THE PARTITION RANK VS. ANALYTIC RANK PROBLEM FOR CYCLIC GROUPS I. EQUIDISTRIBUTION FOR PERIODIC NILSEQUENCES

JAMES LENG

ABSTRACT. We give improved quantitative equidistribution estimates for nilsequences that are periodic modulo a large prime, obtaining bounds single exponential in dimension, improving a result of Green and Tao (who obtained similar results but with losses double exponential in dimension). To do so, we refine Green and Tao's argument and overcome the "induction on dimensions" obstruction present in several places in their argument. Our results are enough to imply quasi-polynomial type bounds for certain complexity one polynomial Szemerédi theorems that the author proved in [Len1], improving on the iterated logarithm bound the author obtained in [Len1]. In subsequent work [Len2], we extend these results to general and multiparameter nilsequences. The strength of our quantitative bounds are analogous in higher order Fourier analysis over $\mathbb{F}_p^n$ to the partition rank being polynomial in the analytic rank. In principle, this also gives a new proof of quantitative equidistribution of nilsequences, although there are many similarities to Green and Tao's proof.

1. INTRODUCTION

In 2001, Gowers [Go] introduced *Gowers norms* as a way to measure "pseudorandomness" in Szemerédi's theorem and proved an inverse theorem for the Gowers norm stating that obstructions to Gowers uniformity norms are roughly constant on short arithmetic progressions. In 2010, Green-Tao-Ziegler, motivated by parallel work in ergodic theory [HK1] and [Z], in a series of works [GT1, GT3, GT4, GTZ1, GTZ2] identified nilsequences as the obstructions to Gowers uniformity, and used this to generalize Vinogradov's approach of using the circle method to count certain linear configurations in the primes. A key result needed in their analysis is a quantitative equidistribution theorem [GT1, Theorem 1.16], which we list as follows (relevant definitions such as $\|\cdot\|_{C^\infty[N]}$ can be found in Section 2):

Theorem 1. *Let $F(g(n)\Gamma)$ be a nilsequence on a nilmanifold G/Γ with a δ^{-1}-rational Mal'cev basis such that F has Lipschitz parameter (which is the sum of the Lipschitz constant and the L^∞ norm) ≤ 1. If*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma) - \int_{G/\Gamma} F d\mu| \ge \delta$$

then either $N \ll_{G/\Gamma,\delta} 1$ or there exists a nonzero homomorphism $\eta : G \to \mathbb{C}$ of modulus at most $\delta^{-O_{G/\Gamma}(1)}$ which annihilates Γ such that $\|\eta \circ g\|_{C^\infty[N]} \le \delta^{-O_{G/\Gamma}(1)}$.

We note that a special case of this theorem is the case of when G is abelian and $g(n) = g^n$ is an orbit, which is as follows:

Proposition 1.1. *Let $\alpha \in \mathbb{R}^d/\mathbb{Z}^d$ and $F : \mathbb{R}^d/\mathbb{Z}^d \to \mathbb{C}$ be a Lipschitz function with Lipschitz parameter ≤ 1. If*

$$|\mathbb{E}_{n\in[N]}F(\alpha n) - \int_{\mathbb{R}^d/\mathbb{Z}^d} F d\mu| \geq \delta$$

then either $N \ll_{d,\delta} 1$ or there exists a vector $\eta \in \mathbb{Z}^d$ of size at most $\delta^{-O_d(1)}$ such that $\|k \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \ll_{d,\delta} 1$.

A corollary of Theorem 1 is the "Ratner-type factorization theorem" of Green and Tao [GT1, Corollary 1.20], which is often used in applications (see [GT1, Definition 1.2] for definition of "δ-equidistributed" and "totally δ-equidistributed"; we do not define these terms here since the purpose of this article is to eliminate these notions in favor of quantitatively superior alternatives):

Theorem 2. *Let G/Γ be a filtered nilmanifold with complexity $M \geq 2$ and $g(n)$ be a polynomial sequence on G. For each $A \geq 2$, there exists some δ with $M \leq \delta^{-1} \leq M^{O_{G/\Gamma,A}(1)}$ and a factorization $g = \epsilon g_1 \gamma$ where*

- *ϵ is (δ^{-1}, N)-smooth, meaning that for all $n \in [N]$, $d(\epsilon(n-1), \epsilon(n)) \leq \frac{\delta^{-1}}{N}$;*
- *γ is δ^{-1}-periodic; and*
- *g_1 is δ^A-equidistributed inside a subnilmanifold $\tilde{G}/\tilde{\Gamma}$ with complexity at most δ^{-1} where $\tilde{G}$ is a subgroup of G with rationality at most δ^{-1}.*

The corresponding abelian degree one case is the following:

Proposition 1.2. *Let $\alpha \in \mathbb{R}^d/\mathbb{Z}^d$. Then given $M \geq 2$ and $A \geq 2$, there exists some δ with $M \leq \delta^{-1} \leq M^{O_{d,A}(1)}$ and a factorization $\alpha = \epsilon + \alpha' + \gamma$ where*

- *$\|\epsilon\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{\delta^{-1}}{N}$;*
- *there exists some integer $k \leq \delta^{-1}$ such that $k\gamma \in \mathbb{Z}^d$; and*
- *$n \mapsto \alpha' n$ is totally δ^A-equidistributed inside a subgroup $\tilde{G}$ of $\mathbb{R}^d/\mathbb{Z}^d$ of rationality at most δ^{-1}.*

In [TT], the authors work out that bounds "$M^{O_{G/\Gamma,A}(1)}$" and "$\delta^{-O_{G/\Gamma}(1)}$" corresponds to quantities double exponential in the dimension of G. They raise the question of whether such theorems can be single exponential in the dimension. In the negative direction, a simple example below shows that even in the abelian degree one case, the Ratner-type factorization theorem cannot have bounds single exponential in dimension.

Example. *Consider the example $g(n) = (\alpha n, \delta^{-1}\alpha n, \delta^{-2}\alpha n, \delta^{-4}\alpha n, \dots, \delta^{-2^d}\alpha n)$ (a linear orbit on $\mathbb{T}^{d+2}$) with α highly irrational. The factorization theorem with $A = \log(1/\delta)$ and $M = 2$, yields that this $\delta^{2^{d+1}}$-equidistributes in the subtorus $(x, \delta^{-1}x, \dots, \delta^{-2^d}x)$, which give losses double exponential in dimension. This is because the threshold for equidistribution in $g(n)$ at each step in the iteration barely keeps up with how fast the rationality of the torus increases, causing us to continue to need to pass to a subtorus. Note that if we worked with $g_1(n) = (\alpha n, \delta^{-2}\alpha n, \delta^{-2^2}\alpha n, \dots, \delta^{-2^{2^d}}\alpha n)$, then the factorization theorem would yield equidistribution in $(x, \delta^{-2}x, \dots, \delta^{-2^{2^{O(1)}}}x, \dots, x_{d+2})$, which give polynomial losses. In this case, the threshold for equidistribution is far less than how fast the rationality of the torus increases, causing us to not need to pass to a subtorus after $O(1)$ many steps.*

We will now use some space to explain why the abelian degree one factorization theorem is formulated as it is; from which it is clear why the general factorization theorem is formulated as it is. More specifically, we explain why (for the sake of applications) the theorem specifies a threshold of equidisttribution of smaller than δ^A, where δ^{-1} upper bounds the rationality of the sub-torus. Given a Lipschitz function $F : \mathbb{R}^d/\mathbb{Z}^d \to \mathbb{C}$ of parameter ≤ 1, $\delta > 0$, and $\alpha \in \mathbb{R}^d/\mathbb{Z}^d$, an application of the abelian Ratner-type factorization theorem gives us a decomposition of $\alpha = \epsilon + \alpha' + \gamma$ where $\epsilon, \alpha', \gamma$ are as specified in Proposition 1.2. Then letting $\tilde{G}$ be the subgroup that α' lies in, we see that along any coset H of $\tilde{G}$, the restriction of F to H has Lipschitz norm at most $d\delta^{-1}$. Letting Q be the linear integer such that $Q\gamma \in \mathbb{Z}^d$, we see that if Q is a subprogression of $[N]$ of size at least $\delta^3/2N$, it follows that choosing A to be $10\log(d)$, we see that

$$\left|\mathbb{E}_{n\in Q}F(\alpha n) - \int_{\tilde{G}} F d\mu\right| \le \left|\mathbb{E}_{n\in Q}\tilde{F}(\alpha' n) - \int_{\tilde{G}} F d\mu\right| + O(\delta^2) \ll \delta^A \|F\|_{Lip(\tilde{G}} + \delta^2 \ll \delta^2$$

where $\tilde{F} = F(\epsilon_0 + \cdot)$ for some element $\epsilon_0 \in \mathbb{R}/\mathbb{Z}$ (i.e., some element ϵ_0 such that ϵn is close to ϵ_0 for each $n \in Q$). We see that if A is too small (i.e., the threshold for equidistribution is too large), then $\delta^A\|F\|_{Lip(\tilde{G})}$ would be too large for equidistribution to hold. This explains why we must have the threshold for equidistribution to be significantly smaller than the rationality of the subgroup. Thus, we have a"structure vs. randomness" theory for degree one nilsequence $F(n\alpha)$: on progressions (structure), the nilsequence $F(n\alpha)$ behaves like a totally equidistributed nilsequence (random). This theory has led to the Green-Tao-Ziegler deduction of the U^{s+1} inverse theorem [GTZ1, GTZ2], was an integral part of their result on linear equations in primes [GT3, GT4], and continues to see use to this day [TT].

For quantitative higher order Fourier analysis, however, this is somewhat inefficient. To describe how to make this more efficient, we briefly review the proof of Proposition 1.2 and point out where the proof can be made more efficient if one merely wanted a structure vs. randomness theory for a *nilsequence* rather than a *polynomial sequence*. The proof of Proposition 1.2 proceeds algorithmically as follows:

(i) If αn is δ-equidistributed, we are done.
(ii) Otherwise, there exists some $k \le \delta^{-1-o(1)}$ such that $\|k \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \le \frac{\delta^{-O(d)}}{N}$.
(iii) We may thus write $\alpha = \epsilon + \alpha' + \gamma$ where γ is $\delta^{-O(d)}$-rational, ϵ is $d\delta^{-1-o(1)}$-smooth, and α' lies in a lower dimensional subgroup.
(iv) Iterate this procedure along progressions Q for which ϵn is roughly constant, and γn lies in a fixed coset of $\mathbb{Z}^d$ whenever $n \in Q$.

We see that along each step of this iteration, if we were to run the argument with a nilsequence $F(\alpha n)$, we see that upon restricting to a subgroup, F became Lipschitz with parameter roughly $\delta^{-1-o(1)}$. This increase in Lipschitz parameter causes one to need to look at characters of size at most $\delta^{-2-o(1)}$; upon an iteration, the number of characters one needs to look at is on the order of double exponential in dimension. *This* is ultimately what is responsible for the double exponential losses in dimension. We note that since the bound on k in step (ii) are essentially optimal that this procedure *has to* yield double exponential bounds in dimension.

To overcome this, the observation to make is that the abelian Ratner-type factorization theorem is tailored towards the *polynomial sequence* αn rather than the *nilsequence* $F(\alpha n)$. Since a Lipschitz function F can be approximated by finitely many characters, the number of characters we should look at should not dramatically increase under the iteration; rather, it should decrease. This can be done if one worked with the finitely many characters that F approximates, rather than the function F itself as we have done above.[1] The present article then studies the following question:

Question. *What is the analogue of the above heuristic for nilsequences?*

The example above also foreshadows future obstructions we must overcome in our quest to improve quantitative bounds for the equidistribution of nilsequences. Many such arguments related to the equidistribution on tori or nilmanifolds proceed via an induction on dimensions. Induction on dimensions is ultimately inevitable, but we must prevent ourselves from incurring losses double exponential in dimension. This means that if δ is a parameter in the proof, we cannot even allow a seemingly harmless iteration of $\delta \mapsto \delta^2$, since after an induction on dimensions, this would lead to losses of δ^{2^d}.

Despite this example, previous work of Gowers and Wolf [GW], Green and Tao [GT2], and the author [Len3] show that an equidistribution theory for degree two nilsequences with losses single exponential in dimension is both possible and useful. However, these previous works stop frustratingly short of proving a result of the flavor of Theorem 1 with good quantitative bounds, instead only proving that a non-equidistributed degree two nilsequence is (with losses single exponential in dimension) a degree one nilsequence in *some* way. In the above theorem, one can deduce a distributional result for an orbit of a *polynomial sequence*, while these previous methods only allow one to conclude that the *nilsequence* is approximately one-step in some way (in the terminology of [Len3, Section 3], we can only conclude that the nilsequence has bounded *approximate Fourier complexity*). The reason the Green-Tao approach is so powerful is that they show that the orbit can only end up in boundedly many subgroups, each of bounded rationality, whereas in contrast, showing that a nilsequence is one-step does not limit nearly as much the number of nilsequences one can obtain. This extra information is what's necessary for applications such as the complexity one polynomial Szemerédi theorem [Len1], which the previous equidistribution approaches of [GT2], [GW], and [Len3] seem to be unable to replicate. To drive this point home, we note that while [GW] obtains quantitatively stronger results, it isn't clear from their work how to generalize their main results to higher complexity systems, while Altman's [A1, A2], Candela-Sisask's [CS], Green-Tao's [GT5], and Kuca's [K1, K2] powerful approaches seem far more flexible and comparatively "easy."

While being inflexible, these previous approaches suggest that at least for degree two nilsequences, it may at least be possible to deduce a more robust equidistribution theorem in the flavor of Theorem 1 with good bounds. The proof of our main theorem (in particular Theorem 9) makes this a reality, and it turns out that understanding how the previous approaches of [GW], [GT2] and [Len3] relate to the equidistribution on nilmanifolds leads to the key of the proof, Lemma 3.1 which we term as the "refined bracket polynomial lemma." Furthermore, the method used to prove the corresponding two-step statement (Theorem 7) generalizes to arbitrary nilsequences:

[1]Another way and perhaps a more satisfying way to rescue the factorization theorem in the abelian case is to use dilated tori as in [GT2] to counteract the increase in Lipschitz constant when we pass to a rational subgroup. The point is that having a larger Lipschitz constant is more expensive than increasing the volume of the torus as [GT2, Lemma 7.2] shows. Unfortunately, the author was unable to find such a generalization to arbitrary nilsequences.

Theorem 3. *Let N be a prime number, $0 < \delta < \frac{1}{10}$, and $M \geq 1$ real. Let $F(g(n)\Gamma)$ be a periodic nilsequence modulo N (that is, $g(n+N)\Gamma = g(n)\Gamma$ for all n) with dimension d, complexity M, Lipschitz parameter ≤ 1, step s, and degree k. Suppose F is a Lipschitz vertical character with nonzero frequency ξ with $|\xi| \leq (\delta/M)^{-1}$. If*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \geq \delta$$

then either $N \ll (\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$ or else $n \mapsto F(g(n)\Gamma)$ is a nilsequence on some nilmanifold $\tilde{G}/\tilde{\Gamma}$ of degree $\leq k$ and step $\leq s-1$ with Lipschitz parameter and complexity at most $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$ and dimension $\leq d-1$.

It turns out that we can take $\tilde{G}/\tilde{\Gamma}$ to be a $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$-rational subnilmanifold of the projection of G/Γ to the image of ξ. See Theorem 9. The corresponding one-step statement is the following:

Proposition 1.3. *Let N be prime and $p(n) = \alpha_0 + \alpha_1 n + \cdots \alpha_k n^k$ be a polynomial whose coefficients are rational with denominator dividing N. If*

$$|\mathbb{E}_{n\in[N]}\exp(2\pi i p(n))| \geq \delta$$

then either $N \ll \delta^{-O_k(1)}$ or else $\|\alpha_i\|_{\mathbb{R}/\mathbb{Z}} = 0$ for all $i = 1, \dots, k$.

We note that Theorem 3 sidesteps the above example, since the above example illustrates problems that may occur if one considers *all* observables, and here, we only consider *one* observable (where here, "observable" refers to a Lipschitz function). This can be thought of as a possible cyclic group analogue of the "partition rank vs. analytic rank problem" that occurs in the study of structure vs. randomness for obstructions to Gowers uniformity in higher order Fourier analysis over $\mathbb{F}_p^n$. As will be explained below, the quantitative bounds we obtain here are analogous to the partition rank being polynomial in the analytic rank. In fact, our deduction of Theorem 3 gives single exponential bounds for Theorem 1 in the case of periodic nilsequences, as indicated by Theorem 9. Furthermore, in Theorem 10 we iterate Theorem 3 to obtain an analogue of the Ratner-type factorization theorem for a single observable. In subsequent work [Len2], the author proves a similar result for general and multiparameter nilsequences. Since the proof for periodic nilsequences is cleaner and conceptually easier for the reader to follow, we have decided to separate the proof of the periodic nilsequences case here. Indeed, in the periodic nilsequences case, as observed in [CS], [K1, K2], and [Len1] that the smooth term ϵ in the Ratner-type factorization theorem disappears, or rather, becomes a constant term. Furthermore, since our nilsequences are periodic modulo a prime, we may also eliminate the rational part γ via a change of variables without having to pass to subprogressions, using the fact that modular inverses for nonzero elements exist modulo a prime. As will be explained below, the periodic nilsequences case is already interesting and is enough for applications to arithmetic combinatorics. Another benefit of isolating the periodic nilsequences case is that all of the previous literature for periodic nilsequences are *ad hoc* and use various tricks that are derived from the general nilsequences case, while here, we provide a direct argument for the periodic nilsequences case. Additionally, by [Ma2], it is the periodic nilsequences case that is the "partition rank vs. analytic rank problem" for cyclic groups.

1.1. **Relevance to quantitative higher order Fourier analysis.** In 2018, Manners [Ma1] showed a quantitative Gowers inverse theorem, obtaining the following result:

Theorem 4. *Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be one-bounded such that $\|f\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \geq \delta$ with s fixed and N prime. Then there exists a nilsequence $F(g(n)\Gamma)$ with dimension $D(\delta) = \delta^{-O(1)}$, complexity $M(\delta) = \exp(O(\exp(\delta^{-O(1)})))$, and $|F(g(n)\Gamma)| \leq 1$ and having Lispchitz constant ≤ 1 such that*

$$|\langle f, F(g(n)\Gamma)\rangle| \geq c(\delta)$$

where $c(\delta) = \exp(-O(\exp(\delta^{-O(1)})))$.

In fact, a calculation indicates that Manners obtains that the complexity is double exponential in the dimension[2], and $c(\delta) = (\delta/M(\delta))^{O(D(\delta))^{O(1)}}$. The conjectured quasi-polynomial inverse theorem states the following:

Conjecture 1. *One can take $D(\delta) = \log(1/\delta)^{O(1)}$, $M(\delta) = \exp(O(D(\delta))^{O(1)})$, and $c(\delta) = (\delta/M(\delta))^{O(D(\delta))^{O(1)}}$.*

In some sense, Manners's result is two (iterated) exponentials away from the conjectured quasi-polynomial inverse theorem. The only case where the quasi-polynomial inverse theorem is known is the U^3 setting, obtained by Sanders [S]. There, a calculation shows that we can take the complexity of the two-step nilmanifold obtained to be $O(1)$. Thus, a moral consequence of our work is that assuming a quasi-polynomial inverse theorem, if one wanted to further apply an equidistribution theorem to the nilsequence obtained from the inverse theorem, one would only end up with quasi-polynomial losses. In other words, if one were content with quasi-polynomial losses and if one were to assume Conjecture 1, then applying an equidistribution result of nilsequences $O(1)$ many times is inexpensive. This is illustrated in the two applications we give below as well as in [Len3].

One can view the analogous result of Theorem 3 in higher order Fourier analysis in $\mathbb{F}_p^n$ is the result that the partition rank of a tensor is polynomial in the analytic rank of the tensor (see e.g., [M] for definitions). The conjectured quasi-polynomial inverse theorem in this setting is as follows:

Conjecture 2. *Let $f : \mathbb{F}_p^n \to \mathbb{C}$ be one-bounded such that $\|f\|_{U^{s+1}(\mathbb{F}_p^n)} \geq \delta$ with $s \ll p$. Then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most s such that*

$$|\langle f, e(P(n))\rangle| \gg_{s,p} \exp(-\log(1/\delta)^{O(1)}).$$

It was shown by Janzer [J] and Milićević [M] that the partition rank is polynomial in the analytic rank. Like the cyclic group case, a moral consequence of the quasi-polynomial inverse theorem and the result of the partition rank being polynomial in the analytic rank is that applying such a partition rank vs. analytic rank result yields at most quasi-polynomial losses since we would likely be able to work with polynomials with analytic rank at most $\log(1/\delta)^{O(1)}$. An illustration of this is [BL] where the authors are able to show strong orthogonality of the Möbius function with degree two polynomials[3]. Again, this is only a *moral* consequence since applications of these theorems depend on the context. For progress on Conjecture 2, see [GM1, GM2, KLT]. For recent work on the partition rank vs. analytic rank problem, see e.g., [AKZ, CM, KZ, LZ, MZ].

[2]The key obstruction to obtaining better complexity bounds also seems to be the difficulty in overcoming an induction on dimensions. The author is not aware, however, of any connection between the obstruction there and the induction on dimensions obstruction we run into.

[3]Although it is not shown there, one can derive quasi-polynomial bounds for the gowers U^3 norm of the Möbius function over $\mathbb{F}_p[T]$ similar to how the author derived similar bounds for $\|\mu - \mu_{Siegel}\|_{U^3[N]}$ in [Len3].

1.2. **Applications.** In subsequent work, we shall deduce applications of Theorem 3 to relevant problems in arithmetic combinatorics. One such application is the following:

Theorem 5. *Let N be a large prime and $A \subseteq \mathbb{Z}/N\mathbb{Z}$ be a subset lacking the configuration $(x, x+P(y), x+Q(y), x+P(y)+Q(y))$ where $P, Q \in \mathbb{Z}[x]$ are linearly independent and have zero constant coefficient and y ranges over $\mathbb{Z}/N\mathbb{Z}$. Then*

$$|A| \ll_{P,Q} \frac{N}{\exp(\log^{c_{P,Q}}(N))}$$

for some constant $c_{P,Q} > 0$.

This theorem is an improvement to a previous result of the author [Len1] which obtained a similar qualitative result as above but with the quantitative bounds

$$|A| \ll_{P,Q} \frac{N}{\log_{O_{P,Q}(1)}(N)}$$

where $\log_{O_{P,Q}(1)}(N)$ is an iterated logarithm. There, [Len1] remarked that improved bounds for [Len1, Lemma 6.1] (combined with Sanders' work [S]) would yield Theorem 5. We shall deduce an improvement to [Len1, Lemma 6.1] here in Appendix B, but postpone further details elsewhere.

The result of [Len2] for general nilsequences also gives a simpler proof of the main result in [Len3], one which also generalizes if one assumes Conjecture 1:

Theorem 6. *Assume Conjecture 1. Then*

$$\|\mu\|_{U^{s+1}[N]} \ll_A^{ineff} \log^{-A}(N)$$
$$\|\Lambda - \Lambda_Q\|_{U^{s+1}[N]} \ll_A^{ineff} \log^{-A}(N)$$

where Λ_Q is defined in [Len3].

This will be deduced in [Len2] along with the equidistribution theorem for general nilsequences. Additionally, it seems plausible that Theorem 3 can find applications in regularity and counting lemmas. For interesting recent work in this direction, see [A1, A2, K1, K2].

1.3. **Organization of the paper.** In Section 2, we define the notation used in the paper. In Section 3, we consider a warm-up problem for two-step nilsequences, and arrive at a crucial lemma, which is the refined bracket polynomial lemma. We shall deduce this lemma in Section 3 via an argument of T. Tao (private communication). This lemma will be used in subsequent sections. In Section 4, we consider a second warm-up problem for more general two-step polynomial sequences. In Section 5, we deduce the main theorem. In Section 6, we deduce a Ratner-type factorization theorem for a single observable.

