

Diophantine equations in the primes

Brian Cook · Ákos Magyar

Received: 31 August 2013 / Accepted: 30 December 2013 / Published online: 23 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Let $\mathbf{p} = (p_1, \dots, p_r)$ be a system of r polynomials with integer coefficients of degree d in n variables $\mathbf{x} = (x_1, \dots, x_n)$. For a given r -tuple of integers, say $\mathbf{s}$, a general local to global type statement is shown via classical Hardy-Littlewood type methods which provides sufficient conditions for the solubility of $\mathbf{p}(\mathbf{x}) = \mathbf{s}$ under the condition that each of the x_i 's is prime.

Mathematics Subject Classification (1991) 11D72 · 11P32

1 Background and main results

Let $\mathbf{p} = (p_1, \dots, p_r)$ be a system of polynomials with integer coefficients of degree d in the n variables $\mathbf{x} = (x_1, \dots, x_n)$. Our primary concern is finding solutions with each coordinate prime, which we call prime solutions to the system of equations $\mathbf{p}(\mathbf{x}) = \mathbf{s}$, where $\mathbf{s} \in \mathbb{Z}^r$ is a fixed element. In more

Ákos Magyar is supported by NSERC grant 22R44824 and ERC-AdG. 321104.

B. Cook

Department of Mathematics, University of Wisconsin–Madison, Madison,
WI 53706-1325, USA
e-mail: bcook@math.wisc.edu

Á. Magyar (✉)

Department of Mathematics, The University of British Columbia, 1984 Mathematics Road,
Vancouver, BC V6T 1Z2, Canada
e-mail: magyar@math.ubc.ca

geometric terms, our task is to count prime points on the complex affine variety defined by this equation, which we will denote by $V_{\mathbf{p},s}$.

If $\mathbf{p}$ is composed entirely of linear forms the known results may be split into a classical regime and a modern one. With $\mathbf{p}$ being a system of r linear forms, define the rank of $\mathbf{p}$ to be the minimum number of nonzero coefficients in a non-trivial linear combination

$$\lambda_1 \mathbf{p}_1 + \cdots + \lambda_r \mathbf{p}_r,$$

and denote this quantity by $\mathcal{B}_1(\mathbf{p})$. This quantity is of course positive if and only if the forms are linearly independent. The classical results on the large scale distribution of prime points on $V_{\mathbf{p},s}$ are conditional on the rank being sufficiently large in terms of r (for example, $2r + 1$ follows from what is shown here). In this realm are many well known results such as the ones due to Vinogradov [18], van der Corput, and more recently Balog [1]. The modern results are mostly summed up in the work of Green and Tao [7], where the large scale distribution of prime points on $V_{\mathbf{p},s}$ is determined only on the condition that $\mathcal{B}_1(\mathbf{p})$ is at least three, a quantity independent of r . These results cover all scenarios that do not reduce to a binary problem. However, the most recent result is due to Zhang. He has shown, by extending the already stunning results of Goldston et al. [5], that one of the equations¹ $x_1 - x_2 = 2i$, $i = 1, \dots, 35 \times 10^6$ does have infinitely many prime solutions [19].

The scenario for systems involving higher degree forms is certainly less clean cut [4], and even the study of integral points on $V_{\mathbf{p},s}$ is a non-trivial problem. General results for the large scale distribution of integral points are provided by Birch [2] and Schmidt [17], which again require the system to be large with respect to c notions of rank (with respect to the number and degrees of the forms involved). Working within the limitations of these results, one should expect to be able to understand the large scale distribution of prime points as well. For systems of forms which are additive, for instance the single form $a_1 x_1^d + \cdots + a_n x_n^d$, this is something that has been done and the primary result here is due to Hua [11]. On the opposite end, if the system of forms is a bilinear system, or even contains a large bilinear piece, one can also provide similar results, a particular instance of which is given by Liu [15] for a quadratic form. However, there have previously been no results for general systems of nonlinear polynomials. Providing such a result is the aim of this work.

