L'analyse harmonique d'ordre supérieur, et applications

mémoire présenté pour l'obtention de l'habilitation à diriger des recherches par

Julia Wolf

à l'Université Paris-Sud (Orsay)

soutenu le 3 décembre devant un jury composé de

Emmanuel BREUILLARD	Professeur, Université Paris-Sud Orsay
Étienne FOUVRY	Professeur, Université Paris-Sud Orsay
Ben GREEN	Professeur, University of Cambridge (rapporteur)
Harald HELFGOTT	Chargé de Recherche, CNRS/École normale supérieure
Bernard HOST	Professeur, Université Paris-Est Marne-la-Vallée (rapporteur)
Alain PLAGNE	Chercheur, École polytechnique

Contents

Résumé **1**

Summary **7**

1 Introduction **13**

1.1 The discrete Fourier transform. 13

1.2 The need for higher-order Fourier analysis. 14

1.3 The uniformity norms. 15

1.4 Discussion and outlook. 17

2 Advances in discrete harmonic analysis **19**

2.1 Inverse theorems. 19

2.2 Decomposition theorems. 23

2.3 Discussion and outlook. 31

3 Applications in number theory **33**

3.1 The "true" complexity of a system of linear equations. 33

3.2 Further applications of higher-order decompositions. 36

3.3 Polynomial configurations in the primes. 38

3.4 Discussion and outlook. 42

4 Interactions with theoretical computer science **45**

4.1 A quadratic Goldreich-Levin theorem. 46

4.2 Algorithmic versions of almost-periodicity results. 51

4.3 Discussion and outlook. 55

5 Related combinatorial results **57**

5.1 An improvement of Behrend's construction. 58

5.2 A finite field Behrend-type construction for longer progressions. 60

5.3 Discussion and outlook. 62

Remerciements **65**

Références **67**

Résumé

Mes travaux se situent dans le domaine de la combinatoire arithmétique, à l'interface entre la théorie analytique des nombres, l'analyse harmonique et la combinatoire. En particulier, je travaille sur des questions fondamentales concernant ce qu'on appelle l'analyse harmonique d'ordre supérieur, dont les applications les plus célèbres sont le théorème de Szemerédi et le théorème de Green et Tao sur les progressions arithmétiques dans l'ensemble des nombres premiers. Je m'intéresse aussi aux interactions avec la théorie ergodique ainsi qu'aux nombreuses applications en informatique théorique.
Mes contributions et ambitions dans ce domaine sont les suivantes :

- le développement des notions fondamentales et d'outils fortement quantitatifs en analyse harmonique discrète, quadratique et d'ordre supérieur [6, 7, 8] ;
- les applications aux questions concernant les structures additives et polynomiales dans des sous-ensembles d'entiers et dans l'ensemble des nombres premiers [2, 12] ;
- l'analyse combinatoire des ensembles qui ne contiennent pas de telles structures [1, 4, 5] ;
- le renforcement des liens entre la combinatoire arithmétique et la théorie ergodique, la théorie des graphes et l'informatique théorique [3, 9, 11].

Les articles [1, 2, 3] découlent de ma thèse de doctorat. J'ai aussi rédigé un article de survol [10] et la version préliminaire d'un livre [13].

Introduction. L'analyse harmonique d'ordre supérieur est issue des travaux de Gowers [41] sur le théorème de Szemerédi [84], qui affirme que tout sous-ensemble d'entiers suffisamment dense contient une progression arithmétique de longueur k. Ce domaine a vu au moins une autre application révolutionnaire en théorie des nombres, dans les travaux de Green et Tao sur les progressions arithmétiques dans l'ensemble des nombres premiers. Il existe une théorie correspondante en théorie ergodique et en théorie des graphes, ainsi que des applications importantes en informatique théorique.

Gowers a (re)découvert que l'analyse de Fourier classique est insuffisante pour compter des progressions arithmétiques de longueur 4 dans des sous-ensembles denses d'un groupe abélien fini G (parmi les exemples courants on trouve $G = \mathbb{Z}_p$ et $G = \mathbb{F}_p^n$). Plus spécifiquement, il existe un ensemble $A \subseteq G$ dont les coefficients de Fourier sont très petits, mais qui ne contient pas le nombre "attendu" de progressions arithmétiques de longueur 4. Pour contourner ce problème, Gowers a introduit la *norme* U^3 sur l'espace de fonctions $f : G \to \mathbb{C}$, définie

par

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

où $\mathcal{C}^{|\epsilon|} f = f$ si $|\epsilon| = \sum_i \epsilon_i$ est pair et $\mathcal{C}^{|\epsilon|} f = \overline{f}$ sinon. Cette formule n'a aucune interprétation utile en termes de la transformée de Fourier de f, mais elle majore la moyenne des progressions arithmétiques de longueur 4 dans le sens où

$$|\mathbb{E}_{x,d} f(x) f(x+d) f(x+2d) f(x+3d)| \leq \|f\|_{U^3}.$$

La formule ci-dessus pour la norme U^3 se généralise facilement, et il n'est pas difficile de voir que la norme U^k contrôle le nombre de progressions arithmétiques de longueur $k+1$.

La preuve du théorème de Szemerédi procède alors selon la dichotomie suivante : soit la norme U^k de la fonction centrée $f_A = 1_A - \alpha$ d'un sous-ensemble $A \subseteq G$ de densité α est petite, auquel cas nous allons dire que A est *uniforme de degré* $k-1$, et il se trouve que A contient toujours le nombre attendu de progressions arithmétiques de longueur $(k+1)$, soit la norme U^k de f_A est grande. Dans ce dernier cas on arrive à déduire, avec beaucoup d'effort et à l'aide de plusieurs théorèmes profonds de la combinatoire additive, que la fonction f_A possède une structure polynomiale (faible) de degré $k-1$. On appelle un tel énoncé un *théorème inverse* pour la norme U^k, et il est suffisant pour conclure la preuve du théorème de Szemerédi à l'aide d'un argument par récurrence.

Théorèmes de décomposition. En analyse de Fourier classique on décompose souvent une fonction en séparant les grands coefficients de Fourier des plus petits, captant ainsi la dichotomie entre partie structurée et partie à caractère aléatoire. Peut-on donner une telle décomposition pour la dichotomie quadratique décrite ci-dessus ? Autrement dit, peut-on écrire une fonction bornée $f : G \to \mathbb{C}$ comme une somme d'une partie à structure quadratique, et une partie qui est quadratiquement uniforme (petite dans la norme U^3) ?

Puisqu'il n'y a pas de base orthonormée pour l'espace des phases quadratiques, toute décomposition quadratique est en soi non-triviale. En collaboration avec Tim Gowers, nous avons développé des décompositions quadratiques fortement quantitatives, dans le groupe modèle $G = \mathbb{F}_p^n$ [6] ainsi que dans le groupe $G = \mathbb{Z}/N\mathbb{Z}$ [7] qui présente des grandes difficultés techniques. Par exemple, nous avons montré, en utilisant le théorème de Hahn-Banach en analyse fonctionnelle, que pour tout $\delta > 0$, il existe $M(\delta)$ tel que toute fonction $f : \mathbb{F}_p^n \to \mathbb{C}$

vérifiant $\|f\|_2 \leq 1$ puisse s'écrire sous la forme

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

où les q_i sont des formes quadratiques définies sur $\mathbb{F}_p^n$, $\|g\|_{U^3} \leq \delta$, $\|h\|_1 \leq \delta$ et $\sum_i |\lambda_i| \leq M(\delta)$. Mis à part le théorème de Szemerédi sur les progressions arithmétiques de longueur 4, de telles décompositions quadratiques ont plusieurs applications en théorie des nombres, dont nous allons donner deux exemples par la suite.

Nous avons aussi obtenu, en utilisant un théorème inverse très récent pour les normes U^k dans $\mathbb{F}_p^n$ démontré par Bergelson, Tao and Ziegler [20, 87], le théorème de décomposition le plus général dans ce contexte, qui nous permet d'écrire toute fonction $f : \mathbb{F}_p^n \to [-1, 1]$ sous la forme

$$f = \sum_i \lambda_i \omega^{\pi_i} + g + h,$$

où $\|g\|_{U^{k+1}}$ et $\|h\|_1$ sont petites, les π_i sont des polynômes de degré $i \leq k$ et $\sum_i |\lambda_i|$ est une somme bornée. Pour nos applications, nous avons été obligés de démontrer plusieurs raffinements de ces théorèmes de décomposition, en rassemblant les phases polynomiales qui sont proches les unes des autres (mesurées par le rang de leur différence), et en éliminant les phases dont le rang est petit [8].

Applications en théorie des nombres. Notre motivation principale pour les théorèmes de décomposition décrits ci-dessus, et leurs raffinements fortement quantitatifs, représente le sujet commun de la série d'articles [2, 6, 7, 8].

Ayant démontré l'existence de progressions arithmétiques de toutes longueurs dans l'ensemble des nombres premiers, Green et Tao ont étendu leurs arguments au cas des systèmes d'équations linéaires quelconques [51] (à l'exception de ceux qui définissent les premiers jumeaux, et les configurations dans la conjecture de Goldbach). Dans un premier temps, ils étaient obligés de majorer la moyenne sur un système de formes linéaires $L_1, \ldots, L_m$ en d variables par une norme d'uniformité U^k. Plus spécifiquement, Green et Tao ont démontré, en utilisant l'inégalité de Cauchy-Schwarz et un changement de variables approprié, que

$$|\mathbb{E}_{x_1,\ldots,x_d} \prod_{i=1}^m f(L_i(x_1, \ldots, x_d))| \leq \|f\|_{U^{k+1}}$$

pour toute fonction bornée f, pourvu que le système linéaire soit de "complexité" au plus k. Cette notion de complexité était définie par des rélations linéaires entre les formes linéaires,

et elle était suffisante pour Green et Tao. Néanmoins, étant donné que l'analyse de Fourier d'ordre supérieur devient bien plus difficile, et bien moins développée lorsque k augmente, il est souhaitable de déterminer la *plus petite* valeur de k pour laquelle la norme U^k majore la moyenne sur un système linéaire donné – la *vraie complexité*.

Ceci était le point de départ de mes recherches avec Tim Gowers. Chose étonnante, nous avons découvert qu'il existe des systèmes linéaires de complexité 2 (d'après Green et Tao), mais qui étaient en réalité contrôlés par la norme U^2 (et non U^3). Un exemple d'un tel système linéaire est donné par les formes linéaires $x, y, z, x+y+z, x+2y-z, x+2z-y$. Plus généralement, nous avons réussi à classifier les systèmes linéaires pour lesquels l'analyse de Fourier classique est suffisante : un système linéaire est contrôlé par l'analyse de Fourier classique si et seulement si les carrés des formes linéaires sont linéairement indépendants. En général, un système linéaire $L_1, \ldots, L_m$ est contrôlé par la norme U^k si et seulement si les puissances d'ordre k des formes linéaires L_i^k sont linéairement indépendantes. Dans une direction, ce résultat est une généralisation directe du contre-exemple mentionné ci-dessus (qui nous empêche d'utiliser l'analyse de Fourier classique pour les progressions arithmétiques de longueur 4). L'implication inverse dépend de manière décisive des théorèmes de décomposition ci-dessus.

Une autre application de notre théorème de décomposition quadratique a été donnée par Candela [28], qui a montré que l'ensemble des raisons des progressions de longueur 3 contenues dans un sous-ensemble dense d'entiers contient lui-même une très longue progression arithmétique. D'ailleurs, la méthode de Hahn-Banach a été employée par Gowers [43] pour donner une nouvelle preuve du *principe de transfert*, outil fondamental dans la preuve du théorème de Green et Tao.

Le théorème de Green et Tao sur les progressions arithmétiques dans l'ensemble des nombres premiers a été étendu au cas des progressions polynomiales par Tao et Ziegler [86]. En collaboration avec Thái Hoàng Lê [12] nous avons formulé un résultat hybride entre le théorème de Tao et Ziegler et les résultats récents de Wooley et Ziegler [93] et Frantzikinakis, Host et Kra [37], en concluant que tout sous-ensemble suffisamment dense de l'ensemble des nombres premiers contient une progression polynomiale sous la forme $x + P_1(m), x + P_2(m), \ldots, x + P_k(m)$, où la variable m est un nombre premier moins 1.

Interactions avec l'informatique théorique. La théorie de l'uniformité d'ordre supérieur est aussi d'un grand intérêt en informatique théorique : les normes U^k ont été utilisées dans le contexte des preuves vérifiables de manière probabiliste [75], de la complexité de communication [92] ainsi que l'analyse des générateurs pseudo-aléatoires [25]. D'ailleurs, le théorème

inverse pour la norme U^3 a été démontré de façon indépendante dans $\mathbb{F}_2^n$ par Samorodnitsky [74], avec applications aux testeurs des propriétés.

En collaboration avec Madhur Tulsiani [9] nous avons conçu un algorithme probabiliste qui calcule, en temps polynomial (en n), les grands "coefficients de Fourier quadratiques", associés aux théorèmes de décomposition pour la norme U^3. Dans le cas de l'analyse de Fourier classique, cela correspond au célèbre algorithme de Goldreich-Levin [40]. Contrairement aux autres décompositions quadratiques, dont l'existence ne peut être démontrée que d'une manière abstraite, cette décomposition est explicite et constructive, en utilisant une technique provenant de l'apprentissage automatique qui s'appelle "boosting". Pour trouver la phase quadratique corrélante il est nécessaire d'établir des versions algorithmiques de plusieurs résultats de combinatoire additive. Ceci n'est pas élémentaire car les ensembles intervenants sont denses dans $\mathbb{F}_2^n$, ce qui nous empêche (si on cherche un algorithme en temps polynomial en n) même de mémoriser tous leurs éléments, et a fortiori de les manipuler, au moins de manière déterministe.

Dans un travail encore plus récent, en commun avec Eli Ben-Sasson, Noga Ron-Zewi and Madhur Tulsiani [11], nous avons donné une nouvelle preuve d'un théorème de presque-périodicité pour les ensembles somme, d'après le *lemme de Croot et Sisask* [32] qui a vu plusieurs applications remarquables ces deux dernières années. Nous éliminons les normes L^p de cette preuve, nous appuyant plutôt sur des méthodes d'echantillonage qui peuvent se convertir facilement en algorithme probabiliste, afin de déterminer si oui ou non un élément donné appartient à un sous-ensemble presque-périodique de $\mathbb{F}_2^n$. Comme application nous donnons une interprétation très simple (ainsi qu'une version algorithmique) du *lemme fort de Bogolyubov-Ruzsa*, démontré récemment par Sanders [76]. Ceci entraîne une amélioration dans le *théorème quadratique de Goldreich-Levin* mentionné ci-dessus, où le nombre de termes dans la décomposition de Fourier quadratique dépend du paramètre d'uniformité ϵ de façon quasipolynomiale (et non exponentielle). Le temps de calcul est amélioré de même.

Résultats combinatoires apparentés. Les méthodes décrites ci-dessus ont été conçues pour donner des bornes supérieures pour le cardinal d'un sous-ensemble d'entiers qui ne contient pas une structure arithmétique donnée. J'ai aussi travaillé sur les bornes inférieures, en fournissant des constructions explicites de tels ensembles.

La meilleure borne inférieure pour le cardinal maximal d'un ensemble $A \subseteq \{1, 2, \ldots, N\}$ qui ne contient pas de progression arithmétique de longueur 3 est dû à Behrend [18]. N'ayant pas été surpassée pendant plus de 60 ans, elle a été améliorée trés légèrement par Elkin [35]. Dans un travail en commun avec Ben Green [4], nous avons donné une démonstration

différente (et très courte) de ce résultat par un argument probabiliste, qui peut être généralisée au cas des progressions de longueur supérieure à 3 [67].

La construction originale de Behrend avait été généralisée au cas de progressions de longueur supérieure par Rankin [70], et des constructions comparables évitant des progressions de longueur 3 existent dans le cadre de $\mathbb{F}_q^n$ depuis les travaux de Edel [34]. Avec mon stagiaire de recherche Yuncheng Lin au MIT, nous avons donné une construction algébrique d'un sous-ensemble de $\mathbb{F}_q^n$ qui ne contient pas de progressions de longueur k pour $k > 3$, qui reste la meilleure jusqu'à ce jour [5].

Plan du mémoire. Le chapitre 1 sert d'introduction et on y présente les définitions et résultats de base. Le chapitre 2 traite de mes travaux concernant les fondements de l'analyse de Fourier d'ordre supérieur, plus spécifiquement les théorèmes inverses et les théorèmes de décomposition que l'on a développés en vue des applications en théorie des nombres, qui seront présentées au chapitre 3. Le chapitre 4 porte sur mes travaux récents à l'interface avec l'informatique théorique, qui sont en complémentarité avec les approches plus abstraites du chapitre 2. Au chapitre 5, on exposera les bornes inférieures pour le théorème de Szemerédi obtenues par des méthodes combinatoires. Chaque chapitre est conclu par une discussion concise des questions ouvertes.

Summary

My research focuses on arithmetic structures in dense sets of integers and combines Fourier analytic, combinatorial and probabilistic methods. It is part of a thriving area called arithmetic combinatorics, a subject that includes many beautiful results such as Szemerédi's theorem on long arithmetic progressions in dense subsets of the integers and the Green-Tao Theorem on long arithmetic progressions in the primes. Some of my work has close connections with parts of ergodic theory, and more recently I have become interested in applications of higher-order Fourier analysis to theoretical computer science.
My contributions to and ambitions in this field can be broadly classified as follows.

- to develop state-of-the-art tools in quadratic and higher-degree Fourier analysis on finite abelian groups [6, 7, 8];
- to apply these tools to the study of linear and polynomial structure in subsets of the integers and the primes [2, 12];
- to work towards a better combinatorial understanding of sets that lack such structure [1, 4, 5];
- and to exploit and strengthen the connections with ergodic theory, graph theory and theoretical computer science [3, 9, 11].

The papers [1, 2, 3] formed part of my doctoral thesis. I have also written an expository article [10] and a draft for a book [13].

Introduction. Higher-order Fourier analysis has its origins in Gowers's work [41] on Szemerédi's theorem [84], which states that any sufficiently dense subset of the integers contains a k-term arithmetic progression. It has seen at least one further groundbreaking number-theoretic application in the work of Green and Tao on long arithmetic progressions in the primes, and has parallels in ergodic and graph theory as well as important applications in theoretical computer science.

Gowers rediscovered that classical harmonic analysis is insufficient for counting 4-term progressions in dense subsets of a finite abelian group G (useful example of G are $\mathbb{Z}/N\mathbb{Z}$ or $\mathbb{F}_p^n$). Specifically, there is a set $A \subseteq G$ whose Fourier coefficients are very small, but which does not contain the "expected" number of 4-term progressions. To get around this problem, Gowers introduced the U^3 norm on the space of functions $f : G \to \mathbb{C}$, by defining

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

where $\mathcal{C}^{|\epsilon|}f = f$ if $|\epsilon| = \sum_i \epsilon_i$ is even and $\mathcal{C}^{|\epsilon|}f = \overline{f}$ otherwise. This formula has no useful interpretation in terms of the Fourier transform, but it does control the 4-term progression count in the sense that

$$|\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)f(x+3d)| \leq \|f\|_{U^3}.$$

The above formula for the U^3 norm can easily be generalized, and it is not hard to show that the U^k norm thus defined controls the count of $(k+1)$-term arithmetic progressions. The proof of Szemerédi's theorem then proceeds via the following dichotomy: either the U^k norm of the balanced function $f_A = 1_A - \alpha$ of a subset $A \subseteq G$ of density α is small, in which case A is called *uniform of degree* $k-1$ and turns out to contain the expected number of $(k+1)$-term progressions, or the U^k norm of f_A is large. In the latter case, one can, with much effort and recourse to deep theorems from additive combinatorics, infer that f_A has weak polynomial structure of degree $k-1$. Such a statement is known as an *inverse theorem* for the U^k norm, and suffices to complete the proof of Szemerédi's theorem via an iteration argument.

