OPTIMAL MEAN VALUE ESTIMATES BEYOND VINOGRADOV'S MEAN VALUE THEOREM

JULIA BRANDES AND TREVOR D. WOOLEY

Abstract. We establish sharp mean value estimates associated with the number of integer solutions of certain systems of diagonal equations. This is the first occasion on which bounds of this quality have been attained for Diophantine systems not of Vinogradov type. As a consequence of this progress, whenever $u \geqslant 3v$ we obtain the Hasse principle for systems consisting of v cubic and u quadratic diagonal equations in $6v + 4u + 1$ variables, thus attaining the convexity barrier for this problem.

1. Introduction

In recent years, our understanding of systems of diagonal equations and their associated mean values has advanced rapidly. Whilst only a few years ago, such mean values had been comprehensively understood only in the most basic cases, the resolution of the main conjecture associated with Vinogradov's mean value theorem by the second author [13, 14] and Bourgain, Demeter and Guth [1] has transformed the landscape. It now seems feasible to address the challenge of establishing similarly strong results for a much wider class of cognate problems.

In this memoir, we attain the convexity barrier for a family of mean values associated with systems of equations that fail to be translation-dilation invariant, and thus lie outside the scope of the efficient congruencing and ℓ^2-decoupling methods developed by the second author [13, 14] and Bourgain, Demeter and Guth [1]. The most accessible of our results addresses systems of cubic and quadratic diagonal equations. Let $\mathcal{N}_{s,v,u}(X)$ denote the number of integral solutions $\mathbf{x} \in [-X, X]^s$ of the system of equations

$$\begin{aligned} c_{i,1}^{(3)} x_1^3 + \ldots + c_{i,s}^{(3)} x_s^3 &= 0 \qquad (1 \leqslant i \leqslant v) \\ c_{j,1}^{(2)} x_1^2 + \ldots + c_{j,s}^{(2)} x_s^2 &= 0 \qquad (1 \leqslant j \leqslant u), \end{aligned} \tag{1.1}$$

consisting of u quadratic and v cubic equations of diagonal shape. Here and throughout we assume the coefficients $c_{i,j}^{(k)}$ of such systems to be integral. It is clear that the presence of coefficients in such systems necessitates some kind of non-singularity condition, lest the equations interact in some non-generic way. We

2010 *Mathematics Subject Classification.* 11L15, 11D45, 11L07, 11P55.
Key words and phrases. Exponential sums, Hardy–Littlewood method.

refer to an $r \times s$ matrix C as *highly non-singular* if $s \geqslant r$ and any collection of r distinct columns of C forms a non-singular matrix.

Our first result shows that $\mathcal{N}_{s,v,u}(X)$ satisfies the anticipated asymptotic formula for all sets of coefficients in general position, provided that $s \geqslant 6v + 4u + 1$ and $u \geqslant 3v$. This achieves the conjectured convexity barrier.

Theorem 1.1. *Suppose that $u \geqslant 3v$ and that $s \geqslant 6v + 4u + 1$. Assume further that the coefficient matrices*

$$C^{(2)} = (c_{i,j}^{(2)})_{\substack{1 \leqslant i \leqslant u \\ 1 \leqslant j \leqslant s}} \text{ and } C^{(3)} = (c_{i,j}^{(3)})_{\substack{1 \leqslant i \leqslant v \\ 1 \leqslant j \leqslant s}}$$

are highly non-singular. Then there exist constants $\mathcal{C} \geqslant 0$ and $\delta > 0$ such that

$$\mathcal{N}_{s,v,u}(X) = \mathcal{C} X^{s-3v-2u} + O(X^{s-3v-2u-\delta}). \tag{1.2}$$

Moreover, if the system (1.1) *has non-singular real and p-adic solutions for all primes p, then $\mathcal{C} > 0$.*

In general, asymptotic formulæ like the one supplied by (1.2) are expected to hold whenever the number of variables exceeds twice the total degree of the system. However, thus far the validity of such an asymptotic formula has been proved only in a few isolated instances. Arguably the first non-trivial case in which this convexity barrier was achieved occurs in work of Cook [7, 8] concerning pairs and triples of diagonal quadratic equations. Recent work of Brüdern and the second author [5, 6] obtains asymptotic lower bounds at the convexity limit for systems of diagonal cubic forms. In the case of mixed systems of cubic and quadratic equations, work of the second author underlying [12, Theorem 1.2] achieves the convexity limit in the case $u = v = 1$ with $s \geqslant 11$ relating to systems consisting of one cubic and one quadratic diagonal equation. Most recently, investigations of the first author joint with Parsell [3, Theorem 1.4] establish an asymptotic formula tantamount to (1.2) for systems of v cubic and u quadratic diagonal equations, though under the more restrictive hypothesis that $s \geqslant \lfloor 20v/3 \rfloor + 4u + 1$, thus missing the convexity barrier whenever $v \geqslant 2$. In subsequent work [2], the first author proved that an asymptotic formula of the shape (1.2) holds when $v \geqslant 2u$ and $s \geqslant 6v + \lfloor 14u/3 \rfloor + 1$, which misses the convexity barrier when $u \geqslant 2$. Thus, Theorem 1.1 provides the first instance where bounds of the expected quality have been achieved for systems of v cubic and u quadratic equations in settings where both u and v exceed 1.

Theorem 1.1 is in fact a special case of our more general Theorem 1.6 below. Both of these results rest on our new estimates for certain mean values of Vinogradov type. In the most general form, such mean values encode the number of integral solutions of systems of the general shape

$$c_{i,1}^{(l)}(x_1^l - y_1^l) + \ldots + c_{i,s}^{(l)}(x_s^l - y_s^l) = 0 \qquad (1 \leqslant i \leqslant r_l,\ 1 \leqslant l \leqslant k), \tag{1.3}$$

in which $r_1, \ldots, r_k$ are non-negative integers and the coefficients $c_{i,j}^{(l)}$ are integers. When all of the coefficient matrices $C^{(l)} = (c_{i,j}^{(l)})_{\substack{1\leqslant i\leqslant r_l\\ 1\leqslant j\leqslant s}}$ are highly non-singular, then the main conjecture states that the number of integral solutions $\mathbf{x}, \mathbf{y} \in [-X, X]^s$ of the system (1.3) should be at most of order $X^{s+\varepsilon} + X^{2s-K}$, for any $\varepsilon > 0$, where $K = r_1 + 2r_2 + \ldots + kr_k$ denotes the total degree of the system (1.3). A corresponding lower bound, with $\varepsilon = 0$, is provided by an argument akin to that delivering [11, equation (7.4)]. Systems of this kind have previously been studied by the first author together with Parsell [3], where it was shown that the main conjecture for such systems holds when $r_l \geqslant r_{l+1}$ for all $1 \leqslant l \leqslant k-1$, in which case the system (1.3) can be viewed as a superposition of Vinogradov systems of various degrees (see Theorem 2.1 and Corollary 2.2 in that paper). However, bounds of this strength were known hitherto only for systems of quadratic equations and systems of Vinogradov type, as well as superpositions of these two special classes of systems.

The goal of the work at hand is to enlarge the range of systems of type (1.3) for which the main conjecture is known to hold. What we have in mind is a superposition of systems of the shape

$$\sum_{j=1}^{s} b_j^{(l)}(x_j^l - y_j^l) = 0 \qquad (2 \leqslant l \leqslant k),$$

which can be viewed as Vinogradov systems missing the linear slice. We take $k_1 \geqslant k_2 \geqslant \ldots \geqslant k_u$ to be the degrees of these systems, with $k_u = 2$, and we put $k_1 = k$. The number of equations of degree l, with $2 \leqslant l \leqslant k$, is given by

$$r_l = \mathrm{card}\{j : k_j \geqslant l\}.$$

For convenience, we put $v = r_3$, and we note that $r_2 = u$. Thus, we are considering a superposition of the aforementioned modified Vinogradov systems of respective degrees $k_1, \ldots, k_v$, together with $u - v$ additional quadratic equations. We write

$$r = r_2 + \ldots + r_k$$

for the total number of equations. When the coefficient matrices $C^{(l)}$ are highly non-singular for $2 \leqslant l \leqslant k$, we denote by

$$I_{s,\mathbf{k},u}(X) = I_{s,\mathbf{k},u}(X; C^{(2)}, \ldots, C^{(k)}) \tag{1.4}$$

the number of integral solutions $\mathbf{x}, \mathbf{y} \in [-X, X]^s$ of the system (1.3). Here and elsewhere, we think of $\mathbf{k}$ as denoting $(k_1, \ldots, k_v)$. Write further

$$\kappa = \sum_{j=1}^{v} \left(\frac{k_j(k_j+1)}{2} - 3 \right) \quad \text{and} \quad K = \kappa + 2u, \tag{1.5}$$

so that K denotes the total degree of the system. The most general formulation of our result is as follows.

Theorem 1.2. *Suppose that $u \geqslant 2v$ and $u|K$. Then for any integer $s \geqslant u$ and any $\varepsilon > 0$ we have*

$$I_{s,\mathbf{k},u}(X) \ll X^{\varepsilon}(X^s + X^{2s-K}). \tag{1.6}$$

As announced above, Theorem 1.2 delivers an estimate tantamount to the main conjecture for a sizeable family of systems of diagonal equations not of Vinogradov type. We remark further that the choice $u = \kappa$ is permissible for any v-tuple $\mathbf{k}$ with $k_j \geqslant 3$ $(1 \leqslant j \leqslant v)$, so Theorem 1.2 has content for any configuration of degrees $\mathbf{k}$ described in the preamble to the statement of the theorem.

The interested reader may inquire what bounds can be obtained in Theorem 1.2 when the condition $u|K$ is violated. It follows from the arguments presented in the first author's work with Parsell [3, Theorem 2.1] in combination with Hua's inequality for quadratic exponential sums that, should the bound (1.6) be known for some integer $u_0 \geqslant 2v$ dividing κ, it continues to hold for all $u \geqslant u_0$. We summarise this observation in the form of a corollary.

Corollary 1.3. *Suppose that $u_0 \geqslant 2v$ with $u_0|\kappa$. Then for any integer $u \geqslant u_0$, any integer $s \geqslant u$ and any $\varepsilon > 0$ we have*

$$I_{s,\mathbf{k},u}(X) \ll X^{\varepsilon}(X^s + X^{2s-K}).$$

If κ does not have any suitably small divisors, we no longer achieve the convexity barrier for u not dividing K. In such circumstances, however, our methods are still apt to establish the bound (1.6) in the ranges

$$u \leqslant s \leqslant u\lfloor K/u \rfloor = K - w \quad \text{and} \quad s \geqslant u\lceil K/u \rceil = K + u - w, \tag{1.7}$$

where we write w for the remainder of K on division by u. We outline the modifications necessary in our argument to treat these cases at the end of §2. Our results complement older ones that can be obtained by other means. Specifically, it follows from Theorem 2.1 and Corollary 2.2 of the first author's work with Parsell [3] that (1.6) holds in the range

$$s \geqslant \sum_{j=1}^{v} \frac{k_j(k_j+1)}{2} + 2(u-v) = K + v, \tag{1.8}$$

whereas for small s the second author's result [14, Corollary 1.2] can be combined with the arguments of [3, Theorem 2.1] to show that (1.6) holds in the range

$$u \leqslant s \leqslant \sum_{j=1}^{v} \frac{k_j(k_j-1)}{2} + 2(u-v) = K + v - \sum_{j=1}^{v} k_j. \tag{1.9}$$

In either case, it depends on the residue class of K modulo u which of the bounds prevails. We note, however, that in the most interesting cases when $u \geqslant 2v$ is relatively small and in particular smaller than $\sum_{j=1}^{v} k_j$, the exceptional range in (1.7) is shorter than the one that is obtained upon combining (1.8) and (1.9).

