Topics in and around higher-order Fourier analysis



Daniel Altman

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Dedication

To my family.

And to Vicky Neale, who was always very kind to me.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Ben Green. Ben is exceptionally generous with his time and thoughts, and I am very fortunate to have received more than my fair share of both.

I am thankful for many discussions with my friend Zachary Chase during our time in Oxford, and for a nice collaboration leading to [7]. I would also like to thank the other friends, students, postdocs and faculty who have enriched the environment under which I have completed this thesis.

Furthermore, the author would like to thank: Freddie Manners for helpful conversations and valuable feedback pertaining to Chapter 3 and [2]; Terence Tao for helpful comments on an earlier version of Chapter 4 and [5]; Anita Liebenau for helpful comments on an earlier version of Chapter 7 and [4]; Ben Green for helpful conversations pertaining to Chapters 3, 4, 5 and to [2, 3, 5], for valuable feedback on earlier versions of Chapters 6, 7 and [4, 6], and for drawing his attention to the study of the local Sidorenko property in the graph-theoretic setting (in particular [55]). Finally, the author would also like to thank Julia Wolf for a very careful reading of this thesis and for lots of valuable feedback.

Abstract

We present some results in both the theory and application of higher-order Fourier analysis.

In Chapters 1 and 2 we give background and context to the study of higher-order Fourier analysis, and to the questions and problems addressed in this thesis.

Gowers and Wolf conjectured that, given a set of linear forms $\{\psi_i\}_{i=1}^t$ each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$, if s is an integer such that the functions $\psi_i^{s+1}, \dots, \psi_t^{s+1}$ are linearly independent, then averages of the form $\mathbb{E}_{\boldsymbol{x}} \prod_{i=1}^t f(\psi_i(\boldsymbol{x}))$ may be controlled by the Gowers U^{s+1}-norm of f. In Chapter 3 we prove (a stronger version of) this conjecture. This chapter appears in the literature as [2].

Green and Tao's arithmetic regularity lemma and counting lemma together apply to systems of linear forms which satisfy a particular algebraic criterion known as the 'flag condition'. In Chapter 4 we give an arithmetic regularity lemma and counting lemma which applies to all systems of linear forms and provide a couple of applications. This chapter appears on the arXiv as [5].

In Chapter 5 we consider Szemerédi's theorem on k-term arithmetic progressions where the set S of allowed common differences in those progressions is restricted to a set which is chosen randomly. We show that the threshold over finite fields is different to the conjectured threshold over the integers. This chapter appears in the literature as [3].

A system of linear equations in $\mathbb{F}_p^n$ is *Sidorenko* if any subset of $\mathbb{F}_p^n$ contains at least as many solutions to the system as a random set of the same density, asymptotically as $n \to \infty$. A system of linear equations in $\mathbb{F}_p^n$ is *common* if every two-colouring of $\mathbb{F}_p^n$ yields at least as many monochromatic solutions as a random two-colouring, asymptotically as $n \to \infty$. By analogy to the graph-theoretic setting, Alon has asked whether any (non-Sidorenko) system of linear equations can be made uncommon by adding sufficiently many free variables. Fox, Pham and Zhao answered this question in the affirmative among systems which consist of a single equation. In Chapter 6 we answer Alon's question in the negative. Here we also observe that the property of remaining common despite that addition of arbitrarily many free variables is closely related to a notion of commonness in which one replaces the arithmetic mean of the number of monochromatic solutions with the geometric mean, and furthermore resolve questions of Kamčev–Liebenau–Morrison. This chapter appears on the arXiv as [6].

In Chapter 7 we show that a certain generic family of systems of two linear equations is not Sidorenko. In fact, we show that systems in this family are not locally Sidorenko, and that systems in this family which do not contain additive tuples are not weakly locally Sidorenko. This endeavour answers a conjecture and question of Kamčev–Liebenau–Morrison. Insofar as methods are concerned, we observe that the true complexity of a linear system is not maintained under Fourier inversion; our main novelty is the use of higher-order methods in the frequency space of systems which have complexity one. We also give a shorter proof of the recent result of Kamčev–Liebenau–Morrison and independently Versteegen that any linear system containing a four-term arithmetic progression is uncommon. This chapter appears on the arXiv as [4].

Contents

1 Introduction **1**

1.1 Context 1

1.2 Outline 3

2 Background **5**

2.1 Notation 5

2.2 The Fourier transform 5

2.3 Gowers norms 6

2.4 Nilsequences 8

2.5 An example 10

3 On a conjecture of Gowers and Wolf **13**

3.1 Introduction 13

3.2 A strengthened result in the flag case 15

3.3 Proof of Theorem 3.1.1 23

3.4 Appendix: Results from higher order Fourier analysis 24

4 A non-flag arithmetic regularity lemma and counting lemma **27**

4.1 Introduction 27

4.2 Examples, discussion of methods and qualitative results 32

4.3 Description of $\mathfrak{g}^{\Psi}(S_\bullet)$, comparison with $\log G^{\Psi}$ 46

4.4 The counting lemma 48

4.5 The factorisation theorem for linear irrationality 54

4.6 The strongly irrational arithmetic regularity lemma 62

4.7 Applications 68

4.8 Appendix A: Details on polynomial sequences in Lie groups and Lie algebras 70

4.9 Appendix B: Some quantitative integer linear algebra 80

5 On Szemerédi's theorem with differences from a random set 82
5.1 Introduction 82
5.2 Proof of the main result 84

6 On a question of Alon 87
6.1 Introduction 87
6.2 Preliminaries, and an answer to Question 6.1.7 90
6.3 Proof of Theorem 6.1.4 and Theorem 6.1.6 92
6.4 Geometric commonness 96

7 Local aspects of the Sidorenko property for linear systems 99
7.1 Introduction 99
7.2 Preliminaries 104
7.3 Proof of Theorem 7.1.1 107
7.4 Examples 115
7.5 Localising to subsystems with fewest variables 118

Bibliography 120

Chapter 1

Introduction

This thesis is concerned with higher-order Fourier analysis. This area studies the combinatorics of linear and other arithmetic patterns in abelian groups via an analysis of arithmetically-structured functions which underlie it. We will concern ourselves both with this theory (Chapters 3, 4), as well as its application to combinatorial questions (Chapters 5, 6, 7).

1.1 Context

In this introductory section we will give very brief context for the development of higher-order Fourier analysis. We refer a reader who is interested in a more comprehensive account to [65, Chapter 1].

Roth's theorem says that positive density subsets of the integers contain nontrivial 3-term arithmetic progressions (3APs). Roth's original proof [58] is Fourier-analytic and in particular proceeds via the observation that sets with an unexpected number of 3APs (i.e. more or less than what is expected of a random set of the same density) have a large Fourier coefficient, so correlate with a nontrivial linear phase function. Roth's proof as a whole may loosely be described as the union of this Fourier-analytic observation and a 'density increment' strategy (which we will not describe any further).

Unfortunately, a naive attempt to generalise this strategy to prove Szemerédi's theorem, the analogous statement for longer arithmetic progressions, fails for the following reason: there are examples of sets with unexpectedly many four-term APs, and whose characteristic function does not correlate with any nontrivial linear phase function. See Example 2.5.2 below.

It transpires that in fact any such set must, in a certain sense, be quadratically-structured. This observation underpinned Gowers's seminal work [26], where he was

able to obtain a 'Roth-type', density-increment proof of Szemerédi's theorem. Here, a 'higher-order' counterpart to Fourier analysis – sufficient to generalise Roth's theorem to obtain Szemerédi's theorem for longer APs – was born.

In this work Gowers defined a hierarchy of norms, now known as the Gowers U^s-norms for $s = 1, 2, \ldots$. In a finite abelian group G, define the *discrete multiplicative derivative* of a complex-valued function f at a point $h \in G$ by

$$\Delta_h f(x) := f(x+h)\overline{f(x)},$$

where $\bar{f}$ is the complex conjugate of f. Then define

$$||f||_{U^s} := \left(\frac{1}{|G|^{s+1}} \sum_{h_1,\ldots,h_{s-1} \in G} \left| \sum_{x \in G} \Delta_{h_1} \cdots \Delta_{h_{s-1}} f(x) \right|^2 \right)^{1/2^s}.$$

Although it is not immediately obvious, this is indeed a norm for $s \geq 2$, and a seminorm for $s = 1$. We see from the above that the U^s-norm measures the L^2-norms of the $(s-1)$-th derivatives of f. Then, for example, if f is a function which is 'linear' in the sense that for all h, $\Delta_h f(x)$ is constant (i.e. independent of x), we expect $||f||_{U^2}$ to be large, as there will be no cancellation in the inner sum. We may similarly interpret the U^3-norm (respectively, U^s-norm) as measuring the extent to which a function exhibits 'quadratic' (respectively, degree-$(s-1)$ polynomial) behaviour.

Gowers showed that sets S with unexpectedly many kAPs satisfy the property that $||1_S - \alpha||_{U^{k-1}}$ is large, where 1_S is the characteristic function of S and α is the density of S (see (2.3)), giving a particular sense in which sets with unexpectedly many 4APs must be quadratically structured.

In the integers it further transpires that functions with large U^3-norm (respectively, U^s-norm) must correlate with an element of a class of quadratically structured functions known as *2-step nilsequences* (resp. $(s-1)$-step nilsequences). This is the content of the *inverse theorem for the U^3-norm* (resp. U^s-norm); see Section 2.4. A reader wishing to obtain a rough appreciation of the content of this statement may, for the meantime, think loosely of 2-step nilsequences as comprising linear combinations of quadratic phase functions $e(\alpha n^2)$ as well as *bracket quadratic* phases $e(\alpha n \lfloor \beta n \rfloor)$, where here and henceforth $e(\cdot)$ is shorthand for $e^{2\pi i \cdot}$ and $\lfloor \cdot \rfloor$ is the floor function. Similarly, s-step nilsequences may be thought loosely of as linear combinations of degree-s polynomial phase functions. We once again refer to Section 2.4 for more details.

Using the inverse theorem for Gowers norms as a black box, Green and Tao demonstrate a short proof of Szemerédi's theorem via the density increment strategy [37].

It is in this sense that higher-order Fourier analysis, and its central objects sketched briefly above, are precisely the 'right' generalisation of Fourier analysis to the analysis longer arithmetic progressions.

We are now ready to outline the contribution of this thesis to the study higher-order Fourier analysis and the combinatorics of linear patterns.

1.2 Outline

In Chapter 2 we provide some necessary background material on the notation and fundamentals which are used throughout the thesis.

Chapters 3 and 4 of this thesis deal with questions about the Gowers U^s-norms, and about nilsequences.

As mentioned above, Gowers showed [26] that the count of kAPs may be 'controlled' by the U^{k-1}-norm (see (2.3) for a more precise statement). More recently, Green–Tao [36] demonstrated that in fact any suitably non-degenerate linear pattern may be controlled by *some* U^s-norm (see, e.g. [65, Exercise 1.3.23]), and so extremal sets must exhibit polynomial behaviour of some degree. Answering a question of Gowers and Wolf and building on work of Green–Tao, in Chapter 3 we determine the minimal s for which one is able to obtain U^s-norm control on a given linear pattern, and so what is the maximal degree polynomial behaviour that one must consider when analysing extremal sets for this linear pattern.

Green and Tao's arithmetic regularity lemma [35] allows one to reduce certain questions in additive combinatorics to an analysis of nilsequences. Green and Tao complement their arithmetic regularity lemma with a counting lemma to assist in the analysis of nilsequences: via equidistributional results of polynomial sequences on nilmanifolds (see [38]), they reduce the analysis of discrete averages over products of nilsequences to an often more algebraic problem of integrating smooth functions over nilmanifolds. The Green–Tao counting lemma applies to linear patterns which satisfy a particular algebraic criterion known as 'the flag condition'. In Chapter 4, we obtain a version of the Green–Tao arithmetic regularity lemma and counting lemma which applies in the generality of all linear patterns.

Chapters 5, 6 and 7 are concerned with combinatorial questions about linear patterns in subsets of abelian groups.

In Chapter 5 we investigate a variant of Szemerédi's theorem in which one restricts the set of 'allowed' common differences. In particular, we are interested in the minimal size of a randomly chosen set such that Szemerédi's theorem holds almost surely

with common differences restricted to that set. We investigate the question in the integers and in $\mathbb{F}_p^n$, and show that the answer in the latter context is different from the conjectured answer in the former.

Chapters 6 and 7 look at arithmetic analogues of Sidorenko's conjecture, a well-studied but still wide-open conjecture in graph theory, see [61]. The arithmetic Sidorenko question in $\mathbb{F}_p^n$ asks: which systems of linear equations Ψ have the property that amongst all subsets of $\mathbb{F}_p^n$ of fixed density, random subsets minimise the number of solutions to Ψ? We call such systems of linear forms *Sidorenko*. In the arithmetic setting, we seem to be some distance from even a conjectural answer to this question. Linear systems which satisfy the property that amongst all 2-colourings of $\mathbb{F}_p^n$, a random 2-colouring minimises the number of monochromatic solutions to Ψ we shall call *common*. Alon [59, Question 4.1] asked whether a linear system is Sidorenko if and only if it remains common despite the addition of arbitrarily many free variables (indeed in the graph theoretic context the analogous fact is true under Sidorenko's conjecture [46]). In Chapter 6 we answer Alon's question in the negative. In Chapter 7, we investigate a local Sidorenko property: for which systems of linear forms do random subsets locally minimise the number of solutions?

Chapter 2

Background

In this chapter we collect background material which will be useful in forthcoming chapters.

2.1 Notation

For a positive integer N, we let $[N]$ denote the set $\{1, 2, \dots, N\}$. Let $[-N, N] := \{-N, -N+1, \dots, -1, 0, 1, \dots, N\}$.

We use standard asymptotic notation: O, o. The notation $O_A(X)$ is used to denote a quantity which can be upper bounded in magnitude by $c(A)X$, where A is some real-valued parameter and c is some positive real-valued function. The notation $o_{A\to 0;B}(X)$ is used as shorthand for a quantity bounded in magnitude by $c_B(A)X$ where $c_B(A) \to 0$ as $A \to 0$ for B fixed. We may also use slight variants of these notations which may be translated analogously.

Given a set S, the notation $\mathbb{E}_{x\in S}$ is used throughout as shorthand for $\frac{1}{|S|}\sum_{x\in S}$. At times, when the set S is clear from context, we may omit it from the notation. We will let 1_S denote the characteristic function of the set S.

We use $e(\cdot)$ to denote $e^{2\pi i \cdot}$, and $e_N(\cdot)$ as shorthand for $e^{2\pi i \cdot/N}$.

2.2 The Fourier transform

The *Fourier transform* of a complex-valued function f on $\mathbb{Z}/N\mathbb{Z}$ is defined as follows. For $r \in \mathbb{Z}/N\mathbb{Z}$ let

$$\hat{f}(r) := \mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f(x) e_N(-rx).$$

As a consequence of the orthogonality of characters $e_N(r_1\cdot), e_N(r_2\cdot)$ (where $r_1 \neq r_2$), one obtains the following Fourier inversion formula

$$f(x) = \sum_{r \in \mathbb{Z}/N\mathbb{Z}} \hat{f}(r) e_N(rx),$$

and Parseval's identity

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f(x)\overline{g(x)} = \sum_{r \in \mathbb{Z}/N\mathbb{Z}} \hat{f}(r)\overline{\hat{g}(r)},$$

for functions $g : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$.

Similarly, in $\mathbb{F}_p^n$, we define the Fourier transform of $f : \mathbb{F}_p^n \to \mathbb{C}$ at $r \in \mathbb{F}_p^n$ by

$$\hat{f}(r) := \mathbb{E}_{x \in \mathbb{F}_p^n} f(x) e_p(-r \cdot x).$$

One obtains analogous formulas for Fourier inversion and Parseval's identity.

2.3 Gowers norms

In this section we provide technical background on Gowers norms which is sufficient to state the main result of Chapter 3.

Definition 2.3.1. Let G be a finite abelian group. Let $k \geq 1$ be a positive integer. Let $f : G \to \mathbb{C}$. Then define the *Gowers U^k-norm* on G by:

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

where $\mathcal{C}$ denotes complex conjugation, $|\omega|$ is the sum of the elements of ω, and $h = (h_1, \ldots, h_k)$.

It transpires that $|| \cdot ||_{U^k(G)}$ as defined above is indeed a norm for $k \geq 2$, and is a seminorm when $k = 1$. One may check that this definition matches with the one provided in Chapter 1.

We will mostly be interested in case when $G = \mathbb{Z}/N\mathbb{Z}$, or $G = \mathbb{F}_p^n$. In these settings one has the following Fourier-analytic identity

$$||f||_{U^2(G)}^4 := \mathbb{E}_{x, h_1, h_2 \in G} f(x)\overline{f(x + h_1)f(x + h_2)}f(x + h_1 + h_2) = \sum_{r \in G} |\hat{f}(r)|^4. \quad (2.1)$$

We also define corresponding $U^k([N])$ norms on the interval $[N] := \{1, 2, \ldots, N\}$ by

$$||f||_{U^k([N])} := \frac{||f1_{[N]}||_{U^k(\mathbb{Z}/M\mathbb{Z})}}{||1_{[N]}||_{U^k(\mathbb{Z}/M\mathbb{Z})}},$$

where $M := 2^k N$. One may then deduce (see [36, Lemma B.5]) that there are constants c_k, C_k such that

$$c_k ||f 1_{[N]}||_{U^k(\mathbb{Z}/M\mathbb{Z})} \leq ||f||_{U^k([N])} \leq C_k ||f 1_{[N]}||_{U^k(\mathbb{Z}/M\mathbb{Z})}.$$

At times we may simply write U^k in place of $U^k(G)$ when the ambient group (or subset thereof) is clear from context. In a group G, we have the nesting property

$$||f||_{U^2(G)} \leq ||f||_{U^3(G)} \leq ||f||_{U^4(G)} \leq \cdots, \tag{2.2}$$

and in $[N]$ or $[-N, N]$, the same is true up to multiplicative constants.

Gowers showed for 1-bounded functions f_i on $\mathbb{Z}/N\mathbb{Z}$ that

$$|\mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} f_1(x) f_2(x+d) \cdots f_k(x+(k-1)d)| = o_{||f_i||_{U^{k+1}} \to 0;k}(1), \tag{2.3}$$

for all i. Thus, for a set S with density α in $\mathbb{Z}/N\mathbb{Z}$, by writing $1_S = \alpha + (1_S - \alpha)$, expanding appropriately and applying (2.3) many times, we may obtain

$$|\mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} 1_S(x) 1_S(x+d) \cdots 1_S(x+(k-1)d)| = \alpha^k + o_{||1_S - \alpha||_{U^{k+1}} \to 0;k}(1). \tag{2.4}$$

That is, if S is such that $||1_S - \alpha||_{U^{k+1}}$ small, then S contains approximately the number of kAPs that one would expect given its density in $\mathbb{Z}/N\mathbb{Z}$. We used the contrapositive of this statement in Chapter 1.

More generally we have the following fact due to Green–Tao [36]. See also [65, Exercise 1.3.23].

Theorem 2.3.2. *Let $t, D \geq 1$ be integers and let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^D \to \mathbb{Z}^t$ be a sequence of linear forms which are pairwise linearly independent. Then there exists an integer s such that for all 1-bounded functions f_i*

$$\left| \mathbb{E}_{\boldsymbol{x} \in [-N,N]^D} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x})) \right| = o_{||f_i||_{U^{s+1}[N]} \to 0;t,D,\Psi}(1). \tag{2.5}$$

In light of (2.2), it is natural to ask what the minimal s is such that (2.5) holds; we call such s the *complexity* of the system Ψ. Gowers and Wolf [28, Conjecture 2.5] conjectured that the complexity of Ψ is the smallest s such that the functions ψ_i^{s+1} are linearly independent. Building on work of Green–Tao [35], we resolve this conjecture in the integer setting in Chapter 3.

2.4 Nilsequences

In this section we give a brief account of the qualitative aspects of the theory of nilsequences. We intend for this to be sufficient to roughly appreciate their place in the theory. We do not attempt to provide a full technical or quantitative account. Quantitative aspects are developed later throughout the thesis as needed.

Recall that if a positive density subset of the integers does not contain approximately the expected number of 3APs given its density, then its characteristic function correlates with a linear phase function. In the analogous situation for longer progressions we have only so far deduced that the characteristic function has large Gowers norm. Obtaining an analogous statement about correlation with an arithmetically-structured function comes in the form of the following deep theorem due to Green–Tao–Ziegler (imprecisely stated here – see [41, Conjecture 1.2, Theorem 1.3] for a formal statement).

Theorem 2.4.1 (Informal statement of inverse theorem for Gowers norms in $[N]$)**.** *Let $\frac{1}{2} > \delta > 0$. Suppose that $f : [N] \to \mathbb{C}$ is such that $||f||_{U^k[N]} > \delta$. Then f correlates with a nilsequence $\phi : \mathbb{Z} \to \mathbb{C}$ of step at most $k-1$ and complexity bounded in terms of δ, that is,*

$$\left|\mathbb{E}_{x\in[N]} f(x)\phi(x)\right| \geq c_{k,\delta} > 0.$$

Towards defining nilsequences, we recall the following setup. Let G be a connected, simply-connected, nilpotent Lie group. Let Γ be a lattice (discrete, cocompact subgroup) in G, whence G/Γ is a nilmanifold. Let $G_\bullet = (G_i)_{i=0}^{s+1}$ be an s-step Lie filtration (so $G_0 = G_1 \geq G_2 \geq \cdots \geq G_{s+1} = 0$, and $G_{i+j} \geq [G_i, G_j]$ for all i, j) which is rational with respect to Γ in the sense that $\Gamma \cap G_i$ is a lattice in G_i for all i. Let $g : [N] \to G$ be a polynomial sequence with respect to $G_\bullet$, so that we may write $g(n) = g_1^n g_2^{\binom{n}{2}} \cdots g_s^{\binom{n}{s}}$ with $g_i \in G_i$. Let f be a Lipschitz continuous function[1] $G \to \mathbb{C}$ which is Γ-automorphic (we will often abuse notation and consider f as a function on G/Γ). With this setup, defining $\phi(n) := f(g(n)\Gamma)$ yields a nilsequence. This set of definitions is essentially consistent with [41, Definition 4.1]. We direct a reader who wishes to see formal definitions before engaging further with the content of this thesis to [41]. However, such a digression is not necessary; the upcoming chapters are self-sufficient, and and relevant definitions and references are recalled when needed.

We conclude this section by giving a couple of examples of 2-step nilsequences.

[1]This definition presupposes a metric structure on G which we will not discuss here. See, for example, [38, Definition 2.2] for more details.

Example 2.4.2. Let $G = \mathbb{R}$ and let $\Gamma = \mathbb{Z}$. For $\alpha \in \mathbb{R}$ consider the polynomial sequence $g(n) = \alpha n + 2\alpha\binom{n}{2}$, which is adapted to the filtration given by $G_0 = G_1 = G_2 = \mathbb{R}$, $G_3 = G_4 = \ldots = 0$. Consider the function $e(\cdot)$ on $\mathbb{R}$ which is automorphic with respect to $\mathbb{Z}$. This setup demonstrates that the quadratic phase $e(\alpha n^2)$ is a 2-step nilsequence.

Example 2.4.3. In this example we work in the Heisenberg group $G := \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$, where the group operation is matrix multiplication. This is a connected, simply connected Lie group. The subset $\Gamma := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$ is a lattice. To construct a function which is automorphic with respect to the action of this lattice, we will identify its values on a fundamental domain, and then extend it by the action of the lattice. Note that given $x := \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \in G$, we have $\gamma_x := \begin{pmatrix} 1 & -\lfloor a \rfloor & -\lfloor c - a\lfloor b \rfloor \rfloor \\ 0 & 1 & -\lfloor b \rfloor \\ 0 & 0 & 1 \end{pmatrix} \in \Gamma$ such that $x\gamma_x \in \begin{pmatrix} 1 & [0,1) & [0,1) \\ 0 & 1 & [0,1) \\ 0 & 0 & 1 \end{pmatrix}$. This latter set will be our chosen fundamental domain. Indeed, one computes $x\gamma_x = \begin{pmatrix} 1 & a - \lfloor a \rfloor & c - a\lfloor b \rfloor - \lfloor c - a\lfloor b \rfloor \rfloor \\ 0 & 1 & b - \lfloor b \rfloor \\ 0 & 0 & 1 \end{pmatrix}$. We will choose our function on our fundamental domain as follows:

$$\begin{pmatrix} 1 & [0,1) & [0,1) \\ 0 & 1 & [0,1) \\ 0 & 0 & 1 \end{pmatrix} \ni \begin{pmatrix} 1 & A & C \\ 0 & 1 & B \\ 0 & 0 & 1 \end{pmatrix} \mapsto e(-C).$$

Therefore, the function $G \to \mathbb{C}$ defined by $x \mapsto e(-c + a\lfloor b \rfloor)$ is Γ-automorphic (as can indeed be checked by direct computation). We note that this function is not actually continuous on G, so the sequence we ultimately produce will not technically be a nilsequence by the definition above. However, we encourage the reader to overlook this technicality for the purposes of reading this introductory example.

To identify a nilsequence, it remains to compose this function with a polynomial sequence. We will use the filtration $G_0 = G_1 = G$, $G_2 = \begin{pmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $G_3 = G_4 = \ldots = \{\mathrm{id}\}$. We define our polynomial sequence $g(n) = g_1^n g_2^{\binom{n}{2}}$ with $g_1 = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{pmatrix}$ and $g_2 = \begin{pmatrix} 1 & 0 & -\alpha\beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. One computes then that $g(n) = \begin{pmatrix} 1 & \alpha n & 0 \\ 0 & 1 & \beta n \\ 0 & 0 & 1 \end{pmatrix}$. Compos-

ing this with our automorphic function above, we have demonstrated that the bracket quadratic $e(\alpha n\lfloor\beta n\rfloor)$ is a 2-step nilsequence.

2.5 An example

In this section we provide an example to justify the claim we made in Chapter 1 that there exists a set which has unexpectedly many 4APs, but no large nontrivial Fourier coefficients.

We will need the following very standard Gauss sum estimate.

Lemma 2.5.1. *Let $a, b \in \mathbb{Z}/N\mathbb{Z}$ with $a \neq 0$. Then*

$$\left|\sum_{x\in\mathbb{Z}/N\mathbb{Z}} e_N(ax^2+bx)\right| = N^{-1/2}.$$

Proof. We compute that

$$\begin{aligned}\left|\sum_{x\in\mathbb{Z}/N\mathbb{Z}} e_N(ax^2+bx)\right|^2 &= \sum_{x,h\in\mathbb{Z}/N\mathbb{Z}} e_N(a(x+h)^2+b(x+h))\overline{e_N(ax^2+bx)}\\ &= \sum_h e_N(ah^2+bh)\sum_x e_N(2axh).\end{aligned}$$

The inner sum is zero unless $h=0$, in which case it is equal to N. The lemma follows by taking square roots. □

Example 2.5.2. Let $f : \mathbb{Z}/N\mathbb{Z} \to [0,1]$ be defined by

$$f(x) = \frac{1}{2} + \frac{1}{8}\left(e_N(x^2) + e_N(-x^2) + e_N(3x^2) + e_N(-3x^2)\right).$$

We will show that f is Fourier uniform in the sense that $\sup_{r\neq 0}\hat{f}(r) = o_N(1)$. We will furthermore show that the normalised count of 4APs "weighted by f" is larger by a constant than $1/16$, which is what one would expect given its average over $\mathbb{Z}/N\mathbb{Z}$. One may convert this function f to a set S with the same properties by choosing an element x to lie in S at random with probability $f(x)$; these details are standard and we will not provide them here. We direct the reader who is interested in details of the random sampling argument to [21], for example.

We have for $r \neq 0$ that

$$\begin{aligned}|\hat{f}(r)| &= \left|\mathbb{E}_x \frac{1}{2} e_N(-rx) + \frac{1}{8}\mathbb{E}_x \left(e_N(x^2) + e_N(-x^2) + e_N(3x^2) + e_N(-3x^2)\right) e_N(-rx)\right| \\ &\leq \frac{1}{8}(\left|\mathbb{E}_x e_N(x^2 - rx)\right| + \left|\mathbb{E}_x e_N(-x^2 - rx)\right| \\ &\qquad + \left|\mathbb{E}_x e_N(3x^2 - rx)\right| + \left|\mathbb{E}_x e_N(-3x^2 - rx)\right|) \\ &\leq N^{-1/2},\end{aligned}$$

where the final line follows from Gauss sum estimate Lemma 2.5.1. That is, f is Fourier uniform. One may also compute that

$$\mathbb{E}_x f(x) = \hat{f}(0) = \frac{1}{2} + O(N^{-1/2}).$$

For brevity we will use the notation $f_0 := f - \frac{1}{2}$. We may now compute the normalised number of 4APs, "weighted by f":

$$\begin{aligned}&\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)f(x+3d) \\ =&\mathbb{E}_{x,d}(\frac{1}{2} + f_0)(x)(\frac{1}{2} + f_0)(x+d)(\frac{1}{2} + f_0)(x+2d)(\frac{1}{2} + f_0)(x+3d) \\ =&\frac{1}{16} + \text{ 14 terms } + \mathbb{E}_{x,d} f_0(x)f_0(x+d)f_0(x+2d)f_0(x+3d).\end{aligned}$$

In fact, one may show using the Fourier-analytic identities and Gauss sum estimates above, that the middle 14 terms contribute $O(N^{-1/2})$. On the other hand we compute

$$\begin{aligned}&\mathbb{E}_{x,d} f_0(x)f_0(x+d)f_0(x+2d)f_0(x+3d) \\ =&\frac{1}{8^4}\mathbb{E}_{x,d} \left(e_N(x^2) + e_N(-x^2) + e_N(3x^2) + e_N(-3x^2)\right) \cdots \\ &\qquad \cdots \left(e_N((x+3d)^2) + e_N(-(x+3d)^2) + e_N(3(x+3d)^2) + e_N(-3(x+3d)^2)\right) \\ =&\frac{1}{8^4}\sum_i \mathbb{E}_{x,d} e_N(q_i(x,d)),\end{aligned}$$

where the sum is over 256 quadratics $q_i(x,d)$, each of the form $a_1x^2 + a_2(x+d)^2 + a_3(x+2d)^2 + a_4(x+3d)^2$ where $a_i \in \{\pm 1, \pm 3\}$. By our Gauss sum estimates above, we know that the contribution of $\mathbb{E}_{x,d} e_N(q_i(x,d))$ is of size at most $N^{-1/2}$ unless q_i is the zero quadratic. One may check that this happens only when $(a_1, a_2, a_3, a_4) = \pm(1, -3, 3, -1)$. Thus, combining the above, we ultimately have

$$\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)f(x+3d) = \frac{1}{16} + \frac{2}{8^4} + O(N^{-1/2}),$$

which is larger than $\frac{1}{16}$ for large N.

Thus, by the random sampling argument mentioned above, we may produce a set of density approximately $1/2$, which has small Fourier coefficients, and which contains more 4APs than what is expected from a set of this density. The Roth-type, Fourier analytic observation for Roth's theorem on 3APs no longer holds for 4APs.

Chapter 3

On a conjecture of Gowers and Wolf

3.1 Introduction

In [28, Conjecture 2.5], Gowers and Wolf made a conjecture on the minimal Gowers norm which is able to control averages of f on a set of linear forms: $\mathbb{E}_{\boldsymbol{x}} \prod_{i=1}^{t} f(\psi_i(\boldsymbol{x}))$. In $\mathbb{F}_p^n$ (p fixed, n large), the conjecture was resolved by Gowers and Wolf for large enough p [28, 30, 29], and ultimately for all p by Hatami–Hatami–Lovett [44]. Furthermore, Gowers and Wolf resolved the first nontrivial case of the conjecture in $\mathbb{Z}/N\mathbb{Z}$ in [31]. In the integers, the conjecture was considered to be resolved by Green and Tao until recently when it was noted by the author (see [40], [66]) that the proof of [35, Theorem 1.13] requires an assumption on the system of linear forms in question, which we will now describe.

Let $\Psi := (\psi_i)_{i=1}^t$ be a family of linear forms, each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$. For positive integers k, let $\Psi^{[k]}$ be the real vector space $\mathrm{span}_{\mathbb{R}}\{(\psi_1(\boldsymbol{x})^k, \dots, \psi_t(\boldsymbol{x})^k) : \boldsymbol{x} \in \mathbb{Z}^D\} \leq \mathbb{R}^t$. The system of linear forms Ψ satisfies *the flag condition* if the containment of vector spaces $\Psi^{[k]} \leq \Psi^{[l]}$ holds whenever $k < l$.[1]

Green and Tao's result [35, Theorem 1.13] holds for systems of linear forms that satisfy the flag condition. It is perhaps not *a priori* clear to what extent this is a restriction on the space of linear forms and thus to what extent the Gowers-Wolf conjecture ought to be considered open. We begin with the modest remark that indeed there do exist systems of linear forms which do not satisfy the flag condition. Further, there are values of the parameters (t, D) for which a system of linear forms on these parameters is generically not flag. On the other hand, many important examples of systems of linear forms do satisfy the flag condition. For example, systems

[1] We may also, at times, use *flag* as an adjective.

corresponding to arithmetic progressions satisfy the flag condition. More generally, *translation invariant* systems – those for which the vector $(1, 1, \dots, 1)$ lies in $\Psi^{[1]}$ – satisfy the flag condition. For a more careful discussion of the significance of the flag condition, and counterexamples that lead to the restriction of [35, Theorem 1.13], we direct the reader to [40] and [66].

The main result of this chapter is Theorem 3.1.1 below, a full resolution of the Gowers-Wolf conjecture; that is, a generalisation [35, Theorem 1.13] to systems of linear forms which do not necessarily satisfy the flag condition. Further to this, Theorem 3.1.1 is stronger than an affirmative answer to the Gowers-Wolf conjecture in that it achieves control over averages involving t distinct functions $\mathbb{E}_{\boldsymbol{x}\in[-N,N]^D}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x}))$ given the relatively weak information that only one of the functions f_i has small U^{s+1}-norm. A strengthening of the Gowers-Wolf conjecture in this direction has already been achieved in the finite field setting; see [45], [44]. Here and in what follows, $[N] := \{1, \dots, N\}$ and $[-N, N] := \{-N, \dots, N\}$.[2]

Theorem 3.1.1. *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$, and let $s \geq 1$ be an integer such that the polynomials $\psi_1^{s+1}, \dots, \psi_t^{s+1}$ are linearly independent. For $i = 1, \dots t$, let $f_i : [-N, N] \to \mathbb{C}$ be functions bounded in magnitude by 1 (and defined to be zero outside of $[-N, N]$). For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $\min_i ||f_i||_{U^{s+1}[-N,N]} \leq \delta$, then*

$$\left|\mathbb{E}_{\boldsymbol{x}\in[-N,N]^D}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x}))\right| \leq \varepsilon.$$

The strategy of proof employed is the same as the general 'arithmetic regularity lemma and counting lemma' strategy put forward in [35]. Unfortunately, the counting lemma [35, Theorem 1.11] applies only to systems of linear forms which satisfy the flag condition (indeed this is where [40, Theorem 1.13] fails to generalise). We need, in the first instance, a somewhat generalised and strengthened version of [40, Theorem 1.13] for flag systems of linear forms; this is Theorem 3.2.1 which is proven in Section 2. We defer to the beginning of Section 3.2 for a brief discussion on the ways in which Theorem 3.2.1 extends what is done in [40, Theorem 1.13]. Then to recover Theorem 3.1.1 above, we use Theorem 3.2.1 together with the observation that if $(\psi_i)_{i=1}^t$ is not flag, there is a sequence of integers $(a_i)_{i=1}^t$ such that the system $(a_i\psi_i)_{i=1}^t$ is flag; this is implemented in Section 3.

[2]We work in $[-N, N]$ rather than $[N]$ because the set $K := [-N, N]^D \cap \Psi^{-1}([-N, N]^t)$ satisfies $|K| \gg N^D$. This is not necessarily true of the set $[N]^D \cap \Psi^{-1}([N]^t)$, whereupon the upcoming theorem would be trivially true for those systems of linear forms with $|[N]^D \cap \Psi^{-1}([N]^t)| = o(N^D)$. Take, for example, the system $\Psi(x, y) = (-x, -x-y, -x-2y)$, or $\Psi'(x, y, \dots) = (x-2y, y-2x, \dots)$.

3.1.1 Notation, conventions

We reiterate and slightly generalise some notation that was provided in Chapter 2.

As in [36, Appendix B], define Gowers norms on subsets of additive groups as follows. Let Z be an abelian group and let $A \subset Z$. For any function $f : A \to \mathbb{C}$ and for all $s \geq 1$ define the Gowers uniformity norm $||f||_{U^{s+1}(A)}$ by

$$||f||^{2^{s+1}}_{U^{s+1}(A)} = \mathbb{E}_{\substack{x,h: x+\omega\cdot h \in A \\ \text{for all } \omega \in \{0,1\}^{s+1}}} \prod_{\omega \in \{0,1\}^{s+1}} \mathcal{C}^{|\omega|} f(x + \omega \cdot h),$$

where the x are elements of Z, the h are vectors in Z^{s+1}, $\mathcal{C}$ is the complex conjugation operator and $|\omega| = \sum_{i=1}^{s+1} \omega_i$. As is noted in [36, Appendix B], we have the relation

$$||f||_{U^{s+1}(A)} = ||f1_A||_{U^{s+1}(Z)}/||1_A||_{U^{s+1}(Z)}.$$

As in [35], we will use $o_{A\to\infty;M}(X)$ as shorthand for a quantity bounded in magnitude by $c_M(A)X$ where $c_M(A) \to 0$ as $A \to \infty$ for M fixed. We will also use slight variations on this notation which may be translated analogously.

3.2 A strengthened result in the flag case

The goal of this section is to establish the following theorem.

Theorem 3.2.1. *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$ such that Ψ satisfies the flag condition. Let $s \geq 1$ be an integer such that the polynomials $\psi_1^{s+1}, \dots, \psi_t^{s+1}$ are linearly independent. For $i = 1, \dots t$, let $f_i : [N] \to \mathbb{C}$ be functions bounded in magnitude by 1 (and defined to be zero outside of $[N]$). Let K be a subset of $[-N, N]^D$ whose convex hull in $\mathbb{R}^D$ is a polytope with $O_{\Psi,s,D,t}(1)$ faces, and for which $|K| = \Theta_{\Psi,s,D,t}(N^D)$. Let $c \in \mathbb{Z}$ be such that $\Psi(K) + (c, c, \dots, c) \subset [N]^t$. For all $\varepsilon > 0$ there exists $\delta > 0$ (which depends on D, t, s, Ψ and ε, but is uniform in N, K, f_i and c) such that if $\min_i ||f_i||_{U^{s+1}[N]} \leq \delta$, then*

$$\left| \mathbb{E}_{\boldsymbol{x} \in K} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x}) + c) \right| \leq \varepsilon.$$

Theorem 3.2.1 is [35, Theorem 1.13] with some additional features and amendments which we will describe shortly. In particular our proof strategy is taken from [35, Theorem 1.13]. However, the modifications are essentially ubiquitous and so writing out the full proof appears to us to be necessary.

The additional features in Theorem 3.2.1 which will be necessary to prove Theorem 3.1.1 in the next section are as follows:

1. we accommodate averages over large convex polytopes in $[-N, N]^D$,

2. we accommodate shifts of systems of linear forms,

3. we obtain a result for averages comprising t distinct functions f_i given that only one of these has $U^{s+1}[N]$-norm less than δ.

In addition, we amend two parts of the argument in [35, Theorem 1.13]:

1. We give an explicit and quantitative argument for the case in which N is bounded in terms of δ; see Lemma 3.2.2 below.

2. In the proof of [35, Proposition 7.2], it is claimed that "the contribution of $f_{\mathrm{sml}}(n)$ to (7.6) is $O(\varepsilon)$ by the Cauchy-Schwarz inequality", where (7.6) is an average over some $\varepsilon' N$-length subprogressions of $[N]$. It appears to us that one cannot reach this conclusion immediately and that indeed one needs to utilise outer averages over the set of $\varepsilon' N$-length progressions. The argument to bound the analogous term in our proof is somewhat more intricate than an application of the Cauchy-Schwarz inequality and occurs in and around Lemma 3.2.5.

Proof of Theorem 3.2.1. Recall the notation from the statement of Theorem 3.2.1. Let $\varepsilon > 0$ and let $\delta > 0$ be a parameter which we will of course optimise in terms of ε later. For $i = 1, \dots, t$ let $\delta_i := ||f_i||_{U^{s+1}[N]}$ so that $\min_i \delta_i \leq \delta$. In this section $\{t, D, s, \Psi\}$ is fixed, and we will let all constants depend on this set without indicating this in our notation. Furthermore, we will abuse notation and let C be a constant which may change line to line.

We may assume that ε is small because if the condition from Theorem 3.2.1 is true for ε then it is obviously true for all $\varepsilon' > \varepsilon$. To prove Theorem 3.2.1 it is then, of course, sufficient to show that

$$\mathbb{E}_{\boldsymbol{x} \in K} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x}) + c) = O(\varepsilon^{1/2}),$$

whenever δ is sufficiently small depending on ε.

Next, we will claim that we may assume that N is large depending on ε.

Lemma 3.2.2. *The following inequality holds:*

$$\left| \mathbb{E}_{\boldsymbol{x} \in K} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x}) + c) \right| \leq C N^{1/4} \delta.$$

Proof. Let j be chosen so that $||f_j||_{U^{s+1}[N]} = \min_i ||f_i||_{U^{s+1}[N]} \leq \delta$. By Cauchy-Schwarz we may upper bound $|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x})+c)|$ by $(\mathbb{E}_{\boldsymbol{x}\in K}|f_j(\psi_j(\boldsymbol{x})+c)|^2)^{1/2}$. Next, $\#\{\boldsymbol{x}\in K : \psi_j(\boldsymbol{x})+c = n\} = O(N^{D-1}) = O(|K|/N)$ and so $\mathbb{E}_{\boldsymbol{x}\in K}|f_j(\psi_j(\boldsymbol{x})+c)|^2 \ll \mathbb{E}_{n\in[N]}|f_j(n)|^2$. Furthermore, observing that $|f_j(n)|^2 = f_j(n)\overline{f_j(n+0)}$ and then adding the (positive) contribution from all nonzero shifts h, we have

$$\begin{aligned}
|\mathbb{E}_{n\in[N]}|f_j(n)|^2|^2 &\leq \sum_{h\in\mathbb{Z}} |\mathbb{E}_{n\in[N]} f_j(n)\overline{f_j(n+h)}|^2 \\
&= \sum_h \mathbb{E}_{n,m\in[N]} f_j(n)\overline{f_j(m)f_j(n+h)}f_j(m+h) \\
&= \frac{1}{N^2}\sum_{n,h,h'\in\mathbb{Z}} f_j(n)\overline{f_j(n+h')f_j(n+h)}f_j(n+h'+h) \\
&= \frac{|\{(n,h,h')\in\mathbb{Z} : n, n+h', n+h, n+h'+h\in[N]\}|}{N^2}||f_j||^4_{U^2[N]}.
\end{aligned}$$

But $\#\{n,h,h'\in\mathbb{Z} : n, n+h', n+h, n+h'+h\in[N]\} = \Theta(N^3)$ and so we have that

$$\left|\mathbb{E}_{n\in[N]}|f_j(n)|^2\right|^2 \leq CN||f_j||^4_{U^2[N]}.$$

Finally, it follows from the monotonicity of Gowers norms in additive groups and the formula $||f||_{U^{s+1}(A)} = ||f1_A||_{U^{s+1}(Z)}/||1_A||_{U^{s+1}(Z)}$ which we recalled in Subsection 3.1.1 above that $||f_j||_{U^2[N]} \ll ||f_j||_{U^{s+1}[N]}$ for $s\geq 1$. One concludes by combining the above inequalities. □

It follows from the previous lemma that if $N = O_\varepsilon(1)$ then we may simply choose δ to be sufficiently small depending on N (depending on ε) to prove Theorem 3.2.1; the claim follows.

Let s' be the Cauchy-Schwarz complexity of the system Ψ (cf. [65, Definition 1.3.2]). Note that we may assume that $s'\geq s$. Apply the functional decomposition from Theorem 3.4.1 at step s' with parameters ε as above and $\mathcal{F}$ to be determined later to the functions $\{f_i\}_{i=1}^t$. We import notation from Appendix 3.4.1 and so in particular we have $f_i = f_{i,\mathrm{sml}} + f_{i,\mathrm{nil}} + f_{i,\mathrm{unf}}$ for functions $f_{i,\mathrm{sml}}$ with $||f_{i,\mathrm{sml}}||_2 \leq \varepsilon$, $f_{i,\mathrm{unf}}$ with $||f_{i,\mathrm{unf}}||_{U^{s'+1}[N]} \leq 1/\mathcal{F}(M)$ and $f_{i,\mathrm{nil}}$ with $f_{i,\mathrm{nil}}(n) = F_i(g(n)\Gamma, n(\mathrm{mod}\, q), n/N)$ where $1\leq q\leq M = O_{\varepsilon,\mathcal{F}}(1)$ and where $g(n)$ is an $(\mathcal{F}(M),N)$-irrational polynomial sequence with respect to a filtered nilmanifold $(G/\Gamma, G_\bullet)$.

Then we may write

$$\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x})+c) = \mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t (f_{i,\mathrm{nil}}+f_{i,\mathrm{sml}}+f_{i,\mathrm{unf}})(\psi_i(\boldsymbol{x})+c)$$

and expand this as the sum of 3^t terms each of the form $\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_{i,\mathrm{label}(i)}(\psi_i(\boldsymbol{x})+c)$, where $\mathrm{label}(i)\in\{\mathrm{nil},\mathrm{unf},\mathrm{sml}\}$ for each i. Recall that we inherit a quantity $M=O_{\varepsilon,\mathcal{F}}(1)$ from Theorem 3.4.1 which is an upper bound for the complexity of the nilsequences $f_{i,\mathrm{nil}}$.

Of these 3^t terms, we claim that any term with some $\mathrm{label}(i)=\mathrm{sml}$ is of size $O(\varepsilon)$. Indeed by Cauchy-Schwarz we obtain the upper bound $(\mathbb{E}_{\boldsymbol{x}\in K}|f_{i,\mathrm{sml}}(\psi_i(\boldsymbol{x})+c)|^2)^{1/2}$ for such a term. We conclude by recalling that $|K|=\Theta(N^D)$, noting that as $\boldsymbol{x}$ ranges over K, the quantity $\psi_i(\boldsymbol{x})+c$ takes on any particular value in $[N]$ at most N^{D-1} times, and using the L^2 bound inherited from Theorem 3.4.1.

Next, any of these 3^t terms with some $\mathrm{label}(i)=\mathrm{unf}$ is of size $o_{\mathcal{F}(M)\to\infty}(1)+o_{N\to\infty;\mathcal{F}(M)}(1)$ by Theorem 3.4.3. Thus we have

$$\left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x})+c)\right| \le \left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_{i,\mathrm{nil}}(\psi_i(\boldsymbol{x})+c)\right| + O(\varepsilon) + o_{\mathcal{F}(M)\to\infty}(1)+o_{N\to\infty;\mathcal{F}(M)}(1). \tag{3.1}$$

Recall that we may write $f_{i,\mathrm{nil}}(n)=F_i(g(n)\Gamma,n\ (\mathrm{mod}\,q),n/N)$. To deal with the mod q and Archimedean behaviour, we will need to do some volume-packing. Let ε' be a small quantity depending on ε,M; we will decide how small later. In what follows, a 'cube' is a cartesian product of equal length intervals, and a 'dilated cube' is a cartesian product of equal length equal step arithmetic progressions.