Decomposition theorems. In classical Fourier analysis one often decomposes a function by separating large and small Fourier coefficients in order to efficiently capture a dichotomy between structure and randomness. Can we achieve a similar decomposition for the quadratic dichotomy just described, that is, can we write a bounded function $f : G \to \mathbb{C}$ as the sum of a quadratically structured and a quadratically uniform part (which is small in U^3)?

Since there is no canonical orthonormal basis for the space of quadratic phases, any such decomposition is a highly non-trivial statement, even with knowledge of the above-mentioned inverse theorem. In collaboration with Tim Gowers, I developed the strongest available decomposition theorems for the U^3 norm, both in the setting of $\mathbb{F}_p^n$ [6] and the much more demanding context of $\mathbb{Z}/N\mathbb{Z}$ [7]. For example, we showed, using the Hahn-Banach theorem from functional analysis, that given any $\delta > 0$, there exists $M(\delta)$ such that any function $f : \mathbb{F}_p^n \to \mathbb{C}$ that satisfies $\|f\|_2 \leq 1$ has a decomposition of the form

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

where the q_i are quadratic forms on $\mathbb{F}_p^n$, $\|g\|_{U^3} \leq \delta$, $\|h\|_1 \leq \delta$ and $\sum_i |\lambda_i| \leq M(\delta)$. Apart from proving Szemerédi's theorem for progressions of length 4, such strong quantitative decomposition theorems have a number of other applications in number theory which I shall briefly describe below.

We also used a recent inverse theorem for the U^k norm in $\mathbb{F}_p^n$ by Bergelson, Tao and Ziegler [20, 87] to prove the most general decomposition theorem for bounded functions in that context, which allows us to write $f : \mathbb{F}_p^n \to [-1, 1]$ as

$$f = \sum_i \lambda_i \omega^{\pi_i} + g + h,$$

where $\|g\|_{U^{k+1}}$ is small, h is small in L^1, the π_i are polynomials of degree $i \leq k$ and $\sum_i |\lambda_i|$ is bounded. For our applications, we were forced to develop several refinements of this decomposition, by clustering polynomial phases of approximately equal rank together, and eliminating low-rank phases from the decomposition altogether [8].

Applications in number theory. The main motivation for proving the above-mentioned strong quantitative decomposition theorems constitutes the common theme of the series of papers [2, 6, 8, 7].

After proving that there are arbitrarily long arithmetic progressions in the primes, Green and Tao dealt with arbitrary systems of linear equations in the primes [51] (with the exception of the famous twin primes and those constellations related to Goldbach's problem). A first step was to control a system of linear forms $L_1, \ldots, L_m$ in d variables by an appropriate uniformity norm. In particular, Green and Tao proved, using the Cauchy-Schwarz inequality and a suitable reparameterization of the linear system, that

$$|\mathbb{E}_{x_1,\ldots,x_d} \prod_{i=1}^{m} f(L_i(x_1, \ldots, x_d))| \leq \|f\|_{U^{k+1}}$$

for any bounded function f, provided that the linear system has "complexity" at most k. This complexity was defined in terms of the linear dependence between the linear forms, and sufficient for their purpose. However, given that higher-order Fourier analysis becomes significantly more difficult and less well developed as k increases, it seemed desirable to determine the *least* k such that the U^k norm controls the average over the linear system – in other words, the *true complexity*.

This was the starting point of my joint investigations with Tim Gowers. Rather surprisingly, we discovered that there were linear systems that had complexity 2 according to the measure of Green and Tao, but that were in reality controlled by the U^2 (and not the U^3) norm. An example of such a system is given by the linear forms $x, y, z, x+y+z, x+2y-z, x+2z-y$. More generally, we were able to precisely classify those linear systems for which ordinary Fourier analysis suffices: a linear system is governed by ordinary Fourier analysis if and only if the

squares of the linear forms are linearly independent. In general, a linear system $L_1, \ldots, L_m$ is controlled by the U^k norm if and only if the kth powers L_i^k of the linear forms are linearly independent. One direction of this result is a direct generalization of the counterexample mentioned above to using the Fourier transform for counting 4-term progressions, while the implication in the reverse direction makes substantial use of the above-mentioned decomposition theorems.

Another application of our quadratic decomposition theorem was given by Candela [28], who showed that the set of common differences of the 3-term progressions contained in a dense set contains a long arithmetic progression. The Hahn-Banach technique was also used by Gowers [43] to give a new proof of the transference principle, which was instrumental in the proof of the Green-Tao theorem.

The Green-Tao theorem on arithmetic progression in the primes was extended by Tao and Ziegler [86] to the case of polynomial common differences (even though no quantitative such theorem exists in dense subsets of the integers). Together with Thái Hoàng Lê [12] we were able to formulate a hybrid of the Tao-Ziegler theorem and recent results of Wooley and Ziegler [93] and Frantzikinakis, Host and Kra [37], and conclude that any sufficiently dense subset of the primes contains a polynomial progression $x + P_1(m), x + P_2(m), \ldots, x + P_k(m)$, where the variable m itself is a prime minus 1.

Interactions with theoretical computer science. The theory of higher-degree uniformity is also of great interest to computer scientists: the U^k norms have been used in the context of probabilistically checkable proofs [75], communication complexity [92] as well as in the analysis of pseudo-random generators [25]. Moreover, the inverse theorem for the U^3 norm was independently proved by Samorodnitsky in $\mathbb{F}_2^n$ [74], with implications for property testing.

In collaboration with Madhur Tulsiani [9] I gave a probabilistic polynomial-time algorithm that computes the "large quadratic Fourier coefficients" associated with a decomposition theorem for the U^3 norm. In the case of classical Fourier analysis this is accomplished by the celebrated Goldreich-Levin algorithm [40]. Unlike all other known quadratic decompositions, whose existence one can prove abstractly either by the Hahn-Banach theorem or an energy increment argument, ours is constructive, using a technique from machine learning called "boosting". Finding the correlating quadratic phase involves giving algorithmic versions of several results from additive combinatorics. This is not a simple matter since the sets under consideration are usually dense in $\mathbb{F}_2^n$, which, if we are looking to obtain a polynomial-time algorithm (in n), does not even allow us to store the elements of the set, let alone perform operations on them.

In recent joint work with Eli Ben-Sasson, Noga Ron-Zewi and Madhur Tulsiani [11], we give new combinatorial proofs of known almost-periodicity results for sumsets of sets with small doubling in the spirit of Croot and Sisask [32], whose lemma has had far-reaching implications in additive combinatorics. We provide an an alternative (and L^p-norm free) point of view, which allows for proofs to easily be converted to probabilistic algorithms that decide membership in almost-periodic subsets of dense subsets of $\mathbb{F}_2^n$.

As an application, we give a simple interpretation of the proof and a new algorithmic version of the *strong Bogolyubov-Ruzsa lemma* recently proved by Sanders [76]. This implies a version of the above-mentioned *quadratic Goldreich-Levin theorem* in which the number of terms in the quadratic Fourier decomposition of a function is quasipolynomial (as opposed to exponential) in the error parameter ϵ. It also improves the running time of the algorithm to have quasipolynomial dependence on ϵ.

Related combinatorial results. The methods described so far were conceived to provide upper bounds on the size of sets that lack a given arithmetic structure. I have also worked towards a better understanding of the corresponding lower bounds by giving explicit constructions of such sets.

The best known lower bound on the maximum size of a set $A \subseteq \{1, 2, \dots, N\}$ that contains no 3-term progressions is due to Behrend [18]. After having remained unsurpassed for over 60 years, it was improved by a very small factor by Elkin [35]. Ben Green and I [4] gave a different (and very short) proof of Elkin's result, using a neat probabilistic argument which turns out to generalize to longer progressions [67].

Behrend's construction was originally extended to longer progressions by Rankin [70], and analogues for progressions of length 3 in $\mathbb{F}_q^n$ are known through the work of Edel [34]. Together with my summer student Yuncheng Lin at MIT I completed the picture by exploring k-term progression-free sets in $\mathbb{F}_q^n$ for $k > 3$, thereby providing the first construction of large sets of this kind [5].

Outline. Chapter 1 is entirely introductory and gives the necessary background. Chapter 2 covers my work on the theoretical foundations of higher-order Fourier analysis, specifically the inverse and decomposition theorems we developed for the number-theoretic applications in Chapter 3. Chapter 4 focuses on my recent work at the interface with theoretical computer science, which complements the more abstract approach in Chapter 2. Chapter 5 deals with combinatorial lower bounds for Szemerédi's theorem. Each chapter ends with a brief discussion of open problems and further directions for research.

1 Introduction

1.1. The discrete Fourier transform. Harmonic analysis has been used to solve problems in number theory for more than a century. In order to attack a problem about the integers, we often regard the interval $\{1, 2, \dots, N\}$ as a subset of the finite abelian group $\mathbb{Z}/p\mathbb{Z}$, most conveniently with p a prime not too much larger than N. Many times we also consider the equivalent problem in the setting of $\mathbb{F}_p^n$, where p is to be thought of as small and fixed and n as tending to infinity. Unlike $\mathbb{Z}/p\mathbb{Z}$, the vector space $\mathbb{F}_p^n$ has plenty of algebraic substructure and is therefore much easier to deal with from a technical point of view, while retaining the principal characteristics of the problem. This so-called "finite fields model" approach was advocated by Green in the excellent survey [44].

In either case, $G = \mathbb{Z}/p\mathbb{Z}$ or $G = \mathbb{F}_p^n$, we define the *Fourier transform* of a function $f : G \to \mathbb{C}$ by setting, for a character $\gamma \in \widehat{G}$,

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

where $\mathbb{E}_{x \in G}$ simply denotes the normalized sum over elements in G. With this convention, the inversion formula states that $f(x) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\overline{\gamma(x)}$, and Parseval's identity asserts that $\|f\|_2 = \mathbb{E}_x |f(x)|^2 = \sum_\gamma |\widehat{f}(\gamma)|^2 = \|\widehat{f}\|_2$.

In order to understand why the Fourier transform helps us count certain arithmetic structures inside dense subsets of G, let us start by counting a rather simple pattern such as a 3-term arithmetic progression in G, which takes the form of a triple $x, x + d, x + 2d$ with $x, d \in G$. The proportion of 3-term progression inside a subset $A \subseteq G$ can be expressed as

$$\mathbb{E}_{x,d \in G} 1_A(x) 1_A(x + d) 1_A(x + 2d), \tag{1}$$

where 1_A denotes the indicator function of the set A. By expanding each indicator function in terms of its Fourier coefficients and exploiting the orthogonality of the characters, we obtain the expression

$$\sum_{\gamma \in \widehat{G}} \widehat{1_A}(\gamma)^2 \widehat{1_A}(-2\gamma).$$

It is easy to see that the Fourier coefficient at the trivial character is equal to the density α of the set A in G, so that we can split this sum as

$$\alpha^3 + \sum_{\gamma \neq \gamma_0} \widehat{1_A}(\gamma)^2 \widehat{1_A}(-2\gamma).$$

From this it follows by the triangle inequality and Parseval's identity that if all non-trivial Fourier coefficients of A are small (that is, tend to zero as $|G|$ goes to infinity) then A contains $\alpha^3 p^2$ 3-term progressions; in other words, A contains the number of 3-term progressions one would expect if the elements of A were chosen independently and uniformly from the elements of G with probability α. We shall call a set with small non-trivial Fourier coefficients *uniform*. (Since a randomly chosen set has small Fourier coefficients, one sometimes finds the term *quasirandom* in the literature.)

If our aim is to establish the existence of a 3-term progression inside the set A, we need to look no further if the set A is uniform, since we have just shown that in this case we have an ample supply of 3-term progressions. But what can we say when A is non-uniform? By definition this means that for some $\gamma \neq \gamma_0$, the Fourier coefficient $\widehat{f}(\gamma)$ is large. Looking back at how we defined the Fourier transform, this means that our set A (or rather its indicator function) correlates with a linear phase (namely the character $\gamma(x)$). In practice this means that we can show that A has increased density on some linear substructure of G, such as a subspace of $\mathbb{F}_p^n$ or a long arithmetic progression in $\mathbb{Z}/p\mathbb{Z}$.

What we have just described is a very simple case of a ubiquitous dichotomy between structure and randomness. It can be summarized by a *decomposition* of a function f into a uniform and a structured part, which in the case of the Fourier transform is displayed most naturally as

$$f(x) = \sum_{\gamma \in K} \widehat{f}(\gamma)\overline{\gamma(x)} + \sum_{\gamma \notin K} \widehat{f}(\gamma)\overline{\gamma(x)}, \tag{2}$$

where K is the set of characters at which the Fourier transform of f is large.

This approach, which is of course reminiscent of the circle method in analytic number theory, allows one to prove Roth's theorem (which states that any sufficiently dense subset of the integers $\{1, 2, \ldots, N\}$ contains a 3-term progression), and generalizes to any configuration that is defined by a single linear equation.

1.2. The need for higher-order Fourier analysis. It had been known in the ergodic theory community for some time when it was rediscovered by Gowers [41] in his work on Szemerédi's theorem (which states that any sufficiently dense subset of the integers $\{1, 2, \ldots, N\}$ contains a k-term progression) that the Fourier approach described in the preceding section is insufficient for counting more sophisticated arithmetic structures such as arithmetic progressions of length at least 4. It is not difficult to see that the expression for the 4-term progression count corresponding to (1) does not transform as nicely as it did for 3-term progressions above, but something stronger is true: there is, in fact, a counterexample to the Fourier transform

approach sketched above, meaning that there exists a set A that is uniform in the Fourier sense but contains the "wrong" number of 4-term progressions. In $\mathbb{F}_p^n$ such a set is particularly easy to describe:

$$A = \{x \in \mathbb{F}_p^n : x^T x = 0\}. \tag{3}$$

It is a simple exercise in conditional probability to show that this set contains too many 4-term progressions: the identity

$$x^T x - 3(x+d)^T(x+d) + 3(x+2d)^T(x+2d) - (x+3d)^T(x+3d) = 0, \tag{4}$$

which holds for all $x, d \in \mathbb{F}_p^n$, implies that if the first three terms $x, x+d, x+2d$ of a 4-term progression lie in A, then the last term $x+3d$ is guaranteed to lie in A with probability 1 (and not α, which would have to be the case if we wanted to obtain the expected number $\alpha^4 p^2$). We also note that the Gauss sum $\mathbb{E}_x \omega^{x^T x}$ exhibits square-root cancellation, which implies that the Fourier coefficients of A are very small, so that the set A is indeed uniform. Identities of the form (4) will play an important role later on.

1.3. The uniformity norms. Example (3) shows that for counting longer progressions a different approach is required. In his work on Szemerédi's theorem Gowers developed an approach to 3-term progressions that generalizes more easily to longer progressions. It turns out that being uniform in the Fourier sense is equivalent to containing the expected number of additive quadruples, i.e. quadruples $(x, y, z, w) \in G^4$ satisfying $x+y = z+w$. By defining the U^2 norm as the average

$$\|f\|_{U^2}^4 = \mathbb{E}_{x+y=z+w} f(x)f(y)\overline{f(z)f(w)} = \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b),$$

which is easily seen to equal the (4th power of the) ℓ^4 norm of the Fourier transform of f, we obtain a measure of uniformity that is equivalent to the Fourier transform. Viewing the points $x, x+a, x+b, x+a+b$ as the vertices of a parallelogram, a generalization readily suggests itself. Gowers defined the U^3 norm by the formula

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

this time summing over the vertices of a 3-dimensional parallelepiped, taking complex conjugates of f via the operator $\mathcal{C}$ as required. This formula has no useful interpretation in terms

of the Fourier transform, but it does control the 4-term progression count in the sense that

$$|\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)f(x+3d)| \leq \|f\|_{U^3},$$

which follows straight from the definitions and three judicious applications of the Cauchy-Schwarz inequality. In other words, the 4-AP count is stable under small perturbations in the U^3 norm, which implies that a set whose balanced function $f_A = 1_A - \alpha$ has small U^3 norm contains the expected number of 4-APs. We call such sets *quadratically uniform*.

As before, if we aim to establish the existence of a 4-term progression inside the set A, then we are done if this set is quadratically uniform. However, the "structure" part of the dichotomy (2) is much harder to obtain this time, although perhaps not too difficult to guess. Recall that if A is non-uniform in the classical sense, then it correlates with a linear phase, so it is not unreasonable to expect a set which is not quadratically uniform to correlate with a quadratic phase, especially after one has checked that a function of the form $f(x) = \omega^{q(x)}$, where ω is a pth root of unity and q a quadratic form, has U^3 norm 1. Indeed, the bulk of Gowers's work on Szemerédi's theorem [41] goes into showing that a set whose balanced function has large U^3 norm correlates with a quadratic structure on a very long arithmetic progression, which turns out to be deep statement requiring a number of sophisticated tools from additive number theory for its proof. We give a more detailed account of results of this type, now referred to as *inverse theorems*, in Section 2.1.

To conclude this introductory discussion, we give the formal definition of the U^k norm and discuss some of the properties which we shall need later on.

Definition 1.1 (Uniformity norms)**.** *Let $k \geq 2$ be an integer. For any function $f : G \to \mathbb{C}$, define the U^k norm by the formula*

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

where for $y = (y_1, \ldots, y_k) \in G^k$ and $\omega = (\omega_1, \ldots, \omega_k) \in \{0,1\}^k$ we have written $\omega \cdot y = \omega_1 y_1 + \cdots + \omega_k y_k$, as well as $\mathcal{C}^{|\omega|} f = f$ if $|\omega| = \omega_1 + \cdots + \omega_k$ is even and $\overline{f}$ otherwise.

It is not hard to show (but neither is it obvious) that this expression defines a norm for all $k \geq 2$. The main technical device for this purpose is the so-called *Gowers-Cauchy-Schwarz inequality*, which is proved using several applications of the ordinary Cauchy-Schwarz inequality. It states that for any family of functions $g_\omega : \mathbb{Z}_N \to \mathbb{C}$, $\omega \in \{0,1\}^k$, we have the

bound

$$|\mathbb{E}_{x\in G, y\in G^k} \prod_{\omega\in\{0,1\}^k} \mathcal{C}^{|\omega|} g_\omega(x+\omega\cdot y)| \leq \prod_{\omega\in\{0,1\}^k} \|g_\omega\|_{U^k}.$$

With the help of this inequality it is straightforward to verify that the uniformity norms form a nested sequence

$$\|f\|_{U^2} \leq \|f\|_{U^3} \leq \cdots \leq \|f\|_{U^k} \leq \ldots \|f\|_\infty.$$

We also record the fact that the U^k norms can be defined inductively via the formula

$$\|f\|_{U^k}^{2^k} = \mathbb{E}_{h\in G}\|\Delta_h f\|_{U^{k-1}}^{2^{k-1}},$$

where $\Delta_h f(x) = f(x)\overline{f(x+h)}$ should be viewed as a discrete derivative of f. The U^k norm controls the count of $(k+1)$-term progressions in the following sense.