In Appendix A, we give proofs of some auxiliary lemmas for the proof of the main theorem. In Appendix B, we deduce the improved bounds for [Len1, Lemma 6.1]. In Appendix C, we give two additional proofs of the refined bracket polynomial lemma, meant to be used in [Len2]. The first is a more straightforward but more technical proof of the refined bracket polynomial lemma. Readers unfamiliar with the geometry of numbers can refer to the proof there instead of the one in Section 3. The second is a generalization of the proof given in Section 3 adapted to work for arbitrary nilsequences. In Appendix D, we state relevant results in Diophantine approximation and the geometry of numbers.

1.4. **Acknowledgements.** We would like to thank Terry Tao for advisement and for coming up with a simpler proof of the periodic refined bracket polynomial lemma than the proof the author initially came up with. We would also like to thank Jaume de Dios, Zach Hunter, Ben Johnsrude, Borys Kuca, Freddie Manners, Rushil Raghavan, and David Soukup for helpful discussions and suggestions related to this problem. In particular, when the author was on the verge of giving up on this problem, the preparation of giving an explanation of this problem to Borys led the author to the key inspiration needed to solve this problem. We are grateful to Joni Teräväinen as well for helpful comments and raising the problem of formulating an equidistribution theory with better bounds (jointly with Terry Tao) as a remark in [TT] where the author first learned of this problem. The author is supported by an NSF Graduate Research Fellowship Grant No. DGE-2034835.

2. Notation and Conventions

In this section, we will mostly go over notation regarding nilpotent Lie groups. Most of our notation follows [GT1]. First, we set some general notation. Given a function $f : A \to \mathbb{C}$ defined on a finite set A, we set

$$\mathbb{E}_{a\in A} f(a) := \frac{1}{|A|}\sum_{a\in A} f(a).$$

We also denote for $x \in \mathbb{C}$, $e(x) = \exp(2\pi i x)$. Given a real number $r \in \mathbb{R}$, we denote $\{r\} \in (-1/2, 1/2]$ as the difference between r and a nearest integer to r. Given a general vector $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, we denote $\{a\} = (\{a_1\}, \dots, \{a_d\})$.

Nilmanifolds arise naturally as a means to "clean up" bracket polynomials that show up in the inverse theory of Gowers norms. For further reference along these objects and how they relate to higher order Fourier analysis, we refer the reader to [T2] or [HK2]. A nilpotent Lie group G is a group whose *lower central series* $G \supseteq G_{(2)} \supseteq G_{(3)} \cdots$ where $G_{(i)} := [G, G_{(i-1)}]$ is eventually the identity element after finitely many steps. If s is the least integer such that $G_{(s)}$ is the identity, then G is known as a s-step nilpotent Lie group. We shall examine G/Γ where Γ is a discrete cocompact subgroup. Objects $X = G/\Gamma$ are referred to as *nilmanifolds*. $G_{(s)}$ is known as the *vertical component*[4] and $G/G_{(2)}$ as the *horizontal component*. Following a reduction of Leibman [Le], we shall assume that G is connected and simply connected.

A *filtration* for a nilpotent Lie group G is a sequence of normal subgroups $(G_i)_{i=0}^\infty$ such that $G_0 = G_1 = G$, $[G_i, G_j] \subseteq G_{i+j}$ and for which $G_i = 0$ eventually. The largest k for which G_k is nonzero is known as the *degree* of the filtration. Note that each nilpotent Lie group has the *standard filtration* which is defined in the above paragraph as $G_i = [G, G_{i-1}]$, but not all filtrations have to be the standard filtration. It is true that the standard filtration is minimal in the sense that given a filtration (G_i), $G_i \supseteq G_{(i)}$. See [T2, Exercise 1.6.2]. The point of specifying a filtration on a nilpotent Lie group is so one can define general *polynomial sequences* on nilpotent Lie groups. Intuitively, if one thinks of G has a subgroup of the unipotent matrices, then polynomial sequences are sequences of G where all matrix coefficients are polynomials. More specifically, given a filtration $(G_i)_{i=1}^\infty$, we define a *polynomial*

[4]Note that we use a slightly different convention than [GT1], which work with a filtered nilpotent Lie group (G_i) of degree k and the vertical component is G_k.

sequence on G as a sequence

$$g(n) = g_0 g_1^n g_2^{\binom{n}{2}} g_3^{\binom{n}{3}} \cdots$$

where $g_i \in G_i$. We shall denote the set of all polynomial sequences as $\mathrm{poly}(\mathbb{Z}, G)$. It can be shown that $\mathrm{poly}(\mathbb{Z}, G)$ forms a group under pointwise multiplication. See [T2, Section 1.6]. Given a positive integer D, an element g of G is *rational with denominator* D, or D*-rational*, if $g^D \in \Gamma$. If a metric on G is specified (as will be shortly), a *nilsequence* is a sequence of the form $n \mapsto F(g(n)\Gamma)$ where $F : G/\Gamma \to \mathbb{C}$ is Lipschitz. The *Lipschitz parameter* of $F(g(n)\Gamma)$, denoted $\|F\|_{Lip(G)}$, or simply $\|F\|_{Lip}$ if the space it is a Lipschitz function in is clear, is the sum of the Lipschitz constant and the $L^\infty(G)$ norm of F. The point of considering Lipschitz functions is that in the case that G is abelian, they can be quantitatively Fourier approximated. See Lemma *A.6* for generalizations of that fact. We say a polynomial sequence is *periodic modulo* N if $g(n+N)\Gamma = g(n)\Gamma$.

Given a nilpotent Lie group G with a discrete cocompact subgroup Γ, we denote $\mathfrak{g}$ as its Lie algebra with the maps $\exp : \mathfrak{g} \to G$ and $\log : G \to \mathfrak{g}$ the exponential and logarithm maps. Given a filtration (G_i) of G, a *Mal'cev basis* of G/Γ is a basis $\{X_1, \dots, X_d\}$ of $\mathfrak{g}$ satisfying the following:

- For each $j = 0, \dots, d-1$, the subspace $\mathfrak{h}_j = \{X_{j+1}, \dots, X_d\}$ is an ideal of $\mathfrak{g}$ and hence $H_j := \exp(\mathfrak{h}_j)$ is a normal subgroup of G.
- If $d_i = \dim(G_i)$, we have $H_{d-d_i} = G_i$.
- Each $g \in G$ may be written uniquely as $\exp(t_1 X_1)\exp(t_2 X_2)\cdots\exp(t_d X_d)$.
- Γ consists of the elements in G for which $t_i \in \mathbb{Z}$.

It is a result of Mal'cev [Mal] that these coordinates exist on a nilmanifold corresponding to a connected and simply connected nilpotent Lie group. Given a Mal'cev basis, we let $\psi : G \to \mathbb{R}^d$ be the coordinate map. A nilmanifold G/Γ together with a filtration G_i and a Mal'cev basis has complexity M if for all i, j

$$[X_i, X_j] = \sum_k c_{ijk} X_k$$

and c_{ijk} is rational with numerator and denominator at most M. Similarly, a subgroup G' of G is Q-rational if its Lie algebra generators X_i' can be written as a rational combination of $X_i's$ with numerator and denominator at most Q. For convenience, when we say that a nilmanifold G/Γ has *degree* k and *complexity* M, we implicitely specify a filtration of degree k and a Mal'cev basis adapted to that filtration of complexity M.

Given a Mal'cev basis, with a coordinate map ψ, we may define a distance map d to be right invariant and satisfying

$$d_\psi(x, y) := \inf\{\sum_{i=1}^n \min(|\psi(x_i^{-1} x_{i-1})|, |\psi(x_{i-1}^{-1} x_i)|) : x_0 = x, x_n = y\}.$$

This defines a metric on G/Γ via $d(x\Gamma, y\Gamma) := \inf\{d(x', y') : x'\Gamma = x\Gamma, y'\Gamma = y\Gamma\}$. *Horizontal characters* on G/Γ are homomorphisms $\eta : G \to S^1$ which annihilate Γ and hence also $[G, G]$ since S^1 is abelian. They are referred to as such since $G/[G, G]\Gamma$ is known as the *horizontal torus*. Since we were given a Mal'cev basis, each horizontal character η corresponds to a vector $k \in \mathbb{Z}^d$ and thus embeds in Γ. Thus, a horizontal character lifts to a homomorphism $\eta : G \to \mathbb{R}$ such that $\eta(\Gamma) \subseteq \mathbb{Z}$.

Definition. Given a horizontal character η and an element $w \in G/[G,G]$, we define an inner product $\langle \eta, w\rangle := \eta(w)$ and we say that w and η are *orthogonal* (or correspondingly w is orthogonal to η or η is orthogonal to w) if $\langle \eta, w\rangle = 0$.

Since $G/[G,G]$ can be identified with its Lie algebra (and by the orthogonal subspace to $[\mathfrak{g},\mathfrak{g}]$ with respect to the chosen Mal'cev basis), it makes sense to say that a set of elements of $G/[G,G]$ or a set of horizontal characters on G are *linearly independent*, and it makes sense to talk about the *size* of an element $w \in G/[G,G]$.

Given a character $\xi : G_{(s)} \to S^1$, a *vertical character* or *nilcharacter* of *frequency* ξ is $F : G/\Gamma \to \mathbb{C}$ such that for any $g_s \in G_{(s)}$ (G is s-step nilpotent), we have $F(g_s x) = e(\xi(g_s))F(x)$. These terminology "vertical character" is an association to the *vertical torus*, $G_{(s)}/(\Gamma \cap G_{(s)})$, which is isomorphic to $\mathbb{T}^{d_s}$. Similarly, if H is a subgroup of the center of G, we shall define H-characters with frequency ξ as functions $F : G/\Gamma \to \mathbb{C}$ such that $F(hx) = e(\xi(h))F(x)$ for any $h \in H$. Note that given elements $w_1, \dots, w_{s-1}$ in $\Gamma/(\Gamma \cap [G,G])$, that $g \mapsto \xi([[[[g,w_1],w_2],\dots],w_{s-1}])$ defines a horizontal character. Since the notation $[[[[g,w_1],w_2],\dots],w_{s-1}]$ will appear very often, we introduce the following notation as shorthand:

Definition. Given $g_1, \dots, g_k$ in a group G, we define

$$[g_1, g_2, \dots, g_k] := [[[g_1,g_2],g_3],\dots,g_k].$$

The *size* of a horizontal character η, denoted $|k|$ or $\|k\|_\infty$, is the L^∞ norm of the vector k considered above[5]. Similarly, the size of a vertical character ξ is the L^∞ norm of ξ when considered as a character of $\mathbb{T}^{d_s}$ under the isomorphism of $G_d/(G_d \cap \Gamma)$ and $\mathbb{T}^{d_s}$.

We will now define "smoothness norms." Firstly, recall that for $x \in \mathbb{R}$ that $\|x\|_{\mathbb{R}/\mathbb{Z}}$ is the distance from x to the nearest integer. For a polynomial $p(n) = a_0 + a_1 n + a_2 n^2 + \cdots + a_d n^d$, we define

$$\|p(n)\|_{C^\infty[N]} := \max_{1 \le j \le d} N^j \|a_j\|_{\mathbb{R}/\mathbb{Z}}.$$

Notice the presence of the variable "n" in the left hand side of the above equation. When doing calculations with these smoothness norms involving polynomials in more than one variable, we shall perform the calculations of the $C^\infty[N]$ as if the polynomial were a polyomial in n.

One technique used often in the proof is the elimination of a rational part. Specifically, if $g(n)$ is a polynomial sequence of degree $k < N$ and $g(n)\Gamma$ is periodic modulo N, if we can factor $g(n)\Gamma = g_1(n)\gamma(n)\Gamma$ for γ that is P-rational for P relatively prime to N, $\gamma(0) = 1$, and g_1 lying in some subgroup G_1 of G with $\Gamma_1 = G_1 \cap \Gamma$, picking m such that $n \equiv Pk!m \pmod N$, we obtain $g(Pk!m)\Gamma = g_1(Pk!m)\Gamma$ (the use of $k!$ comes from taking a Taylor expansion of $\gamma(n) = \gamma_1^n \gamma_2^{\binom{n}{2}} \cdots \gamma_k^{\binom{n}{k}}$ and using Pólya's classification of integer-valued polynomials [P]). Since $g_1(Pk!m)$ lies inside G_1, we may naturally identify $g_1(Pk!m)\Gamma$ with $g_1(Pk!m)\Gamma_1$. Thus, given such a factorization, we may restrict attention to the sequence $m \mapsto g_1(Pk!m)\Gamma_1$ which is again a periodic modulo N. In the non-periodic case, one would just focus on subprogressions of common difference P, but this is rather cumbersome and is

[5] Our notion of "size" agrees with the notion in "modulus" given in [GT1]

one of the reasons why we have decided to isolate the proof of the periodic case as a separate part. Furthermore, any horizontal character (induced by G) that $g_1(Pk!\cdot)$ satisfies, meaning that $\|\eta\circ g_1(Pk!\cdot)\|_{C^\infty[N]} = 0$, should also be satisfied by $g(\cdot)\Gamma$ since P is relatively prime to N.

Of course, we used the following observation above: if q is an integer relatively prime to N and α is a rational with denominator N, a prime and if

$$\|q\alpha\|_{\mathbb{R}/\mathbb{Z}} = 0$$

then

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z}} = 0.$$

We highlight the use of this observation here since this observation will be used many more times throughout this document. We record this technique in the following lemma:

Lemma 2.1. *Let $g(n)$ be a polynomial sequence on a nilpotent Lie group G with $g(0) = 1$ and let Γ be a discrete cocompact subgroup of G such that $g(n)\Gamma$ is periodic modulo N. Suppose g is degree $k < N$ and that we may factorize $g(n) = g_1(n)\gamma(n)$ with $g_1(0) = 1$ and $\gamma(n)$ is P-rational and $\gamma(n)\Gamma$ is periodic modulo P for some P relatively prime to N. Then $\|\eta\circ g\|_{C^\infty[N]} = 0$ if $\|\eta\circ g_1\|_{C^\infty[N]} = 0$ for any horizontal character η.*

Another technique we use is one used to reduce to the case that $g(0) = 1$ and $|\psi(g(1))| \leq 1/2$. We encapsulte this in the following lemma:

Lemma 2.2. *Given a nilsequence $F(g(n)\Gamma)$, there exists a nilsequence $\tilde{F}(\tilde{g}(n)\Gamma)$ such that*

- *$F(g(n)\Gamma) = \tilde{F}(\tilde{g}(n)\Gamma)$ for all $n \in \mathbb{Z}$*
- *$\|\tilde{F}\|_{Lip(G)} \leq M^{O_{s,k}(1)}\|F\|_{Lip(G)}$*
- *$\tilde{g}(0) = 1$ and $|\psi(\tilde{g}(1))| \leq \frac{1}{2}$*
- *For any horizontal character η, $\|\eta\circ g\|_{C^\infty[N]} = \|\eta\circ\tilde{g}\|_{C^\infty[N]}$*
- *If F is a vertical character of frequency ξ, then $\tilde{F}$ is also a vertical character of frequency ξ.*

Proof. To prove this, we factorize $g(0) = \{g(0)\}[g(0)]$ where $|\psi(\{g(0)\})| \leq 1/2$ and $[g(0)] \in \Gamma$ (which is not necessarily unique). Letting $g_1(n) = \{g(0)\}^{-1}g(n)g(0)^{-1}\{g(0)\}$ and $\tilde{F}(x) = F(\{g(0)\}x)$, it follows that $\tilde{F}$ has Lipschitz constant $M^{O_{s,k}(1)}\|F\|_{Lip}$ and $\tilde{F}$ is a vertical character of frequency ξ. This would allow us to reduce to the case that $g(0) = 1$. To reduce to the case that $|\psi(g(1))| \leq 1/2$, we once again factorize $g_1(1) = \{g_1(1)\}[g_1(1)]$ with $[g_1(1)] \in \Gamma$. Letting $\tilde{g}(n) = g_1(n)[g_1(1)]^{-n}$, we that have that $|\psi(\tilde{g}(1))| \leq 1/2$. Furthermore, we have

$$\eta(\tilde{g}) \equiv \eta(g_1(n)) + \eta([g_1(1)]^{-n}) \equiv \eta(\{g(0)\}^{-1}) + \eta(g(n)) \pmod 1.$$

Since $\eta(\{g(0)\}^{-1})$ and $\eta(g(0)^{-1})$ contribute to the constant term of the polynomial which does not affect the smoothness norm, it follows that $\|\eta\circ g\|_{C^\infty[N]} = \|\eta\circ\tilde{g}\|_{C^\infty[N]}$ as we desired. $\square$

One final technique we will mention here is a trick we can use to reduce to the one dimensional vertical torus case:

Lemma 2.3. *Suppose F is a vertical character with oscillation ξ of size at most L and $F(g(n)\Gamma)$ is a nilsequence on a nilmanifold G/Γ. Then there exists a nilsequence $\tilde{F}(\tilde{g}(n)\tilde{\Gamma})$ on a nilmanifold $\tilde{G}/\tilde{\Gamma}$ such that*

- *The vertical torus of the nilmanifold is one-dimensional.*

- $\tilde{F}(\tilde{g}(n)\tilde{\Gamma}) = F(g(n)\Gamma)$.
- $\tilde{G} = G/ker(\xi)$ *and* $\tilde{\Gamma} = \Gamma/(\Gamma \cap ker(\xi))$.
- *The horizontal component of* G *and the horizontal component of* $\tilde{G}$ *are isomorphic.*
- $\|\tilde{F}\|_{Lip(\tilde{G})} \le L^{O_{s,k}(d)^{O_{s,k}(1)}} \|F\|_{Lip(G)}$ *and* $\tilde{G}/\tilde{\Gamma}$ *has complexity at most* $ML^{O_{s,k}(d)^{O_{s,k}(1)}}$.

Proof. Since F is invariant under the kernel of ξ, F descends to a map $\tilde{F}$ on $\tilde{G}$ via a quotient map $\pi : G \to \tilde{G}$. Thus, defining $\tilde{g} = \pi(g)$, it follows that $\tilde{F}(\tilde{g}(n)\tilde{\Gamma}) = F(g(n)\Gamma)$. The complexity and Lipschitz parameter bounds follow from choosing invoking Cramer's rule (i.e., Lemma A.7) an orthogonal Mal'cev basis to ξ.

To show that the vertical direction is one-dimensional, note that if G is s-step, the map $(g_1, g_2, \dots, g_s) \mapsto [g_1, g_2, \dots, g_s]$ descends to a map on $\tilde{G}$ and is only nonzero if $\xi([g_1, g_2, \dots, g_s])$ is nonzero. Finally, to show that the horizontal directions are isomorphic, we note that (set theoretically) $[\tilde{G}, \tilde{G}] = [G, G]/\ker(\xi)$ by considering both sides as cosets of $\ker(\xi)$. Then by the third isomorphism theorem, it follows that $G/[G, G] \cong \tilde{G}/[\tilde{G}, \tilde{G}]$. Furthermore, in the induced quotient map $G/[G, G] \to (\tilde{G}/[\tilde{G}, \tilde{G}])/(\tilde{\Gamma}/(\tilde{\Gamma} \cap [\tilde{G}, \tilde{G}]))$, an element maps to zero if and only if under the original isomorphism, it maps to an element in $\tilde{\Gamma}/(\tilde{\Gamma} \cap [\tilde{G}, \tilde{G}])$, which happens if and only if the element lies in $\Gamma/(\Gamma \cap [G, G])$. Thus, the horizontal tori are (naturally) isomorphic. □

Remark. We note that it is not necessarily true that if H lies in the center of G, the horizontal component of G/H agrees with the horizontal component of G. This is because the set theoretic equivalence we used gives $[G/H, G/H] = [G, G]H/H$, so $(G/H)/([G/H, G/H]) \cong G/H[G, G]$.

2.1. **Asymptotic notation.** We will specify asymptotic notation here. We say that $f = O(g)$ if there exists some absolute constant C such that $|f| \le C|g|$. If h is a variable, we say that $f = O_h(g)$ if there exists a constant C_h depending on h such that $|f| \le C_h|g|$. We shall also adopt Vinogradov's notation and to $f = O(g)$ as $f \ll g$ and $f = O_h(g)$ as $f \ll_h g$. In this paper, s will often denote "step" of a nilmanifold, and k the "degree" of the nilmanifold. Since we are in the setting where they are bounded, $O(g)$ will actually often be $O_{s,k}(g)$. Since in applications to arithmetic combinatorics, s and k are often constant, we make no effort to specify the explicit losses in terms of s and k, though we anticipate since there is an iteration in s and k that losses are double exponential in those parameters. In an effort to shorthand a lot of the exponentials and quantities in [TT, Appendix A], the authors adopted the use of "$\text{poly}_m(\delta)$" (or denoted in our notation as $\text{poly}_d(\delta)$) as any quantity lower bounded by $\gg \exp(\exp(-d^{O_{s,k}(1)}))\delta^{\exp(-d^{O_{s,k}(1)})}$. Since many of our quantities are bounded above by a similarly cumbersome quantity $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$, we shall adopt a similar practice and instead denote $c_1(\delta)$ as any quantity lower bounded by $\gg (\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$.

2.2. **Parameters used.** Given a filtered nilpotent Lie group G with a discrete cocompact subgroup Γ, we will specify some accompanying parameters that will be often used. The dimension of G will be denoted d, and the dimension of G_i will be denoted d_i. We will also specify that $d_{(j)}$ is the dimension of $G_{(j)}$ where as we defined above $G_{(j)}$ is the jth element of the standard, or lower central filtration. We will also specify the dimension of the horizontal torus as $d_{horiz} := d - d_{(2)}$. As stated above, the step of the nilpotent Lie group will often be denoted s and the degree of the filtration will often be denoted k. The complexity of the filtered nilpotent Lie group will be denoted M.