Lemma 3.2.3. *There exists a 'boundary' subset S of K which contains $O(q\varepsilon' N^D)$ elements of $[N]^D$ such that the set $K\backslash S$ can be partitioned into cubes of side length $q\varepsilon' N$, where each cube is itself a disjoint union of q^D dilated cubes of the form $P_1\times\cdots\times P_D$ where each P_i is an arithmetic progression in $[N]$ of spacing q and length $\varepsilon' N$.*

Proof. The decomposition of cubes with q-divisible side lengths into q^D-many q-dilated cubes is obvious. From here the strategy is essentially to draw K on D-dimensional grid paper with cubes of side length $q\varepsilon' N$ and exclude from K any cubes which are not strictly contained in K. Let Q be a cube such that $Q\cap K\neq\emptyset$ and $Q\cap K^c\neq\emptyset$. The maximum distance between any two points in Q is $\sqrt{D}q\varepsilon' N$ and so all points in Q lie within the $\sqrt{D}q\varepsilon' N$-neighbourhood of the boundary of (the convex hull in $\mathbb{R}^D$ of) K. Now recall that this boundary comprises at most $O(1)$ codimension 1 faces. The $\sqrt{D}q\varepsilon' N$-neighbourhood of any one of these faces may contain at most $O(q\varepsilon' N^D)$ points in $[-N,N]^D$. It follows that there are at most $O(q\varepsilon' N^D)$ points in K which lie in the $\sqrt{D}q\varepsilon' N$-neighbourhood of the boundary of K. □

Remove S from K and partition $K\backslash S$ as per the above lemma. Let $\mathcal{P}$ be the collection of all $P := P_1 \times \cdots \times P_D$. Then we have

$$\left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^{t} f_{i,\mathrm{nil}}(\psi_i(\boldsymbol{x})+c)\right| \leq \left|\mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{\boldsymbol{x}\in P}\prod_{i=1}^{t} f_{i,\mathrm{nil}}(\psi_i(\boldsymbol{x})+c)\right| + O_M(\varepsilon'). \tag{3.2}$$

Observe that for a fixed P, all $\boldsymbol{x} \in P$ yield the same value $b_{P,i} := \psi_i(\boldsymbol{x})+c \pmod q$. Furthermore, there is a number $c_{P,i}$ such that $|(\psi_i(\boldsymbol{x})+c)/N - c_{P,i}| = O(\varepsilon')$ for all $\boldsymbol{x} \in P$. Thus from the fact that F_i is M-Lipschitz we have

$$\mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{\boldsymbol{x}\in P}\prod_{i=1}^{t} f_{i,\mathrm{nil}}(\psi_i(\boldsymbol{x})+c) = \mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{\boldsymbol{x}\in P}\prod_{i=1}^{t} F_i(g(\psi_i(\boldsymbol{x})+c)\Gamma, b_{P,i}, c_{P,i}) + O_M(\varepsilon'). \tag{3.3}$$

Now define $\tilde{g}(n) = g(n+c)$; [35, Lemma A.8] gives that $\tilde{g}$ is also $(\mathcal{F}(M), N)$-irrational. Also recall the definition of the Leibman group $G^\Psi := \langle g_i^{v_i} : g_i \in G_i, v_i \in \Psi^{[i]}\rangle$ (we direct the reader to [35, Chapter 3] for some basic facts pertaining to G^Ψ). Finally, recall that the linear forms Ψ satisfy the flag condition. Thus we may apply Theorem 3.4.2 to obtain

$$\begin{aligned}\mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{\boldsymbol{x}\in P}\prod_{i=1}^{t} F_i(\tilde{g}(\psi_i(\boldsymbol{x}))\Gamma, b_{P,i}, c_{P,i}) = \mathbb{E}_{P\in\mathcal{P}}\int_{(g_1,\ldots,g_t)\Gamma^\Psi\in G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t} F_i(g_i\Gamma, b_{P,i}, c_{P,i}) \\ + o_{\mathcal{F}(M)\to\infty;M}(1) + o_{\varepsilon' N\to\infty;M}(1).\end{aligned} \tag{3.4}$$

Since the forms $\psi_1^{s+1}, \ldots, \psi_t^{s+1}$ are linearly independent, we have $\Psi^{[s+1]} = \mathbb{R}^t$ and so $G_{s+1}^t \leq G^\Psi$. Now, recall from [35, Chapter 7] the notation

$$F_{i,\leq s}(g\Gamma, b, c) := \int_{G_{s+1}/\Gamma_{s+1}} F_i(gg_{s+1}\Gamma, b, c)dg_{s+1}.$$

It follows that

$$\int_{G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t} F_i(g_i\Gamma, b_{P,i}, c_{P,i}) = \int_{G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t} F_{i,\leq s}(g_i\Gamma, b_{P,i}, c_{P,i}).$$

Furthermore,

$$\begin{aligned}\left|\int_{G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t} F_{i,\leq s}(g_i\Gamma, b_{P,i}, c_{P,i})\right| &\leq \int_{G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t} |F_{i,\leq s}(g_i\Gamma, b_{P,i}, c_{P,i})| \\ &\leq \int_{G^\Psi/\Gamma^\Psi}\min_i |F_{i,\leq s}(g_i\Gamma, b_{P,i}, c_{P,i})| \\ &\leq \min_i \int_{G/\Gamma} |F_{i,\leq s}(g\Gamma, b_{P,i}, c_{P,i})| \\ &\leq \min_i \left(\int_{G/\Gamma} |F_{i,\leq s}(g\Gamma, b_{P,i}, c_{P,i})|^2\right)^{1/2}\end{aligned}$$

where we have used in the penultimate line that the projection of G^Ψ onto any of its coordinates is surjective onto G (since each ψ_i must be nonzero) and that this projection maps the Haar measure on G^Ψ to the Haar measure on G.

All in all we have (continuing on from (3.4)):

$$\left|\mathbb{E}_{P\in\mathcal{P}}\int_{G^\Psi/\Gamma^\Psi}\prod_{i=1}^{t}F_i(g_i\Gamma,b_{P,i},c_{P,i})\right| \leq \min_i \mathbb{E}_{P\in\mathcal{P}}\left(\int_{G/\Gamma}|F_{i,\leq s}(g\Gamma,b_{P,i},c_{P,i})|^2\right)^{1/2}. \tag{3.5}$$

For brevity we temporarily abuse notation and use F_i to denote the function $F_i(\cdot,b_{P,i},c_{P,i})$; similarly for $F_{i,\leq s}$. We have

$$\int_{G/\Gamma}|F_{i,\leq s}|^2 = \int_{G/\Gamma}(F_{i,\leq s}-F_i)\overline{F_{i,\leq s}} + \int_{G/\Gamma}F_i\overline{F_{i,\leq s}} = \int_{G/\Gamma}F_i\overline{F_{i,\leq s}},$$

since $F_{i,\leq s}$ is invariant on G_{s+1} cosets and $F_{i,\leq s}-F_i$ integrates to zero on any such coset.

Lemma 3.2.4. *For each $P\in\mathcal{P}, i=1,\dots,t$ there is a subprogression $\tilde{P}_i\subset\psi_i(P)+c$ of size $\Theta(\varepsilon' N)$ such that $n=b_{P,i}\pmod q$ and $|n/N-c_{P,i}|=O(\varepsilon')$ for all $n\in\tilde{P}_i$.*

Proof. Recall that $b_{P,i}$ was defined to be the (unique) residue class of $\psi_i(P)+c\pmod q$ and that $c_{P,i}$ was defined to be a number such that $|(\psi_i(\boldsymbol{x})+c)/N-c_{P,i}|=O(\varepsilon')$ for all $\boldsymbol{x}\in P$. Thus we may let $\tilde{P}_i$ be *any* subprogression of length $\Theta(\varepsilon' N)$ contained in the set $\psi_i(P)+c$. Note that $|\psi_i(P)+c|=O(\varepsilon' N)$ and so we need only concern ourselves with a lower bound on the subset we seek. To see that $\psi_i(P)+c$ does contain a subprogression of length $\gg\varepsilon' N$, let $j\in[D]$ be an index such that the jth coefficient of ψ_i is nonzero. Recall that $P=P_1\times\cdots\times P_D$ is a (dilated) cube in which each side contains $\varepsilon' N$ points, and so we may let $\tilde{P}_i$ be the image under $\psi_i(\cdot)+c$ of the set $(p_0,p_1,\dots,P_j,\dots,p_D)$ where each of the p_k are fixed points in P_k for $k\neq j$. □

Let $\tilde{P}_i$ be as in the previous lemma. Then we may apply Theorem 3.4.2 to obtain

$$\int_{G/\Gamma}F_i\overline{F_{i,\leq s}} = \mathbb{E}_{n\in\tilde{P}_i}F_i\overline{F_{i,\leq s}}(g(n)\Gamma,b_{P,i},c_{P,i}) + o_{\mathcal{F}(M)\to\infty;M}(1) + o_{\varepsilon' N\to\infty;M}(1).$$

Recalling the properties of $\tilde{P}_i$ from Lemma 3.2.4 and that F_i (viewed as a function on $G/\Gamma\times\mathbb{Z}/q\mathbb{Z}\times\mathbb{R}\to\mathbb{C}$, i.e. ceasing the abuse of notation) has Lipschitz constant $\leq M$, we may write the right hand side of the above as

$$\mathbb{E}_{n\in\tilde{P}_i}F_i(g(n)\Gamma,n\pmod q,n/N)\overline{F_{i,\leq s}}(g(n)\Gamma,b_{P,i},c_{P,i}) + o_{\mathcal{F}(M)\to\infty;M}(1) + o_{\varepsilon' N\to\infty;M}(1) + O_M(\varepsilon').$$

Thus since (3.5) we have shown:

$$\begin{aligned}\min_i \mathbb{E}_{P\in\mathcal{P}} \left(\int_{G/\Gamma} |F_{i,\leq s}(g\Gamma, b_{P,i}, c_{P,i})|^2\right)^{1/2} \\ \leq \min_i \mathbb{E}_{P\in\mathcal{P}} \left|\mathbb{E}_{n\in\tilde{P}_i} F_i(g(n)\Gamma, n \pmod q, n/N)\overline{F_{i,\leq s}}(g(n)\Gamma, b_{P,i}, c_{P,i})\right|^{1/2} \\ + o_{\mathcal{F}(M)\to\infty;M}(1) + o_{\varepsilon' N\to\infty;M}(1) + O_M(\varepsilon').\end{aligned} \tag{3.6}$$

Now recall that $F_i(g(n)\Gamma, n \pmod q, n/N) = f_i(n) - f_{i,\mathrm{sml}}(n) - f_{i,\mathrm{unf}}(n)$. Then substituting $F_i(g(n)\Gamma, n \pmod q, n/N) = f_i(n) - f_{i,\mathrm{sml}}(n) - f_{i,\mathrm{unf}}(n)$ into (3.6), we claim that the term with f_i is of size $o_{\delta_i\to 0;M,\varepsilon'}(1) + o_{N\to\infty;M,\varepsilon'}(1)$. Indeed, by Proposition 3.4.4 we have that $||f_i||_{U^{s+1}(\tilde{P}_i)} = O_{N\to\infty;\varepsilon'}(\delta_i) + o_{N\to\infty;\varepsilon'}(1)$. Also, observe that $\overline{F_{i,\leq s}}(g(n)\Gamma, b_{P,i}, c_{P,i})$ is an s-step nilsequence of complexity $O_M(1)$. Thus our claim follows by invoking the converse to the inverse theorem for Gowers norms on $\tilde{P}_i$. Similarly, we may conclude that the term with $f_{i,\mathrm{unf}}$ is of size $o_{\mathcal{F}(M)\to\infty;M,\varepsilon'}(1) + o_{N\to\infty;M,\varepsilon'}(1)$ by the same argument and the additional ingredients that $s' \geq s$ and that Gowers norms are monotonic (up to a constant factor) in s.

For the term with $f_{i,\mathrm{sml}}$ we need to utilise the average over $P \in \mathcal{P}$. First we compute the following using the Cauchy-Schwarz inequality and convexity:

$$\begin{aligned}\min_i \mathbb{E}_{P\in\mathcal{P}} \left|\mathbb{E}_{n\in\tilde{P}_i} f_{i,\mathrm{sml}}(n)\overline{F_{i,\leq s}}(g(n)\Gamma, b_{P,i}, c_{P,i})\right|^{1/2} &\leq \min_i \mathbb{E}_{P\in\mathcal{P}} \left(\mathbb{E}_{n\in\tilde{P}_i}|f_{i,\mathrm{sml}}(n)|^2\right)^{1/4} \\ &\leq \min_i \left(\mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{n\in\tilde{P}_i}|f_{i,\mathrm{sml}}(n)|^2\right)^{1/4}.\end{aligned}$$

Now we need the following lemma which implies that as P varies in $\mathcal{P}$, the progressions $\tilde{P}_i$ are sufficiently well distributed in $[N]$.

Lemma 3.2.5. *For each $n \in [N]$, there are at most $O(\varepsilon'^{-(D-1)})$ elements $P \in \mathcal{P}$ such that $n \in \psi_i(P) + c$.*

Proof. One may assume that $n \in \psi_i(K\backslash S) + c$ because otherwise the statement is trivial. Let $T := \psi_i^{-1}(n - c)$. Any cube (as in the statement of Lemma 3.2.3) which contains a point in T lies entirely in the $\sqrt{D}q\varepsilon' N$-neighbourhood of T; denote this neighbourhood by U. Note that U has volume $O(q\varepsilon' N^D)$. On the other hand, the volume of one of these cubes is $(q\varepsilon' N)^D$, and so U can contain at most $O(\varepsilon'^{-(D-1)}/q^{D-1})$ distinct cubes. Fix such a cube Q and recall that Q comprises q^D different $P \in \mathcal{P}$. Let $\mathcal{P}_Q$ be those $P \in \mathcal{P}$ which are contained in Q; it remains to argue that at most $O(q^{D-1})$ elements of $\mathcal{P}_Q$ can contain a point in T.

Note that $P \cap T \neq \emptyset$ implies that $\psi_i(P) + c = n \pmod q$, which is well-defined by the q-periodicity of P in all coordinate directions. Thus it suffices to bound

the number of $P \in \mathcal{P}_Q$ with $\psi_i(P) = n - c \pmod q$. Every $P \in \mathcal{P}_Q$ has exactly one representative in a fundamental domain F for $\mathbb{Z}^D/(q\mathbb{Z})^D$ which is chosen to be contained in Q. Furthermore the map $\psi_i : \mathbb{Z}^D \to \mathbb{Z}/q\mathbb{Z}$ is a group homomorphism which factors through the quotient $\mathbb{Z}^D/(q\mathbb{Z})^D$; denote the corresponding map from $\mathbb{Z}^D/(q\mathbb{Z})^D$ to $\mathbb{Z}/q\mathbb{Z}$ by $\overline{\psi_i}$. Thus we may view $\overline{\psi_i}$ as a group homomorphism on F and so the number of $P \in \mathcal{P}_Q$ with $\psi_i(P) = n - c \pmod q$ is either 0 or $|\ker \overline{\psi_i}|$. But $|\ker \overline{\psi_i}|$ is easily seen to be of size $O(q^{D-1})$ by invoking the fact that the size of the coefficients of ψ_i are $O(1)$. □

Since $|K| \gg N^D$ and recalling Lemma 3.2.3, we have $|\mathcal{P}| \geq C\varepsilon'^{-D}$. Also, $|\tilde{P}_i| = \Theta(\varepsilon' N)$, and so using the previous lemma in the penultimate line,

$$\begin{aligned}
\mathbb{E}_{P\in\mathcal{P}}\mathbb{E}_{n\in\tilde{P}_i}|f_{i,\mathrm{sml}}(n)|^2 &\leq \frac{\varepsilon'^{(D-1)}}{CN}\sum_{P\in\mathcal{P},n\in\tilde{P}_i}|f_{i,\mathrm{sml}}(n)|^2 \\
&= \frac{\varepsilon'^{(D-1)}}{CN}\sum_{n\in[N]}|f_{i,\mathrm{sml}}(n)|^2\#\{P\in\mathcal{P} : n\in\tilde{P}_i\} \\
&\leq \frac{1}{C}||f_{i,\mathrm{sml}}||_2^2 \\
&= O(\varepsilon^2).
\end{aligned}$$

All in all, since (3.6) we have shown that

$$\begin{aligned}
&\min_i \mathbb{E}_{P\in\mathcal{P}}\left|\mathbb{E}_{n\in\tilde{P}_i}F_i(g(n)\Gamma, n \pmod q, n/N)\overline{F_{i,\leq s}}(g(n)\Gamma, b_{P,i}, c_{P,i})\right|^{1/2} \\
&\qquad \leq \min_i o_{\delta_i\to 0;M,\varepsilon'}(1) + o_{\mathcal{F}(M)\to\infty;M,\varepsilon'}(1) + o_{N\to\infty;M,\varepsilon'}(1) + O(\varepsilon^{1/2}). \qquad (3.7)
\end{aligned}$$

Recall that $\min_i \delta_i \leq \delta$ so that as $\delta \to 0$ we have $\min_i \delta_i \to 0$ and so we may write $\min_i o_{\delta_i\to 0;M,\varepsilon'}(1) = o_{\delta\to 0;M,\varepsilon'}(1)$. Putting together the results from Equations (3.1)–(3.7), we have

$$\begin{aligned}
\left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x})+c)\right| &\leq O(\varepsilon) + o_{N\to\infty;\mathcal{F}(M)}(1) + O_M(\varepsilon') + o_{\mathcal{F}(M)\to\infty;M,\varepsilon'}(1) \\
&\quad + o_{\varepsilon' N\to\infty;M}(1) + o_{\delta\to 0;M,\varepsilon'}(1) + o_{N\to\infty;M,\varepsilon'}(1) + O(\varepsilon^{1/2}).
\end{aligned}$$

We complete the proof of Theorem 3.2.1 by choosing ε' to be sufficiently small depending on M and ε, choosing $\mathcal{F}$ to be a sufficiently rapidly growing function depending on ε' and the function implicit in the $o_{\mathcal{F}(M)\to\infty;M,\varepsilon'}(1)$ notation, choosing δ to be sufficiently small depending on M and ε' and setting N to be sufficiently large depending on ε', M, $\mathcal{F}$ and ε. □

3.3 Proof of Theorem 3.1.1

In what follows we will use notation from the statement of Theorem 3.1.1 without explicitly reintroducing it. In particular, in this section $\Psi = (\psi_1, \dots, \psi_t)$ is a system of linear forms which is not necessarily flag. As in the previous section, we will allow all constants to depend on Ψ, D, s, t and let C be a constant which may change line to line. Note also that we may take N arbitrarily large in terms of ε because the case where N is bounded in terms of ε may be handled with exactly the same argument as it was in the proof of Theorem 3.2.1.

We claim that there is a sequence of scalars $(a_i)_{i=1}^t$ each in $\mathbb{Z}\backslash\{0\}$ and of size $O(1)$ such that the family $\tilde{\Psi} := (a_i\psi_i)_{i=1}^t$ satisfies the flag property. Indeed, since all ψ_i are nonzero, the image of Ψ is a linear space which is not contained in any hyperplane $x_i = 0$, and so it contains a vector which is nonzero in every component; denote a choice of such a vector by $(b_1, \dots, b_t)$. Then setting $a_i = \prod_{j\neq i} b_j$ ensures that the image of $\tilde{\Psi} := (a_i\psi_i)_{i=1}^t$ contains a nonzero scalar multiple of $(1, 1, \dots, 1) \in \mathbb{R}^t$ so that $\tilde{\Psi}$ is translation invariant and hence satisfies the flag property.

Let $(a_i)_{i=1}^t$ be integers as above and let $a = \max_i |a_i|$. Define $\tilde{f_i}$ on $a_i[-N, N]+aN$ by $\tilde{f_i}(x) = f_i((x - aN)/a_i)$. Extend $\tilde{f_i}$ to a 1-bounded function on $[2aN]$ by letting $\tilde{f_i}(x) = 0$ for $x \in [2aN]\backslash(a_i[-N, N] + aN)$.

Lemma 3.3.1. *The following inequality holds:*

$$||\tilde{f_i}||_{U^{s+1}[2aN]} \leq ||f_i||_{U^{s+1}[-N,N]} + o_{N\to\infty}(1).$$

Proof. We compute directly from the definition of Gowers norms (cf. Section 3.1.1 above), recalling that $\tilde{f_i}(x) = 0$ for $x \in [2aN]\backslash(a_i[-N, N] + aN)$:

$$\begin{aligned}
||\tilde{f_i}||^{2^{s+1}}_{U^{s+1}[2aN]} &= \mathbb{E}_{\substack{x,h:x+\omega\cdot h\in[2aN]\\ \text{for all } \omega\in\{0,1\}^{s+1}}} \prod_{\omega\in\{0,1\}^{s+1}} \mathcal{C}^{|\omega|}\tilde{f_i}(x + \omega \cdot h) \\
&= \frac{|\{(x, h) : x + \omega \cdot h \in a_i[-N, N] + aN \text{ for all } \omega \in \{0, 1\}^{s+1}\}|}{|\{(x, h) : x + \omega \cdot h \in [2aN] \text{ for all } \omega \in \{0, 1\}^{s+1}\}|}\times \\
&\quad \mathbb{E}_{\substack{x,h:x+\omega\cdot h\in a_i[-N,N]+aN\\ \text{for all } \omega\in\{0,1\}^{s+1}}} \prod_{\omega\in\{0,1\}^{s+1}} \mathcal{C}^{|\omega|}\tilde{f_i}(x + \omega \cdot h) \\
&\leq \mathbb{E}_{\substack{x,h:x+\omega\cdot h\in a_i[-N,N]+aN\\ \text{for all } \omega\in\{0,1\}^{s+1}}} \prod_{\omega\in\{0,1\}^{s+1}} \mathcal{C}^{|\omega|}\tilde{f_i}(x + \omega \cdot h) + o_{N\to\infty}(1) \\
&= ||\tilde{f_i}||_{U^{s+1}(a_i[-N,N]+aN)} + o_{N\to\infty}(1).
\end{aligned}$$

But one sees easily that $||\tilde{f_i}||_{U^{s+1}(a_i[-N,N]+aN)} = ||f_i||_{U^{s+1}[-N,N]}$ since $f_i = \tilde{f_i} \circ \phi$ where $\phi(x) = a_ix + aN$, which is a Freiman isomorphism on the respective sets. □

Let $K = [-N,N]^D \cap \Psi^{-1}([-N,N]^t)$. Then we have

$$\left|\mathbb{E}_{\boldsymbol{x}\in[-N,N]^D}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x}))\right| \leq \left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x}))\right| = \left|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t \tilde{f}_i(a_i\psi_i(\boldsymbol{x})+aN)\right|. \tag{3.8}$$

By insisting that $\min_i ||f_i||_{U^{s+1}[-N,N]} \leq \delta$ is sufficiently small and N is sufficiently large, we may force $\min_i ||\tilde{f}_i||_{U^{s+1}([2aN])}$ to be as small as we like by Lemma 3.3.1. Now, the system of linear forms $\tilde{\Psi} = (a_i\psi_i)_{i=1}^t$ is flag, and $(a_1\psi_1)^{s+1},\dots,(a_t\psi_t)^{s+1}$ are linearly independent since $(\psi_1^{s+1},\dots,\psi_t^{s+1})$ are. Furthermore, one may check that the volume of K is equal to $|K| + O(N^{D-1})$ (in fact, this statement is proven in [36, Appendix A] for K convex). Also, one may show easily that the volume of K is of size $\Theta(N^D)$. Thus we may invoke Theorem 3.2.1 (with functions on $[2aN]$ rather than $[N]$) to force $|\mathbb{E}_{\boldsymbol{x}\in K}\prod_{i=1}^t \tilde{f}_i(a_i\psi_i(\boldsymbol{x})+aN)|$ to be as small as we like. In turn we may make $|\mathbb{E}_{\boldsymbol{x}\in[-N,N]^D}\prod_{i=1}^t f_i(\psi_i(\boldsymbol{x}))|$ as small as we like from Equation (3.8). This completes the proof of Theorem 3.1.1.

3.4 Appendix: Results from higher order Fourier analysis

In this appendix we collect some results from higher order Fourier analysis which we need at various points in the chapter. All are known to experts and/or are small perturbations of results that already appear in the literature.

Theorem 3.4.1 (Simultaneous arithmetic regularity lemma)**.** *Let $f_1,\dots,f_t : [N] \to [0,1]$ be functions, let $s \geq 1$ be an integer, let $\varepsilon > 0$, and let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a growth function. Then there exists a quantity $M = O_{s,\varepsilon,\mathcal{F},t}(1)$ and decompositions*

$$f_i = f_{i,\mathrm{nil}} + f_{i,\mathrm{sml}} + f_{i,\mathrm{unf}}$$

of each f_i into functions $f_{i,\mathrm{nil}}, f_{i,\mathrm{sml}}, f_{i,\mathrm{unf}} : [N] \to [-1,1]$ where

1. *($f_{i,\mathrm{nil}}$ structured) $f_{i,\mathrm{nil}}$ is a $(\mathcal{F}(M),N)$-irrational virtual nilsequence of degree at most s, complexity at most M and scale N, and we may write $f_{i,nil}(n) = F_i(g(n)\Gamma, n \pmod q, n/N)$, where g is a polynomial sequence adapted to a filtered nilmanifold and Γ is a cocompact lattice on that nilmanifold,*

2. *($f_{i,\mathrm{sml}}$ small) $f_{i,\mathrm{sml}}$ has $L^2[N]$-norm at most ε,*

3. *($f_{i,\mathrm{unf}}$ very uniform) $f_{i,\mathrm{unf}}$ has $U^{s+1}[N]$-norm at most $1/\mathcal{F}(M)$,*

4. *(Nonnegativity) $f_{i,\mathrm{nil}}$ and $f_{i,\mathrm{nil}} + f_{i,\mathrm{sml}}$ take values in $[0,1]$.*

In particular, the functions $\{f_{i,\mathrm{nil}}\}_{i=1}^t$ differ only by the choice of Lipschitz function $\{F_i\}_{i=1}^t$.

Proof. This is [35, Theorem 1.2] together with the extra ingredient that the nilsequences $f_{i,\mathrm{nil}}$ differ only by the choice of Lipschitz function. The existence of this result is well-known to experts.

One derives the above statement from [35] by invoking the non-irrational regularity lemma [35, Proposition 2.7] at each i, forming the product polynomial sequence $g(n) = (g_1(n), \dots, g_t(n))$ on the product nilmanifold $\prod_{i=1}^t G_i / \prod_{i=1}^t \Gamma_i$ and then factorising for irrationality as in [35, Chapter 2]. We omit the details but remark that they are worked out in a similar setting in [51, Appendix A]. □

We refer the reader to [35] for further background on the theorem that follows.

Theorem 3.4.2 (Flag counting lemma, [35, Theorem 1.11])**.** *Let M, D, t, s be positive integers with $D, t, s \leq M$, let $(G/\Gamma, G_\bullet)$ be a degree $\leq s$ filtered nilmanifold of complexity $\leq M$, let $g : \mathbb{Z} \to G$ be an (A,N)-irrational polynomial sequence adapted to $G_\bullet$, let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of linear forms with coefficients of magnitude at most M and which satisfies the flag condition, let P be a convex subset of $[-N,N]^D$ with positive Lebesgue measure, let $\Lambda \leq \mathbb{Z}^D$ be a sublattice of index $[\mathbb{Z}^D : \Lambda] \leq M$ and let $\boldsymbol{x}_0 \in \mathbb{Z}^D$. Then for any sequence of M-Lipschitz functions $F_1, \dots, F_t : G/\Gamma \to \mathbb{C}$, we have*

$$\mathbb{E}_{\boldsymbol{x} \in (\boldsymbol{x}_0 + \Lambda) \cap P} \prod_{i=1}^t F_i(g(\psi_i(\boldsymbol{x}))\Gamma) = \int_{g(0)^\Delta G^\Psi / \Gamma^\Psi} \otimes_{i=1}^t F_i + o_{A \to \infty; M}(1) + o_{N \to \infty; M}(1),$$

where $g(0)^\Delta := (g(0), \dots, g(0)) \in G^t$.

Recall the definition of the *Cauchy-Schwarz complexity* of a system of linear forms (cf. [65, Definition 1.3.2]).

Theorem 3.4.3 (Generalised von Neumann inequality on convex sets)**.** *Let t, D be positive integers and let $L > 0$. Let $f_1, \dots, f_t : [N] \to \mathbb{C}$ be 1-bounded. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms with constant term of size at most LN and linear coefficients of size at most L. Let $K \subset [N]^D$ be convex and such that*

$\Psi(K) \subset [N]^t$. *Then letting s be the Cauchy-Schwarz complexity of Ψ and letting δ be such that $\min_{i\in[t]} ||f_i||_{U^{s+1}[N]} \leq \delta$ we have*

$$\mathbb{E}_{\boldsymbol{x}\in K} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x})) = o_{\delta\to 0;L}(1) + o_{N\to\infty;\delta,L}(1).$$

Proof. This is essentially [36, Proposition 7.1], except we are dealing with functions f_i which are 1-bounded rather than bounded by a pseudorandom measure. The proof of [36, Proposition 7.1] which is given in [36, Appendix C] is easily adapted to recover the statement above. □

Proposition 3.4.4 (Gowers uniformity on subprogressions)**.** *Let $f : [N] \to \mathbb{C}$ be 1-bounded, let $\varepsilon > 0$ and let P be a subprogression of $[N]$ of size εN. Then*

$$||f||_{U^{s+1}(P)} = O_{N\to\infty;\varepsilon}(||f||_{U^{s+1}[N]}) + o_{N\to\infty;\varepsilon}(1).$$

Proof. In this proof, all asymptotic notation is implicitly with respect to the limit $N \to \infty$. From Subsection 3.1.1 we have $||f||_{U^{s+1}(P)} = \Theta_\varepsilon(||1_P f||_{U^{s+1}[N]})$. Let $I_\varepsilon = [-\varepsilon N/2, \varepsilon N/2]$. Write $P = qI_\varepsilon + c$. Let χ_0 be the function $\mathbb{Z} \to \mathbb{R}$ which is 1 on I_ε, which is zero when $|x| > (1 + 1/\sqrt{N})\varepsilon N/2$, and which linearly interpolates otherwise (χ_0 is the Fourier transform of a de la Vallée Poussin kernel). It is standard that $||\hat{\chi}_0||_1 = O_\varepsilon(1)$ (one may see this by writing χ_0 as the convolution of suitably normalised characteristic functions of intervals). Define χ by $\chi(x) = \chi_0((x-c)/q)$ for all $x = c \pmod q$ and zero otherwise. Then we have $||\hat{\chi}||_1 = O_\varepsilon(1)$ and furthermore that $1_P - \chi$ is supported on a set of size $o_\varepsilon(N)$. The former property and Fourier inversion yield that

$$||\chi f||_{U^{s+1}[N]} \leq ||\hat{\chi}||_1 \sup_\theta ||e(\theta\cdot)f||_{U^{s+1}[N]} = O_\varepsilon(||f||_{U^{s+1}[N]}),$$

using that the $U^{s+1}[N]$-norm is linear phase invariant, that is $||e(\theta\cdot)f||_{U^{s+1}[N]} = \Theta\left(||f||_{U^{s+1}[N]}\right)$; see [65, Exercise 1.3.21]. The latter property yields that $||(1_P - \chi)f||_{U^{s+1}[N]} = o_\varepsilon(1)$. We are ready to conclude:

$$||1_P f||_{U^{s+1}[N]} \leq ||\chi f||_{U^{s+1}[N]} + ||(1_P - \chi)f||_{U^{s+1}[N]} = O_\varepsilon(||f||_{U^{s+1}[N]}) + o_\varepsilon(1).$$

□

Chapter 4

A non-flag arithmetic regularity lemma and counting lemma

4.1 Introduction

4.1.1 Context and summary

The use of 'regularity lemmas' in combinatorics dates back to Szeméredi's proof of his eponymous theorem on arithmetic progressions in dense sets of integers [64]. Szemerédi's regularity lemma and developments thereof have since become powerful tools in combinatorics which have been brought to bear on a far wider range of problems than the study of arithmetic progressions in sets of integers. Where Szemerédi's regularity lemma is a structural result about graphs, Green introduced an arithmetic analogue of this brand of regularity results in [32], and used it in that paper to solve some problems in additive combinatorics. From the perspective of applications to additive patterns in the integers, the state of the art is Green and Tao's arithmetic regularity lemma [35].

As we saw in the previous chapter, Green and Tao's arithmetic regularity lemma is phrased as a decomposition result for bounded functions which consequently may be written as the sum of a nilsequence, a Gowers uniform function, and an error. [35] demonstrates a number of applications for which the contributions of the uniform and error components are negligible and so the problem is essentially reduced to an analysis of nilsequences. Complementing the arithmetic regularity lemma in [35] is an arithmetic counting lemma which gives a general formula (via equidistribution on nilmanifolds) for computing averages on linear patterns in the nilsequence component.

It has come to light (see [40], [66]) that this counting lemma [35, Theorem 1.11] applies only to systems of linear forms which satisfy the so-called *flag condition* (see Subsection 4.1.3 for definitions), whence the arithmetic regularity and counting lemma

strategy may be exercised only on this particular set of linear patterns. Further to this, there are examples (see Subsection 4.2.1) which demonstrate that the output of the arithmetic regularity lemma [35, Theorem 1.2] is too coarse to determine the distribution of non-flag linear patterns in any prospective counting lemma.

The goal of this chapter is to give an arithmetic regularity lemma and complementary counting lemma which are able to handle all linear patterns. The main results are a stronger arithmetic regularity lemma Theorem 4.1.4, which feeds into a new counting lemma Theorem 4.1.2, which itself is able to handle all sets of linear forms. We remark that our description of the nilmanifold on which we obtain equidistribution in the counting lemma is less explicit than what is worked out for the Leibman group in [35, Section 3] (though this is partially remedied in Section 4.3 below). Nonetheless, the broadstrokes 'arithmetic regularity and counting' strategy for solving problems in additive combinatorics can now be applied to all linear patterns.

We also give a couple of applications; these are discussed in the next subsection.

4.1.2 Results

Given a flag system of linear forms Ψ, a filtered Lie group $G_\bullet$ and a polynomial sequence $g : \mathbb{Z} \to G$ which is suitably 'irrational' with respect to this filtration, Green and Tao's counting lemma determines (quantitatively) the distribution of the tuple $g^\Psi(\boldsymbol{n}) := ((g(\psi_1(\boldsymbol{n})), \dots, g(\psi_t(\boldsymbol{n})))\Gamma^t)_{\boldsymbol{n}\in\mathbb{Z}^D}$ in G^t/Γ^t from the data $\{G_\bullet, \Psi\}$. Unfortunately, this data is insufficient to determine the distribution of g^Ψ in the general, non-flag setting in the sense that there exist a non-flag system of linear forms Ψ and distinct polynomial sequences g, $\tilde{g}$, both of which are irrational in the same $G_\bullet$, such that g^Ψ distributes differently to $\tilde{g}^\Psi$; see Examples 4.2.2 and 4.2.3. Instead of dealing with a filtered Lie group $G_\bullet$, we will consider a sequence of subspaces $S_\bullet$ of the Lie algebra $\mathfrak{g}$ of G. One may define a notion of irrationality – which we will call 'linear irrationality' – with respect to a sequence of subspaces such that for all linearly irrational polynomial sequences g with respect to $S_\bullet$, one may determine the distribution of g^Ψ from the data $\{S_\bullet, \Psi\}$. It will prove convenient to describe the nilmanifold on which g^Ψ equidistributes as a Lie subalgebra of $\mathfrak{g}^t$. We will denote this Lie subalgebra by $\mathfrak{g}^\Psi(S_\bullet)$, define it in terms of Ψ and $S_\bullet$ in Definition 4.2.14 and then compute it more explicitly in Section 4.3.

Working in the Lie algebra, we will denote group multiplication by $*$.

Definition 4.1.1. Define the operation $* : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ by $x * y := \log(\exp(x)\exp(y))$.

We denote group multiplication in this way partially to avoid the temptation to accidentally distribute over addition. By Baker–Campbell–Hausdorff, we have explicitly that

$$x * y = x + y + \frac{1}{2}[x, y] + \frac{1}{12}[x, [x, y]] + \dots.$$

All that we will need about the Baker–Campbell–Hausdorff formula can be found in [35, Appendix C]. We will now state our counting lemma. Although there are a number of requisite definitions which have not yet been provided (rational basis: Definition 4.4.1, rational subspace: Definition 4.2.12, complexity of a rational subspace: Definition 4.4.6, linear irrationality: Definitions 4.4.7 and 4.4.8, quantitative equidistribution: Definitions 4.8.9 and 4.4.14, Hermite basis: Definition 4.9.2), we hope that the reader may obtain a rough appreciation of the content of the result. It is proven as Theorem 4.4.16 in Section 4.4.

Theorem 4.1.2 (Counting lemma)**.** *Let $A, M \geq 2$ and $d, s, t \geq 1$ be integers. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with rational basis $\mathcal{X}$, dimension d and rational structure constants of height at most M. Let $S_\bullet$ be a sequence of rational subspaces of complexity M in $\mathfrak{g}$. Suppose that $p(n)$ is an (A, N)-linearly irrational polynomial sequence of degree s in $S_\bullet$. Then there are constants $0 < c_1, c_2 = O_{d,s,D,t,\Psi}(1)$ such that the polynomial sequence $p^\Psi(\boldsymbol{x}) := (p(\psi_1(\boldsymbol{x})), \dots, p(\psi_t(\boldsymbol{x}))$ is $O(M^{c_1}/A^{c_2})$-equidistributed on $\mathfrak{g}^\Psi(S_\bullet)$ with respect to $\mathcal{X}^\Psi$, the Hermite basis for $\mathfrak{g}^\Psi(S_\bullet)$ in $(\mathfrak{g}^t, \mathcal{X}^t)$.*

The next task is to obtain a factorisation result for the notion of linear irrationality. The following theorem says that any polynomial sequence may be written as the product of an 'error' sequence, a sequence which is linearly irrational with respect to some sequence of subspaces $S_\bullet$, and a rational sequence. Furthermore, the factorisation process is such that if we begin with a polynomial sequence which is irrational in the Green-Tao sense (we will henceforth call this 'filtration irrational'), then the linearly irrational factorisation is still (though slightly less) Green-Tao (i.e. filtration) irrational. There are a number of additional definitions (filtration: Definition 4.2.4, polynomial sequence adapted to a filtration: Definition 4.2.5, quantitative rationality: Definition 4.5.2, quantitative smallness: Definition 4.5.3, quantitative filtration irrationality: Definitions 4.5.6 and 4.5.7) required for a rigorous interpretation.

Theorem 4.1.3 (Factorisation of polynomial sequences)**.** *Let $A, M \geq 2$ and $d, s, t \geq 1$ be integers such that $AM > d$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with filtration $\mathfrak{g}_\bullet$, rational basis $\mathcal{X}$, dimension d and rational structure constants of height at most*

M. Let p be a polynomial sequence adapted to $\mathfrak{g}_\bullet$ and sequence of subspaces $S_\bullet \leq \mathfrak{g}_\bullet$. Finally, insist that $S_\bullet, \mathfrak{g}_\bullet$ are of complexity $A^{O(1)}$ in $\mathfrak{g}$, where here and henceforth $O(1)$ terms may depend on d, s. Then we may write $p = e*p'*r$ where e, p', r are polynomial sequences in $\mathfrak{g}$ adapted to $\mathfrak{g}_\bullet$, p' is (A, N)-linearly irrational in a sequence of subspaces $S'_\bullet$ of complexity $A^{O(1)}$, r is $(AM)^{O(1)}$-rational, and e is $((AM)^{O(1)}, N)$-small.

This theorem is proven in Section 4.5 as Theorem 4.5.10 (where we actually prove a bit more about the factorisation). Our final goal is to provide an arithmetic regularity lemma which produces a virtual nilsequence which is both filtration irrational and linearly irrational; we call such a nilsequence 'strongly irrational' (cf. Definition 4.5.11). Though filtration irrationality is not strictly necessary to determine distribution, it is useful in applications and furthermore means that our arithmetic regularity lemma is more compatible with, and explicitly strengthens, what is done in [35].

Theorem 4.1.4 (Strongly irrational arithmetic regularity lemma)**.** *Let $f : [N] \to [0,1]$, let $s \geq 1$, let $\varepsilon > 0$, and let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a growth function. Then there exists a quantity $M = O_{s,\varepsilon,\mathcal{F}}(1)$ and a decomposition*

$$f = f_{\mathrm{nil}} + f_{\mathrm{sml}} + f_{\mathrm{unf}}$$

of f into functions $f_{\mathrm{nil}}, f_{\mathrm{unf}} : [N] \to [-1,1]$ such that

1. *(f_{nil} structured) f_{nil} is a degree $\leq s$, $(\mathcal{F}(M), N)$-strongly irrational virtual polynomial nilsequence of complexity $\leq M$,*
2. *(f_{sml} small) $||f_{\mathrm{sml}}||_{L^2[N]} \leq \varepsilon$,*
3. *(f_{unf} very uniform) $||f_{\mathrm{nil}}||_{U^{s+1}[N]} \leq 1/\mathcal{F}(M)$,*
4. *(Nonnegativity) f_{nil} and f_{sml} take values in $[0,1]$.*

We also provide a couple of applications. The first is the modest task of recovering (with polynomial bounds) the flag counting lemma for strongly irrational nilsequences.

Theorem 4.1.5 (Equidistribution in the Leibman group for flag systems)**.** *Let $A, M \geq 2$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a flag system of linear forms, each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra of dimension d and step s. Let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$ which passes through $\mathfrak{g}_\bullet$ and with respect to which $\mathfrak{g}$ has rational structure constants of height at most M. Let $S_\bullet \leq \mathfrak{g}_\bullet$ be a sequence of subspaces of complexity at most M. If p is (A, N)-strongly irrational in $(\mathfrak{g}_\bullet, S_\bullet)$ then there are constants $0 < c_1, c_2 = O_{d,s,D,t,\Psi}(1)$ such that p^Ψ is $O(M^{c_1}/A^{c_2})$-equidistributed in $\log G^\Psi$, where G^Ψ is the Leibman group for $G_\bullet, \Psi$.*

This is proven in Subsection 4.7.1, where we show that under appropriate assumptions, $\mathfrak{g}^{\Psi}(S_{\bullet}) = \log G^{\Psi}$.

Next, in Subsection 4.7.2 we give another resolution of a conjecture of Gowers and Wolf [28, Conjecture 2.5], which we recovered in the previous chapter in the generality of all linear forms by employing a somewhat specialised linear algebraic trick. The point of Subsection 4.7.2 is to provide a less ad-hoc argument. Indeed, Green and Tao's proof in the flag setting [35, Theorem 1.13] may be followed after doing some modest algebraic computations (Proposition 4.7.5) and employing the strongly irrational arithmetic regularity lemma Theorem 4.1.4 and counting lemma Theorem 4.1.2 in place of the original irrational arithmetic regularity lemma [35, Theorem 1.2] and counting lemma [35, Theorem 1.11] respectively.

Theorem 4.1.6 (Gowers-Wolf Conjecture)**.** *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$, and let $s \geq 1$ be an integer such that the polynomials $\psi_1^{s+1}, \dots, \psi_t^{s+1}$ are linearly independent. For $i = 1, \dots t$, let $f_i : [-N, N] \to \mathbb{C}$ be functions bounded in magnitude by 1 (and defined to be zero outside of $[-N, N]$). For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $\min_i ||f_i||_{U^{s+1}[-N,N]} \leq \delta$, then*

$$\left| \mathbb{E}_{\boldsymbol{x} \in [-N,N]^D} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x})) \right| \leq \varepsilon.$$

4.1.3 On the flag condition

Let $\Psi = (\psi_1, \dots, \psi_t)$ where each ψ_i maps $\mathbb{Z}^D$ to $\mathbb{Z}$. For each $j \in \mathbb{N}^+$, one may write $\Psi^j(\boldsymbol{x}) = \sum_{m \in \mathcal{M}} v_{m,j} m(\boldsymbol{x})$, where $v_{m,j} \in \mathbb{Z}^t$, $m \in \mathcal{M}$ is a 'monomial map' which maps $\boldsymbol{x} = (x_1, \dots, x_D)$ to some monomial in the variables $\{x_1, \dots, x_D\}$, and $v_{m,j}$ will be nonzero only for monomials of degree j. To elucidate the notation, we compute with the specific example

$$\Psi(x, y) := (\psi_1(x, y), \psi_2(x, y), \psi_3(x, y), \psi_4(x, y)) := (y, 2x + 2y, x + 3y, x),$$

that

$$\Psi^1(x, y) = (0, 2, 1, 1)x + (1, 2, 3, 0)y,$$

and

$$\Psi^2(x, y) = (0, 4, 1, 1)x^2 + (1, 4, 9, 0)y^2 + (0, 8, 6, 0)xy.$$

For $i \in \mathbb{N}^+$ we define the vector spaces $V_i = \mathrm{span}_{\mathbb{R}}\{v_{m,i}\}_{m \in \mathcal{M}}$ and let $V = V_1$. In fact, it is not difficult to see that $V_i = V^i$ where V^i is the smallest vector space containing i-fold products of elements of V, where multiplication is conducted coordinatewise.

A system of linear forms Ψ satisfies the *flag condition* if $V^i \leq V^j$ whenever $i \leq j$. One may check that the system of linear forms in the example above does not satisfy the flag condition (i.e. *is not flag*) because V is not contained in V^2.

In what follows, if context is sufficiently clear, we will use V to denote the vector subspace of $\mathbb{R}^t$ formed in this way from Ψ. If context is less clear, we may also sometimes use the notation V_Ψ.

4.1.4 Outline of this chapter

In Section 4.2 we begin by providing examples which motivate the transition from Lie group to Lie algebra. Then we work out the qualitative theory in the Lie algebra setting, giving a qualitative counting lemma Theorem 4.2.16 and qualitative factorisation theorem Proposition 4.2.28. We encourage a reader who wishes to understand the ideas behind Theorems 4.1.2 and 4.1.3 to at least glean from this section the main algebraic strategy before progressing to the quantitative sections that follow. The theory is substantially cleaner here and we hope that this section convinces the reader that this is the right setup in general. This section may also be of independent interest.

The qualitative counting lemma in Section 4.2 describes a particular Lie algebra $\mathfrak{g}^\Psi(S_\bullet)$ on which we obtain equidistribution. This is the same Lie algebra that appears in the statement of Theorem 4.1.2. In Section 4.3 we give more explicit and tenable algebraic descriptions of this Lie algebra.

In Section 4.4, we prove our quantitative counting lemma, Theorem 4.1.2.

In Section 4.5, we prove our quantitative factorisation of polynomial sequences for linear irrationality, Theorem 4.1.3.

In Section 4.6, we prove our quantitative 'strongly irrational' arithmetic regularity lemma, Theorem 4.1.4.

In Section 4.7 we give applications as discussed above.

4.2 Examples, discussion of methods and qualitative results

In this section we begin with some examples which demonstrate key phenomena in the qualitative setting. This motivates our methods and setup and in particular how they deviate from those in [35]. Then we work out the infinitary distribution of polynomial sequences on linear patterns. This includes most of the key ideas and phenomena that will appear in the quantitative setting (which is what we will

ultimately use for our applications in Section 4.7), but is not muddied by quantitative bounds or the technical baggage introduced in their pursuit. This section will also be interspersed with examples. We hope that this section is edifying in preparation for the more arduous task of reading Sections 4.4 and 4.5.

Recall that a filtration on a nilpotent simply-connected Lie group G is a decreasing sequence of simply-connected Lie subgroups $G = G_0 = G_1 \geq G_2 \geq \cdots$ with the property that $[G_i, G_j] \leq G_{i+j}$ for all i, j. A polynomial sequence $g(n) := \prod_i g_i^{\binom{n}{i}}$ on G is said to be adapted to a filtered Lie group $G_\bullet$ if $g_i \in G_i$ for all i. As a general disclaimer, in this section we may use some standard terminology from this area without explicit introduction. We refer the reader to [35] or [38] for the development of the basic theory used here (on such an expedition, readers really need only acquaint themselves with basic qualitative definitions to make this rather informal section legible; we will provide our own definitions when doing things rigorously later on.)

4.2.1 Non-flag examples

In the flag case, Green and Tao's counting lemma computes averages of the form $\mathbb{E}_{\boldsymbol{x}} \prod_{i=1}^t \phi(\psi_i(\boldsymbol{x}))$ where $\phi = f \circ g$ is a suitably 'irrational' nilsequence in some filtered nilmanifold $G_\bullet$. Roughly speaking, if $G_\bullet$ is the 'correct' filtered nilmanifold for the polynomial sequence g (i.e. if g cannot be written as a polynomial sequence on cosets of some smaller filtered subnilmanifold), then the distribution of $g^\Psi(\boldsymbol{x}) := (g(\psi_1(\boldsymbol{x})), \ldots, g(\psi_t(\boldsymbol{x})))$ in G^t/Γ^t may be determined from the data $\{G_\bullet, \Psi\}$.