Proposition 1.2 (U^k norm controls $(k+1)$-APs)**.** *Let $k \geq 2$ be an integer, and let $f_0, f_1, \ldots, f_k : G \to \mathbb{C}$ be functions satisfying $\|f_j\|_\infty \leq 1$ for $j = 0, 1, \ldots, k$. Then*

$$|\mathbb{E}_{x,d\in G} f_0(x) f_1(x+d) f_2(x+2d) \ldots f_k(x+kd)| \leq \min_{0\leq j\leq k} \|f_j\|_{U^k}.$$

Having defined the U^k norms, and more importantly realized their potential for controlling progressions, the proof of this statement is not very difficult (if a little tedious). It involves a number of applications of the Cauchy-Schwarz inequality, combined with a suitable reparametrization of the progression itself.

1.4. Discussion and outlook. There are some fundamental open problems concerning the uniformity norms themselves. For example, can they be classified in a sense comparable to the L^p norms [16]? For example, in [16] the L^p norms were shown to be the only norms which are both invariant under coordinate permutation and multiplicative with respect to tensor products.

The uniformity norms are perfectly suited to counting arithmetic progressions, but turn out to be less ideal in situations involving more general linear systems, a phenomenon which we shall discuss in detail in Chapter 3. Recent work of Hatami [59] gives examples of interesting classes of norms on graphs, whose analogues for sets may have further applications.

Above we saw that set which is uniform in the Fourier sense does not always contain the expected number of 4-term progressions. A beautiful question due to Ruzsa asks the following: if $A \subseteq G$ is a uniform subset of density α, does it always contain at least $\alpha^C|G|^2$ 4-term progressions, for some constant C? For densities close to $1/2$ some conclusions can

be drawn from the results in [3].

2 Advances in discrete harmonic analysis

Much of this thesis is about developing higher-order generalizations of the Fourier decomposition in (2) suitable for the various applications we have in mind. This involves first of all a deep understanding of the structure of functions whose U^k norms is large. In Section 2.1 I discuss the main results in this direction, together with some important observations made in the series of joint papers with Tim Gowers [6, 7, 8], which will be of crucial importance for developing efficient decompositions in Section 2.2. I also mention a result which is implicit in [9] and which will be used again in Chapter 4.

2.1. Inverse theorems. Because it is technically simpler, we first discuss the finite field case. Following the work of Gowers [42] and adding an additional "symmetry argument", Green and Tao proved the following inverse theorem for U^3 on $\mathbb{F}_p^n$ [49], which they then applied it to give a proof of Szemerédi's theorem for progressions of length 4.

Theorem 2.1 (Inverse theorem for U^3 on $\mathbb{F}_p^n$)**.** *Let $p > 2$, and 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*

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

where C_p is a constant that depends on p only.

A recent breakthrough of Sanders [76], which gives improved bounds on one of the central ingredients in the inverse theorem, namely Bogolyubov's lemma (see Section 4.2) implies quasi-polynomial correlation. It was shown in [53] that the conjectured polynomial bounds are equivalent to the polynomial Freiman-Ruzsa conjecture (PFR).

The case $p = 2$ of Theorem 2.1 was independently proved by Samorodnitsky [74] in the context of theoretical computer science (see Chapter 4). The structure of the proof is roughly the same as that in [49], but Green and Tao's proof includes division by 2, which is of course forbidden in characteristic 2. We shall see a bit further along that the case of low characteristic continues to present somewhat unexpected problems in this area.

It was observed but not exploited by Green and Tao [49] that a slightly stronger form of the inverse theorem holds, which we state as Theorem 2.2 below. If V is a subspace of $\mathbb{F}_p^n$ and $y \in \mathbb{F}_p^n$, then one can define a seminorm $\|.\|_{u^3(y+V)}$ on functions from $\mathbb{F}_p^n$ to $\mathbb{C}$ by setting

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

where the supremum is taken over all quadratic forms q on $y+V$ and ω denotes a pth root of unity. This semi-norm clearly measures the correlation over a coset of the subspace V.

Theorem 2.2 (Local inverse theorem for U^3 on $\mathbb{F}_p^n$)**.** *Let $p > 2$, and let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$. Then there exists a subspace V of $\mathbb{F}_p^n$ such that* $\mathrm{codim} V \leq \delta^{-C_p}$ *and*

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

In fact, with Sanders's improvement [76] the codimension of this subspace is now polylogarithmic in δ. One can deduce the usual inverse theorem from this version without too much effort by extending the locally defined quadratic phase to the whole of $\mathbb{F}_p^n$, recovering the afore-mentioned exponential bound.

For $p = 2$, the equivalent of Theorem 2.2 follows directly neither from Green and Tao's nor Samorodnitsky's approach but instead requires a merging of the two. This was implemented in [9].

Theorem 2.3 (Tulsiani-Wolf)**.** *Theorem 2.2 holds in characteristic 2.*

Actually, one can be even more precise: it follows from the proof that the quadratic parts of the quadratic phase functions q_y, defined on $y+V$ for each y and having correlation at least δ^{C_p} each, are all identical. In other words, each $q_y(x)$ has the form $q(x-y)+l_y(x-y)$ for a single quadratic function $q : V \to \mathbb{F}_p$ (that is independent of y) and some Freiman 2-homomorphisms $l_y : V \to \mathbb{F}_p$. In order to take full advantage of this parallel correlation property, we made the following definition in [6].

Definition 2.4 (Quadratic average)**.** *Let V be a subspace of $\mathbb{F}_p^n$ and let q be a quadratic form on V. A* quadratic average with base (V, q) *is a function of the form $Q(x) = \mathbb{E}_{y \in x-V} \omega^{q_y(x)}$, where each function q_y is a quadratic map from $y+V$ to $\mathbb{F}_p$ defined by a formula of the form $q_y(x) = q(x-y) + \phi_y(x-y)$ for some Freiman homomorphism $\phi_y : V \to \mathbb{F}_p$. The* rank *of Q is the rank of the quadratic form q, and the* complexity *of Q is the codimension of V.*

It is not difficult to deduce the following variant involving quadratic averages from Theorem 2.2, which we shall use in some of our decomposition theorems later on.

Corollary 2.5 (Gowers-Wolf)**.** *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 quadratic average Q of complexity at most $(2/\delta)^{C_p}$ such that*

$$|\langle f, Q \rangle| \geq (\delta/2)^{C_p}/2.$$

Finally, let us discuss the higher-order generalization of Theorem 2.1. It should not come as a surprise that a function whose U^k norm is large correlates with a polynomial phase of degree $k-1$, even though this took some time to prove. Bergelson, Tao and Ziegler [20, 87] had to resort to ergodic theoretic methods and a correspondence principle specifically designed for this purpose to obtain the following result.

Theorem 2.6 (Inverse theorem for U^k on $\mathbb{F}_p^n$)**.** *Let $0 < \delta \leq 1$ and let p be a prime. Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function with $\|f\|_\infty \leq 1$ and $\|f\|_{U^{k+1}} \geq \delta$. Then there exists a polynomial $\pi : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree k and a constant $\gamma(\delta)$ such that*

$$|\mathbb{E}_{x \in \mathbb{F}_p^n} f(x) \omega^{\pi(x)}| \geq \gamma(\delta),$$

provided that $p \geq d$.

Tao and Ziegler also proved an inverse theorem for the U^k norm in the case where the degree of the polynomial exceeds the characteristic of the field [88]. The original conjecture in this regard had been shown to be false [65, 54], and it turns out that a polynomial phase needs to be defined slightly differently to include, for example, functions such as ω^π, where ω is now a p^dth root of unity for some $d = d(k)$. We shall not consider this case in applications and shall therefore not discuss it further.

Let us now turn our attention to the case of functions defined on $\mathbb{Z}_N$. An example of Furstenberg and Weiss [85] shows that it is *not* true that a function whose U^3 norm is large correlates with a quadratic phase function of the form ω^q on the whole of $\mathbb{Z}_N$. Gowers showed that it does so locally, however, for example on a long arithmetic progression, and Green and Tao refined his argument to prove that it does so on a so-called *Bohr set* [85, 13]. In order to give the definition of a Bohr set, we write $\|\beta\|$ for the distance of $\beta \in \mathbb{R}$ from the nearest integer.

Definition 2.7 (Bohr set)**.** *Given $K \subseteq \widehat{\mathbb{Z}_N}$ and $0 < \rho < 1/2$, the set*

$$B(K, \rho) = \{x \in \mathbb{Z}_N : \|xt/N\| < \rho \text{ for all } t \in K\}$$

is referred to as a Bohr set *of dimension $|K|$ and width ρ.*

Setting $\rho = 0$ and thinking of the product xt as a scalar product of vectors in $\mathbb{F}_p^n$, we see that $B(K, \rho)$ is an approximate version of a subspace whose orthogonal complement is given by K. So-called *regular* Bohr sets, introduced by Bourgain in [26], can indeed act as "approximate subgroups" in $\mathbb{Z}_N$, and take the role of the subspace V in $\mathbb{F}_p^n$. Of course Bohr

sets as defined above are not actually additively closed. In fact, if $B = B(K, \rho)$, then $B + B$ can be as large $2^{|K|}|B|$. Instead, one has to work with pairs B, B' where B' is much smaller than B, satisfying the approximate additive closure condition $B + B' \approx B$. The techniques for dealing with regular Bohr sets are well understood by a handful of experts, but because they are extremely technical we say no more about them here.

We can now state Green and Tao's result concisely, which represents the $\mathbb{Z}_N$ analogue of Theorem 2.2 above.

Theorem 2.8 (Inverse theorem for U^3 on $\mathbb{Z}_N$)**.** *Let $f : \mathbb{Z}_N \to \mathbb{C}$ be a function such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$, and let $C = 2^{24}$. Then there exists a regular Bohr set $B = B(K, \rho)$ with $|K| \leq (2/\delta)^C$ and $\rho \geq (\delta/2)^C$ such that $\mathbb{E}_y \|f\|_{u^3(B+y)} \geq (\delta/2)^C$.*

As in the finite field case, we can also deduce a variant involving quadratic averages, which are defined in the obvious analogous way. The only difference is that the complexity of a quadratic average in this context has two parameters, corresponding to the dimension and width of the base Bohr set.

It is only very recently that Green, Tao and Ziegler [57], in a major breakthrough, were able to prove an inverse theorem for functions on $\mathbb{Z}_N$ for higher values of k. Both statement and proof draw inspiration from ergodic theory, and in particular the deep structure theory of characteristic factors induced by the analogue of the U^k norms in the ergodic context, developed in a seminal article by Host and Kra [61]. Even though we shall not use the result in the sequel, we include its statement here in order to illustrate the difference with the highly quantitative statements we have seen so far. We shall need the following definition, which made its first appearance in [21].

Definition 2.9 (Nilsequence)**.** *Let G be a k-step nilpotent group, i.e. a connected, simply connected Lie group with central series $G = G_1 \supseteq ... \supseteq G_{k+1} = \{1\}$. Let $\Gamma \subseteq G$ be a discrete co-compact subgroup. Then the quotient G/Γ is a k-*step nilmanifold*, and a sequence of the form $(F(g^n x))_{n \in \mathbb{N}}$ with $g \in G, x \in G/\Gamma$ and continuous $F : G/\Gamma \to \mathbb{R}$ is called a k-*step nilsequence.

Note that $G/\Gamma = \mathbb{R}/\mathbb{Z}$ is an example of a 1-step nilmanifold, and that any function of the form $n \mapsto F(x + n\alpha)$ for $\alpha \in \mathbb{R}$ and continuous F is a 1-step nilsequence. In particular, the linear characters $n \mapsto \exp(2\pi i n\alpha)$ are examples of basic 1-step nilsequences. For higher values of k, including $k = 2$, a k-step nilsequence displays an undeniably non-commutative character. A simple example of a 2-step nilmanifold is given by the quotient of the matrix

group G and its discrete subgroup Γ defined by

$$G = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix},$$

which can be identified topologically with the 2-torus. Now let $g \in G$ be given by

$$g = \begin{pmatrix} 1 & m & \beta \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix},$$

where $m \in \mathbb{Z}$ and $\alpha, \beta \in \mathbb{R}$. Then a shift of $(x, y) \in \mathbb{T}^2$ by g is given by $(x, y) \mapsto (x+\alpha, y+\beta+mx)$, and the nilsequence $F(g^n(x, y))$ for $n \in \mathbb{N}$ is given by $F(x + n\alpha, y + n\beta + \frac{1}{2}mn(n+1)\alpha)$, clearly exhibiting the claimed quadratic behaviour. Observe in particular that a quadratic phase function such as $n \mapsto \exp(\pi i n(n+1)\alpha)$ belongs to the family of basic 2-step nilsequences.

We are now able to state the inverse theorem for the U^k norm, which asserts that a function whose U^k norm is large correlates with a $(k-1)$-step nilsequence.

Theorem 2.10 (Inverse theorem for U^k on $\mathbb{Z}_N$)**.** *Let $0 < \delta \leq 1$ and $k \geq 1$ be an integer. Then there exists a finite collection $\mathcal{M}_{k,\delta}$ of k-step nilmanifolds G/Γ, each equipped with some smooth Riemannian metric $d_{G/\Gamma}$ as well as constants $C(k, \delta)$, $c(k, \delta) > 0$ with the following property. Whenever $N \geq 1$ and $f : [N] \to \mathbb{C}$ with $\|f\|_\infty \leq 1$ is a function such that $\|f\|_{U^{k+1}[N]} \geq \delta$, then there exist a nilmanifold $G/\Gamma \in \mathcal{M}_{k,\delta}$, some $g \in G$ and a function $F : G/\Gamma \to \mathbb{C}$ with Lipschitz constant at most $C(k, \delta)$ with respect to the metric $d_{G/\Gamma}$, such that*

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F(g^n x)}| \geq c(k, \delta).$$

The dependence of c on k and δ is ineffective.

2.2. Decomposition theorems. While an inverse theorem suffices to prove a statement such as Szemerédi's theorem via a dichotomy (either $\|f\|_{U^k}$ is small, or it is not), it is often more convenient to rephrase this dichotomy in terms of a decomposition of the function f into a structured and a uniform part. An example of such a decomposition for the U^2 norm was given in (2): any bounded function f can be written as a small sum of some linear phases (f_1, consisting of the characters where the Fourier transform of f is large), plus a part which is small in U^2 (f_2, consisting of the characters where the Fourier transform of f is small).

The aim of the series of joint papers with Tim Gowers [6, 7, 8] was to develop such decomposition theorems for the quadratic and higher-order case. In this section we shall focus on the harmonic analysis aspects of the theory. While we postpone our discussion of applications to Chapter 3, it is important to bear in mind that the more sophisticated (and technically challenging) versions of the decomposition theorems mentioned here were driven by the need for strong quantitative results in these applications.

Note first that the reason this decomposition is easy to come by in the linear case is that the Fourier characters form an orthonormal basis, giving rise to a unique expansion of any bounded function in terms of this basis. In the quadratic case already, the system is overdetermined: there are far more than p^n functions of the form ω^q for a quadratic form q. A simple inductive argument which tries to extract the quadratic phases from f one by one does not work: while the inverse theorem gives us a quadratic phase with which f correlates, after subtracting it from f with a coefficient that ensures an energy decrease, we can no longer be sure that the new function is bounded by 1 in L^∞ (and worse, there doesn't seem to be any way to control the potential increase).

One way of getting around this problem is by using averaging projections, which decrease both the L^2 and the L^∞ norm. This was implemented by Green and Tao in [45], in what can legitimately be called the first quadratic decomposition theorem. We briefly describe their approach here for historical reasons, and for comparison with later decomposition statements.

Observe that in the linear case, we called f_1 "structured" because it is constant on affine subspaces of low codimension, namely the simultaneous level sets of the characters where the Fourier transform of f is large. If we would like to make this idea quadratic, we should consider simultaneous level sets of quadratic forms. Let us briefly introduce the terminology of Green and Tao to make this idea more precise. Let $\Gamma_1 : \mathbb{F}_p^n \to \mathbb{F}_p^{d_1}$ be a surjective linear map, and for each $a \in \mathbb{F}_p^{d_1}$ set V_a to equal $\Gamma_1^{-1}(\{a\})$. 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 $\{x \in \mathbb{F}_p^n : \Gamma_2(x) = b\}$. 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.

More precisely, 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), in close analogy with the "characteristic factors" that occur in ergodic theory. Writing $\mathbb{E}(f|\mathcal{B}_2)$ for the conditional expectation, or averaging projection, of f, we can now state a first decomposition theorem.

Theorem 2.11 (Green-Tao)**.** *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}.

This decomposition is proved via an energy increment argument: using the inverse theorem one refines the trivial factor in successive stages in such a way that the energy of the projection of f (the L^2 norm of $\mathbb{E}(f|\mathcal{B}_2)$) increases at each step, a process which has to terminate because the energy is bounded above by 1. Green and Tao used this decomposition to give a generalization of Khintchine's recurrence theorem in [45]: they proved that for any dense subset of $\mathbb{F}_p^n$, there exists at least one common difference which occurs almost the expected number of times in the 4-term progressions of A. In [2] Tim Gowers and I used a refinement of Theorem 2.11 to prove the first case of conjecture regarding the true complexity of a system of linear equations (see Chapter 3).

However, it turned out that there were certain technical aspects of this decomposition (in particular, rank considerations which are discuss in more detail towards the end of this chapter) that made it necessary to search for alternative approaches. One of the main innovations that [6] brought to the subject is the use of the Hahn-Banach theorem from functional analysis to prove the existence of higher-order decompositions. This has turned out to be quite an influential approach and led to a number of further applications (see Section 3.2). Moreover, the decompositions thus obtained look and feel much closer to what one would expect a higher-order generalization of (2) to look like, giving added legitimacy to calling the subject "higher-order Fourier analysis".

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 this section we shall refer to an inner product, which is just the standard inner product on $\mathbb{C}^N = \mathbb{C}^{\mathbb{Z}_N}$ (or equivalently $\mathbb{C}^{\mathbb{F}_p^n}$).

Theorem 2.12 (Finite-dimensional Hahn-Banach theorem)**.** *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 $|\langle x, z\rangle| > 1$ and such that $|\langle y, z\rangle| \leq 1$ whenever $\|y\| \leq 1$.*

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

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

which for technical reasons we shall have to generalize 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 straightforward to prove.

Lemma 2.13. *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

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

We quickly deduce the following corollary of Theorem 2.12.

Corollary 2.14. *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.*

We now illustrate in a simple case how Corollary 2.14 can be used to deduce a decomposition theorem from a given inverse theorem.

Theorem 2.15 (Gowers-Wolf)**.** *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 $\exp(C_p(\eta\delta)^{-C_p})$, where C_p is the constant in Theorem 2.1.

The proof proceeds by contradiction. Suppose that the function f cannot be decomposed in the way desired. 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 2.14 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 (Theorem 2.1) to $\eta\phi$, we find that there is a quadratic form q such that $|\langle \phi, \omega^q\rangle| \geq \exp(-C_p(\eta\delta)^{-C_p})$, contradicting the fact that it has to be at most M^{-1}.

This gives the desired decomposition into a quadratically structured plus a quadratically uniform part. The L^1 error arises as a consequence of the fact that the inverse theorem can only be proved for functions bounded in L^∞.

Using the refined inverse theorem involving quadratic averages (Theorem 2.5) instead of Theorem 2.1, we are able to reduce the complexity M of the structured part to polynomial instead of exponential in the error parameters, at the cost of making the description of the quadratic objects involved slightly more complicated.

Theorem 2.16 (Gowers-Wolf)**.** *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*

$$f(x) = \sum_i \lambda_i Q_i(x) + g(x) + h(x),$$

where the Q_i are quadratic averages on $\mathbb{F}_p^n$ of complexity at most $(2/\delta)^{C_p}$, and

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

Here M can be taken to be $(2/\eta\delta)^{C_p}/2$.

Unfortunately the measure of the complexity of the structured part, here taken to be the sum of the absolute values of the coefficients of the quadratic phases, is not strong enough

for applications. It turns out that we need to be able to bound the *number* of quadratic phases, not just the L^1 norm of their coefficients.

This was done in [6] by clustering those quadratic phases which are close in rank together in a crude way. Basically, if two quadratic forms are close in rank, then they differ by a multiplicative factor exhibiting strongly linear behaviour, leading to the introduction of the functions U_i in the next decomposition.

Theorem 2.17 (Gowers-Wolf)**.** *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be a function such that $\|f\|_2 \leq 1$, and let $\delta > 0$. Let $d = (2/\delta)^{C_p}$ and $C = (2/\delta^2)^{C_p}$. Then f has a decomposition of the form*

$$f(x) = \sum_{i=1}^{k} Q_i(x)U_i(x) + g(x) + h(x),$$

where $k \leq C^2/\delta^2$, the Q_i are quadratic averages on $\mathbb{F}_p^n$, $\sum_{i=1}^k \|U_i\|_{U^2}^* \leq 2^{3/2}C^4\delta^{-3}p^{3d/4}$, $\sum_{i=1}^k \|U_i\|_\infty \leq C$, $\|g\|_1 \leq 2\delta$ and $\|h\|_{U^3} \leq \delta$.

The fact that one has functions U_i (as opposed to simple complex coefficients λ_i) does not cause any actual problems in applications because of their strongly linear nature, but it is extremely inconvenient and leads to a number of technical issues. It was observed by Ernie Croot and Olof Sisask (personal communication) that one can obtain a decomposition with a comparable bound on the number of phases involved while retaining the nature of the coefficients directly from Theorem 2.16 using their almost-periodicity result, which we discuss in depth in Section 4.2. This also follows from the "boosting" method of decomposition in Section 4.1.

It turns out that in our main application, which will be described in Chapter 3, we have some additional information on the function f that we are decomposing, namely that it is strongly linearly uniform (small in U^2). Here then is the most sophisticated quadratic decomposition theorem that we proved for $\mathbb{F}_p^n$.

Theorem 2.18 (Gowers-Wolf)**.** *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 C such that for every R_0 there exists a constant c with the following property. Let $d = (2/\delta)^{C_p}$ and $C = (2/\delta^2)^{C_p}$. Suppose that $\|f\|_{U^2} \leq c$. Then f can be written as*

$$f(x) = \sum_{i=1}^{k} Q_i(x)U_i(x) + g(x) + h(x),$$

where $k \leq C^2/\delta^2$, the Q_i are quadratic averages on $\mathbb{F}_p^n$ of complexity at most d, $\sum_{i=1}^k \|U_i\|_{U^2}^* \leq 2^{3/2}C^4\delta^{-3}p^{3d/4}$, $\sum_{i=1}^k \|U_i\|_\infty \leq C$, $\|g\|_1 \leq 7\delta$ and $\|h\|_{U^3} \leq 2\delta$. In addition, each quadratic average Q_i has rank at least R_0 provided that c satisfies the inequality $c \leq p^{-p^{16d}R_0}$.

It stands to reason that the quadratic phases in the decomposition of linearly uniform functions should be of high rank as low rank phases exhibit strongly linear behaviour, contradicting the assumption of uniformity. Note also that the rank condition here is absolutely essential in applications, since the exponential sum $\mathbb{E}_x\omega^{q(x)}$ is small if and only if the rank of q is large, and it is precisely these exponential sums that we need to compute in Chapter 3.

The next challenge is to adapt the ideas just described to $\mathbb{Z}_N$, which turned out to be rather challenging. This is because we no longer have the exact algebraic notion of a subspace available to us. While in $\mathbb{Z}_N$ we do have regular Bohr sets (introduced in Section 2.1) at our disposal, the technicalities involved in making the proofs "approximate" are quite formidable. The other algebraic notion that immediately becomes problematic, glancing back at the statement of Theorem 2.18, is that of the *rank* of a quadratic phase. Remember that our quadratic phases in $\mathbb{Z}_N$ are only defined locally, and may not look anything like an actual quadratic polynomial. Instead, they only exhibit piecewise quadratic behaviour, for example as in the bracket quadratic $\{\alpha n\{\beta n\}\}$, where $\{\gamma\}$ denotes the fractional part of a real number γ. How does one deal with the highly algebraic notion of rank in this context? The answer to this question is one of the important innovations in [7, 8].

Let us examine the case of quadratic forms in the finite field setting more closely to see what can be salvaged. What we actually use in applications is the following dichotomy: either q has large rank, in which case the exponential sum $\mathbb{E}_{x\in V}\omega^{q(x)}$ is small, or its rank is large, in which case q is constant on translates of a subspace of low codimension. Indeed, suppose that we have a quadratic form q defined on a subspace V of $\mathbb{F}_p^n$. Then a simple calculation shows that $|\mathbb{E}_{x\in V}\omega^{q(x)}| = p^{-r/2}$, where r is the rank of the bilinear form $\beta(u,v) = (q(u+v) - q(u) - q(v))/2$ associated with q. Now

$$|\mathbb{E}_{x\in V}\omega^{q(x)}|^2 = \mathbb{E}_{x,y\in V}\omega^{q(x)-q(y)} = \mathbb{E}_{x,y\in V}\omega^{\beta(x+y,x-y)} = \mathbb{E}_{u,v\in V}\omega^{\beta(u,v)}.$$

For each u, the expectation over v is 0 unless $\beta(u,v) = 0$ for every v, in which case it is 1. But the set of u such that $\beta(u,v)$ vanishes is a subspace of V of codimension r, so it has density p^{-r}, which proves the result.

With this in mind, observe that we could, if we wanted, define the rank of q to be $\log_p(\alpha^{-1})$, where $\alpha = |\mathbb{E}_x\omega^{q(x)}|^2$. It turns out that this gives us a definition that can be carried over much more easily to functions defined on Bohr sets in $\mathbb{Z}_N$.

Definition 2.19 (Analytic rank)**.** *Let B be a Bohr set and let q be a quadratic form on B. Let B' be a Bohr set such that $2B' - 2B' \subset B$ and let P be a subset of B'. The* rank *of the local quadratic phase function $h(x) = 1_B(x)\omega^{q(x)}$* relative to *$P$ is $\log(1/\alpha)$, where α is the*

quantity

$$|\mathbb{E}_{a,a',b,b' \in P} \omega^{q(a+b-a'-b')-q(a-a')-q(b-b')}|$$

If Q is a generalized quadratic average with base (B, q), then we define the rank *of Q* relative to *P to be the rank of the local quadratic phase function $1_B(x)\omega^{q(x)}$ relative to P.*

We do of course pay a price for such a move. If we define rank in this way then it becomes true by definition that averages over quadratic phase functions of high rank are small. But we clearly cannot avoid doing any work: it is now *not* obvious that rank is subadditive or that a quadratic function of low rank has linear structure. In fact, neither of these statements is exactly true, but with some effort we were able to prove usable approximations to them.

The resulting high-rank quadratic decomposition theorem for uniform functions on $\mathbb{Z}_N$ (the analogue of Theorem 2.18) is too technical to comfortably state in this exposition, so we refer the courageous reader to Theorem 8.16 in [7]. We remark that the idea of analytic rank was taken up by Tao and Ziegler in [88].

Finally, let us say a few words about higher-order decomposition theorems, which we developed in [8] for $\mathbb{F}_p^n$. It turns out that the notion of analytic rank (Definition 2.19) which was so useful in $\mathbb{Z}_N$ can also be used for higher-degree polynomials and their associated multilinear forms over $\mathbb{F}_p^n$. However, this requires an iterative procedure. A naive application of Hahn-Banach and the U^4 inverse theorem, for example, gives us a decomposition into a part consisting of cubic phase functions, plus a part that is small in U^4. If we only have the additional assumption that the function we are decomposing is linearly uniform, then we cannot insist that the cubic phases have high rank. Instead, we need to take any low-rank cubic phase functions and express them in terms of quadratic ones, creating an additional (quadratically) structured part, which fortunately can be shown to be high-rank by the uniformity assumption.

At this point an additional problem arises: in applications we shall estimate averages over products of functions which have been decomposed in this way, each one consisting of a cubically and a quadratically structured part. If we want all resulting exponential sums to be small, then we need to be able to say, for example, that the product of a quadratically structured with a cubically structured part is small. This is achieved if the rank of the cubic forms can be made arbitrarily small compared with the complexity of the quadratic part, leading to what is known as an *arithmetic-regularity type* decomposition. This type of decomposition, which involves arbitrary growth functions R_i allowing us to lower bound the rank of each structured part, necessarily comes with tower-type bounds and is thus of a more qualitative nature.

Theorem 2.20 (Gowers-Wolf)**.** *Let s and k be positive integers with $s \leq k$, let $\epsilon > 0$, and*

let $\eta : \mathbb{R}_+^{k-s+1} \to \mathbb{R}_+$ be a function that is strictly decreasing in each variable. Let $R_s, \dots, R_k$ be functions from $\mathbb{R}_+^{k-s+1}$ to $\mathbb{R}_+$ that are strictly increasing in each variable. Then there are functions $M_{s,0}, \dots, M_{k,0}$, where $M_{i,0}$ is a function from $\mathbb{R}_+^{i-s}$ to $\mathbb{R}_+$ (that depends on ϵ, η and the functions $R_i, \dots, R_k$) and a constant $c' = c'(\epsilon, \eta, R_s, \dots, R_k) > 0$, such that if f is any function that takes values in $[-1, 1]$ and satisfies $\|f\|_{U^s} \leq c'$, then there are real numbers $M_s, \dots, M_k$ and a decomposition

$$f = f'_s + \cdots + f'_k + g + h$$

with the following properties.

- *We can write $f'_i = \sum_j \lambda_{i,j} \omega^{\pi_{i,j}}$, where the functions $\pi_{i,j}$ are polynomials of degree i and the $\lambda_{i,j}$ are real coefficients with $\sum_j |\lambda_{i,j}| = M_i \leq M_{i,0}(M_s, \dots, M_{i-1})$.*
- *For each i, each polynomial $\pi_{i,j}$ has rank at least $R_i(M_s, \dots, M_k)$.*
- *$\|g\|_{U^{k+1}} \leq \eta(M_s, \dots, M_k)$.*
- *$\|h\|_2 \leq \epsilon$.*

A qualitative higher-order decomposition on $\mathbb{Z}_N$ in terms of nilsequences was obtained by Green and Tao in [55].

2.3. Discussion and outlook. Despite a number of impressive results in recent years, the theory of higher-order Fourier analysis remains poorly understood. The beginnings of an algebraic theory, which combines Host and Kra's ergodic theoretic approach [61] with the work of Gowers, Green and Tao, are emerging in the work of Szegedy (see [83] and sequels) but are considered unsatisfactory at the time of writing.

While ergodic theory has proved an extremely useful source of inspiration and technical tools, there is an strong need for non-ergodic, strongly quantitative results. In particular, the fact that the current proof of the inverse theorem for U^4 over $\mathbb{F}_p^n$, which has a completely combinatorial statement, is of an infinitary ergodic nature is a highly disturbing state of affairs that urgently requires resolution.

More explicit analogues of statements that are known for traditional Fourier analysis are needed. An example is Parseval's identity, which can be used to give an upper bound on the number of large Fourier coefficients in a linear decomposition. In this weak sense the bound on k in Theorem 2.17 can be considered a quadratic analogue, but it is certainly not the formula we are after. In the finite field setting, an explicitly (or perhaps randomly) chosen

subset $\mathcal{Q}$ of quadratic phases ought to satisfy an approximate Parseval identity, in the sense that the U^3 norm of f can be expressed as the sum over all $q \in \mathcal{Q}$ of an appropriate power of $\mathbb{E}_x f(x)\omega^{q(x)}$.

The decompositions in this chapter have all been of an abstract nature – the Hahn-Banach theorem implies the existence of a decomposition, but gives us no way of writing one down explicitly. An obvious question to ask is whether, given a bounded function, we can actually compute its quadratic Fourier coefficients. We shall address this question in detail in Chapter 4.

3 Applications in number theory

Listing Szemerédi's theorem as one of the applications of the tools of higher-order Fourier analysis developed in Chapter 2 would of course be strictly speaking correct. But it would boldly misrepresent the order of events – the basics of higher-order Fourier analysis were invented by Gowers for the sole purpose of giving an analytic proof of Szemerédi's theorem, not out of some abstract interest in quadratic phase decompositions.

As mentioned in Section 2.2, the first truly new application of Green and Tao's quadratic decomposition theorem (Theorem 2.11) was a Khintchine-type recurrence result for dense sets [45]. Undoubtedly the most celebrated fruit of the higher-order analytic theory, however, is the Green-Tao theorem on long arithmetic progressions in the primes [50].

Theorem 3.1 (Green-Tao theorem)**.** *The primes contain arbitrarily long arithmetic progressions. Moreover, the same is true of any subset of the primes of positive relative density.*

In Section 3.1 I will explain the main application of the sophisticated decomposition theorems from Section 2.2, which was inspired by a generalization of Theorem 3.1 but brought with it some surprising consequences for our understanding of the uniformity norms. In collaboration with Tim Gowers, I studied the counting of solutions to general linear systems of equations in dense uniform subsets of $\mathbb{Z}_N$ and $\mathbb{F}_p^n$. I shall describe some further applications in Section 3.2, which are *not* due to myself but rather have directly used or at least been influenced by the methods developed in Section 2.2. In Section 3.3 I present very recent joint work with Thái Hoàng Lê [12], in which we prove that any subset of the primes of positive relative upper density contains a polynomial configuration with prime step, extending a result of Tao and Ziegler [86].

3.1. The "true" complexity of a system of linear equations. Following the proof of Theorem 3.1, Green and Tao established established in [51] the existence of arbitrary linear patterns in primes (with the exception of the famous twin primes and those constellations related to Goldbach's problem), at the time conditional on a U^k inverse theorem and a conjecture regarding the pseudorandomness properties of the Möbius function, which has since been resolved [52].

Not surprisingly, a first step in this paper was to control a linear configuration, that is the image of a system of linear forms $L_1, \dots, L_m$ in d variables, by an appropriate uniformity norm. In particular, Green and Tao proved, using the Cauchy-Schwarz inequality and a suitable

reparameterization of the linear system, that

$$|\mathbb{E}_{x_1,\dots,x_d} \prod_{i=1}^{m} f(L_i(x_1,\dots,x_d))| \leq \|f\|_{U^{k+1}} \tag{5}$$

for any bounded function f, provided that the linear system has "Cauchy-Schwarz complexity" at most k. This complexity was defined in terms of the linear dependence between the linear forms, and sufficient for their purpose.

Definition 3.2 (Cauchy-Schwarz complexity)**.** *Let $L_1, ..., L_m$ be a system of m linear forms in d variables. For $1 \leq i \leq m$ and $s \geq 0$, this system is said to be 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 *of the system $L_1, ..., L_m$ 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.*

For example, it is not hard to check that the system $x, x+d, \dots, x+(k-1)d$ defining a k-term progression has Cauchy-Schwarz complexity $k-2$. Given that higher-order Fourier analysis becomes significantly more difficult and less well developed as k increases, it seemed desirable to determine the *least* k such that the U^k norm controls the average over the linear system in (5).

As we saw in Section 1.2, (5) is tight for arithmetic progressions, but is it tight in general? This question was the starting point of my joint investigations with Tim Gowers, the first result of which already formed part of my doctoral thesis. Rather surprisingly, we discovered that there were linear systems that had complexity 2 according to the measure of Green and Tao, but that were in reality controlled by the U^2 (and not the U^3) norm. An example of such a system is given by the linear forms $x, y, z, x+y+z, x+2y-z, x+2z-y$. We made the following definition to make the notion of "control" quantitatively precise.

Definition 3.3 (True complexity)**.** *Let $L_1, L_2, \dots, L_m$ be a system of m distinct linear forms in d variables. Its* true complexity *is the least integer 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*

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

Given the obvious necessary condition which follows from a generalization of the construction in Section 1.2, we made the following conjecture in [2].

Conjecture 3.4 (Gowers-Wolf, 2007)**.** *The true complexity of a system of linear forms $L_1, \dots, L_m$ is equal to the least integer k such that the functions L_i^{k+1} are linearly independent.*

In other words, we conjectured that a linear system $L_1, \dots, L_m$ is controlled by the U^k norm if and only if the kth powers L_i^k of the linear forms are linearly independent. In particular, a linear system is governed by ordinary Fourier analysis if and only if the squares of the linear forms are linearly independent. Note that this condition is satisfied for a generic system of linear forms.

In [2] we resolved the simplest non-trivial case in the finite field setting: we showed that if $L_1, \dots, L_m$ is a linear system over $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity at most 2, then it has true complexity 1 if and only if the linear forms are square-independent, i.e. if and only if the forms $L_1^2, \dots, L_m^2$ are linearly independent. This first paper used the quadratic factor decomposition of Green and Tao (Theorem 2.11). It gave tower-type bounds for the dependence of ϵ on δ in Definition 3.3, and should therefore be considered more of a qualitative result.

The main purpose of the strong decomposition theorems in Section 1.2 therefore was to develop strong quantitative bounds for this problem, remove the hypothesis of Cauchy-Schwarz complexity 2 in the quadratic case and prove the higher-order analogues of Conjecture 3.4, as well as extend the above-mentioned result to the technically much more challenging setting of $\mathbb{Z}_N$.

Indeed, we were able to address all of these points, that is, we were able to precisely classify those linear systems for which ordinary Fourier analysis suffices. One direction of this result is a relatively straightforward generalization of the example in (3). There the U^3 norm was shown to be necessary because the squares of the linear forms $x, x+d, x+2d, x+3d$ representing a 4-term arithmetic progression were linearly dependent (4). In order to prove the implication in the reverse direction, we decomposed the balanced function f_A of $A \subseteq G$ into a part that is small in a certain U^j norm (where j is determined by how well the Cauchy-Schwarz inequality can be applied to control the average in (5)) plus a part that has polynomial structure of degree between k and $j-1$. One then performs an explicit computation over the structured parts, exploiting the fact that the functions L_i^k (and hence all higher powers) are linearly independent.

In [6] and [7] we obtained the following strongly quantitative results in the quadratic case. Note in particular that the paper [7] would have ended up being half its current length if we had settled for an ever so slightly weaker bound.

Theorem 3.5 (Gowers-Wolf, 2010)**.** *Let $L_1, \dots, L_m$ be a square-independent system of m*

linear forms in d variables in $G = \mathbb{F}_p^n$ or $G = \mathbb{Z}_N$ of Cauchy-Schwarz complexity at most 2. Then for every $\epsilon > 0$ there exists $c > 0$ with the following property. If $f : G \to [-1, 1]$ is such that $\|f\|_{U^2} \leq c$, then

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

Moreover, the uniformity parameter c can be taken to be

$$\exp(-\exp(c_m \epsilon^{-C_m})),$$

where c_m and C_m are constants depending on m only (and on p, in the case $G = \mathbb{F}_p^n$).

It seems unlikely that this doubly exponential dependence is best possible. In the higher-order case, we obtained the following result of a more qualitative flavour [8].

Theorem 3.6 (Gowers-Wolf, 2010)**.** *Let $L_1, \dots, L_m$ be a system of m linear forms in d variables in $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity $k \leq p$. Suppose that $L_1, \dots, L_m$ are degree-s independent for some $s \leq p$. Then for every $\epsilon > 0$ there exists $c > 0$ with the following property.*

If $f : \mathbb{F}_p^n \to [-1, 1]$ is such that $\|f\|_{U^s} \leq c$, then

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

In other words, $L_1, \dots, L_m$ has true complexity at most $s - 1$.

Since the proof of this theorem uses the arithmetic-regularity type decomposition in Theorem 2.20, the dependence of c on ϵ is very bad. However, since the higher-order inverse theorem (Theorem 2.6) itself is of a non-quantitative form, this is probably a forgivable weakness.

Hatami and Lovett treated the off-diagonal case of Theorem 3.6, that is the case where we allow m different functions $f_1, \dots, f_m$ to appear in the average to be estimated, using a factor-type decomposition theorem. The higher-order case in $\mathbb{Z}_N$ was settled by Green and Tao in [55], using the machinery of nilsequences in an entirely non-quantitative fashion.

A similar phenomenon was discovered simultaneously and independently by Leibman [64] in the context of ergodic theory.

3.2. Further applications of higher-order decompositions. In this section I briefly mention some other results that have followed from our higher-order decomposition theorems and their

method of proof.

Firstly, a modification of Theorem 2.17 was used by Candela [28] to prove that if A is a dense subset of $\{1, 2, \dots, N\}$ then the set of all d such that A contains an arithmetic progression of length 3 and common difference d must itself contain an arithmetic progression of length at least $(\log\log N)^c$. Here the exponent c depends on the density of A only.

Since the set of gaps of 3-term progressions of a dense set A is itself a dense set, the existence of a long arithmetic progression follows directly from Szemerédi's theorem. However, the length of such a progression that is implied by Gowers's bound is of the order of $\log\log\log\log\log N$, which is much weaker than Candela's result. To do better, Candela observes that the function

$$S_3(d) = \mathbb{E}_{x\in\mathbb{Z}_N} 1_A(x)1_A(x+d)1_A(x+2d),$$

whose support is precisely the set of 3-AP gaps of A, is quadratically anti-uniform in the sense that $\|S_3\|^*_{U^3} \leq \alpha^{3/2}$, where α is the density of A. Such quadratically anti-uniform functions are quadratically structured, and can therefore be decomposed similarly to Theorem 2.17, without the quadratically uniform part g.

Shortly after we used the Hahn-Banach theorem for proving the first decomposition theorems in [6], it was observed by Gowers [43] that this method can also be applied to provide an alternative proof of one of the crucial ingredients of the Green-Tao theorem (Theorem 3.1), namely the so-called transference principle. This observation was made simultaneously and completely independently by Reingold, Trevisan, Tulsiani and Vadhan [71] in the context of theoretical computer science, where the Hahn-Banach theorem is known under the name *duality of linear programming*.

The main idea of the transference principle is that if a theorem (such as Szemerédi's theorem or the inverse theorem) is true for bounded functions, then it should be true for functions that are majorized by a so-called "pseudorandom measure" (we shall not give the precise definition here since it is rather technical). Indeed, a first step in this direction was taken by Kohayakawa, Łuczak and Rödl [62], who proved the existence of 3-term progressions in dense subsets of a (truly) random set. Since then, a number of results from additive combinatorics have been transferred to the context of dense subsets of random sets (see for example [58]). With the benefit of hindsight therefore, establishing a similar result for dense subsets of *pseudo*random sets seems rather natural, even though at the time it represented a clear conceptual leap that marks the main advance in the work of Green and Tao (and the earlier paper by Green on Roth's theorem in the primes [47]). The transference principle was

not stated explicitly in [50] but was for the first time isolated in [86]. Both papers proceed via a decomposition into an almost periodic and a weakly mixing part, reminiscent of Furstenberg's ergodic structure theorem [39]. Here we state the more abstract version that follows from the Hahn-Banach technique in Gowers's language.

Theorem 3.7 (Transference principle)**.** *For every $\delta, \eta > 0$, there exists $\epsilon > 0$ with the following property. Let ν be a $(k+1)$-pseudorandom measure satisfying $\|\nu - 1\|_{U^k} \leq \epsilon$, and suppose that $f : \mathbb{Z}_N \to \mathbb{R}$ is a function satisfying $0 \leq f \leq \nu$. Then there exists a function g such that $0 \leq g \leq (1-\delta)^{-1}$ and $\|f - g\|_{U^k} \leq \eta$.*

The existence of such a measure ν in the case where f equals the prime-counting von Mangoldt function follows from sieve theory methods. The main additional difficulty in the proof of Theorem 3.7 (over some of the decomposition theorems from Section 2.2 say) is that one requires an additional lemma, which states that products of degree-k dual functions are bounded in the U_k^* norm.

For completeness, let us very briefly sketch how Theorem 3.7, together with Szemerédi's theorem, implies the Green-Tao theorem. Recall that we wish to estimate from below the average

$$\mathbb{E}_{x,d} f(x) f(x+d) \dots f(x+kd),$$

where $f(n) = \Lambda(n)$ (or rather a slightly modified version thereof), is bounded above by (a constant times) ν. Theorem 3.7 gives us a function $0 \leq g \leq 1 + \delta$ such that $\|f - g\|_{U^k} \leq \eta$, and so by Proposition 1.2 the above average is, up to small error terms, equal to (a constant multiple of)

$$\mathbb{E}_{x,d} g(x) g(x+d) \dots g(x+kd).$$

Now g is an (almost) bounded function, so Szemerédi's theorem applies, implying that the latter average is bounded below by a constant depending only on the density of g. But the density of g is strictly positive since $\mathbb{E}g = \mathbb{E}f - \mathbb{E}(f-g)$, and $|\mathbb{E}(f-g)| \leq \|f-g\|_{U^k} \leq \eta$, concluding the proof for an appropriate choice of the parameters δ and η.

For a more detailed discussion of the transference principle see [10]. The transference principle and the underlying Hahn-Banach approach has also found important, more combinatorial applications in recent work of Conlon and Gowers [31].

3.3. Polynomial configurations in the primes. It is natural to ask for polynomial generalizations of Szemerédi-type theorems: is it true that any sufficiently dense subset of the first N integers contains a given polynomial configuration? For example, is it true that any

sufficiently dense subset of $[N]$ contains a configuration of the form $x, x+n^2$ for $x, n \in \mathbb{N}$? The latter question was answered in the affirmative by Sárközy [81], stated here with the best known bound due to Pintz, Steiger and Szemerédi [69].

Theorem 3.8 (Sárközy's theorem)**.** *Suppose that $A \subseteq \{1, \dots, N\}$ is a subset of density α which contains no two distinct elements whose difference is a perfect square. Then*

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

An analogous statement is easily seen to be false for a difference of the form n^2+1: since there are no squares congruent to 2 mod 3, we can take the set of multiples of 3 as a (very dense) counterexample. The proof of Sárközy's theorem proceeds via classical Fourier analysis on $\mathbb{Z}$, following the circle method approach of Hardy and Littlewood. It can be extended to configurations of the form $x, x+P(n)$, where P is an "intersective" polynomial, that is, a polynomial which has a root modulo q for every $q \in \mathbb{N}$. Unfortunately, no such quantitative results are known for more general polynomial configurations in one variable. What we do have is a result of Bergelson and Leibman [19], proved entirely within the realm of ergodic theory, which states in a purely qualitative fashion that any subset of the integers of positive upper density contains the translate of a simultaneous image of a system of polynomials with zero constant coefficient.

Theorem 3.9 (Bergelson-Leibman theorem)**.** *Let $A \subseteq \mathbb{Z}$ be a set of positive upper density, and let $P_1, \dots, P_k$ be polynomials with rational coefficients satisfying $P_i(0) = 0$ for $i = 1, \dots, k$. Then there exist $x, m \in \mathbb{Z}$ such that $x + P_i(m) \in A$ for $i = 1, \dots, k$.*

Theorem 3.9 follows from an abstract result about the convergence of multiple ergodic averages in a measure-preserving dynamical system, via Furstenberg's correspondence principle [39]. A combinatorial proof of the Bergelson-Leibman theorem is yet to be found. Quantitatively, even the case of 3-term arithmetic progressions with square common difference remains unresolved.

Even so, Tao and Ziegler [86] managed to extend the Green-Tao theorem (Theorem 3.1) to the case of polynomial common differences.

Theorem 3.10 (Tao-Ziegler theorem)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbb{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the primes of positive relative upper density contains a configuration $x + P_1(m), \dots, x + P_k(m)$, where x, m are integers, $d \neq 0$.*

Tao and Ziegler used a pseudo-quantitative Bergelson-Leibman theorem as a crucial ingredient in their proof (to which they then apply a "transference principle", see Section

3.2), and hence do not obtain asymptotics. However, their result is quantitative in principle and even shows that these configurations exist in the expected order of magnitude.

Recently Wooley and Ziegler [93] and Frantzikinakis, Host and Kra [37] extended the Bergelson-Leibman theorem in another direction, by showing that the variable m in Theorem 3.9 can be taken to be a prime minus 1.

Theorem 3.11 (Wooley-Ziegler, Frantzikinakis-Host-Kra)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbb{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the integers of positive relative upper density contains a configuration $x + P_1(p-1), \dots, x + P_k(p-1)$, where x is an integer and p is prime.*

Together with Thái Hoàng Lê [12] we were able to formulate a hybrid of Theorem 3.10 and Theorem 3.11.

Theorem 3.12 (Lê-Wolf, 2012)**.** *Let $P_1, \dots, P_k$ be polynomials in $\mathbb{Z}[x]$ such that $P_i(0) = 0$ for $i = 1, \dots, k$. Then any subset of the primes of positive relative upper density contains a configuration of the form $x + P_1(p-1), \dots, x + P_k(p-1)$, where x is an integer and p is prime. The same is true if we replace $p - 1$ with $p + 1$.*

In other words, we conclude that any subset of the primes of positive relative upper density contains a polynomial progression $x + P_1(m), x + P_2(m), \dots, x + P_k(m)$, where the variable m itself is a prime minus 1 (or a prime plus 1). Again, our proof actually yields the correct order of magnitude for these configurations, but no information on the asymptotics.

The method of proof follows that in [37] and involves a comparison of the average along the shifted primes with the corresponding average along the natural numbers. More specifically, we prove the following proposition, using a now standard inductive process called *PET induction* on the polynomial system that linearizes it in successive steps, until we end up with an average that resembles a Gowers norm.

Proposition 3.13 (Lê-Wolf, 2012)**.** *Let ν_1, ν_2 be a pair of pseudorandom measures on $X = \mathbb{Z}/N\mathbb{Z}$ satisfying the extra condition. Suppose that $f_1, \dots, f_k$ are functions on X with $|f_i| \leq \nu_1$ for $i = 1, \dots, k$, and that g is a weight on X with support in $[M]$ such that $|g| \leq 1 + \nu_2$. Then*

$$\mathbb{E}_{m \in [M]} \int_X g(m) T^{P_1(Wm)/W} f_1 \cdots T^{P_k(Wm)/W} f_k = O(\|g\|_{U^d(\mathbb{Z}_{(2d+1)M})}) + o(1),$$

where d is an integer depending only on the system of polynomials.

Such a proposition was proved in [37] for functions that were bounded above by a constant, and we adapt the proof to cover functions that are majorized by a pseudorandom measure, similar to the generalized von Neumann theorem in [50].

In order to deduce Theorem 3.12 from Proposition 3.13 we want to set the function g equal to a suitable variant of the von Mangoldt function minus the constant function, f roughly equal to the indicator function of the subset A of positive relative density in the primes, and finally show that the U^d norm of g is small. The latter fact holds as a result of Green and Tao's deep programme of counting linear patterns in primes [51, 52, 57], whose main result can be stated as follows [37, Theorem 2.2].

Theorem 3.14 (Green-Tao, Green-Tao-Ziegler)**.** *Let $w : \mathbb{Z} \to \mathbb{R}$ be a function tending to infinity more slowly than $\log\log\log N$, and set $W = \prod_{p<w} p$ be the product of primes less than w. Then for every $d \in \mathbb{Z}^+$, we have*

$$\lim_{N\to\infty} \max_{\substack{1\le b<W,\\ (b,W)=1}} \|\Lambda_{W,b;N} - 1_{[N]}\|_{U^d(\mathbb{Z}_{(2d+1)N})} = 0.$$

Here $\Lambda_{W,b;N}$ is the von Mangoldt function, modified so as to get around the fact that the primes are not equidistributed with respect to small moduli, specifically

$$\Lambda_{W,b;N}(n) = \begin{cases} \frac{\phi(W)}{W}\log(Wn+b), & \text{if } Wn+b \text{ is prime and } 1 \le n \le N, \\ 0 & \text{otherwise.} \end{cases}$$

Thus Theorem 3.12 follows from Proposition 3.13 by setting $g = \Lambda_{W,b;M} - 1_{[M]}$ and $f_1 = \cdots = f_k = f$, where

$$f(x) = \begin{cases} \frac{\phi(W)}{W}\log R & \text{if } R \le x \le N/2 \text{ and } Wx+b \in A, \\ 0 & \text{otherwise,} \end{cases}$$

and b is chosen by the pigeonhole principle such that the subset A of the primes of positive relative density has positive relative density in the residue class $b \mod W$. Then both f and g are (essentially) controlled by one of the pseudorandom measures

$$\nu_{W,b}(n) = \frac{\phi(W)}{W}\log R\left(\sum_{m|Wn+b} \mu(m)\chi\left(\frac{\log m}{\log R}\right)\right)^2,$$

where μ is the Möbius function and χ is an even smooth function supported on $[-1,1]$

satisfying

$$\int_0^1 |\chi'(t)|^2 dt = 1.$$

(This definition was used already by Green and Tao in [51], replacing a slightly less convenient variant in [50].) Indeed, we need to take $\nu_1 = \nu_{W,b}$ and $\nu_2 = \nu_{W,1}$, so that the functions f, g are controlled by ν_1, ν_2, respectively.

It turns out that in order to be able to prove results about polynomial systems the pseudorandom measure needs to satisfy more stringent conditions than in the case of linear systems. In particular, in [86] Tao and Ziegler showed that $\nu_{W,b}$ as defined above satisfies the so-called *polynomial forms* and the *polynomial correlation condition*. In order to prove Proposition 3.13, we need a further hypothesis that is very similar to the polynomial forms condition and that we termed the *extra condition*. It does not immediately follow from the properties of $\nu_{W,b}$ stated by Tao and Ziegler, so in order to show that it is satisfied by we had to go relatively deep into the number theoretic part of [86].

3.4. Discussion and outlook. Conjecture 3.4 suggests that the uniformity norms may not be the perfect tool to control all types of arithmetic structures in dense sets, even though they are perfectly suited to the study of arithmetic progressions. Are there other norms that capture "square independence", and for which a linear inverse theorem can be proved? The fact that we have to pass through quadratic Fourier analysis to prove that the classical Fourier transform suffices to control a square-independent linear system is disconcerting. It seems plausible that a careful analysis of the proof of the U^3 inverse theorem might allow one to use the square-independence hypothesis directly with a linear outcome.

As we remarked in Section 3.3, very few results about polynomial systems are known. Even the case of 3-term arithmetic progressions with square common difference remains wide open. In a beautiful paper [48] Green showed that any sufficiently dense subset of $\{1, \dots, N\}$ contains a 3-term progression with common difference $x^2 + y^2$ (a configuration that occurs with much higher frequency in the integers than x^2 itself).

Theorem 3.15 (Green)**.** *Let $A \subseteq \{1, \dots, N\}$ be a subset of density α containing no 3-term progressions with common difference of the form $x^2 + y^2$. Then*

$$\alpha \ll (\log \log N)^{-c}$$

for some constant c.

The proof of Theorem 3.15 crucially relies on the fact that every prime congruent 1 mod 4

can be written as a sum of two squares. Sieve theory methods then allow one to count various configurations of such primes, which in turn reduces the problem to a mild modification of the inverse question in quadratic Fourier analysis.

My masters student Eric Naslund has been exploring the extent to which this proof generalizes to more general polynomial configurations [66]. Results from algebraic number theory on the representation of primes by so-called *norm forms* (related to Chebotarev's density theorem, see for example [68]) allow him to establish a new non-trivial class of polynomial common differences. In addition, he is able to improve Green's original doubly logarithmic bound to one of the form $\exp(-c\sqrt{\log\log N})$ by applying methods for simultaneous linearization of quadratics developed by Green and Tao in the context of Szemerédi's theorem for progressions of length 4 [56].

The study of polynomial (as opposed to linear) structures using analytic tools will play an important role in years to come. A first step, with the introduction and careful analysis of *local* uniformity norms, was taken by Tao and Ziegler in [86], where they showed that the primes contain arbitrarily long polynomial progressions. However, just as in the original proof of the Green-Tao theorem, they got away without proving an inverse theorem for these norms (which would be necessary if one wanted to give asymptotics for the number of such polynomial progressions in the sense of the Bateman-Horn conjecture). Obtaining such an inverse theorem would be an important step forward in our understanding of polynomial patterns.

I have also begun collaborating on a problem in sieve theory with Harald Helfgott. The large sieve is an analytic tool for bounding from above the density of a set S of integers that is ill-distributed modulo p for many primes p. It is optimal for at least one example, namely the set of squares, but is likely to be far from optimal in general. The aim is to give strong upper bounds for the density of ill-distributed sets that are not strongly algebraic. Some steps in this direction were taken in [60], and recent (as yet unpublished) work of Harper and Green implies the desired result in the case where the set of excluded residues modulo each prime is rigidly structured. The uniformity norms are another tool for detecting algebraicity which may be effectively employed in this context, but the project is still in its initial stages.

4 Interactions with theoretical computer science

Many of the aforementioned questions in higher-order Fourier analysis arise naturally in theoretical computer science. In particular, the uniformity norms and associated analysis have been used in the context of probabilistically checkable proofs [75], pseudo-random generators [25] and communication complexity [92]. Here we shall focus on the implications for property testing and decoding algorithms.

In property testing one is interested in determining whether a given object (such as a graph, or a function) has a certain property (such as containing a triangle, or being linear), or whether the object is "far" from having that property in an appropriate sense, with high probability. One wants to do so by sampling the object at very few instances, usually logarithmically many in the size of the object. In order to be more precise, let us say that two functions $f, g : \mathbb{F}_2^n \to \{0,1\}$ are ϵ-far from each other if their fractional Hamming distance (the fraction of inputs on which they differ) exceeds ϵ.

For example, in order to test whether a function $f : \mathbb{F}_2^n \to \{0,1\}$ is close to a linear function, one simply asks whether $f(x+y)$ equals $f(x)+f(y)$ (modulo 2) for randomly chosen points x and y in the domain of f. The basic *linearity test*, due to Blum, Luby, and Rubinfeld [24], accepts this function as close to linear if the answer to this question is positive after having queried the function a certain number of times, and rejects otherwise. The analysis of this test relies on the discrete Fourier transform.

We have already seen that the U^3 norm is a natural measure for how far a function is from being quadratic. For functions $\mathbb{F}_2^n \to \{0,1\}$, the U^3 norm can be estimated fairly accurately by the Hoeffding bound (Lemma 4.8 below) using only few samples. The obvious "quadraticity" test will therefore reject the function as ϵ-close to a quadratic if its U^3 norm is below a certain threshold parameter $c(\epsilon)$, and accept otherwise. Showing that functions with large U^3 norm possess quadratic structure is then a crucial step in the analysis of this test, which aims to determine the relationship between ϵ, $c(\epsilon)$, the probability of false rejection and the number of samples needed. A tight analysis was carried out by Samorodnitsky in [74] (see also the important predecessor [14]).

The question of testing low-degree polynomials over $\mathbb{F}_2^n$ is closely related to the decoding problem for Reed-Muller codes of order d, which are simply the evaluation tables of all degree-d polynomials in n variables. Decoding such a code amounts to the following task: given $f : \mathbb{F}_2^n \to \{0,1\}$ which is known to be within distance $\delta > 0$ of a codeword of the Reed-Muller code of order d, actually find such a codeword. The unique decoding radius (within which there exists precisely one such codeword) is known to be 2^{-d} for Reed-Muller codes of oder d.

The question of decoding therefore turns from an algebraic into an analytic one for distances beyond the unique decoding radius, where there may be exponentially many codewords near f. Samorodnitsky's U^3 test [74] gives a way of determining, with high probability, whether there is such a codeword at distance $1/2 - \epsilon$ in the case $d = 2$.

In Section 4.1 I describe work with Madhur Tulsiani [9] in which we turned Samorodnitsky's tester into a *local self-corrector*, that is, we gave a probabilistic algorithm which retrieves a codeword of the Reed-Muller code of order 2 when the function is in fact $(1/2 - \epsilon)$-close to such a codeword. To our knowledge, this is the first decoding procedure for any class of codes beyond the list-decoding radius. Using the "gradient descent" or "boosting" technique from machine learning, we used this algorithm to give a procedure that explicitly and efficiently computes a quadratic decomposition for a given function in the sense of Section 2.2. In Section 4.2 I outline an alternative proof of an important almost-periodicity result of Croot and Sisask [32], together with several algorithmic applications. This covers recent work done in collaboration with Eli Ben-Sasson, Noga Ron-Zewi and Madhur Tulsiani [11], which significantly improved the efficiency of the above-mentioned quadratic decomposition algorithm.

4.1. A quadratic Goldreich-Levin theorem. Our aim in this section will be to show how to efficiently compute a quadratic decomposition of a given function $f : \mathbb{F}_2^n \to [-1, 1]$ with small error, somewhat comparable to Theorem 2.15 and its refinements. The linear version of this statement is the celebrated theorem of Goldreich and Levin [40], which was originally conceived for the construction of pseudorandom generators from one-way functions. It gives an algorithm that outputs a list of large Fourier coefficients for the given function f in time logarithmic in 2^n (hence much faster than the Fast Fourier Transform).

Theorem 4.1 (Goldreich-Levin theorem)**.** *Let $\gamma, \delta > 0$. There is a randomized algorithm which, given oracle access to a function $f : \mathbb{F}_2^n \to \pm 1$, runs in time $O(n^2 \log n \cdot \mathtt{poly}(1/\gamma, \log(1/\delta)))$ and outputs a decomposition*

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

with the following guarantee.

- $k = O(1/\gamma^2)$.
- $\mathbb{P}[\exists i \ |c_i - \widehat{f}(\alpha_i)| > \gamma/2] \leq \delta$.

- $\mathbb{P}[\forall \alpha \text{ such that } |\widehat{f}(\alpha) \geq \gamma|, \quad \exists i \ \alpha_i = \alpha] \geq 1 - \delta$.

In [9] Madhur Tulsiani and I proved the following quadratic analogue.

Theorem 4.2 (Tulsiani-Wolf)**.** *Let $\epsilon, \delta > 0$, $n \in \mathbb{N}$ and $B > 1$. Then there exists $\eta = \exp((B/\epsilon)^C)$ and a randomized algorithm running in time $O(n^4 \log n \cdot \mathtt{poly}(1/\eta, \log(1/\delta)))$ which, given any function $f : \mathbb{F}_2^n \to [-1, 1]$ as an oracle, outputs with probability at least $1-\delta$ a decomposition into quadratic phases*

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

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

This is the first explicit decomposition theorem of its kind, as all previous such results (in particular the ones described in Section 2.2) where of a rather abstract nature: the existence of such a decomposition was proved, for example, via the Hahn-Banach theorem or an energy-increment strategy, but the coefficients c_i and quadratic phases q_i could not be determined. There are, however, some precedents of algorithmic decomposition theorems in computer science, for example the weak regularity lemma of Frieze and Kannan [38], which decomposes a matrix as a small sum of cut matrices and has found numerous applications in approximately solving constraint satisfaction problems.

The algorithm in Theorem 4.2 naturally splits into two subtasks. First one has to develop a procedure to address the following question: given a function $f : \mathbb{F}_2^n \to \{-1, 1\}$, which is at distance at most $1/2 - \epsilon$ from an unknown quadratic phase $(-1)^{q'}$, find (efficiently) a quadratic phase $(-1)^q$ which is at distance at most $1/2 - \eta$ from f, for some $\eta = \eta(\epsilon)$. This amounts to a (relaxed) self-corrector for the Reed-Muller code of order 2, which is hard to obtain for the reasons described in the introduction of this chapter. In particular, the list-decoding radius of Reed-Muller codes of order 2 is $1/4$, and at distance greater than $1/4$ there may be exponentially many (in n) such functions q. However, the problem is still be tractable as we are required to find only *one* such q (which might be at a slightly larger distance than q'). We discuss this question in more detail below.

The second ingredient, given an algorithm `Find-Quadratic` which performs the self-correcting task just described, is a procedure that assembles the quadratic phases with which the function correlates into an actual decomposition. The general difficulty encountered in doing this was already discussed in detail at the start of Section 2.2. Here we take an approach which is very different from that of Section 2.2 and which has its origins in machine learning theory.

The procedure is actually quite simple to state. Given a function f, an error parameter ϵ and algorithm `Find-Quadratic`, we perform the following routine.

- Define functions $g_1 = h_1 = f$. Set $t = 1$.
- While $\|g_t\|_{U^3} \geq \epsilon$
  - let q_t be the output of `Find-Quadratic` when called with the function g_t
  - $h_{t+1} := h_t - \eta q_t$
  - $g_{t+1} := \mathtt{Truncate}_{[-B,B]}(h_{t+1}) = \max\{-B, \min\{B, h_{t+1}\}\}$
  - $t := t + 1$.

When the algorithm ends after k steps, set $g = g_k$ and $h = h_k - g_k$. It is not too difficult to show that the resulting decomposition satisfies the properties required for Theorem 4.2.

Note that the main objection we raised in Section 2.2 to a naive iterative extraction of quadratic phases from the function f, namely the loss of the L^∞ bound required for an application of the U^3 inverse theorem (Theorem 2.1), is dealt with by truncation here. Observe also that in [7] we had to work much harder to obtain a bound on the number of terms in the decomposition, rather than just the ℓ^1 norm of its coefficients (see Theorem 2.17). The above decomposition approach gives such a bound immediately and is equivalent from a quantitative point of view: we can bound the number of terms here by $1/\eta^2$, which is exponential in $1/\epsilon$.

Previous decomposition theorems have used similar procedures [38, 91], generally referred to as *gradient descent* or *boosting*. However, they required that the quadratic phase found at each step have correlation $\eta = O(\epsilon)$, if one exists with correlation ϵ. In particular, they required the fact that if one scales f to change its L_∞ norm, the quantities η and ϵ scale the same way (which is not true if, say, $\eta = \epsilon^2$). Our procedure works even as η degrades arbitrarily in $1/\epsilon$. This requires a somewhat more sophisticated analysis and the introduction of a third error term, for which we bound the L_1 norm.

We now turn our attention to the self-correction procedure `Find-Quadratic`, whose properties we first spell out in detail.

Theorem 4.3 (Tulsiani-Wolf)**.** *Given $\epsilon, \delta > 0$, there exists $\eta = \exp(-1/\epsilon^C)$ and a randomized algorithm* `Find-Quadratic` *running in time $O(n^4 \log n \cdot \mathtt{poly}(1/\epsilon, 1/\eta, \log(1/\delta)))$ which, given oracle access to a function $f : \mathbb{F}_2^n \to \pm 1$, either outputs a quadratic form $q(x)$ or $\perp$. The algorithm satisfies the following guarantee.*

- *If $\|f\|_{U^3} \geq \epsilon$, then with probability at least $1 - \delta$ it finds a quadratic form q such that $\langle f, (-1)^q \rangle \geq \eta$.*

- *The probability that the algorithm outputs a quadratic form q with $\langle f, (-1)^q \rangle \leq \eta/2$ is at most δ.*

The algorithm `Find-Quadratic` is an algorithmic version of the inverse theorem (Theorem 2.1) in disguise. We did not say anything about the proof of the inverse theorem in Section 2.1, and shall not describe it in detail here either as it is rather lengthy. We confine ourselves to stating that it uses several deep results from additive combinatorics in succession, notably the Balog-Szemerédi-Gowers theorem (Theorem 4.4 below) and the Freiman-Ruzsa theorem [85, 13].

What is important (and highly problematic) about these theorems from the algorithmic point of view is that they all deal with dense subsets of $\mathbb{F}_2^n$ – we are looking to obtain an algorithm of running time which is polylogarithmic in 2^n, so we do not have time to store such a set, let alone perform sophisticated operations on it. We get around this problem by building efficient sampling procedures or procedures for efficiently deciding membership in such sets, which lead to new algorithmic proofs of the above-mentioned combinatorial theorems.

A subtlety arises when one tries to construct such testing procedures. The estimates obtained may be erroneous, leading to some noise in the decision of such an algorithm. The output therefore is a noisy version of the set (or rather, a distribution over sets). In order to feed this noisy version into the next part of the algorithm, we needed to develop robust versions of the relevant theorems, for which we can "sandwich" the output between two sets with desirable properties.

We give an example of what we mean by this in the case of the Balog-Szemerédi-Gowers theorem [17, 41]. It states that if a set A contains many additive quadruples, that is, elements $a_1, a_2, a_3, a_4 \in A$ such that $a_1 + a_2 = a_3 + a_4$, then a large subset of it must have small sumset. (The sumset $A + A$ of a set $A \subseteq \mathbb{F}_2^n$ is defined to be the set of elements $a + a'$ such that $a, a' \in A$.)

Theorem 4.4 (Balog-Szemerédi-Gowers theorem)**.** *Suppose that $A \subseteq \mathbb{F}_2^n$ contains at least $|A|^3/K$ additive quadruples. Then there exists a constant C depending only on K and a subset $A' \subseteq A$ of size $|A'| \geq K^{-C}|A|$ with the property that $|A' + A'| \leq K^C|A'|$.*

A robust algorithmic version of Theorem 4.4 as it is needed in the proof of the U^3 inverse theorem is the following [9].

Proposition 4.5 (Tulsiani-Wolf)**.** *Let $\delta, \epsilon > 0$ and let the parameters $\gamma(\epsilon)$ and $\rho(\epsilon)$ be chosen appropriately. Let ϕ be a function $\mathbb{F}_2^n \to \mathbb{F}_2^n$, and let A_ϕ be a subset of the graph $\{(x, \phi(x)) : x \in \mathbb{F}_2^n\}$ containing at least a proportion γ of additive quadruples.*

Then there exists a procedure `BSG-Test`*, making a number of queries depending on* ρ *and* δ*, with the following property. For every* $u = (x, \phi(x))$ *and every appropriate choice of* η*, there exist two sets* $A_\phi^{(1)}(u) \subseteq A_\phi^{(2)}(u)$ *such that the output of* `BSG-Test` *satisfies the following with probability at least* $1 - \delta$*.*

- $\texttt{BSG-Test}(u, v, \eta) = 1 \implies v \in A_\phi^{(2)}(u)$.
- $\texttt{BSG-Test}(u, v, \eta) = 0 \implies v \notin A_\phi^{(1)}(u)$.

Moreover, with probability $\rho^3/24$ *over the choice of* u *and* η*, we have*

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

In other words, the elements on which `BSG-Test` answers 1 are sandwiched between a large set and a set with small doubling. Moreover, the procedure is very efficient in that it takes at most $\texttt{poly}(1/\epsilon, \log(1/\delta))$ many steps.

In Section 2.1 we already mentioned that a more precise, "local" version of the inverse theorem exists: disregarding for the moment Sanders's improvement [76] to a crucial ingredient in the proof, Theorem 2.2 says that a function of U^3 norm at least ϵ correlates with a quadratic phase on a subspace of codimension polynomial in ϵ^{-1}, and the correlation itself is polynomial in nature if one allows more complicated objects such as quadratic averages as output (see Definition 2.4). Combining the Green-Tao approach with Samorodnitsky's symmetry argument in characteristic 2, we obtained an analogue of the local inverse theorem in characteristic 2 (Theorem 2.3). Moreover, we were able to prove the following algorithmic version. (Recall that the *complexity* of a quadratic average is the codimension of the subspace on which the quadratic phases are defined.)

Theorem 4.6 (Tulsiani-Wolf)**.** *Given* $\epsilon, \delta > 0$ *and* $n \in \mathbb{N}$*, there exist* $K, C = O(1)$ *and a randomized algorithm* `Find-QuadraticAverage` *running in time* $O(n^4 \log^2 n \cdot \exp(1/\epsilon^K) \cdot \log(1/\delta))$*, which, given oracle access to a function* $f : \mathbb{F}_2^n \to \pm 1$*, either outputs a quadratic average* $Q(x)$ *of complexity* $O(\epsilon^{-C})$*, or the symbol* $\perp$*. The algorithm satisfies the following guarantee.*

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

Of course this theorem can be combined with the iterative procedure described above for assembling a decomposition. The resulting algorithm now outputs a decomposition into only polynomially many quadratic objects (of which each has polynomial complexity).

We briefly outline the key modifications to the proof of Theorem 4.3 that allow us to obtain Theorem 4.6. An important ingredient in the proof of the inverse theorem is the Freiman-Ruzsa theorem [72], which states that a set A with small sumset $A + A$ is efficiently contained in a subspace. This is applied immediately after the Balog-Szemerédi-Gowers theorem, and is relatively costly. Instead, Green and Tao [49] employ a result known as *Bogolyubov's lemma* (see Theorem 4.11 in the next section) to locate a large subspace inside the iterated sumset $2A - 2A$. In [9] we therefore also give a robust algorithmic version of Bogolyubov's lemma (which turns out to be much easier than Proposition 4.5).

The main motivation for the work described in the next section [11] was to strengthen the quantitative aspects of Theorems 4.3 and 4.6 (and hence 4.2) by incorporating Sanders's improvements to Bogolyubov's lemma [76] into the algorithm.

4.2. Algorithmic versions of almost-periodicity results. When Croot and Sisask introduced "a probabilistic technique for finding almost-periods of convolution" a couple of years ago [32], it created quite a splash in the additive combinatorics community. Roughly speaking, their main result says that if a set A has bounded sumset $|A + A| \leq K|A|$, then there exists a dense set T such that $1_A * 1_A(\cdot)$ and $1_A * 1_A(\cdot + t)$ are almost indistinguishable in the L^2 norm (or higher L^p norms) for all $t \in T$. Since it is valid for general groups G, we write it here in multiplicative notation.

Proposition 4.7 (Croot-Sisask lemma)**.** *Let $\epsilon > 0$ and let $m \geq 1$ be an integer. Let G be a group, let $A, B \subseteq G$ be finite subsets, and suppose $S \subseteq G$ is a set such that $|B \cdot S| \leq K|B|$. Then there is a set $T \subseteq S$ of size $|T| \geq |S|/(2K)^{50m/\epsilon}$ such that for each $t \in TT^{-1}$,*

$$\|1_A * 1_B(xt) - 1_A * 1_B(x)\|_{2m}^{2m} \leq \max\{\epsilon^m |AB||B|^m, \|1_A * 1_B\|_m^m\}\epsilon^m |B|^m.$$

Croot and Sisask initially used Proposition 4.7 to establish the existence of long arithmetic progressions in sumsets of dense sets, and to give a new (albeit quantitatively weak) proof of Roth's theorem on 3-term progressions. They also proved a weak Bogolyubov-Ruzsa lemma, which asserts the existence of an iterated sumset of a dense set inside $2A - 2A$. Their original proof of Proposition 4.7 used a simple sampling technique combined with tail bounds for a multinomial distribution, which Sanders [76] replaced by the Marcinkiewicz-Zygmund inequality. Both made crucial use of L^p norms, where in applications p would be taken very

large (a function of the density α of a set $A \subseteq G$ under investigation, such as $\log \alpha^{-1}$).

In [11] we gave a different proof of Proposition 4.7 which proceeds entirely without recourse to L^p norms, instead only relying on Chernoff-type tail estimates for sampling.

Lemma 4.8 (Hoeffding bound for sampling)**.** *If X is a random variable with $|X| \leq 1$ and $\hat{\mu}$ is the empirical average obtained from t samples, then*

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

Of course it is well known that L^p bounds and Chernoff inequalities such as the one in Lemma 4.8 are, in a certain sense, equivalent. Specifically, a random variable X obeys a Chernoff-type tail bound of the form

$$\Pr[|X| \geq t\|X\|_2] \leq C\exp(-\Omega(t^2))$$

if and only if its L^p norm satisfies

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

for all $p \in [2, \infty)$, the latter representing a Khinchine-type inequality (from which Marcinkiewicz-Zygmund can be derived). As a consequence, we do not claim that our proof of the Croot-Sisask lemma is entirely new. However, we do believe it writes itself very naturally in the special case of $\mathbb{F}_2^n$, and lends itself more readily to applications in that setting. Moreover, we show how these applications can be "algorithmified" in a straightforward way.

To state our result concisely, we define, given subsets $A, B \subset \mathbb{F}_2^n$ where A is finite, the *measure of additive containment* $\rho_{A\to B} : \mathbb{F}_2^n \to [0,1]$ by

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

for each $y \in \mathbb{F}_2^n$. Notice that $\rho_{A\to B}(y) = 1$ when $y + A \subseteq B$ and $\rho_{A\to B}(y) = 0$ when $(y + A) \cap B = \emptyset$. Finally, for two real numbers α, β, we write $\alpha \approx_\epsilon \beta$ to denote $|\alpha - \beta| \leq \epsilon$ and if $|\alpha - \beta| > \epsilon$ we write $\alpha \not\approx_\epsilon \beta$.

Proposition 4.9 (Ben-Sasson–Ron-Zewi–Tulsiani–Wolf)**.** *If $A \subseteq \mathbb{F}_2^n$ satisfies $|2A| \leq K|A|$, then for every integer t and set $B \subseteq \mathbb{F}_2^n$, there exists a set X with the following properties.*

1. *The set X is contained in an affine shift of A.*
2. *The size of X is at least $|A|/(2K^{t-1})$.*

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

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

We briefly sketch the proof. To obtain X we replace $\rho_{A\to B}$ by an estimator function computed by taking a sequence of t independent random samples distributed uniformly over A. Denoting the sample sequence by $\alpha = (a_1, \ldots, a_t)$, we estimate $\rho_{A\to B}(y)$ by the fraction of $a_i \in \alpha$ satisfying $y + a_i \in B$ and denote the estimator function corresponding to α by $\hat{\rho}_\alpha$. Fixing y, the Chernoff-Hoeffding bound (Lemma 4.8) says that the probability that $\hat{\rho}_\alpha(y)$ differs from $\rho_{A\to B}(y)$ by more than ϵ, i.e., the probability of the event "$\rho_{A\to B}(y) \not\approx_\epsilon \hat{\rho}_\alpha(y)$", is at most $\exp\left(-\Omega\left(\epsilon^2 t\right)\right)$.

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

$$\rho_{A\to B}(y) \approx_\epsilon \hat{\rho}_\alpha(y) \quad \text{and} \quad \rho_{A\to B}(y+x) \approx_\epsilon \hat{\rho}_{\alpha'}(y+x). \tag{6}$$

Then if $\alpha' = x + \alpha$ we have

$$\hat{\rho}_\alpha(y) = \hat{\rho}_\alpha(y+x+x) = \hat{\rho}_{\alpha+x}(y+x) = \hat{\rho}_{\alpha'}(y+x)$$

and combining this with (6) implies that for each "good" y we have $\rho_{A\to B}(y) \approx_{2\epsilon} \rho_{A\to B}(y+x)$. So to prove the lemma we only need to bound from below the number of "special" elements x, which is done based on the assumption that A has small doubling constant.

For applications one would like a version of Proposition 4.9 in which the set X of periods is in fact a subspace. It was observed by Sanders [76] that one can use iterated almost-periodicity statements such as Proposition 4.9, combined with a little Fourier analysis, in order to obtain this property. We follow Sanders's argument to deduce the following statement from Proposition 4.9.

Corollary 4.10 (Ben-Sasson–Ron-Zewi–Tulsiani–Wolf)**.** *If $A \subseteq \mathbb{F}_2^n$ is a subset of density α, then for every integer t and set $B \subseteq \mathbb{F}_2^n$ there exists a subspace V with the following properties.*

1. The size of V is at least $\left(\alpha^t/2\right)^{32} \cdot 2^n$.

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

$$\Pr_{y \in S}\left[\rho_{A\to B}(y) \approx_{\epsilon'} \rho_{A\to B}(y+v)\right] \geq 1 - 16\frac{\ell}{\eta}\frac{|A+B|}{|S|} \cdot \exp\left(-2\epsilon^2 t\right),$$

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

The idea of the proof of Corollary 4.10 is the following. Iterating Proposition 4.9 shows that if $|A+A| \leq K|A|$, then there is a dense set X such that $\rho_{A\to B}(y) \approx_{2\epsilon} \rho_{A\to B}(y+x)$ for almost every $y \in S$ whenever $x \in \ell X$ (as before, ℓX is the ℓ-times iterated sumset of X). Take the set X corresponding to the trivial value $K = 1/\alpha$, and define the subspace V by setting $V = \mathrm{Spec}_{1/2}(X)^{\perp}$. Now if X were a subspace, then $\mathrm{Spec}_{1/2}(X) = V^{\perp}$, and hence $V = X$. Thus V serves as an "approximate subspace" for X. Chang's theorem [29], which tells us that the set of ρ-large Fourier coefficients of a set X of density χ is contained in a subspace of dimension at most $C\rho^{-2}\log\chi^{-1}$, implies that the subspace V is also dense in $\mathbb{F}_2^n$.

We now turn to applications. It was quickly recognized by Sanders [76] that the Croot-Sisask lemma could be used to prove a strong version of the so-called *Bogolyubov-Ruzsa lemma*. In its original form, this states that if $A \subseteq \mathbb{F}_2^n$ is a set of density α, then $2A - 2A$ contains a subspace of codimension at most $2\alpha^{-2}$. It is a crucial ingredient in the proof of Freiman's theorem which describes the structure of sets with small sumsets, and is instrumental in the inverse theorem for the U^3 norm along the lines of Green and Tao.

Passing via a statement similar to Proposition 4.10 above, Sanders showed that $2A - 2A$ contains a subspace whose codimension is polylogarithmic in the density of the set A. This was also a crucial ingredient in his groundbreaking upper bound of $C(\log\log N)^5/\log N$ for the density of a subset of $[N]$ not containing any 3-term arithmetic progressions [77].

Theorem 4.11 (Quasipolynomial Bogloyubov-Ruzsa lemma)**.** *Let $A \subseteq \mathbb{F}_2^n$ be a subset of density α. Then there exists a subspace V of $\mathbb{F}_2^n$ of codimension*

$$\mathrm{codim}(V) = O(\log^4(1/\alpha))$$

with the property that $V \subseteq 4A$.

In [11] we gave an alternative proof of Theorem 4.11, based on Proposition 4.10. Moreover, we develop an algorithmic version which can be tied seamlessly into the algorithm for finding a quadratic decomposition in Section 4.1. This can be done in two ways: either as an improvement to Theorem 4.2 in which the number of true quadratic phases in the decomposition now becomes quasipolynomial in the uniformity parameter, or as an improvement to

Theorem 4.6, leading to a decomposition into polynomially many quadratic averages, each of which now has a description which is quasipolynomial in size. The running time of the algorithms is similarly improved.

Theorem 4.12 (Ben-Sasson–Ron-Zewi–Tulsiani–Wolf)**.** *There exists a randomized algorithm* `Strong-Bogolyubov` *with input parameters α and γ which, given oracle access to a function $h : \mathbb{F}_2^n \to \{0,1\}$ with $\mathbb{E}h \geq \alpha$, outputs a subspace $V \leqslant \mathbb{F}_2^n$ (by giving a basis for $V^\perp$) of codimension at most $O(\log^4(1/\alpha))$ such that with probability at least $1-\gamma$, we have $h * h * h * h(v) > 0$ for each $v \in V$. The algorithm runs in time $2^{O(\log^4(1/\alpha))} \cdot \mathtt{poly}(\log(1/\gamma)) \cdot n^3 \log n$.*

Let us say a couple of words about what needs to be done to make the proof of Theorem 4.11 algorithmic. In the combinatorial proof of Proposition 4.9 we considered the measure $\mu(y) = \mathbb{P}_{a \in A}[y + a \in 2A]$, and the subspace in Theorem 4.11 is defined in terms of a set X which is described using this measure. However, in the setting of the quadratic Goldreich-Levin theorem this measure is difficult to compute since we do not have an unambiguous membership oracle for the set A (see the discussion around Proposition 4.5). Instead, we note that $y + a \in 2A$ is equivalent to saying that $1_A * 1_A(a + y) > 0$, where the latter formulation is robust, so we can test whether $1_A * 1_A(a + y) \geq \eta\alpha^2$ for some $\eta > 0$.

Our second application concerns finding large subspaces within sumsets of a dense set. Inspired by the question whether dense subsets of $\{1, \dots, N\}$ contain long arithmetic progressions, which had received extensive coverage in the literature [27, 46, 79], Ben Green asked the corresponding question in the finite field setting and showed in [44] that if $A \subseteq \mathbb{F}_2^n$ has density α, then $A + A$ contains a subspace of dimension $\Omega(\alpha^2 n)$. Sanders proved in [80], using a Fourier-iteration lemma, that this subspace must be of dimension at least $\Omega(\alpha n)$, and remarked that a bound of roughly similar strength follows implicitly from the techniques of Croot-Łaba-Sisask [33].

In [11] we give a simplified version of this latter bound, avoiding Fourier analysis and using instead a slightly modified version of our sampling approach to Croot-Sisask almost-periodicity (the variance needs to be taken into account in Hoeffding's bound).

4.3. Discussion and outlook. Further quantitative improvements in the decomposition algorithm, and more specifically in the quadratic self-corrector (Theorem 4.3) and the closely connected U^3 inverse theorem (Theorem 2.1) appear difficult, but should not be ruled out given Sanders's recent breakthrough (Theorem 4.11).

The obvious question about higher-order analogues of Theorem 4.3, for example explicit decompositions into cubic phases, relies on progress on a combinatorial/analytic proof of the

U^4 inverse theorem in the finite field setting, a question which was already mentioned in the closing remarks of Chapter 2. Testers for cubic and higher-degree polynomials, based on the uniformity norms, were developed by Alon et al. [14] and Bhattacharyya et al. [22], but they do not even work up to the unique decoding radius (let alone beyond the list-decoding radius). In the language of arithmetic combinatorics, these results correspond to inverse theorems in which the U^k norm is assumed to be very large (close to 1). There is some hope that the very concrete understanding of the behaviour of low-degree polynomials which has been developed in the theoretical computer science community in recent years will accelerate progress on this problem.

It also seems highly likely that the "boosting" method described in Section 4.1 has further applications in arithmetic combinatorics, in places where a Hahn-Banach-type or energy increment argument is traditionally used. An important example is the so-called *triangle removal lemma*, which has immediate implications for graph testing. It states that for any $\epsilon > 0$, there exists $\delta > 0$ with the property that in any graph G on n vertices with fewer than δn^3 triangles, one can remove fewer than ϵn^2 edges to make it triangle-free. Despite a recent improvement by Fox [36], the bounds are of tower-type (while the conjecture is that δ depends polynomially on ϵ), and the problem is widely considered to be very difficult.

So far I have described a number of different ways of obtaining a quadratic decomposition theorem, abstractly or algorithmically. In either case, there seems to be no way of telling or stipulating in advance which quadratic phases will be used in the decomposition. A judiciously (perhaps randomly) chosen subset of all quadratic phases ω^q for a quadratic form q should form a *frame* (a weakened notion of basis). Each coefficients should be computable by algorithms already available in this relatively well-developed branch of harmonic analysis. This approach may offer an alternative to higher-order decomposition theorems, in which would one would get an approximate Parseval identity for free. I intend to pursue this together with Pablo Candela during his postdoctoral stay in Paris.

In the language of coding theory, one can ask a closely related question: for what distributions D over the codewords of the Reed-Muller code of order 2 within a ball of radius r from a given function f is it possible to sample a random codeword according to D? For example, what (if any) effect does the rank of the quadratic polynomial have on the probability of sampling it?

5 Related combinatorial results

The early chapters of this thesis dealt with analytic techniques for obtaining upper bounds on the density of sets that lack a given arithmetic structure, such as an arithmetic progression of length k or more generally solutions to a given system of linear equations. I have also considered lower bounds for this problem in various settings.

Let me start by describing the simplest case in the integers: how large a subset of the interval $\{1, \dots, N\}$ can we construct, while avoiding all non-trivial 3-term progressions? For over 60 years, the best answer to this question has been given by Behrend [18].

Theorem 5.1 (Behrend example)**.** *There exists a subset $A \subseteq [N]$ of size*

$$|A| \gg \exp(-c\sqrt{\log N})N$$

containing no 3-term arithmetic progressions.

We sketch the argument and leave the reader to fill in the details. The construction is geometric in nature: consider the d-dimensional integer grid $[m]^d$ with parameters m and d, to be optimized later. By a simple averaging argument we show that at least one of the spheres

$$S_r = \{x \in [m]^d : x_1^2 + \cdots + x_d^2 = r\}$$

for radii $r = 1, 2, \dots, dm^2$ must contain a lot of integer points. Indeed, since every point in $[m]^d$ lies on one of these spheres, there must exist $r \in [dm^2]$ such that S_r has size at least m^d/dm^2.

Clearly a sphere such as S_r contains no three points on a line, and therefore certainly no 3-term arithmetic progressions. So we have constructed a large set which is 3-term progression-free, except that unfortunately it is not a subset of the integers. We get around this problem by projecting S_r into the interval $[N]$ while preserving the property that it is 3-AP free.

To be more concrete, let Φ be the map $x \mapsto \sum_{i=1}^{d} x_i(2m)^{d+1-i}$, thereby considering x as the coefficients of the base $2m$ expansion of $\Phi(x)$. It is straightforward to check that Φ preserves the property that S_r is progression-free. The range of Φ is $[(2m)^d]$, and in order to maximize the cardinality of $A = \Phi(S_r)$ we set $d \sim \sqrt{\log N}$ and $m \sim N^{1/\sqrt{\log N}}$, leading to a lower bound on A of the form

$$m^d/dm^2 \sim \exp(-c\sqrt{\log N})N.$$

This construction was extended to longer progressions by Rankin [70], and recently rediscovered by Lacey and Łaba [63]. They showed that there exist k-term progression free sets of size at least $\Omega(N\exp(-c(\log N)^{1/K+1}))$, where $k = 2^K + 1$, thereby providing a lower bound on Szemerédi's theorem for long progressions.

It should not come as a surprise that this question can also be asked in the context of finite fields: can we construct large subsets of $\mathbb{F}_3^n$ which do not contain any 3-term progressions? The best construction in this setting is due to Edel [34], who showed that there exists $A \subseteq \mathbb{F}_3^n$ of size $\Omega(N^{0.7249})$. It is an important and wide open problem to determine whether or not a 3-term progression free set can be as large as $(3 - o(1))^n$.

In Section 5.1 we describe a refinement of Behrend's basic construction which was obtained in collaboration with Ben Green [4]. The method combines Behrend's geometric idea with the probabilistic method. In Section 5.2 I summarize work done in collaboration with Yuncheng Lin, whom I mentored during the MIT Summer Program for Undergraduate Research (SPUR) in the summer of 2008 [5]. The method of proof is entirely algebraic in nature.

5.1. An improvement of Behrend's construction. Write $r_3([N])$ for the cardinality of the largest subset of $[N] = \{1, \dots, N\}$ not containing three distinct elements in arithmetic progression. The above-mentioned construction of Behrend (Theorem 5.1) shows, when analysed carefully, that

$$r_3(N) \gg \frac{1}{\log^{1/4} N} \cdot \frac{N}{2^{2\sqrt{2}\sqrt{\log_2 N}}}.$$

In 2008 Elkin [35] was able to improve this then 62-year old bound by a factor of $\log^{1/2} N$. Elkin's argument was nearly 40 pages long and rather involved. Together with Ben Green, I gave a simple 3-page probabilistic proof in [4].

Theorem 5.2 (Elkin, Green-Wolf)**.**

$$r_3(N) \gg \log^{1/4} N \cdot \frac{N}{2^{2\sqrt{2}\sqrt{\log_2 N}}}.$$

Since it is so short, we are able to briefly sketch the argument. Instead of considering the surface of a sphere, we consider a slightly thickened annulus. This no longer has the property that it contains no 3-points on a line, but for appropriately chosen parameters it won't contain too many. By a standard concentration result, there must be an annulus with many grid points in it. We project randomly this time, and destroy the few remaining progressions by removing one point from each, which doesn't decrease the density by too much.

To make this more precise, let us identify $\mathbb{T}^d$ with $[0,1)^d$ in the obvious way. For each $r \leq \frac{1}{2}\sqrt{d}$, write $S(r)$ for the region

$$\{x \in [0,1/2]^d : r - \delta \leq \|x\|_2 \leq r\},$$

where δ will be chosen later.

Lemma 5.3. *There is some choice of r for which* $\mathrm{vol}(S(r)) \geq c\delta 2^{-d}$.

The lemma follows immediately from the pigeonhole principle, together with the fact that if $(x_1, \dots, x_d)$ is chosen at random from $[0,1/2]^d$ then, with probability at least c, we have $|\|x\|_2 - \sqrt{d/12}| \leq C$. This is a consequence of standard tail estimates for sums of independent identically distributed random variables, of which $\|x\|_2^2 = \sum_{i=1}^d x_i^2$ is an example.

Since there is no "wraparound", we can regard S as a subset of $\mathbb{R}^d$ for the purpose of counting 3-term progressions in S. We next show that S doesn't contain too many of these. Indeed, suppose that (x,y) is a pair for which $x-y, x$ and $x+y$ lie in S. By the parallelogram law we have

$$2\|x\|_2^2 + 2\|y\|_2^2 = \|x+y\|_2^2 + \|x-y\|_2^2,$$

and straightforward algebra gives

$$\|y\|_2 \leq \sqrt{r^2 - (r-\delta)^2} \leq \sqrt{2\delta r}.$$

It follows from the formula for the volume of a sphere in $\mathbb{R}^d$ that the volume of the set $B \subseteq \mathbb{T}^d \times \mathbb{T}^d$ in which each such pair (x,y) must lie is at most $\mathrm{vol}(S)C^d(\delta/\sqrt{d})^{d/2}$.

Finally, we consider the pullback to the integers. It is easy to show the following.

Lemma 5.4. *Suppose that N is even. Define $A_{\theta,\alpha} := \{n \in [N] : \theta n + \alpha \mod 1 \in S\}$. Then*

$$\mathbb{E}_{\theta,\alpha}|A_{\theta,\alpha}| = N\mathrm{vol}(S)$$

whilst the expected number of nontrivial 3-term arithmetic progressions in $A_{\theta,\alpha}$ is

$$\mathbb{E}_{\theta,\alpha}T_3(A_{\theta,\alpha}) = \frac{1}{4}N(N-5)\mathrm{vol}(B),$$

where $T_3(A)$ denotes the number of 3-term progressions in A.

Putting everything together, we want to choose our parameters so that

$$\frac{1}{3}\mathrm{vol}(S) \geq \frac{1}{4}(N-5)\mathrm{vol}(B),$$

for in that case we have

$$\mathbb{E}\big(\frac{2}{3}|A_{\theta,\alpha}| - T(A_{\theta,\alpha})\big) \geq \frac{1}{3}N\mathrm{vol}(S).$$

In particular there is a specific choice of $A := A_{\theta,\alpha}$ for which both $T(A) \leq 2|A|/3$ and $|A| \geq \frac{1}{2}N\mathrm{vol}(S)$. Deleting up to two thirds of the elements of A, we are left with a set of size at least $\frac{1}{6}N\mathrm{vol}(S)$ that is free of 3-term arithmetic progressions.

To do this it suffices to have $C^d(\delta/\sqrt{d})^{d/2} \leq c/N$, which can certainly be achieved by taking $\delta := c\sqrt{d}N^{-2/d}$. For this choice of parameters we have, by the earlier lower bound on $\mathrm{vol}(S)$, that

$$|A| \geq \frac{1}{6}N\mathrm{vol}(S) \geq c\sqrt{d}2^{-d}N^{1-2/d}.$$

Choosing $d := \lceil \sqrt{2\log_2 N} \rceil$ we recover Elkin's bound.

This construction was extended to longer progressions by O'Bryant [67]. Further progress on this problem seems extremely difficult.

5.2. A finite field Behrend-type construction for longer progressions. In the finite field setting the geometric intuition which we so heavily exploited in the preceding section is no longer available. Indeed, Edel's lower bound on $r_3(\mathbb{F}_3^n)$ which we mentioned in the introduction is of a fundamentally different flavour, and relies crucially on small algebraic examples from coding theory.

Moving from small examples to very large ones is simple in this context: note that if A and B are subsets of $\mathbb{F}_q^m$ and $\mathbb{F}_q^n$, respectively, containing no k-term arithmetic progression, then their Cartesian product $A \times B := \{(a,b) : a \in A, b \in B\}$ is a subset of $\mathbb{F}_q^{m+n}$ that contains no k-term arithmetic progression. More sophisticated product constructions are known, and indeed are crucial to Edel's bound.

This problem has also been studied in the case where the characteristic of the underlying field is greater than 3. Bierbrauer [23] showed that for any finite field $\mathbb{F}_q$ with at least 3 elements, there exists a subset of $\mathbb{F}_q^3$ with q^2 elements that does not contain 3 points on a line.

By taking a Cartesian product of this set with itself sufficiently many times, it is clear that for any prime power $q \geq 3$ and any n divisible by 3, we have

$$r_3(\mathbb{F}_q^n) = \Omega((q^2)^{n/3}).$$

This can be improved: it is shown in [23] that there exists a subset of $\mathbb{F}_q^6$ of size $q^4 + q^2 - 1$

that does not contain 3 points on a line, and hence that for any prime power $q \geq 3$ and any n divisible by 6, we have

$$r_3(\mathbb{F}_q^n) = \Omega((q^4 + q^2 - 1)^{n/6}). \tag{7}$$

With my student Yuncheng Lin I set out to construct large subsets of $\mathbb{F}_q^n$ that contain no k-term arithmetic progressions for $q \geq k > 3$. Generalizing Bierbrauer's method, our main result is the following [5].

Theorem 5.5 (Lin-Wolf)**.** *Let k be a positive integer. Let $\mathbb{F}_q$ be the finite field of q elements such that $q \geq k$. For any n divisible by $2k$, we have*

$$r_k(\mathbb{F}_q^n) = \Omega((q^{2(k-1)} + q^{k-1} - 1)^{n/2k}).$$

It represents the first construction of a large k-term progression free subset of $\mathbb{F}_q^n$ (and to this date, as far as I am aware, the only one). To get a feel for the strength of this bound I will discuss some background here. First note that by a brute force argument, it is not hard to see that we have

$$r_4(\mathbb{F}_5^2) = 11,$$

which implies that $r_4(\mathbb{F}_5^n) = \Omega(N^{\log 11/(2 \log 5)}) = \Omega(N^{0.7449})$. For longer progressions, examples even in small dimension are hard to come by. As a special case of Proposition 5.6 below we first prove that there exists a subset of $\mathbb{F}_5^4$ of size 125 which does not contain any 4-term arithmetic progressions.

Proposition 5.6. *Let k be positive integer. Let $\mathbb{F}_q$ be the finite field of q elements such that $q \geq k$. Then there is a subset of $\mathbb{F}_q^k$ of size q^{k-1} containing no k points on a line, and hence no k-term arithmetic progressions.*

By taking the product of the set thus constructed with itself many times, we obtain the following corollary.

Corollary 5.7. *Let k be a positive integer and $\mathbb{F}_q$ be the finite field with q elements. Whenever $q \geq k$ and n is divisible by k, we have*

$$r_k(\mathbb{F}_q^n) = \Omega((q^{k-1})^{n/k}) = \Omega(N^{1-1/k}).$$

In particular, $r_4(\mathbb{F}_5^n) = \Omega(N^{\log 125/(4 \log 5)}) = \Omega(N^{0.75})$. But we can do better.

Theorem 5.8. *Let k be a positive integer. Let $\mathbb{F}_q$ be the finite field of q elements such that $q \geq k$. Then there is a subset of $\mathbb{F}_q^{2k}$ of size $q^{2(k-1)} + q^{k-1} - 1$ that contains no k points on a line, and hence no k-term arithmetic progression.*

Taking Cartesian products gives the statement of Theorem 5.5. For fixed q and k, this result improves upon the more trivial lower bound of $\Omega(N^{1-1/k})$ obtained in Corollary 5.7 asymptotically, where $N = q^n$. In particular, we find that $r_4(\mathbb{F}_5^n) = \Omega(N^{\log 15749/(8\log 5)}) = \Omega(N^{0.7506})$, an improvement in the fourth digit of the exponent!

5.3. Discussion and outlook. Lower bound problems appear to be extremely difficult, and it is quite possible in the case of Behrend that the known construction is essentially optimal. I suspect that Edel's construction in the finite field setting can be improved (indeed, a tiny improvement was suggested by Yunchen Lin while we were working on [5]). An interesting question is whether the type of product construction in [34] can be adapted to longer progressions. The analogous problem in a ring such as $\mathbb{Z}/4\mathbb{Z}$ may be more tractable. An upper bound for $r_3((\mathbb{Z}/4\mathbb{Z})^n)$ was obtained by Sanders [78], and is of interest since it beats the $O(N(\log N)^{-1})$ bound which in the case of a field is considered to be the limit of the Fourier analytic method. A computer search in [78] revealed that $r_3((\mathbb{Z}/4\mathbb{Z})^3) = 16$, which gives rise to a lower bound on $r_3((\mathbb{Z}/4\mathbb{Z})^n)$ in the usual way. Can Theorem 5.8 be adapted to $\mathbb{Z}/4\mathbb{Z}$? It is also worth noting that the progression-free sets in Theorem 5.8 as well as those in [18], [70], [67], [34] obey the stronger property that they contain no k points on a line. Is there any construction that actually uses the fact that the set contains no *progressions*?

Another lower bound that is notoriously difficult to improve upon is that for Sárközy's theorem on square differences. Sárközy's theorem states that any sufficiently dense subset of $\{1, \dots, N\}$ contains two elements x, y such that $x - y = d^2$ for some integer d, and the best known lower bound is due to Ruzsa [73]. Upper bounds are also available for more general polynomial differences such as $d^2 - 1$, but as far as I am aware no lower bounds are known in this case, and Ruzsa's construction for squares does not seem to generalize in a straightforward way.

Looking further afield, there is a close connection between the notions of uniformity in subsets of $\mathbb{Z}_N$ and that in graphs and hypergraphs. Indeed, Gowers's definition of the U^2 norm is inspired by the count of of C_4s in graphs, which is a well-known measure of quasirandomness that goes back to Thomason [89] and Chung et al. [30]. Let us give an example of an interesting direction of research in this context.

It was conjectured by Erdős that the 2-colouring of the edges of a complete graph K_n on n vertices with the smallest number of monochromatic K_4s is the random colouring. This was disproved by Thomason [90], who exhibited colourings that beat the random number (the earliest of which was, interestingly but rather mysteriously, defined using a quadratic form on $\mathbb{F}_2^n$). On the other hand, Goodman gave an ingenious combinatorial argument that led

to a lower bound on the number of monochromatic K_4s in any 2-colouring. In my doctoral thesis I investigated the analogous problem for $\mathbb{Z}/p\mathbb{Z}$ [3]. What is the minimum number of monochromatic 4-term progressions in any 2-colouring of $\mathbb{Z}/p\mathbb{Z}$? Improving an argument by Cameron, Cilleruelo and Serra I gave a combinatorial lower bound, and using a previously known construction by Gowers I was able to exhibit a 2-colouring that beats the random case. This colouring is based on a set which is uniform but not quadratically uniform in the sense described earlier.

There are several other known examples of graphs (apart from K_4) that may occur in less than the expected number in a 2-colouring of K_n. Is this phenomenon the result of a similar disparity between different degrees of uniformity, as it can be shown to be in sets? This cannnot be true by trivial analogy as a quasirandom graph contains all small subgraphs in the right order. It would be very interesting to be able to explain Thomason's ad-hoc and computer-aided constructions by providing a more unified theory of uniformity. Even a well-founded conjecture regarding what I would call the *true colouring complexity* in graphs, which takes care of all known examples, would be a most satisfying result from my point of view (the original conjecture by Burr and Rosta was disproved in 1989).

Remerciements

Tout d'abord je voudrais remercier les rapporteurs d'avoir rendu possible la soutenance de cette thèse, et les membres du jury d'être présents le jour de ma soutenance. Je suis consciente que pour les premiers cela a répresenté un travail considérable, et que pour les derniers ce n'était pas facile de se libérer dans un délai aussi court. Merci !

Je suis très reconnaissante à mes collègues à l'École polytechnique de m'avoir accueillie aussi chaleureusement à mon arrivée en France : Frank Pacard, François Golse, mes frères Hadamard Philippe Gravejat, Julien Marché et Romain Dujardin, ainsi que Karine Beauchard, Gaëtan Chenevier, Charles Favre, Fabrice Orgogozo, et surtout Michèle Lavallette pour son soutien infatigable. Je tiens également à remercier Alain Plagne et toute l'équipe CAESAR d'avoir créé une plateforme pour la combinatoire additive en France, y compris Éric Balandraud, Benjamin Girard, Wolfgang Schmid, Emmanuel Breuillard et Harald Helfgott. Je suis très reconnaissante à Étienne Fouvry, Bernard Host et Régis de la Bretèche de m'avoir poussée à rédiger ce mémoire cette année, et à Nessim Sibony, Emmanuel Breuillard et Valérie Lavigne pour leur soutien administratif à Orsay. Enfin, un grand merci à Eric Balandraud pour avoir relu la partie française de ce mémoire.

My warmest thanks goes to all my teachers, collaborators and students. Above all, I would like to thank Tim Gowers for his inspiration, and the time and energy he invested in our joint projects. I am deeply appreciative of all the support that Ben Green, Mel Nathanson and Endre Szemerédi have given me over the years, as well as the challenges they have put in my way. I am indebted to Terry Tao for making my head spin on numerous occasions, and for ensuring that arithmetic combinatorics remains a field that I am excited to be working in. I thank Avi Wigderson for getting me interested in theoretical computer science and introducing me to Madhur Tulsiani, who is a great pleasure to work with and has taught me an enormous amount. I am grateful to Harald Helfgott and Tom Sanders for stimulating mathematical discussions in Paris and on Skype, respectively, to be continued. Finally, I have enjoyed and learnt a lot from teaching and working with Yuncheng Lin, Cordelia Link and Eric Naslund.

I am grateful to my family and my friends, near and far, for entertaining, supporting and grounding me in this world, and to Ruben Portugues for always being by my side.

Références

Mes publications.

[1] Julia Wolf. The structure of popular difference sets. *Israel J. Math.*, 179:253–278, 2010.

[2] Timothy Gowers and Julia Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010.

[3] Julia Wolf. The minimum number of monochromatic 4-term progressions in $\mathbb{Z}_p$. *J. Comb.*, 1(1):53–68, 2010.

[4] Ben Green and Julia Wolf. A note on Elkin's improvement of Behrend's construction. In *Additive number theory*, pages 141–144. Springer, New York, 2010.

[5] Yuncheng Lin and Julia Wolf. On subsets of $\mathbb{F}_q^n$ containing no k-term progressions. *European J. Combin.*, 31(5):1398–1403, 2010.

[6] Timothy Gowers and Julia Wolf. Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$. *Mathematika*, 57(2):215–237, 2012.

[7] Timothy Gowers and Julia Wolf. Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$. *J. Anal. Math.*, 115:121–186, 2011.

[8] Timothy Gowers and Julia Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, 21(1):36–69, 2011.

[9] Madhur Tulsiani and Julia Wolf. Quadratic Goldreich-Levin Theorems. *IEEE 52nd Annual Symposium on Foundations of Computer Science FOCS 2011*, 619–628, 2011. Long version submitted by invitation to special issue of SIAM J. Comput., available at *arXiv:1105.4372*, 37 pages, 2012.

[10] Julia Wolf. Arithmetic and polynomial progressions in the primes, d'apres Gowers, Green, Tao and Ziegler. To appear in *Astérisque*, 37 pages, 2012.

[11] Eli Ben-Sasson, Noga Ron-Zewi, Madhur Tulsiani and Julia Wolf. Sampling-based proofs of almost-periodicity results and applications. Submitted, available at *arXiv:1210.6917*, 28 pages, 2012.

[12] Thái Hoàng Lê and Julia Wolf. Polynomial configurations in the primes. Submitted, available at *arXiv:1210.4659*, 22 pages, 2012.

En préparation.

[13] Julia Wolf. An introduction to arithmetic combinatorics. *Book draft*, 160 pages, 2012.

Autres références.

[14] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Low-Degree Polynomials over GF(2). In *Approximation, randomization, and combinatorial optimization*, volume 2764 of LNCS, pp. 188–199. Springer-Verlag, Berlin, 2003.

[15] Guillaume Aubrun and Ion Nechita. The multiplicative property characterizes ℓ_p and L_p norms. *arXiv:1102.2618v1*, 2011.

[16] Guillaume Aubrun and Ion Nechita. The multiplicative property characterizes ℓ_p and L_p norms. *arXiv:1102.2618v1*, 2011.

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

[18] F. A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. U. S. A.*, 32:331–332, 1946.

[19] Vitaly Bergelson and Alexander Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[20] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010.

[21] Vitaly Bergelson, Bernard Host and Bryna Kra. Multiple recurrence and nilsequences. *Invent. Math.*, 160(2):261–303, 2005.

[22] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan and David Zuckerman. Optimal Testing of Reed-Muller Codes. 6390:269–275, 2011.

[23] J. Bierbrauer. Large caps. *J. Geom.*, 76:16–51, 2003.

[24] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990)*, pages 549–595, 1993.

[25] Andrej Bogdanov and Emanuele Viola. Pseudorandom Bits for Polynomials. In *Foundations of Computer Science, 2007. FOCS '07. 48th Annual IEEE Symposium on*, pages 41–51, 2007.

[26] Jean Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9(5), 968–984, 1999.

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

[28] Pablo Candela. On the structure of steps of three-term arithmetic progressions in a

dense set of integers. *Bull. Lond. Math. Soc.*, 42(1):1–14, 2010.

[29] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke Math. J.*, 113(3):399–419, 2002.

[30] F.R.K. Chung, R. Graham and R.M. Wilson. Quasirandom graphs. *Proceedings of the National Academy of Sciences of the United States of America*, 85(4), 969–970, 1988.

[31] David Conlon and Timothy Gowers. Combinatorial theorems in sparse random sets. *arXiv:1011.4310*, 2010.

[32] Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geom. Funct. Anal.*, 20(6):1367–1396, 2010.

[33] Ernie Croot, Izabella Łaba, and Olof Sisask. Arithmetic progressions in sumsets and L^p-almost-periodicity. *arXiv:1103.6000v1*, 2011.

[34] Yves Edel. Extensions of generalized product caps. *Des. Codes Cryptogr.*, 31(1):5–14, 2004.

[35] Michael Elkin. An improved construction of progression-free sets. *Israel J. Math.*, 184:93–128, 2011.

[36] Jacob Fox. A new proof of the graph removal lemma. *Ann. Math. (2)*, 174(1):561–579, 2011.

[37] Nikos Frantzikinakis, Bernard Host and Bryna Kra. The polynomial multidimensional Szemeredi theorem along shifted primes. To appear in *Israel. J. Math.*, 2012.

[38] Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[39] Hillel Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.*, 31:204–256, 1977.

[40] O. Goldreich and L.A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 25–32, New York, NY, USA, 1989. ACM.

[41] Timothy Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

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

[43] Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010.

[44] Ben Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.

[45] Ben Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*,

pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[46] Ben Green. Arithmetic progressions in sumsets. *Geom. Funct. Anal.*, 12(3):584–597, 2002.

[47] Ben Green. Roth's theorem in the primes. *Ann. of Math. (2)*, 161(3):1609–1636, 2005.

[48] Ben Green. On arithmetic structures in dense sets of integers. *Duke Math. J.*, 114(2):215–238, 2002.

[49] Ben Green and Terence Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc. (2)*, 51(1):73–153, 2008.

[50] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[51] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[52] B. Green and Terence Tao. The Möbius function is strongly orthogonal to nilsequences. *To appear, Ann. Math.*, 2012.

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

[54] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009.

[55] Ben Green and Terence Tao. An arithmetic regularity lemma, an associated counting lemma, and applications. *An Irregular Mind*, pages 261–334, 2010.

[56] Ben Green and Terence Tao. New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$. In *Analytic number theory*, pages 180–204. Cambridge Univ. Press, Cambridge, 2009.

[57] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *To appear, Ann. Math.*, 2010.

[58] Mariah Hamel and Izabella Łaba. Arithmetic structures in random sets. *Integers*, 8:A04–21, 2008.

[59] Hamed Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*,175(1):125–150, 2010.

[60] Harald Helfgott and Akshay Venkatesh. How small must ill-distributed sets be? In *Analytic number theory*, pages 224–234. Cambridge Univ. Press, Cambridge, 2009.

[61] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[62] Yoshiharu Kohayakawa, Tomasz Łuczak, and Vojtěch Rödl. Arithmetic progressions of length three in subsets of a random set. *Acta Arith.*, 75(2):133–163, 1996.

[63] Michael Lacey and Izabella Łaba. On sets of integers not containing long arithmetic progressions. *Available at http://www.math.ubc.ca/∼ilaba/preprints*, 2001.

[64] Sasha Leibman. Orbit of the diagonal in the power of a nilmanifold. *Trans. AMS*, 362(3):1619–1658, 2010.

[65] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *STOC'08*, pages 547–556. ACM, New York, 2008.

[66] Eric Naslund. A note on arithmetic progressions of length 3 with polynomial common difference. *In preparation*, 2012.

[67] Kevin O'Bryant. Sets of integers that do not contain long arithmetic progressions. *Electron. J. Combin.*, 18(1):P59, 2011.

[68] R.W.K. Odoni. The distribution of integral and prime-integral values of systems of full-norm polynomials and affine-decomposable polynomials. *Mathematika*, 26(1):80–87, 1979.

[69] János Pintz, W. L. Steiger, and Endre Szemerédi. On sets of natural numbers whose difference set contains no squares. *J. Lond. Math. Soc. (2)*, 37(2):219–231, 1988.

[70] R. A. Rankin. Sets of integers containing not more than a given number of terms in arithmetical progression. *Proc. Roy. Soc. Edinburgh Sect. A*, 65:332–344, 1960.

[71] Omer Reingold, Luca Trevisan, Madhur Tulsiani and Salil Vadhan. Dense Subsets of Pseudorandom Sets. In *Foundations of Computer Science, 2007. FOCS '07. 48th Annual IEEE Symposium on*, pages 76–85, 2008.

[72] Imre Ruzsa. An analog of Freiman's theorem in groups. *Astérisque*, (258):xv, 323–326, 1999.

[73] Imre Ruzsa. Difference sets without squares. *Period. Math. Hungar.*, 15(3):205–209, 1984.

[74] Alex Samorodnitsky. Low-degree tests at large distances. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 506–515. ACM, New York, 2007.

[75] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. *SIAM J. Comput.*, 39(1):323–360, 2009.

[76] Tom Sanders. On the Bogolyubov-Ruzsa lemma. *To appear, Anal. PDE*, November 2010.

[77] Tom Sanders. On Roth's theorem on progressions. *Ann. of Math. (2)*, 174(1):619–636, 2011.

[78] Tom Sanders. Roth's Theorem in $(\mathbb{Z}/4\mathbb{Z})^n$. *Anal. PDE 2, no. 2, 211–234*, 2009.

[79] Tom Sanders. Additive structures in sumsets. *Math. Proc. Cambridge Philos. Soc.*, 144(2):289–316, 2008.

[80] Tom Sanders. Green's sumset problem at density one half. *Acta Arith.*, 146(1):91–101, 2011.

[81] András Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978.

[82] Tomasz Schoen. Near optimal bounds in Freiman's theorem. *Duke Math. J.*, 158(1):1–12, May 2011.

[83] Balazs Szegedy. Higher order Fourier analysis as an algebraic theory I. *arXiv:0903.0897*, 2009.

[84] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[85] Terence Tao and Van Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[86] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.*, 201(2):213–305, 2008.

[87] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Anal. PDE*, 3(1):1–20, 2010.

[88] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. To appear in *Ann. Comb.*, 2011.

[89] Andrew Thomason. Pseudo-random graphs. *Proceedings of Random Graphs, Pozn?an 1985* (M. Karonski, ed.), Annals of Discrete Mathematics 33, 307–331, 1985.

[90] Andrew Thomason. A disproof of a conjecture of Erdős in Ramsey theory. *J. Lond. Math. Soc. (2)*, 39(2):246–255, 1989.

[91] Luca Trevisan, Madhur Tulsiani and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 126–136. IEEE, 2009.

[92] Emanuele Viola and Avi Wigderson. Norms, XOR Lemmas, and Lower Bounds for GF(2) Polynomials and Multiparty Protocols. In *Computational Complexity, 2007. CCC '07. Twenty-Second Annual IEEE Conference on*, pages 141–154, 2007.

[93] Trevor Wooley and Tamar Ziegler. Multiple recurrence and convergence along the primes. To appear in *American J. of Math.*, 2012.