On generating functions in additive number theory, II: lower-order terms and applications to PDEs

Citation for the original published paper (version of record):
Brandes, J., Parsell, S., Poulias, C. et al (2021). On generating functions in additive number theory, II: lower-order terms and applications to
PDEs. Mathematische Annalen, 379(1-2): 347-376. http://dx.doi.org/10.1007/s00208-020-02107-0

Mathematische Annalen
https://doi.org/10.1007/s00208-020-02107-0

Mathematische Annalen

On generating functions in additive number theory, II: lower-order terms and applications to PDEs

J. Brandes[1] · S. T. Parsell[2] · C. Poulias[3] · G. Shakan[4] · R. C. Vaughan[5]

Received: 16 January 2020 / Revised: 16 October 2020 / Accepted: 17 October 2020

Abstract
We obtain asymptotics for sums of the form

$$\sum_{n=1}^{P} e\left(\alpha_k n^k + \alpha_1 n\right),$$

involving lower order main terms. As an application, we show that for almost all $\alpha_2 \in [0,1)$ one has

$$\sup_{\alpha_1 \in [0,1)} \left| \sum_{1 \le n \le P} e\left(\alpha_1\left(n^3+n\right)+\alpha_2 n^3\right)\right| \ll P^{3/4+\varepsilon},$$

and that in a suitable sense this is best possible. This allows us to improve bounds for the fractal dimension of solutions to the Schrödinger and Airy equations.

Mathematics Subject Classification 11L15 · 11P55 · 35Q53 · 35Q55

1 Introduction

Exponential sums are a ubiquitous tool throughout analytic number theory, and have been studied in their own right at least since the 1920s. When $\mathbf{k} = (k_1, \ldots, k_t)$ is a tuple of pairwise distinct natural numbers and P is a large positive integer, the Weyl sum of multidegree $\mathbf{k}$ is given by

$$f_{\mathbf{k}}\left(\alpha_{k_1}, \ldots, \alpha_{k_t}\right) = \sum_{n=1}^{P} e\left(\alpha_{k_1} n^{k_1} + \cdots + \alpha_{k_t} n^{k_t}\right). \tag{1.1}$$

Communicated by Kannan Soundararajan.

Extended author information available on the last page of the article

Published online: 23 December 2020

Such sums regularly feature in applications of the Hardy-Littlewood circle method in connection with diophantine systems of the shape

$$x_1^{k_j} + \cdots + x_s^{k_j} = y_1^{k_j} + \cdots + y_s^{k_j} \quad (1 \le j \le t). \tag{1.2}$$

Whilst the theory of systems of the kind (1.2) has recently seen significant advances in the work of Wooley [22,23] and Bourgain, Demeter and Guth [3] on Vinogradov's mean value theorem, our grasp of the cases involving lacunary degrees remains insufficient. One of the simplest such systems is that corresponding to $\mathbf{k} = (1, 3)$, a variant of which is given by

$$x_1^3 + \cdots + x_s^3 = x_1 + \cdots + x_s = 0. \tag{1.3}$$

Although recent progress on this system has been achieved by Brüdern and Robert [5] and Wooley [20], a full understanding of the system (1.3) remains tantalisingly out of reach. In both papers, the authors apply the circle method in order to derive asymptotic formulæ for the number of solutions of such systems, and they succeed in doing so as soon as $s \ge 10$. On the basis of standard heuristics, one would expect to be able to extend the range in which formulæ of this kind are valid to at least $s \ge 9$, but unfortunately we lack a sufficiently precise understanding of the underlying Weyl sum $f_{1,3}(\boldsymbol{\alpha})$ to achieve such a bound. A similar phenomenon occurs in forthcoming work of Hughes and Wooley [11], which deals with moments of a weighted version of $f_{1,3}(\boldsymbol{\alpha})$. Trying to make some headway towards a better understanding of these exponential sums is the main motivation behind the paper at hand.

The main motif underpinning the Hardy-Littlewood method is that sums of the shape (1.1) should be small unless all components of the coefficient vector $\boldsymbol{\alpha}$ lie in the vicinity of fractions with a small denominator, in which case they can be well approximated by certain generating functions that are easier to handle and encode the adelic information inherent in the associated system (1.2). To make this notion precise, we introduce some notation. Suppose that the entries of $\boldsymbol{\alpha}$ have a rational approximation of the shape

$$\alpha_{k_j} = a_{k_j}/q + \beta_{k_j} \quad (1 \le j \le t) \tag{1.4}$$

with a common denominator q satisfying $\gcd(q, a_{k_1}, \ldots, a_{k_t}) = 1$, and define

$$S_{\mathbf{k}}(q; \mathbf{a}) = \sum_{x=1}^{q} e\left(q^{-1} \sum_{j=1}^{t} a_{k_j} x^{k_j}\right) \quad \text{and} \quad I_{\mathbf{k}}(\boldsymbol{\beta}) = \int_0^P e\left(\sum_{j=1}^{t} \beta_{k_j} x^{k_j}\right) \mathrm{d}x.$$

In this notation, we anticipate that $q^{-1} S_{\mathbf{k}}(q; \mathbf{a}) I_{\mathbf{k}}(\boldsymbol{\beta})$ should be a good approximation to $f_{\mathbf{k}}(\boldsymbol{\alpha})$, and we denote the difference by

$$\Delta_{\mathbf{k}}(q, \mathbf{a}; \boldsymbol{\beta}) = f_{\mathbf{k}}(\mathbf{a}/q + \boldsymbol{\beta}) - q^{-1} S_{\mathbf{k}}(q; \mathbf{a}) I_{\mathbf{k}}(\boldsymbol{\beta}). \tag{1.5}$$

There is a considerable body of work related to Weyl sums of the type (1.1) and their approximations (1.5). When $t = 1$ so that $\mathbf{k} = k$, it is known from [17, Theorem 4.1] that

$$\Delta_k(q, a; \beta) \ll q^{1/2+\varepsilon}(1 + P^k|\beta|)^{1/2}, \tag{1.6}$$

and Daemen [7, Theorem 2] and Brüdern and Daemen [4, Theorem 1] showed that this bound is sharp up to at most a factor of q^{ε}. For general multidegrees $\mathbf{k}$ we have the weaker bound

$$\Delta_{\mathbf{k}}(q, \mathbf{a}; \boldsymbol{\beta}) \ll q \left(1 + \sum_{j=1}^{t} P^{k_j}|\beta_{k_j}|\right)$$

from [17, Theorem 7.2]. This bound has been improved for $\mathbf{k} = (1, k)$ by Brüdern and Robert [5, Theorem 3], who obtained the estimate

$$\Delta_{1,k}(q, \mathbf{a}; \boldsymbol{\beta}) \ll q^{1-1/k+\varepsilon} \left(1 + P^k|\beta_k|\right)^{1/2}. \tag{1.7}$$

Whilst their result holds for all $k \geq 2$, an epsilon-free version is available in the quadratic case due to the last author [18, Theorem 8]. This invites the question of how optimal the bound in (1.7) is.

The primary objective of this memoir is to examine the exponential sum $f_{1,k}(\alpha_1, \alpha_k)$ and its associated error term $\Delta_{1,k}(q, \mathbf{a}; \boldsymbol{\beta})$ more closely. Our main result is the following.

Theorem 1.1 *Let $k \geq 2$. Assume (1.4) with $(q, a_k) = 1$ and $|\beta_1| \leq (2q)^{-1}$. Then*

$$\begin{aligned} f_{1,k}(\alpha_1, \alpha_k) = q^{-1} \sum_{d|q}^{\dagger} S_{1,k}\left(q; d\,[[a_1/d]], a_k\right) I_{1,k}\left(\alpha_1 - \frac{[[a_1/d]]}{q/d}, \beta_k\right) \\ + O\left(q^{1/2+\varepsilon}(1 + |\beta_k|P^k)^{1/2} \log P\right), \end{aligned}$$

where $[[x]]$ denotes a closest integer to x, and the notation $\sum^{\dagger}$ indicates that the sum runs over all distinct values of $d\,[[a_1/d]]$ satisfying $([[a_1/d]], q/d) = 1$.

If, moreover, we have

$$|\beta_k| \leq (4kqP^{k-1})^{-1}, \tag{1.8}$$

then the error term in the above asymptotic may be replaced by $O(q^{1/2+\varepsilon})$.

Thus, by extracting additional main terms, we are able to obtain an error term of the same quality as in (1.6), which is essentially optimal. We note that the factor $\log P$ in the error term can be eliminated by means of a more careful analysis. Observe also that the coprimality condition $([[a_1/d]], q/d) = 1$ implies that the fractions $[[a_1/d]]/(q/d)$ are reduced and therefore pairwise distinct.

In the case when $k = 3$, it follows from Dirichlet's approximation theorem that every $\alpha \in [0, 1]$ has an approximation $\alpha = a/q + \beta$ with $q(1 + P^3|\beta|) \le 2P^{3/2}$. Thus, in the cubic case we obtain the following.

Corollary 1.1 *Assume* (1.4) *with* $q(1 + P^3|\beta_3|) \le 2P^{3/2}$*, where* $(q, a_3) = 1$ *and* $|\beta_1| \le (2q)^{-1}$.

(a) *We have*

$$f_{1,3}(\alpha_1, \alpha_3) = q^{-1} \sum_{d|q}^{\dagger} S_{1,3}(q; d\,[[a_1/d]], a_3)\, I_{1,3}\left(\alpha_1 - \frac{[[a_1/d]]}{q/d}, \beta_3\right) + O(P^{3/4+\varepsilon}).$$

(b) *Moreover, we have the bound*

$$f_{1,3}(\alpha_1, \alpha_3) \ll \frac{P^{1+\varepsilon}}{\left(q + q|\beta_3|P^3\right)^{1/3}} + P^{3/4+\varepsilon}$$

uniformly in α_1.

For general degree k, an analogous chain of reasoning would replace the error term $O(P^{3/4+\varepsilon})$ by $O(P^{k/4+\varepsilon})$, which is trivial when $k \ge 4$. Thus, our result in Theorem 1.1 is strongest in the cubic case, and for higher degrees should be viewed as a bound for the major arcs only.

