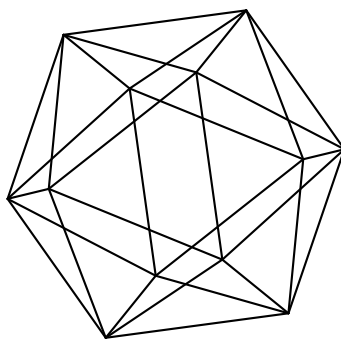# Max-Planck-Institut für Mathematik Bonn

A transference principle for systems of linear equations, and applications to almost twin primes

by

Pierre-Yves Bienvenue
Xuancheng Shao
Joni Teräväinen

A transference principle for systems of linear equations, and applications to almost twin primes

by

Pierre-Yves Bienvenue
Xuancheng Shao
Joni Teräväinen

Max-Planck-Institut für Mathematik
Vivatsgasse 7
53111 Bonn
Germany

Department of Mathematics
University of Kentucky
Lexington, KY 40506
USA

Mathematical Institute
Radcliffe Observatory Quarter
Woodstock Road
Oxford OX2 6GG
United Kingdom

A TRANSFERENCE PRINCIPLE FOR SYSTEMS OF LINEAR EQUATIONS, AND APPLICATIONS TO ALMOST TWIN PRIMES

PIERRE-YVES BIENVENU, XUANCHENG SHAO, AND JONI TERÄVÄINEN

ABSTRACT. The transference principle of Green and Tao enabled various authors to transfer Szemerédi's theorem on long arithmetic progressions in dense sets to various sparse sets of integers, mostly sparse sets of primes. In this paper, we provide a transference principle which applies to general affine-linear configurations of finite complexity.

We illustrate the broad applicability of our transference principle with the case of *almost twin primes*, by which we mean either Chen primes or "bounded gap primes", as well as with the case of primes of the form $x^2 + y^2 + 1$. Thus, we show that in these sets of primes the existence of solutions to finite complexity systems of linear equations is determined by natural local conditions. These applications rely on a recent work of the last two authors on Bombieri–Vinogradov type estimates for nilsequences.

1. INTRODUCTION

1.1. **The problem and its background.** Green and Tao famously proved [9] that the primes contain arbitrarily long arithmetic progressions. Their proof introduced an influential transference principle, stating that if a set of integers is dense inside a *pseudorandom* set, then it contains arbitrarily long arithmetic progressions. This is called a transference principle, since it transfers Szemerédi's theorem, which states that any dense subset of $\mathbb{Z}$ contains arbitrarily long arithmetic progressions, to a sparse setting. In fact, the proof of Green and Tao relied on Szemerédi's theorem as a black box.

More generally, given any admissible[1] affine-linear map $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$, and a subset $\mathcal{P}$ of the primes, one may ask whether $\mathcal{P}^t$ contains a tuple of the form $\Psi(\mathbf{n})$ with $\mathbf{n} \in \mathbb{Z}^d$. Equivalently, this may be formulated as the problem of determining which linear systems of equations can be solved inside $\mathcal{P}$.

Since the Green–Tao theorem, a lot of research has been devoted to this question. Note that k-term arithmetic progressions correspond to the map $\Psi(n, d) = (n, n + d, \ldots, n + (k - 1)d)$, so this case is handled by the Green–Tao theorem, which actually holds for dense subsets of the primes (or even not too sparse subsets, see [26]). Further, since Szemerédi's theorem holds for any given translation-invariant linear configuration in place of arithmetic progressions (that is, for homogeneous linear maps Ψ such that $(1, \ldots, 1) \in \Psi(\mathbb{Z}^d)$), the Green–Tao theorem also holds for these linear configurations.

Regarding general linear configurations, under the mere assumption that $\Psi = (\psi_1, \ldots, \psi_t)$ has finite complexity, (that is, no two of the forms ψ_i are affinely related), Green and Tao [11] provided a complete answer in the case where $\mathcal{P}$ is the full set $\mathbb{P}$ of primes, in fact giving an asymptotic formula for the number of $\mathbf{n} \in [N]^d$ for which $\Psi(\mathbf{n}) \in \mathcal{P}^t$ as $N \to \infty$.

[1]We say that $\Psi = (\psi_1, \ldots, \psi_t)$ is *admissible* if $(\psi_i(\mathbf{n}))_{\mathbf{n} \in \mathbb{Z}^d}$ has no fixed prime divisor for each $i \in [t]$.

Regarding subsets of the primes, it is known that a number of interesting sparse subsets of the primes contain arbitrarily long arithmetic progressions (or again, any given translation-invariant linear configuration). Indeed, the Chen primes

$$\mathcal{P}_{\text{Chen}} := \{p \in \mathbb{P} : \ p + 2 \in P_2\},$$

where P_2 is the set of integers which have at most two prime factors (counted with multiplicity), have this property by [32], and the *bounded gap primes*

$$\mathcal{P}_{\text{bdd},H} := \{n \in \mathbb{P} : \ |[n, n + H] \cap \mathbb{P}| \geq 2\}$$

for large H have this property by [24, 25]. Primes of the form $x^2 + y^2 + 1$ by [29] and Piatetski–Shapiro primes $\lfloor n^c \rfloor$ with $1 < c \leq c_0$ by [21] have this property as well.

However, very little is known when Ψ is not translation-invariant and simultaneously $\mathcal{P}$ is not the full set of the primes. When $\mathcal{P} \subset \mathbb{P}$ is the (dense) set of the shifted squarefree primes (i.e. primes p such that $p - 1$ is squarefree), for any finite complexity Ψ, an asymptotic for the number of $\mathbf{n} \in [N]^d$ for which $\Psi(\mathbf{n}) \in \mathcal{P}$ was proven by the first author [1]. When it comes to non-translation-invariant configurations in subsets of the primes, previous research has concentrated on the ternary Goldbach system, that is, the affine-linear map $\Psi_N(n, m) = (n, m, N - n - m)$. The subsets of the primes where it was studied include subsets of relative density above a certain threshold [20, 27], the set of primes of the form $x^2 + y^2 + 1$ [31], the set of primes in a given Chebotarev class [19], the set of Fouvry–Iwaniec primes $x^2 + p^2$ with p prime [16], and the set of primes admitting a given primitive root [7]. More relevantly for the present study, Matomäki and the second author [22] showed that any sufficiently large odd integer (resp. integer congruent to three modulo six) is a sum of three bounded gap primes (resp. three Chen primes). Both of these types of primes have properties akin to those of twin primes, and are therefore referred to as *almost twin primes.*

We mention that all of the papers [7, 16, 19, 22, 27, 31] rely on classical Fourier analysis, which is considerably simpler than higher order Fourier analysis, and hence the proofs do not adapt to any systems Ψ of complexity at least 2. The papers [21, 24, 25, 29, 32], in turn, all use the Green–Tao transference principle, and hence the proofs do not adapt to any non-translation-invariant configurations Ψ. Our main result handles the case of arbitrary finite complexity systems Ψ when $\mathcal{P}$ is the set of almost twin primes.

1.2. **Results on linear equations.** Now we state our results on linear equations in almost twin primes precisely. Let $\mathcal{H} = \{h_1, \ldots, h_m\}$ be an admissible m-tuple: for every prime p, there exists $n \in \mathbb{N}$ such that $\prod_{i \in [m]} (n + h_i) \not\equiv 0 \,(\text{mod}\, p)$. Let $\mathcal{P}_{\mathcal{H}} := \{n \in \mathbb{N} : |(n + \mathcal{H}) \cap \mathbb{P}| \geq 2\}$. Note that $\mathcal{P}_{\mathcal{H}}$ is actually not a subset of the primes; in fact $\mathcal{P}_{\text{bdd},H} = \mathbb{P} \cap \bigcup_{\mathcal{H} \subset [0,H]} \mathcal{P}_{\mathcal{H}}$. Define the weighted indicator functions of the Chen primes and $\mathcal{P}_{\mathcal{H}}$ by

$$\theta_1(n) := (\log n)^2 1_{\mathcal{P}_{\text{Chen}}}(n) 1_{p|n(n+2) \Longrightarrow p \geq n^{1/10}}, \quad \theta_2(n) := (\log n)^m 1_{\mathcal{P}_{\mathcal{H}}}(n) 1_{p|\prod_{i=1}^m (n+h_i) \Longrightarrow p \geq n^\rho},$$

where $m = |\mathcal{H}| \geq 2$ and $\rho \in (0, 1)$.

Then we know by Chen's theorem [2] and Maynard's theorem [23], upon assuming that m is large enough and ρ is small enough, that $\sum_{n \leq N} \theta_i(n) \gg N$ for $i \in \{1, 2\}$ (and we have upper bounds of the same order of magnitude by Selberg's sieve). Throughout this paper, we fix such m, ρ, and we also fix an admissible m-tuple $\mathcal{H}$ in the definition of θ_2.

Theorem 1.1 (Arbitrary linear configurations weighted by almost twin primes). *Let $i \in \{1, 2\}$. Let $\eta > 0$, $N, d, t, L \geq 1$, and let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of finite complexity whose homogeneous parts have coefficients bounded in modulus by L.*

Then there exists a constant $C_i(\Psi) \geq 0$ such that the following holds. Let $K \subset [-N, N]^d$ be a convex body satisfying $\text{Vol}(K) \geq \eta N^d$ and $\Psi(K) \subset [1, N]^t$. Then, for $N \geq N_0(L, \eta, d, t)$,

we have

$$(1.1) \qquad \sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{j \in [t]} \theta_i(\psi_j(\mathbf{n})) \gg_{L,\eta,d,t} C_i(\Psi) \mathrm{Vol}(K).$$

Further $C_i(\Psi) > 0$ unless there is an obstruction modulo some prime p. More precisely, $C_1(\Psi) > 0$ as soon as for every prime p there exists $\mathbf{n} \in \mathbb{Z}^d$ such that $\prod_{i \in [t]} \psi_i(\mathbf{n})(\psi_i(\mathbf{n}) + 2) \not\equiv 0 \,(\mathrm{mod}\, p)$ and $C_2(\Psi) > 0$ as soon as for every prime p there exists $\mathbf{n} \in \mathbb{Z}^d$ such that $\prod_{i \in [t]} \prod_{j \in [m]} (\psi_i(\mathbf{n}) + h_j) \not\equiv 0 \,(\mathrm{mod}\, p)$.

Note that the hypotheses imply that the non-homogeneous coefficients of Ψ are bounded in modulus by $(dL + 1)N$. It turns out that $C_i(\Psi) \gg_{d,t,L} 1$ whenever $C_i(\Psi) > 0$; therefore the right-hand side of the estimate (1.1) is $\gg_{d,t,L} \mathrm{Vol}(K)$ whenever it is not 0.

We can also obtain an analogous result for primes of the form $x^2 + y^2 + 1$. Let

$$(1.2) \qquad \theta_3(n) := (\log(2n))^{3/2} 1_{\mathbb{P}}(n) 1_{n = x^2 + y^2 + 1, x, y \in \mathbb{Z}}$$

be the weighted indicator function of primes of the form $x^2 + y^2 + 1$. By a result of Iwaniec [17], we have $\sum_{n \le N} \theta_3(n) \gg N$ (and we have an upper bound of the same order of magnitude from Selberg's sieve).

Theorem 1.2 (Arbitrary linear configurations weighted by primes of the form $x^2 + y^2 + 1$). *Theorem 1.1 continues to hold with θ_3 in place of θ_i. Moreover, $C_3(\Psi) > 0$ as soon as for every prime p there exists $\mathbf{n} \in \mathbb{Z}^d$ such that $\prod_{i \in [t]} \psi_i(\mathbf{n})(\psi_i(\mathbf{n}) + a(p)) \not\equiv 0 \,(\mathrm{mod}\, p)$, where $a(p) = -1$ if $p \equiv -1 \,(\mathrm{mod}\, 4)$ and $a(p) = 0$ otherwise.*

Theorem 1.1 has an immediate corollary to linear systems of equations within the Chen or bounded gap primes.

Corollary 1.3 (Linear equations in almost twin primes). *Let $L_1, \ldots, L_t : \mathbb{Z}^d \to \mathbb{Z}$ be linear forms. Consider the linear system of equations*

$$(1.3) \qquad L_i(\mathbf{n}) = 0 \quad \forall\, i \in 1, \ldots, t.$$

Suppose that the system has a solution in the positive real numbers. Then

  (i) *The system (1.3) has a solution in $\mathcal{P}_{\mathrm{Chen}}^d$, provided that it has a solution in A_p^d for every prime p, where $A_p = \{x \in \mathbb{Z}/p\mathbb{Z} : x(x + 2) \not\equiv 0 \,(\mathrm{mod}\, p)\}$.*
  (ii) *The system (1.3) has a solution in $\mathcal{P}_{\mathcal{H}}^d$, provided that it has a solution in B_p^d for every prime p, where $B_p = \{x \in \mathbb{Z}/p\mathbb{Z} : \prod_{j \in [m]} (x + h_j) \not\equiv 0 \,(\mathrm{mod}\, p)\}$.*

Proof. We may assume that each L_i is primitive (i.e., its coefficients have no common prime factor) and that the linear forms are linearly independent (so $t \le d$ and the system has full rank t). Since our system is homogeneous, we may assume that the span of the linear forms L_i does not contain a linear form which has only two or fewer nonzero coefficients (indeed, otherwise we may eliminate variables to obtain an equivalent system with fewer variables).

Therefore, the lattice of integer solutions of the system has a multiplicity-free parametrization of the form $\Psi(\mathbb{Z}^{t-d})$, where Ψ is a system of linear forms. The system Ψ has finite complexity, since no two forms of Ψ are linearly dependent, owing to the assumption about the span of the L_i's not containing linear forms with at most two nonzero coefficients.

Further, the local conditions (i) and (ii) imply that $C_1(\Psi) > 0$ and $C_2(\Psi) > 0$ respectively. We can then apply Theorem 1.1 to the convex body $K = \{\mathbf{x} \in \mathbb{R}^d : L_i(\mathbf{x}) \in [1, N] \,\forall i \in [t]\}$ with $N \to \infty$ conclude the proof. $\qquad \square$

As we will see, our method works more generally for $\mathcal{P}_{\mathcal{H},k} := \{n \in \mathbb{N} : |(n + \mathcal{H}) \cap \mathbb{P}| \ge k\}$ instead of $\mathcal{P}_{\mathcal{H}}$ whenever the admissible tuple $\mathcal{H}$ is sufficiently large in terms of k.

1.3. **Transference principles.** Given a finite complexity affine-linear map $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ (often referred to as a system of finite complexity), a function $f : [N] \to \mathbb{R}_{\geq 0}$ (typically a weighted indicator function of a set of arithmetic interest) and a convex body $K \subset \mathbb{R}^d$, the Ψ-count of f in K is given by

$$T_\Psi(f, K) := \sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} f(\psi_i(\mathbf{n})).$$

Thus Theorem 1.1 is about lower-bounding $T_\Psi(\theta_i, K)$ for $i \in \{1, 2\}$. Assume that

$$\sum_{n \in [N]} f(n) \geq \delta N$$

for some $\delta > 0$ and infinitely many integers N. If f takes its values in $[0, 1]$ and if additionally Ψ is a homogeneous translation-invariant linear map, a functional version of Szemerédi's theorem (see [30, Theorem 11.1]) allows one to prove that $T_\Psi(f, K) \gg_\delta \mathrm{Vol}(K)$. Now, if f is instead unbounded (for example, the von Mangoldt function), the Green–Tao transference principle consists in approximating f (assuming that $f \leq \nu$ for some "pseudorandom measure" ν), by a bounded function $\tilde{f} : [N] \to [0, 1]$ (called a dense model of f) in such a way that $T_\Psi(f, K) \approx T_\Psi(\tilde{f}, K)$, and invoking Szemerédi's theorem.

In our case, however, as we are interested in non-translation-invariant systems, we will need a different dense model and hence a different transference principle. To see the need for a different transference principle, note for example that the configuration $(x, x + y, x + 2y, y)$ does not occur in the dense set $A = \{n \in \mathbb{Z} : \|\sqrt{2}n^2\|_{\mathbb{R}/\mathbb{Z}} \in [1/3, 1/3 + 1/100]\}$, whereas any translation-invariant configuration can be found inside this set since it is dense.

In the case of the ternary Goldbach system $\Psi = \Psi_N : (n, m) \mapsto (n, m, N - n - m)$, the Matomäki–Shao transference principle [22] provides, under a Fourier-type condition, an approximating function $\tilde{f}$ to f satisfying again $T_\Psi(f, K) \approx T_\Psi(\tilde{f}, K)$, which is lower bounded pointwise: $\tilde{f}(n) \gg_\delta 1$; however, this does not generalize to the higher complexity case, as the set A above (which is Fourier uniform) demonstrates.

The proof of our main theorem, Theorem 1.1, produces more generally a lower bounded dense model for any system of finite complexity. This results in a transference principle (Theorem 3.2) of independent interest, which allows us to lower bound $T_\Psi(f, K)$ as desired for any function f which is majorized by a pseudorandom measure and dense in every "higher order Bohr set" (to be defined precisely later).

We then check these two conditions for our weighted indicator functions of almost twin primes, i.e. functions θ_1 and θ_2. This will follow by working out a reduction to the case of equidistributed higher order Bohr sets (Section 5) and then applying weighted versions of Chen's and Maynard's sieves (Section 6), and adapting a Bombieri–Vinogradov theorem for primes twisted by nilsequences [28], proven by the last two authors. We also note that our transference principle is slightly stronger than the Green–Tao transference principle even for translation-invariant systems in the sense that our pseudorandomness requirement is weaker (we do not need the correlation condition from [9]); we achieve this relaxation by applying the work of Dodos and Kanellopoulos [5].

The proof of Theorem 1.2 follows the same strategy, but the sieve input is now different (see Proposition 6.4), and we need a bit more arithmetic information for the Bombieri–Vinogradov estimates (see Lemma 7.6). This proof is given in Section 9.

2. Notation and preliminary definitions

Throughout the paper, we will use bold face characters to denote vectors or tuples. The expectation notation $\mathbb{E}_{x \in X}$ shall mean, for a finite set X, the averaging operator $\frac{1}{|X|} \sum_{x \in X}$. Further we will use the Vinogradov notation $f \ll g$ or $g \gg f$ whenever two functions f

and g from $\mathbb{N}$ to $\mathbb{R}$ satisfy $|f| \leq Cg$ for some constant $C > 0$; the parameters on which the implied constant C depends may be indicated as subscripts. The conjunction $f \ll g$ and $g \ll f$ will be denoted by $f \asymp g$. For any assertion A, the number 1_A is 1 if A is true and 0 if it is false. The indicator function of a set X will also be denoted by 1_X, which should generate no ambiguity. As usual, we denote by Λ, φ, d_k the von Mangoldt, Euler, and k-fold divisor functions, respectively. The greatest common divisor of two integers n and m will be denoted by (n, m). A vector or tuple of numbers will usually be denoted in bold font and its components in regular font. Given an integer N, we denote the interval of integers $\{1, \ldots, N\}$ by $[N]$. We will often identify the sets $[N]$ and $\mathbb{Z}/N\mathbb{Z}$, which we always implicitly do in the natural way (reduction modulo N). Thus a function f defined on $[N]$ may naturally be seen as a function on $\mathbb{Z}/N\mathbb{Z}$ and vice versa. When x is a positive real number, we define $[x] = [N]$ where $N = \lfloor x \rfloor$ is the integral part of x.

2.1. **Systems of linear forms.** Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms. We first define a quantity that captures the local behavior of Ψ modulo a prime p.

Definition 2.1 (Local factors). For each prime p, define the *p-adic local factor* of Ψ as

$$\beta_p(\Psi) := \mathbb{E}_{\mathbf{a} \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i \in [t]} \frac{p}{\varphi(p)} 1_{\psi_i(\mathbf{a}) \not\equiv 0 \,(\mathrm{mod}\, p)}.$$

Observe that Ψ is admissible as defined in footnote 1 if and only if $\beta_p(\Psi) \neq 0$ for each p.

We need to control the asymptotic behavior of β_p as p approaches infinity, whence the following easy variant of [11, Lemma 1.3].