To further illustrate the strength of our result in Theorem 1.2, we discuss in more detail some of the most relevant special cases. Among the systems of diagonal equations not of Vinogradov type, the most well-studied ones are systems of cubic equations and systems of cubic and quadratic equations, such as we considered in our motivating example in Theorem 1.1. Regarding such systems, it is immediate from work of the second author [12, Theorem 1.1] that for every $\varepsilon > 0$ one has $I_{5,3,1}(X) \ll X^{31/6+\varepsilon}$, and this bound implies via [3, Theorem 2.1] that $I_{3+2u,3,u}(X) \ll X^{3+2u+1/6+\varepsilon}$ for all $u \geqslant 1$. Theorem 1.2 now allows us to improve this result whenever $u \geqslant 3$.

Corollary 1.4. *Suppose that $s \geqslant u \geqslant 3$. Then for any $\varepsilon > 0$ we have*

$$I_{s,3,u}(X) \ll X^{\varepsilon}(X^s + X^{2s-3-2u}).$$

This follows from Corollary 1.3 by specialising $k = 3$ and $v = 1$, so that $u_0 = 3$ is permissible. Corollary 1.4 is only the second time, after the second author's successful treatment of the cubic case of Vinogradov's mean value theorem [13], that the convexity barrier has been attained for a system of diagonal equations in general position involving cubic equations. In particular, one now has the main conjecture for mean values that correspond to systems consisting of one cubic and three quadratic diagonal equations. One should regard the latter as providing the main new input that enables us to prove Theorem 1.1.

Our second special case concerns systems of higher degree k. In a recent paper [4], we studied Vinogradov systems lacking the linear slice and established diagonal behaviour for the mean value $I_{s,k,1}(X)$ for $s \leqslant (k^2-1)/2$, thus missing the critical point $s = k(k+1)/2 - 1$ only by a term linear in k. It turns out, however, that under suitable congruence conditions the full main conjecture is true when one adds in one additional quadratic equation.

Corollary 1.5. *Suppose that $k \equiv 1$ or $2 \pmod 4$. Then for any integer $s \geqslant 2$ and any $\varepsilon > 0$ we have the bound*

$$I_{s,k,2}(X) \ll X^{\varepsilon}(X^s + X^{2s-(k^2+k+2)/2}).$$

This follows easily from Theorem 1.2 upon noting that $\frac{1}{2}k(k+1) - 3$ is even under the stated congruence requirements.

Mean value estimates like that of Theorem 1.2 have long been employed to establish asymptotic formulæ for the number of solutions of simultaneous diagonal equations. For $\mathbf{r}$ as in the preamble to Theorem 1.2 and highly non-singular coefficient matrices $C^{(l)}$ for $2 \leqslant l \leqslant k$, denote by $N_{s,\mathbf{k},u}(X)$ the number of integral solutions of the system of equations

$$c_{i,1}^{(l)} x_1^l + \ldots + c_{i,s}^{(l)} x_s^l = 0 \qquad (1 \leqslant i \leqslant r_l,\ 2 \leqslant l \leqslant k), \tag{1.10}$$

with $|x_j| \leqslant X$ $(1 \leqslant j \leqslant s)$. It is well known that, if s is sufficiently large in terms of $\mathbf{k}$ and u, there is an asymptotic formula of the shape

$$N_{s,\mathbf{k},u}(X) = (\mathcal{C} + o(1))X^{s-K}, \tag{1.11}$$

where $\mathcal{C}$ is a non-negative constant encoding the local solubility data for the system (1.10). The relevant question is how large s has to be for an asymptotic formula like that of (1.11) to hold. Theorem 1.1 of [3] provides a bound for s that is somewhat unwieldy, but can likely be reduced to

$$s \geqslant \sum_{i=1}^{v} k_i(k_i + 1) + 4(u - v) + 1 = 2K + 2v + 1$$

by accounting for our revised treatment of the major arcs described in §4 below. On the other hand, unless fundamentally new methods become available that avoid the use of mean values, we cannot expect to be able to establish asymptotic estimates when $s \leqslant 2K$. Thanks to our new mean value estimate in Theorem 1.2, we are now able to attain this theoretical barrier.

Theorem 1.6. *Suppose that $u \geqslant u_0 \geqslant 2v$ with $u_0|\kappa$. Then, provided that $s \geqslant 2K + 1$, the asymptotic formula* (1.11) *holds with $\mathcal{C} \geqslant 0$. If, furthermore, the system* (1.10) *has non-singular solutions in $\mathbb{R}$ as well as in the fields $\mathbb{Q}_p$ for all p, then the constant $\mathcal{C}$ is positive.*

Observe that Theorem 1.1 is a special case of Theorem 1.6.

The proofs of our results rest on an idea that played a crucial role in the second author's work on pairs of quadratic and cubic diagonal equations [12], and which has been explored further in the authors' recent work on incomplete Vinogradov systems [4]. In these papers, the missing linear equation is artificially added in, which makes it possible to exploit the strong bounds on Vinogradov's mean value theorem. By taking advantage of the translation-dilation invariance of the newly completed Vinogradov systems, we then generate an auxiliary mean value related to that of Vinogradov. Whilst our understanding of these auxiliary mean values remains unsatisfactory for general degree k, in the case $k = 2$ they may be comprehensively understood in terms of quadratic Vinogradov systems. This observation plays a pivotal role in our argument.

Throughout, the letters s, u, v, and k, as well as the entries of the vectors $\mathbf{k}$ and $\mathbf{r}$, will denote positive integers with $u \geqslant 2v$. The letter ε will be used to denote an arbitrary, but sufficiently small positive number, and we adopt the convention that whenever it appears in a statement, we assert that the statement holds for all sufficiently small $\varepsilon > 0$. We take X to be a large positive number which, just like the implicit constants in the notations of Landau and Vinogradov, is permitted to depend at most on s, $\mathbf{k}$, u, the coefficient matrices $C^{(l)}$ $(2 \leqslant l \leqslant k)$, and ε. We employ the non-standard notation that when $G : [0,1)^n \to \mathbb{C}$ is integrable for some $n \in \mathbb{N}$, then

$$\oint G(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha} = \int_{[0,1)^n} G(\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha}.$$

Here and elsewhere, we use vector notation liberally in a manner that is easily discerned from the context. In particular, when $\mathbf{b}$ denotes the integer tuple $(b_1, \ldots, b_n)$, we write $(q, \mathbf{b}) = \gcd(q, b_1, \ldots, b_n)$.

Acknowledgements. Both authors thank the Fields Institute in Toronto for excellent working conditions and support that made this work possible during the Thematic Program on Unlikely Intersections, Heights, and Efficient Congruencing. This work was further facilitated by subsequent visits of the first author to the University of Bristol, and of the second author to the University of Waterloo. The authors gratefully acknowledge the hospitality of both institutions. The work of both authors was supported by the National Science Foundation under Grant No. DMS-1440140 while they were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester. The first author's work was supported in part by Starting Grant 2017-05110 from Vetenskapsrådet. The second author's work was supported by a European Research Council Advanced Grant under the European Union's Horizon 2020 research and innovation programme via grant agreement No. 695223.

2. The underlying mean value

Our goal in this section is the proof of Theorem 1.2. Before delving into the core of the argument, we pause to introduce some notation and establish a mean value estimate that will be of use in our subsequent discussion. For $2 \leqslant l \leqslant k$ we define the exponential sum $K_l(\boldsymbol{\alpha}; X, H)$ by putting

$$K_l(\boldsymbol{\alpha}; X, H) = \sum_{|h| \leqslant H} \sum_{|z| \leqslant X} e(h\alpha^{(1)} + 2hz\alpha^{(2)} + \ldots + lhz^{l-1}\alpha^{(l)}), \tag{2.1}$$

and we write

$$f_l(\boldsymbol{\alpha}; X) = \sum_{|x| \leqslant X} e(\alpha^{(1)}x + \alpha^{(2)}x^2 + \ldots + \alpha^{(l)}x^l).$$

Then, with the standard notation associated with Vinogradov's mean value theorem in mind, we put

$$J_{s,l}(X) = \oint |f_l(\boldsymbol{\alpha}; X)|^{2s}\,\mathrm{d}\boldsymbol{\alpha}.$$

We note that the main conjecture associated with Vinogradov's mean value theorem is now known to hold for all degrees. This is classical when $l = 2$, it is a consequence of work of the second author [13] for degree $l = 3$, and for degrees exceeding three it follows from the work of Bourgain, Demeter and Guth, and of the second author (see [1, Theorem 1.1] and [14, Corollary 1.3]). Thus, for all $\sigma > 0$ one has

$$J_{\sigma,l}(X) \ll X^{\varepsilon}(X^{\sigma} + X^{2\sigma - l(l+1)/2}). \tag{2.2}$$

For future reference, we record the trivial inequality

$$|a_1 \cdots a_n| \leqslant |a_1|^n + \ldots + |a_n|^n, \tag{2.3}$$

which is valid for all $a_1, \ldots, a_n \in \mathbb{C}$.

Lemma 2.1. *Let H and X be large real numbers. Then one has*

$$\oint |f_2(\boldsymbol{\alpha}; X) K_2(\boldsymbol{\alpha}; X, H)|^2 \, \mathrm{d}\boldsymbol{\alpha} \ll X^2 (X + H)^{1+\varepsilon}. \tag{2.4}$$

Proof. Upon considering the underlying system of equations, we see that the mean value on the left hand side of (2.4) is given by the number of integer solutions of the system of equations

$$\begin{aligned} x_1^2 - x_2^2 &= 2(h_1 z_1 - h_2 z_2) \\ x_1 - x_2 &= h_1 - h_2, \end{aligned} \tag{2.5}$$

with $|h_i| \leqslant H$ and $|z_i|, |x_i| \leqslant X$ for $i = 1, 2$. The second of these equations permits the substitution $h_2 = h_1 - x_1 + x_2$ into the first, whence

$$(x_1 - x_2)(x_1 + x_2 - 2z_2) = 2h_1(z_1 - z_2).$$

Suppose first that $h_1(z_1 - z_2)$ is non-zero. Then for each of the $O(HX^2)$ possible choices for h_1, z_1 and z_2 fixing the latter integer in such a manner, an elementary divisor function estimate shows there to be $O((HX)^{\varepsilon})$ possible choices for the integers $x_1 - x_2$ and $x_1 + x_2 - 2z_2$, and hence also for x_1 and x_2. These choices also fix $h_2 = h_1 - x_1 + x_2$, so we see that there are $O(H^{1+\varepsilon} X^{2+\varepsilon})$ solutions of this first type. Meanwhile, if $h_1(z_1 - z_2) = 0$, then $h_1 = 0$ or $z_1 = z_2$, and at the same time either $x_1 = x_2$ or $x_1 = 2z_2 - x_2$. In any case, therefore, each of the $O(X^2)$ possible choices for z_2 and x_2 determine x_1 and either h_1 or z_1. Since there are $O(X + H)$ possible choices left by this constraint for the latter, and h_2 is again fixed by these choices just as before, we find that there are $O(X^2(X + H))$ solutions of this second type. The conclusion of the lemma follows by summing the contributions from both types of solutions. □

We remark that the system (2.5) can be interpreted as being of Vinogradov shape of degree two by means of the substitution $h_i = u_i - v_i$ and $z_i = u_i + v_i$ for $i = 1, 2$. Viewed in this way, Lemma 2.1 amounts to no more than a rephrasing of the classical elementary proof of the quadratic case in Vinogradov's mean value theorem.