When $d = (a_1, q)$ we have $d\,[[a_1/d]] = a_1$; this is the leading term in Theorem 1.1 and corresponds to the approximation in (1.5). We can thus rephrase the conclusion of Theorem 1.1 in the form

$$\Delta_{1,k}(q, \mathbf{a}; \boldsymbol{\beta}) = q^{-1} \sum_{\substack{d|q \\ d \ne (q,a_1)}}^{\dagger} S_{1,k}(q; d\,[[a_1/d]], a_k)\, I_{1,k}\left(\alpha_1 - \frac{[[a_1/d]]}{q/d}, \beta_k\right) + O\left(q^{1/2+\varepsilon}(1 + |\beta_k|P^k)^{1/2}\log P\right). \tag{1.9}$$

When $d \ne (a_1, q)$ the fractions a_1/d are all non-integral. In particular, when a_1/d is half an odd integer, there are two choices for $[[a_1/d]]$ and both may occur in the sum. Note further that the sum in (1.9) is empty if and only if a_1 is a multiple of q. When $q \nmid a_1$ we have $(a_1, q) \ne q$, and thus it will always contain at least the term $S_{1,k}(q; 0, a_k) I_{1,k}(\alpha_1, \beta_k)$ corresponding to $d = q$, and in many cases this is the only one. For instance, it is not hard to see that when $a_1 = 1$ and $q > 1$ is odd, the sum in Theorem 1.1 contains precisely the terms corresponding to $d = 1$ and $d = q$, and so in this case the asymptotic formula reads

$$f_{1,k}(\alpha_1, \alpha_k) = q^{-1}\left(S_{1,k}(q; 1, a_k)\, I_{1,k}(\beta_1, \beta_k) + S_{1,k}(q; 0, a_k) I_{1,k}(\alpha_1, \beta_k)\right) + O\left(q^{1/2+\varepsilon}\left(1 + |\beta_k|P^k\right)^{1/2}\log P\right).$$

This behaviour, which occurs in many generic situations, indicates that we cannot expect the secondary terms to be subject to any significant cancellation. In fact, we have the following result on the size of the error term.

Theorem 1.2 *Suppose that a_k and q are coprime integers satisfying $|\alpha_k - a_k/q| < q^{-2}$, and assume that (1.4) holds with $|\beta_1| < (2q)^{-1}$.*

(a) *We have the upper bound*

$$\Delta_{1,k}(q, \mathbf{a}; \boldsymbol{\beta}) \ll q^{1/2+\varepsilon} \left(1 + \sum_{\substack{d|q \\ d \neq (q,a_1)}}^{\dagger} \frac{|S_{1,k}(q; d\,[[a_1/d]]\,, a_k)|}{q^{1/2}} \right) \times \left(1 + |\beta_k| P^k\right)^{1/2} \log P.$$

(b) *Suppose now that $\beta_k = 0$ and $a_1 = 1$ with $q = p^k$, where p is an odd prime. Then, whenever $\|\alpha_1 P\| \geq \delta$ for some suitable real number $\delta > 0$, we have the lower bound*

$$|\Delta_{1,k}(q, 1, a_k; \beta_1, 0)| \geq \frac{4\delta}{3\pi} q^{1-1/k} + O\left(q^{1/2+\varepsilon} \log P\right).$$

Thus Theorem 1.2 shows that the bound (1.7) of Brüdern and Robert is sharp at least when β_k is small and q is a perfect k-th power of a square-free number. Here, the term of size $q^{1-1/k} = p^{k-1}$ arises from the exponential sum $S_{1,k}(q; 0, a_k)$ via Lemma 4.4 in [17], and as noted above, unless $q|a_1$ this term will always appear in the sum in (1.9). Thus, large values of $\Delta_{1,k}(q, \mathbf{a}; \boldsymbol{\beta})$ cannot be considered an exceptional occurrence when β_k is small.

As a consequence of Theorem 1.1, we are able to make progress on a problem concerning the fractal dimension of solutions of certain partial differential equations. The motivation for this problem goes back to optical experiments by Talbot [15] in the 1830s concerning the diffraction of light passing through a grating. Berry [2] later initiated the theoretical investigation of the problem and has in particular made predictions regarding the fractal dimension of the diffraction pattern along certain slices in space. The reader is referred to Chapter 2 of [9] and the introduction of [8] for an introduction to the general topic as well as a more thorough history of this particular problem.

In this paper, we shall focus in particular on the family of partial differential equations given by

$$i\partial_t q(t, x) - i^k \partial_x^{(k)} q(t, x) = 0 \quad (k \in \mathbb{N}), \tag{1.10}$$

where $t \in \mathbb{R}$ and $x \in \mathbb{R}/2\pi\mathbb{Z}$. When $k = 2$, the reader will recognize (1.10) as the linear Schrödinger equation, while the case $k = 3$ corresponds to the linear part of the Korteweg-de Vries (KdV) equation, also known as Airy's equation. For any natural number k, given initial data $g_k(n) \in L^2(\mathbb{R}/2\pi\mathbb{Z})$, the evolution of g_k under (1.10) is given by

$$q_k(t, x) = \sum_{n\in\mathbb{Z}} \hat{g}_k(n) e^{itn^k + ixn}.$$

Clearly, q_k is periodic in both t and x with period 2π.

We are interested in the restriction of q_k to linear subsets of $(\mathbb{R}/2\pi\mathbb{Z}) \times (\mathbb{R}/2\pi\mathbb{Z})$. Given $c \in \mathbb{R}$ and $r \in \mathbb{Q} \setminus \{0\}$, as well as initial data g_k, let

$$q_{k;r,c}(x) = \sum_{n\in\mathbb{Z}} \hat{g}_k(n) e^{i(c-rx)n^k + ixn}$$

denote the restriction of q_k to the oblique line $t + rx = c$. Recall that the fractal (also known as upper Minkowski or upper box-counting) dimension of a bounded set E is given by

$$\overline{\dim}(E) = \limsup_{\varepsilon\to 0} \frac{\log\left(\mathcal{N}(E, \varepsilon)\right)}{\log\left(1/\varepsilon\right)},$$

where $\mathcal{N}(E, \varepsilon)$ is the minimum number of ε-balls required to cover E. Assuming that g_k is a suitably well-behaved function, we would like to know the the fractal dimension of the real and imaginary parts of the graph of $q_{k;r,c}$ for a typical c. Note here that it is possible for either the real or the imaginary part of the graph to vanish, so we are really interested in the size of the larger of the two. The simplest non-constant choices for g_k are step functions, and in such situations, we see that in order to make progress, it is imperative to understand the distribution of large values of exponential sums. As it is convenient to work with dyadic sums in this context, we modify our definition (1.1) by writing

$$f_{\mathbf{k}}\left(\alpha_{k_1}, \ldots, \alpha_{k_t}; Q\right) = \sum_{Q<n\le 2Q} e\left(\alpha_{k_1} n^{k_1} + \cdots + \alpha_{k_t} n^{k_t}\right), \tag{1.11}$$

where Q is a positive number. Let Θ_k denote the set of all $\theta \in \mathbb{R}$ such that for almost all $\gamma \in [0, 1)$ one has

$$\sup_{Q\ge 1} Q^{-\theta} \sup_{\alpha\in[0,1)} |f_{1,k}(\alpha, \alpha+\gamma; Q)| \ll_{\theta,\gamma} 1, \tag{1.12}$$

and set $\theta_k = \inf \Theta_k$. The size of θ_k and related quantities has recently been studied by Chen and Shparlinski [6], building on work by Wooley [21]. Clearly, one sees that

$$\lfloor 2Q\rfloor - \lfloor Q\rfloor = \int_0^1 |f_{1,k}\left(\alpha, \alpha+\gamma; Q\right)|^2 \mathrm{d}\alpha \le \left(\sup_{\alpha\in[0,1)} |f_{1,k}(\alpha, \alpha+\gamma; Q)|\right)^2 \le Q^2,$$

whence we have the trivial bounds

$$1/2 \le \theta_k \le 1 \tag{1.13}$$

valid for all $k \ge 2$ and for the entire range $\gamma \in [0, 1)$. Moreover, we have the trivial bound $\sup_{\alpha,\gamma} |f_{1,k}(\alpha, \alpha+\gamma; Q)| \asymp Q$, and it is known (see e.g. [6, Corollary 2.2]) that for independent variables γ, α we have $|f_{1,k}(\alpha, \alpha+\gamma; Q)| \ll Q^{1/2+\varepsilon}$ almost everywhere. It turns out that in our case where only one of the variables is restricted to lie in the complement of a thin set while the other one ranges freely, the bound is appreciably larger.

Theorem 1.3 *We have $\theta_2 = \theta_3 = 3/4$.*

Theorem 1.3 may be a bit surprising as one naively expects square root cancellation in exponential sums. As will transpire from the proof, it turns out that for almost every γ in (1.12) the supremum is obtained for a special choice of α on what can be considered a major arc. One might speculate that $\theta_k = 3/4$ for all $k \ge 2$. Indeed, one might hope to adapt the proof of Theorem 1.3 above to show that for almost all γ this gives the correct extremal value on a suitable set of major arcs and that for almost all γ the sum is smaller on the corresponding minor arcs. This latter speculation would be consistent with the main result of the last author and Wooley [19].

With the help of Theorem 1.3, we can address our motivating problem.

Theorem 1.4 *For $k = 2, 3$ let g_k be a step function, and fix $r \in \mathbb{Q} \setminus \{0\}$. Set $\alpha_2 = 1/8$ and $\alpha_3 = 1/12$. Then, for almost every $c \in \mathbb{R}$, the function $q_{k;r,c}$ satisfies the Hölder condition C^α for every $\alpha < \alpha_k$. In particular, the fractal dimension of the graph of the real and imaginary parts of $q_{k;r,c}$ is at most $2 - \alpha_k$.*

This improves the values of $\alpha_2 = 1/10$ and $\alpha_3 = 1/27$ of the fourth author with Erdoğan [8, Theorem 1.1]. The proof is identical to that of [8, Corollary 3.5], but inputs our Theorem 1.3 instead of [8, Proposition 3.3]. Moreover, the results of Theorem 1.4 can be transferred to non-linear partial differential equations, in particular the non-linear Schrödinger and KdV equations, by the same methods as Theorems 1.2 and 1.3 are derived from Theorem 1.1 in [8].

Note that Theorem 1.1 in [8] also gives a lower bound for the fractal dimensions in question. Specifically, the authors show that the graph of at least one of the real and imaginary parts of $q_{k;r,c}$ has fractal dimension of at least $2 - 1/(2k)$, and this is sharp at least in the Schrödinger case $k = 2$. Moreover, they remark that, if it were true that $\theta_k = 1/2$, their argument could be adapted to show that this lower bound reflects the actual value. Our Theorem 1.3 rules out this approach at least for the cases $k = 2$ and $k = 3$. Meanwhile, if our speculation that $\theta_k = 3/4$ could be substantiated for all k, it would imply that the maximum dimension of the respective graphs of the real and imaginary parts would lie in the range $[2 - \frac{1}{2k}, 2 - \frac{1}{4k}]$. It is worth noting that Lemma 2 of [12] along with [8] imply that for special combinations of initial data and oblique lines the fractal dimension in Theorem 1.4 for $k = 2$ is precisely 7/4 (see [8, Footnote 3] for more details).