Definition 4.2.1 (The Leibman group)**.** Let $G_\bullet$ be a filtered nilmanifold and let Ψ be a family of linear forms. Then define $G^\Psi := \langle g_i^{v_i}, g_i \in G_i, v_i \in V^i \rangle$, where $V = V_\Psi$ is as defined in Subsection 4.1.3.

The flag counting lemma says that if g is irrational in $G_\bullet$ then g^Ψ equidistributes in $G^\Psi/G^\Psi \cap \Gamma^t$, and so the average $\mathbb{E}_{\boldsymbol{x}} \prod_{i=1}^t \phi(\psi_i(\boldsymbol{x}))$ is asymptotically equal to the integral $\int_{G^\Psi/G^\Psi \cap \Gamma^t} f^{\otimes t}$.

When moving to non-flag systems of forms, not only does g^Ψ not not necessarily equidistribute on the Leibman group, but the notion of irrationality originally considered by Green and Tao does not capture enough information about the polynomial sequence g to determine its distribution. The following example appears in [40] and [66].

Example 4.2.2. Let $\Psi : \mathbb{Z}^2 \to \mathbb{Z}^4$ be our non-flag system of linear forms from above:

$$\Psi(x, y) = (\psi_1(x, y), \psi_2(x, y), \psi_3(x, y), \psi_4(x, y)) = (y, 2x + 2y, x + 3y, x).$$

Then, in $\mathbb{R}^4$,

$$V = \text{span}\{(0,2,1,1),(1,2,3,0)\} = \ker\{(3,0,-1,1),(2,-1,0,2)\},$$

$$V^2 = \text{span}\{(0,4,1,1),(1,4,9,0),(0,4,3,0)\} = \ker\{(24,3,-4,-8)\},$$

and $V^3 = \mathbb{R}^4$.

Let $\alpha, \beta \in \mathbb{R}$ be such that $\{1, \alpha, \beta\}$ satisfy no nontrivial linear relations over $\mathbb{Q}$. Let $f : \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{C}$ be smooth and let $g(n) = (\alpha n, \beta n^2)$ determine a polynomial sequence on $\mathbb{R}^2$. Then (suppressing in our notation the quotient $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$) we have that $\phi(n) := f(\alpha n, \beta n^2)$ is an irrational nilsequence in the filtered Lie group $\mathbb{R}^2$ with filtration $G_1 = \mathbb{R}^2$, $G_2 = 0 \times \mathbb{R}$, essentially because there is no filtered subnilmanifold with respect to which g is a polynomial sequence. We would like our counting lemma to determine the average

$$\mathbb{E}_{x,y}(f \circ g)(\psi_1(x,y))(f \circ g)(\psi_2(x,y))(f \circ g)(\psi_3(x,y))(f \circ g)(\psi_4(x,y)). \tag{4.1}$$

To do so we find where $g^\Psi(x,y) := (g(\psi_1(x,y)), \dots, g(\psi_4(x,y)))$ distributes within in $(\mathbb{R}^2/\mathbb{Z}^2)^4$. Thereafter, the average (4.1) may be computed by the relatively easy task of integrating over a subtorus.

Here the Leibman group is

$$\{(a,b) : a \in V, b \in V + V^2\},$$

suitably interpreted as a subgroup of $(\mathbb{R}^2)^4$. On the other hand, one may compute that in fact g^Ψ is constrained to the following subgroup of G^Ψ:

$$\{(a,b) : a \in V, b \in V^2\}.$$

Since $V^2 \subsetneq V + V^2$ we have that g^Ψ takes values in a strict subgroup of the Leibman group and so will certainly not equidistribute there. (Note on the other hand that if the flag condition was satisfied by the system of linear forms, i.e. $V \leq V^2$ so $V^2 = V + V^2$, then the subgroup is equal to the Leibman group and indeed Green and Tao's counting lemma holds.)

Now define $\tilde{g}(n) = (\alpha n, \gamma n + \beta n^2)$, where $\{1, \alpha, \beta, \gamma\}$ is a $\mathbb{Q}$-linearly independent set. Then $\tilde{g}$ is irrational in the same filtration as g and in this instance $\tilde{g}^\Psi$ is not constrained to the same subgroup of G^Ψ as g^Ψ was. In fact, $\tilde{g}^\Psi$ equidistributes in G^Ψ. The upshot is that in the non-flag setting, the irrationality of a polynomial sequence in $G_\bullet$ is not enough to determine the distribution of g^Ψ.

In light of this example, one might conjecture in general that the subgroup of G^t on which g^Ψ equidistributes may be expressed in terms of the sequence of subgroups on which the distinct degree parts of g equidistribute. In our example with $g(n) = (\alpha, 0)n + (0, \beta)n^2$ and $\tilde{g}(n) = (\alpha, \gamma)n + (0, \beta)n^2$, we have $(\alpha, 0)n$ equidistributing in $\mathbb{R} \times 0$ whereas $(\alpha, \gamma)n$ equidistributes in $\mathbb{R}^2$ and indeed this precisely accounts for how the distribution of g^Ψ differs from that of $\tilde{g}^\Psi$. It turns out that when the nilmanifold in question is abelian, one can indeed compute the distribution of g^Ψ in G^t/Γ^t if the distribution of the coefficients of g are known. However, this is not the case in the nonabelian setting, as the following example of linear sequences on the Heisenberg group demonstrates.

Example 4.2.3. Let G be the Heisenberg group with filtration determined by $G_2 = [G, G]$. Let Ψ, ψ_i, V be as above. Let $v_1 = (1, 2, 3, 0)$ and $v_2 = (0, 2, 1, 1)$ so $V = \text{span}(v_1, v_2)$. Let

$$g_1 = \begin{pmatrix} 1 & \alpha & \gamma \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{pmatrix}, \quad \tilde{g}_1 = \begin{pmatrix} 1 & \alpha & \frac{1}{2}\alpha\beta \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{pmatrix},$$

where $1, \alpha, \beta, \alpha\beta, \gamma$ satisfy no nontrivial linear relations over $\mathbb{Q}$. Then both $g(n) = g_1^n$ and $\tilde{g}(n) = \tilde{g}_1^n$ equidistribute in G/Γ, but it transpires that g^Ψ distributes differently to $\tilde{g}^\Psi$ in G^4/Γ^4. (Furthermore, both polynomial sequences are irrational by Definition A.6 of [35]).

Both g^Ψ and $\tilde{g}^\Psi$ take values in the Leibman group G^Ψ. Define η on $G^\Psi \leq G^4$ by

$$\eta(h_1, h_2, h_3, h_4) = w \cdot (h_{1,2}, h_{2,2}, h_{3,2}, h_{4,2}),$$

where

$$h_i = \begin{pmatrix} 1 & h_{i,0} & h_{i,2} \\ 0 & 1 & h_{i,1} \\ 0 & 0 & 1 \end{pmatrix} \in G$$

for each i and $w = (24, 3, -4, -8) \in V^{2\perp}$. An easy but mundane computation (using crucially that $w \in V^{2\perp}$) reveals that in fact η is a character on G^Ψ. We compute (with slight abuse of notation where we use a single matrix with vector-valued entries to denote a 4-tuple of matrices)

$$g^\Psi(x, y) = g_1^{\Psi(x,y)} = g_1^{xv_1 + yv_2} = \begin{pmatrix} 1 & \alpha(xv_1 + yv_2) & \gamma(xv_1 + yv_2) + \binom{xv_1 + yv_2}{2}\alpha\beta \\ 0 & 1 & \beta(xv_1 + yv_2) \\ 0 & 0 & 1 \end{pmatrix}$$

and so

$$\eta(g^{\Psi}(x,y)) = w\cdot v_1(\gamma - \frac{1}{2}\alpha\beta)x + w\cdot v_2(\gamma - \frac{1}{2}\alpha\beta)y + \\ + w\cdot v_1^2(\frac{1}{2}\alpha\beta)x^2 + 2w\cdot v_1v_2(\frac{1}{2}\alpha\beta)xy + w\cdot v_2^2(\frac{1}{2}\alpha\beta)y^2.$$

Since $w \in V_2^{\perp}$, we have that $\eta(g^{\Psi}(x)) = w\cdot v_1(\gamma - \frac{1}{2}\alpha\beta)x + w\cdot v_2(\gamma - \frac{1}{2}\alpha\beta)y$, where $w\cdot v_1, w\cdot v_2 \neq 0$. Of course, completing the same computation with $\tilde{g}$ yields the same result with $\frac{1}{2}\alpha\beta$ in place of γ and so in particular $\eta(\tilde{g}^{\Psi}(x)) = 0$. That is, $\tilde{g}^{\Psi}$ is constrained to the subgroup of G^{Ψ} given by $\ker\eta$ whereas g^{Ψ} is not. (As an application of our qualitative counting lemma Theorem 4.2.16, we will see in Example 4.2.19 that g^{Ψ} in fact equidistributes on $G^{\Psi}/G^{\Psi}\cap\Gamma^4$ and $\tilde{g}^{\Psi}$ equidistributes on $\ker\eta/\ker\eta\cap\Gamma^4$.)

4.2.2 Moving to the Lie algebra

Let us now interpret the previous example in the Lie algebra $\mathfrak{g}$ of the Heisenberg group G. Recall that $\log : G \to \mathfrak{g}$ is a bijection given by

$$\log\begin{pmatrix}1 & a & c\\ 0 & 1 & b\\ 0 & 0 & 1\end{pmatrix} = \begin{pmatrix}0 & a & c-\frac{1}{2}ab\\ 0 & 0 & b\\ 0 & 0 & 0\end{pmatrix}.$$

For convenience and brevity we will use the coordinates $(a, b, c - \frac{1}{2}ab)$ to denote the latter matrix in the previous sentence. The Lie bracket in these coordinates is

$$[(a,b,c),(x,y,z)] = (0,0,ay-bx). \tag{4.2}$$

The algebra $\mathfrak{g}$ inherits a filtration $\mathfrak{g}_0 = \mathfrak{g}_1 = \mathfrak{g}$ and $\mathfrak{g}_2 = [\mathfrak{g},\mathfrak{g}] = \{(0,0,\mathbb{R})\}$. In these coordinates, we have that $p(n) := \log g(n) = (\alpha,\beta,\gamma - \frac{1}{2}\alpha\beta)n$ and $\tilde{p}(n) := \log\tilde{g}(n) = (\alpha,\beta,0)n$.

Let us digress slightly and consider a general polynomial sequence p on some Lie algebra $\mathfrak{g}$ with linear forms Ψ; we are interested in determining where p^{Ψ} equidistributes within $\mathfrak{g}^t$.

Definition 4.2.4 (Filtration in the Lie algebra)**.** A *filtration* on $\mathfrak{g}$ of *step* at most s a sequence $(\mathfrak{g}_i)_{i=1}^s$ of Lie subalgebras of $\mathfrak{g}$ such that $\mathfrak{g}_0 = \mathfrak{g}_1 \geq \mathfrak{g}_2 \geq \cdots \geq \mathfrak{g}_s$ and such that for all i, j, $[\mathfrak{g}_i, \mathfrak{g}_j] \leq \mathfrak{g}_{i+j}$ where we define $\mathfrak{g}_{s+1} = \mathfrak{g}_{s+2} = \cdots = 0$.

Definition 4.2.5 (Polynomial sequence adapted to a filtration)**.** A polynomial sequence defined by $p(n) := \sum_{i=1}^s a_i n^i$ in $\mathfrak{g}$ is *adapted to* $\mathfrak{g}_\bullet$ if $a_i \in \mathfrak{g}_i$ for all i.

It will prove convenient to deal with $\mathfrak{g} \otimes \mathbb{R}^t$ rather than $\mathfrak{g}^t$, where the map from the former to the latter on elementary tensors is given by $\varphi : a \otimes v \mapsto (v_1 a, v_2 a, \ldots, v_t a)$. This gives a vector space isomorphism and we may obtain a Lie algebra isomorphism by endowing $\mathfrak{g} \otimes \mathbb{R}^t$ with a Lie bracket as follows:

$$\begin{aligned}
[a \otimes v, b \otimes w] &:= \varphi^{-1}([(v_1 a, \ldots, v_t a), (w_1 b, \ldots, w_t b)]) \\
&= \varphi^{-1}(([v_1 a, w_1 b], \ldots, [v_t a, w_t b])) \\
&= \varphi^{-1}((v_1 w_1 [a, b], \ldots, v_t w_t [a, b])) \\
&= [a, b] \otimes vw
\end{aligned} \tag{4.3}$$

where multiplication in the second entry is conducted coordinatewise. With this setup, and henceforth suppressing our φ notation and implicitly identifying $\mathfrak{g}^t$ with $\mathfrak{g} \otimes \mathbb{R}^t$, if $p(n) = \sum a_i n^i$ then $p^\Psi(\boldsymbol{x}) := (p(\psi_1(\boldsymbol{x})), \ldots, p(\psi_t(\boldsymbol{x}))) = \sum a_i \otimes \Psi^i(\boldsymbol{x}) \in \mathfrak{g} \otimes \mathbb{R}^t$.

Returning to our example, $\tilde{p}^\Psi$ is contained in the subspace $(\mathbb{R}, \mathbb{R}, 0) \otimes V$ at all inputs. A straightfoward computation using the Lie bracket defined above shows that $(\mathbb{R}, \mathbb{R}, 0) \otimes V + \mathfrak{g}_2 \otimes V^2$ is the smallest Lie algebra containing the subspace $(\mathbb{R}, \mathbb{R}, 0) \otimes V$ and indeed we will see later as an application of our qualitative counting lemma Theorem 4.2.16 in Example 4.2.19 that $\tilde{p}^\Psi$ equidistributes in this subalgebra (modulo an appropriate lattice). Importantly, $(\mathbb{R}, \mathbb{R}, 0) \otimes V + \mathfrak{g}_2 \otimes V^2$ is a proper subalgebra of $\mathfrak{g} \otimes V + \mathfrak{g}_2 \otimes V^2 = \log G^\Psi$ and indeed p^Ψ equidistributes in $\log G^\Psi$ (again, modulo an appropriate lattice).

4.2.3 Notions of irrationality

In this subsection we will restrict our attention to polynomial sequences p with $p(0) = 0 \in \mathfrak{g}$. In particular, notions of irrationality will only be defined for polynomial sequences satisfying this property.

We begin by introducing some definitions, in particular that of a *rational structure* on $\mathfrak{g}$. A *rational structure* on $\mathfrak{g}$ is a rational Lie subalgebra $\mathfrak{g}_\mathbb{Q}$ of $\mathfrak{g}$ such that the vector space isomorphism $\mathfrak{g} \cong \mathfrak{g}_\mathbb{Q} \otimes \mathbb{R}$ holds. Then, let $\mathfrak{g}^*_\mathbb{Q} \subset \mathfrak{g}^*$ be the linear maps on $\mathfrak{g}$ which map $\mathfrak{g}_\mathbb{Q}$ to $\mathbb{Q}$. In the coordinates in our Heisenberg example above, we have implicitly used the rational structure $\mathbb{Q}^3 \leq \mathbb{R}^3$. In what follows, let $\mathrm{Hom}_\mathbb{Q}(\mathfrak{g}, \mathbb{R})$ denote the Lie algebra homomorphisms from $\mathfrak{g}$ to $\mathbb{R}$ which map $\mathfrak{g}_\mathbb{Q}$ to $\mathbb{Q}$. Finally, many statements in this chapter are made with respect to a rational structure $\mathfrak{g}_\mathbb{Q}$ on $\mathfrak{g}$; at times the underlying rational structure may not be explicitly referenced.

Definition 4.2.6 (Qualitative additive irrationality)**.** A polynomial sequence p in $\mathfrak{g}$ is *additively irrational* in $\mathfrak{g}$ with respect to $\mathfrak{g}_{\mathbb{Q}}$ if for all nontrivial $\eta \in \mathrm{Hom}_{\mathbb{Q}}(\mathfrak{g}, \mathbb{R})$ we have that $(\eta \circ p)(\boldsymbol{x}) \notin \mathbb{Q}$ for some $\boldsymbol{x}$.

In the following we invoke Lie's third theorem to glean the simply-connected Lie group G from the Lie algebra $\mathfrak{g}$.

Definition 4.2.7 (Qualitative equidistribution with respect to a rational structure)**.** A polynomial sequence p in $\mathfrak{g}$ *equidistributes* with respect to a rational structure $\mathfrak{g}_{\mathbb{Q}}$ if $\exp \circ p$ equidistributes in G/Γ (as per Definition 4.8.4), where Γ is *any* lattice in G such that $\mathrm{span}_{\mathbb{Q}} \log \Gamma = \mathfrak{g}_{\mathbb{Q}}$.

It is not at all *a priori* clear that the above is well defined (in particular that such a lattice exists and furthermore that the property of equidistribution is independent of the choice of such a lattice). This is addressed in Appendix 4.8.2.

The following theorem of Leibman is our main tool for determining qualitative equidistribution. It says that the above two definitions are equivalent.

Theorem 4.2.8 (Leibman's criterion for qualitative equidistribution in the Lie algebra)**.** *Let D be a positive integer, let $\mathfrak{g}$ be a real, finite dimensional, nilpotent Lie algebra and let $\mathfrak{g}_{\mathbb{Q}}$ be a rational structure on $\mathfrak{g}$. Let $(p(\boldsymbol{n}))_{\boldsymbol{n} \in \mathbb{N}^D}$ be a polynomial sequence on $\mathfrak{g}$. Then $(p(\boldsymbol{n}))_{\boldsymbol{n} \in \mathbb{Z}^D}$ equidistributes in $\mathfrak{g}$ with respect to $\mathfrak{g}_{\mathbb{Q}}$ if and only if p is additively irrational in $\mathfrak{g}$ with respect to $\mathfrak{g}_{\mathbb{Q}}$.*

We deduce the above statement from Leibman's original theorem ([53]) in Theorem 4.8.8.

For appropriately (i.e. *linearly*) irrational p, it transpires that p^{Ψ} is additively irrational in an explicit subalgebra of $\mathfrak{g}^t$, and so we have determined its distribution by Theorem 4.2.8. Before defining linear irrationality, we will transfer Green and Tao's notion of irrationality [35, Definition A.6] to the Lie algebra. For the purposes of our current discussion, we give the following qualitative translation in which one removes the complexity requirement on characters and takes the limit $N \to \infty$.

Definition 4.2.9 (Qualitative Green-Tao irrationality)**.** An element $g \in G_i$ is i-irrational if for all nontrivial characters $\eta : G_i \to \mathbb{R}$ which vanish on G_{i+1} and $[G_j, G_{i-j}]$ for all j, and which map the lattice Γ to $\mathbb{Q}$, we have $\eta(g) \notin \mathbb{Q}$.

(A direct translation works with characters that map Γ to $\mathbb{Z}$, but the above is equivalent and more natural in the qualitative setting.) By basic Lie theory one may interpret this in the Lie algebra with filtration $\mathfrak{g}_\bullet := (\log G_i)_{i=1}^s$.

Definition 4.2.10. Let $\mathfrak{g}$ have filtration $\mathfrak{g}_\bullet = (\mathfrak{g}_i)_{i=0}^s$. For $i = 1, \dots, s$, define the subalgebra $\mathfrak{h}_i$ to be the smallest Lie algebra containing $\mathfrak{g}_{i+1}$ and $[\mathfrak{g}_j, \mathfrak{g}_{i-j}]$ for all j.

The following definition is what one obtains when one lifts qualitative Green-Tao irrationality to the Lie algebra. We will call it 'filtration irrationality'.

Definition 4.2.11 (Qualitative filtration irrationality)**.** We say $a \in \mathfrak{g}_i$ is *i-filtration irrational* if for all $l \in \mathfrak{g}_{\mathbb{Q}}^*$ which are nontrivial on $\mathfrak{g}_i$ and vanish on $\mathfrak{h}_i$, we have $l(a) \notin \mathbb{Q}$. A polynomial sequence $p(n) = \sum_{i=1}^s a_i n^i$ is *filtration irrational* in $\mathfrak{g}_\bullet$ if a_i is i-filtration irrational for $i = 1, \dots, s$.

It is standard ([18, Theorem 5.1.8]) that for any lattice $\Gamma \leq G$, $\mathrm{span}_{\mathbb{Q}} \log \Gamma$ is a rational structure in $\mathfrak{g}$. With the rational structure $\mathfrak{g}_{\mathbb{Q}}$ on $\mathfrak{g}$ chosen in this way, one may check that every l as in Definition 4.2.11 induces a character η on G_i with the properties in Definition 4.2.9 (and conversely every η induces an $l \in \mathfrak{g}_{\mathbb{Q}}^*$ which vanishes on $\mathfrak{h}_i$).

In the examples in the previous subsection, we were somewhat vague about the notion irrationality. One is now in a position to verify that all polynomial sequences that feature in Examples 4.2.2 and 4.2.3 are filtration irrational and that ultimately filtration irrationality is insufficient to determine the distribution of p^Ψ in the non-flag setting.

Definition 4.2.12 (Rational subspace)**.** A subspace S of $\mathfrak{g}$ is *rational* with respect to a rational structure $\mathfrak{g}_{\mathbb{Q}}$ if it may be written as the kernel of a set of elements of $\mathfrak{g}_{\mathbb{Q}}^*$.

Definition 4.2.13 (Qualitative linear irrationality)**.** Let S be a rational subspace of $\mathfrak{g}$. An element $a \in S$ is *linearly irrational* in S if for all $l \in \mathfrak{g}_{\mathbb{Q}}^*$ which are nontrivial on S, we have $l(a) \notin \mathbb{Q}$. A polynomial sequence $p(n) := \sum_{i=1}^s a_i n^i$ is *linearly irrational* in a sequence of subspaces $S_\bullet = (S_i)_{i=1}^s$ if a_i is linearly irrational in S_i for $i = 1, \dots, s$.

The notion of linear irrationality does not refer to the Lie algebra structure on $\mathfrak{g}$ and so may be defined analogously for vector spaces with fixed rational structures. We note that linear irrationality is robust under rational (vector space) automorphisms.

In our Heisenberg example above, recalling that α, β, γ and 1 are linearly independent over $\mathbb{Q}$, we have that $\tilde{p}(n) = (\alpha, \beta, 0)n$ is linearly irrational in $S_\bullet = (S_1 := \mathbb{R} \times \mathbb{R} \times 0, S_2 := 0)$, but is not linearly irrational in $(\mathbb{R}^3, 0)$, or in $(\mathbb{R} \times \mathbb{R} \times 0, 0 \times 0 \times \mathbb{R})$. Define

$$p'(n) = (\alpha, \beta, 2\alpha + \beta)n + (0, 0, \gamma)n^2,$$
$$p''(n) = (\alpha, \beta, 2\alpha + \beta + 1/3)n + (0, 0, \gamma)n^2.$$

Then p' is linearly irrational in $(\ker(2,1,-1), 0 \times 0 \times \mathbb{R})$. On the other hand, p'' is not linearly irrational in any sequence of subspaces of $\mathfrak{g}$ because there is $l \in \mathfrak{g}_{\mathbb{Q}}^*$ with $0 \neq l(a_1) \in \mathbb{Q}$.

In general, any polynomial sequence which has a coefficient a_i with $0 \neq l(a_i) \in \mathbb{Q}$ for some $l \in \mathfrak{g}_{\mathbb{Q}}^*$ is not linearly irrational in any sequence of subspaces. If, on the other hand, all coefficients $\{a_i\}_i$ of p satisfy the property that $l(a_i) \in \mathbb{Q}$ implies $l(a_i) = 0$, then p is linearly irrational in the sequence of rational subspaces $S_\bullet = (S_i)_{i=1}^s$ defined by $S_i = (a_i^\perp \cap \mathfrak{g}_{\mathbb{Q}}^*)^\perp$.

Linear irrationality is a stronger notion than filtration irrationality in the sense that if p is linearly irrational with respect to some $S_\bullet$ then there exists some $\mathfrak{g}'_\bullet$ (potentially on a subalgebra $\mathfrak{g}' \leq \mathfrak{g}$) with respect to which p is filtration irrational (see Lemma 4.3.3).

The converse is not true, as is demonstrated by the example of p'' above which is filtration irrational with respect to the lower central series filtration. Finally, also note that if p is adapted to a rational filtration $\mathfrak{g}_\bullet$ and is linearly irrational with respect to $S_\bullet$ then we automatically have that $S_i \leq \mathfrak{g}_i$ for all i and so $[S_i, S_j] \subset \mathfrak{g}_{i+j}$ for all i, j.

Unfortunately the notions of linear and filtration irrationality are not quite as simple in the quantitative setting. For example, a polynomial sequence may be quantitatively filtration/linearly irrational in two distinct sequences of subspaces – see Example 4.4.2. The relationship between the two notions of irrationality is also not so clear cut in the quantitative setting (though they are related – see Lemma 4.7.1). For this reason we will explicitly keep track of the sequence of subspaces with respect to which we are asserting quantitative linear irrationality and independently the filtered Lie algebra with respect to which we are asserting quantitative filtration irrationality.

4.2.4 The counting lemma

We can now prove a qualitative counting lemma for polynomial sequences on linear patterns. Again we will restrict our attention to polynomial sequences with $p(0) = 0 \in \mathfrak{g}$; in the general case one simply obtains equidistribution on a coset of the algebra studied in this section. Taking for granted our algebraic preliminaries, Theorem 4.2.16 is rather straightfoward.

Definition 4.2.14. Given a system of t linear forms Ψ and a sequence of rational subspaces $S_\bullet = (S_i)_{i=1}^s$ of $\mathfrak{g}_\bullet$, define the Lie algebra $g^\Psi(S_\bullet)$ as the smallest Lie subalgebra of $\mathfrak{g}^t$ which contains $S_i \otimes V^i$ for all i (where recall $V = V_\Psi$ is defined in Subsection 4.1.3).

The qualitative counting lemma may be stated informally as follows: if a polynomial sequence p is linearly irrational in $S_\bullet$ then p^Ψ equidistributes on $\mathfrak{g}^\Psi(S_\bullet)$. We will offer a more explicit description of $g^\Psi(S_\bullet)$ in Section 4.3.

Lemma 4.2.15. *Let $p : \mathbb{R}^D \to \mathbb{R}$ be a polynomial such that $p(\mathbb{Z}^D) \subset \mathbb{Q}$. Then the coefficients of p (with respect to the monomial basis) lie in $\mathbb{Q}$.*

Proof. For univariate polynomials ($D = 1$), the coefficients may be recovered by evaluating the polynomial at sufficiently distinct points and basic linear algebra over $\mathbb{Q}$. The multivariate case follows from an easy induction on the number of variables. □

We are ready to prove the qualitative counting lemma.

Theorem 4.2.16 (Qualitative counting lemma)**.** *Suppose that p is a linearly irrational polynomial sequence in $S_\bullet$. Then the polynomial sequence $p^\Psi(\boldsymbol{x})$ equidistributes on $\mathfrak{g}^\Psi(S_\bullet)$ with respect to the rational structure $\mathfrak{g}^\Psi(S_\bullet) \cap \mathfrak{g}_\mathbb{Q}^t$.*

Proof. We use Theorem 4.2.8 and so will show that p^Ψ is additively irrational in $\mathfrak{g}^\Psi(S_\bullet)$. Let $p(n) := \sum_{i=1}^s a_i n^i$. Let $\eta \in \mathrm{Hom}_\mathbb{Q}(\mathfrak{g}^\Psi(S_\bullet), \mathbb{R})$.[1] In $\mathfrak{g}^\Psi(S_\bullet) \leq \mathfrak{g} \otimes \mathbb{R}^t$, we have that

$$\eta(g^\Psi(\boldsymbol{x})) = \sum_{i=1}^s \eta(a_i \otimes \Psi(\boldsymbol{x})^i) = \sum_{m \in \mathcal{M}} \sum_{i=1}^s \eta(a_i \otimes v_{m,i}) m(\boldsymbol{x}), \tag{4.4}$$

where the $\{v_{m,i}\}_{m\in\mathcal{M}}$ span V^i as in Subsection 4.1.3. By Lemma 4.2.15, if $\eta(g^\Psi(\boldsymbol{x})) \in \mathbb{Q}$ for all $\boldsymbol{x} \in \mathbb{Z}^D$ then each $\eta(a_i \otimes v_{m,i}) \in \mathbb{Q}$. But each $\eta(\cdot \otimes v_{m,i})$ lies in $\mathfrak{g}_\mathbb{Q}^*$ and so by the linear irrationality of a_i in S_i we have that in fact $\eta(a_i \otimes v_{m,i}) = 0$. The kernel of a rational homomorphism is a rational subspace and so $\eta(S_i \otimes v_{m,i}) = 0$, since S_i is necessarily the smallest rational subspace containing a_i. Next, the $\{v_{m,i}\}_{m\in\mathcal{M}}$ span V^i, so we have $\eta(S_i \otimes V^i) = 0$ for all i. Furthermore, since η is a Lie algebra homomorphism to an abelian Lie algebra, η vanishes on the smallest Lie algebra containing the subspaces $S_i \otimes V^i$ and so η is trivial on $\mathfrak{g}^\Psi(S_\bullet)$. This completes the proof. □

[1] Recall that we use the notation $\mathrm{Hom}_\mathbb{Q}$ to refer to homomorphisms of Lie algebras which map $\mathfrak{g}_\mathbb{Q}$ to $\mathbb{Q}$.

Remark 4.2.17. In the flag setting, if η is a character on G^t (or a subgroup thereof), the 'pullback' $\eta(\cdot^{v_{m,i}})$ is an i-horizontal character on G_i ([35, Definition A.5]). However, in the non-flag setting, $\eta(\cdot^{v_{m,i}})$ may fail to possess obviously discernible structure as a map on G_i (cf. Example 4.2.3 and the character η from there). The key technical point in the above is that characters on $\mathfrak{g}^t$ (or subalgebras thereof) pull back to linear maps on $\mathfrak{g}$. That is, $\eta(\cdot \otimes v_{m,i})$ is a rational linear map on $\mathfrak{g}$ whence linear irrationality is sufficient to determine the distribution of p^Ψ.

Note that Theorem 4.2.16 yields equidistribution statements for polynomial sequences which are not filtration irrational.

Example 4.2.18. In the Heisenberg group with the usual coordinates, let $p(n) = (\alpha, \alpha, 0)n$. Then p is linearly irrational in $S_\bullet := (S_1 := \{(x, x, 0) : x \in \mathbb{R}\}, S_2 := \{0\})$. Furthermore, we have from Equation (4.2) that $[S_1, S_1] = \{0\}$ (as is expected since this example is really coming from an abelian group). Theorem 4.2.16 says that p^Ψ equidistributes on $S_1 \otimes V$.

We are also able to prove the equidistribution claims we made in Examples 4.2.2 and 4.2.3, and in Subsection 4.2.2.

Example 4.2.19 (Equidistribution for Examples 4.2.2 and 4.2.3)**.** In the abelian setting we may identify the Lie group with the Lie algebra. Let $p(n) = (\alpha, 0)n + (0, \beta)n^2$ on $\mathbb{R}^2$ be the first polynomial sequence from Example 4.2.2. Theorem 4.2.16 yields that p^Ψ equidistributes on $(\mathbb{R} \times 0) \otimes V + (0 \times \mathbb{R}) \otimes V^2$. On the other hand, the second polynomial sequence from Example 4.2.2 $\tilde{p}(n) = (\alpha, \gamma)n + (0, \beta)n^2$ equidistributes on $\mathbb{R}^2 \otimes V + (0 \times \mathbb{R}) \otimes V^2$.

Moving to Example 4.2.3, we saw in Subsection 4.2.2 that the relevant polynomial sequences may be transferred to the Lie algebra as $p(n) = (\alpha, \beta, \gamma - \frac{1}{2}\alpha\beta)n$ and $\tilde{p}(n) = (\alpha, \beta, 0)n$. Theorem 4.2.16 says that p^Ψ equidistributes in the smallest Lie algebra containing $\mathbb{R}^3 \otimes V$, which is $\mathbb{R}^3 \otimes V + (0 \times 0 \times \mathbb{R}) \otimes V^2 = \mathfrak{g} \otimes V + \mathfrak{g}_2 \otimes V^2$. We also obtain that $\tilde{p}^\Psi$ equidistributes in $(\mathbb{R} \times \mathbb{R} \times 0) \otimes V + \mathfrak{g}_2 \otimes V^2$. One may furthermore check that this subalgebra maps to the subgroup $\ker \eta$ from Example 4.2.3 under the exponential map.

In each of the examples above, computing explicitly the Lie algebra $\mathfrak{g}^\Psi(S_\bullet)$ is not a difficult task. Of course, this may not be the case in general. We will address this in Section 4.3.

4.2.5 Factorisation

We have demonstrated how to obtain an equidistribution result for polynomial sequences which satisfy a special property: linear irrationality. Ultimately, one wants to understand the equidistribution of arbitrary polynomial sequences (we will again focus on p with $p(0) = 0$). The purpose of this section is to show (Proposition 4.2.28) that one may write an arbitrary p as a product $p = p' * r$ where p' is linearly irrational and r is rational (and as before, $*$ is group multiplication conducted in the Lie algebra). Then, roughly speaking, one may partition $\mathbb{Z}^D$ into appropriately-spaced subprogressions where we understand the equidistribution of p on each one.

Part of the power of Green and Tao's counting lemma is that it requires a surprisingly small amount of data from a polynomial sequence g to determine the distribution of g^Ψ in the flag case. In the non-flag case, we require more data to determine distribution and so the task of reducing a general nilsequence to something that our counting lemma can handle (i.e. factorising for linear irrationality) is inevitably a little more involved. This subsection may be viewed as a warm up for the quantitative factorisation theory in Section 4.5.

Linear irrationality is a linear algebraic notion and indeed one may, by natural linear algebraic arguments, write an arbitrary $a \in \mathfrak{g}$ as the sum of a linearly irrational element and a rational element.

Lemma 4.2.20. *Let $a \in \mathfrak{g}$. Then there exists a_p which is linearly irrational in some rational subspace $S \leq \mathfrak{g}$ and $a_r \in \mathfrak{g}_\mathbb{Q}$ such that $a = a_p + a_r$.*

One may apply this to each coefficient of an arbitrary polynomial sequence p to obtain an additive decomposition $p = p' + r$, where p' is linearly irrational and r is rational. This additive decomposition does not, however, immediately determine a factorisation with respect to group multiplication, which is ultimately what is required.

To later obtain our multiplicative factorisation, we will need a stronger additive decomposition, Lemma 4.2.22, of which Lemma 4.2.20 is a special case. First we will need the following definition of linear irrationality relative to a subspace.

Definition 4.2.21 (Relative linear irrationality)**.** Let U, T be rational subspaces of $\mathfrak{g}$. An element $a \in U$ is *linearly irrational in U* mod T if for all $l \in \mathfrak{g}^*_\mathbb{Q}$ which are nontrivial on U and which vanish on $U \cap T$ we have $l(a) \notin \mathbb{Q}$.

In what follows we use $T_{\mathbb{Q}}$ to denote the space $T \cap \mathfrak{g}_{\mathbb{Q}}$, and similarly for other subspaces. If T is rational with respect to $\mathfrak{g}_{\mathbb{Q}}$, then $T_{\mathbb{Q}}$ is a rational vector space of full dimension (and indeed topologically dense) inside T.

Lemma 4.2.22 (Qualitative additive decomposition for linear irrationality)**.** *Let U, T be rational subspaces of $\mathfrak{g}$ and let $a \in U$ be linearly irrational in $U \bmod T$. Then there exist $a_p \in U$ and $a_r \in (U \cap T)_{\mathbb{Q}}$ such that a_p is linearly irrational in some rational subspace $U' \leq U$ and such that $a = a_p + a_r$.*

Proof. Let $L \leq \mathfrak{g}_{\mathbb{Q}}^*$ be the subspace of elements which are nontrivial on U and such that $l(a) \in \mathbb{Q}$. Choose a basis $\mathcal{B}$ for L. The elements of $\mathcal{B}$ remain linearly independent when restricted to $U \cap T$ (if not, choose a nontrivial linear combination of them which is the trivial map on $U \cap T$ to obtain a contradiction with the linear irrationality of a in $U \bmod T$). Furthermore, since U, T are rational subspaces, so too is $U \cap T$ and so $\mathcal{B}$ is linearly independent when restricted to a set of maps on the $\mathbb{Q}$-vector space $(U \cap T)_{\mathbb{Q}}$. Thus there exists $a_r \in (U \cap T)_{\mathbb{Q}}$ such that $l(a) = l(a_r)$ for all $l \in \mathcal{B}$, and therefore for all $l \in L$. Note that since $a_r \in U_{\mathbb{Q}}$, the subspace of elements which are nontrivial on U and such that $l(a - a_r) \in \mathbb{Q}$ is also equal to L. Ultimately $l(a - a_r) = 0$ whenever $l(a - a_r) \in \mathbb{Q}$ for all $l \in \mathfrak{g}_{\mathbb{Q}}^*$ (this is trivially true for maps which are trivial on U). Thus $a - a_r$ is linearly irrational in $\ker\{l \in \mathfrak{g}_{\mathbb{Q}}^* : l(a - a_r) \in \mathbb{Q}\} \leq U$. □

Note that the above lemma does not depend at all on the Lie bracket structure and so holds in vector spaces.

Definition 4.2.23. Let $\mathfrak{g}_\bullet$ be a filtration of step s on $\mathfrak{g}$ and let $U_\bullet, T_\bullet \leq \mathfrak{g}_\bullet$ be two sequences of rational subspaces with $U_i, T_i \leq \mathfrak{g}_i$ for all i. A polynomial sequence $p := \sum_{i=1}^s a_i n^i$ in $\mathfrak{g}$ is *linearly irrational in $U_\bullet \bmod T_\bullet$* if a_i is linearly irrational in $U_i \bmod T_i$ for all i.

Definition 4.2.24. A polynomial sequence $p(n) := \sum_{1=1}^s a_i n^s$ adapted to $\mathfrak{g}_\bullet$ is *rational* if $a_i \in (\mathfrak{g}_i)_{\mathbb{Q}}$ for all i.

In what follows, let $(p)_i$ denote the ith coefficient of a polynomial sequence p. Applying Lemma 4.2.22 to each coefficient of a polynomial sequence we obtain the following.

Corollary 4.2.25. *Let $U_\bullet, T_\bullet$ be sequences of rational subspaces of $\mathfrak{g}_\bullet$ and let p be a polynomial sequence which is linearly irrational in $U_\bullet \bmod T_\bullet$. Then there are polynomials p' and r adapted to $U_\bullet$ and $(U_\bullet \cap T_\bullet)_{\mathbb{Q}} := ((U_i \cap T_i)_{\mathbb{Q}})_{i=1}^s$ respectively such that p' is linearly irrational in some sequence of rational subspaces $U'_\bullet \leq U_\bullet$ and $p = p' + r$.*

Now we pursue a factorisation with respect to the $*$ operation. We will need the fact that the set of polynomial maps adapted to $\mathfrak{g}_\bullet$ form a group. This result (in the Lie group, and in greater generality) is originally due to Leibman [52]. See also [38, Proposition 6.2] for a related statement and different proof. The result that we need in the Lie algebra is more obvious and follows from Baker-Campbell-Hausdorff.

Lemma 4.2.26. *The set of polynomials adapted to $\mathfrak{g}_\bullet$ forms a group under the $*$ operation.*

Lemma 4.2.27. *The product of two rational polynomial sequences is rational.*

Proof. This follows immediately from the Baker-Campbell-Hausdorff formula $x * y = x + y + \frac{1}{2}[x, y] + \dots$ and the fact the rational subspaces $(\mathfrak{g}_i)_{\mathbb{Q}}$ are rational subalgebras which obey the same filtration as $\mathfrak{g}_\bullet$. □

Proposition 4.2.28 (Qualitative multiplicative factorisation for linear irrationality)**.** *Let p be a polynomial sequence in $\mathfrak{g}$ adapted to $\mathfrak{g}_\bullet$. We may write $p = p' * r$ where p' is linearly irrational with respect to some sequence of rational subspaces $S'_\bullet \leq \mathfrak{g}_\bullet$ and r is a rational polynomial sequence, both adapted to $\mathfrak{g}_\bullet$.*

Proof. Invoke Corollary 4.2.25 with $U_\bullet = \mathfrak{g}_\bullet$ and $T_\bullet = (\mathfrak{g}, \dots, \mathfrak{g})$ to obtain p_1 which is linearly irrational in some sequence $S^{(1)}_\bullet$ and r_1 adapted to $(S_\bullet)_{\mathbb{Q}}$ such that $p = p_1 + r_1$. By Baker-Campbell-Hausdorff and the fact that $\mathfrak{g}/\mathfrak{g}_2$ is abelian we have that

$$p = p_1 * r_1 \mod \mathfrak{g}_2.$$

Set $T^{(2)}_\bullet := (\mathfrak{g}_2, \mathfrak{g}_2, \dots, \mathfrak{g}_2)$. Then $p * r_1^{-1}$ is a polynomial sequence adapted to $\mathfrak{g}_\bullet$ by Lemma 4.2.26, and is linearly irrational in $(S^{(1)}_\bullet + T^{(2)}_\bullet) \bmod T^{(2)}_\bullet$.[2] Next apply Corollary 4.2.25 with $U_\bullet = S^{(1)}_\bullet + T^{(2)}_\bullet$ and $T_\bullet = T^{(2)}_\bullet$ to $p * r_1^{-1}$ to find p_2 which is linearly irrational in some $S^{(2)}_\bullet$ and r_2 which is adapted to $(S^{(1)}_\bullet \cap T^{(2)}_\bullet)_{\mathbb{Q}}$ such that $p * r_1^{-1} = p_2 + r_2$. By Baker-Campbell-Hausdorff and the fact that $[\mathfrak{g}, \mathfrak{g}_2] \subset \mathfrak{g}_3$ we then have $p * r_1^{-1} = p_2 * r_2 \bmod \mathfrak{g}_3$ and so

$$p = p_2 * r'_2 \mod \mathfrak{g}_3,$$

where $r'_2 := r_2 * r_1$ is rational by Lemma 4.2.27. Continue in this way to eventually find $p' := p_s$, $S'_\bullet := S^{(s)}_\bullet$ and $r := r'_s$. □

[2] Here we recall that all elements of the filtration $\mathfrak{g}_\bullet$ are rational in $\mathfrak{g}$, so in particular $T_\bullet$ is a sequence of rational subspaces.

4.3 Description of $\mathfrak{g}^\Psi(S_\bullet)$, comparison with $\log G^\Psi$

Our counting lemma Theorem 4.1.2 says that if p is a polynomial sequence which is quantitatively linearly irrational in $S_\bullet$, then p^Ψ quantitatively equidistributes in a Lie subalgebra $\mathfrak{g}^\Psi(S_\bullet)$ of $\mathfrak{g}^t$. We have proven a qualitative version of this result Theorem 4.2.16 which gives qualitative equidistribution on the same Lie algebra $\mathfrak{g}^\Psi(S_\bullet)$. So far, however, we have only described $\mathfrak{g}^\Psi(S_\bullet)$ as the smallest Lie subalgebra of $\mathfrak{g}^t$ which contains $S_i \otimes V_\Psi^i$ for all i. The goal of this section is to give descriptions of $\mathfrak{g}^\Psi(S_\bullet)$ which are more explicit and amenable to applications.

Let $S_\bullet := (S_i)_{i=1}^s$ be a sequence of rational subspaces of $\mathfrak{g}$ with respect to a fixed rational structure $\mathfrak{g}_\mathbb{Q}$. Let $W_1 = S_1$ and define $W_i = \text{span}\{S_j, [W_j, W_{i-j}] \text{ for } j = 1, \dots, i-1\}$ for $i = 2, \dots, s$. That is, W_i comprises all degree i commutators in the S_j, where the degree of a commutator is the sum of the subscripts it contains. Note that if $\mathfrak{g}_\bullet$ is a filtration such that $S_i \leq \mathfrak{g}_i$ for all i then $W_i \leq \mathfrak{g}_i$ for all i.

Proposition 4.3.1 (General description of $\mathfrak{g}^\Psi(S_\bullet)$)**.** *Let $S_\bullet$ be a sequence of rational subspaces of $\mathfrak{g}_\bullet$. Let Ψ be a system of linear forms. Then $\mathfrak{g}^\Psi(S_\bullet) = \sum_{i=1}^s W_i \otimes V_\Psi^i$.*

Proof. Recall from Equation (4.3) the Lie bracket operation on $\mathfrak{g} \otimes \mathbb{R}^t$. It follows immediately that for all i, j we have $[W_i \otimes V^i, W_j \otimes V^j] \subset W_{i+j} \otimes V^{i+j}$ (where $W_{s+1} = W_{s+2} = \cdots = 0$) and so $\sum_{i=1}^s W_i \otimes V^i$ is indeed a Lie subalgebra of $\mathfrak{g}^t$. It is also immediate from definitions that $\sum_{i=1}^s W_i \otimes V^i$ is generated by $\{S_i \otimes V^i\}_{i=1}^s$ under commutators and linear spans. □

Recall that we use G^Ψ to denote the Leibman group (Definition 4.2.1); we will call $\log G^\Psi = \sum_{i=1}^s \mathfrak{g}_i \otimes V_\Psi^i$ the *Leibman algebra*. We now pursue Proposition 4.3.6, a description of $\log G^\Psi$ which is more readily comparable to that of $\mathfrak{g}^\Psi(S_\bullet)$ in the above proposition. In so doing we will make the assumption that $S_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i (recall Definition 4.2.10). The goal of the upcoming lemmas is to motivate why this is reasonable.

For a sequence of subspaces $S_\bullet \leq \mathfrak{g}_\bullet$, we will say that a filtration $\mathfrak{g}'_\bullet \leq \mathfrak{g}_\bullet$ is *minimal* for $S_\bullet$ if $S_\bullet \leq \mathfrak{g}'_\bullet$ and there is no proper subfiltration $\mathfrak{g}''_\bullet < \mathfrak{g}'_\bullet$ with $S_i \leq \mathfrak{g}''_i$ for all i.

Lemma 4.3.2. *For a sequence of subspaces $S_\bullet \leq \mathfrak{g}_\bullet$, if $\mathfrak{g}_\bullet$ is minimal for $S_\bullet$ then $S_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i.*[3]

[3]The converse is also true though we do not need such a statement.

Proof. If not, for the smallest i such that $S_i + \mathfrak{h}_i \neq \mathfrak{g}_i$, set $\mathfrak{g}'_i = S_i + \mathfrak{h}_i$, and otherwise set $\mathfrak{g}'_j = \mathfrak{g}_j$ for all $j \neq i$. This yields a proper subfiltration $\mathfrak{g}'_\bullet < \mathfrak{g}_\bullet$. □

Lemma 4.3.3. *Let p be linearly irrational with respect to a sequence of subspaces $S_\bullet \leq \mathfrak{g}_\bullet$. Then p is filtration irrational with respect to the filtration $\mathfrak{g}'_\bullet \leq \mathfrak{g}_\bullet$ which is minimal for $S_\bullet$.*

Proof. This follows easily from definitions. We omit the details. □

Lemma 4.3.4. *Let $p(n) := \sum_{i=1}^{s} a_i n^i$ be filtration irrational in $\mathfrak{g}_\bullet$. Then for any sequence of rational subspaces $T_\bullet$ with $a_i \in T_i \leq \mathfrak{g}_i$, we have $T_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i.*

In particular, if p is linearly irrational in $S_\bullet$ then $S_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i.

Proof. That $T_i + \mathfrak{h}_i \leq \mathfrak{g}_i$ is immediate. For all $l \in \mathfrak{g}^*_{i\mathbb{Q}}$ which vanish on $\mathfrak{h}_i$ we have $l(a_i) \notin \mathbb{Q}$. In particular, we have $\mathfrak{g}_i = (a_i^\perp \cap \mathfrak{g}^*_{i\mathbb{Q}} \cap \mathfrak{h}_i^\perp)^\perp = (a_i^\perp \cap \mathfrak{g}^*_{i\mathbb{Q}})^\perp + \mathfrak{h}_i$ where orthogonal complements are taken in $\mathfrak{g}_i$ and $\mathfrak{g}_i^*$. But $(a_i^\perp \cap \mathfrak{g}^*_{i\mathbb{Q}})^\perp$ is the smallest rational subspace of $\mathfrak{g}_i$ containing a_i. Thus $T_i + \mathfrak{h}_i \geq \mathfrak{g}_i$. □

The following is the main ingredient in the proof of Proposition 4.3.6.

