

Arbitrarily long strings of consecutive primes in special sets

By SAI SANJEEV BALAKRISHNAN, FÉLIX HOUE,
VAHAGN HOVHANNISYAN, MARYNA MANSKOVA and YIQING WANG

August 26, 2025

Abstract

Let $F(x)$ be a function of the form $\sum_{i=1}^r d_i x^{\rho_i}$ where $d_1, \dots, d_r \in \mathbb{R}$, $0 \leq \rho_1 < \dots < \rho_r$, $\rho_r \notin \mathbb{Z}$, $\rho_i \in \mathbb{R}$ for $1 \leq i \leq r$ and $d_r \neq 0$. We prove that sets of the form $\{n \in \mathbb{N} : \{F(n)\} \in U\}$ for any non-empty open set $U \subseteq [0, 1)$ contain arbitrarily long strings of consecutive primes.

Contents

1	Introduction	1
1.1	Some notation	3
2	Main ideas of the proof	4
3	Uniform distribution of values of polynomials with real exponents in arithmetic progressions	7
4	Uniform distribution of values of polynomials with real exponents in arithmetic progressions along primes	17
5	Proof of the main theorem	29

1. Introduction

A natural question that arises in the study of the distribution of prime numbers is determining which sets $\mathcal{A} \subseteq \mathbb{N}$ contain arbitrarily long strings of consecutive primes $p_{k+1}, p_{k+2}, \dots, p_{k+m}$, where p_n is the n th prime number in $\mathbb{N}$. When $\mathcal{A}$ is an arithmetic progression of the form $a\mathbb{N}+b$ with $\gcd(a, b) = 1$, Shiu famously showed that $\mathcal{A}$ contains arbitrarily long strings of consecutive primes and gave a lower bound on the number of such strings in $(a\mathbb{N}+b) \cap [1, x)$ (see Theorem 1 and 2 of [Shi00]).

The recent breakthroughs of Zhang [Zha14] and Maynard ([May15] and [May16]) on small gaps between primes introduced new approaches to investigate these questions, that yielded a strengthening of Shiu's result in [May16, Theorem 3.3]. Using these

2020 *Mathematics Subject Classification*: Primary 11N05; Secondary 11N36.

Key words and phrases: consecutive primes, Maynard sieve, equidistribution mod 1.

methods, Shiu's result was further strengthened in [Rad21, Corollary 1.2], where it was shown that there exist arbitrarily long strings of primes in $(a\mathbb{N} + b)$, such that the difference of the largest and the smallest prime in the string is bounded by a fixed number. Another interesting application of the method yielded the existence of arbitrarily long strings of consecutive primes having a fixed integer $g \neq -1$, not a square, as a primitive root modulo those primes ([Pol14, Theorem 4.1]), conditional on the generalized Riemann hypothesis.

The existence of arbitrarily long strings of consecutive primes has also been investigated when $\mathcal{A}$ is a Beatty sequence satisfying certain conditions. For an irrational number $\alpha > 0$, the associated Beatty sequence is the set of integers

$$\mathcal{A} = \{\lfloor n\alpha \rfloor : n \in \mathbb{N}\},$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x . Baker and Zhao showed that the sequence of primes in any Beatty sequence associated with an irrational α has bounded gaps by using a variant of Maynard's method ([BZ16, Theorem 2]). Similarly, in [CPS15, Theorem 1.1], the same result was shown when α is of finite type, and, in addition, the proof implies that Beatty sequences for such α contain arbitrarily long strings of consecutive primes (see the remark after the proof of [CPS15, Theorem 1.1]). Yet another interesting result along these lines is [MPY13, Corollary 1.2], which shows that when α is of finite type, the consecutive primes can be chosen such that they all lie in a single arithmetic progression. Surprisingly, a proof that for all irrational $\alpha > 0$ the set $\{\lfloor n\alpha \rfloor : n \in \mathbb{N}\}$ contains arbitrarily long strings of consecutive primes has not appeared explicitly in the literature. However, as we will discuss next, we believe an even more general statement can be derived by modifying the recent results of Shao and Teräväinen.

Observe that any Beatty sequence can be written as a Bohr set, i.e.,

$$\{\lfloor n\alpha \rfloor : n \in \mathbb{N}\} = \{n \in \mathbb{N} : \{\alpha^{-1}n\} \in (1 - \alpha^{-1}, 1)\},$$

where we define $\{y\} = y - \lfloor y \rfloor$ for a real number y . Therefore, a natural question is whether one can generalize results of the above kind to sets of the form

$$\mathcal{A} = \{n \in \mathbb{N} : \{F(n)\} \in U\},$$

where F belongs to a family of polynomials and U is, say, an open subset of $[0, 1)$. Using a weighted variant of the Maynard sieve (see [MS17, Theorem 6.2]), Shao and Teräväinen showed that for every open set U , the primes in the set $\mathcal{A}$ have bounded gaps whenever F is a real polynomial with at least one irrational coefficient besides the constant term (see [ST21, Theorem 2.3]). In this case, $\mathcal{A}$ belongs to the family of so-called Nil-Bohr sets. Their method shows that for every $m \geq 1$, there are infinitely many intervals of bounded length containing m primes that lie inside the set $\mathcal{A}$, as noted in the footnote of [ST21, page 46]. Although not mentioned explicitly in their paper, it seems that their method also shows that those primes can be taken to be consecutive. Indeed, under some additional hypotheses, Maynard's theorem [May16, Theorem 3.1] gives a lower bound on the number of strings of consecutive primes found in a set of integers, and we believe that this should also carry over to the weighted

version [MS17, Theorem 6.2] used by Shao and Teräväinen.

Inspired by this, we prove a similar result for sets $\mathcal{A}$ of the form

$$\mathcal{A} = \{n \in \mathbb{N} : \{F(n)\} \in U\},$$

where U is an open set, and F belongs to a family of polynomial-like functions that we refer to as *polynomials with real exponents*.

Definition 1.1. A *polynomial with non-integer exponents*³ is a function of the form

$$f(x) = \sum_{i=1}^r d_i x^{\rho_i}$$

where $r \geq 1$, $0 < \rho_1 < \dots < \rho_r$, $\rho_i \notin \mathbb{Z}$, $\rho_i \in \mathbb{R}$ and $d_i \in \mathbb{R}$ for $1 \leq i \leq r$. A *polynomial with real exponents* is a function $F(x)$ of the form $f(x) + P(x)$, where $P(x)$ is a real polynomial and $f(x)$ is a polynomial with non-integer exponents.

We now state our main result.

Theorem 1.2. Let $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ with $0 \leq \rho_1 < \dots < \rho_r$ be a polynomial with real exponents, with $\rho_r \notin \mathbb{Z}$ and $d_r \neq 0$. Let $\mathcal{A} := \{n \in \mathbb{N} : \{F(n)\} \in U\}$ for some non-empty open set $U \subseteq [0, 1)$. Then $\mathcal{A}$ contains arbitrarily long strings of consecutive primes. Moreover, if $p_1 < p_2 < p_3 < \dots$ is the increasing enumeration of the prime numbers, then, for every positive integer m , there exists a constant $K = K(F, m)$ such that there are infinitely many n with $p_{n+1}, \dots, p_{n+m} \in \mathcal{A}$ and $p_{n+m} - p_{n+1} \leq K$.

By passing to a subset of $\mathcal{A}$ if necessary, it follows from the theorem that one can take U to be any subset of $[0, 1)$ with a nonempty interior. The restriction $\rho_r \notin \mathbb{Z}$ is technical and is explained in Section 5.

Acknowledgments

This work was done as part of the 2023 edition of the Young Researchers in Mathematics Program. We would like to thank the Bernoulli Center for Fundamental Studies for funding and hosting the program. We are very grateful to Florian Richter for creating this opportunity and guiding us throughout the process. We would also like to thank Nadia Kaiser and Carole Weissenberger, who helped organize our stay in Lausanne. We also want to thank the anonymous reviewer for their careful reading of the paper and their useful comments on the presentation.

1.1. Some notation

In what follows, we denote by $\mathbb{N}$ the set $\{1, 2, 3, \dots\}$ of positive integers, and by $\mathbb{P}$ the set of all prime numbers. We use the letter p in sums and products to indicate that they are over prime numbers. We let $(p_n)_{n \in \mathbb{N}}$ be the sequence of all prime numbers in increasing order. We let $\pi(N)$ be the number of prime numbers up to N , and, for

³In the literature one might find many other names for slightly different versions of these functions, *e.g.* fractional polynomials, exponential polynomials, polynomials with real exponents, signomials, etc.

$M \geq N$, we write $\pi(N, M) = \pi(M) - \pi(N)$. We write $\pi(N; a, q)$ for the number of prime numbers congruent to $a \pmod q$ up to N . Similarly, $\pi(N, M; a, q)$ denotes the number of primes between N and M that are congruent to $a \pmod q$. We write the greatest common divisor of some integers a and b as (a, b) . We write $\varphi(n)$ for Euler's totient function evaluated at n , and $\tau_j(n)$ for the number of ways to decompose n as a product of j positive integers. We also write $\Lambda(n)$ for the von Mangoldt function evaluated at n , that is

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k; \\ 0 & \text{otherwise.} \end{cases}$$

The indicator function of a set S evaluated at n is denoted by $\mathbf{1}_S(n)$. We shorten $e^{2\pi i x}$ as $e(x)$. We denote the ℓ -dimensional Lebesgue measure of a measurable set $E \subseteq \mathbb{R}^\ell$ by $\lambda_\ell(E)$. We write $\langle \cdot, \cdot \rangle$ for the dot product of $\mathbb{R}^\ell$ and denote $\bar{0} \in \mathbb{R}^\ell$ the zero vector. For real-valued functions F and G , we write $F(x) \ll G(x)$ or $F(x) = O(G(x))$ for $x \in S$ if there exists some constant $C > 0$ such that $|F(x)| \leq CG(x)$ for all $x \in S$. We refer to the constant C as the implied constant. If the constant C depends on some parameters $\alpha, \beta, \dots$, we write $F \ll_{\alpha, \beta, \dots} G$ or $F = O_{\alpha, \beta, \dots}(G)$. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ and $\lceil x \rceil$ to denote the largest integer no larger than x and the smallest integer no smaller than x respectively. We write $\{x\}$ for $x - \lfloor x \rfloor$. We also write $\|x\|$ for the distance between x and the closest integer. We denote the cardinality of a set I by $\#I$.

2. Main ideas of the proof

The main ingredient of our proof of Theorem 1.2 is a result of Maynard in [May16] about bounded gaps between (consecutive) primes in sets satisfying some technical hypotheses. We first state a useful definition.

Definition 2.1. A family of linear polynomials $\{L_1, \dots, L_d\}$ with $L_i(n) = a_i n + b_i$, $a_i \in \mathbb{N}$, $b_i \in \mathbb{N} \cup \{0\}$ is *admissible* if for every prime p , there exists some n such that $\prod_{i=1}^d L_i(n) \not\equiv 0 \pmod p$. We say that set of integers $\{h_1, \dots, h_d\}$ is admissible if the family of linear polynomials $\{n + h_1, \dots, n + h_d\}$ is admissible.

Theorem 2.2 ([May16], Theorem 3.1). *Let $\alpha > 0$ and $0 < \theta < 1$. Let $\mathcal{B}$ be a set of integers, $\mathcal{P}$ a subset of $\mathbb{P}$, and $\mathcal{L} = \{L_1, \dots, L_d\}$ an admissible family of d linear polynomials, where $L_i(n) = a_i n + b_i$, $a_i \in \mathbb{N}$, $b_i \in \mathbb{N} \cup \{0\}$. Let B and N be positive integers. Assume that $B \leq N^\alpha$, $a_i, b_i \leq N^\alpha$ for $1 \leq i \leq d$, and $d \leq (\log N)^\alpha$. Then, there exists a constant C depending only on α and θ such that the following holds when $d \geq C$:*

If there exists some $\delta > 1/\log d$ such that,

1. *The set $\{a_i n + b_i : n \in \mathcal{B}, 1 \leq i \leq d\}$ contains a large enough proportion of primes:*

$$\frac{1}{d} \frac{\varphi(B)}{B} \sum_{L_i \in \mathcal{L}} \frac{\varphi(a_i)}{a_i} \sum_{\substack{N \leq n < 2N \\ a_i n + b_i \in \mathcal{P}}} \mathbf{1}_{\mathcal{B}}(n) \geq \frac{\delta}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).$$

2. $\mathcal{B}$ is well-distributed in arithmetic progressions:

$$\sum_{\substack{q \leq N^\theta \\ (q, B)=1}} \max_{0 \leq c \leq q-1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll \frac{\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)}{(\log N)^{101d^2}}.$$

3. Primes represented by the linear forms in $\mathcal{L}$ are well-distributed in arithmetic progressions, that is, for any $L \in \mathcal{L}$, the following holds:

$$\sum_{q \leq N^\theta} \max_{c: (L(c), q)=1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q} \\ L(n) \in \mathcal{P}}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{\varphi_L(q)} \sum_{\substack{N \leq n < 2N \\ L(n) \in \mathcal{P}}} \mathbf{1}_{\mathcal{B}}(n) \right| \ll \frac{\sum_{N \leq n < 2N, L(n) \in \mathcal{P}} \mathbf{1}_{\mathcal{B}}(n)}{(\log N)^{101d^2}}.$$

where $\varphi_L(q) = \varphi(aq)/\varphi(a)$ for $L(n) = an + b$.

