyperedge. For instance, let u be a random symmetric

function from X^2 to $\{-1,1\}$ and let $H(x,y,z) = (3 + u(x,y) + u(y,z) + u(x,z))/6$. Then H is vertex uniform and has density $1/2$, but it is a simple exercise to show that $\mathbb{E}_{x,y,z,w}H(x,y,z)H(x,y,w)$ is about $5/18$ instead of the expected $1/4$.

However, this is perhaps not the right question to be asking. If g is uniform, then f has a stronger property than just vertex uniformity: one can prove that it does not correlate with any function of the form $u(x,y)w(x,z)$, $u(x,y)v(y,z)$ or $v(y,z)w(x,z)$. If we take *this* as our definition of "weak quasirandomness" for functions (and call the hypergraph H weakly quasirandom if the function $H-\alpha$ is), then which hypergraphs appear with the right frequency (or with "frequency zero" if we are talking about functions rather than sets)? The answer turns out to be that a sum over copies of a small hypergraph H' will have the "right" value if and only if there is a pair $\{x,y\}$ that belongs to exactly one hyperedge $\{x,y,z\}$ of H'. The proof in the "if" direction is an easy exercise. In particular, it does not involve any interesting results about decomposing hypergraphs, which suggests that the main result of this chapter is, in a certain sense, truly arithmetical.

As for the "only if" direction, here is a quick indication of how to produce an example (in the complex case, for simplicity). Suppose that no pair $\{x,y\}$ belongs to more than m hyperedges in H'. For each k between 2 and m let $f_k : X^2 \to \mathbb{C}$ be a function whose values are randomly chosen kth roots of unity. Then let $f(x,y,z)$ be the sum of all functions of the form $u(x,y)v(y,z)w(x,z)$, where each of u, v and w is some f_k with $2 \leq k \leq m$. When one expands out the relevant sum for this function f, one finds that most terms cancel, but there will be some that don't and they will all make a positive contribution. To find such a term, the rough idea is to choose for each face F of H' a triple of functions $(f_{k_1}, f_{k_2}, f_{k_3})$, where k_1, k_2 and k_3 are the number of faces of H' that include each of the three edges that make up the face F. For this term, each time a kth root of unity appears in the product, it is raised to the power k, so the term is large.

2.4 Improved Bounds

In this section we derive improved bounds for Theorem 2.19. For the sake of clarity and continuity, we briefly recall its statement.

Theorem 2.19. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let $f : \mathbb{F}_p^n \to [-1,1]$ be a c-uniform function. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, with Cauchy-Schwarz complexity*

at most 2. Then

$$\left|\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d}\prod_{i=1}^{m} f(L_i(\mathbf{x}))\right| \leq \epsilon.$$

In Section 2.3, the bounds obtained for the uniformity parameter c in terms of the error ϵ were of tower-type as a result of using Theorem 2.11 as a black box. In the present section we obtain the following improvement.

Theorem 2.21. *In Theorem 2.19, the uniformity parameter c can be taken to be a tower of exponentials of height $m+1$ in the error ϵ^{-1}.*

Of course, Theorem 2.21 immediately translates into a bound on the number of solutions in any uniform subset $A \subseteq \mathbb{F}_p^n$ as in Corollary 2.20.

Recall that the core of the proof of Theorem 2.19 consisted of the decomposition of the function f into a quadratically structured and a quadratically uniform part in the form of Theorem 2.11. In the next subsection, we shall give an alternative decomposition for a general bounded function f which is more in the spirit of classical harmonic analysis. In Section 2.4.2 we make use of the additional assumption that f is uniform in order to eliminate low-rank quadratic phases from this decomposition. In Section 2.4.3 we show how square-independence of the linear system comes into the equation, and finally Section 2.4.4 completes the proof.

2.4.1 Decomposing f into a Sum of Quadratic Phases

Let us start almost completely from scratch and state the U^3-inverse theorem [GT05b] on which the decomposition result Theorem 2.11 is based, and whose history we have already discussed in Section 2.3.

Theorem 2.22. *Let $0 < \delta \leq 1$ and let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function with $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$. Then there exists a quadratic form $q : \mathbb{F}_p^n \to \mathbb{F}_p$ such that*

$$|\mathbb{E}_x f(x)\omega^{q(x)}| \geq \exp(-C\delta^{-C}).$$

Here, C is a constant that depends on p only.

As we saw in Section 2.3, Green and Tao use the above theorem to decompose an arbitrary function f into two parts, f_1 and f_2, where f_2 is quadratically uniform and f_1 is quadratically structured, in the sense that one can partition $\mathbb{F}_p^n$ into a small number of quadratic subvarieties on each of which f_1 is constant. In this section, we shall take a somewhat different approach, more closely analogous to the way conventional

Fourier analysis is used to prove Roth's theorem. That is, we shall simply decompose f into a sum of functions of the form ω^{q_i}, where the q_i are quadratic forms, plus an error that we can afford to ignore, and then calculate directly using this expansion of f.

A big difference between the expansion we shall obtain and the expansion of a function into Fourier coefficients is that there does not seem to be a canonical way of doing it, because there are far more than p^n different functions of the form ω^q. (In harmonic-analysis terms, we are dealing with an "overdetermined" system.) This creates difficulties, which Green and Tao dealt with by projecting onto "quadratic factors". Here we shall deal with them by applying the Hahn-Banach theorem for finite-dimensional normed spaces.

Before we can explain why the Hahn-Banach theorem is useful, we must state both it and one or two other simple results about duality in normed spaces. Throughout the next few results, we shall refer to an inner product: this is just the standard inner product on $\mathbb{C}^n$ (or later $\mathbb{C}^{\mathbb{F}_p^n}$).

Theorem 2.23. *Let $X = (\mathbb{C}^n, \|.\|)$ be a normed space and let $x \in X$ be a vector with $\|x\| \geq 1$. Then there is a vector z such that $|\langle x, z\rangle| \geq 1$ and such that $|\langle y, z\rangle| \leq 1$ whenever $\|y\| \leq 1$.*

Recall that the dual norm $\|.\|^*$ of a norm $\|.\|$ on $\mathbb{C}^n$ is defined by the formula

$$\|z\|^* = \sup\{|\langle x, z\rangle| : \|x\| \leq 1\}$$

For technical reasons, we shall generalize this concept to the situation where the norm $\|.\|$ is defined on a subspace V of $\mathbb{C}^n$. Then the dual is a seminorm, given by the formula

$$\|z\|^* = \sup\{|\langle x, z\rangle| : x \in V, \|x\| \leq 1\}$$

The next lemma is a standard fact in Banach space theory.

Lemma 2.24. *Let k be a positive integer, and for each i between 1 and k let $\|.\|_i$ be a norm defined on a subspace V_i of $\mathbb{C}^n$. Suppose that $V_1 + \cdots + V_k = \mathbb{C}^n$, and define a norm $\|.\|$ on $\mathbb{C}^n$ by the formula*

$$\|x\| = \inf\{\|x_1\|_1 + \cdots + \|x_k\|_k : x_1 + \cdots + x_k = x\}$$

Then this formula does indeed define a norm, and its dual norm $\|.\|^*$ is given by the formula

$$\|z\|^* = \max\{\|z\|_1^*, \ldots, \|z\|_k^*\}$$

Proof. It is a simple exercise to check that the expression does indeed define a norm.

Let us begin by supposing that $\|z\|_i^* \geq 1$ for some i. Then there exists $x \in V_i$ such that $\|x\|_i \leq 1$ and $|\langle x, z\rangle| \geq 1$. But then $\|x\| \leq 1$ as well, from which it follows that $\|z\|^* \geq 1$. Therefore, $\|z\|^*$ is at least the maximum of the $\|z\|_i^*$.

Now let us suppose that $\|z\|^* > 1$. This means that there exists x such that $\|x\| \leq 1$ and $|\langle x, z\rangle| \geq 1 + \epsilon$ for some $\epsilon > 0$. Let us choose $x_1, \dots, x_k$ such that $x_i \in V_i$ for each i, $x_1 + \cdots + x_k = x$, and $\|x_1\|_1 + \cdots + \|x_k\|_k < 1 + \epsilon$. Then

$$\sum_i |\langle x_i, z\rangle| > \|x_1\|_1 + \cdots + \|x_k\|_k$$

so there must exist i such that $|\langle x_i, z\rangle| > \|x_i\|_i$, from which it follows that $\|z\|_i^* > 1$. This proves that $\|z\|^*$ is at most the maximum of the $\|z\|_i^*$. □

Corollary 2.25. *Let k be a positive integer and for each $i \leq k$ let $\|.\|_i$ be a norm defined on a subspace V_i of $\mathbb{C}^n$, and suppose that $V_1 + \cdots + V_k = \mathbb{C}^n$. Let $\alpha_1, \dots, \alpha_k$ be positive real numbers, and suppose that it is not possible to write the vector x as a linear sum $x_1 + \cdots + x_k$ in such a way that $x_i \in V_i$ for each i and $\alpha_1\|x_1\|_1 + \cdots + \alpha_k\|x_k\|_k \leq 1$. Then there exists a vector $z \in \mathbb{C}$ such that $|\langle x, z\rangle| \geq 1$ and such that $\|z\|_i^* \leq \alpha_i$ for every i—or equivalently, $|\langle y, z\rangle| \leq \alpha_i$ for every i and every $y \in V_i$ with $\|y\|_i \leq 1$.*

Proof. Let us define a norm $\|.\|$ by the formula

$$\|x\| = \inf\{\alpha_1\|x_1\|_1 + \cdots + \alpha_k\|x_k\|_k : x_1 + \cdots + x_k = x\}$$

Then our hypothesis is that $\|x\| \geq 1$. Therefore, by Theorem 2.23 there is a vector z such that $|\langle x, z\rangle| \geq 1$ and $|\langle y, z\rangle| \leq 1$ whenever $\|y\| \leq 1$.

The second condition tells us that $\|z\|^* \leq 1$, and Lemma 2.24, applied to the norms $\alpha_i\|.\|_i$, tells us that $\|z\|^*$ is the maximum of the numbers $\alpha_i^{-1}\|z\|_i^*$. Therefore, $\|z\|_i^* \leq \alpha_i$ for every i, as stated. □

Recall that the difficulty we are trying to deal with is that there is no (known) canonical way of decomposing a function into functions of the form ω^q. Corollary 2.25 is an extremely useful tool for proving the existence of decompositions under these circumstances. Instead of trying to find a decomposition explicitly, one assumes that there is no decomposition and uses Corollary 2.25 to derive a contradiction. The next result illustrates the technique.

Theorem 2.26. *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_2 \leq 1$. Then for every $\delta > 0$ and $\eta > 0$ there exists M such that f has a decomposition of the form*

$$f(x) = \sum_i \lambda_i \omega^{q_i(x)} + g(x) + h(x),$$

where the q_i are quadratic forms on $\mathbb{F}_p^n$, and

$$\eta^{-1}\|g\|_1 + \delta^{-1}\|h\|_{U^3} + M^{-1}\sum_i |\lambda_i| \leq 1.$$

In fact, M can be taken to be $\exp(C(\eta\delta)^{-C})$.

Proof. Suppose not. Then for every quadratic form q on $\mathbb{F}_p^n$ let $V(q)$ be the one-dimensional subspace of $\mathbb{C}^{\mathbb{F}_p^n}$ generated by the function ω^q, with the obvious norm: the norm of $\lambda\omega^q$ is $|\lambda|$.

Applying Corollary 2.25 to these norms and subspaces, and also to the L_1-norm and U^3-norm defined on all of $\mathbb{C}^{\mathbb{F}_p^n}$, we deduce that there is a function $\phi : \mathbb{F}_p^n \to \mathbb{C}$ such that $|\langle f, \phi\rangle| \geq 1$, $\|\phi\|_\infty \leq \eta^{-1}$, $\|\phi\|_{U^3}^* \leq \delta^{-1}$ and $|\langle \phi, \omega^q\rangle| \leq M^{-1}$ for every quadratic form q.

Now the fact that $|\langle f, \phi\rangle| \geq 1$ implies, by Cauchy-Schwarz, that $\|\phi\|_2 \geq 1$. But we also know that $\langle \phi, \phi\rangle \leq \|\phi\|_{U^3}\|\phi\|_{U^3}^*$, so $\|\phi\|_{U^3} \geq \delta$. Applying the inverse theorem to $\eta\phi$, we find that there is a quadratic form q such that $|\langle \phi, \omega^q\rangle| \geq \exp(-C(\eta\delta)^{-C})$, contradicting the fact that it has to be at most M^{-1}. □

Just before we continue, let us briefly discuss a more obvious approach to Theorem 2.26 and why it does not work. Theorem 2.22 tells us that every bounded function f with large U^3-norm correlates well with some function of the form ω^q. So one might try a simple inductive argument along the following lines. If $\|f\|_{U^3}$ is large, then Theorem 2.26 gives us a quadratic form q_1 such that f correlates with ω^{q_1}. So choose λ_1 such that $\|f - \lambda_1\omega^{q_1}\|_2$ is minimized, and let $f_1 = f - \lambda_1\omega^{q_1}$. Because of the correlation, $\|f_1\|_2$ is substantially less than $\|f\|_2$. Now repeat for f_1, and keep going until you reach some k for which $\|f_{k+1}\|_{U^3}$ is small.

The problem with this argument is that we lose control of the boundedness of f. As we continually subtract the functions $\lambda_i\omega^{q_i}$, the L_2-norm goes down, but the L_∞-norm can go up. And L_2 control is not enough for Theorem 2.22. (Green and Tao's approach to quadratic Fourier analysis uses averaging projections, which decrease both the L_2- and L_∞-norms.)

2.4.2 Eliminating Low-Rank Quadratic Phases

Our next task is to show that if the function f in Theorem 2.26 is sufficiently uniform, then a decomposition can be found such that all the quadratic forms q_i have high rank. This is not at all surprising, since f does not correlate with functions ω^q for which q has low rank, but it is not as easy to prove as one might expect, and requires us to look in detail at sums of the form $\sum_i \lambda_i \omega^{q_i}$ for which all the quadratic forms q_i have rank bounded above by some R.

The rough idea of what we shall do is this. It turns out that technical problems arise when large numbers of the quadratic forms $q_i - q_j$ have rank smaller than some r, which will typically be considerably smaller than R. However, in this situation another argument can be used. So we shall prove a couple of preliminary lemmas, one about very low rank forms and one about sums where most of the pairs $q_i - q_j$ have reasonably high rank. Later we will combine these two lemmas into one that applies to all sums. Before all this, however, we prove three very basic technical lemmas.

Lemma 2.27. *Let $\mathcal{B}_1$ be a linear factor on $\mathbb{F}_p^n$ and let f be a function from $\mathbb{F}_p^n$ to $\mathbb{C}$. Let $\|.\|$ be any translation-invariant norm defined on such functions, and let $g = \mathbb{E}(f|\mathcal{B}_1)$. Then $\|g\| \leq \|f\|$.*

Proof. Let V be the subspace of $\mathbb{F}_p^n$ whose translates are the atoms of $\mathcal{B}_1$. Then $g(x) = \mathbb{E}_{v \in V} f(x+v)$ for every x. Therefore, if we write $f_v(x)$ for $f(x+v)$, we find that $g = \mathbb{E}_v f_v$ and we know that all the functions f_v have the same norm as f. The lemma therefore follows from the triangle inequality. □

Lemma 2.28. *Let $\mathcal{B}_1$ be a linear factor of complexity r on $\mathbb{F}_p^n$ and let f be a function from $\mathbb{F}_p^n$ to $\mathbb{C}$ that is constant on the atoms of $\mathcal{B}_1$. Then $\|f\|_{U^2} \geq p^{-r/4}\|f\|_2$ and $\|f\|_{U^2}^* \leq p^{r/4}\|f\|_2$.*

Proof. Again let V be the subspace of $\mathbb{F}_p^n$ whose translates are the atoms of $\mathcal{B}_1$ and let Γ be a linear map from $\mathbb{F}_p^n$ to $\mathbb{F}_p^r$ with kernel V. Let $g : \mathbb{F}_p^r \to \mathbb{C}$ be defined by the formula $f(x) = g(\Gamma x)$, which is well-defined since f is constant on translates of V. It is easy to see that $\|f\|_2 = \|g\|_2$ and $\|f\|_{U^2} = \|g\|_{U^2}$. But

$$\|g\|_{U^2}^4 = \mathbb{E}_y |\mathbb{E}_x g(x)\overline{g(x+y)}|^2 \geq p^{-r}(\mathbb{E}_x |g(x)|^2)^2 = p^{-r}\|g\|_2^4.$$

This proves the first part.

For the second part, we know that $\|f\|_{U^2}^*$ is the maximum of $\langle f, g\rangle$ over all functions g such that $\|g\|_{U^2} \leq 1$. Now replacing g by $\mathbb{E}(g|\mathcal{B}_1)$ does not affect the inner product $\langle f, g\rangle$ and does not increase $\|g\|_{U^2}$. Therefore, the maximum must be achieved by a function g that is constant on the atoms of $\mathcal{B}_1$. But then $|\langle f, g\rangle| \leq \|f\|_2\|g\|_2$, which is at most $p^{r/4}\|f\|_2\|g\|_{U^2}$, by the first part. This completes the proof of the lemma. □

The next lemma is a standard fact about Gauss sums, which we have already encountered as Lemma 2.29.

Lemma 2.29. *Let q be a quadratic form of rank r. Then $|\mathbb{E}_x\omega^{q(x)}| = p^{-r/2}$.*

As is to be expected, the rank of the quadratic form determines the U^2- (and hence the $(U^2)^*$-) norm of the corresponding quadratic phase.

Lemma 2.30. *Let q be a quadratic form of rank r. Then $\|\omega^q\|_{U^2} = p^{-r/4}$ and $\|\omega^q\|_{U^2}^* = p^{r/4}$.*

Proof. Let β be as in Lemma 2.29. Then for any x, a and b in $\mathbb{F}_p^n$ we have

$$q(x) - q(x+a) - q(x+b) + q(x+a+b) = \beta(x+b, a) + q(a) - q(a) - \beta(x, a) = \beta(a, b).$$

Therefore,

$$\|\omega^q\|_{U^2}^4 = \mathbb{E}_{x,a,b}\omega^{\beta(a,b)}.$$

Now for each fixed a the function $\beta(a, b)$ is linear in b. It therefore sums to zero unless it is identically zero. But β has rank r, so the subspace of all a such that $\beta(a, b)$ is identically zero has codimension r. Therefore, the expectation on the right hand side is p^{-r}. This proves the first part.

Now let f be an arbitrary function from $\mathbb{F}_p^n$ to $\mathbb{C}$ and let us obtain an upper bound for $|\langle f, \omega^q\rangle|$. Let $\mathcal{B}_1$ be the linear factor whose atoms are the translates of V. Then ω^q is constant on each atom, so if we let $g = \mathbb{E}(f|\mathcal{B}_1)$ then $\langle f, \omega^q\rangle = \langle g, \omega^q\rangle$. Moreover, $\|g\|_{U^2} \leq \|f\|_{U^2}$, by Lemma 2.27. But $\|g\|_2 \leq p^{r/4}\|g\|_{U^2}$, by Lemma 2.28, and therefore, by the Cauchy-Schwarz inequality and the fact that $\|\omega^q\|_2 = 1$,

$$|\langle f, \omega^q\rangle| = |\langle g, \omega^q\rangle| \leq p^{r/4}\|g\|_{U^2} \leq p^{r/4}\|f\|_{U^2}.$$

It follows that $\|\omega^q\|_{U^2}^* \leq p^{r/4}$. Moreover, taking $f = \omega^q$ and using the first part, we see that this inequality is in fact an equality. □

We remark that an alternative argument for Lemma 2.30 is to use Lemma 2.29 to prove that ω^q has p^r non-zero Fourier coefficients, each of magnitude $p^{-r/2}$, and then

use the fact that $\|f\|_{U^2} = \|\widehat{f}\|_4$. However, the argument we have used takes place entirely in physical space and is therefore easier to generalize.

Now we are ready for the case of forms of very low rank.

Lemma 2.31. *Let $f = \sum_i \lambda_i \omega^{q_i}$, where the functions q_i are quadratic forms on $\mathbb{F}_p^n$ of rank at most r, and $\sum_i |\lambda_i| \leq M$. Then for every $\delta > 0$ there is a linear factor $\mathcal{B}_1$ of complexity at most $\delta^{-2}M^4p^r$ such that $\|f - \mathbb{E}(f|\mathcal{B}_1)\|_2 \leq \delta$.*

Proof. By Lemma 2.30 we know that $\|f\|_{U^2}^* \leq Mp^{r/4}$, which is all we need to know about f. The rest of the proof is a standard Bogolyubov-type argument. First of all, since $\|f\|_{U^2} = \|\widehat{f}\|_4$, it follows that $\|f\|_{U^2}^* = \|\widehat{f}\|_{4/3}$. By the Fourier inversion formula, we know that $f(x) = \sum_r \widehat{f}(r)\omega^{-r.x}$. Let $\alpha = \delta^{3/2}p^{-r/2}M^{-2}$ and let $K = \{r : |\widehat{f}(r)| \geq \alpha\}$. Then we can decompose f as a sum $f_1 + f_2$, where $f_1(x) = \sum_{r\in K} \widehat{f}(r)\omega^{-r.x}$ and $f_2(x) = \sum_{r\notin K} \widehat{f}(r)\omega^{-r.x}$.

Let V be the subspace of all x such that $r.x = 0$ for every $r \in K$. Since $\|\widehat{f}\|_{4/3} \leq Mp^{r/4}$, we know that $|K| \leq M^{4/3}p^{r/3}\alpha^{-4/3}$. Therefore, V has codimension at most $M^{4/3}p^{r/3}\alpha^{-4/3} = \delta^{-2}M^4p^r$, which is also an upper bound for the complexity of the linear factor $\mathcal{B}_1$ defined by V. Since f_1 depends only on the values of the functions $\omega^{r.x}$ with $r \in K$, f_1 is constant on the atoms of $\mathcal{B}_1$.

As for f_2, we can bound its L_2 norm as follows.

$$\|f_2\|_2^2 = \|\widehat{f_2}\|_2^2 \leq \|\widehat{f_2}\|_{4/3}^{4/3}\|\widehat{f_2}\|_\infty^{2/3} \leq \alpha^{2/3}M^{4/3}p^{r/3} \leq \delta.$$

To complete the proof it remains to observe that $\mathbb{E}(f|\mathcal{B}_1)$ is the closest function (in L_2) to f that is constant on the atoms of $\mathcal{B}_1$. Therefore, the statement of the lemma follows from our calculations. □

Next, we deal with sums of forms that mostly have differences of high rank.

Lemma 2.32. *Let $f = \sum_i \lambda_i \omega^{q_i}$ be a function with $\sum_i |\lambda_i| \leq M$, let r be an integer and let Z be the set of pairs (i, j) such that the rank of $q_i - q_j$ is at most r. Let $\eta > 0$ and suppose that $\sum_{(i,j)\in Z} |\lambda_i||\lambda_j| \leq \eta$. Then $\|f\|_2^2 \leq \eta + p^{-r/2}M^2$.*

Proof. By Lemma 2.29,

$$\|f\|_2^2 = \sum_{i,j} \lambda_i\overline{\lambda_j}\mathbb{E}_x\omega^{q_i(x)-q_j(x)} \leq \sum_{(i,j)\in Z} |\lambda_i||\lambda_j| + p^{-r/2}\sum_{(i,j)\notin Z} |\lambda_i||\lambda_j|.$$

By hypothesis, this is at most $\eta + p^{-r/2}M^2$, as claimed. □

We are now ready for a combined lemma that will deal with arbitrary sums of forms of rank at most R.

Lemma 2.33. *Let $M \geq 1$ and let $f = \sum_i \lambda_i \omega^{q_i}$ be a function with $\sum_i |\lambda_i| \leq M$, where each q_i is a quadratic form on $\mathbb{F}_p^n$ of rank at most R. Let $\delta > 0$ and let $s = 2^{18}(M/\delta)^{12}$. Then there is a linear factor $\mathcal{B}$ of complexity at most $8(M/\delta)^2(R+s)$ such that $\|f - \mathbb{E}(f|\mathcal{B})\|_2 \leq \delta$.*

Proof. Let $\eta = \delta^2/8M$, let r be such that $p^{-r/2}M^2 = \delta^2/8$ and let $s = \eta^{-4}M^4p^r$. (It can be checked that this definition of s agrees with the definition in the statement. These numbers are chosen so that we get the right bounds out of Lemmas 2.32 and 2.31, as will become clear.)

Let us define a vertex-weighted graph G as follows. The vertices of G are the quadratic forms q_i, and the weight of q_i is $|\lambda_i|$. And q_i is joined to q_j if and only if the rank of $q_i - q_j$ is at most r. Let V be the vertex set of G, and define the *degree* of a vertex q_i to be the sum of the weights of those q_j that are joined to q_i. (We will allow G to have loops, so q_i is joined to itself.)