We now initiate the proof of Theorem 1.2, assuming the hypotheses of its statement. For $l \geqslant 2$ let

$$g_l(\boldsymbol{\alpha}; X) = \sum_{|x| \leqslant X} e(\alpha^{(2)} x^2 + \alpha^{(3)} x^3 + \ldots + \alpha^{(l)} x^l),$$

define $\boldsymbol{\alpha}^{(l)} = (\alpha_i^{(l)})_{1 \leqslant i \leqslant r_l}$ $(2 \leqslant l \leqslant k)$, and put

$$\gamma_j^{(l)} = \sum_{i=1}^{r_l} c_{i,j}^{(l)} \alpha_i^{(l)} \qquad (1 \leqslant j \leqslant s). \tag{2.6}$$

Also, set $\boldsymbol{\gamma}_j = (\gamma_j^{(l)})_{2 \leqslant l \leqslant k}$ for $1 \leqslant j \leqslant s$ and $\boldsymbol{\gamma}^{(l)} = (\gamma_j^{(l)})_{1 \leqslant j \leqslant s}$ for $2 \leqslant l \leqslant k$, and put $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_s) = (\boldsymbol{\gamma}^{(2)}, \ldots, \boldsymbol{\gamma}^{(k)})^T$. Then by orthogonality we have

$$I_{s,\mathbf{k},u}(X) = \oint \prod_{j=1}^{s} |g_k(\boldsymbol{\gamma}_j; X)|^2 \, \mathrm{d}\boldsymbol{\alpha}.$$

Set $t = K/u$, and observe that under the hypotheses of the theorem, this is an integer. Further, let $\mathcal{I}$ denote the set of all integral u-tuples $(j_1, \ldots, j_u)$ with $1 \leqslant j_1 < j_2 < \ldots < j_u \leqslant s$, and put

$$G_{\mathbf{k},u}(X) = \max_{\mathbf{j} \in \mathcal{I}} \oint \prod_{m=1}^{u} |g_k(\boldsymbol{\gamma}_{j_m}; X)|^{2t} \, \mathrm{d}\boldsymbol{\alpha}. \tag{2.7}$$

We can bound $I_{s,\mathbf{k},u}(X)$ in terms of $G_{\mathbf{k},u}(X)$.

Lemma 2.2. *We have the bound*

$$I_{s,\mathbf{k},u}(X) \ll \begin{cases} (G_{\mathbf{k},u}(X))^{s/K} & (u \leqslant s \leqslant K), \\ X^{2s-2K} G_{\mathbf{k},u}(X) & (s > K). \end{cases}$$

Proof. When $s > K$, the trivial bound $g_k(\boldsymbol{\gamma}_j; X) = O(X)$ delivers the estimate

$$I_{s,\mathbf{k},u}(X) \ll X^{2s-2K} \oint \prod_{j=1}^{K} |g_k(\boldsymbol{\gamma}_j; X)|^2 \, \mathrm{d}\boldsymbol{\alpha},$$

and the claim for this case follows from (2.3). Suppose now that $u \leqslant s \leqslant K$. Then by applying the bound (2.3) again, we find that

$$\begin{aligned} I_{s,\mathbf{k},u}(X) &\ll \max_{\mathbf{j} \in \mathcal{I}} \oint \prod_{m=1}^{u} |g_k(\boldsymbol{\gamma}_{j_m}; X)|^{2s/u} \, \mathrm{d}\boldsymbol{\alpha} \\ &\ll \left(\max_{\mathbf{j} \in \mathcal{I}} \oint \prod_{m=1}^{u} |g_k(\boldsymbol{\gamma}_{j_m}; X)|^{2t} \, \mathrm{d}\boldsymbol{\alpha} \right)^{s/K}, \end{aligned}$$

where in the last step we used Hölder's inequality. Thus the lemma is established in both cases. $\square$

Suppose that the maximum in (2.7) is assumed at some tuple $\mathbf{j} \in \mathcal{I}$, which we consider fixed for the remainder of the analysis. We then relabel $d_{i,m}^{(l)} = c_{i,j_m}^{(l)}$ for $1 \leqslant m \leqslant u$, $1 \leqslant i \leqslant r_l$, and $2 \leqslant l \leqslant k$, and we define the coefficient matrices

$$D^{(l)} = (d_{i,m}^{(l)})_{\substack{1 \leqslant i \leqslant r_l \\ 1 \leqslant m \leqslant u}}.$$

In addition, we define $D^{(1)}$ to be the $u \times u$ identity matrix. In this section, we henceforth adopt the convention that $r_1 = u$. We then define $\delta_i^{(l)}$ via the relations $\boldsymbol{\delta}^{(l)} = (D^{(l)})^T \boldsymbol{\alpha}^{(l)}$ for $1 \leqslant l \leqslant k$, so that $\delta_i^{(l)} = \gamma_{j_i}^{(l)}$ for $1 \leqslant i \leqslant u$ and $2 \leqslant l \leqslant k$, and we put $\boldsymbol{\delta}_m = (\delta_m^{(l)})_{1 \leqslant l \leqslant k}$ for $1 \leqslant m \leqslant u$. Here, we have in mind notational conventions analogous to those in the sequel to (2.6).

Next, we define

$$H_{\mathbf{k},u}(X) = \oint \prod_{m=1}^{u} |f_k(\boldsymbol{\delta}_m; 2X)|^{2t} K_k(-\boldsymbol{\delta}_m; X, 2tX) \, \mathrm{d}\boldsymbol{\alpha}. \tag{2.8}$$

We can bound $G_{\mathbf{k},u}(X)$, and hence $I_{s,\mathbf{k},u}(X)$, in terms of $H_{\mathbf{k},u}(X)$.

Lemma 2.3. *One has $G_{\mathbf{k},u}(X) \ll X^{-u} H_{\mathbf{k},u}(X)$.*

Proof. Define ω_l to be 1 when $l = 1$, and 0 otherwise. We decompose the set $\{1, \ldots, K\}$ into the blocks $\mathcal{B}_m = \{(m-1)t + 1, \ldots, mt\}$ for $1 \leqslant m \leqslant u$. We start by noting that the mean value $G_{\mathbf{k},u}(X)$ counts the number of integral solutions of the system of equations

$$\sum_{m=1}^{u} d_{j,m}^{(l)} \xi_m^{(l)} = \omega_l h_j \qquad (1 \leqslant j \leqslant r_l,\ 1 \leqslant l \leqslant k), \tag{2.9}$$

where

$$\xi_m^{(l)} = \sum_{i \in \mathcal{B}_m} (x_i^l - y_i^l) \qquad (1 \leqslant m \leqslant u,\ 1 \leqslant l \leqslant k),$$

with $-X \leqslant x_i, y_i \leqslant X$ $(1 \leqslant i \leqslant K)$ and $|h_j| \leqslant 2tX$ $(1 \leqslant j \leqslant u)$. Note that the constraints on the $\xi_j^{(1)}$ $(1 \leqslant j \leqslant u)$ imposed by the linear equations in (2.9) are void, since the ranges for the new variables h_j automatically accommodate all possible values for the expressions $\xi_j^{(1)}$ within (2.9).

Before embarking into the core of the argument, we pause to make a small simplification of the system (2.9). Consider the subsystem of (2.9) associated with $l = 2$. By our initial assumptions, the $u \times s$ matrix $C^{(2)}$ is highly non-singular, so it follows that the $u \times u$ matrix $D^{(2)}$ is non-singular and can be diagonalised by elementary row operations. Consequently, by taking suitable linear combinations of the quadratic equations, we may assume without loss of generality that the coefficients $d_{j,m}^{(2)}$ in (2.9) vanish except when $j = m$.

We now consider the effect of shifting every variable with index in a given block $\mathcal{B}_m$ by an integer z_m with $|z_m| \leqslant X$. By the binomial theorem, for any family of shifts $\mathbf{z}$, one finds that $(\mathbf{x}, \mathbf{y})$ is a solution of (2.9) if and only if it is also a solution of the system

$$\sum_{m=1}^{u} d_{j,m}^{(l)} \zeta_m^{(l)} = \sum_{m=1}^{u} d_{j,m}^{(l)} l h_m z_m^{l-1} \qquad (1 \leqslant j \leqslant r_l,\ 1 \leqslant l \leqslant k),$$

where

$$\zeta_m^{(l)} = \sum_{i \in \mathcal{B}_m} \big((x_i + z_m)^l - (y_i + z_m)^l \big) \qquad (1 \leqslant m \leqslant u,\ 1 \leqslant l \leqslant k).$$

Thus, for each fixed integer u-tuple $\mathbf{z}$ with $|z_m| \leqslant X$ $(1 \leqslant m \leqslant u)$, the mean value $G_{\mathbf{k},u}(X)$ is bounded above by the number of integral solutions of the system

$$\sum_{m=1}^{u} d_{j,m}^{(l)} \Big(\sum_{i \in \mathcal{B}_m} (v_i^l - w_i^l) \Big) = \sum_{m=1}^{u} d_{j,m}^{(l)} l h_m z_m^{l-1} \qquad (1 \leqslant j \leqslant r_l,\ 1 \leqslant l \leqslant k),$$

with $|\mathbf{v}|, |\mathbf{w}| \leqslant 2X$ and $|\mathbf{h}| \leqslant 2tX$. On applying orthogonality and averaging over all possible choices for $\mathbf{z}$, we therefore infer that

$$G_{\mathbf{k},u}(X) \ll X^{-u} \sum_{|\mathbf{z}| \leqslant X} \oint \prod_{m=1}^{u} |f_k(\boldsymbol{\delta}_m; 2X)|^{2t} \mathfrak{k}(-\boldsymbol{\delta}_m; z_m)\, \mathrm{d}\boldsymbol{\alpha},$$

where

$$\mathfrak{k}(\boldsymbol{\alpha}; z) = \sum_{|h| \leqslant 2tX} e(h\alpha^{(1)} + 2hz\alpha^{(2)} + \ldots + khz^{k-1}\alpha^{(k)}).$$

The proof of the lemma is completed by reference to (2.1) and (2.8). □

We can now turn to the task of estimating $H_{\mathbf{k},u}(X)$.

Lemma 2.4. *We have* $H_{\mathbf{k},u}(X) \ll X^{K+u+\varepsilon}$.

Proof. Let $\sigma \in \mathcal{S}_u$ be given by $(1, 2, \ldots, u)$, where $\mathcal{S}_u$ denotes the group of permutations on u elements, and for $1 \leqslant n \leqslant u$ set

$$\mathfrak{G}_{1,n}(\boldsymbol{\delta}) = \prod_{m=1}^{v} |f_k(\boldsymbol{\delta}_{\sigma^n(m)}; 2X)|^{k_m(k_m+1)}. \tag{2.10}$$

Also, observe that the relations (1.5) together with our definition $t = K/u$ imply that

$$2tu - \sum_{m=1}^{v} k_m(k_m+1) = 2tu - (2K + 6v - 4u) = 4u - 6v.$$

Thus, if we set

$$\mathfrak{G}_{2,n}(\boldsymbol{\delta})=\prod_{m=v+1}^{u}|f_k(\boldsymbol{\delta}_{\sigma^n(m)};2X)|^{\frac{4u-6v}{u-v}}|K_k(-\boldsymbol{\delta}_{\sigma^n(m)};X,2tX)|^{\frac{u}{u-v}},$$

then it follows from Hölder's inequality in conjunction with (2.3) and (2.8) that

$$H_{\mathbf{k},u}(X)\ll\prod_{n=1}^{u}\left(\oint\mathfrak{G}_{1,n}(\boldsymbol{\delta})\mathfrak{G}_{2,n}(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\alpha}\right)^{1/u}\ll\max_{1\leqslant n\leqslant u}\oint\mathfrak{G}_{1,n}(\boldsymbol{\delta})\mathfrak{G}_{2,n}(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\alpha}.$$

By relabelling indices here, we may assume without loss of generality that the maximum is taken at $n=u$, so that $\sigma^u=\mathrm{id}\in\mathcal{S}_u$. Thus we obtain

$$H_{\mathbf{k},u}(X)\ll\oint\mathfrak{G}_1(\boldsymbol{\delta})\mathfrak{G}_2(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\alpha},\tag{2.11}$$

where

$$\mathfrak{G}_1(\boldsymbol{\delta})=\prod_{m=1}^{v}|f_k(\boldsymbol{\delta}_m;2X)|^{k_m(k_m+1)}\tag{2.12}$$

and

$$\mathfrak{G}_2(\boldsymbol{\delta})=\prod_{m=v+1}^{u}|f_k(\boldsymbol{\delta}_m;2X)|^{\frac{4u-6v}{u-v}}|K_k(-\boldsymbol{\delta}_m;X,2tX)|^{\frac{u}{u-v}}.$$

Recall now that we had arranged for the coefficient matrices $D^{(1)}$ and $D^{(2)}$ to be diagonal. Consequently, the variables $\boldsymbol{\delta}_m$ with $1\leqslant m\leqslant v$ are independent of those $\alpha_m^{(l)}$ having $l\in\{1,2\}$ and $v+1\leqslant m\leqslant u$. Similarly, since for $l\geqslant 3$ the matrices $D^{(l)}$ are of format $r_l\times u$ with $r_l\leqslant v$, the first r_l entries of each vector $\boldsymbol{\delta}^{(l)}$ suffice to uniquely recover the corresponding vector $\boldsymbol{\alpha}^{(l)}$, and thus in turn completely determine those entries $\delta_m^{(l)}$ having $m>r_l$. Set $\boldsymbol{\eta}_1=(\boldsymbol{\delta}_m)_{1\leqslant m\leqslant v}$ and $\boldsymbol{\eta}_2=(\alpha_m^{(1)},\alpha_m^{(2)})_{v+1\leqslant m\leqslant u}$, then it follows from our discussion that $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ completely determine all entries of $\boldsymbol{\delta}$. With this notation, the quantity $\mathfrak{G}_1$ depends only on $\boldsymbol{\eta}_1$. Keeping in mind the implicit linear relations between the entries of $\boldsymbol{\eta}_1$ that arise from the change of variables (2.6), we can thus, with some abuse of notation, rewrite the integral on the right hand side of (2.11) as

$$\oint\mathfrak{G}_1(\boldsymbol{\delta})\mathfrak{G}_2(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\alpha}=\oint\mathfrak{G}_1(\boldsymbol{\eta}_1)\mathfrak{H}(\boldsymbol{\eta}_1)\,\mathrm{d}\boldsymbol{\eta}_1,\tag{2.13}$$

where

$$\mathfrak{H}(\boldsymbol{\eta}_1)=\oint\mathfrak{G}_2(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)\,\mathrm{d}\boldsymbol{\eta}_2.$$

Since $u\geqslant 2v$, it now follows from an application of Hölder's inequality that

$$\mathfrak{H}(\boldsymbol{\eta}_1)\ll U_1(\boldsymbol{\eta}_1)^{\frac{u-2v}{2u-2v}}U_2(\boldsymbol{\eta}_1)^{\frac{u}{2u-2v}},\tag{2.14}$$

where

$$U_1(\boldsymbol{\eta}_1) = \oint \prod_{m=v+1}^{u} |f_k(\boldsymbol{\delta}_m; 2X)|^6 \, \mathrm{d}\boldsymbol{\eta}_2$$

and

$$U_2(\boldsymbol{\eta}_1) = \oint \prod_{m=v+1}^{u} |f_k(\boldsymbol{\delta}_m; 2X) K_k(-\boldsymbol{\delta}_m; X, 2tX)|^2 \, \mathrm{d}\boldsymbol{\eta}_2.$$

Upon considering the underlying system of equations, we see that U_1 and U_2 count solutions to the associated systems of linear and quadratic equations, where each solution is counted with a unimodular weight depending on $\boldsymbol{\eta}_1$. It thus follows from the triangle inequality that

$$U_i(\boldsymbol{\eta}_1) \leqslant U_i(\mathbf{0}) \quad \text{for } i = 1, 2. \tag{2.15}$$

Using the fact that the coefficient matrices $D^{(1)}$ and $D^{(2)}$ are both diagonal, upon recalling (2.2) we discern further that

$$U_1(\mathbf{0}) \ll \oint \prod_{m=v+1}^{u} |f_2(\delta_m^{(1)}, \delta_m^{(2)}; 2X)|^6 \, \mathrm{d}\boldsymbol{\eta}_2 \ll (J_{3,2}(2X))^{u-v} \ll X^{3(u-v)+\varepsilon}. \tag{2.16}$$

By an analogous chain of reasoning, we derive from Lemma 2.1 the corresponding bound

$$\begin{aligned} U_2(\mathbf{0}) &\ll \oint \prod_{m=v+1}^{u} |f_2(\delta_m^{(1)}, \delta_m^{(2)}; 2X) K_2(-(\delta_m^{(1)}, \delta_m^{(2)}); X, 2tX)|^2 \, \mathrm{d}\boldsymbol{\eta}_2 \\ &\ll \left(\oint |f_2(\boldsymbol{\alpha}; 2X) K_2(\boldsymbol{\alpha}; X, 2tX)|^2 \, \mathrm{d}\boldsymbol{\alpha} \right)^{u-v} \ll X^{3(u-v)+\varepsilon}. \end{aligned} \tag{2.17}$$

Thus, from (2.14)–(2.17) we have $\mathfrak{H}(\boldsymbol{\eta}_1) \ll X^{3(u-v)+\varepsilon}$, and hence (2.11) and (2.13) imply that

$$H_{\mathbf{k},u}(X) \ll X^{3(u-v)+\varepsilon} \oint \mathfrak{G}_1(\boldsymbol{\eta}_1) \, \mathrm{d}\boldsymbol{\eta}_1.$$

On recalling (2.12), the integral on the right hand side is easily bounded by reference to the first author's work with Parsell (Theorem 2.1 and Corollary 2.2 in [3]), where it was shown that

$$\oint \mathfrak{G}_1(\boldsymbol{\eta}_1) \, \mathrm{d}\boldsymbol{\eta}_1 \ll \prod_{m=1}^{v} J_{k_m(k_m+1)/2, k_m}(2X) \ll X^{\kappa+3v+\varepsilon}. \tag{2.18}$$

Here, we applied (2.2), and in the last step we made use of the notation (1.5). Altogether, we obtain

$$H_{\mathbf{k},u}(X) \ll X^{\kappa+3v+\varepsilon} X^{3(u-v)+\varepsilon} \ll X^{K+u+2\varepsilon},$$

as claimed. □

The proof of Theorem 1.2 is now completed by combining Lemmata 2.2, 2.3 and 2.4.

Before concluding the section, we briefly detail the changes to the argument that are necessary in order to establish the bound (1.6) subject to (1.7) when the condition $u|K$ is violated. From the proof of Lemma 2.2 we see that it is sufficient to consider the cases $s = s_0$ where $s_0 = u\lfloor K/u\rfloor$ or $s_0 = u\lceil K/u\rceil$, respectively.

When $s_0 = u\lceil K/u\rceil$, we set $\tau = s_0 - K$, and then define $\boldsymbol{\tau}$ by putting $\tau_1 = \tau$ and $\tau_m = 0$ $(2 \leqslant m \leqslant v)$. In (2.10) and (2.12), we replace the exponents $k_m(k_m+1)$ with $2\sigma_m$, where $\sigma_m = k_m(k_m+1)/2 + \tau_m$, and we leave the definitions of $\mathfrak{G}_{2,n}(\boldsymbol{\delta})$ and $\mathfrak{G}_2(\boldsymbol{\delta})$ unchanged. Then the argument applied above proceeds identically up to (2.18), where instead we obtain

$$\oint \mathfrak{G}_1(\boldsymbol{\eta}_1)\,\mathrm{d}\boldsymbol{\eta}_1 \ll \prod_{m=1}^{v} J_{\sigma_m,k_m}(2X),$$

and the desired conclusion follows from (2.2) as before.

When $s_0 = u\lfloor K/u\rfloor$, we proceed just as in the previous case, though we adjust the definition of the v-tuple $\boldsymbol{\tau}$. Observe that the conclusion of Theorem 1.2 holds for $u = u_0$ when $u_0 \geqslant 2v$ and $u_0|\kappa$, and by Corollary 1.3 it holds also for any $u \geqslant u_0$. Thus in particular, it holds for $u \geqslant \kappa$, as we have already noted. We may therefore suppose that

$$u \leqslant \kappa - 1 = \sum_{m=1}^{v} \frac{k_m(k_m+1)}{2} - 3v - 1,$$

but we maintain the condition that $u \geqslant 2v$. Define s_0 and τ as before, noting that

$$0 > \tau \geqslant -(u-1) \geqslant 2 - \kappa > 2 - \sum_{m=1}^{v} \frac{k_m(k_m+1)}{2}.$$

In this instance we define $\boldsymbol{\tau}$ to be any v-tuple of non-positive integers having the property that $|\tau_m| \leqslant k_m(k_m+1)/2$ $(1 \leqslant m \leqslant v)$ and $\tau_1 + \ldots + \tau_m = \tau$. We may then replace the exponents $k_m(k_m+1)$ with the non-negative integers $2\sigma_m$ in (2.10) and in (2.12), and proceed just as in the previous case.

3. The Hardy-Littlewood method

We can now derive Theorem 1.6 from the mean value estimate of Theorem 1.2. We make use of the notation introduced in §2, and recall in particular (2.6) and its sequel. Throughout this and the next section we will assume that $s \geqslant 2K+1$. When $\mathfrak{B} \subseteq [0,1)^r$ is a measurable set, put

$$N_{s,\mathbf{k},u}(X;\mathfrak{B}) = \int_{\mathfrak{B}} \prod_{j=1}^{s} g_k(\boldsymbol{\gamma}_j;X)\,\mathrm{d}\boldsymbol{\alpha}. \tag{3.1}$$

Our Hardy–Littlewood dissection is defined as follows. When Y and Q are parameters with $1 \leqslant Q \leqslant Y$, we take the major arcs $\mathfrak{M}_Y = \mathfrak{M}_Y(Q)$ to be the union of the boxes

$$\mathfrak{M}_Y(q,\mathbf{a}) = \left\{ \boldsymbol{\alpha} \in [0,1)^r : |\alpha_j^{(l)} - a_j^{(l)}/q| \leqslant QY^{-l} \quad (1 \leqslant j \leqslant r_l,\, 2 \leqslant l \leqslant k) \right\}, \tag{3.2}$$

with $0 \leqslant \mathbf{a} \leqslant q \leqslant Q$ and $(q,\mathbf{a}) = 1$. The corresponding set of minor arcs is defined to be $\mathfrak{m}_Y = \mathfrak{m}_Y(Q)$, where $\mathfrak{m}_Y(Q) = [0,1)^r \setminus \mathfrak{M}_Y(Q)$. Unless indicated otherwise, we fix $Y = X$ and $Q = X^{6r}$, and abbreviate $\mathfrak{M}_X$ to $\mathfrak{M}$ and $\mathfrak{m}_X$ to $\mathfrak{m}$.

We require certain auxiliary functions in order to analyse the major arcs contribution $N_{s,\mathbf{k},u}(X;\mathfrak{M})$. Write

$$S_k(q,\mathbf{a}) = \sum_{x=1}^{q} e((a^{(2)}x^2 + \ldots + a^{(k)}x^k)/q),$$

and recall that the argument of [11, Theorem 7.1] gives

$$S_k(q,\mathbf{a}) \ll (q,\mathbf{a})^{1/k} q^{1-1/k+\varepsilon}. \tag{3.3}$$

Further, set

$$v_k(\boldsymbol{\beta};X) = \int_{-X}^{X} e(\beta^{(2)}z^2 + \ldots + \beta^{(k)}z^k)\,\mathrm{d}z,$$

and recall from the arguments of [11, Theorem 7.3] the estimate

$$v_k(\boldsymbol{\beta};X) \ll X\left(1 + |\beta^{(2)}|X^2 + \ldots + |\beta^{(k)}|X^k\right)^{-1/k}. \tag{3.4}$$

We put

$$\Lambda_j^{(l)} = \sum_{i=1}^{r_l} c_{i,j}^{(l)} a_i^{(l)} \quad \text{and} \quad \vartheta_j^{(l)} = \sum_{i=1}^{r_l} c_{i,j}^{(l)} \beta_i^{(l)} \qquad (1 \leqslant j \leqslant s,\, 2 \leqslant l \leqslant k). \tag{3.5}$$

Following the same convention regarding vector notation as we applied for $\boldsymbol{\gamma}$ in (2.6) and its sequel, we have $\boldsymbol{\vartheta} = \boldsymbol{\gamma} - \boldsymbol{\Lambda}/q$. Then as a consequence of [11, Theorem 7.2], we find that when $\boldsymbol{\alpha} = \mathbf{a}/q + \boldsymbol{\beta} \in \mathfrak{M}$, one has

$$g_k(\boldsymbol{\gamma}_j;X) = q^{-1}S_k(q,\boldsymbol{\Lambda}_j)v_k(\boldsymbol{\vartheta}_j;X) + O(Q^2). \tag{3.6}$$

Finally, define

$$\mathfrak{S}(Q) = \sum_{q \leqslant Q} \sum_{\substack{1 \leqslant \mathbf{a} \leqslant q \\ (q,\mathbf{a})=1}} \prod_{j=1}^{s} q^{-1} S_k(q,\boldsymbol{\Lambda}_j) \tag{3.7}$$

and

$$\mathfrak{J}_X(Q) = \int_{\mathcal{I}(X,Q)} \prod_{j=1}^{s} v_k(\boldsymbol{\vartheta}_j;X)\,\mathrm{d}\boldsymbol{\beta},$$

where

$$\mathcal{I}(X,Q)=\prod_{l=2}^{k}[-QX^{-l},QX^{-l}]^{r_l}.$$

The preliminary conclusion of our major arcs analysis is summarised in the following lemma.

Lemma 3.1. *There is a positive number ω for which*

$$N_{s,\mathbf{k},u}(X;\mathfrak{M})=X^{s-K}\mathfrak{S}(Q)\mathfrak{J}_1(Q)+O(X^{s-K-\omega}).$$

Proof. Since $\mathrm{vol}(\mathfrak{M})\ll Q^{2r+1}X^{-K}$, it follows from (3.6) that

$$N_{s,\mathbf{k},u}(X;\mathfrak{M})=\mathfrak{S}(Q)\mathfrak{J}_X(Q)+O(X^{s-K-1}Q^{2r+3}). \tag{3.8}$$

Furthermore, by a change of variables we see that

$$\mathfrak{J}_X(Q)=X^{s-K}\mathfrak{J}_1(Q).$$

The conclusion of the lemma therefore follows from our choice $Q=X^{1/(6r)}$. □

In order to address the contribution of the minor arcs, we need the following Weyl-type estimate.

Lemma 3.2. *Suppose that $\boldsymbol{\alpha}\in\mathfrak{m}$. There exists $\tau>0$ such that for each u-tuple $(j_1,\ldots,j_u)$ of distinct indices there exists an index j_i for which one has*

$$|g_k(\boldsymbol{\gamma}_{j_i};X)|\leqslant XQ^{-\tau}.$$

Proof. This is the content of [3, Lemma 3.1]. Note that the minor arcs in our setting are a subset of the minor arcs defined in the context of that lemma. □

We now complete the analysis of the minor arcs for Theorem 1.6. In particular, we will establish the following minor arcs bound.

Lemma 3.3. *There is a positive number ω for which*

$$N_{s,\mathbf{k},u}(X;\mathfrak{m})\ll X^{s-K-\omega}.$$

Proof. Set $\sigma_0=2K+1$. We begin by estimating the last $s-\sigma_0$ exponential sums in the product (3.1) trivially, so that

$$N_{s,\mathbf{k},u}(X;\mathfrak{m})\ll X^{s-\sigma_0}\int_{\mathfrak{m}}\prod_{j=1}^{\sigma_0}|g_k(\boldsymbol{\gamma}_j;X)|\,\mathrm{d}\boldsymbol{\alpha}. \tag{3.9}$$

Given a measurable set $\mathfrak{B}\subseteq[0,1)^r$, we write

$$N^*(X;\mathfrak{B})=\int_{\mathfrak{B}}\prod_{j=1}^{\sigma_0}|g_k(\boldsymbol{\gamma}_j;X)|\,\mathrm{d}\boldsymbol{\alpha}.$$

For $1 \leqslant i \leqslant u$ and $\tau > 0$ sufficiently small, let $\mathfrak{m}^{(i)}$ denote the set of $\boldsymbol{\alpha} \in [0,1)^r$ for which $|g_k(\boldsymbol{\gamma}_i; X)| \leqslant XQ^{-\tau}$. In view of (2.3), there is a subset

$$\mathscr{J}_i \subseteq \{1, \ldots, \sigma_0\} \setminus \{i\}$$

with card $\mathscr{J}_i = K$ for which

$$N^*(X; \mathfrak{m}^{(i)}) \ll XQ^{-\tau} \oint \prod_{j \in \mathscr{J}_i} |g_k(\boldsymbol{\gamma}_j; X)|^2 \, \mathrm{d}\boldsymbol{\alpha}.$$

Write $C_i^{(l)}$ for the submatrix of $C^{(l)}$ having columns indexed by $\mathscr{J}_i$. The condition that the coefficient matrices $C^{(l)}$ be highly non-singular implies that the submatrices $C_i^{(l)}$ of $C^{(l)}$ are also highly non-singular. Thus, by orthogonality, we see from the definition (1.4) of the mean value $I_{K,\mathbf{k},u}(X)$ that that

$$N^*(X; \mathfrak{m}^{(i)}) \ll XQ^{-\tau} I_{K,\mathbf{k},u}(X; C_i^{(2)}, \ldots, C_i^{(k)}).$$

Consider a fixed $\boldsymbol{\alpha} \in \mathfrak{m}$. If τ has been chosen sufficiently small, Lemma 3.2 ensures that we can find an index $1 \leqslant j \leqslant u$ with $\boldsymbol{\alpha} \in \mathfrak{m}^{(j)}$. Thus we see that

$$\mathfrak{m} \subseteq \mathfrak{m}^{(1)} \cup \cdots \cup \mathfrak{m}^{(u)},$$

whence

$$N^*(X; \mathfrak{m}) \ll XQ^{-\tau} \max_{1 \leqslant i \leqslant u} I_{K,\mathbf{k},u}(X; C_i^{(2)}, \ldots, C_i^{(k)}). \tag{3.10}$$

Now recall that $Q = X^{1/(6r)}$. Upon combining the estimate in (3.10) with Theorem 1.2, we obtain the bound

$$N^*(X; \mathfrak{m}) \ll XQ^{-\tau} X^{K+\varepsilon} \ll X^{\sigma_0 - K} Q^{-\tau/2}.$$

Together with the trivial bound (3.9), this establishes the conclusion of the lemma. □

Upon combining the results of Lemmata 3.1 and 3.3, we infer that for some $\omega > 0$ one has the asymptotic formula

$$N_{s,\mathbf{k},u}(X) = X^{s-K} \mathfrak{S}(Q) \mathfrak{J}_1(Q) + O(X^{s-K-\omega}). \tag{3.11}$$

This completes our analysis of the minor arcs.

4. Analysis of the major arcs

It remains to show that the singular series $\mathfrak{S}(Q)$ and singular integral $\mathfrak{J}_1(Q)$ converge as Q tends to infinity. Recall from the definition (1.5) that

$$K = \kappa + 2u = \sum_{m=1}^{u} \left(\tfrac{1}{2} k_m (k_m + 1) - 1 \right)$$

denotes the total degree of this system. Throughout this section, we work under the assumption that $s \geqslant 2K + 1$.

We first complete the singular series. Put

$$A(q)=q^{-s}\sum_{\substack{1\leqslant \mathbf{a}\leqslant q\\ (q,\mathbf{a})=1}}\prod_{j=1}^{s}S_k(q,\mathbf{\Lambda}_j). \tag{4.1}$$

By applying (2.3), we find that for some choice of distinct indices $j_1,\ldots,j_u\in\{1,\ldots,s\}$ we have the asymptotic bound

$$A(q)\ll q^{2K-s}\max_{\substack{1\leqslant \mathbf{a}\leqslant q\\ (q,\mathbf{a})=1}}\biggl(\prod_{m=1}^{u}|S_k(q,\mathbf{\Lambda}_{j_m})|\biggr)^{(s-2K)/u}A_1(q), \tag{4.2}$$

where

$$A_1(q)=q^{-2K}\sum_{\substack{1\leqslant \mathbf{a}\leqslant q\\ (q,\mathbf{a})=1}}\prod_{m=1}^{u}|S_k(q,\mathbf{\Lambda}_{j_m})|^{k_m(k_m+1)-2}.$$

Note that both $A(q)$ and $A_1(q)$ are multiplicative in q. For this reason, the key to understanding the singular series is to maintain good control over the multiplicative quantity

$$B_1(q)=\sum_{d|q}A_1(d) \tag{4.3}$$

as q runs over the prime powers.

Define t_m by setting $2t_m=k_m(k_m+1)-2$ for $1\leqslant m\leqslant u$, and write $T_m=t_1+\ldots+t_m$, so that $T_u=K$. For consistency we also set $T_0=0$. Now, adopting a notation similar to that of Section 2, when $2\leqslant l\leqslant k$ we write $D^{(l)}$ for the submatrices

$$\bigl(d_{i,m}^{(l)}\bigr)_{\substack{1\leqslant i\leqslant r_l\\ 1\leqslant m\leqslant u}}=\bigl(c_{i,j_m}^{(l)}\bigr)_{\substack{1\leqslant i\leqslant r_l\\ 1\leqslant m\leqslant u}}$$

of the coefficient matrices $C^{(l)}$ consisting of the columns indexed by $j_1,\ldots,j_u$. Note that the hypothesis that each $C^{(l)}$ is highly non-singular ensures that the same is true for each $D^{(l)}$. Next, for $1\leqslant m\leqslant u$ and $2\leqslant l\leqslant k$ we set $\Delta_m^{(l)}=\Lambda_{j_m}^{(l)}$, and we employ the same conventions regarding vector notation as in (3.5) and (2.6) and its sequel. Thus, we write $\mathbf{\Delta}_m=(\Delta_m^{(l)})_{2\leqslant l\leqslant k}$ and $\mathbf{\Delta}^{(l)}=(\Delta_m^{(l)})_{1\leqslant m\leqslant u}$, so that

$$\mathbf{\Delta}^{(l)}=(D^{(l)})^T\mathbf{a}^{(l)}\qquad (2\leqslant l\leqslant k). \tag{4.4}$$

In this notation, it follows from standard orthogonality relations that

$$q^{2K-r}B_1(q)=q^{-r}\sum_{1\leqslant \mathbf{a}\leqslant q}\prod_{m=1}^{u}|S_k(q,\mathbf{\Delta}_m)|^{2t_m}$$

counts the number of solutions $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}/q\mathbb{Z})^K$ of the system of congruences

$$\sum_{m=1}^{u} d_{j,m}^{(l)} \left(\sum_{i=T_{m-1}+1}^{T_m} (x_i^l - y_i^l) \right) \equiv 0 \pmod{q}, \tag{4.5}$$

where $1 \leqslant j \leqslant r_l$ and $2 \leqslant l \leqslant k$.

Our first goal is to apply a procedure inspired by the proof of Theorem 2.1 in [3] in order to disentangle the congruences in (4.5). This will enable us to replace the sum $B_1(q)$ by a related expression in which for all indices m the degree k in the exponential sum $S_k(q, \mathbf{\Delta}_m)$ is replaced by k_m. Since k_m is typically smaller than k, we will reap the rewards of this preparatory step when the reduced degrees allow us to exert greater control on the size of the exponential sums in question.

Given a $(k-1)$-tuple of variables $\xi^{(2)}, \dots, \xi^{(k)}$, we adopt the convention that $\boldsymbol{\xi}^{[l]} = (\xi^{(2)}, \dots, \xi^{(l)})$ for $2 \leqslant l \leqslant k$. Also, when $\mathbf{d} = (d_2, \dots, d_k)$ is a coefficient vector, we abbreviate the vector $(d_2\xi^{(2)}, \dots, d_k\xi^{(k)})$ to $\mathbf{d}\boldsymbol{\xi}$, and we appropriate the notation $\mathbf{d}^{[l]}$ and $(\mathbf{d}\boldsymbol{\xi})^{[l]}$ to denote the corresponding subvectors whose entries are indexed by $2 \leqslant i \leqslant l$. The following observation will play a part in our ensuing arguments.

Lemma 4.1. *Let l, q and t be natural numbers, with $2 \leqslant l \leqslant k-1$. Suppose that $d_2, \dots, d_k$ and $c_2, \dots, c_k$ are fixed integers, and put*

$$\Gamma_q(\mathbf{d}^{[l]}) = \prod_{j=2}^{l} (q, d_j).$$

Then for any fixed integers $a^{(l+1)}, \dots, a^{(k)}$ we have

$$\sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_k(q, \mathbf{d}\mathbf{a} + \mathbf{c})|^{2t} \leqslant \Gamma_q(\mathbf{d}^{[l]}) \sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_l(q, \mathbf{a}^{[l]})|^{2t}.$$

Proof. By standard orthogonality relations, the sum

$$T = q^{1-l} \sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_k(q, \mathbf{d}\mathbf{a} + \mathbf{c})|^{2t} \tag{4.6}$$

counts solutions $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}/q\mathbb{Z})^t$ of the system of congruences

$$d_j \sum_{i=1}^{t} (x_i^j - y_i^j) \equiv 0 \pmod{q} \qquad (2 \leqslant j \leqslant l), \tag{4.7}$$

where each solution is counted with a unimodular weight depending on the inert variables $a^{(l+1)}, \dots, a^{(k)}$, together with the coefficients $\mathbf{d}$ and $\mathbf{c}$. Thus, by the triangle inequality, one finds that

$$T \leqslant q^{1-l} \sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_l(q, (\mathbf{d}\mathbf{a})^{[l]})|^{2t}.$$

We therefore discern that T is bounded above by the number of solutions of (4.7) counted without weights, and hence by the number of solutions $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}/q\mathbb{Z})^t$ of the system of congruences

$$\sum_{i=1}^t (x_i^j - y_i^j) \equiv 0 \pmod{q/(q, d_j)} \qquad (2 \leqslant j \leqslant l).$$

We interpret the latter as the number of solutions of the system

$$\sum_{i=1}^t (x_i^j - y_i^j) \equiv \frac{m_j q}{(q, d_j)} \pmod{q} \qquad (2 \leqslant j \leqslant l),$$

with $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}/q\mathbb{Z})^t$ and $1 \leqslant m_j \leqslant (q, d_j)$ for $2 \leqslant j \leqslant l$. Thus, by orthogonality and the triangle inequality, one sees that

$$\begin{aligned} T &\leqslant \sum_{\substack{1 \leqslant m_j \leqslant (q,d_j) \\ (2 \leqslant j \leqslant l)}} q^{1-l} \sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_l(q, \mathbf{a}^{[l]})|^{2t} e\left(-\sum_{j=2}^l \frac{m_j a^{(j)}}{(q, d_j)} \right) \\ &\leqslant \Gamma_q(\mathbf{d}^{[l]}) q^{1-l} \sum_{1 \leqslant \mathbf{a}^{[l]} \leqslant q} |S_l(q, \mathbf{a}^{[l]})|^{2t}. \end{aligned}$$

The conclusion of the lemma is now immediate from (4.6). □

Lemma 4.2. *Let q be a natural number, and suppose that the matrices $D^{(l)}$ are all highly non-singular. Then there exists a finite set of primes $\Omega(D)$ and a natural number $\mathcal{R}(q) = \mathcal{R}(q, D)$, both depending at most on the coefficient matrices $D^{(l)}$ and in the latter case also q, with the property that*

$$B_1(q) \leqslant \mathcal{R}(q) q^{-2K} \sum_{1 \leqslant \mathbf{a} \leqslant q} \prod_{m=1}^u |S_{k_m}(q, \mathbf{a}_m^{[k_m]})|^{2t_m}.$$

The constant $\mathcal{R}(q)$ is bounded above uniformly in q, and one can take $\mathcal{R}(q) = 1$ whenever $(q, p) = 1$ for all $p \in \Omega(D)$.

Proof. Recall that $q^{2K-r} B_1(q)$ counts the number of solutions $\mathbf{x}, \mathbf{y} \in (\mathbb{Z}/q\mathbb{Z})^K$ of the system of congruences (4.5) for $1 \leqslant j \leqslant r_l$ and $2 \leqslant l \leqslant k$. Since $B_1(q)$ is a multiplicative function of q, it is apparent that it suffices to establish the conclusion of the lemma in the special case in which q is a prime power, say $q = p^h$ for a given prime p. By applying suitable elementary row operations within the coefficient matrices $D^{(l)}$ for $2 \leqslant l \leqslant k$ that are invertible over $\mathbb{Z}/p^h\mathbb{Z}$, we may suppose without loss of generality that each coefficient matrix $D^{(l)}$ is in upper row echelon form. This operation corresponds to taking appropriate linear combinations of the congruences comprising (4.5). Here, we stress that the property that each $D^{(l)}$ is highly non-singular implies that the first $r_l \times r_l$ submatrix of $D^{(l)}$ is now upper triangular. We denote this matrix by $D_0^{(l)}$. Note that the power of p dividing the diagonal entries of $D_0^{(l)}$ depends only on the first $r_l \times r_l$ submatrices of the original coefficient matrices $D^{(l)}$. In particular, by defining $\Omega(D)$ to be the set of all primes dividing any of the determinants of the latter submatrices, we ensure that when $p \notin \Omega(D)$, then none of the diagonal entries of $D_0^{(l)}$ is divisible by p.

We now employ an inductive argument in order to successively reduce the degrees of the exponential sums occurring within the mean value

$$B_1(p^h) = p^{-2Kh} \sum_{1 \leqslant \mathbf{a} \leqslant p^h} \prod_{m=1}^{u} |S_k(p^h, \mathbf{\Delta}_m)|^{2t_m}.$$

Observe that, as a result of our preparatory manipulations, the $u \times u$ coefficient matrix $D^{(2)}$ is upper triangular. Thus, the only exponential sum within the above formula for $B_1(p^h)$ that depends on $a_u^{(2)}$ is the one involving $\mathbf{\Delta}_u$. In order to save clutter, we temporarily drop the modulus p^h in our exponential sums $S_k(p^h, \mathbf{\Delta}_j)$. We may thus write

$$B_1(p^h) = p^{-2Kh} \sum_{\substack{1 \leqslant \mathbf{a}_j^{[k_j]} \leqslant p^h \\ (1 \leqslant j \leqslant u-1)}} \left(\prod_{j=1}^{u-1} |S_k(\mathbf{\Delta}_j)|^{2t_j} \right) \sum_{1 \leqslant a_u^{(2)} \leqslant p^h} |S_k(\mathbf{\Delta}_u)|^{2t_u}.$$

The inner sum is of the shape considered in Lemma 4.1 with $l = 2$. On writing $\mathbf{d}_m = (d_{m,m}^{(l)})_{2 \leqslant l \leqslant k_m}$ $(1 \leqslant m \leqslant u)$, we thus obtain the bound

$$B_1(p^h) \leqslant p^{-2Kh} \Gamma_{p^h}(\mathbf{d}_u) \sum_{\substack{1 \leqslant \mathbf{a}_j^{[k_j]} \leqslant p^h \\ (1 \leqslant j \leqslant u-1)}} \left(\prod_{j=1}^{u-1} |S_k(\mathbf{\Delta}_j)|^{2t_j} \right) \sum_{1 \leqslant a_u^{(2)} \leqslant p^h} |S_2(a_u^{(2)})|^{2t_u}.$$

Now suppose that for some index m with $1 \leqslant m \leqslant u-1$ we have the bound

$$B_1(p^h) \leqslant p^{-2Kh} \Upsilon_m \prod_{j=m+1}^{u} \Gamma_{p^h}(\mathbf{d}_j) \sum_{1 \leqslant \mathbf{a}_j^{[k_j]} \leqslant p^h} |S_{k_j}(\mathbf{a}_j^{[k_j]})|^{2t_j}, \tag{4.8}$$

where

$$\Upsilon_m = \sum_{\substack{1 \leqslant \mathbf{a}_j^{[k_j]} \leqslant p^h \\ (1 \leqslant j \leqslant m)}} \left(\prod_{j=1}^{m} |S_k(\mathbf{\Delta}_j)|^{2t_j} \right).$$

Again, since we may assume all coefficient matrices $D^{(l)}$ to be in upper row echelon form, the only exponential sum within the mean value defining Υ_m that depends on the vector $\mathbf{a}_m^{[k_m]}$ is the one involving $\mathbf{\Delta}_m$. Thus, as in the case $m = u$ considered above, we may isolate the exponential sum indexed by m and apply Lemma 4.1.

As a result, we find that

$$\begin{aligned}\Upsilon_m &= \sum_{\substack{1\leqslant \mathbf{a}_j^{[k_j]}\leqslant p^h\\ (1\leqslant j\leqslant m-1)}} \left(\prod_{j=1}^{m-1} |S_k(\boldsymbol{\Delta}_j)|^{2t_j}\right) \sum_{1\leqslant \mathbf{a}_m^{[k_m]}\leqslant p^h} |S_k(\boldsymbol{\Delta}_m)|^{2t_m}\\ &\leqslant \Gamma_{p^h}(\mathbf{d}_m) \sum_{1\leqslant \mathbf{a}_m^{[k_m]}\leqslant p^h} |S_{k_m}(\mathbf{a}_m^{[k_m]})|^{2t_m} \sum_{\substack{1\leqslant \mathbf{a}_j^{[k_j]}\leqslant p^h\\ (1\leqslant j\leqslant m-1)}} \prod_{j=1}^{m-1} |S_k(\boldsymbol{\Delta}_j)|^{2t_j}.\end{aligned}$$

Inserting this bound into (4.8) reproduces (4.8) with m replaced by $m-1$. We may clearly iterate, and after u steps we find that

$$B_1(p^h) \leqslant p^{-2Kh} \prod_{j=1}^{u} \Gamma_{p^h}(\mathbf{d}_j) \sum_{1\leqslant \mathbf{a}_j^{[k_j]}\leqslant p^h} |S_{k_j}(\mathbf{a}_j^{[k_j]})|^{2t_j}.$$

Clearly, the vectors $\mathbf{a}_j^{[k_j]}$ with $1\leqslant j\leqslant u$ together list the coordinates of $\mathbf{a}$. Since $B_1(q)$ is multiplicative, the assertion of the lemma is now confirmed upon taking $\mathcal{R}(q)$ to be the multiplicative function defined via the formula

$$\mathcal{R}(p^h) = \prod_{j=1}^{u} \Gamma_{p^h}(\mathbf{d}_j).$$

In particular, we note that $\mathcal{R}(p^h)$ depends at most on the coefficient matrices $D^{[l]}$, and one has $\mathcal{R}(p^h)=1$ whenever $p\notin\Omega(D)$. □

With Lemma 4.2 we are now equipped to engage with our goal to show that the singular series $\mathfrak{S}=\lim_{Q\to\infty}\mathfrak{S}(Q)$ converges absolutely. In this context, for each prime number p we define the p-adic factor

$$\chi_p = \sum_{i=0}^{\infty} A(p^i). \tag{4.9}$$

Lemma 4.3. *Suppose that the coefficient matrices $C^{(l)}$ associated with the system* (1.10) *are highly non-singular, and that $r_l\geqslant r_{l+1}$ for $2\leqslant l\leqslant k-1$. Furthermore, assume that $s\geqslant 2K+1$. Then the p-adic densities χ_p exist, the singular series $\mathfrak{S}$ is absolutely convergent, and $\mathfrak{S}=\prod_p\chi_p$. Furthermore, one has $\mathfrak{S}(Q)=\mathfrak{S}+O(Q^{-\delta})$ for some $\delta>0$. Moreover, if the system* (1.10) *has a non-singular p-adic solution for all primes p, then $\mathfrak{S}\gg 1$.*

Proof. On recalling (3.7) and (4.1), we see that $\mathfrak{S}(Q)=\sum_{1\leqslant q\leqslant Q}A(q)$, and so the estimation of the quantity $A(q)$ is our central focus. The multiplicativity of $A(q)$

allows us to restrict our attention to the cases where q is a prime power. Set $\chi_p(h) = \sum_{i=0}^h A(p^i)$ and $L_p(Q) = \lfloor \log Q / \log p \rfloor$. If the product

$$\prod_{p \leqslant Q} \chi_p(L_p(Q))$$

converges absolutely as $Q \to \infty$, then so does $\mathfrak{S}(Q)$ with the same limit. In such circumstances, one has $\mathfrak{S} = \prod_p \chi_p$. It is therefore sufficient to show that for all primes p the limit

$$\chi_p = \lim_{h \to \infty} \chi_p(h)$$

exists, and moreover that there exists a positive number δ having the property that $\chi_p = 1 + O(p^{-1-\delta})$ for all but at most a finite set of primes p.

On recalling (4.2), we find from (3.3) that

$$\begin{aligned} A(p^i) &\ll (p^i)^{2K-s} \max_{\substack{1 \leqslant \mathbf{a} \leqslant p^i \\ (\mathbf{a},p)=1}} \left(\prod_{m=1}^u |S_k(p^i, \mathbf{\Delta}_m)| \right)^{(s-2K)/u} A_1(p^i) \\ &\ll \max_{\substack{1 \leqslant \mathbf{a} \leqslant p^i \\ (\mathbf{a},p)=1}} \left(\prod_{m=1}^u (p^i)^{-1/k+\varepsilon} (p^i, \mathbf{\Delta}_m)^{1/k} \right)^{(s-2K)/u} A_1(p^i). \end{aligned}$$

The invertibility of the coordinate transform (4.4) implies that when $(\mathbf{a}, p^i) = 1$, then there is at least one index m with $1 \leqslant m \leqslant u$ such that $(p^i, \mathbf{\Delta}_m) \ll 1$ with an implied constant depending at most on the coefficient matrices $C^{(l)}$. Since $s - 2K \geqslant 1$ and ε may be taken arbitrarily small, we deduce that there is a positive number c_1, depending at most on the coefficient matrices $C^{(l)}$, having the property that

$$A(p^i) \leqslant c_1 p^{-i/(2ku)} A_1(p^i). \tag{4.10}$$

We now wish to apply Lemma 4.2. To this end, we first recall (4.3) and observe that a summation by parts yields the relation

$$\sum_{i=0}^L p^{-\frac{i}{2ku}} A_1(p^i) = p^{-\frac{L}{2ku}} B_1(p^L) + \sum_{j=0}^{L-1} \left(p^{-\frac{j}{2ku}} - p^{-\frac{j+1}{2ku}} \right) B_1(p^j). \tag{4.11}$$

Since all coefficients on the right hand side are positive, and also both $B_1(p^j)$ and

$$B_1^*(p^j) = p^{-2Kj} \sum_{1 \leqslant \mathbf{a} \leqslant p^j} \prod_{m=1}^u |S_{k_m}(p^j, \mathbf{a}_m^{[k_m]})|^{2t_m}$$

are non-negative for all non-negative integers j, it follows from Lemma 4.2 that we may majorise the right hand side of (4.11) by replacing $B_1(p^j)$ with $\mathcal{R}(p^j) B_1^*(p^j)$ for $0 \leqslant j \leqslant L$. Set $\mathcal{R}_p = \max_{j \geqslant 0} \mathcal{R}(p^j)$, noting that this maximum exists as $\mathcal{R}(p^j)$

is an integer which is bounded uniformly for all non-negative integers j. Also, in analogy to the definition of $B_1^*(p^j)$, we put

$$A_1^*(p^j)=p^{-2Kj}\sum_{\substack{1\leqslant \mathbf{a}\leqslant p^j\\ (\mathbf{a},p)=1}}\prod_{m=1}^{u}|S_{k_m}(p^j,\mathbf{a}_m^{[k_m]})|^{2t_m}.$$

Thus, another summation by parts shows that the right hand side of (4.11) is no larger than

$$\mathcal{R}_p\left(p^{-\frac{L}{2ku}}B_1^*(p^L)+\sum_{j=0}^{L-1}\left(p^{-\frac{j}{2ku}}-p^{-\frac{j+1}{2ku}}\right)B_1^*(p^j)\right)=\mathcal{R}_p\sum_{i=0}^{L}p^{-\frac{i}{2ku}}A_1^*(p^i).$$

We have therefore established the bound

$$\sum_{i=0}^{L}p^{-\frac{i}{2ku}}A_1(p^i)\leqslant \mathcal{R}_p\sum_{i=0}^{L}p^{-\frac{i}{2ku}}A_1^*(p^i). \tag{4.12}$$

Since $2t_m=k_m(k_m+1)-2\geqslant k_m^2$ for all m, we can infer further from (3.3) that there exists a positive number c_2, depending at most on ε, such that

$$\begin{aligned}A_1^*(p^i)&\leqslant c_2p^{i\varepsilon}\sum_{\substack{1\leqslant \mathbf{a}\leqslant p^i\\ (\mathbf{a},p)=1}}\prod_{m=1}^{u}\left(p^{-i/k_m}(p^i,\mathbf{a}_m^{[k_m]})^{1/k_m}\right)^{2t_m}\\ &\leqslant c_2p^{i\varepsilon}\sum_{\substack{1\leqslant \mathbf{a}\leqslant p^i\\ (\mathbf{a},p)=1}}\prod_{m=1}^{u}p^{-ik_m}(p^i,\mathbf{a}_m^{[k_m]})^{k_m}.\end{aligned}$$

For a fixed vector $\mathbf{e}\in\mathbb{Z}_{\geqslant 0}^u$ denote by $\Xi(p^i,\mathbf{e})$ the number of vectors $\mathbf{a}\in\mathbb{Z}^r$ satisfying $1\leqslant \mathbf{a}\leqslant p^i$ and $(p^i,\mathbf{a}_m^{[k_m]})=p^{e_m}$ for $1\leqslant m\leqslant u$. Then one has

$$A_1^*(p^i)\leqslant c_2p^{i\varepsilon}\sum_{\mathbf{e}}\Xi(p^i,\mathbf{e})\prod_{m=1}^{u}p^{(e_m-i)k_m},$$

where the sum is over all vectors $\mathbf{e}\in\mathbb{Z}^u$ satisfying $0\leqslant e_m\leqslant i$ and having the property that $e_m=0$ for at least one index m. For any fixed m, the number of choices for $\mathbf{a}_m^{[k_m]}\in\mathbb{Z}^{k_m-1}$ having $1\leqslant \mathbf{a}_m^{[k_m]}\leqslant p^i$ and $(p^i,\mathbf{a}_m^{[k_m]})=p^{e_m}$ is at most $p^{(i-e_m)(k_m-1)}$. It follows that

$$\Xi(p^i,\mathbf{e})\leqslant\prod_{m=1}^{u}p^{(i-e_m)(k_m-1)},$$

and hence

$$A_1^*(p^i)\leqslant c_2p^{i\varepsilon}\sum_{\substack{0\leqslant \mathbf{e}\leqslant i\\ e_1\dots e_u=0}}\prod_{m=1}^{u}p^{e_m-i}\leqslant c_2p^{i\varepsilon-i}\left(\sum_{e=0}^{i}p^{e-i}\right)^{u-1}\leqslant 2^uc_2(p^i)^{-1+\varepsilon}. \tag{4.13}$$

On recalling (4.10), (4.12) and (4.13) we find that

$$\begin{aligned}\sum_{i=1}^{\infty}|A(p^i)| &\leqslant c_1\biggl(-1+\sum_{i=0}^{\infty}p^{-i/(2ku)}A_1(p^i)\biggr)\\ &\leqslant c_1\biggl(-1+\mathcal{R}_p\sum_{i=0}^{\infty}p^{-i/(2ku)}A_1^*(p^i)\biggr)\\ &\leqslant c_1(\mathcal{R}_p-1)+2^u c_1c_2\mathcal{R}_p\sum_{i=1}^{\infty}p^{-i(1+1/(2ku)-\varepsilon)}\ll 1.\end{aligned}$$

It follows that the p-adic density χ_p defined in (4.9) exists. Since ε may be taken sufficiently small, we deduce in particular that, when $\mathcal{R}_p=1$, there is a positive number c_3, depending at most on the coefficient matrices $C^{(l)}$, such that

$$\sum_{i=1}^{\infty}|A(p^i)|\leqslant c_3p^{-1-1/(3ku)}. \tag{4.14}$$

On recalling the conclusion of Lemma 4.2, one sees that $\mathcal{R}_p=1$ for all but a finite set of primes depending at most on the coefficient matrices $C^{(l)}$, namely those primes p with $p\notin\Omega(D)$, and thus

$$\prod_p\biggl(\sum_{i=0}^{\infty}|A(p^i)|\biggr)\ll\prod_p\bigl(1+p^{-1-1/(3ku)}\bigr)^{c_3}<\zeta(1+1/(3ku))^{c_3}.$$

Hence, the singular series $\mathfrak{S}$ converges absolutely and one has $\mathfrak{S}=\prod_p\chi_p$.

Furthermore, a standard argument yields

$$\chi_p=\lim_{i\to\infty}p^{-i(s-r)}M(p^i),$$

where $M(q)$ denotes the number of solutions $\mathbf{x}\in(\mathbb{Z}/q\mathbb{Z})^s$ of the congruences

$$c_{j,1}^{(l)}x_1^l+\ldots+c_{j,s}^{(l)}x_s^l\equiv 0\pmod q\qquad(1\leqslant j\leqslant r_l,\ 2\leqslant l\leqslant k),$$

corresponding to the equations (1.10). Using again the observation that $\mathcal{R}_p=1$ for all sufficiently large primes p, we discern from (4.14) that there exists an integer p_0 with the property that

$$1/2\leqslant\prod_{\substack{p>p_0\\ p\text{ prime}}}\chi_p\leqslant 3/2.$$

For the remaining finite set of primes, a standard application of Hensel's lemma shows that $\chi_p>0$ whenever the system (1.10) possesses a non-singular solution in $\mathbb{Q}_p$. We thus conclude that under the hypotheses of the lemma we have $\mathfrak{S}\gg 1$ as claimed. □

We next demonstrate the existence of the limit

$$\chi_\infty = \lim_{Q\to\infty} \mathfrak{J}_1(Q).$$

With this goal in mind, when W is a positive real number, we introduce the auxiliary mean value

$$\mathfrak{J}_1^*(W) = \int_{[-W,W]^r} \prod_{j=1}^s |v_k(\boldsymbol{\vartheta}_j;1)|\,\mathrm{d}\boldsymbol{\beta}.$$

Lemma 4.4. *Under the hypotheses of Theorem 1.6, there is a positive number δ for which one has $\mathfrak{J}_1^*(2Q)-\mathfrak{J}_1^*(Q)\ll Q^{-\delta}$, and hence the limit χ_∞ exists. In particular, one has*

$$\mathfrak{J}_1(Q) = \chi_\infty + O(Q^{-\delta}).$$

Furthermore, if the system (1.10) *has a non-singular solution inside the real unit cube $(-1,1)^s$, then the singular integral χ_∞ is positive.*

Proof. The first part of the proof is inspired by a singular series argument of Heath-Brown and Skorobogatov (see [9, pages 173 and 174]). Let $\mathscr{J}$ denote the set of K-element subsets $\{j_1,\ldots,j_K\}$ of $\{1,\ldots,s\}$. When $J\in\mathscr{J}$, define

$$\mathscr{S}_J(Q) = \sum_{1\leqslant q\leqslant Q}\sum_{\substack{1\leqslant \mathbf{a}\leqslant q\\ (q,\mathbf{a})=1}} q^{-2K}\prod_{j\in J}|S_k(q,\boldsymbol{\Lambda}_j)|^2 \tag{4.15}$$

and

$$\mathscr{I}_J(Q) = \int_{[-Q,Q]^r}\prod_{j\in J}|v_k(\boldsymbol{\vartheta}_j;1)|^2\,\mathrm{d}\boldsymbol{\beta}.$$

Set $Y=Q^{6r}$, and define the major arcs $\mathfrak{M}_Y(Q)$ via (3.2). By making the necessary modifications to our initial analysis of the major arcs, we see from (3.8) that for any $J\in\mathscr{J}$ one has

$$\int_{\mathfrak{M}_Y(Q)}\prod_{j\in J}|g_k(\boldsymbol{\gamma}_j;Y)|^2\,\mathrm{d}\boldsymbol{\alpha} = Y^K\mathscr{S}_J(Q)\mathscr{I}_J(Q)+O(Y^KQ^{-1}). \tag{4.16}$$

Note that we have $S_k(1,\mathbf{1})=1$ for the term corresponding to $q=1$ in (4.15). Since all other summands are non-negative, it follows that for any $Q\geqslant 1$ and any $J\in\mathscr{J}$, one has

$$\mathscr{S}_J(Q)\geqslant 1. \tag{4.17}$$

On the other hand, for $Y=Q^{6r}$ the major arcs $\mathfrak{M}_Y(Q)$ are disjoint, and we conclude from Theorem 1.2 that

$$\int_{\mathfrak{M}_Y(Q)}\prod_{j\in J}|g_k(\boldsymbol{\gamma}_j;Y)|^2\,\mathrm{d}\boldsymbol{\alpha} \ll I_{K,\mathbf{k},u}(Y)\ll Y^{K+\varepsilon}.$$

In combination with (4.16) and (4.17) it follows that

$$\max_{J\in\mathcal{J}} \mathscr{I}_J(Q) \ll Y^{\varepsilon}. \tag{4.18}$$

Since Y is a power of Q, we discern from (3.4) and (4.18) via (2.3) that for any $Q>1$ we have

$$\begin{aligned}|\mathfrak{J}_1^*(2Q)-\mathfrak{J}_1^*(Q)| &\ll \Bigg(\sup_{|\boldsymbol{\beta}|>Q}\prod_{j=1}^{s}|v_k(\boldsymbol{\vartheta}_j;1)|\Bigg)^{1-\frac{2K}{s}} \int_{[-2Q,2Q]^r}\prod_{j=1}^{s}|v_k(\boldsymbol{\vartheta}_j;1)|^{\frac{2K}{s}}\,\mathrm{d}\boldsymbol{\beta}\\ &\ll Q^{-1/(ks)}\max_{J\in\mathcal{J}}\mathscr{I}_J(2Q) \ll Q^{-1/(ks)+\varepsilon}.\end{aligned}$$

Here, we exploited the fact that, since the coefficient matrices $C^{(l)}$ are highly non-singular, the condition $|\boldsymbol{\beta}|>Q$ implies that $|\boldsymbol{\vartheta}_j|\gg Q$ for some index j with $1\leqslant j\leqslant s$. This implies the first statement of the lemma. In particular, the singular integral χ_∞ converges absolutely.

In order to establish the second claim, we follow an argument of Schmidt [10]. When $T\geqslant 1$, define

$$w_T(y)=\begin{cases}T(1-T|y|) & \text{when } |y|\leqslant T^{-1},\\ 0 & \text{otherwise,}\end{cases}$$

and recall that

$$w_T(y)=\int_{-\infty}^{\infty} e(\beta y)\left(\frac{\sin(\pi\beta/T)}{\pi\beta/T}\right)^2\mathrm{d}\beta, \tag{4.19}$$

where the integral converges absolutely. Set

$$\Phi_j^{(l)}(\mathbf{x})=c_{j,1}^{(l)}x_1^l+\ldots+c_{j,s}^{(l)}x_s^l \qquad (1\leqslant j\leqslant r_l,\ 2\leqslant l\leqslant k),$$

and put

$$W_T=\int_{[-1,1]^s}\prod_{l=2}^{k}\prod_{j=1}^{r_l} w_T(\Phi_j^{(l)}(\mathbf{z}))\,\mathrm{d}\mathbf{z}.$$

We adapt the argument of §11 in Schmidt's work [10] to show that $W_T\to\chi_\infty$ as $T\to\infty$.

Set

$$\psi_T(\boldsymbol{\beta})=\prod_{l=2}^{k}\prod_{j=1}^{r_l}\left(\frac{\sin(\pi\beta_j^{(l)}/T)}{\pi\beta_j^{(l)}/T}\right)^2.$$

Then in the light of (4.19) a change of the order of integration shows that

$$W_T=\int_{\mathbb{R}^r}\left(\prod_{i=1}^{s}v_k(\boldsymbol{\vartheta}_i;1)\right)\psi_T(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta},$$

and hence

$$W_T - \chi_\infty = \int_{\mathbb{R}^r} \left(\prod_{i=1}^s v_k(\boldsymbol{\vartheta}_i; 1) \right) (\psi_T(\boldsymbol{\beta}) - 1) \, \mathrm{d}\boldsymbol{\beta}. \tag{4.20}$$

In order to analyse the integral on the right hand side of (4.20), it is convenient to consider two domains separately. Write $U_1 = [-\sqrt{T}, \sqrt{T}]^r$, and set $U_2 = \mathbb{R}^r \setminus U_1$. From the power series expansion of ψ_T we find that

$$0 \leqslant 1 - \psi_T(\boldsymbol{\beta}) \ll \min \left\{ 1, \sum_{l=2}^k \sum_{j=1}^{r_l} (|\beta_j^{(l)}|/T)^2 \right\},$$

whence we discern that the domain U_1 contributes at most

$$\sup_{\boldsymbol{\beta} \in U_1} |1 - \psi_T(\boldsymbol{\beta})| \int_{\mathbb{R}^r} \prod_{i=1}^s |v_k(\boldsymbol{\vartheta}_i; 1)| \, \mathrm{d}\boldsymbol{\beta} \ll T^{-1}.$$

Note that in the last step we used our previous insight that the singular integral converges absolutely. Meanwhile, the contribution from U_2 is bounded above by

$$\sum_{i=1}^\infty |\mathfrak{J}_1^*(2^i \sqrt{T}) - \mathfrak{J}_1^*(2^{i-1} \sqrt{T})| \ll \sum_{i=1}^\infty (2^i \sqrt{T})^{-\delta} \ll T^{-\delta/2},$$

for some positive number δ with $\delta < 1$, where again we took advantage of our earlier findings. Thus we infer from (4.20) that

$$|W_T - \chi_\infty| \ll T^{-\delta/2} \tag{4.21}$$

for all $T \geqslant 1$, and hence W_T does indeed converge to χ_∞, as claimed.

Suppose now that the system (1.10) has a non-singular solution inside $(-1, 1)^s$. Then it follows from the implicit function theorem that the real manifold described by the equations in (1.10) has positive $(s - r)$-dimensional volume inside $(-1, 1)^s$. In such circumstances, Lemma 2 of Schmidt [10] shows that $W_T \gg 1$ uniformly in T. We therefore deduce from (4.21) that χ_∞ is indeed positive, confirming the second claim of the lemma. □

Upon combining our results of (3.11) with Lemmata 4.3 and 4.4, we conclude that

$$\begin{aligned} N_{s,\mathbf{k},u}(X) &= X^{s-k} \left(\mathfrak{S} + O(Q^{-\delta}) \right) \left(\chi_\infty + O(Q^{-\delta}) \right) + O(X^{s-k-\omega}) \\ &= (\mathcal{C} + o(1)) X^{s-k}, \end{aligned}$$

where $\mathcal{C} = \chi_\infty \prod_p \chi_p$. Moreover, the constant $\mathcal{C}$ is positive whenever the system (1.10) possesses non-singular solutions in all local fields. This confirms the asymptotic formula (1.11), and completes our proof of Theorem 1.6.

References

[1] J. Bourgain, C. Demeter and L. Guth, *Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three*, Ann. of Math. (2) **184** (2016), no. 2, 633–682.

[2] J. Brandes, *The Hasse principle for systems of quadratic and cubic diagonal equations*, Q. J. Math. **68** (2017), no. 3, 831–850.

[3] J. Brandes and S. T. Parsell, *Simultaneous additive equations: repeated and differing degrees*, Canad. J. Math. **69** (2017), no. 2, 258–283.

[4] J. Brandes and T. D. Wooley, *Vinogradov systems with a slice off*, Mathematika **63** (2017), no. 3, 797–817.

[5] J. Brüdern and T. D. Wooley, *The Hasse principle for pairs of diagonal cubic forms*, Ann. of Math. (2) **166** (2007), no. 3, 865–895.

[6] J. Brüdern and T. D. Wooley, *The Hasse principle for systems of diagonal cubic forms*, Math. Ann. **364** (2016), no. 3-4, 1255–1274.

[7] R. J. Cook, *Simultaneous quadratic equations*, J. London Math. Soc. (2) **4** (1971), 319–326.

[8] R. J. Cook, *Simultaneous quadratic equations II*, Acta Arith. **25** (1973/74), 1–5.

[9] R. Heath-Brown and A. Skorobogatov, *Rational solutions of certain equations involving norms*, Acta Math. **189** (2002), no. 2, 161–177.

[10] W. M. Schmidt, *Simultaneous rational zeros of quadratic forms*, in: Seminar on Number Theory (Paris 1980-81), Progr. Math. 22, Birkhäuser, Boston, MA, 1982, 281–307.

[11] R. C. Vaughan, *The Hardy–Littlewood Method*. Second ed., Cambridge University Press, Cambridge, 1997.

[12] T. D. Wooley, *Rational solutions of pairs of diagonal equations, one cubic and one quadratic*, Proc. London Math. Soc. (3) **110** (2015), no. 2, 325–356.

[13] T. D. Wooley, *The cubic case of the main conjecture in Vinogradov's mean value theorem*, Adv. Math. **294** (2016), 532–561.

[14] T. D. Wooley, *Nested efficient congruencing and relatives of Vinogradov's mean value theorem*, Proc. London Math. Soc. (*in press*); https://doi.org/10.1112/plms.12204; preprint available as arXiv:1708.01220.

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

E-mail address: brjulia@chalmers.se

School of Mathematics, University of Bristol, University Walk, Clifton, Bristol BS8 1TW, United Kingdom

E-mail address: matdw@bristol.ac.uk