Lemma 2.2. *Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms, and let p be a prime. Suppose that there are t_p linear forms among $\psi_1, \ldots, \psi_t$ modulo p such that no two of them are linearly dependent over $\mathbb{F}_p$, and that t_p is maximal for this property. Then*

$$\beta_p(\Psi) = \left(\frac{p}{\varphi(p)} \right)^{t - t_p} \left(1 + O_{d,t}(p^{-2}) \right).$$

Proof. Without loss of generality, assume that no two of $\psi_1, \ldots, \psi_{t_p}$ are proportional modulo p, and let $\Psi_p = (\psi_1, \ldots, \psi_{t_p})$. By maximality of t_p, $\beta_p(\Psi) = (p/\varphi(p))^{t - t_p} \beta_p(\Psi_p)$. Since no two of ψ_i, ψ_j with $1 \leq i < j \leq t_p$ can be linearly dependent modulo p, one can follow the proof of [11, Lemma 1.3] to conclude that $\beta_p(\Psi_p) = 1 + O_{d,t}(p^{-2})$. $\square$

The next crucial condition on linear systems that we will require is the aforementioned *finite complexity* condition, which we now quantify. For an affine-linear form ψ, let $\dot{\psi}$ be its linear part.

Definition 2.3 (Complexity of a system). For $A \subset [t]$, let V_A be the set of linear forms on $\mathbb{Z}^d$ generated by $\{\dot{\psi}_i \mid i \in A\}$. Let $i \in [t]$. A system Ψ of linear forms is said to be of *complexity at most k at i* if there exists a partition of $[t] \setminus \{i\}$ into at most $k + 1$ parts such that $\dot{\psi}_i \notin V_A$ for each part A of the partition. It is said to be of *complexity at most k* if it is of complexity at most k at any $i \in [t]$. The complexity is the minimum k such that the complexity is at most k, if there is any such $k \in \mathbb{N}$, in which case $k \leq t - 2$. Otherwise, it is said to be infinite.

A convenient parametrization of a system of finite complexity is the normal form (cf. [11, Definition 4.2]), which we now define; it facilitates multiple applications of the Cauchy–Schwarz inequality, yielding the generalized von Neumann theorem [11, Proposition 7.1] which we will use later. Let $\mathbf{e_1}, \ldots, \mathbf{e_d}$ be the canonical basis of $\mathbb{Z}^d$.

Definition 2.4 (Normal form of a system). The system Ψ is in *s-normal form at $i \in [t]$* if there exists a set $J_i \subset [t] \setminus \{i\}$ of cardinality at most $s + 1$ such that $\prod_{j \in J_i} \dot{\psi}_i(\mathbf{e_j}) \neq 0$ whereas

for all $k \in [t] \setminus \{i\}$, we have $\prod_{j \in J_i} \dot{\psi}_k(\mathbf{e_j}) = 0$. The system Ψ is in *s-normal form* if it is in s-normal form at each $i \in [t]$.

One may assume that $s \leq t - 2$. Clearly, a system in s-normal form has complexity at most s. Due to a simple linear-algebraic argument from [11, Theorem 4.5], we may assume in Theorems 1.1 and 1.2 that the system Ψ is in s-normal form for some $s \leq t - 2$.

2.2. Gowers norms.

Definition 2.5 (Gowers norms over abelian groups). Let Z be a finite abelian group. Let $g : Z \to \mathbb{C}$ be a function and $k \geq 1$ an integer. The *Gowers U^k norm* of g is the expression

$$\|g\|_{U^k(Z)} := \left(\mathbb{E}_{x \in Z} \mathbb{E}_{\mathbf{h} \in Z^k} \prod_{\boldsymbol{\omega} \in \{0,1\}^k} \mathcal{C}^{|\boldsymbol{\omega}|} g(x + \boldsymbol{\omega} \cdot \mathbf{h}) \right)^{2^{-k}},$$

where $\mathcal{C}$ is the conjugation operator and $|\boldsymbol{\omega}| := \sum_{i \in [k]} \omega_i$.

Definition 2.6 (Gowers norms over intervals). Given a function $f : \mathbb{Z} \to \mathbb{C}$ and an integer N, we define its Gowers norm $\|f\|_{U^k[N]}$ over the interval $[N]$ as

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

where $N' > 2N$ (say $N' = 2N + 1$ for concreteness) and $f \cdot 1_{[N]}$ and $1_{[N]}$ are extended to $\mathbb{Z}/N'\mathbb{Z}$ in the natural way. By [6, Lemma A.2], this definition is independent of the choice of N'.

Observe that if N and N' are two integers satisfying $\alpha N' \leq N \leq N'$ for some $\alpha > 0$ and a function $f : [N] \to \mathbb{C}$ is extended to $\mathbb{Z}/N'\mathbb{Z}$ by setting $f(n) = 0$ for $n \in \mathbb{Z}/N'\mathbb{Z} \setminus [N]$, then $\|f\|_{U^s[N]} \asymp_{\alpha,s} \|f\|_{U^s(\mathbb{Z}/N'\mathbb{Z})}$ (see [11, Lemma B.5]). Another norm that we will need is the L^p norm on $[N]$ equipped with the uniform probability measure, thus

$$\|f\|_{L^p[N]} := \left(\mathbb{E}_{x \in [N]} |f(x)|^p \right)^{1/p}$$

for $p \geq 1$ a real number. Finally, we define the dual Gowers norm over an interval by

$$\|f\|_{U^k[N]^*} := \sup_{\|g\|_{U^k[N]} = 1} |\mathbb{E}_{x \in [N]} f(x) g(x)|.$$

2.3. Nilsequences.

Definition 2.7 (Nilsequences). Let G be a connected, simply-connected nilpotent Lie group, and let $\Gamma \leq G$ be a lattice. A *filtration* $G_\bullet = (G_i)_{i=0}^\infty$ on G is an infinite sequence of subgroups of G (which are also connected, simply-connected nilpotent Lie groups) satisfying

$$G = G_0 = G_1 \supset G_2 \supset \cdots,$$

and such that the commutators satisfy $[G_i, G_j] \subset G_{i+j}$, and with the additional conditions that $\Gamma_i := \Gamma \cap G_i$ is a lattice in G_i for $i \geq 0$ and $G_{s+1} = \{\mathrm{id}\}$ for some s.

The least such s is called the *degree* of $G_\bullet$ and the manifold G/Γ is called a *nilmanifold*.

A *polynomial sequence* on G (adapted to the filtration $G_\bullet$) is a sequence $g : \mathbb{Z} \to G$ satisfying the derivative condition

$$\partial_{h_1} \cdots \partial_{h_k} g(n) \in G_k$$

for all $k \geq 0$, $n \in \mathbb{Z}$ and $h_1, \ldots, h_k \in \mathbb{Z}$, where $\partial_h g(n) := g(n+h)g(n)^{-1}$ denotes the discrete derivative with shift h.

Now fix a nilmanifold G/Γ, a filtration $G_\bullet$ of degree s and a polynomial sequence $g : \mathbb{Z} \to G$. Further, assume that the nilmanifold is equipped with a Mal'cev basis $\mathcal{X}$ (see [13,

Definition 2.1, Definition 2.4]). If $F : G/\Gamma \to \mathbb{C}$ is Lipschitz with respect to the metric on G/Γ induced by $\mathcal{X}$, we call a sequence of the form $n \mapsto F(g(n)\Gamma)$ a *nilsequence*. The degree of the nilsequence is then s, and it is said to be of *complexity* at most M if each of the degree s, the dimension of G/Γ, the rationality of $\mathcal{X}$ and the Lipschitz constant of F is at most M.

We now introduce a class of nilsequences of bounded degree and controlled complexity.

Definition 2.8. Let $s \geq 1$ and $\Delta, K \geq 2$. Define $\Psi_s(\Delta, K)$ to be the collection of all nilsequences $\psi : \mathbb{Z} \to \mathbb{C}$ of the form $\psi(n) = F(g(n)\Gamma)$, where
  (1) G/Γ is a nilmanifold of dimension at most Δ, equipped with a filtration $G_\bullet$ of degree $\leq s$ and a K-rational Mal'cev basis $\mathcal{X}$;
  (2) $g : \mathbb{Z} \to G$ is a polynomial sequence adapted to $G_\bullet$;
  (3) $F : G/\Gamma \to \mathbb{C}$ is a Lipschitz function satisfying $\|F\|_{\mathrm{Lip}(\mathcal{X})} \leq 1$.

Definition 2.9 (Equidistributed nilsequences)**.** For $\eta \in (0, 1)$ and $x \geq 2$, define $\Psi_s(\Delta, K; \eta, x)$ to be the collection of those nilsequences $\psi \in \Psi_s(\Delta, K)$ of the form $\psi(n) = F(g(n)\Gamma)$ that fulfill the additional condition that the sequence $(g(n)\Gamma)_{1 \leq n \leq 10x}$ is totally η-equidistributed in G/Γ (as defined in [13, Definition 1.2]).

We will loosely call such nilsequences η-*equidistributed*. We caution that this notation is slightly different from [28], in that we do not require $\int_{G/\Gamma} F = 0$ (where the integral is taken with respect to the unique Haar measure on G/Γ). We shall use $\Psi_s^0(\Delta, K; \eta, x)$ to denote the set of η-equidistributed nilsequences in $\Psi_s(\Delta, K; \eta, x)$ satisfying the additional condition that $\int_{G/\Gamma} F = 0$.

3. A transference principle for arbitrary systems of linear equations

A fundamental notion related to transference principles is that of *pseudorandom measures*, the basic philosophy being that if a function is bounded by such a measure, it behaves as if it was bounded by 1.

Definition 3.1. A function $\nu : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geq 0}$ is said to satisfy the $(d_0, t_0, L_0, \varepsilon)$-*linear forms conditions* if it satisfies the following. Let $1 \leq d \leq d_0$ and $1 \leq t \leq t_0$. For every finite complexity system of affine-linear forms $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ with linear coefficients bounded by L_0 in modulus, the following estimate holds :

$$(3.1) \qquad \left| \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/N\mathbb{Z})^d} \prod_{i \in [t]} \nu(\psi_i(\mathbf{n})) - 1 \right| \leq \varepsilon.$$

If it satisfies the (M, M, M, ϵ)-linear forms conditions, it is said to be (M, ε)-pseudorandom.

Observe that $\|\nu - 1\|_{U^k(\mathbb{Z}/N\mathbb{Z})} = O(\varepsilon^{1/2^k})$ as soon as ν satisfies the $(k+1, 2^k, 1, M, \varepsilon)$-linear forms conditions. Note that our definition is less restrictive than that of Green and Tao [11], since we do not require the so-called correlation condition.

The aim of this section is to prove the following theorem.

Theorem 3.2 (Transference principle for linear systems)**.** *Let $t, d, L, s \geq 1$, and let $\delta, \eta > 0$ be small. Then there exist constants $M \geq 1$ depending on d, t, L and $Y, \varepsilon > 0$ depending on d, t, L, δ such that the following holds. Let $\Psi = (L_1, \ldots, L_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms in s-normal form whose homogeneous parts have coefficients bounded by L.*

Let N be a large enough prime and α be small enough (both in terms of t, d, L, η). Let $K \subset [-N/4, N/4]^d$ be a convex body satisfying $\mathrm{Vol}(K) \geq \eta N^d$ and $\Psi(K) \subset [1, N]^t$. Lastly, for each $i \in [t]$, let $\lambda_i : [N] \to \mathbb{R}_{\geq 0}$ be a function. Assume that the following additional hypotheses hold.

(i) *There exists an (M, α)-pseudorandom measure $\nu : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geq 0}$ such that*

$$\lambda_i(n) \leq \nu(n) \quad \text{for each } i \in [t] \text{ and } n \in [N]$$

where we identify $[N]$ and $\mathbb{Z}/N\mathbb{Z}$ in the natural way.

(ii) *Each function λ_i is dense in higher order Bohr sets in the sense that*

$$\mathbb{E}_{n\in[N]}\lambda_i(n)\psi(n) \geq \delta\mathbb{E}_{n\in[N]}\psi(n)$$

for every nilsequence $\psi : \mathbb{Z} \to [0,1]$ of degree $\leq s$ and of complexity at most Y that satisfies $\mathbb{E}_{n\leq N}\psi(n) \geq \varepsilon$.

Then we have

(3.2) $$\mathbb{E}_{\mathbf{n}\in K\cap\mathbb{Z}^d}\lambda_1(L_1(\mathbf{n}))\lambda_2(L_2(\mathbf{n}))\cdots\lambda_t(L_t(\mathbf{n})) \geq 0.99\delta^t.$$

We now present the tools which will enable us to prove this. We need a notion of higher order Bohr sets, which are roughly speaking sets that are approximated by level sets of nilsequences to any given accuracy. The following definitions of s-measurable sets and s-factors are from [14, Section 2].

Definition 3.3 (s-measurable sets)**.** Let $s \geq 1$ and let $\Phi : \mathbb{R} \to \mathbb{R}$ be a growth function. A subset $E \subset [N]$ is called *s-measurable* with growth function Φ if, for any $M \geq 1$, there exists a degree $\leq s$ nilsequence $\psi : \mathbb{Z} \to [0,1]$ of complexity at most $\Phi(M)$ such that $\|1_E - \psi\|_{L^2[N]} \leq 1/M$.

Definition 3.4. If $\mathcal{B}$ is a partition of $[N]$, we call *atoms* its parts $E \in \mathcal{B}$. The *conditional expectation* of a function $f : [N] \to \mathbb{R}$ with respect to $\mathcal{B}$ is the function $\mathbb{E}[f|\mathcal{B}]$ which is constant on each atom, equal to the average of f on the atom.

Definition 3.5 (s-factors)**.** Let $s \geq 1$ and let $\Phi : \mathbb{R} \to \mathbb{R}$ be a growth function. A partition $\mathcal{B}$ of $[N]$ is called an *s-factor* of complexity at most M and growth function Φ if $\mathcal{B}$ contains at most M atoms and each atom is s-measurable of growth function Φ.

The following two propositions will be important in our proof of Theorem 3.2. The first one is the weak regularity lemma[2] proved in [10, Corollary 2.6].

Proposition 3.6 (Weak regularity lemma)**.** *Let $s \geq 1$.*

Let $f : [N] \to \mathbb{R}$ be a function with $|f(n)| \leq 1$ pointwise. Let $\varepsilon > 0$ be small. There exists an s-factor $\mathcal{B}$ of complexity $O_{s,\varepsilon}(1)$ and growth function depending only on s, ε such that $\|f - \mathbb{E}(f|\mathcal{B})\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon$.

In the just cited the reference, the Gowers norms are interval Gowers norms, but this makes no difference since the $U^{s+1}(\mathbb{Z}/N\mathbb{Z})$ and $U^{s+1}[N]$ norms are equivalent on bounded functions (see [6, Lemma A.4] for instance). We also state the dense model theorem from the work of Dodos and Kanellopoulos [5, Corollary 4.4].

Proposition 3.7 (Dense model theorem)**.** *Let $s \geq 1$ and let Z be a finite abelian group. Let $0 < \eta \leq 1$. Suppose that $\nu : Z \to \mathbb{R}^+$ satisfies $\|\nu - 1\|_{U^{2s}(Z)} \leq \eta$, and that $f : Z \to \mathbb{R}$ is a function such that $|f(n)| \leq \nu(n)$ pointwise. Then we may decompose $f = f_1 + f_2$, where $\sup_{n\in Z}|f_1(n)| \leq 1$ and $\|f_2\|_{U^s(Z)} = o_{\eta\to 0;s}(1)$.*

Further, if f is nonnegative, so is f_1.

---

[2]It was discovered recently that the reference [10] contains a slight error. See the arXiv version [14] for details and a correction. Nevertheless, the regularity lemma part of that reference is unaltered, only the counting lemma (and what depends on it) was not entirely correct.

We note that this version of the dense model theorem has weaker hypotheses than the one in [11] (it does not require the so-called correlation condition), a fact that will be important for us. A dense model for arithmetic progressions was also achieved without correlation conditions by Conlon, Fox and Zhao [3, 4], but their dense model is not as strong as we need since it is not close in the Gowers norms topology to the function to be modeled.

Finally, we state a version of the generalized von Neumann theorem [11, Proposition 7.1'].

Proposition 3.8 (Generalised von Neumann theorem). *Let t, d, L, s be positive integer parameters. Let δ, ε be in $(0,1)$ and $N \geq 1$. Then there is are positive constants D, depending on t, d and L such that the following holds. Let $\nu : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geq 0}$ be a (M, ε)-pseudorandom measure, and suppose that $f_1, \ldots, f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ are functions with $|f_i(x)| \leq \nu(x)$ for all $i \in [t]$ and $x \in \mathbb{Z}/N\mathbb{Z}$. Suppose that $\Psi = (\psi_1, \ldots, \psi_t)$ is a system of affine-linear forms in s-normal forms whose linear coefficients are bounded by L. Let $K' \subset (\mathbb{Z}/N\mathbb{Z})^d$ be identified with $K \cap \mathbb{Z}^d$ where $K \subset [-N/4, N/4]^d$ is a convex set. Finally, suppose that*

$$(3.3) \qquad \min_{1 \leq j \leq t} \|f_j\|_{U^{s+1}[N']} \leq \delta.$$

Then we have

$$\mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/N\mathbb{Z})^d} 1_{K'} \prod_{i \in [t]} f_i(\psi_i(\mathbf{n})) = o_{\delta \to 0}(1) + o_{N \to \infty; \delta}(1) + o_{\varepsilon \to 0; \delta}(1)$$

where the $o(1)$ terms may also depend on d, t, L.

In the cited reference, the parameter ε is itself $o_{N \to \infty}(1)$ but we make it independent here, whence the slightly different statement.

We are now ready to state and prove a crucial lemma.

Proposition 3.9 (Decomposition into a uniformly lower bounded and Gowers uniform components). *Let $s \geq 1$ and let $N \geq 1$ be an integer. Let $\delta, \varepsilon, \rho$ be real constants in the interval $(0,1)$. Then there exist quantities $\iota \in (0,1)$ depending on s, ε and $Y > 0$ depending on $s, \varepsilon, \rho, \delta$ such that the following holds. Suppose that $\nu : [N] \to \mathbb{R}_{\geq 0}$ satisfies $\|\nu - 1\|_{U^{2s+2}(\mathbb{Z}/N\mathbb{Z})} \leq \eta$, where we naturally identify $[N]$ with $\mathbb{Z}/N\mathbb{Z}$ and η is sufficiently small in terms of ε, ρ, s. Let $f : [N] \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq \nu(n)$ pointwise. Further, suppose that*

$$\mathbb{E}_{n \in [N]} f(n) \psi(n) \geq \delta \mathbb{E}_{n \in [N]} \psi(n)$$

for every nilsequence $\psi : \mathbb{Z} \to [0,1]$ of degree $\leq s$ and of complexity at most Y that satisfies $\mathbb{E}_{n \leq N} \psi(n) \geq \iota N$.

Then there exists a decomposition $f = f_3 + f_4$ where $f_3 \geq (1 - \rho)\delta$ pointwise and $\|f_4\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon$.

Proof. Without loss of generality, we may assume that ε is small enough in terms of δ and ρ. Let $\varepsilon' \leq \varepsilon/3$ be a sufficiently small constant, to be determined later. Applying Proposition 3.7, we may write $f = f_1 + f_2$ where f_1 takes its values in $[0,1]$ and $\|f_2\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon'$. Then using Proposition 3.6 on f_1, we decompose $f_1 = h + g$ where $h = \mathbb{E}[f_1 | \mathcal{B}]$ and $\mathcal{B}$ is an s-factor of complexity $O_{s,\varepsilon}(1)$ and growth function Φ, and $\|g\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon/3$. The growth function Φ of $\mathcal{B}$ depends only on ε and s.

Note that $h = \sum_{E \in \mathcal{B}} c_E 1_E$ where $c_E = \frac{1}{|E|} \sum_{n \in [N]} f_1(n) 1_E(n)$ for any atom E of $\mathcal{B}$. Fix $c = (\varepsilon/3)^{2^{s+1}} / |\mathcal{B}|$, so $c^{-1} = O_{s,\varepsilon}(1)$. We effect the splitting

$$h = \sum_{E \in \mathcal{B}} (c_E + \delta 1_{|E| < cN}) 1_E - \sum_{\substack{E \in \mathcal{B} \\ |E| < cN}} \delta 1_E.$$

We denote by h_1 the first sum and h_2 the second one.

Since $\delta \in [0,1]$ for any $E \in \mathcal{B}$, we see that $\|h_2\|_{L^1[N]} \leq c|\mathcal{B}| = (\varepsilon/3)^{2^{s+1}}$. Crudely estimating by the triangle inequality, this implies that $\|h_2\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon/3$. Now write $f_3 = h_1$ and $f_4 = g + f_2 + h_2$. By the triangle inequality for Gowers norms, we have

$$\|f_4\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq 3 \cdot \varepsilon/3 = \varepsilon.$$

Our aim is then to show that

(3.4) $$c_E \geq (1-\rho)\delta, \quad \text{whenever} \quad |E| \geq cN,$$

after which $f_3 = h_1 \geq (1-\rho)\delta$ pointwise follows.

Fix a large enough constant $M > 0$ in terms of c, δ, ρ (explicitly, we may take $M = 4/(c\delta\rho)$) and an atom $E \in \mathcal{B}$ satisfying $|E| \geq cN$. By definition of an s-factor, we may write $1_E = \psi + g_{\text{sml}}$ where $\|g_{\text{sml}}\|_{L^2[N]} \leq 1/M$ and ψ (depending on E) is a nilsequence of degree at most s whose complexity is bounded by $Y := \Phi(M) = O_{s,\varepsilon,\rho,\delta}(1)$. By the Cauchy–Schwarz inequality, we have $|\sum_{n \in [N]} f_1(n)g_{\text{sml}}(n)| \leq N/M$. Therefore,

$$c_E \geq \frac{1}{|E|} \sum_{n \in [N]} f_1(n)\psi(n) - 1/(cM).$$

We now recall that $f = f_1 + f_2$, so that (3.4) follows once we show that

(3.5) $$\sum_{n \in [N]} f(n)\psi(n) \geq \delta |E| (1 - 1/(cM)) \quad \text{and} \quad \left| \sum_{n \in [N]} f_2(n)\psi(n) \right| \leq \rho\delta |E| /2,$$

upon setting $M = 4/(c\delta\rho)$. First we bound the correlation of f_2 and ψ. Since f_2 has small U^{s+1} norm, it would be convenient to replace ψ by a function of bounded dual U^{s+1} norm. To achieve this, we invoke[3] [11, Proposition 11.2], which yields for any $\kappa > 0$ a splitting $\psi = \psi_1 + \psi_2$ where $\|\psi_1\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})^*} \leq K$ for some $K = O_{\kappa,Y}(1)$, while $\|\psi_2\|_\infty \leq \kappa$. We infer that

(3.6) $$\left| \sum_{n \in [N]} f_2(n)\psi_1(n) \right| \leq \|f_2\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})}\|\psi_1\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})^*}N \leq \varepsilon'KN.$$

Further, since $|f_2| \leq \nu + 1$ pointwise,

(3.7) $$\left| \sum_{n \in [N]} f_2(n)\psi_2(n) \right| \leq \|\psi_2\|_\infty \sum_{n \in [N]} (\nu(n) + 1) \leq N(2 + o(1))\|\psi_2\|_\infty \leq 3\kappa N \leq 3\kappa c^{-1}N.$$

Now, if we choose $\kappa = c\rho\delta/12$ and $\varepsilon' = \rho\delta c/(4K)$ (thus $(\varepsilon')^{-1} = O_{\rho,s,\delta,\varepsilon}(1)$), we have $\varepsilon'K + 3\kappa < c\rho\delta/2$, so by combining (3.6) and (3.7) we obtain the required bound (3.5) for f_2.

It remains to be shown that the correlation of f and ψ obeys the lower bound in (3.5). By the definition of ψ, we have

$$\sum_{n \in [N]} \psi(n) = |E| - \sum_n g_{\text{sml}}(n) \geq |E| (1 - 1/(cM)) \geq |E|/2 \geq cN/2,$$

so recalling the "denseness in higher order Bohr sets" hypothesis of Proposition 3.9 (letting $\iota = c/2$ there) and the fact that the complexity of ψ is at most Y, the desired estimate (3.5) follows. This was enough to complete the proof. $\square$

Now we may prove Theorem 3.2.

---

[3]In the cited reference, written at a time where the theory of nilsequences was just emerging, the result is stated for linear nilsequences. However, nowadays we know that any polynomial nilsequence may be realized as a linear one, see [15, Appendix C]. Also the result is stated in terms of interval Gowers norms, but the proof naturally yields $\|\psi_1\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})^*} \leq K$ first as it moves from intervals to cyclic groups.

Proof of Theorem 3.2. Fix $\kappa > 0$ small enough (to be determined later). Let ρ be small enough in terms of t (say $\rho = 1/(10000t)$). By hypothesis (i), if N is large enough, we may apply Proposition 3.9, thus obtaining a decomposition $\lambda_i = \lambda_i^{(1)} + \lambda_i^{(2)}$ for each $i \in [t]$ where $\lambda_i^{(1)} \geq (1 - \rho)\delta$ pointwise and $\|\lambda_i^{(2)}\|_{U^{s+1}(\mathbb{Z}/N\mathbb{Z})} \leq \kappa$. Inserting this decomposition in the left-hand side of (3.2), we obtain a splitting of the average into 2^t terms:

$$\mathbb{E}_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{j=1}^t \lambda_j(L_j(\mathbf{n})) = \sum_{a_1, a_2, \ldots, a_t \in \{1,2\}} \mathbb{E}_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{j=1}^t \lambda_j^{(a_j)}(L_j(\mathbf{n})).$$

One of the 2^t terms involves only the functions $\lambda_i^{(1)}$; since $\lambda_i^{(1)}$ is pointwise lower bounded by $(1 - \rho)\delta$, this term is at least

$$(1 - \rho)^t \delta^t \geq 0.999\delta^t,$$

since $\rho = 1/(10000t)$. Any other term involves at least one copy of a uniform function $\lambda_i^{(2)}$. Let $\mathbf{a} \in \{1,2\}^t \setminus \{(1, \ldots, 1)\}$. Since $K \subset [-N/4, N/4]^d$ and $\Psi(K) \subset [1, N]^t$, one may identify K with a subset of $(\mathbb{Z}/N\mathbb{Z})^d$, which we also denote by K, and write

$$(3.8) \qquad \mathbb{E}_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{j=1}^t \lambda_j^{(a_j)}(L_j(\mathbf{n})) = \mathrm{Vol}(K)^{-1} \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/N\mathbb{Z})^d} 1_K(\mathbf{n}) \prod_{j=1}^t \lambda_j^{(a_j)}(L_j(\mathbf{n})).$$

According to Proposition 3.8 (for which we need N to be a prime, M to be sufficiently large in terms of d, t, L and the fact that Ψ is in s-normal form), the right-hand side of equation (3.8) is bounded by $o_{N \to \infty; \kappa}(1) + o_{\alpha \to 0; \kappa} + o_{\kappa \to 0}(1)$. Therefore, choosing first κ appropriately, and then N sufficiently large and α sufficiently small, we conclude the proof. $\square$

The rest of the paper is devoted to establishing the hypotheses (i) and (ii) of Theorem 3.2 for the functions θ_1 and θ_2 (and θ_3 in Section 9); we start with hypothesis (i).

4. W-trick and Pseudorandom Majorants

4.1. W-trick. We wish to apply Theorem 3.2 to prove our main theorem, but an initial problem is that the indicator functions of almost twin primes are not bounded by a pseudorandom majorant, as they are biased modulo small primes. We will first have to remove these biases modulo small primes to obtain a pseudorandomly majorized function.

We introduce the general framework we will work with in this section. Let $\rho \in (0, 1)$, $r \geq 1$, and let $\mathcal{H} = \{h_1, \ldots, h_r\} \subset \mathbb{N}$ be a set of r pairwise distinct integers and let $\theta = \theta_{\mathcal{H}} : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be any function supported on the set

$$\{n \in \mathbb{N} : p \mid n + h_j \implies p > w \text{ for all } j \in [r]\}$$

and satisfying the upper bound

$$\theta(n) \leq \log^r(n + 2)$$

for all $n \geq 0$. Observe that the functions θ_1 and θ_2 our main theorem deals with have these properties (with $r = 2$ in the case of θ_1 and $r = m$ in the case of θ_2).

Whenever w is an integer, let $W = W(w) = \prod_{p \leq w} p$. Given integers $q > 0$ and b, let

$$(4.1) \qquad \theta_{q,b}(n) := \left(\frac{\varphi(q)}{q}\right)^r \theta(qn + b).$$

Proposition 4.1 (Reduction to W-tricked sums). *Let the notation be as above. Also let $\eta > 0$, $\gamma > 0$, $N, L, d, t, w \geq 1$. Let $\Psi = (\psi_1, \ldots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a finite complexity system of affine-linear forms. Suppose that $\Psi_{\mathcal{H}} := (\psi_i + h_j)_{i \in [t], j \in [r]}$ is admissible and that the linear*

coefficients of Ψ as well as the elements of $\mathcal{H}$ are bounded by L. Let $K \subset [-N, N]^d$ be a convex body satisfying $\mathrm{Vol}(K) \geq \eta N^d$ and $\Psi(K) \subset [1, N]^t$. Suppose that θ satisfies

$$(4.2) \qquad \sum_{\substack{\mathbf{n} \in \mathbb{Z}^d \\ W\mathbf{n}+\mathbf{a} \in K}} \prod_{i=1}^{t} \theta_{W,c_i(\mathbf{a})}(\psi_i'(\mathbf{n})) \geq \gamma W^{-d} \mathrm{Vol}(K),$$

for each $\mathbf{a} \in A$, where

$$A = A_{\Psi,\mathcal{H}} = \{\mathbf{a} \in [W]^d : \forall (i,j) \in [t] \times [r] \quad (\psi_i(\mathbf{a}) + h_j, W) = 1\}$$

and for each $i \in [t]$, the integer $c_i(\mathbf{a}) \in [W]$ and the form $\psi_i' : \mathbb{Z}^d \to \mathbb{Z}$ are uniquely defined by the relation $\psi_i(W\mathbf{n} + \mathbf{a}) = W\psi_i'(\mathbf{n}) + c_i(\mathbf{a})$. Then, provided that w is large enough in terms of d, t, L, we have

$$(4.3) \qquad \sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^{t} \theta(\psi_i(\mathbf{n})) \geq \gamma/2 \cdot \prod_{p} \beta_p \cdot \mathrm{Vol}(K),$$

where the local factors $\beta_p = \beta_p(\Psi_{\mathcal{H}})$ are as defined in Definition 2.1.

Proof. We write

$$\mathbb{Z}^d \cap K = \bigcup_{\mathbf{a} \in [W]^d} (\mathbb{Z}^d \cap (WK_{\mathbf{a}} + \mathbf{a})),$$

where

$$K_{\mathbf{a}} := \{\mathbf{x} \in \mathbb{R}^d : W\mathbf{x} + \mathbf{a} \in K\}$$

is again a convex body. Putting

$$F(\mathbf{n}) := \prod_{j=1}^{t} \theta(\psi_j(\mathbf{n}))$$

we can write the left-hand side of (4.3) as

$$(4.4) \qquad \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} F(\mathbf{n}) = \sum_{\mathbf{a} \in [W]^d} \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K_{\mathbf{a}}} F(W\mathbf{n} + \mathbf{a}).$$

We note that if $\psi_i(\mathbf{a}) + h_j$ is not coprime to p for some $i \in [t]$, some $j \in [r]$ and some prime $p \leq w$ and w is large enough, then for each $\mathbf{n} \in K_{\mathbf{a}} \cap \mathbb{Z}^d$ we have $F(W\mathbf{n} + \mathbf{a}) = 0$: indeed, if $(\psi_i(\mathbf{a}) + h_j, p) > 1$, the integer $\psi_i(W\mathbf{n} + \mathbf{a}) + h_j$ has a prime factor $p < w$. Thus, the residues $\mathbf{a}$ which bring a nonzero contribution to the right-hand side of (4.4) are all mapped by Ψ to tuples $(b_1, \ldots, b_t)$ for which $b_i + h_j$ is coprime to W for all $i \in [t], j \in [r]$.

Recalling the definitions of $A = A_{\Psi,\mathcal{H}}$, the integers $c_i(\mathbf{a})$, the forms ψ_i' and $\theta_{W,b}$, we can then rewrite equation (4.4) as

$$(4.5) \qquad \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} F(\mathbf{n}) = \left(\frac{W}{\varphi(W)}\right)^{rt} \sum_{\mathbf{a} \in A} \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K_{\mathbf{a}}} \prod_{i=1}^{t} \theta_{W,c_i(\mathbf{a})}(\psi_i'(\mathbf{n})).$$

By our assumption (4.2), we have

$$(4.6) \qquad \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K_{\mathbf{a}}} \prod_{i=1}^{t} \theta_{W,c_i(\mathbf{a})}(\psi_i'(\mathbf{n})) \geq \gamma W^{-d} \mathrm{Vol}(K) = \gamma \mathrm{Vol}(K_{\mathbf{a}})$$

for each $\mathbf{a} \in A$, so to obtain the conclusion (4.3) it suffices to prove that

$$(4.7) \qquad \left(\frac{W}{\varphi(W)}\right)^{rt} |A_{\Psi,\mathcal{H}}| \geq \frac{1}{2} W^d \prod_{p} \beta_p.$$

Note that by the Chinese remainder theorem we have

$$\left(\frac{W}{\varphi(W)}\right)^{rt} |A_{\Psi,\mathcal{H}}| = W^d \prod_{p \leq w} \beta_p.$$

Lemma 2.2 implies that $\beta_p = 1 + O_{d,t,L}(p^{-2})$ whenever $|\mathcal{H} \pmod p| = r$ (which is the case whenever $w > H$), and $\beta_p > 0$ for any p since $\Psi_{\mathcal{H}}$ is an admissible system. Therefore $\prod_p \beta_p$ is convergent and if $w > H$ we have $\prod_{p \leq w} \beta_p = (1 + O_{d,t,L}(1/w)) \prod_p \beta_p$. Taking w large enough in terms of d, t, L, this concludes the proof of Proposition 4.1. $\square$

In fact Lemma 2.2 implies that $\beta_p = 1 + O(p^{-1})$ and $\beta_p = 1 + O(p^{-2})$ except when $p \mid \prod_{i,j}(h_i - h_j)$. Combining this with the fact that $\prod_{w < p \mid q}(1 + O(p^{-1})) \leq \prod_{w < p < w + \omega(q)}(1 + O(p^{-1})) = O(1 + \log\log q / \log w)$ we have

$$\prod_{p > w} \beta_p \ll O(1 + \log\log H / \log w)$$

so the weaker hypothesis $H \leq \exp(w^{O(1)})$ could suffice instead of $w > H$ at the cost of replacing $1/2$ by a worse constant.

Also we note that the system Ψ' introduced above differs of Ψ only in the constant term, and so it is of finite complexity whenever Ψ is.

4.2. Pseudorandom majorants.

In order to prove Theorem 1.1, it remains to establish the lower bound (4.2) when θ is either θ_1 or θ_2, which we will do by invoking Theorem 3.2. In order to appeal to this theorem, we need to supply a pseudorandom majorant for the function $\theta_{W,b}$ where b is coprime to W. The only properties of θ that we need for this construction are that it is supported on the set

$$\{n \in \mathbb{N} : p \mid n + h_j \implies p > n^\rho \text{ for all } j \in [r]\}$$

and satisfies $0 \leq \theta(n) \leq \log^r(n+2)$ (and these are satisfied for θ_1, θ_2 with $r = 2, m$, respectively). Let

(4.8)
$$B_{\mathcal{H}} := \{b \in \mathbb{N} : \forall j \in [r] \ (b + h_j, W) = 1\}.$$

The next proposition provides us with a pseudorandom majorant.

Proposition 4.2 (Pseudorandom majorants). *Let $M \geq 1$ and $s \in [t_0]$ be integers. Let $\epsilon > 0$. Assume that N and w are large enough in terms of (M, ϵ) and satisfy $w \leq \log\log N$. Let $\mathbf{b} = (b_1, \ldots, b_s)$ be in $B_{\mathcal{H}}^s$ satisfy $|b_i - b_j| \leq M$ for any $(i, j) \in [s]^2$. Suppose also that θ is as above. Then there is an (M, ϵ)-pseudorandom measure $\nu_{\mathbf{b}} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geq 0}$ and a constant $c \in (0, 1)$ depending on M only such that*

$$\theta_{W,b_1}(n) + \cdots + \theta_{W,b_s}(n) \ll_M \nu_{\mathbf{b}}(n)$$

for all $n \in [N^c, N] \subset \mathbb{Z}/N\mathbb{Z}$.

The fact that the bound does not necessarily hold on the full interval $[N]$ is not a serious restriction, as we may impose θ to be supported in $[N^c, N]$ without changing the left-hand side of (4.2) and (4.3) by more than $O(N^{cd} \log^{O(1)} N) = o(N^d)$. Zhou [32] and Pintz [24] already constructed pseudorandom majorants for Chen and bounded gap primes respectively. We provide here a similar construction. Let $R = N^\gamma$ for some small $\gamma > 0$ to be chosen appropriately. Relying on Green and Tao's "smoothed" approach [11], we define

(4.9)
$$\Lambda_{\chi,\gamma}(n) := \log R \Big(\sum_{\ell \mid n} \mu(\ell) \chi \Big(\frac{\log \ell}{\log R} \Big) \Big)^2,$$

where $\chi : \mathbb{R} \to [0,1]$ is a smooth, even function supported on $[-2,2]$ satisfying $\chi(0) = 1 = \int_0^2 |\chi'|^2$. Finally let $\Lambda_{\chi,\gamma,\mathcal{H}}(n) := \prod_{h \in \mathcal{H}} \Lambda_{\chi,\gamma}(n+h)$. Note that this function is periodic (of period $\prod_{\ell \leq R^2} \ell$ for instance), so we extend it on $\mathbb{Z}$ as a periodic function. Once W-tricked, this will be a pseudorandom measure. Ultimately, this is a consequence of the following proposition.

Proposition 4.3 (Correlations of sieve weights). *Let $d, t \geq 1$ be integers. Let $D, \eta > 0$. Let $\Psi = (L_1, \ldots, L_t)$ be a finite complexity system of affine-linear forms in d variables. Suppose that $\gamma > 0$ is sufficiently small in terms of d, t. Suppose that the linear coefficients, as well as the integers $h_1, \ldots, h_r$, are bounded in magnitude by D. Assume that w is sufficiently large in terms of d, t, D. Let $K \subset [-N,N]^d$ satisfy $\mathrm{Vol}(K) \geq \eta N^d$. Suppose $b_1, \ldots, b_t$ are in $B_{\mathcal{H}}$ (with $B_{\mathcal{H}}$ as in (4.8)). Then*

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda_{\chi,\gamma,\mathcal{H}}(WL_i(\mathbf{n}) + b_i) = \mathrm{Vol}(K) \left(\frac{W}{\varphi(W)}\right)^{rt} (1 + o_{w\to\infty}(1) + O(e^{\sqrt{w}}/\log^{1/20} N))$$

where the error terms above may depend on d, t, D, η only.

Proof. The left-hand side equals

$$(4.10) \qquad \sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i \in [t], j \in [r]} \Lambda_{\chi,\gamma}(WL_i(\mathbf{n}) + b_i + h_j).$$

We then apply [11, Theorem D.3] to the system $\mathcal{L} = (WL_i + b_i + h_j)_{i \in [t], j \in [r]}$, whereby we assume that γ is small enough in terms of d, t. Recalling that $\int_0^2 |\chi'|^2 = 1$, [11, Theorem D.3] gives for (4.10) an estimate

$$(4.11) \qquad \mathrm{Vol}(K) \prod_p \beta_p(\mathcal{L}) + O(N^d e^X / (\log R)^{1/20}),$$

where $X = \sum_{p \in \mathcal{P}} p^{-1/2}$, and $\mathcal{P}$ is the set of those primes p such that modulo p some two of our linear forms are proportional.

Note that the primes in $\mathcal{P}$ are either $\leq w$, or $O_{d,t}(D)$. Assuming w is large enough in terms of d, t, D, we can assume that they are all $\leq w$, and therefore $X \ll \sqrt{w}/\log w$, so the error term in (4.11) becomes $O_{d,t,D,\eta}(\mathrm{Vol}(K) e^{\sqrt{w}}/\log^{1/20} N))$. Moreover, we have $\beta_p = \beta_p(\mathcal{L}) = (p/(p-1))^{rt}$ for $p \leq w$ and $\beta_p = 1 + O_{d,t,D}(p^{-2})$ as $p > w > D$ tends to infinity thanks to Lemma 2.2, whence $\prod_p \beta_p = (W/\varphi(W))^{rt}(1 + o_{w\to\infty}(1))$. This concludes the proof. $\square$

Now we are ready to prove Proposition 4.2.

Proof of Proposition 4.2. In this proof, for any $n \in \mathbb{Z}_N$ or any $n \in \mathbb{Z}$, we will denote by $\tilde{n}$ the unique element of $[N]$ such that $\tilde{n} \equiv n \,(\mathrm{mod}\, N)$.

For $b \in \mathbb{N}$ we define $\nu_b : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geq 0}$ to be the function $n \mapsto \left(\frac{\varphi(W)}{W}\right)^r \Lambda_{\chi,\gamma,\mathcal{H}}(W\tilde{n} + b)$ where $\gamma \in (0, \rho/2)$. This definition naturally gives rise to a function denoted again by ν_b on $\mathbb{Z}$. Further we set $\nu_{\mathbf{b}}(n) := \frac{1}{s} \sum_{i=1}^{s} \nu_{b_i}(n)$. Note that whenever $N^{2\gamma/\rho} \leq n \leq N$ satisfies $\theta(n) > 0$, we have $\theta(n) \leq \log^r(n+2) \ll \log^r R = \Lambda_{\chi,\gamma}(n)$. Hence, whenever $\mathbf{b} = (b_1, \ldots, b_s)$ is in $B_{\mathcal{H}}^s$, we have $\theta_{W,b_i}(n) \ll_\gamma \Lambda_{\chi,\gamma,\mathcal{H}}(Wn + b_i) \ll \nu_{\mathbf{b}}(n)$ for each $i \in [s]$ and $n \in [N^{\gamma/\rho}, N]$.

Let us verify that $\nu_{\mathbf{b}}$ is a (M, ε)-pseudorandom measure. Hence, let $d \leq M$ and $t \leq M$ and $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a finite complexity system of affine-linear forms whose linear coefficients are bounded by M. We work in the regime where w and N tend to infinity while $w \leq \log \log N$. It suffices to verify that in this regime

$$(4.12) \qquad \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/N\mathbb{Z})^d} \prod_{i \in [t]} \nu_{c_i}(\psi_i(\mathbf{n})) = 1 + o_M(1)$$

for any fixed $\mathbf{c} \in \{b_1, \ldots, b_s\}^t$. We cannot apply Proposition 4.3 to prove equation (4.12) at this stage, since this equation effectively concerns a linear system over $\mathbb{Z}/N\mathbb{Z}$ and not over $\mathbb{Z}$. In other words, there are wrap-around issues. To be able to apply Proposition 4.3, we first rewrite the left-hand side of (4.12) as

$$(4.13) \qquad \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/N\mathbb{Z})^d} \prod_{i \in [t]} \nu_{c_i}(\psi_i(\mathbf{n})) = \mathbb{E}_{\mathbf{n} \in [N]^d} \prod_{i \in [t]} \nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}).$$

Observe that the map $\mathbf{n} \mapsto \widetilde{\psi_i(\mathbf{n})}$ is not an affine-linear map, so that we still cannot invoke Proposition 4.3. However, it is piecewise affine-linear. To exploit this property, we decompose $[N]^d$ in boxes of the form

$$B_{\mathbf{u}} = \left\{ \mathbf{x} \in [N]^d \ : \ x_j \in \left(\left\lfloor \frac{(u_j - 1)N}{Q} \right\rfloor, \left\lfloor \frac{u_j N}{Q} \right\rfloor \right], j \in [d] \right\}$$

where $\mathbf{u}$ ranges over $[Q]^d$, and Q is some function of N, to be determined later, that tends slowly to infinity with N. Assuming that N/Q tends to infinity, we have

$$(4.14) \qquad \mathbb{E}_{\mathbf{n} \in [N]^d} \prod_{i \in [t]} \nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}) = \mathbb{E}_{\mathbf{u} \in [Q]^d} \mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} \nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}) + o(1),$$

where the $o(1)$ accounts for the fact that all boxes are not exactly of the same size; they are all of size $(N/Q + O(1))^d = (N/Q)^d(1 + o(1))$ though. Call $\mathbf{u}$ and the corresponding box $B_{\mathbf{u}}$ *nice* if for every $i \in [t]$ the number $\lceil \psi_i(\mathbf{n})/N \rceil$ is constant as $\mathbf{n}$ ranges in $B_{\mathbf{u}}$; that is, there exists $k = k_{i,\mathbf{u}} \in \mathbb{Z}$ such that $\psi_i(\mathbf{n}) \in (kN, (k+1)N]$ for every $\mathbf{n} \in B_{\mathbf{u}}$. When $\mathbf{u}$ is nice, $\widetilde{\psi_i(\mathbf{n})} = \psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N \in [N]$ for every $i \in [t]$ and $\mathbf{n} \in B_{\mathbf{u}}$. Therefore,

$$\mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} \nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}) = \mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} \nu_{c_i}(\psi_{i,\mathbf{u}}(\mathbf{n}))$$

where the affine-linear map $\Psi_{\mathbf{u}} : \mathbb{Z}^d \to \mathbb{Z}^t$ is defined by setting $\psi_{i,\mathbf{u}} : \mathbf{n} \mapsto \psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N$. It is clear that this map, having the same homogeneous part as Ψ, is still of finite complexity and has bounded homogeneous coefficients.

Thus, we may now apply Proposition 4.3 to conclude that

$$\mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} \nu_{c_i}(\psi_{i,\mathbf{u}}(\mathbf{n})) = 1 + o_{w \to \infty; M}(1) + o_{N \to \infty; M}(1).$$

It remains to handle the other boxes. Suppose that $\mathbf{u}$ is not nice. Thus, there exists $i \in [t]$ and two vectors $\mathbf{x}, \mathbf{y}$ in $B_{\mathbf{u}}$ such that $\lceil \psi_i(\mathbf{x})/N \rceil \neq \lceil \psi_i(\mathbf{y})/N \rceil$. However, we have $|\psi_i(\mathbf{x}) - \psi_i(\mathbf{y})| \leq 2dM(N/Q + 1) < N$, if $Q > 3dM$. Therefore, both $\psi_i(\mathbf{x})$ and $\psi_i(\mathbf{y})$ are close to the same multiple of N; thus $\psi_i(\mathbf{n}) = O_M(N/Q) \pmod{N}$ for all $\mathbf{n} \in B_{\mathbf{u}}$. Further, there exists an integer $k = k_{i,\mathbf{u}}$ such that for all $\mathbf{n} \in B_{\mathbf{u}}$ and $i \in [t]$ either $\psi_i(\mathbf{n}) - kN \in [N]$ or $\psi_i(\mathbf{n}) - (k+1)N \in [N]$; consequently,

$$\nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}) = \nu_{c_i}(\psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N)1_{\psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N \in [N]} + \nu_{c_i}(\psi_i(\mathbf{n}) - (k_{i,\mathbf{u}} + 1)N)1_{\psi_i(\mathbf{n}) - (k_{i,\mathbf{u}}+1)N \in [N]}$$
$$\leq \nu_{c_i}(\psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N) + \nu_{c_i}(\psi_i(\mathbf{n}) - (k_{i,\mathbf{u}} + 1)N).$$

Whence the bound

$$(4.15) \quad \mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} \nu_{c_i}(\widetilde{\psi_i(\mathbf{n})}) \leq \mathbb{E}_{\mathbf{n} \in B_{\mathbf{u}}} \prod_{i \in [t]} (\nu_{c_i}(\psi_i(\mathbf{n}) - k_{i,\mathbf{u}}N) + \nu_{c_i}(\psi_i(\mathbf{n}) - (k_{i,\mathbf{u}} + 1)N)).$$

Expanding the product makes the right-hand side of inequality (4.15) the sum of 2^t averages, each of which equals $1 + o(1)$ by Proposition 4.3. So the left-hand side of inequality (4.15) is $O(1)$. It remains to prove that non-nice boxes are rare. Suppose that $\mathbf{u}$ is not nice. As pointed

out above, if Q is large enough, there exists $i \in [t]$ such that $\psi_i(\mathbf{n}) = O_M(N/Q) \, (\mathrm{mod}\, N)$ for all $\mathbf{n} \in B_\mathbf{u}$. On the other hand,

$$\begin{aligned}
\psi_i(\mathbf{n}) &= \psi_i(\lfloor N\mathbf{u}/Q \rfloor) + O_M(N/Q) \\
&= \psi_i(N\mathbf{u}/Q + O(1)) + O_M(N/Q) \\
&= N\dot{\psi}_i(\mathbf{u})/Q + \psi_i(0) + O_M(N/Q).
\end{aligned}$$

Dividing by N/Q yields

$$(4.16) \qquad\qquad \dot{\psi}_i(\mathbf{u}) + Q\psi_i(0)/N = O(1) \, (\mathrm{mod}\, Q).$$

Now $\dot{\psi}_i \neq 0$ and when Q is large enough $\dot{\psi}_i \neq 0 \, (\mathrm{mod}\, Q)$ as well, so the number of solutions $\mathbf{u} \in [Q]^d$ to the estimate (4.16) is $O(Q^{d-1})$. Multiplying by t, the proportion of bad boxes among all boxes is therefore $O(Q^{d-1}) = o(1)$, the implied constant depending on M only. This concludes the proof of the estimate (4.14). □

5. Almost twin primes in generalized Bohr sets

Now we want to prove hypothesis (ii) in Theorem 3.2 for the W-tricked functions θ_1 and θ_2. Thus we need to find almost twin primes in s-measurable sets. We start with bounded gap primes. The next proposition establishes hypothesis (ii) in Theorem 3.2 for a function of the form $\theta_{W,b}$ from Section 4, upon letting $b_i = b + h_i$.

Proposition 5.1 (Bounded gap primes with nilsequences). *Fix positive integers m, d, Δ, and some $\varepsilon > 0$, $K \geq 2$. Also let $w \geq 1$ be sufficiently large in terms of $m, d, \Delta, \varepsilon, K$ and let $W = \prod_{p \leq w} p$. There exist $\rho = \rho(m) > 0$, and a positive integer $k = k(m)$, such that the following statement holds for sufficiently large $x \geq x_0(m, d, \Delta, \varepsilon, K, w)$.*

Let $\psi \in \Psi_d(\Delta, K)$ be a nilsequence taking values in $[0, 1]$. Let $b_1, \cdots, b_k$ be distinct integers satisfying $(b_i, W) = 1$ and $|b_i| \leq \log x$ for each $i \in [k]$. Then

$$\sum_{\substack{n \leq x \\ |\{Wn+b_1, \ldots, Wn+b_k\} \cap \mathbb{P}| \geq m \\ p | \prod_{i=1}^k (Wn+b_i) \implies p > x^\rho}} \psi(n) \gg_m \left(\prod_p \beta_p \right) \frac{1}{(\log x)^k} \left(\sum_{n \leq x} \psi(n) - \varepsilon x \right),$$

where $\beta_p = \beta_p(\mathcal{L})$ is the local factor defined as in Definition 2.1 for the system of affine-linear forms $\mathcal{L} = \{L_1, \ldots, L_k\}$ with $L_i(n) = Wn + b_i$.

We remark that if $p \leq w$ then $\beta_p = (p/\varphi(p))^k$, and if $p > w$ then, writing a_p for the number of distinct residue classes among $b_1, \ldots, b_k \, (\mathrm{mod}\, p)$,

$$(5.1) \qquad \beta_p = \left(\frac{p}{\varphi(p)} \right)^k \left(1 - \frac{a_p}{p} \right) = \left(\frac{p}{\varphi(p)} \right)^{k-a_p} \left(1 + O_k(p^{-2}) \right).$$

In particular, if $w > |b_i - b_j|$ for all i, j, we infer that $\prod_p \beta_p \gg (W/\phi(W))^k$.

We now turn to the corresponding statement for Chen primes.

Proposition 5.2 (Chen primes with nilsequences). *Fix positive integers d, Δ and some $\varepsilon > 0$, $K \geq 2$. Also let $w \geq 1$ be sufficiently large in terms of $d, \Delta, \varepsilon, K$ and $W = \prod_{p \leq w} p$. The following statement holds for sufficiently large $x \geq x_0(d, \Delta, \varepsilon, K, w)$.*

Let $\psi \in \Psi_d(\Delta, K)$ be a nilsequence taking values in $[0,1]$. Then for some absolute constant $\delta_0 > 0$ and any $1 \le b \le W$ with $(b, W) = (b + 2, W) = 1$ we have

$$\sum_{\substack{n \le x \\ Wn + b \in \mathbb{P} \\ Wn + b + 2 \in P_2 \\ p \mid Wn + b + 2 \implies p \ge x^{1/10}}} \psi(n) \ge \left(\frac{W}{\varphi(W)}\right)^2 \frac{\delta_0}{(\log x)^2} \left(\sum_{n \le x} \psi(n) - \varepsilon x\right).$$

We will prove Propositions 5.1 and 5.2 by reducing to the case when the underlying polynomial sequence is equidistributed, Propositions 5.3 and 5.4, using a factorization theorem for nilsequences.

Proposition 5.3 (Bounded gap primes weighted by equidistributed nilsequences). *Fix positive integers m, d, Δ, and some $\varepsilon > 0$, $A \ge 2$. There exist $\rho = \rho(m) > 0$, a positive integer $k = k(m)$, and $C = C(m, d, \Delta) > 0$, such that the following statement holds for sufficiently large $x \ge x_0(m, d, \Delta, \varepsilon, A)$.*

Let $K \ge 2$ and $\eta \in (0, 1/2)$ be parameters satisfying the conditions

$$\eta \le K^{-C}(\log x)^{-CA}, \quad K \le (\log x)^C.$$

Let $\psi \in \Psi_d(\Delta, K; \eta, x)$ be a nilsequence taking values in $[0,1]$. Let $\mathcal{L} = \{L_1, \ldots, L_k\}$ be an admissible k-tuple of linear functions with $L_i(n) = a_i n + b_i$ and $1 \le a_i \le (\log x)^A$, $|b_i| \le x$. Then

$$(5.2) \qquad \sum_{\substack{n \le x \\ |\{L_1(n), \ldots, L_k(n)\} \cap \mathbb{P}| \ge m \\ p \mid \prod_{i=1}^{k} L_i(n) \implies p > x^\rho}} \psi(n) \gg_m \frac{\mathfrak{S}(\mathcal{L})}{(\log x)^k} \left(\sum_{n \le x} \psi(n) - \varepsilon x\right),$$

where the singular series is given by $\mathfrak{S}(\mathcal{L}) := \prod_p \beta_p(\mathcal{L})$, i.e.

$$(5.3) \qquad \mathfrak{S}(\mathcal{L}) = \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{|\{n \in \mathbb{Z}/p\mathbb{Z} : L_1(n) \cdots L_k(n) \equiv 0 \,(\mathrm{mod}\, p)\}|}{p}\right) > 0.$$

Proposition 5.4 (Chen primes weighted by equidistributed nilsequences). *Fix positive integers d, Δ and some $\varepsilon > 0$, $A \ge 2$. There exists $C = C(d, \Delta) > 0$, such that the following statement holds for sufficiently large $x \ge x_0(d, \Delta, \varepsilon, A)$.*

Let $K \ge 2$ and $\eta \in (0, 1/2)$ be parameters satisfying the conditions

$$\eta \le K^{-C}(\log x)^{-CA}, \quad K \le (\log x)^C.$$

Let $\psi \in \Psi_d(\Delta, K; \eta, x)$ be a nilsequence taking values in $[0,1]$. Let $\mathcal{L} = \{L_1, L_2\}$ be an admissible set of two linear functions with $L_1(n) = an + b$ and $L_2(n) = an + b + 2$, where $1 \le a \le \log x$, $|b| \le x$. Then for some absolute constant $\delta_0 > 0$ we have

$$\sum_{\substack{n \le x \\ L_1(n) \in \mathbb{P} \\ L_2(n) \in P_2 \\ p \mid L_2(n) \implies p \ge x^{1/10}}} \psi(n) \ge \delta_0 \frac{\mathfrak{S}(\mathcal{L})}{(\log x)^2} \left(\sum_{n \le x} \psi(n) - \varepsilon x\right),$$

where the singular series is given by (5.3).

The purpose of this section is to deduce Propositions 5.1 and 5.2 from the equidistributed case, Propositions 5.3 and 5.4. We will collect some sieve lemmas in Section 6 and some analytic inputs of Bombieri–Vinogradov type in Section 7 before proving Propositions 5.3 and 5.4 in Section 8.

5.1. **Dealing with the periodic case.** In the deduction process, we need to deal with a local (modulo q) version of Propositions 5.1 and 5.2, where the requirement that all of the $Wn + b_i$'s are almost primes is replaced by the local conditions that $(Wn + b_i, q) = 1$.

Lemma 5.5. *Fix positive integers k, d, Δ, and some $\varepsilon > 0$, $K \geq 2$. Also let $w \geq 1$ be sufficiently large in terms of $k, d, \Delta, \varepsilon, K$ and $W = \prod_{p \leq w} p$. Then the following statement holds for sufficiently large $x \geq x_0(k, d, \Delta, \varepsilon, K, w)$.*

Let $\psi \in \Psi_d(\Delta, K)$. Let $\{b_1, \ldots, b_k\}$ satisfy $(b_i, W) = 1$ for every $i \in [k]$. Let $q \leq x^{0.9}$ be a positive integer with $(q, W) = 1$. Then

$$\left| \beta^{-1} \left(\frac{q}{\varphi(q)} \right)^k \sum_{\substack{n \leq x \\ (\prod_{i=1}^k (Wn+b_i), q)=1}} \psi(n) - \sum_{n \leq x} \psi(n) \right| \leq \varepsilon x,$$

where $\beta = \prod_{p|q} \beta_p$, and β_p is defined as in Proposition 5.1.

Proof. Let X be the set of $n \leq x$ such that $(Wn + b_i, q) = 1$ for each $1 \leq i \leq k$. Consider the function

$$f(n) = \beta^{-1} \left(\frac{q}{\varphi(q)} \right)^k 1_X(n) - 1.$$

Let us prove first that

$$\|f\|_{U^{d+1}[x]} = o_{x \to \infty; k, d}(1) + o_{w \to \infty; k, d}(1).$$

Expanding out $\|f\|_{U^{d+1}[x]}$ and letting $g : \mathbb{Z}^{d+2} \to \mathbb{Z}^{\{0,1\}^{d+1}}$ be the Gowers norm system $(x, \mathbf{h}) \mapsto (x + \boldsymbol{\omega} \cdot \mathbf{h})_{\boldsymbol{\omega} \in \{0,1\}^{d+1}}$, we are left with the task of proving that

$$(5.4) \qquad \sum_{\mathbf{n} \in K \cap \mathbb{Z}^{d+2}} \prod_{\boldsymbol{\omega} \in \{0,1\}^{d+1}} f(g_{\boldsymbol{\omega}}(\mathbf{n})) = o_{x \to \infty; k, d}(x^{d+2}) + o_{w \to \infty; k, d}(x^{d+2}),$$

where $K = \{\mathbf{y} \in \mathbb{R}^{d+2} : g_{\boldsymbol{\omega}}(\mathbf{y}) \in [1, x], \ \forall \boldsymbol{\omega} \in \{0,1\}^{d+1}\}$. Expanding further, the left-hand side of (5.4) equals

$$(5.5) \qquad \sum_{\Omega \subset \{0,1\}^{d+1}} (-1)^{|\Omega|} \sum_{\mathbf{n} \in K \cap \mathbb{Z}^{d+2}} \prod_{\boldsymbol{\omega} \in \Omega} \beta^{-1} \left(\frac{q}{\varphi(q)} \right)^k 1_X(g_{\boldsymbol{\omega}}(\mathbf{n})) = \sum_{\Omega \subset \{0,1\}^{d+1}} (-1)^{|\Omega|} S_\Omega,$$

where, after a change of variables, we have

$$(5.6) \qquad S_\Omega = \sum_{\mathbf{a} \in [q]^{d+2}} \sum_{\substack{\mathbf{n} \in \mathbb{Z}^{d+2} \\ q\mathbf{n}+\mathbf{a} \in K}} \prod_{\boldsymbol{\omega} \in \Omega} \beta^{-1} \left(\frac{q}{\varphi(q)} \right)^k 1_X(g_{\boldsymbol{\omega}}(q\mathbf{n} + \mathbf{a})).$$

Now the summand of the inner sum actually does not depend on $\mathbf{n}$ since $1_X(g_{\boldsymbol{\omega}}(q\mathbf{n} + \mathbf{a})) = 1_X(g_{\boldsymbol{\omega}}(\mathbf{a}))$. Let $K_{\mathbf{a}} = \{\mathbf{n} \in \mathbb{R}^{d+2} : q\mathbf{n} + \mathbf{a} \in K\}$, which is a convex body of volume $(1 + o(1))q^{-(d+2)}\text{Vol}(K)$. Since $K_{\mathbf{a}} \subset [1, x/q]^{d+2}$ and $\text{Vol}(K) \gg_d x^{d+2}$, the number of integral points $\mathbf{n} \in K_{\mathbf{a}} \cap \mathbb{Z}^{d+2}$ is

$$\text{Vol}(K_{\mathbf{a}}) + O_d \left((x/q)^{d+1} \right) = q^{-(d+2)}\text{Vol}(K) \left(1 + O_d(q/x) \right).$$

It follows that

$$S_\Omega = (1 + O_d(q/x)) \text{Vol}(K) \cdot \beta^{-|\Omega|} \mathbb{E}_{\mathbf{a} \in [q]^{d+2}} \prod_{\boldsymbol{\omega} \in \Omega} \left(\frac{q}{\varphi(q)} \right)^k 1_X(g_{\boldsymbol{\omega}}(\mathbf{a})).$$

By multiplicativity and the definition of the singular series (Definition 2.1), the average over $\mathbf{a}$ above can be written as

$$\prod_{p|q}\left(\mathbb{E}_{\mathbf{a}\in(\mathbb{Z}/p\mathbb{Z})^{d+2}}\prod_{\omega\in\Omega}\prod_{i=1}^{k}\frac{p}{\varphi(p)}1_{(Wg_\omega(\mathbf{a})+b_i,p)=1}\right)=\prod_{p|q}\beta_p(\mathcal{G}_\Omega),$$

where $\mathcal{G}_\Omega$ is the system of affine-linear forms $\mathcal{G}_\Omega$ consisting of $\mathbf{a}\mapsto Wg_\omega(\mathbf{a})+b_i$ for $\omega\in\Omega$ and $1\le i\le k$. If there are a_p distinct residue classes among $b_1,\cdots,b_k\,(\mathrm{mod}\,p)$, then $\mathcal{G}_\Omega$ consists of $a_p|\Omega|$ distinct affine-linear forms modulo p, no two of which are linearly dependent (over $\mathbb{F}_p$). Hence, Lemma 2.2 implies that

$$\beta_p(\mathcal{G}_\Omega)=\left(\frac{p}{\varphi(p)}\right)^{(k-a_p)|\Omega|}\left(1+O_{k,d}(p^{-2})\right).$$

Since $p\mid q\Rightarrow p>w$, we have $\prod_{p|q}(1+O_{k,d}(p^{-2}))=1+O_{k,d}(w^{-1})$. Putting things together, we have

$$S_\Omega=\left(1+O_d(q/x)+O_{k,d}(w^{-1})\right)\mathrm{Vol}(K)\cdot\beta^{-|\Omega|}\prod_{p|q}\left(\frac{p}{\varphi(p)}\right)^{(k-a_p)|\Omega|}.$$

From (5.1) we deduce that

$$S_\Omega=\left(1+O_{k,d}(q/x+w^{-1})\right)\mathrm{Vol}(K)$$

for each $\Omega\subset\{0,1\}^{d+1}$, and this establishes (5.4).

By [11, Proposition 11.2] (see also footnote 3), the nilsequence ψ can be decomposed as

$$\psi=\psi_1+\psi_2,$$

where $\|\psi_1\|_{U^{d+1}[x]^*}=O_{d,\Delta,K,\varepsilon}(1)$ and $\|\psi_2\|_\infty\le\varepsilon/4$. Hence,

$$\left|\sum_{n\le x}f(n)\psi_1(n)\right|\le x\|f\|_{U^{d+1}[x]}\cdot\|\psi_1\|_{U^{d+1}[x]^*}\le\tfrac{\varepsilon}{2}x,$$

provided that w and x are large enough in terms of k,d,Δ,K,ε, and

$$\left|\sum_{n\le x}f(n)\psi_2(n)\right|\le\|\psi_2\|_\infty\sum_{n\le x}|f(n)|\le\tfrac{\varepsilon}{2}x.$$

Combining the two inequalities above gives

$$\left|\sum_{n\le x}f(n)\psi(n)\right|\le\varepsilon x,$$

as desired.   □

5.2. **Reducing to the equidistributed case.** We now complete the proof of Propositions 5.1 and 5.2 assuming Propositions 5.3 and 5.4. Let $f:\mathbb{N}\to\mathbb{C}$ be any function satisfying $|f(n)|\le 1$ for any $n\in\mathbb{N}$; we will specialize later to the case where f is the indicator functions of the sets over which the summations in these propositions run. Let $\psi:\mathbb{N}\to[0,1]$ be a nilsequence in $\Psi_d(\Delta,K)$. We want to estimate $\sum_{n\le x}f(n)\psi(n)$. By Definition 2.8, there exists a nilmanifold G/Γ of dimension at most Δ, equipped with a filtration $G_\bullet$ of degree $\le d$ and a K-rational Mal'cev basis $\mathcal{X}$, a polynomial sequence $g:\mathbb{Z}\to G$ adapted to $G_\bullet$ and a Lipschitz function $F:G/\Gamma\to\mathbb{C}$ satisfying $\|F\|_{\mathrm{Lip}(\mathcal{X})}\le 1$, such that $\psi(n)=F(g(n)\Gamma)$.

Let

$$\mu:=\frac{1}{x}\sum_{n\le x}\psi(n).$$

We may assume that $\mu\ge\varepsilon$, as otherwise there is nothing to prove.

Let $B = B(m, d, \Delta) > 0$ be sufficiently large. To reduce Propositions 5.1 and 5.2 to the case when g is equidistributed, we apply the factorization theorem [13, Theorem 1.19] to obtain some parameter $M \in [\log x, (\log x)^{O_{B,d,\Delta}(1)}]$ and a decomposition $g = \epsilon g' \gamma$ into polynomial sequences $\epsilon, g', \gamma : \mathbb{Z} \to G$ with the following properties:

(1) ϵ is (M, x)-smooth, i.e. $d(\epsilon(n), \mathrm{id}_G) \leq M$ and $d(\epsilon(n), \epsilon(n-1)) \leq M/x$ for all $n \in [x]$, with $d = d_\mathcal{X}$ the metric used on G.

(2) g' takes values in a rational subgroup $G' \subseteq G$, equipped with a Mal'cev basis $\mathcal{X}'$ in which each element is an M-rational combination of the elements of $\mathcal{X}$, and moreover $\{g'(n)\}_{n \leq x}$ is totally M^{-B} equidistributed in $G'/\Gamma \cap G'$.

(3) γ is M-rational (so that $\gamma(n)\Gamma$ is an M-rational point for every $n \in \mathbb{Z}$), and moreover $\{\gamma(n)\Gamma\}_{n \in \mathbb{Z}}$ is periodic with period some $q \leq M$.

In the case of Proposition 5.1, we may make the following additional assumption on q by enlarging q and M if necessary: If $b_i \equiv b_j \pmod{p}$ for some $i \neq j$ and some prime p, then p divides q. By (5.1), this implies that if $p \nmid qW$, then $\beta_p = 1 + O_k(p^{-2})$.

Let $\mathcal{Q}$ be a collection of arithmetic progressions of step q and length $\asymp (x/qM)(\log x)^{-100}$ such that $[x] = \bigcup_{P \in \mathcal{Q}} P$. For each $P \in \mathcal{Q}$, let γ_P be the (constant) value of γ on P and consider

$$(5.7) \qquad \sum_{n \in P} f(n)\psi(n) = \sum_{n \in P} f(n)F(\epsilon(n)g'(n)\gamma_P\Gamma).$$

We shall first dispose of the smooth part $\epsilon(n)$. Pick an arbitrary $n_P \in P$, and let $\epsilon_P = \epsilon(n_P)$. If $n \in P$, then $|n - n_P| \ll (x/M)(\log x)^{-100}$. Since the Lipschitz norm of F is bounded by 1, we have

$$\begin{aligned} |F(\epsilon(n)g'(n)\gamma_P\Gamma) - F(\epsilon_P g'(n)\gamma_P\Gamma)| &\leq d_\mathcal{X}(\epsilon(n)g'(n)\gamma_P, \epsilon(n_P)g'(n)\gamma_P) \\ &= d_\mathcal{X}(\epsilon(n), \epsilon(n_P)) \\ &\leq \frac{M}{x}|n - n_P| \ll (\log x)^{-100}. \end{aligned}$$

Hence (5.7) equals

$$\sum_{n \in P} f(n)F(\epsilon_P g'(n)\gamma_P\Gamma) + O(|P|(\log x)^{-100}).$$

Let H_P be the conjugate $H_P = \gamma_P^{-1}G'\gamma_P$, let $\Gamma_P = H_P \cap \Gamma$, let $F_P : H_P \to [0,1]$ be the Γ_P-automorphic function defined by $F_P(x) = F(\epsilon_P\gamma_P x)$ (the conjugation is necessary to make F_P automorphic), and let $g_P : \mathbb{Z} \to H_P$ be the polynomial sequence defined by $g_P(n) = \gamma_P^{-1}g'(n)\gamma_P$. Thus

$$F(\epsilon_P g'(n)\gamma_P\Gamma) = F_P(g_P(n)\Gamma_P).$$

Some routine arguments (see the Claim at the end of Section 2 in [12]) produce the following properties:

(1) The sub-nilmanifold H_P/Γ_P is equipped with a Mal'cev basis $\mathcal{X}_P$ in which each element is an $M^{O_{d,\Delta}(1)}$-rational combination of the elements of $\mathcal{X}$.

(2) $\{g_P(n)\}_{n \leq x}$ is totally M^{-cB}-equidistributed for some constant $c = c(d, \Delta) > 0$. By choosing B large enough we may ensure that $cB \geq C$, the constant from Proposition 5.3 or Proposition 5.4.

(3) $\|F_P\|_{\mathrm{Lip}} \leq M^{O_{d,\Delta}(1)}$.

Let $y = |P| \gg (x/qM)(\log x)^{-100} \geq x^{1/2}$, and write $P = \{qn + t : n \leq y\}$ for some $t \in \mathbb{Z}$. Let $g'_P : \mathbb{Z} \to H_P$ be the polynomial sequence defined by $g'_P(n) = g_P(qn + t)$, so that $\{g'_P(n)\}_{n \leq y}$

is still totally M^{-cB}-equidistributed (after possibly reducing the constant c). Then

$$(5.8) \qquad \sum_{n \in P} f(n) F(\epsilon_P g'(n) \gamma_P \Gamma) = \sum_{n \leq y} f(qn + t) F_P(g'_P(n) \Gamma_P).$$

Case of Proposition 5.1. Now we specialize to the function f relevant for Proposition 5.1. Write $L_i(n) = Wn + b_i$, and let $\mathcal{L}' = \mathcal{L}'_P = \{L'_1, \cdots, L'_k\}$, where L'_i is the linear function defined by $L'_i(n) = L_i(qn + t)$. Let f be the indicator function of the set of integers n such that $\#(\{L_1(n), \ldots, L_k(n)\} \cap \mathbb{P}) \geq m$ and $p \mid L_1(n) \cdots L_k(n) \implies p \geq x^\rho$. Thus

$$\sum_{n \leq y} f(qn + t) F_P(g'_P(n) \Gamma_P) = \sum_{\substack{n \leq y \\ \#(\{L'_1(n), \ldots, L'_k(n)\} \cap \mathbb{P}) \geq m \\ p \mid L'_1(n) \cdots L'_k(n) \implies p \geq x^\rho}} F_P(g'_P(n) \Gamma_P).$$

If $\mathcal{L}'$ remains admissible, then by Proposition 5.3 the right-hand side above is

$$\gg_k \frac{\mathfrak{S}(\mathcal{L}')}{(\log x)^k} \left(\sum_{n \leq y} F_P(g'_P(n) \Gamma_P) - \varepsilon/4 \cdot y \right) = \frac{\mathfrak{S}(\mathcal{L}')}{(\log x)^k} \left(\sum_{n \in P} F(\epsilon_P g'(n) \gamma_P \Gamma) - \varepsilon/4 \cdot |P| \right)$$

$$\geq \frac{\mathfrak{S}(\mathcal{L}')}{(\log x)^k} \left(\sum_{n \in P} F(\epsilon(n) g'(n) \gamma_P \Gamma) - \varepsilon/2 \cdot |P| \right),$$

where the last inequality follows once again from the smoothness of ϵ. By the definition of $\mathfrak{S}(\mathcal{L}')$, we see that $\mathcal{L}'$ is admissible if any only if $(Wt + b_i, q) = 1$ for each $1 \leq i \leq k$, and in this case we have

$$\mathfrak{S}(\mathcal{L}') = \prod_{p \mid qW} \left(1 - \frac{1}{p} \right)^{-k} \prod_{p \nmid qW} \left(1 - \frac{k}{p} \right) \left(1 - \frac{1}{p} \right)^{-k} \asymp \left(\frac{qW}{\varphi(qW)} \right)^k.$$

Putting everything together, under the assumption that $\mathcal{L}'_P$ is admissible, we have proven that

$$\sum_{\substack{n \in P \\ |\{Wn+b_1, \ldots, Wn+b_k\} \cap \mathbb{P}| \geq m \\ p \mid \prod_{i=1}^k (Wn+b_i) \implies p > x^\rho}} \psi(n) \gg_m \left(\frac{qW}{\varphi(qW)} \right)^k \frac{1}{(\log x)^k} \sum_{n \in P} \left(\psi(n) - \frac{\varepsilon}{2} \right).$$

Summing this estimate over all $P \in \mathcal{Q}$, we get

$$\sum_{\substack{n \leq x \\ |\{Wn+b_1, \ldots, Wn+b_k\} \cap \mathbb{P}| \geq m \\ p \mid \prod_{i=1}^k (Wn+b_i) \implies p > x^\rho}} \psi(n) \gg_m \left(\frac{qW}{\varphi(qW)} \right)^k \frac{1}{(\log x)^k} \sum_{\substack{n \leq x \\ (\prod_{i=1}^k (Wn+b_i), q)=1}} \left(\psi(n) - \frac{\varepsilon}{2} \right)$$

$$= \left(\frac{W}{\varphi(W)(\log x)} \cdot \frac{\frac{q}{(q,W)}}{\varphi(\frac{q}{(q,W)})} \right)^k \sum_{\substack{n \leq x \\ (\prod_{i=1}^k (Wn+b_i), \frac{q}{(q,W)})=1}} \left(\psi(n) - \frac{\varepsilon}{2} \right).$$

Finally, applying Lemma 5.5 to the summation on the right-hand side, with q replaced by $q/(q, W))$ and ψ replaced by $\psi - \varepsilon/2$, we get that the right-hand side above is at least

$$\left(\frac{W}{\varphi(W)(\log x)} \right)^k \left(\prod_{p \mid q/(q,W)=1} \beta_p \right) \left(\sum_{n \leq x} \psi(n) - \varepsilon x \right) = \left(\prod_{p \mid qW} \beta_p \right) \frac{1}{(\log x)^k} \left(\sum_{n \leq x} \psi(n) - \varepsilon x \right).$$

The conclusion of Proposition 5.1 follows, since $\beta_p = 1 + O_k(p^{-2})$ for $p \nmid qW$ by our assumption on q.

Case of Proposition 5.2. Finally, we address Proposition 5.2. Thus we return to equation (5.8) and now specialize to the case where f is the indicator function of the set of integers n such that $L_1(n) \in \mathbb{P}$ and $L_2(n) \in P_2$ and $p \mid L_2(n) \implies p \geq x^{1/10}$ where $L_1(n) = Wn + b$ and $L_2(n) = Wn + b + 2$. Let $\mathcal{L}' = \mathcal{L}'_P = \{L'_1, L'_2\}$ where for $i \in \{1, 2\}$, the linear function L'_i is defined by $L'_i(n) = L_i(qn + t)$. Then we have

$$\sum_{n \leq y} f(qn + t) F_P(g'_P(n)\Gamma_P) = \sum_{\substack{n \leq y \\ L'_1(n) \in \mathbb{P} \\ L'_2(n) \in P_2 \\ p \mid L'_2(n) \implies p \geq x^{1/10}}} F_P(g'_P(n)\Gamma_P).$$

If $\mathcal{L}'$ remains admissible (which happens precisely when $(Wt + b, q) = (Wt + b + 2, q) = 1$), then Proposition 5.4 and the same argument as above prove that the right-hand side above is at least

$$\delta_0 \left(\frac{qW}{\varphi(qW)} \right)^2 \frac{1}{(\log x)^2} \sum_{n \in P} (\psi(n) - \varepsilon/2).$$

This time the summation over all $P \in \mathcal{Q}$ yields, after applying Lemma 5.5,

$$\sum_{\substack{n \leq x \\ L_1(n) \in \mathbb{P} \\ L_2(n) \in P_2 \\ p \mid L_2(n) \implies p \geq x^{1/10}}} \psi(n) \geq \delta_0 \left(\prod_{p \mid q/(q,W)} \beta_p \right) \left(\frac{W}{\varphi(W)} \right)^2 \frac{1}{(\log x)^2} \left(\sum_{n \leq x} \psi(n) - \varepsilon x \right).$$

The conclusion of Proposition 5.2 follows, since $\beta_p = 1 + O_k(p^{-2})$ for $p > w$ by (5.1).

6. Three sieve lemmas

In the proof of Theorems 1.1 and 1.2, we will need weighted versions of Maynard's sieve for bounded gap primes, Chen's sieve for almost twin primes, and Iwaniec's sieve for primes of the form $x^2 + y^2 + 1$.

Proposition 6.1 (Maynard's sieve). *For any $\theta \in (0, 1)$, $k \in \mathbb{N}$, there exist constants $C = C(\theta)$, $\rho = \rho(\theta, k)$ such that the following holds.*

Let $(\omega_n)_{n \leq x}$ be any nonnegative sequence, and let $\mathcal{L} = (L_1, \ldots, L_k)$ be an admissible k-tuple of linear functions with $L_i(n) = a_i n + b_i$ and $1 \leq a_i, b_i \leq x$. Suppose that (ω_n) obeys the following hypotheses.

(i) *(Prime number theorem) For each $1 \leq i \leq k$ and some $\delta > 0$, we have*

$$\frac{\varphi(a_i)}{a_i} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \omega_n \geq \frac{\delta}{\log x} \sum_{n \leq x} \omega_n.$$

(ii) *(Well-distribution in arithmetic progressions) For some $C_0 > 0$ we have*

$$\sum_{r \leq x^\theta} \max_{c \, (\mathrm{mod} \, r)} \left| \sum_{\substack{n \leq x \\ n \equiv c \, (\mathrm{mod} \, r)}} \omega_n - \frac{1}{r} \sum_{n \leq x} \omega_n \right| \leq C_0 \frac{\sum_{n \leq x} \omega_n}{(\log x)^{101k^2}}.$$

(iii) *(Bombieri–Vinogradov) For each $1 \leq i \leq k$ we have*

$$\sum_{r \leq x^\theta} \max_{(L_i(c), r) = 1} \left| \sum_{\substack{n \leq x \\ n \equiv c \, (\mathrm{mod} \, r) \\ L_i(n) \in \mathbb{P}}} \omega_n - \frac{\varphi(a_i)}{\varphi(a_i r)} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \omega_n \right| \leq C_0 \frac{\sum_{n \leq x} \omega_n}{(\log x)^{101k^2}}.$$

(iv) *(Brun–Titchmarsh)* We have

$$\max_{c\,(\mathrm{mod}\,r)} \sum_{\substack{n \leq x \\ n \equiv c\,(\mathrm{mod}\,r)}} \omega_n \leq \frac{C_0}{r} \sum_{n \leq x} \omega_n,$$

uniformly for $r \leq x^\theta$.

Then, for $x \geq x_0(\theta, k, C_0)$, we have

$$\sum_{\substack{n \leq x \\ |\{L_1(n),\dots,L_k(n)\} \cap \mathbb{P}| \geq C^{-1}\delta \log k \\ p | \prod_{i=1}^k L_i(n) \implies p > x^\rho}} \omega_n \gg_{k,\theta,\delta} \frac{\mathfrak{S}(\mathcal{L})}{(\log x)^k} \sum_{n \leq x} \omega_n,$$

where the singular series $\mathfrak{S}(\mathcal{L})$ is given by (5.3).

Proof. This is [22, Theorem 6.2] (with $\alpha = 1$ there), which adds weights to the corresponding statement in [23]. $\qquad\square$

Proposition 6.2 (Chen's sieve). *Let $\varepsilon > 0$ be a small enough absolute constant. Let $(\omega_n)_{n \leq x}$ be any nonnegative sequence, and let $\mathcal{L} = \{L_1, L_2\}$ with $L_i(n) = a_i n + b_i$, $1 \leq a_i \leq \log x$, $|b_i| \leq x$. Suppose that (ω_n) satisfies the following hypotheses:*

(i) *(Bombieri–Vinogradov with well-factorable weights) We have*

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r, a_2(a_1 b_2 - a_2 b_1))=1}} \lambda(r) \left(\sum_{\substack{n \leq x \\ L_2(n) \equiv 0\,(\mathrm{mod}\,r) \\ L_1(n) \in \mathbb{P}}} \omega_n - \frac{a_1}{\varphi(a_1 r)} \sum_{n \leq x} \frac{\omega_n}{\log L_1(n)} \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{10}}$$

for any well-factorable sequence λ of level $x^{1/2-\varepsilon}$, and also for $\lambda = 1_{p \in [P,P')} * \lambda'$ with λ' any well-factorable sequence of level $x^{1/2-\varepsilon}/P$ and $2P \geq P' \geq P \in [x^{1/10}, x^{1/3-\varepsilon}]$.

(ii) *(Bombieri–Vinogradov for almost primes with well-factorable weights) For $j \in \{1, 2\}$ we have*

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r, a_1(a_1 b_2 - a_2 b_1))=1}} \lambda(r) \left(\sum_{\substack{n \leq x \\ L_1(n) \equiv 0\,(\mathrm{mod}\,r) \\ L_2(n) \in B_j}} \omega_n - \frac{\varphi(a_2)}{\varphi(a_2 r)} \sum_{\substack{n \leq x \\ L_2(n) \in B_j}} \omega_n \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{10}},$$

where $\lambda(r)$ is as above and

$$B_1 = \{p_1 p_2 p_3 : \ x^{1/10} \leq p_1 \leq x^{1/3-\varepsilon}, \ x^{1/3-\varepsilon} \leq p_2 \leq (2x/p_1)^{1/2}, \ p_3 \geq x^{1/10}\},$$

$$B_2 = \{p_1 p_2 p_3 : \ x^{1/3-\varepsilon} \leq p_1 \leq p_2 \leq (2x/p_1)^{1/2}, \ p_3 \geq x^{1/10}\}.$$

(iii) *(Upper bound on almost primes): For $j \in \{1, 2\}$ we have*

$$\sum_{\substack{n \leq x \\ L_2(n) \in B_j}} \omega_n \leq (1 + \varepsilon) \cdot \frac{|B_j \cap [1, L_2(x)]|}{\varphi(a_2) x} \sum_{n \leq x} \omega_n.$$

Then, for $x \geq x_0$, we have

$$\sum_{\substack{n \leq x \\ L_1(n) \in \mathbb{P} \\ L_2(n) \in P_2 \\ p | L_2(n) \implies p \geq x^{1/10}}} \omega_n \geq \delta_0 \frac{\mathfrak{S}(\mathcal{L})}{(\log x)^2} \sum_{n \leq x} \omega_n - O(x^{0.9} \max_n \omega_n),$$

for some absolute constant $\delta_0 > 0$, where the singular series $\mathfrak{S}(L)$ is given by (5.3).

Proof. This is [22, Theorem 6.4] (which adds weights to Chen's sieve), with the slight modification that $|b_i|$ may be as large as x (as opposed to $x^{o(1)}$). However, this restriction on $|b_i|$ was not used in the proof. Also, in [22, Theorem 6.4] $\lambda(r)$ was replaced with $\mu(r)^2\lambda(r)$, but since in the proof the sequence $\lambda(r)$ is always a sieve coefficient supported on squarefree numbers, this makes no difference. $\square$

For stating the weighted sieve for primes of the form $x^2 + y^2 + 1$, we need a notion slightly different from admissibility, which we call amenability, following [31, Definition 3.1].

Definition 6.3. We say that a linear function $L(n) = Kn + b$ with $K \geq 1$ and $b \in \mathbb{Z}$ is *amenable* if

  (i) $6^3 \mid K$;
  (ii) $(b, K) = (b - 1, s(K)) = 1$, where $s(n) := \prod_{p|n, p \equiv -1 \,(\mathrm{mod}\, 4), p \neq 3} p$;
  (iii) $b - 1 = 2^j 3^{2t}(4h + 1)$ for some $h \in \mathbb{Z}$ with $3 \nmid 4h + 1$, and $j, t \geq 0$ with $2^{j+2}3^{2t+1} \mid K$.

Here condition (ii) guarantees that there are no local obstructions to $L(n)$ being a prime of the form $x^2 + y^2 + 1$. Conditions (i) and (iii) are introduced for technical reasons to do with sieves in [31], but they are not very restrictive.

Proposition 6.4 (Weighted sieve for primes of the form $x^2 + y^2 + 1$). *There exists some small $\varepsilon > 0$ such that the following holds. Let $(\omega_n)_{n \leq x}$ be any nonnegative sequence, and let $L(n) = Kn + b$ be amenable with $1 \leq K \leq \log x$ and $1 \leq b \leq x$. Suppose that (ω_n) obeys the following hypotheses:*

  (i) *For any sequence $(g(\ell))_\ell$ supported on $[1, x^{0.9}]$ and of the form $g = \alpha * \beta$ with α supported on $[x^{1/(3+\varepsilon)}, x^{1-1/(3+\varepsilon)}]$ and $|\alpha(n)|, |\beta(n)| \leq 1$, we have*

(6.1)
$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,K)=1}} \lambda_d^{+,\mathrm{LIN}} \sum_{\substack{\ell \leq x^{0.9} \\ (\ell,K)=\delta \\ (\ell,r)=1}} g(\ell) \left(\sum_{\substack{n \leq x \\ p \leq x \\ L(n)=\ell p+1 \\ L(n) \equiv 0 \,(\mathrm{mod}\, r)}} \omega_n - \frac{1}{\varphi(r)} \frac{K}{\varphi(\frac{K}{\delta})} \sum_{n \leq x} \frac{\omega_n}{\ell \log \frac{Kn}{\ell}} \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{100}},$$

   where $\delta := (b - 1, K)$ and $\lambda_r^{+,\mathrm{LIN}}$ are the upper bound linear sieve coefficients of level $x^{1/2-\varepsilon}$ and sifting parameter $x^{\frac{1}{5}}$.
  (ii) *We have*

(6.2)
$$\left| \sum_{\substack{r \leq x^{3/7-\varepsilon} \\ (r,K)=1}} \lambda_r^{-,\mathrm{SEM}} \left(\sum_{\substack{n \leq x \\ L(n) \in \mathbb{P} \\ L(n) \equiv 1 \,(\mathrm{mod}\, r)}} \omega_n - \frac{1}{\varphi(r)} \frac{K}{\varphi(K)} \sum_{n \leq x} \frac{\omega_n}{\log(Kn)} \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{100}},$$

   where $\lambda_r^{-,\mathrm{SEM}}$ are the lower bound semilinear sieve coefficients of level $x^{3/7-\varepsilon}$ and sifting parameter $x^{1/(3+\varepsilon)}$.

Then, for some absolute constant $\delta_0 > 0$, we have

(6.3)
$$\sum_{\substack{n \leq x \\ L(n) \in \mathbb{P} \\ p|L(n)-1 \implies p \not\equiv -1 \,(\mathrm{mod}\, 4)}} \omega_n \geq \delta_0 \frac{\mathfrak{S}(L)}{(\log x)^{3/2}} \sum_{n \leq x} \omega_n - O(x^{1/2}),$$

where the singular series $\mathfrak{S}(L)$ *is given by*

$$\mathfrak{S}(L) := \prod_{\substack{p \equiv -1 \,(\mathrm{mod}\,4) \\ p \neq 3}} \left(1 - \frac{|\{n \in \mathbb{Z}/p\mathbb{Z} : \ L(n) \equiv 0 \text{ or } 1\,(\mathrm{mod}\,p)\}|}{p}\right)\left(1 - \frac{2}{p}\right)^{-1}$$

(6.4)

$$\cdot \prod_{p \not\equiv -1 \,(\mathrm{mod}\,4)} \left(1 - \frac{|\{n \in \mathbb{Z}/p\mathbb{Z} : \ L(n) \equiv 0\,(\mathrm{mod}\,p)\}|}{p}\right)\left(1 - \frac{1}{p}\right)^{-1}.$$

Proof. This follows from [31, Theorem 6.5], taking $\rho_1 = 1/2 - \varepsilon$, $\rho_2 = 3/7 - \varepsilon$ and $\sigma = 3 + \varepsilon$ there and using the fact that hypothesis $H(\rho_1, \rho_2, \sigma)$ there holds with these parameters (the n summation in [31, Theorem 6.5] is over a dyadic interval, but this clearly makes no difference). $\square$

7. Bombieri–Vinogradov and Type I/II estimates for nilsequences

In this section, we collect Bombieri–Vinogradov type estimates for nilsequences from [28] that we shall need. Theorems 7.1 and 7.2 below are slight generalizations of [28, Theorem 4.3] and [28, Theorem 4.4], respectively.

Theorem 7.1. *Let an integer $s \geq 1$, a large real number $\Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/3)$ be given. Let $L(n) = an + b$ for some $1 \leq a \leq x^{\varepsilon/2}$ and $|b| \leq x$ with $(a, b) = 1$. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for any $x \geq 2$, $\eta > 0$ and any nilsequence $\psi \in \Psi_s^0(\Delta, \eta^{-\kappa}; \eta, x)$ we have*

$$\sum_{d \leq x^{1/3 - \varepsilon}} \max_{(L(c), d) = 1} \left| \sum_{\substack{n \leq x \\ n \equiv c\,(\mathrm{mod}\,d)}} \Lambda(L(n))\psi(n) \right| \ll \eta^\kappa ax(\log x)^2.$$

Theorem 7.2. *Let integers $s \geq 1$, $c \neq 0$, a large real number $\Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. Let $L(n) = an + b$ for some $1 \leq a \leq x^{\varepsilon/2}$ and $|b| \leq x$ with $(a, b) = 1$. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for any well-factorable sequence (λ_d) of level $x^{1/2 - \varepsilon}$ with $x \geq 2$ and any nilsequence $\psi \in \Psi_s^0(\Delta, \eta^{-\kappa}; \eta, x)$ with $\eta > 0$, we have*

$$\left| \sum_{\substack{d \leq x^{1/2 - \varepsilon} \\ (d, c) = 1}} \lambda_d \sum_{\substack{n \leq x \\ L(n) \equiv c\,(\mathrm{mod}\,d)}} \Lambda(L(n))\psi(n) \right| \ll \eta^\kappa ax(\log x)^2.$$

Proofs of Theorem 7.1 and 7.2. We deduce Theorem 7.1 from [28, Theorem 4.3]; the deduction of Theorem 7.2 from [28, Theorem 4.4] is completely similar.

Write $\psi(n) = \varphi(g(n))$. One can find a polynomial sequence g' such that $g'(L(n)) = g(n)$, for example, by examining the Taylor coefficients of g in coordinates (see [13, Lemma 6.7]).

We claim that $\{g'(n)\}_{n \leq ax}$ is totally η^c-equidistributed for some small constant $c = c(s, \Delta) > 0$. Suppose that this is not the case. Then by the quantitative Kronecker theorem for nilsequences (see [13, Theorem 2.9]), there is a nontrivial horizontal character χ with $\|\chi\| \ll \eta^{-O_{s,\Delta}(c)}$ such that

$$\|\chi \circ g'\|_{C^\infty(ax)} \ll \eta^{-O_{s,\Delta}(c)}.$$

Hence if we write

$$\chi \circ g'(n) = \alpha_0 + \alpha_1 n + \cdots + \alpha_s n^s,$$

then there is a positive integer $q \ll_s 1$ such that the coefficients satisfy

$$\|q\alpha_i\| \ll_s (ax)^{-i} \eta^{-O_{s,\Delta}(c)}$$

for each $1 \leq i \leq s$. Now

$$\chi \circ g(n) = \chi \circ g'(an + b) = \sum_{i=0}^{s} \alpha_i(an + b)^i.$$

If we write β_j for the coefficient of n^j in $\eta \circ g$, then one can establish that

$$\|q\beta_j\| \ll_s x^{-j}\eta^{-O_{s,\Delta}(c)}$$

for each $1 \leq j \leq s$. Hence

$$\|q\chi \circ g\|_{C^\infty(x)} \ll_s \eta^{-O_{s,\Delta}(c)}.$$

It now follows (from [28, Lemma 3.6]) that $\{g(n)\}_{n \leq x}$ is not totally $\eta^{-O_{s,\Delta}(c)}$-equidistributed, which is a contradiction if c is chosen small enough.

Let $\psi'(n) = \varphi(g'(n))$. Then $\psi' \in \Psi_s^0(\Delta, \eta^{-\kappa}; \eta^c, ax)$. After a change of variables, we can write

$$\sum_{\substack{n \leq x \\ n \equiv c \,(\mathrm{mod}\, d)}} \Lambda(L(n))\psi(n) = \sum_{\substack{L(0) < n \leq L(x) \\ n \equiv L(c)\,(\mathrm{mod}\, ad)}} \Lambda(n)\psi'(n).$$

It follows that

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(L(c),d)=1} \Big| \sum_{\substack{n \leq x \\ n \equiv c\,(\mathrm{mod}\, d)}} \Lambda(L(n))\psi(n)\Big| \leq \sum_{d' \leq ax^{1/3-\varepsilon}} \max_{(c',d')=1} \Big| \sum_{\substack{L(0) < n \leq L(x) \\ n \equiv c'\,(\mathrm{mod}\, d')}} \Lambda(n)\psi'(n)\Big|.$$

By [28, Theorem 4.3], the right-hand side above is $\ll \eta^{c\kappa}ax(\log x)^2$ for some constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$. The conclusion follows. $\square$

We will also need a few type I and type II estimates appearing in [28].

Lemma 7.3 (Type I Bombieri–Vinogradov estimate). *Let $x \geq 2$ and $\varepsilon > 0$. Let $1 \leq M \leq x^{1/2}$ and $1 \leq D \leq x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < 1/2$. Let $L(n) = an + b$ for some $1 \leq a \leq x^{\varepsilon/2}$ and $|b| \leq x$ with $(a,b) = 1$. Let $\psi \in \Psi_s^0(\Delta, \delta^{-1}; \delta^C, x)$ for some sufficiently large constant $C = C(s, \Delta, \varepsilon)$. Then*

$$\sum_{D \leq d \leq 2D} \max_{c\,(\mathrm{mod}\, d)} \sum_{\substack{M \leq m \leq 2M \\ (m,ad)=1}} \Big| \sum_{\substack{mn \leq L(x) \\ mn \equiv c\,(\mathrm{mod}\, d) \\ mn \equiv b\,(\mathrm{mod}\, a)}} \psi(L^{-1}(mn))\Big| \ll \delta x.$$

Proof. The case $L(n) = n$ is [28, Proposition 5.5]. We shall quickly reduce the general case to this case.

From the arguments in the proof of Theorems 7.1 and 7.2, there exists a nilsequence $\psi' \in \Psi_s^0(\Delta, \delta^{-1}; \delta^{C'}, ax)$ for some large constant $C' = C'(s, \Delta, \varepsilon)$, such that $\psi'(n) = \psi(L^{-1}(n))$ if $n \equiv b\,(\mathrm{mod}\, a)$. Then use the identity

$$1_{mn \equiv b\,(\mathrm{mod}\, a)} = \frac{1}{\varphi(a)} \sum_{\chi\,(\mathrm{mod}\, a)} \chi(m)\chi(n)\overline{\chi(b)}$$

to reduce matters to

$$\sum_{D \leq d \leq 2D} \max_{c\,(\mathrm{mod}\, d)} \sum_{\substack{M \leq m \leq 2M \\ (m,ad)=1}} \Big| \sum_{\substack{mn \leq L(x) \\ mn \equiv c\,(\mathrm{mod}\, d)}} \chi(mn)\psi'(mn)\Big| \ll \delta x$$

for characters $\chi \pmod a$. Splitting mn into residue classes $\pmod a$, it suffices to show for all u coprime to a that

$$\sum_{D \leq d \leq 2D} \max_{c \, (\mathrm{mod}\, d)} \sum_{\substack{M \leq m \leq 2M \\ (m,ad)=1}} \Big| \sum_{\substack{mn \leq L(x) \\ mn \equiv c \, (\mathrm{mod}\, d) \\ mn \equiv u \, (\mathrm{mod}\, a)}} \psi'(mn) \Big| \ll \delta x$$

Now, applying the Chinese remainder theorem to combine the congruences on mn, and making the change of variables $d' = [d, a] \leq 2aD$, the conclusion then follows from the case $L(n) = n$ that was already established. $\qquad\square$

The type II estimate we need involves well-factorable weights.

Definition 7.4. We say that a sequence $\lambda : [N] \to \mathbb{R}$ is *well-factorable* of level $D \geq 1$, if for any $R, S \geq 1$ satisfying $D = RS$, we can write $\lambda = \lambda_1 * \lambda_2$ for some sequences $|\lambda_1|, |\lambda_2| \leq 1$ supported on $[1, R]$ and $[1, S]$, respectively, with $*$ denoting Dirichlet convolution.

Lemma 7.5 (Well-factorable type II Bombieri–Vinogradov estimate)**.** *Let $\varepsilon > 0$ be a small constant. Let $x \geq 2$ and $M \in [x^{1/4}, x^{3/4}]$ be large and let $c \neq 0, k$ be fixed integers. Suppose that either*

  (i) *λ is well-factorable of level $x^{1/2-\varepsilon}$, or*
  (ii) *$\lambda = 1_{p \in [P,P')} * \lambda'$, where λ' is well-factorable of level $x^{1/2-\varepsilon}/P$ and $2P \geq P' \geq P \in [x^{1/10}, x^{1/3-\varepsilon}]$.*

Let $s \geq 1$, $\Delta \geq 2$, $0 < \delta < 1/2$. Let $L(n) = an + b$ for some $1 \leq a \leq x^{\varepsilon/2}$ and $|b| \leq x$ with $(a, b) = 1$. Let $\psi \in \Psi_s^0(\Delta, \delta^{-1}; \delta^C, x)$ for some sufficiently large constant $C = C(s, \Delta, \varepsilon)$. Then

$$\Big| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d,ac)=1}} \lambda_d \sum_{\substack{L(x) \leq mn \leq L(2x) \\ M \leq m \leq 2M \\ mn \equiv c \, (\mathrm{mod}\, d) \\ mn \equiv b \, (\mathrm{mod}\, a)}} \alpha(m)\beta(n)\psi(L^{-1}(mn)) \Big| \ll \delta a x (\log x)^{O_k(1)},$$

uniformly for sequences $\{\alpha(n)\}$ and $\{\beta(n)\}$ satisfying $|\alpha(n)|, |\beta(n)| \leq d_k(n)$.

This is a consequence of the following somewhat more general statement.

Lemma 7.6. *Let $\varepsilon > 0$ be a small constant. Let $x \geq 2$ and $M \in [x^{1/4}, x^{3/4}]$ be large and let $c \neq 0, k$ be fixed integers. Let $R_1, R_2 \geq 1$ be such that $R_1 \leq x^{1-\varepsilon}/M$, $R_1 R_2 \leq x^{1/2-\varepsilon}$ and $R_1 R_2^2 \leq M x^{-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, $0 < \delta < 1/2$. Let $L(n) = an + b$ for some $1 \leq a \leq x^{\varepsilon/2}$ and $|b| \leq x$ with $(a, b) = 1$. Let $\psi \in \Psi_s^0(\Delta, \delta^{-1}; \delta^C, x)$ for some sufficiently large constant $C = C(s, \Delta, \varepsilon)$. Then*

$$\sum_{\substack{R_1 \leq r_1 \leq 2R_1 \\ R_2 \leq r_2 \leq 2R_2 \\ (r_1 r_2, ac)=1}} \Big| \sum_{\substack{L(x) \leq mn \leq L(2x) \\ M \leq m \leq 2M \\ mn \equiv c \, (\mathrm{mod}\, r_1 r_2) \\ mn \equiv b \, (\mathrm{mod}\, a)}} \alpha(m)\beta(n)\psi(L^{-1}(mn)) \Big| \ll \delta a x (\log x)^{O_k(1)},$$

uniformly for sequences $\{\alpha(n)\}$ and $\{\beta(n)\}$ satisfying $|\alpha(n)|, |\beta(n)| \leq d_k(n)$.

To see that Lemma 7.5 follows from Lemma 7.6, it suffices to show that the well-factorable sequence λ can be decomposed into convolutions of the form $\gamma * \theta$, where the sequences γ and θ are 1-bounded sequences supported on $[1, 2R_1]$ and $[1, 2R_2]$, respectively, with $R_1 = x^{1-\varepsilon}/M$ and $R_2 = Mx^{-1/2}$. This is evidently true in case (i) of Lemma 7.5 since λ is well-factorable. In case (ii), since $P \leq R_1$, we can write $\lambda' = \lambda_1 * \lambda_2$ for some sequences λ_1, λ_2 supported on $[1, R_1/P]$ and $[1, R_2]$, respectively. Then we can take $\gamma = 1_{p \in [P,P')} * \lambda_1$ and $\theta = \lambda_2$.

Proof of Lemma 7.6. By switching the roles of m and n if necessary, we may assume that $M \in [x^{1/2}, x^{3/4}]$ The case $L(n) = n$ follows from [28, Proposition 6.6]. The reduction of the general case to this case is very similar to the corresponding reduction in the proof of Lemma 7.3. $\square$

In the special case when $\lambda_d = 1_{d=1}$, Lemma 7.5 implies that

$$(7.1) \qquad \Big| \sum_{\substack{mn \leq L(x) \\ M \leq m \leq 2M \\ mn \equiv b \,(\mathrm{mod}\,a)}} \alpha(m)\beta(n)\psi(L^{-1}(mn)) \Big| \ll_k \delta a x (\log x)^{O_k(1)}.$$

In the case $L(n) = n$, this is also the type II information required in Green and Tao's proof [12, Section 3] that the Möbius function is orthogonal to nilsequences.

8. Dealing with the equidistributed case

The goal of this section is to prove Propositions 5.3 and 5.4. We shall apply the sieve lemmas in Section 6 to reduce matters to certain Bombieri–Vinogradov type equidistribution results about primes weighted by nilsequences in arithmetic progressions, which follow from results in Section 7.

8.1. **Proof of Proposition 5.3.** We may assume that $\varepsilon > 0$ is fixed, since x is large enough in terms of ε. In what follows, let B be a large enough constant depending on m, d, Δ. We may assume that C is large enough in terms of B. Recall Definition 2.9 and the notation from that definition, thus $\psi(n) = F(g(n)\Gamma)$, where G/Γ is a nilmanifold equipped with a filtration of degree at most d, etc. Let $\mu = \int_{G/\Gamma} F$, so that the η-equidistribution of $\{g(n)\}_{n \leq x}$ implies

$$(8.1) \qquad \Big| \sum_{n \leq x} (\psi(n) - \mu) \Big| \ll x/(\log x)^B.$$

We may assume that $\mu \geq \varepsilon/2$, since otherwise (5.2) is trivial.

We will apply Maynard's sieve method in the form of Proposition 6.1. We need to verify hypotheses (i)–(iv) there for the sequence $\omega_n = \psi(n)$ (with $\delta = 1/2$ and $\theta = 1/10$, say), and then the claim follows.

Hypothesis (i) (with $\delta = 1/2$ in its statement) asserts that

$$\frac{\varphi(a_i)}{a_i} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \psi(n) \geq \frac{1}{2(\log x)} \sum_{n \leq x} \psi(n).$$

Note that $\psi' := \psi - \mu$ is an equidistributed nilsequence lying in $\Psi_d^0(\Delta, K; \eta, x)$. By partial summation, we have

$$(8.2) \qquad \Big| \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \psi'(n) \Big| \ll \sup_{2 \leq y \leq x} \frac{1}{\log L_i(y)} \Big| \sum_{n \leq y} \Lambda(L_i(n))\psi'(n) \Big| + O(x^{1/2}).$$

we may apply (the $d = 1$ case of) Theorem 7.1 to bound the right-hand side of (8.2) by $\ll \eta^\kappa x (\log x)^{A+10}$ for some constant $\kappa = \kappa(d, \Delta) > 0$, which can be made $\ll x(\log x)^{-B}$ by our assumption on η. Hypothesis (i) now follows from the prime number theorem and (8.1).

We turn to hypothesis (ii), which (taking $\theta = 1/10$ there) states that

$$(8.3) \qquad \sum_{r \leq x^{1/10}} \max_{c \,(\mathrm{mod}\,r)} \Big| \sum_{\substack{n \leq x \\ n \equiv c \,(\mathrm{mod}\,r)}} \psi(n) - \frac{1}{r} \sum_{n \leq x} \psi(n) \Big| \ll x/(\log x)^B.$$

We may clearly replace $\psi(n)$ by $\psi'(n) = \psi(n) - \mu$ here; the new nilsequence ψ' lies in $\Psi_d^0(\Delta, K; \eta, x)$. Recalling (8.1), our task is to show that

$$\sum_{r \leq x^{1/10}} \max_{c \,(\mathrm{mod}\, r)} \Big| \sum_{\substack{n \leq x \\ n \equiv c \,(\mathrm{mod}\, r)}} \psi'(n) \Big| \ll x/(\log x)^B.$$

But this follows from Lemma 7.3 with $M = 1$ and $L(n) = n$.

Next we consider hypothesis (iii), which (with $\theta = 1/10$) states that

$$(8.4) \qquad \sum_{r \leq x^{1/10}} \max_{(L_i(c),r)=1} \Big| \sum_{\substack{n \leq x \\ n \equiv c \,(\mathrm{mod}\, r) \\ L_i(n) \in \mathbb{P}}} \psi(n) - \frac{\varphi(a_i)}{\varphi(a_i r)} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \psi(n) \Big| \ll x/(\log x)^B.$$

Applying the Bombieri–Vinogradov theorem, we may replace $\psi(n)$ by $\psi'(n) = \psi(n) - \mu$ on the left-hand side of (8.4). By the argument we used to verify hypothesis (i), we have

$$(8.5) \qquad \Big| \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \psi'(n) \Big| \ll x/(\log x)^{B+1}.$$

Hence, by partial summation, (8.4) reduces to

$$(8.6) \qquad \sum_{r \leq x^{1/10}} \max_{(L_i(c),r)=1} \Big| \sum_{\substack{n \leq y \\ n \equiv c \,(\mathrm{mod}\, r)}} \Lambda(L_i(n)) \psi'(n) \Big| \ll x/(\log x)^B$$

for $y \in [x(\log x)^{-10B}, x]$. This last claim follows from Theorem 7.1.

Finally, hypothesis (iv) states that

$$\max_{c \,(\mathrm{mod}\, r)} \sum_{\substack{n \leq x \\ n \equiv c \,(\mathrm{mod}\, r)}} \psi(n) \ll \frac{1}{r} \sum_{n \leq x} \psi(n).$$

However, this is trivial, since the left-hand side is $O(x/r)$ and the right-hand side is $\geq \varepsilon x/r$ by the consideration at the beginning of the proof and the fact that $\varepsilon > 0$ is fixed.

This concludes the proof of Proposition 5.3.

8.2. **Proof of Proposition 5.4.** We now turn to Chen primes. Let $\mu = \int_{G/\Gamma} F$. Similarly as in the proof of Proposition 5.3, we may assume that $\mu \geq \varepsilon/2$, and we have (8.1).

We apply a weighted version of Chen's sieve from Proposition 6.2. We see from it that the claim follows once we verify hypotheses (i)–(iii) there.

Hypothesis (i) states that

$$(8.7) \qquad \Big| \sum_{\substack{r \leq x^{1/2-\varepsilon'} \\ (r,2a)=1}} \lambda(r) \Big(\sum_{\substack{n \leq x \\ L_2(n) \equiv 0 \,(\mathrm{mod}\, r) \\ L_1(n) \in \mathbb{P}}} \psi(n) - \frac{a}{\varphi(ar)} \sum_{n \leq x} \frac{\psi(n)}{\log L_1(n)} \Big) \Big| \ll x/(\log x)^{10}$$

for some small enough constant $\varepsilon' > 0$, with $\lambda(r)$ either well-factorable of level $x^{1/2-\varepsilon'}$ or a convolution of the shape $1_{p \in [P,P')} * \lambda'$, with λ' a well-factorable function of level $x^{1/2-\varepsilon'}/P$ and $2P \geq P' \geq P \in [x^{1/10}, x^{1/3-\varepsilon'}]$. Note first that by the Bombieri–Vinogradov theorem we may replace ψ with $\psi' = \psi - \mu$ on the left-hand side of (8.7) up to negligible error. Note also that

$$\Big| \sum_{n \leq x} \frac{\psi'(n)}{\log L_1(n)} \Big| \ll x/(\log x)^B$$

by (8.1) and partial summation. Applying partial summation to replace $1_{\mathbb{P}}(L_1(n))$ with the von Mangoldt function, we are left with showing

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon'} \\ (r,2a)=1}} \lambda(r) \sum_{\substack{n \leq y \\ L_2(n) \equiv 0 \,(\mathrm{mod}\, r)}} \Lambda(L_1(n))\psi'(n) \right| \ll x/(\log x)^{100}$$

for all $y \in [x/(\log x)^{101}, x]$. Since $L_2(n) = L_1(n) + 2$, the condition $L_2(n) \equiv 0 \,(\mathrm{mod}\, r)$ is equivalent to $L_1(n) \equiv -2 \,(\mathrm{mod}\, r)$. So this follows from Theorem 7.2.

The statement of hypothesis (ii) is that, for $j \in \{1,2\}$, we have

$$(8.8) \qquad \left| \sum_{\substack{r \leq x^{1/2-\varepsilon'} \\ (r,2a)=1}} \lambda(r) \left(\sum_{\substack{n \leq x \\ L_1(n) \equiv 0 \,(\mathrm{mod}\, r) \\ L_2(n) \in B_j}} \psi(n) - \frac{\varphi(a)}{\varphi(ar)} \sum_{\substack{n \leq x \\ L_2(n) \in B_j}} \psi(n) \right) \right| \ll x/(\log x)^{10},$$

where $\lambda(r)$ is as in hypothesis (i) and

$$B_1 = \{p_1 p_2 p_3 :\ x^{1/10} \leq p_1 \leq x^{1/3-\varepsilon'},\ x^{1/3-\varepsilon'} \leq p_2 \leq (2x/p_1)^{1/2},\ p_3 \geq x^{1/10}\},$$
$$B_2 = \{p_1 p_2 p_3 :\ x^{1/3-\varepsilon'} \leq p_1 \leq p_2 \leq (2x/p_1)^{1/2},\ p_3 \geq x^{1/10}\}.$$

First note that $1_{n \in B_j}$ splits into a sum of $(\log x)^{10}$ type II convolutions $\alpha * \beta(n)$, where $|\alpha(n)|, |\beta(n)| \leq 1$ and α is supported on an interval $[M, 2M] \subset [x^{1/3-\varepsilon'}, x^{1/2}]$. Now, we decompose $\psi(n) = \psi'(n) + \mu$ and note that the contribution of the μ term to (8.8) is $\ll x/(\log x)^B$ by a type II Bombieri–Vinogradov estimate [18, Theorem 17.4] and the previous observation about $1_{n \in B_j}$ being a sum of type II convolutions. Now we shall prove that

$$(8.9) \qquad \left| \sum_{\substack{n \leq x \\ L_2(n) \in B_j}} \psi'(n) \right| \ll x/(\log x)^B.$$

Again by the fact that $1_{L_2(n) \in B_j}$ is of type II, it suffices to prove after a change of variables that

$$\left| \sum_{\substack{mn \leq L_2(x) \\ mn \equiv L_2(0) \,(\mathrm{mod}\, a)}} \alpha(m)\beta(n)\psi'(L_2^{-1}(mn)) \right| \ll x/(\log x)^{2B}$$

for any $|\alpha(n)|, |\beta(n)| \leq 1$, where $\alpha(n)$ is supported on an interval $[M, 2M] \subset [x^{1/3-\varepsilon'}, x^{1/2}]$. But this estimate follows from (7.1) as a special case of Lemma 7.5. Now we have reduced (8.8) to proving

$$(8.10) \qquad \left| \sum_{\substack{r \leq x^{1/2-\varepsilon'} \\ (r,2a)=1}} \lambda(r) \sum_{\substack{n \leq x \\ L_1(n) \equiv 0 \,(\mathrm{mod}\, r) \\ L_2(n) \in B_j}} \psi'(n) \right| \ll x/(\log x)^{10}.$$

The condition $L_1(n) \equiv 0 \,(\mathrm{mod}\, r)$ above is equivalent to $L_2(n) \equiv 2 \,(\mathrm{mod}\, r)$. Once again recalling the type II nature of $1_{L_2(n) \in B_j}$ and using the well-factorable type II estimate of Lemma 7.5, we obtain (8.10).

We are left with hypothesis (iii), which states that

$$(8.11) \qquad \sum_{\substack{n \leq x \\ L_2(n) \in B_j}} \psi(n) \leq (1+\varepsilon') \frac{|B_j \cap [1, L_2(x)]|}{\varphi(a)x} \sum_{n \leq x} \psi(n)$$

for $j \in \{1,2\}$ and for $\varepsilon' > 0$ a small enough constant. This claim follows simply by decomposing $\psi(n) = \psi'(n) + \mu$ and using (8.1), (8.9), and the prime number theorem in arithmetic progressions.

All the hypotheses have now been verified, so the proposition follows.

9. Proof of the main theorem

We now present the proof of our main theorem by combining the work in the previous sections.

Proof of Theorem 1.1. We seek to apply Proposition 4.1 to the functions $\theta = \theta_1$ (in which case $\mathcal{H} = \mathcal{H}_1 = \{0, 2\}, r = 2$) and $\theta = \theta_2$ (in which case $\mathcal{H} = \mathcal{H}_2$ is the tuple fixed when we defined θ_2 and $r = m$). To apply this theorem, we need to establish (4.2).

Thus let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of finite complexity whose linear coefficients are bounded in modulus by some constant L. Recall that by an easy linear algebraic argument, we may assume that Ψ is in s-normal form for some s. Also suppose that $\mathcal{H} \subset [0, L]$. Let $x \geq 1$ and $K \subset [-x, x]^d$ be a convex body such that $\Psi(K) \subset [1, x]^t$ and $\mathrm{Vol}(K) \geq \eta x^d$ for some constant $\eta > 0$. Let $w \geq 1$ be chosen later (sufficiently large in terms of d, t, L) and let $W = \prod_{p \leq w} p$. Let $(b_1, \ldots, b_t) \in B_{\mathcal{H}}^t$. It suffices to prove that

$$(9.1) \qquad \sum_{\mathbf{n} \in \mathbb{Z}^d \cap K} \prod_{i \in [t]} \theta_{j, W, b_i}(\psi_i(\mathbf{n})) \gg_{d, t, L, \eta} \mathrm{Vol}(K)$$

for $j \in \{1, 2\}$, where, recalling (4.1), $\theta_{j, W, b}$ is defined as $\theta'_{W, b}$ for $\theta' = \theta_j$. We will prove (9.1) by appealing to Theorem 3.2. Let $M = M(d, t, L)$ be the constant produced by this theorem, and suppose that x is large enough and α small enough as in this theorem. Without loss of generality (upon using Bertrand's postulate, dilating x by a factor of at most 8 and shrinking η by a factor at most 8^d), we may assume that x is prime and $K \subset [-x/4, x/4]^d$. By Proposition 4.2, there exists $c \in (0, 1)$ and $C > 0$ depending only on d, t, L, M (and therefore ultimately on (d, t, L) only) and an (M, α)-pseudorandom measure $\nu_{\mathbf{b}} : \mathbb{Z}/x\mathbb{Z} \to \mathbb{R}_{\geq 0}$ such that $\theta_{j, W, b_i}(n) \leq C\nu(n)$ whenever $i \in [t]$ and $n \in [x^c, x]$. Define then $\lambda_i : \mathbb{Z}/x\mathbb{Z} \to \mathbb{R}_{\geq 0}$ by $\lambda_i = \theta_{j, W, b_i} 1_{[x^c, x]} / C$ where as usual we identify $[x]$ and $\mathbb{Z}/x\mathbb{Z}$ in the natural way. Therefore we have $\lambda_i \leq \nu$ on $\mathbb{Z}/x\mathbb{Z}$ by construction, so hypothesis (i) of Theorem 3.2 is satisfied.

We now turn to hypothesis (ii). Let $\delta_1 = \delta_0/(3C)$ where δ_0 is the absolute constant of Proposition 5.2 and δ_2 be the implied constant of Proposition 5.1, for our choice of m, divided by $3C$. Therefore δ_j depends at most on d, t, L for each $j \in [t]$. Let Y_j, ε_j be the corresponding constants given by Theorem 3.2, which are functions of d, t, L, δ_j for $j \in \{1, 2\}$. Fix $i \in [t]$. For $j \in \{1, 2\}$, denote by f_j the function λ_i constructed above from the function $\theta = \theta_j$. We intend to show that

$$(9.2) \qquad \sum_{n \leq x} f_j(n) \psi(n) \geq \delta_j \sum_{n \leq x} \psi(n)$$

whenever $\psi : \mathbb{Z} \to [0, 1]$ is a nilsequence of complexity at most Y_j satisfying $\sum_{n \leq x} \psi(n) \geq \varepsilon_j x$. Since $\sum_{n \leq x^c} \theta_{W, b}(n) \ll x^{c+o(1)}$, it suffices to show that $\sum_{n \leq x} \theta_{W, b}(n) \psi(n) \geq 2C\delta_j \sum_{n \leq x} \psi(n)$. But this follows from Propositions 5.2 and 5.1 assuming that x is large enough, and that the diameter of $\mathcal{H}_2$ is smaller than w.

Therefore, the hypotheses of Theorem 3.2 are met, and applying this theorem yields equation (9.1) and we are done.

Thus we obtain Theorem 1.1 with $C_i(\Psi) = \prod_p \beta_p(\Psi_{\mathcal{H}_i})$. $\qquad \square$

9.1. The case of primes of the form $x^2 + y^2 + 1$.
We now turn to the proof of Theorem 1.2. We shall be brief with the arguments at places, since for the most part they closely resemble those used to prove Theorem 1.1.