Suppose G has a vertex q_i of degree at least η. Then let V_1 be the neighbourhood of q_i, and remove V_1 from the vertex set of G. Now repeat this process with the induced subgraph with vertex set $V \setminus V_1$ (and the same weights). Continuing in this way we find disjoint sets $V_1, V_2, \dots, V_k$ and W such that the weight of each V_i is at least η and every vertex in W has degree less than η.

Let us now focus on V_1. If q_i is the form of which V_1 is the neighbourhood, then the rank of $q_i - q_j$ is at most r for every $q_j \in V_1$. Therefore, if we set g_1 to equal $\sum_j \overline{\lambda_j}\omega^{q_i - q_j}$, Lemma 2.31 tells us that there is a linear factor $\mathcal{B}_1$ of complexity at most s such that $\|g_1 - \mathbb{E}(g|\mathcal{B}_1)\|_2 \leq \eta^2$. It follows that $\|g_1 - \mathbb{E}(g_1|\mathcal{B})\|_2 \leq \eta^2$ for any linear factor $\mathcal{B}$ that refines $\mathcal{B}_1$.

Now let $f_1 = \sum_{q_j \in V_1} \lambda_j \omega^{q_j} = \omega^{q_i}\overline{g_1}$. Since q_i has rank at most R, there is a linear factor $\mathcal{B}_1'$ of complexity at most R on the atoms of which q_i is constant. Therefore, if $\mathcal{B}_1''$ is the smallest common refinement of $\mathcal{B}_1$ and $\mathcal{B}_1'$, we find that $\mathcal{B}_1''$ has complexity at most $R + s$ and that $\|f_1 - \mathbb{E}(f_1|\mathcal{B}_1'')\|_2 \leq \eta^2$. (Here we have also used the fact that the modulus of ω^{q_i} is everywhere 1.) Again, this statement is also true for every refinement $\mathcal{B}$ of $\mathcal{B}_1''$.

Let us do the same for each V_i. That is, $f_i = \sum_{q_j \in V_i} \lambda_j \omega^{q_j}$ and $\mathcal{B}_i''$ is a linear factor of complexity at most $R + s$ such that $\|f_i - \mathbb{E}(f|\mathcal{B}_i'')\|_2 \leq \eta^2$. Let $\mathcal{B}_1$ be a common refinement of all the linear factors $\mathcal{B}_i''$. Then $\|f_i - \mathbb{E}(f|\mathcal{B})\|_2 \leq \eta^2$ as well, and $\mathcal{B}$ has complexity at most $\eta^{-1}M(R + s)$.

Finally, let $g = f - (f_1 + \cdots + f_k)$. Then g is a sum of the form $\sum_{i \in W} \lambda_i \omega^{q_i}$, where each q_i has degree less than η in the subgraph of G induced by W. If we let Z be the set of all pairs $(i,j) \in W^2$ such that $q_i - q_j$ has rank at most r, then $\sum_{(i,j) \in Z} |\lambda_i||\lambda_j| \leq \eta M \leq \delta^2/8$. From Lemma 2.32 and our choice of r, it follows that $\|g\|_2^2 \leq \delta^2/4$.

We are now more or less done. We have

$$\|f - \mathbb{E}(f|\mathcal{B})\|_2 \leq \sum_i \|f_i - \mathbb{E}(f_i|\mathcal{B})\|_2 + \|g - \mathbb{E}(g|\mathcal{B})\|_2 \leq \eta^{-1} M \eta^2 + \delta/2 \leq \delta.$$

It remains to note that $\eta^{-1}M(R+s)$, our upper bound for the complexity of $\mathcal{B}$, is at most $8M^2(R+s)/\delta^2$, as stated. □

Needless to say, the precise form of the bound for the complexity of $\mathcal{B}$ is not important. What does matter, however, is the way it depends on R. In particular, for large R the bound is significantly better than the bound of $\delta^{-2}M^4 p^R$ that we could read out of Lemma 2.31. If we regard δ and M as fixed, then that bound is exponential in R, whereas we have just proved a linear bound. (However, M and s will be rather large constants, so this is not quite as good as it sounds.)

Theorem 2.34. *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_2 \leq 1$. Then for every $\delta > 0$, there exists a constant M such that for every R_0 there exists a constant $c > 0$ with the following property. If $\|f\|_{U^2} \leq c$ then f has a decomposition of the form*

$$f(x) = \sum_i \lambda_i \omega^{q_i(x)} + g(x) + h(x),$$

where the q_i are quadratic forms on $\mathbb{F}_p^n$, all of which have rank at least R_0, and

$$\delta^{-1}\|g\|_1 + \delta^{-1}\|h\|_{U^3} + M^{-1}\sum_i |\lambda_i| \leq 17.$$

Moreover, M can be taken to be $\exp(C(1/\delta^2)^C)$ and c can be taken to be $\delta p^{-R/4}$, where $R \leq (2^{23}(M/\delta)^{12})^{M/\delta} R_0$.

Proof. Let us begin by applying Theorem 2.26 with η replaced by δ. Then we obtain a decomposition of the required kind, except that we do not know anything about the ranks of the quadratic forms q_i and we know that

$$\delta^{-1}\|g\|_1 + \delta^{-1}\|h\|_{U^3} + M^{-1}\sum_i |\lambda_i| \leq 1.$$

However, now we have the extra hypothesis that $\|f\|_{U^2} \leq c$.

The next step is to find a number $R_1 \geq R_0$ such that

$$\sum \{|\lambda_i| : R_1 \leq r(q_i) < \theta R_1 + t\} \leq \delta,$$

where we have written $r(q_i)$ to stand for the rank of q_i, we have set $\theta := 2^{21}(M/\delta)^{14}$ and t is chosen so that $p^{t/4} > M/\delta$. Since t is much less than θ and we know that $\sum_i |\lambda_i| \leq M$, we must be able to find such an R_1 with $R_1 \leq (2\theta)^{M/\delta} R_0$. Let us define R to be θR_1. It is not hard to check that R satisfies the inequality stated in the theorem.

Now let $S = \{i : r(q_i) < R_1\}$ and $L = \{i : r(q_i) \geq R + t\}$. (These letters are chosen to stand for "small" and "large", respectively.) Then

$$\sum_i \lambda_i \omega^{q_i} = \sum_{i \in S} \lambda_i \omega^{q_i} + \sum_{i \in L} \lambda_i \omega^{q_i} + g_1,$$

where $\|g_1\|_1 \leq \delta$. Let $f_S = \sum_{i \in S} \lambda_i \omega^{q_i}$ and $f_L = \sum_{i \in L} \lambda_i \omega^{q_i}$. Then we have shown that f has a decomposition of the form $f_S + f_L + g + h$, where f_S is made out of functions ω^q with q of rank at most R_1, the forms used for f_L have rank at least $R+t$, the function g (which is the new name we have given to the old $g + g_1$) has L_1-norm at most 2δ, and $\|h\|_{U^3} \leq \delta$.

Now Lemma 2.33 gives us a linear factor $\mathcal{B}$ of complexity at most $8(M/\delta)^2(R_1 + s)$, where $s = 2^{18}(M/\delta)^{12}$, such that $\|f_S - \mathbb{E}(f_S|\mathcal{B})\|_2 \leq \delta$. In order to simplify matters, let us bound this complexity above by $\theta R_1 = R$.

So now we have a decomposition $f = \mathbb{E}(f_S|\mathcal{B}) + f_L + g + h$, where f_S, f_L and h are as before and $\|g\|_1 \leq 3\delta$. (We have added $f_S - \mathbb{E}(f_S|\mathcal{B})$ to the old g and used the fact that its L_1-norm is at most its L_2-norm.)

Without the term $f'_S := \mathbb{E}(f_S|\mathcal{B})$, we would be done. To complete the proof we shall show that f_S can be absorbed into the error term $g + h$. More precisely, let us suppose that we cannot write f'_S as a sum $g' + h'$ with $\|g'\|_1 \leq 6\delta$ and $\|h'\|_{U^3} \leq 6\delta$. Then by Corollary 2.25 there exists a function ϕ such that $|\langle f'_S, \phi\rangle| \geq 1$ and $6\delta\|\phi\|_\infty + 6\delta\|\phi\|^*_{U^3} \leq 1$.

Next, we apply Lemma 2.27, which allows us to replace ϕ by $\mathbb{E}(\phi|\mathcal{B})$. The reason for this is that the averaging projection does not increase the L_∞- or $(U^3)^*$- norms and does not change the inner product $\langle f'_S, \phi\rangle$. Let us therefore assume that ϕ is constant on the atoms of $\mathcal{B}$.

We are a short step away from the contradiction we are looking for. Since $\|\phi\|_\infty \leq 1/6\delta$ and $\|g\|_1 \leq 3\delta$, it follows that $|\langle g, \phi\rangle| \leq 1/2$. Similarly, $|\langle h, \phi\rangle| \leq 1/6$ because $\|\phi\|_{U^3}^* \leq 1/6\delta$ and $\|h\|_{U^3} \leq \delta$.

Lemma 2.28 implies that $\|\phi\|_{U^2}^* \leq p^{R/4}\|\phi\|_2$, which is at most $p^{R/4}\|\phi\|_\infty$, which we know is at most $p^{R/4}/6\delta$. Therefore, $|\langle f, \phi\rangle| \leq cp^{R/4}/6\delta$. By our choice of c, this is at most $1/6$.

Finally, Lemma 2.30 and the triangle inequality imply between them that $\|f_L\|_{U^2} \leq p^{-(R+t)/4}M$. Therefore, $|\langle f_L, \phi\rangle| \leq p^{R/4}p^{-(R+t)/4}M/6\delta$, which, by our choice of t, is strictly less than $1/6$. This is a contradiction because $f = f'_S + f_L + g + h$ and we have now shown that $|\langle f'_S, \phi\rangle| > |\langle f, \phi\rangle| + |\langle f_L, \phi\rangle| + |\langle g, \phi\rangle| + |\langle h, \phi\rangle|$.

This contradiction shows that we can after all write f'_S as a sum $g' + h'$ with $\|g'\|_1 \leq 6\delta$ and $\|h'\|_{U^3} \leq 6\delta$. Therefore, we can write $f = f_L + g + h$ with $\|g\|_1 \leq 9\delta$ and $\|h\|_{U^3} \leq 7\delta$, which implies the result. □

Once again, the exact bounds we obtain are not too important, but we do care about their rough order of magnitude and the constants on which they depend. Since M is exponential in δ^{-2}, R is exponential in M and c depends exponentially on R, the dependence of c on δ in Theorem 2.34 has the form $c \leq p^{-\exp\exp(C/\delta^2)}$ for some absolute constant C. That is, it has a trebly exponential dependence on δ.

Theorem 2.34 will be our main tool. Before we apply it to systems of square-independent linear forms, we need a couple of lemmas to help us with our calculations.

2.4.3 Identifying a High-Rank Bilinear Form

The first lemma is simply a useful version of the statement that the function $\omega^{\beta(x,y)}$ is quasirandom if the bilinear form β has sufficiently high rank.

Lemma 2.35. *Let β be a bilinear form of rank at least r and let g and h be two functions with $\|g\|_\infty$ and $\|h\|_\infty$ at most 1. Then $|\mathbb{E}_{x,y}\omega^{\beta(x,y)}g(x)h(y)| \leq p^{-r/2}$.*

Proof. This result can be proved quite easily, either directly (as we shall do) or indirectly, by first estimating the rectangle norm of the function and applying standard results in the theory of quasirandomness. Either way, the proof is a standard application of the Cauchy-Schwarz inequality. We have

$$|\mathbb{E}_{x,y}\omega^{\beta(x,y)}g(x)h(y)|^2 \leq \mathbb{E}_x|\mathbb{E}_y\omega^{\beta(x,y)}g(x)h(y)|^2 \leq \mathbb{E}_x|\mathbb{E}_y\omega^{\beta(x,y)}h(y)|^2,$$

where the latter inequality uses the boundedness of g, which in turn precisely equals

$$\mathbb{E}_{y,y'}h(y)h(y')\mathbb{E}_x\omega^{\beta(x,y-y')} \leq \mathbb{E}_{y,y'}|\mathbb{E}_x\omega^{\beta(x,y-y')}|.$$

Now $\beta(x, y - y')$ depends linearly on x, so $\mathbb{E}_x\omega^{\beta(x,y-y')}$ is zero unless $\omega^{\beta(x,y-y')}$ is constant. Let x_0 be an arbitrary element of $\mathbb{F}_p^n$ and let $\beta'(x, y) = \beta(x, y) - \beta(x_0, y)$. Then β' also has rank at least r, and if $\beta(x, y)$ is constant in x then $\beta'(x, y)$ is zero for every x.

Therefore, $\mathbb{E}_x\omega^{\beta(x,y-y')}$ is zero unless $y-y'$ belongs to the annihilator of β'. Otherwise, it has modulus 1. Since β has rank at least r, the probability, for each y, that $y - y'$ belongs to the annihilator is at most p^{-r}. Therefore,

$$\mathbb{E}_{y,y'}|\mathbb{E}_x\omega^{\beta(x,y-y')}| \leq p^{-r}.$$

The result follows on taking square roots. □

In Section 2.3 we had a condition on the rank of the quadratic factor appearing in the decomposition of f, which enabled us to say that any non-trivial linear combination of quadratic forms had high rank. Forcing the factor to satisfy this condition was precisely what lead to tower-type bounds in Theorem 2.19. Here we do better because we have shown that the additional assumption of uniformity allows us to consider sums of high-rank quadratic phases only. But when we compute the average along a linear system, we need to multiply out a product of sums of high-rank quadratic phases, and hence need to consider their linear combinations. What can we say about their ranks? Precisely nothing, as the example of the two high-rank quadratic phases $\sum_{i=1}^n x_i^2$ and $\sum_{i=1}^n x_i^2 - x_1^2$ shows whose difference has rank 1. However, fortunately it suffices to be able to pick out one bilinear form of high rank in order to evaluate an average. This is the content of the next lemma.

Lemma 2.36. *For each pair $(u, v) \in [d]^2$ let β_{uv} be a bilinear form on $\mathbb{F}_p^n$, and suppose that the rank of β_{uv} is at least r for at least one pair (u, v). Then*

$$\left|\mathbb{E}_{\mathbf{x}}\omega^{\sum_{u,v}\beta_{uv}(x_u,x_v)}\right| \leq p^{-r/2}.$$

Proof. Let us assume first that β_{uu} has rank at least r for some u. If we fix the values of x_v for every $v \neq u$, then the sum in the exponent takes the form $\beta_{uu}(x_u, x_u)+\gamma(x_u)$ for some linear functional γ. Therefore, by Lemma 2.29 the expectation over x_u has modulus at most $p^{-r/2}$. Since this is true for every choice of the other x_v, the whole expectation has modulus at most $p^{-r/2}$.

Now let us assume that β_{uv} has rank at least r for some pair (u, v) with $u \neq v$. This time, let us fix all the variables apart from x_u and x_v. Now the sum in the exponent takes the form

$$2\beta_{uv}(x_u, x_v) + \phi(x_u) + \psi(x_v)$$

so by Lemma 2.35 the expectation over x_u and x_v is at most $p^{-r/2}$. Again, since this is true for every possible choice of the other variables, the whole expectation is at most $p^{-r/2}$. □

In the next two lemmas we shall show that if we have a square-independent system $\mathcal{L}$ and a set of bilinear forms of high rank, then at least one linear combination of these bilinear forms resulting from the average over $\mathcal{L}$ must have fairly high rank. This is the only place in the argument where we use the fact that $\mathcal{L}$ is square-independent.

Lemma 2.37. *Let $\beta_1, \dots, \beta_m$ be bilinear forms of rank at least r. Let B be an invertible $m \times m$ matrix with entries $B_{ij} \in \mathbb{F}_p$. Then at least one of the bilinear forms $\eta_j = \sum_{i=1}^m B_{ij}\beta_i$ has rank at least r/m.*

Proof. It follows from the assumption that B is invertible that $\beta_j = B^{-1}_{ij}\eta_j$ for all $j = 1, \dots, m$. But the rank of a linear combination of the η_j is clearly at most the sum of the ranks of the η_j. Hence there must exist an index j for which η_j has rank at least r/m. □

Lemma 2.38. *Suppose that $\mathcal{L}$ is a square-independent system consisting of linear forms $L_i(\mathbf{x}) = \sum_{u=1}^d c_{iu}x_u$ for $i = 1, \dots, m$. Suppose that for each $i = 1, \dots, m$, each of the (not necessarily distinct) bilinear forms β_i has rank at least r. Then at least one of the forms $\beta_{uv} := \sum_{i=1}^m c_{iu}c_{iv}\beta_i$ has rank at least r/m.*

Proof. For each $i = 1, \dots, m$, let C_i denote the matrix $(c_{iu}c_{iv})_{u,v}$. Square-independence implies that the matrices C_i are linearly independent over $\mathbb{F}_p$. Now consider the $d^2 \times m$ matrix whose $((u, v), i)$ entry is $c_{iu}c_{iv}$. The columns of this matrix are the matrices $C_1, \dots, C_m$. The rows are the vectors $C_{uv} = (c_{1u}c_{1v}, c_{2u}c_{2v}, \dots, c_{mu}c_{mv})$. Since row-rank equals column-rank, we can find a collection of m linearly independent vectors C_{uv}. It now suffices to apply Lemma 2.37 to the corresponding bilinear forms $\beta_{uv} = \sum_{i=1}^m (C_{uv})_i\beta_i$ to obtain the result. □

2.4.4 Proof of Theorem 2.21

We are now in a position to put the technical results from the preceding two subsections together to give an improved bound for Theorem 2.19.

Proof of Theorem 2.21. Given $\epsilon > 0$ and $f : \mathbb{F}_p^n \to [-1,1]$ with $\|f\|_{U^2} \leq c$ (where c will be chosen in terms of the parameter ϵ later), we first apply Theorem 2.34 with $\delta_1 := \epsilon/(68m)$ to obtain a decomposition

$$f = f_1 + g_1 + h_1,$$

where $f_1 = \sum_j \lambda_j \omega^{q_j^{(1)}}$ with $\sum_j |\lambda_j| \leq 17M_1$, $\|g_1\|_1 \leq 17\delta_1$ and $\|h_1\|_{U^3} \leq 17\delta_1$. We have carefully ensured that each form $q_j^{(i)}$ has rank at last R_0 for some R_0 to be chosen later. M_1 is a function of δ_1 only, and can be taken to equal $\exp(C\delta^{-2C})$. Recall that we want to show that

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m f(L_i(\mathbf{x}))$$

is bounded above in modulus by ϵ for a sufficiently uniform function f. We first replace the last f in the product by $g_1 + h_1$. The product involving g_1 yields an error term of $17\delta_1$ since all the remaining factors have L_∞-norm bounded by 1, while the product involving h_1 yields an error of $17\delta_1$ by Theorem 2.4. Our choice of δ_1 implies that the sum of these two errors is at most $\epsilon/(2m)$.

Now we apply Theorem 2.34 again, this time with $\delta_2 := \epsilon/(68mM_1)$, to obtain a decomposition

$$f = f_2 + g_2 + h_2,$$

where $f_2 = \sum_j \lambda_j \omega^{q_j^{(2)}}$ with $\sum_j |\lambda_j| \leq 17M_2$, $\|g_2\|_1 \leq 17\delta_2$ and $\|h_2\|_{U^3} \leq 17\delta_2$. When replacing f with $g_2 + h_2$, the product involving g_2 now contributes an error term of at most $17\delta_2 M_1$ (since $\|f_1\|_\infty \leq M_1$). In order to estimate the contribution from the product involving h_2, we require a slight generalization of Theorem 2.4 to functions whose L_∞-norm is bounded, but not necessarily by the constant 1. The following statement follows straightforwardly by applying Theorem 2.4 to the functions $g_i := f_i/\|f_i\|_\infty$.

Theorem 2.39. *Let $f_1, \dots, f_m$ be functions from $\mathbb{F}_p^n$ with $\|f_i\|_\infty \leq \kappa_i$ for each i, and let $\mathcal{L}$ be a linear system of CS-complexity 2 consisting of m forms in d variables. Then*

$$\left| \mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m f_i(L_i(\mathbf{x})) \right| \leq \min_i \|f_i\|_{U^3} \prod_{j\neq i} \kappa_j.$$

It follows that the contribution from the product involving h_2 is bounded above by $17\delta_2 M_1$. Therefore the total error incurred is at most $34\delta_2 M_1$, which is at most $\epsilon/(2m)$ by our choice of δ_2.

When we apply Theorem 2.34 to the kth instance of f in the product, we need to do

so with δ_k satisfying $34\delta_k M_1 \dots M_{k-1} \leq \epsilon/(2m)$ for $k = 2, \dots, m$. This ensures that the initial average can be replaced by

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m f_i(L_i(\mathbf{x}))$$

with an error of at most $\epsilon/2$. Since each M_k is exponential in δ_k^{-1}, and since δ_1 was chosen proportional to ϵ, it is easy to see that M_m will be bounded above by a tower of exponentials of ϵ^{-1} of height $m-1$.

We now concentrate on estimating the average over the product of the f_i, which we recall was of the form $\sum_j \lambda_j^{(i)} \omega^{q_j^{(i)}}$ with $\sum_j |\lambda_j^{(i)}| \leq 17M_i$. Moreover, each $q_j^{(i)}$ had rank at least R_0. We can therefore write

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m f_i(L_i(\mathbf{x})) = \sum_{j_1,\dots,j_m} \lambda_{j_1}^{(1)} \dots \lambda_{j_m}^{(m)} \, \mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \, \omega^{\sum_{i=1}^m q_{j_i}^{(i)}(L_i(\mathbf{x}))}.$$

From now on we shall fix a choice of $j_1, \dots, j_m$, and simply write Q_i for the quadratic form $q_{j_i}^{(i)}$ as well as β_i for the associated bilinear form. Using the co-ordinatewise representation $\sum_{u=1}^d c_{iu}x_u$ of $L_i(\mathbf{x})$, the expectation over $\mathbf{x}$ in the preceding expression becomes

$$\mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \, \omega^{\sum_{u,v=1}^d \sum_{i=1}^m c_{iu}c_{iv}\beta_i(x_u,x_v)} = \mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \, \omega^{\sum_{u,v=1}^d \beta_{uv}(x_u,x_v)},$$

where we have set $\beta_{uv} := \sum_{i=1}^m c_{iu}c_{iv}\beta_i$. At this point we are going to make use of our results from Section 2.4.3. In particular, since each β_i has rank at least R_0 and the $L_1, \dots, L_m$ are square-independent, we can apply Lemma 2.37 to conclude that at least on the forms β_{uv} has rank at least R_0/m. Lemma 2.36 then tells us that the expectation is bounded above in modulus by $p^{-R_0/(2m)}$. It therefore follows that

$$\left| \mathbb{E}_{\mathbf{x}\in(\mathbb{F}_p^n)^d} \prod_{i=1}^m f_i(L_i(\mathbf{x})) \right| \leq p^{-R_0/(2m)} \prod_{i=1}^m M_i,$$

and with hindsight we choose R_0 to be such that $2M_m^m \leq \epsilon p^{R_0/(2m)}$ in every application of Theorem 2.34. In order to do so, we require that f satisfy $\|f\|_{U^2} \leq c$ with $c = \delta_m p^{R/4}$, where $R \leq (2^{23}(M_m/\delta_m)^{12})^{M_m/\delta_m} R_0$. We observed earlier that M_m was bounded above by a tower of exponentials of height $m-1$ in ϵ^{-1}, and conclude that c can therefore be taken to be a tower of exponentials of height $m+1$ in ϵ^{-1}. □

2.4.5 Remarks

The improvement in the bounds for Theorem 2.19 which we derived in this section is due to the fact that we were able to make use of the additional assumption that f is uniform to produce a decomposition of f in which only the high-rank quadratic phases played a rôle.

In a forthcoming paper [GW07a], we improve the bound for Theorem 2.19 even further. We show that c can in fact be taken to be a double exponential in ϵ^{-1}. This is achieved by making more refined use of Theorem 2.22, or more precisely, a slightly stronger form of the inverse theorem which Green and Tao [GT05a] mention but do not formally state. What they do state in [GT05a] is the following "localised" version.

Theorem 2.40. *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$. Then there exists a subspace V of $\mathbb{F}_p^n$ of codimension at most $(2/\delta)^{C_p}$, where C_p is a constant that depends only on p, such that*

$$\mathbb{E}_y \|f\|_{u^3(y+V)} \geq (\delta/2)^{C_p},$$

where $\|f\|_{u^3(y+V)}$ denotes the maximum of $|\mathbb{E}_{x \in y+V} f(x)\omega^{-q(x)}|$ taken over all quadratic forms q on $y + V$.