3. The two-step case

The proofs of our results will proceed similarly as the proofs in [GT1], which follow the strategy of "apply the van der Corput inequality and see what happens." We apply a van der Corput inequality and end up with a nilsequence on the joining $G \times_{G_2} G$. It will happen that our Lipschitz function will be invariant under the $(G \times_{G_2} G)_k$-torus, so we may reduce the degree of our nilsequence and proceed by induction. We will divert proofs of these facts of $G \times_{G_2} G$ to Appendix A. Morally speaking, these computations involving van der Corput and considering the joining are equivalent to what one would arrive at if one worked with a bracket polynomial, used the van der Corput inequality, and performed various reductions in the flavor of [GTZ2, Appendix E]. We in fact encourage the reader to work out the degree two bracket polynomial example, despite the inelegance of the computations involved, and compare with the proofs given here and in [GT1]. In principle, all of the proofs in this paper can be rewritten in terms of bracket polynomials, but we have opted out from doing so due to additional technical annoyances that come from working with bracket polynomials. In this section, we shall prove the following theorem (see section 2 for definition of orthogonality):

Theorem 7. *Let N be a prime, $0 < \delta < \frac{1}{10}$, and $g(n)\Gamma$ be a periodic modulo N polynomial sequence on a two-step nilmanifold G/Γ with complexity M equipped with the standard filtration. Let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz nilcharacter of nonzero frequency ξ and Lipschitz parameter ≤ 1. Suppose*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \geq \delta.$$

Then either $N \ll (\delta/M)^{-O(d)^{O(1)}}$ or there exists some integer $d_{horiz} \geq r \geq 0$ and elements $w_1, \dots, w_r \in \Gamma/(\Gamma \cap [G,G])$ and horizontal characters $\eta_1, \dots, \eta_{d_{horiz}-r}$ all bounded by $(\delta/M)^{-O(d)^{O(1)}}$ such that

- *w_i's are linearly independent of each other and η_j's are linearly independent of each other and $\langle \eta_j, w_i\rangle = 0$ for all i and j.*
- *We have*

$$\|\xi([w_i, g])\|_{C^\infty[N]} = 0$$
$$\|\eta_j \circ g\|_{C^\infty[N]} = 0.$$

It's worth noting that the subgroup $\tilde{G} = \{g \in G : \eta_j(g) = 0, [w_i, g] = 0 \forall i, j\}$ is an abelian subgroup of G since given two elements $g, h \in \tilde{G}$, we see that since $\eta_j(h) = 0$ and w_i and η_j are orthogonal, it follows that the horizontal component of h can be spanned by the w_i's, so to verify that $[g, h] = 0$, it just suffices to verify that $[w_i, g] = 0$, which is true by definition. In fact, as we show in Lemma *A.8*, each abelian rational subgroup of G is a subgroup of some group of this form. Combining this lemma with Lemma *A.1*, we obtain

Corollary 3.1. *Let N be a prime, $0 < \delta < \frac{1}{10}$, G/Γ be a two-step nilpotent Lie group with the standard filtration. Let $g(n)\Gamma$ be a periodic modulo N with $g(n)$ a polynomial sequence on G. Let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz nilcharacter of nonzero frequency ξ and Lipschitz parameter ≤ 1 with $|\xi| \leq (\delta/M)^{-1}$. Suppose G/Γ has a one-dimensional vertical torus, and*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \geq \delta.$$

Then either $N \ll (\delta/M)^{-O(d)^{O(1)}}$ *or we can write* $g(n)\Gamma = \epsilon(n)g_1(n)\gamma(n)\Gamma$ *where* ϵ *is constant,* $g_1(n)$ *lies on an abelian subgroup of* G *with rationality* $(\delta/M)^{-O(d)^{O(1)}}$ *and* γ *is* $(\delta/M)^{-O(d)^{O(1)}}$*-rational.*

We now prove Theorem 7. Suppose

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \geq \delta.$$

Using the van der Corput inequality, we see that there are $\delta^{O(1)}N$ many h's such that for each such h,

$$|\mathbb{E}_{n\in[N]}F(g(n+h)\Gamma)\overline{F(g(n)\Gamma)}| \geq \delta^{O(1)}.$$

By an argument in Section 2, we can reduce to the case that $g(0) = id$, and $|\psi(g(1))| \leq \frac{1}{2}$. By defining $\tilde{F}_h(x,y) = F(\{g(1)^h\}x)\overline{F(y)}$, the nonlinear part $g_2(n) = g(n)g(1)^{-n}$, and $\tilde{g}_h(n) = (\{g(1)^h\}^{-1}g_2(n+h)g(1)^n\{g(1)\}^n, g_2(n)g(1)^n)$, we see that

$$|\mathbb{E}_{n\in[N]}\tilde{F}_h(\tilde{g}_h(n)\Gamma)| \geq \delta^{O(1)}.$$

Since $\tilde{g}_h(n)$ is contained in the group $G\times_{G_2} G$, and $\tilde{F}_h$ is invariant under $G_2^{\triangle}$, the diagonal subgroup of G_2^2, and since $[G\times_{G_2} G, G\times_{G_2} G] = G_2^{\triangle}$ it follows that $\tilde{F}_h$ descends to a function on $G\times_{G_2} G/G_2^{\triangle}$, which is a one-step nilpotent group. Since $\tilde{F}_h$ is a nontrivial character in the direction $(x,-x)$ in the G_2^2 coordinate (since it is frequency ξ), it follows that the integral of $\tilde{F}_h$ is zero on the quotient it descended to. By Lemma A.3, we may decompose $\eta(g_1g_2, g_1) = \eta_1(g_1) + \eta_2(g_2)$ with η_1 and η_2 bounded by $c_1(\delta)^{-1}$. In this case, we see that we can take $\eta_2 = \xi$, so

$$|\mathbb{E}_{n\in[N]}e(n\eta_1(g) + \xi(g_2(n+h)g_2(n)^{-1}) + \xi([\{g(1)^h\}, g(1)^n]))| \geq (\delta/M)^{O(d)^{O(1)}} \tag{1}$$

for $\delta^{O(1)}N$ many elements $h \in [N]$. The next lemma is a refinement of what Green and Tao refer to as the "bracket polynomial lemma" (see [GT1, Proposition 5.3]).

Lemma 3.1 (refined bracket polynomial lemma)**.** *Let* $\frac{1}{10} > \delta > 0$ *and* N *be a positive integer. Suppose*

$$|\mathbb{E}_{n\in[N]}e(n\beta + an\cdot\{\alpha h\})| \geq K^{-1}$$

for $\delta N'$ *many* $h \in [N']$ *with* $|a| \leq M$, $\alpha, a \in \mathbb{R}^d$. *Then there exists a quantity* $c_2(\delta, K, M, d) = (\delta/KM)^{O(d)^{O(1)}}$ *such that the following holds. Either one of* $N, N' \ll c_2^{-O(1)}$, *or else there exists linearly independent vectors* $w_1, \dots, w_r$ *and* $\eta_1, \dots, \eta_{d-r}$ *in* $\mathbb{Z}^d$, *all having size less than* c_2^{-1} *such that* $\langle w_i, \eta_j\rangle = 0$ *and*

$$|w_i \cdot a| \leq c_2^{-1}/N', \quad \|\eta_i \circ \alpha\|_{\mathbb{R}/\mathbb{Z}} \leq c_2^{-1}/N'.$$

The theorem will then follow from this lemma. We will supply at least two proofs of this fact and we will give the cleaner proof of the refined bracket polynomial lemma here. Appendix C will contain two additional proofs, one which is a generalization of the following proof (which only works for certain real vectors with denominator N) to arbitrary real vectors, and a different more elementary but cumbersome proof.

3.1. **First proof of the refined bracket polynomial lemma.** The first proof of the refined bracket polynomial lemma is due to T. Tao (private communication) and proceeds via Minkowski's second theorem. Readers unfamiliar with the geometry of numbers may want to either consult [TV] or refer to an alternative proof of this fact given in Appendix C. While we assume here that parameters a and α are rational with denominator N, this proof can be generalized to drop this requirement. See Section C.1. For our purposes, this case is already enough for the proof of our main theorem and for applications for Theorem 5.

It turns out that because $\xi([g(1), \{g(1)^h\}])$ has denominator N (since $\xi([g(1), \{g(1)^h\}]) = \xi([g(1), [g(1)^h]])$ and since $\xi([w_1, w_2]) = 0$ for any w_1, w_2 in Γ) that the following version of the bracket polynomial lemma is sufficient for our purposes.

Lemma 3.2 (periodic refined bracket polynomial lemma)**.** *Let $\frac{1}{10} > \delta > 0$ and N be a prime. Suppose $\alpha, a \in \mathbb{R}^d$ are of denominator N, $|a| \leq M$,*

$$\|\beta + a \cdot \{\alpha h\}\|_{\mathbb{R}/\mathbb{Z}} = 0$$

for δN many $h \in [N]$. The either $N \ll (\delta/M)^{-O(d)^{O(1)}}$ or else there exists linearly independent $w_1, \dots, w_r$ and $\eta_1, \dots, \eta_{d-r}$ in $\mathbb{Z}^d$ with size at most $(\delta/M)^{-O(d)^{O(1)}}$ such that $\langle w_i, \eta_j \rangle = 0$ and

$$\|\eta_j \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} = 0, \quad |w_i \cdot a| = 0.$$

Proof. We first reduce to the case of when $\beta = 0$. Using the pigeonhole principle, there exists at least $\delta N/M$ many h's such that

$$\beta + a \cdot \{\alpha h\} = k$$

for some $k \in \mathbb{Z}$. By using the pigeonhole principle again, there exists a sign pattern in $\{-1, 1\}^d$ such that for proportion of $\delta/2^d M$ many h's all satisfying the sign pattern, we have

$$\beta + a \cdot \{\alpha h\} = k.$$

Taking the difference of any two h's, we see that

$$a \cdot \{\alpha(h - h')\} = 0.$$

Let B denote the convex body $\{|a \cdot x| = 0, |x_i| < \frac{1}{2}\}$[6] and let Γ denote the lattice $\alpha\mathbb{Z} + \mathbb{Z}^d$. By the pigeonhole principle, at least $\delta \frac{N}{2^d M}$ many elements of the lattice lie in B. By Minkowski's second theorem (or rather Proposition $D.1$), there exists linearly independent vectors $v_1, \dots, v_{d'}$ of the lattice corresponding to successive minima $|\lambda_1, \dots, \lambda_{d'}| \leq 100^{-1}$ such that denoting $N_i = (2d\lambda_i)^{-1}$, we have $N_1 \cdots N_d = \frac{\delta N}{2^d M} d^{-O(d)}$ and $\{\ell_1 v_1 + \cdots \ell_{d'} v_{d'} : \ell_i \in [N_i]\} \subseteq B$. Each v_i is of the form $\{\alpha h_i\}$. Let Γ' denote the sublattice of Γ generated by the vectors $v_1, \dots, v_{d'}$ and let V denote the vector space generated by $v_1, \dots, v_{d'}$. The ball of radius $\frac{1}{2}$ around zero contains at most N points of Γ and therefore at most N points of Γ'. Thus, by placing a fundamental parallelopiped of Γ' around each lattice point in the ball of radius $1/2$ (similar in flavor to the argument of Gauss's circle problem), we see that the volume of the ball of radius $1/2$ in V is at most $N(\frac{\delta}{2^d M} d^{-O(d)})^{-O(1)}$ times the the volume of the fundamental parallelopiped of Γ' in V. By Ruzsa's covering lemma (Lemma $D.1$), using the fact that the ball of radius $1/2$ is connected, it follows that the the dilation of $O(d\delta/2M)^{O(d)^2}$ of the ball of radius $\frac{1}{2}$ in V lies in $V \cap B$.

[6]actually we should take $|a \cdot x| < N^{-3}$ to make B have nonempty interior to apply Minkowski's second theorem with but this makes no difference in the proof since a has denominator a large prime

Let P_K denote the generalized arithmetic progression $\{n_1v_1+\cdots+n_{d'}v_{d'} : |n_i| \leq KN_i\}$. This is contained in a ball of radius dK. Letting $v_j = \{h_j\alpha\}$, we see that if $n_1h_1+n_2h_2+\cdots n_{d'}h_{d'} \equiv 0 \pmod N$, then $n_1v_1+\cdots+n_{d'}v_{d'}$ is a point in $\mathbb{Z}^d$. Thus, by the pigeonhole principle, P_K contains at least $K^{d'}\frac{\delta}{2^dM}d^{-O(d)}$ many points in $\mathbb{Z}^d$. Consequently, letting K go to infinity, $\mathbb{Z}^d\cap\Gamma'$ is a lattice in V with covolume at most $(d\delta/2^dM)^{-O(1)}$. This implies, via Minkowski's second theorem again (using the fact that the product of successive minima of the unit ball is bounded by $(\delta/2^dM)^{-O(d)}$ and the fact that all successive minima are at least 1), that the lattice $\mathbb{Z}^d\cap\Gamma'$ has d' linearly independent integer vectors $w_1,\dots,w_{d'}$ of size at most $(d\delta/2^dM)^{-O(d)}$. Since the ball of radius $(d\delta/2^dM)^{O(d)^2}$ is contained in B, it follows that

$$|w_i\cdot a| = 0.$$

By Lemma A.7 (basically Cramer's rule), we may pick $\eta_1,\dots,\eta_{d-d'}$ orthogonal linearly independent integer vectors of size at most $O(d\delta/M)^{-O(d)^2}$, it follows that $\eta_j\cdot\{\alpha h\} = 0$ for a proportion of $(\delta/2^dM)^{O(d)^2}$ many $h\in[N]$ and hence

$$\|\eta_j\cdot\alpha\|_{\mathbb{R}/\mathbb{Z}} = 0.$$

□

A corollary of this is as follows:

Corollary 3.2. *Let N be a prime and $0<\delta<\frac{1}{10}$. Suppose $\alpha, a\in\mathbb{R}^d$ are of denominator N, $|a|\leq M$, $\beta,\gamma\in\mathbb{R}$ and β is of denominator N, and*

$$\|\gamma + a\cdot\{\alpha h\} + \beta h\|_{\mathbb{R}/\mathbb{Z}} = 0$$

for δN many $h\in[N]$. Then either $N \ll (d\delta/M)^{-O(d)^{O(1)}}$ or there exists linearly independent $w_1,\dots,w_r$ and $\eta_1,\dots,\eta_{d-r}$ such that $\langle w_i,\eta_j\rangle = 0$ and

$$\|\eta_j\cdot\alpha\|_{\mathbb{R}/\mathbb{Z}} = 0, \quad \|w_i\cdot a\|_{\mathbb{R}/\mathbb{Z}} = 0.$$

Proof. We define $\tilde{a} = (a,1)$ and $\tilde{\alpha} = (\alpha,\beta)$. Invoking Lemma 3.2, there exists $w_1,\dots,w_r$ and $\eta_1,\dots,\eta_{d+1-r}$ such that $|w_i(\tilde{a})| = 0$, $\|\eta_j(\tilde{\alpha})\|_{\mathbb{R}/\mathbb{Z}} = 0$. We denote $w_i = (u_i,v_i)$ and $\eta_j = (\alpha_j,\beta_j)$ where the second component represents the component of the 1 and the β. Suppose $\beta_1\neq 0$. Let $\tilde{\eta_j} = \beta_j\alpha_1 - \alpha_j\beta_1$. We see that $\|\tilde{\eta_j}(\alpha)\|_{\mathbb{R}/\mathbb{Z}} = 0$. We claim that the $\tilde{\eta_j}$'s are independent of each other. Suppose there exists some a_i such that

$$\sum_{i\neq 1} a_i(\beta_i\alpha_1 - \alpha_i\beta_1) = 0.$$

We can rewrite this sum as

$$\alpha_1\left(\sum_{i\neq 1} a_i\beta_i\right) + \sum_{i\neq 1}(-a_i\beta_1)\alpha_i = 0.$$

Letting these coefficients of α_i be c_i, we see that

$$\sum_i c_i\beta_i = \beta_1\left(\sum_{i\neq 1} a_i\beta_i\right) - \sum_{i\neq 1} a_i\beta_1\beta_i = 0.$$

Thus, each of these coefficients are zero, and since β_1 is nonzero, $a_i = 0$. Thus, $\tilde{\eta_j}$'s are independent of each other. We next claim that $\tilde{\eta_j}$ are orthogonal to the u_i's. This follows since

$$\tilde{\eta_j} \cdot u_i = \beta_j \alpha_1 \cdot u_i - \beta_1 \alpha_j \cdot u_i$$
$$\eta_j \cdot w_i = \alpha_j \cdot u_i + \beta_j \cdot v_i = 0$$
$$\eta_1 \cdot w_i = \alpha_1 \cdot u_i + \beta_1 \cdot v_i = 0$$

so subtracting the second and third equations gives that the first expression is equal to zero. Finally, we claim that the u_i's are linearly independent of each other. To see this, note that (u_i, v_i) are orthogonal to $(\tilde{\eta_j}, 0)$ and (α_1, β_1). Since $(0, 1)$ is not orthogonal to (α_1, β_1), it follows that (u_i, v_i) cannot span $(0, 1)$, so $(u_i, v_i), (0, 1)$ are linearly independent of each other, which implies that u_i are linearly independent of each other.

If $\beta_i \neq 0$ for some i, we let β_i play the role of β_1 in the above argument. If $\beta_i = 0$ for all i, we simply let $\tilde{\eta_j} = \eta_j$ and proceed as follows: we see that $\|\tilde{\eta_j} \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} = 0$, and $\|u_i \cdot a\|_{\mathbb{R}/\mathbb{Z}} = 0$ with $\tilde{\eta_j}$ orthogonal to u_i. □

Let us pause now and describe what happens to the "induction on dimensions" obstruction present. In the proof above, the induction on dimensions got pushed to Minkowski's second theorem. The first proof in Appendix C has a more direct approach of handling the induction on dimensions. There, we split variables into several different variables, all but one of the variables increases linearly under each iteration. The last variable (denoted K) does not iteratively increase. Instead, it increases based on the step of the iteration we are.

It turns out that the remaining induction on dimensions that [GT1] had to run into completely disappear under the use of the refined bracket polynomial lemma. Instead, the rest of the argument relies on an induction on step and degree. This will be clearer to the reader upon reading the proofs in the next two Sections.

3.2. **Finishing the proof of Theorem 7.** We claim that $\xi([g(1), [g(1)^h])$ is rational with denominator N. To see this, note that $\xi([g(1), \{g(1)^h\}]) = \xi([g(1), [g(1)^h]])$, and since for $v, w \in \Gamma$, $\xi([v, w]) = 0$, and since $\xi([g(1), [g(1)^h]])^N = \xi([g(1)^N, [g(1)^h]]) = 1$, it follows that $\xi([g(1), [g(1)^h])$ is rational with denominator N. From (1) and Lemma *A.10*, we have for some $\beta, \gamma \in \mathbb{R}/\mathbb{Z}$ with β having denominator $O(N)$,

$$\|\beta n + \gamma + \xi([g(1), \{g(1)^h\}])\|_{\mathbb{R}/\mathbb{Z}} = 0.$$

We can write

$$\xi([g(1), \{g(1)^h\}]) = \langle (C - C^t) g(1), \{g(1)^h\} \rangle = \langle a, \{\alpha h\} \rangle$$

where $C - C^t$ is the antisymmetric matrix representing the commutator identity in the horizontal torus. Note that $C - C^t$ has numerator and denominator at most M. By a change of variables $n \mapsto M_1 n$ in the original hypothesis (using the fact that M has a modular inverse mod N since N is prime) for some $M_1 \leq O(M)$ which cancels out with the denominator of $C - C^t$,

$$|\mathbb{E}_{n \in [N]} F(g(M_1 n)\Gamma)| \geq \delta.$$

Thus, after applying the change of variables, we can assume that a, α, and β have denominator N. Applying Corollary 3.2 and noticing that $C - C^t$ is antisymmetric, we obtain w_i's and η_j's which are linearly independent, $\eta_j(w_i) = 0$, $\|M_1 \xi([w_i, g])\|_{\mathbb{R}/\mathbb{Z}} = 0$, and $\|\eta_j(g)\|_{\mathbb{R}/\mathbb{Z}} = 0$. Using the fact that N is prime, we obtain $\|\xi([w_i, g])\|_{\mathbb{R}/\mathbb{Z}} = 0$. This completes the proof.

4. The two-step polynomial sequence case

As a second warm-up problem, we shall prove the two-step polynomial sequences case:

Theorem 8. *Let N be a prime, $0 < \delta < \frac{1}{10}$, and $g(n)$ be a polynomial sequence on a two-step nilpotent Lie group that is periodic modulo N. Let $F : G/\Gamma \to \mathbb{C}$ be a vertical character with frequency ξ with $|\xi| \le (\delta/M)^{-1}$. Suppose*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \ge \delta.$$

Then either $N \ll (\delta/M)^{-O_k(d)^{-O_k(1)}}$ or else there exists some integer $d_{horiz} \ge r \ge 0$ and linearly independent elements $w_1, \dots, w_r \in \Gamma/(\Gamma \cap [G,G])$, linearly independent horizontal characters $\eta_1, \dots, \eta_{d_{horiz}-r}$ with $|w_i|, |\eta_j| \le (\delta/M)^{-O_k(d)^{O_k(1)}}$, $\langle w_i, \eta_j\rangle = 0$, and

$$\|\xi([w_i, g])\|_{C^\infty[N]}, \|\eta_j \circ g\|_{C^\infty[N]} = 0.$$

It turns out that the two-step polynomial case breaks down into two cases: one where the filtration is redundant, i.e. when $\xi([G, G_2]) \neq 0$, and one where the filtration isn't redundant. We shall take care of the latter case first:

Lemma 4.1. *Let N be a prime, $0 < \delta < \frac{1}{10}$, and $g(n)$ be a polynomial sequence on a two-step nilpotent Lie group of degree k and complexity at most M with G_2 lying in the center of G. Let $F : G/\Gamma \to \mathbb{C}$ be a nilcharacter with frequency ξ with $|\xi| \le (\delta/M)^{-1}$. Suppose*

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)| \ge \delta.$$

Then either $N \ll (\delta/M)^{-O_k(d)^{O_k(1)}}$ or else there exists some integer $d_{horiz} \ge r \ge 0$, linearly independent elements $w_1, \dots, w_r \in \Gamma/(\Gamma \cap [G,G])$, and linearly independent horizontal characters $\eta_1, \dots, \eta_{d_{horiz}-r}$ with $|w_i|, |\eta_j| \le (\delta/M)^{-O_k(d)^{O_k(1)}}$, $\langle w_i, \eta_j\rangle = 0$, and

$$\|\xi([w_i, g])\|_{C^\infty[N]}, \|\eta_j \circ g\|_{C^\infty[N]} = 0.$$

Proof. Instead of Fourier expanding along the vertical torus, we Fourier expand along the G_2-torus, that is, $G_2/(G_2 \cap \Gamma)$, so we may assume that F is a G_2 character of frequency $\xi + \xi'$ with ξ' orthogonal to the vertical $[G,G]$ direction. By taking a quotient by the kernel of $\xi + \xi'$, we may assume that G has one dimensional nonlinear component and complexity at most $(\delta/dM)^{-O(d)}$. By Lemma 2.2, we may reduce to the case of when $g(0) = 1$ and $|\psi(g(1))| \le 1/2$.

Once again, we apply the van der Corput inequality

$$|\mathbb{E}_{n\in[N]}F(g(n)\Gamma)\overline{F(g(n+h)\Gamma)}| \ge \delta^2$$

for $\delta^2 N$ many $h \in [N]$. Then letting $g_2(n) := g(n)g(1)^{-n}$ be the nonlinear part of g, we denote

$$g_h(n) = (\{g(1)^h\}^{-1}g_2(n+h)g(1)^n\{g(1)^h\}, g(n))$$

and $F_h(x,y) = \overline{F}(\{g(1)^h\}x)F(y)$. By Lemma A.3, we see that g_h lies inside $G \times_{G_2} G$ and $F_h(g_h(n))$ descends to $\tilde{F}_h(\tilde{g}_h(n)\Gamma^\square)$ a nilsequence on $G^\square$. Since $g(n)\Gamma$ is periodic modulo N, it follows that $g_h(n)\Gamma \times \Gamma$ is periodic modulo N and thus $g_h(n)\Gamma \times_{G_2\cap\Gamma} \Gamma$ is also periodic modulo N, and so $\tilde{g}_h(n)\Gamma^\square$ is periodic modulo N. The hypothesis then rearranges to

$$|\mathbb{E}_{n\in[N]}\tilde{F}_h(\tilde{g}_h(n)\Gamma^\square)| \ge \delta^2$$

for $\delta^2 N$ many $h \in [N]$. Making a change of variables for some integer $1 \le M_1 \le (10^k k)!M$

$$|\mathbb{E}_{n\in[N]}\tilde{F}_h(\tilde{g_h}(M_1 n)\Gamma^\square)| \ge \delta^2$$

By Lemma *A.3*, it follows that F_h has parameter at most $M^{O(1)}$ on $G^\square$, $G^\square$ is abelian, and so Fourier expanding $\tilde{F}_h$ into characters η, we note by Lemma *A.3* that η decomposes into $\zeta + \zeta_2$ where ζ is a horizontal character on G which annihilates G_2 and ζ_2 is a character on G_2. Note that we can take ζ_2 to be $\xi + \xi'$ because F is a G_2-character of frequency $\xi + \xi'$. Note also that $(\xi + \xi')([g, h]) = \xi([g, h])$ since ξ' is orthogonal to the vertical direction. It thus follows from the one-step case that

$$\|\zeta(g(M_1 n)) + (\xi + \xi')(g_2(M_1 n + h)) - (\xi + \xi')(g_2(M_1 n)) + \xi([g(1)^{M_1 n}, \{g(1)^h\}])\|_{C^\infty[N]} = 0.$$

The point of making the change of variables is so that by Lemma *A.10*, the coefficients of $g_2(M_1\cdot)$ have denominator N, and $\xi([g(1)^{M_1 n}, \{g(1)^h\}])$ consists of $\langle an, \{\alpha h\}\rangle$ where a and α have denominator N. Applying Corollary 3.2 and using the fact that N is prime yields w_i's and η_j's which satisfy the conclusions of the lemma. □

Proof of Theorem 8. We will once again quotient by the kernel of ξ. Suppose G_2 does not lie in the center of G. If it does, we may apply Lemma 4.1 to finish. If F was not already a G_k-character, we may Fourier expand via Lemma *A.6* and pigeonhole in one character, we may replace F with a character on $G_{(s)}G_k$ with frequency $\xi + \xi'$ with ξ' orthogonal to the vertical direction in G. By Lemma 2.2, we may reduce to the case of when $g(0) = 1$ and $|\psi(g(1))| \le \frac{1}{2}$. Then by van der Corput, we have for $\delta^{O(1)}N$ many h's

$$|\mathbb{E}_{n\in[N]}F(g(n+h)\Gamma)\overline{F(g(n)\Gamma)}| \ge \delta^{O(1)}.$$