Notation Throughout the paper, we make use of the following conventions. All statements involving the letter ε are claimed to be true for all (sufficiently small) $\varepsilon > 0$.

Thus, the precise ‘value’ of ε is allowed to change from one line to the next. Moreover, P always denotes a large positive number. We use the Vinogradov and Bachmann–Landau notations liberally, and here the implied constants are allowed to depend on k and ε, but never on P, Q or $\boldsymbol{\alpha}$.

2 Preliminary lemmata

In this section, we briefly collect some technical lemmata that will be of use in our arguments later. All of these results pertain to the case $\mathbf{k} = (1, k)$, and in order to avoid clutter, we will in our arguments below drop the multidegree $(1, k)$ in our notation. Throughout, Q denotes a positive number. For easier reference, we begin by stating a few results from the literature.

Lemma 2.1 *Let $a_1, a_k \in \mathbb{Z}$ and $q \in \mathbb{N}$, and suppose that $(a_k, q) = 1$.*

(a) *Uniformly in a_1, we have $S(q; a_1, a_k) \ll q^{1-1/k+\varepsilon}$.*
(b) *Moreover, $S(q; a_1, a_k) \ll q^{1/2+\varepsilon}(q, a_1)$.*

Proof These are Theorem 7.1 and Lemma 4.1 in [17], respectively. □

We also record an elementary average bound for exponential sums.

Lemma 2.2 *For any positive integer q and any integer a_k we have*

$$\sum_{b=1}^{q} |S(q; b, a_k)| \leq q^{3/2}.$$

Proof By the Cauchy-Schwarz inequality we see that

$$\sum_{b=1}^{q} |S\,(q; b, a_k)\,| \leq q^{1/2} \left(\sum_{b=1}^{q} |S(q; b, a_k)|^2 \right)^{1/2},$$

and expanding the square yields

$$\sum_{b=1}^{q} |S\,(q; b, a_k)\,|^2 = \sum_{b=1}^{q} \sum_{x,y=1}^{q} e\left(\frac{b(x-y) + a_k\left(x^k - y^k\right)}{q} \right) = q \sum_{x=1}^{q} 1 = q^2.$$

This completes the proof. □

The next result is a direct consequence of [17, Lemma 4.2].

Lemma 2.3 *Suppose that ϕ is a twice continuously differentiable function on an interval I and let $H \geq 2$ be a number such that $|\phi'(x)| \leq H$ for all $x \in I$. Suppose further that ϕ'' has at most finitely many zeros in the interval I. Then*

$$\sum_{x \in I \cap \mathbb{Z}} e(\phi(n)) = \sum_{|h| \leq H} \int_I e(\phi(x) - hx)\mathrm{d}x + O(\log H).$$

Proof This is immediate upon partitioning I into subintervals on which ϕ' is monotonic, and then applying Lemma 4.2 of [17] on each of these finitely many intervals. □

We continue with bounds on oscillating integrals. For a measurable subset $\mathcal{A}$ we define

$$I(\beta_1, \beta_k; \mathcal{A}) = \int_{\mathcal{A}} e\left(\beta_1 x + \beta_k x^k\right) \mathrm{d}x.$$

We then have the following bounds for $I(\beta_1, \beta_k; \mathcal{A})$.

Lemma 2.4 *Let $k \geq 2$ and suppose that $\mathcal{A}$ is a finite union of pairwise disjoint intervals.*

(a) *Let $\tau > 0$ be a parameter satisfying $|\beta_1 + k\beta_k x^{k-1}| \geq \tau$ for all $x \in \mathcal{A}$. Then*

$$I(\beta_1, \beta_k; \mathcal{A}) \ll \tau^{-1}.$$

(b) *Assume that $\mathcal{A} \subseteq [Q, 2Q]$ for some $Q > 0$. Then, whenever $\beta_k \neq 0$ we have*

$$I(\beta_1, \beta_k; \mathcal{A}) \ll \left(|\beta_k| Q^{k-2}\right)^{-1/2}.$$

Proof These bounds are Lemmata 4.2 and 4.4 in [16], respectively, applied to the function $F(x) = \beta_1 x + \beta_k x^k$.

Lemma 2.5 *Assume that $k \geq 2$ and $\beta_1 \neq 0$. Suppose further that $\mathcal{A}$ is a union of finitely many pairwise disjoint intervals contained inside $[Q, 2Q]$ for some $Q > 0$. Then we have the bound*

$$I(\beta_1, \beta_k; \mathcal{A}) \ll |\beta_1|^{-1} \left(1 + Q^k |\beta_k|\right)^{1/2}.$$

Proof Suppose first that the relation

$$|\beta_1 - k\beta_k x^{k-1}| \geq \tfrac{1}{2}|\beta_1| \tag{2.1}$$

holds for all $x \in \mathcal{A}$. Then we see from Lemma 2.4(a) that

$$I(\beta_1, \beta_k; \mathcal{A}) \ll |\beta_1|^{-1}, \tag{2.2}$$

which is sufficient to prove the lemma in this case. We may thus concentrate on the opposite case where the inequality (2.1) is violated for some $x \in \mathcal{A}$. It follows from the triangle inequality that any such x must satisfy the inequalities $\frac{1}{2}|\beta_1| \leq k|\beta_k|x^{k-1} \leq \frac{3}{2}|\beta_1|$. Since $Q \leq x \leq 2Q$, this can happen only if

$$|\beta_1| \asymp Q^{k-1}|\beta_k| \tag{2.3}$$

and in particular only when $\beta_k \neq 0$. We can thus deploy Lemma 2.4(b) and obtain

$$I(\beta_1, \beta_k; \mathcal{A}) \ll \left(Q^{k-2}|\beta_k|\right)^{-1/2} \\ \ll \left(|\beta_k|Q^{k-1}\right)^{-1}\left(|\beta_k|Q^k\right)^{1/2} \ll |\beta_1|^{-1}\left(|\beta_k|Q^k\right)^{1/2}, \qquad (2.4)$$

where in the last step we used (2.3) again. The full statement now follows upon combining (2.2) and (2.4). □

3 Proof of Theorem 1.1

For the proof of our first main result it is convenient to work over dyadic ranges. Recalling our notation (1.11), we make the analogous definition

$$I(\beta_1, \beta_k; Q) = I_{1,k}(\beta_1, \beta_k; Q) = \int_Q^{2Q} e\left(\beta_1 x + \beta_k x^k\right) \mathrm{d}x.$$

Thus, if we can show that

$$f(\alpha_1, \alpha_k; Q) = q^{-1} \sum_{d|q}^{\dagger} S(q; d[[a_1/d]], a_k) I\left(\alpha_1 - \frac{[[a_1/d]]}{q/d}, \beta_k; Q\right) \\ + O(q^{1/2+\varepsilon}\left(1 + |\beta_k|Q^k)^{1/2}\right), \qquad (3.1)$$

for any $Q \geq 1/2$, the conclusion of Theorem 1.1 will follow upon dyadic summation, as

$$f(\alpha_1, \alpha_k) = \sum_{i=1}^{\lceil \log P/\log 2 \rceil} f\left(\alpha_1, \alpha_k; 2^{-i}P\right).$$

The initial stages of our argument follow along the lines of the proof of [5, Theorem 3], which in turn is an adaptation of the argument found in [17, pp. 43–44].

Lemma 3.1 *Assume* (1.4) *with* $(a_1, q) = 1$*, and set*

$$H = 2^{k-1}q\left(1 + kQ^{k-1}|\beta_k|\right).$$

Then

$$f(\alpha_1, \alpha_k; Q) = q^{-1} \sum_{|h| \leq H} S(q; a_1 + h, a_k) I(\beta_1 - h/q, \beta_k; Q) + O\left(q^{1/2}\log H\right).$$

Proof By sorting the terms of $f(\alpha_1, \alpha_k; Q)$ into congruence classes modulo q and encoding the congruence condition in an exponential sum, we find that

$$\begin{aligned} f(\alpha_1, \alpha_k; Q) &= \sum_{r=1}^{q} e\left(\frac{a_1 r + a_k r^k}{q}\right) \sum_{\substack{Q<n\leq 2Q \\ n\equiv r \pmod q}} e\left(\beta_1 n + \beta_k n^k\right) \\ &= \frac{1}{q} \sum_{-q/2<b\leq q/2} S(q; a_1+b, a_k)\, f(\beta_1 - b/q, \beta_k; Q). \end{aligned} \tag{3.2}$$

We treat the sum $f(\beta_1 - b/q, \beta_k; Q)$ by Lemma 2.3, where we take

$$\phi(x) = (\beta_1 - b/q)x + \beta_k x^k$$

and set

$$H_1 = 2^{k-1}(1 + kQ^{k-1}|\beta_k|) - 1/2.$$

Then we have $|\phi'(x)| \leq H_1$ for all $x \leq 2Q$ and thus Lemma 2.3 yields

$$f(\beta_1 - b/q, \beta_k; Q) = \sum_{|j|\leq H_1} I(\beta_1 - b/q - j, \beta_k; Q) + O(\log H_1).$$

Using this within (3.2) and applying Lemma 2.2 in the error term yields

$$\begin{aligned} f(\alpha_1, \alpha_k; Q) = \frac{1}{q} &\sum_{-q/2<b\leq q/2} \sum_{|j|\leq H_1} S(q; a_1+b+jq, a_k)\, I(\beta_1 - (b+jq)/q, \beta_k; Q) \\ &+ O\left(q^{1/2}\log H_1\right). \end{aligned}$$

The proof is complete upon making the change of variables $b + qj = h$, noting that under the summation conditions this is in fact a bijection into the set of integers h satisfying $-H < h \leq H$ where $H = q(H_1 + 1/2)$. □

We now distinguish two cases according to which term in H is larger. Suppose first that $k|\beta_k|Q^{k-1} > 1$, so that

$$1 \ll H/q \ll |\beta_k|Q^{k-1}. \tag{3.3}$$

In such a situation, we discern from Lemma 2.4(b) and Lemma 2.2 that

$$\begin{aligned} \sum_{|h|\leq H} S(q; a_1+h, a_k) I(\beta_1 - h/q, \beta_k; Q) &\ll (Q^{k-2}|\beta_k|)^{-1/2}\left(\frac{H}{q}+1\right)\sum_{a=1}^{q}|S(q; a, a_k)| \\ &\ll q^{3/2}Q^{k/2}|\beta_k|^{1/2}, \end{aligned}$$

where in the last step we used (3.3). Thus, in this situation, we find that

$$f(\alpha_1, \alpha_k; Q) \ll q^{1/2}\left(Q^{k/2}|\beta_k|^{1/2} + \log H\right) \ll q^{1/2+\varepsilon} Q^{k/2}|\beta_k|^{1/2}, \tag{3.4}$$

which is satisfactory for the purposes of Theorem 1.1.