Note that the average maximum correlation we obtain on the coset of a subspace is polynomial in δ^{-1}, at the cost of a polynomial loss in the codimension of this subspace. It is not difficult to see that this implies Theorem 2.22 by extending the quadratic phase to all of $\mathbb{F}_p^n$, which can be achieved in a number of ways.

On closer inspection, Theorem 2.40 says that for each y, we can find a local quadratic phase q_y defined on $y + V$ such that $|\mathbb{E}_{x \in y+V} f(x)\omega^{q_y(x)}|$ is at least $(\delta/2)^{C_p}$. In fact, it turns out (and is remarked upon in [GT05a]) that we can do this in such a way that the quadratic parts of the quadratic phase functions q_y are all the same.

This observation allows us to prove a version of Theorem 2.34 in which each quadratic phase is replaced by a so-called *quadratic average* of the form $Q(x) = \mathbb{E}_{y \in x-V} \omega^{q_y(x)}$, where each $q_y(x)$ has the form $q(x - y) + \phi_y(x - y)$ for some quadratic function $q : V \to \mathbb{F}_p$ and some linear functionals $\phi_y : V \to \mathbb{F}_p$. By arguments similar to the ones in Section 2.4.2 each such average can be taken to be of high rank.

It turns out that this more local approach also generalizes more easily to $\mathbb{Z}_N$ as global correlation is too much to hope for in this setting (see the remarks in Section 2.3.2). In $\mathbb{Z}_N$ we do not have the vector space structure of $\mathbb{F}_p^n$ and its plentiful supply of

subspaces at our disposal. Instead, we need to make do with so-called *Bohr sets*, which mimic an approximate subgroup structure. Almost all local versions of the lemmas proved in this section have analogues in $\mathbb{Z}_N$ when one replaces a subspace V by a Bohr neighbourhood B. In particular, one can define quadratic averages with Bohr sets as their bases. One of the main difficulties is to establish a meaningful definition of the rank of a quadratic phase relative to the Bohr set on which it is defined. The details are due to appear in [GW07a].

Acknowledgements. The author would like to thank Tim Gowers for his commitment to the collaboration that led to the results discussed in this chapter.

2.5 The Ergodic Analogue

In this expository section we outline the analogies between two recent preprints by Leibman [Lei07] and Gowers and the author [GW07b]. Both papers independently describe two manifestations of the same phenomenon, the former in the context of ergodic theory and the latter in arithmetic combinatorics. In their respective settings, they address the question after the degree of the minimal characteristic factor of a multiple ergodic average along a system of linear forms, or the minimal degree of uniformity needed to accurately count solutions to the corresponding system of linear equations. The exposition is aimed at readers with a combinatorics background and limited prior exposure to ergodic theory.

In [GW07b] (and indeed, the earlier sections of this chapter) we investigated the following question: for which types of systems of linear equations can we guarantee that any subset of $\mathbb{F}_p^n$ which is uniform of degree k contains the "expected" number of solutions, that is, the number of solutions one would expect in a random subset of the same density. By *uniform of degree k* we mean that the balanced function of the set is small in the so-called U^{k+1}-norm, which originated in the work of Gowers on Szemerédi's Theorem for long arithmetic progressions [Gow01] and will be recalled at the start of Section 2.5.2.

To make this question more precise, we developed a new notion of complexity of a linear system which we called the *true complexity*. For example, we defined a system of linear forms $\mathcal{L} = (L_1, \ldots, L_m)$ on $(\mathbb{F}_p^n)^d$ to have *true complexity* 1 if and only if it contains the "correct" number of solutions in any uniform set. More generally, we say a system has *true complexity k* if k is the least integer such that the average over the linear forms is governed by the U^{k+1}-norm.

We then proceeded to show, under one additional assumption, that linear systems of true complexity 1 are precisely those for which the squares of the linear forms defining the system are linearly independent. It is straightforward to show that square-independence is a necessary condition for true complexity 1 by adapting a well-known construction used to show that there are uniform sets which contain too many 4-term progressions. This was achieved in Section 2.3.1. The precise qualitative version of the fact that square-independence is also a sufficient condition for true complexity 1 (again, under one additional assumption) was stated as follows.

Theorem 2.19. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let $f : \mathbb{F}_p^n \to [-1, 1]$ satisfy $\|f\|_{U^2} \leq c$. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables of Cauchy-Schwarz complexity at most 2. Then*

$$\left| \mathbb{E}_{x_1,\dots,x_d \in \mathbb{F}_p^n} \prod_{i=1}^{m} f(L_i(x_1, \dots, x_d) \right| \leq \epsilon.$$

In other words, $\mathcal{L}$ has true complexity 1.

Recall that the *Cauchy-Schwarz complexity* of a linear system described precisely the condition that enabled us to prove the following theorem via a simple Cauchy-Schwarz argument.

Theorem 2.4. *Let $f_1, \dots, f_m$ be functions from $\mathbb{F}_p^n$ to $[-1, 1]$, and let $\mathcal{L}$ be a linear system of Cauchy-Schwarz complexity k consisting of m forms in d variables. Then*

$$\left| \mathbb{E}_{x_1,\dots,x_d \in \mathbb{Z}_N} \prod_{i=1}^{m} f_i(L_i(x_1, ..., x_d)) \right| \leq \min_i \|f_i\|_{U^{k+1}}.$$

The additional hypothesis of Cauchy-Schwarz complexity 2 in Theorem 2.19 is a technical yet important condition. It stems from the fact that when considering an average such as the one in Theorem 2.19, it is convenient to decompose the function f into a quadratically structured part and a part that is small in U^3, and then Theorem 2.4 tells us that for systems of Cauchy-Schwarz complexity 2, only the contribution from the structured part needs to be considered. Unfortunately, we do not currently have such a decomposition for higher-order U^k-norms, hence the restriction to systems of Cauchy-Schwarz complexity 2.

Example 2.41. *Linear systems that were previously thought to require quadratic Fourier analysis but that have been shown to be governed by the U^2-norm by Theorem 2.19 include the systems $\mathcal{L}_1 = (x, n, m, x + n + m, x + 2n - m, x + 2m - n)$ and the translation-invariant $\mathcal{L}_2 = (x, x + n, x + m, x + n + m, x + n - m, x + m - n)$.*

From Theorem 2.19 we deduced the following corollary concerning the number of solutions of a square-independent linear system in uniform subsets of $\mathbb{F}_p^n$.

Corollary 2.20. *For every $\epsilon > 0$ there exists $c > 0$ with the following property. Let A be a subset of $\mathbb{F}_p^n$ of density α whose balanced function has U^2-norm bounded by c. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a square-independent system of linear forms in d variables, with Cauchy-Schwarz complexity at most 2. Let $(x_1, \dots, x_d)$ be a random element of $(\mathbb{F}_p^n)^d$. Then the probability that $L_i(x_1, \dots, x_d) \in A$ for every i differs from α^m by at most ϵ.*

For a detailed discussion of the context of these results and their (conjectured) higher-order generalizations the reader is referred to the introduction of this chapter.

Let us now have a look at the ergodic world. Ergodic theorists are concerned with the convergence (in L^∞, L^1 or L^2) of *multiple ergodic averages* of the form

$$\frac{1}{N^d} \sum_{n_1,\dots,n_d=1}^{N} T^{p_1(n_1,\dots,n_d)} f_1(x)\ T^{p_2(n_1,\dots,n_d)} f_2(x)\ \dots\ T^{p_m(n_1,\dots,n_d)} f_m(x),$$

where T is a measure preserving transformation on a probability measure space $(X, \mathcal{B}, \mu)$, the functions f_i belong to $L^\infty(\mu)$ and the p_i are polynomials on $\mathbb{Z}^d$. For example, the case where $d = 1$, $p_j(n) = jn$ for $j = 1, \dots, k$ and f_i equals the indicator function 1_A of a set $A \in \mathcal{B}$ with $\mu(A) > 0$ appeared in Furstenberg's proof of Szemerédi's Theorem [Fur77], which states that any subset of $\mathbb{Z}$ of positive upper density contains an arithmetic progression of length k. More precisely, Furstenberg proved that the $\liminf_{N\to\infty}$ of the average

$$\frac{1}{N^d} \sum_{n_1,\dots,n_d=1}^{N} \int 1_A(x) T^n 1_A(x)\ T^{2n} 1_A(x)\ \dots\ T^{kn} 1_A(x) d\mu(x), \tag{2.1}$$

was strictly greater than 0. Ergodic theorists were the first to prove a multi-dimensional Szemerédi Theorem, as well as polynomial extensions [BL96] which remain beyond the reach of arithmetic combinatorics to date. However, the fact that only translation-invariant systems can be studied using such averages and, more importantly, the lack of quantitative bounds (but see [Tao06]) pose serious limitations and more than justify the search for alternative approaches via arithmetic combinatorics.

The question in ergodic theory which is analogous to the one we have been studying in this chapter concerns so-called *characteristic factors* for ergodic averages of the

form

$$\frac{1}{N^d} \sum_{n_1,\dots,n_d=1}^{N} T^{L_1(n_1,\dots,n_d)} f_1(x)\; T^{L_2(n_1,\dots,n_d)} f_2(x)\; \dots\; T^{L_m(n_1,\dots,n_d)} f_m(x),$$

where T is a measure-preserving map on a probability measure space $(X, \mathcal{B}, \mu)$, the functions f_i belong to $L^\infty(\mu)$ and the L_i are linear forms on $\mathbb{Z}^d$. Very roughly speaking, a characteristic factor is a system onto which one can project without losing any information about the convergence of the average under consideration. The aim is to find characteristic factors which possess enough structure to allow one to establish convergence of the above average in a rather explicit way. For example, it was shown by Host and Kra [HK05] and Ziegler [Zie07] independently that when the linear forms $L_1, \dots, L_m$ describe an arithmetic progression of length m, there exists a characteristic factor for the corresponding average which is isomorphic to an inverse limit of a sequence of $(m-2)$-step nilsystems. For $m = 4$, these very structured objects are closely related to the quadratic factor introduced in Section 2.3.2, on which computations can be performed rather straightforwardly. After these remarks it should not be surprising that there is a notion of *degree* associated with a characteristic factor. What we have called the true complexity of a linear system is closely analogous to the degree of the minimal characteristic factor.

In a recent preprint [Lei07], Leibman characterizes the degree of the minimal characteristic factor for general linear as well as certain polynomial systems. Using his examples and our terminology, the system given by $\mathcal{L}_3 = (x+n+m, x+2n+4m, x+3n+9m, x+4n+16m, x+5n+25m, x+6n+36m)$ has true complexity strictly greater than 1 (in fact, equal to 2), while the ever so slightly different $\mathcal{L}_4 = (x+n+m, x+2n+4m, x+3n+9m, x+4n+16m, x+5n+25m, x+6n+37m)$ has true complexity 1. The crucial distinguishing factor of $\mathcal{L}_5$ is that its squares are independent, or, as Leibman puts it, that the six vectors $(1,1,1,1,1,1)$, $(1, c_1, c_2, \dots, c_5)$, $(1, d_1, d_2, \dots, d_5)$, $(1, c_1^2, c_2^2, \dots, c_5^2)$, $(1, d_1^2, d_2^2, \dots, d_5^2)$ and $(1, c_1 d_1, c_2 d_2, \dots, c_5 d_5)$ span $\mathbb{R}^6$. (Here c_i, d_i are the cofficients of n, m, respectively, in the linear form $i+1$. Note that the special form of the ergodic average forces one to consider translation-invariant systems only, which leads to a formulation of square-independence that is particular to systems where one variable has coefficient 1 in all linear forms.)

In his proof of Szemerédi's Theorem, Furstenberg [Fur77] developed an important tool known as the *Correspondence Principle*, which allowed him to deduce Szemerédi's combinatorial statement from the recurrence properties of a dynamical system. While the Correspondence Principle has allowed us to deduce many a combinatorial appli-

cation from results in ergodic theory, our result in the $\mathbb{Z}_N$ case does not appear to follow from Leibman's result by a standard application. We shall briefly discuss this issue in the final section.

For an excellent introduction to ergodic theory and its connections with additive combinatorics we refer the interested reader to [Kra06]. In this short note, we make no attempt to give a comprehensive overview of the subject but confine ourselves to describing the concepts needed to understand the parallels between [Lei07] and [GW07b].

2.5.1 Basic Concepts in Ergodic Theory

Ergodic theory is the study of the dynamical behaviour of certain *probability measure preserving systems*.

Definition 2.42. *A* probability measure-preserving system *is a quadruple* $(X, \mathcal{X}, \mu, T)$ *where* (X, μ) *is a probability space and* $T : X \to X$ *is a bijective, measurable, measure-preserving transformation. This means that for all* $A \in \mathcal{X}$, $T^{-1}A \in \mathcal{X}$ *and* $\mu(T^{-1}A) = \mu(A)$.

For our purposes, we may always assume that the system $(X, \mathcal{X}, \mu, T)$ is an *ergodic* system, which means that the only sets which are left invariant under the action of T have measure 0 or are in fact the whole space. This assumption is justified by a principle called *ergodic decomposition*, which says, in very rough terms, that one can decompose any measure preserving system into a number of ergodic ones. For a clear explanation see page 17 of [CFS82].

Example 2.43. *Let* $X = \mathbb{T}$ *be equipped with the Borel* σ*-algebra* $\mathcal{X}$ *and Haar measure* μ. *Take* $T : X \to X$ *to be the rotation* $Tx = x + \alpha \bmod 1$ *for some* $\alpha \in \mathbb{R}$. *The measure preserving system* $(X, \mathcal{X}, \mu, T)$ *is ergodic if and only if* α *is irrational.*

The next important notion we need is that of a *factor* of a measure preserving system, that is, a subsystem which has the obvious desirable properties.

Definition 2.44. *A* factor *of a system* $(X, \mathcal{X}, \mu, T)$ *can be defined in several equivalent ways. Any* T*-invariant sub-*σ*-algebra* $\mathcal{Y}$ *of* $\mathcal{X}$ *is a factor of* $\mathcal{X}$. *A factor can also be thought of as a system* $(Y, \mathcal{Y}, \nu, S)$ *and a measurable map* $\pi : X \to Y$, *the* factor map, *such that* $\mu \circ \pi^{-1} = \nu$ *and* $S \circ \pi = \pi \circ T$ *for* μ*-almost every* $x \in X$.

We shall be using the same letter T to denote both the transformation in the original system and the transformation on the factor.

Example 2.45. *Let $X = \mathbb{T} \times \mathbb{T}$ be equipped with the Borel σ-algebra $\mathcal{X}$ and Haar measure μ. Take $T : X \to X$ to be the transformation $T(x, y) = (x + \alpha, y + x)$ for some $\alpha \in \mathbb{R}$. Then $\mathbb{T}$ together with the rotation $x \mapsto x + \alpha$ is a factor of X.*

As already mentioned in the introduction, in order to study multiple ergodic averages it is useful to work on a so-called *characteristic factor*. A factor is said to be characteristic for an ergodic average if we can study the average of the projection onto the factor without losing any information about the convergence of the average. In other words, focusing on L^2-convergence we make the following definition.

Definition 2.46. *We say a factor Y of X is* characteristic *for the average*

$$\frac{1}{N^d} \sum_{n_1,\dots,n_d=1}^{N} T^{p_1(n_1,\dots,n_d)} f_1(x) T^{p_2(n_1,\dots,n_d)} f_2(x) \dots T^{p_m(n_1,\dots,n_d)} f_m(x)$$

if and only if the difference with

$$\frac{1}{N^d} \sum_{n_1,\dots,n_d=1}^{N} T^{p_1(n_1,\dots,n_d)} \mathbb{E}(f_1|\mathcal{Y})(x) T^{p_2(n_1,\dots,n_d)} \mathbb{E}(f_2|\mathcal{Y})(x) \dots T^{p_m(n_1,\dots,n_d)} \mathbb{E}(f_m|\mathcal{Y})(x)$$

tends to 0 in $L^2(\mu)$.

Equivalently, Y is characteristic for X if the average converges to 0 whenever $\mathbb{E}(f_i|\mathcal{Y}) = 0$ for at least one $i = 1, 2, \dots, m$. Here we have written $\mathbb{E}(f|\mathcal{Y})$ for the *conditional expectation* of f with respect to the factor $\mathcal{Y}$, that is, the usual Hilbert space projection of f onto the sub-σ-algebra $\mathcal{Y}$.

We have already mentioned that in arithmetic combinatorics, in order to show that a given linear system is governed by *some* uniformity norm all one needs is the Cauchy-Schwarz Inequality, multiple applications of which yield Theorem 2.4. We shall see that in ergodic theory, in order to show that *some* factor is characteristic for a particular average, one uses a multi-dimensional version of *Van der Corput's Lemma*, which is essentially an infinitary version of the Cauchy-Schwarz Inequality (see page 13 of [Kra06] for a standard proof which makes this obvious). We shall state Van der Corput's Lemma in dimension 3 only, for simplicity and because it is sufficient to deal with the examples we shall focus on shortly.

Proposition 2.47. *Suppose that $\{u_{n_1,n_2,n_3} : n_1, n_2, n_3 \in \mathbb{Z}\}$ form a bounded triple sequence of vectors in a Hilbert space. If*

$$\lim_{K\to\infty} \frac{1}{K^3} \sum_{k_1,k_2,k_3=0}^{K-1} \left| \lim_{N-M\to\infty} \frac{1}{(N-M)^3} \sum_{n_1,n_2,n_3=M}^{N-1} < u_{n_1,n_2,n_3}, u_{n_1+k_1,n_2+k_2,n_3+k_3} > \right|$$

equals zero, then

$$\lim_{N-M\to\infty} \left\| \frac{1}{(N-M)^3} \sum_{n_1,n_2,n_3=M}^{N-1} u_{n_1,n_2,n_3} \right\| = 0.$$

This concludes the preliminaries. In the next section we will have a closer look at how to define characteristic factors for linear systems, and collect some results about their structural properties.

2.5.2 Gowers Norms, Host-Kra Factors and Nilmanifolds

Recall the definition of higher-degree uniformity norms in arithmetic combinatorics, which originated in Gowers's work on Szemerédi's Theorem for longer progressions [Gow01].

Definition 2.2. *Let G be a 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*

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

where $C^{|\boldsymbol{\omega}|} f = f$ if $\sum_i \omega_i$ is even and $\overline{f}$ otherwise.

By a special case of Proposition 2.4, which was in fact proved in [Gow01], the U^{k+1}-norm governs the average over arithmetic progressions of length k (this is because progressions of length k have Cauchy-Schwarz complexity $k-2$). A family of semi-norms analogous to the U^k-norms have recently appeared in the work of Host and Kra [HK05].

Definition 2.48. *For $f \in L^\infty(\mu)$ and $k \in \mathbb{N}$, we define the* Host-Kra semi-norms *as*

$$\|\!|f|\!\|_k := \left(\int_{X^{[k]}} f \otimes \cdots \otimes f d\mu^{[k]} \right)^{1/k}.$$

Of course we haven't actually defined the measure $\mu^{[k]}$ yet, nor the space $X^{[k]}$ over which we integrate. The definition below looks rather off-putting, and we invite the reader to skip the details on first reading. However, even on more superficial inspection it can be intuited that the construction of the measure μ_k encodes the structure of combinatorial cubes of dimension k.

Definition 2.49. *Let $X^{[k]} = X^{2^k}$ and define $T^{[k]}\colon X^{[k]} \to X^{[k]}$ by $T^{[k]} = T \times \cdots \times T$ (2^k times). We write a point $\mathbf{x} \in X^{[k]}$ as $\mathbf{x} = (x_\epsilon)_{\epsilon \in \{0,1\}^k}$ and make the natural identification of $X^{[k+1]}$ with $X^{[k]} \times X^{[k]}$, writing $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$ for a point of $X^{[k+1]}$, with $\mathbf{x}', \mathbf{x}'' \in X^{[k]}$. By induction, we define a measure $\mu^{[k]}$ on $X^{[k]}$ invariant under $T^{[k]}$. Set $\mu^{[0]} := \mu$. Let $\mathcal{I}^{[k]}$ be the invariant σ-algebra of $(X^{[k]}, \mathcal{X}^{[k]}, \mu^{[k]}, T^{[k]})$. Then $\mu^{[k+1]}$ is defined to be the* relatively independent joining *of $\mu^{[k]}$ with itself over $\mathcal{I}^{[k]}$, meaning that if F and G are bounded functions on $X^{[k]}$,*

$$\int_{X^{[k+1]}} F(\mathbf{x}') \cdot G(\mathbf{x}'')\, d\mu^{[k+1]}(\mathbf{x}) = \int_{X^{[k]}} \mathbb{E}(F|\mathcal{I}^{[k]})(\mathbf{y}) \cdot \mathbb{E}(G|\mathcal{I}^{[k]})(\mathbf{y})\, d\mu^{[k]}(\mathbf{y}) \ .$$

Since $(X, \mathcal{X}, \mu, T)$ is assumed to be ergodic, $\mathcal{I}^{[0]}$ is trivial and $\mu^{[1]} = \mu \times \mu$. Just like the U^k-norms in arithmetic combinatorics, these seminorms are nested, in the sense that they satisfy

$$\|\!|f|\!\|_1 \leq \|\!|f|\!\|_2 \leq \cdots \leq \|\!|f|\!\|_k \leq \cdots \leq \|f\|_\infty,$$

and a *Gowers-Cauchy-Schwarz*-type inequality holds, that is,

$$\Big| \int \prod_{\epsilon \in \{0,1\}^k} f_\epsilon(x_\epsilon) d\mu^{[k]} \Big| \leq \prod_{\epsilon \in \{0,1\}^k} \|\!|f_\epsilon|\!\|_k,$$

which can be used to show that $\|\!|.|\!\|_k$ is indeed a semi-norm on $L^\infty(\mu)$. Moreover, it can be checked that just like the U^k-norms, the semi-norms $\|\!|.|\!\|_k$ can be defined inductively via the formula

$$\|\!|f|\!\|_{k+1}^{2^{k+1}} = \int_{I_k} \mathbb{E}(f^{\otimes 2^k}|\mathcal{I}^{[k]})^2 d\mu^{[k]}.$$

Together with the *Von Neumann Ergodic Theorem*, which states that for an ergodic system $(X, \mathcal{X}, \mu, T)$ and $f \in L^2(\mu)$, the L^2-limit as N tends to infinity of $\frac{1}{N}\sum_{n=1}^N f(T^n x)$ is the constant function $\int f d\mu$, this can be rewritten as

$$\|\!|f|\!\|_{k+1}^{2^{k+1}} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \|\!|f \cdot T^n f|\!\|_k^{2^k}.$$

This fact in turn is a useful ingredient in the proof of Proposition 2.50 below, which represents the analogue of Theorem 2.4 and will be discussed in more detail at the start of Section 2.5.3. We refer the keen reader to page 20 of [Kra06] for a proof in the case of arithmetic progressions.

Proposition 2.50. *Assume that $(X, \mathcal{X}, \mu, T)$ is ergodic and let $d, k, m \in \mathbb{N}$. Suppose*

$\|f_i\|_\infty \leq 1$ *for all* $i = 1, 2, \ldots, m$, *and that the system* $\mathcal{L} = (L_1, L_2, \ldots, L_m)$ *in* d *variables has Cauchy-Schwarz complexity* k. *Then*

$$\limsup_{N\to\infty} \left\| \frac{1}{N^d} \sum_{n_1,n_2,\ldots,n_d=0}^{N-1} T^{L_1(n_1,\ldots,n_d)} f_1(x) \ \ldots \ T^{L_m(n_1,\ldots,n_d)} f_m(x) \right\|_2 \ll \min_{l=1,2,\ldots,m} \|\!| f_l |\!\|_{k+1}.$$

With the definitions in place, it is now straightforward to define the sequence of so-called *Host-Kra factors*, which first appeared in [HK05].

Definition 2.51. *Given a measure-preserving system* $(X, \mathcal{X}, \mu, T)$, *there is a nested sequence of factors* $\mathcal{Z}_k$ *of* X *such that for any bounded function* f *on* X

$$\|\!| f |\!\|_{k+1} = 0 \text{ if and only if } \mathbb{E}(f|\mathcal{Z}_k) = 0.$$

It follows straight from this definition combined with Proposition 2.50 that the factors $\mathcal{Z}_k$ are characteristic for systems of Cauchy-Schwarz complexity k. In particular, $\mathcal{Z}_1$ is characteristic for the average along 3-term progressions, while the factor $\mathcal{Z}_2$ controls 4-term progressions.