Before continuing it is worth discussing this problem from the point view of some recent results in sieve theory. Let $\Gamma \subset GL_n(\mathbb{Z})$ be a finitely generated subgroup, $\mathcal{O} = \Gamma \mathbf{b}$ be the orbit for some $\mathbf{b} \in \mathbb{Z}^n$, and g be an integral polynomial in n variables. Bourgain et al. in [3] initiated the study of a sieve method

¹ The number of needed equations has been reduced by a poly-math project.

for finding points $\mathbf{y} \in \mathcal{O}$ such that $g(\mathbf{y})$ has few prime factors, provided of course there are no obvious constraints. In particular they prove for suitable groups that the almost prime points of the orbit form a Zariski dense subset, taking g to be the monomial form $g(\mathbf{y}) = y_1 \dots y_n$. In case if a form f is preserved by a sufficiently large group of transformations $\Gamma \subset GL_n(\mathbb{Z})$, the level sets $V_{f,s}$ are partitioned into orbits of Γ and the methods of [3] show the existence of almost prime solutions. The advantage is that one does not need the largeness of the rank of the form, however, apart from quadratic forms, examples of such forms are quite rare (e.g. determinant forms), see [16] for the study of integer solutions for such forms. Liu and Sarnak in [14] carry out this idea for indefinite quadratic forms in three variables, with Γ being the subgroup of the special orthogonal group which preserves f . In particular, they prove that one can find points on certain level sets of such quadratic forms where each coordinate has at most 30 prime factors.

Returning to the problem at hand, for a fixed system of polynomials $\mathbf{p}$ let us define for each prime p the quantity

$$\mu_p = \lim_{t \rightarrow \infty} \frac{(p^t)^r M(p^t)}{\phi^n(p^t)},$$

provided the limit exists, where $M(p^t)$ represents the number of solutions to the equation $\mathbf{p}(\mathbf{x}) = \mathbf{s}$ in the multiplicative group of reduced residue classes $\text{mod } p^t$, denoted by $U_{p^t}^n$, and ϕ is Euler's totient function. A general heuristic argument suggests that we should have

$$\begin{aligned} \mathcal{M}_{\mathbf{p},\mathbf{s}}(N) &:= \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) \mathbf{1}_{V_{\mathbf{p},\mathbf{s}}}(\mathbf{x}) \\ &\approx \mu_\infty(N, \mathbf{s}) \prod_{p \text{ prime}} \mu_p(\mathbf{s}) N^{n-D}, \end{aligned} \quad (1.1)$$

where $\mu_\infty(N, \mathbf{s})$ is the singular integral that appears in the study of integral points on $V_{\mathbf{p},\mathbf{s}}$, see [2], [17], Λ denotes the von Mangoldt function, $D = dr$, and $\Lambda(\mathbf{x}) = \Lambda(x_1) \dots \Lambda(x_n)$.

What is actually shown here is a precise result of this form for systems of polynomials of common degree provided that the system has large rank in the sense of Birch for the nonlinear forms and in the sense described above for linear forms. Let us be given a system of homogeneous polynomials $\mathbf{f} = (f_1, \dots, f_r)$ with integer coefficients of degree exactly d . Define the singular variety, over $\mathbb{C}^n$, associated to the forms $\mathbf{f}$ to be the collection of $\mathbf{x}$ such that

the Jacobian of $\mathbf{f}$ at $\mathbf{x}$, given by the matrix of partial derivatives

$$Jac_{\mathbf{f}}(\mathbf{x}) = \left[\frac{\partial f_k}{\partial x_j}(\mathbf{x}) \right]_{k=1, j=1}^{r, n},$$

has rank strictly less than r . This collection is labeled as $V_{\mathbf{f}}^*$. The Birch rank $\mathcal{B}_d(\mathbf{f})$ is defined for $d > 1$ and is given by $\text{codim}(V_{\mathbf{f}}^*)$, provided that $r \neq 0$. If $r = 0$ then we simply assign the value ∞ . This notion is extended to a general polynomial system $\mathbf{p}$ of degree d by defining the rank by $\mathcal{B}_d(\mathbf{p}) = \mathcal{B}_d(\mathbf{f})$, where $\mathbf{f}$ is the system of forms consisting of the highest degree homogeneous parts of the polynomials $\mathbf{p}$. In particular, if the a system of forms has positive Birch rank then the systems of forms $\mathbf{f}$ is a linearly independent system.

The main result that is shown here is the following.

Theorem 1 *For given positive integers r and d , there exists a constant $\chi(r, d)$ such that the following holds:*

Let $\mathbf{p} = \mathbf{p}^{(d)}$ be a given system of integral polynomials with r polynomials of degree d in n variables, and set $D = dr$. If we have $\mathcal{B}_d(\mathbf{p}) \geq \chi(r, d)$ then for the equation $\mathbf{p}(\mathbf{x}) = \mathbf{s}$ we have an asymptotic of the form

$$\mathcal{M}_{\mathbf{p}, \mathbf{s}}(N) = \prod_p \mu_p(\mathbf{s}) \mu_{\infty}(N, \mathbf{s}) N^{n-D} + O\left(N^{n-D} (\log N)^{-1}\right)$$