It remains to study the behaviour of $f(\alpha_1, \alpha_k; Q)$ when $k|\beta_k|Q^{k-1} \leq 1$, or in other words,

$$H \ll q. \tag{3.5}$$

Set $d = (a_1 + h, q)$ and $e = (a_1 + h)/d$. In this notation, we have $h = de - a_1$ and $(e, q/d) = 1$, and the conclusion of Lemma 3.1 reads

$$\begin{aligned} &f(\alpha_1, \alpha_k; Q) \\ &\quad = q^{-1}\sum_{d|q} \sum_{\substack{e\in\mathbb{Z} \\ |de-a_1|\leq H \\ (e,q/d)=1}} S(q; de, a_k)\, I\left(\beta_1 - \frac{de-a_1}{q}, \beta_k; Q\right) + O\left(q^{1/2+\varepsilon}\right). \end{aligned} \tag{3.6}$$

We expect the sum on the right hand side of (3.6) to be dominated by the terms corresponding to small values of h. In particular, whenever $|h| \leq d/2$ we have that $|e-a_1/d| \leq 1/2$ and hence $e = [[a_1/d]]$. These terms will form our main term. Write

$$E(q, \mathbf{a}; \boldsymbol{\beta}) = q^{-1}\sum_{d|q} \sum_{\substack{e\in\mathbb{Z} \\ d/2<|de-a_1|\leq H}} \left| S(q; de, a_k) I\left(\alpha_1 - \frac{de}{q}, \beta_k; Q\right)\right|$$

for the sum over all the remaining terms where $h > d/2$. Then (3.6) may be rephrased as

$$\begin{aligned} f(\alpha_1, \alpha_k; Q) = q^{-1}&\sum_{d|q}^{\dagger} S(q; d\,[[a_1/d]], a_k)\, I\left(\alpha_1 - \frac{[[a_1/d]]}{q/d}, \beta_k; Q\right) \\ &+ O\left(E(q, \mathbf{a}; \boldsymbol{\beta}) + q^{1/2+\varepsilon}\right). \end{aligned} \tag{3.7}$$

Thus, it suffices to bound $E(q, \mathbf{a}; \boldsymbol{\beta})$. By Lemma 2.5 we see that

$$E(q, \mathbf{a}; \boldsymbol{\beta}) \ll q^{-1}\left(1 + Q^k|\beta_k|\right)^{1/2} \sum_{d|q} \sum_{\substack{e\in\mathbb{Z} \\ d/2<|de-a_1|\leq H}} \frac{|S(q; de, a_k)|}{|\alpha_1 - de/q|}. \tag{3.8}$$

Now, the condition $de \neq a_1$ together with our bound $|\beta_1| \leq (2q)^{-1}$ implies that

$$\left|\alpha_1 - \frac{de}{q}\right| \geq \frac{|a_1 - de|}{q} - |\beta_1| \geq \frac{|a_1 - de|}{q} - \frac{1}{2q} \geq \frac{|a_1 - de|}{2q}. \tag{3.9}$$

Using this within (3.8) and applying Lemma 2.1(b) yields the bound

$$\begin{aligned} E\left(q,\mathbf{a};\boldsymbol{\beta}\right) &\ll \left(1+Q^k|\beta_k|\right)^{1/2}\sum_{d|q}\sum_{\substack{e\in\mathbb{Z}\\ d/2<|de-a_1|\le H}}\frac{|S\left(q;de,a_k\right)|}{|a_1-de|}\\ &\ll q^{1/2+\varepsilon}\left(1+Q^k|\beta_k|\right)^{1/2}\sum_{d|q}d\sum_{\substack{e\in\mathbb{Z}\\ d/2<|de-a_1|\le H}}|a_1-de|^{-1}\\ &\ll q^{1/2+\varepsilon}\left(1+Q^k|\beta_k|\right)^{1/2}, \end{aligned} \tag{3.10}$$

where in the last step we used (3.5) together with standard bounds for the divisor function. The proof of (3.1) under the assumption (3.5) is now complete upon inserting (3.10) into (3.7), and the unconditional statement follows upon combining this with the bound (3.4).

The second statement of Theorem 1.1 is proved in a similar manner, and we only briefly detail the changes that need to be effected. Here, we do not need to consider a dyadic dissection of the interval, so all of our arguments will involve the exponential sum $f(\alpha_1,\alpha_k)$ and integral $I(\beta_1,\beta_k)$ instead of their dyadic analogues $f(\alpha_1,\alpha_k;Q)$ and $I(\beta_1,\beta_k;Q)$, and will have P instead of Q. We now observe that in the proof of Lemma 3.1, the condition (1.8) implies that

$$|\phi'(x)|=|\beta_1-b/q+k\beta_k x^{k-1}|\le\frac{1}{2q}+\frac{1}{2}+\frac{1}{4q}\le 2.$$

We may thus take $H_1=2$. The argument then proceeds as above, with the difference that we may skip the discussion of the case (3.3) and can continue directly with the hypothesis (3.5). From this point we arrive, *mutatis mutandis*, at (3.7). In order to bound the error term $E(q,\mathbf{a};\boldsymbol{\beta})$ we now note that (3.9) and (1.8) combine to show that, for $1\le x\le P$, one has

$$\left|\alpha_1-\frac{de}{q}+k\beta_k x^{k-1}\right|\ge\left|\alpha_1-\frac{de}{q}\right|-\frac{1}{4q}\ge\frac{|a_1-de|}{4q}.$$

It thus follows from Lemma 2.4(a) that (3.8) can be replaced by

$$E\left(q,\mathbf{a};\boldsymbol{\beta}\right)\ll\sum_{d|q}\sum_{\substack{e\in\mathbb{Z}\\ d/2<|de-a_1|\le H}}\frac{|S\left(q;de,a_k\right)|}{|a_1-de|},$$

and the desired bound $E(q,\mathbf{a};\boldsymbol{\beta})\ll q^{1/2+\varepsilon}$ follows as above. This completes the proof of Theorem 1.1. Moreover, Corollary 1.1(a) is now immediate, and part (b) follows easily upon applying Theorem 7.3 of [17] and Lemma 2.1(a) within Theorem 1.1.

We now turn to the proof of Theorem 1.2. When $d\neq(a_1,q)$, the fractions a_1/q and $d\,[[a_1/d]]\,/q$ are distinct, and thus the latter one corresponds to a non-optimal rational approximation to α_1. In particular, we have $|\alpha_1-q^{-1}d\,[[a_1/d]]\,|\ge 1/(2q)$.

Thus, we can apply Lemma 2.5 within the sum (1.9) much as above, and the statement of Theorem 1.2(a) follows immediately.

For the second statement of the theorem, we begin by observing that the hypotheses imply that the sum in Theorem 1.1 has exactly two main terms, corresponding to the values $d = 1$ and $d = q$, respectively. We will focus on the latter. Note that when $\beta_k = 0$, we can explicitly compute

$$I(\alpha_1, 0) = \int_0^P e(\alpha_1 x)\,\mathrm{d}x = \frac{e(\alpha_1 P) - 1}{2\pi i \alpha_1} = e\left(\frac{\alpha_1 P}{2}\right)\frac{\sin(\pi\alpha_1 P)}{\pi\alpha_1}.$$

When $\|\alpha_1 P\| > \delta$, we have $2\delta \le |\sin(\pi\alpha_1 P)| \le 1$, and thus

$$\frac{4q\delta}{3\pi} \le \frac{2\delta}{\pi|\alpha_1|} \le |I(\alpha_1, 0)| \le \frac{1}{\pi|\alpha_1|} \le \frac{2q}{\pi}$$

under our assumption that $1/(2q) \le \alpha_1 \le 3/(2q)$. Thus, the main term corresponding to $d = q$ is given by

$$q^{-1}|S(q; 0, a_k)\, I(\alpha_1, 0)| \ge \frac{4\delta}{3\pi}|S(q; 0, a_k)|.$$

Finally, we note that when $q = p^k$, Lemma 4.4 in [17] shows that $|S(q; 0, a_k)| = q^{1-1/k}$, which implies the result.

4 Additional lemmata for the proof of Theorem 1.3

For the proof of our second main result, we need some more detailed information about cubic complete exponential sums.

Lemma 4.1 *Let $q \in \mathbb{N}$ and set*

$$q_2 = \prod_{\substack{p^t \| q \\ t=1 \text{ or } 2}} p^t, \quad q_3 = \prod_{\substack{p^t \| q \\ t \ge 3}} p^t, \quad \kappa(q) = q_2^{1/2} q_3^{1/3}.$$

Furthermore, suppose that $(q, a) = 1$. Then

$$S_{1,3}(q; b, a) \ll q^{1+\varepsilon}\kappa(q)^{-1}.$$

Proof Suppose that $q = rs$ with $(r, s) = 1$, and write $a = a_2 r + a_1 s$ and $b = b_2 r + b_1 s$. Then it follows by standard arguments that

$$S_{1,3}(q; b, a) = S_{1,3}(r; b_1, a_1)\, S_{1,3}(s; b_2, a_2).$$

We are therefore free to restrict our focus to prime power moduli. By Lemma 2.1(a) when $(q_3, a) = 1$ we have

$$S_{1,3}\left(q_3; b, a\right) \ll q_3^{2/3+\varepsilon}.$$

Thus it suffices to show that whenever $(a, p) = 1$ and $t = 1$ or 2 we have

$$S_{1,3}\left(p^t; b, a\right) \ll p^{t/2}.$$

When $t = 1$ this follows from Corollary II.2F of Schmidt [13]. Suppose $t = 2$. Then when $p = 2$ or 3 this bound is trivial, so we can suppose that $p > 3$. Thus

$$\begin{aligned} S_{1,3}(p^2; b, a) &= \sum_{v=0}^{p-1}\sum_{u=1}^{p} e\left(\frac{b(pv+u)+a(pv+u)^3}{p^2}\right) \\ &= \sum_{u=1}^{p} e\left(\frac{bu+au^3}{p^2}\right)\sum_{v=0}^{p-1}\left(\frac{b+3au^2}{p}v\right) \\ &= p \sum_{\substack{u=1 \\ b+3au^2\equiv 0 \ (\mathrm{mod}\ p)}}^{p} e\left(\frac{bu+au^3}{p^2}\right). \end{aligned}$$

The congruence $b + 3au^2 \equiv 0 \pmod p$ has at most two solutions, and it follows that $|S_{1,3}(p^2; b, a)| \leq 2p$. The claim of the lemma follows upon collecting our results. □

Lemma 4.2 *Let $q \in \mathbb{N}$ be odd and $c \in \mathbb{Z}$ with $(q, c) = 1$. Then there exists $a \in \mathbb{Z}$ such that $(q, a) = 1$ and for every $\varepsilon > 0$ we have*

$$|S_{1,3}(q; a - c, a)| \gg q^{1/2-\varepsilon}.$$

Proof Before embarking on the argument, we note that the lemma is trivial for $q = 1$, and hence we may assume $q > 1$ and consequently $c \neq 0$. Again we use the multiplicative property of the Gauss sum as described at the start of the proof of the previous lemma. Thus it suffices to establish the lemma for prime powers.