Let us pause for a moment to compare this situation with our combinatorial approach: In order to concentrate on the structured part in arithmetic combinatorics, we needed a deep U^3*-inverse theorem* which allowed us to decompose any bounded function into a quadratically structured and a quadratically uniform part. In ergodic theory, the fact that the factors $\mathcal{Z}_k$ are characteristic for systems of Cauchy-Schwarz complexity k follows straight from the definition and Proposition 2.50. The real difficulty lies in giving a geometric description of the factors defined in this very "soft" way.

Having said that, it is not hard to see that the first factor in this sequence $\mathcal{Z}_1$ corresponds to the classical *Kronecker factor*. There are many equivalent ways of describing the Kronecker factor $\mathcal{K}$ of a measure-preserving system which do not use the semi-norm $\|\!|.|\!\|_2$.

- $\mathcal{K}$ is the largest abelian group rotation factor.
- $\mathcal{K}$ is the smallest sub-σ-algebra of $\mathcal{X}$ with the property that every member of $\mathcal{I}^{[1]}$ is measurable with respect to $\mathcal{K} \otimes \mathcal{K}$.
- The measure $\mu^{[2]}$ is relatively independent with respect to $\mathcal{K}^4$ and the factor $\mathcal{K}$ of X is minimal with this property.

Example 2.52. *Let* $X = \mathbb{T} \times \mathbb{T}$ *be equipped with the Borel* σ*-algebra and Haar*

measure. Fix $\alpha \in \mathbb{T}$ *and define* $T\colon X \to X$ *by*

$$T(x, y) = (x + \alpha, y + x)$$

The system is ergodic if and only if $\alpha \notin \mathbb{Q}$*, and it is not isomorphic to a group rotation. The Kronecker factor of* X *is the factor* $\mathbb{T}$ *equipped with the rotation* $x \mapsto x + \alpha$*. We say* X *is a* skew extension *of* $\mathbb{T}$ *by another copy of* $\mathbb{T}$*.*

It is not hard to see directly that $\|\!|f|\!\|_2$ equals the l^4-norm of the Fourier transform of f projected onto the Kronecker factor, and that the Kronecker factor is characteristic for studying ergodic averages along 3-term progressions (see page 21 of [Kra06]). This corresponds to saying that ordinary Fourier analysis suffices in this case.

In order to study longer progressions, higher-order factors are needed. The *Conze-Lesigne factor*, which in modern terminology represents the second level in the series of Host-Kra factors, was introduced by Conze and Lesigne in a series of papers [CL84], [CL87], [CL88]. Equivalent and more explicit descriptions were given by Rudolph [Rud95] and Host and Kra [HK01], and we refer the interested reader to these works for more detail.

It turns out that every Conze-Lesigne system is the inverse limit of a sequence of 2-step nilsystems (see Theorem 18 in [HK04]). More generally, Host and Kra proved the following deep structure theorem in [HK05]:

Theorem 2.53. *For each integer* k*, the factor* $\mathcal{Z}_k$ *is isomorphic to an inverse limit of* k*-step nilsystems.*

In order to make use of this structure theorem, we need to understand what a *k-step nilsystem* is, as well as what it means to be an *inverse limit* of a sequence of such systems.

Definition 2.54. *Let* G *be a group. If* $g, h \in G$*, let* $[g, h] = g^{-1}h^{-1}gh$ *denote the* commutator *of* g *and* h*. If* $A, B \subset G$*, we write* $[A, B]$ *for the subgroup of* G *spanned by* $\{[a, b] : a \in A,\ b \in B\}$*. The* lower central series

$$G = G_1 \supset G_2 \supset \cdots \supset G_j \supset G_{j+1} \supset \ldots$$

of G *is defined by setting* $G_1 = G$ *and* $G_{j+1} = [G, G_j]$ *for* $j \geq 1$*. We say that* G *is* k-step nilpotent *if* $G_{k+1} = \{1_G\}$*. If* G *is a* k*-step nilpotent Lie group and* Γ *is a discrete co-compact subgroup, the compact manifold* $X = G/\Gamma$ *is a* k-step nilmanifold*. The group* G *acts naturally on* X *by left translation, that is if* $a \in G$ *and* $x \in X$*, then the*

translation T_a by a is given by $T_a(x\Gamma) = (ax)\Gamma$. There is a unique Borel probability measure μ (the Haar measure*) on X that is invariant under this action. For a fixed element $a \in G$, we say that the system $(G/\Gamma, \mathcal{G}/\Gamma, T_a, \mu)$ is a k-*step nilsystem.

Important examples of nilsystems include the circle nilflow (Example 2.43, easily seen to be a 1-step nilsystem by setting $G = \mathbb{R}$ and $\Gamma = \mathbb{Z}$ in the above definition), the skew torus (Example 2.52, a primitive 2-step nilsystem), and the *Heisenberg nilflow*, which we shall discuss in Example 2.55 below. More information on these basic examples can be found in both [Kra06] and [GT06c].

Example 2.55. *Let G be the Heisenberg group $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ with multiplication given by*

$$(x, y, z) * (u, v, w) = (x + u, y + v, z + w + xv),$$

which is a 2-step nilpotent Lie group (and is perhaps more easily thought of as the group of upper-diagonal real matrices with 1s on the diagonal). Take the discrete co-compact subgroup $\Gamma = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$, so that $X = G/\Gamma$ is a 2-step nilmanifold. Then the transformation T defined as translation by $(g_1, g_2, g_3) \in G$ together with the Borel σ-algebra $\mathcal{X}$ and Haar measure μ defines a 2-step nilsystem. This system is ergodic if and only if g_1 and g_2 are rationally independent. The compact abelian group $G/G_2\Gamma$ is isomorphic to $\mathbb{T}^2$, and the rotation by (g_1, g_2) on $\mathbb{T}^2$ is ergodic. This factor of X represents the Kronecker factor $\mathcal{Z}_1$.

It is not terribly important to us to know what exactly an *inverse limit* is, since it behaves well enough to always allow us to concentrate on a single nilmanifold, but for the sake of completeness we present the definition below.

Definition 2.56. *The system $(X, \mathcal{X}, \mu, T)$ is an* inverse limit *of a sequence of factors $\{(X_j, \mathcal{X}_j, \mu_j, T)\}_{j\in\mathbb{N}}$ if $\{\mathcal{X}_j\}_{j\in\mathbb{N}}$ is an increasing sequence of T-invariant sub-σ-algebras such that $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to sets of measure. If each system $(X_j, \mathcal{X}_j, \mu_j, T)$ is isomorphic to a k-step nilsystem, then $(X, \mathcal{X}, \mu, T)$ is an* inverse limit of k-step nilsystems.

As indicated earlier, nilmanifolds possess an enormous amount of structure, so by reducing to the study of averages on nilmanifolds via Proposition 2.50 and Theorem 2.53, many questions about the convergence of ergodic averages on abstract measure-preserving systems become explicit computations. Before we look at the general case, however, let us consider in more detail the simple 2-step nilsystem that is the skew torus.

2.5.3 A Square-Independent System on the Skew Torus

From now on we shall focus our attention on one of the examples of square-independent systems which was mentioned in the introduction of Section 2.5, namely

$$\mathcal{L}_2 = (x, x+n, x+m, x+n+m, x+n-m, x+m-n).$$

It is easy to check that this linear system has Cauchy-Schwarz complexity 2 and is translation invariant. First, we shall see that the factor $\mathcal{Z}_2$ is characteristic for the average along $\mathcal{L}_2$, which is a special case of Proposition 2.50. The proof uses Van der Corput's Lemma 2.47 and the inductive definition of the semi-norm $\|\!|.|\!\|_2$ given at the start of Section 2.5.2. It follows from a refined analysis of Proposition 5 in [Lei04] and is left as an exercise.

Proposition 2.57. *Suppose $(X, \mathcal{X}, \mu, T)$ is an ergodic measure-preserving system, and let $E = \{(1,0), (0,1), (1,1), (1,-1), (-1,1)\}$. If $\|f\|_\infty \leq 1$, then*

$$\limsup_{N\to\infty} \left\| \frac{1}{N^2} \sum_{n,m=1}^{N} \prod_{\epsilon\in E} f \circ T^{\epsilon\cdot n} \right\|_{L^2(\mu)} \ll \|\!|f|\!\|_2.$$

In other words, the factor $\mathcal{Z}_2$ is characteristic for the system $\mathcal{L}_2$.

By Proposition 2.50 and Theorem 2.53 we are now in the fortunate position to know that we can reduce to the case where our system is a 2-step nilmanifold. Our next aim would be to show that, in fact, the Kronecker factor is the minimal characteristic factor for $\mathcal{L}_2$. For illustrative purposes we now focus on the case of the simplest possible 2-step nilsystem only, the skew torus discussed in Example 2.52 of the previous section.

Recall that the skew torus was defined by setting $X = \mathbb{T} \times \mathbb{T}$, equipped with Borel σ-algebra $\mathcal{X}$ and Haar measure μ. We take $T : X \to X$ to be the transformation $T(x,y) = (x+\alpha, y+x)$ for some $\alpha \in \mathbb{R}$. This is a 2-step nilsystem, whose Kronecker factor is $\mathbb{T}$ together with the rotation $x \mapsto x+\alpha$. Iterating the transformation T, we find that the n^{th} iterate is given by the formula

$$T^n(x,y) = (x+n\alpha, y+nx+\frac{n(n+1)}{2}\alpha).$$

It is now not difficult to compute the average explicitly. For example, suppose f is a Riemann integrable function. Standard approximation arguments allow us to reduce to the case of a continuous function, and by Weierstrass approximation and linearity

we are in fact justified in thinking of f as a simple exponential. Inserting the formula for T^n in the average

$$\frac{1}{N^2}\sum_{n,m=1}^{N}\prod_{\epsilon\in E} f\circ T^{\epsilon\cdot n}$$

and replacing each instance of f by an appropriate exponential function, we find that the square-independence of $\mathcal{L}_2$ implies that there is always a non-zero quadratic coefficient of α. This fact combined with the uniform distribution of the fractional part of $n^2\alpha$ allows us to conclude that the orbit of the diagonal $\Delta_X = \{(x, x, \ldots, x) : x \in X\}$ is uniformly distributed on the fibres over the Kronecker factor (in this case, the second co-ordinate). This in turn means that it is in fact possible to project down to the Kronecker factor without affecting the convergence of the limit of the average. Since it would not be instructive to include the full details of this computation, we leave them to the interested reader.

2.5.4 A General 2-Step Nilmanifold

The purpose of this section is to provide some intuition for the general case of a 2-step nilmanifold, and to illustrate what we mean by "parameterising" such a manifold. We shall not attempt to reproduce any proofs, but rather provide a tourist's guide to [Lei07] for the interested reader. We shall assume that we have proved Proposition 2.57 and are therefore able to restrict our attention to a 2-step nilmanifold.

Given an s-step nilmanifold $X = G/\Gamma$, there exists a sequence of natural factors $X = X_s \to X_{s-1} \to \ldots \to X_1 \to X_0 = \{1_X\}$ defined by $X_j = G/(\Gamma G_{j-1})$. For each j, X_j is a j-step nilmanifold. This comes with a sequence of projections $\pi_j : X \to X_j$. In our case $s = 2$, so we are looking at the sequence $X_2 = X \to X_1 = G/(\Gamma G_2) \to X_0 = \{1\}$. The projection π_1 takes the simple form $G/\Gamma \to G/\Gamma G_2$.

We want to show that the factor X_1 is characteristic for the average along $\mathcal{L}_2$ which we were studying in the preceding section. In fact, it is possible to give a completely explicit description of the orbit of the diagonal $\Delta_X = \{(x, x, \ldots, x) : x \in X\}$ under a system of linear actions. For example, for a simple linear system of 5 forms in 2 variables such as $\mathcal{L}_2$, it can be shown that the orbit of the diagonal Δ_X is of the form $\pi^5(H)$ with H a rational subgroup of the form

$$\left\langle \begin{pmatrix} b_0\ b_1^{c_1}\ b_2^{d_1}\ b_3^{c_1^2}\ b_4^{c_1 d_1}\ b_5^{d_1^2} \\ \vdots \\ b_0\ b_1^{c_5}\ b_2^{d_5}\ b_3^{c_5^2}\ b_4^{c_5 d_5}\ b_5^{d_5^2} \end{pmatrix} : b_0, b_1, b_2 \in G, b_3, b_4, b_5 \in G_2 \right\rangle,$$

where we have written $E = \{(1,0),(0,1),(1,1),(1,-1),(-1,1)\} = \{(c_i,d_i) : i = 1,2,\ldots,5\}$ for the coefficients of n and m in $\mathcal{L}_2$. This is the main content of Proposition 6.3 in [Lei07], which we have illustrated using an adaptation of Example 6.7 in that paper. But for all $i = 1,2,\ldots,5$, we can now rewrite

$$b_0\, b_1^{c_i}\, b_2^{d_i}\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}$$

as a product

$$b_0\, b_1^{c_i}\, b_2^{d_i}\, (a_2^{-d_i}\, a_1^{-c_i}\, a_0^{-1}\, a_0\, a_1^{c_i}\, a_2^{d_2})\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}$$

with $a_0, a_i, a_2 \in G_2$, which in turn can be expressed as

$$(a_0^{-1}b_0)(a_1^{-1}b_1)^{c_i}(a_2^{-1}b_2)^{d_i} a_0\, a_1^{c_i}\, a_2^{d_i}\, b_3^{c_i^2}\, b_4^{c_i d_i}\, b_5^{d_i^2}.$$

This reparametrisation takes place in Corollary 5.8 of [Lei07]. Finally, we know that because the system $\mathcal{L}_2$ is square-independent, the matrix of coefficients

$$\begin{pmatrix} 1 & c_1 & d_1 & c_1^2 & c_1 d_1 & d_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & c_5 & d_5 & c_5^2 & c_5 d_5 & d_5^2 \end{pmatrix}$$

has full rank, and hence the rational subgroup H takes the form

$$\left\langle \begin{pmatrix} b_0\, b_1^{c_1}\, b_2^{d_1} \\ \vdots \\ b_0\, b_1^{c_5}\, b_2^{d_5} \end{pmatrix} : b_0, b_1, b_2 \in G \right\rangle \cdot G_2^5.$$

Since the projection $\pi_1 : X \to X_1$ amounted to nothing more than quotienting out by the commutator subgroup G_2, we see that in fact the factor X_1 is characteristic for a square-independent average.

Note that a very similar parametrisation can be carried out for polynomial orbits, details of which can be found in the later sections of [Lei07].

2.5.5 The Correspondence Principle

It is not clear whether Leibman's ergodic theoretic result has any number theoretic consequences of the form we saw in Corollary 2.20. In general, one uses the following standard tool for transfering ergodic theoretic to combinatorial statements, which

originated in Furstenberg's proof of Szemerédi's Theorem [Fur77] and is now known as the *Correspondence Principle*.

Proposition 2.58. *Let E be a set of integers of positive upper density. Then there exist an ergodic system $(X, \mathcal{X}, \mu, T)$ and a set $A \in \mathcal{X}$ with $\mu(A) = d^*(E)$ such that*

$$\mu(T^{m_1}A \cap \cdots \cap T^{m_k}A) \leq d^*((E + m_1) \cap \cdots \cap (E + m_k))$$

for all integers $k \geq 1$ and all $m_1, ..., m_k \in \mathbb{Z}$.

While it is easily seen that this proposition implies Szemerédi's Theorem for progressions of length k once a positive limit for the ergodic average (2.1) is established, when one attempts to transfer Leibman's result to a statement such as Corollary 2.20, one only obtains a lower bound on the number of solutions rather than an asymptotically exact statement.

2.5.6 Remarks

Leibman [Lei07] is able to determine the true complexity of all translation-invariant linear systems, not just those of Cauchy-Schwarz complexity 2. The main reason for this level of generality is that Host and Kra's structure theorem (Theorem 2.53) is available for all k, unlike the situation in arithmetic combinatorics where the decomposition theorem depends on the existence of a suitable U^k-inverse theorem, which has only been proved for $k \leq 3$. The fact that ergodic theorists are able to deal with polynomial systems is another point of envy. Indeed, it turns out that the seminorms $\|.\|_k$ also control polynomial averages when combined with *PET induction* (a linearization method which originated in [BL96]). In the finite combinatorial world, on the other hand, so-called "local" U^k-norms will be required in order to control polynomial averages. The reason for this is that when we consider polynomials such as $x + n^2$ inside an interval $1, 2, \ldots, N$, we are forced to restrict the range of the parameter n to $\sqrt{N}$. These local uniformity norms are currently much less well understood, but see [TZ06] for more details on the emerging theory of local uniformity.

Acknowledgements. The author would like to thank Bryna Kra for valuable comments and discussions.

Chapter 3

The Minimum Number of Monochromatic 4-Term Progressions in $\mathbb{Z}_p$

3.1 Introduction

In this short chapter we improve the lower bound given by Cameron, Cilleruelo and Serra [CCS05] for the minimum number of monochromatic 4-term progressions contained in any 2-colouring of $\mathbb{Z}_p$ with p a prime. We also exhibit a colouring with significantly fewer than the random number of monochromatic 4-term progressions, which is based on an a recent example in additive combinatorics by Gowers [Gow06b]. In the second half of this chapter we discuss the corresponding problem in graphs, which has received a great deal more attention to date. We give a simplified proof of the best known lower bound on the minimum number of monochromatic K_4s contained in any 2-colouring of K_n by Giraud [Gir79], and briefly discuss the analogy between the upper-bound graph constructions of Thomason [Tho89] and ours for subsets of $\mathbb{Z}_p$.

Let p be a prime. It is a pretty and well-known fact that in any 2-colouring of the cyclic group $\mathbb{Z}_p$ the number of monochromatic 3-term arithmetic progressions depends only on the densities of the colour classes R and B. Using discrete Fourier analysis, specifically the fact that $\widehat{1_R}(t) = -\widehat{1_B}(t)$ for $t \neq 0$, one easily obtains the result that the number of monochromatic 3-term progressions in any colouring equals

$$1/2(1 - 3\alpha + 3\alpha^2)p^2,$$

where one of the colour classes, R say, has size αp (see also [Dat03]). Note that this

is precisely the number of 3-term progressions we would expect if we were to choose the elements of the red colour class independently at random from $\mathbb{Z}_p$ with density α. Throughout this chapter, we shall be counting progressions without orientation, that is we shall be considering $3, 5, 7 \bmod 13$ as identical to $7, 5, 3 \bmod 13$. Our results will always be asymptotic in the order p of the group.

While a similar formula holds for other equations in three variables, for example Schur triples of the form $x + y \equiv z$, this is not the case for longer progressions. It is not difficult to see that the number of monochromatic 4-term progressions in a given 2-colouring does not just depend on the density ratio of the colour classes. Instead, we will ask for the minimum number of monochromatic 4-APs in any 2-colouring of $\mathbb{Z}_p$, a quantity which we shall denote by $M_4(p)$. Bounding the more convenient normalised quantity $m_4(p) := 2M_4(p)/p^2$ is the aim we shall be concerned with throughout the first two sections of this chapter.

An easy bound on $m_4(p)$ can be derived from Van der Waerden's Theorem. We know that the Van der Waerden number $W(4)$ equals 35, that is, any 2-colouring of 35 numbers in arithmetic progression is guaranteed to contain a monochromatic 4-AP. By averaging, we obtain a lower bound on $m_4(p)$ of the form

$$m_4(p) \geq \frac{1}{185} + o(1).$$

Here $o(1)$ denotes a quantity that tends to zero as p tends to infinity through the primes.

This primitive estimate was significantly improved by Cameron, Cilleruelo and Serra [CCS05] by observing that although the number 35 cannot be reduced when searching for monochromatic 4-APs, we only need to colour 7 points in arithmetic progression before we are guaranteed to find a monochromatic 4-AP or one which is evenly coloured, i.e. one in which precisely 2 points are red and 2 points are blue. This fact together with one additional ingredient, which we shall inspect in more detail in Lemma 3.3 below, gives their bound

$$m_4(p) \geq \frac{1}{20} + o(1).$$

A further computational improvement yields their best effort of

$$m_4(p) \geq \frac{2}{33} + o(1).$$

In this short chapter we prove the following small improvement.

Theorem 3.1. *Any 2-colouring of $\mathbb{Z}_p$ with p a prime contains at least $p^2/32$ monochromatic 4-term progressions. In other words,*

$$m_4(p) \geq \frac{1}{16} + o(1).$$

In the other direction, it is clear that a random colouring with probability $1/2$ will contain $p^2/16$ monochromatic 4-APs, so that $m_4(p) \leq 1/8 + o(1)$. In Section 3.3 we exhibit a colouring with fewer than this random number of monochromatic 4-APs, which shows that the critical constant must lie strictly below $1/8$. More precisely, we shall prove the following theorem.

Theorem 3.2. *There exists a colouring of $\mathbb{Z}_p$ with p a prime containing fewer than $1/16(1 - 1/2025)p^2$ monochromatic 4-term progressions. In other words,*

$$m_4(p) \leq \frac{1}{8}\left(1 - \frac{1}{2025}\right) + o(1).$$

A gap between the upper and lower bound remains. Perhaps we shouldn't be too surprised at this state of affairs in view of the fact that the corresponding problem in graphs, where one wants to determine the minimum number of monochromatic K_4s in any 2-colouring of the complete graph K_n, has resisted a complete resolution for quite some time and for similar reasons. In Section 3.4 we give a simplified version of Giraud's argument [Gir79] which yields the best known lower bound for this problem. In the final section we review some constructions by Thomason which yield 2-colourings of graphs with fewer than the random number of monochromatic K_4s and discuss the analogy between graphs and sets. Perhaps the gap is accounted for by the different methods used to prove the upper and lower bounds in both cases: while the lower bounds are obtained by simple (if somewhat ingenious) counting, the upper bound constructions rely on Fourier analytic techniques.

3.2 A Lower Bound on the Number of Monochromatic 4-APs

Given any 2-colouring C of $\mathbb{Z}_p$, let $m_4(C, p)$ denote the number of monochromatic 4-term progressions in C, divided by $p^2/2$. For $i = 0, 1, 2, 3, 4$, let $c_i := c_i(C, p)$ denote the number of 4-term progressions in $\mathbb{Z}_p$ which have precisely i red elements, divided by $p^2/2$. We immediately note that $\sum_{i=0}^{4} c_i = 1$ and $m_4(C, p) = c_0 + c_4$.

Write $E := c_0 + c_2 + c_4$ and $O := c_1 + c_3$ for the normalised number of even- and odd-coloured progressions, respectively.

A simple counting argument yields a further relation between the c_i in terms of the density α of the red colour class. The following lemma is borrowed from [CCS05], although for the sake of self-containedness we give our own, more direct proof here.

Lemma 3.3. *With the c_i defined as above, we have that*

$$4(c_0 + c_4) + (c_1 + c_3) = 4(1 - 3\alpha + 3\alpha^2)$$

for any colouring of $\mathbb{Z}_p$ in which the red colour class has size αp.

Proof. We will perform double-counting on the edges of a bipartite graph with vertex sets $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2 \cup Y_3$. Here X_1 consists of all 4-APs counted by $c_0 + c_4$, and X_2 of all those counted by $c_1 + c_3$. Y_1 denotes the set of all monochromatic 3-APs, while Y_2 and Y_3 denote the sets of all monochromatic configurations of the form $x, x + d, x + 3d$ and $x, x + 2d, x + 3d$ respectively. Elements $x \in X$ and $y \in Y$ are joined by an edge if and only if x contains the configuration y. It is now easy to see that the total (normalised) out-degree of X equals $4(c_0 + c_4) + (c_1 + c_3)$, while the total out-degree of Y equals twice the number of monochromatic 3-APs plus the number of other monochromatic configurations in Y. By the second paragraph of the introduction, the number of such monochromatic 3-term configurations equals $1/2(1 - 3\alpha + 3\alpha^2)p^2$. Therefore the normalised out-degree of Y equals $4(1 - 3\alpha + 3\alpha^2)$. □

The preceding lemma together with the identity $\sum_{i=0}^{4} c_i = 1$ now implies that

$$c_0 + c_4 = \frac{1}{3}c_2 + (1 - 4\alpha + 4\alpha^2).$$

Cameron, Cilleruelo and Serra immediately discard the second term on the right-hand side, which is indeed equal to zero for $\alpha = 1/2$ and hence doesn't appear to be of much use. However, it will be vital for us to keep the dependence on the density of the red colour class. From the above formula it is straightforward to see that any lower bound on the number of even-coloured 4-APs (which we denoted by E) results in a lower bound on $m_4(C, p)$ via the formula

$$m_4(C, p) = \frac{1}{2}E + \frac{3}{4}(1 - 4\alpha + 4\alpha^2).$$

For the remainder of this section we shall aim to bound E from below. In the process we shall focus on a method that works for densities close to $1/2$, since for densities bounded away from $1/2$ the second term in the lower bound for $m_4(C,p)$ already provides a fairly good estimate.