Defining

$$g_h(n) = (\{g(1)^h\}^{-1}g_2(n+h)g(1)^n\{g(1)^h\}, g_2(n)g(1)^n)$$

and $\tilde{F}_h(x, y) = F(\{g(1)^h\}x)\overline{F(y)}$, it follows that

$$|\mathbb{E}_{n\in[N]}\tilde{F}_h(g_h(n))| \ge \delta^{O(1)}.$$

It follows that $g_h(n)$ lies in $G \times_{G_2} G = \{(g, g') : g'g^{-1} \in G_2\}$. This group has the filtration $(G \times_{G_2} G)_i = G_i \times_{G_{i+1}} G_i$ with $(G \times_{G_2} G)_k = G_k^\triangle$. However, $\tilde{F}_h$ is $G_k^\triangle$-invariant, so descends via a quotient by $G_k^\triangle$ to a degree $d-1$ nilsequence. In order to apply the induction hypothesis, we shall need to Fourier expand $\tilde{F}_h$ into nilcharacters on $G^\square$. Since ξ doesn't annihilate $[G, G_2]$, it follows that $G^\square$ is two-step with vertical direction $G_{(2)} \times_{G_{(2)}} G_{(2)}$. Thus, we can write each vertical character χ there as

$$\chi(gh, g) = \chi^1(g) + \chi^2(h)$$

for all $g \in G_{(2)}$ and $h \in [G, G_2]$. To emphasize that χ^2 comes from the $x_1 - x_2$ direction, we shall write χ^2 as $\chi^2 \otimes \overline{\chi^2}$. Here, since F is a nilcharacter of frequency ξ, it follows that $\tilde{F}_h$ is already a nilcharacter of frequency $\xi \otimes \overline{\xi}$. Pigeonholing in h and applying the induction hypothesis we find w_i's and η_j's such that for $c_1(\delta)N$ many $h \in [N]$ that

$$\|\xi \otimes \overline{\xi}([w_i, \tilde{g_h}(n)])\|_{C^\infty[N]} = 0$$

$$\|\eta_j \circ \tilde{g_h}(n)\|_{C^\infty[N]} = 0$$

where $\tilde{g_h}$ is the projection of g_h to $G^\square$. Here, we remind the reader that $c_1(\delta)$ as defined in Section 2.1 is any quantity lower bounded by $\gg (\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$. Letting $H = G \times_{G_2} G$, we see that the horizontal component of H is $H/[H, H]$ while the horizontal component of

$G^\square = H/G_k^\triangle$ can be identified with $H/[H,H]G_k^\triangle$ (since $[H/G_k^\triangle, H/G_k^\triangle] = [H,H]G_k^\triangle/G_k^\triangle$ by viewing the coset equivalence). It follows that each η_j can be identified with $\tilde{\eta_j}$, a horizontal character on H which annihilates $G_k^\triangle$. In addition, each $w \in (\Gamma \times_{G_2} \Gamma)/([H,H] \cap \Gamma \times_{G_2} \Gamma)$ which descends to an element orthogonal to the η_j's are orthogonal to the $\tilde{\eta_j}$'s since $\tilde{\eta_j}$'s all annihilate $G_k^\triangle$. Also, we see that $w_i \mapsto \xi \otimes \overline{\xi}([w_i, \tilde{g_h}(n)])$ does not depend on the $G_k^\triangle$ component of w_i. Thus, we may find linearly independent horizontal characters η_j (which are the $\tilde{\eta_j}$'s defined above that annihilate $G_k^\triangle$) on H and linearly independent w_i on the lattice of H such that $\eta_j(w_i) = 0$ and

$$\|\xi \otimes \overline{\xi}([w_i, g_h(n)])\|_{C^\infty[N]} = 0$$

$$\|\eta_j \circ g_h(n)\|_{C^\infty[N]} = 0.$$

Since the two-step part only consists of $\xi \otimes \overline{\xi}$, we in fact have

$$\|\xi \otimes \overline{\xi}([w_i, g_h(n)])\|_{C^\infty[N]} = 0.$$

Note that $(g(n), g(n)g_2(n+h)g_2(n)^{-1}p(n,h))$ where p is a (bracket) polynomial in n and h which is an element in $[G,G]$. We write $w_i = (u_iv_i, u_i)$ and $\eta_j = (\alpha_j, \beta_j)$ so

$$\begin{aligned}
\eta_j(g_h(n)) &= \alpha_j(g(n)) + \beta_j(g_2(n+h)g_2(n)^{-1}) \\
\eta_j(w_i) &= \alpha_j(u_i) + \beta_j(v_i) = 0 \\
\xi \otimes \overline{\xi}([w_i, g_h(n)]) &= \xi \otimes \xi^{-1}(([u_i, g(n)], [u_iv_i, g(n)][u_iv_i, g_2(n+h)g_2(n)^{-1}]) \\
&= \xi([v_i, g(n)] + [u_iv_i, g_2(n+h)g_2(n)^{-1}])
\end{aligned}$$

where $\alpha_j \in G$, $\beta_j \in G_2$ and $v_i \in G_2$, and $u_i \in G$ (note that after quotienting by ξ that since β_j annihilates $[G, G_2] = [G, G]$, it follows that β_j annihilates the bracket polynomial part). Using Vinogradov's lemma (Lemma *D.2*), it follows that

$$\alpha_j(g(n)), \xi([v_i, g(n)]), \beta_j(g_2(n)), \xi([u_iv_i, g_2(n)]) \equiv 0 \pmod 1.$$

Note that here we must use the fact that N is prime and the fact that g_2 has horizontal component with denominator N to eliminate the binomial coefficients that come from expanding out $g_2(n+h)g_2(n)^{-1}$. Let $\tilde{G} = \{g \in G : \alpha_j \circ g = 0, [v_i, g] = 0\}$ and $\tilde{G}_2 = \{g \in \tilde{G}, \beta_j(g) = 0, [u_iv_i, g] = 0\}$. We claim that $[\tilde{G}, \tilde{G}_2] = 0$. To show this, we let $\tilde{H} = \tilde{G} \times_{G_2} \tilde{G}/[G,G]^\triangle$ and we claim is that $\tilde{H}$ is Abelian. To see this, note that each element of $\tilde{H}$ of the form (gg_2, g) satisfies $\alpha_j(g) + \beta_j(g_2) = 0$ and $[v_i, g] + [u_iv_i, g_2] = 0$. Since $\alpha_j(u_i) + \beta_j(v_i) = 0$ for each i, j, it follows that (gg_2, g) can be generated by (u_iv_i, u_i) modulo $[G,G]^2$. However, for any other (hh_2, h) in $\tilde{H}$, we have $\xi \otimes \overline{\xi}[(u_iv_i, u_i), (hh_2, h)] = 0$. Hence $\tilde{H}$ is Abelian. Finally, we have $\xi \otimes \overline{\xi}([(g,g), (hg_2, h)]) = \xi([g, g_2]) = 0$ whenever $g \in \tilde{G}$ and $g_2 \in \tilde{G}_2$. This shows that $\xi([\tilde{G}, \tilde{G}_2]) = 0$, so we may make our filtration finer by restricting to $\tilde{G}$ and $\tilde{G}_2$. Since we've quotiented by the kernel of ξ, we work with $[\tilde{G}, \tilde{G}_2] = 0$. Thus, by Lemma *A.1* and Lemma *A.2*, we can write $g(n)\Gamma = g_1(n)\gamma(n)\Gamma$ with g_1 being a polynomial sequence with the filtration $\tilde{G}_i$ with $\tilde{G}_i$ satisfying $\tilde{G}_i = \tilde{G}_2 \cap G_i$ for $i \geq 2$. Finally, we apply Lemma *A.11*, Lemma 4.1, and Lemma *A.8* to $g_1(Pk!n)$ (with P the period of γ) to find $\alpha_1, \dots, \alpha_{d'}$ with the property that we can factorize $g_1(Pk!n) = g_2(n)\gamma(n)$ such that g_2 lies inside $\bigcap_i \ker(\alpha_i)$, which is 1-step and γ is P'-rational. Thus, $g(k!PP'n)\Gamma$ lies inside $\bigcap_i \ker(\alpha_i)$, and so invoking Lemma 2.1, the first conclusion is satisfied. To satisfy the second conclusion, we see that $g(k!PP'n)$ can be generated by vectors orthogonal to α_i's and Γ.

Thus, for any $w \in \Gamma/(\Gamma \cap [G,G])$ orthogonal to all of the α_i's, it follows from Lemma 2.1 that

$$\|\xi([w,g])\|_{C^\infty[N]} = 0.$$

Using Lemma *A.7* we may pick a basis for all such w with size at most $c_1(\delta)^{-1}$. □

5. Periodic Polynomial Nilsequences

In this section, we deduce Theorem 3, which will be deduced from Theorem 9. An example in the beginning of Section 7 in [GT1] indicates that an obstruction that may occur is that our filtration may be redundant. To overcome the obstruction, [GT1] proceeds via an induction on dimensions, lowering the dimension of G_2 one-by-one until no such obstruction exists. The downside to using induction on dimensions is that it has a habit of incurring losses double exponential in dimension, which is what happens if one proceeds along the lines of [GT1]. Our proof here completely sidesteps the induction on dimensions obstruction the [GT1] runs into, and instead shifts the induction on dimension to an induction on degree or step. We do, however, run into a similar phenomenon of the filtration being redundant, causing us to have to divide our proof into two cases. In the two-step polynomial sequence case, we also needed to consider two cases. The first case is where G_2 lies in the center of G, that is, when $[G_2, G] = 0$. The second case is when this does not occur. Roughly speaking, the second case corresponds to the obstruction that Green and Tao run into of the filtration being redundant. In the proof of the second case, we reduce to the first case. Similarly, in the general polynomial nilsequences case on a s-step nilmanifold, we divide our proof into two cases: one where $[G_2, G, G, \dots, G] = 0$ where $s-1$ commutators are taken and one where this does not occur. Similar to the two-step nilmanifold case, we shall reduce the second case to the first case (see Section 2.1 for asymptotic notation such as $c_1(\delta)$).

Theorem 9. *Let $0 < \delta < \frac{1}{10}$, N a prime, $M \geq 1$, and $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz vertical character of nonzero frequency ξ on an s-step nilmanifold G/Γ and Lipschitz parameter ≤ 1 with $|\xi| \leq (\delta/M)^{-1}$. Suppose $g(n)$ is a polynomial sequence in G, and that*

$$|\mathbb{E}_{n\in[N]} F(g(n)\Gamma)| \geq \delta$$

with $g(n)\Gamma$ being N-periodic. Then either $N \ll (\delta/M)^{-O_{k,s}(d)^{O_{k,s}(1)}}$ or else there exists some integer $0 \leq r \leq d_{horiz}$ and a set of linearly independent $\eta_1, \dots, \eta_r$ of size at most $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$ such that

$$\|\eta_i \circ g\|_{C^\infty[N]} = 0$$

and such that for any $w_1, \dots, w_{s-1} \in \Gamma/(\Gamma \cap [G,G])$ orthogonal to all of the η's,

$$\|\xi([g, w_1, \dots, w_{s-1}])\|_{C^\infty[N]} = 0.$$

In other words, after taking a quotient by the kernel of ξ, g is contained (up to some periodic part) in an $s-1$-step nilmanifold of some higher complexity.

Remark. See section two for the definition of "orthogonal." Note that $\tilde{G} := \{\eta_i(g) = 0, [g, w_1, \dots, w_{s-1}] = 0$ for all such $w_i's\}$ is $s-1$-step. This is because if $g_1, \dots, g_s \in \tilde{G}$, then each $g_1, \dots, g_{s-1}$ is generated by w_j's modulo $G_{(2)}$ since they are orthogonal to the η_i's. Thus, $[g_1, [g_2, \dots, [g_{s-1}, g_s]]]$ can be written as a combination of $[w_{j_1}, [w_{j_2}, \dots, [w_{j_{s-1}}, g_s]]] = 0$ by definition.

Thus, Theorem 3 follows from Theorem 9. To prove this, we will need the following preliminary lemma:

Lemma 5.1. *Suppose $\xi([G_2, G, \cdots, G]) = 0$ (with the commutator being taken $s-1$ times). Then the same is true if we swap any one of the G's with the G_2.*

Proof. Our main tool for proving this is the Hall-Witt identity, which states that if $x^y = x[x,y]$, then
$$[[x,y],z^x][[z,x],y^z][[y,z],x^y] = 1.$$
Since $[G_2, G, \dots, G]$ lies inside $G_{(s)}$, the commutator term $[x,y]$ is a "lower order term" and so the Hall-Witt identity we morally work with is
$$[[x,y],z][[z,x],y][[y,z],x]\text{``}="1$$
or rather
$$[[x,y],z][[z,x],y][[y,z],x] = 1 \pmod{G_{(4)}}.$$
We claim by induction that for $0 \le r \le k-1$ that if $g_k \in G_2$ and $g_j \in G$ for all other j that
$$[[g_1, \dots, g_{k-r-1}], [g_k, g_{k-1}, \dots, g_{k-r}]] \in [G_2, G, \dots, G] \pmod{G_{(k+1)}}.$$
By taking commutators outside, this would imply that if any of the j's lie inside G_2, then $[g_1, \dots, g_k] \in [G_2, \dots, G] \pmod{G_{(k+1)}}$ from which the lemma follows. The case of $k = 1, 2$ are trivial and $k = 3$ follows from the Hall-Witt identity. Suppose this holds for $1, \dots, k$. We wish to show this for $k+1$. The case of $r = k$ follows straight from the fact that $[g_k, \dots, g_1]$ lies inside $[G_2, \dots, G]$. We will show this by backwards induction on r as well, i.e., assuming the case for $\ge r$, we wish to show
$$[[g_1, \dots, g_{k-r}], [g_k, g_{k-1}, \dots, g_{k-r+1}]] \in [G_2, G, \dots, G] \pmod{G_{(k+1)}}.$$
By the Hall-Witt identity, it suffices to show that
$$[[[g_1, \dots, g_{k-r-1}], [g_k, g_{k-1}, \dots, g_{k-r+1}]], g_{k-r}] \in [G_2, G, \dots, G] \pmod{G_{(k+1)}}$$
and
$$[[g_1, \dots, g_{k-r-1}], [g_k, \dots, g_{k-r}]] \in [G_2, G, \dots, G] \pmod{G_{(k+1)}}$$
both of which are true by induction hypothesis.

∎

The start of the proof for either case is the same. We first reduce via Lemma 2.2 to the case of when $g(0) = 1$ and $|\psi(g(1))| \le \frac{1}{2}$ and quotient out by the kernel of ξ to make a one-dimensional vertical torus. We then apply van der Corput's inequality to obtain
$$|\mathbb{E}_{n \in [N]} F(g(n)\Gamma)\overline{F(g(n+h)\Gamma)}| \ge \delta^2$$
for $\delta^2 N$ many $h \in [N]$. Then letting $g_2(n) := g(n)g(1)^{-n}$ be the nonlinear part of g, we denote
$$g_h(n) = (\{g(1)^h\}^{-1} g_2(n+h) g(1)^n \{g(1)^h\}, g(n))$$
and $F_h(x,y) = \overline{F}(\{g(1)^h\}x)F(y)$. The hypothesis then rearranges to
$$|\mathbb{E}_{n \in [N]} F_h(g_h(n))| \ge \delta^2$$
for $\delta^2 N$ many $h \in [N]$. Next, we make a change of variables $n \mapsto M_1 n$ with some $M_1 \le M^{O_{s,k}(1)}$ to eliminate all of the rationality coming from the Lie bracket and so that $g_2(M_1 n)$ has denominator N. This can be done via Lemma *A.10*. Note that since $g(n)\Gamma$ is periodic modulo N, that $g_h(n)(\Gamma \times \Gamma)$ is also periodic modulo N. Define $G \times_{G_2} G$ as the group

$\{(g,g') : g^{-1}g' \in G_2\}$. This is an $\leq s$-step nilpotent Lie group and has a natural filtration of $G_i \times_{G_{i+1}} G_i$. Since $[G,G] \subseteq G_2$, it follows that $g_h(n) \in G \times_{G_2} G$. Since F is a nilcharacter on G/Γ of frequency ξ, and since G_k lies in the center of G and F_h is invariant under $G_k^{\triangle}$, it follows that F_h descends to a function on $G \times_{G_2} G/G_k^{\triangle} = G^{\square}$. By Lemma *A.3*, it follows that $G \times_{G_2} G$ has complexity at most $M^{O_{s,k}(1)}$, and given a horizontal character $\eta : G \times_{G_2} G \to S^1$, we may decompose it as $\eta(gg_2, g) = \eta_1(g) + \eta_2(g_2)$ where $g \in G$ and $g_2 \in G_2$ where η_1 is a horizontal character on G and η_2 is a horizontal character on G_2. The same lemma tells us that if our horizontal character has size at most $c_1(\delta)^{-1}$, then η_1 and η_2 has size at most $c_1(\delta)^{-1}$.

5.1. **Intuition for the argument.** Here, we provide a brief sketch of the argument. The argument proceeds via an induction on degree or step. If $[G_2, G, \dots, G] = 0$, then $G^{\square}$ is $s-1$-step, so we may use induction to show that g_h lies in an $s-2$-step subgroup. A candidate for such a subgroup is $\tilde{G}^{\square} := \tilde{G} \times_{\tilde{G}_2} \tilde{G}/\tilde{G}_{(s-1)}^{\triangle}$ where $\tilde{G}$ is $s-1$-step and $\tilde{G}_2 = \tilde{G} \cap G_2$. It's possible that not all $s-2$-step subgroups of $\tilde{G}$ are of this form (and it's also possible that this is not even $s-2$-step), but we can show via Vinogradov's lemma and the refined bracket polynomial lemma that after pigeonholing in h, g_h lies inside a subgroup of bounded rationality of that form. We can then use this to show that g lies inside $\tilde{G}$ which is $s-1$-step.[7]

If $[G_2, G, \dots, G] \neq 0$ (where the commutator is taken $s-1$ times), then $G^{\square}$ is not $s-1$-step, but rather s-step so an induction on degree tells us that $g_h(n)$ lies in an $s-1$-step subgroup of $G^{\square}$. A candidate subgroup is $\tilde{G}^{\square} := \tilde{G} \times_{\tilde{G}_2} \tilde{G}/\tilde{G}_{(s)}^{\triangle}$ where $[\tilde{G}_2, \dots, \tilde{G}] = 0$. These are probably not the only $s-1$-step subgroups of $G^{\square}$, but the structure of $g_h(n)$ allows us to use Vinogradov's lemma to guarantee that after pigeonholing in h, $g_h(n)$ actually does lie inside some group of the form $\tilde{G}^{\square}$ with bounded rationality. Using this, we can show that g lies inside the filtered nilpotent Lie group $\tilde{G}$ where $\tilde{G}_0 = \tilde{G}_1 = \tilde{G}$ and $\tilde{G}_i = \tilde{G}_2 \cap G_i$ for $i \geq 2$.

5.2. **Case 1.** Suppose first that

$$\xi([G_2, G, G, \cdots, G]) = 0.$$

Since we took a quotient of the vertical component by the kernel of ξ, by Lemma 2.3, we are in the case that $[G_2, G, G, \cdots, G] = 0$. We will be inducting on step rather than degree for our argument. Replacing G_ℓ for all $k \geq \ell \geq 2$ with $G_\ell G_{(s)}$ (and noting that this still preserves normality and the filtration property since $G_{(s)}$ lies in the center of G), and taking a Fourier expansion to G_k-characters as in Lemma *A.6* and pigeonholing in one character, we may replace F with a character on G_k with frequency $\xi + \xi'$ with ξ' orthogonal to the vertical direction in G. In order to apply induction for our argument, we must analyze nilcharacters on $G \times_{G_2} G$. First note that since $\xi([G_2, G, G, \cdots, G]) = 0$ and ξ' is orthogonal to the vertical direction on G, it follows that F_h descends to a $\leq s-1$-step nilsequence on $G^{\square}$. In order to apply induction in this case, we must understand nilcharacters on $G^{\square}$. The vertical torus is $G_{(s-1)} \times_K G_{(s-1)}/G_{(s)}^{\triangle}$ where $K = [G_2, G, G, \cdots, G]$ where the commutator is taken $s-2$ times. Thus, if χ is a vertical character on $G_{(s-1)} \times_K G_{(s-1)}/G_{(s)}$, it follows that

$$\chi(gh, g) = \chi^1(g) + \chi^2(h)$$

[7]Technically, this description is a bit too strong for us to hope for, and is not exactly what happens in our proof. Instead, what is shown is the bit weaker fact that g lies inside some $\tilde{G}$ which is $s-1$-step.

for all $g \in G_{(s-1)}$ and $h \in K$. To emphasize that χ^2 lies in a $x_1 - x_2$ direction, we write χ^2 as $\chi^2 \otimes \overline{\chi^2}$. We may further decompose $\chi^2(h)$ as the sum of a vertical component χ^2_{vert} and a complementary component χ^2_{comp}. Going back to the problem, since F is a vertical character with frequency ξ, it follows that in the Fourier expansion of F_h in $G^\square$, we can take χ^2_{vert} to be ξ. By induction (the base case being two-step polynomial sequences), Fourier expanding F_h via Lemma A.6 into $\xi \otimes \overline{\xi} + \zeta$ (where ζ is a horizontal character on $G_{(s-1)}$ that annihilates $[G, G_{(s-1)}]$), and pigeonholing in h, there exists horizontal characters $\eta_1, \dots, \eta_r$ of size at most $c_1(\delta)^{-1}$ such that

$$\|\eta_i(\tilde{g_h}(M_1n))\|_{C^\infty[N]} = 0$$

and such that for any $w_1, \dots, w_{s-2} \in \Gamma^\square/(\Gamma^\square \cap [G^\square, G^\square])$ which are orthogonal to the η_i's, we have

$$\|(\zeta + \xi \otimes \bar{\xi})([\tilde{g_h}(M_1n), w_1, w_2, \dots, w_{s-2}])\|_{C^\infty[N]} = 0.$$

Letting $H = G \times_{G_2} G$, we see that the horizontal component of H is $H/[H, H]$ while the horizontal component of $G^\square = H/G_k^\triangle$ can be identified with $H/[H, H]G_k^\triangle$ (since $[H/G_k^\triangle, H/G_k^\triangle] = [H, H]G_k^\triangle/G_k^\triangle$ by viewing the coset equivalence). It follows that each η_j can be identified with a horizontal character $\tilde{\eta_j}$ on H which annihilates $G_k^\triangle$. Each $w \in (H \cap \Gamma \times \Gamma)/([H, H] \cap \Gamma \times \Gamma)$ that descends to an element orthogonal to each of the η_j's will be annihilated by $\tilde{\eta_j}$. In addition, the map $(w_1, \dots, w_{s-1}) \mapsto \xi \otimes \bar{\xi}([g_h(M_1n), w_1, \dots, w_{s-1}])$ does not depend on the $G_k^\triangle$ component of the w_i's. Thus, we may find $\eta_1, \dots, \eta_r$ (which are the $\tilde{\eta_j}$'s defined above) which are horizontal characters on $G \times_{G_2} G$ such that for any orthogonal $w_1, \dots, w_{s-1}$ to the η_j's, we have for $c_1(\delta)N$ many $h \in [N]$ that

$$\|\eta_i(\tilde{g_h}(M_1n))\|_{C^\infty[N]} = 0$$

$$\|(\zeta + \xi \otimes \bar{\xi})([\tilde{g_h}(M_1n), w_1, w_2, \dots, w_{s-2}])\|_{C^\infty[N]} = 0.$$

Writing $\eta_i = \eta_i^1 + \eta_i^2$ via Lemma A.3, and $w_i = (u_iv_i, u_i)$, we have

$$\|\eta_i^1(g(M_1n)) + \eta_i^2([g(1)^{M_1n}, \{g(1)^h\}]) + \eta_i^2(g_2(M_1n + h)) - \eta_i^2(g_2(M_1n))\|_{C^\infty[N]} = 0$$

$$\|[\deg_n \neq 1 \text{ terms}] + \alpha nh + \xi([g(1)^{M_1n}, \{g(1)^h\}, u_1, u_2, \dots, u_{s-2}]) + M_1\beta n\|_{C^\infty[N]} = 0$$

for some $\alpha, \beta \in \widehat{\mathbb{Z}/N\mathbb{Z}}$. Here, $[\deg_n \neq 1 \text{ terms}]$ denotes a polynomial in n and h where there are no terms in the n variable which are degree one. By linear algebraic manipulations and Cramer's rule (in particular Lemma A.7), we may assume without a loss of generality that the η_i's are consists of orthogonal vectors to u_j's and have a zero η_i^2 component (since those are certainly orthogonal to the w_i's), then to v_j's and have a zero η_i^1 component (again since those are orthogonal to the w_i's), and the remaining have both η_i^1 and η_i^2 components, with everything having size bounded by $c_1(\delta)^{-1}$. This implies via various applications of Vinogradov's lemma and Lemma A.10 that for all η_i^1 orthogonal to all of the u_j's (and have zero η_i^2 component) that

$$\|\eta_i^1 \circ g\|_{C^\infty[N]} = 0$$

$$\|\alpha h + \xi([g(1)^{M_1}, \{g(1)^h\}, u_1, u_2, \dots, u_{s-2}])\|_{\mathbb{R}/\mathbb{Z}} = 0$$

for $c_1(\delta)N$ many h. We now shift our focus to the linear dimensions G/G_2. Let $\tilde{u_i}$ be the linear dimension part of u_i and let $\tilde{\eta_i^1}$ the linear part of η_i^1. By Lemma 5.1,

$$\xi([g(1)^{M_1n}, \{g(1)^h\}, \widetilde{u_1}, \widetilde{u_2}, \dots, \widetilde{u_{s-2}}])$$

only depends on the linear part of the u_i's and the linear part of the $g(1)^n$ and $\{g(1)^h\}$. Applying Corollary 3.2, for each choice of $\widetilde{u_1}, \widetilde{u_2}, \dots, \widetilde{u_{s-2}}$, there exists $x_1, \dots, x_r$ horizontal characters that annihilate G_2 and $y_1, \dots, y_{d_{lin}-r}$ inside $\Gamma/(\Gamma \cap G_2)$ (with d_{lin} the linear dimension $\dim(G/G_2)$) which are independent, x_i's are orthogonal to y_j's (i.e. $x_i(y_j) = 0$ for all i and j) and such that

$$x_i \circ g \equiv 0 \pmod 1, \xi([y_i, g, \widetilde{u_1}, \widetilde{u_2}, \dots, \widetilde{u_{s-2}}]) \equiv 0 \pmod 1.$$

We see that by Cramer's rule (i.e., Lemma A.7), we can pick linearly independent $(u_i v_i, u_i)$ which span the subspace orthogonal to the η_j's with u_i and v_i having size at most $c_1(\delta)^{-1}$. Define

$$\tilde{G} = \{g \in G : x_\ell(g) = 0, [y_k, g, \widetilde{u_1}, \widetilde{u_2}, \dots, \widetilde{u_{s-2}}] = 0 \text{ for all such } u_i, \eta_j^1(g) = 0 \text{ for all j, k, } \ell\}.$$

We see from u_i and v_i having size at most $c_1(\delta)^{-1}$ implies that $\widetilde{u_j}$ also has size at most $c_1(\delta)^{-1}$ and hence $\tilde{G}$ has rationality at most $c_1(\delta)^{-1}$ and therefore by Lemma A.11 has complexity at most $c_1(\delta)^{-1}$. We claim that $\tilde{G}$ is $s-1$-step. To see this, letting $g_1, \dots, g_{s-1}$ be elements in $\tilde{G}$, we wish to show that

$$[g_{s-1}, g_{s-2}, \dots, g_1] = 0.$$

Note that the right hand side of the equation only depends on the linear dimensions of g_i, which by definition and orthogonality can be generated by vectors in the orthogonal complement of $\tilde{\eta_j^1}$'s, and for $\tilde{u}_i$ an annihilator of the $\tilde{\eta_j^1}$'s, (u_i, u_i) lies inside the annihilator of η_j. In addition, g_{s-2} is orthogonal to x_i, so can be generated by y_i's modulo G_2. Thus, this amounts to checking that

$$[y_i, g_{s-1}, \widetilde{u_1}, \widetilde{u_2}, \dots, \widetilde{u_{s-2}}] = 0$$

which is true by definition. This completes the proof of Case 1 since $\|\eta_j \circ g\|_{C^\infty[N]} = 0$ implies that $\|\tilde{\eta_j^1}(g(1))\|_{\mathbb{R}/\mathbb{Z}} = 0$ so by Lemma A.1, and since we quotiented out by the kernel of ξ, $g(n)\Gamma$ is contained, up to a periodic part in an $s-1$-step nilmanifold and we can write $g(n) = g_1(n)\gamma(n)$ with $\gamma(n)\Gamma$ being P-rational with some $P \le c_1(\delta)^{-1}$. We can then apply Lemma A.8 obtaining characters $\alpha_1, \dots, \alpha_r$ such that $\|\alpha_i \circ g(Pk!\cdot)\|_{C^\infty[N]} = 0$ and since $Pk! \ll N$, it is relatively prime to N, we have $\|\alpha_i \circ g\|_{C^\infty[N]} = 0$. Furthermore, it follows that if $\beta_1, \dots, \beta_{s-1}$ are orthogonal horizontal characters to α_i, it follows that

$$\|\xi([g(Pk!\cdot)\Gamma, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = \|\xi([g_1(Pk!\cdot)\Gamma, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = 0$$

so

$$\|\xi([g, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = 0.$$

5.3. **Case 2.** Suppose now that

$$\xi([G_2, G, G, \dots, G]) \neq 0.$$

Our goal for case two is to reduce to case one. In this case, we shall need to replace G_ℓ with $G_\ell G_{(s)}$ for $k \ge \ell \ge s$ so that G_k contains $G_{(s)}$. Note that this preserves normality and the filtration property since $G_{(s)}$ lies in the center of G. Fourier expanding F_h via Lemma A.6 and pigeonholing in one of the Fourier coefficients, we may assume (at the cost of replacing δ with $c_1(\delta)$ and M with $M^{O_{s,k}(1)}$) that F_h is a G_k-character with frequency $\xi + \xi'$. This time, while $G \times_{G_2} G$ is no longer $s-1$-step, F_h now annihilates $(G \times_{G_2} G)_k = G_k^\triangle$, so $F_h(g_h(n)\Gamma)$ descends to a degree $k-1$ nilsequence on $G^\square/\Gamma^\square$. The vertical torus of $G_{(s)} \times_K G_{(s)}$ where $K = [G_2, G, \dots, G]$ with the commutator being taken $s-1$ times, F_h is a nilcharacter on

$G^\square$ of frequency $\xi \otimes \overline{\xi}$. By our induction on degree and pigeonholing in h, we find $\eta_1, \dots, \eta_r$ such that for any orthogonal $w_1, \dots, w_{s-1} \in \Gamma^\square/(\Gamma^\square \cap [G^\square, G^\square])$ to the η_j's we have

$$\|\eta_i(\tilde{g_h}(M_1 n))\|_{C^\infty[N]} = 0$$

$$\|\xi \otimes \bar{\xi}([\tilde{g_h}(M_1 n), w_1, \dots, w_{s-1}])\|_{C^\infty[N]} = 0$$

where $\tilde{g_h}$ is the projection of g_h to $G^\square$. Letting $H = G \times_{G_2} G$, we see that the horizontal component of H is $H/[H,H]$ while the horizontal component of $G^\square = H/G_k^\triangle$ can be identified with $H/[H,H]G_k^\triangle$ (since $[H/G_k^\triangle, H/G_k^\triangle] = [H,H]G_k^\triangle/G_k^\triangle$ by viewing the coset equivalence). It follows that each η_j can be identified with a horizontal character $\tilde{\eta_j}$ on H which annihilates $G_k^\triangle$. Each $w \in (H \cap \Gamma \times \Gamma)/([H,H] \cap \Gamma \times \Gamma)$ that descends to an element orthogonal to each of the η_j's will be annihilated by $\tilde{\eta_j}$. In addition, the map $(w_1, \dots, w_{s-1}) \mapsto \xi \otimes \bar{\xi}([g_h(M_1 n), w_1, \dots, w_{s-1}])$ does not depend on the $G_k^\triangle$ component of the w_i's. Thus, we may find $\eta_1, \dots, \eta_r$ (which are the $\tilde{\eta_j}$'s defined above) which are horizontal characters on $G \times_{G_2} G$ such that for any orthogonal $w_1, \dots, w_{s-1}$ to the η_j's, we have for $c_1(\delta)N$ many $h \in [N]$ that

$$\|\eta_i(g_h(M_1 n))\|_{C^\infty[N]} = 0$$

$$\|\xi \otimes \bar{\xi}([g_h(M_1 n), w_1, \dots, w_{s-1}])\|_{C^\infty[N]} = 0.$$

Decomposing the η_i's as in Lemma A.3, we obtain

$$\|\eta_i^1(g(M_1 n)) + \eta_i^2([g(1)^{M_1 n}, \{g(1)^h\}]) + \eta_i^2(g_2(M_1 n + h)) - \eta_2(g_2(M_1 n))\|_{C^\infty[N]} = 0.$$

We choose (via Cramer's rule or Lemma A.7) linear independent $(\tilde{u}_i\tilde{v}_i, \tilde{u}_i)$ of size at most $c_1(\delta)^{-1}$ to be orthogonal vectors in $G^\square$ that are orthogonal to the η_i's and span the subspace orthogonal to the η_i's. Letting where $\tilde{v}_i^1 = \psi_{horiz}(v_i)$, we see that using Lemma A.7 that we may replace η_i's with orthogonal vectors to $(\tilde{u}_i\tilde{v}_i, \tilde{u}_i)$ with a property that a subset of them contains $(\eta_i^1, \eta_i^{2,1})$ and that $(\eta_i^1, \eta_i^{2,1})$ annihilates all of $(\tilde{u}_i\tilde{v}_i{}^1, \tilde{u}_i)$ and $\eta_i^{2,1}$ is a horizontal character on G_2 that annihilates $[G,G]$ and such that $(\eta_i^1, \eta_i^{2,1})$ span the subspace of the dual space of $G \times_{G_2} G/[G,G]^2$ orthogonal to the $(\tilde{u}_i\tilde{v}_i{}^1, \tilde{u}_i)$'s, and that each of the η_i's have size at most $c_1(\delta)^{-1}$. Vinogradov's lemma (applied to the constant term in the polynomial in h) and the fact that N is prime gives us that for $\eta_i = (\eta_i^1, \eta_i^{2,1})$ that

$$\|\eta_i^1(g)\|_{C^\infty[N]} = 0.$$

The second hypothesis rearranges to (observing that the $[g(1)^n, \{g(1)^h\}]$ term disappears under $s-1$ commutators) for $c_1(\delta)N$ many $h \in [N]$ that for some horizontal character ζ (corresponding to the $2^{s-1}-1$ leftover terms of containing the v_i's in the commutators),

$$\|\zeta(g(M_1 n)) + \xi([g_2(M_1 n + h)g_2(n)^{-1}, \widetilde{u_{j_1}}\widetilde{v_{j_1}}, \dots, \widetilde{u_{j_{s-1}}}\widetilde{v_{j_{s-1}}}])\|_{C^\infty[N]} = 0$$

for each $j : \{1, \dots, s-1\} \to \{1, \dots, d'-r\}$ where d' is the dimension of the horizontal torus of $G \times_{G_2} G$. Applying Vinogradov's lemma and using the fact that N is prime (so that we can eliminate the binomial coefficient and the power of M_1 in front of each coefficient when expanding out $g_2(M_1 n + h)g_2(n)^{-1}$ in coordinates) yields

$$\|\xi([g_2, \widetilde{u_{j_1}}\widetilde{v_{j_1}}, \dots, \widetilde{u_{j_{s-1}}}\widetilde{v_{j_{s-1}}}])\|_{C^\infty[N]} = 0.$$

We now define

$$\tilde{G} = \{g \in G : \eta_i^1(g) = 0, \forall i \text{ such that } \eta_i = (\eta_i^1, \eta_i^{2,1})\}$$

$$\tilde{G}_2 = \{g \in \tilde{G} : [g, \widetilde{u_{j_1}}\widetilde{v_{j_1}}, \widetilde{u_{j_2}}\widetilde{v_{j_2}}, \dots, \widetilde{u_{j_{s-1}}}\widetilde{v_{j_{s-1}}}] = 0 \forall j\}.$$

We claim that $[\tilde{G}_2, \tilde{G}, \tilde{G}, \dots, \tilde{G}] = 0$ where the commutator is taken $s-1$ times. This amounts to showing that for any $g_1, \dots, g_{s-2} \in \tilde{G}$ and any $h \in \tilde{G}_2$, that

$$[h, g_1, \dots, g_{s-2}] = 0.$$

This follows from the fact that (x, x) for $x \in \tilde{G}$ is orthogonal to all of the η_i of the form $(\eta_i^1, \eta_i^{1,2})$ and so can be generated by $(\tilde{u}_j \tilde{v}_j^1, \tilde{u}_j)$'s modulo $[G, G]^2$ (noting that $[G, G]^2$ contains $[G \times_{G_2} G, G \times_{G_2} G]$ and is a normal subgroup of $G \times_{G_2} G$). We then use the fact that the above expression only depends on the horizontal components of $\tilde{v}_j$'s. Using Lemma *A.1* and Lemma *A.2*, we write $g(n)\Gamma = g_1(n)\gamma(n)\Gamma$ with $g_1 \in \tilde{G}$ with the filtration $\tilde{G}_i = \tilde{G}_2 \cap G_i$ for $i \geq 2$, we can then apply Lemma *A.11*, Case 1, and Lemma *A.8* to $g_1(Pk!n)$ (with P the period of γ), obtaining characters $\alpha_1, \dots, \alpha_r$ such that $\|\alpha_i \circ g(Pk!\cdot)\|_{C^\infty[N]} = 0$ and since $Pk!$ is relatively prime to N, we have $\|\alpha_i \circ g\|_{C^\infty[N]} = 0$. Furthermore, it follows that if $\beta_1, \dots, \beta_{s-1}$ are orthogonal to the α_i's, it follows that

$$\|\xi([g(Pk!\cdot)\Gamma, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = \|\xi([g_1(Pk!\cdot)\Gamma, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = 0$$

so

$$\|\xi([g, \beta_1, \dots, \beta_{s-1}])\|_{C^\infty[N]} = 0.$$

5.4. **Deducing the main theorem.** Finally, to deduce Theorem 3 from Theorem 9, we quotient out by the kernel of ξ and apply Lemma 2.2 to assume that $g(0) = 1$. We then apply apply Lemma *A.1* to the horizontal characters obtained from Theorem 9 to obtain $g(n) = g_1(n)\gamma(n)$ where ϵ is constant, g_1 lies on a $\leq s-1$-step nilmanifold G_1/Γ_1 with rationality at most $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$ and γ is $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$-rational. By Lemma *A.11* G_1/Γ_1 has complexity at most $(\delta/M)^{-O_{s,k}(d)^{O_{s,k}(1)}}$. By making a change of variables $n \equiv Pk!m \pmod N$, we obtain

$$F(g(n)\Gamma) = \tilde{F}(g_1(Pk!m)\Gamma_1)$$

where $\tilde{F}$ is the restriction of F to $\tilde{G}_1$. Making another change of variables $Q \equiv (Pk!)^{-1} \pmod N$, we may write

$$F(g(n)\Gamma) = \tilde{F}(g_1(Pk!Qn)\Gamma_1).$$

Thus, $\tilde{F}(\epsilon g_1(PQn)\Gamma_1)$ satisfies the required conclusions of Theorem 3 and we are done.

Remark. The reader may have picked up on the fact that our argument isn't sharp, in the sense that there is information gained during the proof that was not fully used in the proof. For instance, in the proof of case 1, we only fully used the η_i's which had zero η_i^2 component. Thus, one can speculate whether one can induct on even more quantities than we did so here to obtain a qualitatively stronger result with good quantitative bounds. We did not attempt to do so here, since these improvements aren't relevant to us in obtaining an equidistribution theory with losses single exponential in dimension, but it could be possible that understanding the equidistribution theory of nilsequences better could lead to a better understanding of other problems in higher order Fourier analysis such as the inverse theorem (though, for purposes of the inverse theorem, it seems that a total understanding of linear nilsequences is sufficient).

6. A Ratner-type factorization theorem for a single observable

To illustrate the quantitative strength of our main theorem, we iterate our main theorem to deduce an analogous Ratner-type factorization theorem for a single observable. We do want to emphasize, though, that the actual Ratner-type factorization theorem does at least appear significantly stronger and more flexible than the theorem we prove below, though it remains to be seen if for the sake of applications, the theorem below can or cannot be used as a substitute for the Ratner-type factorization theorem. We need the following definition before we state our result:

Definition. We say that a nilsequence $F(g(n)\Gamma)$ is δ-equidistributed on scale N if

$$\left|\mathbb{E}_{n\in[N]}F(g(n)\Gamma) - \int_{G/\Gamma} F d\mu\right| < \delta\|F\|_{Lip(G)}$$

Theorem 10. *For any periodic nilsequence $F(g(n)\Gamma)$ of Lipschitz parameter ≤ 1, degree k, complexity M_0, dimension d, and step s, and $\delta > 0$, $A \geq 100$, there exists some M with $M_0 \leq M \leq (M_0/\delta)^{O(Ad)^{O(1)}}$ such that*

$$F(g(n)\Gamma) = \sum_{i=1}^{L} F_i(g_i(n)\Gamma_i) + h(n)$$

with $L \leq M$, g_i is a polynomial sequence in G_i, which is the projection of some subgroup of G, has complexity $\leq M$, and $F_i(g_i(n)\Gamma_i)$ is M^{-A}-equidistributed in G_i/Γ_i, and $\|h\|_{L^\infty[N]} \leq \delta$.

This theorem can be thought of as a "Fourier approximation of a nilsequence."

Proof. Let $1/M_0^A = \delta_1 > \delta_2 > \cdots$ and $M_0 \leq M_1 \leq M_2 \leq \dots$ be a sequence of parameters to be specified later on. Next, we invoke Lemma 2.2 to assume that $g(0) = 1$ and $|\psi(g(1))| \leq \frac{1}{2}$. If $F(g(n)\Gamma)$ is M_0^A-equidistributed, then we are done. Otherwise, by Lemma *A.6*, we may Fourier approximate

$$F(g(n)\Gamma) = \sum_{i=1}^{L_1} F_i(g(n)\Gamma) + O_{L^\infty}(\delta)$$

with $L_1 \leq (\delta/M_0)^{-O(Ad)} := \delta_1$. Applying the our main theorem to each $F_i(g(n)\Gamma)$, we see that they are all either δ_1^2-equidistributed or there exists a factorization $g(n) = g_i(n)\gamma_i(n)$. We obtain

$$F(g(n)\Gamma) = \sum_{i=1}^{L_1} F_i(g_i(n)\gamma_i(n)\Gamma) + O_{L^\infty}(\delta).$$

For each term $F_i(g_i(n)\gamma_i(n)\Gamma)$, we make a change of variables $F_i(g_i(Pn)\Gamma) = F_i(g_i(Pn)\Gamma_i)$ where Γ_i is the lattice of the subgroup that g_i lies in and P is the period of γ_i, everything having complexity and rationality bounded by $M_1 \leq (\delta_1/M_0)^{-O(Ad)^{O(1)}}$. We then take another Fourier expansion of F_i in the vertical expansion in G_i to obtain

$$F_i(x\Gamma_i) = \sum_{j=1}^{(\delta/M_1)^{-O(d)}} F_{i,j}(x\Gamma_i) + O_{L^\infty}((\delta/M_1)^{2A}).$$

We can then apply our main theorem to $F_{i,j}(g_i(Pn)\Gamma_i)$ with $\delta_2 := (\delta_1/M_1)^{2A}$-equidistribution. Iterating this procedure $O_{s,k}(1)$ many times gives the desired result. □

Appendix A. Auxiliary Lemmas

In this section, we shall state auxiliary lemmas we use in the proof of our main theorem. Most of these results come from [GT1]. The first two lemmas are [GT1, Proposition 9.2] and [GT1, Lemma 7.9], respectively.

Lemma A.1 (Factorization lemma I)**.** *Let $g(n)\Gamma$ be a periodic nilsequence of step s, degree k, dimension d, and complexity M and suppose $\eta_1, \dots, \eta_r$ are a set of linearly independent nonzero horizontal characters of size at most L. Suppose $\|\eta_i \circ g\|_{C^\infty[N]} = 0$ for each character i. Then we may write $g(n) = \epsilon(n)g_1(n)\gamma(n)$ where $\epsilon(n)$ is constant, $g_1(n)$ is a periodic nilsequence in*

$$\tilde{G} = \bigcap_{i=1}^{r} \ker(\eta_i)$$

which has complexity at most $M(dL)^{O_{s,k}(r)}$, γ is $O((dL)^{O(r)})$-rational, and $g_1(0) = \gamma(0) = 1$.

Remark. The important point here is that losses are at most single exponential in dimension.

Proof. By setting $\epsilon(n) = g(0)$, we may work with the assumption that $g(0) = 1$. We may thus write in coordinates that

$$\psi(g(n)) = \sum_i \binom{n}{i} t_i$$

where t_i are vectors representing the coordinates of the degree i component of g in Mal'cev coordinates. By Cramer's rule, we may pick a rational vector v with denominator at most $(dL)^{O(r)}$ such that $\eta_i \cdot v = \eta_i \cdot \psi(g(n))$ and such that the nonlinear component of v is zero. We define $\gamma(n)$ to be the preimage of ψ of $\sum_i \binom{n}{i} v_i$. Thus, the polynomial sequence $g_1(n) := g(n)\gamma(n)^{-1}$ lies inside $\tilde{G}$ and by construction $\gamma(n)\Gamma$ is $(dL)^{O(r)}$-periodic. ∎

Lemma A.2 (Factorization lemma II)**.** *Let $g(n)\Gamma$ be a periodic nilsequence of step s, degree k, dimension d, and complexity M and suppose $\eta_1, \dots, \eta_r$ are a set of linearly independent nonzero horizontal characters on G_2 which annihilate $[G, G_2]$ of size at most L. Suppose $\|\eta_i \circ g_2\|_{C^\infty[N]} = 0$ for each character i. Then we may write $g(n) = g_1(n)\gamma(n)$ where $g_1(n)$ is a periodic nilsequence with nonlinear part in*

$$\tilde{G}_2 = \bigcap_{i=1}^{r} \ker(\eta_i)$$

which has complexity at most $M(dL)^{O_{s,k}(r)}$, γ is $O((dL)^{O(r)})$-rational, and $g_1(0) = \gamma(0) = 1$.

Proof. We write in Mal'cev coordinates that

$$\psi_{G_2}(g_2(n)) = \sum_{i \geq 2} \binom{n}{i} t_i$$

where t_i are vectors representing the coordinates of the degree i component of g in Mal'cev coordinates. By Cramer's rule, we may pick a rational vector v in G_2 with denominator at most $(dL)^{O(r)}$ such that $\eta_i \cdot v = \eta_i \cdot \psi(g(n))$ and such that the nonlinear component of v is zero. We define $\gamma(n)$ to be the preimage of ψ of $\sum_i \binom{n}{i} v_i$. Thus, the polynomial sequence $g_2^1(n) := g_2(n)\gamma(n)^{-1}$ lies inside $\tilde{G}_2$ and by construction $\gamma(n)\Gamma$ is $(dL)^{O(r)}$-rational. Defining $\tilde{g}(n) = g(1)^n g_2(n)\gamma(n)\Gamma$, we obtain the desired properties. ∎

Let G be a nilpotent Lie group of step s with the finite degree k filtration G_i satisfying $[G_i, G_j] \subseteq G_{i+j}$. Suppose G has complexity M and dimension d. In many of the proofs above, we work with the group $G \times_{G_2} G$. The next lemma is a lemma regarding properties of this group.

Lemma A.3 (Properties of $G^\square$)**.** *The following properties hold for $G \times_{G_2} G$:*

- *$(G \times_{G_2} G)_i = G_i \times_{G_{i+1}} G_i$ forms a filtration of $G \times_{G_2} G$.*
- *If G has complexity M, then $G \times_{G_2} G$ has complexity at most $(dM)^{O_s(1)}$.*
- *Furthermore, if F is a Lipschitz function on G with norm L. Define $F^\square(x, y) = F(gx)\overline{F}(y)$ where $d_G(g, e) \leq 1$, then $F^\square$ has Lipschitz parameter at most $(dML)^{O_{s,k}(1)}$. Furthermore, if F is a nilcharacter of frequency ξ, then F annihilates $G_k^\triangle$.*
- *For each horizontal character η on $G \times_{G_2} G$, we may decompose η uniquely as*
$$\eta(g', g) = \eta_1(g) + \eta_2(g'g^{-1})$$
where η_1 is a horizontal character on G, η_2 is a horizontal character on G_2 which annihilates $[G, G_2]$. Furthermore, if $|\eta|$ is bounded by K, then $|\eta_1|, |\eta_2|$ are bounded above by $K(dM)^{O_{s,k}(1)}$.

Proof. These properties will ultimately follow from [GT1, Proposition 7.2, Lemma 7.4, Lemma 7.5]. For convenience to the reader, we will sketch out an argument here. For the first point, note that if $(g_i, g_{i+1}g_i)$ and $(h_i, h_{i+1}h_i)$ are two elements in $(G \times_{G_2} G)_i$, then
$$[(g_i, g_{i+1}g_i), (h_i, h_{i+1}h_i)] = ([g_i, h_i], [g_{i+1}g_i, h_{i+1}h_i])$$
and various commutator identities show that
$$[g_{i+1}g_i, h_{i+1}h_i] = [g_i, h_i][g_{i+1}, h_i][g_i, h_{i+1}][g_{i+1}, h_{i+1}] = [g_i, h_i] \pmod{G_{i+j+1}}.$$
Hence, $(G \times_{G_2} G)_i$ forms a filtration. To show the second point, denoting $\{X_1, \dots, X_d\}$ as the Mal'cev basis, consider
$$\{(X_1, 0), (0, X_1), \dots, (X_d, 0), (0, X_d)\}.$$
This is a Mal'cev basis for $G/\Gamma \times G/\Gamma$, and by Cramer's rule, $G \times_{G_2} G$ is $(dM)^{O(d)}$-rational with respect to this basis. By [GT1, Proposition A.10], it follows that there exists a Mal'cev basis on $G \times_{G_2} G$ which is an $(dM)^{O_{s,k}(d)}$-rational combination of (X_i, X_j). For the third point, note that F restricted to $G/\Gamma \times G/\Gamma$ has Lipschitz constant L^2. We see that $G \times_{G_2} G$ has rationality $(dM)^{O_{s,k}(1)}$, so if $x, y \in G \times_{G_2} G$, then $d_{G/\Gamma \times G/\Gamma}(x, y) \leq (dM)^{O_{s,k}(1)} d_{G \times_{G_2} G}(x, y)$. The third point follows from $F(gxg_s)\overline{F}(g_s y) = F(gx)\overline{F}(y)$ where $g_s \in G_{(s)}$. Finally, for the fourth point, we define $\eta_1(g) = \eta(g, g)$ and $\eta_2(h) = \eta(h, 1)$. Since η annihilates $[G \times_{G_2} G, G \times_{G_2} G]$, this contains $[G^\triangle . G^\triangle] = [G, G]^\triangle$ and $[G_2 \times 1, G^\triangle] = [G_2, G] \times 1$, we see that η_2 must annihilate $[G, G_2]$ and η_1 must annihilate $[G, G]$. We also see that since η annihilate $\Gamma \times_{G_2 \cap \Gamma} \Gamma$, this contains both $\Gamma^\triangle$ and $(\Gamma \cap G_2) \times 1$, so η_1 annihilates Γ and η_2 annihilates $\Gamma \cap G_2$.

To check the boundedness conditions, we see that the Mal'cev coordinates of $G \times_{G_2} G$ are rational combinations of $G \times G$ with coefficients that have denominator at most $(dM)^{O_{s,k}(1)}$. It follows that η_1, η_2 are bounded by $K(dM)^{O_{s,k}(1)}$. $\square$

The next lemma is another auxiliary lemma used in the proof of the main theorem above. In that setting, we have $g(n)$ is a polynomial sequence in a filtered nilpotent Lie group. For $h \in \mathbb{Z}$, we define $g_h(n) = (\{g(1)^h\}^{-1} g_2(n+h) g(1)^{-n} \{g(1)^h\}, g(n))$ and the filtered nilpotent Lie group $(G \times_{G_2} G)_i = G_i \times_{G_{i+1}} G_i$.

Lemma A.4. *The sequence $g_h(n)$ is a polynomial sequence in $G_i \times_{G_{i+1}} G_i$.*

Proof. This will once again follow from [GT1, Proposition 7.2], which uses the fact that $\mathrm{poly}(\mathbb{Z}, G\times_{G_2} G)$ is a group and is a normal subgroup of $\mathrm{poly}(\mathbb{Z}, G\times G)$. Since $(\{g(1)^h\}, 1)$ lies inside $G \times G$, it follows that $g_h(n)$ is a polynomial sequence if $(g_2(n+h)g(1)^{-n}, g_2(n)g(1)^{-n})$ is a polynomial sequence. Since $(g(1)^{-n}, g(1)^{-n})$ is a polynomial sequence in $G \times_{G_2} G$, it follows that it suffices to show that $(g_2(n+h), g_2(n))$ is a polynomial sequence in $G \times_{G_2} G$. This amounts to checking that if $g_i \in G_i$, then $(g_i^{\binom{n+h}{i}}, g_i^{\binom{n}{i}})$ is a polynomial sequence on $G \times_{G_2} G$. Taking j derivatives, this amounts to checking that $(g_i^{\binom{n+h}{i-j}}, g_i^{\binom{n}{i-j}})$ is a polynomial sequence in $G_j \times_{G_{j+1}} G_j$. For $j \geq i$, this becomes that $(1, 1)$ is a polynomial sequence, which is true. For $j < i$, this follows from the fact that $g_i \in G_i$. □

The next lemma, is a quantitative Fourier expansion lemma for Abelian groups, which will be useful for the following lemma:

Lemma A.5 (Fourier/Fejer Expansion lemma)**.** *Let $f\colon \mathbb{T}^d \to \mathbb{C}$ be a continuous function with Lipschitz parameter at most L, meaning that $\|f\|_{L^\infty(\mathbb{T}^d)} + \|f\|_{Lip(\mathbb{T}^d)} \leq L$. Then we may write*

$$f = \sum_{i=1}^{k} a_i e(n_i x) + g$$

where $\sum_i^k |a_i| \leq C^{d^2} L\delta^{-2d^2-d}$ and $\|g\|_\infty \leq 3\delta$.

Proof. Let $\phi\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a smooth compactly supported function supported in $[-1, 1]$ and with integral one. Let $Q_\delta(x) = \prod_{i=1}^d \delta^{-1}\phi(x_i/\delta)$ and let $K = Q_\delta * Q_\delta$ be a Fejer-type kernel. Since $|\hat{\phi}(\xi)| \ll_k |\xi|^{-k}$,

$$|\hat{K}(\xi)| \leq C^d \delta^{2d} \xi^{-2}$$

for some constant C it follows that the Fourier coefficients of K larger than M contributes at most $C^d M^{-1}\delta^{2d}$. In addition, for $M = C^{-d}\delta^{-2d-1}\|f\|_{Lip}^{-1}$

$$\|f - f * K\|_\infty \leq \int |f(x) - f(y)|K(x-y)dy = \int zK(z)dz \leq 2\delta$$

since K has integral one and is supported on $|x| \leq 2\delta$. Set

$$h(x) = \sum_{k\in\mathbb{Z}^d : |k|\leq M} \hat{f}(k)\hat{K}(k)e(kx).$$

Then by Fourier inversion formula, it follows that

$$\|h - f * K\|_\infty \leq \delta.$$

Thus, $\|h - f\|_\infty \leq 3\delta$. The sum of the Fourier coefficients of H is at most $LC^{-d^2}\delta^{-2d^2-d}$. □

which shows that we may Fourier expand a Lipschitz function on G/Γ to nilcharcters:

Lemma A.6 (Quantitative Fourier Expansion Lemma)**.** *Let G/Γ be a nilmanifold with dimension d, complexity M, degree k, and step s. Let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz function with Lipschitz constant L. Then we may Fourier expand*

$$F(x) = \sum_{|\xi|\leq(\delta/L)^{-O(d)}} F_\xi(x) + O(\delta)$$

where F_ξ is a nilcharacter of frequency ξ. More generally, given a connected and simply connected subgroup H of the center of G with rationality bounded by M, we can Fourier expand

$$F(x) = \sum_{|\xi| \le (\delta/L)^{-O(d)}} F_\xi(x) + O(\delta).$$

Proof. Let K denote the kernel constructed before adapted to $G_{(s)}$. We define

$$\tilde{F}(x) = \int_{G_{(s)}/(\Gamma \cap G_{(s)})} F(g_s x) K(g_s) dg_s.$$

Write $G_{(s)} \cong \mathbb{R}^{d_s}/\mathbb{Z}^{d_s}$. For $M = C^{-d}\delta^{-2d-1}\|F\|_{Lip}^{-1}$, we have

$$\|F - \tilde{F}\|_\infty \le \int_{\mathbb{T}^{d_s}} |f(gx) - f(hx)| K(g-h) dh \le \int_{\mathbb{T}^{d_s}} hK(h)dh \le 2\delta$$

since K has integral 1 and is supported on $|x|_{G_{(s)}} \le 2\delta$. Let

$$G(x) = \sum_{k \in \mathbb{Z}^{d_s}, |k| \le M} \hat{F}(k)\hat{K}(k)e(kx).$$

Then by Fourier inversion, it follows that

$$\|G - \tilde{F}\|_\infty \le \delta.$$

Thus, $\|G - \tilde{F}\|_\infty \le 3\delta$ and the first part of the lemma follows from this. For the second part of the lemma, we have by the rationality bounds that the Lipschitz constant of F is bounded by ML. We can then follow the proof of the first part of the lemma. □

The following linear algebraic lemma is used often in our paper:

Lemma A.7 (Corollary of Cramer's rule)**.** *Let $v_1, \dots, v_r$ be integral vectors of $\mathbb{R}^d$ size at most $M \ge 2$. Then there exists integral $\eta_1, \dots, \eta_{d-r}$ of size at most $(dM)^{O(d)}$ such that $v_1, \dots, v_r, \eta_1, \dots, \eta_{d-r}$ span $\mathbb{R}^d$ and $\langle v_i, \eta_j \rangle = 0$.*

Proof. This is a simple application of Cramer's rule. Let $e_1, \dots, e_d$ be the unit coordinate vectors in $\mathbb{R}^d$. Then there exists a subset, say, $E = \{e_{j_1}, \dots, e_{j_{d-r}}\}$ such that $\mathrm{span}(E) \oplus \mathrm{span}(v_1, \dots, v_r) = \mathbb{R}^d$. Let A be the matrix who's rows are consisting of the elements $v_1, \dots, v_r$ and $e_{j_1}, \dots, e_{j_{d-r}}$. Then in the matrix A^{-1} has columns that are linearly independent, and letting $\eta_1, \dots, \eta_{d-r}$ be the last $d-r$ columns, we see that $\langle v_i, \eta_j \rangle = 0$. Multiplying η_j by some integer bounded by $(dM)^{O(d)}$ gives the result. □

Lemma A.8. *Suppose G' is a subgroup of G (which is nilpotent of step s) with rationality Q such that G' is nilpotent of step $\le s-1$. Then there exists $\eta_1, \dots, \eta_r$ horizontal characters with size bounded by $Q^{O(d)^2}$ of G such that $\eta_i(G') = 0$, and if $w_1, \dots, w_s$ are vectors on the horizontal component orthogonal to η_i, then $[w_1, w_2, \dots, w_{s-1}, w_s] = 0$.*

Proof. This is a consequence of the Lie algebra of G' being a subspace of the Lie algebra of G. Let $\pi_{horiz} : G \to G/[G,G]$ be the projection from G to the horizontal component. Since G' is Q-rational, it follows from Cramer's rule (or rather Lemma A.7) that there exists horizontal characters $\eta_1, \dots, \eta_r$ of size at most $Q^{O(d)^2}$ such that

$$\pi_{\mathrm{horiz}}\left(\bigcap_{i=1}^r \ker(\eta_i)\right) = \pi_{\mathrm{horiz}}(G').$$

Furthermore, given $w_1, \dots, w_s$ vectors inside $\bigcap \ker(\eta_i)$, we see that the expression

$$[w_1, w_2, \dots, w_{s-1}, w_s]$$

only depends on the horizontal components of w_i, so we may replace them with elements inside G', in which case the expression above is zero. □

We will also need the following two lemmas about periodic nilsequences:

Lemma A.9. *Let $g(n)\Gamma$ be a periodic modulo N on an abelian nilmanifold with N larger than the degree of g. Then all of the coordinates of g are rational with denominator N.*

Proof. This follows from [T2, Lemma 1.4.1]. □

Lemma A.10. *Let $g(n)\Gamma$ be a periodic modulo N with N larger than $(10^k k)!$ where k is the degree of g. Then the horizontal Mal'cev coordinates of $g(n)\Gamma$ have denominator N (i.e., for any horizontal character η, $\eta \circ g$ has denominator dividing N). In addition, if we define the nonlinear part $g_2(n) := g(n)g(1)^{-n}$, then $g_2((10^k k)!n)$ has horizontal ($G_2/[G, G_2]$) Mal'cev coordinates with denominator N.*

Remark. Note that it is not necessarily true that $g((10^k k)!n)$ has all of its coordinates with denominator N as the following example shows:

$$g(n) = \begin{pmatrix} 1 & a\alpha & a\beta & a\alpha\beta \\ 0 & 1 & 0 & \beta \\ 0 & 0 & 1 & \alpha \\ 0 & 0 & 0 & 1 \end{pmatrix}^n$$

with $a \in \mathbb{Z}$, $\alpha, \beta \in \widehat{\mathbb{Z}/N\mathbb{Z}}$. One can check that if G is the group where we replace $a\alpha, b\beta, \alpha, \beta, a\alpha\beta$ with an arbitrary real numbers, and Γ is the subgroup where all of the corresponding entries are integers, then $g(n)\Gamma$ is N-periodic if α and β have denominator N. Note that this example corresponds to the N-periodic bracket polynomial $n \mapsto e(a\{\alpha n\}\{\beta n\})$. The example also shows that if we were to add a quadratic part to the upper right corner of g, in order for the nilsequence to stay N-periodic, the coefficient of the quadratic part must have denominator N. This is equivalent to saying that if we were to add higher order parts to g, then the horizontal component of g_2 must have denominator N, as the lemma claims.

Proof. The first part of the statement follows from expanding out horizontal Mal'cev coordinates (i.e., projecting to the orthogonal subspace to $[\mathfrak{g}, \mathfrak{g}]$ in the Lie algebra with respect to the chosen Mal'cev coordinates) and applying Lemma A.9. For the second part, we work with the joining $G \times_{G_2} G$. For $h \in \mathbb{N}$, we define $g_h(n) = (\{g(1)^h\}^{-1} g_2(n+h) g(1)^{-n} \{g(1)^h\}, g(n))$. It follows that $g_h(n)\tilde{\Gamma}$ is N-periodic on $G \times_{G_2} G$, and so it is N-periodic on $G \times_{G_2} G/G^{\triangle}_{(s)}$, and so the horizontal component of $[g(1)^n, \{g(1)^h\}] g_2(n+h) g_2(n)^{-1}$ is N-periodic modulo 1. Letting the horizontal part of g_2 be

$$\sum_{i=2}^{k} \alpha_i n^i$$

we see that the horizontal part of $g_2(n+h)g_2(n)^{-1}$ is

$$\sum_{i=2}^{k} \alpha_i \sum_{j=0}^{i-1} n^j h^{i-j} \binom{i}{j} = \sum_i h^i \alpha_i + \sum_{i=1}^{k} n^i \left(\sum_{j=i+1}^{k-1} \alpha_j h^{j-i} \binom{j}{i} \right).$$

Furthermore, we claim that the horizontal component of $[g(1), \{g(1)^h\}]$ in $G_2/[G, G_2]$ has denominator N. To see this, note that it suffices to show that for any character $\xi : G_2 \to S^1$ which annihilates $[G, G_2]$, and $\Gamma \cap G_2$ that $\xi([g(1), \{g(1)^h\}])$ has denominator N, or that $\xi([g(1)^N, \{g(1)^h\}]) = 0 \pmod 1$. We may factor $\{g(1)^h\} = g(1)^h[g(1)^h]^{-1}$ and using the fact that powers of $g(1)$ commute with each other, we see that

$$\xi([g(1)^N, \{g(1)^h\}]) = \xi([g(1)^N, [g(1)^h]^{-1}]) = 0$$

since $[g(1)^N, [g(1)^h]^{-1}] \in \Gamma \cap G_2$. The horizontal component of $[g(1)^n, \{g(1)^h\}]g_2(n+h)g_2(n)^{-1}$ is

$$n\psi_{horiz}([g(1)^n, \{g(1)^h\}]) + \sum_i h^i \alpha_i + \sum_{i=1}^{k} n^i \left(\sum_{j=i+1}^{k-1} \alpha_j h^{j-i} \binom{j}{i} \right).$$

By Lemma A.9 and the fact that $\psi_{horiz}([g(1)^n, \{g(1)^h\}])$ has denominator N, we see that for each $i \in [k]$, that

$$\sum_{j=i+1}^{k-1} \alpha_j h^{j-i} \binom{j}{i} \equiv 0 \pmod{1/N}$$

for all $h \in \mathbb{N}$. Applying Vinogradov's lemma, it follows that $(10^k)!\alpha_j \binom{j}{i} \equiv 0 \pmod{1/N}$ for each $j > i$, from whence the lemma follows.

□

Remark. In fact, if one uses Pólya's classification of integer-valued polynomials [P], one can eliminate the factor of 10^k.

Lemma A.11 (Complexity of Rational Subgroups). *Let G be a filtered nilpotent Lie group of dimension d and let G' be a Q-rational subgroup of G. Then G' has complexity at most $Q^{O_{s,k}(d)^{O_{s,k}(1)}}$.*

Proof. This follows from [GT1, Proposition A.9]. The one property we must check is that (keeping the same notation as in [GT1, Proposition A.9]) the c_i have numerator and denominator bounded by $Q^{O_{s,k}(d)^{O_{s,k}(1)}}$. Recall that c_j were constructed so that

$$(\mathrm{Span}(Y_j, \dots, Y_m) \cap \Gamma) \setminus \mathrm{Span}(Y_{j+1}, \dots, Y_m)$$

is generated by $\overline{\exp(c_j Y_j)}$ for some choice of $c_j \in \mathbb{Q}$. Here, we see that c_j must be the rational with largest denominator and smallest numerator for which $\exp(c_j Y_j)$ lies in Γ, because otherwise, either $\exp(c_j Y_j)$ is not an element of Γ or $\exp(c_j Y_j)$ cannot generate $(\mathrm{Span}(Y_j, \dots, Y_m) \cap \Gamma) \setminus \mathrm{Span}(Y_{j+1}, \dots, Y_m)$. From this property, we see that each c_j must have height $Q^{O_{s,k}(d)^{O_{s,k}(1)}}$ and thus the lemma follows. □

Appendix B. Deduction of the equidistribution lemma for Theorem 5

In this section, we deduce the following, which is a refinement of [Len1, Lemma 6.1]:

Lemma B.1. *Let G/Γ be a two-step nilmanifold equipped with the standard filtration and let $g(n)$ be a polynomial sequence on G that is periodic modulo N with one dimensional vertical component. Let F_1, F_2, F_3 be Lipschitz functions on G/Γ with the same nonzero frequency ξ. If*

$$|\mathbb{E}_{n\in[N]} F_1(g(P(y))\Gamma) F_2(g(Q(y))\Gamma) \overline{F_3(g(P(y)+Q(y))\Gamma)} e(\alpha P(y) + \beta Q(y))| \geq \delta$$

for some frequencies $\alpha, \beta \in \widehat{\mathbb{Z}/N\mathbb{Z}}$, then either $N \ll (\delta/M)^{-O_{P,Q}(d)^{O_{P,Q}(1)}}$ or there exists $w_1, \dots, w_r \in \Gamma/(\Gamma \cap [G,G])$ and horizontal characters $\eta_1, \dots, \eta_{d-1-r}$ such that $|w_i|, |\eta_j| \le (\delta/M)^{-O_{P,Q}(d)^{O_{P,Q}(1)}}$, $\langle \eta_j, w_i \rangle = 0$ for all i, j, and

$$\|\xi([w_i, g])\|_{C^\infty[N]} = \|\eta_j \circ g\|_{C^\infty[N]} = 0.$$

Proof. The first part of the argument is similar to the first part of the argument in [Len1, Lemma 6.1]. We first use the fact that F is a nilcharacter of nonzero frequency to absorb $\alpha P(y)$ and $\beta Q(y)$ to the vertical component of $g(P(y))$ and $g(Q(y))$, respectively to obtain $g_1(P(y))$ and $g_2(Q(y))$. Since the conclusion only depends on the horizontal component of g, and since the horizontal component of g_1 and g_2 agree with g, it follows that we can assume that both g_1 and g_2 are g and $\alpha, \beta = 0$. Let H denote the subgroup of G^3 consisting of elements $\{(g_1, g_2, g_3) : g_1 g_2 g_3^{-1} \in [G,G]\}$. We claim that $[H,H] = [G,G]^3$. By definition, we see that for $h \in [G,G]$ that $(1,h,h)$ and $(h,1,h)$, and (h,h,h^4) lies inside $[H,H]$ (the last fact is true because $[(g_1, g_1, g_1^2), (h_1, h_1, h_1^2)] = ([g_1,h_1],[g_1,h_1],[g_1,h_1]^4))$. This yields that $(1,1,h^2)$ lies inside $[H,H]$, and because of connectedness and simple connectedness, it follows that $(1,1,h) \in [H,H]$. We can verify from there that $[H,H] = [G,G]^3$.

We were given the polynomial sequence

$$(g(P(y)), g(Q(y)), g(P(y) + Q(y)))$$

on H. Since F_i are nilcharacters of frequency ξ on H, $F_1 \otimes F_2 \otimes F_3$ is a nilcharacter on H of frequency $(\xi, \xi, -\xi)$. Taking a quotient of H by the kernel of $(\xi, \xi, -\xi)$, which is $(x, y, x+y)$, we obtain that the center is of the form $(x, x, -x)$ with (x, y, z) being projected to $(x+y-z)(1,1,1)$. Let H_1 denote the subgroup with the one dimensional vertical directions. Applying Theorem 8, we obtain $w_1, \dots, w_r$ and $\eta_1, \dots, \eta_{d-r}$ such that $\langle w_i, \eta_j \rangle = 0$ and $\eta_j \circ (g(P(y)), g(Q(y)), g(P(y)+Q(y))) \equiv 0 \pmod 1$, and $\xi([w_i, g(P(y)), g(Q(y)), g(P(y) + Q(y))]) \equiv 0 \pmod 1$. Denoting $\eta_j = (\alpha_j, \beta_j)$ and $w_i = (u_i, v_i, u_i v_i)$ and the action $\eta_j(w_i) := \alpha_j(u_i) + \beta_j(v_i)$, we see that

$$\|\xi([v_i, g(P(y))]) + \xi([u_i, g(Q(y))])\|_{C^\infty[N]} \equiv 0 \pmod 1$$
$$\|\alpha_j(g(P(y))) + \beta_j(g(Q(y)))\|_{C^\infty[N]} \equiv 0 \pmod 1.$$

Since P and Q are linearly independent, it follows that there exists some coefficients $c_k x^k$, $c_\ell x^\ell$ of P, and $d_k x^k$ and $d_\ell x^\ell$ of Q such that $c_k d_\ell - d_k c_\ell \neq 0$. Thus, the conditions become

$$c_k \xi([u_i, g(1)]) + d_k \xi([v_i, g(1)]) \equiv 0 \pmod 1$$
$$c_\ell \xi([u_i, g(1)]) + d_\ell \xi([v_i, g(1)]) \equiv 0 \pmod 1$$

which implies since α has denominator N which is prime that $\xi([u_i, g(1)]) \equiv 0 \pmod 1$ and $\xi([v_i, g(1)]) \equiv 0 \pmod 1$. Similarly, we have $\alpha_j(g(1)) \equiv 0 \pmod 1$ and $\beta_j(g(1)) \equiv 0 \pmod 1$. Let $\tilde{G} := \{g \in G : \xi([v_i, g]) = 0, \xi([u_i, g]) = 0, \alpha_j(g) = 0, \beta_j(g) = 0 \forall i, j\}$. We claim that $\tilde{G}$ is abelian, from whence the Lemma would follow from an application of Lemma *A.8*. This amounts to showing that for any $g, h \in \tilde{G}$ that $[g,h] = 1$. For such g, (g,g) is annihilated by (α_j, β_j), and since $\alpha_j(u_i) + \beta_j(v_i) = 0$, it follows that (g,g) can be written as a combination of (u_i, v_i) modulo $[G,G]^2$. It follows that $[(g,g),(h,h)] = 1$, and thus $[g,h] = 1$. □

One can insert the proof of this lemma into the argument of [Len1], using the Sanders U^3 inverse theorem [S] instead of the Green-Tao inverse theorem used there. This would yield quasi-polynomial bounds on the inverse theorem proven there, which would yield Theorem 5. Again, the formal deduction will appear in [Len2]

Appendix C. Proof of the general refined bracket polynomial lemma

In this section, we shall provide two additional proofs of the refined bracket polynomial lemma, this time adapted for the case where we do not assume that various real numbers have denominator N. The first proof given here appears to be different from the proof given in section 3 and the second proof given in this section is a generalization of the one given in Section 3. The second proof of the refined bracket polynomial lemma given here proceeds via a more straightforward induction on dimensions and Minkowski's first theorem[8] (or the uncertainty principle) but is quite a bit more cumbersome. As such, we shall describe the main idea of the proof here in a bit more before we dive into the details. The reader may want to first assume that α is a rational with denominator N, a large prime. In this setting, it's much more clear that the induction on dimensions closes with bounds single exponential in dimension.

The key idea which prevents the size of a from growing too much in the bracket polynomial lemma is that the characters η should morally speaking lie close to the direction of a, which when we project a in the direction orthogonal to η to reduce the dimension should actually make a smaller. We can actually prove that η satisfies such a property either by using Minkowski's first theorem (in a similar way as it's used in Pólya's orchard problem) or by the uncertainty principle (which states that a function with essential support in a box has essential Fourier support in the dual box) in a Fourier argument in a flavor closer to the argument of Green and Tao. The rest of the proof involves carefully tracking how the constants are used so as to not run into bad behavior. This involves splitting the role of δ as in [GT1, Proposition 5.3] into several different variables and carefully tracking how they iterate under an induction on dimensions. We need the following lemma (see [GTZ2, Appendix E] for similar manipulations). Before we state it, we would need a definition:

We define a notion of *Fourier complexity* as in [Len3]. Specifically, we define the $L^p[N]$ δ-*Fourier complexity* (likewise $L^p([N]\times[H])$ δ-Fourier complexity) of a function $f:[N]\to\mathbb{C}$ to be the infimum of all L such that

$$f(n) = \sum_i a_i e(\xi_i n) + g$$

where $\|g\|_{L^p[N]} \le \delta$ and $\sum_i |a_i| = L$.

Lemma C.1 (Fourier complexity lemma). *Either $N, H \ll (\delta/k)^{-O(d)^2}$, or else*

$$e(k_1\{\alpha_1 h\}\{\beta_1 n\} + k_2\{\alpha_2 h\}\{\beta_2 n\} + \cdots k_d\{\alpha_d h\}\{\beta_d h\})$$

has $L^1([N]\times[H])$-δ-Fourier complexity at most $(\delta/2^d k)^{-O(d^2)}$ for $|k_i| \le k$ integers.

Proof. Let

$$\phi(n,h) = k_1\{\alpha_1 h\}\{\beta_1 n\} + k_2\{\alpha_2 h\}\{\beta_2 n\} + \cdots k_d\{\alpha_d h\}\{\beta_d n\}.$$

[8]readers unfamiliar with Minkowski's first theorem can refer to [TV, Theorem 3.28] (or simply doing an internet search) for a proof

The function $e(\phi(n,h))$ resembles a degree one nilsequence on a torus $\mathbb{T}^{2d}$ except with possible discontinuities at the endpoints $x_i = \frac{1}{2} \pmod 1$. To remedy this possible issue, we must set ourselves in a position so that the endpoints contribute very little to the $L^1([N] \times [H])$ norm. In the periodic nilsequences case, this will end up being true since $N = H$ are prime and α_i and β_i will always have denominator N. For readers only interested in the periodic case, one can multiply $e(\phi(n,h))$ by smooth cutoff supported supported in $[-1/2, 1/2]$ and equal to one on $[-1/2 + (\delta/2^d k), 1/2 - (\delta/2^d k)]^{2d}$ of derivative at most $(\delta/2^d k)^{-1}$ and Fourier expand everything via Lemma A.5. The terms for which we have ignored from taking cutoff contribute an $L^1([N] \times [H])$ norm at most $\delta/2^{d-1}$ since $\{n \in [N] : \|\alpha_i n - \frac{1}{2}\|_{\mathbb{R}/\mathbb{Z}} < \delta\}$ has at most $2\delta N$ many elements (and similarly for each β_i) since each α_i has denominator N. Thus, we have obtained the desired Fourier complexity bounds. The reader can skip the rest of the proof which only applies in the nonperiodic case.

In the general case, we would need an induction on dimensions procedure so our goal is to show that this does not lead to double exponential losses in parameters. This involves carefully quantifying the procedure in [GTZ2, Appendix E] and making sure that losses are only single exponential in dimension. Let $1 > \epsilon > 0$ be a quantity to be determined later. If $\{n \in [N] : \|\alpha_1 n - \frac{1}{2}\|_{\mathbb{R}/\mathbb{Z}} < \epsilon\}$ has more than $\sqrt{\epsilon}N$ many elements, then an application of Vinogradov's lemma shows that there exists some integer $q_1 \leq \sqrt{\epsilon}^{-1}$ such that $\|2q_1\alpha_1\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\sqrt{\epsilon}}{N}$. We now divide $[N]$ into progressions of common difference $2q_1$ such that on each progression, $\{\alpha_1 n\}$ varies by at most $\epsilon^{1/2-o(1)}$. This can be done as follows:

- First, divide $[N]$ into progressions of common difference $2q_1$.
- On each progression, we have that $\alpha_1 n$ varies by at most $\sqrt{\epsilon}$ modulo 1; however, it's possible that $\{\alpha_1 n\}$ "lapped over" $1/2$ along the progression, so we must divide each progression into two subprogressions, one before $\{\alpha_1 n\}$ lapped over $1/2$ and one after it lapped over $1/2$.
- The number of progressions we obtain is at most $4q_1$.

Thus, we may divide $[N] = \bigsqcup P_i$ into such progressions. We shall want to "prune out" the small progressions and write $[N] = \bigsqcup P_i \sqcup E$ where P_i' are sufficiently large and E is a small error. If $|P_i| \leq \epsilon^{1+1/2}N$, then we include it in E and otherwise, we do nothing. Thus, we can ensure that $|P_i| \geq \epsilon^{1+1/2}N$ and $|E| \leq 4q_1\epsilon^{1+1/2}N \leq 2\epsilon N$.

We now iterate in the following manner: if $\{n \in [N] : \|\alpha_2 n - \frac{1}{2}\|_{\mathbb{R}/\mathbb{Z}} < \epsilon\}$ has more than $\sqrt{\epsilon}N$ many elements, then (assuming that ϵ is sufficiently small) there exists a progression P (as in the decomposition above) such that $\{n \in P : \|\alpha_2 n - \frac{1}{2}\|_{\mathbb{R}/\mathbb{Z}} < \epsilon\}$ has more than $\frac{\sqrt{\epsilon}}{2}|P|$ many elements, then writing $P = q_1 \cdot [N_1] + r_1$, it follows that there are more than $\frac{\sqrt{\epsilon}}{2}N_1$ many elements in $\{n \in [N_1] : \|q_1\alpha_2 n + \gamma\|_{\mathbb{R}/\mathbb{Z}} < \epsilon\}$ for some γ. Letting n and m be two such elements, it follows that $\|q_1\alpha_2(n-m)\|_{\mathbb{R}/\mathbb{Z}} \leq \|-\gamma - q_1\alpha_2 m\|_{\mathbb{R}/\mathbb{Z}} + \|q_1\alpha_2 n + \gamma\|_{\mathbb{R}/\mathbb{Z}} \leq 2\epsilon$, so there are more than $\frac{\sqrt{\epsilon}}{2}|P|$ many elements such $n \in [N_1]$ such that $\|q_1\alpha_2\|_{\mathbb{R}/\mathbb{Z}} < 2\epsilon$. Thus, (assuming that N is sufficiently large with respect to ϵ) there exists some $q_2 \leq 2\sqrt{\epsilon}^{-1}$ such that $\|q_1q_2\alpha_2\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{2\sqrt{\epsilon}}{|P|}$. We now continue down a similar path, dividing each P_i into subprogressions instead of dividing $[N]$ into subprogressions, this time obtaining an error set of size at most $4\epsilon|P_i|$.

Under an iteration (making sure to choose ϵ so that $\frac{\sqrt{\epsilon}}{2^{2d}} \gg \epsilon$), we would obtain at most $(\epsilon/2^d)^{-O(d)}$ many progressions each of size at least $(\epsilon/2^d)^{O(d)}N$ and common difference at most $(\epsilon/2^d)^{-O(d)}$ such that

- either $n \mapsto \{\alpha_i n\}$ or $h \mapsto \{\beta_i h\}$ varies by at most $\epsilon^{1/2-o(1)}$ on each progression;
- or the set $\{n \in [N] : \|\alpha_i n - \frac{1}{2}\|_{\mathbb{R}/\mathbb{Z}} < \epsilon\}$ has at most $\sqrt{\epsilon}N$ many elements

and an error set E of size at most $2^{O(d)}\epsilon N$. For the α_i's or β_i's that lie in the second case, we introduce a smooth cutoff φ_j of derivative at most ϵ^{-1} in the dimensions that correspond to the α_i's or β_i's. Taking a product of $e(\phi(n,h))$ and the smooth cutoff and summing over each arithmetic progressions, we have obtained a sum of degree one nilseqences with Lipschitz parameter at most ϵ^{-1} and dimension at most d along progressions. We may approximate the terms that vary very little on the subprogressions as a constant function plus an error of at most $4kd\epsilon^{1/2}$. Fourier expanding the progression via the procedure in [GTZ2, Appendix E] and the Lipschitz function via Lemma A.5, we obtain at most $(\epsilon/2^d)^{-O(d^2)}$ many terms and an $L^1([N] \times [H])$ error of at most $2^{Cd}kd\sqrt{\epsilon}$ for some absolute constant C. Choosing ϵ so that $\delta \geq 2^{Cd}kd\sqrt{\epsilon}$, we see that we have a $L^1([N] \times [H])$-Fourier complexity at most $(\delta/2^d k)^{-O(d^2)}$. □

The important point about this is that we experience at most single exponential losses in the number of bracket phases we need in an application of the above lemma. Our proof involves an iteration scheme via the following lemma:

Lemma C.2. *Suppose there are at least δN_1 many $h \in J$ where J is an interval of size N_1 such that*

$$\|\beta + \gamma h + a \cdot \{\alpha h\}\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{K}{N}$$

with $|\gamma| \leq L/N_1$. Then either $N \ll L^{O(1)}(K\delta/2^d M)^{-O(d)^{O(1)}}$ or $N_1 \ll L^{O(1)}(K\delta/2^d M)^{-O(d)^{O(1)}}$ or $(\delta/2^d M)^{4d}\|a\|_\infty \leq K/N$ or there exists an integer vector v of size at most $(\delta/2^d M)^{-O(d)}$ in a $(\delta/2^d M)$-tube in the direction of a such that $\|v \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \leq L(\delta/2^d M)^{-O(d)}/N_1$.

Remark. The fact that v lies in a small tube in the direction of a is crucial for our iteration and is the chief difference between this lemma and the bracket polynomial lemma [GT1, Proposition 5.3] of Green and Tao.

Proof. We may assume that $(\delta/2^d M)^{4d}\|a\|_\infty \geq K/N$ since otherwise the lemma is proven. The assumption then implies that

$$\|\beta + \gamma h\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{K}{N} + d\|a\|_\infty \leq (d+1)\|a\|_\infty.$$

If $(d+1)\|a\|_\infty \gg \delta$, it follows that $|\gamma| \ll (d+1)\delta^{-1}L\|a\|_\infty/N_1$. Otherwise, Vinogradov's Lemma implies that here exists some $q \leq \delta^{-C}$ such that

$$\|q\gamma\|_{\mathbb{R}/\mathbb{Z}} \leq \|a\|_\infty (d+1)\delta^{-C}/N_1.$$

Since $|\gamma| \leq \frac{L}{N_1}$, it follows that $|\gamma| \leq L(d+1)\delta^{-C}\|a\|_\infty/N_1$. Thus, if we pigeonhole and replace N_1 with $N_2 = \frac{\delta^C(\delta/2^d M)^{O(d)}N_1}{(d+1)L}$ it follows that there exists some θ such that for $\delta|I|$ many $h \in I$ with I an interval of size at least $N_2/2$, we have

$$\|\theta + a \cdot \{\alpha h\}\|_{\mathbb{R}/\mathbb{Z}} \leq (\delta/2^d M)^{3d}\|a\|_\infty.$$

Following the first proof, we use the pigeonhole principle to isolate a single integer j such that

$$\theta + a \cdot \{\alpha h\} = j + O((\delta/2^d M)^{3d} \|a\|_\infty)$$

and we pigeonhole in a sign pattern so there exists $\delta|I|/2^d M$ many integers $h \in [|I|]$ such that

$$|a \cdot \{\alpha h\}| = O((\delta/2^d M)^{3d} \|a\|_\infty). \tag{1}$$

Let T be the tube of width $(\delta/2^d M)^2$ and length $2^{d+1}(\delta/2^d M)^{-2d}$ in the direction of a. By Minkowski's first theorem, this tube contains a nonzero lattice point, η. Multiplying (1) by t so that ta is $(\delta/2^d M)^2$-close to η, we see that

$$|\eta \cdot \{\alpha h\}| = O(t(\delta/2^d M)^{3d} \|a\|_\infty + (\delta/2^d M)^2).$$

We see that $t \leq 2^{d+1}(\delta/2^d M)^{-2d} d/\|a\|_\infty$. Thus

$$\|\eta \cdot \alpha h\|_{\mathbb{R}/\mathbb{Z}} \leq 2(\delta/2^d M)^2.$$

Thus, by Vinogradov's Lemma, there exists some nonzero integer $q \leq (\delta/2^d M)^{-1}$ such that

$$\|q\eta \cdot \alpha h\|_{\mathbb{R}/\mathbb{Z}} \leq 2(\delta/2^d M)/|I| \leq L(\delta/2^d M)^{-O(d)}/N_1.$$

□

We will now explain the iteration step as follows: at stage j, we let K_j, δ_j, M_j, L_j, N_j be the K, δ, M, L, and N_1 we apply Lemma C.2 with. At a first read through, the reader can take the case of when a and α have denominator N, and γ in the above iteration lemma is zero. The reader can then take $K_j = L_j = 0$, $N_j = N$. At step j of the iteration, we will also need to pigeonhole to a progression, of which the common difference we will denote as q_j. Again, at a first read through, with the assumption that a and α have denominator N, we can simply take $q_j = 1$. The quantity q_j will not be relevant during the use of the lemma, but rather, will be relevant for passing from the hypothesis of the bracket polynomial lemma, which is

$$|\mathbb{E}_{n \in [N]} e(\beta n + an \cdot \{\alpha h\})| \geq K^{-1}. \tag{2}$$

We iterate this lemma as follows: at the first step of the iteration, we apply the lemma with $K_1 = K$, $\delta_1 = \delta$, $M_1 = M$, $L_1 = 1$, $N_1 = N$ to obtain some nonzero η. Suppose without a loss of generality that $\eta_1 \neq 0$, that is, the first component of η is nonzero. By pigeonholing in progressions of common difference η_1, it follows that there exists a subprogression $\eta_1 \cdot [N_2] + r$ of common difference η_1 with size at least $\lfloor N/\eta_1 \rfloor$ such that at least $\delta N_2/2$ many h's in the subprogression that satisfy (2). Letting $q_2 := \eta_1$, and writing $h = q_2 k + r$, and noting that

$$\begin{aligned} an \cdot \{\alpha h\} &= an \cdot \{\alpha q_1 k + \alpha r\} \pmod 1 \\ &= an \cdot \{\alpha q_2 k\} + an \cdot \{\alpha r\} + \{an\}(-\{\alpha(q_2 k + r)\} + \{\alpha q_2 k\} + \{\alpha r\}) \pmod 1 \end{aligned}$$

we note that the last few terms, while may look daunting, will be eliminated via the Fourier complexity lemma (Lemma C.1). Since the Fourier complexity lemma only guarantees an L^1 Fourier complexity, in which case we only get a statement for *almost all* k, we will have to adjust $\delta/2$ by $\delta/2 - \delta^2/4$, or for simplicity, $\delta_2 = \delta/4$. Writing

$$\alpha_1 q_2 \equiv -\eta_2 \alpha_2 - \cdots - \eta_d \alpha_d + O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right)$$

and using a similar bracket polynomial expansion as above (and passing to a further subintervals so N_2 is size at least $\gg (\delta/2^d M)^{-O(d)} N_1/\eta_1$), we will obtain

$$an \cdot \{\alpha q_2 k\} = \tilde{a}n \cdot \{\alpha k\} + an \cdot O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k + [\text{Lower order bracketed terms}] \pmod 1$$

where $\tilde{a}_j = a_j \eta_1 - \eta_j a_1$ with $\tilde{a}_1 = 0$, and the lower order bracketed terms consist of at most $O(d)$ many bracketed terms such as $\{\ell \alpha k\}\{\beta n\} - \ell\{\alpha k\}\{\beta n\}$. The point is that

$$\{\alpha_1 q_2 k\} - \eta_2\{\alpha_2 k\} - \cdots \eta_d\{\alpha_d k\} - \left\{O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k\right\}$$

is an integer, so

$$an \cdot \left(\{\alpha_1 q_2 k\} - \eta_2\{\alpha_2 k\} - \cdots \eta_d\{\alpha_d k\} - \left\{O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k\right\}\right)$$

is equal to

$$\{an\} \cdot \left(\{\alpha_1 q_2 k\} - \eta_2\{\alpha_2 k\} - \cdots \eta_d\{\alpha_d k\} - \left\{O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k\right\}\right)$$

modulo one. Furthermore, since we pigeonholed N_2 into further subintervals, it follows that $\left[O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k\right]$ is constant as k ranges over the subinterval we pigeonholed into, so we may replace

$$an \cdot \left\{O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k\right\}$$

with

$$an \cdot O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k$$

at the cost of a single Fourier phase in n (which we also classify as a "lower order bracketed term"). Applying the Fourier complexity lemma to the threshold $\min(\delta^2/4, K^{-1}/2)$ and pigeonholing in one of the Fourier coefficients, we obtain that for at least $\delta N_2/4$ many $k \in [N_2]$,

$$(3) \qquad |\mathbb{E}_{n \in [N]} e(\tilde{a}n \cdot \{\alpha k\} + an \cdot O\left(\frac{(\delta/2^d M)^{-O(d)}}{N}\right) k + \alpha' k + \beta' n)| \geq (4q_2 K/2^d)^{-Cd^2}.$$

for some absolute constant C. Thus, we let $K_2 := (4q_2 K)^{4d^2}$ and $L_2 = (\delta/2^d M)^{-O(d)}$. The very important point of the Fourier complexity lemma is that it pushes all of the potential losses of δ to K. If we were to pigeonhole in k (i.e., using the fact that something like $\{\ell \alpha k\}\{\beta n\} - \ell\{\alpha k\}\{\beta n\}$ takes at most $O(\ell)$ many values and pigeonholing in one of those values), we would experience an iteration of $\delta \mapsto \delta^2$, which is disastrous. While $K \mapsto (4q_2 K/2^d)^{Cd^2}$ seems similarly disastrous, we will refer to step (2) at each step of the iteration (and redo the pigeonholing using the information gathered from the previous steps of the iteration) rather than step (3), so the effect of the iteration is $K_j = O(4q_j K_1/2^d)^{Cjd^2}$ rather than $K_j = O(4q_j K_{j-1}/2^d)^{Cd^2}$. Since η lies in a $(\delta/2^d M)$-tube around a, it follows that we may write

$$\eta = ta + O(\delta/2^d M)$$

and so

$$\eta_1 = ta_1 + O(\delta/2^d M)$$

$$\eta_j = ta_j + O(\delta/2^d M).$$

Multiplying the first equation by a_j and the second equation by a_1 and subtracting the two, we obtain

$$|\eta_1 a_j - \eta_j a_1| \leq 2|a|\delta/2^d M$$

and so we may take $M_2 = M$. This is more or less how the iteration will proceed at step j as well. Again, as emphasized before, we should refer back to the original hypothesis (2) in the iteration (and redo the pigeonholing in the progression using the information gathered from previous steps of the iteration) rather than an intermediate hypothesis (or basically the step before) as in (3). This will prevent K_j from increasing too quickly. At stage j, we will have j linearly independent vectors $\eta^1, \dots, \eta^j$ of size at most $(\delta/2^d M)^{-O(d)}$. Applying the bracket polynomial lemma yields either than $\tilde{a}_j$ is too small, in which case we are done, or there exists some η^{j+1} of size at most $(\delta_j/2^d M)^{-O(d)}$ such that

$$\|\eta^{j+1} \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \leq L_j(\delta_j/2^d M)^{-O(d)}/N_j.$$

Doing a similar computation as above with projecting in the directions orthogonal to $\eta^1, \dots, \eta^j$, we will once again obtain $\tilde{a}_{j+1}$ that is bounded above by M. We may take $q_{j+1} = q_j \|\eta^{j+1}\|_\infty$. Applying the Fourier complexity lemma to the $O(jd)$ many lower order bracketed terms yields that we can take $K_{j+1} = (2q_{j+1}K/2^d)^{O(jd^2)}$ and finally, we can simply take $L_{j+1} = jL_j(\delta_j/2^d M)^{-O(d)}$, and $N_{j+1} = N_j/L_j q_j$ and $\delta_{j+1} = \delta_j/4$. Thus, the iteration looks like[9]:

$$(\delta_j, M_j, K_j, N_j, L_j, q_j)$$

$$= (\delta_{j-1}/4, M_{j-1}, (2q_{j-1}K/2^d)^{O(jd^2)}, N_{j-1}/(L_{j-1}q_{j-1}), jL_{j-1}(\delta_{j-1}/2^d M)^{-O(d)}, (\delta_{j-1}/2^d M)^{-O(d)} q_{j-1}).$$

In the periodic nilsequences case, where we can assume that a and α have denominator N, the iteration becomes

$$(\delta_j, M_j, K_j, N_j, L_j, q_j) = (\delta_{j-1}/4, M_{j-1}, 0, N, 0, 1).$$

In either of these iterations, it's easy to see that all of the terms are increasing linearly except for K, which doesn't increase iteratively, but rather increases based on which step of the iteration we are on. Hence, we have ensured that all of the terms are bounded in the appropriate way.

Finally, we verify the orthogonality properties of the w_i's and the η_j's. Note that by Cramer's rule, that we can ensure that the w_i's are also bounded by $(\delta/2^d M)^{-O(d)^{O(1)}}$. More explicitely, we can express $\tilde{a}_j$ as a projection of a in the direction orthogonal to the η_j's. The first conclusion of Lemma C.2 that states that the components of $\tilde{a}_j$ are small are then equivalent to the orthogonal vector w_i applied to a is small, and hence by construction w_i must be orthogonal to η_j.

[9]If one used the uncertainty principle to ensure that η lies in a small tube around a instead of Minkowski's first theorem, δ would iterate to $\delta_{j+1} = \delta_j \log^2(1/\delta_j)$ and similarly $M_{j+1} = M_j \log^2(M_j)$, but the iteration in K and L would be slightly more efficient; it turns out that this would also work.

C.1. **Generalization of the first proof.** Since the above proof is rather cumbersome, we also provide a generalization of the proof of the refined bracket polynomial lemma given in Section 3 to the setting where a and α are real. We shall work with the following hypothesis:

Lemma C.3. *Suppose*

$$\|\beta + a \cdot \{\alpha h\}\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{K}{N}$$

for δN many $h \in [N]$ with $|a| \leq M$, $\alpha, a \in \mathbb{R}^d$. Then either $N \ll (\delta/KM)^{-O(d)^{O(1)}}$ or else there exists linearly independent vectors $w_1, \dots, w_r$ and $\eta_1, \dots, \eta_{d-r}$ in $\mathbb{Z}^d$ all having norm less than $(\delta/KM)^{-O(d)^{O(1)}}$ such that $\langle w_i, \eta_j \rangle = 0$ and

$$|w_i \cdot a| \leq (\delta/KM)^{-O(d)^{O(1)}}/N, \quad \|\eta_i \circ \alpha\|_{\mathbb{R}/\mathbb{Z}} \leq (\delta/KM)^{-O(d)^{O(1)}}/N.$$

Proof. First, we modify α by replacing it with a rational with denominator prime that is of size between N^2 and $2N^2$. We proceed similarly as in the first proof, reducing to the case of when $\beta = 0$, at the case of changing δ to $\frac{\delta}{2^d M}$ and changing K to $2K$.

This time, however, our lattice and convex set is slightly different, since we can no longer use that α has denominator N. Let Γ denote the lattice $(\alpha, 1/N)\mathbb{Z} + \mathbb{Z}^{d+1}$ and $B = \{|x_i| \leq \frac{1}{2}, |a \cdot x| \leq \frac{K}{N}, |x_{d+1}| \leq 1/2\}$. Since α has prime denominator, it follows that $|B \cap \Gamma| \geq \frac{\delta N}{2^d M}$. By Minkowski's second theorem (or rather Proposition $D.1$), there exists vectors $v_1, \dots, v_{d'}$ and corresponding $N_1, \dots, N_r$ such that $N_1 \cdots N_r = N\frac{\delta d^{-O(d)}}{2^d M}$ such that $P := \{\ell_1 v_1 + \cdots + \ell_r v_r : \ell_i \in [N_i]\}$ which lies inside B. Let V be the vector subspace generated by $v_1, \dots, v_r$. Since the set $X := \{|x_i| \leq 1/2\}$ has at most N points in Γ, by putting a fundamental parallelopiped of Γ at each lattice point of Γ in X, we see that X has volume at most N times $\delta(\frac{\delta d^{-O(d)}}{2^d M})^{-O(1)}$, and so the wedge product of $|v_1 \wedge \cdots \wedge v_r|$ has size $\frac{(d\delta/2^d M)^{O(d)}}{N}$ so by Ruzsa's covering lemma (Lemma $D.1$) X can be covered by $(\frac{\delta d^{-O(d)}}{2^d M})^{-O(1)}$ many translates of P, and so using the fact that X is connected, it follows that a dilation of $(\frac{\delta d}{2^d M})^{O(d)^{O(1)}}$ times the unit ball lies inside B.

Let $P_L = \{\ell_1 v_1 + \cdots + \ell_r v_r : \ell_i \in [LN_i]\}$. This lies in the intersection of a ball of radius dL and the set $\{|x_i| \leq \frac{K}{2}, |a \cdot x| \leq \frac{KL}{N}, |x_{d+1}| \leq L/2\}$. By the pigeonhole principle, at least $L^r \frac{\delta}{2^d M} d^{-O(d)}$ many elements in the ball that are a distance of at most $2/N$ from a lattice point in $\mathbb{Z}^{d+1}$. Since the elements of P_L lie in a lattice itself and the generators $v_1, \dots, v_r$ have wedge product $\frac{(d\delta/2^d M)^{O(d)}}{N}$ for some C, it follows that there are at least $L^r (d\delta/2^d M)^{O(d)}$ many distinct points $\mathbb{Z}^{d+1}$ inside $B_{2/N}(B_{dL}(0) \cap V)$ where $B_r(p)$ is the open ball around p with radius r. By Minkowski's second theorem again (or rather Proposition $D.1$) applied to $\mathbb{Z}^{d+1}$ and $B_{2/N}(B_{dL}(0) \cap V)$[10], choosing L to be larger than $(\frac{\delta}{2^d M} d^{-O(d)})^{O(1)}$, it follows that there exists r' linearly independent integer vectors $w_1, \dots, w_{r'}$ that lie inside the $B_K(0) \cap B_{2/N}(V)$ where $B_{2/N}(V)$ is the $2/N$-neighborhood of V and $M_1, \dots, M_{r'}$ such that $Q := \{m_1 w_1 + \cdots + m_{r'} w_{r'} : m_i \in [M_i]\}$ lies inside $B_K(0) \cap B_{2/N}(V)$

[10] This is primarily the difference between the periodic nilsequences case and the general case; in the periodic nilsequences case, we got a lattice for free while restricting to the integer vectors on the vector space we obtain. Here, the integer vectors don't have to lie on the vector subspace we obtain so we must apply Minkowski's second theorem one extra time.

with $|M_1 \cdots M_{r'}| \sim L^r \frac{\delta}{2^d M} d^{-O(d)}$. A volume packing argument and N and L being sufficiently large shows that, projecting Q to V, we obtain a r-dimensional generalized arithmetic progression whose generators have wedge product at most $(d\delta/2^d M)^{-O(d)^{O(1)}}$. Pulling back, we may thus assume that $r' = r$ and that the generators have wedge product at most $(d\delta/2^d M)^{-O(d)^{O(1)}}$. Applying Minkowski's second theorem a third time where the lattice is the lattice generated by $w_1, \dots, w_r$, the ambient space is the vector space generated by $w_1, \dots, w_r$, and the convex body is the ball of radius 1, we then obtain that there exists generators $w_1', \dots, w_r'$ of the lattice generated by Q, all of which have size at most $(d\delta/2^d M)^{-O(d)^{O(1)}}$. We claim that in some sense many elements of P_L lie close to the real span of the vectors in Q.

To prove this, let $\tilde{Q}$ be $B_{2/N}(\{m_1 w_1 + \cdots + m_r w_r : m_i \in \mathbb{R}, |m_i| \le M_i\})$. By Ruzsa's covering lemma, it follows that we may cover $B_{2/N}(V \cap B_{dL}(0))$ (since that aforementioned set contains at most $(2L)^r$ lattice points) with at most $(\frac{\delta}{2^d M} d^{-O(d)})^{-O(d)^{O(1)}}$ many translations of $\tilde{Q}$, and since $B_{2/N}(V \cap B_{dL}(0))$ is connected, symmetric, and convex, it follows that a $(d\delta/2^d M)^{O(d)^{O(1)}}$-dilation of it lies inside $\tilde{Q}$ and thus at least $N(d\delta/2^d M)^{O(d)^{O(1)}}$ many elements in P is distance of $\frac{2}{N}$ from an element in the real (actually, $(-1/2, 1/2)$-span) of $w_1, \dots, w_r$.

Let $\eta_1, \dots, \eta_{d+1-r}$ be integral orthogonal vectors to $w_1', \dots, w_r'$ (and so they are orthogonal to $w_1, \dots, w_r$). We claim that for at least $N(d\delta/2^d M)^{O(d)^{O(1)}}$ many elements v in P that $|\eta_i \cdot v| \le \frac{(d\delta/2^d M)^{-O(d)^{O(1)}}}{N}$. This would hold for elements v that are $\frac{2}{N}$ away from being in the real span of $w_1, \dots, w_r$. Unravelling everything and using Vinogradov's lemma, we obtain that

$$\|\eta_i \cdot (\alpha, 1/N)\|_{\mathbb{R}/\mathbb{Z}} \le \frac{(d\delta/2^d M)^{-O(d)^{O(1)}}}{N}.$$

Since the ball of radius $(d\delta/2^d M)^{O(d)^{O(1)}}$ with the intersection of V lies inside B (or rather, inside the interior of the generalized arithmetic progression P), and the dilation by $L(d\delta/2^d M)^{O(d)^{O(1)}}$ of P lies inside $\tilde{Q}$ (and the components of Q generate the vector space that all of the w_i''s lie in) and since the components of w_i' have size at most $(d\delta/2^d M)^{-O(d)^{O(1)}}$, it follows that w_i' is a distance of at most $(d\delta/2^d M)^{-O(d)^{O(1)}}/N$ away from a point in V and so consequently,

$$|w_i' \cdot a| \le \frac{K(d\delta/2^d M)^{-O(d)^{O(1)}}}{N}.$$

For the sake of cleanliness, we shall relabel w_i' as w_i. Finally, we eliminate the last component of the η_i's. The remaining computations will be quite similar to those in Corollary 3.2. To do so, suppose that they are not all zero. Then without a loss of generality, we may suppose that η_r has last component η_r^{d+1} nonzero. We then define for $i = 1, \dots, r-1$ $\tilde{\eta}_i = \eta_r^{d+1}\eta_i - \eta_i^{d+1}\eta_r$, and $\tilde{w}_i$ is equal to w_i except at the last coordinate where it is zero. We claim that

- $\tilde{w}_i$ and $\tilde{\eta}_j$ are orthogonal to each other
- $\tilde{w}_i$ and $\tilde{\eta}_j$ are linearly independent.

To show the first point, we see that

$$\langle w_i, \eta_j \rangle = \langle \tilde{w}_i, \eta_j \rangle + w_i^{d+1}\eta_j^{d+1} = 0$$

$$\langle w_i, \eta_r\rangle = \langle \tilde{w}_i, \eta_r\rangle + w_i^{d+1}\eta_r^{d+1} = 0.$$

Multiplying the first equation by η_r^{d+1} and the second by η_j^{d+1} and subtracting the two yields

$$\langle \tilde{w}_i, \tilde{\eta}_j\rangle = 0.$$

To verify the second point, we first verify that $\tilde{\eta}_j$ are linearly independent. Suppose there is some vector $\vec{c}$ with

$$\sum_j c_j\tilde{\eta}_j = \sum_j c_j(\eta_r^{d+1}\eta_j - \eta_j^{d+1}\eta_r) = 0.$$

Then rearranging the coefficients, we see that

$$\sum_{j<r} c_j\eta_r^{d+1}\eta_j + \left(\sum_{j<r} c_j\eta_j^{d+1}\right)\eta_r = 0.$$

Hence $c_j\eta_r^{d+1} = 0$ and since $\eta_r^{d+1} \neq 0$, it follows that $c_j = 0$ for all j. Hence $\tilde{\eta}_j$'s are linearly independent. To verify that $\tilde{w}_i$ are linearly independent, note that w_i are orthogonal to $(\tilde{\eta}_j, 0)$ and η_r. Since $(0,1)$ is not orthogonal to η_r, it follows that w_i cannot span $(0,1)$, so w_i, $(0,1)$ are linearly independent of each other, which implies that $\tilde{w}_i$ are linearly independent of each other. We can then verify that

$$|\tilde{w}_i \cdot a| \leq \frac{K(d\delta/2^dM)^{-O(d)^{O(1)}}}{N}$$

and

$$\|\tilde{\eta}_j \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{(d\delta/2^dM)^{-O(d)^{O(1)}}}{N}$$

as desired. □

Appendix D. Diophantine approximation and the geometry of numbers

In this section, we shall state relevant theorems from the geometry of numbers used mostly in the proofs of the refined bracket polynomial lemma. The relevant geometry of numbers theorems we'll be listing below can be found along with their proofs in [TV]. The first is Minkowski's first theorem, which one can find a statement and proof in [TV, Theorem 3.28].

Theorem 11 (Minkowski's first theorem)**.** *Let Γ be a lattice of $\mathbb{R}^d$. Let X be a convex body symmetric about the origin such that vol$(X) > 2^d$vol$(\mathbb{R}^d/\Gamma)$. Then X contains a nonzero vector $v \in \Gamma$.*

A generalization of this theorem (which is obtained from carefully iterating Minkowski's first theorem) is Minkowski's second theorem below. Before we state it, we shall need some terminology. Given a lattice Γ of $\mathbb{R}^d$ and a convex body X, the *successive minima* of X with respect to Γ, denoted λ_i, are defined as

$$\lambda_k := \inf\{\lambda > 0 : \lambda \cdot X \text{ contains } k \text{ independent vectors of } \Gamma\}.$$

We have $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq \cdots \leq \lambda_d < \infty$. Minkowski's second theorem [TV, Theorem 3.30] states that the successive minima satisfy the following property:

Theorem 12 (Minkowski's second theorem)**.** *Let Γ be a lattice of full rank in $\mathbb{R}^d$, B a symmetric convex body with successive minima $\lambda_1, \dots, \lambda_d$. Then there exists vectors $v_1, \dots, v_d$ such that*

- *For each $1 \leq j \leq d$, v_j lies on the boundary of $\lambda_j \cdot B$, but $\lambda_j \cdot B$ does not contain any vectors in Γ outside the span of $v_1, \dots, v_{j-1}$.*
- *The high dimensional polyhedra with vertices $\pm v_i$ contain no elements of Γ other than 0.*
- *We have*
$$\frac{2^d |\Gamma/(Span_{\mathbb{Z}}(v_1, \dots, v_d))|}{d!} \leq \frac{\lambda_1 \cdots \lambda_d vol(B)}{vol(\mathbb{R}^d/\Gamma)} \leq 2^d.$$

The next two statements below show, as a corollary of Minkowski's second theorem that such an intersection must contain a *generalized arithmetic progression* (see below for definition) which has size roughly $d^{-O(d)}$ times the intersection of the convex body and the lattice. The next lemma we'll need, a statement and proof of which can be found in [TV, Lemma 3.14], is Ruzsa's covering lemma:

Lemma D.1 (Ruzsa's covering lemma)**.** *For any bounded subsets A and B of $\mathbb{R}^d$ with positive measure, we may cover B by at most $min\left(\frac{vol(A+B)}{vol(A)}, \frac{vol(A-B)}{vol(A)}\right)$ many translates of $A - A$.*

We use Ruzsa's covering lemma in the various proofs of the refined bracket polynomial lemma to convex subsets of $\mathbb{R}^d$. In the case of convex and symmetric A and B, Ruzsa's covering lemma is extra powerful, since $A + A$ and $A - A$ are both just dilations of A.

Given a group G, a *generalized arithmetic progression* is a subset of G of the form $\{\ell_1 v_1 + \cdots + \ell_d v_d : \ell_i \in [N_i]\}$. The generalized arithmetic progression is *proper* if each of the elements $\ell_1 v_1 + \cdots + \ell_d v_d$ is distinct as ℓ_i ranges in $[N_i]$. The *rank* of a proper generalized arithmetic progression is the quantity d. As a consequence of Minkowski's second theorem and Ruzsa's covering lemma, we have the following, which is [TV, Lemma 3.33]:

Proposition D.1. *Let B be a symmetric convex body and let Γ be a lattice in $\mathbb{R}^d$. Then there exists a proper generalized arithmetic progression P in $B \cap \Gamma$ of rank at most d such that $|P| \geq O(d)^{-7d/2} |B \cap \Gamma|$.*

We will also need Vinogradov's lemma, a statement and proof of which can be found in [T1, Lemma 6]:

Lemma D.2 (Vinogradov's Lemma)**.** *Let $I \subseteq [N]$ be an interval and $P \colon \mathbb{Z} \to \mathbb{R}$ a polynomial of degree d of the form $P(n) = \sum_{i=0}^{d} \alpha_i n^i$. Suppose that $\|P(n)\| \leq \epsilon$ for δN many values of $n \in I$ with $0 < \delta, \epsilon < 1$. Then either*
$$N \ll \delta^{-\exp(O(d)^{O(1)})}$$
or
$$\epsilon \ll O(\delta)^{\exp(O(d)^{O(1)})}$$
or there exists some $q \ll O(\delta)^{-\exp(O(d)^{O(1)})}$ such that
$$\|q\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\delta^{-O(1)}\epsilon}{N^i}.$$

A consequence of Vinogradov's lemma is a periodic version:

Corollary D.1. *Let N be a large prime and $P \colon \mathbb{R} \to \mathbb{R}$ a polynomial of degree d of the form $P(n) = \sum_{i=0}^{d} \alpha_i n^i$ where α_i has denoinator N. Suppose $N \gg \delta^{-O_d(1)}$ (though with the*

Weil bounds one can perhaps even take $N \gg \delta^{-2}$) and there are at least δN many elements $n \in [N]$ such that

$$\|P(n)\|_{\mathbb{R}/\mathbb{Z}} \leq \epsilon.$$

Then either $\epsilon \gg \delta^{O_d(1)}$ or $\|\alpha_i\|_{\mathbb{R}/\mathbb{Z}} = 0$.

References

[A1] D. Altman. *On a conjecture of Gowers and Wolf.* Discrete Analysis 2022:10, 13 pp.

[A2] D. Altman. *A non-flag arithmetic regularity lemma and counting lemma.* https://arxiv.org/abs/2209.14083.

[AGH] L. Auslander, L. Green and F. Hahn. *Flows on homogeneous spaces.* Annals of Mathematics Studies 53, Princeton University Press, Princeton, N.J. 1963 vii+107 pp.

[AKZ] K. Adiprasito, D. Kazhdan, and T. Ziegler. *On the Schmidt and analytic ranks for trilinear forms.* https://arxiv.org/abs/2102.03659.

[BL] P.-Y. Bienvenu and T.H. Le. *Linear and quadratic uniformity of the Möbius function over $\mathbb{F}_q[t]$.* Mathematika 65 (2019), no. 3, 505-529.

[CM] A. Cohen and G. Moshkovitz. *Partition and Analytic Rank are Equivalent over Large Fields.* to appear in Duke Mathematical Journal.

[CS] P. Candela and O. Sisask. *Convergence results for systems of linear forms on cyclic groups, and periodic nilsequences.* SIAM J. Discrete Math. 28 (2) (2014), 786–810.

[FK] N. Frantzikinakis and B. Kuca. *Degree lowering for ergodic averages along arithmetic progressions.* https://arxiv.org/abs/2212.09819.

[G] L. Green. *Spectra of nilflows.* Bull. Amer. Math. Soc. 67 1961 414–415.

[Go] W.T. Gowers. *A new proof of Szemerédi's theorem.* GAFA, Geom. funct. anal. 11, 465–588 (2001).

[GM1] W.T. Gowers and L. Milićević. *A quantitative inverse theorem for the U4 norm over finite fields.* https://arxiv.org/abs/1712.00241.

[GM2] W.T. Gowers and L. Milićević. *An inverse theorem for Freiman multi-homomorphisms.* https://arxiv.org/abs/2002.11667.

[GT1] B. Green and T. Tao. *The quantitative behaviour of polynomial orbits on nilmanifolds.* Ann. Math. 175 (2012), no. 2, 465–540.

[GT2] B. Green and T. Tao. *New bounds for Szemerédi's theorem, III: A polylogarithmic bound for $r_4(N)$.* (2017) Mathematika. 63.944-1040.10.1112/S0025579317000316. https://arxiv.org/abs/1705.01703.

[GT3] B. Green and T. Tao, *Linear Equations in Primes. Ann. of Math.* 171:1753-1850, 2010

[GT4] B. Green and T. Tao, *The Mobius function is strongly orthogonal to nilsequences. Ann. of Math.* 175:541-566, 2012.

[GT5] B. Green and T. Tao. *An Arithmetic Regularity Lemma, An Associated Counting Lemma, and Applications.* In: Bárány, I., Solymosi, J., Sági, G. (eds) An Irregular Mind. Bolyai Society Mathematical

Studies, vol 21. Springer, Berlin, Heidelberg.

[GTZ1] B. Green, T. Tao, and T. Ziegler *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*. Ann. of Math. 176: 1231-1372 (2012)

[GTZ2] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers U^4 norm*. Glasgow Mathematical Journal, 53(1), 1-50. doi:10.1017/S0017089510000546

[GW] W.T. Gowers and J. Wolf, *Linear forms and quadratic uniformity for functions on $\mathbb{Z}/N\mathbb{Z}$*. Journal d'Analyse Mathématique. Vol 115.

[HK1] B. Host and B. Kra. *Nonconventional Ergodic Averages and Nilmanifolds*. Annals of Mathematics, vol. 161, no. 1, 2005, pp. 397–488.

[HK2] B. Host and B. Kra. *Nilpotent Structures in Ergodic Theory*. Mathematical Surveys and Monographs. Volume: 236; 2018; 427 pp.

[J] O. Janzer. *Polynomial bound for the partition rank vs the analytic rank of tensors*. Discrete Analysis 2020:7, 18 pp

[K1] B. Kuca. *True complexity of polynomial progressions in finite fields*. Proceedings of the Edinburgh Mathematical Society 64.3 (2021), 448-500.

[K2] B. Kuca. *On several notions of complexity of polynomial progressions*. Ergodic Theory and Dynamical Systems (2022), 1-55.

[KLT] D. Kim, A. Li, and J. Tidor. *Cubic Goldreich-Levin*. https://arxiv.org/abs/2207.13281.

[KZ] D. Kazhdan and T. Ziegler. *Approximate Cohomology*. Sel. Math. New Ser. 24, 499–509 (2018).

[L] M. Lazard. *Sur les groupes nilpotents et les anneaux de Lie*. Ann. Sci. Ecole Norm. Sup. (3) 71 (1954), 101–190.

[Le] A. Leibman. *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*. Ergodic Theory and Dynamical Systems 25 (2005), no. 1, 201–213.

[Len1] J. Leng. *A Quantitative Bound for Szemeredi's Theorem for a Complexity One Polynomial Progressions over Z/NZ*. https://arxiv.org/abs/2205.05540

[Len2] J. Leng. *The partition rank vs. analytic rank problem for cyclic groups II. Equidistribution for general and multiparameter nilsequences and applications*. In preparation.

[Len3] J. Leng. *Improved Quadratic Gowers Uniformity for the Möbius Function*. arxiv.org/abs/2212.09635

[LZ] A. Lampert and T. Ziegler. *Relative rank and regularization*. https://arxiv.org/abs/2106.03933.

[M] L. Milićević. *Polynomial bound for partition rank in terms of analytic rank*. Geometric and Functional Analysis 29 (2019), no. 5, 1503-1530.

[Ma1] F. Manners. *Quantitative bounds in the inverse theorem for the Gowers U^{s+1}-norms over cyclic groups*. https://arxiv.org/abs/1811.00718. 2018.

[Ma2] F. Manners. *Periodic nilsequences and inverse theorems on cyclic groups*. https://arxiv.org/abs/1404.7742.

[Mal] A. Mal'cev. *On a class of homogeneous spaces*. Izvestiya Akad. Nauk SSSR, Ser Mat. 13 (1949), 9–32.

[May] J. Maynard. *Simultaneous Small Fractional Parts of Polynomials*. Geom. Funct. Anal. 31, 150–179 (2021).

[MZ] G. Moshkovitz and D. Zhu. *Quasi-linear relation between partition and analytic rank*. https://arxiv.org/abs/2211.05780.

[P] G. Pólya. *Über ganzwertige ganze Funktionen*, Palermo Rend. 40: 1–16 (1935).

[S] T. Sanders. *On the Bogolyubov-Ruzsa lemma*. Analysis & PDE, Anal. PDE 5(3), 627-655, (2012).

[Sc] W. M. Schmidt. *Small fractional parts of polynomials*. American Mathematical Society, Providence, R.I., 1977. Regional Conference Series in Mathematics, No. 32.

[T1] T. Tao. *Equidistribution for Multidimensional Polynomial Phases*. https://terrytao.wordpress.com/2015/08/06/equidistribution-for-multidimensional-polynomial-phases/.

[T2] T. Tao, *Higher Order Fourier Analysis*. Graduate Studies in Mathematics. Vol 42. AMS, Providence, RI, 2012.

[TT] T. Tao and J. Teräväinen, *Quantitative Bounds for Gowers Uniformity of the Möbius and von Mangoldt Functions*. https://arxiv.org/abs/2107.02158.

[TV] T. Tao and V. Vu. *Additive Combinatorics*. (Cambridge Studies in Advanced Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511755149.

[Z] T. Ziegler. *Universal Characteristic Factors and Furstenberg Averages*. J. Amer. Math. Soc. 20 (2007), 53-97.

(James Leng) Department of Mathematics, UCLA, Los Angeles, CA 90095, USA.
Email address: jamesleng@math.ucla.edu