Let p be an odd prime, $t \in \mathbb{N}$ and $c \in \mathbb{Z}$ be such that $p \nmid c$. The lemma follows if we are able to show that there is an absolute constant $\xi > 0$ having the property that

$$|S_{1,3}(p^t; a - c, a)| \geq \xi p^{t/2}$$

for all odd prime powers p^t.

First we deal with the case $t = 1$. Clearly, the desired statement follows if we can show that

$$\sum_{a=1}^{p-1} |S_{1,3}(p; a - c, a)|^2 \geq \xi p^2. \tag{4.1}$$

In general, we have

$$\sum_{a=1}^{p-1} |S_{1,3}\,(p; a-c, a)\,|^2 = p \sum_{\substack{m,n=1 \\ n+n^3 \equiv m+m^3 \pmod{p}}}^{p} e\left(\frac{c(n-m)}{p}\right) - \left|\sum_{m=1}^{p} e\left(\frac{cm}{p}\right)\right|^2 .$$

Clearly, the second sum vanishes. Hence

$$\begin{aligned}\sum_{a=1}^{p-1} |S_{1,3}\,(p; a-c, a)\,|^2 &= p \sum_{m=1}^{p} \sum_{\substack{n=1 \\ n+n^3 \equiv m+m^3 \pmod{p}}}^{p} e\left(\frac{c(n-m)}{p}\right) \\ &= p^2 + p \sum_{m=1}^{p} \sum_{\substack{h=1 \\ 3m^2+3hm+h^2+1 \equiv 0 \pmod{p}}}^{p-1} e\left(\frac{ch}{p}\right). \qquad (4.2)\end{aligned}$$

When $p = 3$ the congruence in the inner sum becomes $h^2 \equiv -1 \pmod 3$ which is insoluble, so that sum vanishes, and (4.2) reads

$$\sum_{a=1}^{2} |S_{1,3}(3; a-c, a)|^2 = 3^2.$$

When $p > 3$, on the other hand, we use that

$$12\left(3m^2 + 3hm + h^2 + 1\right) = (6m + 3h)^2 + 3h^2 + 12,$$

whence we obtain

$$p \sum_{m=1}^{p} \sum_{\substack{h=1 \\ 3m^2+3hm+h^2+1 \equiv 0 \pmod{p}}}^{p-1} e\left(\frac{ch}{p}\right) = p \sum_{h=1}^{p-1} e\left(\frac{ch}{p}\right)\left(1 + \left(\frac{-3h^2-12}{p}\right)_L\right),$$

where $\left(\frac{a}{p}\right)_L$ denotes the Legendre symbol. Thus, upon extending the sum on the right hand side to include the term $h = p$ also, while noting that $-3p^2 - 12 \equiv 2^2(-3) \pmod p$, we obtain

$$\begin{aligned}\sum_{a=1}^{p-1} |S_{1,3}(p; a-c, a)|^2 = p^2 - p \left(1 + \left(\frac{-3}{p}\right)_L\right) \\ + p \sum_{h=1}^{p} e\left(\frac{ch}{p}\right)\left(\frac{-3h^2-12}{p}\right)_L . \qquad (4.3)\end{aligned}$$

At this point, we see from Theorem II.2G in [13] that

$$\sum_{h=1}^{p} e\left(\frac{ch}{p}\right)\left(\frac{-3h^2-12}{p}\right)_L \leq 2p^{1/2}, \tag{4.4}$$

whence we find that

$$\sum_{a=1}^{p-1} |S_{1,3}(p; a-c, a)|^2 \geq p^2 - 2p - 2p^{3/2}.$$

When $p > 7$, this is already sufficient for (4.1), so it remains to analyse the cases when $p = 5$ and $p = 7$.

Let now $p = 5$. Since $\left(\frac{-3}{5}\right)_L = -1$, the desired bound (4.1) follows with $\xi = 1 - 2/\sqrt{5}$ upon deploying (4.4) within the expression in (4.3). Finally, consider the case $p = 7$. Since $\left(\frac{-3}{7}\right)_L = 1$, we have in (4.3) that

$$\sum_{a=1}^{6} |S_{1,3}(7; a-c, a)|^2 = 42 + 7\sum_{h=1}^{6} e\left(\frac{ch}{7}\right)\left(\frac{-3h^2-12}{7}\right)_L,$$

where we observed that the summand corresponding to $h = 7$ is 1. The remaining sum over h is

$$-e\left(\frac{c}{7}\right) + e\left(\frac{2c}{7}\right) - e\left(\frac{3c}{7}\right) - e\left(\frac{4c}{7}\right) + e\left(\frac{5c}{7}\right) - e\left(\frac{6c}{7}\right),$$

which has absolute value smaller than 6 for all $c \in \{1, \ldots, 6\}$. It follows that (4.1) holds in this case also. Thus whenever $t = 1$ there is at least one a satisfying the necessary requirements.

Now consider the case $t \geq 2$. Suppose first that $p > 3$. Write $t = 3v + u$ with $v \geq 0$ and $1 \leq u \leq 3$. When $u \neq 1$ choose $a = c$. Then by iteratively applying Lemma 4.4 of [17] we find that

$$S_{1,3}(p^t; a-c, a) = S_{1,3}(p^t; 0, a) = p^{2v+u-1} \geq p^{t/2},$$

since in the notation of that lemma and (2.25) *ibidem* we have $l = t > 1 = \gamma$.

When $u = 1$, we may now assume that $v \geq 1$. Choose a so that $a \equiv c + p^{2v}(a' - c) \pmod{p^t}$ where a' is at our disposal. Put $m = p^{3v}x + y$. Then our sum is

$$S_{1,3}(p^{3v+1}; p^{2v}(a'-c), a) = \sum_{y=1}^{p^{3v}} \sum_{x=0}^{p-1} e\left(\frac{3ay^2x}{p} + \frac{(a'-c)y}{p^{v+1}} + \frac{ay^3}{p^{3v+1}}\right).$$

The sum over x is 0 unless $p|y$ in which case it sums to p, and thus the above is

$$p\sum_{z=1}^{p^{3v-1}} e\left(\frac{(a'-c)z}{p^v}+\frac{az^3}{p^{3(v-1)+1}}\right) = p^2 S_{1,3}(p^{3(v-1)+1}; p^{2v-2}(a'-c), a).$$

Iterating this argument gives

$$S_{1,3}(p^t; a-c, a) = p^{2v} S_{1,3}(p; a'-c, a).$$

Now we choose a' in accordance with the case $t=1$ above. Thus

$$|S_{1,3}(p^t; a-c, a)| \geq \xi p^{2v+1/2} \geq \xi p^{t/2}.$$

When $p=3$ we can apply a slightly modified argument. Now in the notation (2.25) of [17] we have $\gamma = 2$. When $t = 3v+u$ with $u=2$ or 3 we again take $a=c$ and obtain, by Lemma 4.4 *ibidem*,

$$S_{1,3}\left(3^t; a-c, a\right) = 3^{2v} S_{1,3}\left(3^u; 0, a\right)$$

and this is 3^{2v+2} when $u=3$. When $u=2$, we have instead

$$S_{1,3}\left(3^u; 0, a\right) = 3 + 6\cos\left(2\pi a/9\right).$$

Since a is not divisible by 3, the cosine cannot be $-\frac{1}{2}$. Thus

$$|S_{1,3}(3^t; a-c, a)| \geq \xi 3^{t/2}.$$

When $u=1$ we follow the recipe for general p and obtain

$$S_{1,3}(3^t; a-c, a) = 3^{2v} S_{1,3}(3; a'-c, a)$$

and again appeal to the case $t=1$. □

5 Prolegomena to the proof of Theorem 1.3

Before proceeding to the various parts of the proof of Theorem 1.3, it is useful to review some measure theoretic aspects of approximation of real numbers by rational numbers. In view of the periodicity of our functions, we concentrate on the interval [0, 1].

As is well known, Dirichlet's approximation theorem states that every real number γ has the property that there are arbitrarily large $q \in \mathbb{N}$ and $c \in \mathbb{Z}$ such that $(q, c) = 1$ and $|\gamma - c/q| \leq q^{-2}$. Moreover, it follows from Khinchine's theorem (see e.g. Theorem

III.3A in [14]) that for almost every such γ there is a positive number $C(\gamma)$ such that whenever $(q,c)=1$ we have

$$\frac{C(\gamma)}{q^2(\log 2q)^2} \le \left|\gamma - \frac{c}{q}\right|. \tag{5.1}$$

In particular there is a subset Γ of $(\mathbb{R}\setminus\mathbb{Q})\cap[0,1]$ having these properties and with $\operatorname{meas}\Gamma = 1$.

For the upper bound when $k=3$ we need to refine this further. Let Γ_0 denote the subset of Γ with the property that for every $\delta>0$ and $\gamma\in\Gamma_0$ there are, in the notation of Lemma 4.1, at most a finite number of q and c with

$$\left|\gamma - \frac{c}{q}\right| \le q_2^{-2-\delta} q_3^{-4/3-\delta}. \tag{5.2}$$

Lemma 5.1 *The set Γ_0 has full measure in* $[0,1]$.