Given any 3-term progression S, let p_S denote the number of evenly coloured 4-APs which contain S. Let q_S be the number of 4-APs containing S which are not evenly coloured. It is obvious from these definitions that $0 \leq p_S, q_S \leq 2$ and $p_S + q_S = 2$. Another second's thought confirms that

$$\mathbb{E}_S\, p_S = 2E \text{ while } \mathbb{E}_S\, q_S = 2O,$$

where the expectation operator $\mathbb{E}_S$ denotes the sum $\sum_S$ divided by $p^2/2$. Any 3-term progression S of the form $x, x+d, x+2d$ determines a unique (unordered) pair of points (a,b) such that the five points and each of the quadruples $a, x, x+d, x+2d$ and $x, x+d, x+2d, b$ lie in arithmetic progression. We shall call the pair (a,b) a *frame pair*. It is straightforward to see that each frame pair belongs to a unique 3-term progression. Note that in these statements we have used the assumption that p is prime.

The crucial observation is that the two 4-APs containing S have different colour parities if and only if the frame pair of S is bichromatic. Therefore, $p_S - q_S$ is not equal to zero if and only if S has a monochromatic frame pair. For densities close to $1/2$, the total number of monochromatic pairs is at its minimum, which will enable us to get an acceptable estimate in this density regime. As remarked before, for densities bounded away from $1/2$ the second term in the lower bound for $m_4(C,p)$ will take over.

We find a trivial lower bound on E of the form

$$2E = 2O + \mathbb{E}_S(p_S - q_S) \geq 2(1-E) - \mathbb{E}_S|p_S - q_S|.$$

But $\mathbb{E}_S\ |p_S - q_S|$ precisely equals 2 times the appropriately normalised number of monochromatic pairs in the colouring. Now the number of monochromatic (unordered) pairs in a colouring of $\mathbb{Z}_p$ in which one colour class has density α is precisely $(\alpha^2 + (1-\alpha)^2)p^2/2$, which yields

$$E \geq 1/2(1 - (1 - 2\alpha + 2\alpha^2)) = \alpha(1-\alpha),$$

which in turns produces the bound

$$m_4 \geq 1/4(\alpha(1-\alpha) + 3(1 - 4\alpha + 4\alpha^2)).$$

The minimum of this function is easily seen to be $1/16$, attained at $\alpha = 1/2$, concluding the proof of Theorem 3.1. The next section shows that we have at least found the correct bound to within a factor of 2.

3.3 A Colouring with Few Monochromatic 4-APs

Very recently Gowers [Gow06b] gave an example of a uniform subset of $\mathbb{Z}_p$ which contains fewer than the number of 4-APs expected in a random subset of $\mathbb{Z}_p$ of the same density. By *uniform* we mean that the largest non-trivial Fourier coefficient of the indicator function of the set is $o(1)$ in modulus. It is easy to establish that uniform sets always contain the number of 3-term progressions expected in a random subset of $\mathbb{Z}_p$ of the same density. It had been known for quite some time that ordinary Fourier analysis was insufficient when it came to counting longer progressions. Indeed, it is not too difficult to construct uniform sets that contain significantly *more* than the expected number of progressions of length 4. It was Gowers's intention to show that it is possible to achieve a negative 4-AP count (compared with random) while retaining the uniformity of the set. (For progressions of length strictly greater than 4 this is significantly easier, see the remarks in [Gow06b]).

In this section we observe that this construction immediately gives rise to a 2-colouring of $\mathbb{Z}_p$ with strictly fewer than $p^2/16$ monochromatic 4-APs. For $A \subseteq \mathbb{Z}_p$, let $p_4(A, p)$ denote the number of 4-term progressions in A, divided by $p^2/2$, and let $m_4(A, p)$ denote the number of monochromatic 4-APs in the colouring C which is induced by A (that is, we take $R = A$ and $B = A^C$), normalised in the same way. For uniform sets, the quantities $m_4(A, p)$ and $p_4(A, p)$ are related as follows.

Lemma 3.4. *Given a uniform set $A \subseteq \mathbb{Z}_p$ of density α, we have the relation*

$$m_4(A, p) = \frac{1}{2}((1-\alpha)^4 - \alpha^4) + 2p_4(A, p) + o(1).$$

Proof. Writing 1_A for the characteristic function of the set A, we find that

$$\begin{aligned} 2m_4(A,p) &= \mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+2d)1_A(x+3d) \\ &\quad +\mathbb{E}_{x,d}1_{A^C}(x)1_{A^C}(x+d)1_{A^C}(x+2d)1_{A^C}(x+3d) \\ &= 1-4\alpha+6\alpha^2+4p_4(A,p) \\ &\quad -\mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+2d)-\mathbb{E}_{x,d}1_A(x+d)1_A(x+2d)1_A(x+3d) \\ &\quad -\mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+3d)-\mathbb{E}_{x,d}1_A(x)1_A(x+2d)1_A(x+3d). \end{aligned}$$

It is easily seen that all the 3-term configurations appearing in the above sum appear in the expected number, by writing, for example,

$$\mathbb{E}_{x,d}1_A(x)1_A(x+d)1_A(x+2d)=\sum_t \widehat{1_A}(t)^2\widehat{1_A}(-2t)=\alpha^3+o(1),$$

assuming that the subset A is sufficiently uniform, that is $\sup_{t\neq 0}|\widehat{1_A}(t)|=o(1)$. □

We shall briefly sketch Gowers's construction with a slight numerical improvement over the original version. It is included for the sake of completeness and for purposes of comparison with the graphs case later on. We shall conclude the section with a statement of the exact bound we obtain.

We start off by constructing a function g taking values ± 1 on the cube $\{1,2,3,4\}^3$ which satisfies

$$\sum_{x,d} g(x)g(x+d)g(x+2d)g(x+3d)=-72.$$

This is done on the basis of a geometric argument. One then proceeds to project the cube into the interval $[1,300]$ using a map ϕ. (For the reader familiar with this kind of argument, the map ϕ is a standard Freiman isomorphism.) Next we define a new function f, which takes values ± 1 and 0 on the interval $[1,300]$, by setting $f(x)=g(\phi^{-1}(x))$ if x lies in the image of the projection ϕ of the cube, and $f(x)=0$ otherwise. By construction, f also satisfies

$$\sum_{x,d} f(x)f(x+d)f(x+2d)f(x+3d)=-72,$$

that is, it has negative relative 4-AP count. Elegant and neat as this example is, it is also rather inefficient. By an exhaustive numeric search for small examples we found the interval $[1,18]$, on which we let f take successive values

$$-1,-1,-1,1,-1,-1,1,-1,-1,-1,-1,1,1,1,-1,1,-1,-1$$

and for which

$$\sum_{x,d} f(x)f(x+d)f(x+2d)f(x+3d) = -36.$$

Unfortunately at this point we cannot rule out the existence of even more efficient small examples on intervals of length greater than 25.

The next stage is to blow up this small example to one that lives inside $\mathbb{Z}_p$ for large p. To this end, we define a function $F : \mathbb{Z}_p \to \{-1, 0, 1\}$ by setting $F(x) = f(t)$ whenever $x \in I_t$, where I_t stands for the interval $[(2t-1)m, 2tm]$ and m is a positive integer between $p/(5 \times 18)$ and $p/(4 \times 18)$. It is easy to check that F is well-defined, and that the 4-AP counts of F and f are related via

$$\sum_{x,d} F(x)F(x+d)F(x+2d)F(x+3d) = s \sum_{x,d} f(x)f(x+d)f(x+2d)f(x+3d)$$

where $s \geq m^2/9$. It remains to ensure that F is uniform, and to convert the ± 1 function into a subset of $\mathbb{Z}_p$. The former is achieved by multiplying F by an appropriate sum of quadratic exponentials, giving rise to a function G defined by

$$G(x) := F(x)(\omega^{x^2} + \omega^{3x^2} + \omega^{-3x^2} + \omega^{-x^2}),$$

where ω is a pth root of unity (note that the negative exponents are needed to make the resulting function G real). Since F essentially behaves like the indicator function of a union of intervals, its Fourier transform has bounded l^1-norm, and because of the large amount of cancellation coming from the quadratic phases we can conclude that all non-trivial Fourier coefficients of G are tiny. Finally, turning the function G into a subset of $\mathbb{Z}_p$ is a completely standard procedure in which, roughly speaking, we choose an element x to lie in $A \subseteq \mathbb{Z}_p$ with probability $(1 + G(x))/2$. With high probability the resulting set A has density $1/2$ and is uniform by construction but contains at most

$$\frac{1}{16}\left(1 - \frac{36}{9(5 \times 18)^2}\right)p^2$$

4-term progressions. In conjunction with Lemma 3.4, this discussion concludes our proof of Theorem 3.2.

Incidentally, it is also interesting to combine (via Lemma 3.4) this approach with the lower bound on $m_4(p)$ we obtained in Section 3.2: It tells us that any uniform subset of density $1/2$ must contain at least $p^2/64$ 4-term progressions. This is related to a question Gowers asks in [Gow06b] and which can be traced back to I. Ruzsa: If $A \subseteq \mathbb{Z}_p$ is uniform of density α, must A contain at least α^c progressions of length 4 for some large constant c? Of course, for densities away from $1/2$ the considerations

described here yield no results.

Although transferring results from the density to the colouring world and vice versa has proved fruitful in this instance, we doubt that the correct constant for the problem of counting the minimal number of monochromatic 4-term progressions can be obtained in this way.

3.4 Giraud's Lower Bound for the Number of Monochromatic K_4s

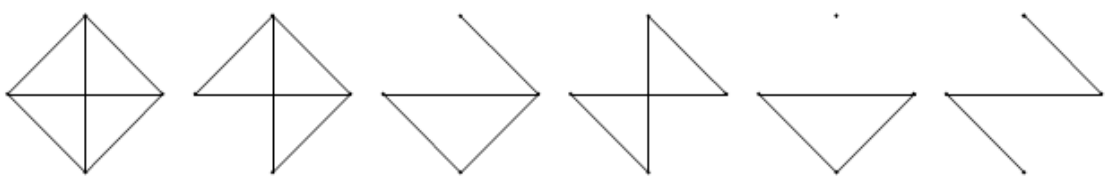
In this section we give a simplified proof of Giraud's lower bound [Gir79] on the minimum number of monochromatic K_4s which we are guaranteed to find in any 2-colouring of the edges of the complete graph K_n on n vertices. We call this quantity $M_{K_4}(n)$, and its normalised sibling $m_{K_4}(n) := M_{K_4}(n)/\binom{n}{4}$. Since we are only concerned with asymptotics, we shall for the remainder take a rather relaxed approach to equalities: $x \asymp 1$ will always mean $x = 1 + o(1)$. We shall again be using the expectation operator $\mathbb{E}$, denoting the sum over edges or triangles normalised by $\binom{n}{2}$ or $\binom{n}{3}$, respectively.

As in the case of 4-APs, a simple lower bound can be given by averaging using Ramsey's Theorem. Giraud proved the much superior lower bound

$$m_{K_4}(n) > \frac{1}{46} + o(1),$$

and we shall give a concise exposition of his work here, including a simplification of his argument. The original presentation in [Gir79] is rather convoluted, and, *en plus*, in French.

Throughout, we shall fix a colouring of K_n and colour-blindly count the following configurations on four vertices: the complete graph on 4 vertices denoted by K, the double triangle (or K with one edge missing) DT, the triangle with a pendant edge TE, the 4-cycle C, the ordinary triangle T and the path of length 3 denoted by P. We shall abuse notation and use the acronyms to denote the number of occurrences of these structures divided by $\binom{n}{4}$. We call the collection of these substructures $\mathcal{Q}$.



Following Giraud, we define for a given edge e the quantities $m_e :=$ # monochromatic triangles containing e, divided by $n-2$, $b_e :=$ normalised # bichromatic triangles containing e in which e is the only edge of its colour, and $c_e :=$ normalised # bichromatic triangles containing e in which e is not the only edge of its colour.

We immediately see that $m_e + b_e + c_e = 1$. Considering the number of mono- and bichromatic triangles in members of $\mathcal{Q}$, we find the following system of equations:

$$\begin{array}{rrcl} 12 & \mathbb{E}_e \binom{m_e}{2} & \asymp & 6K + DT \\ 12 & \mathbb{E}_e\, m_e b_e & \asymp & TE + 3T \\ 12 & \mathbb{E}_e\, m_e c_e & \asymp & 4DT + 2TE \\ 12 & \mathbb{E}_e\, b_e c_e & \asymp & 2TE + 4P \\ 12 & \mathbb{E}_e \binom{b_e}{2} & \asymp & DT + 2C \end{array}$$

Thus, taking suitable linear combinations of these equations, we obtain

$$\begin{array}{rcl} K + DT + TE + C + T + P & \asymp & 1 \\ K + TE + C - (DT + T + P) & \asymp & 32K - 3\mathbb{E}_e b_e + 11\mathbb{E}_e m_e - 48\mathbb{E}_e m_e^2 + 6\mathbb{E}_e (m_e - b_e)^2. \end{array}$$

Still following Giraud, we define for given a triangle t the parameters $p_t :=$ # even-edged configurations in $\mathcal{Q}$ containing t, divided by n, and $q_t :=$ normalised # odd-edged configurations in $\mathcal{Q}$ containing t. Double-counting again, we find that

$$\mathbb{E}_t\, p_t = K + TE + C \quad \text{and} \quad \mathbb{E}_t\, q_t = DT + T + P.$$

Combining this with the two previous equations, we obtain

$$32K = 1 - 12\mathbb{E}_e\, m_e + 48\mathbb{E}_e\, m_e^2 - 6\mathbb{E}_e (m_e - b_e)^2 + \mathbb{E}_t (p_t - q_t) \tag{3.1}$$

We observe the similarities with our work in Section 3.2, and note the increased level of difficulty here due to the increased complexity of the substructures.

The remainder of the proof consists in bounding the final two terms above in modulus by a suitable application of the Cauchy-Schwarz Inequality, and then performing an optimization over what is essentially the mean and variance of the variables m_e. Before carrying out this plan we shift our variables by setting

$$\mu_e := 4m_e - 1 \quad \text{and} \quad \delta_e := 2(m_e - b_e).$$

With these definitions, the parameters with respect to which we optimize later are

$$s := \mathbb{E}_e\,(\mu_e - \delta_e) = \frac{1}{3}\mathbb{E}_e\,\mu_e, \quad r := \mathbb{E}_e\,(\mu_e - \delta_e)^2, \quad p := \mathbb{E}_e\,\mu_e^2.$$

It is now possible to rewrite equation (3.1) as

$$32K = 1 + 3\mathbb{E}_e\,\mu_e + 3\mathbb{E}_e\,\mu_e^2 - \frac{3}{2}\mathbb{E}_e\,\delta_e^2 + \mathbb{E}_t(p_t - q_t). \tag{3.2}$$

Our first task is to find an upper bound for $\mathbb{E}_e\,\delta_e^2$: Let B denote the number of bichromatic triangles in the graph K_n, and set $b := B/\binom{n}{3}$. Using the fact that every non-monochromatic triangle is bichromatic, it is straightforward to compute that $b = 3/4(1-s)$. We need some additional notation: For any vertex i, let r_i denote the number of red edges incident with i, divided by $n-2$, and b_i the normalised number of blue edges. For an edge $e = ij$, we let c'_e denote the number of bichromatic triangles including e where the two edges of the same colour meet in vertex i. Let $c''_e = c_e - c'_e$. If the edge $e = ij$ is coloured red, it is clear that $r_i \asymp m_e + c'_e$ and $r_j \asymp m_e + c''_e$. Now the total proportion b of bichromatic triangles can also be expressed as

$$3/2\mathbb{E}_{e=ij} r_i b_i + r_j b_j \asymp 3/2\mathbb{E}_e(m_e + c'_e)(1 - (m_e + c'_e)) + (m_e + c''_e)(1 - (m_e + c''_e)),$$

the right-hand side of which can be bounded above by $3/2\mathbb{E}_e(2m_e + c_e) - 1/2(2m_e + c_e)^2$. This in turn can be expressed in terms of δ_e to give the bound $\mathbb{E}_e\,\delta_e^2 \leq 4s$.

We now turn to bounding $\mathbb{E}_t(p_t - q_t)^2$. We note that, given a triangle t and an edge $e = ij$, the structures induced by $t \cup i$ and $t \cup j$ have different colour parities if and only if an odd number of edge pairs $(iv, jv)_{v \in t}$ differ in colour. It follows that

$$\mathbb{E}_t\,p_t q_t \asymp 3\mathbb{E}_e\left(\binom{c_e}{3} + c_e\binom{m_e + b_e}{2}\right),$$

which can be rearranged to give

$$\mathbb{E}_t(p_t - q_t)^2 \asymp \mathbb{E}_e(1 - 2c_e)^3.$$

The latter expectation can be bounded via the following simple lemma for cubes, several rather laboured versions of which appear in [Gir79]:

Lemma 3.5. *Suppose we have N variables $x_i \in [-1, 1]$ with $\mathbb{E}_i x_i = s$ and $\mathbb{E}_i x_i^2 = r$. Then we have the bound*

$$\mathbb{E}_i x_i^3 \leq \frac{r(1+s) - (r^2 + s^2)}{1 - s} =: g(r, s).$$

Proof. Let $y_i = 1 - x_i \in [0, 2]$, and observe see that $\mathbb{E}_i y_i = (1 - s), \mathbb{E}_i y_i^2 = (1 + r - 2s)$ as well as $\mathbb{E}_i y_i^3 = (1 + 3r - 3s) - \mathbb{E}_i x_i^3$. Bounding $\mathbb{E}_i y_i^3$ below by Cauchy-Schwarz

$$\mathbb{E}_i y_i^3 \geq \frac{(\mathbb{E}_i y_i^2)^2}{\mathbb{E}_i y_i} = \frac{(1 + r - 2s)^2}{1 - s}$$

gives the desired result after rearranging. □

It follows straightforwardly from Lemma 3.5 with x_i replaced by $\mu_e - \delta_e$ and the preceding discussion that

$$\mathbb{E}_t (p_t - q_t)^2 \leq g(r, s).$$

Inserting this bound via the Cauchy-Schwarz Inequality into (3.2), we find that

$$32K \geq 1 + 3s + 3p - \sqrt{g(r, s)}. \tag{3.3}$$

For the purpose of optimizing this expression, we observe that for $r \leq (1 + s)/2$, the function $g(r, s)$ is decreasing in r. In general, we can bound r (using nothing but the Cauchy-Schwarz Inequality and the definitions) by

$$r \leq (2\sqrt{s} + \sqrt{p})^2, \tag{3.4}$$

so we distinguish the cases $(2\sqrt{s} + \sqrt{p})^2 \geq (1 + s)/2$ and $(2\sqrt{s} + \sqrt{p})^2 \leq (1 + s)/2$. In the first case, we can set $r = (1 + s)/2$, which implies that (3.4) gives a lower bound for p in terms of s, and we are left to find the minimum of the right-hand side of (3.3) as a function of the single variable s. In the second case, we set $r = (2\sqrt{s} + \sqrt{p})^2$, and then minimize the right-hand side of (3.3) as a function of r and s.

This is a question of seconds using a computer, and the minimum value thus obtained turns out to be $0.0217514\ldots$, which lies between $1/46$ and $1/45$.

3.5 Thomason's Upper Bound for the Number of Monochromatic K_4s

In 1989 Thomason [Tho89] disproved a conjecture by Erdös which claimed that there are always at least the random number of monochromatic K_4s in every 2-colouring of K_n. Even though there exists a wealth of counterexamples by now, this conjecture didn't seem quite so unreasonable back then. Indeed, the result is true if one replaces K_4s by triangles (see [Goo59], although he makes it seem like rather hard work) or by ordinary 4-cycles. The initial construction Thomason gave to disprove Erdös'

conjecture was rather obscurely phrased in terms of quadratic forms over a finite field, but equivalent and much clearer formulations have since appeared in [JŠT96] and [Tho97].

It is interesting to note that the graphs constructed in these follow-up papers are quite similar in structure to the set constructed by Gowers which we described in Section 3.3. One takes a small example exhibiting a strong bias and then uses a product construction to produce a biased example of size growing asymptotically in n. In the case of Gowers's example we producted by long intervals, whereas Thomason uses a *tensor product* of graphs.

Definition 3.6. *Given two graphs J_1 and J_2, let their* tensor product $J_1 \otimes J_2$ *be the graph with vertex set $V(J_1 \otimes J_2) = V(J_1) \times V(J_2)$. The edges of $J_1 \otimes J_2$ are determined by $(v, w)(v', w') \in E(J_1 \otimes J_2)$ if either $vv' \in E(J_1)$ or $ww' \in E(J_2)$ but not both.*

(Many authors refer to this product as the *Cartesian product* of graphs. According to their definition, the *tensor product* requires both coordinates to be edges in the factor graphs, but we shall stick with Thomason's notation in order to minimise confusion.)

Note also that the tensor product is commutative and associative, and observe that if J_2 is the empty graph $\overline{K_m}$, then $J_1 \otimes \overline{K_m}$ is the usual *m-fold cover* of J_1. We now rephrase the tensor product in terms of the balanced adjacency matrices of the graphs involved. We associate with J the matrix $A(J) = (a(u, v))_{u,v \in V(J)}$ whose entries are indexed by the vertices of J and are defined by $a(u, v) = -1$ if $uv \in E(J)$ and $a(u, v) = 1$ otherwise. It is important to note that the diagonal entries of $A(J)$ are all equal to 1.

Definition 3.7. *Given two square matrices $A = (a_{ij})_{i,j=1}^{n}$ and $B = (b_{ij})_{k,l=1}^{m}$, their* tensor product $A \otimes B$ *is defined to be the $nm \times nm$ square matrix with entries $(A \otimes B)_{(i,k)(j,l)} = a_{ij} b_{kl}$.*

It is straightforward to see that the matrix $A(J_1 \otimes J_2)$ associated with the graph tensor product $J_1 \otimes J_2$ is just the matrix tensor product $A(J_1) \otimes A(J_2)$.

We are now in a position to count the number of monochromatic K_4s occurring in a given colouring of the complete graph K_n, or equivalently, the number of K_4s occurring in a given graph $J \subseteq K_n$ and its complement, which we shall denote by $m_{K_4}(J)$. It is easy to check that $m_{K_4}(J)$ equals

$$\sum_{u_1,\ldots,u_4 \in V(J)} \prod_{ij \in E(K_4)} (1 - a(u_i, u_j)) + \sum_{u_1,\ldots,u_4 \in V(J)} \prod_{ij \in E(K_4)} (1 + a(u_i, u_j)).$$

Writing

$$\Psi(J,F) = |J|^{-4} \sum_{u_1,\dots,u_4 \in V(J)} \prod_{ij \in E(F)} a(u_i,u_j)$$

for each spanning subgraph F of K_4, we can rewrite $m_{K_4}(J)$ as

$$2^{-5}|J|^4(1+O(|J|^{-1})) \sum_{F \subseteq K_4} \Psi(J,F), \tag{3.5}$$

where the sum is over all spanning subgraphs of K_4 with an even number of edges. It is clear from (3.5) that any graph J with $\sum_{F \subseteq K_4} \Psi(J,F) < 1$ will have fewer than the expected number of monochromatic K_4s. Since $\Psi(\overline{K_m},F) = 1$ for all $F \subseteq K_4$, we see that in order to find a sequence of graphs with too few monochromatic K_4s of order tending to infinity, it suffices to find a small graph J with $\sum_{F \subseteq K_4} \Psi(J,F) < 1$. One can then set $J_m = J \otimes \overline{K_m}$ to obtain the desired family.

The function $\Psi(J,F)$ has the very useful property that it is multiplicative with respect to the tensor product of graphs defined above, in the sense that

$$\Psi(J_1 \otimes J_2, F) = \Psi(J_1,F)\Psi(J_2,F).$$

This will enable us to compute the number of monochromatic K_4s inside graph products with small factors very easily, since $\Psi(J,F)$ can be evaluated explicitly with little computational effort for small graphs J. (At this point we would like to draw the reader's attention to how Ψ relates to the Fourier transform on $\mathbb{F}_2^n$.)

According to the computer investigations conducted in [Tho97], the example which exhibits the largest relative bias amongst all tensor products of small graphs is the graph product $K_4 \otimes M \otimes \overline{(K_3 \otimes K_3 \otimes \overline{K_2})}$, where M stands for the graph on 4 vertices with two non-adjacent edges. More precisely, computations result in the bound

$$m_{K_4}(K_4 \otimes M \otimes \overline{(K_3 \otimes K_3 \otimes \overline{K_2})} \otimes \overline{K_m}) < \frac{1}{33} + o(1),$$

where $o(1)$ stands for a quantity which tends to 0 as m tends to infinity. It is observed in the final paragraph of [Tho97] that it is possible to improve this construction by an absolutely tiny amount using a random perturbation.