Lemma 4.3.5. *Suppose that the sequence of subspaces $(S_i)_{i=1}^s \leq \mathfrak{g}_\bullet$ satsifes $S_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i. Then the sequence of subspaces $(W_i)_{i=1}^s$ satisfies*

$$W_i + \mathfrak{g}_{i+1} = \mathfrak{g}_i,$$

for all i.

Proof. The proof proceeds by induction and the base case is easy from definitions. Expanding the definition of W_i, we will show that

$$S_i + \sum_{j=1}^{i-1} [W_j, W_{i-j}] + \mathfrak{g}_{i+1} = \mathfrak{g}_i. \tag{4.5}$$

Now for $j = 1, \ldots, i-1$ we have $W_j \leq \mathfrak{g}_j$ and so by the filtration $\mathfrak{g}_\bullet$ we get

$$[W_j + \mathfrak{g}_{j+1}, W_{i-j} + \mathfrak{g}_{i+1-j}] + \mathfrak{g}_{i+1} = [W_j, W_{i-j}] + \mathfrak{g}_{i+1}.$$

By induction, therefore, Equation (4.5) is equivalent to

$$S_i + \sum_{j=1}^{i-1} [\mathfrak{g}_j, \mathfrak{g}_{i-j}] + \mathfrak{g}_{i+1} = \mathfrak{g}_i.$$

But the left hand side of the above is just $S_i + \mathfrak{h}_i$. □

Proposition 4.3.6 (Comparison of Leibman algebra and $\mathfrak{g}^{\Psi}(S_\bullet)$)**.** *Suppose $S_\bullet$ is a sequence of rational subspaces of $\mathfrak{g}_\bullet$ such that $S_i + \mathfrak{h}_i = \mathfrak{g}_i$. Then*

$$\sum_{i=1}^{s} \mathfrak{g}_i \otimes V_\Psi^i = \sum_{i=1}^{s} W_i \otimes \left(\sum_{j=1}^{i} V_\Psi^j \right) = \mathfrak{g}^{\Psi}(S_\bullet) + \sum_{i=1}^{s} W_i \otimes \left(\sum_{j=1}^{i-1} V_\Psi^j \right).$$

Proof. From Lemma 4.3.5 we have $W_i + \mathfrak{g}_{i+1} = \mathfrak{g}_i$ and so $\mathfrak{g}_i = \sum_{j=i}^{s} W_j$. The result follows by switching the order of summation. □

Corollary 4.3.7. *If Ψ satisfies the flag property and $\mathfrak{g}_\bullet$ is minimal for $S_\bullet$ then $\mathfrak{g}^{\Psi}(S_\bullet)$ is equal to the Leibman algebra.*

Proof. If Ψ satisfies the flag property then for all i we have $\sum_{j=1}^{i} V_\Psi^j = V_\Psi^i$; invoke Proposition 4.3.1 and 4.3.6. □

4.4 The counting lemma

Our first goal in the quantitative setting will be to obtain a counting lemma for polynomial sequences which are (quantitatively) linearly irrational. Throughout this section and the next, we will assume that the dimension d of $\mathfrak{g}$, the step s of $\mathfrak{g}$, the number of linear forms t and the number of variables D in these linear forms are all of bounded size; in particular we will always allow constants to depend on these parameters without necessarily indicating this in our notation.

We will need to fix a basis for $\mathfrak{g}$ in order to quantify various notions. (Indeed, in the absence of a choice of basis for $\mathfrak{g}$ we do not even have a canonical metric or (Lebesgue) measure on $\mathfrak{g}$.)

Definition 4.4.1 (Rational basis)**.** A basis $\mathcal{X}$ for $\mathfrak{g}$ is *rational* if its structure constants lie in $\mathbb{Q}$.

The $\mathbb{Q}$-span of a rational basis is a rational structure on $\mathfrak{g}$. In what follows, we fix a rational basis $\mathcal{X}$ for $\mathfrak{g}$ and the corresponding rational structure $\mathfrak{g}_{\mathbb{Q}} := \mathrm{span}_{\mathbb{Q}}(\mathcal{X})$. If statements whose meaning implicitly depends upon a basis are made without reference to one, it may be assumed that $\mathcal{X}$ is the chosen basis. Furthermore, we will denote by $\langle \mathcal{X} \rangle$ the abelian group generated by $\mathcal{X}$, that is $\mathrm{span}_{\mathbb{Z}} \mathcal{X}$. We will also need to conduct quantitative arithmetic on subspaces of $\mathfrak{g}$ and so for essentially the same reasons, we need to identify $\mathbb{Z}$-bases for sublattices obtained by intersecting $\langle \mathcal{X} \rangle$ with a rational subspace S. This may be done canonically by invoking the Hermite normal form. For details and definitions see Appendix 4.9, but for the purposes of a more casual reading

of this section, one may interpret *Hermite basis* to mean 'a canonically-chosen basis $\mathcal{X}' = \mathcal{X}'_{S,\mathcal{X}}$ for S such that $\mathcal{X}'$ is also a $\mathbb{Z}$-basis for the lattice $S \cap \langle \mathcal{X} \rangle$.' We caution that the action of taking a Hermite basis of a subspace is not transitive. In the case that $S = \mathfrak{g}$, we have $\mathcal{X}' = \mathcal{X}$.

In the qualitative setting in Section 4.2, if a polynomial sequence was linearly irrational with respect to any sequence of rational subspaces, then it was linearly irrational with respect to the sequence of minimal rational subspaces of $\mathfrak{g}$ which contained (respectively) the sequence of polynomial coefficients. We need to be a little more careful in the quantitative setting. Even in the single variable case (that is, for the trivial sequence of linear forms $\Psi := (\psi_1(n))$ where $\psi_1(n) = n$) and for A arbitrarily large, (A, N)-irrationality (in the original Green-Tao sense [35, Definition A.6]) does not preclude a polynomial sequence from equidistributing on more than one nilmanifold of 'complexity' $A^{O(1)}$.

Example 4.4.2. Let A be a large integer and let $N > 100A^3$. Define the polynomial sequence $p(n) := (\frac{1}{10A}, \frac{1}{10A-1})n$ in the abelian Lie algebra $\mathfrak{g} := \mathbb{R}^2$ with basis $\mathcal{X} := \{(1,0),(0,1)\}$ and lattice $\langle \mathcal{X} \rangle = \mathbb{Z}^2$.

Firstly, let us first verify that in the limit $A \to \infty$, the quantitative distribution obtained from the Green-Tao counting lemma agrees with our qualitative counting lemma Theorem 4.2.16. Here, $a_1 = (\frac{1}{10A}, \frac{1}{10A-1})$ and the smallest rational subspace containing a_1 is $S_1 := \{(x,y) \in \mathbb{R}^2 : (10A-1)x = 10Ay\}$. Furthermore, S_1 is a subalgebra of $\mathbb{R}^2$ and has 'complexity' $A^{O(1)}$ because it is defined by linear relations of height $A^{O(1)}$.[4] We claim that p is $o_{A\to\infty}(1)$ equidistributed in S_1. Clearly S_1 is isomorphic to $\mathbb{R}$, and $\varphi : (x,y) \mapsto x/(10A)$ is the isomorphism which maps $\mathcal{X}' = \{(10A, 10A-1)\}$ to $1 \in \mathbb{R}$. Thus we may analyse $p'(n) = \frac{n}{100A^2}$ in $\mathbb{R}/\mathbb{Z}$.[5] Set $a'_1 := \varphi(a_1) = \frac{1}{100A^2}$. For any nonzero $m \in \mathbb{Z}$ with $|m| \leq A$ we have $||ma'_1||_{\mathbb{R}/\mathbb{Z}} \geq \frac{1}{100A^2} > A/N$. Thus, by [35, Definition A.6], p' is (A,N)-irrational in $\mathbb{R}/\mathbb{Z}$, so p is (A,N)-irrational in S_1 with respect to $\mathcal{X}'$ and so invoking the single variable version of Green and Tao's counting lemma [35, Lemma 3.7] we have that p is $o_{A\to\infty}(1)$ equidistributed in S_1, as is predicted by Theorem 4.2.16.

On the other hand, given any nonzero linear map $l : \mathfrak{g} \to \mathbb{R}$ which maps $\mathcal{X}$ to $\mathbb{Z}$ and has height at most A, we have by the coprimality of $10A$ and $10A-1$ that $||l(a_1)||_{\mathbb{R}/\mathbb{Z}} \geq \frac{1}{(10A)(10A-1)} > \frac{A}{N}$, and so p is (A,N)-irrational in $\mathfrak{g}$. Therefore by [35, Lemma 3.7], p is $o_{A\to\infty}(1)$ equidistributed in all of $\mathbb{R}^2/\mathbb{Z}^2$.

[4] The complexity of a subspace is defined rigorously soon. See Definition 4.4.6.

[5] We transfer this example to $\mathbb{R}$ with lattice $\mathbb{Z}$ where it is clearer what the definition of the complexity of a character should be. One may check that transferring in this way is equivalent to our upcoming formal definition 4.4.5.

Remark 4.4.3. Green and Tao's quantitative factorisation of polynomial sequences [35, Lemma 2.9] states that given a polynomial sequence g on a nilmanifold, *at least one* of the following is true: (1) g is (A, N)-irrational in the nilmanifold, or (2) one may find a smooth sequence β and a rational sequence γ such that $g = \beta * g' * \gamma$, where g' takes values in a complexity $O_A(1)$ subnilmanifold of strictly smaller total dimension. Example 4.4.2 exhibits a situation where these two eventualities are not mutually exclusive (note: the smooth and rational parts in our example are trivial).

Definition 4.4.4 (Integer linear map)**.** We call an element of $\mathfrak{g}^*$ *integer* (with respect to $\mathcal{X}$) if it maps $\mathcal{X}$ to $\mathbb{Z}$. We will denote by $\mathfrak{g}^*_{\mathcal{X}}$ the set of integer linear maps with respect to $\mathcal{X}$ and we will denote by $\mathrm{Hom}_{\mathcal{X}}(\mathfrak{g}, \mathbb{R})$ those elements of $\mathfrak{g}^*_{\mathcal{X}}$ which vanish on $[\mathfrak{g}, \mathfrak{g}]$ (i.e. are Lie algebra homomorphisms).

In the following we refer to the Hermite basis $\mathcal{X}'_{S,\mathcal{X}}$ whose formal definition is given in Definition 4.9.2.

Definition 4.4.5 (Complexity of an integer linear map)**.** The *complexity* of an integer linear map $l : \mathfrak{g} \to \mathbb{R}$ on a rational subspace $S \leq \mathfrak{g}$ with respect to $\mathcal{X}$ is the ℓ_∞ norm of l with respect to the dual basis of the Hermite basis $\mathcal{X}'_{S,\mathcal{X}}$.

Definition 4.4.6 (Complexity of a subspace)**.** Given a Lie algebra $\mathfrak{g}$ with rational basis $\mathcal{X}$ and a proper rational subspace $S \leq \mathfrak{g}$, we say that S has *complexity at most C in $\mathfrak{g}$ (with respect to $\mathcal{X}$)* if S may be written as the kernel of a set of integer linear maps, each of which have complexity at most C with respect to $\mathcal{X}$. We will say that $\mathfrak{g}$ has complexity 1 in itself.

Definition 4.4.7 (Quantitative linear irrationality)**.** For a Lie algebra $\mathfrak{g}$ and a rational subspace $S \leq \mathfrak{g}$ we will say that $a \in S$ is (A, ε)-*linearly irrational in S* with respect to $\mathcal{X}$ if, for all integer linear maps $l : \mathfrak{g} \to \mathbb{R}$ which are nontrivial on S and have complexity at most A on S, we have $||l(a)||_{\mathbb{R}/\mathbb{Z}} \geq \varepsilon$.

Definition 4.4.8 (Quantitative linear irrationality for polynomial sequences)**.** For a Lie algebra $\mathfrak{g}$ and a sequence of rational subspaces $S_\bullet := (S_i)_{i=1}^s$, $S_i \leq \mathfrak{g}$, a polynomial sequence $(p(\boldsymbol{n}))_{\boldsymbol{n} \in [N]^D}$ in $\mathfrak{g}$ is (A, N)-*linearly irrational in $S_\bullet$* with respect to $\mathcal{X}$ if for all i, all coefficients of all degree-i terms of p are $(A, A/N^i)$-linearly irrational in S_i.

We will record some easy consequences of these definitions for more convenient use later on. The proofs are just linear algebra. We refer the reader to Appendix 4.9 for some further details. Recall that throughout this section we are assuming that $\dim \mathfrak{g} = O(1)$.

Lemma 4.4.9 (Complexity of subspaces under operations)**.** *Let $C > 1$. Let S, T be rational subspaces of $\mathfrak{g}$ of complexity at most C with respect to $\mathcal{X}$ and let U be a rational subspace of a distinct, dimension-$O(1)$ Lie algebra $\mathfrak{h}$ of complexity at most C with respect to a rational basis $\mathcal{Y}$. Then*

1. *$S + T$ has complexity at most $C^{O(1)}$,*
2. *$[S, T]$ has complexity at most $M^{O(1)}C^{O(1)}$, where M is an upper bound on the height of the largest (rational) structure constant of $\mathfrak{g}$ with respect to $\mathcal{X}$,*
3. *$S \otimes U$ has complexity at most $C^{O(1)}$ in $\mathfrak{g} \otimes \mathfrak{h}$ with respect to $\mathcal{X} \otimes \mathcal{Y}$.*

Lemma 4.4.10 (Complexity of maps under lifting and restriction)**.** *Let $A, B, C > 1$ be integers and let S, T be rational subspaces of $\mathfrak{g}$ of complexity at most C with respect to $\mathcal{X}$ such that $S \leq T$. Then*

1. *$\mathcal{X}'_{S,\mathcal{X}}$ has index $C^{O(1)}$ in $\mathcal{X}$,*
2. *if l is an integer linear map on T of complexity at most A, then $l|_S$ is an integer linear map of complexity at most $C^{O(1)}A$ on S,*
3. *if l is an integer linear map on S of complexity at most B then there is an integer linear map $\tilde{l}$ on T of complexity at most $C^{O(1)}B$ such that $\tilde{l}|_S = Kl$, where $K = C^{O(1)}$.*

Example 4.4.2 demonstrated that there are elements which are (A, ε)-linearly irrational with respect to more than one subspace. In this example, the nontrivial subspace was of complexity $A^{O(1)}$. The following lemma gives a sense in which this is the only way in which such a phenomenon can occur.

Lemma 4.4.11. *Let $A > 1$. Let $a \in \mathfrak{g}$ and let S, T be subspaces of $\mathfrak{g}$ of complexity at most A which contain a. There is a constant $k > 1$ (which may depend on $\dim \mathfrak{g}$) such that for every $\varepsilon > 0$ if a is (A^k, ε)-linearly irrational in T then $T \leq S$.*

Proof. We show that $S^\perp \leq T^\perp$. Let $l : \mathfrak{g} \to \mathbb{R}$ have complexity at most A and lie in $S^\perp$. Since $a \in S$ we have $l(a) = 0$. Furthermore l has complexity at most $A^{O(1)}$ with respect to $\mathcal{X}'_{\mathcal{X},T}$ by Lemma 4.4.10 and so for $k = O(1)$ sufficiently large, this contradicts the (A^k, ε)-linear irrationality of a in T, unless $l \in T^\perp$. Letting $\mathfrak{g}^*_{\mathcal{X}}(A)$ denote the set of integer linear maps of complexity at most A with respect to $\mathcal{X}$, we have so far that $S^\perp \cap \mathfrak{g}^*_{\mathcal{X}}(A) \subset T^\perp$ and taking the vector space closure on both sides we have $\mathrm{span}_{\mathbb{R}}\{S^\perp \cap \mathfrak{g}^*_{\mathcal{X}}(A)\} \subset T^\perp$. But by the definition of the complexity of S, we

have that $S^\perp$ may be written as the span of maps of complexity at most A and so $S^\perp = \mathrm{span}_{\mathbb{R}}\{S^\perp \cap \mathfrak{g}^*_{\mathcal{X}}(A)\}$. This completes the proof. □

In particular, there is a polynomial threshold above which linear irrationality may be obtained on at most one rational subspace of bounded complexity. In general, however, the sequence of subspaces on which we are asserting linear irrationality is not determined by the polynomial sequence and so our counting lemma will take this as additional input.

Similarly to the use of Leibman's criterion for equidistribution in the qualitative setting, we will introduce here Green and Tao's criterion for quantitative equidistribution. First we present a quantitative version of additive irrationality; see Subsection 4.8.1 for a definition of the $||\cdot||_{C^\infty[\boldsymbol{N}]}$-norm.

Definition 4.4.12 (Quantitative additive irrationality)**.** Let $A > 1$, let D be a positive integer, let $N_1, \dots, N_D \geq 1$ be integers and let $[\boldsymbol{N}]$ denote the set $[N_1]\times\cdots\times[N_D]$. A polynomial sequence p on $\mathfrak{g}$ is $(A, \boldsymbol{N})$-additively irrational with respect to $\mathcal{X}$ if for all nontrivial $\eta \in \mathrm{Hom}_{\mathcal{X}}(\mathfrak{g}, \mathbb{R})$ of complexity at most A, we have $||\eta \circ p||_{C^\infty[\boldsymbol{N}]} > A$.

Remark 4.4.13. Though we have not done so here, one may develop the various notions of irrationality in a transparently analogous way to highlight that these notions essentially only differ by the subalgebras of $\mathfrak{g}$ on which the set of maps under consideration are required to vanish ($[\mathfrak{g},\mathfrak{g}]$, $\mathfrak{h}_i$, none for additive, filtration, linear irrationality respectively). One may define quantitative additive irrationality for elements of (rather than just polynomial sequences in) $\mathfrak{g}$, and define quantitative filtration/linear irrationality for multivariate polynomial sequences; these definitions are not required in this chapter and so are not provided.

Next we define quantitative equidistribution in $(\mathfrak{g}, \mathcal{X})$.

Definition 4.4.14 (Quantitative equidistribution in the Lie algebra)**.** Let $\mathcal{X}$ be a rational basis in $\mathfrak{g}$. Then a polynomial sequence $(p(\boldsymbol{n}))_{\boldsymbol{n}\in[\boldsymbol{N}]}$ is δ-equidistributed in $(\mathfrak{g}, \mathcal{X})$ if $(\exp(p(\boldsymbol{n})))_{\boldsymbol{n}\in[\boldsymbol{N}]}$ is δ-equidistributed in G/Γ, where Γ is the smallest multiplicative lattice containing $\langle\mathcal{X}\rangle = \mathrm{span}_{\mathbb{Z}}\,\mathcal{X}$.

Green and Tao showed [39, Theorem p.3] that quantitative additive irrationality implies quantitative equidistribution. Their result concerns equidistribution in a nilmanifold; we defer to Appendix 4.8 for a deduction of the following Lie algebra version.

Theorem 4.4.15 (Green and Tao's criterion for quantitative equidistribution in the Lie algebra)**.** *Let $0 < \delta < 1/2$, let $N_1, \dots, N_D \geq 1$ and let $[\boldsymbol{N}]$ denote the set $[N_1] \times \cdots \times [N_D]$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with dimension d, step s and basis $\mathcal{X}$ which is $1/\delta$-rational. Let $(p(\boldsymbol{n}))_{\boldsymbol{n} \in \mathbb{Z}^D}$ be a polynomial sequence in $\mathfrak{g}$. There are positive constants $c_{d,s,D}, c'_{d,s,D}$ such that at least one of the following is true:*

1. *there is some $N_i < \delta^{-c'_{d,s,D}}$,*
2. *$(p(\boldsymbol{n}))_{\boldsymbol{n} \in [\boldsymbol{N}]}$ is δ-equidistributed in $(\mathfrak{g}, \mathcal{X})$,*
3. *p is not $(\delta^{-c_{d,s,D}}, \boldsymbol{N})$-additively irrational in $(\mathfrak{g}, \mathcal{X})$.*

Furthermore, when $N_1 = \cdots = N_D$, the first option above may be removed.

We are ready to prove the main theorem of this section. We only prove a result for averages over a cube $[N]^D$ (i.e. when $N_1 = \cdots = N_D = N$), but it should be clear how to obtain a result for the more general case in which not all sides have equal length.

Theorem 4.4.16 (Quantitative counting lemma, Theorem 4.1.2)**.** *Let $N, A, M \geq 2$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with rational basis $\mathcal{X}$, dimension d and rational structure constants of height at most M. Let $S_\bullet$ be a sequence of rational subspaces of complexity M in $\mathfrak{g}$. Suppose that $p(n)$ is an (A, N)-linearly irrational polynomial sequence of degree s in $S_\bullet$. Then there are constants $0 < c_1, c_2 = O_{d,s,D,t,\Psi}(1)$ such that the polynomial sequence $p^\Psi(\boldsymbol{x}) := (p(\psi_1(\boldsymbol{x})), \dots, p(\psi_t(\boldsymbol{x}))$ is $O(M^{c_1}/A^{c_2})$-equidistributed on $\mathfrak{g}^\Psi(S_\bullet)$ with respect to $\mathcal{X}^\Psi$, the Hermite basis for $\mathfrak{g}^\Psi(S_\bullet)$ in $(\mathfrak{g}^t, \mathcal{X}^t)$.*

Proof. We let constants depend on s, d, D, t, Ψ. Suppose that p is (A, N)-linearly irrational in $(S_i)_{i=1}^s$. Write $p(n) = \sum_{i=1}^s a_i n^i$, where a_i is $(A, A/N^i)$-linearly irrational in S_i (see Definition 4.4.8).

Since M bounds the complexity of $S_\bullet$ and the height of the structure constants of $\mathfrak{g}$, the description of $\mathfrak{g}^\Psi(S_\bullet)$ in Proposition 4.3.1 and repeated application of Lemma 4.4.9 yields that $\mathfrak{g}^\Psi(S_\bullet)$ is of complexity at most $M^{O(1)}$ in $\mathfrak{g} \otimes \mathbb{R}^t$ with respect to $\mathcal{X} \otimes \mathcal{Z}$, where $\mathcal{Z}$ is the standard basis for $\mathbb{Z}^t$. Denote by $\mathcal{X}^\Psi$ the Hermite basis for $\mathfrak{g}^\Psi(S_\bullet)$ and $\mathcal{X} \otimes \mathcal{Z}$ and note that $\mathcal{X}^\Psi$ has rational structure constants of height at most $M^{O(1)}$. Let $\delta > 0$ be a parameter which we will choose later and suppose that p^Ψ is not δ-equidistributed in $\mathfrak{g}^\Psi(S_\bullet)$ with respect to $\mathcal{X}^\Psi$.

Invoking Theorem 4.4.15 with respect to the basis $\mathcal{X}^\Psi$, we obtain a nontrivial $\eta \in \mathrm{Hom}_{\mathcal{X}^\Psi}(\mathfrak{g}^\Psi(S_\bullet), \mathbb{R})$ which has complexity at most $M^{O(1)}\delta^{-O(1)}$ and such that

$||\eta \circ p^\Psi||_{C^\infty[N]^D} < M^{O(1)}\delta^{-O(1)}$. We have then, recalling the definition of the $C^\infty_{\mathcal{M}}$ norms from Definition 4.8.2 and invoking Lemma 4.8.3, that $||s!\eta \circ p^\Psi||_{C^\infty_{\mathcal{M}}[N]^D} < M^{O(1)}\delta^{-O(1)}$. Now we compute

$$s!\eta(p^\Psi(\boldsymbol{x})) = \sum_{i=1}^{s} s!\eta(a_i \otimes \Psi(\boldsymbol{x})^i) = \sum_{m\in\mathcal{M}} \sum_{i=1}^{s} s!\eta(a_i \otimes v_{m,i})m(\boldsymbol{x}), \tag{4.6}$$

where, as in Subsection 4.1.3, $v_{m,i} \in \mathbb{Z}^t$ is the coefficient of the monomial $m(\boldsymbol{x})$ in the polynomial $\Psi^i(\boldsymbol{x})$. Thus, by the definition of the $C^\infty_{\mathcal{M}}[N]^D$ norm Definition 4.8.2, we have for all m, i that

$$||s!\eta(a_i \otimes v_{m,i})||_{\mathbb{R}/\mathbb{Z}} < M^{O(1)}\delta^{-O(1)}/N^i. \tag{4.7}$$

By Lemma 4.4.10, there is $\tilde{\eta} \in (\mathfrak{g} \otimes \mathbb{R}^t)^*$ of complexity $M^{O(1)}\delta^{-O(1)}$ with respect to $\mathcal{X} \otimes \mathcal{Z}$ such that $\tilde{\eta}|_{\mathfrak{g}^\Psi(S_\bullet)} = M^{O(1)}s!\eta$. It follows then that $\tilde{\eta}(\cdot \otimes v_{m,i})$ is an integer linear map on $\mathfrak{g}$ of complexity $M^{O(1)}\delta^{-O(1)}$ with respect to $\mathcal{X}$ and so invoking Lemma 4.4.10 once again, we have that $\tilde{\eta}(\cdot \otimes v_{m,i})|_{S_i}$ has complexity $M^{O(1)}\delta^{-O(1)}$ with respect to $\mathcal{X}_i'$, the Hermite basis for S_i.

Now note that Equation (4.7) implies that $||\tilde{\eta}(a_i \otimes v_{m,i})||_{\mathbb{R}/\mathbb{Z}} < M^{O(1)}\delta^{-O(1)}/N^i$ for all i, m. Therefore, if $A \geq M^{O(1)}\delta^{-O(1)}$ (the complexity of the map $\tilde{\eta}(\cdot \otimes v_{m,i})|_{S_i}$) and $A \geq M^{O(1)}\delta^{-O(1)}$ so that $||\tilde{\eta}(a_i \otimes v_{m,i})||_{\mathbb{R}/\mathbb{Z}} < A/N^i$, then the $(A, A/N^i)$-linear irrationality of a_i in S_i implies that $\tilde{\eta}(\cdot \otimes v_{m,i})$ is trivial on S_i, i.e. $\tilde{\eta}(S_i \otimes v_{m,i}) = 0$. Since the $\{v_{m,i}\}_{m\in\mathcal{M}}$ span V^i, we have $\tilde{\eta}(S_i \otimes V^i) = 0$ and so from definitions $\eta(S_i \otimes V^i) = 0$. Since η is a Lie algebra homomorphism to an abelian Lie algebra, η vanishes on the smallest Lie subalgebra of $\mathfrak{g}^\Psi(S_\bullet)$ containing the subspaces $S_i \otimes V^i$, which is $\mathfrak{g}^\Psi(S_\bullet)$ itself. That is, η is trivial on $\mathfrak{g}^\Psi(S_\bullet)$, a contradiction. So p^Ψ is indeed (upon solving for δ) $(M^{O(1)}/A^{O(1)})$-equidistributed in $\mathfrak{g}^\Psi(S_\bullet)$. □

4.5 The factorisation theorem for linear irrationality

As in the qualitative setting, a counting lemma is of little use without a method to reduce a general polynomial sequence to something it can handle. Our main result of this section is Theorem 4.5.10, a factorisation theorem for linear irrationality in much the same spirit as [35, Lemma 2.10] and [38, Theorem 1.19], which may be viewed as factorisation theorems for the notions of filtration and additive irrationality respectively. Throughout this section we assume d, s, t, D are of bounded size and let constants depend on them. As we did in the qualitative setting, we begin with an additive decomposition.

4.5.1 Additive decomposition for linear irrationality

In this subsection we prove an additive decomposition of vector space elements into linearly irrational, rational and small parts (Proposition 4.5.4), and as a corollary obtain a similar decomposition for polynomial sequences (Corollary 4.5.5).

Definition 4.5.1 (Quantitative relative linear irrationality)**.** Let $A > 1$ and $\varepsilon > 0$, let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$, and let U, T be rational subspaces. An element $a \in U$ is (A, ε)-*linearly irrational in* $U \bmod T$ *with respect to* $\mathcal{X}$ if, for all integer linear maps l which have complexity at most A in U with respect to $\mathcal{X}'_{\mathcal{X},U}$, which are nontrivial on U and which vanish on $U \cap T$, we have that $||l(a)||_{\mathbb{R}/\mathbb{Z}} \geq \varepsilon$.

Definition 4.5.2 (Quantitative rationality)**.** Let $a \in \mathfrak{g}$, $A \in \mathbb{N}^+$. We say that a is A-rational with respect to $\mathcal{X}$ if $a \in \mathfrak{g}_{\mathbb{Q}}$ and, when written with respect to the basis $\mathcal{X}$, the vector for a has entries in $\frac{1}{A}\mathbb{Z}$. A polynomial sequence p in $\mathfrak{g}$ is A-rational if all of its coefficients are A-rational.

Definition 4.5.3 (Quantitative smallness)**.** Let $\varepsilon > 0$. An element $a \in \mathfrak{g}$ is ε-small with respect to $\mathcal{X}$ if $||a||_\infty \leq \varepsilon$, where $||\cdot||_\infty$ is the ℓ_∞ norm of a with respect to the basis $\mathcal{X}$. A polynomial sequence $(p(\boldsymbol{n}))_{\boldsymbol{n} \in [N]^D}$ in $\mathfrak{g}$ is (A, N)-*small* with respect to $\mathcal{X}$ if, for all monomials $m \in \mathcal{M}$, we have that c_m is $A/N^{\deg m}$-small, where c_m is the coefficient of $m(\boldsymbol{n})$ in p.

The following decomposition of vector space elements into linearly irrational, rational and error terms allows us to perturb an element which is linearly irrational in a quotient space by elements of the quotient to obtain an element which is linearly irrational in the original space. Note that this result includes the case where the quotient space is trivial (that is, $T = U$ in the upcoming notation), whereupon we obtain a decomposition of an arbitrary element in the vector space.

Proposition 4.5.4 (Quantitative decomposition of vector space elements)**.** *Let $A > 1$ be an integer, let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$, let U, T be rational subspaces of $\mathfrak{g}$, each of complexity $A^{O(1)}$. Then there is a constant $c > 1$ (which may depend on $\dim U, \dim T$) such that the following is true. Let $0 < \varepsilon < 1/2$ and $a \in U$, and suppose that a is $(A^c, A^c\varepsilon)$-linearly irrational in $U \bmod T$. Then there are elements $a_s \in U \cap T$, $a_r \in (U \cap T)_{\mathbb{Q}}$ such that a_r is $A^{O(1)}$-rational, a_s is $A^{O(1)}\varepsilon$-small and $a - a_r - a_s$ is (A, ε)-linearly irrational in a subspace $U' \leq U$ of complexity $A^{O(1)}$ with respect to $\mathcal{X}$.*

Proof. If a is (A,ε)-linearly irrational in U then we are done and so supposing it is not, we have an integer linear map l of complexity at most A with respect to the Hermite basis $\mathcal{X}'_{U,\mathcal{X}} = \mathcal{X}'$ which is nontrivial on U such that $||l(a)||_{\mathbb{R}/\mathbb{Z}} < \varepsilon$. Note also that l must be nontrivial as a map on $U \cap T$ due to the linear irrationality of a in $U \bmod T$. This is the $m = 1$ case of the following setup; our proof proceeds inductively on m.

Let $L = L_m$ be an $m \times \dim U$ integer matrix whose rows are linearly independent, integer linear maps $U \to \mathbb{R}$ with respect to the basis $\mathcal{X}' := \mathcal{X}'_{U,\mathcal{X}}$ and which possess the following properties. Firstly, the projection of the rows are linearly independent as maps in $(U \cap T)^*$. Furthermore, for each row l, we have that l has complexity $A^{O(1)}$ with respect to $\mathcal{X}'$ and that $||l \cdot a||_{\mathbb{R}/\mathbb{Z}} \leq A^{O(1)}\varepsilon$. We will show below that either:

1. there are elements $a_s \in U \cap T$, $a_r \in (U \cap T)_{\mathbb{Q}}$ such that with respect to $\mathcal{X}'$, a_r is $A^{O(1)}$-rational, $||a_s||_\infty = A^{O(1)}\varepsilon$ and $a - a_r - a_s$ is (A,ε)-linearly irrational in $\ker L \leq U$ (i.e. the proof is complete upon observing that with respect to $\mathcal{X}$, a_s is still $A^{O(1)}\varepsilon$-small and a_r is still $A^{O(1)}$-rational), or

2. we may find some L_{m+1} which fits the description above with $m+1$ in place of m.

Note that in situation (1), we set $U' = \ker L$. One may use the fact that the rows of L are linearly independent as maps in $(U \cap T)^*$ to show that $\dim(\ker L + T) = \dim(U + T)$. In situation (2), the constants implicit in the $O(1)$ notation may change when we pass from m to $m+1$. However this can occur only finitely many times since $\dim(U \cap T) = O(1)$ and the rows of each L_m are linearly independent in $(U \cap T)^*$. In particular, we must, at some point, find ourselves in situation (1), completing the proof. It remains to prove the above claim.

Let $L = L_m$ be as above. For $i = 1, \ldots, m$, we have by supposition that $l_i \cdot a = n_i + e_i$, for some integer n_i and $0 < |e_i| < \varepsilon$. We may write this simultaneously for $i = 1, \ldots, k$ as $La = n + e$ where n, e are vectors consisting of the n_i, e_i respectively.

We claim that there is a solution to $Lx = n$ with $x \in U \cap T$ and such that each entry of x with respect to $\mathcal{X}'$ lies in $(\frac{1}{A^{O(1)}})\mathbb{Z}$; we will let this be a_r. Recall that by the induction statement L has full rank as a family of linear maps on $U \cap T$. Let M be the change of basis matrix such that $[L]_{\mathcal{X}'} = [LM]_{\mathcal{X}''}$, where $\mathcal{X}'' = \mathcal{X}''_{\mathcal{X},U\cap T}$ is the Hermite basis for $U \cap T$ in $\mathcal{X}$. Note that M is an integer matrix since $\langle \mathcal{X}' \rangle = U \cap \langle \mathcal{X} \rangle \supset \mathcal{X}''$, and further, complexity bounds ensure that M has entries of size at most $A^{O(1)}$. Let L' be a square submatrix of LM obtained by choosing m leading columns. Then it is a standard linear algebra fact that the simultaneous equations $L'x' = n$ are solvable

in $\frac{1}{|\det L'|}\mathbb{Z}$. Observing that $|\det L'| = A^{O(1)}$ and then padding the solution x' with $\dim(U \cap T) - m$ zero entries to form an appropriate vector $\tilde{x}$ yields a solution $LM\tilde{x} = n$ in $((\frac{1}{A^{O(1)}})\mathbb{Z})^{\dim U \cap T}$. Setting $x = M\tilde{x}$ proves the claim made at the beginning of the paragraph.

Next, we claim that there is a solution to $Lx = e$ with $x \in U \cap T$ and $||x||_\infty = A^{O(1)}\varepsilon$. This may be proven, for example, by choosing L' as above, setting $x' = L'^{-1}e$, upper bounding $||L'^{-1}||_\infty$ directly using using Cramer's rule and then augmenting x' with zero entries to form a vector $\tilde{x}$ with $LM\tilde{x} = e$. Let $a_s = M\tilde{x} \in U \cap T \subset U$ so that $La_s = e$ and $||a_s||_\infty = A^{O(1)}\varepsilon$ (where here the ℓ_∞ norm is of course with respect to the basis $\mathcal{X}'$).

Now, $a_p := a - a_r - a_s$ is contained in $\ker L$. If a_p is (A, ε)-linearly irrational in $\ker L$ then we are in situation (1) and we are done, so suppose not. Then there is some nontrivial integer linear map l of complexity at most A in $\ker L$ such that $||l(a_p)||_{\mathbb{R}/\mathbb{Z}} < \varepsilon$. Also, it follows from definitions that $\ker L$ is of complexity $A^{O(1)}$ and so by Lemma 4.4.10 we may find an integer linear map $\tilde{l} \in U^*$ of complexity $A^{O(1)}$ such that $\tilde{l}|_{\ker L} = A^{O(1)}l$. Furthermore since l is nontrivial on $\ker L$, we have that $\tilde{l}$ does not lie in the row span of L. We claim that in fact (the projection of) the rows of L and $\tilde{l}$ are linearly independent in $(U \cap T)^*$. Suppose not, for a contradiction with the $(A^c, A^c\varepsilon)$-linear irrationality of a in $U \bmod T$. Then there is some linear combination $b\tilde{l} + \sum_{i=1}^m b_i l_i = 0 \in (U \cap T)^*$, where $b \neq 0$ because $\{l_1, \ldots, l_m\}$ are linearly independent in $(U \cap T)^*$. In fact, since the entries of each $l_i, \tilde{l}$ are integers of size $A^{O(1)}$, one may check e.g. by standard Gaussian elimination that we can insist that the constants b_i, b are themselves integers of size $A^{O(1)}$ (with a different constant $O(1)$ which may depend on the dimension of the space). For these b, b_i, let $l' = b\tilde{l} + \sum_{i=1}^m b_i l_i \in (U \cap T)^\perp \subset U^*$. Then $l'(a) = l'(a_p)$ (since $a_r, a_s \in U \cap T$), $l'(a_p) = b\tilde{l}(a_p)$ (since $a_p \in \ker L$), and $b\tilde{l}(a_p) = bA^{O(1)}l(a_p)$ (by the construction of $\tilde{l}$ and since $a_p \in \ker L$). Thus

$$||l'(a)||_{\mathbb{R}/\mathbb{Z}} = ||bA^{O(1)}l(a_p)||_{\mathbb{R}/\mathbb{Z}} \leq bA^{O(1)}||l(a_p)||_{\mathbb{R}/\mathbb{Z}} < A^{O(1)}\varepsilon.$$

However, l' is of complexity $A^{O(1)}$ with respect to $\mathcal{X}'$. For c chosen suitably large, this contradicts the $(A^c, A^c\varepsilon)$-linear irrationality of a in $U \bmod T$. Thus we have that (the projection of) the rows of L and $\tilde{l}$ are linearly independent in $(U \cap T)^*$.

Finally, let C be the smallest positive integer such that $C\tilde{l}(a_r) \in \mathbb{Z}$; note that $C = A^{O(1)}$. Then we compute

$$\begin{aligned} ||C\tilde{l}(a)||_{\mathbb{R}/\mathbb{Z}} &\leq ||C\tilde{l}(a_p)||_{\mathbb{R}/\mathbb{Z}} + ||C\tilde{l}(a_r)||_{\mathbb{R}/\mathbb{Z}} + ||C\tilde{l}(a_s)||_{\mathbb{R}/\mathbb{Z}} \\ &\leq CA^{O(1)}\varepsilon + 0 + CA^{O(1)}\varepsilon \leq A^{O(1)}\varepsilon. \end{aligned}$$

One may then form L_{m+1} by adding the row $C\tilde{l}$ to the bottom of $L = L_m$. This puts us in situation (2) and completes the proof. □

Corollary 4.5.5 (Quantitative additive decomposition of polynomial sequences)**.** *Let $A, N > 1$ be integers, let $U_\bullet = (U_i)_{i=1}^s$ be a sequence of rational subspaces of $\mathfrak{g}$ each of complexity $A^{O(1)}$ and let T be a rational subspace of $\mathfrak{g}$ which is also of complexity $A^{O(1)}$. There is a constant $c > 1$ (which may depend on $\dim U$, $\dim T$) such that the following is true. Let p be a polynomial sequence defined by $p(n) = \sum_{i=1}^s a_i n^i$ such that $a_i \in U_i$ for all i and which is (A^c, N)-linearly irrational in $U_\bullet \bmod T$.*[6] *Then there are polynomial sequences e, r in $U_\bullet \cap T$ of degree s and a sequence of subspaces $(U_i')_{i=1}^s \leq U_\bullet$ with each U_i' of complexity $A^{O(1)}$ such that e is $(A^{O(1)}, N)$-small, r is $A^{O(1)}$-rational, and $p - e - r$ is (A, N)-linearly irrational in $U_\bullet' := (U_i')_{i=1}^s$. Furthermore, $\dim(U_i' + T) = \dim(U_i + T)$ for all i.*

Proof. Let $p(n) = \sum_{i=1}^s a_i n^i$. If p is (A, N)-linearly irrational in $U_\bullet$ then we are done so suppose it is not and conduct the following argument for each index i such that a_i fails to be $(A, A/N^i)$-linearly irrational in U_i. By definition we have that each such a_i is $(A^c, A^c/N^i)$-linearly irrational in $U_i \bmod T$. Trivially then, a_i is $(A^{c-1}, A^c/N^i)$-linearly irrational in $U_i \bmod T$. Now invoke Proposition 4.5.4 for a_i and let c_i be a constant (obtained from the statement of Proposition 4.5.4) so that we may find $a_{e,i} \in U_i \cap T$ which is $(A^{O(1)}/N^i)$-small and $a_{r,i} \in U \cap T$ which is $A^{O(1)}$-rational such that $a_i' := a_i - a_{r,i} - a_{e,i}$ is $(A^{(c-1)/c_i}, A/N^i)$-linearly irrational in some subspace $U_i' \leq U_i$ of complexity $A^{O(1)}$. If $(c-1)/c_i \geq 1$ then a_i' is $(A, A/N^i)$-linearly irrational in U_i'; insist that c is chosen so that this is the case. Let $r(n) = \sum_{i=1}^s a_{r,i} n^i$ and $e(n) = \sum_{i=1}^s a_{e,i} n^i$. The result follows. □

4.5.2 Multiplicative factorisation for strong irrationality

The goal of this subsection is to prove Theorem 4.5.10. This may be viewed as analogous to Green and Tao's factorisation theorems for polynomial sequences (see [35], [38]) for the notion of linear irrationality. We will also want to control the following notion of irrationality which comes from [35] (see in particular [35, Appendix A]).

In the following one will need to recall the definition of the subalgebras $\mathfrak{h}_i$: Definition 4.2.10.

[6] The operations are conducted pointwise on the sequence i.e. this may be translated to mean that each coefficient a_i is $(A^c, A^c/N^i)$-linearly irrational in $U_i \bmod T$.

Definition 4.5.6 (Filtration irrationality)**.** Let $A > 1, \varepsilon > 0$. We say $a \in \mathfrak{g}_i$ is (A, ε)-filtration irrational in $\mathfrak{g}_i$ with respect to a rational basis $\mathcal{X}$ for $\mathfrak{g}$ if $||l(a)||_{\mathbb{R}/\mathbb{Z}} \geq \varepsilon$ for all $l \in \mathfrak{g}_{\mathbb{Q}}^*$ which are nontrivial on $\mathfrak{g}_i$, have complexity at most A on $\mathfrak{g}_i$ and which vanish on $\mathfrak{h}_i$.

One may equivalently define $a \in \mathfrak{g}_i$ to be (A, ε)-filtration irrational in $\mathfrak{g}_i$ if and only if it is (A, ε)-linearly irrational in $\mathfrak{g}_i \bmod \mathfrak{h}_i$, as per Definition 4.5.1.

Definition 4.5.7 (Filtration irrationality for polynomial sequences)**.** Let $A, N > 1$ and let $\mathfrak{g}$ have rational basis $\mathcal{X}$. Let p be a polynomial sequence adapted to $\mathfrak{g}_\bullet$ which is defined by $p(n) := \sum_{i=1}^{s} a_i n^i$. We say that p is (A, N)-*filtration irrational* if a_i is $(A, A/N^i)$-filtration irrational in $\mathfrak{g}_i$ for every i. Furthermore we say that p is (A, N)-*filtration irrational in* $\mathfrak{g} \bmod \mathfrak{g}_j$ if a_i is $(A, A/N^i)$-filtration irrational in $\mathfrak{g}_i$ for $i = 1, \ldots, j-1$.

Before proving Theorem 4.5.10, we need some lemmas. Henceforth, for a univariate polynomial sequence q, let $(q)_j$ denote the coefficient of the term of degree j in q. The following equation is an easy consequence of definitions.

Lemma 4.5.8. *For polynomials q, q' adapted to $\mathfrak{g}_\bullet$ we have*

$$(q * q')_j = (q)_j + (q')_j \mod \mathfrak{h}_j,$$

for all j.

Lemma 4.5.9. *Let $A > R > 1$ be integers and let $\delta, \varepsilon > 0$ be such that $\delta - (\dim \mathfrak{g}_i)A\varepsilon > 0$. Suppose that $a, r, s \in \mathfrak{g}_i$ are such that a is (A, δ)-filtration irrational in $\mathfrak{g}_i$, r is R-rational in $\mathfrak{g}_i$ and s is ε-small. Then $a - r - s$ is $(A/R, (\delta - (\dim \mathfrak{g}_i)A\varepsilon)/R)$-filtration irrational in $\mathfrak{g}_i$.*

Proof. Let $\beta = (\delta - (\dim \mathfrak{g}_i)A\varepsilon)/R$ and $b = a - r - s$. We prove the contrapositive: suppose b is not $(A/R, \beta)$-filtration irrational. We may thus choose an integer linear map l, nontrivial on $\mathfrak{g}_i$, which vanishes on $\mathfrak{h}_i$ and with complexity A/R in $\mathfrak{g}_i$ such that $||l(b)||_{\mathbb{R}/\mathbb{Z}} < \beta$. Then Rl has complexity A on $\mathfrak{g}_i$ and

$$||Rl(a)||_{\mathbb{R}/\mathbb{Z}} \leq ||Rl(b)||_{\mathbb{R}/\mathbb{Z}} + ||Rl(r)||_{\mathbb{R}/\mathbb{Z}} + ||Rl(s)||_{\mathbb{R}/\mathbb{Z}} < R\beta + 0 + (\dim \mathfrak{g}_i)A\varepsilon = \delta,$$

so a is not (A, δ)-filtration irrational. □

Recall that we have already proven Proposition 4.2.28, a qualitative version of the following theorem which may be of interest to a reader wishing to extricate the important parts of the upcoming proof.

Theorem 4.5.10 (Quantitative factorisation of polynomial sequences)**.** *Let $A, M \geq 2$ and $d, s \geq 1$ be integers such that $AM > d$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with filtration $\mathfrak{g}_\bullet$, rational basis $\mathcal{X}$, dimension d and rational structure constants of height at most M. Let p be a polynomial sequence adapted to $\mathfrak{g}_\bullet$ and sequence of subspaces $S_\bullet \leq \mathfrak{g}_\bullet$. Finally, insist that $S_\bullet, \mathfrak{g}_\bullet$ are of complexity $A^{O(1)}$ in $\mathfrak{g}$, where here and henceforth $O(1)$ terms may depend on d, s. Then we may write $p = e * p' * r$ where e, p', r are polynomial sequences in $\mathfrak{g}$ adapted to $\mathfrak{g}_\bullet$, p' is (A, N)-linearly irrational in a sequence of subspaces $S'_\bullet$ of complexity $A^{O(1)}$, r is $(AM)^{O(1)}$-rational, and e is $((AM)^{O(1)}, N)$-small. Furthermore:*

- *writing $p(n) = \sum_{i=1}^{s} a_i n^i$, if j is the smallest integer such that a_j is not (A, N)-linearly irrational in S_j, then we may take $S'_k = S_k$ for all $k = 1, \ldots, j-1$ and have S'_j strictly contained in S_j,*
- *there is a constant $c > 1$ such that if $B > AM$ and p is (B^c, N)-filtration irrational then p' is (B, N)-filtration irrational.*

Proof. The factorisation process proceeds inductively. We will demonstrate the first couple of steps, from which the general process should be clear. The process terminates because $\mathfrak{g}$ is nilpotent of finite step s.