Proof Let $\Upsilon_0 = \bigcup_{\delta>0}\Upsilon(\delta)$ where $\Upsilon(\delta)$ denotes the set of $\gamma\in[0,1]$ having the property that (5.2) holds for infinitely many pairs (q,c). Let further $N\in\mathbb{N}$. Then

$$\Upsilon(\delta) \subseteq \bigcup_{q>N}\bigcup_{0\le c\le q}\left\{\gamma : \left|\gamma - \frac{c}{q}\right| \le q_2^{-2-\delta} q_3^{-4/3-\delta}\right\}.$$

Hence

$$\operatorname{meas}\Upsilon(\delta) \le \sum_{q_2q_3>N} 4q_2^{-1-\delta}q_3^{-1/3-\delta}.$$

We have

$$\sum_{q_3\le X} 1 \le \sum_{\substack{r^3 l\le X\\ l\mid r^2}} 1 \le \sum_{r\le X^{1/3}} d(r^2) \ll X^{1/3+\varepsilon}.$$

Therefore for any $Y>0$ we have

$$\sum_{q_3>Y} q_3^{-1/3-\delta} = \sum_{k=0}^{\infty}\sum_{2^kY<q_3\le 2^{k+1}Y} q_3^{-1/3-\delta} \ll Y^{-\delta/2},$$

and so

$$\begin{aligned}\sum_{q_2q_3>N} 4q_2^{-1-\delta}q_3^{-1/3-\delta} &\ll \sum_{q_2\le N} q_2^{-1-\delta}\sum_{q_3>N/q_2} q_3^{-1/3-\delta} + \sum_{q_2>N} 4q_2^{-1-\delta}\\ &\ll \sum_{q_2\le N} q_2^{-1-\delta/2}N^{-\delta/2} + N^{-\delta}.\end{aligned}$$

Thus

$$\text{meas}\Upsilon(\delta) \ll N^{-\delta/2}$$

and this holds for every $N \in \mathbb{N}$. □

6 Theorem 1.3: the upper bound when $k = 2$

Let $\alpha \in \mathbb{R}$ and $\gamma \in \varGamma$. By Dirichlet's theorem on diophantine approximation we can choose a_2, q with $(a_2, q) = 1$, $q \leq Q$ and

$$\left|\alpha + \gamma - \frac{a_2}{q}\right| \leq \frac{1}{qQ}.$$

Then choose a_1 so that

$$\left|\alpha - \frac{a_1}{q}\right| \leq \frac{1}{2q}.$$

Set $\beta_1 = \alpha - a_1/q$ and $\beta_2 = \alpha + \gamma - a_2/q$. Hence, by Theorem 8 of [18], we have

$$f_{1,2}(\alpha, \alpha + \gamma; Q) = q^{-1} S_{1,2}(q; a_1, a_2)\, I_{1,2}(\beta_1, \beta_2; Q) + O\left(Q^{1/2}\right).$$

Since $(a_2, q) = 1$ we have $|S_{1,2}(q; a_1, a_2)| \ll \sqrt{q}$ by classical bounds on the Gauss sum. Thus

$$f_{1,2}(\alpha, \alpha + \gamma; Q) \ll |I_{1,2}(\beta_1, \beta_2; Q)|q^{-1/2} + Q^{1/2}.$$

Therefore, by Theorem 7.1 in [17] we have the bound

$$f_{1,2}(\alpha, \alpha + \gamma; Q) \ll \frac{Q}{\left(q + qQ|\beta_1| + qQ^2|\beta_2|\right)^{1/2}} + Q^{1/2}. \tag{6.1}$$

We may assume that

$$q \leq \left(\tfrac{1}{2}C(\varepsilon, \gamma)\, Q\right)^{1/(2+\varepsilon)}, \tag{6.2}$$

for otherwise in the denominator in (6.1) we have

$$q + qQ|\beta_1| + qQ^2|\beta_2| \geq \left(\tfrac{1}{2}C(\varepsilon, \gamma)\, Q\right)^{1/(2+\varepsilon)}$$

trivially. Since $\gamma \in \varGamma$, we infer from (5.1) via (6.2) that

$$|\beta_1| \geq \left|\gamma - \frac{a_2 - a_1}{q}\right| - |\beta_2| \geq \frac{C(\varepsilon, \gamma)}{q^{2+\varepsilon}} - \frac{1}{qQ} \geq \frac{1}{2}C(\varepsilon, \gamma)q^{-2-\varepsilon},$$

provided that $Q \geq 2/C(\varepsilon, \gamma)$, which we may certainly assume. Thus, we find

$$q + qQ|\beta_1| \geq q + \tfrac{1}{2}QC(\varepsilon, \gamma)\, q^{-1-\varepsilon} \geq \left(\frac{1}{2}C(\varepsilon, \gamma)\, Q\right)^{1/(2+\varepsilon)}$$

in this case as well, and (6.1) becomes

$$f_{1,2}\,(\alpha, \alpha + \gamma;\, Q) \ll_{\varepsilon,\gamma} Q^{1-1/(4+2\varepsilon)}.$$

We conclude that if $\theta > \frac{3}{4}$, then

$$Q^{-\theta} \sup_{\alpha \in [0,1)} |f_{1,2}\,(\alpha, \alpha + \gamma;\, Q)| \ll_{\theta,\gamma} 1$$

as required.

7 Theorem 1.3: the upper bound when $k = 3$

This follows the pattern set in §6. Again we use (5.1), but now we suppose that $\gamma \in \Gamma_0$, and hence given γ we may suppose that for any fixed $\delta > 0$ the inequality (5.2) holds for at most a finite number of q and c. It will be convenient to also suppose that δ is sufficiently small. For such a γ we show that for arbitrarily large Q we have

$$f_{1,3}(\alpha, \alpha + \gamma;\, Q) \ll_{\delta,\gamma} Q^{3/4+\delta} \tag{7.1}$$

uniformly for $\alpha \in \mathbb{R}$.

Let $\alpha \in \mathbb{R}$. By Dirichlet's theorem on diophantine approximation we can choose a_3, q with

$$(a_3, q) = 1, \quad q \leq Q^{3/2} \quad \text{and} \quad \left|\alpha + \gamma - \frac{a_3}{q}\right| \leq \frac{1}{qQ^{3/2}}. \tag{7.2}$$

Then choose a_1 so that

$$-\frac{1}{2q} < \alpha - \frac{a_1}{q} \leq \frac{1}{2q},$$

and let $\beta_3 = \alpha + \gamma - a_3/q$ and $\beta_1 = \alpha - a_1/q$. By Corollary 1.1(a) we have

$$f_{1,3}\,(\alpha, \alpha + \gamma;\, Q) = q^{-1} \sideset{}{^\dagger}\sum_{d|q} S_{1,3}\,(q;\, d\,[[a_1/d]]\,, a_3)\, I_{1,3}\left(\alpha - \frac{d\,[[a_1/d]]}{q}, \beta_3;\, Q\right) + O\left(Q^{3/4+\varepsilon}\right). \tag{7.3}$$

Since $(a_3, q) = 1$, by Theorem 7.3 of [17] together with Lemma 4.1 we have

$$q^{-1} S_{1,3}(q; d\,[[a_1/d]]\,, a_3) I_{1,3}\left(\alpha - \frac{d\,[[a_1/d]]}{q}, \beta_3; Q\right) \\ \ll Q^{1+\varepsilon}\kappa(q)^{-1}\left(1 + Q\left|\alpha - \frac{d\,[[a_1/d]]}{q}\right| + Q^3|\beta_3|\right)^{-1/3}.$$

It follows that the contribution arising from those $d|q$ for which

$$\kappa(q)^3\left(1 + Q\left|\alpha - \frac{d\,[[a_1/d]]}{q}\right| + Q^3|\beta_3|\right) \geq Q^{3/4-2\delta} \tag{7.4}$$

is satisfactory in view of (7.1).

Thus it remains to deal with any possible terms for which (7.4) fails to hold. Suppose that $d \neq (a_1, q)$ for some d violating (7.4), so that

$$\left|\alpha - \frac{d\,[[a_1/d]]}{q}\right| \geq \frac{1}{2q}.$$

In that case we would have

$$q_2^{1/2} Q = q_2^{3/2} q_3 q^{-1} Q \leq 2\kappa(q)^3 Q\left|\alpha - \frac{d\,[[a_1/d]]}{q}\right| \leq 2Q^{3/4-2\delta},$$

which is impossible for Q large. Hence the only term in (7.3) that could possibly violate (7.4) is the one corresponding to $d = (q, a_1)$, for which

$$\alpha - \frac{d\,[[a_1/d]]}{q} = \alpha - \frac{a_1}{q} = \beta_1.$$

For this term the negation of (7.4) reads

$$\kappa(q)^3\left(1 + Q|\beta_1| + Q^3|\beta_3|\right) < Q^{3/4-2\delta}, \tag{7.5}$$

and we observe that in this instance

$$q \leq q_2^{3/2} q_3 = \kappa(q)^3 \leq Q^{3/4-2\delta}. \tag{7.6}$$

We assume there is such a term and show that it contradicts the assumption $\gamma \in \Gamma_0$. Since $\Gamma_0 \subseteq \Gamma$, we see from (5.1) and (7.2) that

$$\frac{C(\gamma)}{q^2(\log 2q)^2} < \left|\gamma - \frac{a_3 - a_1}{q}\right| \leq |\beta_1| + |\beta_3| \leq |\beta_1| + q^{-1}Q^{-3/2}.$$

Since by (7.6) we may suppose that Q is large enough so that

$$q^{-1}Q^{-3/2} \le \frac{C(\gamma)}{2q^2(\log 2q)^2} \le \frac{1}{2}\left|\gamma - \frac{a_3 - a_1}{q}\right|,$$

it follows from our assumption (7.5) that

$$\left|\gamma - \frac{a_3 - a_1}{q}\right| \le 2|\beta_1| \le 2\kappa(q)^{-3}Q^{-1/4-2\delta}. \tag{7.7}$$

We also have

$$\frac{C(\gamma)}{q^2(\log 2q)^2} \le 2\kappa(q)^{-3}Q^{-1/4-2\delta}$$

and therefore

$$C(\gamma)Q^{1/4+2\delta} \le 2q_2^{1/2}q_3(\log 2q)^2 \le 2q(\log 2q)^2.$$

Thus, as Q can be taken large enough so that $(\log 2q)^2 < \frac{1}{2}C(\gamma)Q^{2\delta}$, we have $q > Q^{1/4}$. By (7.6) we have

$$\begin{aligned} Q^{-1/4-2\delta} &= \left(Q^{-3/4+2\delta}\right)^{1/3+\frac{32\delta}{9-24\delta}} \\ &\le \left(q_2^{-3/2}q_3^{-1}\right)^{1/3+\frac{32\delta}{9-24\delta}} \\ &\le q_2^{-1/2-2\delta}q_3^{-1/3-2\delta} \\ &\le \tfrac{1}{2}q_2^{-1/2-\delta}q_3^{-1/3-\delta}. \end{aligned}$$

Hence, by (7.7) we find that

$$\left|\gamma - \frac{a_3 - a_1}{q}\right| \le q_2^{-2-\delta}q_3^{-4/3-\delta}.$$

However, by the definition of Γ_0 in Sect. 5 this is expressly excluded for large q, and so this establishes as promised that (7.5) is impossible. This completes the proof of (7.1), which gives the conclusion $\theta_3 \le \frac{3}{4}$.