Moreover, if $\mathbf{p}(\mathbf{x}) = \mathbf{s}$ has a nonsingular solution in $\mathbb{U}_p$, the p -adic integer units, for all primes p , then

$$\prod_p \mu_p(\mathbf{s}) > 0.$$

Note that the singular integral $\mu_{\infty}(N, \mathbf{s})$ is the same as in the work of Birch and Schmidt (see [17], Sec.9 and [2], Sec.6). It is positive provided that the variety $V_{\mathbf{f}, N^{-d}\mathbf{s}}$ has a nonsingular real point in the open cube $(0, 1)^n$, and is bounded from below independently of N if there is a nonsingular real point in the cube $(\varepsilon, 1 - \varepsilon)^n$ for some $\varepsilon > 0$. Indeed, the singular integral has the representation

$$\int_{V_{\mathbf{f}, \mathbf{t}} \cap [0, 1]^n} d\sigma_{\mathbf{f}}(x), \quad (1.2)$$

where $\mathbf{t} = N^{-d}\mathbf{s}$, and $d\sigma_{\mathbf{f}}(x)$ is a positive measure on $V_{\mathbf{f}, \mathbf{t}} \setminus V_{\mathbf{f}}^*$ which is absolutely continuous with respect to the Euclidean surface area measure, see Birch [2], Sec.6.

The quantitative aspects of the constants $\chi(r, d)$ are in general extremely poor. The terms $\chi(r, 1)$ may be taken to be $2r + 1$. The case for quadratic forms is still somewhat reasonable, for systems of quadratics one can show $\chi(r, 2) \leq 2^{2^{Cr^2}}$ (to be compared to $r(r+1)$ for the integral analogue). However, the constants $\chi(1, d)$ already exhibit tower type behavior in d (to be compared to $d2^d$ for the integral analogue), and the situation further worsens from there.

A more detailed analysis of the singular series can be carried out in certain instances to provide more concrete statements. This is not done here, however there are some circumstances that have previously been considered. Borrowing the results in this direction from Hua and Vinogradov gives the following known results (which are weaker than the best known for $d > 1$) for diagonal forms as a corollary. Note that for a diagonal form f of degree d in n variables we have $\mathcal{B}_d(f) = n$.

Corollary 1 *Theorem 1 applies to diagonal integral forms $f(\mathbf{x}) = a_1x_1^d + a_2x_2^d + \dots + a_nx_n^d$, $a_1a_2 \dots a_n \neq 0$, when $n \geq 3 + (d-1)2^d$. In particular if $d = 1$ and $a_1 = a_2 = a_3 = 1$ one recovers the well known fact that every sufficiently large odd integer is a sum of three primes; if $d = 2$ and $a_1 = a_2 = \dots = a_7 = 1$ then every sufficiently large integer congruent to 7 modulo 24 is a sum of seven squares of primes.*

Liu provides a general scenario to guarantee that the singular series is positive for a single quadratic form. A quadratic form Q has the representation $\langle \mathbf{x}, A\mathbf{x} \rangle$ for some symmetric integral $n \times n$ matrix A . The Jacobian of the form is $2A\mathbf{x}$, giving that the singular variety V_Q^* is the null space of A , and so $\mathcal{B}(Q) = \text{rank}(A)$. A general condition for Q to be well behaved with respect to each prime modulus is that p^{n-2} does not divide $\det(A)$ for any prime p , as then a simple consideration of eigenvalues shows that the matrix A has rank at least three when viewed over the finite field with p elements.

Corollary 2 *Let $Q(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle$ be an indefinite integral quadratic form in n variables with $\text{rank}(A) \geq \chi(2, 1)$ and $p^{n-2} \nmid \det(A)$ for any prime p . Then $Q(\mathbf{x}) = 0$ has a solution in $\mathbb{U}_p$ for all p if and only if $Q(\mathbf{x}) = 0$ has a solution with x_i prime for each i .*

The value of $\chi(2, 1)$ can be worked out directly with what is done in this paper with minimal difficulty, and previously a value of 34 has been provided by the first author. Applying the results of Liu, as mentioned in the following section, is however more efficient and gives a value of 21. For the related problem of finding integer solutions to a single quadratic equation $Q(\mathbf{x}) = 0$ in a set of positive upper density $A \subseteq \mathbb{Z}$, Keil in [13] has recently obtained the rank condition $\text{rank}(Q) \geq 17$. The analogous statement for the integers is the famous Hasse–Minkowski Theorem which requires no conditions on the rank or the determinant, and of course only looks at solubility in $\mathbb{Z}_p$ and not $\mathbb{U}_p$.