3.6 Remarks

Section 3.5 completed the fourth corner of the square defined by the axes *graphs - sets* and *upper bound - lower bound* which we have discussed in this chapter. It would be

of great interest to close the gap between the upper and lower bound in both cases.

While we have pointed out some tentative analogies between the world of graphs and sets, their exact nature remains somewhat elusive. In particular, Gowers's set is uniform yet contains the wrong number of 4-APs. In the world of graphs, a uniform (that is, *quasirandom*) graph contains the correct number of K_4s and will therefore be of no use in constructing a bad example. In view of this breakdown of analogies, it seems likely that in order to fully understand Thomason's constructions, one needs to instead consider notions of uniformity which have been developed in the context of hypergraphs.

Acknowledgements. The author would like to thank Tim Gowers for making the preprint [Gow06b] available, and Andrew Thomason for helpful comments and discussions.

Chapter 4

The Structure of Popular Difference Sets

4.1 Introduction

Let G be a finite Abelian group of order N. Suppose that A is a subset of G of cardinality linear in N, and define the set of γ-*popular* differences of A to be

$$D_\gamma(A) := \{x \in G : A * -A(x) \geq \gamma\},$$

where we have written A for the indicator function of the subset A. In other words, $D_M(A)$ is the set of elements of G which can be written as a difference of elements of A in at least γN different ways. Because we are considering subsets of G of size linear in N, we shall take γ to be a small constant throughout this chapter. Is it true that $D_\gamma(A)$ always contains the complete difference set $A_0 - A_0$ for some large set A_0? Our aim in this chapter is to show that this is not always so. More precisely, when $G = \mathbb{F}_2^n$ and $G = \mathbb{Z}_N$ with N a prime, we prove that there exists a set A of linear size such that any set A_0 whose difference set is contained in $D_\gamma(A)$ has density $o(1)$. Here $o(1)$ denotes a quantity tending to 0 as the order N of the group G tends to infinity.

Theorem 4.1. *Let $G = \mathbb{F}_2^n$ or $G = \mathbb{Z}_N$. Then there exists a set $A \subseteq G$ of size greater than $N/3$ with the property that any set A_0 whose difference set is contained in the set $D_\gamma(A)$ of γ-popular differences of A has density $o(1)$.*

Apart from being an interesting question in its own right, this problem has arisen in the context of counting the number of sum-free subsets of an Abelian group G,

notably in the work of Lev, Łuczak and Schoen [LŁS01] and Green and Ruzsa [GR05]. The first team of authors pursued the following strategy: Suppose every sum-free set A contained a small subset E with large difference set. The small cardinality of E implies that there are relatively few such sets, and from the fact that the difference set is large it follows that there are only few sets A corresponding to a given E, since for a sum-free set A we have $A \subseteq G\backslash(A-A) \subseteq G\backslash(E-E)$. By taking a random subset of A with suitable probability, one can obtain a small set E which has the property that its difference set contains the set $D_\gamma(A)$ of popular differences of A. Therefore the argument we just sketched implies an upper bound on the number of sum-free sets A whenever $D_\gamma(A)$ is large. For those A with few popular differences, the following proposition from [LŁS01] can be used in conjunction with Kneser's Theorem to obtain an upper bound in the remaining case. Its proof consists of a simple averaging argument on the Cayley graph on $\mathbb{Z}_N$ generated by $D_\gamma(A)$.

Proposition 4.2. *Let X be a subset of G, and let γ be a positive constant. Suppose that the set of γ-popular differences $D_\gamma(X)$ satisfies*

$$|D_\gamma(X)| \leq 2|X| - 5\sqrt{\gamma N|X-X|}.$$

Then there exists a subset $X' \subseteq X$ such that

$$|X\backslash X'| \leq \sqrt{\gamma N|X-X|} \quad \textit{and} \quad X'-X' \subseteq D_\gamma(X).$$

Green and Ruzsa [GR05] used this proposition to show that it suffices to remove ϵN elements from a set of size greater than $(1/3+\epsilon)N$ with few (more precisely, up to $\epsilon^3 N^2/27$) Schur triples in order to make it sum-free, which allows them to strengthen the result of Lev, Łuczak and Schoen on the number of sum-free subsets of G.

The result we present in this chapter shows that the condition on the size of the set of popular differences in Proposition 4.2 cannot be removed, which by the preceding discussion rules out simpler approaches to counting sum-free sets of Abelian groups. Before dealing with the case of the group $G = \mathbb{Z}_N$ with N a prime in Section 4.3, we first describe a combinatorial approach in the model setting of $G = \mathbb{F}_2^n$.

4.2 Vector Spaces over Finite Fields

The case where G is a finite-dimensional vector space over the field of two elements is often a good model for what happens in the cyclic groups $\mathbb{Z}_N$, and generally easier to deal with as we have additional geometric structure available. We refer the reader

to the excellent survey [Gre05a] for a plentiful supply of examples confirming this assertion.

For $x \in \mathbb{F}_2^n$, let $|x|$ denote the number of non-zero coordinates of the vector x. In this section we shall show that in the model setting $\mathbb{F}_2^n$, the set $A \subseteq \mathbb{F}_2^n$ defined by

$$A := \left\{ x \in \mathbb{F}_2^n : |x| \geq \frac{n}{2} + \frac{\sqrt{3n}}{2} \right\}$$

is an example of a set whose popular difference set does not contain the complete difference set of any other large set.

The set A described above can be viewed as the finite field analogue of a so-called *niveau set*, which was originally introduced by Ruzsa in [Ruz87] and later used in [Ruz91] to show that there exists a subset of $\mathbb{Z}_N$ whose sumset does not contain any long arithmetic progressions. It is a versatile construction that has received a fair amount of attention since. For example, a modified version of such a set can be used to show that Chang's Theorem on the structure of the large Fourier spectrum of a function is tight [Gre03]. We shall discuss the original construction in more detail in Section 4.3.

First we need to show that the set A thus constructed has the required size, that is, that it contains a positive proportion of all elements of $\mathbb{F}_2^n$. The proof of this well-known fact uses only very standard probabilistic estimates, but we include it for the sake of completeness. For the remainder of this section, we write $N := 2^n$ for the size of the group.

Lemma 4.3. *The set $A \subseteq \mathbb{F}_2^n$ as defined above has size at least $(1 - \exp(-1/2))N$.*

Proof. By definition, the size of A can be written as

$$|A| = \sum_{j=0}^{\frac{n}{2}+\frac{\sqrt{3n}}{2}} \binom{n}{j},$$

which equals the probability that a random variable X with binomial distribution $B(n, 1/2)$ takes values at most $\sqrt{3n}/2$ above its mean. We use a standard Chernov-type tail estimate, details of which can be found in [JŁR00] or Appendix A of [AS00].

Lemma 4.4. *Suppose X is a random variable with binomial distribution. Then for any $0 \leq \epsilon \leq 1$, we have the estimates*

$$\mathbb{P}(X \leq (1 - \epsilon)\mathbb{E}X) \leq \exp(-\epsilon^2 \mathbb{E}X/2)$$

and

$$\mathbb{P}(X \geq (1+\epsilon)\mathbb{E}X) \leq \exp(-\epsilon^2 \mathbb{E}X/3).$$

It follows immediately from the second inequality that the density of A is at least $1 - \exp(-1/2)$, which means that A contains more than a third of all elements of $\mathbb{F}_2^n$. □

Next we show that the set of popular differences $D_\gamma(A)$ is contained in a very structured subset of the discrete cube $\mathbb{F}_2^n$. More precisely, $D_\gamma(A)$ is contained in the complement of a *Hamming ball* centred at 1, which is defined to be

$$B_t(1) := \{x \in \mathbb{F}_2^n : |x| \geq n - t\}.$$

Note that our finite field niveau set A is in fact itself a Hamming ball.

Lemma 4.5. *Let the set $A \subseteq \mathbb{F}_2^n$ and the Hamming ball $B_t(1)$ be defined as above. Then for any real $t \leq 3n/4 \log(\gamma^{-1})$, we have*

$$D_\gamma(A) \subseteq B_t(1)^C.$$

Proof. We shall show that if $z \in \mathbb{F}_2^n$ is such that $|z| = n - t$, then the number of ways of writing z as a difference (or, equivalently, as a sum since we are performing addition modulo 2) of two elements of A is bounded above by $N \exp(-3n/4t)$. So suppose that z is the sum of two vectors x and y which both lie in A. Without loss of generality, we can assume that the first t coordinates of z are 0s, and the remaining $n - t$ coordinates are 1s. Writing

$$(z_1, z_2, ...z_t, z_{t+1}, ..., z_n) \equiv (x_1, x_2, ...x_t, x_{t+1}, ..., x_n) + (y_1, y_2, ...y_t, y_{t+1}, ..., y_n),$$

we observe (again without loss of generality) that the number of 1s amongst the coordinates $x_{t+1}, ..., x_n$ is bounded above by $(n-t)/2$. But we require that x be an element of A, so that the number of 1s amongst $x_1, ..., x_t$ is at least $n/2 + \sqrt{3n}/2 - (n-t)/2 = t/2 + \sqrt{3n}/2$. Hence the number of possible vectors x, which for fixed z in turn immediately determine y, is bounded above by

$$2 \sum_{i=\frac{t}{2}+\frac{\sqrt{3n}}{2}}^{t} \binom{t}{i} \sum_{j=0}^{\frac{1}{2}(n-t)} \binom{n-t}{j}.$$

The first sum can be bounded above by $2^t \exp(-(\sqrt{3n}/t)^2 t/4) = 2^t \exp(-3n/4t)$ by the first inequality of Lemma 4.4, and the second sum clearly equals 2^{n-t-1} by the

binomial theorem. The result follows. □

Finally, we need to exploit the geometric information we have just gathered. It is not unreasonable to expect to be able to bound the size of *any* set whose difference set is contained in the complement of a large Hamming ball. For this purpose we shall use a simple instance of *measure concentration* on the discrete cube. More background on the concentration of measure phenomenon in general compact metric groups will be presented in Section 3.3.

Lemma 4.6. *Let A_0 be any subset of $\mathbb{F}_2^n$ with the property that $A_0 - A_0 \subseteq B_t(1)^C$. Then the density of A_0 is bounded above by* $\exp(-t^2/4n)$.

Proof. For ease of notation let us also define the Hamming ball centred at 0 in the obvious way by setting

$$B_t(0) := \{x \in \mathbb{F}_2^n : |x| < t\}.$$

This is just the usual ball associated with the so-called *Hamming metric* on $\mathbb{F}_2^n$ defined by setting $d(x, y) = |x - y|$. In other words, the distance between x and y equals the number of coordinates in which they differ. It is easy to see that

$$A_0 - A_0 \subseteq B_t(1)^C \quad \Rightarrow \quad A_0 + B_t(1) \cap A_0 = \emptyset,$$

which in turn implies that

$$\overline{A_0} + B_t(0) \cap A_0 = \emptyset,$$

where we have used the bar to denote the set $(1, 1, \ldots, 1) + A_0$ of *antipodal vectors* of A_0. But the set $\overline{A_0} + B_t(0)$ is just the set of elements of $\mathbb{F}_2^n$ at Hamming distance less than t from some element in A_0. It is this observation which inspires us to use the following classical measure concentration result in the discrete cube, which can be found on page 172 of [McD89] or page 31 of [Led01].

Theorem 4.7. *Let μ denote the uniform measure on $\mathbb{F}_2^n$. Given any subset C of $\mathbb{F}_2^n$, we have the inequality*

$$\mu(C + B_t(0)) \geq 1 - \frac{\exp(-t^2/2n)}{\mu(C)}.$$

We remark that it was already shown by Harper [Har66] that this inequality is sharp if the set C in Theorem 4.7 is a Hamming ball. Applying Theorem 4.7 to the set $\overline{A_0}$,

we immediately deduce that

$$\mu(\overline{A_0} + B_t(0)) \geq 1 - \frac{\exp(-t^2/2n)}{\mu(A_0)},$$

but the fact that $\overline{A_0} + B_t(0) \cap A_0 = \emptyset$ implies that

$$1 - \frac{\exp(-t^2/2n)}{\mu(A_0)} + \mu(A_0) \leq 1,$$

which after rearranging concludes the proof. □

Combining Lemma 4.5 and Lemma 4.6, we have proved the main result of this section. It asserts that $D_\gamma(A)$ only contains the complete difference set of sets of density $o(1)$.

Theorem 4.8. *There exists a set $A \subseteq \mathbb{F}_2^n$ of size greater than $N/3$ with the property that the set $D_\gamma(A)$ of γ-popular differences does not contain the complete difference set of any set of density greater than*

$$\exp(-9n/64\log^2(\gamma^{-1})).$$

4.3 From the Model Case to $\mathbb{Z}_N$

We now focus our attention on the finite Abelian group $\mathbb{Z}_N$ with N a large prime, whose characters are of the form $x \mapsto e(rx/N) := \exp(2\pi irx/N)$. In this more general context, we define a *niveau set* $A \subseteq \mathbb{Z}_N$ as the set

$$A := \left\{ x \in \mathbb{Z}_N : \Re \sum_{i=1}^{k} \gamma_i(x) \geq \epsilon\sqrt{k} \right\},$$

for some judiciously chosen set of characters $\gamma_1, \gamma_2, ..., \gamma_k$. The precise value of the parameters ϵ and k will be determined in the course of the argument, but ϵ should always be thought of as a fixed constant and k as growing roughly like $\log N$ to some small power.

As already mentioned in Section 4.2, this construction was originally introduced by Ruzsa in [Ruz87] and later used in [Ruz91] to give an example of a subset of $\mathbb{Z}_N$ whose sumset does not contain any long arithmetic progressions. We shall follow his analysis of the properties of such a set very closely in Section 4.3.1, where we show that A contains a positive proportion of all elements of $\mathbb{Z}_N$. In order to be

able to give an estimate for the size of A, we need the characters to behave roughly "independently" in the following sense:

Definition 4.9. *We say that a set of characters $(\gamma_i(x) = e(r_i x/N))_{i=1}^k$ is* K-independent *if $\sum_{i=1}^k \lambda_i r_i \equiv 0 \mod N$ has no solutions satisfying $\sum_i |\lambda_i| \leq K$. We shall also sometimes refer to the corresponding k-tuple $(r_i)_{i=1}^k \subseteq \mathbb{Z}_N^k$ as K-independent.*

We first of all need to make sure that such a set of characters actually exists, otherwise Definition 4.9 would be rather pointless.

Lemma 4.10. *The number of k-tuples in $\mathbb{Z}_N$ which are* not *K-independent is bounded above by*

$$(2K+1)^k N^{k-1}.$$

In other words, there exists a set of k characters with the K-independence property provided that K satisfies the inequality $K < N^{1/k}/4$.

Proof. A very crude but effective counting argument will do the job: Every k-tuple which is not K-independent satisfies by definition an equation in k variables with coefficients between $-K$ and K. There are at most $(2K+1)^k$ such equations. □

From now on we assume that we are dealing with a set of K-independent characters whenever we make reference to the niveau set A. Having set up the basics, we now turn to proving the analogues of Lemmas 4.3, 4.5 and 4.6 in Sections 4.3.1, 4.3.2 and 4.3.3, respectively.

4.3.1 Estimating the Size of the Niveau Set

The following lower bound on the cardinality of the niveau set A is proved in [Ruz91]. It is the analogue of Lemma 4.3 in the case $G = \mathbb{Z}_N$.

Proposition 4.11. *Let $\epsilon \leq 1/4$ and suppose $k \ll \log N / \log\log N$. Then the set A with parameters ϵ and k as defined above has cardinality at least $N/3$.*

For the sake of clarity, self-containedness and because we want to use a very similar argument later on, we give a concise exposition of Ruzsa's proof in this section. We shall proceed in two steps. First, we compare the character sum appearing in the definition of A to a sum of independent random variables distributed uniformly on the unit circle. Second, we approximate this sum of independent random variables by a normal distribution, which allows us to perform explicit computations.

A crucial tool in proving the first step is the following theorem in probability theory, which is known as *Esseen's Inequality*. It dates back to Esseen [Ess45] and independently Berry [Ber41], but see Shiryayev [Shi84] for a general introductory reference.

Theorem 4.12. *Let F_1, F_2 be probability distribution functions with corresponding characteristic functions ϕ_1, ϕ_2. Assume F_1' exists and is pointwise bounded by a constant V. Then*

$$\sup_x |F_1(x) - F_2(x)| \ll \frac{V}{T} + \int_0^T \frac{|\phi_1(t) - \phi_2(t)|}{t} dt.$$

We briefly recall that the *characteristic function* ϕ_X of a random variable X is defined to be $\phi_X(t) := \mathbb{E}\exp(itX)$, and that therefore the probability density function of a random variable is the inverse Fourier transform of its characteristic function. From now on we shall be using the notation $a \ll b$ to indicate that there exists an absolute constant c such that $a \leq cb$.

A special case of Theorem 4.12, also known as the *Berry-Esseen Inequality*, will help us complete the second step. It measures the total variation distance between a sum of independent identically distributed random variables and the normal distribution, in other words, it gives us information about the rate of convergence in the Central Limit Theorem. More precisely, let $X_1, X_2, \ldots, X_k$ be independent random variables, each distributed uniformly on the unit circle, and define their sum to be

$$X := \sum_{j=1}^{k} X_j \text{ with real part } \widetilde{X} := \Re X.$$

Let $\sigma := \sqrt{k/2}$ denote the standard deviation of $\widetilde{X}$. The following formulation of the Berry-Esseen Inequality is taken from page 374 of [Shi84].

Theorem 4.13. *Let $\widetilde{X}$ be defined as above, and let Φ denote the standard normal distribution function. Then*

$$\sup_x |F_{\widetilde{X}/\sigma}(x) - \Phi(x)| \ll \frac{\mathbb{E}|\widetilde{X}|^3}{\sigma^4},$$

provided that the third absolute moment $\mathbb{E}|\widetilde{X}|^3$ is finite.

In order to estimate the difference between two characteristic functions effectively using Theorem 4.12, we need to consider the moments of the corresponding random variables. Given a random variable $\widetilde{X}$ as defined above, we can express its l^{th} moment

$\widetilde{\mu}_l := \mathbb{E}\widetilde{X}^l$ as

$$\widetilde{\mu}_l = \frac{1}{2^l}\sum_{i=0}^{l}\binom{l}{i}\widetilde{\mu}_{i,l-i} \quad \text{by writing} \quad \widetilde{\mu}_{i,j} := \mathbb{E}X^i\overline{X}^j.$$

We set up analogous expressions for the character sum defining A by writing

$$f(x) := \sum_{j=1}^{k}\gamma_j(x) \text{ with real part } \widetilde{f}(x) := \Re f(x) \text{ and } l^{th} \text{ moment } \widetilde{\nu}_l := \frac{1}{N}\sum_{x=1}^{N}\widetilde{f}(x)^l.$$

The l^{th} moment of $\widetilde{f}$ can likewise be expanded as

$$\widetilde{\nu}_l = \frac{1}{2^l}\sum_{i=0}^{l}\binom{l}{i}\widetilde{\nu}_{i,l-i} \quad \text{upon setting} \quad \widetilde{\nu}_{i,j} := \frac{1}{N}\sum_{x=1}^{N} f(x)^i\overline{f(x)^j}.$$

Let $F_{\widetilde{X}}$, $F_{\widetilde{f}}$ denote the obvious distribution functions, and write $\phi_{\widetilde{X}}$, $\phi_{\widetilde{f}}$ for the corresponding characteristic functions.

We are interested in the distribution of $\widetilde{f}$. More precisely, in order to estimate the size of A we want to count the number of elements $x \in \mathbb{Z}_N$ such that $\widetilde{f}(x) \geq \epsilon\sqrt{k}$. This means that $1 - F_{\widetilde{f}}(\epsilon\sqrt{k})$ is the quantity we are ultimately interested in.

Our first lemma shows that K-independence guarantees that the lower moments of $\widetilde{f}$ and $\widetilde{X}$ are equal.

Lemma 4.14. *With the moments $\widetilde{\mu}_l$ and $\widetilde{\nu}_l$ defined as above and the characters $\gamma_1, \gamma_2, \ldots, \gamma_k$ assumed to be K-independent, we have $\widetilde{\nu}_l = \widetilde{\mu}_l$ for all $l = 1, 2, \ldots, K$.*

Proof. Under the assumption of K-independence, it is not too difficult to compute the mixed moments explicitly. Indeed, we can rewrite $\widetilde{\nu}_{i,j}$ as

$$\frac{1}{N}\sum_{x=1}^{N}\left(\sum_{m=1}^{k}\gamma_m(x)\right)^i\left(\sum_{n=1}^{k}\overline{\gamma_n(x)}\right)^j = \frac{1}{N}\sum_{\substack{m_1,\ldots,m_i\\ n_1,\ldots,n_j}}\sum_{x=1}^{N} e((r_{m_1}+\ldots+r_{m_i}-r_{n_1}-\ldots-r_{n_j})x/N).$$

Whenever $i+j \leq K$, the latter sum equals zero by K-independence unless $m_1, \ldots, m_i$ is a permutation of $n_1, \ldots, n_j$, in which case it equals N. We compare this with

$$\widetilde{\mu}_{i,j} = \mathbb{E}\left(\sum_{m=1}^{k}X_m\right)^i\left(\sum_{n=1}^{k}\overline{X_n}\right)^j = \sum_{m_1,\ldots,m_i=1}^{k}\sum_{n_1,\ldots,n_j=1}^{k}\mathbb{E}X_{m_1}\ldots X_{m_i}\overline{X}_{n_1}\ldots\overline{X}_{n_j}.$$

Again, since X_i is independent of X_j for $i \neq j$, the expectation is non-zero only when $m_1, \ldots, m_i$ is a permutation of $n_1, \ldots, n_j$, in which case it equals 1. Hence $\widetilde{\nu}_{i,j} = \widetilde{\mu}_{i,j}$ for all $i + j \leq K$, and the result follows as stated. □

In order to usefully estimate the difference between the two characteristic functions we also need to infer a decent bound on the l^{th} moment $\widetilde{\mu}_l$.

Lemma 4.15. *For any even integer $l \leq K$ and $\widetilde{\mu}_l$ defined as above, we have the upper bound*

$$\widetilde{\mu}_l \leq \min\left\{k^l, \frac{l!}{2^l(l/2)!}k^{l/2}\right\}.$$

Proof. The first part of the bound is obvious, and the second follows from the fact that the only non-zero mixed moments $\widetilde{\mu}_{i,l-i}$ are those for which $i = l/2$, when they are of magnitude $k^{l/2}(l/2)!$. □

We are now ready to carry out the first step of the argument, namely showing that $\widetilde{f}$ and $\widetilde{X}$ are close in distribution using Theorem 4.12.

Proposition 4.16. *Under the same assumptions as before, $\widetilde{f}$ and $\widetilde{X}$ are close in distribution in the sense that*

$$\sup_x |F_{\widetilde{X}}(x) - F_{\widetilde{f}}(x)| \ll \min\left\{\frac{1}{\sqrt{K}}, \frac{\sqrt{k}}{K}\right\}.$$

Proof. In order to apply Esseen's Inequality, we first need to verify that $F'_{\widetilde{X}}$ exists and is bounded above by a suitable constant. As we have already mentioned, it is a well-known fact in probability theory that the density function of a random variable is the inverse Fourier transform of its characteristic function, hence

$$F'_{\widetilde{X}}(x) \leq \int_{-\infty}^{\infty} |\phi_{\widetilde{X}}(t)|dt.$$

We thus require the following bounds on the characteristic function $\phi_{\widetilde{X}}$ of $\widetilde{X}$, which we state here without proof. The interested reader is referred to [Ruz91] for details.