8 Theorem 1.3: the lower bound

Let $\delta > 0$ be sufficiently small, and let $k = 2$ or 3. We show that for all $\gamma \in \mathbb{R} \setminus \mathbb{Q}$ there are arbitrarily large Q such that

$$\sup_{\alpha \in [0,1)} \left| \sum_{Q < n \leq 2Q} e\left(\alpha n + (\alpha + \gamma) n^k\right) \right| \gg Q^{3/4-\delta}.$$

The continued fraction algorithm for γ gives q and c with q arbitrarily large, $(q, c) = 1$, and

$$|\gamma - c/q| \leq q^{-2}.$$

Note that any two successive convergents c/q and c'/q' of the continued fraction satisfy $cq' - c'q = \pm 1$ and so $(q, q') = 1$. Thus either q or q' is odd. For an arbitrary odd convergent q and a fixed small parameter $\delta > 0$ set

$$Q = q^{2/(1+2\delta)}. \tag{8.1}$$

Let a_k be any integer with $(q, a_k) = 1$, and if $k = 3$ assume additionally that a_3 is such that $S_{1,3}(q; a_3 - c, a_3) \gg q^{1/2-\delta}$. The existence of such an a_3 is guaranteed by Lemma 4.2. Take further $\alpha = -\gamma + a_k/q$ and $a_1 = a_k - c$, and define $\alpha_k = \alpha + \gamma$, $\alpha_1 = \alpha$ and $\beta_j = \alpha_j - a_j/q$ for $j = 1, k$. Thus

$$\beta_k = \alpha + \gamma - \frac{a_k}{q} = 0 \quad \text{and} \quad \beta_1 = -\gamma + \frac{c}{q},$$

and so in particular

$$|\beta_1| \leq q^{-2} \leq Q^{-1-2\delta}.$$

We have

$$f_{1,k}(\alpha, \alpha + \gamma; Q) = q^{-1} S_{1,k}(q; a_1, a_k)\, I_{1,k}(-\gamma + c/q, 0; Q) + O\left(Q^{1/3+\delta}\right). \tag{8.2}$$

When $k = 2$, this is Theorem 8 in [18], and for $k = 3$ it is a consequence of the second statement of Theorem 1.1. Indeed, if $d \neq (a_1, q)$ we have $d\,[[a_1/d]] \neq a_1$, and thus it follows that $|\alpha - d\,[[a_1/d]]\,/q| > 1/(2q)$ and therefore

$$\left| I_{1,3}\left(\alpha - \frac{d\,[[a_1/d]]}{q}, 0; Q\right) \right| \ll q$$

by Lemma 2.4(a). Moreover, by Lemma 2.1(a) we have

$$S_{1,3}\left(q; d\,[[a_1/d]]\,, a_3\right) \ll q^{2/3+\varepsilon}.$$

Hence we see from (8.1) that

$$\sideset{}{^\dagger}\sum_{\substack{d|q\\ d\neq(a_1,q)}} S_{1,3}(q; d\,[[a_1/d]]\,, a_3) I_{1,3}\left(\alpha - \frac{d\,[[a_1/d]]}{q}, 0; Q\right) \ll \sum_{d|q} q^{5/3+\varepsilon} \ll q Q^{1/3+\delta/2},$$

and consequently (3.1) implies that

$$\begin{aligned}
&f_{1,3}\left(\alpha, \alpha+\gamma; Q\right) - q^{-1} S_{1,3}\left(q; a_1, a_3\right) I_{1,3}\left(-\gamma + c/q, 0; Q\right)\\
&\quad = \frac{1}{q} \sideset{}{^\dagger}\sum_{\substack{d|q\\ d\neq(a_1,q)}} S_{1,3}\left(q; d\,[[a_1/d]]\,, a_3\right) I_{1,3}\left(\alpha - \frac{d\,[[a_1/d]]}{q}, 0; Q\right) + O\left(q^{1/2+\varepsilon}\right)\\
&\quad \ll Q^{1/3+\delta/2} + q^{1/2+\varepsilon} \ll Q^{1/3+\delta}
\end{aligned}$$

as claimed.

Note that

$$I_{1,k}\left(\beta_1, 0; Q\right) = \int_Q^{2Q} e\left(\beta_1 x\right) \mathrm{d}x = Qe\left(3Q\beta_1/2\right) \frac{\sin\left(\pi\beta_1 Q\right)}{\pi\beta_1 Q} \gg Q, \tag{8.3}$$

where in the last step we used that $Q|\beta_1| = Q|\gamma - c/q| \leq Q^{-2\delta}$ and thus

$$\frac{\sin\left(\pi\beta_1 Q\right)}{\pi\beta_1 Q} \gg 1.$$

Moreover, since q is odd and $(a_k, q) = 1$, we have

$$q^{-1}|S_{1,k}(q; a_1, a_k)| \gg q^{-1/2-\delta} \gg Q^{-1/4-2\delta}. \tag{8.4}$$

This follows from the classical bound for quadratic Gauss sums in the case $k = 2$, and is a consequence of our choice of a_3 via Lemma 4.2 when $k = 3$. Hence upon combining (8.2), (8.3) and (8.4), we see that

$$|\, f_{1,k}\left(\alpha, \alpha+\gamma; Q\right)| \gg Q^{3/4-2\delta},$$

and the theorem follows.

9 Concluding remarks

In conclusion, we make some remarks about what might be the real truth with regard to estimates for Weyl sums on suitable sets of minor arcs. Consider sums of the kind (1.1) with multidegree $\mathbf{k} = (1, \ldots, k)$, where $\boldsymbol{\alpha}$ is ‘sufficiently random’. Then we might guess that in this circumstance $f_{\mathbf{k}}(\boldsymbol{\alpha})$ behaves like a sum of independent random unimodular variables and so the central limit theorem suggests that

$$f_{\mathbf{k}}(\boldsymbol{\alpha}) \ll P^{1/2+\varepsilon}. \tag{9.1}$$

In fact, the received wisdom states that for (9.1) to be true, it should suffice that $\boldsymbol{\alpha}$ is such that for any $a_1, \ldots, a_k, q$ satisfying (1.4) with $(q, a_1, \ldots, a_k) = 1$ we have

$$q + qP|\alpha_1 - a_1/q| + \cdots + qP^k|\alpha_k - a_k/q| > P. \tag{9.2}$$

In the one-dimensional case, $\mathbf{k} = k$, the last author and Wooley [19] show that there is a connection between the size of $f_k(\alpha)$ and the number of solutions of

$$x_1^k + \cdots + x_b^k = y_1^k + \cdots + y_b^k$$

with $0 < x_j, y_j \leq P$, and that the values of $|f_k(\alpha)|$ will even have a normal distribution on the minor arcs. For $k = 2$ we have more precise bounds (see [18, Theorem 7)]) since we can cover all of $[0, 1]^2$ by choosing q, a_2 with $(q, a_2) = 1$, $|\alpha_2 - a_2/q| \leq 1/(5qP)$, $q \leq 5P$ and then taking a_1 with $|\alpha_1 - a_1/q| \leq 1/(2q)$. Thus the main interest lies in the case $k \geq 3$. In order to make a proper comparison with the current state of play we first review what is known.

Take $\mathbf{k} = (1, \ldots, k)$. The Vinogradov mean value theorem (Bourgain, Demeter, Guth [3] and Wooley [22,23]) combined with Theorem 5.2 of Vaughan [17] give

$$f_{\mathbf{k}}(\boldsymbol{\alpha}) \ll P^{1+\varepsilon}\left(\frac{1}{q_j} + \frac{1}{P} + \frac{q_j}{P^j}\right)^{\frac{1}{k(k-1)}} \tag{9.3}$$

when for some $j \geq 2$ there are coprime q_j and a_j such that $|\alpha_j - a_j/q_j| \leq q_j^{-2}$. At least when $q_j + P^j|\alpha_j - a_j| > P^2$ for every j and choice of q_j, a_j, this is the best that we know when $k \geq 6$, and is perhaps also the best that we know when $3 \leq k \leq 5$ and we need to approximate some (presumably non-zero) α_j with $2 \leq j \leq k - 1$. When we have q, a_k with $(q, a_k) = 1$ and $|\alpha_k - a_k/q| \leq q^{-2}$, Weyl's inequality [17, Lemma 2.4] gives

$$f_{\mathbf{k}}(\boldsymbol{\alpha}) \ll P^{1+\varepsilon}\left(\frac{1}{q} + \frac{1}{P} + \frac{q}{P^k}\right)^{2^{1-k}} \tag{9.4}$$

and this is superior to (9.3) when $3 \leq k \leq 5$.

There is an underlying problem when dealing with a sum as general as (1.1). It seems that one ought to consider a general rational approximation to $\boldsymbol{\alpha}$ of the kind $|q\alpha_j - a_j| \leq Q_j^{-1}$ with $(q, a_1, \ldots, a_k) = 1$, but then to cover the whole of $[0, 1]^k$ one needs that q can be as large as $cQ_1 \cdots Q_k$. This means that either q might be much larger than P^k or the intervals about each rational are too large to be able to deduce anything useful. The alternative, as in the statement of (9.3) above, is to deal with one j at a time. If $q_j > P^{\theta_j}$ for some $\theta_j \le 1$ then we are done and that leaves the situation when $q_j \le P^{\theta_j}$ for every j. Now one can take $q = \operatorname{lcm}(q_1, \dots, q_k)$ and the numerators of the rational approximation to $\boldsymbol{\alpha}$ becomes $(a_1q/q_1, \dots, a_kq/q_k)$. However, the likelihood is that any useful bound will still need q fairly constrained in terms of P and so the θ_j will have to be rather small. An example of this process is given in the proof of [17, Theorem 7.4]. Some aspects of methods to overcome this are described in Chapter 5 of Baker [1].