4. The set $\mathcal{B}$ is not too concentrated in any single arithmetic progression: for any $q \leq N^\theta$ and $0 \leq c \leq q-1$, the following holds:

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) \ll \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).$$

Then,

$$\#\{n \in \mathcal{B} \cap [N, 2N) : \#\{L_1(n), \dots, L_d(n)\} \cap \mathcal{P} \geq C^{-1} \delta \log d\} \gg \frac{\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)}{(\log N)^d \exp(Cd)}.$$

If, in addition, $\mathcal{P} = \mathbb{P}$, $k \leq (\log N)^{1/5}$, and $a_i = a \ll 1, b_i \leq (\log x)d^{-2}$, we have

$$\begin{aligned} \#\{n \in \mathcal{B} \cap [N, 2N) : \text{at least } \lceil C^{-1} \delta \log d \rceil \text{ of } \{L_1(n), \dots, L_d(n)\} \text{ are consecutive primes}\} \\ \gg \frac{\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)}{(\log N)^d \exp(Cd^5)}. \end{aligned}$$

In both estimates, the implied constants do not depend on N and d .

When $B = 1$, $a_i = 1$ for all i , $0 < \theta < 1/2$, $\mathcal{B} = \mathbb{N}$ and $\mathcal{P} = \mathbb{P}$, Hypotheses 1-4 of Theorem 2.2 are well known. Indeed, Hypothesis 1 is a consequence of the prime number theorem, Hypotheses 2 and 4 are obvious, and Hypothesis 3 follows from a form of the Bombieri-Vinogradov theorem.

Theorem 2.3 (Bombieri-Vinogradov). *For all real numbers $0 < \theta < \frac{1}{2}$ and $D > 0$, and every integer $N \in \mathbb{N}$ large enough in terms of θ and D , we have*

$$\sum_{q \leq N^\theta} \max_{(a, q)=1} \left| \pi(N; a, q) - \frac{1}{\varphi(q)} \pi(N) \right| \ll_D \frac{N}{(\log N)^D}. \quad (2.1)$$

Proof. This follows from [IK04, Theorem 17.1] by partial summation. $\square$

To prove Theorem 1.2, we show that for every $m \in \mathbb{N}$, we can find some $d > m$ and some integers $h_1 < h_2 < \dots < h_d$ for which there exists a subset $\mathcal{B} \subseteq \{n \in \mathbb{N} : \{F(n + h_1)\}, \dots, \{F(n + h_d)\} \in U\}$ which satisfies the hypotheses of Theorem 2.2 for the linear polynomials $L_i(n) = n + h_i$, $1 \leq i \leq d$. In particular, for every $N \in \mathbb{N}$ large enough, there exists some

$$n \in [N, 2N) \cap \mathcal{B}$$

for which $\{n + h_1, \dots, n + h_d\}$ contains at least m consecutive primes. Notice that $n \in \mathcal{B}$ implies $n \in \{n \in \mathbb{N} : \{F(n + h_1)\}, \dots, \{F(n + h_d)\} \in U\}$ which in turn implies that these primes will necessarily lie in the set $\mathcal{A}$ from the statement of Theorem 1.2. We show that we can choose the integers $h_1, \dots, h_d$ so that the aforementioned set $\mathcal{B}$ can be taken to be

$$\mathcal{B} = \{n \in \mathbb{N} : (\{F(n + h_1)\}, \dots, \{F(n + h_\ell)\}) \in U_\ell\},$$

where U_ℓ is a semi-open box in $[0, 1)^\ell$, and ℓ is a well chosen integer less than d . To verify the hypotheses of Theorem 2.2, we use the fact that for this choice of ℓ , the sequence $(\{F(n + h_1)\}, \dots, \{F(n + h_\ell)\})$ is well distributed in $[0, 1)^\ell$. A quantitative measure of how well a finite sequence is distributed in $[0, 1)^\ell$ is given by its discrepancy.

Definition 2.4. The discrepancy of a finite sequence $(\mathbf{x}_n)_{n=1}^N$ in $[0, 1)^\ell$ is the quantity

$$D((\mathbf{x}_n)_{n=1}^N) = \sup_{B \subseteq [0, 1)^\ell} \left| \frac{1}{N} \sum_{n=1}^N \mathbf{1}_B(\mathbf{x}_n) - \lambda_\ell(B) \right|,$$

where the supremum is taken over semi-open boxes of the form $B = \prod_{i=1}^\ell [a_i, b_i)$, where $0 \leq a_i < b_i < 1$ for each i .

The Erdős-Turán-Koksma inequality gives a bound on the discrepancy of a sequence in terms of exponential sums.

Theorem 2.5 (Erdős-Turán-Koksma inequality, [DT97], Theorem 1.21). *For every $H \in \mathbb{N}$ and for every finite sequence $(\mathbf{x}_n)_{n=1}^N$ in $[0, 1)^\ell$, we have*

$$D((\mathbf{x}_n)_{n=1}^N) \ll_\ell \left(\frac{1}{H+1} + \sum_{\substack{\mathbf{m} \in \mathbb{Z}^\ell \setminus \{\mathbf{0}\} \\ \|\mathbf{m}\|_\infty \leq H}} \frac{1}{r(\mathbf{m})} \left| \frac{1}{N} \sum_{n=1}^N e(\langle \mathbf{x}_n, \mathbf{m} \rangle) \right| \right),$$

where if $\mathbf{m} = (m_1, \dots, m_\ell)$, we define $\|\mathbf{m}\|_\infty := \max_{1 \leq i \leq \ell} |m_i|$, and $r(\mathbf{m}) := \prod_{i=1}^\ell \max\{|m_i|, 1\}$.

In Section 3, we study the distribution of sequences of the form $(\{F(n + h_1)\}, \dots, \{F(n + h_\ell)\})$ in $[0, 1)^\ell$, where F is as in the statement of Theorem 1.2, $\ell = \lfloor \rho_r \rfloor + 1$, and where n runs over the integers in some dyadic interval $[N, 2N)$ or over the integers in $[N, 2N)$ that lie in an arithmetic progression. This allows us to verify Hypotheses 2 and 4 of Theorem 2.2 in the proof of Theorem 1.2. Then, in Section 4, we develop the tools to verify the Hypotheses 1 and 3 by studying the distribution of the

same sequence but over primes in $[N, 2N)$, as well as the primes $[N, 2N)$ restricted to an arithmetic progression. The restriction $\rho_r \notin \mathbb{Z}$ to the polynomial with real exponents $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ ensures the existence of good upper bounds on exponential sums of the form

$$\sum_{N \leq n < 2N} e(\langle (F(n+h_1), \dots, F(n+h_\ell)), \mathbf{m} \rangle)$$

for $\mathbf{m}$ a vector of ℓ integers, which is crucial for our argument. This is explained in more detail at the beginning of Section 5, and especially in Remark 5.1. Finally, in Section 5, we combine the results of the preceding sections to prove Theorem 1.2.

3. Uniform distribution of values of polynomials with real exponents in arithmetic progressions

In this section, we prove results that will help us verify Hypotheses 2 and 4 of Theorem 2.2 (Proposition 3.5 and Corollary 3.6). To do this, we study the distribution of sequences of the form $\overline{F}(n) = (\{F(n+h_1)\}, \dots, \{F(n+h_\ell)\})$ in $[0, 1]^\ell$ for F as in the statement of Theorem 1.2. Theorem 2.5 allows us to bound the discrepancy of such sequences by exponential sums over sequences of the form $\langle \mathbf{m}, \overline{F}(n) \rangle$ where $\mathbf{m}$ is a nonzero vector of integers. We first prove in Proposition 3.1 that $\langle \mathbf{m}, \overline{F} \rangle$ can be approximated by polynomial-like functions with real exponents. Then, in Lemma 3.2, we reprove a bound on exponential sums from [BKM⁺14], and we specialize it to our setting in Proposition 3.3 and Corollary 3.4.

Proposition 3.1. *Let $\mathcal{H} = \{h_1, \dots, h_\ell\}$ be a set of real numbers, with $h_1 < h_2 < \dots < h_\ell$, and let $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ with $0 \leq \rho_1 < \dots < \rho_r, \rho_r \notin \mathbb{Z}, d_r \neq 0$ be a polynomial with real exponents, such that $\ell = \lfloor \rho_r \rfloor + 1$. Define $\overline{F}(x) := (F(x+h_1), \dots, F(x+h_\ell))$. Then, for any $\eta > 0$, any integer N and any $\mathbf{m} = (m_1, \dots, m_\ell) \in \mathbb{Z}^\ell \setminus \{\overline{0}\}$ with $\max_{1 \leq j \leq \ell} |m_j| \leq N^\eta$, we can write*

$$\langle \mathbf{m}, \overline{F}(x) \rangle = F_1(x) + O_{F, \mathcal{H}}(x^{\{\rho_r\} + \eta - 1}),$$

where F_1 is a function of the form $F_1(x) = \sum_{n=n_0}^{\ell-1} \sum_{i=1}^r b_{n,i} x^{\rho_i - n}$, with the coefficient of the largest power of x in the expression being $b_{n_0, r} = \left(\sum_{j=1}^\ell m_j h_j^{n_0} \right) d_r \binom{\rho_r}{n_0}$ for some integer $0 \leq n_0 \leq \ell - 1$. Moreover, the coefficients $b_{n,j}$ are all uniformly bounded by

$$\ell^2 r \max_{\substack{0 \leq n \leq \ell-1 \\ 1 \leq j \leq \ell \\ 1 \leq i \leq r}} \left(\left| h_j^n d_i \binom{\rho_i}{n} \right| \right) N^\eta,$$

where

$$\binom{\rho}{0} = 1, \quad \binom{\rho}{n} = \frac{\rho(\rho-1)\dots(\rho-n+1)}{n!} \quad \text{for } n \geq 1.$$

Proof. By Taylor's theorem, we have that that for $|y| < 1$ and $\rho \in \mathbb{R}$,

$$(1+y)^\rho = \sum_{n=0}^{\ell-1} \binom{\rho}{n} y^n + O_\rho(y^\ell).$$

It follows that for $j = 1, 2, \dots, \ell$ and $x > h_\ell$, $x \geq 1$,

$$\begin{aligned} F(x+h_j) &= \sum_{i=1}^r d_i (x+h_j)^{\rho_i} \\ &= \sum_{i=1}^r d_i x^{\rho_i} \left(\sum_{n=0}^{\ell-1} \binom{\rho_i}{n} \left(\frac{h_j}{x} \right)^n + O_{\rho_i} \left(\left(\frac{h_j}{x} \right)^\ell \right) \right) \\ &= \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} h_j^n x^{\rho_i-n} + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}). \end{aligned}$$

For a fixed $\mathbf{m} = (m_1, \dots, m_\ell) \in \mathbb{Z}^\ell \setminus \{\bar{0}\}$, we can write

$$\begin{aligned} \langle \mathbf{m}, \bar{F}(x) \rangle &= \sum_{j=1}^{\ell} m_j F(x+h_j) \\ &= \sum_{j=1}^{\ell} m_j \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} h_j^n x^{\rho_i-n} + O_{F,\mathcal{H}} \left(x^{\{\rho_r\}-1} \sum_{j=1}^{\ell} |m_j| \right) \\ &= \sum_{n=0}^{\ell-1} \left(\sum_{j=1}^{\ell} m_j h_j^n \right) \sum_{i=1}^r d_i \binom{\rho_i}{n} x^{\rho_i-n} + O_{F,\mathcal{H}}(x^{\{\rho_r\}+\eta-1}). \end{aligned}$$

Define

$$\begin{aligned} F_1(x) &= \sum_{n=0}^{\ell-1} \left(\sum_{j=1}^{\ell} m_j h_j^n \right) \sum_{i=1}^r d_i \binom{\rho_i}{n} x^{\rho_i-n} \\ &= \sum_{n=1}^{\ell-1} \sum_{i=1}^r b_{n,i} x^{\rho_i-n}. \end{aligned}$$

The term with the highest power of x in this expression is

$$b_{n_0,r} x^{\rho_r-n_0} = \left(\sum_{j=1}^{\ell} m_j h_j^{n_0} \right) d_r \binom{\rho_r}{n_0} x^{\rho_r-n_0},$$

where n_0 is the smallest integer $0 \leq n_0 \leq \ell-1$ such that $\sum_{j=1}^{\ell} m_j h_j^{n_0}$ is nonzero. This

n_0 exists since $(m_1, \dots, m_\ell)$ is a non-zero vector of integers and if we write

$$\nu(h) = \begin{pmatrix} 1 \\ h \\ \vdots \\ h^{\ell-1} \end{pmatrix} \in \mathbb{R}^\ell,$$

then the vectors $\nu(h_1), \dots, \nu(h_\ell)$ are all linearly independent. This follows, for example, from the fact that they are the columns of the Vandermonde matrix

$$V = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ h_1 & h_2 & h_3 & \cdots & h_\ell \\ h_1^2 & h_2^2 & h_3^2 & \cdots & h_\ell^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_1^{\ell-1} & h_2^{\ell-1} & h_3^{\ell-1} & \cdots & h_\ell^{\ell-1} \end{pmatrix},$$

and $\det(V) = \prod_{1 \leq i < j \leq \ell} (h_j - h_i) \neq 0$.

Also, the absolute value of any of the coefficients of $F_1(x)$ is certainly bounded by the sum of the absolute values of all the coefficients, which in turn, by the triangle inequality, is bounded by

$$\sum_{n=0}^{\ell-1} \left(\sum_{j=1}^{\ell} |m_j| |h_j|^n \right) \sum_{i=1}^r |d_i| \left| \binom{\rho_i}{n} \right| \leq \ell^2 r \max_{\substack{0 \leq n \leq \ell-1 \\ 1 \leq j \leq \ell \\ 1 \leq i \leq r}} \left(|h_j^n d_i \binom{\rho_i}{n}| \right) N^\eta.$$

□

Next, we prove an upper bound on exponential sums of the form

$$\sum_{X \leq x < 2X} e(f(x) + P(x)),$$

where P is a polynomial whose degree is bounded, and f is a function whose derivatives satisfy certain growth conditions.

Lemma 3.2. *Let $k, q \in \mathbb{N}$, and set $K = 2^k$ and $Q = 2^q$. Let $0 < c_1 < c_2$. For every $X \in \mathbb{N}$, $X \geq K + 1$, every real number $g \in [1, X^{q+2}]$, every real-valued function $f(x)$ that is $(q + k + 2)$ -times continuously differentiable on $[X, 2X]$ and satisfies $c_1 g X^{-r} \leq |f^{(r)}(x)| \leq c_2 g X^{-r}$ for $r = 1, \dots, q + k + 2$, and every real polynomial $P(x)$ of degree at most k , one has*

$$\left| \sum_{X \leq x < 2X} e(f(x) + P(x)) \right| \ll_{c_1, c_2, q, k} X^{1 - \frac{1}{K}} + X \left(\frac{(\log X)^k}{g} \right)^{1/K} + X \left(\frac{g}{X^{q+2}} \right)^{\frac{1}{4QK-2K}}.$$

Proof. This lemma is a reformulation of [BKM⁺14, Lemma 2.5]. We reprove it here for the convenience of the reader. Lemma 2.1 of [BKM⁺14] with $H_i = \frac{X}{K}$ for $i = 1, \dots, k$

and $S = \left| \sum_{X \leq x < 2X} e(f(x) + P(x)) \right|$ gives us

$$\begin{aligned} \left(\frac{S}{X-1} \right)^K &\leq 8^{K-1} \left(\frac{1}{H_1^{K/2}} + \dots + \frac{1}{H_k} + \frac{1}{H_1 \dots H_k (X-1)} \sum_{h_1=1}^{H_1} \dots \sum_{h_k=1}^{H_k} \left| \sum_{x \in I(\mathbf{h})} e(f_1(x)) \right| \right) \\ &\ll_k \frac{1}{X} + \frac{1}{H_1 \dots H_k (X-1)} \sum_{h_1=1}^{H_1} \sum_{h_2=1}^{H_2} \dots \sum_{h_k=1}^{H_k} \left| \sum_{x \in I(\mathbf{h})} e(f_1(x)) \right|. \end{aligned}$$

where $f_1(x) = h_1 \dots h_k \left[\int_0^1 \dots \int_0^1 \frac{\partial^k}{\partial x^k} f(x + \langle \mathbf{h}, \mathbf{t} \rangle) d\mathbf{t} + a_k k! \right]$, $\mathbf{h} = (h_1, \dots, h_k)$, a_k is the coefficient of x^k in $P(x)$ (if the degree of $P(x)$ is less than k , then $a_k = 0$), and $I(\mathbf{h}) = (X, 2X - 1 - h_1 - \dots - h_k]$. Here, the variable $\mathbf{t}$ is a vector in $[0, 1]^k$, and the measure is the usual Lebesgue measure.

We have that $f_1(x)$ has $q+2$ continuous derivatives on $I(\mathbf{h})$, and

$$f_1^{(r)}(x) = h_1 \dots h_k \left[\int_0^1 \dots \int_0^1 \frac{\partial^{k+r}}{\partial x^{k+r}} f(x + \langle \mathbf{h}, \mathbf{t} \rangle) d\mathbf{t} \right]$$

for $r = 1, \dots, q+2$. This, combined with the hypothesis that $c_1 g X^{-r} \leq |f^{(r)}(x)| \leq c_2 g X^{-r}$, gives us the inequalities

$$c_1 g h_1 \dots h_k X^{-(k+r)} \leq |f_1^{(r)}(x)| \leq c_2 g h_1 \dots h_k X^{-(k+r)}$$

for $r = 1, \dots, q+2$. Hence, the function $f_1(x)$ satisfies the conditions of Lemma 2.2 in [BKM⁺14], with $G = \frac{g h_1 \dots h_k}{X^k}$ and $I = I(\mathbf{h})$. It follows that

$$\left| \sum_{x \in I(\mathbf{h})} e(f_1(x)) \right| \ll_{c_1, c_2, q} G^{\frac{1}{4Q-2}} X^{1-\frac{q+2}{4Q-2}} + \frac{X}{G}.$$

Hence, we obtain

$$\left(\frac{S}{X-1} \right)^K \ll_{c_1, c_2, q, k} \frac{1}{X} + \frac{1}{H_1 \dots H_k (X-1)} \sum_{h_1=1}^{H_1} \sum_{h_2=1}^{H_2} \dots \sum_{h_k=1}^{H_k} \left(G^{\frac{1}{4Q-2}} X^{1-\frac{q+2}{4Q-2}} + \frac{X}{G} \right).$$

Now, we have $G = \frac{g h_1 \dots h_k}{X^k} \leq \frac{g H_1 \dots H_k}{X^k} = \frac{g}{(K)^k}$. With this upper bound, the first term in the innermost sum is independent of the h_i 's. To evaluate the second term, we have

$$\sum_{h_1=1}^{H_1} \dots \sum_{h_k=1}^{H_k} \frac{1}{G} = \frac{X^k}{g} \sum_{h_1=1}^{H_1} \dots \sum_{h_k=1}^{H_k} \frac{1}{h_1 \dots h_k} \ll_k \frac{X^k}{g} \log H_1 \dots \log H_k \leq \frac{X^k}{g} (\log X)^k.$$

By using all of these bounds and substituting $H_i = \frac{X}{K}$, we get

$$\begin{aligned} \left(\frac{S}{X-1} \right)^K &\ll_{c_1, c_2, q, k} \frac{1}{X} + \left(\frac{g}{K^k X^{q+2}} \right)^{\frac{1}{4Q-2}} + \frac{K^k}{g} (\log X)^k \\ &\ll_{c_1, c_2, q, k} \frac{1}{X} + \left(\frac{g}{X^{q+2}} \right)^{\frac{1}{4Q-2}} + \frac{(\log X)^k}{g}. \end{aligned}$$

Thus,

$$S \ll_{c_1, c_2, q, k} X^{1-\frac{1}{K}} + X \left(\frac{g}{X^{q+2}} \right)^{\frac{1}{4QK-2K}} + X \left(\frac{(\log X)^k}{g} \right)^{1/K}.$$

□

For the rest of this paper, we fix $\eta = \min\{\frac{1-\{\rho_r\}}{2}, \rho_r - \rho_{r-1}\}/2$. We now use Proposition 3.1 together with Lemma 3.2 to get an upper bound on the exponential sums that appear when we apply the Erdős-Turán-Koksma inequality (Theorem 2.5) later in the section.

Proposition 3.3. *Let $k, r \in \mathbb{N}$ and fix a polynomial with non-integer exponents $f(x) = \sum_{j=1}^r d_j x^{\rho_j}$ with $d_r \neq 0$ and $0 < \rho_1 < \rho_2 < \dots < \rho_r$. Let $\ell = \lfloor \rho_r \rfloor + 1$, and let $\mathcal{H} = \{h_1, \dots, h_\ell\}$ be a set of integers with $h_1 < h_2 < \dots < h_\ell$. Define $\bar{f}(x) := (f(x+h_1), \dots, f(x+h_\ell))$. Then, there exists some $\alpha > 0$ depending only on k and ρ_r such that uniformly over all real polynomials P of degree at most k , all $N \in \mathbb{N}$ large enough in terms of $f, \mathcal{H}$ and k , and all $\mathbf{m} = (m_1, \dots, m_\ell) \in \mathbb{Z}^\ell \setminus \{0\}$ with $\max_{1 \leq j \leq \ell} |m_j| \leq N^\eta$,*

$$\left| \sum_{N \leq n < 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| \ll_{f, \mathcal{H}, k} N^{1-\alpha}. \quad (3.1)$$

Proof. By Proposition 3.1, we can write

$$\langle \mathbf{m}, \bar{f}(x) \rangle = f_1(x) + R(x),$$

where f_1 is of the form $\sum_{n=n_0}^{\ell-1} \sum_{i=1}^r b_{n,i} x^{\rho_i-n}$, with $b := b_{n_0, r} = \left(\sum_{j=1}^\ell m_j h_j^{n_0} \right) d_r \binom{\rho_r}{n_0} \neq 0$, and

$$R(x) \ll_{f, \mathcal{H}} x^{\{\rho_r\}-1+\eta}.$$

Notice that our choice of η ensures that $\{\rho_r\} - 1 + \eta < 0$. By the triangle inequality, we have that for any polynomial P ,

$$\begin{aligned} \left| \sum_{N \leq n < 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| &\leq \left| \sum_{N \leq n < 2N} e(f_1(n) + P(n)) \right| \\ &\quad + \left| \sum_{N \leq n < 2N} e(f_1(n) + P(n) + R(n)) - \sum_{N \leq n < 2N} e(f_1(n) + P(n)) \right| \\ &= \left| \sum_{N \leq n < 2N} e(f_1(n) + P(n)) \right| \\ &\quad + \left| \sum_{N \leq n < 2N} e(f_1(n) + P(n)) (e(R(n)) - 1) \right|. \end{aligned}$$

Call the first term S_1 and the second term S_2 . We have that for any $N \geq 1$,

$$\begin{aligned}
S_2 &\leq \sum_{N \leq n < 2N} |e(f_1(n) + P(n))| |e(R(n)) - 1| \\
&= \sum_{N \leq n < 2N} |e(R(n)) - 1| \\
&= \sum_{N \leq n < 2N} \left| \sum_{k=1}^{\infty} \frac{(2\pi i)^k}{k!} (R(n))^k \right| \\
&\ll_{f, \mathcal{H}} \sum_{k=1}^{\infty} \frac{(2\pi)^k}{k!} \sum_{N \leq n < 2N} n^{\{\rho_r\} - 1 + \eta} \\
&\ll_{f, \mathcal{H}} N^{\{\rho_r\} + \eta} \\
&\ll_{f, \mathcal{H}} N^{\frac{1 + \{\rho_r\}}{2}}.
\end{aligned}$$

To get an upper bound on S_1 , we use Lemma 3.2. The function f_1 can be written as

$$f_1(x) = bx^{\rho_r - n_0} + \sum_{i=1}^{r'} b_i x^{\rho'_i},$$

where $\rho'_1 < \rho'_2 < \dots < \rho'_{r'} < \rho_r - n_0$, $|b|, |b_i| \ll_{f, \mathcal{H}} N^\eta$, and $r' \leq \ell r$. This function is smooth on $(0, \infty)$, and for every $s \in \mathbb{N} \cup \{0\}$, we have that

$$f_1^{(s)}(x) = b P_s^{\rho_r - n_0} x^{\rho_r - n_0 - s} + \sum_{i=1}^{r'} b_i P_s^{\rho'_i} x^{\rho'_i - s},$$

where $P_n^\rho = \rho(\rho - 1) \dots (\rho - n + 1)$ for $\rho \in \mathbb{R}$ and $n \in \mathbb{Z}^+$. Using this fact together with the uniform bound for the b_i 's from the conclusion of Proposition 3.1, we get that for each $0 \leq s \leq \ell + k + 1$,

$$\left| \sum_{i=1}^{r'} b_i P_s^{\rho'_i} x^{\rho'_i - s} \right| \ll_{f, \mathcal{H}, k} x^{\max\{\rho_{r-1} - n_0, \rho_r - n_0 - 1\} + \eta - s}.$$

Since $\eta < \min\{\frac{1 - \{\rho_r\}}{2}, \rho_r - \rho_{r-1}\}$, there exists some integer $N_0 = N_0(f, \mathcal{H}, \eta)$ such that for any $N_0 \leq N \leq x \leq 2N$ and any $0 \leq s \leq \ell + k + 1$, we have

$$c_1 |b| N^{\rho_r - n_0} N^{-s} \leq |f_1^{(s)}(x)| \leq c_2 |b| N^{\rho_r - n_0} N^{-s},$$

where

$$c_1 = \frac{1}{2} \min_{\substack{0 \leq n \leq \ell - 1 \\ 0 \leq t \leq \ell + k + 1}} |P_t^{\rho_r - n}|, \quad c_2 = 2^{\rho_r + 1} \max_{\substack{0 \leq n \leq \ell - 1 \\ 0 \leq t \leq \ell + k + 1}} |P_t^{\rho_r - n}|.$$

Notice that since $\rho_r \notin \mathbb{Z}$, we have that $c_1, c_2 > 0$. Also, notice that c_1 and c_2 depend only on f and k . Therefore, by Lemma 3.2 with k as the degree of the polynomial

$P(x)$ and $q = \ell - 1$, there exists some constant $C' > 0$ that depends only on f and k such that for every polynomial P of degree at most k and every $N \geq N_0$,

$$S_1 \ll_{f, \mathcal{H}, k} \left(N^{1-\frac{1}{2k}} + N \left(\frac{(\log N)^k}{bN^{\{\rho_r\}}} \right)^{1/2k} + N \left(\frac{|b|N^{\rho_r}}{N^{\ell+1}} \right)^{\frac{1}{2k+\ell+1-2k}} \right). \quad (3.2)$$

Since $b = \left(\sum_{j=1}^{\ell} m_j h_j^{n_0} \right) d_r \left(\frac{\rho_r}{n_0} \right) \neq 0$ for some $n_0 \in \{0, \dots, \ell-1\}$ where $\mathbf{m} = (m_1, \dots, m_{\ell}) \in \mathbb{Z}^{\ell} \setminus \{\bar{0}\}$ and $\max_{1 \leq j \leq \ell} |m_j| \leq N^{\eta}$, we have

$$1 \ll_{f, \mathcal{H}} |b| = \left(\sum_{j=1}^{\ell} m_j h_j^n \right) d_r \left(\frac{\rho_r}{n} \right) \ll_{f, \mathcal{H}} N^{\eta} < N.$$

Also, for any $\varepsilon > 0$, $(\log N)^{k/K} \ll_{k, \varepsilon} N^{\varepsilon}$. Combining all of this, we get

$$S_1 \ll_{f, \mathcal{H}, k, \varepsilon} \left(N^{1-\frac{1}{2k}} + N^{1-\left(\frac{\{\rho_r\}}{2k}\right)-\varepsilon} + N^{1-\frac{1-\{\rho_r\}}{2k+\ell+1-2k+1}} \right).$$

Choosing $\varepsilon > 0$ small enough and combining this with the estimate on S_2 yields the desired conclusion for some $\alpha > 0$ depending only on f and k . $\square$

The uniformity over all polynomials of a bounded degree in Proposition 3.3 allows us to restrict the sum on the left side of (3.1) to a certain congruence class without changing the upper bound.

Corollary 3.4. *Under the hypotheses of Proposition 3.3, uniformly over all real polynomials P of degree at most k , all $N \in \mathbb{N}$ large enough depending on f , $\mathcal{H}$ and k , all $\mathbf{m} = (m_1, \dots, m_{\ell}) \in \mathbb{Z}^{\ell} \setminus \{\bar{0}\}$ with $\max_{1 \leq j \leq \ell} |m_j| \leq N^{\eta}$, all $q \in \mathbb{N}$, and all $a \in \{0, 1, \dots, q-1\}$, we have*

$$\left| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| \ll_{f, \mathcal{H}, k} N^{1-\alpha},$$

where α and the implied constant are the same as in Proposition 3.3.

Proof. Recall that

$$\mathbf{1}_{q\mathbb{Z}+a}(n) = \frac{1}{q} \sum_{0 \leq t \leq q-1} e\left(\frac{(n-a)t}{q}\right).$$

Therefore, we have

$$\left| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| = \left| \sum_{N \leq n < 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \mathbf{1}_{q\mathbb{Z}+a}(n) \right|$$

$$\begin{aligned}
&= \left| \sum_{N \leq n < 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \left(\frac{1}{q} \sum_{0 \leq t \leq q-1} e\left(\frac{(n-a)t}{q}\right) \right) \right| \\
&\leq \frac{1}{q} \sum_{0 \leq t \leq q-1} \left| \sum_{N \leq n < 2N} e\left(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n) + \frac{(n-a)t}{q}\right) \right| \\
&= \frac{1}{q} \sum_{0 \leq t \leq q-1} \left| \sum_{N \leq n < 2N} e\left(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n) + \frac{tn}{q}\right) \right|.
\end{aligned}$$

Hence, the result follows from Proposition 3.3. $\square$

Now, we prove some key results that will allow us to prove Hypotheses 2 and 4 of Theorem 2.2 in Section 5. In fact, the following result is stronger than what is needed for Hypothesis 2.

Proposition 3.5. *Let $P(x)$ be a real polynomial of degree at most $k \geq 1$ and let $f(x) = \sum_{j=1}^r d_j x^{\rho_j}$ be a polynomial with non-integer exponents with $0 < \rho_1 < \rho_2 < \dots < \rho_r$ and $d_r \neq 0$. Define $\ell = \lfloor \rho_r \rfloor + 1$. Then, there exist some constants $0 < \beta, \theta < 1$, depending only on $f(x)$ and k such that the following holds:*

Let $\mathcal{H} = \{h_1, \dots, h_\ell\}$ be a set of integers with $h_1 < \dots < h_\ell$, and let $U_\ell = \prod_{i=1}^\ell [u_i, v_i)$ with $0 \leq u_i < v_i \leq 1$. Then, for any integer $N \in \mathbb{N}$ such that $h_\ell \leq N$, if we write

$$\mathcal{B} = \{n \in \mathbb{N} : (\{f(n+h_1) + P(n+h_1)\}, \dots, \{f(n+h_\ell) + P(n+h_\ell)\}) \in U_\ell\},$$

then,

$$\sum_{q \leq N^\theta} \max_{0 \leq c \leq q-1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll_{f, \mathcal{H}, k, U_\ell} \frac{1}{N^\beta} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n). \quad (3.3)$$

Proof. Let $\mathbf{x}_n = (\{f(n+h_1) + P(n+h_1)\}, \dots, \{f(n+h_\ell) + P(n+h_\ell)\})$. By ℓ -dimensional Erdős-Turán-Koksma inequality (Theorem 2.5), we have

$$D((\bar{f}(n) + \bar{P}(n))_{n \in [N, 2N)}) \ll_\ell \left(\frac{1}{H+1} + \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \left| \frac{1}{N} \sum_{N \leq n \leq 2N} e(\langle \mathbf{m}, \mathbf{x}_n \rangle) \right| \right)$$

for any integer H , where J is the set of all semi open rectangles of the form $\prod_{i=1}^\ell [a_i, b_i)$. We take $H = \lfloor N^\eta \rfloor$. Notice that we have the following:

$$\begin{aligned}
\left| \sum_{N \leq n \leq 2N} e(\langle \mathbf{m}, \mathbf{x}_n \rangle) \right| &= \left| \sum_{N \leq n \leq 2N} e(\langle \mathbf{m}, \bar{f}(n) + \bar{P}(n) \rangle) \right| \\
&= \left| \sum_{N \leq n \leq 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + \langle \mathbf{m}, \bar{P}(n) \rangle) \right|,
\end{aligned}$$

where $\bar{f}(n) = (f(n+h_1), \dots, f(n+h_\ell))$ and $\bar{P}(n) = (P(n+h_1), \dots, P(n+h_\ell))$. Since $\langle \mathbf{m}, \bar{P}(n) \rangle$ is again a polynomial of degree at most k , by using Proposition 3.3 and the fact that $0 < \|\mathbf{m}\|_\infty \leq H \leq N^\eta$, we get that

$$\begin{aligned} D((\bar{f}(n) + \bar{P}(n))_{n \in [N, 2N]}) &\ll_{f, \mathcal{H}, k} \left(\frac{1}{H+1} + \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \frac{1}{N} N^{1-\beta''} \right) \\ &= \left(\frac{1}{H+1} + N^{-\beta''} \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \right) \end{aligned}$$

for some $0 < \beta'' < 1/2$, depending only on $f(x)$ and the degree of $P(x)$.

Now, it remains to bound the sum $\sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})}$. We have that

$$\begin{aligned} \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} &= \sum_{\substack{L \subseteq \{1, \dots, \ell\} \\ L \neq \emptyset}} \sum_{\substack{0 < |m_j| \leq H \\ j \in L}} \prod_{j \in L} \frac{1}{|m_j|} \\ &= \sum_{\substack{L \subseteq \{1, \dots, \ell\} \\ L \neq \emptyset}} \prod_{j \in L} \sum_{0 < |m_j| \leq H} \frac{1}{|m_j|} \\ &\ll \sum_{\substack{L \subseteq \{1, \dots, \ell\} \\ L \neq \emptyset}} (\log H)^{|L|}, \end{aligned}$$

so, we have

$$\sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \ll_\ell (\log H)^\ell \ll_{f, k} N^{\beta''/2}.$$

Putting everything together, we get

$$D((\bar{f}(n) + \bar{P}(n))_{n \in [N, 2N]}) \ll_{f, \mathcal{H}, k} N^{-\beta'}$$

for $\beta' = \min\{\eta, \beta''/2\}$. In particular, by the definition of $\mathcal{B}$ and since $U_\ell = \prod_{i=1}^\ell [u_i, v_i]$, it follows from the definition of the discrepancy that

$$\left| \frac{1}{N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) - \lambda_\ell(U_\ell) \right| \ll_{f, \mathcal{H}, k} N^{-\beta'},$$

i.e.

$$\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) = N \lambda_\ell(U_\ell) + O_{f, \mathcal{H}, k}(N^{1-\beta'}). \quad (3.4)$$

Fix some $0 < \theta < 1$ to be determined later. Let q, c be integers with $0 \leq c < q \leq N^\theta$, and define $\mathbf{y}_n = (\{f(qn+c+h_1) + P(qn+c+h_1)\}, \dots, \{f(qn+c+h_\ell) + P(qn+c+h_\ell)\})$.

Note that we have

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} \mathbf{1}_{\mathcal{B}}(n) = \sum_{N \leq nq + c < 2N} \mathbf{1}_{U_\ell}(\mathbf{y}_n) = \sum_{n \in I} \mathbf{1}_{\mathcal{B}'}(n),$$

where $\mathcal{B}' = \{n \in \mathbb{N} : \mathbf{y}_n \in U_\ell\}$ and

$$I = \left\{ \left\lceil \frac{N-c}{q} \right\rceil, \left\lceil \frac{N-c}{q} \right\rceil + 1, \dots, \left\lceil \frac{2N-c}{q} \right\rceil - 1 \right\}.$$

Hence, by applying the Erdős-Turán-Koksma inequality, but this time to the sequence $(\mathbf{y}_n)_{n \in I}$, we get that for any $H \in \mathbb{N}$,

$$\begin{aligned} D \left((\bar{f}(n) + \bar{P}(n))_{\substack{n \in [N, 2N] \\ n \equiv c \pmod q}} \right) &\ll_\ell \left(\frac{1}{H+1} + \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \left| \frac{1}{\#I} \sum_{n \in I} e(\langle \mathbf{m}, \mathbf{y}_n \rangle) \right| \right) \\ &= \left(\frac{1}{H+1} + \sum_{0 < \|\mathbf{m}\|_\infty \leq H} \frac{1}{r(\mathbf{m})} \left| \frac{1}{\#I} \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} e(\langle \mathbf{m}, \mathbf{x}_n \rangle) \right| \right). \end{aligned}$$

Again, by taking $H = \lfloor N^\eta \rfloor$, using Corollary 3.4 and following the exact same treatment as above, we get that

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} \mathbf{1}_{\mathcal{B}}(n) = \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}'}(n) = \#I \lambda_\ell(U_\ell) + O_{f, \mathcal{H}, k}(N^{1-\beta'}) \quad (3.5)$$

for $\beta' > 0$. Using that and the fact that $\#I - \frac{N}{q} = O(1)$, we get

$$\left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll_{f, \mathcal{H}, k, U_\ell} N^{1-\beta'}.$$

Hence, we get

$$\sum_{q \leq N^\theta} \max_{0 \leq c \leq q-1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll_{f, \mathcal{H}, k, U_\ell} N^{1-\beta'+\theta}.$$

Taking $\theta = \beta = \frac{\beta'}{2}$ yields

$$\sum_{q \leq N^\theta} \max_{0 \leq c \leq q-1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod q}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll_{f, \mathcal{H}, k, U_\ell} N^{1-\beta} \ll_{f, \mathcal{H}, k, U_\ell} \frac{1}{N^\beta} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).$$

□

Finally, the following result will help us prove Hypothesis 4 of Theorem 2.2 in Section 5.

Corollary 3.6. *Let the notation and hypotheses be as in Proposition 3.5. Then, we have*

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) \ll_{f, \mathcal{H}, k, U_\ell} \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)$$

for all $q < N^\theta$.

Proof. By (3.4) and (3.5), we have

$$\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) = \lambda_\ell(U_\ell)N + O_{f, \mathcal{H}, k}(N^{1-\beta'})$$

and

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) = \lambda_\ell(U_\ell) \frac{N}{q} + O_{f, \mathcal{H}, k}(N^{1-\beta'}),$$

and therefore

$$\sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) \ll_{f, \mathcal{H}, k, U_\ell} \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).$$

□

4. Uniform distribution of values of polynomials with real exponents in arithmetic progressions along primes

In this section, we prove results that allow us to verify Hypotheses 1 and 3 of Theorem 2.2 in Section 5. To do this, we study the distribution of sequences of the form $(\{F(p + h_1)\}, \dots, \{F(p + h_\ell)\})$ in $[0, 1)^\ell$, where p runs over the primes in an interval, or the primes in an interval lying in some arithmetic progression. Proposition 4.2 and Corollary 4.3 are analogues of Proposition 3.3 and Corollary 3.4 for exponential sums over the primes. Using these results, we prove Lemma 4.4, which will help us verify Hypothesis 3, and Lemma 4.5, which will help us verify Hypothesis 1. But first, we prove an easy lemma that is useful in the proof of Proposition 4.2.

Lemma 4.1. *Let $\alpha > 0$, $W \geq 0$. Let $f : \mathbb{N} \rightarrow \mathbb{R}$ be such that $|f(n)| \leq 1$ for all $n \leq 2W$. If*

$$\left| \sum_{N \leq n < 2N} f(n) \right| \leq CN^\alpha$$

uniformly for $N \geq W$, then we have that

$$\sup_U \sup_V \left| \sum_{\substack{N \leq n < 2N \\ U \leq n < V}} f(n) \right| \ll_\alpha CN^\alpha + W,$$

where the supremums are over all the positive real numbers.

Proof. Fix any $U, V \in \mathbb{R}$. Without loss of generality, we can assume that $N \leq U < V \leq 2N$. By using the triangle inequality, we get that

$$\left| \sum_{\substack{N \leq n < 2N \\ U \leq n < V}} f(n) \right| = \left| \sum_{U \leq n < V} f(n) \right| \leq \left| \sum_{1 \leq n < U} f(n) \right| + \left| \sum_{1 \leq n < V} f(n) \right|.$$

Let $k \geq 0$ be the smallest integer such that $U/2^k \leq 2W$. We have

$$\begin{aligned} \left| \sum_{1 \leq n < U} f(n) \right| &\leq \sum_{j=0}^{k-1} \left| \sum_{U/2^{j+1} \leq n < U/2^j} f(n) \right| + \left| \sum_{1 \leq n < U/2^k} f(n) \right| \\ &\leq CU^\alpha \sum_{j=0}^{k-1} \frac{1}{(2^\alpha)^{j+1}} + W \\ &\ll_\alpha CN^\alpha + W. \end{aligned}$$

Similarly, $\left| \sum_{1 \leq n < V} f(n) \right| \ll_\alpha CN^\alpha + W$, so

$$\left| \sum_{\substack{N \leq n < 2N \\ U \leq n < V}} f(n) \right| \ll_\alpha CN^\alpha + W.$$

□

Next, we use the von Mangoldt function and partial summation to compare the sums

$$\left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \quad \text{and} \quad \left| \sum_{N \leq n < 2N} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right|.$$

This allows us to deduce the proof of Proposition 4.2 using Lemma 3.2 and some properties of the von Mangoldt function (see [BKM⁺14, Lemma 2.3]).

Proposition 4.2. *Let $k, r \in \mathbb{N}$ and fix some polynomial with non-integer exponents $f(x) = \sum_{j=1}^r d_j x^{\rho_j}$ with $0 < \rho_1 < \rho_2 < \dots < \rho_r$ and $d_r \neq 0$. Let $\ell = \lfloor \rho_r \rfloor + 1$ and let η be the same number as in Section 3. Let $\mathcal{H} = \{h_1, \dots, h_\ell\}$ be a set of integers with $h_1 < \dots < h_\ell$, and define $\bar{f}(n) = (f(n+h_1), \dots, f(n+h_\ell))$. There exists a real number $\zeta > 0$ depending only on f and k such that uniformly over all real polynomials P of degree at most k , all $N \in \mathbb{N}$ and all vectors $\mathbf{m} = (m_1 \dots m_\ell) \in \mathbb{Z}^\ell \setminus \{\bar{0}\}$ with $\max_{1 \leq j \leq \ell} |m_j| \leq N^\eta$,*

$$\left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \ll_{f, \mathcal{H}, k} N^{1-\zeta}. \quad (4.1)$$

Proof. We follow the proof of [BKM⁺14, Proposition 2.1]. Let S denote the left side of (4.1). By using partial summation, we have

$$\begin{aligned}
S &= \left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \\
&= \left| \sum_{N \leq p < 2N} \frac{\Lambda(p)}{\log p} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \\
&= \left| \sum_{N \leq n < 2N} \frac{\Lambda(n)}{\log n} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(n)) - \sum_{N \leq p^k < 2N, k \geq 2} \frac{\Lambda(p)}{\log p} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \\
&\leq \left| \sum_{N \leq n < 2N} \frac{\Lambda(n)}{\log n} e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| + O(N^{1/2} \log N) \\
&\ll \frac{1}{\log N} \left| \sum_{N \leq n < N_1} \Lambda(n) e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right| + O(N^{1/2} \log N)
\end{aligned}$$

for the integer $N_1 \in [N, 2N]$ such that

$$\left| \sum_{N \leq n < N_1} \Lambda(n) e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n)) \right|$$

is maximal. By Lemma 4.1 with $W = 0$, to get an upper bound as in (4.1), it is sufficient to find such an upper bound for the sum

$$S_1 = \sum_{N \leq n < 2N} \Lambda(n) e(\langle \mathbf{m}, \bar{f}(n) \rangle + P(n))$$

uniformly for $N \in \mathbb{N}$. To do this, we apply [BKM⁺14, Lemma 2.3], with $F(x) = e(\langle \mathbf{m}, \bar{f}(x) \rangle + P(x)) \mathbf{1}_{N \leq x < 2N}$, $|F(x)| \leq 1 = F_0$, $U = \frac{1}{4}N^{1/5}$, $V = 4N^{1/3}$, and Z a number in $\frac{1}{2} + \mathbb{N}$ at a minimal distance from $\frac{1}{4}N^{2/5}$. For all N large enough (say $N \geq N_0$), we have $3 \leq U < V < Z < 2N$, $Z \geq 4U^2$, $2N \geq 64Z^2U$, $V^3 = 64N$, so the conditions of the lemma are satisfied. Then, by the conclusion of the lemma, we have

$$|S_1| \ll 1 + L (\log N)^8 + K \log N$$

for

$$\begin{aligned}
K &= \max_M \sum_{t=1}^{\infty} \tau_3(t) \left| \sum_{\substack{N \leq nt < 2N \\ Z < n \leq M}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|, \\
L &= \sup_b \sum_{t=1}^{\infty} \tau_4(t) \left| \sum_{\substack{N \leq nt < 2N \\ U < n < V}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|,
\end{aligned}$$

where the supremum is taken over all arithmetic functions $b(n)$ satisfying $|b(n)| \leq \tau_3(n)$. Fix $\varepsilon > 0$. Note that in the expression for K , the innermost sum is zero when $t \geq 2N/Z$, so we have

$$K = \max_M \sum_{t < \frac{2N}{Z}} \tau_3(t) \left| \sum_{\substack{\frac{N}{t} \leq n < \frac{2N}{t} \\ Z < n \leq M}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|$$

$$\ll_{\varepsilon} N^{\varepsilon} \sum_{t < \frac{2N}{Z}} \max_M \left| \sum_{\substack{\frac{N}{t} \leq n < \frac{2N}{t} \\ Z < n \leq M}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|,$$

where we have used the trivial upper bound $\tau_3(t) \ll_{\varepsilon} t^{\varepsilon} \ll N^{\varepsilon}$.

As in the proof of Proposition 3.3, we can write $\langle \mathbf{m}, \bar{f}(x) \rangle = f_1(x) + R(x)$, where $R(x) \ll_{f, \mathcal{H}} x^{\{\rho_r\} - 1 + \eta}$ and f_1 is of the form

$$f_1(x) = bx^{\rho_r - n_0} + \sum_{i=1}^{r'} b_i x^{\rho'_i},$$

where $n_0 \in \{0, \dots, \ell - 1\}$, $\rho'_1 < \rho'_2 < \dots < \rho'_{r'} < \rho_r - n_0 - \eta$, $b = \left(\sum_{j=1}^{\ell} m_j h_j^{n_0} \right) d_r \binom{\rho_r}{n_0}$, $|b|, |b_i| \ll_{f, \mathcal{H}} N^{\eta}$ and $r' \leq \ell r$. By a variant of the argument in Proposition 3.3, we have

$$\left| \sum_{\substack{\frac{N}{t} \leq n < \frac{2N}{t} \\ Z < n \leq M}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right| \ll_{f, \mathcal{H}} \left| \sum_{\substack{\frac{N}{t} \leq n < \frac{2N}{t} \\ Z < n \leq M}} e(f_1(nt) + P(nt)) \right| + \frac{N^{\frac{1+\{\rho_r\}}{2}}}{t}$$

for any M . To get an upper bound on the sum on the right-hand side, we start by bounding

$$S_2 := \left| \sum_{\frac{N}{t} \leq n < \frac{2N}{t}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|.$$

As in the proof of Proposition 3.3, for some constants $0 < c_1 < c_2$ depending only on f and k , one has

$$c_1 |b| N^{\rho_r - n_0} \left(\frac{N}{t} \right)^{-s} \leq \left| \frac{d^s}{dx^s} f_1(tx) \right| \leq c_2 |b| N^{\rho_r - n_0} \left(\frac{N}{t} \right)^{-s} \quad (4.2)$$

for any integers $0 \leq s \leq 5(\ell + 1) + k + 2$, $1 \leq t < N^{4/5}$ and any real number $N/t \leq x \leq 2N/t$, as long as N is large enough in terms of f and $\mathcal{H}$. Therefore, by applying Lemma 3.2 to the function $x \mapsto f_1(xt)$ and with $q = 5(\ell + 1)$, we get that

uniformly over the integers $1 \leq t \leq N^{4/5}$,

$$\begin{aligned}
\left| \sum_{\frac{N}{t} \leq n < \frac{2N}{t}} e(f_1(nt) + P(nt)) \right| &\ll_{f, \mathcal{H}, k} \left(\frac{N}{t} \right)^{1 - \frac{1}{2^k}} + \frac{N}{t} \left(\frac{(\log(N/t))^k}{|b| N^{\rho_r - n_0}} \right)^{\frac{1}{2^k}} \\
&\quad + \frac{N}{t} \left(\frac{|b| N^{\rho_r - n_0}}{(N/t)^{5(\ell+1)+2}} \right)^{\frac{1}{2^{k+5}(\ell+1)+2-2^{k+1}}} \\
&\ll_{f, \mathcal{H}, k, \varepsilon} \left(\frac{N}{t} \right)^{1 - \frac{1}{2^k}} + \frac{N^{1 - \frac{\{\rho_r\}}{2^k} + \frac{\varepsilon}{2}}}{t} + \frac{N^{1 - \frac{1 - \{\rho_r\}}{2^{k+5}(\ell+1)+2-2^{k+1}}}}{t} \\
&\ll_{f, \mathcal{H}, k, \varepsilon} \left(\frac{N}{t} \right)^{1 - \alpha'},
\end{aligned}$$

where $\alpha' = \min \left\{ \frac{1}{2^k}, \frac{\{\rho_r\}}{2^k} - \varepsilon, \frac{1 - \{\rho_r\}}{2^{k+5}(\ell+1)+2-2^{k+1}} - \varepsilon \right\}$. For any fixed t large enough in terms of f and $\mathcal{H}$, this estimate holds as long as $N/t \geq N^{1/5}$, so, by applying Lemma 4.1 with $W = N^{1/5}$, we get

$$\max_M \left| \sum_{\substack{\frac{N}{t} \leq n < \frac{2N}{t} \\ Z < n \leq M}} e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right| \ll_{f, \mathcal{H}, k, \varepsilon} \left(\frac{N}{t} \right)^{1 - \alpha'} + N^{1/5}$$

and so

$$K \ll_{f, \mathcal{H}, k, \varepsilon} N^{1 - \zeta_1},$$

where

$$\zeta_1 = \min \left\{ \frac{3}{5} \alpha', \frac{1}{5} \right\} - \varepsilon.$$

Next, we find an upper bound for L . Let

$$S_3 = \sum_{t=1}^{\infty} \tau_4(t) \left| \sum_{\substack{N \leq nt < 2N \\ U < n < V}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|$$

for a fixed arithmetic function $b(n)$ satisfying $|b(n)| \leq \tau_3(n)$. Again, since $\tau_4(n) \ll_{\varepsilon} n^{\varepsilon}$, we have

$$S_3 \ll_{\varepsilon} N^{\varepsilon} \sum_{t=1}^{\infty} \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|.$$

The innermost sum is zero when $t \leq N/V$ or $t \geq 2N/U$, so, in fact, we have

$$S_3 \ll_{\varepsilon} N^{\varepsilon} \sum_{\frac{N}{V} < t < \frac{2N}{U}} \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|.$$

By replacing the lower bound N/V with some integer between $N/(2V)$ and N/V , the sum over t can be decomposed into $\ll \log N$ sums over t in dyadic intervals $[X, 2X)$. Let

$$S_4 = \sum_{X \leq t < 2X} \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|$$

be one of those sums. Then, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |S_4|^2 &= \left(\sum_{X \leq t < 2X} 1 \cdot \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right| \right)^2 \\ &\leq \left(\sum_{X \leq t < 2X} 1^2 \right) \left(\sum_{X \leq t < 2X} \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|^2 \right) \\ &\leq X \sum_{X \leq t < 2X} \left| \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right|^2 \\ &= X \sum_{X \leq t < 2X} \left(\sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} b(n) e(\langle \mathbf{m}, \bar{f}(nt) \rangle + P(nt)) \right) \\ &\quad \times \left(\sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} \overline{b(n)} e(-\langle \mathbf{m}, \bar{f}(nt) \rangle - P(nt)) \right) \\ &\ll X \sum_{X \leq t < 2X} \sum_{\substack{U < n < V \\ \frac{N}{t} \leq n < \frac{2N}{t}}} |b(n)|^2 \end{aligned}$$

$$+ X \left| \sum_{X \leq t < 2X} \sum_{\substack{A < n_1, n_2 < B \\ n_1 \neq n_2}} b(n_1) \overline{b(n_2)} e(\langle \mathbf{m}, \bar{f}(n_1 t) - \bar{f}(n_2 t) \rangle + P(n_1 t) - P(n_2 t)) \right|,$$

where $A = \max\{U, \lceil N/t \rceil - 1\}$ and $B = \min\{V, 2N/t\}$. By changing the order of summation and using the bound $|b(n)| \leq \tau_3(n) \ll_\varepsilon n^\varepsilon \ll_\varepsilon N^\varepsilon$ for $n \leq 2N$, we get

$$|S_4|^2 \ll_\varepsilon N^\varepsilon X \left(N + \sum_{\substack{A' \leq n_1, n_2 < B' \\ n_1 \neq n_2}} \left| \sum_{\substack{X \leq t < 2X \\ \frac{N}{n_1} \leq t < \frac{2N}{n_2}}} e(\langle \mathbf{m}, \bar{f}(n_1 t) - \bar{f}(n_2 t) \rangle + P(n_1 t) - P(n_2 t)) \right| \right), \quad (4.3)$$

with $A' = \max\{U, \lceil N/(2X) \rceil - 1\}$ and $B' = \min\{V, 2N/X\}$. For fixed n_1 and $n_2 \neq n_1$ in that range, we have that $Q(t) = P(n_1 t) - P(n_2 t)$ is a polynomial of degree at most $k = \deg P(x)$. Also, we can write

$$\langle \mathbf{m}, \bar{f}(n_1 t) - \bar{f}(n_2 t) \rangle = f_1(n_1 t) - f_1(n_2 t) + R(n_1 t) - R(n_2 t),$$

where f_1 and R are the same functions as above. As before,

$$\begin{aligned} & \left| \sum_{\substack{X \leq t < 2X \\ \frac{N}{n_1} \leq t < \frac{2N}{n_2}}} e(\langle \mathbf{m}, \bar{f}(n_1 t) - \bar{f}(n_2 t) \rangle + P(n_1 t) - P(n_2 t)) \right| \\ & \ll_{f, \mathcal{H}} \left| \sum_{\substack{X \leq t < 2X \\ \frac{N}{n_1} \leq t < \frac{2N}{n_2}}} e(f_1(n_1 t) - f_1(n_2 t) + P(n_1 t) - P(n_2 t)) \right| + n_2^{\frac{\{\rho_r\}-1}{2}} X^{\frac{1+\{\rho_r\}}{2}}. \end{aligned}$$

Now, there are some constants c'_1, c'_2 depending only on f and k such that

$$\begin{aligned} \left| \frac{d^s}{dt^s} (f_1(n_1 t) - f_1(n_2 t)) \right| & \geq c_1 |b(n_1^{\rho_r - n_0} - n_2^{\rho_r - n_0})| X^{\rho_r - n_0} X^{-s} \\ & \geq c'_1 \left| b \frac{(n_1 - n_2)}{n_1} \right| N^{\rho_r - n_0} X^{-s}, \end{aligned}$$

$$\begin{aligned} \left| \frac{d^s}{dt^s} (f_1(n_1 t) - f_1(n_2 t)) \right| & \leq c_2 |b(n_1^{\rho_r - n_0} - n_2^{\rho_r - n_0})| X^{\rho_r - n_0} X^{-s} \\ & \leq c'_2 \left| b \frac{(n_1 - n_2)}{n_1} \right| N^{\rho_r - n_0} X^{-s} \end{aligned}$$

for all $0 \leq s \leq 3(\ell + 1) + k + 2$ as long as N is large enough in terms of f and $\mathcal{H}$, so by applying Lemma 3.2 to the function $t \mapsto f_1(n_1 t) - f_1(n_2 t)$ and with $q = 3(\ell + 1)$, we get

$$\begin{aligned} & \left| \sum_{X \leq t < 2X} e(f_1(n_1 t) - f_1(n_2 t) + P(n_1 t) - P(n_2 t)) \right| \\ & \ll_{f, \mathcal{H}, k} X^{1 - \frac{1}{2^k}} + X \left(\frac{n_1 (\log X)^k}{|b(n_1 - n_2)| N^{\rho_r - n_0}} \right)^{1/2^k} \\ & \quad + X \left(\frac{|b(n_1 - n_2)| N^{\rho_r - n_0}}{n_1 X^{3(\ell+1)+2}} \right)^{\frac{1}{2k+3\ell+5-2k+1}} \\ & \ll_{f, \mathcal{H}, k, \varepsilon} X^{1 - \frac{1}{2^k}} + X^{1+\varepsilon} \left(\frac{n_1}{|n_1 - n_2| N^{\rho_r - n_0}} \right)^{1/2^k} + X N^{-\frac{1 - \{\rho_r\}}{2k+3\ell+5-2k+1}}. \end{aligned}$$

Here we used the fact that $X \geq N/2V \gg N^{2/3} \geq N^{1/3}$ and $1 \ll_{f, \mathcal{H}} b \ll_{f, \mathcal{H}} N^\eta < N$. In fact, the result holds for $X \geq N^{1/3}$, so by applying Lemma 4.1 with $W = N^{1/3}$, we deduce that

$$\begin{aligned} & \left| \sum_{\substack{X \leq t < 2X \\ \frac{N}{n_1} \leq t < \frac{2N}{n_2}}} e(f_1(n_1 t) - f_1(n_2 t) + P(n_1 t) - P(n_2 t)) \right| \\ & \ll_{f, \mathcal{H}, k, \varepsilon} X^{1 - \frac{1}{2^k}} + X^{1+\varepsilon} \left(\frac{n_1}{|n_1 - n_2| N^{\rho_r - n_0}} \right)^{1/2^k} + X N^{-\frac{1 - \{\rho_r\}}{2k+3\ell+5-2k+1}} + N^{\frac{1}{3}}. \end{aligned}$$

Notice that

$$\begin{aligned} & \sum_{\substack{A' \leq n_1, n_2 < B' \\ n_1 \neq n_2}} \left(\frac{n_1}{|n_1 - n_2|} \right)^{1/2^k} \ll_k \left(\sum_{A' \leq n_1 < B'} n_1^{\frac{1}{2^k}} \right) \left(\sum_{1 \leq n_3 \leq B'} n_3^{-\frac{1}{2^k}} \right) \\ & \ll_k (B')^2, \end{aligned}$$

so by inserting these upper bounds into (4.3), we get

$$\begin{aligned} |S_4|^2 & \ll_{f, \mathcal{H}, k, \varepsilon} N^{2\varepsilon} (XB')^2 N^{-\frac{\{\rho_r\}}{2^k}} + N^\varepsilon (B'X)^2 X^{-\frac{1}{2^k}} \\ & \quad + (B'X)^2 N^\varepsilon N^{-\frac{1 - \{\rho_r\}}{2k+3\ell+5-2k+1}} + XB' N^{\frac{1 + \{\rho_r\}}{2} + \varepsilon} + XN^{1+\varepsilon} + X(B')^2 N^{\frac{1}{3} + \varepsilon} \\ & \ll_{f, k, \mathcal{H}, \varepsilon} N^{2 - \frac{\{\rho_r\}}{2^k} + 2\varepsilon} + N^{2 - \frac{1}{3 \cdot 2^{k-1}} + \varepsilon} + N^{2 - \frac{1 - \{\rho_r\}}{2k+3\ell+5-2k+1} + \varepsilon} + N^{\frac{3 + \{\rho_r\}}{2} + \varepsilon} + N^{\frac{5}{3} + \varepsilon} \\ & \ll_{f, k, \mathcal{H}, \varepsilon} N^\beta \end{aligned}$$

where

$$\beta = \max \left\{ 2 - \frac{\{\rho_r\}}{2^k} + 2\varepsilon, 2 - \frac{1 - \{\rho_r\}}{2k+3\ell+5-2k+1} + \varepsilon, \frac{3 + \{\rho_r\}}{2} + \varepsilon, \frac{5}{3} + \varepsilon \right\}.$$

Here, we used the fact that $B'X \ll N$, $X \gg \frac{N}{V} \gg N^{\frac{2}{3}}$ and $X \ll \frac{N}{U} \ll N^{4/5}$. Note that $\beta < 2$ if ε is small enough. Now, summing over all subintervals, we get

$$L \ll_{f, \mathcal{H}, k, \varepsilon} N^{1-\zeta_2}$$

where $\zeta_2 = 1 - \frac{\beta}{2} - \varepsilon$. For $\varepsilon > 0$ small enough and $\zeta := \min\{\zeta_1, \zeta_2\}$ we get the desired result. $\square$

Again, as in Corollary 3.4, we get the same bound if we restrict the sum to primes in some arithmetic progression.

Corollary 4.3. *Under the hypotheses of Proposition 4.2, uniformly over all real polynomials P of degree at most k , all $N \in \mathbb{N}$, all $\mathbf{m} = (m_1, \dots, m_\ell) \in \mathbb{Z}^k \setminus \{\bar{0}\}$ with $\max_{1 \leq j \leq \ell} |m_j| \leq N^\eta$, all $q \in \mathbb{N}$, and all $a \in \{0, 1, \dots, q-1\}$ one has*

$$\left| \sum_{\substack{N \leq p < 2N \\ p \equiv a \pmod{q}}} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| \ll_{f, \mathcal{H}, k} N^{1-\zeta}, \quad (4.4)$$

where ζ and the implied constant are the same as in Proposition 4.2.

Proof. As in the proof of Corollary 3.4, we use the fact that

$$\mathbf{1}_{q\mathbb{Z}+a}(n) = \frac{1}{q} \sum_{0 \leq t \leq q-1} e\left(\frac{(n-a)t}{q}\right).$$

Therefore, we have

$$\begin{aligned} \left| \sum_{\substack{N \leq p < 2N \\ p \equiv a \pmod{q}}} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \right| &= \left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \mathbf{1}_{q\mathbb{Z}+a}(p) \right| \\ &= \left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p)) \left(\frac{1}{q} \sum_{0 \leq t \leq q-1} e\left(\frac{(p-a)t}{q}\right) \right) \right| \\ &\leq \frac{1}{q} \sum_{0 \leq t \leq q-1} \left| \sum_{N \leq p < 2N} e\left(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p) + \frac{(p-a)t}{q}\right) \right| \\ &= \frac{1}{q} \sum_{0 \leq t \leq q-1} \left| \sum_{N \leq p < 2N} e\left(\langle \mathbf{m}, \bar{f}(p) \rangle + P(p) + \frac{tp}{q}\right) \right|. \end{aligned}$$

The claim now follows from Proposition 4.2. $\square$

Now, we prove a version of the Bombieri-Vinogradov theorem (Theorem 2.3) for the sets of our interest (cf. [Shu20, Theorem 2]). We start by following the steps of

the proof of Proposition 3.5 (Proposition 4.2 and Corollary 4.3 play a role analogous to Proposition 3.3 and Corollary 3.4).

Lemma 4.4. *Let $k, r \in \mathbb{N}$, and fix some polynomial with non-integer exponents $f(x) = \sum_{j=1}^r d_j x^{\rho_j}$ with $0 < \rho_1 < \rho_2 < \dots < \rho_r$ and $d_r \neq 0$. Then, there exists some $\theta > 0$ depending only on $f(x)$ and k such that the following holds:*

Let $h_1 < h_2 < \dots < h_\ell$ be integers. Let P be a real polynomial of degree at most k . Let $U_\ell = \prod_{i=1}^\ell [u_i, v_i)$ with $0 \leq u_i < v_i < 1$ for all i , and define

$$\mathcal{B} := \{n \in \mathbb{N} : (\{f(n + h_1) + P(n + h_1)\}, \dots, \{f(n + h_\ell) + P(n + h_\ell)\}) \in U_\ell\}.$$

Fix a $D > 0$. Then, for all N large enough depending on all that precedes, we have

$$\sum_{q < N^\theta} \max_{(q,a)=1} \left| \sum_{\substack{N \leq p < 2N \\ p \equiv a \pmod{q}}} \mathbf{1}_{\mathcal{B}}(p) - \frac{1}{\varphi(q)} \sum_{N \leq p < 2N} \mathbf{1}_{\mathcal{B}}(p) \right| \ll_{f, \mathcal{H}, D, k, U_\ell} \frac{1}{(\log N)^D} \sum_{N \leq p < 2N} \mathbf{1}_{\mathcal{B}}(p). \quad (4.5)$$

Proof. Define

$$\bar{f}(x) := (f(x + h_1), \dots, f(x + h_\ell))$$

and

$$\bar{P}(x) := (P(x + h_1), \dots, P(x + h_\ell)).$$

By the Erdős-Turán-Koksma inequality (Theorem 2.5), the discrepancy mod 1 of $(\bar{f}(p) + \bar{P}(p))_{p \in [N, 2N]}$ verifies

$$D((\bar{f}(p) + \bar{P}(p))_{p \in [N, 2N] \cap \mathbb{P}}) \ll_\ell \frac{1}{H} + \sum_{\substack{\mathbf{m} \in \mathbb{Z}^k \setminus \{\bar{0}\} \\ \|\mathbf{m}\|_\infty \leq H}} \frac{1}{r(\mathbf{m})} \frac{1}{\pi(N, 2N)} \left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) + \bar{P}(p) \rangle) \right|$$

for any $H \in \mathbb{N}$. Now, observe that

$$\left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) + \bar{P}(p) \rangle) \right| = \left| \sum_{N \leq p < 2N} e(\langle \mathbf{m}, \bar{f}(p) \rangle + \langle \mathbf{m}, \bar{P}(p) \rangle) \right| \ll_{f, \mathcal{H}, k} N^{1-\zeta}$$

by Proposition 4.2. Moreover, by the prime number theorem, $\frac{1}{\pi(N, 2N)} \ll \frac{\log N}{N}$ and, as in the proof of Lemma 3.5, we also have

$$\sum_{\substack{m \in \mathbb{Z}^k \setminus \{\bar{0}\} \\ \|m\|_\infty \leq H}} \frac{1}{r(m)} \ll (\log H)^\ell.$$

Thus,

$$D((\bar{f}(p) + \bar{P}(p))_{p \in [N, 2N] \cap \mathbb{P}}) \ll_{f, \mathcal{H}, k} \frac{1}{H} + (\log H)^\ell \frac{N^{-\zeta}}{\log N}.$$

Taking $H = \lfloor N^\eta \rfloor$, we get

$$D \left((\bar{f}(p) + \bar{P}(p))_{p \in [N, 2N) \cap \mathbb{P}} \right) \ll_{f, \mathcal{H}, k, \gamma} N^{-\gamma}$$

for all $0 < \gamma < \min\{\eta, \zeta\}$. For concreteness, let $\gamma = \frac{\min\{\eta, \zeta\}}{2}$. It follows that

$$\left| \frac{1}{\pi(N, 2N)} \sum_{N \leq p < 2N} \mathbf{1}_{\mathcal{B}}(p) - \lambda_\ell(U_\ell) \right| \leq D \left((\bar{f}(p) + \bar{P}(p))_{p \in [N, 2N) \cap \mathbb{P}} \right) \ll_{f, \mathcal{H}, k} N^{-\gamma},$$

which implies that

$$\sum_{N \leq p < 2N} \mathbf{1}_{\mathcal{B}}(p) = \pi(N, 2N) \lambda_\ell(U_\ell) + O_{f, \mathcal{H}, k}(N^{1-\gamma}).$$

Now, consider the sequence $(\bar{f}(p) + \bar{P}(p))_{\substack{p \in [N, 2N) \cap \mathbb{P} \\ p \equiv c \pmod{q}}}$ for integers $0 \leq a < q$. Again, as in the proof of Proposition 3.5, using Corollary 4.3, we get

$$D \left((\bar{f}(p) + \bar{P}(p))_{\substack{p \in [N, 2N) \cap \mathbb{P} \\ p \equiv c \pmod{q}}} \right) \ll_{f, \mathcal{H}, k} N^{-\gamma}.$$

Therefore, we have

$$\sum_{\substack{N \leq p < 2N \\ p \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(p) = \pi(N, 2N; c, q) \lambda_\ell(U_\ell) + O_{f, \mathcal{H}, k}(N^{1-\gamma}).$$

Now, the left side of (4.5) is of the form

$$S := \sum_{q < N^\theta} \max_{(q, c)=1} \left| \lambda_\ell(U_\ell) \pi(N, 2N; c, q) - \lambda_\ell(U_\ell) \frac{\pi(N, 2N)}{\varphi(q)} + O_{f, \mathcal{H}, k}(N^{1-\gamma}) \right|$$

The triangle inequality together with Theorem 2.3 yield

$$\begin{aligned} S &\leq \lambda_\ell(U_\ell) \sum_{q < N^\theta} \max_{(q, c)=1} \left| \pi(2N; c, q) - \frac{\pi(2N)}{\varphi(q)} \right| \\ &\quad + \lambda_\ell(U_\ell) \sum_{q < N^\theta} \max_{(q, c)=1} \left| \pi(N; c, q) - \frac{\pi(N)}{\varphi(q)} \right| + O_{f, \mathcal{H}, k}(N^{1-\gamma+\theta}) \\ &\ll_{D, U_\ell} \frac{N}{(\log N)^{D+1}} + O_{f, \mathcal{H}, k}(N^{1-\gamma+\theta}) \end{aligned}$$

for all N large enough. But since

$$\frac{N}{(\log N)^{D+1}} \ll \frac{\pi(N, 2N)}{(\log N)^D} \ll_{f, \mathcal{H}, k, U_\ell} \frac{1}{(\log N)^D} \sum_{N \leq p < 2N} \mathbf{1}_{\mathcal{B}}(p),$$

by choosing, say, $\theta = \frac{\gamma}{2}$, we get the desired result. $\square$

Finally, we show a result that will help us obtain Hypothesis 1 in Section 5. It will

follow from the estimates

$$\sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) = \lambda_{\ell}(U_{\ell})\pi(N, 2N) + O_{f, \mathcal{H}, k}(N^{1-\gamma})$$

and

$$\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) = \lambda_{\ell}(U_{\ell})N + O_{f, \mathcal{H}, k}(N^{1-\beta'})$$

obtained in the proofs of Lemma 4.4 and Proposition 3.5, respectively.

Lemma 4.5. *Under the hypotheses of Lemma 4.4, for all $0 < \delta < 1$ and N large enough in terms of all that precedes, we have*

$$\sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) \geq \frac{\delta}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).$$

Proof. As in the proof of Lemma 4.4, for $i = 1, \dots, \ell$, we can write

$$\sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) = \lambda_{\ell}(U_{\ell})\pi(N, 2N) + O_{f, \mathcal{H}, k}(N^{1-\sigma'})$$

for some $\sigma' > 0$, so, by the prime number theorem,

$$\sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) \sim \lambda_{\ell}(U_{\ell}) \frac{N}{\log N}.$$

As in the proof of Proposition 3.5, we have

$$\sum_{N \leq n \leq 2N} \mathbf{1}_{\mathcal{B}}(n) = \lambda_{\ell}(U_{\ell})N + O_{f, \mathcal{H}, k}(N^{1-\sigma})$$

for some $\sigma > 0$, so

$$\frac{1}{\log N} \sum_{N \leq n \leq 2N} \mathbf{1}_{\mathcal{B}}(n) \sim \lambda_{\ell}(U_{\ell}) \frac{N}{\log N}.$$

Therefore,

$$\lim_{N \rightarrow \infty} \frac{\frac{\sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n)}{\frac{1}{\log N}}}{\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)} = 1.$$

In particular, for all N large enough, the ratio on the left side is bigger than δ .

□

5. Proof of the main theorem

In this section, we prove Theorem 1.2. Fix F as in the statement of Theorem 1.2. We show that for any integer m , there is some integer $d \geq m$ as well as integers $h_1 < \dots < h_d$ such that for any N large enough, there exists some n in the set

$$\{n \in \mathbb{N} : \{F(n + h_j)\} \in U \text{ for } 1 \leq j \leq d\} \cap [N, 2N) \quad (5.1)$$

with the property that at least m elements of the set $\{n + h_1, \dots, n + h_d\}$ are consecutive primes. An important ingredient in establishing the main results of Sections 3 and 4 was that if ρ is the largest exponent appearing in the expression for F and $\ell = \lfloor \rho \rfloor + 1$, then the sequence $(\{F(n + h_1)\}, \dots, \{F(n + h_\ell)\})$ is uniformly distributed in $[0, 1)^\ell$. This, however, does not hold if ℓ is replaced by any integer $d > \ell$, because, in that case, we can find a $\mathbb{Z}$ -linear combination of $F(n + h_1), \dots, F(n + h_d)$ that goes to zero as n goes to infinity. In Proposition 5.2, we show that we can choose $h_1, \dots, h_d$ in such a way that for $\ell + 1 \leq j \leq d$, the function $F(n + h_j)$ can be written as a $\mathbb{Z}$ -linear combination of $F(n + h_1), \dots, F(n + h_\ell)$, up to a small error term. This is the key ingredient of our argument, as, for this choice of admissible set, the value of $(\{F(n + h_1)\}, \dots, \{F(n + h_d)\})$, which is not uniformly distributed in $[0, 1)^d$, is completely determined by the value of the vector $(\{F(n + h_1)\}, \dots, \{F(n + h_\ell)\})$, which is uniformly distributed in $[0, 1)^\ell$.

Remark 5.1. This is the reason why we require that the largest exponent of F is not an integer, as opposed to some exponent not being an integer. Indeed, if F is a function of the form $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ where $d_1, \dots, d_r \in \mathbb{R}$, $0 \leq \rho_1 < \dots < \rho_r$, $\rho_i \in \mathbb{R}$ for $1 \leq i \leq r$ and i_0 is the largest integer such that ρ_{i_0} is not an integer and $d_{i_0} \neq 0$, then similar arguments to those of Sections 3 and 4 would imply that, for $\ell_0 = \lfloor \rho_{i_0} \rfloor + 1$, the vector $(\{F(x + h_1)\}, \dots, \{F(x + h_{\ell_0})\})$ is uniformly distributed in $[0, 1)^{\ell_0}$ for every choice of distinct $h_1, \dots, h_{\ell_0}$, but our argument would fail, as for every integer $h \notin \{h_1, \dots, h_{\ell_0}\}$, we would not be able to express $F(x + h)$ as a $\mathbb{Z}$ -linear (or even $\mathbb{Q}$ -linear) combination of the $F(x + h_1), \dots, F(x + h_{\ell_0})$.

Proposition 5.2. *Let $\ell \in \mathbb{N}$. Then, for every $d \geq \ell$, there exists an admissible set $\mathcal{H} = \{h_1, \dots, h_d\}$ of d integers such that the following holds:*

Let $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ with $0 \leq \rho_1 < \dots < \rho_r$ be a polynomial with real exponents, with $\rho_r \notin \mathbb{Z}$, $d_r \neq 0$ and $\ell = \lfloor \rho_r \rfloor + 1$. Then, for every $\ell + 1 \leq s \leq d$, there is some vector $(b_{s,1}, \dots, b_{s,\ell}) \in \mathbb{Z}^\ell$ such that

$$F(x + h_s) = \sum_{j=1}^{\ell} b_{s,j} F(x + h_j) + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}).$$

Proof. Again, for $h \in \mathbb{Q}$, write

$$\nu(h) = \begin{pmatrix} 1 \\ h \\ \vdots \\ h^{\ell-1} \end{pmatrix} \in \mathbb{Q}^\ell.$$

Let

$$h_j = (j-1) \prod_{p \leq d} p$$

for $j = 1, \dots, \ell$, so that $0 = h_1 < h_2 < \dots < h_\ell$ and $\{h_1, \dots, h_\ell\}$ is admissible. Then, the vectors $\nu(h_1), \dots, \nu(h_\ell)$ are all linearly independent, by the same argument as in Proposition 3.1. Therefore, $\nu(h_1), \dots, \nu(h_\ell)$ form a basis of $\mathbb{Q}^\ell$. In particular, for $k = 1, \dots, \ell$, if $e_k \in \mathbb{Q}^\ell$ is the vector with a 1 in the k^{th} position and 0's in all the other positions, we can write

$$e_k = \sum_{j=1}^{\ell} \frac{p_{k,j}}{q_{k,j}} \nu(h_j),$$

where $p_{k,j} \in \mathbb{Z}$, $q_{k,j} \in \mathbb{N}$ and $(p_{k,j}, q_{k,j}) = 1$. Note that $e_1 = \nu(0) = \nu(h_1)$, so $q_{1,1} = 1$ and $p_{1,j} = 0$ for $j \geq 2$.

Denote $q := \prod_{k,j=1}^{\ell} q_{k,j}$. Then, for any integer t , we have

$$\nu(qt) = \sum_{k=1}^{\ell} (qt)^{k-1} e_k = \sum_{k=1}^{\ell} t^{k-1} q^{k-1} \left(\sum_{j=1}^{\ell} \frac{p_{k,j}}{q_{k,j}} \nu(h_j) \right) = \sum_{j=1}^{\ell} \left(\sum_{k=1}^{\ell} t^{k-1} p_{k,j} \frac{q^{k-1}}{q_{k,j}} \right) \nu(h_j).$$

But $q_{1,1} = 1, p_{1,j} = 0$, and $q_{k,j} | q^{k-1}$ for $k > 1$, so $\sum_{k=1}^{\ell} t^{k-1} p_{k,j} \frac{q^{k-1}}{q_{k,j}} \in \mathbb{Z}$. Hence, by defining, say,

$$h_s = qs \prod_{p \leq d} p$$

for $\ell < s \leq d$, we have that $\{h_1, \dots, h_d\}$ is admissible, $h_1 < h_2 < \dots < h_d$, and we can write

$$\nu(h_s) = \sum_{j=1}^{\ell} b_{s,j} \nu(h_j)$$

for

$$b_{s,j} = \sum_{k=1}^{\ell} \left(s \prod_{p \leq d} p \right)^{k-1} p_{k,j} \frac{q^{k-1}}{q_{k,j}} \in \mathbb{Z}.$$

Now, as in the proof of Proposition 3.1, by Taylor's theorem, for $j = 1, 2, \dots, d$, and $x \geq h_d$, $x \geq 1$, we can write

$$F(x + h_j) = \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} h_j^n x^{\rho_i - n} + O_{F, \mathcal{H}}(x^{\{\rho_r\}-1}). \quad (5.2)$$

In particular, we have that for $s = \ell + 1, \ell + 2, \dots, d$,

$$\begin{aligned}
F(x + h_s) &= \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} h_s^n x^{\rho_i-n} + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}) \\
&= \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} \left(\sum_{j=1}^{\ell} b_{s,j} h_j^n \right) x^{\rho_i-n} + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}) \\
&= \sum_{j=1}^{\ell} b_{s,j} \sum_{i=1}^r d_i \sum_{n=0}^{\ell-1} \binom{\rho_i}{n} h_j^n x^{\rho_i-n} + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}) \\
&= \sum_{j=1}^{\ell} b_{s,j} F(x + h_j) + O_{F,\mathcal{H}}(x^{\{\rho_r\}-1}),
\end{aligned}$$

where in the last line we use Equation (5.2) and the fact that $b_{s,j}$ depends only on $\ell, s, p_{k,j}, q_{k,j}$, which, in turn, depend only on $\mathcal{H}$. □

As a consequence of Proposition 3.1, assuming U is of the form $(1 - \varepsilon, 1) \cup [0, \varepsilon)$ for some $\varepsilon > 0$, then for all N large enough, we can find a subset of (5.1) that can be shown to satisfy the hypotheses of Theorem 2.2 by using the results of Sections 3 and 4.

Proposition 5.3. *Let $F(x) = \sum_{i=1}^r d_i x^{\rho_i}$ with $0 \leq \rho_1 < \dots < \rho_r$ be a polynomial with real exponents, with $\rho_r \notin \mathbb{Z}, d_r \neq 0$ and $\ell = \lfloor \rho_r \rfloor + 1$. Let $d \geq \ell$ be an integer and let $\varepsilon > 0$ be any real number. For these $d, \ell, F(x)$, let $(b_{s,1}, \dots, b_{s,\ell}) \in \mathbb{Z}^\ell, \ell + 1 \leq s \leq d$ be the vectors obtained in the conclusion of Proposition 5.2. Denote*

$$\Delta = \frac{\varepsilon}{2\ell \max\{|b_{s,j}| : \ell < s \leq d, 1 \leq j \leq \ell\}}.$$

Then, there exists an integer N_0 such that

$$\begin{aligned}
\mathcal{B} &:= \{n \in \mathbb{N} : n \geq N_0, \{F(n + h_j)\} \in [0, \Delta) \text{ for } 1 \leq j \leq \ell\} \\
&\subseteq \{n \in \mathbb{N} : n \geq N_0, \|F(n + h_j)\| < \varepsilon \text{ for } 1 \leq j \leq d\}.
\end{aligned}$$

Proof. Notice that if

$$\{F(n + h_j)\} \in [0, \Delta)$$

for $j = 1, \dots, \ell$, then for every $\ell < k \leq d$ and every $1 \leq j \leq \ell$,

$$\|b_{k,j} F(n + h_j)\| < \frac{\varepsilon}{2\ell}.$$

Thus, if n is large enough in terms of F , the h_i 's, d , and ℓ (say $n \geq N_0$), we have

$$\|F(n + h_k)\| < \sum_{j=1}^{\ell} \|b_{k,j} F(n + h_j)\| + \frac{\varepsilon}{2} < \varepsilon.$$

It follows that

$$\begin{aligned}\mathcal{B} &:= \{n \in \mathbb{N} : n \geq N_0, \{F(n + h_j)\} \in [0, \Delta) \text{ for } 1 \leq j \leq \ell\} \\ &\subseteq \{n \in \mathbb{N} : n \geq N_0, \|F(n + h_j)\| < \varepsilon \text{ for } 1 \leq j \leq d\}.\end{aligned}$$

□

Proof of Theorem 1.2. First, notice that it suffices to prove the theorem for U a non-empty open set of the form $(1 - \varepsilon, 1) \cup [0, \varepsilon)$ for some $\varepsilon > 0$. For any other open sets, we can choose a constant c such that $U + c \bmod 1$ contains a set U' of the above form, and apply the theorem to $F(x) + c$ and U' . In this case, the relevant set $\mathcal{A}$ takes the form $\mathcal{A} := \{n \in \mathbb{N} : \|F(n)\| < \varepsilon\}$. Let $d \geq \max\{\ell, 2\}$ be an integer. By Proposition 5.3, there exists an admissible set of polynomials $\mathcal{L} = \{n + h_1, \dots, n + h_d\}$ such that $h_1 < \dots < h_d$ and with

$$\begin{aligned}\mathcal{B} &:= \{n \in \mathbb{N} : n \geq N_0, \{F(n + h_j)\} \in [0, \Delta) \text{ for } 1 \leq j \leq \ell\} \\ &\subseteq \{n \in \mathbb{N} : n \geq N_0, \|F(n + h_j)\| < \varepsilon \text{ for } 1 \leq j \leq d\}\end{aligned}$$

for some $0 < \Delta < \varepsilon$ and some integer $N_0 \in \mathbb{N}$. We show that $\mathcal{B}$ together with the set of admissible polynomials $\mathcal{L}$ satisfy the hypotheses of Theorem 2.2 for $B = \alpha = 1$ and some $0 < \theta < 1$ depending only on F .

For Hypothesis 1, let $\delta = 0.98 > 1/\log d$. Let h be any element of $\{h_1, \dots, h_d\}$. Then, notice that $n \in \mathcal{B}$ if and only if $n + h$ is in

$$\mathcal{C} = \{n \in \mathbb{N} : n \geq N_0 + h, \{F(n - h + h_j)\} \in [0, \Delta) \text{ for } 1 \leq j \leq \ell\}.$$

Therefore, by Lemma 4.5 applied to h , for all N large enough in terms of F and $\mathcal{L}$, we have that

$$\begin{aligned}\sum_{\substack{N \leq n < 2N \\ n+h \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) &= \sum_{\substack{N+h \leq n < 2N+h \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{C}}(n) \\ &\geq \frac{0.99}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{C}}(n) - h \\ &\geq \frac{0.99}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) - \left(1 + \frac{0.99}{\log N}\right) h \\ &\geq \frac{\delta}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n).\end{aligned}$$

In particular, we have that

$$\frac{1}{d} \sum_{j=1}^d \sum_{\substack{N \leq n < 2N \\ n+h_j \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) \geq \frac{\delta}{\log N} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)$$

as long as N is large enough.

For Hypotheses 2 and 3, it follows from Proposition 3.5 and Lemma 4.4 that there are some $0 < \theta_1, \theta_2 < 1$ depending only on F such that

$$\sum_{q \leq N^{\theta_1}} \max_{0 \leq c \leq q-1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv c \pmod{q}}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{q} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n) \right| \ll_{F,d,\varepsilon} \frac{1}{(\log N)^{101d^2}} \sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)$$

and, for every $h \in \{h_1, \dots, h_d\}$,

$$\begin{aligned} & \sum_{q < N^{\theta_2}} \max_{(q,a+h)=1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q} \\ n+h \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) - \frac{1}{\varphi(q)} \sum_{\substack{N \leq n < 2N \\ n+h \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) \right| \\ & \ll \sum_{q < N^{\theta_2}} \max_{(q,a)=1} \left| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q} \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{C}}(n) - \frac{1}{\varphi(q)} \sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{C}}(n) \right| + hN^{\theta_2} \\ & \ll_{F,d,\varepsilon} \frac{1}{(\log N)^{101d^2}} \sum_{\substack{N \leq n < 2N \\ n \in \mathbb{P}}} \mathbf{1}_{\mathcal{C}}(n) + N^{\theta_2} \\ & \ll_{F,d,\varepsilon} \frac{1}{(\log N)^{101d^2}} \sum_{\substack{N \leq n < 2N \\ n+h \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n), \end{aligned}$$

where, for the last step, we have used the fact that $\theta_2 < 1$ and the fact that

$$\sum_{\substack{N \leq n < 2N \\ n+h \in \mathbb{P}}} \mathbf{1}_{\mathcal{B}}(n) \gg_{F,d,\varepsilon} \pi(N, 2N),$$

which follows from the proof of Lemma 4.4. The result obviously also holds if θ_1 and θ_2 are replaced by $\theta = \min\{\theta_1, \theta_2\}$. This θ is again dependent on F only, and, in particular, is independent of the choice of d .

Finally, Hypothesis 4 is Corollary 3.6.

It follows from the conclusion of Theorem 2.2 that there exists some constant C depending only on θ , and, therefore, only on F , such that for all d and N large enough, the number of $n \in \mathcal{B} \cap [N, 2N)$ such that the set $\{n + h_1, \dots, n + h_d\}$ contains a string of at least $\lceil 0.98C^{-1} \log d \rceil$ consecutive primes is

$$\gg \frac{\sum_{N \leq n < 2N} \mathbf{1}_{\mathcal{B}}(n)}{(\log N)^d \exp(Cd^5)},$$

where the implied constant may depend on F and ε , but not on N and d . It follows

from the proof of Proposition 3.5 that

$$\sum_{N \leq n < 2N} 1_{\mathcal{B}}(n) \gg_{F,d,\varepsilon} N$$

for all N large enough, so, for any fixed d , for all N large enough in terms of F , d and ε , there will be at least one $n \in \mathcal{B} \cap [N, 2N)$ such that the set $\{n + h_1, \dots, n + h_d\}$ contains a string of at least $\lceil 0.98C^{-1} \log d \rceil$ consecutive primes, and, by our choice of $\mathcal{B}$, they are all contained in the set $\mathcal{A}$. Therefore, for any $m \in \mathbb{N}$, by taking d large enough (so that all the conditions in Theorem 2.2 are satisfied) and with $\lceil 0.98C^{-1} \log d \rceil \geq m$, there are infinitely many strings of m consecutive primes of bounded length (depending on F and m) inside $\mathcal{A}$. □

References

- [BKM⁺14] Vitaly Bergelson, Grigori Kolesnik, Manfred Madritsch, Younghwan Son, and Robert Tichy. Uniform distribution of prime powers and sets of recurrence and van der Corput sets in $\mathbb{Z}^k$. *Israel Journal of Mathematics*, 201(2):729–760, April 2014.
- [BZ16] Roger C. Baker and Liangyi Zhao. Gaps between primes in Beatty sequences. *Acta Arithmetica*, 172(3):207–242, 2016.
- [CPS15] Lynn Chua, Soohyun Park, and Geoffrey Smith. Bounded gaps between primes in special sequences. *Proceedings of the American Mathematical Society*, 143(11):4597–4611, 2015.
- [DT97] Michael Drmota and Robert F. Tichy. *Sequences, Discrepancies and Applications*. Springer Berlin Heidelberg, 1997.
- [IK04] Henryk Iwaniec and Emmanuel Kowalski. *Analytic Number Theory*, volume 53. American Mathematical Society, 2004.
- [May15] James Maynard. Small gaps between primes. *Annals of Mathematics*, 181(1):383–413, 2015.
- [May16] James Maynard. Dense clusters of primes in subsets. *Compositio Mathematica*, 152(7):1517–1554, April 2016.
- [MPY13] Keenan Monks, Sarah Peluse, and Lynnelle Ye. Strings of special primes in arithmetic progressions. *Archiv der Mathematik*, 101(3):219–234, sep 2013.
- [MS17] Kaisa Matomäki and Xuancheng Shao. Vinogradov’s three primes theorem with almost twin primes. *Compositio Mathematica*, 153(6):1220–1256, April 2017.
- [Pol14] Paul Pollack. Bounded gaps between primes with a given primitive root. *Algebra and Number Theory*, 8(7):1769–1786, 2014.
- [Rad21] Artyom O. Radomskii. Consecutive Primes in Short Intervals. *Proceedings of the Steklov Institute of Mathematics*, 314(1):144–202, sep 2021.

- [Shi00] Daniel Shiu. Strings of congruent primes. *Journal of the London Mathematical Society*, 61(2):359–373, 2000.
- [Shu20] A. V. Shubin. Fractional parts of noninteger powers of primes. *Mathematical Notes*, 108(3-4):394–408, 2020.
- [ST21] Xuancheng Shao and Joni Teräväinen. The Bombieri-Vinogradov theorem for nilsequences. *Discrete Analysis*, oct 11 2021.
- [Zha14] Yitang Zhang. Bounded gaps between primes. *Annals of Mathematics*, pages 1121–1174, 2014.

Sai Sanjeev Balakrishnan
 DEPARTMENT OF MATHEMATICS,
 UNIVERSITY OF CALIFORNIA, BERKELEY,
 EVANS HALL, BERKELEY, CALIFORNIA,
 UNITED STATES OF AMERICA - 94720-3840
saisanjeev@berkeley.edu

Félix Houde
 DEPARTMENT OF MATHEMATICS AND STATISTICS,
 CONCORDIA UNIVERSITY,
 1455 DE MAISONNEUVE BLVD WEST, MONTREAL, QUEBEC, CANADA, H3G 1M8
f.houd@live.concordia.ca

Vahagn Hovhannisyan
 FACULTY OF MATHEMATICS AND MECHANICS,
 YEREVAN STATE UNIVERSITY,
 1 ALEK MANUKYAN ST, YEREVAN, ARMENIA,
vahagn.hovhannisyan3@edu.y-su.am

Maryna Manskova
 INSTITUTE OF ANALYSIS AND NUMBER THEORY,
 GRAZ UNIVERSITY OF TECHNOLOGY,
 STEYRERGASSE 30, 8010 GRAZ, AUSTRIA,
maryna.manskova@tugraz.at

Yiqing Wang
 DEPARTMENT OF MATHEMATICS,
 UNIVERSITY OF WISCONSIN–MADISON,
 480 LINCOLN DR, MADISON, WI 53706, USA
wang3498@wisc.edu