Lemma 4.17. *There exist constants $a, b > 0$ and $T_0 > 1$ such that $\phi_{\widetilde{X}}$ satisfies*

$$|\phi_{\widetilde{X}}(t)| \leq \begin{cases} \exp\left(-akt^2\right) & |t| \leq T_0\sigma \\ (b|t|)^{-k/2} & |t| > T_0\sigma \end{cases}.$$

It is immediate to deduce that $F'_{\widetilde{X}}(x)$ is bounded above by a constant times the standard deviation σ. Next we observe that by Taylor's Theorem with remainder we can write

$$\phi_{\widetilde{X}}(t) = \sum_{j=1}^{l-1} \frac{\widetilde{\mu}_j}{j!}(it)^j + \delta\widetilde{\mu}_l \frac{|t|^l}{l!},$$

and similarly

$$\phi_{\widetilde{f}}(t) = \sum_{j=1}^{l-1} \frac{\widetilde{\nu}_j}{j!}(it)^j + \delta\widetilde{\nu}_l \frac{|t|^l}{l!}$$

for some $|\delta| \leq 1$. With the benefit of hindsight, this allows us to justify why we were so keen to compare moments in the first place. K-independence gave us through Lemma 4.14 that all moments $\widetilde{\mu}_j$ and $\widetilde{\nu}_j$ up to order K were equal, and thus

$$|\phi_{\widetilde{X}}(t) - \phi_{\widetilde{f}}(t)| \leq 2\widetilde{\mu}_K \frac{|t|^K}{K!}.$$

It now follows from Theorem 4.12 that for any $T > 1$,

$$\sup_x |F_{\widetilde{X}}(x) - F_{\widetilde{f}}(x)| \ll \frac{\sigma}{T} + \widetilde{\mu}_K \frac{T^K}{K!K}.$$

Using the bound on $\widetilde{\mu}_K$ derived in Lemma 4.15 and setting $T = \sigma(K!/\widetilde{\mu}_K)^{1/(K+1)}$ followed by a short computation concludes the proof of Proposition 4.16. □

We have thus successfully approximated $\widetilde{f}$ by $\widetilde{X}$. It remains to compare a suitably normalized version of $\widetilde{X}$ to a standard normal random variable. The following proposition states that $\widetilde{X}$ is close to a normal distribution with mean 0 and standard deviation σ.

Proposition 4.18. *Let $\widetilde{X}$ be defined as above, and let Φ denote the standard normal distribution function. Then*

$$\sup_x |F_{\widetilde{X}/\sigma}(x) - \Phi(x)| \ll \frac{1}{\sigma}.$$

Proof. This is a straightforward application of Theorem 4.13. The third absolute moment $\mathbb{E}|\widetilde{X}|^3$ can be bounded by the Cauchy-Schwarz Inequality as

$$\mathbb{E}|\widetilde{X}|^3 \leq (\mathbb{E}|\widetilde{X}|^2)^{1/2}(\mathbb{E}|\widetilde{X}|^4)^{1/2}.$$

Splitting X_j into real and imaginary parts $X_j = R_j + iI_j$, we first observe that $\mathbb{E}R_j^2 = 1/2$ and $\mathbb{E}I_j^2 = 1/2$ as well as $\mathbb{E}R_j^4 = 3/8$. It is not hard to see that X_i and

X_j are independent if and only if the pairs (R_i, I_i) and (R_j, I_j) are independent (but see page 273 of [Shi84] for a justification of this claim), which yields

$$\mathbb{E}\widetilde{X}^2 = \mathbb{E}\sum_{j,l=1}^{k} R_j R_l = \sum_{j=1}^{k} \mathbb{E}R_j^2 + \sum_{j\neq l=1}^{k} \mathbb{E}R_j R_l = \frac{k}{2}$$

and

$$\mathbb{E}\widetilde{X}^4 = \mathbb{E}\sum_{j=1}^{k} R_j^4 + \sum_{j,l=1}^{k} \mathbb{E}R_j^2 \mathbb{E}R_l^2 = \left(\frac{k}{2}\right)^2 + \frac{3}{4}k.$$

This implies that $\mathbb{E}|\widetilde{X}|^3 \ll \sigma^3$, and the result follows as claimed from Theorem 4.13. □

We remark that in fact Ruzsa [Ruz91] proves the slightly stronger error term of σ^{-2}, but we shall not need to do so here. Proposition 4.18 completes the second step of the argument, so we are now in a position to estimate the size of the niveau set A.

Proof of Proposition 4.11. Bearing in mind that by definition of the distribution function $F_{\widetilde{X}/\sigma}(x) = F_{\widetilde{X}}(\sigma x)$, we deduce from Propositions 4.16 and 4.18 the existence of two constants c and c' such that

$$F_{\widetilde{f}}(\epsilon\sqrt{k}) \leq F_{\widetilde{X}}(\epsilon\sqrt{k}) + c\min\left\{\frac{1}{\sqrt{K}}, \frac{\sqrt{k}}{K}\right\} \leq \Phi(\sqrt{2}\epsilon) + c\min\left\{\frac{1}{\sqrt{K}}, \frac{\sqrt{k}}{K}\right\} + c'\frac{1}{\sqrt{k}}.$$

It is easy to compute that for $\epsilon \leq 1/4$, the value of the standard normal distribution function Φ at $\sqrt{2}\epsilon$ is bounded above by $2/3$, so that the size of the set A is at least $N/3$. In fact, the density can be made arbitrarily close to $1/2$ by choosing ϵ small enough. We also need to ensure that the error term $\sqrt{k}/K$ tends to 0 as N tends to infinity, and that K satisfies $K \ll N^{1/k}$. We therefore require that k grow at most like a constant times $\log N/\log\log(N)$. This proves Proposition 4.11 for N sufficiently large. □

4.3.2 Counting the Number of Representations in $A - A$

This section is devoted to proving the analogue of Lemma 4.5 for the finite Abelian group $\mathbb{Z}_N$. More precisely, we shall show that the popular difference set $D_\gamma(A)$ is contained in the complement of a ball $B_t(1)$, which in this context will be defined as

$$B_t(1) := \left\{x \in \mathbb{Z}_N : \sum_{i=1}^{k} |\gamma_i(x) + 1| \leq t\right\} = \left\{x \in \mathbb{Z}_N : \sum_{i=1}^{k} \cos(2\pi x r_i/N) \leq -k + t\right\}$$

using the same set $\gamma_1, \ldots, \gamma_k$ of K-independent characters as the niveau set A. Of course we hope to be able to take the radius t as large as possible.

Proposition 4.19. *For every $\gamma > 0$ there exist constants $\epsilon > 0$ and $\beta > 0$ with the following property. Suppose that $k \ll \log N / \log\log N$, write $t := \beta k$ and let the niveau set A with parameters ϵ and k be defined as above. Then we have the inclusion*

$$D_\gamma(A) \subseteq B_t(1)^C.$$

Let us first observe, as is done in Ruzsa's original paper [Ruz91], that the complete difference set $A - A$ is contained in the complement of the ball $B_{4\epsilon\sqrt{k}}(1)$. Indeed, for arbitrary $x, y \in A$, we have

$$2\epsilon\sqrt{k} \leq \Re\left[\sum_{i=1}^{k} \gamma_i(x) + \sum_{i=1}^{k} \gamma_i(y)\right],$$

which in turn is bounded above by

$$\left|\sum_{i=1}^{k} \gamma_i\left((x+y)/2\right)\left(\gamma_i\left((x-y)/2\right) + \gamma_i\left(-(x-y)/2\right)\right)\right| \leq \sum_{i=1}^{k} \left|\cos(\pi(x-y)r_i/N)\right|.$$

This proves our claim. It stands to reason that the set of popular differences $D_\gamma(A)$ should be contained in the complement of a much larger ball around 1. However, a trivial adaptation of the method we used in the model setting $\mathbb{F}_2^n$, that is, coordinate-wise counting, falls short of what is required.

Recall that we would like to show that for fixed $z \in B_t(1)$, the number of representations of z as a difference $x - y$ with x and y in A is strictly less than γN. In other words, our aim is to establish that for fixed $z \in B_t(1)$, there are few elements x such that both $x \in A$ and $x - z \in A$. This condition is equivalent to counting the number of elements $x \in \mathbb{Z}_N$ that satisfy both $\Re\sum_{j=1}^{k} \gamma_j(x) > \epsilon\sqrt{k}$ and $\Re\sum_{j=1}^{k} \gamma_j(x-z) > \epsilon\sqrt{k}$, under the assumption that $\sum_{j=1}^{k} |\gamma_j(z) + 1| = \beta k$ with $\beta = t/k$. As before, we write

$$f(x) := \sum_{j=1}^{k} \gamma_j(x) \quad \text{with real part} \quad \widetilde{f}(x) := \Re f(x),$$

but now we also need

$$g(x) = \sum_{j=1}^{k} \gamma_j(x - z) \quad \text{with real part} \quad \widetilde{g}(x) = \Re g(x).$$

Thus we are interested in an upper bound on the probability that both $\widetilde{f}$ and $\widetilde{g}$ are greater than $\epsilon\sqrt{k}$, under the hypothesis that $\sum_{j=1}^{k} |\gamma_j(z) + 1| = \beta k$. It turns out that when the parameter β is small enough, the functions $\widetilde{f}$ and $\widetilde{g}$ are sufficiently negatively correlated for this probability to be less than γ.

In order to prove this, we shall use techniques very similar to the ones we used to establish a lower bound on the size of A in the preceding section. We shall first compare the joint distribution of $(\widetilde{f}, \widetilde{g})$ with the joint distribution of two sums of appropriately defined independent random variables, and then compare their distribution to a suitable bi-variate normal.

It should be obvious at this point that we will need a 2-dimensional analogue of Esseen's Inequality, which can be found in [Sad66] and [Ber45] (with better bounds in the former).

Theorem 4.20. *Let F_1, F_2 be 2-dimensional distribution functions, and let ϕ_1, ϕ_2 be the corresponding characteristic functions. Write $\widetilde{\phi_i}(s,t) = \phi_i(s,t) - \phi_i(s,0)\phi_i(0,t)$ for $i = 1, 2$, and set*

$$\gamma_1 := \sup_{x,y} \frac{\partial F_2(x,y)}{\partial x} \quad , \quad \gamma_2 := \sup_{x,y} \frac{\partial F_2(x,y)}{\partial y}.$$

Then for any $T > 0$, the total variation distance $\sup_{x,y} |F_1(x,y) - F_2(x,y)|$ is bounded above by

$$\frac{2}{(2\pi)^2} \int_{-T}^{T} \int_{-T}^{T} \left| \frac{\widetilde{\phi_1}(s,t) - \widetilde{\phi_2}(s,t)}{st} \right| dsdt$$

plus an additional error term of the form

$$\frac{2}{\pi} \int_{-T}^{T} \left| \frac{\phi_1(s,0) - \phi_2(s,0)}{s} \right| ds + \frac{2}{\pi} \int_{-T}^{T} \left| \frac{\phi_1(0,t) - \phi_2(0,t)}{t} \right| dt + \frac{(6\sqrt{2} + 8\sqrt{3})(\gamma_1 + \gamma_2)}{T}.$$

As a more or less immediate corollary we have the 2-dimensional Berry-Esseen Inequality, the precise statement of which is taken from [Sad66].

Theorem 4.21. *Let $\widetilde{X}$ and $\widetilde{Z}$ be sums of k independent identically distributed mean-zero random variables $\widetilde{X}_i$, $\widetilde{Z}_i$, respectively. Let Φ_ρ denote the distribution function of a standard bi-variate normal distribution with correlation ρ. Suppose that $\widetilde{X}$ and $\widetilde{Z}$ have correlation ρ, and denote their joint distribution function by $F_{(\widetilde{X},\widetilde{Z})}$. Then*

$$\sup_{x,z} |F_{(\widetilde{X}/\sigma, \widetilde{Z}/\sigma)}(x,z) - \Phi_\rho(x,z)| \ll \frac{\widetilde{\mu}_{3,0}^{abs} + \widetilde{\mu}_{0,3}^{abs}}{\sigma^2(1-\rho^2)^2 \min\{\widetilde{\mu}_{2,0}^{3/2}, \widetilde{\mu}_{0,2}^{3/2}\}},$$

where we have written

$$\widetilde{\mu}_{i,j} := \mathbb{E}\widetilde{X}^i\widetilde{Z}^j \quad \textit{and} \quad \widetilde{\mu}^{abs}_{i,j} := \mathbb{E}|\widetilde{X}^i\widetilde{Z}^j|.$$

Let us put our idea into practice and first compare the joint distribution of $\widetilde{f}$ and $\widetilde{g}$ to the joint distribution of two sums of sequences of independent random variables with correlation ρ. In addition to

$$X := \sum_{j=1}^{k} X_j \quad \text{with real part} \quad \widetilde{X} := \Re X,$$

we now also define

$$Z := \sum_{j=1}^{k} \gamma_j(-z)X_j \quad \text{with real part} \quad \widetilde{Z} := \Re Z,$$

where the X_i are independently and uniformly distributed on the unit circle as in Section 4.3.1. We first show that $(\widetilde{f},\widetilde{g})$ and $(\widetilde{X},\widetilde{Z})$ are close in distribution using Theorem 4.20.

Proposition 4.22. *Let $(\widetilde{X},\widetilde{Z})$ and $(\widetilde{f},\widetilde{g})$ be defined as above, and let their joint distribution functions be denoted by $F_{(\widetilde{X},\widetilde{Z})}$ and $F_{(\widetilde{f},\widetilde{g})}$, respectively. Then the total variation distance satisfies*

$$\sup_{x,z} |F_{(\widetilde{X},\widetilde{Z})}(x,z) - F_{(\widetilde{f},\widetilde{g})}(x,z)| \ll \min\left\{\frac{1}{\sqrt{K}}, \frac{\sqrt{k}}{K}\right\}.$$

Proof. We need to consider the characteristic functions

$$\phi_{(\widetilde{f},\widetilde{g})}(s,t) = \frac{1}{N}\sum_{x=1}^{N} \exp\left(i(s\widetilde{f}(x) + t\widetilde{g}(x))\right) \quad \text{and} \quad \phi_{(\widetilde{X},\widetilde{Z})}(s,t) = \mathbb{E}\exp\left(i(s\widetilde{X} + t\widetilde{Z})\right).$$

It is easy to check that the partial derivatives of $F_{(\widetilde{X},\widetilde{Z})}$ are bounded above by a constant times the standard deviation σ. Indeed, let $\eta(s,t)$ denote the joint probability density function of $(\widetilde{X},\widetilde{Z})$. By definition, we have

$$\sup_{x,z} \frac{\partial F_{(\widetilde{X},\widetilde{Z})}(x,z)}{\partial x} = \int_{-\infty}^{z} \eta(x,t)dt,$$

which by positivity of the probability density function η is bounded above by

$$\int_{-\infty}^{\infty} \eta(x,t)dt = F'_{\widetilde{X}}(x).$$

The final expression is exactly the same term as in the 1-dimensional case, which we bounded by a constant times σ using Lemma 4.17. An analogous inequality holds for the partial derivative with respect to z.

The second and third term in the bound in Theorem 4.20 are bounded above just as in the 1-dimensional case. It remains to estimate the main error term, and we shall proceed as before by comparing moments. As in the proof of Proposition 4.11, we can write

$$\phi_{(\widetilde{X},\widetilde{Z})}(s,t) = \sum_{j=1}^{l-1} \frac{i^j}{j!}\mathbb{E}(s\widetilde{X}+t\widetilde{Z})^j + \delta\frac{\mathbb{E}|s\widetilde{X}+t\widetilde{Z}|^l}{l!}$$

with $|\delta| \leq 1$, and similarly with $(\widetilde{X},\widetilde{Z})$ replaced by $(\widetilde{f},\widetilde{g})$. Let's have a closer look at $\mathbb{E}(s\widetilde{X}+t\widetilde{Z})^l$, which can be expressed as

$$\sum_{i=1}^{l}\binom{l}{i}s^i t^{l-i}\mathbb{E}\widetilde{X}^i\widetilde{Z}^{l-i} = \frac{1}{2^l}\sum_{i=1}^{l}\binom{l}{i}s^i t^{l-i}\sum_{c=1}^{i}\sum_{d=1}^{l-i}\binom{i}{c}\binom{l-i}{d}\mathbb{E}X^c\overline{X}^{i-c}Z^d\overline{Z}^{l-i-d}.$$

After defining the mixed moments

$$\xi_{i,j,c,d} := \mathbb{E}X^c\overline{X}^{i-c}Z^d\overline{Z}^{j-d} \quad \text{and} \quad \theta_{i,j,c,d} := \mathbb{E}f(x)^c\overline{f(x)}^{i-c}g(x)^d\overline{g(x)}^{j-d},$$

the expression for the l^{th} moment becomes

$$\mathbb{E}(s\widetilde{X}+t\widetilde{Z})^l = \frac{1}{2^l}\sum_{i=1}^{l}\binom{l}{i}s^i t^{l-i}\sum_{c=1}^{i}\sum_{d=1}^{l-i}\binom{i}{c}\binom{l-i}{d}\xi_{i,l-i,c,d}.$$

As in the 1-dimensional case, we need a lemma saying that for independent characters, the low mixed moments $\xi_{i,j,c,d}$ and $\theta_{i,j,c,d}$ are equal.

Lemma 4.23. *For all $1 \leq c \leq i, 1 \leq d \leq j$ and $i+j \leq K$, we have that $\xi_{i,j,c,d} = \theta_{i,j,c,d}$.*

Proof. It is easily checked that under the given conditions both expressions reduce to the number of sequences $(m_1,\ldots,m_c,n_1,\ldots,n_c)$ and $(m'_1,\ldots,m'_c,n'_1,\ldots,n'_c)$ that are permutations of each other. □

We also need to prove a bound on $\mathbb{E}|s\widetilde{X}+t\widetilde{Z}|^l$ for even l in the style of Lemma 4.15.

Lemma 4.24. *For any even integer $l \leq K$ and $\widetilde{X}$, $\widetilde{Z}$ defined as above, we have*

$$\mathbb{E}|s\widetilde{X} + t\widetilde{Z}|^l \leq \frac{k^{l/2} l!}{2^l (l/2)!}(|s| + |t|)^l.$$

Proof. This is a straightforward computation just as in the 1-dimensional case. The moment $\xi_{i,j,c,d}$ is easily to be seen non-zero only when $2(c+d) = i+j$, in which case its absolute value is bounded above by $k^{c+d}(c+d)!$. The l^{th} moment is therefore bounded by

$$\frac{1}{2^l} \sum_{i=1}^{l} \binom{l}{i} s^i t^{l-i} \sum_{c=1}^{i} \binom{i}{c} \binom{l-i}{l/2-c} k^{l/2} (l/2)!.$$

The sum over c in this expression is no greater than

$$\sum_{c=1}^{l/2} \binom{i}{c} \binom{l-i}{l/2-c} k^{l/2} (l/2)!$$

and by Vandermonde convolution, the sum over the binomial coefficients actually equals $\binom{l}{l/2}$. The statement of the lemma now follows as claimed. □

We have now gathered enough information to estimate the main error term in Theorem 4.20. A not too lengthy computation using Lemmas 4.23 and 4.24 concludes the proof of Proposition 4.22 for the appropriate choice of the parameter T. □

It remains to compare the joint distribution of $(\widetilde{X}, \widetilde{Z})$ to a bi-variate standard normal distribution, and we shall do so using Theorem 4.21 in the following proposition.

Proposition 4.25. *Let $\widetilde{X}$ and $\widetilde{Z}$ be defined as above, and write $F_{\widetilde{X},\widetilde{Z}}$ for their joint distribution function. Let Φ_ρ denote the standard bi-variate normal distribution function with correlation ρ. Then*

$$\sup_{x,z} |F_{(\widetilde{X}/\sigma, \widetilde{Z}/\sigma)}(x,z) - \Phi_{-1+\beta}(x,z)| \ll \frac{1}{\sigma^{1/2}}.$$

Proof. We have already seen in Proposition 4.18 that the third absolute moment of $\widetilde{X}$ is bounded above by σ^3. A similar analysis can be carried out for $\widetilde{Z}$. For instance, writing $z_j = -zr_j/N$ for $r_1, \ldots, r_k \in \mathbb{Z}_N$ corresponding to the characters $\gamma_1, \ldots, \gamma_k$, we find that

$$\mathbb{E}\widetilde{Z}^2 = \mathbb{E}(\sum_{j=1}^{k} \cos 2\pi z_j R_j - \sin 2\pi z_j I_j)^2 = \sum_{j=1}^{k} (\cos 2\pi z_j)^2 \mathbb{E}R_j^2 + (\sin 2\pi z_j)^2 \mathbb{E}I_j^2 = \frac{k}{2}.$$

Therefore the third absolute moments $\widetilde{\mu}_{3,0}^{abs}$ and $\widetilde{\mu}_{0,3}^{abs}$ are both bounded by σ^3. Finally, we need to check that $\widetilde{X}$ and $\widetilde{Z}$ have the required correlation, so we compute the covariance

$$\mathbb{E}\widetilde{X}\widetilde{Z} = \mathbb{E}\sum_{j=1}^{k} R_j \sum_{l=1}^{k} \cos 2\pi z_l R_l - \sin 2\pi z_l I_l = \sum_{j=1}^{k} \cos 2\pi z_j \mathbb{E}R_j^2 = (-1+\beta)\frac{k}{2}$$

by the condition we imposed on the $(z_j)_{j=1}^k$ by requiring that $z \in B_t(1)$. Thus the correlation, which is always a dimension-less quantity, of the two random variables $\widetilde{X}/\sigma$ and $\widetilde{Z}/\sigma$ with mean 0 and variance 1 is

$$\rho = \frac{\mathbb{E}\widetilde{X}\widetilde{Z}}{\sqrt{\mathbb{E}\widetilde{X}^2\mathbb{E}\widetilde{Z}^2}} = -1+\beta.$$

Proposition 4.25 now follows from Theorem 4.21. □

Last but not least, now that we have successfully approximated the distribution of $(\widetilde{f},\widetilde{g})$ by a bi-variate normal distribution, we turn to computing the corresponding bi-variate probability.

Lemma 4.26. *For every $\gamma > 0$ there exist constants $\epsilon > 0$ and $\beta > 0$ with the following property. Let Φ_ρ denote the standard bi-variate normal distribution function with correlation ρ. Then*

$$\Phi_{-1+\beta}(\sqrt{2}\epsilon, \sqrt{2}\epsilon) \leq \gamma.$$

Proof. This is a straightforward computation. The probability we would like to bound can be calculated as

$$\Phi_\rho(\sqrt{2}\epsilon,\sqrt{2}\epsilon) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\sqrt{2}\epsilon}\int_{-\infty}^{\sqrt{2}\epsilon} \exp\left(-\frac{1}{2(1-\rho^2)}\left(y^2 - 2\rho yw + w^2\right)\right)dydw,$$

with $\rho = -1+\beta$ as before. One could use standard approximations to the bivariate normal such as [Pol46] to obtain explicit estimates, but we shall confine ourselves to asserting that the probability in question is less than γ provided that β and ϵ are sufficiently small. □

Summarising our work in this section, we have shown that $D_\gamma(A)$ is contained in the complement of a ball $B_t(1)$, where the parameter $\beta = t/k$ can be taken to be a small constant depending on γ, that is, the radius t can be taken to be of order k. This compares favourably with the statement of Lemma 4.5 in the model setting $\mathbb{F}_2^n$, where $n = \log N$ played the rôle of the parameter k.

4.3.3 Using Concentration of Measure on the Torus

In this final section we prove the $\mathbb{Z}_N$-analogue of Lemma 4.6, that is, we show that for an appropriately chosen parameter t the complement of a ball $B_t(1)$ contains only difference sets of sets of density $o(1)$.

Proposition 4.27. *Let β be a constant and write $t = \beta k$ with $k \ll \sqrt{\log N}$. Let A_0 be any subset of $\mathbb{Z}_N$ with the property that $A_0 - A_0 \subseteq B_t(1)^C$. Then the density of A_0 is bounded above by $\exp(-\beta^2 k/72)$.*

By considering the map

$$\Psi : \mathbb{Z}_N \to \mathbb{T}^k,$$

which takes $x \mapsto (\arg\gamma_1(x), \arg\gamma_2(x), ..., \arg\gamma_k(x))/2\pi$, we move the problem to the k-dimensional torus $\mathbb{T}^k$, where appropriate measure concentration results are known. For an exhaustive survey of all aspects of measure concentration we recommend the book [Led01], and in particular Chapter 4 on concentration in product spaces. It puts into context as well as generalizes the classical probabilistic inequalities by Talagrand, which in turn are based on martingale results by Hoeffding (1963) and Azuma (1967). The precise statement of Theorem 4.28 below can be taken from page 71 of [Led01], or page 173 of [McD89], whose excellent survey article emphasizes applications to combinatorial and discrete structures.

Theorem 4.28. *Let G be a compact metric group with a translation invariant metric d and let*

$$G = G_0 \supseteq G_1 \supseteq ... \supseteq G_n = \{1_G\}$$

be a decreasing sequence of closed subspaces of G. Let $a_i = diam(G_{i-1}/G_i)$, and write $l = (\sum_{i=1}^n a_i^2)^{1/2}$. Let μ be Haar measure on G. Then for any measurable subset E of G, we have

$$\mu(E + B_d(0,t)) \geq 1 - \frac{\exp(-t^2/2l^2)}{\mu(E)}.$$

For the application we have in mind, let $G = \mathbb{T}^k$ be equipped with normalised product measure μ and metric $d(s,t) = \sum_{i=1}^k |\sin\pi(s_i - t_i)|$. It is easily checked that d is indeed a translation invariant metric on G which encapsulates the antipodal concept. Setting $G_i = \mathbb{T}^{k-i}$, the diameter a_i of each quotient G_{i-1}/G_i equals 1, whence $l^2 = k$. Denote by $C_t(1)$ the ball

$$C_t(1) := \left\{ x \in \mathbb{T}^k : \sum_{j=1}^k |\gamma_i(x) + 1| \leq t \right\}.$$

The reader may care to verify that $C_t(1)$ coincides with a ball in the metric d as defined above of radius $t/2$ about the point $(1/2, 1/2, \ldots, 1/2) \in \mathbb{T}^k$. We thus have the following quantitative statement of measure concentration in $\mathbb{T}^k$ with respect to the special metric d, which arises from the definition of the niveau set.