Whilst our result of Theorem 1.3 does not strictly contradict such heuristics as detailed in the opening paragraph of this section, it does raise the question of whether these heuristics might not be too naive in some cases. When $\{p_1, \dots, p_t\}$ is a set of polynomials with a non-vanishing Wronskian, consider the associated exponential sum

$$f_{\mathbf{p}}(\alpha_1, \dots, \alpha_t) = \sum_{1 \le n \le P} e\left(\alpha_1 p_1(n) + \cdots + \alpha_t p_t(n)\right). \tag{9.5}$$

We know from standard bounds that $\sup_{\boldsymbol{\alpha}} |f_{\mathbf{p}}(\boldsymbol{\alpha})| = f_{\mathbf{p}}(\mathbf{0}) = \lfloor P \rfloor$, whilst on the other hand it follows from [6, Corollary 2.2] that whenever the polynomials p_j have a non-vanishing Wronskian, the bound (9.1) holds for a set of $\boldsymbol{\alpha} \in [0,1]^t$ of full measure. Meanwhile, the analogous inequality to (9.2) is not sufficient for the bound (9.1) to hold. This is evidenced in our Theorem 1.3 with the choices $p_1(n) = n^k$ and $p_2(n) = n^k + n$, where $k = 2$ or $k = 3$. Here, it transpires from our arguments that even if α_1 lacks a good rational approximation, for certain choices of α_2 the contributions of the degree k part of the polynomial more or less cancel out, leading to a less random behaviour. It is not clear to the authors whether this behaviour is particular to the presence and influence of the linear term in p_2 or whether there is some underlying phenomenon at work. In the latter scenario, one might now be inclined to guess that if only r of the t coefficients are restricted to lie in a set of full measure whilst the other $t - r$ coefficients are allowed to range over the entire unit interval, that the ensuing bound would interpolate between the two extremes. Thus, one might speculate that when the polynomials p_j are all of the same degree one has the bound

$$\sup_{\alpha_1, \dots, \alpha_r} |f_{\mathbf{p}}(\boldsymbol{\alpha})| \ll P^{1 - r/(2t) + \varepsilon} \quad \text{for almost all } \alpha_{r+1}, \dots, \alpha_t,$$

and that this might in some cases even be sharp (up to epsilon) for a sequence of values P tending to infinity. The exponent here is $1 - r/(2t) = (r/2 + (t - r))/t$ and interpolates between r contributions of $1/2$ and $t - r$ contributions of 1. Whilst this is compatible with the bounds of our Theorem 1.3, there is still hope that stronger bounds may be available if the polynomials in question differ by more than a linear term.

Our understanding is better in the one-dimensional case. On page 43 of Vaughan [17] it was stated (we have changed the notation to be consistent with this memoir) that when $(q, a) = 1$ and $\beta = \alpha - a/q$ it would be very interesting to decide whether the relation

$$f_k(\alpha) = q^{-1}S_k(q,a)I_k(\beta) + O\left((q + qP^k|\beta|)^\theta\right)$$

holds with an exponent θ smaller than $1/2$, and it was even speculated that θ might be as small as $1/k$. This was shown to be false by Daemen [7] and Brüdern and Daemen [4]. One other result has come to our attention. Heath-Brown [10] has shown on the assumption of the *abc* conjecture that if α is a quadratic irrational then

$$\sum_{n\le P} e(\alpha n^3) \ll P^{\frac{5}{7}+\varepsilon}.$$

It may be worthwhile to note that quadratic irrationals are badly approximable numbers so that Heath-Brown's result can be viewed as applying to an 'extreme minor arc' situation.

Finally, we briefly outline an argument, versions of which in other contexts are quite well known, that shows that one cannot expect to bound the exponential sum $f_k(\alpha)$ by anything smaller than $P^{1/2}$ on the minor arcs. Let P be large and choose $R = P^{1+\phi}$ and $Q = P^{k-1-\psi}$, where ϕ and ψ are positive numbers at our disposal and $\phi < \psi$ so that $RQ < P^{k-\delta}$ for some substantial $\delta > 0$. There are various wrinkles that could be introduced to enable a quite large δ.

Let $\mathfrak{M}$ denote the union of the intervals $[a/q - 1/(qQ), a/q + 1/(qQ)]$ with $1 \le a \le q \le R$ and $(q,a) = 1$ and let $\mathfrak{m} = (1/Q, 1 + 1/Q] \setminus \mathfrak{M}$. The total measure of $\mathfrak{M}$ is $\le 2RQ^{-1}$ and so

$$\int_{\mathfrak{M}} |f_k(\alpha)|^2 \mathrm{d}\alpha \ll P^2RQ^{-1} \ll P^{4+\phi+\psi-k}.$$

When $k \ge 4$ this is $\ll P^{1-\delta}$ for some $\delta > 0$ as long as $\phi + \psi < 1 - \delta$. When $k = 3$ this argument can be refined by using the approximations for $f_k(\alpha)$ given by (4.13), Theorem 4.2 and the integral version of Lemma 2.8 of Vaughan [17]. Then we obtain the bound

$$\begin{aligned}\int_{\mathfrak{M}} |f_k(\alpha)|^2 \mathrm{d}\alpha &\ll \sum_{q\le R} q \int_0^{1/(qQ)} \left(\frac{P^2}{(q+qP^3\beta)^{2/3}} + q^\varepsilon(q+qP^3\beta)\right)\mathrm{d}\beta \\ &\ll P^{\frac{1}{3}+\frac{4\phi}{3}} + P^{\frac{1}{3}+\phi+\frac{\psi}{3}} + P^{2\phi+\psi+\varepsilon} + P^{\phi+2\psi+\varepsilon} \ll P^{1-\delta}\end{aligned}$$

whenever $\phi + \psi < \frac{1}{2} - \delta$. Hence by Parseval's identity, for all $k \ge 3$ we have

$$\int_{\mathfrak{m}} |f_k(\alpha)|^2 \mathrm{d}\alpha = P + O(P^{1-\delta}), \tag{9.6}$$

which gives the desired conclusion. Similar arguments may easily be implemented for multidimensional exponential sums, and it is also quite feasible to consider higher moments than the second.

Acknowledgements This paper emerged as part of a group discussion at the Heilbronn Focused Research Workshop on Decoupling and Efficient Congruencing, University of Bristol, 17th–21st June 2019. The authors are very grateful to the Heilbronn Institute for Mathematical Research, as well as the European Union's Horizon 2020 research and innovation programme via grant agreement No. 695223, for funding this workshop, and to the University of Bristol for its hospitality. The work was further facilitated by a visit of three of the authors to Oberwolfach in November 2019. In addition, JB was supported by Starting Grant no. 2017-05110 of the Swedish Science Foundation (Vetenskapsrådet), CP was supported by a studentship sponsored by a European Research Council Advanced Grant under the European Union's Horizon 2020 research and innovation programme via grant agreement No. 695223, and GS was supported by Simons Investigator Grant no. 376201 in the name of Ben Green.

Funding Open access funding provided by University of Gothenburg.

References

1. Baker, R.C.: Diophantine inequalities. London Math. Soc. Monographs, New Series, vol. 1. The Clarendon Press, Oxford, pp. xii+275 (1986)
2. Berry, M.: Quantum fractals in boxes. J. Phys. A Math. Gen. **29**, 6617–6629 (1996)
3. Bourgain, J., Demeter, C., Guth, L.: Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three. Ann. Math. (2) **184**, 633–682 (2016)
4. Brüdern, J., Daemen, D.: Imperfect mimesis of Weyl sums. Int. Math. Res. Not. (IMRN) **16**, 3112–3126 (2009)
5. Brüdern, J., Robert, O.: Rational points on linear slices of diagonal hypersurfaces. Nagoya Math. J. **218**, 51–100 (2015)
6. Chen, C., Shparlinski, I.E.: New bounds of Weyl sums. Int. Math. Res. Not. (IMRN) **(to appear)**
7. Daemen, D.: The asymptotic formula for localized solutions in Waring's problem and approximations to Weyl sums. Bull. Lond. Math. Soc. **42**, 75–82 (2010)
8. Erdoğan, M.B., Shakan, G.: Fractal solutions of dispersive partial differential equations on the torus. Selecta Math. (N.S.) **25**(1), Art. 11 (2019)
9. Erdoğan, M.B., Tzirakis, N.: Dispersive Partial Differential Equations: Wellposedness and Applications, London Mathematical Society Student Texts 86. Cambridge University Press, Cambridge (2016)
10. Heath-Brown, D.R.: Bounds for the cubic Weyl sums. J. Math. Sci. **171**, 813–823 (2010)
11. Hughes, K., Wooley, T.D.: Discrete restriction for (x, x^3) and related topics. arXiv:1911.12262
12. Oskolkov, K.I., Chakhkiev, M.A.: Traces of the discrete Hilbert transform with quadratic phase (Russian). Tr. Mat. Inst. Steklova 280 (2013), Ortogonalnye Ryady, Teoriya Priblizheni i Smezhnye Voprosy, pp. 255–269 [translation in Proc. Steklov Inst. Math. **280**(1), 248–262 (2013)]
13. Schmidt, W.M.: Equations Over Finite Fields—An Elementary Approach. Lecture Notes in Mathematics, vol. 536. Springer, Berlin (1976)
14. Schmidt, W.M.: Diophantine Approximation. Lecture Notes in Mathematics, vol. 785. Springer, Berlin (1980)
15. Talbot, H.F.: Facts related to optical science, No. IV. Philos. Mag. **9**, 401–407 (1836)
16. Titchmarsh, E.C.: The Theory of the Riemann Zeta-Function, 2nd edn, Revised by D. R. Heath-Brown. Oxford University Press, Oxford (1986)
17. Vaughan, R.C.: The Hardy–Littlewood Method. Cambridge Tracts in Mathematics, vol. 125. Cambridge University Press, Cambridge (1997)

18. Vaughan, R.C.: On generating functions in additive number theory, I. In: Chen, W.W.L., Gowers, W.T., Halberstam, H., Schmidt, W.M., Vaughan, R.C. (eds.) Analytic Number Theory, Essays in Honour of Klaus Roth. Cambridge University Press, Cambridge (2009)
19. Vaughan, R.C., Wooley, T.D.: On the distribution of generating functions. Bull. Lond. Math. Soc. **30**, 113–122 (1998)
20. Wooley, T.D.: Mean value estimates for odd cubic Weyl sums. Bull. Lond. Math. Soc. **47**(6), 946–957 (2015)
21. Wooley, T.D.: Perturbations of Weyl Sums. Int. Math. Res. Not. (IMRN) **2016**(9), 2632–2646 (2016)
22. Wooley, T.D.: The cubic case of the main conjecture in Vinogradov's mean value theorem. Adv. Math. **294**, 532–561 (2016)
23. Wooley, T.D.: Nested efficient congruencing and relatives of Vinogradov's mean value theorem. Proc. Lond. Math. Soc. (3) **118**(4), 942–1016 (2019)

Affiliations

J. Brandes[1] · S. T. Parsell[2] · C. Poulias[3] · G. Shakan[4] · R. C. Vaughan[5]

✉ J. Brandes
brjulia@chalmers.se

S. T. Parsell
sparsell@wcupa.edu

C. Poulias
kp17312@bristol.ac.uk

G. Shakan
george.shakan@gmail.com

R. C. Vaughan
rcv4@psu.edu

[1] Mathematical Sciences, University of Gothenburg and Chalmers Institute of Technology, 412 96 Göteborg, Sweden

[2] Department of Mathematics, West Chester University, West Chester, PA 19383, USA

[3] School of Mathematics, University Walk, Clifton, Bristol BS8 1TW, UK

[4] Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

[5] Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA