

Partition regularity of homogeneous quadratics: Current trends and challenges

Nikos Frantzikinakis

Abstract. Suppose we partition the integers into finitely many cells. Can we always find a solution of the equation $x^2 + y^2 = z^2$ with x, y, z on the same cell? What about more general homogeneous quadratic equations in three variables? These are basic questions in arithmetic Ramsey theory, which have recently been partially answered using ideas inspired by ergodic theory and tools such as Gowers uniformity properties and concentration estimates of bounded multiplicative functions. The aim of this article is to provide an introduction to this exciting research area, explaining the main ideas behind the recent progress and some of the important challenges that lie ahead.

1. Introduction

A general principle in arithmetic Ramsey theory is that if you finitely partition $\mathbb{N} := \{1, 2, \dots\}$, then a cell of the partition will have a lot of structure. By structure, we mean solutions to algebraic equations or systems of algebraic equations. This is nicely illustrated by a theorem of Schur [53] and van der Waerden [59] and its extension by Rado [51], which covers systems of linear equations. In the case of a single linear equation, it takes a particularly nice form that we will explain next.

Definition 1.1. We say that the equation $P(x, y, z) = 0$ is *partition regular* if for every finite partition of $\mathbb{N}$ there exist distinct x, y, z on the same cell that satisfy the equation.

Rado's theorem states that for $a, b, c \in \mathbb{N}$, the equation

$$ax + by = cz$$

Mathematics Subject Classification 2020: 05D10 (primary); 11N37, 11B30, 37A44 (secondary).

Keywords: partition regularity, Pythagorean triples, multiplicative functions, concentration inequalities, Gowers uniformity.

is partition regular if and only if either a, b , or $a + b$ equals c ; we call such a triple of positive integers (a, b, c) a *Rado triple*. To prove the necessity of the conditions, take a prime $p > a + b + c$ and partition the positive integers into $p - 1$ cells according to the last non-zero digit in their base p expansion. Such colorings arise from level sets of some modified Dirichlet characters (see Section 5.2) and will be our main source of counterexamples in this article.

Higher-degree polynomial equations have proved much harder to tackle. Perhaps the most well-known problem in this area is that of Erdős and Graham [26, 27], which asks whether Pythagorean triples are partition regular, i.e., whether the equation

$$x^2 + y^2 = z^2 \quad (1.1)$$

is partition regular. Graham in [26] places the origin of the problem in the late 1970s and offered \$250 for its solution, noting that “there is actually very little data (in either direction) to know which way to guess”. While this may have been true a decade ago, there have been some positive developments in recent years. The case where one allows only partitions of $\mathbb{N}$ into two sets was verified in 2016 with the help of a computer search [39]; this endeavor was hailed as the “longest mathematical proof” at the time, consuming 200 terabytes of data [45]. Also, partition regularity of Pythagorean triples for level sets of finite-valued multiplicative functions was recently established [23].

The seminal work of Furstenberg [24] and Sárközy [52], culminating in the influential polynomial Szemerédi theorem of Bergelson and Leibman [7], allows us to deal with shift-invariant systems of polynomial equations. Regarding equations that resemble (1.1), here are some examples that are known to be or not to be partition regular:

- (i) $x^2 + y = z$ is partition regular (Bergelson [4]).
- (ii) $x + y = z^2$ is not partition regular (Csikvári–Gyarmati–Sárközy [14]).
- (iii) $x^2 + y^2 = z$ is not partition regular (variant of Green–Lindqvist [36, Section 2]).
- (iv) $x^2 + y = z^2$ is partition regular (Moreira [48]).

The second equation is partition regular if we consider partitions with only two cells [36, 49] or if we only require x, y to belong to the same cell [41]. There are a variety of tools used to prove these results, such as elementary combinatorial techniques, Fourier analysis, topological dynamics, and ergodic theory. Other partition regularity results of similar flavor can be found in [2, 3, 6, 16–18, 46].

However, in the fully non-linear setting, the only known results involve equations in more than three variables. For example, a result of Chow–Lindqvist–Prendiville [13] establishes that the equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_5^2$$

is partition regular (see also [9–11, 50] for related results).

Here, we focus on homogeneous quadratic equations in three variables, for which the methodology developed to deal with previous results has so far proved inadequate. A problem that drives much recent research in this area is the following.

Conjecture 1. *If $a, b, c \in \mathbb{N}$, then the equation*

$$ax^2 + by^2 = cz^2 \tag{1.2}$$

is partition regular if and only if (a, b, c) is a Rado triple.

The necessity of the conditions follows by combining Rado’s theorem [51] with the fact that partition regularity of the equation $P(x^2, y^2, z^2) = 0$ implies partition regularity of the equation $P(x, y, z) = 0$. Regarding sufficiency, unfortunately there is no instance of a Rado triple for which the partition regularity of (1.2) has been proved. Three typical cases to keep in mind are the equations

$$x^2 + y^2 = z^2, \quad x^2 + 2y^2 = z^2, \quad x^2 + y^2 = 2z^2,$$

which, as we will see, have particular difficulties that are representative of the general case. The last equation was conjectured to be partition regular by Gyarmati–Ruzsa [37] and may even be density regular with respect to the natural density, i.e., it may have non-trivial solutions on any subset of the integers with positive upper density.

Although we are not in a position to address the question of partition regularity for triples, we will describe a systematic method, based on recent work by the author, Klurman, and Moreira [22, 23] and older work by the author and Host [21], which allows us to address the problem for pairs in several cases.

Definition 1.2. Given non-zero $a, b, c \in \mathbb{Z}$, we say that equation (1.2) is *partition regular with respect to x, y* if for every finite coloring of $\mathbb{N}$ there exist distinct $x, y \in \mathbb{N}$, with the same color, and $z \in \mathbb{N}$ that satisfy (1.2). Similarly, we define the partition regularity of (1.2) with respect to the variables x, z and y, z .

One of our main results in [23] is that Pythagorean pairs are partition regular, which answers a question that appeared in [18, 19, 21, 31].

Theorem 1.3. *The Pythagorean equation (1.1) is partition regular with respect to all pairs of variables.*

Our starting point is to link our problem with one concerning the asymptotic properties of bounded multiplicative functions via a representation result of Bochner–Herglotz; the details of this link are explained in Section 3.3. Let us just say that multiplicative functions play the role of characters in multiplicative Fourier analysis, which is an alternative to additive Fourier analysis that is worth exploring for the Pythagorean equation, or any other equation that is dilation invariant but not translation invariant. As we explain in Section 7.3 below, there is no known analogous “Bochner–Herglotz representation result” in the context of the partition regularity of Pythagorean triples, and an approach based on ergodic theory may be better suited to this problem. This is similar to the situation with shift-invariant patterns, where patterns of length two can be handled by a variety of tools, including Fourier analysis, but ergodic theory tools are better suited for patterns of length three and above, especially when these patterns are sparse.

In [22], we extended our methodology to cover partition regularity with respect to the variables x, y of more general equations of the form $ax^2 + by^2 = cz^2$, where a, b, c are non-zero integers such that either ac or bc is a square. We will discuss why equations like $x^2 \pm y^2 = 2z^2$ are harder to handle. In all of these problems, our approach depends on an in-depth study of the asymptotic behavior of the averages

$$\frac{1}{N^2} \sum_{1 \leq m, n \leq N} f(P_1(m, n)) \cdot \overline{f(P_2(m, n))},$$

where f is an arbitrary completely multiplicative function taking values on the unit circle, and P_1, P_2 are binary quadratic forms used to parametrize the solutions of (1.2). We use Gowers uniformity properties to deal with “minor arc” multiplicative functions, which is the aperiodic case discussed in Section 4, and we use concentration estimates to deal with “major arc” multiplicative functions, which is the pretentious case discussed in Section 5.

Before these works, partition regularity of pairs was proved by the author and Host in [21], which covered, for example, partition regularity of the equation $16x^2 + 9y^2 = z^2$ with respect to x, y , but missed the case of Pythagorean pairs for reasons we will explain later. Extending these ideas, Sun in [54, 55] established partition regularity with respect to x, y for the equation $x^2 - y^2 = z^2$ when $\mathbb{N}$ is replaced by the ring of integers of a larger number field, such as the Gaussian integers. However, the methods used there do not apply to $\mathbb{N}$.

One may also pursue similar partition regularity questions on the rationals, but as it turns out, in the case of homogeneous equations (or dilation-invariant patterns), partition regularity over $\mathbb{Q}$ and $\mathbb{Z}$ is an equivalent problem [18, Proposition 2.13].

This is not the case for non-homogeneous patterns, and there are several interesting patterns that are known to be partition regular over the rationals, but it is not known whether they are partition regular over the integers; see, for example, some recent results in [1, 8].

Notation. We let $\mathbb{N} := \{1, 2, \dots\}$, $\mathbb{Z}_+ := \{0, 1, 2, \dots\}$, $\mathbb{R}_+ := [0, +\infty)$, $\mathbb{S}^1 \subset \mathbb{C}$ be the unit circle, and $\mathbb{U}$ be the closed complex unit disk. With $\mathbb{P}$ we denote the set of primes, and throughout we use the letter p to denote prime numbers.

For $t \in \mathbb{R}$, we let $e(t) := e^{2\pi it}$. For $z \in \mathbb{C}$, with $\Re(z)$ and $\Im(z)$ we denote the real and imaginary parts of z , respectively, and we also let $\exp(z) := e^z$.

For $N \in \mathbb{N}$, we let $[N] := \{1, \dots, N\}$. We often denote sequences $a: \mathbb{N} \rightarrow \mathbb{U}$ by $(a(n))$ instead of $(a(n))_{n \in \mathbb{N}}$.

If A is a finite non-empty subset of the integers and $a: A \rightarrow \mathbb{C}$, we let

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

We write $a(n) \ll b(n)$ if for some $C > 0$ we have $a(n) \leq C b(n)$ for every $n \in \mathbb{N}$.

Throughout this article, the letter f is typically used for multiplicative functions and the letter χ for Dirichlet characters. With $\mathcal{M}$ we denote the set of completely multiplicative functions with modulus 1.

2. Partition regularity of generalized Pythagorean pairs

We state here the main results concerning the partition regularity of pairs for homogeneous quadratic equations and also make a crucial reduction to a density regularity statement involving parametrizations of the solution sets of the respective pairs.

2.1. Main results

A variant of Conjecture 1 that covers partition regularity of pairs for (1.2) is the following one.

Conjecture 2. *Suppose $a, b, c \in \mathbb{Z}$ are non-zero and at least one of the integers $ac, bc, (a + b)c$ is a square, then (1.2) is partition regular with respect to x, y .*

If (1.2) is partition regular with respect to all three variables, then it is obviously partition regular for any pair of variables. However, the converse is not true: [23, Theorem 1.1] implies that the equation $x^2 + y^2 = 4z^2$ is partition regular with respect to any pair of variables, but this equation is not partition regular with respect to all three variables since $(1, 1, 4)$ is not a Rado triple.

In [22], we prove the following result, which partially answers Conjecture 2.

Theorem 2.1. *If $a, b, c \in \mathbb{Z}$ are non-zero and ac or bc is a square, then (1.2) is partition regular with respect to x, y .*

Taking $a = b = c = 1$ and $a = -b = c = 1$, we get that the equation $x^2 + y^2 = z^2$ is partition regular with respect to all pairs of variables x, y, z . Taking $a = c = 1, b = 2$, and $a = c = 1, b = -2$, we get that the equation $x^2 + 2y^2 = z^2$ is partition regular with respect to the pairs x, y and y, z . We will see later why the pair x, z is harder to handle for this equation, and why a similar problem arises with the equation $x^2 + y^2 = 2z^2$, this time with respect to all pairs of variables. In this regard, we can only give a conditional result, assuming a difficult Chowla–Elliott-type conjecture (see Section 6.5.2) for a class of “sufficiently aperiodic” completely multiplicative functions.

Theorem 2.2. *Suppose that Conjecture 3 below holds. If $a, b, c \in \mathbb{Z}$ are non-zero and $(a + b)c$ is a square (perhaps zero), then (1.2) is partition regular with respect to x, y .*

Although the conditionality of this result is an unpleasant feature, we can cover, unconditionally, partitions generated by pretentious multiplicative functions (see [22, Theorem 1.4] for the precise statement), which are the only known sources of counterexamples to the partition regularity of pairs of homogeneous quadratic equations.

An immediate consequence of Theorems 2.1 and 2.2 is that conditionally on the truth of Conjecture 3 in Section 6.5.2, we have that if (a, b, c) is a Rado triple, then equation (1.2) is partition regular with respect to all pairs of variables x, y, z . We get this result unconditionally if $a = c$ or $b = c$.

2.2. Parametric reformulation

To prove our main results, we will work with the parametric form of the solutions of equation (1.2). We give three representative examples of such parametrizations (below, k, m, n are arbitrary integers):

- (i) $x^2 + y^2 = z^2$: $x = k(m^2 - n^2)$, $y = k(2mn)$, $z = k(m^2 + n^2)$.
- (ii) $x^2 + 2y^2 = z^2$: $x = k(m^2 - 2n^2)$, $y = k(2mn)$, $z = k(m^2 + 2n^2)$.
- (iii) $x^2 + y^2 = 2z^2$: $x = k(m^2 - n^2 + 2mn)$, $y = k(m^2 - n^2 - 2mn)$, $z = k(m^2 + n^2)$.

To give the reader a sense of how certain features of these parametrizations affect our analysis, we make a few remarks that will be properly justified when we go deeper into the proof outline of our main results.

In the case of (i), in order to prove partition regularity with respect to x, y , it suffices to show that for every finite partition of $\mathbb{N}$, there exist $k, m, n \in \mathbb{N}$ such that the integers

$$k(m^2 - n^2) \quad \text{and} \quad k(2mn)$$

are distinct and belong to the same partition cell. The fact that both polynomials $m^2 - n^2$ and $2mn$ are products of linear forms simplifies the argument substantially. On the other hand, the parameter z uses the irreducible quadratic form $m^2 + n^2$, which takes values on a zero density subset of $\mathbb{N}$, and this makes the corresponding analysis of partition regularity with respect to the variables x, z (or y, z) much more complicated. On the positive side, the polynomial $m^2 + n^2$ is a norm form (higher-degree irreducible polynomials that do not have this property are out of reach at the moment), and the corresponding ring of integers one needs to work with in part of our analysis is the Gaussian integers, which is a principal ideal domain with finitely many units. All these things reduce the complexity of our argument.

In the case of (ii), handling the parameters y, z is similar to the case of (i). However, handling the parameters x, y is much harder, since the ring of integers associated with the polynomial $m^2 - 2n^2$ is $\mathbb{Z}[\sqrt{2}]$, which has infinitely many units; this adds another layer of problems to overcome. Finally, the variables x, z depend on two irreducible quadratic forms, which is the most difficult case to handle, and we can offer only conditional results in this case. The same problem appears in equation (iii); in this case, all pairs depend on two irreducible quadratic forms, so we can only prove conditional results in these cases.

2.3. Density regularity

It will be more convenient for us to prove stronger versions of Theorems 2.1 and 2.2, which deal with density regularity. Roughly, the stronger statements will state that every set with positive multiplicative density contains the pairs we are interested in, written in parametric form.

We recall some standard notions. A *multiplicative Følner sequence* in $\mathbb{N}$ is a sequence $\Phi = (\Phi_K)_{K=1}^\infty$ of finite subsets of $\mathbb{N}$ that is asymptotically invariant under dilation, in the sense that

$$\lim_{K \rightarrow \infty} \frac{|\Phi_K \cap (x \cdot \Phi_K)|}{|\Phi_K|} = 1 \quad \text{for all } x \in \mathbb{N}.$$

An example of a multiplicative Følner sequence is given by

$$\Phi_K := \left\{ \prod_{p \leq K} p^{a_p} : 0 \leq a_p \leq K \right\}, \quad K \in \mathbb{N}.$$

The *upper multiplicative density* of a set $\Lambda \subset \mathbb{N}$ with respect to a multiplicative Følner sequence $\Phi = (\Phi_K)_{K=1}^\infty$ is the quantity

$$\bar{d}_\Phi(\Lambda) := \limsup_{K \rightarrow \infty} \frac{|\Phi_K \cap \Lambda|}{|\Phi_K|}.$$

For example, for every $r \in \mathbb{N}$, the set $r\mathbb{Z} + j$ has multiplicative density 1 if $j = 0$ and multiplicative density 0 if $j = 1, \dots, r-1$, so a set with positive multiplicative density should contain many elements that are highly divisible. The set of squares has multiplicative density 0, while the set of integers $n \in \mathbb{N}$ such that $2^{2k} \mid n$ but $2^{2k+1} \nmid n$ for some $k \in \mathbb{Z}_+$ has multiplicative density $1/2$.

Definition 2.3. We say that the pair $P_1, P_2 \in \mathbb{Z}[m, n]$ is *good for density regularity* if whenever $\Lambda \subset \mathbb{N}$ satisfies $\bar{d}_\Phi(\Lambda) > 0$ for some multiplicative Følner sequence Φ , there exist $m, n \in \mathbb{N}$ such that $P_1(m, n)$ and $P_2(m, n)$ are distinct positive integers and

$$\bar{d}_\Phi((P_1(m, n))^{-1}\Lambda \cap (P_2(m, n))^{-1}\Lambda) > 0, \quad (2.1)$$

where $r^{-1}\Lambda := \{n \in \mathbb{N} : rn \in \Lambda\}$.