If p is (A, N)-linearly irrational in $S_\bullet$ then we are done trivially. Suppose not, let $b \geq 1$ be a constant to be chosen later, and write $p(n) = \sum_{i=1}^{s} a_i n^i$ and let j be the smallest integer such that a_j is not (A^b, N)-linearly irrational in S_j (since $b \geq 1$, such a j certainly exists). We begin by applying Corollary 4.5.5 to the degree $\geq j$ terms of p with $U_\bullet = S_\bullet$ and $T_\bullet = (\mathfrak{g}, \ldots, \mathfrak{g})$ (note that p is arbitrarily linearly irrational in $U_\bullet$ mod T) to obtain $p_1 := p - \tilde{e}_1 - \tilde{r}_1$, where $\tilde{e}_1, \tilde{r}_1$ are polynomial sequences adapted to $\mathfrak{g}_\bullet$ with all terms of degree at least j, $\tilde{e}_1$ is $((AM)^{O(1)}, N)$-small, $\tilde{r}_1$ is $A^{O(1)}$-rational and p_1 is (A^b, N)-linearly irrational in some sequence of subspaces $S^{(1)}_\bullet$ with $S^{(1)}_k = S_k$ for $k < j$ and $S^{(1)}_j < S_j$. Then set

$$\tilde{p}_2 := \tilde{e}_1^{-1} * p * \tilde{r}_1^{-1} = p - \tilde{e}_1 - \tilde{r}_1 - \frac{1}{2}([\tilde{e}_1, p] + [\tilde{e}_1, r_1] + [p, \tilde{r}_1]) - \ldots.$$

Note that $\tilde{p}_2 = p_1 \mod \mathfrak{g}_2$ and furthermore that $(\tilde{p}_2)_k = (p_1)_k$ for all $k \leq j$ (since every term coming from a Lie bracket will have degree at least $j+1$). The former fact implies that $\tilde{p}_2$ is (A^b, N)-linearly irrational in $(S^{(1)}_\bullet + \mathfrak{g}_2) \mod \mathfrak{g}_2$.

Next we apply Corollary 4.5.5 to the degree $> j$ terms of $\tilde{p}_2$ with $U_\bullet = S^{(1)}_\bullet + \mathfrak{g}_2$ and $T = (\mathfrak{g}_2, \ldots, \mathfrak{g}_2)$. We obtain $p_2 := \tilde{p}_2 - \tilde{e}_2 - \tilde{r}_2$, where $\tilde{e}_2, \tilde{r}_2$ are polynomial sequences in $\mathfrak{g}_2$, adapted to $\mathfrak{g}_\bullet$ and with all terms of degree $> j$, $\tilde{e}_2$ is $(A^{O(1)}, N)$-small, $\tilde{r}_2$ is

$A^{O(1)}$-rational, and p_2 is $(A^{b/c_2}, N)$-linearly irrational (where c_2 is the constant from the statement of Corollary 4.5.5) in a sequence of subspaces $S_\bullet^{(2)}$ with $S_k^{(2)} = S_k^{(1)} = S_k$ for $k < j$ and $S_j^{(2)} = S_j^{(1)} < S_j$.

If $s = 2$ (i.e. $\mathfrak{g}_3 = 0$), we conclude:

$$\begin{aligned} p &= \tilde{e}_1 * \tilde{p}_2 * \tilde{r}_1 \\ &= \tilde{e}_1 * (\tilde{e}_2 + p_2 + \tilde{r}_2) * \tilde{r}_1 \\ &= \tilde{e}_1 * \tilde{e}_2 * p_2 * \tilde{r}_2 * \tilde{r}_1, \end{aligned}$$

where in the final line we have used the fact that $\tilde{e}_2, \tilde{r}_2 \in \mathfrak{g}_2$, Baker-Campbell-Hausdorff and that $[\mathfrak{g}, \mathfrak{g}_2] \subset \mathfrak{g}_3$. Furthermore Lemma 4.8.15 yields that $e_2 := \tilde{e}_1 * \tilde{e}_2$ is $((AM)^{O(1)}, N)$-small and $r_2 := \tilde{r}_2 * \tilde{r}_1$ is $(AM)^{O(1)}$-rational.

Otherwise, set

$$\tilde{p}_3 := \tilde{e}_2^{-1} * \tilde{p}_2 * \tilde{r}_2^{-1} = \tilde{p}_2 - \tilde{e}_2 - \tilde{r}_2 - \frac{1}{2}([\tilde{e}_2, \tilde{p}_2] + [\tilde{e}_2, \tilde{r}_2] + [\tilde{p}_2, \tilde{r}_2]) - \dots.$$

Note that $\tilde{p}_3 = p_2 \mod \mathfrak{g}_3$ and furthermore that $(\tilde{p}_3)_k = (\tilde{p}_2)_k = (p_1)_k$ for all $k \leq j$. Apply Corollary 4.5.5 to the degree $> j$ terms of $\tilde{p}_3$ to obtain a linearly irrational p_3 and proceed as above.

If we begin with $b \geq \prod_{i=2}^{s} c_i$, then the process terminates with the polynomial sequence p_s which is (A, N)-linearly irrational in $S_\bullet^{(s)}$, a sequence of complexity $A^{O(1)}$ subspaces of $\mathfrak{g}$ with $S_k^{(s)} = S_k$ for all $k < j$ and $S_j^{(s)} = S_j^{(1)} < S_j$.

Now we proceed to the statement about filtration irrationality. Let c be a large constant whose value we will decide upon later and let $b_1 := c$. Suppose that p_i is (B^{b_i}, N)-filtration irrational in $\mathfrak{g}$ mod $\mathfrak{g}_{i+1}$ and recall our (implicit) definitions

$$p_{i+1} := \tilde{p}_{i+1} - \tilde{e}_{i+1} - \tilde{r}_{i+1} := \tilde{e}_i^{-1} * (\tilde{e}_i + p_i + \tilde{r}_i) * \tilde{r}_i^{-1} - \tilde{e}_{i+1} - \tilde{r}_{i+1}, \tag{4.8}$$

where $\tilde{e}_i, \tilde{r}_i \in \mathfrak{g}_i$ and $\tilde{e}_{i+1}, \tilde{r}_{i+1} \in \mathfrak{g}_{i+1}$, and

$$e_i := \prod_{j=1}^{i} \tilde{e}_j, \qquad r_i := \prod_{j=1}^{i} \tilde{r}_j. \tag{4.9}$$

It follows from (4.8) that $p_{i+1} = p_i$ modulo $\mathfrak{g}_{i+1}$. For $j = 1, \dots, i$ we have that $\mathfrak{h}_j \geq \mathfrak{g}_{i+1}$ and so the jth coefficient of p_{i+1} inherits $(B^{b_i}, B^{b_i}/N^j)$-filtration irrationality from that of p_i. To check the filtration irrationality of p_{i+1} in $\mathfrak{g}$ mod $\mathfrak{g}_{i+2}$, it remains to check the filtration irrationality of the $(i+1)$th coefficient of p_{i+1}. To this end we

analyse

$$\begin{aligned}(p_{i+1})_{i+1} &= (e_i^{-1} * p * r_i^{-1} - \tilde{e}_{i+1} - \tilde{r}_{i+1})_{i+1} \\ &= (e_i^{-1} * p * r_i^{-1})_{i+1} - (\tilde{e}_{i+1})_{i+1} - (\tilde{r}_{i+1})_{i+1} \\ &= (p)_{i+1} - (e_i)_{i+1} - (r_i)_{i+1} - (\tilde{e}_{i+1})_{i+1} - (\tilde{r}_{i+1})_{i+1} \mod \mathfrak{h}_{i+1},\end{aligned}$$

where we have used Lemma 4.5.8 in the second line. It follows from definitions and Lemma 4.8.15 that both $(e_i)_{i+1}$ and $(\tilde{e}_{i+1})_{i+1}$ are $\frac{(AM)^{O(1)}}{N^{i+1}}$-small so that $(e_i)_{i+1} + (\tilde{e}_{i+1})_{i+1}$ is $\frac{(AM)^{O(1)}}{N^{i+1}}$-small by the triangle inequality. Also, both $(r_i)_{i+1}$ and $(\tilde{r}_{i+1})_{i+1}$ are $(AM)^{O(1)}$-rational so $(r_i)_{i+1} + (\tilde{r}_{i+1})_{i+1}$ is $(AM)^{O(1)}$-rational. We have by supposition that $(p)_{i+1}$ is $(B^c, B^c/N^{i+1})$-filtration irrational, and so is also trivially $(B^{c-k}, B^c/N^{i+1})$-filtration irrational for any constant $k > 0$. Invoking Lemma 4.5.9 (and to do so recalling that $B > AM > \dim \mathfrak{g}$), we may conclude that $(p_{i+1})_{i+1}$ is $(\frac{B^{c-k}}{(AM)^{O(1)}}, \frac{B^c/N^{i+1} - (\dim \mathfrak{g})B^{c-k}(AM)^{O(1)}/N^{i+1})}{(AM)^{O(1)}})$-filtration irrational in $\mathfrak{g}_{i+1}$, which in turn is also $(B^{c-k-O(1)}, \frac{B^{c-O(1)} - B^{c-k+O(1)}}{N^{i+1}})$-filtration irrational since $B > AM > \dim \mathfrak{g}$. Setting $k = O(1)$ large enough depending on the other $O(1)$ terms, we may arrange for this coefficient to be $(B^{c-O(1)}, \frac{B^{c-O(1)}}{N^{i+1}})$-irrational. Therefore, p_{i+1} is $(B^{b_{i+1}}, N)$-filtration irrational where $b_{i+1} := \min(b_i, c - O(1))$ and $c = O(1)$ has been chosen sufficiently large. By induction, if $c = O(1)$ is chosen sufficiently large, p_s will indeed be (B, N)-filtration irrational. This completes the proof. □

The previous result says that if we start with a sequence that is sufficiently filtration irrational, then we can factorise to a sequence which is both filtration irrational and linearly irrational.

Definition 4.5.11 (Strong irrationality)**.** If a polynomial sequence p is both (A, N)-filtration irrational in $\mathfrak{g}_\bullet$ and (A, N)-linearly irrational in some sequence of rational subspaces $S_\bullet$ then we will say that it is (A, N)-*strongly irrational* in $(\mathfrak{g}_\bullet, S_\bullet)$.

4.6 The strongly irrational arithmetic regularity lemma

In this section we prove our strongly irrational arithmetic regularity lemma Theorem 4.1.4. The main ingredients are Green and Tao's non-irrational regularity lemma and the factorisation theory we developed in the previous section.

Theorem 4.6.1 ([35, Proposition 2.7], Non-irrational arithmetic regularity lemma)**.** *Let $N > 1$, let $f : [N] \to [0,1]$, let $s \geq 1$, let $\varepsilon > 0$, and let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a growth function. Then there exists a quantity $M = O_{s,\varepsilon,\mathcal{F}}(1)$ and a decomposition*

$$f = f_{\mathrm{nil}} + f_{\mathrm{sml}} + f_{\mathrm{unf}}$$

of f into functions $f_{\mathrm{nil}}, f_{\mathrm{unf}} : [N] \to [-1,1]$ such that

1. *(f_{nil} structured) f_{nil} is a degree $\leq s$ polynomial nilsequence of complexity $\leq M$,*
2. *(f_{sml} small) $||f_{\mathrm{sml}}||_{L^2[N]} \leq \varepsilon$,*
3. *(f_{unf} very uniform) $||f_{\mathrm{unf}}||_{U^{s+1}[N]} \leq 1/\mathcal{F}(M)$,*
4. *(Nonnegativity) f_{nil} and $f_{\mathrm{nil}} + f_{\mathrm{sml}}$ take values in $[0,1]$.*

In fact, Theorem 4.1.4 produces a (virtual) nilsequence which is both linearly irrational and filtration irrational. Although the latter is not strictly necessary to determine distribution and indeed is omitted in Section 4.2, it is useful in applications (where otherwise one might have to replicate a similar factorisation on a case-by-case basis). Throughout this section we assume that d, s, D are of size $O(1)$.

The first step is to bootstrap our factorisation theory from the previous section to obtain a factorisation where p' is (much) more strongly irrational (by an arbitrary growth function, say) than e is small and r is rational. To do so, we first factorise for (very) amplified irrationality using [35, Lemma 2.10] and then repeatedly factorise for linear irrationality using Proposition 4.5.10 and a reduction argument on the lexicographical ordering of the sequence $\dim(S_\bullet) := (\dim S_1, \dots, \dim S_s)$.

The following is what is needed from [35].

Proposition 4.6.2 (Factorising for amplified filtration irrationality, [35, Lemma 2.10])**.** *Let $M_0 > 1$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra of dimension at most M_0 with filtration $\mathfrak{g}_\bullet$ of length at most M_0. Let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$ with structure constants of height at most M_0. Let p be a polynomial sequence adapted to $\mathfrak{g}_\bullet$ with $p(0) = 0$. For any growth function $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$, there exists $M_0 \leq M \leq O_{M_0,\mathcal{F}}(1)$ and a factorisation $p = e * p' * r$, where e is $(O_M(1), N)$-small, r is $O_M(1)$-rational, and p' is $(\mathcal{F}(M), N)$-filtration irrational in a subfiltration $\mathfrak{g}'_\bullet \leq \mathfrak{g}_\bullet$ of complexity $O_M(1)$.[7] This filtration irrationality is measured with respect to $\mathcal{X}'$, a Mal'cev basis[8] for $\mathfrak{g}'_\bullet$ which may be written as a linear combination of $\mathcal{X}$ with rational coefficients of height $O_M(1)$.*

[7] That is, for all i, $\mathfrak{g}'_i$ is a subspace of $\mathfrak{g}_i$ of complexity $O_M(1)$.

[8] See [38, Definition 2.1].

The result [35, Lemma 2.10] is phrased in slightly different language and there is some checking to be done to show that their statement implies ours. We do this in Appendix 4.8.4.

The following proposition is the "strongly irrational" analogue of the previous proposition.

Proposition 4.6.3 (Factorising for amplified strong irrationality)**.** *Let $M_0 > 1$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra of dimension at most M_0 with filtration $\mathfrak{g}_\bullet$ of length at most M_0. Let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$ with structure constants of height at most M_0. Let p be a polynomial sequence adapted to $\mathfrak{g}_\bullet$ with $p(0) = 0$. For any growth function $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$, there exists $M_0 \leq M \leq O_{M_0,\mathcal{F}}(1)$ and a factorisation $p = e * \tilde{p} * r$, e is $(O_M(1), N)$-small, where r is $O_M(1)$-rational, and $\tilde{p}$ is $(\mathcal{F}(M), N)$-strongly irrational in some $(\mathfrak{g}'_\bullet, S_\bullet)$, where $\mathfrak{g}'_\bullet$ is a subfiltration of $\mathfrak{g}_\bullet$ of complexity $O_M(1)$ in $\mathfrak{g}$ and $S_\bullet$ is a sequence of subspaces in $\mathfrak{g}'_\bullet$ of complexity $O_M(1)$. The filtration irrationality is measured with respect to $\mathcal{X}'$, a Mal'cev basis for $\mathfrak{g}'_\bullet$ where each element may be written as a linear combination of $\mathcal{X}$ with rational coefficients of height at most $O_M(1)$.*

Proof. Let $\mathcal{F}'$ be a growth function depending on $\mathcal{F}$ and M_0 to be determined later. For the meantime, insist that $\mathcal{F}' \geq (\mathcal{F} \circ \mathcal{F})^c$ at all inputs, where c is the constant from the statement of Theorem 4.5.10. Invoke Proposition 4.6.2 to factorise $p = e * p' * r$ whereupon there is $M'_0 = O_{M_0,\mathcal{F}'}(1) = O_{M_0,\mathcal{F}}(1)$ such that p' is $(\mathcal{F}'(M'_0), N)$-filtration irrational in some $(\mathfrak{g}'_\bullet, \mathcal{X}')$ of complexity $O_{M'_0}(1)$, e is $O_{M'_0}(1)$-small and r is $O_{M'_0}(1)$-rational. Note also that $\mathfrak{g}'$ has rational structure constants of height $O_{M'_0}(1)$ with respect to $\mathcal{X}'$ since the elements of $\mathcal{X}'$ are rational linear combinations of the elements of $\mathcal{X}$ with coefficients of height $O_{M'_0}(1)$.

If p' is $(\mathcal{F}(M'_0), N)$-linearly irrational in $S_\bullet = \mathfrak{g}'_\bullet$ then we are done upon setting $M := M'_0$. Otherwise, set $M_1 := \mathcal{F}(M'_0) = O_{M_0,\mathcal{F}}(1)$ and invoke Theorem 4.5.10 to obtain a factorisation $p' = e_1 * p_1 * r_1$ where e_1 is $(O_{M_1}(1), N)$-small, p_1 is (M_1, N)-linearly irrational in some sequence of subspaces $S^{(1)}_\bullet \leq \mathfrak{g}'_\bullet$ of complexity $O_{M_1}(1)$, and r_1 is $O_{M_1}(1)$-rational. Note that p_1 is (M_1, N)-linearly irrational in $S^{(1)}_\bullet$ but is not (M_1, N)-linearly irrational in $\mathfrak{g}'_\bullet$, and so we may conclude from the first bullet point of Theorem 4.5.10 that $\dim(S_\bullet)$ (that is, the vector of dimensions) is lexicographically strictly less than $\dim(\mathfrak{g}'_\bullet)$. Recall also that $\mathcal{F}'(M'_0) \geq \mathcal{F}(\mathcal{F}(M'_0))^c = \mathcal{F}(M_1)^c$ and so the $(\mathcal{F}'(M'_0), N)$-filtration irrationality of p' allows us to conclude as per the second bullet point of Theorem 4.5.10 that p_1 is $(\mathcal{F}(M_1), N)$-filtration irrational in $\mathfrak{g}'_\bullet$.

If in fact p_1 is $(\mathcal{F}(M_1), N)$-linearly irrational in $S^{(1)}_\bullet$ then observe that e_1 and r_1 are $(O_{M_0,\mathcal{F}}(1), N)$-small and $O_{M_0,\mathcal{F}}(1)$-rational respectively with respect to $\mathcal{X}$ and so we

are done with $M = M_1$. Otherwise, set $M_2 := \mathcal{F}(M_1)$, insist that $\mathcal{F}' \geq (\mathcal{F} \circ \mathcal{F} \circ \mathcal{F})^c$ at all inputs, and proceed as before to obtain some p_2 which is (M_2, N)-linearly irrational on a sequence of subspaces $S_\bullet^{(2)}$ with $\dim(S_\bullet^{(2)})$ lexicographically strictly less than $\dim(S_\bullet^{(1)})$. Now we have $p = e_1 * e_2 * p_2 * r_2 * r_1$, where one may use Lemma 4.8.15 to see that $e_1 * e_2$ and $r_2 * r_1$ are are $(O_{M_0,\mathcal{F}}(1), N)$-small and $O_{M_0,\mathcal{F}}(1)$-rational respectively with respect to $\mathcal{X}$. If in fact p_2 is $(\mathcal{F}(M_2), N)$-linearly irrational then we are done; if not, proceed as above. Since $\mathfrak{g}'$ has dimension at most M_0 and filtration of length at most M_0, the vector $\dim(S_\bullet^{(i)})$ can lexicographically decrease at most $O_{M_0}(1)$ times. Thus the process must terminate at some p_i which is $(\mathcal{F}(M_i), N)$-strongly irrational in some $(\mathfrak{g}', S_\bullet^{(i)})$ of complexity $O_{M_i}(1)$, with e_i which is $(O_{M_i}(1), N)$-small, and r_i which is $O_{M_i}(1)$-rational. This proves the claim. □

For convenience we will deal in this section only with (virtual) nilsequences for which the rational basis 'passes through' the filtration $\mathfrak{g}_\bullet$. Since the non-irrational arithmetic regularity lemma Theorem 4.6.1 produces nilsequences for which this restriction holds, we do not practically lose anything by making this simplifying assumption.

Definition 4.6.4. We will say that a basis $\mathcal{X} := \{X_1, \ldots, X_n\}$ for a vector space S *passes through* a sequences of rational subspaces $S_1 \geq S_2 \geq \ldots$ if $\{X_{n-\dim S_i+1}, \ldots, X_n\}$ is a basis for S_i for all i. We may also say that each S_i passes through $\mathcal{X}$.

Definition 4.6.5 (Virtual nilsequence)**.** A *virtual nilsequence* of complexity M at scale N is a function $\phi : [N] \to \mathbb{C}$ of the form $\phi(n) = f(p(n) * \Lambda, n \pmod q, n/N)$ where $q \leq M$ is a positive integer, p is a polynomial sequence with respect a filtered nilpotent Lie algebra $\mathfrak{g}_\bullet$ of dimension $d \leq M$ and step $s \leq M$, Λ is an additive lattice in $\mathfrak{g}_\bullet$ with $\mathbb{Z}$-basis $\mathcal{X}$ which passes through $\mathfrak{g}_\bullet$ and for which $\mathfrak{g}$ has rational structure constants of height at most M, and f is Lipschitz with Lipschitz constant at most M.[9]

Let $S_\bullet$ be a sequence of subspaces of $\mathfrak{g}_\bullet$ of complexity at most M. We will furthermore say that ϕ is (A, N)-filtration (resp. linearly/strongly) irrational in $\mathfrak{g}_\bullet$ (resp. $(\mathfrak{g}_\bullet, S_\bullet)$) if p is (A, N)-filtration (resp. linearly/strongly) irrational in $\mathfrak{g}_\bullet$ (resp. $(\mathfrak{g}_\bullet, S_\bullet)$) with respect to $\mathcal{X}$.

[9] As in [35], the precise choice of underlying metric is not very important. It may be taken to be the direct sum of the metrics on each of the factors, where the metric on $\mathfrak{g}$ is induced by the ℓ_∞ norm with respect to $\mathcal{X}$ and the metric on $\mathbb{Z}/q\mathbb{Z}$ is obtained by embedding this group into $\mathbb{R}/\mathbb{Z}$.

Remark 4.6.6. In the above, Λ may not be a multiplicative subgroup (i.e. with respect to the $*$ operation), in which case this definition is equivalent to asking that f is automorphic with respect to the multiplicative closure of Λ in $\mathfrak{g}$. We see in Appendix 4.8.2 that this multiplicative closure is indeed a multiplicative lattice which is not too much larger than Λ.

Our conclusion to the argument follows a similar path to that of [35, Theorem 1.2] and may essentially be viewed as a Lie algebra analogue of what is done there. We note/caution that in the following proposition, the notion of complexity of a nilsequence used is that of Green and Tao, whereas the notion of complexity of a virtual nilsequence is that of our Definition 4.6.5.

Proposition 4.6.7. *Let $s > 0$, $M_0 > 1$, let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a growth function, and let $\phi : \mathbb{Z} \to [0,1]$ be a degree $\leq s$ nilsequence of complexity $\leq M_0$. Then there exists $M = O_{s,M_0,\mathcal{F}}(1)$ such that ϕ (when restricted to $[N]$) is also a $(\mathcal{F}(M), N)$-strongly irrational degree $\leq s$ virtual nilsequence of complexity $\leq M$.*

Proof. As usual we let all constants depend on s. Write $\phi(n) = f(p(n)) = f(p(n) * \log \Gamma)$ where p is a polynomial sequence on some filtered nilpotent Lie algebra $\mathfrak{g}_\bullet$ and where f is automorphic with respect to some lattice $\Gamma \leq G$. Let $\mathcal{X}_{\mathrm{Mal}}$ be the Mal'cev basis which exhibits ϕ as having complexity $\leq M_0$. By [38, Lemma A.8], there is an integer $q_0 = M_0^{O(1)}$ such that $\mathcal{X} := q_0 \mathcal{X}_{\mathrm{Mal}}$ yields $\langle \mathcal{X} \rangle \subset \log \Gamma$. Since $\mathcal{X}_{\mathrm{Mal}}$ has structure constants of height at most M_0, we have that $\mathcal{X}$ has structure constants of height $M_0^{O(1)}$. Write $p = p(0) * (p(0)^{-1} * p)$, invoke Proposition 4.6.3 on the bracketed polynomial (with M_0 replaced by $M_0^{O(1)}$ and for an unspecified growth function $\mathcal{F}'$) and inherit the notation from the statement of that proposition so $\phi(n) = f(p(0) * e(n) * p'(n) * r(n) * \log \Gamma)$. From Lemma 4.8.16 we have that $r(n) * \log \Gamma$ is q-periodic, where $q = O_M(1)$ and we recall that $M = O_{M_0,\mathcal{F}'}(1)$ from Proposition 4.6.3. Now define $\tilde{f} : \mathfrak{g} \times \mathbb{Z}/q\mathbb{Z} \times \mathbb{R} \to [0,1]$ by

$$\tilde{f}(x,s,y) := f(p(0) * e(Ny) * x * r(s) * \log \Gamma),$$

whenever $y \in \frac{1}{N}\mathbb{Z}$ and by Lipschitz extension to $\mathbb{R}$. Note that the q-periodicity of the orbit $r(n) * \log \Gamma$ yields that $\tilde{f}$ is indeed a well-defined function on its second argument.

Next we claim that there exists a positive integer $l = O_M(1)$ such that $\tilde{f}(x * (l\langle \mathcal{X} \rangle), s, y) = \tilde{f}(x,s,y)$, so that $\tilde{f}$ is automorphic with respect to the additive lattice $l\langle \mathcal{X} \rangle$ in its first argument. Since $\langle \mathcal{X} \rangle \subset \log \Gamma$, it suffices to find l such that for all $n \in \mathbb{Z}$ we have $(l\langle \mathcal{X} \rangle) * r(n) \subset r(n) * \langle \mathcal{X} \rangle$. But this follows easily from the fact that

$r(n)$ is $O_M(1)$-rational with respect to $\mathcal{X}$, the Baker-Campbell-Hausdorff formula and our bound on the heights of the rational structure constants of $\mathfrak{g}$ with respect to the basis $\mathcal{X}$. Let $\Lambda = l\langle\mathcal{X}\rangle = \langle l\mathcal{X}\rangle$. Since p' is $(\mathcal{F}'(M), N)$-strongly irrational in $(\mathfrak{g}_\bullet, S_\bullet)$ with respect to $\mathcal{X}$, it is $(\gg_M \mathcal{F}'(M), N)$-strongly irrational in $(\mathfrak{g}_\bullet, S_\bullet)$ with respect to $l\mathcal{X}$.

Using the Lipschitz property of f, it is also not difficult to show that $\tilde{f}$ is $O_M(1)$-Lipschitz, but we omit the details. The result then follows after replacing M with a suitable quantity $O_M(1)$ and letting $\mathcal{F}$ be sufficiently rapidly growing depending on $\mathcal{F}'$. □

We are finally ready to recover the main result of this section.

Theorem 4.6.8 (Strongly irrational arithmetic regularity lemma, Theorem 4.1.4)**.** *Let $f : [N] \to [0,1]$, let $s \geq 1$, let $\varepsilon > 0$, and let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a growth function. Then there exists a quantity $M = O_{s,\varepsilon,\mathcal{F}}(1)$ and a decomposition*

$$f = f_{\text{nil}} + f_{\text{sml}} + f_{\text{unf}}$$

of f into functions $f_{\text{nil}}, f_{\text{unf}} : [N] \to [-1,1]$ such that

1. *(f_{nil} structured) f_{nil} is a degree $\leq s$, $(\mathcal{F}(M), N)$-strongly irrational virtual nilsequence of complexity $\leq M$,*
2. *(f_{sml} small) $||f_{\text{sml}}||_{L^2[N]} \leq \varepsilon$,*
3. *(f_{unf} very uniform) $||f_{\text{unf}}||_{U^{s+1}[N]} \leq 1/\mathcal{F}(M)$,*
4. *(Nonnegativity) f_{nil} and $f_{\text{nil}} + f_{\text{sml}}$ take values in $[0,1]$.*

Proof. This is an easy corollary of Theorem 4.6.1 and Proposition 4.6.7. □

Remark 4.6.9. Technically a strongly irrational virtual nilsequence consists of data $(p, \mathfrak{g}, \mathcal{X}, \mathfrak{g}_\bullet, S_\bullet)$. However, recalling Lemma 4.4.11, if $\mathcal{F}$ is sufficiently rapidly growing, we may drop the data $S_\bullet$ since it may be uniquely recovered from the polynomial sequence p. Furthermore, by Lemma 4.7.1 and Lemma 4.3.5, one may also recover the filtration $\mathfrak{g}_\bullet$ from p. Thus, for sufficiently strongly irrational virtual nilsequences, the data $(p, \mathfrak{g}, \mathcal{X})$ suffices.

4.7 Applications

4.7.1 Recovering the flag counting lemma

This subsection is not an application as much as it is a sanity check (Corollary 4.7.2) that in the flag case, Green and Tao's arithmetic regularity lemma and counting lemma may be recovered by the results in this chapter. Its purpose is also to prove Lemma 4.7.1 which is useful in other applications.

Recall Definition 4.2.10, that of the subalgebras $\mathfrak{h}_i$ associated to the filtered Lie algebra $\mathfrak{g}_\bullet$. The following is a quantitative version of Lemma 4.3.4.

Lemma 4.7.1. *Suppose that p is (A, N)-filtration irrational in $\mathfrak{g}_\bullet$ where the rational basis $\mathcal{X}$ passes through $\mathfrak{g}_\bullet$. Then, for each $i = 1, \dots, s$, for any subspace $T_i \leq \mathfrak{g}_i$ which contains a_i and has complexity at most A, we have $T_i + \mathfrak{h}_i = \mathfrak{g}_i$.*

Proof. From Definition 4.2.10 and the filtration on $\mathfrak{g}_\bullet$ we have $\mathfrak{h}_i \leq \mathfrak{g}_i$ and so $T_i + \mathfrak{h}_i \leq \mathfrak{g}_i$. The slightly less trivial containment to show is $T_i + \mathfrak{h}_i \supseteq \mathfrak{g}_i$. Let $Q = Q(A)$ be the $\mathbb{R}$-span of all complexity $\leq A$ linear maps on $\mathfrak{g}_i$.[10] Since a_i is $(A, A/N^i)$-filtration irrational, we in particular have that $l(a_i) \neq 0$ for all nontrivial $l \in Q$ which vanish on $\mathfrak{h}_i$. Thus in $\mathfrak{g}_i^*$ we have $a_i^\perp \cap Q \cap \mathfrak{h}_i^\perp = \{0\}$ so $(a_i^\perp \cap Q)^\perp + \mathfrak{h}_i = \mathfrak{g}_i$. But $\mathfrak{g}_{i\mathcal{X}}^*(A) \subset Q$ where $\mathfrak{g}_{i\mathcal{X}}^*(A)$ is the set of integer linear maps on $\mathfrak{g}_i$ of complexity at most A, and so $(a_i^\perp \cap Q)^\perp \subset (a_i^\perp \cap \mathfrak{g}_{i\mathcal{X}}^*(A))^\perp$. However, $(a_i^\perp \cap \mathfrak{g}_{i\mathcal{X}}^*(A))^\perp$ is, by definition, the smallest rational subspace of $\mathfrak{g}_i$ which contains a_i and has complexity at most A. Therefore, $T_i \supset (a_i^\perp \cap \mathfrak{g}_{i\mathcal{X}}^*(A))^\perp$ and ultimately, $T_i + \mathfrak{h}_i \supset \mathfrak{g}_i$. □

Corollary 4.7.2 (Theorem 4.1.5)**.** *Let $A, M \geq 2$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a flag system of linear forms, each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra of dimension d and step s. Let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$ which passes through $\mathfrak{g}_\bullet$ and with respect to which $\mathfrak{g}$ has rational structure constants of height at most M. Let $S_\bullet \leq \mathfrak{g}_\bullet$ be a sequence of subspaces of complexity at most M. If p is (A, N)-strongly irrational in $(\mathfrak{g}_\bullet, S_\bullet)$ then there are constants $0 < c_1, c_2 = O_{d,s,D,t,\Psi}(1)$ such that p^Ψ is $O(M^{c_1}/A^{c_2})$-equidistributed in $\log G^\Psi$, where G^Ψ is the Leibman group for $G_\bullet, \Psi$.*

Proof. This follows from our counting lemma Theorem 4.1.2, Lemma 4.7.1 and Proposition 4.3.6. □

[10] Since $\mathcal{X}$ passes through $\mathfrak{g}_\bullet$ the complexity of a map on $\mathfrak{g}$ is equal to that of its restriction on any $\mathfrak{g}_i$.

4.7.2 Another proof of the Gowers-Wolf conjecture

In their original 'arithmetic regularity lemma and counting lemma' paper, Green and Tao proved the Gowers-Wolf conjecture [28, Conjecture 2.5] for flag systems of linear forms.

Theorem 4.7.3 ([35, Theorem 1.13])**.** *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a collection of linear forms each mapping $\mathbb{Z}^D$ to $\mathbb{Z}$ which satisfy the flag condition, and let $s \geq 1$ be an integer such that the polynomials $\psi_1^{s+1}, \dots, \psi_t^{s+1}$ are linearly independent. For $i = 1, \dots t$, let $f_i : [-N, N] \to \mathbb{C}$ be functions bounded in magnitude by 1 (and defined to be zero outside of $[-N, N]$). For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $\min_i ||f_i||_{U^{s+1}[-N,N]} \leq \delta$, then*

$$\left| \mathbb{E}_{\boldsymbol{x} \in [-N,N]^D} \prod_{i=1}^{t} f_i(\psi_i(\boldsymbol{x})) \right| \leq \varepsilon.$$

The restriction to flag patterns in the above theorem is a consequence of the use of the arithmetic regularity lemma and complementary counting lemma from [35], which itself is restricted to flag patterns. The author ([2] and Chapter 3 above) resolved the full result by using a linear algebraic trick to reduce to the flag case whereupon the flag arithmetic regularity lemma and counting lemma from [35] sufficed.[11]

Theorem 4.7.4 ([2, Theorem 1.1])**.** *Theorem 4.7.3 holds without the restriction that Ψ satisfies the flag condition.*

Our arithmetic regularity lemma Theorem 4.1.4 and counting lemma Theorem 4.1.2 mean that original strategy from [35] may be applied directly to resolve the Gowers-Wolf conjecture in full generality. We will not reiterate the proof here since it may be fairly cleanly decomposed as a union of what is is already done in [35, Theorem 1.13] and the upcoming Proposition. We do note that along the way one will need to obtain a version of Theorem 4.1.2 which allow for averages over shifted sublattices of $\mathbb{Z}^D$. This is directly analogous to how [35, Theorem 3.6] differs from [38, Theorem 8.6]. Such a pursuit is routine but technical and tedious; we do not pursue it here to avoid adding further length to this chapter.

Recall the definition of the sequence of subspaces $(W_i)_{i=1}^s$ given a sequence of subspaces $(S_i)_{i=1}^s$: let $W_1 = S_1$ and iteratively define $W_i = \text{span}\{S_j, [W_j, W_{i-j}]$ for $j = 1, \dots, i-1\}$ for $i = 2, \dots, s$. Recall also the definition of the subalgebras $\mathfrak{h}_i$ given a filtration $\mathfrak{g}_\bullet$: Definition 4.2.10.

[11] The reader might also be interested to consult [57] where, even more recently, Manners resolved the problem over finite fields with polynomial bounds [57, Theorem 1.1.5].

Proposition 4.7.5. *Suppose $A > M \geq 1$. Suppose that p is (A, N)-strongly irrational in $(\mathfrak{g}_\bullet, S_\bullet)$ where $\mathfrak{g}$ has rational structure constants of height at most M with respect to a $\mathcal{X}$ which passes through $\mathfrak{g}_\bullet$, that the elements of $S_\bullet$ have complexity at most M with respect to $\mathcal{X}$, and that Ψ has the property that $V^j = \mathbb{R}^t$. Then there are constants $0 < c_1, c_2 = O_{d,s,D,t,\Psi}(1)$ such that p^Ψ is $O(M^{c_1}/A^{c_2})$-equidistributed on $\mathfrak{g}^\Psi(S_\bullet) \leq \mathfrak{g}^t \cong \mathfrak{g} \otimes \mathbb{R}^t$, where $\mathfrak{g}^\Psi(S_\bullet)$ contains $\mathfrak{g}_j \otimes \mathbb{R}^t$.*

Proof. The counting lemma Theorem 4.1.2 and the (A, N)-linear irrationality of p yields that p^Ψ is $O(M^{c_1}/A^{c_2})$-equidistributed on $\mathfrak{g}^\Psi(S_\bullet)$; we need to show that $\mathfrak{g}^\Psi(S_\bullet)$ contains $\mathfrak{g}_j \otimes \mathbb{R}^t$. Firstly, $V^j = \mathbb{R}^t$ implies that $V^i = \mathbb{R}^t$ for all $i \geq j$ and so by Proposition 4.3.1, $\mathfrak{g}^\Psi(S_\bullet)$ contains $\sum_{i=j}^s W_j \otimes \mathbb{R}^t$. Also, the (A, N)-filtration irrationality of p ensures by Lemma 4.7.1 that $S_i + \mathfrak{h}_i = \mathfrak{g}_i$ for all i, and so $W_i + \mathfrak{g}_{i+1} = \mathfrak{g}_i$ for all i by Lemma 4.3.5. Thus $\sum_{i=j}^s W_i = \mathfrak{g}_j$ and so $\mathfrak{g}^\Psi(S_\bullet)$ does indeed contain $\mathfrak{g}_j \otimes \mathbb{R}^t$. □

Corollary 4.7.6. *Theorem 4.7.3 holds without the restriction that Ψ satisfies the flag condition.*

Proof. Apply the argument from the proof of [35, Theorem 1.13] with our strongly-irrational arithmetic regularity lemma Theorem 4.1.4 in place of [35, Theorem 1.2], and use Proposition 4.7.5 in place of the [35, Theorem 1.11]. □

4.8 Appendix A: Details on polynomial sequences in Lie groups and Lie algebras

In this appendix we may use terminology from [35, Appendix A] and [38, Appendix A] without further introduction. As we have throughout this chapter, we assume here that the dimension d and step s of $\mathfrak{g}$ are of size $O(1)$.

4.8.1 Smoothness norms

Green and Tao introduced the following 'smoothness' norms.

Definition 4.8.1 ([38, Definition 8.2])**.** Let $N_1, \dots, N_D \geq 1$ be integers and let $[\boldsymbol{N}]$ denote the set $[N_1] \times \cdots \times [N_D]$. Let $f : \mathbb{Z}^D \to \mathbb{R}/\mathbb{Z}$ be a polynomial map with Taylor expansion $f(\boldsymbol{n}) = \sum_{\boldsymbol{j}} \alpha_{\boldsymbol{j}} \binom{\boldsymbol{n}}{\boldsymbol{j}}$. Then define

$$||f||_{C^\infty[\boldsymbol{N}]} := \sup_{\boldsymbol{j} \neq 0} \prod_{i=1}^D N_i^{j_i} ||\alpha_{\boldsymbol{j}}||_{\mathbb{R}/\mathbb{Z}}.$$

The main theorems in [35] and [38] are written with respect to these smoothness norms. Note that one may naturally extend this definition to polynomials $f : \mathbb{R}^D \to \mathbb{R}$. It will be convenient for us to work with a slightly different definition. Recall that we use $\mathcal{M}$ to denote the set of 'monomial maps' which take $\boldsymbol{x} = (x_1, \dots, x_D)$ to a monomial in the variables $\{x_1, \dots, x_D\}$. The monomial map which takes $\boldsymbol{x}$ to the constant 1 will be denoted by 1.

Definition 4.8.2 (Smoothness norms with respect to monomial basis)**.** Let $f : \mathbb{R}^D \to \mathbb{R}$ have monomial expansion $f(\boldsymbol{x}) = \sum_{m \in \mathcal{M}} \beta_m m(\boldsymbol{x})$. Then define

$$||f||_{C^\infty_{\mathcal{M}}[N]^D} := \sup_{1 \neq m \in \mathcal{M}} N^{\deg m} ||\beta_m||_{\mathbb{R}/\mathbb{Z}}.$$

Lemma 4.8.3. *Let $f : \mathbb{R}^D \to \mathbb{R}$ be a multivariate polynomial of degree s. Then $||s!f||_{C^\infty_{\mathcal{M}}[N]^D} \ll_{s,D} ||f||_{C^\infty[N]^D}$.*

Proof. Let $\varepsilon = ||f||_{C^\infty[N]^D}$. By definition, we may write $f = f_n + f_\varepsilon$, where, when developed with respect to the multinomial basis, the coefficients of f_n are integers and the coefficients of the degree d terms of f_ε are of magnitude at most ε/N^d. Writing f_n with respect to the monomial basis yields coefficients in $\frac{1}{s!}\mathbb{Z}$ and so $s!f_n$ has integer valued coefficients (with respect to the monomial basis). Writing $s!f_\varepsilon$ with respect to the monomial basis yields coefficients of degree d terms with magnitude at most $O_{s,D}(\varepsilon/N^d)$. □

4.8.2 Equidistribution in the Lie algebra

The goal of this subsection is to explain what is meant by, and to justify, Theorems 4.2.8 and 4.4.15. We deduce them from theorems in [53] and [38] respectively. We will first deal with the qualitative setting (i.e. deducing Theorem 4.2.8 from [53]).

When we speak of qualitative equidistribition on a nilmanifold, we mean the following.

Definition 4.8.4 (Qualitative nilmanifold equidistribution)**.** Let G/Γ be a nilmanifold. Recall that this comes with a unique normalised Haar measure, with respect to which we will integrate. A sequence $(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n} \in \mathbb{Z}^D}$ on G/Γ *equidistributes* if for all continuous $F : G/\Gamma \to \mathbb{C}$ we have

$$\lim_{N \to \infty} \mathbb{E}_{\boldsymbol{n} \in [N]^D} F(g(\boldsymbol{n})\Gamma) \to \int_{G/\Gamma} F.$$

Leibman's theorem says that a polynomial sequence equidistributes if and only if it does not lie in (a coset of) the kernel of a nontrivial horizontal character on G.

Theorem 4.8.5 (Leibman's criterion for qualitative equidistribution, [53])**.** *Let D be a positive integer, let G/Γ be a nilmanifold and let $g : \mathbb{Z}^D \to G$ be a polynomial sequence. Then exactly one of the following is true:*

1. *$(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n}\in\mathbb{Z}^D}$ equidistributes in G/Γ,*
2. *there exists a nontrivial horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ which maps Γ to $\mathbb{Z}$ and such that $\eta \circ g$ is constant.*

Given two lattices Γ_0 and Γ_1 in G, we say that they are *commensurable* if $\Gamma_0 \cap \Gamma_1$ is a lattice in G of finite index in both Γ_0 and Γ_1. It is not difficult to check directly from the above definition that if Γ_0 and Γ_1 are commensurable then a polynomial sequence g equidistributes on G/Γ_0 if and only if it equidistributes on G/Γ_1. Therefore, we may extend the definition of equidistribution on a particular nilmanifold G/Γ to consider simultaneously all lattices commensurable to Γ. This may be neatly understood in the Lie algebra $\mathfrak{g}$ by the following result.

Theorem 4.8.6 ([18, Theorem 5.1.12])**.** *Let Γ_0, Γ_1 be lattices in G. Then Γ_0 and Γ_1 determine the same rational structure on $\mathfrak{g}$ (i.e. $\operatorname{span}_{\mathbb{Q}}(\log \Gamma_0) = \operatorname{span}_{\mathbb{Q}}(\log \Gamma_1)$) if and only if Γ_0 and Γ_1 are commensurable.*

We remark also that given any rational structure $\mathfrak{g}_{\mathbb{Q}}$ on $\mathfrak{g}$, there is a lattice Γ in $G = \exp \mathfrak{g}$ (invoke Lie's third theorem) such that $\mathfrak{g}_{\mathbb{Q}} = \operatorname{span}_{\mathbb{Q}}(\log \Gamma)$ ([18, Theorem 5.1.8]). Finally, any polynomial sequence G yields (under the logarithm map) a polynomial sequence in $\mathfrak{g}$ and conversely. By the discussion above, we may define the property of equidistribution of a polynomial sequence in $\mathfrak{g}$ with respect to a rational structure $\mathfrak{g}_{\mathbb{Q}}$ as that of the corresponding polynomial sequence in G with respect to any lattice Γ such that $\operatorname{span}_{\mathbb{Q}}(\log \Gamma) = \mathfrak{g}_{\mathbb{Q}}$.

Definition 4.8.7 (Equidistribution with respect to a rational structure)**.** A polynomial sequence p in $\mathfrak{g}$ *equidistributes in* $(\mathfrak{g}, \mathfrak{g}_{\mathbb{Q}})$ if $\exp p$ equidistributes in G/Γ, where Γ is any lattice in G such that $\operatorname{span}_{\mathbb{Q}} \log \Gamma = \mathfrak{g}_{\mathbb{Q}}$.

By our discussion above, this definition is indeed well-defined. We are ready to prove the following. The proof is very straightforward, but we include it to reiterate what is meant by equidistribution in the Lie algebra. Recall that in the following (and indeed in all discussions of irrationality above) we have restricted to polynomial sequences with $p(0) = 0$.

Theorem 4.8.8 (Leibman's criterion for qualitative equidistribution in the Lie algebra, Theorem 4.2.8)**.** *Let D be a positive integer, let $\mathfrak{g}$ be a real, finite dimensional, nilpotent Lie algebra and let $\mathfrak{g}_{\mathbb{Q}}$ be a rational structure on $\mathfrak{g}$. Let $(p(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{N}^D}$ be a polynomial sequence on $\mathfrak{g}$. Then $(p(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D}$ equidistributes in $(\mathfrak{g},\mathfrak{g}_{\mathbb{Q}})$ if and only if p is additively irrational in $(\mathfrak{g},\mathfrak{g}_{\mathbb{Q}})$.*

Proof. Suppose p is not additively irrational in $(\mathfrak{g},\mathfrak{g}_{\mathbb{Q}})$, so Definition 4.2.6, there exists a nontrivial $\eta \in \mathrm{Hom}_{\mathbb{Q}}(\mathfrak{g},\mathbb{R})$ such that $((\eta\circ p)(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D} \subset \mathbb{Q}$. Let Γ be a lattice in $G = \exp\mathfrak{g}$ such that $\mathrm{span}_{\mathbb{Q}} \log\Gamma = \mathfrak{g}_{\mathbb{Q}}$ (cf. [18, Theorem 5.1.8] for its existence). Then $\eta(\log\Gamma) \subset \mathbb{Q}$. In fact, since $\log\Gamma$ is contained in some additive lattice itself contained in $\mathfrak{g}_{\mathbb{Q}}$ ([18, Theorem 5.4.2]), there is a positive integer M such that $\eta(\log\Gamma) \subset \frac{1}{M}\mathbb{Z}$. Thus there exists η', a large integer multiple of η, so that $\eta'(\log\Gamma)\subset\mathbb{Z}$ and $((\eta'\circ p)(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D}\subset\mathbb{Z}$. By standard Lie theory there is a unique Lie group homomorphism $\eta_0 : G\to\mathbb{R}/\mathbb{Z}$ such that $\eta' = d\eta_0$ and so $\eta_0\circ\exp = \exp\circ\eta'$ (where here the second $\exp:\mathbb{R}\to\mathbb{R}/\mathbb{Z}$ is just the quotient map). Writing this out we have

$$\bar{0} = ((\exp\circ\eta'\circ p)(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D} = ((\eta_0\circ\exp\circ p)(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D},$$

where $\bar{0}$ is the residue class of 0 in $\mathbb{R}/\mathbb{Z}$. Note η_0 is not trivial. Thus the polynomial sequence $\exp\circ p$ does not equidistribute in G/Γ by Theorem 4.8.5, and so p does not equidistribute in $(\mathfrak{g},\mathfrak{g}_{\mathbb{Q}})$ by definition.

In the other direction, suppose that $(p(\boldsymbol{n}))_{\boldsymbol{n}\in\mathbb{Z}^D}$ does not equidistribute in $(\mathfrak{g},\mathfrak{g}_{\mathbb{Q}})$, so $\exp\circ p$ does not equidistribute in G/Γ where Γ is a lattice such that $\mathrm{span}_{\mathbb{Q}}(\log\Gamma) = \mathfrak{g}_{\mathbb{Q}}$. By Theorem 4.8.5, there is a nontrivial horizontal character $\eta_0 : G\to\mathbb{R}/\mathbb{Z}$ which maps Γ to $\mathbb{Z}$ and such that $\eta_0\circ\exp\circ p$ is constant in $\mathbb{R}/\mathbb{Z}$. Standard Lie theory yields that the differential $d\eta_0$ is a nontrivial Lie algebra homomorphism $\mathfrak{g}\to\mathbb{R}$ and $\eta_0\circ\exp = \exp\circ d\eta_0$. Set $\eta = d\eta_0$. Thus $\exp\circ\eta\circ p$ is constant (in fact, zero) in $\mathbb{R}/\mathbb{Z}$, so $\eta\circ p\subset\mathbb{Z}$, and p is not additively irrational. □

Now we proceed to the matter of quantitative equidistribution. In the body of the chapter, we used the notation $\langle\mathcal{X}\rangle$ to refer to the additive group generated by a basis $\mathcal{X}$, that is, $\mathrm{span}_{\mathbb{Z}}\mathcal{X}$. In this appendix we will also have to deal with $\langle\mathcal{X}\rangle_*$, the $*$-multiplicative group generated by $\mathcal{X}$, and so we will generally use $\mathrm{span}_{\mathbb{Z}}\mathcal{X}$ to refer to the additive group in order to make the distinction more notationally clear.