Corollary 4.29. *Let the metric d be defined as above, and let E be a measurable subset of $\mathbb{T}^k$. We have the bound*

$$\mu(\overline{E} + C_t(1)) \geq 1 - \frac{\exp(-t^2/8k)}{\mu(E)},$$

where the bar indicates translation by $(1/2, 1/2, \ldots, 1/2) \mod 1$.

Recall that in the model setting $\mathbb{F}_2^n$ in Section 4.2, we used the fact that for any subset $A_0 \subseteq \mathbb{F}_2^n$,

$$A_0 - A_0 \subseteq B_t(1)^C \quad \Rightarrow \quad \overline{A_0} + B_t(0) \cap A_0 = \emptyset.$$

In the group $\mathbb{Z}_N$ it follows from the fact that Ψ is linear and injective that any subset $A_0 \subseteq \mathbb{Z}_N$ with the property that $A_0 - A_0 \subseteq B_t(1)$ satisfies

$$\Psi(A_0) - \Psi(A_0) \subseteq \Psi(B_t(1)^C) = \Psi(\mathbb{Z}_N) \setminus \Psi(B_t(1)) = \Psi(\mathbb{Z}_N) \cap C_t(1)^C \subseteq C_t(1)^C,$$

and further that

$$\Psi(A_0) + C_t(1) \cap \Psi(A_0) = \emptyset \quad \Rightarrow \quad (\Psi(A_0) + C_{t/3}(1)) + C_{t/3}(1) \cap (\Psi(A_0) + C_{t/3}(1)) = \emptyset.$$

The set $\Psi(A_0) + C_{t/3}(1)$ is a union of balls in $\mathbb{T}^k$ centred at the image points of A_0 under the map Ψ. Corollary 4.29 now gives us a bound on the measure of this set of the form

$$\mu(\overline{\Psi(A_0)} + C_{t/3}(1)) \leq \exp(-t^2/72k). \tag{4.1}$$

We are almost done. Because the characters $\gamma_1, \ldots, \gamma_k$ are K-independent, we expect the image of $\mathbb{Z}_N$ under the map Ψ to be roughly uniformly distributed in $\mathbb{T}^k$. As we shall see shortly, this implies that the translates of the ball $C_{t/3}(1)$ generate a set of measure proportional to the density of A_0, so that we will be able to infer a bound on this density from the bound on the measure of $\overline{\Psi(A_0)} + C_{t/3}(1)$. The remainder of this section serves to make these remarks more precise.

We first turn to the equidistribution of $\mathbb{Z}_N$ under the map Ψ. We have already seen in the preceding sections that K-independence of the characters $\gamma_1, \ldots, \gamma_k$ gives us

rather precise information about their distribution, and we are about to exploit this fact yet again. Let us define the *discrepancy* of a set of points $y_1, ..., y_N$ in $\mathbb{T}^k$ by

$$\mathrm{disc}(y_1, \ldots, y_N) := \sup_{B^\infty \subseteq \mathbb{T}^k} \left| \frac{|\{i : y_i \in B^\infty\}|}{N} - \mu(B^\infty) \right|,$$

where the supremum is taken over all L^∞-balls $B^\infty \subseteq \mathbb{T}^k$ and μ is, of course, Lebesgue measure as before. We shall be able to give a bound on the discrepancy of the set $\Psi(\mathbb{Z}_N)$ using the following proposition known as the *Erdős-Turán-Koksma Inequality*. It can be viewed as a quantitative version of Kronecker's Equidistribution Theorem and is taken from page 15 of [DT97].

Proposition 4.30. *Let $y_1, ..., y_N$ be points in $\mathbb{T}^k$, and let $K \in \mathbb{N}$. Then the discrepancy* $\mathrm{disc}(y_1, \ldots, y_N)$ *satisfies the bound*

$$\mathrm{disc}(y_1, \ldots, y_N) \leq \left(\frac{3}{2}\right)^k \left(\frac{2}{K+1} + \sum_{0<\|h\|_\infty \leq K} \frac{1}{r(h)} \left| \frac{1}{N} \sum_{i=1}^{N} e(h \cdot y_i) \right| \right),$$

where $r(h) = \prod_{i=1}^k \max\{1, |h_i|\}$ for $h = (h_1, ..., h_k) \in \mathbb{Z}^k$.

It should be noted (and is discussed at length in [NP73]) that Proposition 4.30 is very closely related to the Berry-Esseen Inequality. Its proof is again purely Fourier analytic, and we use it here as a black box for pure convenience. As an immediate corollary we have the following result for K-independent characters, once again illustrating the principle that K-independence of characters is the Fourier analytic (and quantitative) analogue of the notion of independence of random variables.

Corollary 4.31. *Given the map Ψ defined as above by a set $\gamma_1, \ldots, \gamma_k$ of K-independent characters, we have the bound*

$$\mathrm{dics}(\Psi(\mathbb{Z}_N)) \ll \left(\frac{3}{2}\right)^k \frac{1}{K}.$$

In other words,

$$|\{x \in \mathbb{Z}_N : \Psi(x) \in B_\eta^\infty\}| = \mu(B_\eta^\infty)N + O((3/2)^k N/K)$$

for all L^∞-balls $B_\eta^\infty \in \mathbb{T}^k$ of side length $\eta \gg K^{-1/k}$.

Recall that in Section 4.3 we were forced to choose $K \ll N^{1/k}$ in order for a set of K-independent characters of cardinality k to exist. This implies that we are able to

resolve down to subcubes of side length $\eta \gg N^{-1/k^2}$. It is this restriction that is chiefly responsible for our bound in Theorem 4.1 in the case $G = \mathbb{Z}_N$.

Finally, we are able to make the transition from a bound on the measure of $\overline{\Psi(A_0)} + C_{t/3}(1)$ to a bound on the density of A_0.

Lemma 4.32. *Let $k \ll \sqrt{\log N}$, and let $\gamma_1, \dots, \gamma_k$ be a set of K-independent characters. Let the map Ψ be defined as above. Then for any set $A_0 \subseteq \mathbb{Z}_N$ we have*

$$|A_0| \leq \mu(\Psi(A_0) + C_{t/3}(1))N.$$

Proof. First note that $C_{t/3}(1)$ always contains the L^∞-ball $B^\infty_{t/3k}$ of side length $t/3k = \beta/3$, which implies

$$\mu(\Psi(A_0) + C_{t/3}(1)) \geq \mu(\Psi(A_0) + B^\infty_{\beta/3}).$$

Now divide $\mathbb{T}^k$ into η^{-k} subcubes of sidelength η satisfying $\eta \gg N^{-1/k^2}$ and $\eta < \beta/3$. This determines the constant required in the growth rate of k. By averaging and Corollary 4.31, at least $|\Psi(A_0)|/\eta^k N$ of these subcubes contain at least one point of $\Psi(A_0)$. Suppose these non-empty subcubes are indexed by the set $I \subseteq [\eta^{-k}]$, so that $|I| \gg |\Psi(A_0)|/\eta^k N$. But by our choice of η the subcubes B_i are smaller than the L^∞-balls $B^\infty_{\beta/3}$. It follows that

$$\mu(\Psi(A_0) + B^\infty_{\beta/3}) \geq \mu\left(\cup_{i\in I} B_i\right) = \sum_{i\in I} \mu(B_i) \gg \frac{|A_0|}{\eta^k N}\eta^k,$$

and therefore we obtain the lemma as stated. □

Lemma 4.32 and Equation (4.1) combine to conclude the proof of Proposition 4.27. We now bring together Propositions 4.11, 4.19 and 4.27 in order to state the main result of this chapter.

Theorem 4.33. *There exists a set $A \subseteq \mathbb{Z}_N$ of size greater than $N/3$ with the property that any set A_0 whose difference set is contained in the set $D_\gamma(A)$ of γ-popular differences of A has density*

$$\exp\left(-c_\gamma \sqrt{\log N}\right),$$

where c_γ is a small constant depending on γ.

4.4 Remarks

Our analysis in Section 4.3 only relied on measure concentration in the k-dimensional torus and our ability to pick a set of independent characters. Therefore, it is evident that our methods will yield the statement of Theorem 4.1 in any finite Abelian group.

It would be interesting to establish whether the bounds in Theorem 4.1 could be improved to give a power-type decay as in Theorem 4.8.

Acknowledgements. The author would like to thank Ben Green for posing the problem and many helpful discussions. She is also indebted to Geoffrey Grimmett for sharing his insights into Esseen's Inequality.

Chapter A

Appendix: Estimates for the Weighted Squares

The material in this section is entirely standard and we give barely enough detail to make this exposition self-contained. For an introduction to the circle method, see [Vau81].

By Dirichlet's Theorem, $t/N \in I(a/q, (qQ)^{-1})$ for some $1 \leq a \leq q \leq Q, (a,q) = 1$. Call the set of those t for which $q \leq R$ the *major arcs* and the set of those t for which $R < q \leq Q = N/K$ the *minor arcs*. It is a typical feature of the Hardy-Littlewood method that the exact values of the boundaries between the arcs need to be determined in the course of the proof. Throughout, R will be of the order of magnitude of $K = e^{l^2}$ defined in the introduction to Chapter 1.

We define the generating function of the weighted squares by

$$F_S(\theta) = \sum_{x^2 \leq N_1} \frac{2x}{\sqrt{N_1}} e(x^2\theta).$$

Note that $F_S(t/N)/N$ coincides with our earlier definition of $\widehat{S}(t)$ used throughout the proof.

We would like to stress that although the estimates presented here are classical, one could alternatively view them as a manifestation of the fact that it is possible to decompose any bounded function into a structured and a random-looking part. In the case of the set of squares we can be very explicit about the structure we obtain.

We start off by considering simple weighted exponential sum estimates for the squares.

Lemma A.1. *Let θ belong to the interval $I(a/q,\eta)$. Then we have the bound*

$$|F_S(\theta)| \ll \frac{\sqrt{\log q}}{\sqrt{q}}|F_S(\eta)| + \sqrt{q\log q}(1+|\eta|N).$$

Proof. Consider the truncated version $F_S(\theta,m) = \sum_{x\leq m} 2xe(x^2\theta)/\sqrt{N_1}$ of F_S, as well as the Gauss sum $B(a/q,m) = \sum_{x\leq m} e(x^2a/q)$. If $m \leq q$, we have $B(a/q,m) \ll \sqrt{q\log q}$. Using Abel's Inequality, which says that if g is a monotone function, then $|\sum_{x\leq m} g(x)f(x)|$ is bounded above by $\sup_{x\leq m}|g(x)|\sup_{j\leq m}|\sum_{x\leq j} f(x)|$, we conclude that $F_S(a/q,m) \ll m\sqrt{q\log q/N}$. It follows that $F_S(a/q) \ll \sqrt{q\log q}$. In the case where $m > q$, we find $F_S(a/q,m) = B(a/q,q)m^2/(q\sqrt{N}) + O(m\sqrt{q\log q/N})$ by splitting into segments of length q, and so $F_S(a/q) = B(a/q,q)\sqrt{N}/q + O(\sqrt{q\log q})$. Now let $\theta = a/q+\eta$ with $(a,q)=1$. By partial summation, we obtain $F_S(\theta,m) - B(a/q,q)F_S(\eta,m)/q = O(m\sqrt{q\log q/N}(1+|\eta|m^2))$, whence the final estimate $F_S(\theta) = B(a/q,q)F_S(\eta)/q + O(\sqrt{q\log q}(1+|\eta|N))$. □

For small values of η, we can give a fairly good estimate for $F_S(\eta)$. Note that without weighting the exponential sum, we would have a bound of $\sigma^{-1}|h|^{-1/2}$ here, which is not good enough for the purposes of this paper.

Lemma A.2. *Let $\frac{1}{10} < h = \eta N \leq H = N^{1/8}$. Then we obtain the estimate*

$$|F_S(\eta)| \ll \frac{\sigma^{-1}}{|h|}.$$

Proof. Let us split the range of summation for F_S into intervals

$$R_{ij} = \{x : x^2 \in [N(i+j/H)/h, N(i+(j+1)/H)/h)\}.$$

Now break up the sum

$$F_S(h/N) = \sum_{i=1}^{\lfloor h/2\rfloor -1}\sum_{j=0}^{H}\sum_{x\in R_{ij}} 2xe(x^2h/N)/\sigma + O(\sigma^{-1}/|h|).$$

On R_{ij}, x^2h/N is equal to an integer plus a small remainder of at most H^{-1}, so the sum becomes

$$\sum_{i=1}^{h/2}\sum_{j=0}^{T} e(j/H)\sigma\sum_{x\in R_{ij}} 2x + \sum_{x^2\leq N_1} 2x/(H\sigma^{-1}).$$

It is easily shown that $\sum_{x\in R_{ij}} 2x = N/(Hh) + O(\sigma^{-1})$, and hence the sum is bounded by $O(hH + \sigma^{-1}/H)$ □

We next describe the behaviour of the weighted squares on what we called the major arcs.

Lemma A.3. *For $t \in I(a/q, (qQ)^{-1})$ with $q \leq R$, we have the* major arcs *estimate*

$$|F_S(t/N)| \ll \frac{\sigma^{-1}}{\sqrt[3]{q}}.$$

Proof. If $q \ll K$, then $h > 1/10$ and putting together the previous two lemmas yields $|F_S(t/N)| = \sqrt{\log q/q}/\sigma|h| + O(\sqrt{q \log q}N/(qQ))$. The first term clearly dominates and thus, if $q \leq R$, we have $F_S(t/N) \ll \sigma^{-1}q^{-1/3}$. □

We also need to investigate the behaviour on the minor arcs in more detail, which is done in the following lemma.

Lemma A.4. *For $t \in I(a/q, (qQ)^{-1})$ with $R < q \leq Q$, we have the* minor arcs *estimate*

$$|F_S(t/N)| \ll \frac{\sigma^{-1}}{\sqrt{K/L}}.$$

Proof. If q ranges between R and $N^{1/8}$, the result follows from the methods used above. For very large q, that is for $q > N^{1/8}$, it follows from Weyl's Inequality that $|F_S(t/N)| \ll \sqrt{N \log N}(q^{-1/2} + \sqrt{Q/N})$, which is clearly bounded above by $\sqrt{QL}$ provided that $q \gg K$. □

Finally, we need the following variant of Hua's Lemma, which is a classical ingredient in the solution of Waring's problem by Hardy and Littlewood.

Lemma A.5.

$$\sum_{t=1}^{N} |F_S(t/N)|^6 \ll \sigma^6.$$

We omit the proof but point out that the lemma corresponds to (a weighted version of) the well-known fact that the number of representations of an integer n as the sum of six squares is asymptotic to n^2.

Bibliography

[AS00] N. Alon and J. Spencer. *The probabilistic method*. Wiley, 2000.

[Beh46] F.A. Behrend. On sets of integers which contain no three elements in arithmetic progression. *Proc. Nat. Acad. Sci.*, 23:331–332, 1946.

[Ber41] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941.

[Ber45] H. Bergstrom. On the central limit theorem in the case $\mathbb{R}^k$, $k > 1$. *Skand. Aktuarietidskr.*, 2(3):106–127, 1945.

[BL96] V. Bergelson and A. Leibman. Polynomial extensions of Van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9:725–753, 1996.

[Bou06] J. Bourgain. Roth's theorem on progressions revisited. Preprint, 2006.

[BPPS94] A. Balog, J. Pelikán, J. Pintz, and E. Szemerédi. Difference sets without κth powers. *Acta Math. Hungar.*, 65(2):165–187, 1994.

[CCS05] P. Cameron, J. Cilleruelo, and O. Serra. On monochromatic solutions of equations in groups. Preprint, 2005.

[CFS82] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.

[CG90] F. R. K. Chung and R. L. Graham. Quasi-random hypergraphs. *Random Structures Algorithms*, 1(1):105–124, 1990.

[CL84] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112:143–175, 1984.

[CL87] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Publications de l'Institut de Recherche de Mathématiques de Rennes, Probabilités*, 1987.

[CL88] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris, Série I*, 306:491–493, 1988.

[Dat03] B.A. Datskovsky. On the number of monochromatic Schur triples. *Advances in Applied Mathematics*, 31:193–198, 2003.

[DT97] M. Drmota and R.F. Tichy. *Sequences, Discrepancies and Applications.* Springer, 1997.

[Ess45] C.-G. Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.*, 77:1–125, 1945.

[Fur77] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[FW96] H. Furstenberg and B. Weiss. A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, volume 5 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 193–227. de Gruyter, Berlin, 1996.

[Gir79] G. Giraud. Sur le problème de Goodman pour les quadrangles et la majoration des nombres de Ramsey. *J. Combin. Theory Ser. B*, 27(3):237–253, 1979.

[Goo59] A.W. Goodman. On sets of acquaintances and strangers at any party. *Amer. Math. Monthly*, 66:778–783, 1959.

[Gow98] W.T. Gowers. A new proof of Szemerédi's theorem for progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[Gow01] W.T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11:465–588, 2001.

[Gow06a] W. T. Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combin. Probab. Comput.*, 15(1-2):143–184, 2006.

[Gow06b] W.T. Gowers. Two examples in additive combinatorics. Unpublished, 2006.

[GR05] B.J. Green and I. Ruzsa. Sum-free sets in abelian groups. *Israel J. Math.*, 147:157–189, 2005.

[Gre02] B.J. Green. On arithmetic structure in dense sets of integers. *Duke Math. Journal*, 114:215–238, 2002.

[Gre03] B.J. Green. Some constructions in the inverse spectral theory of cyclic groups. *Combin. Probab. Comput.*, 12(2):127–138, 2003.

[Gre05a] B.J. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.

[Gre05b] B.J. Green. Montreal lecture notes on quadratic Fourier analysis. Available at arXiv:math.CA/0604089, 2005.

[GT04] B.J. Green and T. Tao. There are arbitrarily long arithmetic progressions in the primes. Available at arXiv:math.NT/0404188, 2004.

[GT05a] B.J. Green and T. Tao. An inverse theorem for the Gowers U^3-norm. To appear. Available at arXiv:math.NT/0503014, 2005.

[GT05b] B.J. Green and T. Tao. New bounds for Szemerédi's theorem, I: Progressions of length 4 in finite field geometries. Submitted. Available at arXiv:math.CO/0509560, 2005.

[GT06a] B.J. Green and T. Tao. Linear equations in primes. Submitted. Available at arXiv:math.NT/0606088, 2006.

[GT06b] B.J. Green and T. Tao. New bounds for Szemerédi's theorem, II: A new bound for $r_4(N)$. Submitted. Available at arXiv:math.NT/0610604, 2006.

[GT06c] B.J. Green and T. Tao. Quadratic uniformity of the Möbius function. Available at arXiv:math.NT/0606087, 2006.

[GW07a] W.T. Gowers and J. Wolf. Decompositions into polynomial phase functions. In preparation, 2007.

[GW07b] W.T. Gowers and J. Wolf. The true complexity of a system of linear equations. Available at arXiv:math.NT/0711.0185, 2007.

[Har66] L.H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combinatorial Theory*, 1:385–393, 1966.

[HB87] D.R. Heath-Brown. Integer sets containing no arithmetic progressions. *J. London Math. Soc. (2)*, 35:385–394, 1987.

[HK01] B. Host and B. Kra. Convergence of Conze-Lesigne averages. *Erg. Th. & Dyn. Sys.*, 21:493–509, 2001.

[HK04] B. Host and B. Kra. Averaging along cubes. In *Dynamical systems and related topics*. Cambridge University Press, Cambridge, 2004.

[HK05] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.*, 161(1):397–488, 2005.

[JŁR00] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[JŠT96] C. Jagger, P. Štovíček, and A. Thomason. Multiplicities of subgraphs. *Combinatorica*, 16(1):123–141, 1996.

[Kra06] B. Kra. Ergodic methods in additive combinatorics. Available at arXiv:math.DS/0608105, 2006.

[Led01] M. Ledoux. *The concentration of measure phenomenon*. AMS Mathematical Surveys and Monographs, 2001.

[Lei04] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. Available at http://www.math.ohio-state.edu/∼Leibman/preprints, 2004.

[Lei07] A. Leibman. Orbit of the diagonal of a power of a nilmanifold. Available at http://www.math.ohio-state.edu/∼Leibman/preprints, 2007.

[LŁS01] V. Lev, T. Łuczak, and T. Schoen. Sum-free sets in abelian groups. *Israel J. Math.*, 125:347–367, 2001.

[LOS82] J.C. Lagarias, A.M. Odlyzko, and J.B. Shearer. On the density of sequences of integers the sum of no two of which is a square. I. Arithmetic progressions. *J. Comb. Theory, Series A.*, 33:167–185, 1982.

[Luc07] J. Lucier. Difference sets and shifted primes. Preprint. Available at arxiv:math.NT/0705.3749, 2007.

[McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[Mes95] R. Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A*, 71(1):168–172, 1995.

[NP73] H. Niederreiter and W. Philipp. Berry-Esseen bounds and a theorem of Erdős and Turán on uniform distribution mod 1. *Duke Math. J.*, 40(3):633–649, 1973.

[NRS06] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[Pol46] G. Polya. Remarks on computing the probability integral in one and two dimensions. *Proc. Berkeley Symp.*, pages 63–78, 1946.

[PSS88] J. Pintz, W.L. Steiger, and E. Szemerédi. On sets of natural numbers whose difference set contains no squares. *J. London Math. Soc. (2)*, 37:219–231, 1988.

[Rot53] K.F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[RS04] V. Rödl and J. Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.

[RS07] I.Z. Ruzsa and T. Sanders. Difference sets and the primes. Preprint. Available at arxiv:math.NT/, 2007.

[Rud95] D.J. Rudolph. Eigenfunctions of T×S and the Conze-Lesigne algebra. In *Ergodic theory and its Connections with Harmonic Analysis*, pages 369–432. Cambridge Univ. Press, New York, 1995.

[Ruz84] I.Z. Ruzsa. Difference sets without squares. *Periodica Math. Hungar.*, 15:205–209, 1984.

[Ruz87] I.Z. Ruzsa. Essential components. *Proc. London Math. Soc. (3)*, 54(1):38–56, 1987.

[Ruz91] I.Z. Ruzsa. Arithmetic progressions in sumsets. *Acta Arith.*, 2:191–202, 1991.

[Sad66] S.M. Sadikova. Two-dimensional analogues of an inequality of Esseen with applications to the Central Limit Theorem. *Theory of Probability and Its Applications*, 11:325–335, 1966.

[Sár78a] A. Sárközy. On difference sets of sequences of integers I. *Acta Math. Acad. Sci. Hungar.*, 31:125–149, 1978.

[Sár78b] A. Sárközy. On difference sets of sequences of integers II. *Ann. Univ. Sci. Budapest*, 21:45–53, 1978.

[Sár78c] A. Sárközy. On difference sets of sequences of integers III. *Acta Math. Acad. Sci. Hungar.*, 31:355–386, 1978.

[Shi84] A.N. Shiryayev. *Probability.* Springer, 1984.

[Sze75] E. Szemerédi. On integer sets containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[Sze90] E. Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Hungar.*, 56(1-2):155–158, 1990.

[Tao05] T. Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. Available at arXiv:math.NT/0512114, 2005.

[Tao06] T. Tao. A quantitative ergodic theory proof of Szemerédi's theorem. Available at arXiv:math.CO/0405251, 2006.

[Tho89] A. Thomason. A disproof of a conjecture of Erdős in Ramsey theory. *J. London Math. Soc. (2)*, 39(2):246–255, 1989.

[Tho97] A. Thomason. Graph products and monochromatic multiplicities. *Combinatorica*, 17(1):125–134, 1997.

[TZ06] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. Available at arXiv:math/0610050, 2006.

[Vau81] R.C. Vaughan. *The Hardy-Littlewood Method*. Cambridge University Press, 1981.

[Wol03] J. Wolf. Arithmetic structure in difference sets. Part III Essay. University of Cambridge, 2003.

[Wol05] J. Wolf. The structure of popular difference sets. Preprint, 2005.

[Wol07] J. Wolf. The minimum number of monochromatic 4-term progressions. Preprint, 2007.

[Zie07] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20:53–97, 2007.