If the pair $P_1, P_2 \in \mathbb{Z}[m, n]$ is good for density regularity, then for any $\Lambda \subset \mathbb{N}$ that satisfies $\bar{d}_\Phi(\Lambda) > 0$ for some multiplicative Følner sequence Φ , the intersection $(P_1(m, n))^{-1}\Lambda \cap (P_2(m, n))^{-1}\Lambda$ has positive multiplicative density and in particular is non-empty for some $m, n \in \mathbb{N}$. Taking any $k \in (P_1(m, n))^{-1}\Lambda \cap (P_2(m, n))^{-1}\Lambda$, it follows that $kP_1(m, n), kP_2(m, n)$ are distinct and they both belong to Λ .

By considering appropriate parametrizations of solutions of (1.2) when ac or bc is a square, it turns out that in order to prove Theorem 2.1 it suffices to establish the following result.

Theorem 2.4. Suppose that for some $\alpha, \gamma \in \mathbb{N}$ and non-zero $\beta \in \mathbb{Z}$, we have

$$P_1(m, n) = \alpha m^2 + \beta n^2, \quad P_2(m, n) = \gamma mn.$$

Then the pair of polynomials P_1, P_2 is good for density regularity.

More generally, our argument shows density regularity if at least one of the two polynomials is reducible and not a square, and the two polynomials are not multiples of each other.

On the other hand, appropriate parametrizations of solutions of (1.2) when $(a + b)c$ is a square show that Theorem 2.2 is an immediate consequence of the following result.

Theorem 2.5. *Suppose that for some $\alpha, \beta, \alpha', \beta' \in \mathbb{Z}$, the polynomials*

$$P_1(m, n) = m^2 + \alpha mn + \beta n^2, \quad P_2(m, n) = m^2 + \alpha' mn + \beta' n^2$$

are distinct and irreducible, and Conjecture 3 holds for P_1, P_2 (see Section 6.5.2). Then the pair of polynomials P_1, P_2 is good for density regularity.

More generally, our argument shows density regularity if P_1, P_2 are irreducible quadratic forms, not multiples of each other, such that either $P_1(1, 0) = P_2(1, 0)$ or $P_1(0, 1) = P_2(0, 1)$. If neither of the last two conditions on the coefficients of m^2 and n^2 is met, then the corresponding pair of polynomials is often not partition regular.

We note that our arguments also give that the positivity property (2.1) occurs quite often; in fact, it holds for a set of $m, n \in \mathbb{N}$ with positive lower density (taken over the squares $[N] \times [N]$).

2.4. Other results

Our methodology is quite flexible and can be applied to other partition regularity problems of pairs of dilation-invariant patterns and related results. Here is a list of some additional examples:

- (i) (More general homogeneous quadratic equations) The methods of this article can be used to prove partition and density regularity of more general equations $P(x, y, z) = 0$, where P is a homogeneous quadratic form. For example, take $a, b \in \mathbb{N}$, $d \in \mathbb{Z}$, and consider the equation

$$ax^2 + by^2 + dxy = az^2. \quad (2.2)$$

Direct computation shows that (2.2) is satisfied when

$$x = k(bm^2 - an^2), \quad y = km(-dm + 2an), \quad z = k(bm^2 - dmn + an^2),$$

where $k, m, n \in \mathbb{Z}$. Using the remark following Theorem 2.4, we get that (2.2) is density regular with respect to x, y and y, z . We also get density regularity with respect to x, z if the polynomial $bm^2 - an^2$ is reducible, i.e., if ab is a square. The last assumption can be removed if we assume that Conjecture 3 in Section 6.5.2 holds; to see this, use the remark following Theorem 2.5.

- (ii) (Pythagorean triples on level sets of multiplicative functions) In [23, Theorem 1.7] we showed that for any completely multiplicative function $f: \mathbb{N} \rightarrow \mathbb{S}^1$ taking finitely many values, there exist $x, y, z \in \mathbb{N}$ such that

$$x^2 + y^2 = z^2 \quad \text{and} \quad f(x) = f(y) = f(z) = 1.$$

- (iii) (A question from [18]) Answering a question of Donoso–Le–Moreira–Sun from [18], we showed in [23, Theorem 1.8] (our argument also uses the main result from [56]) that the patterns $(km^2, k(n^2 + n))$ are density regular.
- (iv) (An extension of a result of Klurman–Mangerel) Recently, Charamaras–Mountakis–Tsinas [12] used a variant of the Q -trick, which we explain in Section 6, to show, among other things, that for every $a \in \mathbb{N}$ and every pair of completely multiplicative functions $f, g: \mathbb{N} \rightarrow \mathbb{S}^1$, we have

$$\liminf_{n \rightarrow \infty} |f(an + 1) - g(an)| = 0, \quad (2.3)$$

and we cannot replace 1 by 2 even when $a = 1$. When $f = g$, equation (2.3) was previously established by Klurman–Mangerel [43, Theorem 1.1] for $a = 1$ and for general $a \in \mathbb{N}$ by Donoso–Le–Moreira–Sun [18, Corollary 1.7].

3. A model problem and connection with multiplicative functions

3.1. An additive model problem

Our approach to solving the density regularity problems of Theorems 2.4 and 2.5 tries to model Furstenberg’s ergodic-theoretic proof of Sárközy’s theorem. Let us briefly describe this argument. The goal is to show that if Λ is a positive density subset of the integers, i.e.,

$$\overline{d}(\Lambda) := \limsup_{N \rightarrow \infty} \frac{|\Lambda \cap [N]|}{N} > 0,$$

then there exists $n \in \mathbb{N}$ such that

$$\overline{d}(\Lambda \cap (\Lambda - n^2)) > 0. \quad (3.1)$$

Using a correspondence principle of Furstenberg [24], it suffices to show that for every “additive” measure-preserving system (X, μ, T) , i.e., probability space $(X, \mathcal{X}, \mu)$, measure-preserving transformation $T: X \rightarrow X$, and measurable set A with $\mu(A) > 0$, there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n^2}A) > 0$. In fact, we show the stronger property

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \mu(A \cap T^{-n^2}A) > 0. \quad (3.2)$$

To prove (3.2), we have two ways to proceed. One is to first show that the rational Kronecker factor is characteristic (by appealing to the Hilbert space version of van der Corput’s lemma) and then to deal with the rational Kronecker factor of the system using the good divisibility properties of n^2 . Unfortunately, this approach cannot be modeled very well in our multiplicative setting, since our patterns use a

mixture of addition and multiplication which does not allow cancellation when using a multiplicative version of van der Corput's lemma.

The second way to prove (3.2) is to use a representation theorem for sequences of the form $n \mapsto \mu(A \cap T^{-n}A)$ or more generally positive density sequences in $\mathbb{Z}$. It follows from Herglotz's theorem that there exists a (positive) bounded Borel measure σ on $[0, 1)$ (depending on A) such that

$$\mu(A \cap T^{-n}A) = \int e(nt) d\sigma(t) \quad (3.3)$$

for every $n \in \mathbb{Z}$. Averaging over $n \in [N]$, taking the limit as $N \rightarrow \infty$, and using the ergodic theorem to lower bound the left side and the bounded convergence theorem to compute the right side, we deduce that $\sigma(\{0\}) \geq (\mu(A))^2 > 0$.

It follows that in order to establish (3.2), it suffices to show that if σ is a bounded (positive) measure on $[0, 1)$ such that $\sigma(\{0\}) = \delta > 0$ and $\hat{\sigma}(n) \geq 0$ for every $n \in \mathbb{N}$, then

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \int e(n^2 t) d\sigma(t) > 0. \quad (3.4)$$

The advantage of this new setting is that exponential sums along squares can be estimated efficiently, and this helps us analyze the left-hand side in (3.4). Indeed, it is a well-known consequence of Weyl's theorem that if t is irrational, we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} e(n^2 t) = 0. \quad (3.5)$$

Using the bounded convergence theorem, this essentially allows us to reduce matters to handling the “rational” part of the measure σ , which can be done using the divisibility properties of n^2 as follows: For $A \subset [0, 1)$, let

$$\sigma_A := \sigma|_A$$

and choose $Q \in \mathbb{N}$ so that

$$\sigma_{\mathbb{Q}}([0, 1) \setminus A_Q) \leq \frac{\delta^2}{2}, \quad \text{where } A_Q := \left\{0, \frac{1}{Q}, \dots, \frac{(Q-1)}{Q}\right\}. \quad (3.6)$$

Using the positivity property $\hat{\sigma}(n) \geq 0, n \in \mathbb{N}$, we get that (3.4) would follow from

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \int e(Q^2 n^2 t) d\sigma(t) > 0; \quad (3.7)$$

indeed, the limit in (3.4) is at least $1/Q$ times the limit in (3.7). Using (3.5) and the bounded convergence theorem, we get that it suffices to show that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \int e(Q^2 n^2 t) d\sigma_{\mathbb{Q}}(t) > 0. \quad (3.8)$$

Since $Q^2 t^2 = 0 \pmod{1}$ for every $t \in A_Q$, we have

$$\int e(Q^2 n^2 t) d\sigma_{A_Q}(t) = \int 1 d\sigma_{A_Q}(t) = \sigma(A_Q) \geq \sigma(\{0\}) = \delta^2 \quad \text{for every } n \in \mathbb{N}. \quad (3.9)$$

Combining (3.6) and (3.9), it follows that the limit in (3.8) is at least $\delta^2/2$, and the proof is complete.

We also note that the use of ergodic theory in this argument is rather superficial, since it was only employed to reduce the density positivity property (3.1) to the analytic positivity property (3.4). This step could also be done directly by showing that the sequence $n \mapsto d^*(\Lambda \cap (\Lambda - n))$ is positive definite, where d^* is a density that agrees with $\bar{d}$ on Λ and is defined along a subsequence of the integers so that all relevant limits exist.

3.2. Bird's eye view of the modified strategy

We plan to use the general principles of the strategy explained in the previous subsection to establish the density statements of Theorems 2.4 and 2.5, with some necessary modifications that we will briefly summarize now. In the following sections, we will give a much more detailed account of how to implement these modifications, and it will become clear why the exact form of the homogeneous polynomials we have to work with greatly affects the difficulty of the problem.

- (i) (Representation result) We can replace the “additive” measure-preserving system (X, μ, T) by a measure-preserving system (X, μ, T_n) that has multiplicative structure, but we opt for a more direct approach and represent directly some multiplicative densities that are naturally linked to our problem. The fact that these densities are dilation invariant and not translation invariant leads to a representation result analogous to (3.3), where the place of the linear exponential sequences $n \mapsto e(nt)$ is taken by sequences with multiplicative structure, i.e., sequences $f: \mathbb{N} \rightarrow \mathbb{S}^1$ that satisfy $f(mn) = f(m) \cdot f(n)$ for every $m, n \in \mathbb{N}$. See (3.10) for the exact identity.
- (ii) (Reduced analytic statement) Using the previous representation result, we reduce the proof of Theorems 2.4 and 2.5 to showing a positivity property analogous to (3.4), which involves averages of integrals of multiplicative functions evaluated over homogeneous polynomials in two variables. See Theorems 3.2 and 3.3 below for the exact statements. It is the case, however, that we have a much larger variety of multiplicative functions than we have linear exponential sequences, and studying their asymptotic behavior turns out to be a much more delicate and interesting task. Similar to the additive case, we consider two complementary cases, the “structured” and the

“random” cases, to handle the “major arcs” and “minor arcs” of the spectral measure.

- (iii) (Minor arcs) In the additive setting, we used Weyl’s theorem to show that for irrational numbers, we have the vanishing property (3.5). In the multiplicative setting, the place of irrational numbers is taken by the “aperiodic” multiplicative functions, for which we prove, using a rather involved argument, several vanishing statements about their correlations; see Section 4 for details.
- (iv) (Major arcs) In the additive setting, we were able to handle the part of the spectral measure σ that was supported on rational numbers by passing to a subprogression and then using identity (3.9). In the multiplicative setting, the place of rational numbers is taken by the “pretentious” multiplicative functions for which we are able to prove concentration estimates as a substitute for identity (3.9). The exact statements appear in Section 5.
- (v) (The Q -trick) In the additive setting, it was easy to combine the information about the major and minor arcs to simplify expression (3.7) to a point where the positivity property is obvious. In the multiplicative setting, the simpler expression we arrive at has no obvious positivity property, so it will be much less clear how to conclude. A key manipulation here is what we call the Q -trick, which allows us to obtain a property analogous to (3.8) for some highly divisible Q . We explain in Section 6 how this key maneuver is used to complete the proof.

3.3. From partition regularity to multiplicative functions

The starting point for the proofs of Theorems 2.4 and 2.5 is to reformulate them as positivity properties involving integrals of averages of completely multiplicative functions. As we also mentioned before, this key step can only be performed for partition and density regularity results of pairs and has no useful analog we know of for triples (though see Section 7.3 for an ergodic reformulation).

3.3.1. Multiplicative functions. A function $f: \mathbb{N} \rightarrow \mathbb{U}$, where $\mathbb{U}$ is the complex unit disk, is called *multiplicative* if

$$f(mn) = f(m) \cdot f(n) \quad \text{whenever } (m, n) = 1.$$

If necessary, we extend f to an even function in $\mathbb{Z}$ by letting $f(0) := 0$ and $f(-n) := f(n)$ for $n \in \mathbb{N}$. It is called *completely multiplicative* if the previous identity holds for all $m, n \in \mathbb{N}$. We let

$$\mathcal{M} := \{f: \mathbb{N} \rightarrow \mathbb{S}^1 \text{ is a completely multiplicative function}\},$$

where $\mathbb{S}^1$ is the unit circle. Throughout, we assume that $\mathcal{M}$ is equipped with the topology of pointwise convergence; then $\mathcal{M}$ is a metrizable compact space with this topology.

3.3.2. Reduction to a statement about multiplicative functions. We will use a representation result of Bochner–Herglotz for positive definite functions on locally compact Abelian groups. We need it only for the discrete group $\mathbb{Q}_+$ with multiplication, in which case we can identify $\mathcal{M}$ with the Pontryagin dual of $(\mathbb{Q}_+, \times)$. We say that $A: \mathbb{Q}_+ \rightarrow \mathbb{C}$ is positive definite on $(\mathbb{Q}_+, \times)$ if for every $n \in \mathbb{N}$, all $r_1, \dots, r_n \in \mathbb{Q}^+$, and all $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ we have

$$\sum_{i,j=1}^n \lambda_i \overline{\lambda_j} A(r_i r_j^{-1}) \geq 0.$$

Theorem 3.1 (Bochner–Herglotz). *Let $A: \mathbb{Q}_+ \rightarrow \mathbb{C}$ be positive definite on $(\mathbb{Q}_+, \times)$. Then there exists a finite Borel measure σ on $\mathcal{M}$ such that*

$$A\left(\frac{r}{s}\right) = \int_{\mathcal{M}} f(r) \cdot \overline{f(s)} d\sigma(f) \quad \text{for all } r, s \in \mathbb{N}.$$

To use this representation result in our setting, we can argue as follows (an alternative way is via ergodic theory). Given a Følner sequence $\Phi = (\Phi_N)$ of subsets of $\mathbb{N}$, we can pass to a subsequence $\Phi' = (\Phi'_N)$ along which $d_{\Phi'}(\Lambda) = d_{\Phi}(\Lambda)$, and the limits

$$d_{\Phi'}((r^{-1}\Lambda) \cap (s^{-1}\Lambda)) := \lim_{N \rightarrow \infty} \frac{|(r^{-1}\Lambda) \cap (s^{-1}\Lambda) \cap \Phi'_N|}{|\Phi'_N|}$$

exist for all $r, s \in \mathbb{Z}$. Then the sequence $r/s \mapsto d_{\Phi'}((r^{-1}\Lambda) \cap (s^{-1}\Lambda))$ from $(\mathbb{Q}_+, \times)$ to $[0, 1]$ is well defined (we use that $d_{\Phi'}$ is invariant under dilations), and a direct computation shows that it is positive definite. Using Theorem 3.1, we get that there exists a finite Borel measure σ on $\mathcal{M}$ such that

$$d_{\Phi'}((r^{-1}\Lambda) \cap (s^{-1}\Lambda)) = \int_{\mathcal{M}} f(r) \cdot \overline{f(s)} d\sigma(f) \quad \text{for all } r, s \in \mathbb{N}. \quad (3.10)$$

Furthermore, it is not hard to show (see [21, Section 10.2]) that

$$\sigma(\{1\}) \geq (d_{\Phi'}(\Lambda))^2.$$

Using the previous two properties, it is easy to see that in order to prove Theorem 2.4, it suffices to prove the following result.

Theorem 3.2. *Let σ be a bounded Borel measure on $\mathcal{M}$ such that $\sigma(\{1\}) > 0$ and*

$$\int_{\mathcal{M}} f(r) \cdot \overline{f(s)} d\sigma(f) \geq 0 \quad \text{for every } r, s \in \mathbb{N}. \quad (3.11)$$

Let also $\alpha, \gamma \in \mathbb{N}$ and non-zero $\beta \in \mathbb{Z}$. Then

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \int_{\mathcal{M}} f(\alpha m^2 + \beta n^2) \cdot \overline{f(\gamma mn)} d\sigma(f) > 0. \quad (3.12)$$

In order to prove Theorem 2.5, it suffices to establish the following result.

Theorem 3.3. *Let σ be as in Theorem 3.2 and $\alpha, \beta, \alpha', \beta' \in \mathbb{Z}$ be such that*

$$P_1(m, n) = m^2 + \alpha mn + \beta n^2, \quad P_2(m, n) = m^2 + \alpha' mn + \beta' n^2$$

are distinct and irreducible quadratic forms. Suppose that either Conjecture 3 in Section 6.5.2 holds for P_1, P_2 or the measure σ is supported on the class of pretentious multiplicative functions (see Section 5.1). Then

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \int_{\mathcal{M}} f(P_1(m, n)) \cdot \overline{f(P_2(m, n))} d\sigma(f) > 0. \quad (3.13)$$

Both results were established in [22], with special cases previously obtained in [21, 23].

One may wonder about our choice of averaging over $m, n \in \mathbb{N}$ in (3.12) and (3.13). Why not take iterated limits $\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]}$, $\liminf_{M \rightarrow \infty} \mathbb{E}_{m \in [M]}$ or perhaps take multiplicative averages over m, n ? Taking iterated limits would lead to much harder problems for “minor arc” multiplicative functions. Although in some cases the required vanishing property can be obtained for logarithmic averages using the work of Tao–Teräväinen [57], the exceptional class of “minor arc” multiplicative functions for which this method is inapplicable would render this argument not very useful. Finally, taking multiplicative averages over the parameters m, n leads to problems that we are not able to handle, both for “major arc” and “minor arc” multiplicative functions.

3.4. Proof strategy

In order to establish the necessary positivity properties of Theorems 3.2 and 3.3, we use the theory of completely multiplicative functions and proceed in two independent steps, each using a distinctly different methodology.

- (i) (Minor arcs) We first study the part of σ that is supported on aperiodic multiplicative functions. The key step here is to establish Gowers uniformity properties of arbitrary aperiodic multiplicative functions and then use

them show that the averages in (3.12) vanish (this last step is trickier when $P(m, n) := \alpha m^2 + \beta n^2$ is irreducible). To prove the uniformity properties, we combine the inverse theorem of Green–Tao–Ziegler [35] with some quite delicate equidistribution results of variable nilsequences. Our approach is described in more detail in Section 4.

- (ii) (Major arcs) Next, we study the complementary case, covering the part of σ that is supported on pretentious multiplicative functions (see Definition 5.1). In this case, we can establish concentration estimates, a feature of pretentious multiplicative functions that is not shared by the aperiodic ones. The details are given in Section 5. We use these concentration estimates to simplify substantially the averages in (3.12) when the range of m, n is restricted to a suitable positive density subset of $\mathbb{N} \times \mathbb{N}$ depending on some parameter Q . We then show that a suitable multiplicative average, over the parameter Q , of these simplified expressions is zero for all but an exceptional set of multiplicative functions and it is positive for the rest. We call this the Q -trick, and it easily implies the required positivity. The details are given in Section 6.

4. Aperiodic multiplicative functions: Vanishing of correlations

4.1. Definition and characterization

The first class of multiplicative functions that will play a crucial role in our arguments is the aperiodic ones.

Definition 4.1. The multiplicative function $f: \mathbb{N} \rightarrow \mathbb{U}$ is *aperiodic* if for every $a, b \in \mathbb{N}$ we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} f(an + b) = 0. \quad (4.1)$$

Sometimes in the literature, the term *non-pretentious* is used instead of the term aperiodic. It can be shown that $f: \mathbb{N} \rightarrow \mathbb{U}$ is aperiodic if and only if

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} f(n) \cdot \chi(n) = 0 \quad \text{for every Dirichlet character } \chi. \quad (4.2)$$

Examples of multiplicative functions that are aperiodic are the Liouville and the Möbius function; this follows from a variant of the prime number theorem on arithmetic progressions.

A rather amazing and very useful fact for our purposes is that correlations of aperiodic multiplicative functions have very strong vanishing properties, which we will describe in the following subsections.

4.2. Gowers uniformity and inverse theorem

We will define some seminorms on $\ell^\infty(\mathbb{N})$ that will be crucial for our study. Roughly, they are given by the limit as $N \rightarrow \infty$ of the Gowers norms of the sequence, restricted to the interval $[N]$.

Definition 4.2 (Gowers norms on $\mathbb{Z}_N$ [25]). Let $N \in \mathbb{N}$ and $a: \mathbb{Z}_N \rightarrow \mathbb{C}$. For $s \in \mathbb{N}$, the *Gowers $U^s(\mathbb{Z}_N)$ -norm* $\|a\|_{U^s(\mathbb{Z}_N)}$ is defined inductively as follows: For every $h \in \mathbb{Z}_N$, we write $a_h(n) := a(n + h)$. We let

$$\|a\|_{U^1(\mathbb{Z}_N)} := |\mathbb{E}_{n \in \mathbb{Z}_N} a(n)|,$$

and for every $s \geq 1$, we let

$$\|a\|_{U^{s+1}(\mathbb{Z}_N)} := (\mathbb{E}_{h \in \mathbb{Z}_N} \|a \cdot \bar{a}_h\|_{U^s(\mathbb{Z}_N)}^{2^s})^{\frac{1}{2^{s+1}}}.$$

For example,

$$\|a\|_{U^2(\mathbb{Z}_N)}^4 = \mathbb{E}_{n, h_1, h_2 \in \mathbb{Z}_N} a(n) \overline{a(n + h_1)} \overline{a(n + h_2)} a(n + h_1 + h_2),$$

and for the $U^s(\mathbb{Z}_N)$ norms for $s \geq 3$, a similar but more complicated closed formula can be given.

Definition 4.3 (Gowers seminorms on $\mathbb{N}$). If $a: \mathbb{N} \rightarrow \mathbb{C}$ is a bounded sequence, we let

$$\|a\|_{U^s(\mathbb{N})} = \limsup_{N \rightarrow \infty} \|a_N\|_{U^s(\mathbb{Z}_N)},$$

where $a_N: \mathbb{Z}_N \rightarrow \mathbb{C}$ is the periodic extension of $a \cdot \mathbf{1}_{[N]}$ to $\mathbb{Z}_N$.

It can be shown that $\|\cdot\|_{U^s(\mathbb{N})}$ satisfies the triangle inequality, thus defining a seminorm on $\ell^\infty(\mathbb{N})$, and for every $s \in \mathbb{N}$, we have

$$\|a\|_{U^s(\mathbb{N})} \leq \|a\|_{U^{s+1}(\mathbb{N})}.$$

We say that the bounded sequence $a: \mathbb{N} \rightarrow \mathbb{C}$ is U^s -uniform if

$$\|a\|_{U^s(\mathbb{N})} = 0.$$

For example, if α is an irrational number and $s \in \mathbb{N}$, it can be shown that the sequence $(e(n^s \alpha))$ is U^s -uniform but not U^{s+1} -uniform, while if α is a rational non-integer, it is U^1 -uniform but not U^2 -uniform. It can also be shown that if c is a non-integer positive real number, then the sequence $(e(n^c))$ is U^s -uniform for all $s \in \mathbb{N}$.

Using Fourier inversion and the Parseval identity in $\mathbb{Z}_N$, for $N \in \mathbb{N}$, it can be shown that if $a: \mathbb{N} \rightarrow \mathbb{U}$ is a sequence, then

$$\|a\|_{U^2(\mathbb{N})} \leq \limsup_{N \rightarrow \infty} \max_{\xi \in \mathbb{Z}_N} |\widehat{a_N}(\xi)|^{\frac{1}{2}},$$

where the Fourier transform is taken in $\mathbb{Z}_N$. It is easy to deduce from this the equivalence

$$\|a\|_{U^2(\mathbb{N})} = 0 \iff \lim_{N \rightarrow \infty} \sup_{\alpha \in \mathbb{R}} |\mathbb{E}_{n \in [N]} a(n) \cdot e(n\alpha)| = 0. \quad (4.3)$$

In order to give a generalization of this result that covers U^s -uniformity for $s \geq 3$, we have to first account for additional obstructions to uniformity. For example, when $s = 3$, it can be seen that

$$\|a\|_{U^3(\mathbb{N})} > 0 \quad \text{when } a(n) = e(n^2\alpha + n\beta) \text{ or } a(n) = e([n\alpha]n\beta), \quad (4.4)$$

where $\alpha, \beta \in \mathbb{R}$. These turn out to be examples of nilsequences, which are defined as follows: If X is an s -step nilmanifold, i.e., $X = G/\Gamma$, where G is an s -step nilpotent Lie group and Γ is a discrete cocompact subgroup, we let

$$a(n) = F(g^n \cdot e_X), \quad n \in \mathbb{N},$$

where $g \in G$, $e_X := \Gamma$, and $F \in C(X)$. Moreover, to cover the second example in (4.4), we must allow a function F that is not continuous but is Riemann integrable in the orbit closure of g . It is an amazing fact that this class of sequences suffices to give a generalization of (4.3) that covers U^s -uniformity. This is a consequence of the inverse theorem of Green–Tao–Ziegler [35], first proved in [32] for $s = 3$, which is the only case we will need for applications to partition regularity. We state it in the following convenient form.

Theorem 4.4. *Let $a: \mathbb{N} \rightarrow \mathbb{U}$ be a sequence and $s \in \mathbb{N}$. Then the following two properties are equivalent:*

- (i) $\|a\|_{U^{s+1}(\mathbb{N})} = 0$.
- (ii) *For every s -step nilmanifold $X = G/\Gamma$ and $F \in C(X)$, we have*

$$\lim_{N \rightarrow \infty} \sup_{g \in G} |\mathbb{E}_{n \in [N]} a(n) \cdot F(g^n \cdot e_X)| = 0.$$

Moreover, in checking the second condition, one can assume that F is smooth, and using what is known as the “vertical Fourier decomposition”, one can further assume that F is a non-trivial vertical nilcharacter, i.e., assuming that G is s -step

nilpotent but not $(s - 1)$ -step nilpotent, we have that there exists $\chi: X_s \rightarrow \mathbb{S}^1$, a non-trivial character of the compact Abelian group $X_s := G_s / (G_s \cap \Gamma)$, such that

$$F(ux) = \chi(u) \cdot F(x)$$

for every $x \in X$ and $u \in G_s$. It easily follows from this that $\int F \, dm_X = 0$. We note that in the case where $X = \mathbb{T}$, non-trivial vertical nilcharacters are functions of the form $F(x) = e(kx)$ for some non-zero $k \in \mathbb{Z}$.

We will see in the next subsection how one can use this result to establish the U^s -uniformity of aperiodic multiplicative functions.

4.3. Gowers uniformity of aperiodic multiplicative functions

A very strong and perhaps a priori unexpected property is that aperiodic multiplicative functions are Gowers uniform of all orders.

To verify that they are U^2 -uniform, we will use the following orthogonality criterion due to Daboussi–Kátai [15, 40].

Theorem 4.5. *Let $a: \mathbb{N} \rightarrow \mathbb{U}$ be a sequence such that*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} a(pn) \cdot \overline{a(qn)} = 0 \quad (4.5)$$

for all distinct primes p, q . Then for every multiplicative function $f: \mathbb{N} \rightarrow \mathbb{U}$, we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} f(n) \cdot a(n) = 0.$$

Applying Theorem 4.5 for $a(n) := e(n\alpha)$, it follows that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} f(n) \cdot e(n\alpha) = 0 \quad (4.6)$$

for every irrational α . Since aperiodic multiplicative functions, by definition, do not correlate with $e(n\alpha)$ for α rational, it follows that for those multiplicative functions, (4.6) holds for every $\alpha \in \mathbb{R}$. It is only slightly more difficult to show that a similar property holds even if we allow the phase α to depend on N [21, Lemma 9.1], namely

$$\lim_{N \rightarrow \infty} \sup_{\alpha \in \mathbb{R}} |\mathbb{E}_{n \in [N]} f(n) \cdot e(n\alpha)| = 0. \quad (4.7)$$

A direct consequence of this identity and the characterization of U^2 -uniformity (see (4.3)) is that

$$\|f\|_{U^2(\mathbb{N})} = 0.$$

It is also true, but much harder to prove, that a similar property holds for higher-order uniformity. This was first done for the Liouville and the Möbius functions for $s = 3$

by Green–Tao [33] and then by Green–Tao–Ziegler [34, 35] for general $s \in \mathbb{N}$. Their approach used a Type I and Type II sums analysis to verify property (ii) of Theorem 4.4, and this required explicit quantitative control for averages of multiplicative functions along arithmetic progressions, not shared by general aperiodic multiplicative functions (and the exceptional class is too large to be ignored in our applications). Using a somewhat different argument, the previous results were extended in a joint paper by the author and Host in [21, Theorem 2.5] to general aperiodic multiplicative functions (see also Matthiesen [47] for an alternative approach to a related result).

Theorem 4.6. *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be an aperiodic multiplicative function. Then for every $s \in \mathbb{N}$, we have*

$$\|f\|_{U^s(\mathbb{N})} = 0.$$

Let us briefly sketch some key ideas of the proof of this result. We start by using Theorem 4.4. It reduces the problem to establishing a strengthening of (4.7) where the place of $e(n\alpha)$ is taken by nilsequences. Namely, we need to show that if $X = G/\Gamma$ is a fixed s -step nilmanifold and $F \in C(X)$ is a non-trivial vertical nilcharacter, then

$$\lim_{N \rightarrow \infty} \sup_{b \in G} |\mathbb{E}_{n \in [N]} f(n) \cdot F(b^n \cdot e_X)| = 0. \quad (4.8)$$

Verifying this property is a rather laborious task, so let us just explain how it works in the much simpler setting where we do not take the supremum over b , the nilmanifold X is connected, and the sequence $(b^n \cdot e_X)$ is equidistributed on X . So in this case, we use the above-mentioned orthogonality criterion of Daboussi–Kátai for $a(n) := F(b^n \cdot e_X)$. We get that it suffices to verify the vanishing property (4.5) for all distinct primes p, q or, equivalently, that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} F(b^{pn} \cdot e_X) \cdot \overline{F(b^{qn} \cdot e_X)} = 0. \quad (4.9)$$

It can be shown that the sequence

$$((b^{pn} \cdot e_X, b^{qn} \cdot e_X))_{n \in \mathbb{N}}$$

is equidistributed in a connected subnilmanifold $Y = H/\Delta$ of $X \times X$. Equation (4.9) would thus follow if we showed that

$$\int F(x) \cdot \overline{F(x')} dm_Y(x, x') = 0. \quad (4.10)$$

This is not a simple task, since the structure of the possible nilmanifolds Y depends on b in a way that is not easy to determine explicitly. The key property that allows us

to prove (4.10) is that no matter how complicated Y may be, it always satisfies the following:

$$\{(g^p, g^q): g \in G\} \subset H \cdot (G_2 \times G_2).$$

This follows by using that the sequences $(b^{p^n} \cdot e_X)$ and $(b^{q^n} \cdot e_X)$ are both equidistributed in X . By taking iterated commutators of elements on the left, we deduce the following key property:

$$(u^{p^s}, u^{q^s}) \in H \quad \text{for every } u \in G_s.$$

This invariance property easily implies that $(x, x') \mapsto F(x) \cdot \overline{F(x')}$ is a vertical nilcharacter of Y with non-zero frequency, so (4.10) holds.

If the place of the sequence b^n takes a more general polynomial sequence, say $a^n b^{n^2}$ for some $a, b \in G$, we first establish that there exist normal subgroups G^1, G^2 of G such that $G^1 \cdot G^2 = G$ and

$$\{(g_1^p g_2^{p^2}, g_1^q g_2^{q^2}): g_1 \in G^1, g_2 \in G^2\} \subset H \cdot (G_2 \times G_2),$$

and by taking iterated commutators of elements on the left, we deduce that

$$\text{the set } \{u \in G_s: (u^{p^j}, u^{q^j}) \in H \text{ for some } j \in \mathbb{N}\} \text{ generates } G_s. \quad (4.11)$$

From this, we can again prove (4.10).

The actual argument is substantially more complicated because in order to verify (4.8), we have to prove a variant of (4.9) involving variable polynomial sequences, in which case the required invariance property, analogous to (4.11), is much harder to identify and prove. However, the previous summary accurately summarizes the skeleton of the proof of Theorem 4.6.

4.4. Vanishing of correlations

An important component in the proof of Theorem 3.2, which covers the case of Pythagorean pairs $(m^2 \pm n^2, 2mn)$, is that for aperiodic multiplicative functions, several correlations vanish, a result obtained in [21, Theorem 2.6].

Theorem 4.7. *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be an aperiodic multiplicative function. Then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m, n \in [N]} f(m^2 - n^2) \cdot \overline{f(mn)} = 0. \quad (4.12)$$

To prove this, one first uses some standard maneuvering to bound the left side by a multiple of $\|f\|_{U^3(\mathbb{Z})}$, which vanishes by Theorem 4.6. A similar argument leads to the following more general result: If $f_1: \mathbb{N} \rightarrow \mathbb{U}$ is an aperiodic multiplicative function and $f_2, \dots, f_\ell: \mathbb{N} \rightarrow \mathbb{U}$ are arbitrary sequences, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m, n \in [N]} f_1(L_1(m, n)) \cdots f_\ell(L_\ell(m, n)) = 0$$

for all linear forms $L_1, \dots, L_\ell: \mathbb{N}^2 \rightarrow \mathbb{N}$ such that L_1 is not a multiple of $L_2, \dots, L_\ell$.

To handle Pythagorean pairs of the form $(2mn, m^2 + n^2)$, we need the following result.

Theorem 4.8. *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be an aperiodic completely multiplicative function. Then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} f(mn) \cdot \overline{f(m^2 + n^2)} = 0. \quad (4.13)$$

To prove this, we first represent the sequence $(m, n) \mapsto f(m^2 + n^2)$ as a multiplicative function on the Gaussian integers. To do this, we let $g: \mathbb{Z}[i] \rightarrow \mathbb{U}$ by $g(z) := \overline{f(|z|^2)}$, and note that for $z = m + in$, we have $g(z) = \overline{f(m^2 + n^2)}$. Then (4.13) takes the following form:

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} f(\Re(z)) \cdot f(\Im(z)) \cdot g(z) = 0.$$

Using a variant of Theorem 4.5 for bounded sequences on the Gaussian integers, it turns out that the last identity follows if we show that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{z \in R_N} f(\Re(pz)) \cdot f(\Im(pz)) \cdot \overline{f(\Re(qz))} \cdot \overline{f(\Im(qz))} = 0$$

holds for all but finitely many distinct non-real Gaussian primes p, q , where

$$R_N := \{z = m + in : m, n \in [N]\}.$$

Using some elementary reasoning on the ring $\mathbb{Z}[i]$, it is not hard to deduce that this follows from

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} f(L_1(m, n)) \cdot f(L_2(m, n)) \cdot \overline{f(L_3(m, n))} \cdot \overline{f(L_4(m, n))} = 0,$$

where f is aperiodic and $L_1, L_2, L_3, L_4: \mathbb{N}^2 \rightarrow \mathbb{N}$ are pairwise independent linear forms (recall that we extend f to the negative integers as an even function if needed). Using standard arguments, this follows from the U^3 -uniformity of f , i.e., the $s = 3$ case of Theorem 4.6.

4.5. More general binary quadratic forms

The results of the previous subsection can be extended to a more general setting, in which one considers the correlations of an aperiodic completely multiplicative function involving several linear forms and at most one irreducible quadratic form. A good representative example is the identity

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} f(mn) \cdot \overline{f(m^2 + dn^2)} = 0,$$

where $d \in \mathbb{Z}$ is not a square of an integer. The case in which the expression $m^2 + dn^2$ is replaced by an arbitrary irreducible binary quadratic form $P(m, n)$ is only marginally more complicated.

Suppose first that $d > 0$. If $d \equiv 1, 2 \pmod{4}$, then $m^2 + dn^2$ is a norm form on the ring of integers $\mathbb{Z}[\sqrt{-d}]$, which is a unique factorization domain. The argument employed in the case of $d = 1$ can be applied without significant alterations to this case, even in the context of a non-principal ideal domain. If $d \equiv 3 \pmod{4}$, then $\mathbb{Z}[\sqrt{-d}]$ is not a unique factorization domain, and this causes problems for our argument. When, say $d = 3$, this issue can be circumvented by working with the binary form $m^2 + mn + n^2$ instead of the form $m^2 + 3n^2$. The advantage now is that this is a norm form on the ring $\mathbb{Z}[(1 + \sqrt{-3})/2]$, which is a unique factorization domain. The identity

$$m^2 + 3n^2 = (m - n)^2 + (m - n)(2n) + (2n)^2$$

and a “change of variables” argument are then employed to transfer results from one binary form to the other.

The case $d < 0$ presents a more substantial challenge. Let us assume that $d = -2$ and explain what the issue is in this case. An element $z \in \mathbb{Z}[\sqrt{2}]$ is said to be *C-regular* if, for all $N \in \mathbb{N}$, the following condition is satisfied:

$$z^{-1}R_N \subset R_{C|\mathcal{N}(z)|^{-\frac{1}{2}}N},$$

where $R_N := \{m + n\sqrt{2} : m, n \in [N]\}$. In order to prove a useful variant of Theorem 4.5, we would like to know that for some fixed $C > 0$, all elements of the ring $\mathbb{Z}[\sqrt{2}]$ are *C-regular*. Unfortunately, this property does not hold, for the reason that $\mathbb{Z}[\sqrt{2}]$ has infinitely many units, so every element has infinitely many associates, and not all of them have good regularity properties. What allows us to overcome this obstacle is a result proved by Sun in [55], namely that for a fixed $C > 0$, every element has a *C-regular* associate. This enables us to prove a variant of Theorem 4.5 for $\mathbb{Z}[\sqrt{2}]$ and more general quadratic integer rings, which is applicable for our purposes.

By combining the previous ideas, we prove the following result in [22, Theorem 3.1].

Theorem 4.9. *Let $s \in \mathbb{N}$ and $f_1, \dots, f_s, g: \mathbb{N} \rightarrow \mathbb{S}^1$ be completely multiplicative functions that are extended to even sequences in $\mathbb{Z}$ and suppose that f_1 is aperiodic. Let also $L_1, \dots, L_s$ be linear forms in two variables with integer coefficients and suppose that either $s = 1$ and L_1 is non-trivial or $s \geq 2$ and the forms L_1, L_j are linearly independent for $j = 2, \dots, s$. Finally, suppose that the form*

$$P(m, n) := \alpha m^2 + \beta mn + \gamma n^2,$$

where $\alpha, \beta, \gamma \in \mathbb{Z}$, is irreducible. Then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \prod_{j=1}^s f_j(L_j(m, n)) \cdot g(P(m, n)) = 0.$$

For the partition regularity applications we have in mind, we only use the case $s = 2$.

5. Pretentious multiplicative functions: Concentration estimates

Now that we have at our disposal the vanishing property of the correlations of aperiodic multiplicative functions given by Theorem 4.9, we turn our attention to the study of those that are not aperiodic. In this case, we prove an approximate periodicity property that is analytically expressed in the form of the concentration estimates given in Theorems 5.5, 5.6, and 5.8. In the non-aperiodic case, this will allow us to substantially simplify the averages that appear in Theorems 3.2 and 3.3.

5.1. Pretentious distance and characterization of aperiodicity

Our first goal is to give a useful characterization of non-aperiodic multiplicative functions. The following are two potential obstructions to aperiodicity:

- (i) (Dirichlet characters) A *Dirichlet character* is a periodic completely multiplicative function. Then $\chi(Qn + 1) = 1$ whenever Q is a multiple of a period of χ and (4.1) fails for $a = Q$ and $b = 1$.
- (ii) (Archimedean characters) An *Archimedean character* is a completely multiplicative function of the form $n \mapsto n^{it}$, $n \in \mathbb{N}$, where $t \in \mathbb{N}$. Then $\mathbb{E}_{n \in [N]} n^{it}$ is asymptotically equal to $N^{it}/(1 + it)$, and so (4.1) fails for $a = 1$ and $b = 0$.

It is a surprising and very useful fact that these are the only obstructions to aperiodicity. A completely multiplicative function that is not aperiodic is, in a sense that will be made precise in Theorem 5.2 below, close to a product of a Dirichlet character and an Archimedean character.

Following Granville and Soundararajan [28–30], we define the notion of pretentiousness and a related distance between multiplicative functions.

Definition 5.1. If $f, g: \mathbb{N} \rightarrow \mathbb{U}$ are multiplicative functions, we define their distance as

$$\mathbb{D}^2(f, g) := \sum_{p \in \mathbb{P}} \frac{1}{p} (1 - \Re(f(p) \cdot \overline{g(p)})).$$

(Note that for $f, g: \mathbb{N} \rightarrow \mathbb{S}^1$, we have $\mathbb{D}^2(f, g) = 1/2 \sum_{p \in \mathbb{P}} 1/p |f(p) - g(p)|^2$.)

It can be shown (see [29] or [30, Section 2.1.1]) that $\mathbb{D}$ satisfies the triangle inequality

$$\mathbb{D}(f, g) \leq \mathbb{D}(f, h) + \mathbb{D}(h, g)$$

for all $f, g, h: \mathbb{N} \rightarrow \mathbb{U}$. Also, for all $f_1, f_2, g_1, g_2: \mathbb{N} \rightarrow \mathbb{U}$, we have (see [28, Lemma 3.1])

$$\mathbb{D}(f_1 f_2, g_1 g_2) \leq \mathbb{D}(f_1, g_1) + \mathbb{D}(f_2, g_2). \quad (5.1)$$

We say that f *pretends to be* g and write $f \sim g$ if $\mathbb{D}(f, g) < +\infty$. It follows from (5.1) that if $f_1 \sim g_1$ and $f_2 \sim g_2$, then $f_1 f_2 \sim g_1 g_2$.

The next result is a direct consequence of the characterization of aperiodicity given by (4.2) and the Wirsing–Halász mean value theorem [38], which gives necessary and sufficient conditions for a multiplicative function to have mean value 0.

Theorem 5.2. *A completely multiplicative function $f: \mathbb{N} \rightarrow \mathbb{U}$ is not aperiodic if and only if there exist $t \in \mathbb{R}$ and a Dirichlet character χ such that $f \sim \chi \cdot n^{it}$, i.e., $\mathbb{D}(f, \chi \cdot n^{it}) < +\infty$.*

This motivates the following definition.

Definition 5.3. We say that f is *pretentious* if $f \sim \chi \cdot n^{it}$ for some $t \in \mathbb{R}$ and Dirichlet character χ .

The value of t is uniquely determined; this follows from (5.1) and the fact that $n^{it} \not\sim \chi$ for every non-zero $t \in \mathbb{R}$ and Dirichlet character χ (see, for example, [30, Corollary 11.4] or [29, Proposition 7]).

It follows from Theorem 5.2 that whenever a completely multiplicative function $f: \mathbb{N} \rightarrow \mathbb{U}$ is not aperiodic, then it is pretentious.

5.2. Examples of pretentious multiplicative functions

We give several illustrative examples of pretentious multiplicative functions in order to cover all the possible varying behaviors that are relevant to our study. In the examples that follow, we describe the multiplicative function only on prime numbers and assume that it is extended to a completely multiplicative function on the positive integers.

- (i) (Modified Dirichlet characters) If χ is a Dirichlet character, then $\chi(p) = 0$ if and only if p divides the minimal period of χ and $\chi(p)$ is a root of

unity otherwise. We let $\tilde{\chi}: \mathbb{N} \rightarrow \mathbb{S}^1$ be the completely multiplicative function satisfying

$$\tilde{\chi}(p) := \begin{cases} \chi(p) & \text{if } \chi(p) \neq 0, \\ 1 & \text{if } \chi(p) = 0. \end{cases}$$

Then $\tilde{\chi}$ is not necessarily periodic and $\tilde{\chi} \sim \chi$.

- (ii) (Finite support) Let $f(p_1) = \cdots = f(p_\ell) := -1$ and $f(p) = 1$ for $p \notin \{p_1, \dots, p_\ell\}$. Then $f \sim 1$.
- (iii) (Infinite support 1-pretentious) Let $f(p) := -1$ for $p \in \mathbb{P}_0$ and $f(p) = 1$ for $p \notin \mathbb{P}_0$, where $\mathbb{P}_0 \subset \mathbb{P}$ is infinite and $\sum_{p \in \mathbb{P}_0} 1/p < +\infty$. Then $f \sim 1$.
- (iv) (1-Pretentious oscillatory) $f(p) := e(1/\log \log p)$, $p \in \mathbb{P}$. Then $f \sim 1$, but it turns out that f does not have a mean value, in fact, $\mathbb{E}_{n \in [N]} f(n)$ is asymptotically equal to $e(w(N))$, where $w(N) = c \log \log \log N$ for some $c > 0$.
- (v) (Archimedean factor) $f(n) := n^{it} \cdot g(n)$ for some $t \in \mathbb{R}$ and g as in the previous examples. Then $f \sim \chi \cdot n^{it}$ if g is as in (i) and $f \sim n^{it}$ if g is as in (ii)–(iv).

We mention some features of the previous examples that will motivate the concentration results to be proved in the next subsections.

In examples (i) and (ii), there exists Q_0 such that if $Q \mid Q_0$, then

$$f(Qn + 1) = 1 \quad \text{for every } n \in \mathbb{N}.$$

Indeed, take Q_0 be the period of χ in the first case and $Q_0 := p_1 \cdots p_\ell$ in the second.

In example (iii), let $\varepsilon > 0$ and $K = K(\varepsilon)$ be such that $\sum_{p \in \mathbb{P}_0, p > K} 1/p < \varepsilon$. For $Q_0 := \prod_{p \leq K} p$, we have that if $Q_0 \mid Q$, then $\bar{d}(n \in \mathbb{N}: f(Qn + 1) \neq 1) \leq \varepsilon$, and hence

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} |f(Qn + 1) - 1| \ll \varepsilon.$$

In example (iv), let $\varepsilon > 0$ and $K := K(\varepsilon)$ be such that $\sum_{p \in \mathbb{P}, p > K} 1/p(1 - \Re(f(p))) < \varepsilon$. For $Q_0 = \prod_{p \leq K} p$, we have that if $Q_0 \mid Q$, then

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} |f(Qn + 1) - e(w(N))| \ll \varepsilon,$$

where $w(N)$ increases to infinity slowly, in fact, like $c \log \log \log N$ for some $c > 0$.

In example (v), we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} |f(Qn + 1) - (Qn)^{it} \cdot g(Qn + 1)| = 0.$$

Combining this with the previous estimates for g , we get analogous results for f .

The general principle that can be derived from these examples is that if $f \sim \chi$, then for highly divisible values of Q , the values $f(Qn + 1)$ concentrate around 1 or a slowly growing oscillatory factor. The next subsection makes this principle explicit. Later, we also give variants of it that apply to arbitrary binary quadratic forms in place of n .

5.3. Linear concentration estimates: Statements

To get a sense of the concentration estimates we will be using in a simpler setting, let us first consider the case where a pretentious multiplicative function takes on finitely many values on the primes.

Theorem 5.4. *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be a pretentious multiplicative function and suppose that the set $\{f(p): p \in \mathbb{P}\}$ is finite. Then for every $\varepsilon > 0$, there exists $Q_0 = Q_0(\varepsilon)$ such that if $Q \in \mathbb{N}$ is such that $Q_0 \mid Q$, then*

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} |f(Qn + 1) - 1| \leq \varepsilon. \quad (5.2)$$

Such concentration results are distinctive properties of pretentious multiplicative functions. Indeed, for every $Q \in \mathbb{N}$, an aperiodic multiplicative function has average 0 on the arithmetic progression $Q\mathbb{N} + 1$, hence (5.2) fails if $\varepsilon < 1$.

We also note that a related property of “Besicovitch rational almost periodicity”, obtained in [15, Theorem 6] for pretentious multiplicative functions $f: \mathbb{N} \rightarrow \mathbb{U}$ that take finitely many values on primes, states that for every $\varepsilon > 0$, there exists a periodic sequence a_ε such that

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} |f(n) - a_\varepsilon(n)| \leq \varepsilon.$$

Unfortunately, for large values of Q , this estimate alone cannot give us useful information about the values of $f(Qn + 1)$, and so it is not helpful for our applications.

For our purposes, we need an extension of Theorem 5.4 that covers all pretentious multiplicative functions and also allows for some uniformity in our choice of Q . To state it, we have to introduce some notation. For $K \in \mathbb{N}$, let

$$\Phi_K := \left\{ \prod_{p \leq K} p^{a_p} : K < a_p \leq 2K \right\}. \quad (5.3)$$

If $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$ and $K, N \in \mathbb{N}$, we let

$$F_N(f, K) := \sum_{K < p \leq N} \frac{1}{p} (f(p) \cdot \overline{\chi(p)} \cdot p^{-it} - 1). \quad (5.4)$$

Since $f \sim \chi \cdot n^{it}$, we have that the limit $\lim_{N \rightarrow \infty} \Re(F_N(f, 1))$ exists. But there are cases where $\lim_{N \rightarrow \infty} |\Im(F_N(f, 1))| = +\infty$; this holds in example (iv) of Section 5.2.

Theorem 5.5 ([44, Lemma 2.5]). *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be a multiplicative function and suppose that $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$. Then*

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{n \in [N]} |f(Qn + 1) - (Qn)^{it} \cdot \exp(F_N(f, K))| = 0, \quad (5.5)$$

where Φ_K and $F_N(f, K)$ are as in (5.3) and (5.4), respectively.

Note that in the case of multiplicative functions that take finitely many values on primes, we have $t = 0$, and the series defining $F_N(f, K)$ can be shown to converge, so in our statement we can replace the term $F_N(f, K)$ with 0. This gives a statement similar to the one in Theorem 5.4.

5.4. Linear concentration estimates: Proof

There are two main steps in the proof of Theorem 5.5. In the first step, we deal with additive functions on the complex numbers and prove the concentration estimate (5.8) under assumption (5.7). We describe the key steps of this argument; the reader will find the missing details, for example, in [58, Chapter III.3.2]. In the second step, given a pretentious multiplicative function f , we apply the previous results for the additive function h defined in (5.12) to derive the required concentration estimate for f .

Step 1 (Turán–Kubilius inequality). The function $h: \mathbb{N} \rightarrow \mathbb{C}$ is *additive* if it satisfies

$$h(mn) = h(m) + h(n) \quad \text{whenever } (m, n) = 1.$$

For simplicity, we also assume that

$$h(p^k) = 0 \quad \text{for every } p \in \mathbb{P}, k \geq 2, \quad (5.6)$$

and remark that this simplifying assumption can be easily removed. The Turán–Kubilius inequality implies that if

$$\sum_{p \in \mathbb{P}} \frac{|h(p)|^2}{p} < \infty, \quad (5.7)$$

then

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{n \in [N]} |h(Qn + 1) - H_N(h, K)|^2 = 0, \quad (5.8)$$

where Φ_K is as in (5.3), and

$$H_N(h, K) := \sum_{K < p \leq N} \frac{h(p)}{p}.$$

For $Q \in \Phi_K$, $H_N(h, K)$ is approximately equal to $\mathbb{E}_{n \in [N]} h(Qn + 1)$.

We mention here the important ingredients of the proof of (5.8). In the next subsections, we list the essential changes that need to be made to get analogous extensions for binary quadratic forms.

Let

$$w_{Q,N}(p, q) := \frac{1}{N} \sum_{n \in [N], p, q \parallel Qn+1} 1,$$

where $p \parallel Qn + 1$ means that $p \mid Qn + 1$ but $p^2 \nmid Qn + 1$. To prove (5.8), we expand the square and compute it using the identity

$$h(Qn + 1) = \sum_{p \parallel Qn+1} h(p) = \sum_{p > K, p \parallel Qn+1} h(p),$$

which follows from the additivity of h , our simplifying assumption $h(p^k) = 0$ for $k \geq 2$, and the fact that if $Q \in \Phi_K$, we have $p \nmid Qn + 1$ for all primes $p \leq K$. We deduce that

$$\mathbb{E}_{n \in [N]} h(Qn + 1) = \sum_{K < p \leq N} w_{Q,N}(p, p) \cdot h(p) \quad (5.9)$$

and

$$\mathbb{E}_{n \in [N]} |h(Qn + 1)|^2 = \sum_{K < p, q \leq N} w_{Q,N}(p, q) \cdot h(p) \cdot \overline{h(q)}. \quad (5.10)$$

By direct computation, we get the approximate identities

$$w_{Q,N}(p, p) = \frac{1}{p} \left(1 - \frac{1}{p}\right) + O\left(\frac{1}{N}\right)$$

and

$$w_{Q,N}(p, q) = \frac{1}{pq} \left(1 - \frac{1}{p}\right) \left(1 - \frac{1}{q}\right) + O\left(\frac{1}{N}\right),$$

which hold for all primes p, q with $p \neq q$. Importantly, for $p \neq q$, we have

$$w_{Q,N}(p, q) = w_{Q,N}(p, p) \cdot w_{Q,N}(q, q) + O\left(\frac{1}{N}\right). \quad (5.11)$$

(The fact that the error term in (5.11) does not depend on Q is responsible for the uniformity with respect to Q in (5.5).) Using (5.9), (5.10), (5.11), it is possible to show that (5.8) holds. To do this, we use that the terms of the form $\sum_{K < p \leq N} |h(p)|^2 / p$

have negligible contribution in (5.8) because of (5.7). Also, error terms of the form $O(1/N)$ contribute a total error $O(\sum_{p,q: pq \leq QN+1} 1/N) = O(Q \log \log N / \log N)$, which is also negligible for (5.8). This last fact is no longer true for polynomials in two variables and respective averages, because there are too many error terms, which causes significant problems in our treatment of analogous concentration estimates for sums of two squares and more general irreducible binary quadratic forms.

Step 2 (From additive to multiplicative). The second step is to pass the previous result from “pretentious” additive functions to pretentious multiplicative functions. To do this, we first reduce to the case where $f \sim 1$ and then define the additive function $h: \mathbb{N} \rightarrow \mathbb{C}$ on prime powers as follows:

$$h(p^k) := f(p^k) - 1, \quad p \in \mathbb{P}, k \in \mathbb{N}. \quad (5.12)$$

We also make the simplifying assumption $f(p^k) = 1$ for $k \geq 2$, hence (5.6) holds. We can remove this assumption using that $f \sim 1$, which implies that the contribution of prime powers greater than 1 is negligible in (5.5).

Since $f \sim 1$, we get that (5.7) holds, and hence (5.8) holds from Step 1. Our goal is to relate the values of $f(Qn + 1)$ with those of $h(Qn + 1)$ and derive (5.5) from (5.8). We will use the approximate identity

$$z = e^{z-1} + O(|z - 1|^2), \quad (5.13)$$

which holds for $|z| \leq 1$. Note that since $f \sim 1$, we have that $f(p) - 1$ is close to 0 for “most” primes p , so for $z_p := f(p)$, we have that (5.13) gives a sensible approximation. We deduce that

$$f(Qn + 1) = \prod_{p \mid Qn+1, p > K} f(p) = \prod_{p \mid Qn+1, p > K} (\exp(h(p)) + O(|h(p)|^2)),$$

where we used that for $Q \in \Phi_K$, only primes greater than K divide $Qn + 1$. If we ignore the error terms on the right side (this can be done using (5.7)), we get that it is equal to

$$\prod_{p \mid Qn+1} \exp(h(p)) = \exp(h(Qn + 1)),$$

where we used (5.6). Summarizing, we have just shown that in establishing (5.5), we can replace $f(Qn + 1)$ with $\exp(h(Qn + 1))$. So to prove (5.5), it remains to show that (note that since $f \sim 1$, we have $t = 0$)

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{n \in [N]} |\exp(h(Qn + 1)) - \exp(F_N(f, K))| = 0.$$

Using the estimate $|e^z - e^w| \ll |z - w|$, which holds for $\Re(z), \Re(w) \leq 0$, and the Cauchy–Schwarz inequality, it suffices to show that

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{n \in [N]} |h(Qn + 1) - F_N(f, K)|^2 = 0.$$

Since $F_N(f, K) = H_N(h, K)$, this follows from (5.8).

5.5. Concentration estimates along sums of two squares

To prove that the pair of polynomials $2mn, m^2 + n^2$ is density regular (which implies that $x^2 + y^2 = z^2$ is partition regular with respect to y, z), we need to verify Theorem 3.2 when $\alpha = \beta = 1, \gamma = 2$. To do this, we need a variant of Theorem 5.5 that deals with concentration estimates along sums of two squares. We give the precise statement next.

If $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$, and $K, N \in \mathbb{N}$, we let

$$G_N(f, K) := \sum_{K < p \leq N, p \equiv 1 \pmod{4}} \frac{1}{p} (f(p) \cdot \overline{\chi(p)} \cdot p^{-it} - 1). \quad (5.14)$$

This expression is similar to $F_N(f, K)$ in (5.4), but we omit primes that are congruent to 3 (mod 4) in the summation. This is due to the fact that no prime of this form is a sum of two squares, which ultimately makes the contribution of such primes negligible for our purposes.

Theorem 5.6 ([23, Proposition 2.11]). *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be a multiplicative function and suppose that $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$. Then*

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{m, n \in [N]} |f((Qm + 1)^2 + (Qn)^2) - Q^{2it} \cdot (m^2 + n^2)^{it} \cdot \exp(G_N(f, K))| = 0, \quad (5.15)$$

where Φ_K and $G_N(f, K)$ are as in (5.3) and (5.14), respectively.

If f is pretentious and takes finitely many values in the primes, then it can be shown that $f \sim \chi$ for some Dirichlet character χ , and the series defining $G_N(f, K)$ converges as $N \rightarrow \infty$, so the term $Q^{2it} \cdot (m^2 + n^2)^{it} \cdot \exp(G_N(f, K))$ in (5.15) can be replaced by 1.

Our strategy is similar to the one used to prove the linear concentration estimates in Section 5.3. However, to carry out the first step, which is the Turán–Kubilius estimate for additive functions taken along sums of two squares, there are some additional important difficulties that we must resolve.

Our goal is to show that if

$$\sum_{p \in \mathbb{P}} \frac{|h(p)|^2}{p} < \infty, \quad (5.16)$$

and h_N is the restriction of h on primes less than N , i.e.,

$$h_N(p^k) := \begin{cases} h(p^k) & \text{if } p \leq N, \\ 0 & \text{otherwise,} \end{cases}$$

then we have the following variant of the Turán–Kubilius inequality (5.8) that deals with sums of two squares

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{m,n \in [N]} |h_N((Qm+1)^2 + (Qn)^2) - H_N(h, K)|^2 = 0, \quad (5.17)$$

where

$$H_N(h, K) := \sum_{K < p \leq N, p \equiv 1 \pmod{4}} \frac{h(p)}{p}.$$

We could replace h_N with h in (5.17), but because of some error terms that appear in the next step of the argument, such a result would not be more useful. For convenience, we also assume that $h(p) = 0$ for all primes that are congruent to 3 (mod 4) and $h(p^k) = 0$ for all primes p and $k \geq 2$; both conditions can be easily removed.

Following the plan used to prove the linear Turán–Kubilius inequality (5.8), we must compute

$$w_{Q,N}(p, q) := \frac{1}{N^2} \sum_{m,n \in [N], p, q \mid (Qm+1)^2 + (Qn)^2} 1$$

whenever $p, q \equiv 1 \pmod{4}$. Using that -1 is a quadratic residue (mod p) whenever $p \equiv 1 \pmod{4}$ is a quadratic residue and applying the Chinese remainder theorem, we find that

$$w_{Q,N}(p, p) = \frac{2}{p} \left(1 - \frac{1}{p}\right)^2 + O\left(\frac{1}{N}\right),$$

and

$$w_{Q,N}(p, q) = \frac{4}{pq} \left(1 - \frac{1}{p}\right)^2 \left(1 - \frac{1}{q}\right)^2 + O\left(\frac{1}{N}\right), \quad (5.18)$$

which holds for all primes p, q with $p \neq q$ that are congruent to 1 (mod 4). Importantly, for primes $p \neq q$, congruent to 1 (mod 4), we again have the approximate identity

$$w_{Q,N}(p, q) = w_{Q,N}(p, p) \cdot w_{Q,N}(q, q) + O\left(\frac{1}{N}\right).$$

In contrast to the linear case, in our case, because we are averaging over two variables m, n , the errors incurred by terms of the form $O(1/N)$ are too many and cannot be easily neglected. To deal with this problem, we decompose h as a sum $h_1 + h_2$ of two additive functions, “supported” on primes less than $\sqrt{N}$ and primes greater than $\sqrt{N}$, respectively. The $O(1/N)$ errors turn out to be negligible for h_1 . For h_2 , they can be handled using the following simple but important fact.

Lemma 5.7. *If p, q are distinct primes congruent to 1 (mod 4), then*

$$w_{Q,N}(p, p) \ll \frac{Q^2}{p}, \quad w_{Q,N}(p, q) \ll \frac{Q^2}{pq},$$

where the implicit constant is absolute.

The proof of this fact is simple and is based on the elementary estimate

$$\sum_{k \leq n} r_2(k) \ll n,$$

where $r_2(n)$ is the number of representations of k as a sum of two squares. Lemma 5.7 gives better upper bounds than (5.18) when pq is much larger than N , bounds that turn out to be sufficient to show that the contribution of h_2 in (5.15) is negligible ((5.16) with h_2 instead of h is also used here).

The second step of the proof is the transition from additive to multiplicative functions. We follow the same approach as in the linear case described above. We first reduce to the case $f \sim 1$, in which case we get that (5.15) holds with f_N instead of f , where f_N is the restriction of f on primes less than N , i.e.,

$$f_N(p^k) := \begin{cases} f(p^k) & \text{if } p \leq N, \\ 1 & \text{otherwise.} \end{cases}$$

The error terms incurred in this transition allow us to deal only with additive and multiplicative functions supported on primes $p \leq N$; this is the reason why in (5.17) a larger range of primes in the support of h_N would not have been helpful.

Finally, to replace f_N with f in (5.15), we use the factorization $f = f_N \cdot g_N$, where g_N is the restriction of f on primes greater than N , defined in a similar way to f_N . We then use a slight variant of Lemma 5.7 together with our hypothesis $f \sim 1$ to show that the contribution of g_N in (5.15) is negligible and complete the proof of Theorem 5.6.

5.6. Concentration estimates for general binary quadratic forms

To prove Theorems 3.2 and 3.3, which are needed to cover partition and density regularity results of the more general quadratic equations $ax^2 + by^2 = cz^2$, we need

to establish variants of Theorem 5.6 in which the place of $m^2 + n^2$ is taken by the more general irreducible binary quadratic form $P(m, n) = \alpha m^2 + \beta mn + \gamma n^2$, where $\alpha, \beta, \gamma \in \mathbb{Z}$. To simplify the notation a bit, we give the exact statement when $P(m, n) = m^2 + dn^2$ and $d \in \mathbb{Z}$ is a square-free integer, the general case is similar, and the exact statements can be found in [22, Proposition 4.9]. We first have to modify the definition of $G_N(f, K)$ from (5.14). If (a/p) denotes the Legendre symbol and $d \in \mathbb{Z}$ is square-free, we let

$$\mathcal{P}_d := \left\{ p \in \mathbb{P} : \left(\frac{-d}{p} \right) = 1 \right\}. \quad (5.19)$$

For example, $\mathcal{P}_1 := \{p \equiv 1 \pmod{4}\}$, $\mathcal{P}_2 := \{p \equiv 1, 3 \pmod{8}\}$, $\mathcal{P}_{-2} := \{p \equiv 1, 7 \pmod{8}\}$. It can be shown that for square-free d , the set $\mathcal{P}_d$ has relative density $1/2$ on the primes, and it always consists of the set of primes that belong to finitely many congruence classes. For example, using quadratic reciprocity, we get that if q is an odd prime, then

$$\mathcal{P}_q = \{p \equiv \pm k^2 \pmod{4q} : k \in [4q] \text{ is odd and coprime to } 4q\}.$$

If $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$, and $K, N \in \mathbb{N}$, we let

$$G_{d,N}(f, K) := 2 \sum_{K < p \leq N, p \in \mathcal{P}_d} \frac{1}{p} (f(p) \cdot \overline{\chi(p)} \cdot n^{-it} - 1). \quad (5.20)$$

Theorem 5.8 ([22, Proposition 4.9]). *Let $f: \mathbb{N} \rightarrow \mathbb{U}$ be a multiplicative function and suppose that $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$. Then for every square-free $d \in \mathbb{Z}$ such that the form $m^2 + dn^2$ is irreducible, we have*

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{m,n \in [N]} |f((Qm+1)^2 + d(Qn)^2) - Q^{2it} \cdot (m^2 + dn^2)^{it} \cdot \exp(G_{d,N}(f, K))| = 0,$$

where Φ_K and $G_{d,N}(f, K)$ are as in (5.3) and (5.20), respectively.

We also prove variants of concentration estimates along more general lattices of the form $\{(Qm + a, Qn + b) : m, n \in \mathbb{N}\}$ for suitable $a, b \in \mathbb{Z}_+$; such results are not needed for the proof of Theorem 3.2 but turn out to be essential for the proof of Theorem 3.3.

To prove Theorem 5.8, we follow the approach used to prove Theorem 5.6, but there are quite a few additional difficulties. The most serious problem is to find a variant of Lemma 5.7 that allows us to discount the contribution of large primes in the various steps of the proof. For $d > 0$, the most problematic case is when the corresponding quadratic integer ring is not a principal ideal domain, and additional,

more substantial problems arise when $d < 0$. For example, suppose that $d < 0$ is a square-free integer such that $d \equiv 1, 2 \pmod{4}$, and we want to establish for all primes $p \in \mathcal{P}_d$ an estimate of the form

$$|\{m, n \in [N]: p \mid m^2 + dn^2\}| \ll_d \frac{N^2}{p}, \quad (5.21)$$

where the implicit constant depends only on d . A priori, it is not clear that such an estimate holds, because the counting is affected by the number of the solutions of Pell's equation $m^2 + dn^2 = 1$ in the box $[1, N] \times [1, N]$, which grows as $N \rightarrow \infty$, roughly like $\log N$ (unlike the case $d > 0$, where the corresponding number is bounded by 4).

Fortunately, we can perform a careful count and use classical but non-trivial facts from algebraic number theory to show that (5.21) indeed holds. We first note that the left-hand side in (5.21) is bounded by

$$\sum_{r \leq \frac{2N^2}{p}} C_{d,N}(pr),$$

where $C_{d,N}(k) := |\{m, n \in [N]: |m^2 + dn^2| = k\}|$. Using the fact that there exists a unit $u \in \mathbb{Z}[\sqrt{-d}] \subset \mathbb{R}$ such that $u > 1$ and every other unit has the form $\pm u^t$ for some $t \in \mathbb{Z}$ (this is the Dirichlet unit theorem for the ring $\mathbb{Z}[\sqrt{-d}]$), it is possible to show that there exists $c_d > 0$ such that

$$C_{d,N}(k) \ll \log\left(\frac{c_d N}{\sqrt{k}}\right) \cdot \mathcal{C}_d(k), \quad (5.22)$$

where

$$\mathcal{C}_d(k) := |\{\text{ideals in } \mathbb{Z}[\sqrt{-d}] \text{ with norm } k\}|.$$

It is important for our argument that the term $\sqrt{k}$ appears in the denominator on the right-hand side of (5.22), our argument would not work without it. So (5.21) follows if we establish the bound

$$\sum_{r \leq \frac{2N^2}{p}} \log\left(\frac{c_d N}{\sqrt{pr}}\right) \cdot \mathcal{C}_d(pr) \ll_d \frac{N^2}{p}. \quad (5.23)$$

To prove this, we crucially use the estimate

$$\mathcal{C}_d(pr) \leq 2\mathcal{C}_d(p), \quad (5.24)$$

which holds for all $p \in \mathcal{P}_d$. This seemingly simple fact is not easy to prove from first principles, especially in the case where the corresponding ring of integers is

not a principal ideal domain. However, its proof follows easily from well-known but non-trivial facts in algebraic number theory: the multiplicativity of $\mathcal{C}_d(k)$, the estimate $\mathcal{C}_d(p^{k+1}) \leq \mathcal{C}_d(p^k) + 1$, and the fact that $\mathcal{C}_d(p^k) \geq 1$ for all $p \in \mathcal{P}_d$ and $k \in \mathbb{N}$. The proof of (5.23) follows easily by using (5.24) and then applying partial summation combined with the estimate

$$\sum_{r=1}^N \mathcal{C}_d(r) \ll_d N,$$

which follows from the ideal counting theorem.

6. Connecting the pieces (the Q -trick)

In this section, we show how we can combine the ingredients of the previous two sections to prove Theorems 3.2 and 3.3. We start with the simplest case and then explain the necessary changes needed to cover the increasingly more complicated cases.

6.1. The pair $(2mn, m^2 - n^2)$

We start with the proof of Theorem 3.2 when $\alpha = 1$, $\beta = -1$, $\gamma = 2$, which suffices to prove partition and density regularity of $x^2 + y^2 = z^2$ with respect to the variables x, y . Using the positivity property (3.11), it suffices to show that there exists $Q \in \mathbb{N}$ for which

$$\int_{\mathcal{M}} L(f, Q) d\sigma(f) > 0, \quad (6.1)$$

where

$$L(f, Q) := \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} f((Qm + 1)^2 - (Qn)^2) \cdot \overline{f(2(Qm + 1)Qn)}. \quad (6.2)$$

Note that in establishing the above reduction, we used the bounded convergence theorem and the non-trivial fact that the limit defining $L(f, Q)$ exists, which follows from [20, Theorem 1.4]. The proof of the existence of the limit is non-trivial and can be avoided at the cost of making the argument a bit more complex.

We also note that we could have used in our averaging any other lattice of the form $(Qm + a, Qn)$, where a is some fixed number independent of Q ; what is important is that on one of the two coordinates we use Qn (or Qm) in order to later on factor out a term $f(Q)$, which is an essential element of a key maneuver that we call the “ Q -trick”.

It easily follows from Theorem 4.7 that if the multiplicative function f is aperiodic, then

$$L(f, Q) = 0 \quad \text{for every } Q \in \mathbb{N}.$$

Hence, (6.1) would follow if we show that

$$\int_{\mathcal{M}_p} L(f, Q) d\sigma(f) > 0 \quad \text{holds for some } Q \in \mathbb{N}, \quad (6.3)$$

where

$$\mathcal{M}_p := \{f \in \mathcal{M} \text{ is pretentious}\}.$$

Note that the integrals in (6.3) can be shown to be real. To prove (6.3), we will use the linear concentration estimates of Theorem 5.5 to substantially simplify the value of the limit in (6.2) and then use a trick, called the Q -trick, to show that for some particular choice of Q , we get positivity.

6.1.1. Finitely generated case. To better illustrate the argument, we first assume that the measure σ is supported on the subset of pretentious multiplicative functions that are finitely generated, i.e., the set $\{f(p) : p \in \mathbb{P}\}$ is a finite subset of the unit circle. The advantage in this case is that $f \sim \chi$ for some Dirichlet character χ (so n^{it} is no longer needed), and also the oscillatory factor $\exp(F_N(f, K))$ in (5.5) can be taken to be 1. So in this case, the concentration estimate in (5.5) takes the much simpler form

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} \mathbb{E}_{n \in [N]} |f(Qn + 1) - 1| = 0, \quad (6.4)$$

where Φ_K is as in (5.3). Since

$$Qm + 1 \pm Qn = Q(m \pm n) + 1,$$

with a bit of fiddling (see [23, Lemma 3.1]), (6.4) implies that for all practical purposes, the limit $L(f, Q)$ in (6.2) can be approximated by $f(Q) \cdot A(f)$, where

$$A(f) := \lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \overline{f(2n)}. \quad (6.5)$$

More precisely, we have

$$\lim_{K \rightarrow \infty} \max_{Q \in \Phi_K} |L(f, Q) - f(Q) \cdot A(f)| = 0. \quad (6.6)$$

It follows from this that in order to establish (6.3), it suffices to study the integrals

$$\int_{\mathcal{M}_p} f(Q) \cdot A(f) d\sigma(f).$$

A priori, however, these expressions do not satisfy any obvious positivity property, and it seems that we are at a dead end. The simple but crucial observation is that the multiplicative average of this last expression, taken over $Q \in \Phi_K$ and letting $K \rightarrow \infty$, does satisfy a nice positivity property, namely,

$$\lim_{K \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} \int_{\mathcal{M}_p} f(Q) \cdot A(f) d\sigma(f) = \sigma(\{1\}) > 0. \quad (6.7)$$

This is because for every completely multiplicative function $f \neq 1$, we have

$$\lim_{K \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} f(Q) = 0,$$

and so the bounded convergence theorem gives that (6.7) holds. Combining (6.6) and (6.7), we get that

$$\lim_{K \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} \int_{\mathcal{M}_p} L(f, Q) d\sigma(f) = \sigma(\{1\}) > 0,$$

hence (6.3) holds.

6.1.2. General case. We now consider the general case of measures supported by multiplicative functions that are not necessarily finitely generated. In this case, suppose that $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$. Then the concentration estimates of Theorem 4.7 give that

$$\lim_{K \rightarrow \infty} \max_{Q \in \Phi_K} |L(f, Q) - f(Q) \cdot Q^{it} \cdot B(f, K)| = 0,$$

where

$$B(f, K) := \exp(2F_{2N}(f, K)) \cdot \overline{\exp(F_N(f, K))} \cdot \lim_{N \rightarrow \infty} \mathbb{E}_{m, n \in [N]} (m^2 - n^2)^{it} \cdot m^{-it} \cdot \overline{f(2n)}.$$

Although the expressions $B(f, K)$ are much more complicated than the expressions $A(f)$ in (6.5), this is of little concern to us; what matters is that for $Q \in \Phi_K$, they are constant in Q . This allows us to show that as long as $f \neq n^{-it}$, we have

$$\lim_{K \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} f(Q) \cdot Q^{it} \cdot B(f, K) = 0.$$

As before, using the bounded convergence theorem, we conclude that

$$\lim_{K \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} \int_{\mathcal{M}_p} L(f, Q) d\sigma(f) = \int_{\mathcal{A}} B(f) d\sigma(f),$$

where

$$\mathcal{A} := \{(n^{it})_{n \in \mathbb{N}} : t \in \mathbb{R}\}$$

and

$$B(f) := \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} (m^2 - n^2)^{it} \cdot (2mn)^{-it}$$

when $f = n^{it}$. In order to complete the proof, it suffices to prove the positivity of $\int_{\mathcal{A}} B(f) d\sigma(f)$. We do this in two different ways below, the second being the more flexible.

(Endgame I: Miraculous positivity). The limit defining $B(f)$ is easily shown to be

$$\int_0^1 \int_0^1 (x - y)^{it} \cdot (2xy)^{it} dx dy.$$

It can be shown that this double integral has a positive real part for every $t \in \mathbb{R}$ by explicitly evaluating the integral (we did this computation using some computer software). Since $\sigma(\{1\}) > 0$, this implies that $\int_{\mathcal{A}} B(f) d\sigma(f) > 0$ and completes the proof.

However, this miraculous positivity property fails when we replace $2mn$ with mn or $3mn$ and probably only works for kmn when $k = 2$, i.e., only for the case of Pythagorean pairs! Next, we will give an alternative, more complicated, but robust method that works in much higher generality (this is needed for equations other than the Pythagorean).

(Endgame II: A more robust method). We introduce some weights that effectively allow us to replace all non-trivial Archimedean characters with 1. For every $\delta \in (0, 1/2)$, we let

$$w_\delta(m, n) := \mathbf{1}_{I_\delta}((m^2 - n^2)^i \cdot (mn)^{-i}) \cdot \mathbf{1}_{m > n}, \quad m, n \in \mathbb{N}, \quad (6.8)$$

where I_δ is an arc in $\mathbb{S}^1$ around 1 with length δ . It is not difficult to show that for all $\delta > 0$, we have

$$\mu_\delta := \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} w_\delta(m, n) > 0 \quad (6.9)$$

and also

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \sup_{Q \in \mathbb{N}} |w_\delta(Qm + 1, Qn) - w_\delta(m, n)| = 0,$$

so for our purposes, the weights $w_\delta(m, n)$ and $w_\delta(Qm + 1, Qn)$ can be used interchangeably. It is also not hard to show that for aperiodic multiplicative functions f , we have the vanishing property (4.12) even if we insert the weights $w_\delta(m, n)$ in the average. Using these properties, we can work out the previous argument, replacing the expression $L(f, Q)$ in (6.2) by

$$L_\delta(f, Q) := \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} w_\delta(m, n) \cdot f((Qm + 1)^2 - (Qn)^2) \cdot \overline{f(2(Qm + 1)Qn)}.$$

In this way, we reduce matters to showing that for some sufficiently small $\delta > 0$, we have

$$\int_{\mathcal{A}} B_{\delta}(f) d\sigma(f) > 0,$$

where

$$B_{\delta}(n^{it}) := \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} w_{\delta}(m, n) \cdot (m^2 - n^2)^{it} \cdot (2mn)^{-it}$$

when $f = n^{it}$. The advantage of this weighted average is that for any fixed $t \in \mathbb{R}$, because of the definition of the weight in (6.8), if δ is small enough, then $B_{\delta}(f)$ is close to 1. This allows us to show with some fiddling that if δ is small enough, then we have the lower bound

$$\int_{\mathcal{A}} B_{\delta}(f) d\sigma(f) \geq \frac{1}{2} \cdot \mu_{\delta} \cdot \sigma(\mathcal{A}) \geq \frac{1}{2} \cdot \mu_{\delta} \cdot \sigma(\{1\}) > 0,$$

where μ_{δ} is as in (6.9). This completes the proof in a robust way that allows us to deal with positivity properties of more general pairs such as $(\ell mn, \ell'(m^2 - n^2))$, where $\ell, \ell' \in \mathbb{N}$ are arbitrary. The reader will find more details in [23, Section 4].

6.2. Comparison with the strategy from [21]

In [21], we were able to show partition regularity for the pairs $(m^2 - n^2, m(m + 2n))$ but not for the pairs $(m^2 - n^2, 2mn)$. Let us briefly explain why. The main technical tool used in [21] was a structural result for general multiplicative functions which roughly states the following (see [21, Theorem 2.1] for the detailed statement): For every $f \in \mathcal{M}$, $\varepsilon > 0$, and $s \in \mathbb{N}$, there exist $Q \in \mathbb{N}$ and $C > 0$ such that for every $N \in \mathbb{N}$, we have a decomposition

$$f(n) = f_{\text{st}}(n) + f_{\text{un}}(n) + f_{\text{er}}(n), \quad n \in [N],$$

such that

- (i) $|f_{\text{st}}(n + Q) - f_{\text{st}}(n)| \leq C/N$ for every $n \in [N - Q]$;
- (ii) $\|f_{\text{un}}\|_{U^s[N]}$ is “very small” in a sense that will not be made precise here;
- (iii) $\mathbb{E}_{n \in [N]} |f_{\text{er}}(n)| \leq \varepsilon$.

In fact, the applications to partition regularity require some uniformity in the choice of Q and C with respect to $f \in \mathcal{M}$, but that will not concern us here.

Let us see how we can use these properties to get some positiveness for fixed $f \in \mathcal{M}$ and averages of the form

$$\mathbb{E}_{m,n \in [N]} f(m^2 - (Qn)^2) \cdot \overline{f(m(m + 2Qn))}.$$

Using properties (ii) and (iii), it is not hard to see that these averages are $O(\varepsilon)$ -close to the averages

$$\mathbb{E}_{m,n \in [N]} f_{\text{st}}(m - Qn) \cdot f_{\text{st}}(m + Qn) \cdot \overline{f_{\text{st}}(m)} \cdot \overline{f_{\text{st}}(m + 2Qn)}.$$

Assuming for simplicity that $f_{\text{st}}(n + Q) = f_{\text{st}}(n)$ for every $n \in \mathbb{N}$, the last average is equal to

$$\mathbb{E}_{m,n \in [N]} |f_{\text{st}}(m)|^4.$$

This positive lower bound is crucial for the argument in [21] and with some maneuvering can be used to prove partition regularity of the patterns $(m^2 - n^2, m(m + 2n))$. However, it does not appear to be possible to modify this argument in a simple way in order to deal with the pairs $(m^2 - n^2, 2mn)$, the reason being that if $m = 0$, or $n = 0$, the two coordinates of this pair do not coincide.

Instead, one can try to restrict both coordinates m and n to suitable progressions (just as we did in Section 6.1) and work with the averages

$$\mathbb{E}_{m,n \in [N]} f((Qm + 1)^2 - (Qn)^2) \cdot \overline{f((Qm + 1)(Qn))}. \quad (6.10)$$

Unfortunately, the previous decomposition result seems inadequate for analyzing these expressions because we have no useful control over the error term f_{er} along the progression $Q\mathbb{N}$. The important advantage of the concentration estimates described in Section 5 is that for pretentious multiplicative functions, we obtain an error term that is small even when we restrict both variables m, n to arithmetic progressions. This is crucial for the purposes of analyzing the asymptotic behavior of the averages in (6.10). Moreover, we will see in the next subsection that the concentration estimates described in Section 5.5 allow us to effectively analyze averages of the form

$$\mathbb{E}_{m,n \in [N]} f((Qm + 1)^2 + (Qn)^2) \cdot \overline{f((Qm + 1)(Qn))}$$

for pretentious multiplicative functions. Such a task was beyond the limits of the methodology used in the article [21].

6.3. The pair $(2mn, m^2 + n^2)$

To establish Theorem 3.2 in this case ($\alpha = \beta = 1, \gamma = 2$), and thus deduce partition and density regularity of $x^2 + y^2 = z^2$ with respect to the variables x, z , we follow the argument of the previous subsection with some necessary and rather laborious changes. We introduce the averages

$$\begin{aligned} L'_{\delta, N}(f, Q) &:= \mathbb{E}_{m,n \in [N]} \tilde{w}_{\delta}(m, n) \cdot f((Qm + 1)^2 \\ &\quad + (Qn)^2) \cdot \overline{f(2(Qm + 1)Qn)}, \end{aligned} \quad (6.11)$$

where the weights $\tilde{w}_\delta$ are defined similarly to (6.8) but with $m^2 - n^2$ replaced by $m^2 + n^2$. In order to study their asymptotic behavior, as $N \rightarrow \infty$, we first show using Theorem 4.8 that in the aperiodic case, they vanish for every $Q \in \mathbb{N}$.

We then study their asymptotic behavior when f is pretentious. To simplify their form, we use the non-linear concentration estimates of Theorem 5.6. This is an important difference, since the non-linear concentration estimates were not previously known and turned out to be much harder to obtain than the linear ones. Another technical annoyance is that, contrary to what happened in the previous subsection, the limit $\lim_{N \rightarrow \infty} L'_{\delta,N}(f, Q)$ may not exist, the reason being that the oscillatory factors resulting from the concentration estimates for the quadratic and linear terms do not conveniently cancel. In any case, we can show that if $f \sim \chi \cdot n^{it}$ for some Dirichlet character χ and $t \in \mathbb{R}$, then

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \max_{Q \in \Phi_K} |L'_{\delta,N}(f, Q) - f(Q) \cdot Q^{it} \cdot B_{\delta,N}(f, K)| = 0,$$

where

$$B_{\delta,N}(f, K) := \exp(G_N(f, K)) \cdot \overline{\exp(F_N(f, K))} \cdot \mathbb{E}_{m,n \in [N]} \tilde{w}_\delta(m, n) \cdot (m^2 + n^2)^{it} \cdot m^{-it} \cdot \overline{f(2n)},$$

and $F_N(f, K)$ and $G_N(f, K)$ are given by (5.4) and (5.14), respectively. Again, the complexity of the expression $B_{\delta,N}(f, K)$ will not bother us; the only important fact is that for $Q \in \Phi_K$, it is constant in Q . This allows us to show that as long as $f \neq n^{-it}$, we have

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} |\mathbb{E}_{Q \in \Phi_K} f(Q) \cdot Q^{it} \cdot B_{\delta,N}(f, K)| = 0.$$

After this point, arguing as in the previous subsection, we get that for some $\delta_0 > 0$, we have the following positivity property:

$$\liminf_{K \rightarrow \infty} \liminf_{N \rightarrow \infty} \mathbb{E}_{Q \in \Phi_K} \int_{\mathcal{M}} L'_{\delta_0,N}(f, Q) d\sigma(f) > 0, \quad (6.12)$$

where $L'_{\delta_0,N}(f, Q)$ is given by (6.11). It is slightly annoying that $\mathbb{E}_{Q \in \Phi_K}$ and $\liminf_{N \rightarrow \infty}$ are not interchangeable (in the previous subsection, the existence of the limit as $N \rightarrow \infty$ allowed us to do this). To overcome this and finish the argument, we argue as follows. We note that using (6.12) and the fact that $\tilde{w}_\delta(m, n)$ is approximately equal to $\tilde{w}_\delta(Qm + 1, Qn)$, we get that there exist $K_0 \in \mathbb{N}$ and $Q_N \in \Phi_{K_0}$, $N \in \mathbb{N}$, such that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} I(Q_N m + 1, Q_N n) =: a > 0, \quad (6.13)$$

where

$$I(m, n) := \int_{\mathcal{M}} \tilde{w}_\delta(m, n) \cdot f(m^2 + n^2) \cdot \overline{f(2mn)} d\sigma(f).$$

Since $I(m, n) \geq 0$ for every $m, n \in \mathbb{N}$ and Q_N belongs to a finite set bounded by, say, M , we deduce from (6.13) that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m, n \in [N]} I(m, n) = \frac{a}{M^2} > 0.$$

The reader will find the details of this argument in [23, Section 6].

6.4. The pair $(mn, m^2 + dn^2)$

The two main new ingredients needed to cover this more general case are Theorem 4.9, which establishes the vanishing property of correlations in the case of aperiodic multiplicative functions, and Theorem 5.8, which establishes concentration estimates for pretentious multiplicative functions for the binary quadratic form $m^2 + dn^2$ when it is irreducible. Modulo these two results, which require considerable additional effort to prove, the required positivity property goes along the lines described in the previous subsections. The details appear in [22, Section 5].

We remark that employing an analogous approach, we can cover all pairs $(P_1(m, n), P_2(m, n))$, where at least one of the P_1, P_2 is a reducible binary quadratic form that is not a square.

6.5. The pair $(m^2 + 2n^2, m^2 - 2n^2)$

We need to cover this pair to prove partition and density regularity of $x^2 + 2y^2 = z^2$ with respect to the variables x, z , (see the parametrization in Section 2.2). The situation is similar for the equation $x^2 + y^2 = 2z^2$ with respect to any pair of variables. In order to prove the positivity property of Theorem 3.3 for $\alpha = \alpha' = 0, \beta = 2, \beta' = -2$, we need to study the correlations

$$\mathbb{E}_{m, n \in [N]} f(m^2 + 2n^2) \cdot \overline{f(m^2 - 2n^2)} \quad (6.14)$$

for arbitrary $f \in \mathcal{M}$. It is expected that asymptotic properties of f along $m^2 + 2n^2$ and $m^2 - 2n^2$ are only affected by its values on primes in $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, where

$$\mathcal{P}_1 := \{p \equiv 1, 3 \pmod{8}\}, \quad \mathcal{P}_2 := \{p \equiv 1, 7 \pmod{8}\}. \quad (6.15)$$

(Using the notation in (5.19), we should have denoted the two sets by $\mathcal{P}_2$ and $\mathcal{P}_{-2}$, respectively; we do not do this to simplify the notation later.)

Analogously to Definition 5.1, if $\mathcal{P}$ is a subset of the primes, we let

$$\mathbb{D}_{\mathcal{P}}^2(f, g) := \sum_{p \in \mathcal{P}} \frac{1}{p} (1 - \Re(f(p) \cdot \overline{g(p)})),$$

and write $f \sim_{\mathcal{P}} g$ if $\mathbb{D}_{\mathcal{P}}(f, g) < +\infty$. The multiplicative function f is $\mathcal{P}$ -pretentious if $\mathbb{D}_{\mathcal{P}}(f, \chi \cdot n^{it}) < +\infty$ for some Dirichlet character χ and $t \in \mathbb{R}$ and $\mathcal{P}$ -aperiodic if it is not $\mathcal{P}$ -pretentious. Note that for $\mathcal{P} = \mathbb{P}$, we get the same definition as in Section 5.1.

6.5.1. The pretentious case. Let $\mathcal{P}_1, \mathcal{P}_2$ be as in (6.15) and suppose that f is $\mathcal{P}_1$ -pretentious and $\mathcal{P}_2$ -pretentious. Furthermore, for convenience, suppose that

$$f \sim_{\mathcal{P}_1} \chi_1 \quad \text{and} \quad f \sim_{\mathcal{P}_2} \chi_2 \quad (6.16)$$

for some Dirichlet characters χ_1, χ_2 . We will ignore Archimedean characters in this discussion, since they can be covered by introducing weights and using a straightforward variant of the argument given in Section 6.3.

Our aim is to use again the strategy described in the previous subsections. However, substantial new complications arise since we no longer have a reducible polynomial like mn to work with, and this makes the Q -trick harder to implement. Moreover, Theorem 5.8 gives that there exist multiplicative functions which concentrate on different oscillatory factors along $m^2 + 2n^2$ and $m^2 - 2n^2$ that do not cancel each other. This casts doubt on whether it is even possible in this case to establish positivity via concentration estimates. Fortunately, we will be able to do this by adapting the Q -trick, this time averaging over appropriate “translates” rather than “dilates” of an appropriate choice of lattice. Our goal is to ensure that for those multiplicative functions that concentrate along $m^2 + 2n^2$ and $m^2 - 2n^2$ around different oscillatory factors, a certain multiplicative average over Q will vanish, and for the remaining multiplicative functions, we can guarantee that they concentrate on the same oscillatory factor, which leads to a favorable cancellation.

Let us explain in more detail how the previous plan is implemented. For a given $K \in \mathbb{N}$ and $l_p, l'_p \in [1, 3K/2]$, let

$$Q_1 := \prod_{p \equiv 3 \pmod{8}, p \leq K, p \nmid q} p^{l_p}, \quad Q_2 := \prod_{p \equiv 7 \pmod{8}, p \leq K, p \nmid q} p^{l'_p}. \quad (6.17)$$

Let q be a common period of the multiplicative functions χ_1 and χ_2 . We plan to apply concentration estimates on the lattice

$$\{(Q_K m + a, Q_K n + b) : m, n \in \mathbb{N}\},$$

where

$$Q_K := \prod_{p \leq K} p^{2K}$$

and $a, b \in [Q]$ depend on Q_1, Q_2 and satisfy the following two properties:

$$a^2 + 2b^2 \equiv a^2 - 2b^2 \equiv 1 \pmod{q} \quad (6.18)$$

and

$$(a^2 + 2b^2, Q) = Q_1, \quad (a^2 - 2b^2, Q) = Q_2. \quad (6.19)$$

To show that a choice of a, b that satisfy (6.18) and (6.19) is possible, we use the Chinese remainder theorem and the fact that for $p \equiv 3 \pmod{8}$ the congruences $p^{l_p} \parallel a^2 + 2b^2$ and $p \nmid a^2 - 2b^2$ have a solution, and similarly, for $p \equiv 7 \pmod{8}$ the congruences $p^{l_p} \parallel a^2 - 2b^2$ and $p \nmid a^2 + 2b^2$ have a solution.

Suppose now that K is sufficiently large, Q_1, Q_2 are as in (6.17), and $a = a_{Q_1, Q_2}, b = b_{Q_1, Q_2}$ satisfy (6.18) and (6.19). For convenience, we let

$$P_1(m, n) := m^2 + 2n^2, \quad P_2(m, n) := m^2 - 2n^2.$$

Using a variant of Theorem 5.8 that allows us to work with lattices of the form $Q \cdot \mathbb{N} \times \mathbb{N} + (a, b)$, we get that for $j = 1, 2$ and $Q = Q_K$, the values of $f(P_j(Qm + a, Qn + b))$ are concentrated at

$$f(Q_j) \cdot \chi_j \left(\frac{P_j(a, b)}{Q_j} \right) \cdot \exp(G_{j,N}(f, K)), \quad (6.20)$$

(we used that $Q_j = (P_j(a, b), Q)$, $j = 1, 2$ and $q \prod_{p \leq K} p$ divides Q/Q_j) where

$$G_{j,N}(f, K) := 2 \sum_{K < p \leq N, p \in \mathcal{P}_j} \frac{1}{p} (f(p) \cdot \overline{\chi_j(p)} \cdot -1), \quad j = 1, 2.$$

Note that for $j = 1, 2$, we have $(Q_j, q) = 1$, hence $|\chi_j(Q_j)| = 1$, and also since (6.18) holds, we have $\chi_j(P_j(a, b)) = 1$. It follows that the expression in (6.20) is equal to

$$f(Q_j) \cdot \overline{\chi_j(Q_j)} \cdot \exp(G_{j,N}(f, K)).$$

Hence, $f(P_1(Qm + a, Qn + b)) \cdot \overline{f(P_2(Qm + a, Qn + b))}$ concentrates at

$$f(Q_1) \cdot \overline{\chi_1(Q_1)} \cdot \exp(G_{1,N}(f, K)) \cdot \overline{f(Q_2)} \cdot \chi_2(Q_2) \cdot \overline{\exp(G_{2,N}(f, K))}. \quad (6.21)$$

We consider two cases.

Case 1. Suppose that for primes p that satisfy $p \nmid q$, we have

$$f(p) \neq \chi_1(p) \quad \text{for some } p \equiv 3 \pmod{8}$$

or

$$f(p) \neq \chi_2(p) \quad \text{for some } p \equiv 7 \pmod{8}.$$

Then the previous facts easily imply that

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} |\mathbb{E}_{Q_1 \in \Phi_{1,K}, Q_2 \in \Phi_{2,K}} L_N(f, K, Q_1, Q_2)| = 0, \quad (6.22)$$

where

$$L_N(f, K, Q_1, Q_2) := \mathbb{E}_{m,n \in [N]} f(P_j(Q_K m + a_{Q_1, Q_2}, Q_K n + b_{Q_1, Q_2})) \cdot \overline{f(P_j(Q_K m + a_{Q_1, Q_2}, Q_K n + b_{Q_1, Q_2}))}, \quad (6.23)$$

and for $j = 1, 2$,

$$\Phi_{j,K} := \left\{ \prod_{p \in \mathcal{P}_j \cap [K]} p^{l_p} : K < l_p \leq \frac{3K}{2} \right\}.$$

Case 2. Suppose that for primes p that satisfy $p \nmid q$, we have

$$f(p) = \chi_1(p) \quad \text{for all } p \equiv 3 \pmod{8} \quad (6.24a)$$

and

$$f(p) = \chi_3(p) \quad \text{for all } p \equiv 7 \pmod{8}, \quad (6.24b)$$

then the expression in (6.21) equals

$$\exp(G_{1,N}(f, K)) \cdot \overline{\exp(G_{2,N}(f, K))}. \quad (6.25)$$

Using (6.24a) and (6.24b), we get that for K large enough and $N > K$, we have

$$G_{1,N}(f, K) = G_{2,N}(f, K).$$

Thus, the expression in (6.25) equals

$$\exp(2\Re(G_{1,N}(f, K))).$$

Because our standing assumption (6.16) is that $f \sim_{P_1} \chi_1$, this last expression approaches 1 as $N \rightarrow \infty$ uniformly on $N > K$. Combining the above, we get

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \sup_{Q_1 \in \Phi_{1,K}, Q_2 \in \Phi_{2,K}} |L_N(f, K, Q_1, Q_2) - 1| = 0, \quad (6.26)$$

where $L_N(f, K, Q_1, Q_2)$ is as in (6.23).

Combining (6.22) and (6.26), we get the positiveness properties needed to prove Theorem 3.3.

6.5.2. The aperiodic case. In the case where the multiplicative function $f \in \mathcal{M}$ is $\mathcal{P}_1$ -aperiodic or $\mathcal{P}_2$ -aperiodic, we expect the averages in (6.14) to vanish as $N \rightarrow \infty$. Unfortunately, this property turns out to be very difficult to establish because, unlike the cases treated in the previous subsections, we have no way of bounding these averages by a usable substitute for the Gowers uniformity seminorms of the multiplicative function f . In fact, the needed vanishing property is not even known for the Liouville function, i.e., it is not known that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \lambda(m^2 + 2n^2) \cdot \lambda(m^2 - 2n^2) = 0.$$

For the applications we have in mind, we would be equally happy to show this for logarithmic averages in place of Cesàro.

Definition 6.1. If $P \in \mathbb{Z}[m, n]$ is an irreducible binary quadratic form with discriminant d_P , we let

$$\mathcal{P}_P := \left\{ p \in \mathbb{P} : \left(\frac{d_P}{p} \right) = 1 \right\}.$$

We say that the irreducible binary quadratic forms P_1, P_2 are *good for vanishing of correlations of aperiodic multiplicative functions* whenever the following property holds: If $f_1, f_2: \mathbb{N} \rightarrow \mathbb{S}^1$ are completely multiplicative functions such that either f_1 is $\mathcal{P}_{P_1}$ -aperiodic or f_2 is $\mathcal{P}_{P_2}$ -aperiodic, then for every $Q \in \mathbb{N}$ and $a, b \in \mathbb{Z}_+$, we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \mathbf{1}_S(Qm + a, Qn + b) \cdot f_1(P_1(Qm + a, Qn + b)) \\ \cdot f_2(P_2(Qm + a, Qn + b)) = 0, \end{aligned}$$

where $S := \{m, n \in \mathbb{N} : P_1(m, n) > 0, P_2(m, n) > 0\}$.

Conjecture 3. *If $P_1, P_2 \in \mathbb{Z}[m, n]$ are irreducible binary quadratic forms that are not multiples of each other, then they are good for vanishing of correlations of aperiodic multiplicative functions.*

If this conjecture is true, we show in [22, Section 6] how Theorem 3.3 follows using the ideas summarized in the previous subsection.

7. More challenges

In this section, we present some open questions linked to the topics already discussed.

7.1. Vanishing of correlations along two irreducible quadratics

The following problem covers a special case of Conjecture 3 that seems very challenging.

Problem 1. If λ is the Liouville function, show that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \lambda(m^2 + n^2) \cdot \lambda(m^2 + 2n^2) = 0.$$

For the applications of this article, we could replace the Cesàro averages $\mathbb{E}_{m,n \in [N]}$ with the logarithmic averages $1/(\log N)^2 \sum_{m,n \in [N]} 1/mn$. In fact, it appears that no explicit example of two irreducible binary quadratic forms P_1, P_2 is known for which

$$\lim_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \lambda(P_1(m, n)) \cdot \lambda(P_2(m, n)) = 0$$

or a similar property with logarithmic averages.

7.2. Partition regularity of “most” generalized Pythagorean triples

Conjecture 3 asks to show that for all Rado triples (a, b, c) , equation (1.2) is partition regular. At the moment, this seems very hard to prove, so even showing that we have partition regularity for a positive proportion, or even better, for most Rado triples, would be desirable.

Problem 2. Show that for “most” choices of Rado triples $(a, b, c) \in \mathbb{N}^3$, the equation

$$ax^2 + by^2 = cz^2$$

is partition regular.

We do not make explicit what “most” means, but let us say that a subset of Rado’s triples with relative density 1 with respect to the standard density would qualify. At the moment, it is not even known that there is a single Rado triple for which we have partition regularity.

7.3. Multiple recurrence results for multiplicative actions

Although the methods of this article are particularly well suited to partition regularity problems of pairs, they seem inadequate on their own to deal with problems of triples. The reason is that we lack a representation result for sequences such as

$$(x, y, z) \mapsto d(x^{-1} \Lambda \cap y^{-1} \Lambda \cap z^{-1} \Lambda),$$

which was our starting point in our approach to partition regularity of pairs. It is understood that a useful representation similar to the Bochner–Herglotz theorem

should be very hard to prove and that new sequences, not just multiplicative functions, should be used.

An alternative approach, which has been very successful for translation-invariant patterns of arbitrary length, is to use the formalism of ergodic theory to provide sufficient statements for partition regularity and then to develop the necessary toolbox in ergodic theory to prove them. The sufficient statements in ergodic theory are described next.

We say that (X, μ, T_n) is a *multiplicative action* if $(X, \mathcal{X}, \mu)$ is a probability space, and for $n \in \mathbb{Z}$, the transformations $T_n: X \rightarrow X$ are invertible, measure preserving, and satisfy the relation $T_0 = T_1 := \text{id}$ and $T_m \circ T_n = T_{mn}$ for all $m, n \in \mathbb{Z}$. Using a variant of the correspondence principle of Furstenberg [24] that appears in [5], it turns out that in order to show that the equation $x^2 + y^2 = z^2$ is partition regular, it suffices to prove the following multiple recurrence result.

Problem 3. Let (X, μ, T_n) be a multiplicative action and A a measurable set such that $X = \bigcup_{j=1}^k T_j^{-1} A$ for some $k \in \mathbb{N}$. Show that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \mu(T_{m^2-n^2}^{-1} A \cap T_{2mn}^{-1} A \cap T_{m^2+n^2}^{-1} A) > 0. \quad (7.1)$$

Unlike the partition regularity results for pairs, such as those in Theorem 2.4, we cannot expect to have density regularity results for the equation $ax^2 + by^2 = cz^2$ when $a + b \neq c$. If, say, $a = b = c = 1$, an example by Bergelson shows that for suitably chosen $\alpha \in \mathbb{R}$, the set $\Lambda := \{n \in \mathbb{N} : \{n^2\alpha\} \in [1/5, 2/5]\}$ has positive multiplicative density but does not contain a Pythagorean triple (this problem already appears for the equation $x + y = z$). This is the reason why in Problem 3 we cannot expect to have a multiple recurrence property for any set A with positive measure. On the other hand, if the action is finitely generated, i.e., the set $\{T_p, p \in \mathbb{P}\}$ is finite, we expect that (7.1) holds for all sets A with positive measure.

In order to show that the equation $x^2 + y^2 = 2z^2$ is partition regular (in fact, density regular with respect to multiplicative density), it suffices to prove the following.

Problem 4. Let (X, μ, T_n) be a multiplicative action, and let $A \in \mathcal{X}$ with $\mu(A) > 0$. Show that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m,n \in [N]} \mu(T_{m^2-n^2+2mn}^{-1} A \cap T_{m^2-n^2-2mn}^{-1} A \cap T_{m^2+n^2}^{-1} A) > 0.$$

The problem is open even if the action is finitely generated, and we omit one of the three terms in the intersection.

7.4. Higher-degree homogeneous polynomials

As we have seen in Section 5.6, for every irreducible binary quadratic form $P(m, n)$, all pretentious multiplicative functions enjoy strong concentration properties. Does a similar fact hold for not necessarily homogeneous polynomials or higher-degree homogeneous polynomials? By slightly modifying an example of Klurman [42, Lemma 2.1], one can construct pretentious (and aperiodic) multiplicative functions $f: \mathbb{N} \rightarrow \{-1, 1\}$ so that the averages

$$\mathbb{E}_{n \in [N]} f(n^2 + 1)$$

do not converge. It follows that in this case, we cannot have concentration estimates similar to those given in Theorem 5.4. Although, as we have shown in Theorem 5.6, such counterexamples do not work for, say, $P(m, n) = m^2 + n^2$, it seems likely that they can be found for higher-degree irreducible homogeneous polynomials, but this is not so easy to achieve.

Problem 5. Let $a \in \mathbb{Z}$ and $d \geq 3$ be such that the polynomial $m^d + an^d$ is irreducible. Show that there exists a pretentious completely multiplicative function $f: \mathbb{N} \rightarrow \{-1, 1\}$ such that the averages

$$\mathbb{E}_{m, n \in [N]} f(m^d + an^d)$$

do not converge as $N \rightarrow \infty$.

We say that a sequence $a: \mathbb{N}^2 \rightarrow \mathbb{Z}$ is *good for single recurrence of multiplicative actions* if the following holds: For every multiplicative action (X, μ, T_n) and $A \in \mathcal{X}$ with $\mu(A) > 0$, we have

$$\mu(A \cap T_{a(m, n)}^{-1} A) > 0$$

for some $m, n \in \mathbb{N}$ such that $a(m, n) \neq 0$. Is the polynomial $P(m, n) = m^3 + 2n^3$ good for single recurrence of multiplicative actions? The following is a more general problem.

Problem 6. Let $P \in \mathbb{Z}[m, n]$ be a homogeneous polynomial with $P(1, 0) = 1$. Is it true that for every multiplicative action (X, μ, T_n) and $A \in \mathcal{X}$ with $\mu(A) > 0$, we have

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{m, n \in [N]} \mu(A \cap T_{P(m, n)}^{-1} A) > 0?$$

We cannot remove the assumption $P(1, 0) = 1$ since the set $\{2n^2: n \in \mathbb{N}\}$ can easily be seen to be bad for measurable recurrence [18, Example 3.11].

It follows from [18, Theorem 2.1 and Example 5.15] that the answer is positive when $\deg(P) \leq 2$. It seems very likely that, using the methodology developed

in [22], we can also get a positive answer for reducible polynomials of degree 3. But for $P(m, n) = m^3 + 2n^3$, it is not even clear how to show that for every completely multiplicative function $f: \mathbb{N} \rightarrow \{-1, 1\}$, we have $f(m^3 + 2n^3) = 1$ for a set of $m, n \in \mathbb{N}$ with positive lower density.

Acknowledgments. I am grateful to my co-authors on articles [21–23], B. Host, O. Klurman, and J. Moreira; without them, these works would not have been possible. I would also like to thank A. Mountakis and D. Sclosa for a number of useful comments.

Funding. The author was supported by the research grant ELIDEK HFRI-Next-GenerationEU-15689.

References

- [1] R. Alweiss, Monochromatic sums and products over $\mathbb{Q}$. 2023, arXiv:2307.08901v5
- [2] R. Alweiss, [Monochromatic sums and products of polynomials](#). *Discrete Anal.* (2024), article no. 5, 7 pp. Zbl 08092730 MR 4793285
- [3] J. M. Barrett, M. Lupini, and J. Moreira, [On Rado conditions for nonlinear Diophantine equations](#). *European J. Combin.* **94** (2021), article no. 103277, 20 pp. Zbl 1468.11076 MR 4192111
- [4] V. Bergelson, [Ergodic Ramsey theory](#). In *Logic and combinatorics (Proc. AMS-IMS-SIAM Conf., Arcata, Calif., 1985)*, pp. 63–87, Contemp. Math. 65, American Mathematical Society, Providence, RI, 1987 Zbl 0642.10052 MR 0891243
- [5] V. Bergelson, [Multiplicatively large sets and ergodic Ramsey theory](#). *Israel J. Math.* **148** (2005), no. 1, 23–40 Zbl 1093.11003 MR 2191223
- [6] V. Bergelson, J. H. Johnson, Jr., and J. Moreira, [New polynomial and multidimensional extensions of classical partition results](#). *J. Combin. Theory Ser. A* **147** (2017), 119–154 Zbl 1352.05023 MR 3589892
- [7] V. Bergelson and A. Leibman, [Polynomial extensions of van der Waerden’s and Szemerédi’s theorems](#). *J. Amer. Math. Soc.* **9** (1996), no. 3, 725–753 Zbl 0870.11015 MR 1325795
- [8] M. Bowen and M. Sabok, [Monochromatic products and sums in the rationals](#). *Forum Math. Pi* **12** (2024), article no. e17, 12 pp. Zbl 1551.05400 MR 4812188
- [9] T. D. Browning and S. M. Prendiville, [A transference approach to a Roth-type theorem in the squares](#). *Int. Math. Res. Not. IMRN* **2017** (2017), no. 7, 2219–2248 Zbl 1405.11131 MR 3658196
- [10] J. Chapman, [Partition regularity for systems of diagonal equations](#). *Int. Math. Res. Not. IMRN* **2022** (2022), no. 17, 13272–13316 Zbl 1507.11016 MR 4475278

- [11] J. Chapman and S. Chow, [Generalised Rado and Roth criteria](#). *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **26** (2025), no. 3, 1263–1312 Zbl [08129930](#) MR [4991571](#)
- [12] D. Charamaras, A. Mountakis, and K. Tsinas, [On multiplicative recurrence along linear patterns](#). *J. Lond. Math. Soc. (2)* **112** (2025), no. 3, article no. e70292, 53 pp. Zbl [08104426](#) MR [4960398](#)
- [13] S. Chow, S. Lindqvist, and S. Prendiville, [Rado’s criterion over squares and higher powers](#). *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 6, 1925–1997 Zbl [1497.11028](#) MR [4244520](#)
- [14] P. Csikvári, K. Gyarmati, and A. Sárközy, [Density and Ramsey type results on algebraic equations with restricted solution sets](#). *Combinatorica* **32** (2012), no. 4, 425–449 Zbl [1286.11026](#) MR [2965285](#)
- [15] H. Daboussi and H. Delange, [On multiplicative arithmetical functions whose modulus does not exceed one](#). *J. Lond. Math. Soc. (2)* **26** (1982), no. 2, 245–264 Zbl [0499.10052](#) MR [0675168](#)
- [16] M. Di Nasso and L. Luperi Baglini, [Ramsey properties of nonlinear Diophantine equations](#). *Adv. Math.* **324** (2018), 84–117 Zbl [1376.05159](#) MR [3733882](#)
- [17] M. Di Nasso and M. Riggio, [Fermat-like equations that are not partition regular](#). *Combinatorica* **38** (2018), no. 5, 1067–1078 Zbl [1438.11075](#) MR [3884778](#)
- [18] S. Donoso, A. N. Le, J. Moreira, and W. Sun, [Additive averages of multiplicative correlation sequences and applications](#). *J. Anal. Math.* **149** (2023), no. 2, 719–761 Zbl [1522.37065](#) MR [4594405](#)
- [19] N. Frantzikinakis, Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90 Zbl [1425.37004](#) MR [3613710](#)
- [20] N. Frantzikinakis and B. Host, [Asymptotics for multilinear averages of multiplicative functions](#). *Math. Proc. Cambridge Philos. Soc.* **161** (2016), no. 1, 87–101 Zbl [1371.11134](#) MR [3505671](#)
- [21] N. Frantzikinakis and B. Host, [Higher order Fourier analysis of multiplicative functions and applications](#). *J. Amer. Math. Soc.* **30** (2017), no. 1, 67–157 Zbl [1355.11094](#) MR [3556289](#)
- [22] N. Frantzikinakis, O. Klurman, and J. Moreira, Partition regularity of generalized Pythagorean pairs. 2024, arXiv:[2407.08360v2](#)
- [23] N. Frantzikinakis, O. Klurman, and J. Moreira, [Partition regularity of Pythagorean pairs](#). *Forum Math. Pi* **13** (2025), article no. e5, 52 pp. Zbl [1558.05134](#) MR [4862834](#)
- [24] H. Furstenberg, [Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions](#). *J. Anal. Math.* **31** (1977), no. 1, 204–256 Zbl [0347.28016](#) MR [0498471](#)
- [25] W. T. Gowers, [A new proof of Szemerédi’s theorem](#). *Geom. Funct. Anal.* **11** (2001), no. 3, 465–588 Zbl [1028.11005](#) MR [1844079](#)
- [26] R. Graham, [Some of my favorite problems in Ramsey theory](#). In B. Landman, M. B. Nathanson, J. Nešetřil, R. J. Nowakowski and C. Pomerance (Eds.), *Combinatorial number theory*, pp. 229–236, De Gruyter, Berlin, 2007 Zbl [1124.05087](#) MR [2337049](#)

- [27] R. Graham, [Old and new problems and results in Ramsey theory](#). In E. Győri, G. O. H. Katona, L. Lovász, and G. Sági (Eds.), *Horizons of combinatorics*, pp. 105–118, Bolyai Soc. Math. Stud. 17, Springer, Berlin, 2008 Zbl [1178.05094](#) MR [2432529](#)
- [28] A. Granville and K. Soundararajan, [Large character sums: pretentious characters and the Pólya–Vinogradov theorem](#). *J. Amer. Math. Soc.* **20** (2007), no. 2, 357–384 Zbl [1210.11090](#) MR [2276774](#)
- [29] A. Granville and K. Soundararajan, [Pretentious multiplicative functions and an inequality for the zeta-function](#). In J.-M. De Koninck, A. Granville, and F. Luca (Eds.), *Anatomy of integers*, pp. 191–197, CRM Proc. Lecture Notes 46, American Mathematical Society, Providence, RI, 2008 Zbl [1187.11032](#) MR [2437976](#)
- [30] A. Granville and K. Soundararajan, Multiplicative number theory: the pretentious approach. Book manuscript in preparation
- [31] B. Green, 100 open problems. Manuscript
- [32] B. Green and T. Tao, [An inverse theorem for the Gowers \$U^3\(G\)\$ norm](#). *Proc. Edinb. Math. Soc.* (2) **51** (2008), no. 1, 73–153 Zbl [1202.11013](#) MR [2391635](#)
- [33] B. Green and T. Tao, [Quadratic uniformity of the Möbius function](#). *Ann. Inst. Fourier (Grenoble)* **58** (2008), no. 6, 1863–1935 Zbl [1160.11017](#) MR [2473624](#)
- [34] B. Green and T. Tao, [The Möbius function is strongly orthogonal to nilsequences](#). *Ann. of Math.* (2) **175** (2012), no. 2, 541–566 Zbl [1347.37019](#) MR [2877066](#)
- [35] B. Green, T. Tao, and T. Ziegler, [An inverse theorem for the Gowers \$U^{s+1}\[N\]\$ -norm](#). *Ann. of Math.* (2) **176** (2012), no. 2, 1231–1372 Zbl [1282.11007](#) MR [2950773](#)
- [36] B. J. Green and S. Lindqvist, [Monochromatic solutions to \$x + y = z^2\$](#) . *Canad. J. Math.* **71** (2019), no. 3, 579–605 Zbl [1439.11076](#) MR [3952491](#)
- [37] K. Gyarmati and I. Z. Ruzsa, [A set of squares without arithmetic progressions](#). *Acta Arith.* **155** (2012), no. 1, 109–115 Zbl [1293.11015](#) MR [2982433](#)
- [38] G. Halász, [Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen](#). *Acta Math. Acad. Sci. Hungar.* **19** (1968), 365–403 Zbl [0165.05804](#) MR [0230694](#)
- [39] M. J. H. Heule, O. Kullmann, and V. W. Marek, [Solving and verifying the Boolean Pythagorean triples problem via cube-and-conquer](#). In *Proceedings of the 19th International Conference on Theory and Applications of Satisfiability Testing—SAT 2016*, pp. 228–245, Lecture Notes in Comput. Sci. 9710, Springer, Cham, 2016 Zbl [1403.68226](#) MR [3534782](#)
- [40] I. Kátai, [A remark on a theorem of H. Daboussi](#). *Acta Math. Hungar.* **47** (1986), no. 1–2, 223–225 Zbl [0607.10034](#) MR [0836415](#)
- [41] A. Khalfalah and E. Szemerédi, [On the number of monochromatic solutions of \$x + y = z^2\$](#) . *Combin. Probab. Comput.* **15** (2006), no. 1–2, 213–227 Zbl [1093.11017](#) MR [2195583](#)
- [42] O. Klurman, [Correlations of multiplicative functions and applications](#). *Compos. Math.* **153** (2017), no. 8, 1622–1657 Zbl [1434.11202](#) MR [3705270](#)

- [43] O. Klurman and A. P. Mangerel, [Rigidity theorems for multiplicative functions](#). *Math. Ann.* **372** (2018), no. 1–2, 651–697 Zbl [1457.11131](#) MR [3856825](#)
- [44] O. Klurman, A. P. Mangerel, C. Pohoata, and J. Teräväinen, [Multiplicative functions that are close to their mean](#). *Trans. Amer. Math. Soc.* **374** (2021), no. 11, 7967–7990 Zbl [1487.11086](#) MR [4328688](#)
- [45] E. Lamb, [Two-hundred-terabyte maths proof is largest ever](#). *Nature* **534** (2016), no. 7605, 17–18
- [46] H. Lefmann, [On partition regular systems of equations](#). *J. Combin. Theory Ser. A* **58** (1991), no. 1, 35–53 Zbl [0746.05076](#) MR [1119700](#)
- [47] L. Matthiesen, [Generalized Fourier coefficients of multiplicative functions](#). *Algebra Number Theory* **12** (2018), no. 6, 1311–1400 Zbl [1444.11201](#) MR [3864201](#)
- [48] J. Moreira, [Monochromatic sums and products in \$\mathbb{N}\$](#) . *Ann. of Math. (2)* **185** (2017), no. 3, 1069–1090 Zbl [1430.05128](#) MR [3664819](#)
- [49] P. P. Pach, [Monochromatic solutions to \$x + y = z^2\$ in the interval \$\[N, cN^4\]\$](#) . *Bull. Lond. Math. Soc.* **50** (2018), no. 6, 1113–1116 Zbl [1439.11077](#) MR [3891947](#)
- [50] S. Prendiville, [Counting monochromatic solutions to diagonal Diophantine equations](#). *Discrete Anal.* (2021), article no. 14, 47 pp. Zbl [1511.11035](#) MR [4317277](#)
- [51] R. Rado, [Studien zur Kombinatorik](#). *Math. Z.* **36** (1933), no. 1, 424–470 Zbl [0006.14603](#) MR [1545354](#)
- [52] A. Sárközy, [On difference sets of sequences of integers. I](#). *Acta Math. Acad. Sci. Hungar.* **31** (1978), no. 1–2, 125–149 Zbl [0387.10033](#) MR [0466059](#)
- [53] I. Schur, [Über die Kongruenz \$x^m + y^m \equiv z^m \pmod{p}\$](#) . *Jahresber. Dtsch. Math.-Ver.* **25** (1916) 114–117 Zbl [46.0193.02](#)
- [54] W. Sun, [A structure theorem for multiplicative functions over the Gaussian integers and applications](#). *J. Anal. Math.* **134** (2018), no. 1, 55–105 Zbl [1418.11025](#) MR [3771476](#)
- [55] W. Sun, [Sarnak’s conjecture for nilsequences on arbitrary number fields and applications](#). *Adv. Math.* **415** (2023), article no. 108883, 91 pp. Zbl [1525.11105](#) MR [4543074](#)
- [56] T. Tao, [The logarithmically averaged Chowla and Elliott conjectures for two-point correlations](#). *Forum Math. Pi* **4** (2016), article no. e8, 36 pp. Zbl [1383.11116](#) MR [3569059](#)
- [57] T. Tao and J. Teräväinen, [The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures](#). *Duke Math. J.* **168** (2019), no. 11, 1977–2027 Zbl [1436.11115](#) MR [3992031](#)
- [58] G. Tenenbaum, [Introduction to analytic and probabilistic number theory](#). Cambridge Stud. Adv. Math. 46, Cambridge University Press, Cambridge, 1995, 448 pp. Zbl [0831.11001](#) MR [1342300](#)
- [59] B. L. van der Waerden, [Beweis einer Baudetschen Vermutung](#). *Nieuw Arch. Wiskd.* **15** (1927), 212–216 Zbl [53.0073.12](#)

Nikos Frantzikinakis

Department of Mathematics and Applied Mathematics, University of Crete, Voutes Campus,
70013 Heraklion, Greece; frantzikinakis@uoc.gr, frantzikinakis@gmail.com