In the following definition we refer to a Lipschitz function on G/Γ, which obviously presupposes the existence of a metric on this space. We refer the reader to [38] for details on this; they are not important for the purposes of this chapter.

Definition 4.8.9 (Quantitative nilmanifold equidistribution)**.** Let $0 < \delta < 1/2$, let G/Γ be a nilmanifold, let $N_1, \dots, N_D \geq 1$ and let $[\boldsymbol{N}]$ denote the set $[N_1] \times \cdots \times [N_D]$. We say that a polynomial sequence $(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n} \in \mathbb{Z}^D}$ is *δ-equidistributed* in G/Γ if for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$ we have

$$|\mathbb{E}_{\boldsymbol{n} \in [\boldsymbol{N}]} F(g(\boldsymbol{n})\Gamma) - \int_{G/\Gamma} F| \leq \delta ||F||_{\mathrm{Lip}}.$$

Green and Tao proved the following; [38] may be consulted for terminology or notation which has not otherwise been introduced in this chapter.

Theorem 4.8.10 ([39, Theorem p.3])**.** *Let $0 < \delta < 1/2$, let $N_1, \dots, N_D \geq 1$ and let $[\boldsymbol{N}]$ denote the set $[N_1] \times \cdots \times [N_D]$. Let G/Γ be a nilmanifold with dimension d, step s and Mal'cev basis $\mathcal{X}$ which is $1/\delta$-rational. Let $(g(\boldsymbol{n}))_{\boldsymbol{n} \in \mathbb{Z}^D}$ be a polynomial sequence in G. Then either*

1. *there is some $N_i \ll \delta^{-O_{d,s,D}(1)}$, or*
2. *$(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n} \in [\boldsymbol{N}]}$ is δ-equidistributed in G/Γ, or*
3. *there is a nontrivial horizontal character η on G/Γ with $0 < ||\eta|| \ll \delta^{-O_{d,s,D}(1)}$ such that $||\eta \circ g||_{C^\infty[\boldsymbol{N}]} \ll \delta^{-O_{d,s,D}(1)}$.*

Furthermore, when $N_1 = \cdots = N_D$, the first option above may be removed.

Our first task in the quantitative setting is to define equidistribution of a polynomial sequence in the Lie algebra with respect to a choice of (rational) basis.

Definition 4.8.11 (Definition 4.4.14, Quantitative equidistribution in the Lie algebra)**.** Let $\mathcal{X}$ be a rational basis in $\mathfrak{g}$. Then a polynomial sequence $(p(\boldsymbol{n}))_{\boldsymbol{n} \in [\boldsymbol{N}]}$ is δ-equidistributed in $(\mathfrak{g}, \mathcal{X})$ if $(\exp(p(\boldsymbol{n})))_{\boldsymbol{n} \in [\boldsymbol{N}]}$ is δ-equidistributed in G/Γ, where Γ is the smallest multiplicative lattice containing $\mathrm{span}_{\mathbb{Z}} \mathcal{X}$.

Let us check that this definition is sensible. Firstly, the intersection of two multiplicative lattices that contain $\mathcal{X}$ is clearly a discrete multiplicative group which spans $\mathfrak{g}$; this is enough to conclude [18, Corollary 5.4.5] that it is cocompact in G and so indeed itself a multiplicative lattice; thus we may indeed speak of the smallest lattice containing $\mathrm{span}_{\mathbb{Z}} \mathcal{X}$.

Next, in the context of constructing nilsequences, a function on $\mathfrak{g}$ is of course multiplicatively automorphic with respect to the set $\mathrm{span}_{\mathbb{Z}} \mathcal{X}$ if and only if it is multiplicatively automorphic with respect to $\langle \mathrm{span}_{\mathbb{Z}} \mathcal{X} \rangle_*$.

The following lemma offers some further validation for the above definition.

Lemma 4.8.12. *Let Λ be an additive lattice in $\mathfrak{g}$. Then the smallest $*$-multiplicative subgroup $\langle\Lambda\rangle_*$ containing Λ is a multiplicative lattice in $\mathfrak{g}$. Furthermore, if $\mathcal{X}$ is a $\mathbb{Z}$-basis for Λ which has rational structure constants of height at most M, then we have*

$$\Lambda \subset \langle\Lambda\rangle_* \subset \frac{1}{M^{O(1)}}\Lambda,$$

where implicit constants may depend on $\dim \mathfrak{g}$.

Proof. Firstly, it is not difficult to construct a basis $\tilde{\mathcal{X}} := \{\tilde{X}_1, \dots, \tilde{X}_d\}$ for $\mathfrak{g}$ consisting of rational linear combinations of $\mathcal{X}$ with coefficients of height $M^{O(1)}$ which satisfies the nesting property

$$[\mathfrak{g}, \tilde{X}_i] \subset \operatorname{span}_{\mathbb{R}}\{\tilde{X}_{i+1}, \dots, \tilde{X}_d\};$$

see [38, Proposition A.9] for details. Thus we may assume without loss of generality that $\mathcal{X}$ satisfies this nesting property; denote $\mathcal{X} := \{X_1, \dots, X_d\}$. We claim that there are integers $k_1, \dots, k_d$ each of size $M^{O(1)}$ such that upon setting $\mathcal{X}' := \{\frac{1}{k_1}X_1, \dots, \frac{1}{k_d}X_d\}$ we have that $\operatorname{span}_{\mathbb{Z}} \mathcal{X}'$ is a multiplicative lattice in $\mathfrak{g}$. Then

$$\Lambda \subset \langle\Lambda\rangle_* \subset \operatorname{span}_{\mathbb{Z}} \mathcal{X}' \subset \frac{1}{M^{O(1)}}\Lambda,$$

and so we are done.

It follows from Baker-Campbell-Hausdorff that multiplication in coordinates with respect to $\mathcal{X}$ is a polynomial map $* : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ of degree $O(1)$ and with rational coefficients of height $M^{O(1)}$. Let P_i be polynomials such that $a * b = (P_1(a,b), \dots, P_d(a,b))$, with respect to $\mathcal{X}$. Then by Baker-Campbell-Hausdorff and the nesting property of $\mathcal{X}$, we have that

$$P_i(a,b) = a_i + b_i + \tilde{P}_i(a_1, \dots, a_{i-1}, b_1, \dots, b_{i-1}),$$

where $\tilde{P}_i$ is a polynomial with all terms of degree at least 2.

Let $k_1, \dots, k_d$ be positive integers and form $\mathcal{X}'$ as above. Then we have that

$$P'_i(a,b) = k_i P_i(a_1/k_1, \dots, a_d/k_d, b_1/k_1, \dots, b_d/k_d),$$

where the P'_i are the polynomials which define multiplication in $\mathfrak{g}$ with respect to $\mathcal{X}'$. In particular, defining $\tilde{P}'_i$ analgously, we have

$$\tilde{P}'_i(a,b) = k_i \tilde{P}_i(a_1/k_1, \dots, a_{i-1}/k_{i-1}, b_1/k_1, \dots, b_{i-1}/k_{i-1}).$$

It follows that we may inductively choose k_i of size $M^{O(1)}$ depending on P_i and $(k_1, \dots, k_{i-1})$ so that $\tilde{P}'_i$ has integer coefficients and thus so too does P'_i. Choosing

the k_i in this way we have $\operatorname{span}_{\mathbb{Z}} \mathcal{X}' * \operatorname{span}_{\mathbb{Z}} \mathcal{X}' \subset \operatorname{span}_{\mathbb{Z}} \mathcal{X}'$. Also, $\operatorname{span}_{\mathbb{Z}} \mathcal{X}'$ is clearly closed under inverses, and so is a multiplicative subgroup. Furthermore, $\operatorname{span}_{\mathbb{Z}} \mathcal{X}'$ is clearly discrete and its cocompactness in $\mathfrak{g}$ follows from [18, Theorem 5.1.6]. This completes the proof. □

Lemma 4.8.13. *Let Λ be an additive lattice in $\mathfrak{g}$ with $\mathbb{Z}$-basis $\mathcal{X}$, which has rational structure constants of height at most M. Then the multiplicative lattice $\langle \Lambda \rangle_*$ has a Mal'cev basis whose elements are rational linear combinations of $\mathcal{X}$ with coefficients of height $M^{O(1)}$.*

Proof. The existence of a Mal'cev basis for $\langle \Lambda \rangle_*$ follows from Lemma 4.8.12 and [18, Theorem 5.1.6]. It is not difficult to use (the proofs of) these results to recover the quantitative statement of this lemma. We omit the details. □

Deducing the Lie algebra version of Green and Tao's criterion for quantitative equidistribution from its Lie group predecessor is similar to what we have already done for Leibman's criterion in the qualitative setting. We will only sketch the proof below.

Theorem 4.8.14 (Green and Tao's criterion for quantitative equidistribution in the Lie algebra, Theorem 4.4.15)**.** *Let $0 < \delta < 1/2$, let $N_1, \dots, N_D \geq 1$ and let $[\boldsymbol{N}]$ denote the set $[N_1] \times \cdots \times [N_D]$. Let $\mathfrak{g}$ be a real nilpotent Lie algebra with dimension d, step s and basis $\mathcal{X}$ which is $1/\delta$-rational. Let $(p(\boldsymbol{n}))_{\boldsymbol{n} \in \mathbb{Z}^D}$ be a polynomial sequence in $\mathfrak{g}$. There are positive constants $c_{d,s,D}, c'_{d,s,D}$ such that at least one of the following is true:*

1. *there is some $N_i < \delta^{-c'_{d,s,D}}$,*
2. *$(p(\boldsymbol{n}))_{\boldsymbol{n} \in [\boldsymbol{N}]}$ is δ-equidistributed in $(\mathfrak{g}, \mathcal{X})$,*
3. *p is not $(\delta^{-c_{d,s,D}}, \boldsymbol{N})$-additively irrational in $(\mathfrak{g}, \mathcal{X})$.*

Furthermore, when $N_1 = \cdots = N_D$, the first option above may be removed.

Proof. Let all constants depend on s, d, D. Suppose that (1) and (2) do not hold and denote $G = \exp \mathfrak{g}$ (as above, we are implicitly invoking Lie's third theorem) and $\Gamma = \exp(\langle \operatorname{span}_{\mathbb{Z}} \mathcal{X} \rangle_*)$ so that, by definition, $\exp \circ p$ is not δ-equidistributed in G/Γ. Let $\mathcal{X}_{\mathrm{Mal}}$ be the Mal'cev basis for Γ which is produced by Lemma 4.8.13; it is $\delta^{-O(1)}$-rational. By Theorem 4.8.10, there is a horizontal character η_0 on G/Γ with $0 < ||\eta_0|| < \delta^{O(1)}$ such that $||\eta_0 \circ g||_{C^\infty[\boldsymbol{N}]} < \delta^{-O(1)}$. Then one may argue using basic Lie theory as in the qualitative setting to find from η_0 some nontrivial $\eta \in \operatorname{Hom}_{\mathcal{X}_{\mathrm{Mal}}}(\mathfrak{g}, \mathbb{R})$

with complexity (as in our Definition 4.4.5) $\delta^{-O(1)}$ such that $||\eta \circ p||_{C^\infty[\boldsymbol{N}]} < \delta^{-O(1)}$. But the elements of $\mathcal{X}_{\text{Mal}}$ are rational linear combinations of $\mathcal{X}$ of height $\delta^{-O(1)}$, so there is some $C = \delta^{-O(1)}$ such that $\eta' := C\eta$ is a nontrivial element of $\text{Hom}_{\mathcal{X}}(\mathfrak{g}, \mathbb{R})$[12] of complexity $\delta^{-O(1)}$. Thus we have

$$||\eta' \circ p||_{C^\infty[\boldsymbol{N}]} \leq C||\eta \circ p||_{C^\infty[\boldsymbol{N}]} < \delta^{-O(1)},$$

and so p is not $(\delta^{-O(1)}, \boldsymbol{N})$-additively irrational in $(\mathfrak{g}, \mathcal{X})$. □

4.8.3 Rational and small polynomial sequences

Lemma 4.8.15 (Smallness and rationality under products)**.** *Let $\mathcal{X}$ be a rational basis for $\mathfrak{g}$ such that the Lie bracket has rational structure constants of height at most M with respect to $\mathcal{X}$. Let e, e' be polynomial sequences adapted to $\mathfrak{g}_\bullet$ which are (A, N)-small and let r, r' be polynomial sequences adapted to $\mathfrak{g}_\bullet$ which are A-rational. Then $e * e'$ and $r * r'$ are polynomial sequences adapted to $\mathfrak{g}_\bullet$, the former is $((AM)^{O(1)}, N)$-small and the latter is $(AM)^{O(1)}$-rational.*

Proof. That $e * e'$ and $r * r'$ are polynomial sequences adapted to $\mathfrak{g}_\bullet$ is standard. The rest follows from definitions and some easy, but disproportionately tedious, arguments. We omit the details. □

Lemma 4.8.16. *Suppose that r is a polynomial sequence in $\mathfrak{g}$ which is A-rational with respect to $\mathcal{X}$. Suppose that $\mathfrak{g}$ has rational structure constants of height at most M. Then for any $*$-multiplicative subgroup Γ of $\mathfrak{g}$ which contains $\langle \mathcal{X} \rangle$, the orbit $r(n) * \Gamma$ is $(AM)^{O(1)}$-periodic.*

Proof. We show that there is an integer q of size $(AM)^{O(1)}$ such that $p(n)^{-1} * p(n+q) \in \text{span}_{\mathbb{Z}} \mathcal{X}$ for all integers n. For the meantime let q be undetermined. Note $p(n)^{-1} = -p(n)$ and let p' be the polynomial sequence in 2 variables such that $qp'(n, q) = p(n + q) - p(n)$. By Baker-Campbell-Hausdorff we have

$$p(n)^{-1} * p(n + q) = p(n + q) - p(n) - \frac{1}{2}[p(n), p(n + q)] + \cdots .$$

We also have from definitions that $[p(n), p(n+q)] = q[p(n), p'(n, q)]$, and indeed every term in the Baker-Campbell-Hausdorff expansion will be a polynomial in n and q which is divisible by q and which has coefficients in $(\frac{1}{(AM)^{O(1)}}\mathbb{Z})^{\deg \mathfrak{g}}$, so that ultimately the polynomial $p(n)^{-1} * p(n + q)$ has this property. Therefore, a specific value of q may be chosen of size at most $(AM)^{O(1)}$ so that, for all $n \in \mathbb{Z}$, $p(n)^{-1} * p(n + q) \in \text{span}_{\mathbb{Z}} \mathcal{X}$. □

[12] Recall that $\text{Hom}_{\mathcal{X}}(\mathfrak{g}, \mathbb{R})$ denotes the set of Lie algebra homomorphisms from $\mathfrak{g}$ to $\mathbb{R}$ which map $\mathcal{X}$ to $\mathbb{Z}$.

4.8.4 Factorising for filtration irrationality

In this section we sketch how to deduce Proposition 4.6.2 from [35, Lemma A.10]. It is mostly a matter of transferring between related notions of complexity whose details we largely omit. The main additional ingredient is the following.

Lemma 4.8.17. *Let $M > 1$ be a positive integer. Let $\mathfrak{g}$ have rational basis $\mathcal{X}$ which has rational structure constants of height at most M. Then there exists a multiplicative lattice $\Gamma \leq G$ such that $M^{O(1)} \operatorname{span}_{\mathbb{Z}} \mathcal{X} \subseteq \log \Gamma \subseteq \frac{1}{M^{O(1)}} \operatorname{span}_{\mathbb{Z}} \mathcal{X}$.*[13]

Proof. Firstly we argue that there is a strong Mal'cev basis[14] $\mathcal{X}'$ such that

$$M^{O(1)} \operatorname{span}_{\mathbb{Z}} \mathcal{X}' \subseteq \operatorname{span}_{\mathbb{Z}} \mathcal{X} \subseteq \frac{1}{M^{O(1)}} \operatorname{span}_{\mathbb{Z}} \mathcal{X}'. \tag{4.10}$$

This is equivalent to finding a strong Mal'cev basis whose elements may be written as rational linear combinations of $\mathcal{X}$ with coefficients of height at most $M^{O(1)}$. Once one makes the observation that the elements of the descending central series have complexity $M^{O(1)}$ in $\mathfrak{g}$ with respect to $\mathcal{X}$ as a consequence on the bound on the height of the rational structure constants, the rest of the claim is established in the proof of [38, Proposition A.9].

Next, by working through the details of [18, Theorem 5.1.8(b)] (in particular, making note of quantitative bounds), one obtains that there is an integer $K = M^{O(1)}$ such the set $\mathbb{Z}^d \subset \mathbb{R}^d$ in *Mal'cev coordinates*[15] with respect to $K\mathcal{X}'$ is a multiplicative lattice (and so in fact $K\mathcal{X}'$ is a Mal'cev basis for this lattice). Let Γ be this lattice. Then [38, Lemma A.2] yields that

$$M^{O(1)} \operatorname{span}_{\mathbb{Z}} K\mathcal{X}' \subseteq \log \Gamma \subseteq \frac{1}{M^{O(1)}} \operatorname{span}_{\mathbb{Z}} K\mathcal{X}'.$$

Putting this together with Equation (4.10) proves the lemma. □

Next we restate [35, Lemma 2.10] from which we will derive Proposition 4.6.2. The statement is copied verbatim (including notation) from [35] and is included here for the convenience of the reader.

Lemma 4.8.18 ([35, Lemma 2.10])**.** *Let $(G/\Gamma, G_\bullet)$ be a degree $\leq s$ filtered nilmanifold of complexity $\leq M_0$, and let $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. For any growth function $\mathcal{F}'$, we can find a quantity $M_0 \leq M \leq O_{M,\mathcal{F}'}(1)$ and a factorisation $g = \beta g' \gamma$ where:*

[13] In the language of [38], $\mathcal{X}$ is then a *$M^{O(1)}$-rational weak basis* for G/Γ.

[14] Here our usage of 'strong Mal'cev basis' is as in [18]: a basis $\{X_1, \ldots, X_d\}$ for $\mathfrak{g}$ such that $\operatorname{span}_{\mathbb{R}}\{X_i, \ldots, X_d\}$ is an ideal for all i.

[15] Our usage of 'Mal'cev coordinates' is as in [38]. These coordinates are also known in the literature as *coordinates of the second kind*.

$\beta \in poly(\mathbb{Z}, G_\bullet)$ *is* $(O_M(1), N)$*-smooth;*

$g' \in poly(\mathbb{Z}, G_\bullet)$ *is* $(\mathcal{F}'(M), N)$*-irrational in a subnilmanifold* $(G'/\Gamma', G'_\bullet)$ *of* $(G/\Gamma, G_\bullet)$ *of complexity* $O_M(1)$*, and*

$\gamma \in poly(\mathbb{Z}, G_\bullet)$ *is* $O_M(1)$*-periodic.*

We are ready to prove Proposition 4.6.2.

Proposition 4.8.19 (Proposition 4.6.2)**.** *Let* $M_0 > 1$*. Let* $\mathfrak{g}$ *be a real nilpotent Lie algebra of dimension at most* M_0 *with filtration* $\mathfrak{g}_\bullet$ *of length at most* M_0*. Let* $\mathcal{X}$ *be a rational basis for* $\mathfrak{g}$ *with structure constants of height at most* M_0*. Let* p *be a polynomial sequence adapted to* $\mathfrak{g}_\bullet$ *with* $p(0) = 0$*. For any growth function* $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$*, there exists* $M_0 \leq M \leq O_{M_0,\mathcal{F}}(1)$ *and a factorisation* $p = e * p' * r$ *where* e *is* $(O_M(1), N)$*-small, where* r *is* $O_M(1)$*-rational and where* p' *is* $(\mathcal{F}(M), N)$*-filtration irrational in a subfiltration* $\mathfrak{g}'_\bullet \leq \mathfrak{g}_\bullet$ *of complexity* $O_M(1)$*.*[16] *This filtration irrationality is measured with respect to* $\mathcal{X}'$*, a Mal'cev basis*[17] *for* $\mathfrak{g}'_\bullet$ *which may be written as a linear combination of* $\mathcal{X}$ *with rational coefficients of height* $O_M(1)$*..*

Proof. We sketch the argument as a detailed proof is both straightforward and tedious. We refer the reader to [38, Appendix A] for more details on Mal'cev bases and quantitative aspects thereof. We will use terminology from there without further introduction.

Let Γ be the multiplicative lattice produced by Lemma 4.8.17 with input $\mathcal{X}$ and M_0. Then G/Γ will have complexity $M_0^{O(1)}$ as a nilmanifold (cf. [35, Definition 1.4]).[18] Invoke Lemma 4.8.18 for the polynomial sequence $g = \exp \circ p$ and with lattice Γ. One may go inside the proof of [35, Lemma A.10] and also use [38, Lemma A.2] to show that the $(O_M(1), N)$-smooth (resp. $O_M(1)$-periodic) sequence produced by [35, Lemma A.10] may instead by taken to be $(O_M(1), N)$-small (resp. $O_M(1)$-rational) by our definition.[19] Inferring the complexity of $\mathfrak{g}'_\bullet := (\log G'_i)_{i=1}^s$ from that of $G'_\bullet$ is an easy exercise in expanding definitions. Furthermore, the $(\mathcal{F}(M), N)$-filtration irrationality of p' with respect to an appropriate basis in $\mathfrak{g}'$ may be inferred from [38, Proposition A.10] and the irrationality of g' in $G'/(G' \cap \Gamma)$. □

[16] That is, for all i, $\mathfrak{g}'_i$ is a subspace of $\mathfrak{g}_i$ of complexity $O_M(1)$.

[17] See [38, Definition 2.1].

[18] This is not entirely trivial as one needs to construct from $\mathcal{X}$ a Mal'cev basis for Γ to exhibit the complexity of G/Γ. This is done in [38, Lemma A.8, Proposition A.9] together with our Lemma 4.8.17 which yields that $\mathcal{X}$ is a $M^{O(1)}$-rational weak basis for Γ as per [38, Definition A.7]. Alternatively, one may go inside the proof of Lemma 4.8.17 to see that the basis $K\mathcal{X}'$ is in fact an appropriate Mal'cev basis for Γ.

[19] With respect to the Mal'cev basis used in [35, Lemma A.10], but then the same is true with respect to $\mathcal{X}$ since the two bases are related by rational linear combinations of height $M_0^{O(1)}$.

4.9 Appendix B: Some quantitative integer linear algebra

In this appendix we define what a Hermite basis is and collect some (very standard) results that one might use to prove Lemmas 4.4.10 and 4.4.9, which are essentially all of the form 'integer linear algebra engenders at most polynomial growth in the size of the integers with which we are computing'.

4.9.1 Hermite normal form

Given a vector space $\mathfrak{g}$ with basis $\mathcal{X}$ and a rational subspace S, we want to define a basis $\mathcal{X}'$ for S such that $\operatorname{span}_{\mathbb{Z}} \mathcal{X}' = (\operatorname{span}_{\mathbb{Z}} \mathcal{X}) \cap S$. To do so, we will use the theory of the Hermite normal form. Everything in this subsection is completely standard and is included here for the reader's convenience.

Let m, n be positive integers and let M be a $m \times n$ integer valued matrix.

Definition/Theorem 4.9.1. *The Hermite normal form of M is the unique matrix H of the form $H = MU$ where U is a unimodular matrix and H has the following properties:*

1. *H is lower triangular,*
2. *for any column, the highest nonzero entry (the 'pivot') is strictly below that of all columns to its left,*
3. *all pivots are positive,*
4. *for all rows with a pivot, all elements to the right of the pivot are zero and all elements to the left of the pivot are non-negative and strictly smaller than the pivot.*

For $\mathfrak{g}, \mathcal{X}, S$ as above, the rationality of S in $\mathfrak{g}$ with respect to the rational structure induced by $\mathcal{X}$ yields that $\operatorname{span}_{\mathbb{Z}} \mathcal{X} \cap S$ is a full rank lattice in S. Let $\mathcal{Z}$ be any $\mathbb{Z}$-basis for this lattice and let $M_{\mathcal{Z}}$ be the $\dim \mathfrak{g} \times \dim S$ integer matrix whose columns are the elements of $\mathcal{Z}$ with respect to $\mathcal{X}$. Let H be the Hermite normal form of $M_{\mathcal{Z}}$. The above Definition/Theorem yields that in fact H is independent of the choice of basis $\mathcal{Z}$, since any other $M_{\mathcal{Z}'} = M_{\mathcal{Z}} U$ for some unimodular U. Thus, to any $\mathcal{X}, S$ we may associate a unique matrix $H_{\mathcal{X},S}$ in Hermite normal form whose columns are a $\mathbb{Z}$-basis for $\operatorname{span}_{\mathbb{Z}} \mathcal{X} \cap S$.

Definition 4.9.2 (Hermite basis)**.** Let $\mathfrak{g}, \mathcal{X}, S, H_{\mathcal{X},S}$ be as above. We define the *Hermite basis* $\mathcal{X}'_{\mathcal{X},S}$ to be the (ordered) set of columns of the Hermite normal form $H_{\mathcal{X},S}$ with respect to $\mathcal{X}$.

The following lemma is a corollary of any sensible algorithm to compute the Hermite normal form.

Lemma 4.9.3. *Let $m, n = O(1)$. Let M be a $m \times n$ matrix with integer entries of size at most C. Then the Hermite normal form of M has entries of size $C^{O(1)}$.*

As a result, given bounds on the complexity of a subspace, one may bound the size of the entries of its Hermite basis.

The following lemma is essentially all of the linear algebra that is needed to prove Lemmas 4.4.10 and 4.4.9. We omit a proof.

Lemma 4.9.4. *Let $m, n = O(1)$. Let M be a $m \times n$ full rank matrix with integer entries of size at most C. Then* $\ker M \cap \mathbb{Z}^n$ *is a lattice which has a $\mathbb{Z}$-basis consisting of integer vectors whose elements have size at most $C^{O(1)}$.*

Conversely, given a $\mathbb{Z}$-basis for a rank-m integer lattice L in $\mathbb{R}^n$ whose vectors have elements of size at most C, there is an $m \times n$ integer matrix M whose entries have size at most $C^{O(1)}$ such that $\ker M \cap \mathbb{Z}^n = L$.

Chapter 5

On Szemerédi's theorem with differences from a random set

5.1 Introduction

5.1.1 Notation and definitions

For a positive integer N, let $[N]$ denote the set $\{1, \dots, N\}$.

For $\delta > 0$, a subset of A of $[N]$ (respectively $\mathbb{F}_p^n$) will be said to have *δ-positive density* (or be *δ-dense*) if $|A| \geq \delta N$ (respectively $\geq \delta p^n$). A subset B of $\mathbb{N}$ will be said to have *positive upper density* if $\limsup_N |B \cap [N]|/N > 0$.

In a subset A of an abelian group, a *k-term arithmetic progression in A* (also *kAP*) is a pair (x, d) such that $x, x + d, \dots, x + (k-1)d \in A$. For $S \subset \mathbb{N}$, a k-term arithmetic progression has *common difference in S* if, in the above notation, $d \in S$.

When the ambient set is $[N]$ (respectively $\mathbb{F}_p^n$), we will say that *Szemerédi's theorem with common differences in S holds* if, for all $\delta > 0, k \geq 2, N > N_0(k, \delta)$ (respectively $n > n_0(k, \delta)$) and sets $A \subset [N]$ (respectively $\subset \mathbb{F}_p^n$) of δ-positive density, there exists a non-trivial k-term arithmetic progression in A with common difference in S. When the ambient set is $\mathbb{N}$, *Szemerédi's theorem with common differences in S holds* means that all sets $B \subset \mathbb{N}$ with positive upper density contain a non-trivial k-term arithmetic progression with common difference in S.

5.1.2 Context

In 1953, Roth [58] showed that sets of integers with positive upper density contain non-trivial 3-term arithmetic progressions. The result was famously extended to arbitrarily long arithmetic progressions by Szemerédi in 1975 [64]. It is well known

that this is equivalent to the finitary formulation which asserts that, for N sufficiently large in terms of k and δ, all δ-dense subsets of $[N]$ contain non-trivial kAPs.

A natural generalisation is to consider under what conditions Szemerédi's theorem is true when the set S of allowed common differences in arithmetic progressions is restricted. It transpires that Szemerédi's theorem holds with common differences restricted to some fairly sparse sets $S \subset \mathbb{N}$; for example, a result of Bergelson and Leibman [8] says that $S = \{1^{100}, 2^{100}, 3^{100}, \ldots\}$ (or indeed $\{p(n) : n \in \mathbb{N}\}$ for an integer polynomial p with $p(0) = 0$) is sufficient. Of course, the set S under consideration by Bergelson and Leibman is of a special structure.

We are interested in the situation where S is chosen at random. In the finitary model, it is common practice to construct the random set S by selecting each $d \in [N]$ to lie in S independently with equal probability.

For 2APs, it is known [9] that if $\mathbb{P}(d \in S) = \omega(\log N/N)$ then Szemerédi's theorem with common difference in S holds asymptotically almost surely (a.a.s), and conversely that if $\mathbb{P}(d \in S) \leq C \log N/N$ then Szemerédi's theorem with common difference in S a.a.s fails. For kAPs, the current record is due to Briët–Castro-Silva [13], which states that $\mathbb{P}(d \in S) = N^{-\frac{2}{k}+o(1)}$ is sufficient for Szemerédi's theorem to hold a.a.s. with common differences in S. This result extends the breakthrough work of Alrabiah et al. [1] on the $k = 3$ case.

For $k \geq 3$, there is a substantial gap between these results and conjectures found in work of Frantzikinakis and others (see [23, Problem 31], [24, Conjecture 2.5]), which state that in fact the threshold for $k \geq 3$ is still $\log N/N$.

5.1.3 Results

For 2-term arithmetic progressions, $\mathbb{P}(d \in S) = \omega(\log N/N)$ is sufficient for Szemerédi's theorem with common difference in S to hold asymptotically almost surely. It is not difficult to prove that the analogous fact is true over finite fields: that $\mathbb{P}(d \in S) = \omega(\log(p^n)/p^n)$ is sufficient. (One uses similar Fourier-analytic arguments to the $k = 2$ case over the integers; here there is only a discrete set of additive characters so the argument is even easier.)

The main result of this chapter is that the threshold for the random Szemerédi problem for $k = 3$ is larger over $\mathbb{F}_p^n$ than the conjecture threshold over the integers. In fact, if elements are selected to lie in S independently with probability

$$\mathbb{P}(d \in S) = \frac{cn^2}{p^n} = \frac{c \log_p^2 |\mathbb{F}_p^n|}{|\mathbb{F}_p^n|},$$

where $c = 1/2 - o(1)$, then there will almost surely exist a set A with positive density such that A contains no 3APs with common difference in S. (Actually we deal with a slightly different probability model for convenience, but the above statement is an easy consequence of Corollary 5.1.2.)

The reason for the different behavior is that there are far more quadratic obstructions to 3APs in $\mathbb{F}_p^n$. Indeed, for $M \in M_n(\mathbb{F}_p)$, define $A_M = \{x : x^\top M x = 0\}$ and note that A_M has positive density (uniformly in n). One observes that if $x, x+d, x+2d \in A_M$, then $d^\top M d = 0$, that is, $d \in A_M$. It follows that if all A_M are to have 3APs with common difference in S, then S must have the following property: for all $M \in M_n(\mathbb{F}_p)$, there exists $d \in S$ such that $d^\top M d = 0$.

Theorem 5.1.1. *Fix p an odd prime. If $S \subset \mathbb{F}_p^n$ is formed by selecting at most $\binom{n+1}{2} - 11n \log_p n$ elements of $\mathbb{F}_p^n$ independently at random, then almost surely as $n \to \infty$ there exists some $M \in M_n(\mathbb{F}_p)$ such that $d^\top M d \neq 0$ for all $d \in S$.*

Corollary 5.1.2. *Fix p an odd prime. If $S \subset \mathbb{F}_p^n$ is formed by selecting at most $\binom{n+1}{2} - 11n \log_p n$ elements of $\mathbb{F}_p^n$ independently at random, then almost surely as $n \to \infty$ there exists some set $A \subset \mathbb{F}_p^n$ of positive density such that A contains no 3-term arithmetic progression with common difference in S.*

We note that the published version of this chapter [3] shows how the conjecture about the $\log N/N$ threshold in the integers follows from a conjecture about how so-called dual functions are approximated by convex combinations of nilsequences. Since then, however, Briët and Green [10] have shown the latter conjecture to be false. We therefore omit this argument from this thesis and refer the interested reader to [3].

5.2 Proof of the main result

In this section we will prove Theorem 5.1.1.

It suffices to consider symmetric matrices because if $M \in M_n(\mathbb{F}_p)$ then, letting $M' = (M + M^\top)/2 \in S_n(\mathbb{F}_p)$, we have $d^\top M d = d^\top M' d$ for all d. We identify $S_n(\mathbb{F}_p)$ with $\mathbb{F}_p^{\binom{n+1}{2}}$ naturally; we will write M_v for the matrix corresponding to a vector v, and v_M for the vector corresponding to a matrix M. Define $\varphi : \mathbb{F}_p^n \to \mathbb{F}_p^{\binom{n+1}{2}}$ to be the degree 2 Veronese map, that is $(d_1, \ldots, d_n) \mapsto (d_i d_j)_{1 \leq i \leq j \leq n}$. Then $d^\top M d = \varphi(d) \cdot v_M$ and so $d^\top M d \neq 0$ for all $d \in S$ if and only if $v_M \notin \cup_{d \in S} \varphi(d)^\perp$.

The following lemma demonstrates that if $\{\varphi(d) : d \in S\}$ is linearly independent then there exists some matrix M with $v_M \notin \cup_{d \in S} \varphi(d)^\perp$.

Lemma 5.2.1. *Let $\{v_1, \dots, v_k\}$ be linearly independent in an m-dimensional vector space over $\mathbb{F}_p$. Then*

$$\left|\bigcup_{i=1}^{k} v_i^{\perp}\right| = p^m \left(1 - \left(\frac{p-1}{p}\right)^k\right).$$

In particular,

$$\bigcup_{i=1}^{k} v_i^{\perp} \subsetneq \mathbb{F}_p^m.$$

Proof. Linear algebra; we omit the details. □

The goal will now be to show that, almost surely as $n \to \infty$, the elements $\varphi(d), d \in S$ are linearly independent. Let $\mathcal{W}_k$ be the set of all k dimensional subspaces of $\mathbb{F}_p^{\binom{n+1}{2}}$ and let $W_k \in \mathcal{W}_k$ be a subspace such that $|W_k \cap \operatorname{Im}\varphi| = \max_{W \in \mathcal{W}_k} |W \cap \operatorname{Im}\varphi|$.

Lemma 5.2.2. *Let $N = |S|$. The probability that $\varphi(S)$ is linearly independent is bounded below by*

$$\left(1 - \mathbb{P}_{d \in \mathbb{F}_p^n}\left(\varphi(d) \in W_N\right)\right)^N.$$

Proof. Sampling S by selecting elements $d_1, \dots, d_N$ at independently at random, the probability that $\varphi(S)$ is linearly independent is bounded below by

$$\begin{aligned}
&\mathbb{P}(d_1 \neq 0) \prod_{i=2}^{N} \mathbb{P}(\varphi(d_i) \notin \operatorname{span}\{\varphi(d_1), \dots, \varphi(d_{i-1})\}) \\
\geq &\mathbb{P}(d_1 \neq 0) \prod_{i=2}^{N} \mathbb{P}(\varphi(d_i) \notin W_{i-1}) \\
\geq &\left(1 - \mathbb{P}_{d \in \mathbb{F}_p^n}\left(\varphi(d) \in W_N\right)\right)^N.
\end{aligned}$$

□

By Lemma 5.2.2, to show that $\varphi(S)$ is almost surely linearly independent, it suffices now to show that $\mathbb{P}_{d \in \mathbb{F}_p^n}\left(\varphi(d) \in W_N\right) = o(1/n^2)$. As an intermediate step, we will show in Proposition 5.2.5 that

$$\mathbb{P}_{d \in \mathbb{F}_p^n}\left(\varphi(d) \in W_N\right) \leq \mathbb{E}_{v \in W_N^{\perp}} p^{-\frac{1}{2}\operatorname{rank} M_v}.$$

We separate out the main analytic observations in the following two lemmas. The first follows from orthogonality of characters.

Lemma 5.2.3. *Let V be a vector space of functions $\mathbb{F}_p^k \to \mathbb{F}_p$ under pointwise operations. Say $V(x) = 0$ if $v(x) = 0$ for all $v \in V$. Then*

$$\mathbb{P}_x(V(x) = 0) = \mathbb{E}_{x,v} e_p(v(x)).$$

The second is a standard estimate for Gauss sums.

Lemma 5.2.4. *Let M be a symmetric matrix over $\mathbb{F}_p$, where $p \neq 2$. Then,*

$$\left|\mathbb{E}_x e_p(x^\top M x)\right| = p^{-\frac{1}{2}\operatorname{rank} M}.$$

Proof. The result follows from taking square roots after the following computation:
$\left|\mathbb{E}_x e_p(x^\top M x)\right|^2 = \mathbb{E}_{x,h} e_p((x+h)^\top M(x+h) - x^\top M x) = \mathbb{E}_h e_p(h^\top M h) 1_{Mh=0} = p^{-\operatorname{rank} M}$.

□

Proposition 5.2.5. *We can bound $\mathbb{P}_d(\varphi(d) \in W_N)$ as follows:*

$$\mathbb{P}_d(\varphi(d) \in W_N) \leq \mathbb{E}_{M_v} p^{-\frac{1}{2}\operatorname{rank} M_v},$$

where the expectation is taken over all M_v such that $v \in W_N^\perp$.

Proof. Using Lemma 5.2.3 (with V the vector space of quadratic forms defined by $\{M_v\}$) and Lemma 5.2.4 we can compute that

$$\begin{aligned}
\mathbb{P}_d(\varphi(d) \in W_N) &= \mathbb{P}_d(d^\top M_v d = 0 \text{ for all } v \in W_N^\perp) \\
&= \mathbb{P}_d(V(d) = 0) \\
&= \mathbb{E}_{d,M_v} e_p(d^\top M_v d) \\
&\leq \mathbb{E}_{M_v} \left|\mathbb{E}_d e_p(d^\top M_v d)\right| \\
&= \mathbb{E}_{M_v} p^{-\frac{1}{2}\operatorname{rank} M_v}.
\end{aligned}$$

□

Finally, it remains to show that

$$\mathbb{E}_{M_v} p^{-\frac{1}{2}\operatorname{rank} M_v} = o(1/n^2).$$

We do so using the following (crude) observations. Firstly, the number of $n \times n$ matrices of rank at most r is bounded above by p^{2nr} (choose the row space in at most p^{nr} ways and then choose each of the n rows in at most p^r ways). Then, splitting the sum by rank, we have

$$\begin{aligned}
\sum_{M_v} p^{-\frac{1}{2}\operatorname{rank} M_v} &\leq \left|\{M_v : \operatorname{rank} M_v < 5\log_p n\}\right| \\
&\quad + p^{-\frac{5}{2}\log_p n}\left|\{M_v : \operatorname{rank} M_v \geq 5\log_p n\}\right| \\
&\leq p^{10n\log_p n} + p^{-\frac{5}{2}\log_p n}\left|\{M_v\}\right|,
\end{aligned}$$

and so, recalling that $N = \binom{n+1}{2} - 11n\log_p n$ and $|\{M_v\}| = p^{\binom{n+1}{2}-N} = p^{11n\log_p n}$,

$$\mathbb{E}_{M_v} p^{-\frac{1}{2}\operatorname{rank} M_v} \leq p^{-n\log_p n} + p^{-\frac{5}{2}\log_p n} = o(1/n^2).$$

This completes the proof of Theorem 5.1.1.

Chapter 6

On a question of Alon

6.1 Introduction

In graph theory, a graph H is called *common* if every two-colouring of K_n contains at least at many monochromatic copies of H as a random two-colouring does (in the limit $n \to \infty$). Goodman [25] proved that K_3 is common. Erdős conjectured that K_4 is common [20] and subsequently Burr and Rosta conjectured [11] that every graph is common. These conjectures were disproved by Sidorenko [62] and Thomason [67] who demonstrated respectively that a triangle with a pendant edge and K_4 are in fact not common. The classification of common graphs remains very much open to this day.

One observes that if H satisfies the stronger property that for every large graph G, among all subsets of G of fixed density, a random subset of G contains the fewest copies of H (in the limit $|V(G)| \to \infty$), then H is certainly common. Graphs satisfying this stronger property are called *Sidorenko*. Conjecturally [61], all bipartite graphs are Sidorenko; this conjecture remains wide open (despite the resolution of various special cases). Certainly, since Sidorenko's conjecture is not known to be false, all known instances of uncommon graphs are not bipartite. In fact, *every* non-bipartite graph can be made uncommon by the addition of sufficiently many pendant edges (an edge e is pendant to a graph H if $|e \cap V(H)| = 1$).

Theorem 6.1.1 ([46])**.** *If H is a non-bipartite graph with $m \geq 3$ vertices, then there is a positive integer $l_0(m)$ such that any graph obtained by successively adding at least l_0 pendant edges to H is uncommon.*

Alon has asked whether the analogous result is true in the arithmetic setting. A system of linear equations in $\mathbb{F}_p^n$ is *common* if every two-colouring of $\mathbb{F}_p^n$ yields at least as many monochromatic solutions as a random two-colouring, asymptotically

as $n \to \infty$. A system of linear equations in $\mathbb{F}_p^n$ is *Sidorenko* if among all subsets S of $\mathbb{F}_p^n$ of fixed density, the number of solutions in S is minimised when S is a random set, asymptotically as $n \to \infty$. It is clear that if a linear system is Sidorenko, then it is common. Saad and Wolf initiated the study of these properties in the arithmetic setting [59]. For both the Sidorenko and common properties, necessary [21] and sufficient [59] conditions for systems containing a single equation are known (in fact, a single equation in an even number of variables is Sidorenko if and only if it is common, and an equation in an odd number of variables is necessarily common but not Sidorenko). The classification problems for systems containing more than one equation remain wide open, despite recent interest and the resolution of special cases; see [21], [47], [48], [69], [68].

It transpires that adding a free variable to a system of linear equations (equivalently, a free dimension to the solution space) corresponds to the addition of a pendant edge in the graph-theoretic setting. In those cases when one can transfer between the graph-theoretic and arithmetic settings via Cayley graph constructions, these operations are equivalent. See [59, Example 4.2] for an instance of this correspondence and elsewhere in that paper for a brief discussion of the correspondence in general. For $l \geq 0$, let $\Psi^{(l)}$ be the linear system attained by adding l free variables to Ψ.

Question 6.1.2 (Alon, [59, Question 4.1])**.** *Is it true that for all non-Sidorenko Ψ, there exists l_0 such that for all $l \geq l_0$ we have $\Psi^{(l)}$ is uncommon?*

For example, it is known that three term arithmetic progressions, whose solution space is parameterised $\Psi = (x, x+d, x+2d)$, are common. Is it true that the system given by $\Psi^{(l)} = (x, x+d, x+2d, y_1, \ldots, y_l)$ is uncommon for all l sufficiently large? The answer is yes by the following theorem of Fox–Pham–Zhao.

Theorem 6.1.3 ([21, Theorem 1.5])**.** *The answer to Question 1.2 is yes among all Ψ whose image has codimension one.*

In light of Question 6.1.2 and for convenience in this chapter, we will refer to systems Ψ with the property that $\Psi^{(l)}$ is *common* for all sufficiently large l as *Alon*. We have mentioned already that Sidorenko implies common; in fact a similarly easy calculation shows that Sidorenko implies Alon. In this language, the above question essentially (i.e. under the assumption that linear systems cannot oscillate infinitely often between common and uncommon under the addition of free variables) asks whether Sidorenko is equivalent to Alon. We see from our discussion above that this is true in the graph-theoretic setting under Sidorenko's conjecture. It turns out that it is not true in the arithmetic setting.

Theorem 6.1.4. *There exists a system of linear equations which is Alon but not Sidorenko, so the answer to Question 6.1.2 is no.*

The example which demonstrates the negative answer to Question 6.1.2 is the following rank 2 system of equations in 9 variables, which we will henceforth denote by Φ:

$$x_1 - x_2 + x_3 - x_4 = 0$$
$$x_5 - x_6 + x_7 - x_8 + x_9 = 0.$$

Theorem 6.1.4 is proven in Section 6.3. As an intermediate lemma in the proof, we provide a lower bound on the *geometric* mean of the number of monochromatic solutions to Φ. We show in Section 6.4 that a lower bound of this type is in fact also necessary, and indeed that a slightly weaker version of the Alon property is equivalent to a geometric notion of commonness in which one asks that the geometric mean of the number of monochromatic solutions is at least that of a random two-colouring. This is proven in Proposition 6.4.5. The intended utility is that geometric commonness is oftentimes more easily studied than the Alon property.

It transpires that Φ also exhibits a negative answer to a question of Kamčev–Liebenau–Morrison. Recall that a system is said to be *translation-invariant* if it is solved by setting $x_1 = x_2 = x_3 = \cdots$. If a system is not translation invariant, then the set of elements of $\mathbb{F}_p^n$ whose first entry is equal to one has positive density but no solutions, so any system which is not translation-invariant cannot be Sidorenko. The only known examples of systems which are not translation invariant and are common are systems of rank one (i.e. systems comprising a single equation), whereupon commonness follows trivially from a cancellation which does not occur for systems of two or more equations. This inspired the following question of Kamčev–Liebenau–Morrison.

Question 6.1.5. *[48, Question 5.6] Does there exist a system of rank at least two which is common, but not translation-invariant?*

Theorem 6.1.6. *The answer to Question 6.1.5 is yes.*

The system Φ has rank two and is not translation invariant. The proof of Theorem 6.1.6 is completed in Lemma 6.3.1, which shows that Φ is common. Since uploading [6] to the arXiv, the author has been made aware that Question 6.1.5 has been independently resolved in $\mathbb{F}_2^n$ by Král'–Lamaison–Pach in forthcoming work. See [50].

Finally, we also take this opportunity to address the following question of Kamčev–Liebenau–Morrison, which turns out to be related to the addition of free variables.

Question 6.1.7. *[47, Question 6.3] Suppose that Ψ is a linear system such that there is a set $A \subset \mathbb{F}_p^n$ such that the density of monochromatic solutions in (A, A^c) is less than $\alpha^t + (1-\alpha)^t$, where $\alpha = |A|/p^n$. Is Ψ necessarily uncommon if one restricts one's attention to sets of density roughly $1/2$?*

For a suitable interpretation of 'roughly', the answer to [47, Question 6.3] is no, even with the stronger demand that the density of monochromatic solutions in (A, A^c) is less than 2^{1-t}. To see why, we will need to introduce some notation.

6.2 Preliminaries, and an answer to Question 6.1.7

Throughout this chapter we assume that p is an odd prime.

If Ψ is a system of m linear equations in t variables, described by a $m \times t$ matrix M_Ψ with entries in $\mathbb{F}_p$, and $f : \mathbb{F}_p^n \to \mathbb{C}$, then we define

$$T_\Psi(f) := \mathbb{E}_{\boldsymbol{x} \in (\mathbb{F}_p^n)^t : M_\Psi \boldsymbol{x} = 0} f(x_1) f(x_2) \cdots f(x_t).$$

In this chapter we only consider the case of homogeneous linear systems: each equation is of the form $\sum_{i=1}^t a_i x_i = 0$. Furthermore, we always assume that $t > m$ and that the system is of full rank. In particular, Ψ has nontrivial solutions. Recall that $\mathbb{E}$ refers to a normalised sum, so here is shorthand for $\frac{1}{\#\{\boldsymbol{x} \in (\mathbb{F}_p^n)^t : M_\Psi \boldsymbol{x} = 0\}} \sum$. One observes that if $A \subset \mathbb{F}_p^n$ and 1_A is the characteristic function for the set A, then $T_\Psi(1_A)$ is a normalised count for the number of solutions to Ψ in A. Sometimes, when the underlying linear system is clear from context, we may just use the notation $T(f)$ for brevity.

We recall the Fourier transform on $\mathbb{F}_p^n$. Let $e_p(\cdot)$ be shorthand for $e^{\frac{2\pi i \cdot}{p}}$. For $h \in \mathbb{F}_p^n$, we define

$$\hat{f}(h) := \mathbb{E}_{x \in \mathbb{F}_p^n} f(x) e_p(-h \cdot x),$$

whereupon one obtains the Fourier inversion formula:

$$f(x) = \sum_{h \in \mathbb{F}_p^n} \hat{f}(h) e_p(x \cdot h),$$

and Parseval's identity:

$$\mathbb{E}_{x \in \mathbb{F}_p^n} |f(x)|^2 = \sum_{h \in \mathbb{F}_p^n} |\hat{f}(h)|^2.$$

As is also standard, we will deal with a functional version of the notion of commonness in which one studies the set of functions $f : \mathbb{F}_p^n \to [0,1]$ rather than focusing only on the subset of these which are characteristic functions of sets.

Definition 6.2.1. A system of t linear forms Ψ is *common* if, for all $n \geq 1$ and all $f : \mathbb{F}_p^n \to [0,1]$, we have

$$T_\Psi(f) + T_\Psi(1-f) \geq 2^{1-t}.$$

Clearly if a system is common by the above definition then it is common when one restricts only to characteristic functions of sets. The other direction is recovered by a standard argument in which one constructs a random set A from f by choosing $x \in A$ with probability $f(x)$; we will not dwell on this further but remark that details are worked out for example in [21].

Sometimes it is more convenient to describe a linear system by parameterising its solution space. In this case we have, equivalently, that

$$T_\Psi(f) = \mathbb{E}_{y_1,\dots,y_D \in \mathbb{F}_p^n} f(\psi_1(y_1,\dots,y_D)) \cdots f(\psi_t(y_1,\dots,y_D)),$$

where each ψ_i is a linear form mapping $(\mathbb{F}_p^n)^D$ to $\mathbb{F}_p^n$, $x_i = \psi_i(y_1,\dots,y_D)$ for each i and $D = t - m$.

Returning to Question 6.1.7, we exhibit a Ψ, $f : \mathbb{F}_p^n \to [0,1]$ such that $T_\Psi(f) < 2^{1-t}$, but $T_\Psi(f) \geq 2^{1-t}$ for all f with $\mathbb{E}f = 1/2$. By randomly sampling as discussed above, this yields a negative answer to Question 6.1.7.

Let Ψ be the system whose solution space is parameterised by $(x, x+d, x+2d)$, so $t = 3, m = 1, D = 2$ and $\psi_i(x,d) = x + (i-1)d$. Let $\Psi^{(1)}$ be parameterised by $(x, x+d, x+2d, y)$. It is shown in [59, Example 4.1] that $\Psi^{(1)}$ is uncommon (with a sequence of sets A_n for which $\left||A_n|/p^n - \frac{1}{2}\right| > \varepsilon$ for all n and some explicit $\varepsilon > 0$). On the other hand, it is well known that Ψ is common (this goes back to Cameron, Cilleruelo and Serra [12]). For a function $f : \mathbb{F}_p^n \to [0,1]$ with $\alpha := \mathbb{E}_{x \in \mathbb{F}_p^n} f(x)$, one sees that

$$T_{\Psi^{(1)}}(f) + T_{\Psi^{(1)}}(1-f) = \alpha T_\Psi(f) + (1-\alpha) T_\Psi(1-f), \tag{6.1}$$

since we average over the free variable y. In fact, in general, we see that

$$T_{\Psi^{(l)}}(f) + T_{\Psi^{(l)}}(1-f) = \alpha^l T_\Psi(f) + (1-\alpha)^l T_\Psi(1-f); \tag{6.2}$$

this formula will be useful later.

Following on from (6.1), if $\alpha = 1/2$ then

$$\inf_{f:\mathbb{E}f=\frac{1}{2}} T_{\Psi^{(1)}}(f) + T_{\Psi^{(1)}}(1-f) = \inf_{f:\mathbb{E}f=\frac{1}{2}} \frac{1}{2}(T_\Psi(f) + T_\Psi(1-f)) \geq 2^{-3},$$

since Ψ is common. Thus $\Psi^{(1)}$ is common among average $1/2$ functions, but is not common in general, answering a suitable interpretation of Question 6.1.7 in the negative.

This argument works more generally for adding free variables to any single equation in an odd number of variables: use [21, Theorems 1.4, 1.5]. Furthermore, any system of equations which is common among average-1/2 functions but not Alon yields counterexamples to Question 6.1.7 in this way.

6.3 Proof of Theorem 6.1.4 and Theorem 6.1.6

Recall we let Φ be the following rank 2 system of equations in 9 variables

$$x_1 - x_2 + x_3 - x_4 = 0$$
$$x_5 - x_6 + x_7 - x_8 + x_9 = 0.$$

Let A4 be shorthand for (the system containing) the first equation in Φ, an additive 4-tuple. Let A5 be shorthand for the second equation.

Since Φ is not translation-invariant and has rank two, to prove Theorem 6.1.6 we just need to show that Φ is common.

Lemma 6.3.1 (Proof of Theorem 6.1.6)**.** Φ *is common.*

Proof. Let $f : \mathbb{F}_p^n \to [0,1]$ have $\mathbb{E}f = \alpha$ and let $g = f - \alpha$. We observe that $T_\Phi(f) = T_{\mathrm{A4}}(f)T_{\mathrm{A5}}(f)$. Furthermore, since A4 has rank one, one sees that $T_{\mathrm{A4}}(\alpha + g) = \alpha^4 + T_{\mathrm{A4}}(g)$, and similarly for A5. Thus,

$$T_\Phi(f) = T_\Phi(\alpha + g) = \alpha^9 + \alpha^5 T_{\mathrm{A4}}(g) + \alpha^4 T_{\mathrm{A5}}(g) + T_\Phi(g),$$

so one computes

$$T_\Phi(f) + T_\Phi(1-f) = \alpha^9 + (1-\alpha)^9 + \left(\alpha^5 + (1-\alpha)^5\right) T_{\mathrm{A4}}(g) + \left(\alpha^4 - (1-\alpha)^4\right) T_{\mathrm{A5}}(g). \tag{6.3}$$

Recall that $T_{\mathrm{A4}}(g)$ is the (4th power of the) ℓ^4-norm of g on the Fourier side so in particular $T_{\mathrm{A4}}(g) \geq 0$. Thus if $\alpha = 1/2$ then we trivially have $T_\Phi(f)+T_\Phi(1-f) \geq 1/2^8$. Assume without loss of generality $\alpha < 1/2$. We have $T_{\mathrm{A4}}(g) = \sum_h |\hat{g}(h)|^4$ and $T_{\mathrm{A5}}(g) = \sum_h |\hat{g}(h)|^4 \hat{g}(h)$, and so $|T_{\mathrm{A5}}(g)| \leq T_{\mathrm{A4}}(g)||\hat{g}||_\infty \leq T_{\mathrm{A4}}(g)(1-\alpha)$. Thus continuing from (6.3),

$$T_\Phi(f) + T_\Phi(1-f) \geq \alpha^9 + (1-\alpha)^9 + T_{\mathrm{A4}}(g)\left(\alpha^5 + \alpha^4(1-\alpha)\right) \geq \alpha^9 + (1-\alpha)^9 \geq 2^{-8},$$

where the final inequality follows from convexity. □

Now we proceed to the proof of Theorem 6.1.4. Since Φ is not translation-invariant, it is not Sidorenko. To prove Theorem 6.1.4 we will show that Φ is Alon, i.e. $\Phi^{(l)}$ is common for all l sufficiently large. First we compile some properties of Φ.

Recall that we have assumed that p is odd.

Lemma 6.3.2. *There is a constant $c_0 > 0$ such that if $f : \mathbb{F}_p^n \to [0,1]$ has $\mathbb{E}f \geq 0.45$ then $T_\Phi(f) \geq c_0$.*

Proof. Let $\alpha \in (0, 1/2)$ and let $f : \mathbb{F}_p^n \to [0,1]$ have $\mathbb{E}f = \alpha$. Let $g = f - \alpha$. Since g is real-valued, we have $|\hat{g}(h)| = |\hat{g}(-h)|$ for all h, and so invoking the fact that $\hat{g}(0) = 0$, the fact that the characteristic is not 2, and Parseval's identity, we obtain

$$2 \sup_h |\hat{g}(h)|^2 \leq \sum_h |\hat{g}(h)|^2 = \mathbb{E}|g(x)|^2 \leq (1-\alpha)^2. \tag{6.4}$$

Thus $||\hat{g}||_\infty \leq (1-\alpha)/\sqrt{2}$. By Fourier-inversion as above, it follows that $|T_{\mathrm{A5}}(g)| \leq \frac{1-\alpha}{\sqrt{2}} T_{\mathrm{A4}}(g)$, and by Parseval again $T_{\mathrm{A4}}(g) \leq (1-\alpha)^4/2$. Thus

$$\begin{aligned} T_\Phi(f) = T_\Phi(\alpha + g) &= \alpha^9 + \alpha^5 T_{\mathrm{A4}}(g) + \alpha^4 T_{\mathrm{A5}}(g) + T_{\mathrm{A4}}(g) T_{\mathrm{A5}}(g) \\ &\geq \alpha^9 + T_{\mathrm{A4}}(g) \left(\alpha^5 - \frac{(1-\alpha)\alpha^4}{\sqrt{2}} - \frac{(1-\alpha)^5}{2\sqrt{2}} \right). \end{aligned}$$

One checks that the polynomial $q(x) = x^5 - \frac{(1-x)x^4}{\sqrt{2}} - \frac{(1-x)^5}{2\sqrt{2}}$ is negative for $x \in (0, 1/2)$ and so

$$T_\Phi(f) \geq \alpha^9 + \frac{(1-\alpha)^4}{2} \left(\alpha^5 - \frac{(1-\alpha)\alpha^4}{\sqrt{2}} - \frac{(1-\alpha)^5}{2\sqrt{2}} \right).$$

Defining $\tilde{q}(x) = x^9 + \frac{1}{2}(1-x)^4 \left(x^5 - \frac{(1-x)x^4}{\sqrt{2}} - \frac{(1-x)^5}{2\sqrt{2}} \right)$, one checks that $\tilde{q}(0.45) > 0$, which proves the result when $\mathbb{E}f = 0.45$. To conclude we observe that $\inf_{f:\mathbb{E}f=\alpha} T_\Phi(f)$ is monotonic in α. □

Next we show that Φ possesses a 'local Sidorenko' property with respect to the $||\dot{:}||_\infty$ norm. We will study local Sidorenko properties in greater detail in Chapter 7.

Lemma 6.3.3. *There is a constant $c_1 > 0$ such that, upon letting $f : \mathbb{F}_p^n \to [0,1]$, $\alpha := \mathbb{E}f$ and $g := f - \alpha$, if $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$ and $||\hat{g}||_\infty \leq c_1$ then $T_\Phi(f) \geq \alpha^9$.*

Proof. Let $\alpha \in [\frac{1}{3}, \frac{2}{3}]$. For $g : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}_x g(x) = 0$ we have, as we saw in Lemma 6.3.1, that $T_{\mathrm{A5}}(g) \leq ||\hat{g}||_\infty T_{\mathrm{A4}}(g) \leq ||\hat{g}||_\infty^3$, where we have used Parseval and the bound $||g||_\infty \leq 1$. Thus

$$\begin{aligned} T_\Phi(\alpha + g) &= \alpha^9 + \alpha^5 T_{\mathrm{A4}}(g) + \alpha^4 T_{\mathrm{A5}}(g) + T_{\mathrm{A4}}(g) T_{\mathrm{A5}}(g) \\ &\geq \alpha^9 + T_{\mathrm{A4}}(g)(\alpha^5 - \alpha^4 ||\hat{g}||_\infty - ||\hat{g}||_\infty^3). \end{aligned}$$

We have that $T_{\mathrm{A}4}(g) \geq 0$ so it remains to argue that there is a constant c_1 such that for all $0 \leq x \leq c_1$ and all $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$, we have $\alpha^5 - \alpha^4 x - x^3 \geq 0$. This is easily seen to be true. □

In our final lemma before the proof of Theorem 6.1.4, we establish the following inequality.

Lemma 6.3.4. *There are constants $c_2, c_3, C_4 > 0$ such that upon letting $f : \mathbb{F}_p^n \to [0,1]$, $\alpha := \mathbb{E}f$ and $g := f - \alpha$ we have that if $|\alpha - \frac{1}{2}| \leq c_2$ then*

$$T_\Phi(f)T_\Phi(1-f) \geq 2^{-18} + c_3||\hat{g}||_\infty^4 - C_4|\alpha - \frac{1}{2}|.$$

Proof. Letting $T_4 := T_{\mathrm{A}4}(g)$ and $T_5 := T_{\mathrm{A}5}(g)$ for brevity we have

$$\begin{aligned}
T_\Phi(f)T_\Phi(1-f) &= T(\alpha+g)T(1-\alpha-g) \\
&= (\alpha^9 + \alpha^5 T_4 + \alpha^4 T_5 + T_4T_5)\times \\
&\qquad\qquad ((1-\alpha)^9 + (1-\alpha)^5 T_4 - (1-\alpha)^4 T_5 - T_4T_5) \\
&= \left(2^{-9} + 2^{-5}T_4 + 2^{-4}T_5 + T_4T_5\right)\times \\
&\qquad\qquad \left(2^{-9} + 2^{-5}T_4 - 2^{-4}T_5 - T_4T_5\right) + O(|\alpha - \frac{1}{2}|) \\
&= \left(2^{-4} + T_4\right)^2 \left(2^{-10} - T_5^2\right) + O(|\alpha - \frac{1}{2}|).
\end{aligned}$$

We showed in (6.4) that $||\hat{g}||_\infty \leq ||g||_\infty/\sqrt{2}$ and so $||\hat{g}||_\infty \leq \frac{1}{2\sqrt{2}} + O(|\alpha - \frac{1}{2}|)$. Thus here we have

$$T_5 \leq ||\hat{g}||_\infty T_4 \leq \frac{1}{2\sqrt{2}} T_4 + O(|\alpha - \frac{1}{2}|) \leq 2^{-6.5} + O(|\alpha - \frac{1}{2}|).$$

Continuing from above,

$$\begin{aligned}
T_\Phi(f)T_\Phi(1-f) &= \left(2^{-4} + T_4\right)^2 \left(2^{-10} - T_5^2\right) + O(|\alpha - \frac{1}{2}|) \\
&\geq \left(2^{-4} + T_4\right)^2 \left(2^{-10} - 2^{-3}T_4^2\right) + O(|\alpha - \frac{1}{2}|) \\
&= 2^{-18} + 2^{-3}T_4\left(2^{-10} + 2^{-8}T_4 - 2^{-3}T_4^2 - T_4^3\right) + O(|\alpha - \frac{1}{2}|).
\end{aligned}$$

One checks that the polynomial $q(x) = 2^{-10} + 2^{-8}x - 2^{-3}x^2 - x^3$ is positive (in fact, at least 0.00029) for $x \in [0, 0.07]$. But $0 \leq T_4 \leq 2^{-5} + O(|\alpha - \frac{1}{2}|)$, and so for c_2 chosen small enough we have that $T_4 \in [0, 0.07]$. Finally note that $T_4 \geq ||\hat{g}||_\infty^4$. This completes the proof. □

We note that explicit values for the constants c_0, c_1, c_2, c_3 and C_4 could be computed if one so wished. As a consequence, one could extract from the following proof an explicit l_0 such that for all $l \geq l_0$, $\Phi^{(l)}$ is common.

Proof of Theorem 6.1.4. Recalling (6.2), we wish to show that for l sufficiently large

$$\alpha^l T_\Phi(f) + (1-\alpha)^l T_\Phi(1-f) \geq 2^{-l-8},$$

where $\alpha := \mathbb{E}f$. Let $g := f - \alpha$. We split into three cases: the first is when $|\alpha - \frac{1}{2}|$ is suitably small and $||\hat{g}||_\infty$ is suitably small; the second is when $|\alpha - \frac{1}{2}|$ is suitably small and $||\hat{g}||_\infty$ is large; the third is when $|\alpha - \frac{1}{2}|$ is large. In this proof, all asymptotic notation is with respect to l.

Firstly, if $\alpha \in [\frac{1}{3}, \frac{2}{3}]$ and $||\hat{g}||_\infty \leq c_1$, then from Lemma 6.3.3 applied to both f and $1-f$ we have

$$\alpha^l T(f) + (1-\alpha)^l T(1-f) \geq \alpha^{l+9} + (1-\alpha)^{l+9} \geq 2^{-l-8},$$

where we use convexity for the second inequality.

For the second case, let c_5 be a constant depending on c_1, c_3 to be determined shortly, and assume that $|\alpha - \frac{1}{2}| \leq \frac{c_5}{\sqrt{l}}$ and that $||\hat{g}||_\infty > c_1$. By the AM-GM inequality, we have

$$\alpha^l T(f) + (1-\alpha)^l T(1-f) \geq 2(\alpha(1-\alpha))^{l/2} \sqrt{T(f)T(1-f)}.$$

Invoking Lemma 6.3.4 by taking l suitably large so $\frac{c_5}{\sqrt{l}} \leq c_2$, we have

$$T(f)T(1-f) \geq 2^{-18} + c_3 c_1^4 - \frac{C_4 c_5}{\sqrt{l}} \geq 2^{-18} + \frac{1}{2} c_3 c_1^4,$$

where we have again taken l suitably large for the second inequality. Continuing from above, there is a positive constant $c_6 > 0$ which may be computed explicitly in terms c_1, c_3 such that

$$\begin{aligned}
\alpha^l T(f) + (1-\alpha)^l T(1-f) &\geq (\alpha(1-\alpha))^{l/2} (2^{-8} + c_6) \\
&= \left(\frac{1}{4} - |\alpha - \frac{1}{2}|^2\right)^{l/2} (2^{-8} + c_6) \\
&\geq \left(1 - \frac{4c_5^2}{l}\right)^{l/2} 2^{-l} (2^{-8} + c_6) \\
&= \left(1 - \frac{4c_5^2}{l}\right)^{l/2} (1 + 2^8 c_6) 2^{-l-8}.
\end{aligned}$$

For c_5 fixed, $\left(1-\frac{4c_5^2}{l}\right)^{l/2}$ converges to $e^{-2c_5^2}$ as $l\to\infty$. Let c_5 be small enough so that $e^{-2c_5^2}(1+2^8c_6)>1$ and then let l be large enough so that $\left(1-\frac{4c_5^2}{l}\right)^{l/2}(1+2^8c_6)\geq 1$. This completes the proof in this case.

Finally, assume that $|\alpha-\frac{1}{2}|>\frac{c_5}{\sqrt{l}}$. Assume with out loss of generality that $\alpha\geq\frac{1}{2}$. Then we have by Lemma 6.3.2 that

$$\alpha^l T(f)+(1-\alpha)^l T(1-f)\geq \alpha^l T(f)\geq c_0\left(\frac{1}{2}+\frac{c_5}{\sqrt{l}}\right)^l = 2^{-l}c_0\left(1+\frac{2c_5}{\sqrt{l}}\right)^l.$$

Taking l sufficiently large, this is greater than 2^{-l-8}, which completes the proof. □

6.4 Geometric commonness

In the previous section we obtained a lower bound on the geometric mean of $T(f)$ and $T(1-f)$ in order to conclude that Φ is Alon. In this section we will introduce the notion of geometric commonness and see in Proposition 6.4.5 that a bound of this form is in fact necessary to obtain the Alon property.

Definition 6.4.1. A system of t linear forms Ψ is *geometrically common* if for all $n\geq 1$ and all $f:\mathbb{F}_p^n\to[0,1]$ with $\mathbb{E}_x f(x)=1/2$,

$$T_\Psi(f)T_\Psi(1-f)\geq 2^{-2t}.$$

We recall that a system of equations in t variables is Sidorenko if $T_\Psi(f)\geq(\mathbb{E}_x f(x))^t$ for all $f:\mathbb{F}_p^n\to[0,1]$. It is clear that if Ψ is Sidorenko then it is geometrically common. Furthermore the AM-GM inequality yields that if Ψ is geometrically common then it is common among functions with density $1/2$ (though as we have seen this does not imply commonness in general).

Definition 6.4.2. A system of t linear forms Ψ is *weakly Alon* if for all $\varepsilon>0$, for all sufficiently large $l\geq l_\varepsilon$, for all $n\geq 1$ and for all $f:\mathbb{F}_p^n\to[0,1]$,

$$T_{\Psi^{(l)}}(f)+T_{\Psi^{(l)}}(1-f)>2^{-l}\left(2^{1-t}-\varepsilon\right).$$

Conversely, the condition that a system is not weakly Alon asks that the sequence of systems $\Psi^{(l)}$ is not common infinitely often by a suitably uniform margin.

Definition 6.4.3. Let $\alpha\in(0,1)$. A system of linear forms Ψ is *α-prevalent* if there exists $c(\alpha)>0$ such that for all n, for every $f:\mathbb{F}_p^n\to[0,1]$ with $\mathbb{E}f=\alpha$, we have $T_\Psi(f)\geq c(\alpha)$.

We note that prevalence is monotonic in α: if $\alpha < \beta$ and Ψ is α-prevalent then it is β-prevalent.

Remark 6.4.4. By (a generalised) Szemerédi's theorem, and Varnavides' strengthening thereof, all translation-invariant systems are α-prevalent for all $\alpha \in (0,1)$. It is not difficult to see that the converse holds too, so α-prevalence for all $\alpha \in (0,1)$ is equivalent to translation-invariance. Schur triples are $(\frac{1}{3}+\varepsilon)$-prevalent but not $\frac{1}{3}$-prevalent if $p=3$; $(\frac{2}{5}+\varepsilon)$-prevalent but not $\frac{2}{5}$-prevalent if $p=5$; $(\frac{2}{7}+\varepsilon)$-prevalent but not $\frac{2}{7}$-prevalent when $p=7$. See [34], [32] for the results pertaining to Schur triples. We showed in Lemma 6.3.2 that Φ is 0.45-prevalent for odd p. On the other hand, if $p=3$, then Φ is not $\frac{1}{3}$-prevalent since it is not translation-invariant.

We will only prove the reverse implication of the below proposition. The forward direction is essentially a union of the arguments in the second and third cases of the proof of Theorem 6.1.4 and so we omit the details. We note that, for example, Φ is $\frac{1}{2}$-prevalent.

Proposition 6.4.5. *A $\frac{1}{2}$-prevalent system of linear forms Ψ is geometrically common if and only if it is weakly Alon.*

Proof. Suppose Ψ is not geometrically common and let n be such that there exists $f:\mathbb{F}_p^n \to [0,1]$ with $\mathbb{E}f = 1/2$ and $T(f)T(1-f) < 2^{-2t}$. If $T(f) = T(1-f)$ then we have immediately that $T(f)+T(1-f) < 2^{-t+1}$, and indeed there is $\varepsilon > 0$ such that $T(f)+T(1-f) \le 2^{-t+1}-\varepsilon$. Then by (6.2),

$$T_{\Psi^{(l)}}(f) + T_{\Psi^{(l)}}(1-f) = 2^{-l}\left(T(f)+T(1-f)\right) \le 2^{-l}(2^{1-t}-\varepsilon),$$

for all $l \ge 1$. Thus Ψ is not weakly Alon in this case.

Assume by symmetry that $T(1-f) > T(f)$. Define $c := \log\sqrt{T(1-f)/T(f)} > 0$. By the $\frac{1}{2}$-prevalence of Ψ, we have that c is bounded above uniformly in n, f. Now let $l \ge \frac{45c}{4}$ be a large integer to be determined more precisely later. Let $S = \{x : f(x) \le \frac{9}{10}\}$; since $\mathbb{E}f = 1/2$, we have by Markov's inequality that $|S| \ge \frac{4p^n}{9}$. Let $g:\mathbb{F}_p^n \to [0,1]$ take value $\frac{p^n}{|S|}\cdot\frac{c}{2l} \le \frac{1}{10}$ on S and be zero otherwise. Then we have that $\mathbb{E}_x g(x) = \frac{c}{2l}$ and that $f+g$ takes values in $[0,1]$. We claim that $f+g$ demonstrates that Ψ is not weakly Alon.

Note that for functions $f_i:\mathbb{F}_p^n \to [0,1]$ and nonzero linear forms $(\psi_i)_{i=1}^t$ we have that $\mathbb{E}_{\boldsymbol{x}} f_1(\psi_1(\boldsymbol{x}))\cdots f_t(\psi_t(\boldsymbol{x})) \le \mathbb{E}_x f_i(x)$ for any i, and so by the multilinearity of

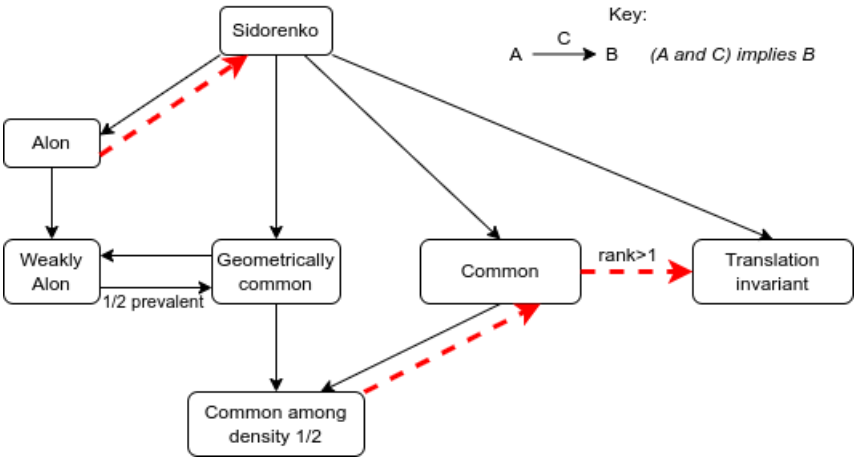


Figure 6.1: Implications between various notions. Red arrows, from left to right, correspond to answers to Questions 6.1.2, 6.1.7, 6.1.5, which show that the corresponding implications do not hold.

the operator T we have that $T(f+g) = T(f) + O(1/l)$. Similarly, $T(1-(f+g)) = T(1-f) + O(1/l)$. Thus,

$$\begin{aligned} T_{\Psi^{(l)}}(f+g) + T_{\Psi^{(l)}}(1-(f+g)) &= \left(\frac{1}{2} + \frac{c}{2l}\right)^l T(f+g) + \left(\frac{1}{2} - \frac{c}{2l}\right)^l T(1-(f+g)) \\ &= \frac{1}{2^l}\left((1+c/l)^l T(f) + (1-c/l)^l T(1-f) + O(1/l)\right) \\ &= \frac{1}{2^l}\left(e^c T(f) + e^{-c} T(1-f) + O(1/l)\right) \\ &= \frac{1}{2^l}\left(2\sqrt{T(f)T(1-f)} + O(1/l)\right). \end{aligned}$$

Now we have that $\sqrt{T(f)T(1-f)} < 1/2^{2t}$ and so there exists $\varepsilon > 0$ such that for l sufficiently large we have $T_{\Psi^{(l)}}(f+g) + T_{\Psi^{(l)}}(1-(f+g)) < 2^{-l}(2^{1-t} - \varepsilon)$. Thus we conclude that Ψ is not weakly Alon.

The proof of the reverse implication is omitted. □

Remark 6.4.6. It is clear that if Ψ is not $\frac{1}{2}$-prevalent then it is not geometrically common. We have not investigated whether the (weak) Alon property implies $\frac{1}{2}$-prevalence.

We conclude with a picture showing the relationships between the various notions discussed in this chapter.

Chapter 7

Local aspects of the Sidorenko property for linear systems

7.1 Introduction

7.1.1 Context

In graph theory, Sidorenko's conjecture [61] predicts that if H is a bipartite graph, then among all graphs G with n vertices and fixed average degree, the number of copies of H in G is asymptotically (as $|V(G)| \to \infty$) minimised when G is random. We will call graphs H which satisfy this property *Sidorenko*. Sidorenko's conjecture has been resolved for certain classes of bipartite graph (for example trees, cycles, complete bipartite graphs [60], see also [14, 43, 54, 63, 49, 16, 15, 17]) though the conjecture in general remains open, and is indeed of great interest. There is a related colouring property which is also of interest: for which graphs H is it true that, among all two-colourings of the edges of K_n, the number of monochromatic copies of H is minimised by a random two-colouring (asymptotically as $n \to \infty$)? Such graphs H are called *common*. It is not difficult to see that if a graph satisfies the Sidorenko property, then it is common. Goodman showed that K_3 is common [25]. Erdős [20] subsequently conjectured that K_4 is common and Burr–Rosta [11] that every graph is common. Sidorenko [62] showed that a triangle with a pendant edge is not common and Thomason [67] showed that in fact K_4 is not common, together disproving both conjectures. In fact, any graph which contains K_4 is uncommon [46]. It is not difficult to see that if H is *not* bipartite then it is not Sidorenko, and so we have a conjectural classification of all graphs. Even a reasonable conjectured classification of common graphs seems out of reach at the moment.

There has been recent interest in obtaining analogous conjectures and results in arithmetic contexts, initiated by Saad and Wolf [59]. Which systems of linear forms

Ψ have the property that, among all subsets S of $\mathbb{F}_p^n$ of fixed size, the number of solutions to Ψ in S is minimised when S is chosen randomly? Such systems are called Sidorenko. Those systems with the property that random two-colourings minimise the number of monochromatic solutions among all two-colourings are called common. Again, Sidorenko implies common. In the arithmetic setting, neither property has a conjectured classification.

One situation in which a classification is known is when the system of linear forms has codimension one (i.e. its image is described by a single equation). Here, an equation is Sidorenko if [59] and only if [21] its coefficients may be partitioned into pairs which sum to zero, and is common and not Sidorenko if and only if it has an odd number of variables. For systems of higher codimension, one makes the trivial observation that the union of two Sidorenko systems on disjoint variable sets is Sidorenko. Kamčev, Liebenau and Morrison give a nontrivial method by which higher codimension Sidorenko systems can be built from Sidorenko equations; see [48, Theorem 1.3]. There are also a number of results in the negative direction [47, 48, 59, 68]. Finally, we note that Král'–Lamaison–Pach have reported on a partial classification of common systems comprising two equations in $\mathbb{F}_2^n$ [50] and have since obtained a near-complete classification of these systems, which is due to appear in forthcoming work (personal communications).

Lovász [55] showed that the graph-theoretic Sidorenko conjecture is locally true: for all bipartite graphs H, there is no suitably small perturbation of the random graph G which increases the number of copies of H in G. He subsequently classified locally Sidorenko graphs: a graph H is locally Sidorenko if and only if it is a forest or has even girth [56, Theorem 16.26] (see also [22] to this end). The study of local notions of commonness has also attracted recent interest: [19, 42, 56]. One of the purposes of this article is to introduce these local notions to the arithmetic setting. Finally, to further motivate the study of these local properties, we note that in the previous chapter and in [6] we have used a local Sidorenko property of a particular linear system as a key ingredient in the main result there (see, in particular, Lemma 6.3.3).

7.1.2 Results

After Fox–Pham–Zhao [21] completed the classification of codimension one systems, the next natural goal ought perhaps to be a classification of codimension two systems.

Codimension one systems are Sidorenko/common if and only if they are locally Sidorenko/common; local notions are only interesting when the codimension is at least

two. We study two different notions of locality. Loosely speaking, the weaker notion ('weakly locally Sidorenko/common') allows the radius of perturbation to depend on the function by which we perturb, and the stronger notion ('locally Sidorenko/common') asks that the radius is uniform in the choice of function.

One of our main theorems shows that a certain generic family of systems of codimension 2 is not Sidorenko, and (nearly) classifies this family by whether they are weakly locally Sidorenko or locally Sidorenko. We direct the reader to Subsection 7.2.1 and in particular Definition 7.2.2 and Definition 7.2.3 for formal definitions of the weak local Sidorenko property and the local Sidorenko property. We note that Sidorenko implies locally Sidorenko, which in turn implies weakly locally Sidorenko.

Let p be an odd prime and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of linear forms, each mapping $\mathbb{F}_p^D$ to $\mathbb{F}_p$, whose image is determined by a $2 \times t$ system of linear equations with coefficients

$$M := \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1t} \\ b_{21} & b_{22} & \cdots & b_{2t} \end{pmatrix}.$$

We will call Ψ *linearly generic* if every 2×2 minor of M has nonzero determinant.[1] An *additive k-tuple* is a tuple $(x_1, x_2, \dots, x_k)$ satisfying the equation $x_1 - x_2 + x_3 - x_4 + \cdots + (-1)^{k+1} x_k = 0$.

Theorem 7.1.1. *Let Ψ be a codimension two system of linear equations in t variables which is linearly generic. Then Ψ is not locally Sidorenko (and so not Sidorenko). If t is even then Ψ is not weakly locally Sidorenko. Furthermore, for p sufficiently large in terms of t, if t is odd and Ψ is weakly locally Sidorenko, then Ψ contains an additive $(t-1)$-tuple.*

Thus, in the situation that p is large in terms of t, we have classified all linearly generic systems by whether they are Sidorenko, locally Sidorenko, or weakly locally Sidorenko, except for those systems with t odd and which contain an additive $(t-1)$-tuple. We note that the condition of containing a $(t-1)$-tuple is more than a reflection of the limitation of our methods: there does exist a linearly generic system with t odd, which contains an additive $(t-1)$-tuple, and which is weakly locally Sidorenko (see Example 7.4.2).

As a consequence of Theorem 7.1.1, we answer the following conjecture and question of Kamčev–Liebenau–Morrison.

[1] We note here that in [48] the same class of systems is considered, albeit in different language. A reader wishing to compare the texts may find it helpful to note that our linearly generic condition is equivalent to the condition that a system has codimension two and (in the notation of [48]) $s(L) = t - 1$.

Conjecture 7.1.2 ([48, Conjecture 5.1]). *Let Ψ be a system of 5 linear forms whose image has codimension 2. If Ψ is linearly generic then Ψ is not Sidorenko.*

Question 7.1.3 ([48, Question 5.3]). *Let $t \geq 7$ be odd. Does there exists a system of t linear forms Ψ whose image has codimension 2, which is linearly generic and which is Sidorenko?*

We note that the above conjecture and question as they appear in [48] do not refer directly to the linearly generic condition, but the formulation there is equivalent. We have the following corollary of Theorem 7.1.1.

Corollary 7.1.4. *Conjecture 7.1.2 is true and the answer to Question 7.1.3 is no.*

Define the *complexity* of $\Psi = (\psi_1, \dots, \psi_t)$, where each ψ_i is a linear map from $\mathbb{F}_p^D$ to $\mathbb{F}_p$, to be the smallest positive integer s such that the functions $(\psi_i^s)_{i=1}^t$ are linearly independent over $\mathbb{F}_p$. Loosely speaking, if the complexity of a system is 1, then its analysis should be controlled by Fourier analysis, and one ought not require the higher-order theory. We direct the reader to [28, 29, 30], for relevant definitions, discussions and results to this end. One may easily compute that for $t \geq 5$ and p sufficiently large, systems whose images have codimension two generically have complexity one. Where it is not difficult, using higher-order constructions, to contrive systems of complexity at least 2 which are weakly locally Sidorenko but not Sidorenko, the situation is less clear for complexity one systems (we will justify this claim in Subsection 7.4.2). We believe therefore that the following corollary of Theorem 7.1.1 is also worth noting.

Corollary 7.1.5. *Among complexity one systems, the weak local Sidorenko property is not equivalent to the local Sidorenko property.*

The explicit example demonstrating Corollary 7.1.5 is Example 7.4.2 in Section 7.4.

Of independent interest is the observation and method underlying Theorem 7.1.1 and its corollaries. We observe that although these systems have complexity one, their 'dual' system (i.e. the linear system obtained in frequency space after Fourier inversion) need not have complexity one. The most involved part of Theorem 7.1.1 utilises higher-order constructions in the frequency space.

We also provide the first example of a system which is locally Sidorenko but not Sidorenko; see Example 7.4.1. (This is the same example as in the previous chapter, though there we prove a local Sidorenko property with respect to the ℓ^∞ norm on the Fourier side.)

Proposition 7.1.6. *In the arithmetic setting, the Sidorenko property is not equivalent to the local Sidorenko property.*

This result is perhaps slightly less trivial than it sounds; for example in the graph-theoretic setting, it remains an open question whether local commonness (appropriately defined, see [19]) is equivalent to commonness. One of the reasons we have chosen to study these local properties is that all existing proofs in the literature that systems are not Sidorenko [21, 48, 59, 68] prove the stronger statement that in fact these systems are not locally Sidorenko.

Finally, we note that the linear genericity condition in Theorem 7.1.1 is necessary. For example, it has been observed by Kamčev–Liebenau–Morrison [48] that the system

$$x_1 + 2x_2 - x_3 - 2x_4 = 0$$
$$3x_1 + x_2 - 3x_3 - x_4 = 0$$

is Sidorenko, and it clearly does not contain an additive tuple.

In another direction, it has been known for some time [46, Theorem 12] that any graph containing K_4 is uncommon. Saad and Wolf [59] asked whether the same is true for four-term arithmetic progressions in the arithmetic context. This question was answered in the affirmative recently by Kamčev–Liebenau–Morrison [47] and independently by Versteegen [68]. Both proofs build on a 'quadratic' construction of Gowers [27]. In Section 7.5, we give (together with a lemma from [47]) a short proof of Saad and Wolf's conjecture which does not rely on higher-order constructions. Our proof is also more amenable to generalisation to rank two systems in $2k$ variables, which would be a significant step in the classification of weakly locally common systems[2].

We set up some notational and technical preliminaries in Section 7.2. We prove Theorem 7.1.1 in Section 7.3. In Section 7.4 we give examples to prove Corollary 7.1.5 and Proposition 7.1.6. Finally in Section 7.5 we give a new proof of the answer to Saad and Wolf's question [59, Question 3.6].

[2]We do not consider weak local commonness in this chapter and so we do not give this concept a formal definition here. However, we hope that the enthusiastic reader may formulate a definition from that of the weak local Sidorenko property Definition 7.2.3 and the common property Definition 7.2.4.

7.2 Preliminaries

7.2.1 Definitions, basic properties

Throughout we let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of $\mathbb{F}_p$-linear forms, each mapping $\mathbb{F}_p^D$ to $\mathbb{F}_p$; we often abuse notation and let them induce linear maps from $(\mathbb{F}_p^n)^D$ to $\mathbb{F}_p^n$ in the natural way. Define the operator T_Ψ acting on functions $f : \mathbb{F}_p^n \to \mathbb{C}$ by the formula

$$T_\Psi(f) := \mathbb{E}_{x_1,\dots,x_D \in \mathbb{F}_p^n} f(\psi_1(x_1, \dots, x_D)) \cdots f(\psi_t(x_1, \dots, x_D)),$$

where throughout $\mathbb{E}$ is a normalised count, so here is shorthand for $\frac{1}{p^{nD}} \sum$. We will always assume that $D \leq t$ and that $\Psi^{-1}(0) = 0$ (it is clear that one may reduce to this case). At times it will be more convenience to describe a system of linear forms by the equations which define its image. Suppose that M_Ψ is an $\mathbb{F}_p$-valued matrix such that $\operatorname{Im} \Psi = \ker M_\Psi$. Then we have

$$T_\Psi(f) = \mathbb{E}_{\boldsymbol{x} \in \ker M_\Psi} f(x_1) \cdots f(x_t),$$

where $\boldsymbol{x} = (x_1, \dots, x_t) \in (\mathbb{F}_p^n)^t$.

Throughout this chapter we will deal with functional versions of the Sidorenko and common properties in which we ask the following of all functions $f : \mathbb{F}_p^n \to [0, 1]$, rather than just restricting to characteristic functions of sets.

Definition 7.2.1. A system of t linear forms Ψ is *Sidorenko* if, for all $n \geq 1$ and all $f : \mathbb{F}_p^n \to [0, 1]$, we have

$$T_\Psi(f) \geq (\mathbb{E}_{x \in \mathbb{F}_p^n} f(x))^t.$$

Note that any function can be written as the sum of a constant function and a function with zero average, whence the inequality in Definition 7.2.1 can be written as $T_\Psi(\alpha + f) \geq \alpha^t$, where the condition must hold for all $\alpha \in [0, 1]$ and all $f : \mathbb{F}_p^n \to [-\alpha, 1 - \alpha]$ with $\mathbb{E}_x f(x) = 0$. A system is locally Sidorenko if the above inequality is satisfied for all sufficiently small perturbations of the constant function.

Definition 7.2.2. A system of t linear forms Ψ is *locally Sidorenko* if for all $\alpha \in [0, 1]$ there exists $\varepsilon > 0$ such that for all $n \geq 1$ and all $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}_x f(x) = 0$, we have

$$T_\Psi(\alpha + \varepsilon f) \geq \alpha^t.$$

We remark that that this notion of locality considers small perturbations of the constant function with respect to the ℓ^∞ norm on the space of functions. One may

also consider analogous notions of locality for different norms. We do not do so in this chapter, but note that for example considering the local Sidorenko property with respect to the ℓ^∞ norm on the Fourier side was an important part of the main result of Chapter 6 (see Lemma 6.3.3).

Above, the size of the neighbourhood of the constant function is fixed independently of n, f. If the size of the neighbourhood is allowed to vary with these parameters, then we will say that the linear system is weakly locally Sidorenko.

Definition 7.2.3. A system of t linear forms Ψ is *weakly locally Sidorenko* if for all $n \geq 1$, for all $\alpha \in [0,1]$ and all $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}_x f(x) = 0$, there exists $\varepsilon_f > 0$ such that for all $\varepsilon < \varepsilon_f$ we have

$$T_\Psi(\alpha + \varepsilon f) \geq \alpha^t.$$

Definition 7.2.4. A system of t linear forms Ψ is *common* if, for all $n \geq 1$ and all $f : \mathbb{F}_p^n \to [0,1]$, we have

$$T_\Psi(f) + T_\Psi(1-f) \geq 2^{1-t}.$$

Local commonness and weak local commonness may be defined analogously.

Let $\Psi = (\psi_1, \ldots, \psi_t)$ be a system of t linear forms. Let α be a constant and f a function on $\mathbb{F}_p^n$ with $\mathbb{E}_x f(x) = 0$. For a subset $S \subset [t]$, we denote the corresponding subsystem of Ψ by $\Psi(S)$. Then, by the multilinearity of the operator T we have

$$T_\Psi(\alpha + f) = \sum_{S \subset [t]} \alpha^{t-|S|} T_{\Psi(S)}(f). \tag{7.1}$$

Note that if f has $\mathbb{E}_x f(x) = 0$ and if the image of $\Psi(S)$ contains a coordinate hyperplane then $T_{\Psi(S)}(f) = 0$. It follows from the linear genericity condition (recall: every 2×2 minor of the matrix defining its equations has nonzero determinant), then for all nonempty $S \subset [t]$ with $|S| \leq t-2$, the subsystem $\Psi(S)$ contains a coordinate hyperplane and so yields $T_{\Psi(S)}(f) = 0$. Thus for linearly generic systems we have

$$T_\Psi(\alpha + f) = \alpha^t + \alpha \sum_{|S|=t-1} T_{\Psi(S)}(f) + T_\Psi(f). \tag{7.2}$$

Furthermore, for such systems, if $|S| = t-1$ then the image of $\Psi(S)$ has codimension one (i.e. its image is described by a single equation).

7.2.2 Fourier inversion

Recall the Fourier transform defined by

$$\hat{f}(h) = \mathbb{E}_{x \in \mathbb{F}_p^n} f(x) e_p(-h \cdot x),$$

where $e_p(\cdot) := e^{\frac{2\pi i \cdot}{p}}$. Recall also the Fourier inversion formula:

$$f(x) = \sum_{h \in \mathbb{F}_p^n} \hat{f}(h) e_p(h \cdot x),$$

and Parseval's identity

$$\mathbb{E}_x |f(x)|^2 = \sum_h |\hat{f}(h)|^2.$$

For a linear system Ψ, let $M = M_\Psi$ have $\operatorname{Im} \Psi = \ker M$. We will make use of the following consequence of the Fourier inversion formula:

$$T_\Psi(f) = \sum_{\boldsymbol{h} \in \operatorname{Im} M^T} \hat{f}(h_1) \cdots \hat{f}(h_t), \tag{7.3}$$

where $\boldsymbol{h} = (h_1, \ldots, h_t) \in (\mathbb{F}_p^n)^t$.

7.2.3 Tensoring

Definition 7.2.5. Let $f_1 : \mathbb{F}_p^{n_1} \to \mathbb{C}$ and $f_2 : \mathbb{F}_p^{n_2} \to \mathbb{C}$. Then the tensor product $f := f_1 \otimes f_2 : \mathbb{F}_p^{n_1+n_2} \to \mathbb{C}$ is the function defined by $f(x) = f_1(\pi_1(x)) f_2(\pi_2(x))$, where π_1 is the projection onto the first n_1 coordinates of $\mathbb{F}_p^{n_1+n_2}$ and π_2 is the projection onto the last n_2 coordinates.

It is clear that $\otimes$ is an associative operation and so one may naturally define the tensor product of a finite collection of functions.

Observation 7.2.6. For a system of linear forms Ψ, we have

$$T_\Psi(f_1 \otimes f_2) = T_\Psi(f_1) T_\Psi(f_2). \tag{7.4}$$

Note that there is a slight abuse of notation here: we interpret T_Ψ as an operator on $\mathbb{F}_p^{n_1+n_2}$, $\mathbb{F}_p^{n_1}$ and $\mathbb{F}_p^{n_2}$ as it appears respectively. The equation above is easily verified by direct computation.

7.3 Proof of Theorem 7.1.1

Recall that we call a system of two equations linearly generic if the $2 \times t$ matrix

$$M := \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1t} \\ b_{21} & b_{22} & \cdots & b_{2t} \end{pmatrix}$$

which defines the equations has the property that all of its 2×2 minors have nonzero determinant.

7.3.1 Systems without additive tuples are not weakly locally Sidorenko

We first work towards proving the following theorem, which is one of the statements in Theorem 7.1.1.

Theorem 7.3.1. *Let Ψ be a system of two linear equations in t variables which is linearly generic. If t is even or if Ψ does not contain an additive tuple and p is sufficiently large in terms of t, then Ψ is not weakly locally Sidorenko.*

We now prove the theorem, modulo the proof of two propositions which will appear subsequently and occupy the remainder of the subsection. In both propositions, we prove something more general than what is needed for the application in this chapter. We do so to illustrate the scope of the constructions and because it is clear that these more general statements can be used to obtain more general results. See also Remark 7.3.4.

Proof of Theorem 7.3.1. Recall from Equation (7.2) that we have

$$T_\Psi(\alpha + \varepsilon f) = \alpha^t + \alpha\varepsilon^{t-1} \sum_{|S|=t-1} T_{\Psi(S)}(f) + \varepsilon^t T_\Psi(f).$$

It suffices to show that there exists f with $\mathbb{E}f = 0$ such that $\sum_{|S|=t-1} T_{\Psi(S)}(f) < 0$; then one makes ε appropriately small to conclude. If $t-1$ is odd, then this is trivial: take f for which $\sum_{|S|=t-1} T_{\Psi(S)}(f) \neq 0$ (it is not difficult to see that such an f must exist), and if $\sum_{|S|=t-1} T_{\Psi(S)}(f) > 0$ then replace f with $-f$. Henceforth we assume that t is odd.