Let $\theta_3(n)$ be the weighted indicator of primes of the form $x^2 + y^2 + 1$ given by (1.2). Also denote the set of sums of two squares by

$$S := \{n \geq 1 : \ n = x^2 + y^2 \text{ for some } x, y \in \mathbb{Z}\}.$$

We follow the proof strategy of Theorem 1.1. Let $W := 6^3 \prod_{3 \leq p \leq w} p$. Define $\theta_{3,W,b}(n) := (W/\varphi(W))^{3/2} \theta(Wn + b)$.

We first claim that Theorem 1.2 follows if

$$(9.3) \qquad \sum_{\substack{\mathbf{n} \in \mathbb{Z}^d \\ W\mathbf{n}+\mathbf{a} \in K}} \prod_{i=1}^{t} \theta_{3,W,\psi_i(\mathbf{a})}(\dot{\psi}_i(\mathbf{n})) \gg W^{-d}\mathrm{Vol}(K),$$

for any convex body K satisfying $\Psi(K) \subset [1, x]^d$, $\mathrm{Vol}(K) \gg x^d$ and for each $\mathbf{a} \in \mathcal{A}$, where $\mathcal{A} := \{\mathbf{a} \in (\mathbb{Z}/W\mathbb{Z})^d : \ \forall\, i \in [t]\ Wn + \psi_i(\mathbf{a}) \text{ is amenable}\}$. The proof of this implication is essentially the same as for Proposition 4.1, i.e., we choose $K = K_{\mathbf{a}}$ as there and sum (9.3) over all $\mathbf{a} \in \mathcal{A}$ and note that $|\mathcal{A}| \gg \prod_p \beta'_p W^d$ by the Chinese remainder theorem, where

$$\beta'_p := \mathbb{E}_{\mathbf{a} \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i \in [t]} (1 - |A_p|/p)^{-1} 1_{\psi_i(\mathbf{a}) \notin A_p \,(\mathrm{mod}\, p)}$$

and $A_p = \{0, 1\}$ for $p \equiv -1\,(\mathrm{mod}\, 4)$ and $A_p = \{0\}$ otherwise.

Thus, by applying Theorem 3.2, it suffices to prove that the following hold for all fixed w and $1 \leq b_i \leq W$ such that $Wn + b_i$ is amenable.

(1) For any $M \geq 1, \alpha > 0$, $t \geq 1$, there exist $0 < c < 1$ and an (M, α)-pseudorandom measure $\nu_{\mathbf{b}} : \mathbb{Z}/x\mathbb{Z} \to \mathbb{R}_{\geq 0}$ such that $\theta_{W,b_i}(n) \ll_{M,t} \nu(n)$ whenever $i \in [t]$ and $n \in [x^c, x]$.

(2) There exists an absolute constant $\delta_0 > 0$ such that, for any $Y \geq 1$, $\varepsilon > 0$ and $x \geq x_0(Y, \varepsilon)$ large enough, we have

$$\sum_{n \leq x} \theta_{3,W,b_i}(n)\psi(n) \geq \delta_0 \sum_{n \leq x} \psi(n)$$

whenever $\psi : \mathbb{Z} \to [0, 1]$ is a nilsequence of complexity at most Y satisfying $\sum_{n \leq x} \psi(n) \geq \varepsilon x$.

Proof of (1). This follows from the work of Sun and Pan [29] (Section 2 and in particular Proposition 2.1 there).

Proof of (2). Let Y and ε be fixed in the statement of (2). For the proof of (2), it suffices to prove the following result, which is a direct analogue of Proposition 5.1 in the case of bounded gap integers or of Proposition 5.2 in the case of Chen primes.

Proposition 9.1 (Primes of the form $x^2 + y^2 + 1$ with nilsequences). *Fix positive integers d, Δ and some $\varepsilon > 0$, $K \geq 2$. Also let $w \geq 1$ be sufficiently large in terms of $d, \Delta, \varepsilon, K$ and $W = 6^3 \prod_{3 \leq p \leq w} p$. The following statement holds for sufficiently large $x \geq x_0(d, \Delta, \varepsilon, K, w)$.*

Let $\psi \in \Psi_d(\Delta, K)$ be a nilsequence taking values in $[0, 1]$. Then for some absolute constant $\delta_0 > 0$ and any $1 \leq b \leq W$ such that $Wn + b$ is amenable, we have

$$\sum_{\substack{n \leq x \\ Wn+b \in \mathbb{P} \\ Wn+b-1 \in S}} \psi(n) \geq \left(\frac{W}{\varphi(W)}\right)^{3/2} \frac{\delta_0}{(\log x)^{3/2}} \left(\sum_{n \leq x} \psi(n) - \varepsilon x\right).$$

By arguments similar to those in Section 5 (with Lemma 5.5 slightly adjusted to handle the local problem in our setting), we can reduce this to the equidistributed case.

Proposition 9.2 (Primes of the form $x^2 + y^2 + 1$ weighted by equidistributed nilsequences).
Fix positive integers d, Δ and some $\varepsilon > 0$, $A \geq 2$. There exists $C = C(d, \Delta) > 0$, such that the following statement holds for sufficiently large $x \geq x_0(d, \Delta, \varepsilon, A)$.

Let $K \geq 2$ and $\eta \in (0, 1/2)$ be parameters satisfying the conditions

$$\eta \leq K^{-C} (\log x)^{-CA}, \quad K \leq (\log x)^C.$$

Let $\psi \in \Psi_d(\Delta, K; \eta, x)$ be a nilsequence taking values in $[0, 1]$. Let $L(n) = an + b$ be an amenable linear function, where $1 \leq a \leq \log x$, $|b| \leq x$. Then for some absolute constant $\delta_0 > 0$ we have

$$(9.4) \qquad \sum_{\substack{n \leq x \\ L(n) \in \mathbb{P} \\ L(n) - 1 \in S}} \psi(n) \geq \delta_0 \frac{\mathfrak{S}(L)}{(\log x)^{3/2}} \left(\sum_{n \leq x} \psi(n) - \varepsilon x \right),$$

where the singular series is given by (6.4).

The remaining task is then to prove this proposition.

Proof of Proposition 9.2. We may assume that $\varepsilon > 0$ is fixed, since x is large enough in terms of ε.

We apply Proposition 6.4. Thus, in order to obtain (9.4), it suffices to verify hypotheses (i)–(ii) of Proposition 6.4 for $\omega_n = \psi(n)$ in order to obtain the claim. Since $\sum_{n \leq x} \omega_n \geq \varepsilon x \gg x$, it in fact suffices to verify versions of hypotheses (i)–(ii) where $(\sum_{n \leq x} \omega_n)/(\log x)^{100}$ is replaced with $x/(\log x)^{100}$ on the right-hand side of the inequalities (6.1), (6.2).

Write $\psi(n) = \psi'(n) + \mu$, where $\mu = \int_{G/\Gamma} F$. Observing that hypotheses (i)–(ii) hold for constant sequences (in the case of (i) by a bilinear Bombieri–Vinogradov type estimate [18, Theorem 17.4] and in the case of (ii) by the classical Bombieri–Vinogradov theorem), it suffices to verify hypotheses (i)–(ii) (with $x/(\log x)^{100}$ on the right-hand side of (6.1), (6.2)) for $\psi'(n)$, which now belongs to $\Psi_d^0(\Delta, K; \eta, x)$ with $\eta \leq (\log x)^{-CA}$ for a large constant C.

Verifying hypothesis (i). Let $\omega_n = \psi'(n)$. First note that by partial summation and the fact that $\psi' \in \Psi_d^0(\Delta, K; \eta, x)$, we have

$$(9.5) \qquad \left| \sum_{n \leq x} \frac{\psi'(n)}{\log(yn)} \right| \ll \frac{x}{(\log x)^{300}},$$

say, uniformly for $x^{-0.99} \leq y \leq x$. Hence, recalling the definition of $g(\ell)$ in hypothesis (i), and denoting $\delta = (b - 1, a)$, our task is to show that

$$\left| \sum_{\substack{r \leq x^{1/2 - \varepsilon} \\ (r, a) = 1}} \lambda_r^{+, \mathrm{LIN}} \sum_{\substack{x^{1/(3+\varepsilon)} \leq \ell_1 \leq x^{1/(3+\varepsilon)} \\ \ell_2 \leq x^{0.9 - 1/(3+\varepsilon)} \\ (\ell_1 \ell_2, a) = \delta \\ (\ell_1 \ell_2, r) = 1}} \alpha(\ell_1) \beta(\ell_2) \left(\sum_{\substack{n \leq x \\ p \leq x \\ L(n) = \ell_1 \ell_2 p + 1 \\ L(n) \equiv 0 \, (\mathrm{mod}\, r)}} \psi'(n) \right) \right| \ll \frac{x}{(\log x)^{100}},$$

uniformly for $|\alpha(n)|, |\beta(n)| \leq 1$. Merging the variables ℓ_2 and p as $m = \ell_2 p$, it suffices to show that

$$\left| \sum_{\substack{r \leq x^{1/2 - \varepsilon} \\ (r, a) = 1}} \lambda_r^{+, \mathrm{LIN}} \sum_{\ell_1 \in I} \alpha(\ell_1) 1_{(\ell_1, r) = 1, (\ell_1, a) = \delta_1} b(m) 1_{(m, r) = 1, (m, a) = \delta_2} \left(\sum_{\substack{n \leq x \\ L(n) = \ell_1 m + 1 \\ L(n) \equiv 0 \, (\mathrm{mod}\, r)}} \psi'(n) \right) \right|$$

$$\ll \frac{x}{(\log x)^{101}},$$

uniformly for $1 \leq \delta_1, \delta_2 \leq \delta$ and $|\alpha(n)|, |b(n)| \leq d_2(n)$, where for brevity we have denoted $I := [x^{1/(3+\varepsilon)}, x^{1-1/(3+\varepsilon)}]$. We make a linear change of variables in the inner sum over n to reduce to

$$\sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,a)=1}} |\lambda_r^{+,\mathrm{LIN}}| \Big| \sum_{\ell_1 \in I} \alpha(\ell_1) 1_{(\ell_1,r)=1, (\ell,a)=\delta_1} b(m) 1_{(m,r)=1,(m,a)=\delta_2} \Big(\sum_{\substack{\ell_1 m \leq L(x) \\ \ell_1 m \equiv 1 \,(\mathrm{mod}\, r) \\ \ell_1 m \equiv b-1 \,(\mathrm{mod}\, a)}} \psi'(\frac{\ell_1 m + 1 - b}{a}) \Big) \Big|$$

$$\ll \frac{x}{(\log x)^{101}}.$$

We can replace the sequence $\lambda_r^{+,\mathrm{LIN}}$ above with a well-factorable sequence using [8, Corollary 12.17], which splits the linear sieve coefficients into a linear combination of boundedly many well-factorable sequences. Now the claimed estimate follows directly from the well-factorable type II estimate given by Lemma 7.5.

Verifying hypothesis (ii). Let $\omega_n = \psi'(n)$. Again applying (9.5), we reduce to

$$\Big| \sum_{\substack{r \leq x^{3/7-\varepsilon} \\ (r,a)=1}} \lambda_r^{-,\mathrm{SEM}} \Big(\sum_{\substack{n \leq x \\ L(n) \in \mathbb{P} \\ L(n) \equiv 1 \,(\mathrm{mod}\, r)}} \psi'(n) \Big) \Big| \ll \frac{x}{(\log x)^{100}}.$$

Making the change of variables $n' = L(n)$, this is equivalent to

$$\Big| \sum_{\substack{r \leq x^{3/7-\varepsilon} \\ (r,a)=1}} \lambda_r^{-,\mathrm{SEM}} \Big(\sum_{\substack{n \leq L(x) \\ n \in \mathbb{P} \\ n \equiv 1 \,(\mathrm{mod}\, r) \\ n \equiv b \,(\mathrm{mod}\, a)}} \psi'(L^{-1}(n)) \Big) \Big| \ll \frac{x}{(\log x)^{100}}.$$

Applying partial summation, it suffices to show

$$\Big| \sum_{\substack{r \leq x^{3/7-\varepsilon} \\ (r,a)=1}} \lambda_r^{-,\mathrm{SEM}} \Big(\sum_{\substack{n \leq y \\ n \equiv 1 \,(\mathrm{mod}\, r) \\ n \equiv b \,(\mathrm{mod}\, a)}} \Lambda(n) \psi'(L^{-1}(n)) \Big) \Big| \ll \frac{x}{(\log x)^{100}},$$

uniformly for $1 \leq y \leq L(x)$. By Vaughan's identity, we can write the von Mangoldt function as a sum of $\ll (\log x)^{10}$ convolutions $a*b(n)$, where $|a(n)|, |b(n)| \leq (\log n) d_2(n)$ and $\mathrm{supp}(a) \subset [M, 2M]$ and one of the following holds:

(1) $M \ll x^{1/3}$ and $b(n) \equiv 1$ or $b(n) \equiv \log n$ (type I case);
(2) $x^{1/2} \ll M \ll x^{2/3}$ (type II case).

Thus, we reduce to proving that

$$\Big| \sum_{\substack{r \leq x^{3/7-\varepsilon} \\ (r,a)=1}} \lambda_r^{-,\mathrm{SEM}} \Big(\sum_{\substack{mn \leq L(x) \\ mn \equiv 1 \,(\mathrm{mod}\, r) \\ mn \equiv b \,(\mathrm{mod}\, a)}} a(m) b(n) \psi'(L^{-1}(mn)) \Big) \Big| \ll \frac{x}{(\log x)^{110}}.$$

In the type I case, we can in fact assume that $b(n) \equiv 1$ by applying partial summation. Now, to handle the type I sums, we can apply the bound $|\lambda_r^{-,\mathrm{SEM}}| \leq 1$, followed by Cauchy–Schwarz to dispose of the $a(m)$ coefficients (this loses a factor of $(\log x)^{10}$, say, so for the resulting sum we need a bound of $x/(\log x)^{120}$). To the resulting sum we can then apply Lemma 7.3 to obtain the desired conclusion.

We then turn to the type I sums. If $M \leq x^{4/7}$, we can directly apply Lemma 7.6 with $R_1 = x^{3/7-\varepsilon}$, $R_2 = 1$, since then $R_1 \leq x^{1-\varepsilon}/M$, $R_1 R_2^2 \leq M x^{-\varepsilon}$. Hence, we may assume for now on that $x^{4/7} \leq M \ll x^{2/3}$.

We now apply a partial factorization of the lower bound semilinear sieve weights from [31, Lemma 9.2, formulas (10.3), (10.4)] (taking $\theta = \varepsilon/2$ and replacing ε with $\varepsilon/10$ in those formulas). Since $x^{1/3-\varepsilon} \ll x^{1-\varepsilon}/M \ll x^{3/7-\varepsilon}$, we conclude that

$$(9.6) \qquad |\lambda_r^{-,\mathrm{SEM}}| \ll (\log x)^2 \max_{R_1,R_2} \sum_{\substack{r=r_1r_2 \\ r_1 \in [R_1, 2R_1] \\ r_2 \in [R_2, 2R_2]}} 1,$$

where the maximum is over those $(R_1, R_2) \in \mathbb{R}_{\geq 1}^2$ satisfying

$$(9.7) \qquad R_1 \leq x^{1-\varepsilon}/M, \quad R_1 R_2^2 \leq M x^{-\varepsilon}, \quad R_1 R_2 \leq x^{3/7-\varepsilon/2}.$$

Hence, applying (9.6), the remaining task is to show that

$$\sum_{\substack{R_1 \leq r_1 \leq 2R_1 \\ R_2 \leq r_2 \leq 2R_2}} \Big| \sum_{\substack{mn \leq L(x) \\ mn \equiv 1 \,(\mathrm{mod}\, r_1 r_2) \\ mn \equiv b \,(\mathrm{mod}\, a)}} a(m)b(n)\psi'(L^{-1}(mn)) \Big| \ll \frac{x}{(\log x)^{200}}$$

under the constraints (9.7). Since the constraints on R_1, R_2 are precisely as in Lemma 7.6, we may appeal to that lemma to conclude. This completes the verification of hypothesis (i)–(ii), and hence the proof of Theorem 1.2. $\qquad\square$

References

[1] P.-Y. Bienvenu. A higher-dimensional Siegel-Walfisz theorem. *Acta Arith.*, 179(1):79–100, 2017.

[2] J. R. Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973.

[3] D. Conlon, J. Fox, and Y. Zhao. The Green-Tao theorem: an exposition. *EMS Surv. Math. Sci.*, 1(2):249–282, 2014.

[4] D. Conlon, J. Fox, and Y. Zhao. A relative Szemerédi theorem. *Geom. Funct. Anal.*, 25(3):733–762, 2015.

[5] P. Dodos and V. Kanellopoulos. Uniformity norms, their weaker versions, and applications. *arXiv e-prints*, page arXiv:1610.00487, October 2016.

[6] N. Frantzikinakis and B. Host. Higher order Fourier analysis of multiplicative functions and applications. *J. Amer. Math. Soc.*, 30(1):67–157, 2017.

[7] C. Frei, P. Koymans, and E. Sofos. Vinogradov's three primes theorem with primes having given primitive roots. *Math. Proc. Cambridge Philos. Soc.*, pages 1–36, 2019.

[8] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.

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

[10] B. Green and T. Tao. An arithmetic regularity lemma, an associated counting lemma, and applications. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 261–334. János Bolyai Math. Soc., Budapest, 2010.

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

[12] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.

[13] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.

[14] B. Green and T. Tao. An arithmetic regularity lemma, associated counting lemma, and applications. *arXiv e-print*, page arXiv:1002.2028, 2020.

[15] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[16] L. Grimmelt. Vinogradov's Theorem with Fouvry-Iwaniec Primes. *arXiv e-prints*, page arXiv:1809.10008, September 2018.

[17] H. Iwaniec. Primes of the type $\phi(x, y) + A$ where ϕ is a quadratic form. *Acta Arith.*, 21:203–234, 1972.

[18] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, Rhode Island, 2004.

[19] D. M. Kane. An asymptotic for the number of solutions to linear equations in prime numbers from specified Chebotarev classes. *Int. J. Number Theory*, 9(4):1073–1111, 2013.

[20] H. Li and H. Pan. A density version of Vinogradov's three primes theorem. *Forum Math.*, 22(4):699–714, 2010.

[21] H. Li and H. Pan. The Green-Tao theorem for Piatetski-Shapiro primes. arXiv:1901.09372, 2019.

[22] K. Matomäki and X. Shao. Vinogradov's three primes theorem with almost twin primes. *Compos. Math.*, 153(6):1220–1256, 2017.

[23] J. Maynard. Dense clusters of primes in subsets. *Compos. Math.*, 152(7):1517–1554, 2016.

[24] J. Pintz. Are there arbitrarily long arithmetic progressions in the sequence of twin primes? In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 525–559. János Bolyai Math. Soc., Budapest, 2010.

[25] J. Pintz. *Patterns of Primes in Arithmetic Progressions*, pages 369–379. Springer International Publishing, Cham, 2017.

[26] L. Rimanić and J. Wolf. Szemerédi's theorem in the primes. *Proc. Edinb. Math. Soc. (2)*, 62(2):443–457, 2019.

[27] X. Shao. A density version of the Vinogradov three primes theorem. *Duke Math. J.*, 163(3):489–512, 2014.

[28] X. Shao and J. Teräväinen. The Bombieri-Vinogradov theorem for nilsequences. arXiv:2006.05954, 2020.

[29] Y.-C. Sun and H. Pan. The Green-Tao theorem for primes of the form $x^2 + y^2 + 1$. *Monatsh. Math.*, 189(4):715–733, 2019.

[30] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, paperback edition, 2010.

[31] J. Teräväinen. The Goldbach problem for primes that are sums of two squares plus one. *Mathematika*, 64(1):20–70, 2018.

[32] B. Zhou. The Chen primes contain arbitrarily long arithmetic progressions. *Acta Arith.*, 138(4):301–315, 2009.

Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn
Email address: `pierre.bienvenu@mpim-bonn.mpg.de`

Department of Mathematics, University of Kentucky, Lexington, KY, 40506, USA
Email address: `xuancheng.shao@uky.edu`

Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom
Email address: `joni.teravainen@maths.ox.ac.uk`