Our strategy will be to find, for each $S \subset [t]$ with $|S| = t-1$, a function f_S with $\mathbb{E}f_S = 0$ such that $T_{\Psi(S)}(f_S) = 0$ and $T_\Psi(f_S) \neq 0$. Then we let $f = \bigotimes_{|S|=t-1} f_S$ and by (7.4) we see that

$$T_\Psi(\alpha + \varepsilon f) = \alpha^t + \varepsilon^t T_\Psi(f),$$

where $T_\Psi(f) = \prod_{|S|=t-1} T_\Psi(f_S) \neq 0$. Recall that t is odd so potentially replacing f with $-f$ depending on the sign of $T_\Psi(f)$, we may conclude that for all ε sufficiently small (indeed, for all ε), we have $T_\Psi(\alpha+\varepsilon f) < \alpha^t$, so Ψ is not weakly locally Sidorenko.

It remains to prove the existence of the functions f_S. Recall that each of the systems $\Psi(S)$ has codimension one, so is described by a single equation. This equation cannot be of the form $\sum_i (-1)^i x_i = 0$ by our assumption that Ψ does not contain an additive tuple. We prove the existence of such an f_S in Proposition 7.3.2 and Proposition 7.3.3. The first deals with the easier case in which there is a pair of coefficients whose ratio is not equal to ± 1, and the second deals with the remaining cases which are not additive tuples. □

Proposition 7.3.2. *Let Ψ be a codimension 2 linear system in t variables which is linearly generic. Let Φ be determined by a single equation $\sum_{i=1}^{l} a_i x_i = 0$, where each a_i is nonzero, such that there are i, j satisfying $a_i \neq \pm a_j$. Then for all $n \geq 2$, there exists $f : \mathbb{F}_p^n \to [-1, 1]$ with $\mathbb{E}_x f(x) = 0$, $T_\Phi(f) = 0$ and $|T_\Psi(f)| \gg_t 1$.*

Proof. Without loss of generality a_1, a_2 are such that $a_1 \neq \pm a_2$. We construct f from its Fourier coefficients. Let $M = M_\Psi = (b_{ij})$ be the $2 \times t$ matrix whose coefficients correspond to the two linear equations of Ψ. Using (7.3) we have that

$$T_\Phi(f) = \sum_{h \in \mathbb{F}_p^n} \prod_{i=1}^{l} \hat{f}(a_i h), \tag{7.5}$$

and

$$T_\Psi(f) = \sum_{h_1, h_2} \prod_{i=1}^{t} \hat{f}(b_{1i} h_1 + b_{2i} h_2). \tag{7.6}$$

We claim that if h_1, h_2 are linearly independent, then $(u_1, \cdots, u_t) := M^T \binom{h_1}{h_2}$ satisfies $u_i \neq 0$ for all i and $a_1 u_i \neq \pm a_2 u_j$ for all i, j. Indeed, that $u_i \neq 0$ follows from linear independence and the fact that M cannot have a zero column, and if $a_1 u_i = \pm a_2 u_j$, then by the linear independence of h_1, h_2 we have $a_1 \binom{b_{1i}}{b_{2i}} = \pm a_2 \binom{b_{1j}}{b_{2j}}$ which contradicts the fact that all 2×2 minors of M are full rank.

Choose any two linearly independent h_1, h_2 and define $u_1, \ldots, u_t$ as above. Now define $\hat{f}(\pm u_i) = 1/(2t)$ for all i and $\hat{f}(h) = 0$ otherwise so that f is clearly real-valued and takes values in $[-1, 1]$ by Fourier inversion and the triangle inequality. Since $u_i \neq 0$ for all i we have $\hat{f}(0) = \mathbb{E}_x f(x) = 0$. Next, since $a_1 \neq \pm a_2$ and since $a_1 u_i \neq \pm a_2 u_j$ for all i, j we have that at least one of $\hat{f}(a_1 h), \hat{f}(a_2 h)$ is zero for all h and so $T_\Phi(f) = 0$ from (7.5). Finally, $T_\Psi(f) \geq 2/(2t)^t$ from (7.6). □

In the following proposition we prove a more general statement than what is needed for the proof of Theorem 7.3.1. Below we consider equations of arbitrary length with ± 1 coefficients. For the theorem above, we just need to consider equations $\sum_{i=1}^{t-1} a_i x_i = 0$ of length $t-1$ (which we recall may be assumed to be even), where $a_i \in \pm 1$. In this case, since $t-1$ is even, we have for parity reasons that $|\#\{a_i = 1\} - \#\{a_i = -1\}|$ must be even, and so the condition that $t \neq (2m+1)|\#\{a_i = 1\} - \#\{a_i = -1\}|$ for all $m \in \mathbb{Z}$ from the statement of the proposition below is automatically satisfied.

Proposition 7.3.3. *Let Ψ be a codimension 2 linear system in t variables which is linearly generic. Let Φ be determined by the equation $\sum_{i=1} a_i x_i = 0$, where each $a_i \in \{-1, 1\}$ and $\#\{a_i = 1\} \neq \#\{a_i = -1\}$. If p is sufficiently large in terms of t and $t \neq (2m+1)|\#\{a_i = 1\} - \#\{a_i = -1\}|$ for any integer m, then there exists $f : \mathbb{F}_p \to [-1, 1]$ with $\mathbb{E}_x f(x) = 0$, with $T_\Phi(f) = 0$ and with $|T_\Psi(f)| \gg_{p,t} 1$.*

Proof. Let $V = \operatorname{Im} M_\Psi^T \leq \mathbb{F}_p^t$. For $i = 1, \dots, t$, let P_i be the ith coordinate hyperplane (i.e. $\{x \in \mathbb{F}_p^t : x_i = 0\}$). Let the symmetric group on t elements $\mathrm{Sym}(t)$ act on $\mathbb{F}_p^t$ by coordinate permutation, and let $(\mathbb{Z}/2\mathbb{Z})^t$ act on $\mathbb{F}_p^t$ by reflection in each coordinate hyperplane (so for example $(1, 0, 1, 0, \dots) \in (\mathbb{Z}/2\mathbb{Z})^t$ maps $(u_1, u_2, u_3, u_4, \dots)$ to $(-u_1, u_2, -u_3, u_4, \dots)$). Let G be the group of permutations of $\mathbb{F}_p^t$ generated by $\mathrm{Sym}(t)$ and $(\mathbb{Z}/2\mathbb{Z})^t$ and let $S = G - \{\mathrm{id}, (1, 1, \dots, 1)\}$.

We claim that $V - \left(\bigcup_{i=1}^t P_i \cup \bigcup_{\sigma \in S} \sigma V\right)$ is non-empty. Indeed,

$$\left| V - \left(\bigcup_{i=1}^t P_i \cup \bigcup_{\sigma \in S} \sigma V \right) \right| \geq |V| - \sum_{i=1}^t |V \cap P_i| - \sum_{\sigma \in S} |V \cap \sigma V|,$$

where each $V \cap P_i$ and $V \cap \sigma V$ is a subspace of V, so has size $1, p$ or p^2. By taking p sufficiently large in terms of t, it suffices to show that $V \neq V \cap P_i$ and $V \neq \sigma V$ for each $i \in [t]$ and $\sigma \in S$. That $V \neq V \cap P_i$ follows from the fact that M cannot have a zero column, and that $V \neq \sigma V$ for every $\sigma \in S$ is an easy exercise in linear algebra, using the fact that each 2×2 minor of M has full rank and that every element of G may be written in the form $\sigma\tau$ for some $\sigma \in \mathrm{Sym}(t)$ and $\tau \in (\mathbb{Z}/2\mathbb{Z})^t$.

Let $u := (u_1, \dots, u_t) \in V - \left(\bigcup_{i=1}^t P_i \cup \bigcup_{\sigma \in S} \sigma V\right)$. We may assume that $\#\{a_i = 1\} > \#\{a_i = -1\}$ so that by Fourier inverting

$$T_\Phi(f) = \sum_{h \in \mathbb{F}_p} |\hat{f}(h)|^{2k} \hat{f}(h)^l, \tag{7.7}$$

where $k \geq 0$, $= \#\{a_i = 1\} - \#\{a_i = -1\} l > 0$. We have by construction that each $\pm u_i$ is distinct and nonzero. Let $\frac{1}{2} \leq c_1, \dots, c_t \leq 1$ be real numbers to be chosen later

and define $\hat{f}(u_i) := c_i e^{\frac{\pi i}{2l}}/(2t)$ and $\hat{f}(-u_i) = \overline{\hat{f}(u_i)}$ for all i. Set $\hat{f}(h) = 0$ for all other h. Then $\mathbb{E}_x f(x) = \hat{f}(0) = 0$, f takes values in $[-1, 1]$ and $\Re(\hat{f}(\pm u_i)^l) = 0$ for all i. Thus we have

$$T_\Phi(f) = \Re\left(T_\Phi(f)\right) = \sum_{h \in \mathbb{F}_p^n} |\hat{f}(h)|^{2k} \Re(\hat{f}(h)^l) = 0.$$

It remains to argue that $|T_\Psi(f)| \gg_{p,t} 1$. Fourier inverting we have

$$T_\Psi(f) = \sum_{y_1, y_2 \in \mathbb{F}_p} \prod_{i=1}^{t} \hat{f}(b_{1i} y_1 + b_{2_i} y_2).$$

Let $U = \{\pm u_1, \ldots, \pm u_t\}$ and let $\mathcal{I}$ be the collection of multisubsets of $[t]$ which have t elements. For a multisubset W of U with t elements, let $i(W) \in \mathcal{I}$ be the multisubset of indices of W (for example if $t = 4$ we may have $i([u_1, u_1, u_3, -u_1]) = [1, 1, 3, 1]$), and let $\Sigma(W)$ be the sum of the signs in W (so $\Sigma([u_1, u_1, u_3, -u_1]) = 1 + 1 + 1 - 1 = 2$). Then

$$T_\Psi(f) = \sum_{I \in \mathcal{I}} \left(\sum_{W : i(W) = I} \#\{y_1, y_2 : M^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = W\} \cdot \frac{1}{(2t)^t} e^{\frac{\Sigma(W)\pi i}{2l}} \right) \prod_{i \in I} c_i, \tag{7.8}$$

where we abuse notation by implicitly 'unordering' the vector $M^T \binom{y_1}{y_2}$ and identifying it with the corresponding multiset. Viewing (7.8) as a polynomial in $c_1, \ldots, c_t$, we may choose $c_1, \ldots, c_t$ in such a way as to force $|T_\Psi(f)| \gg_{p,t} 1$ unless (7.8) is the zero polynomial. Note that when $I = [t]$, we have by our construction of $u = (u_1, \ldots, u_t)$ that the only multisubsets W for which $\#\{y_1, y_2 : M^T \binom{y_1}{y_2} = W\}$ is nonzero are $W = [u_1, u_2, \ldots, u_t]$ and $W = [-u_1, -u_2, \ldots, -u_t]$, and in these cases $\#\{y_1, y_2 : M^T \binom{y_1}{y_2} = W\} = 1$. Furthermore, for these W we have that $\Sigma(W) = t$ and $-t$ respectively. Therefore, the coefficient of the term $\prod_{i=1}^t c_i$ in (7.8) is $\frac{2}{(2t)^t} \Re\left(e^{\frac{t\pi i}{2l}}\right)$, which is nonzero since t is not of the form $(2m+1)l$ for any integer m. Thus we may choose $c_1, \ldots, c_t$ so that $|T_\Psi(f)| \gg_{p,t} 1$, completing the proof. □

Remark 7.3.4. It is not difficult to adapt the construction from the previous proposition to relax the assumption that $t \neq (2m+1)l$ to the assumption that $l \neq 1$ or t is even; one proceeds by defining $\hat{f}(u_i) := c_i e^{\frac{(2n_i+1)\pi i}{2l}}$ for an appropriate choice of integers $\{n_i\}_{i=1}^t$. In fact, it seems likely that a modest perturbation of the above argument may allow us to remove any assumption of this kind completely. We do not pursue the matter in this thesis as we have no immediate need for the more general statement. We hope that the above example is illustrative.

7.3.2 Systems with additive tuples are not locally Sidorenko

Together with Theorem 7.3.1, this completes the proof of Theorem 7.1.1.

Theorem 7.3.5. *Let Ψ be be a system of two linear equations which is linearly generic. Then Ψ is not locally Sidorenko.*

The main observation underpinning the upcoming construction is that the complexity of a system of linear forms is not maintained under Fourier inversion. We have not investigated the extent to which higher-order Fourier analysis may be applied in frequency space, but the upcoming construction (and natural generalisations thereof) at least demonstrates one application of the aforementioned observation.

In light of Theorem 7.3.1 we may assume that t is odd. We note that if $\Psi(S)$ corresponds to an additive $(t-1)$-tuple, then by Fourier inversion, $T_{\Psi(S)}$ is the ($(t-1)$th power of the) ℓ^{t-1} norm in frequency space. In particular, if $T_{\Psi(S)}(f) = 0$ then $f = 0$, so the proof strategy from the previous subsection fails. We will have to settle for a function f for which the ratio $T_{\Psi(S)}(f)/T_{\Psi}(f)$ is arbitrarily large. The existence of such a function is proven in Proposition 7.3.7. First, we need a lemma.

Lemma 7.3.6. *The following facts hold:*

1. *Define the quadratic $q : \mathbb{F}_p^n \to \mathbb{F}_p$ by $q(x) = x^T M x + h^T x$, where M is a matrix of rank r and $h \in \mathbb{F}_p^n$. Then*

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

2. *Let Ψ be a system of linear forms, each mapping $(\mathbb{F}_p^n)^D$ to $\mathbb{F}_p^n$. Let $A \subset \mathbb{F}_p^n$ be the zero set of the quadratic defined by $x^T x$. Then*

$$|T_{\Psi}(1_A) - p^{-k}| \leq p^{-n/2},$$

where k is the dimension of the $\mathbb{F}_p$-span of the functions $\{\psi_i(x)^T \psi_i(x)\}_{i=1}^t$.

Sketch proof. Part 1 is a very standard Gauss sum estimate; a proof may be found in [33], for example. In [28, Proof of Theorem 3.1] it is shown (using Part 1 as the main ingredient) that if a system of linear forms $\tilde{\Psi} := (\tilde{\psi}_i)_{i=1}^l$ has $\{\tilde{\psi}_i(x)^T \tilde{\psi}_i(x)\}_{i \in [l]}$ linearly independent, then $|T_{\tilde{\Psi}}(1_A) - p^{-l}| \leq p^{-n/2}$. Let $S \subset [t]$ be a set such that $\{\psi_i(x)^T \psi_i(x)\}_{i \in S}$ is a basis for $\{\psi_i(x)^T \psi_i(x)\}_{i=1}^t$. Then we observe the equality $\prod_{i \in S} 1_A(\psi_i(x)) = \prod_{i \in [t]} 1_A(\psi_i(x))$ for all x, so $T_{\Psi}(1_A) = T_{\Psi(S)}(1_A)$ and our result then follows from the second sentence of this proof. □

For a $2 \times k$ matrix M with values in $\mathbb{F}_p$, let $M^{(2)}$ be the $3 \times k$ matrix defined by

$$M^{(2)} := \begin{pmatrix} \boldsymbol{r_1}^2 \\ \boldsymbol{r_2}^2 \\ \boldsymbol{r_1 r_2} \end{pmatrix},$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the first and second rows of M respectively, and where multiplication of vectors is conducted coordinatewise.

Proposition 7.3.7. *Let $t \geq 5$ be odd. Let Ψ be a system of t linear forms whose image has codimension 2 and which is linearly generic. Let Φ be the system defined by an additive $(t-1)$-tuple. Then, for all n sufficiently large in terms of p and t, there is a function $f : \mathbb{F}_p^n \to [-1, 1]$ such that $\mathbb{E}_x f(x) = 0$, $|T_\Psi(f)| \gg_{p,t} p^{-n(t-4)/2}$ and $T_\Phi(f) < p^{-n(t-3)/2}$.*

Proof. We construct f by specifying its Fourier transform $\hat{f} : \mathbb{F}_p^n \to \mathbb{C}$. Let A be the zero set of the quadratic form given by $x^T x$. Define $\hat{f}(0) = 0$ and $\hat{f}(h) = \frac{1}{p^{n/2}+1}(1_A(h) - p^{-1})$ otherwise. Firstly, $T_\Phi(f) = \sum_h |\hat{f}(h)|^{t-1} < p^n \cdot \left(p^{-n/2}\right)^{t-1} = p^{-n(t-3)/2}$.

It is clear that for all $h \in \mathbb{F}_p^n$ we have $\hat{f}(h) = \overline{\hat{f}(-h)}$, and so f takes values in $\mathbb{R}$. Furthermore, for all $x \neq 0$:

$$\begin{aligned}
|f(x)| &= \left| \sum_h \hat{f}(h) e_p(x \cdot h) \right| \\
&= \frac{1}{p^{n/2}+1} \left| \sum_h (1_A(h) - p^{-1}) e_p(x \cdot h) - (1 - p^{-1}) \right| \\
&\leq \frac{1}{p^{n/2}+1} \left(\left| p^n \mathbb{E}_h \left(\mathbb{E}_{a \in \mathbb{F}_p} e_p(a h^T h) \right) e_p(x^T h) \right| + (1 - p^{-1}) \right) \\
&\leq \frac{1}{p^{n/2}+1} \left(p^n \mathbb{E}_{a \in \mathbb{F}_p} \left| \mathbb{E}_h e_p(a h^T h + x^T h) \right| + (1 - p^{-1}) \right) \\
&\leq \frac{1}{p^{n/2}+1} \left(p^{n/2} + 1 - p^{-1} \right) \\
&< 1,
\end{aligned}$$

by Fourier inverting in the first line, an expansion of 1_A in terms of quadratic phases (known to experts) in the third line, and Lemma 7.3.6 Part 1 in the fourth. For $x = 0$ we have

$$|f(0)| = \frac{1}{p^{n/2}+1} \left| \sum_h (1_A(h) - p^{-1}) - 1 + p^{-1} \right| < 1,$$

by Lemma 7.3.6 Part 2 with the trivial set of linear forms $x \mapsto (x)$. Thus f takes values in $[-1, 1]$.

By (7.3) and the condition that the image of Ψ has codimension 2 we have

$$|T_\Psi(f)| = p^{2n}|T_{\Psi^\perp}(\hat{f})| \gg_{p,t} p^{2n} \cdot p^{-tn/2} \left|T_{\Psi^\perp}(1_A - p^{-1}) + O_t(p^{-n/2})\right|,$$

where $\Psi^\perp$ denotes the dual set of linear forms which may be determined as in (7.3). It remains to show that $|T_{\Psi^\perp}(1_A - p^{-1})| \gg_{p,t} 1$. We have

$$\begin{aligned} T_{\Psi^\perp}(1_A - p^{-1}) &= \sum_{S\subset[t]} (-p)^{-t+|S|} T_{\Psi^\perp(S)}(1_A) \\ &= \sum_{S\subset[t]} (-p)^{-t+|S|} \mathbb{E}_{x_1,x_2} \prod_{i\in S} 1_0\left(\psi_i^\perp(x_1,x_2)^T \psi_i^\perp(x_1,x_2)\right). \end{aligned}$$

Since the image of Ψ has codimension 2, we have that each $\psi_i^\perp \in \Psi^\perp$ maps $(\mathbb{F}_p^n)^2 \to \mathbb{F}_p^n$. Thus each $\psi_i^\perp(x)^T\psi_i^\perp(x)$ maps $x = (x_1, x_2) \in (\mathbb{F}_p^n)^2$ to a linear combination of $\{x_1^T x_1, x_1^T x_2, x_2^T x_2\}$, so the dimension of the span of the functions $\{\psi_i^\perp(x)^T\psi_i^\perp(x)\}_{i=1}^t$ is at most 3. By Lemma 7.3.6,

$$T_{\Psi^\perp}(1_A - p^{-1}) = \sum_{S\subset[t]} (-p)^{-t+|S|} p^{-\dim \operatorname{span}\{\psi_i^\perp(x)^T\psi_i^\perp(x)\}_{i\in S}} + O_t(p^{-n/2}).$$

One observes that $\dim \operatorname{span}\{\psi_i^\perp(x)^T\psi_i^\perp(x)\}_{i\in S} = \operatorname{rank} M_S^{(2)}$ where M_S is the matrix obtained by restricting the columns of M_Ψ to the coordinates in S. We show that $|T_\Psi(1_A - p^{-1})| \gg_{p,t} 1$ by showing that $\sum_{S\subset[t]}(-p)^{-t+|S|}p^{-\operatorname{rank} M_S^{(2)}}$ is nonzero. A linear algebraic consequence of the fact that every 2×2 minor of M_Ψ has nonzero determinant is that for each $S \subset [t]$ with $|S| \geq 3$, $\operatorname{rank} M_S^{(2)} = 3$. Furthermore it is clear that if $|S| = 1, 2$ then $\operatorname{rank} M_S^{(2)} = 1, 2$ respectively. Thus we have

$$\begin{aligned} \sum_{S\subset[t]} (-p)^{-t+|S|} p^{-\operatorname{rank} M_S^{(2)}} &= p^{-3} \sum_{|S|\geq 3} (-p)^{-t+|S|} + (\binom{t}{2} - t + 1)(-p)^{-t} \\ &= \frac{(1-p)^t + p^3(\binom{t}{2} - t + 1) - (p^2\binom{t}{2} - pt + 1)}{(-p)^t p^3}. \end{aligned}$$

We claim that the numerator is nonzero for pairs (p,t) with p prime and $t \geq 5$. Indeed, for $p = 2$ and t odd we obtain a quadratic whose roots satisfy $t < 5$. The same is true of $p = 2$ and t even. Otherwise, for large values of (p,t) it is clear that the term $(1-p)^t$ dominates, and the claim may be checked by direct computation for small values of (p,t) (the author obtained a bona fide proof when $t \geq 11$, or $t \in \{5,7,9\}$ and $p \geq 11$, and checked the cases $(p,t) \in \{3,5,7\} \times \{5,7,9\}$ by hand; for the sake of brevity we do not include the details). $\square$

Remark 7.3.8. In the above proposition, the assumption that Ψ is linearly generic may be removed quite easily, at least if one allows p to be sufficiently large in terms of t. Indeed one writes $\sum_{S\subset[t]}(-p)^{-t+|S|}p^{-\operatorname{rank} M_S^{(2)}}$ as a p-adic series $\sum_{i=-t}^{-1} a_i p^i$. Each $|a_i|$ may be bounded in terms of t, so if p is large enough in terms of t then this series representation is unique and so the sum is zero if and only if each a_i is zero. One may use some linear algebra to show that at least one of the a_i must be nonzero. It is certainly plausible that the more general result may be used in the further classification of Sidorenko systems in a directly analogous way to how Proposition 7.3.7 is used here, but we have no need for such a statement in this thesis and so we do not write out the details.

Proof of Theorem 7.3.5. Recall that we may assume that t is odd. As in the proof of Theorem 7.3.1, using (7.2), we have

$$T_\Psi(\alpha+\varepsilon f)=\alpha^t+\alpha\varepsilon^{t-1}\sum_{|S|=t-1}T_{\Psi(S)}(f)+\varepsilon^t T_\Psi(f).$$

As we have seen already, each $\Psi(S)$ has codimension one and so $T_{\Psi(S)}(f)$ is of the form $\sum_h\prod_{i=1}^{t-1}\hat{f}(a_ih)$. By Hölder's inequality then

$$|T_{\Psi(S)}(f)|\leq\sum_h|\hat{f}(h)|^{t-1}=T_\Phi(f),$$

where we let Φ denote an additive $(t-1)$-tuple. Thus

$$T_\Psi(\alpha+\varepsilon f)\leq\alpha^t+(t-1)\alpha\varepsilon^{t-1}T_\Phi(f)+\varepsilon^tT_\Psi(f).$$

Let f be the function from Proposition 7.3.7, and if $T_\Psi(f)>0$ then replace f with $-f$, so

$$T_\Psi(\alpha+\varepsilon f)\leq\alpha^t+(t-1)\alpha\varepsilon^{t-1}p^{-n(t-3)/2}-\varepsilon^tc_{p,t}p^{-n(t-4)/2}.$$

For ε,α fixed, we may choose n large enough so that $T_\Psi(\alpha+\varepsilon f)<\alpha^t$, completing the proof. □

This also completes the proof of Theorem 7.1.1, which is a union of Theorem 7.3.1 and Theorem 7.3.5.

7.4 Examples

7.4.1 Locally Sidorenko but not Sidorenko

Recall that a system Ψ comprising t linear forms is locally Sidorenko if for all $\alpha \in [0,1]$ there exists $\varepsilon > 0$ such that for all $n \geq 1$ and all $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}f = 0$ we have $T_\Psi(\alpha + \varepsilon f) \geq \alpha^t$. We begin by providing the following example of a system which is locally Sidorenko but not Sidorenko. This proves Proposition 7.1.6.

Example 7.4.1. Let Φ denote the following system of two linear equations in nine variables:

$$x_1 - x_2 + x_3 - x_4 = 0$$
$$x_5 - x_6 + x_7 - x_8 + x_9 = 0.$$

We firstly observe that Φ is not Sidorenko. Indeed, let $A = \{1\} \subset \mathbb{F}_p$. Then $\mathbb{E}_{x \in \mathbb{F}_p} 1_A(x) = 1/p$, but A contains no solutions to Φ, so $T_\Phi(1_A) = 0$.

We claim that Φ is locally Sidorenko. Let A4 be shorthand for the first equation in Φ and A5 for the second. We observe that for a function f on $\mathbb{F}_p^n$ we have $T_\Phi(f) = T_{\mathrm{A4}}(f)T_{\mathrm{A5}}(f)$. Also, since A4 has codimension one, we get $T_{\mathrm{A4}}(\alpha + f) = \alpha^4 + T_{\mathrm{A4}}(f)$ for functions f with $\mathbb{E}f = 0$, and similarly for A5. Thus, for $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}f = 0$,

$$\begin{aligned} T_\Phi(\alpha + \varepsilon f) &= (\alpha^4 + T_{\mathrm{A4}}(\varepsilon f))(\alpha^5 + T_{\mathrm{A5}}(\varepsilon f)) \\ &= \alpha^9 + \varepsilon^4 \left(\alpha^5 T_{\mathrm{A4}}(f) + \alpha^4 \varepsilon T_{\mathrm{A5}}(f) + \varepsilon^5 T_{\mathrm{A4}}(f) T_{\mathrm{A5}}(f)\right). \end{aligned}$$

Fourier-inverting, we have that $T_{\mathrm{A4}}(f) = \sum_h |\hat{f}(h)|^4$ and $T_{\mathrm{A5}}(f) = \sum_h |\hat{f}(h)|^4 \hat{f}(h)$. In particular we have that $|T_{\mathrm{A5}}(f)| \leq T_{\mathrm{A4}}(f) \leq 1$. Thus,

$$T_\Phi(\alpha + \varepsilon f) \geq \alpha^9 + \varepsilon^4 T_{\mathrm{A4}}(f) \left(\alpha^5 - \alpha^4 \varepsilon - \varepsilon^5\right),$$

and so setting $\varepsilon \leq \alpha/2$ we have that $T_\Phi(\alpha + \varepsilon f) \geq \alpha^9$, so Φ is indeed locally Sidorenko.

7.4.2 Weakly locally Sidorenko but not locally Sidorenko

Recall that a system Ψ comprising t linear forms is weakly locally Sidorenko if for all $n \geq 1$, for all $\alpha \in [0,1]$ and all $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$ with $\mathbb{E}f = 0$, there exists $\varepsilon_f > 0$ such that for all $\varepsilon < \varepsilon_f$ we have $T_\Psi(\alpha + \varepsilon f) \geq \alpha^t$.

We claimed earlier that it is easy to contrive systems of complexity at least two which are weakly locally Sidorenko, but not Sidorenko. One may, for example, begin

with a system of complexity at least two whose shortest subsystems have length at least 5, and take the union of this system with an additive 4-tuple on a separate variable set. The weak local Sidorenko property follows from the fact that its smallest subsystem is an additive 4-tuple, and the (non) Sidorenko property of the entire systems essentially follows from that of the complexity two part by tensoring with an appropriate quadratic phase function. We omit the details.

In this subsection we exhibit a *complexity one*, linearly generic system which is weakly locally Sidorenko, but not locally Sidorenko, proving Corollary 7.1.5. The upcoming example is taken from [48, Example 4.6], as is most of the analysis towards showing that it is weakly locally Sidorenko. That it is not locally Sidorenko follows from Theorem 7.1.1.

Example 7.4.2. Let $p \notin \{2, 3\}$ and let Φ be the following system of two equations in five variables:

$$x_1 - x_2 + x_3 - x_4 = 0$$
$$x_1 + 2x_2 - x_3 - 2x_5 = 0.$$

One sees that Φ is linearly generic. By Theorem 7.1.1, Φ is not locally Sidorenko. We claim that Φ is weakly locally Sidorenko. For f with $\mathbb{E}f = 0$ we have

$$T_\Phi(\alpha + \varepsilon f) = \alpha^5 + \alpha\varepsilon^4 \sum_{|S|=4} T_{\Phi(S)}(f) + \varepsilon^5 T_\Phi(f),$$

so it suffices to show that $\sum_{|S|=4} T_{\Phi(S)}(f) > 0$ whenever $f \neq 0$. One computes using (7.3) and recalling that $T_{\Phi(S)}(f)$ is real-valued that

$$\begin{aligned}
\sum_{|S|=4} T_{\Phi(S)}(f) &= \sum_h |\hat{f}(h)|^4 + 2|\hat{f}(h)|^2|\hat{f}(2h)|^2 + 2\hat{f}(-h)\hat{f}(2h)^2\hat{f}(-3h) \\
&= \sum_h \frac{1}{2}|\hat{f}(2h)|^4 + \frac{1}{2}|\hat{f}(3h)|^4 + 2|\hat{f}(h)|^2|\hat{f}(2h)|^2 \\
&\qquad + 2\Re\left(\hat{f}(-h)\hat{f}(2h)^2\hat{f}(-3h)\right) \\
&\geq \sum_h |\hat{f}(-2h)\hat{f}(3h)|^2 + 2|\hat{f}(h)|^2|\hat{f}(2h)|^2 \\
&\qquad + 2\Re\left(\hat{f}(-h)\hat{f}(2h)^2\hat{f}(-3h)\right) \\
&= \sum_h \left|\hat{f}(h)\hat{f}(-2h) + \overline{\hat{f}(-2h)\hat{f}(3h)}\right|^2 + |\hat{f}(h)|^2|\hat{f}(2h)|^2 \\
&\geq \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2.
\end{aligned}$$

This is positive unless $|\hat{f}(h)|^2|\hat{f}(2h)|^2 = 0$ for all h. But in this case $\sum_{|S|=4} T_{\Phi(S)}(f) = \sum_h |\hat{f}(h)|^4$, which is positive unless f is zero. This completes the proof of the claim that f is weakly locally Sidorenko.

7.4.3 Locally Sidorenko, not Sidorenko, and no additive tuple

Recall that if Φ comprises only an additive $(2k)$-tuple, the corresponding functional T_Φ is a norm on the space of functions and in particular is positive definite: $T_\Phi(f) = 0 \implies f = 0 \implies T_\Psi(f) = 0$, for any Ψ. This is the main obstruction to the proof of Theorem 7.3.1 going through for systems with an additive $(2k)$-tuple. It transpires that there exists a non-linearly generic Ψ with a subsystem Φ whose corresponding functional is not positive definite but with the property that $T_\Phi(f) = 0 \implies T_{\Psi(S)}(f) = 0$ for all subsystems $\Psi(S) \subseteq \Psi$. This motivates the upcoming example.

Example 7.4.3. Let Φ be the following system of two equations in eight variables:

$$\begin{aligned} x_1 - x_2 + 2x_3 - 2x_4 + x_5 + x_6 + x_7 + x_8 &= 0 \\ x_5 - x_6 + 2x_7 - 2x_8 &= 0. \end{aligned}$$

We claim that Φ is locally Sidorenko (and it clearly does not contain any additive tuples). Fix $\alpha \in (0,1)$ and let $f : \mathbb{F}_p^n \to [-\alpha, 1-\alpha]$. Invoking (7.3) and with some computation one finds that

$$\begin{aligned} T_\Phi(\alpha + \varepsilon f) = \alpha^8 &+ \alpha^4\varepsilon^4 \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2 \\ &+ \alpha\varepsilon^7 \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2\hat{f}(-2h)\hat{f}(h)\hat{f}(-3h) \\ &+ \alpha\varepsilon^7 \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2\hat{f}(2h)\hat{f}(3h)\hat{f}(-h) \\ &+ \alpha\varepsilon^7 \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2\hat{f}(-h)\hat{f}(-3h)\hat{f}(-4h) \\ &+ \alpha\varepsilon^7 \sum_h |\hat{f}(h)|^2|\hat{f}(2h)|^2\hat{f}(3h)\hat{f}(h)\hat{f}(4h) \\ &+ \varepsilon^8 \sum_{h_1,h_2} |\hat{f}(h_1)|^2|\hat{f}(2h_1)|^2 \\ &\qquad \cdot \hat{f}(h_1+h_2)\hat{f}(h_1-h_2)\hat{f}(h_1+2h_2)\hat{f}(h_1-2h_2). \end{aligned}$$

The absolute value of the final sum is at most

$$\sum_{h_1} |\hat{f}(h_1)|^2 |\hat{f}(2h_1)|^2 \sum_{h_2} \left| \hat{f}(h_1+h_2)\hat{f}(h_1-h_2)\hat{f}(h_1+2h_2)\hat{f}(h_1-2h_2) \right|$$

$$\leq \sum_{h_1} |\hat{f}(h_1)|^2 |\hat{f}(2h_1)|^2 \sum_{h_2} \left| \hat{f}(h_1+h_2)\hat{f}(h_1-h_2) \right|,$$

by the bound $||\hat{f}||_\infty \leq 1$. Applying Cauchy–Schwarz to the sum over h_2 and subsequently Parseval's identity yields an upper bound of $\sum_h |f(h)|^2 |f(2h)|^2$. Thus

$$T_\Phi(\alpha + \varepsilon f) \geq \alpha^8 + \varepsilon^4 \left(\sum_h |\hat{f}(h)|^2 |\hat{f}(2h)|^2 \right) \left(\alpha^4 - 4\alpha\varepsilon^3 - \varepsilon^4 \right).$$

Thus setting $\varepsilon \leq \alpha/2$, we have that $T_\Phi(\alpha + \varepsilon f) \geq \alpha^8$, and so Φ is locally Sidorenko.

We note that the previous example is not Sidorenko, so gives another, albeit less simple, example verifying Proposition 7.1.6.

7.5 Localising to subsystems with fewest variables

In this section we observe how to 'localise' to subsystems with the fewest variables. As an application we outline a relatively short proof of the recent result of Kamčev–Liebenau–Morrison [47] and independently Versteegen [68] that any system containing a four-term arithmetic progression (or indeed any system of two independent equations in four variables) is uncommon. Our proof does not use any quadratic methods, which is also new here.

Previously we have localised by considering small perturbations of a constant function $\alpha + \varepsilon f$, whereupon, defining s to be the size of smallest subsystem of Ψ for which some $T_{\Psi(S)}(f)$ is nonzero, we have

$$T_\Psi(\alpha + \varepsilon f) \approx \alpha^t + \varepsilon^s \sum_{|S|=s} T_{\Psi(S)}(f).$$

For fixed positive integers n, k and $f : \mathbb{F}_p^n \to [-1,1]$, define the function $f^{(k)} : \mathbb{F}_p^{n+k} \to [-1,1]$ by $f^{(k)} := f \otimes 1_0$ where 1_0 is the characteristic function for $0 \in \mathbb{F}_p^k$ (that is, $f^{(k)} = (f \circ \pi_1)(1_0 \circ \pi_2)$, where π_1 projects onto the first n coordinates, π_2 projects onto the final k coordinates). Then for an injective system $\Psi = (\psi_1, \ldots, \psi_t)$ in D variables, we have recalling (7.4) that

$$T_\Psi(f^{(k)}) = T_\Psi(f) T_\Psi(1_0) = p^{-kD} T_\Psi(f).$$

Thus, expanding $T_\Psi(\alpha + f \otimes 1_0)$ multilinearly and letting k be large we have that

$$T_\Psi(\alpha + f \otimes 1_0) = \alpha^t + p^{-kd} \sum_{\dim \Psi(S)=d} T_{\Psi(S)}(f) + O_t(p^{-k(d+1)}),$$

where $\dim \Psi(S)$ denotes the dimension of the image of $\Psi(S)$ and d is the minimal such quantity for a subsystem for which $T_{\Psi(S)}(f)$ is nonzero.

Now we outline an alternative proof of the fact that any system containing a 4AP is uncommon. In fact we will prove the stronger result (as [47] did) that any system containing a rank two system with four linear forms is uncommon. We note that some parts of the argument remain the same. In particular we need the result that rank 2 systems in 4 variables can have $T_\Psi(f) < 0$. Rather than reproving it we just quote the following weak version of [47, Lemma 4.1] whose proof we note does not use higher-order methods.

Lemma 7.5.1. *[47, Lemma 4.1] Let p be a prime. There exists an integer $n \geq 1$ such that the following is true. Let $(\Psi_i)_{i=1}^k$ be a collection of linear systems each in four variables with codimension two. Suppose that Ψ_i is linearly generic for all i. Then there exists $f : \mathbb{F}_p^n \to [-\frac{1}{2}, \frac{1}{2}]$ with $\mathbb{E}f = 0$ such that $T_{\Psi_i}(f) < 0$ for $i = 1, \ldots, k$.*

Theorem 7.5.2 ([47, Theorem 1.1], generalisation of [68, Theorem 1.4])**.** *Any linear system of distinct, nonzero forms which contains a dimension 4 subsystem of rank 2 is uncommon.*

Proof. Let t be the number linear forms in Ψ and let $f : \mathbb{F}_p^n \to [-\frac{1}{2}, \frac{1}{2}]$ have $\mathbb{E}f = 0$ and be otherwise unspecified for the meantime. Let $k \geq 0$ to be determined later, $\varepsilon = p^{-k}$ and $1_0 : \mathbb{F}_p^k \to \{0, 1\}$ be the characteristic function of zero. Then

$$T_\Psi\left(\frac{1}{2} + (\varepsilon f) \otimes 1_0\right) + T_\Psi\left(\frac{1}{2} - (\varepsilon f) \otimes 1_0\right) = \frac{1}{2^{t-1}} + \sum_{\substack{\emptyset \neq S \subset [t] \\ |S| \text{ even}}} \frac{1}{2^{t-|S|-1}} p^{-k(|S| + \dim \Psi(S))} T_{\Psi(S)}(f).$$

Letting k be arbitrarily large, we see it suffices to show that there is a choice of f such that $\sum_{S:\dim \Psi(S)+|S| \text{ minimal}} T_{\Psi(S)}(f) < 0$, where the minimum is taken over subsystems whose image does not contain a coordinate hyperplane (if the image contains a coordinate hyperplane then $T_{\Psi(S)}(f) = 0$ since $\mathbb{E}f = 0$). Note that this minimum is at most 6 since Ψ contains a system of four nonzero linear forms with codimension two, and for the minimum we must have that $\dim \Psi(S) < |S|$.

If the minimum is equal to 6, then the minimal systems $\Psi(S)$ must have $|S| = 4, \dim \Psi(S)) = 2$. If this is the case then we must have that $\Psi(S)$ is linearly generic for all of these minimal systems, or else by an easy linear algebraic check they would give rise to a further subsystem S' with $|S'| + \dim \Psi(S')$ smaller. Thus we may conclude by invoking Lemma 7.5.1 (with an appropriate choice of n) in this case.

Otherwise any $\Psi(S)$ contributing to the minimum must have $\dim \Psi(S) = 1$, and indeed cannot have $|S| = 4$ since any system $\Psi(S)$ with $|S| = 4, D = 1$ gives rise to $\binom{4}{2}$ subsystems $\Psi(S_i)$ with $\dim \Psi(S_i) = 1, |S_i| = 2$. In the remaining case that the minimal systems have $\dim \Psi(S) = 1, |S| = 2$ we note that these systems are described by a single equation which cannot have its two coefficients summing to zero since this would mean that Ψ has a pair of repeated forms. Thus we may conclude by the Fox–Pham–Zhao random Fourier sampling argument to find f such that $\sum_{S:\dim \Psi(S)+|S| \text{ minimal}} T_{\Psi(S)}(f) < 0$.[3] This completes the proof. □

Remark 7.5.3. Kamčev–Liebenau–Morrison conjecture [47, Conjecture 6.1] that an analogue of Lemma 7.5.1 holds for systems comprising an even number of linear forms whose image has codimension two. They note that their proof of [47, Theorem 1.1] however does not generalise to the case of higher number of variables, even if [47, Conjecture 6.1] is confirmed. We note that our proof above would generalise should [47, Conjecture 6.1] be confirmed. This would go some way towards the classification of weakly locally common systems.

[3]In fact this is a very easy case of the Fox–Pham–Zhao argument since the equations comprise only two variables. It is likely an even simpler argument exists, but this is a known (and nice) argument, so we quote it for brevity.

Bibliography

[1] O. Alrabiah, V. Guruswami, P. Kothari, and P. Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom csp refutation. In *Electron. Colloquium Comput. Complex*, 2022.

[2] D. Altman. On a conjecture of Gowers and Wolf. *Discrete Analysis*, 2022:10, 13pp.

[3] D. Altman. On Szemerédi's theorem with differences from a random set. *Acta Arith.*, 195(1):97–108, 2020.

[4] D. Altman. Local aspects of the sidorenko property for linear equations, 2022. arXiv:2210.17493.

[5] D. Altman. A non-flag arithmetic regularity lemma and counting lemma, 2022. arXiv:2209.14083.

[6] D. Altman. On a question of Alon, 2022. arXiv:2210.13515.

[7] D. Altman and Z. Chase. On the smallest gap in a sequence with Poisson pair correlations, 2022. arXiv:2210.17466.

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

[9] J. Bourgain. Ruzsa's problem on sets of recurrence. *Israel J. Math.*, 59(2):150–166, 1987.

[10] J. Briët and B. Green. Multiple correlation sequences not approximable by nilsequences. *Ergodic Theory Dynam. Systems*, 42(9):2711–2722, 2022.

[11] S. A. Burr and V. Rosta. On the Ramsey multiplicities of graphs—problems and recent results. *J. Graph Theory*, 4(4):347–361, 1980.

[12] P. Cameron, J. Cilleruelo, and O. Serra. On monochromatic solutions of equations in groups. *Rev. Mat. Iberoam.*, 23(1):385–395, 2007.

[13] D. Castro-Silva and J. Briët. On the threshold for Szemerédi's theorem with random differences, 2023. arXiv:2304.03234.

[14] D. Conlon, J. Fox, and B. Sudakov. An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.*, 20(6):1354–1366, 2010.

[15] D. Conlon, J. H. Kim, C. Lee, and J. Lee. Some advances on Sidorenko's conjecture. *J. Lond. Math. Soc. (2)*, 98(3):593–608, 2018.

[16] D. Conlon and J. Lee. Finite reflection groups and graph norms. *Adv. Math.*, 315:130–165, 2017.

[17] D. Conlon and J. Lee. Sidorenko's conjecture for blow-ups. *Discrete Anal.*, pages Paper No. 2, 13, 2021.

[18] L. J. Corwin and F. P. Greenleaf. *Representations of nilpotent Lie groups and their applications. Part I*, volume 18 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990. Basic theory and examples.

[19] E. Csóka, T. Hubai, and L. M. Lovász. Locally common graphs. *Journal of Graph Theory*, 2022.

[20] P. Erdős. On the number of complete subgraphs contained in certain graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 7:459–464, 1962.

[21] J. Fox, H. T. Pham, and Y. Zhao. Common and Sidorenko linear equations. *Q. J. Math.*, 72(4):1223–1234, 2021.

[22] J. Fox and F. Wei. On the local approach to Sidorenko's conjecture. *Electron. Notes Discret. Math.*, 61:459–465, 2017.

[23] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.*, 60:41–90, 2016.

[24] N. Frantzikinakis, E. Lesigne, and M. Wierdl. Random differences in Szemerédi's theorem and related results. *Journal d'Analyse Mathématique*, 130(1):91–133, Nov 2016.

[25] A. W. Goodman. On sets of acquaintances and strangers at any party. *Amer. Math. Monthly*, 66:778–783, 1959.

[26] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[27] W. T. Gowers. A uniform set with fewer than expected arithmetic progressions of length 4. *Acta Math. Hungar.*, 161(2):756–767, 2020.

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

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

[30] W. T. Gowers and J. Wolf. Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$. *Mathematika*, 57(2):215–237, 2011.

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

[32] B. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.

[33] B. Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

[34] B. Green and I. Z. Ruzsa. Sum-free sets in abelian groups. *Israel J. Math.*, 147:157–188, 2005.

[35] B. Green and T. Tao. An arithmetic regularity lemma, an associated counting lemma, and applications. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 261–334. János Bolyai Math. Soc., Budapest, 2010.

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

[37] B. Green and T. Tao. Yet another proof of Szemerédi's theorem. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 335–342. János Bolyai Math. Soc., Budapest, 2010.

[38] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.

[39] B. Green and T. Tao. On the quantitative distribution of polynomial nilsequences - erratum, 2015. arxiv:1311.6170.

[40] B. Green and T. Tao. An arithmetic regularity lemma, associated counting lemma, and applications, 2020. arXiv:1002.2028v3.

[41] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[42] R. Hancock, D. Král, M. Krnc, and J. Volec. Toward characterizing locally common graphs. *Random Structures & Algorithms*, 2022.

[43] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[44] H. Hatami, P. Hatami, and S. Lovett. General systems of linear forms: equidistribution and true complexity. *Adv. Math.*, 292:446–477, 2016.

[45] H. Hatami and S. Lovett. Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms. *Geom. Funct. Anal.*, 21(6):1331–1357, 2011.

[46] C. Jagger, P. Šťovíček, and A. Thomason. Multiplicities of subgraphs. *Combinatorica*, 16(1):123–141, 1996.

[47] N. Kamčev, A. Liebenau, and N. Morrison. On uncommon systems of equations, 2021. arXiv:2106.08986.

[48] N. Kamčev, A. Liebenau, and N. Morrison. Towards a characterization of Sidorenko systems. *The Quarterly Journal of Mathematics*, page haad013, 03 2023.

[49] J. H. Kim, C. Lee, and J. Lee. Two approaches to Sidorenko's conjecture. *Trans. Amer. Math. Soc.*, 368(7):5057–5074, 2016.

[50] D. Král', A. Lamaison, and P. P. Pach. Common systems of two equations over the binary field. 2022. Discrete Math Days, conference proceedings. Extended abstracts available at `https://dmd2022.unican.es/978-84-19024-03-9.pdf`, pages 169–173.

[51] B. Kuca. True complexity of polynomial progressions in finite fields. *Proc. Edinb. Math. Soc. (2)*, 64(3):448–500, 2021.

[52] A. Leibman. Polynomial mappings of groups. *Israel J. Math.*, 129:29–60, 2002.

[53] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):215–225, 2005.

[54] J. L. X. Li and B. Szegedy. On the logarithmic calculus and Sidorenko's conjecture, 2011. arXiv:1107.1153.

[55] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.

[56] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[57] F. Manners. True complexity and iterated Cauchy–Schwarz, 2021. arXiv:2109.05731.

[58] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[59] A. Saad and J. Wolf. Ramsey multiplicity of linear patterns in certain finite abelian groups. *Q. J. Math.*, 68(1):125–140, 2017.

[60] A. F. Sidorenko. Inequalities for functionals generated by bipartite graphs. *Diskret. Mat.*, 3(3):50–65, 1991.

[61] A. F. Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.

[62] Alexander F. Sidorenko. Cycles in graphs and functional inequalities. *Mat. Zametki*, 46(5):72–79, 104, 1989.

[63] B. Szegedy. An information theoretic approach to Sidorenko's conjecture, 2014. arXiv:1406.6738.

[64] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[65] T. Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[66] T. Tao. A correction to "an arithmetic regularity lemma, an associated counting lemma, and applications", 2020. `https://terrytao.wordpress.com/2020/11/26/a-correction-to-an-arithmetic-regularity-lemma-an-associated-counting-lemma-and-applications/`.

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

[68] L. Versteegen. Linear configurations containing 4-term arithmetic progressions are uncommon, 2021. arXiv:2106.06846.

[69] L. Versteegen. Common and Sidorenko equations in Abelian groups. *J. Comb.*, 14(1):53–67, 2023.