2 Overview

The primary technique used in the proof of Theorem 1 is the circle method, and the argument is an adaptation of the following mean value approach. If a single integral form $\mathfrak{F}$ of degree d in n variables takes the shape

$$\mathfrak{F}(\mathbf{x}) = x_1^d + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}), \quad (2.1)$$

where $\mathbf{x} = (x_1, \mathbf{y}, \mathbf{z})$, then we have the representation

$$\begin{aligned} \mathcal{M}_{\mathfrak{F},0} &= \int_0^1 \left(\sum_{x_1 \in [N]} \Lambda(x_1) e(\alpha x_1^d) \right) \left(\sum_{\mathbf{y} \in [N]^m} \Lambda(\mathbf{y}) e(\alpha \mathfrak{F}_1(\mathbf{y})) \right) \\ &\quad \times \left(\sum_{\mathbf{z} \in [N]^{n-1-m}} \Lambda(\mathbf{z}) e(\alpha \mathfrak{F}_2(\mathbf{z})) \right) d\alpha := \int_0^1 S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha. \end{aligned}$$

An application of the Cauchy–Schwarz inequality then gives

$$\mathcal{M}_{\mathfrak{F},0}^2 \leq \|S_0\|_\infty^2 \|S_1\|_2^2 \|S_2\|_2^2 \leq \|S_0\|_\infty^2 (\log N)^{2n-2} Y(N) Z(N),$$

where $Y(N)$ is the number of solutions to the equation $\mathfrak{F}_1(\mathbf{y}) = \mathfrak{F}_1(\mathbf{y}')$ with $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $Z(N)$ is the number of solutions to the equation $\mathfrak{F}_2(\mathbf{z}) = \mathfrak{F}_2(\mathbf{z}')$ with $\mathbf{z}, \mathbf{z}' \in [N]^{n-1-m}$. If $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are assumed to have large rank, then $Y(N)Z(N) = O(N^{2m-d}) O(N^{2(n-1-m)-d}) = O(N^{2n-2-2d})$. More generally, for any measurable subset $u \subset [0, 1]$ we have

$$\int_u S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha = O\left(\|S_0\|_{\infty(u)} (\log N)^{n-1} N^{n-1-d}\right), \quad (2.2)$$

where $\|S_0\|_{\infty(u)}$ denotes the supremum of $|S_0(\alpha)|$ for $\alpha \in u$.

Then a partition into so-called major arcs $\mathfrak{M}(C)$ and minor arcs $\mathfrak{m}(C)$, depending on a parameter C (see Sect. 5 for the precise definitions), becomes useful due to the following fundamental estimate of Hua and Vinogradov

Lemma 1 [11] *Given $c > 0$, there exists a constant $C > 0$ such that $\|S_0\|_{\infty(\mathfrak{m}(C))} \leq N(\log N)^{-c}$.*

This, together with Eq. 2.2, in turn gives the bound

$$\int_{\mathfrak{m}(C)} S_0(\alpha) S_1(\alpha) S_2(\alpha) d\alpha = O\left((\log N)^{-1} N^{n-d}\right), \quad (2.3)$$

and one is left with the task of approximating the integral over the major arcs

$$\int_{\mathfrak{M}(C)} \sum_{\mathbf{x} \in [N]^n} e(\alpha \mathfrak{F}(\mathbf{x})) d\alpha.$$

Without going into the details here, let us remark that the major arcs consist of very small intervals (or boxes) centered at rational points with small denominators, and the integral essentially depends only on the distribution of the values of the polynomials $\mathfrak{F}(x)$ in small residue classes. Then the approximation can be done via standard methods in this area, similarly as in the diagonal case [11].

Now let us look at the case of a general form $\mathfrak{F}$ of degree 2. If we introduce a splitting of the variables $\mathbf{x} = (x_1, \mathbf{y}, \mathbf{z})$, we induce a decomposition of the shape

$$\mathfrak{F}(\mathbf{x}) = ax_1^2 + \mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z})x_1 + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}) + \mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z})$$

for a form $\mathfrak{g}^{(2)}$ which is bilinear in $\mathbf{y}$ and $\mathbf{z}$, and a linear form $\mathfrak{g}^{(1)}$. There are two possible approaches to adapting the above argument to this case.

The first involves a dichotomized argument based on the rank of $\mathfrak{g}^{(2)}$. If we have that $\mathfrak{g}^{(2)}$ has large rank, one can obtain good bounds on the exponential sum

$$\sum_{\mathbf{y} \in [N]^m} \sum_{\mathbf{z} \in [N]^{n-1-m}} \Lambda(\mathbf{y}) \Lambda(\mathbf{z}) e(\alpha (\mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}) + \mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z}) + x_1 \mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z})))$$

by simply removing the contribution of the von Mangoldt function with two applications of the Cauchy–Schwarz inequality. During this process the x_1 variable differences out, and the x_1 summation may be treated trivially. In this case the methods of Birch [2] are applicable (and the rank bounds are comparable, see [15]). If $\mathfrak{g}^{(2)}$ has small rank, then it must be the case that $\mathfrak{F}_1$ and $\mathfrak{F}_2$ each have large rank for appropriately chosen splitting of the variables. Write $\mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z}) = \langle \mathbf{y}, B\mathbf{z} \rangle$ for an appropriately sized matrix B whose rank is small, and split $\mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z}) = l_1(\mathbf{y}) + l_2(\mathbf{z})$. The above argument can then be run on the intersection of the level sets of $l_1(\mathbf{y})$, $l_2(\mathbf{z})$, and $B\mathbf{z}$, as the form $\mathfrak{F}(\mathbf{x})$ takes the diagonal shape given in (2.1), and both $\mathfrak{F}_1$ and $\mathfrak{F}_2$ have large rank on this small co-dimensional affine linear space. On such an intersection we get an extra power gain, which is equal to its codimension which then compensates for the loss of originally applying the Cauchy–Schwarz inequality on each level set. Thus summing the estimates over all such level sets gives an appropriate bound.

The second approach, to be fair, is simply a streamlined version of the first which removes the need for a dichotomized approach, and this is the one we

shall follow. The main requirement here is an appropriate decomposition of $\mathfrak{F}$ in the form

$$\mathfrak{F}(\mathbf{x}) = ax_1^2 + \mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z})x_1 + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}) + \mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z})$$

such that the rank of $\mathfrak{F}_1 + \mathfrak{g}^{(2)}$ is sufficiently large, the number of variables of composing $\mathbf{y}$, say m , is controlled, and the rank of $\mathfrak{F}_2$ is large with respect to m . As before, we wish to fix $l_1(\mathbf{y})$, $l_2(\mathbf{z})$, and $B\mathbf{z}$. The difference is that we have no assumption on the rank of B . However, by controlling the value of m , we have a way to control the number of linear equations in $\mathbf{z}$. Running the argument as before and summing over the level sets reduces the minor arcs estimates to providing an appropriate bound for the number of solutions to the system

$$\begin{aligned}\mathfrak{F}_1(\mathbf{y}) + \mathfrak{g}^{(2)}(\mathbf{y}, \mathbf{z}) &= \mathfrak{F}_1(\mathbf{y}') + \mathfrak{g}^{(2)}(\mathbf{y}', \mathbf{z}) \\ l_1(\mathbf{y}) &= l_1(\mathbf{y}') \\ \mathfrak{F}_2(\mathbf{z}) &= \mathfrak{F}_2(\mathbf{z}') \\ l_2(\mathbf{z}) &= l_2(\mathbf{z}') \\ B\mathbf{z} &= B\mathbf{z}'\end{aligned}\tag{2.4}$$

with $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathbf{z}, \mathbf{z}' \in [N]^{n-1-m}$. This is achieved by the rank assumptions of $\mathfrak{F}_1 + \mathfrak{g}^{(2)}$ and $\mathfrak{F}_2$ in the original decomposition.

The strategy for forms of higher degree starts by a similar decomposition of the form

$$\begin{aligned}\mathfrak{F}(\mathbf{x}) &= ax_1^d + \mathfrak{g}^{(1)}(\mathbf{y}, \mathbf{z})x_1^{d-1} + \cdots + \mathfrak{g}^{(d-1)}(\mathbf{y}, \mathbf{z})x_1 \\ &\quad + \mathfrak{F}_1(\mathbf{y}) + \mathfrak{F}_2(\mathbf{z}) + \mathfrak{g}^{(d)}(\mathbf{y}, \mathbf{z}),\end{aligned}$$

where the $\mathfrak{g}^{(i)}$ are forms of degree i , and $\mathfrak{F}_1(\mathbf{y})$, $\mathfrak{F}_2(\mathbf{z})$, $\mathfrak{g}^{(d)}$ are forms of degree d . Again we require that the rank of $\mathfrak{F}_1 + \mathfrak{g}^{(d)}$ is large with respect to $\mathfrak{g}^{(i)}$ for each $i < d$, the number of variables m composing $\mathbf{y}$ is small, and the rank of $\mathfrak{F}_2$ is large with respect to m . That such a decomposition is possible is the subject of Sect. 4. Then we view each form $\mathfrak{g}^{(i)}$, $1 \leq i \leq d$, as a sum of forms in $\mathbf{y}$ with coefficients that are forms in $\mathbf{z}$, and the number of these coefficients is bounded in terms of m and d . On each of the level sets of this new system of forms in $\mathbf{z}$ we have a system of forms in $\mathbf{y}$, the number of which is bounded in terms of d . Now passing to the further level of sets of these forms $\mathbf{y}$ provides a place to carry out the simple Cauchy–Schwarz argument at the beginning of this section. Summing back over the level sets then provides a system analogous to the one above.

The only problem with this so far is that the system we end up with contains at least a portion of each form $\mathfrak{g}^{(i)}(\mathbf{y}, \mathbf{z})$ for $i = 1, \dots, d$, and we have no

control on the rank of these forms at all and therefore have no way of dealing with the terminal system. The solution to this problem is found in the work of Schmidt [17]. His results provide a way of partitioning the level sets of a form by the level sets of a system of forms that does have high rank in each degree. Section 3 is dedicated to this. Working with this more regular system as opposed to the $g^{(i)}$'s does provide a manageable terminal system, and allows for a bound on the minor arc integral.

Extending this method to systems of forms is relatively straightforward at this point, and is of course carried out below. The major technical difficulty here is the need to isolate larger number of suitable variables $x_1, \dots, x_K$ to get the logarithmic gain on the minor arcs, as opposed to a randomly chosen single variable x_1 .

2.1 Outline and notation

The outline for the rest of the paper is as follows. Sections 3 and 4 are as described above. The completion of the bound for the integral over the minor arcs is going to be carried out in Sect. 5. The major arcs are dealt with in Sect. 6, where the asymptotic formula is shown. Section 7 is dedicated to the proof of Theorem 1. The final section concludes the work with a few further remarks.

Remarks on notation The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ denote the integers, the rational numbers, the real numbers, and the complex numbers, respectively. The r -dimensional flat torus $\mathbb{R}^r/\mathbb{Z}^r$ is denoted by $\mathbb{T}^r$. The p -adic integers are denoted by $\mathbb{Z}_p$, and the units of $\mathbb{Z}_p$ are denoted by $\mathbb{U}_p$. The symbol Z_N represents shorthand for the groups $\mathbb{Z}/N\mathbb{Z}$. Also, the shorthand for the multiplicative group reduced residue classes Z_N^* is U_N .

For a given measurable set $X \subseteq \mathbb{T}^r$ we shall use the notation $\|f\|_{p(X)}$ to denote the L^p norm of the function $\mathbf{1}_X f$ with the normalized Lebesgue measure on the r -dimensional flat torus. If X is omitted it is assumed that $X = \mathbb{T}^r$. Here, and in general, $\mathbf{1}_X$ denotes a characteristic function for X in a specified ambient space, and, on occasion, the set X is replaced by a conditional statement which defines it.

The Landau o and O notation is used throughout the work. The notation $f \lesssim g$ is sometimes used to replace $f = O(g)$. The implied constants are independent of N and $\mathbf{s}$, but may depend on all other parameters such as d , r , n , and p .

3 A regularity lemma

In [17], Schmidt provides an alternative definition of rank for a form. For a single form $\mathfrak{F}$ of degree at least 2 defined over a field k , define the *Schmidt rank* $h_k(\mathfrak{F})$ to be the minimum value of l such that there exists a decomposition

$$\mathfrak{F} = \sum_{i=1}^l U_i V_i,$$

where U_i and V_i are forms defined over k of degree at least one. For a system $\mathfrak{f}^{(d)} = (\mathfrak{f}_1^{(d)}, \dots, \mathfrak{f}_{r_d}^{(d)})$ of forms of degree d we define $h_k(\mathfrak{f})$ to be

$$\min \left\{ h_k \left(\lambda_1 \mathfrak{f}_1^{(d)} + \dots + \lambda_{r_d} \mathfrak{f}_{r_d}^{(d)} \right) : \lambda_i \neq 0 \text{ for some } i \right\}.$$

The following basic properties of the Schmidt rank will be used later.

- If $\mathfrak{f}$ is defined over a field k , and k' is an extension of k , then $h_{k'}(\mathfrak{f}) \leq h_k(\mathfrak{f})$
- The Schmidt rank is invariant under invertible linear transformations of k , i.e. $h_k(\mathfrak{f} \circ A) = h_k(\mathfrak{f})$ for $A \in GL_n(k)$.
- If $\mathfrak{f}'(x_2, \dots, x_n) = \mathfrak{f}(0, x_2, \dots, x_n)$, then $h_k(\mathfrak{F}') \geq h_k(\mathfrak{F}) - 1$.

The first two are clear from the definition, and the third simply follows from the fact that $\mathfrak{f}(\mathbf{x}) - \mathfrak{f}'(\mathbf{x})$ is of the form $x_1 \mathfrak{g}(\mathbf{x})$ for some $d-1$ degree system of forms $\mathfrak{g}$. Also, the second and third imply that the rank cannot drop on a codimension j subspace of k^n by more than j .

As observed by Schmidt (see [17], Lemma 16.1), the Birch rank $\mathcal{B}_d(\mathfrak{F})$ and the complex Schmidt rank $h_{\mathbb{C}}(\mathfrak{F})$ are essentially equivalent for a form of degree d , each being bounded by a constant times the other. For example, if $\mathcal{Q}(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle$ is a quadratic form, then $\mathcal{B}(\mathcal{Q})$ is $\text{rank}(A)$ (as pointed out in the opening section) and $h_{\mathbb{C}}(\mathcal{Q}) = l$ is the smallest integer greater than or equal to $\text{rank}(A)/2$ (which follows from the fact that $\mathcal{Q}$ is equivalent over $\mathbb{C}$ to the form $x_1^2 + \dots + x_l^2$ and $x_1^2 + x_2^2$ factors over $\mathbb{C}$). Of course the same phenomenon is true for systems as well. The rational Schmidt rank $h_{\mathbb{Q}}$ on the other hand is not equivalent to the Birch rank and we need the following result which is a weakened version of a central result in [17]. A definition is required.

Definition 1 Let $\mathfrak{p} = (\mathfrak{p}^{(d)}, \dots, \mathfrak{p}^{(1)})$ be a graded system of polynomials with rational coefficients, meaning $\mathfrak{p}^{(i)}$ is the subfamily of $\mathfrak{p}$ of polynomials of degree precisely i . Assume that $\mathfrak{p}^{(i)}$ consists of r_i polynomials for each $1 \leq i \leq d$, and set $D = \sum_i i r_i$.

The system $\mathfrak{p}$ is said to be regular if $|V_{\mathfrak{p},0} \cap [N]^n| = O(N^{n-D})$, as $N \rightarrow \infty$.

Theorem A (Schmidt [17]) *For a given positive integers R and d , there exists constants $\rho_i(R, d)$ for $2 \leq i \leq d$ such that the following holds:*

Let $\mathfrak{p} = (\mathfrak{p}^{(d)}, \dots, \mathfrak{p}^{(2)})$ be a graded system of rational polynomials and let $\mathfrak{f}^{(i)}$ be the system of forms consisting of the homogeneous parts of the polynomials $\mathfrak{p}^{(i)}$. Let $r_i = |\mathfrak{f}^{(i)}|$, $R = r_2 + \dots + r_d$ the total number of forms, and $D = 2r_2 + \dots + dr_d$ the total degree of the system. If we have $h_{\mathbb{Q}}(\mathfrak{f}^{(i)}) \geq \rho_i(R, d)$ for each i , then the system $\mathfrak{p}$ is regular.

Indeed this follows immediately for homogeneous systems, that is when $\mathfrak{p} = \mathfrak{f}$, from Theorem II, or alternatively from Propositions I, II and III in [17]. For non-homogeneous systems all propositions and hence Theorem II continues to hold, see Sec.9 and the last paragraph of Sec.10 in [17]. The constants $\rho_i(r, d)$ are given explicitly satisfying the bound $\rho_i(r, d) \leq Rr_i 2^{Cd \log d}$.

One of the key observations of Schmidt is that his definition of rank has a very nice reductive quality with respect to the degree, in the sense that forms of small rank may be replaced by a small number of forms of lesser degree. The next result captures this idea.

Proposition 1 (*Regularization of systems*). *Let $d > 1$ be a fixed integer, and let F be any collection of non-decreasing functions $F_i(R)$ for $i = 2, \dots, d$ mapping the nonnegative integers into themselves. For a collection of non-negative integers $r_1, \dots, r_d$, there exist constants*

$$C_1(r_1, \dots, r_d, F), \dots, C_d(r_1, \dots, r_d, F)$$

such that the following holds:

Given a system of integral forms $\mathfrak{f} = (\mathfrak{f}^{(d)}, \mathfrak{f}^{(d-1)}, \dots, \mathfrak{f}^{(1)})$, where each of the $\mathfrak{f}^{(i)}$ is a system of r_i forms of degree i , there exists a system of rational forms $\mathfrak{g} = (\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, \dots, \mathfrak{g}^{(1)})$ satisfying:

- (1) *Each form of the system $\mathfrak{f}$ can be written as a rational polynomial expression of the forms of the system $\mathfrak{g}$. In particular, the level sets of $\mathfrak{g}$ partition those of $\mathfrak{f}$.*
- (2) *The number of forms in each subsystem of $\mathfrak{g}^{(i)}$, say r'_i , is at most $C_i(r_1, \dots, r_d, F)$ for each $1 \leq i \leq d$.*
- (3) *The system $((\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, \dots, \mathfrak{g}^{(2)}))$ satisfies the rank condition $h_{\mathbb{Q}}(\mathfrak{g}^{(i)}) \geq F_i(R')$ for each $2 \leq i \leq d$, with $R' = r'_1 + \dots + r'_d$, moreover the system $\mathfrak{g}^{(1)}$ is a linearly independent family of linear forms.*

If a system $\mathfrak{g}$ satisfying properties (1) and (3) then we call it a *regularization* of the system $\mathfrak{f}$ with respect to the family of functions F .

It is worth noting that results of this type have been previously obtained over fields with positive characteristic. See [9] for such a result over fields with high characteristic, i.e. larger than d where d is as stated in the proposition. See [12] for a result in the low characteristic case.

Proof The proof is carried out by a double induction on the parameters. First for a fixed d we show that the case r_d with any choice of $r_{d-1}, \dots, r_1$ implies the similar scenario for the case $r_d + 1$. Then the induction on d is carried out.

The initial case we need to consider is $d = 2$ with a given function $F_2(R)$. Take a system of forms $\mathfrak{f} = (\mathfrak{f}^{(2)}, \mathfrak{f}^{(1)})$ with $r_2 = 1$ and any value of r_1 . If $h_{\mathbb{Q}}(\mathfrak{f}^{(2)}) \geq F_2(R)$, with $R = r_1 + 1$, then we may simply take $\mathfrak{g} = \mathfrak{f}$,

otherwise $\mathfrak{f}^{(2)} = \sum_{i=1}^l U_i V_i$ for some rational linear forms U_i and V_i where $l < F_2(R)$. We may then adjoin the linear forms $U_1, \dots, V_l$ to the system $\mathfrak{f}^{(1)}$ to obtain the system $\mathfrak{g}^{(1)}$, and let $\mathfrak{g}$ be a maximal linearly independent subsystem of $\mathfrak{g}^{(1)}$. Properties (1) and (2) are easily verified for this system, and property (3) is immediate.

Now for a fixed value of d assume that the result holds for all systems with maximal degree d for any given collection of functions F when $r_d = j$ and $r_{d-1}, \dots, r_1$ are arbitrary. Consider now a fixed collection of functions F and a system $\mathfrak{f} = (\mathfrak{f}^{(d)}, \dots, \mathfrak{f}^{(1)})$ with $r_d = j + 1$. Let $\mathfrak{f}'$ be the system $(\mathfrak{f}^{(d-1)}, \dots, \mathfrak{f}^{(1)})$. By the induction hypothesis, there is a system $\mathfrak{g}'$ of rational forms which is a regularization of $\mathfrak{f}'$ with respect to $F'_i(R) := F_i(R + (j + 1))$ for $i = 2, \dots, (d - 1)$.

Now let $\tilde{\mathfrak{g}}' = (\mathfrak{f}^{(d)}, \mathfrak{g}')$. If $\tilde{\mathfrak{g}}'$ fails to be the regularization of $\mathfrak{f}$ with respect to the family of functions F , then $h_{\mathbb{Q}}(\mathfrak{f}^{(d)}) < F_d(R_{\tilde{\mathfrak{g}}'} + (j + 1))$, where $R_{\tilde{\mathfrak{g}}'}$ is the number of forms of the system $\tilde{\mathfrak{g}}'$. As before, in this case there must exist homogeneous rational polynomials U_i and V_i , $i = 1, \dots, l < F_d(R_{\tilde{\mathfrak{g}}'} + (j + 1))$, such that

$$\lambda_1 \mathfrak{f}_1^{(d)} + \dots \lambda_{j+1} \mathfrak{f}_{j+1}^{(d)} = \sum_{i=1}^l U_i V_i,$$

where without loss of generality we may assume that $\lambda_{j+1} \neq 0$. Now let $\mathfrak{g}''$ be $\mathfrak{g}'$ adjoined with the those forms U_i and V_i which are not linear combinations of forms already in $\mathfrak{g}'$, and set

$$\tilde{\mathfrak{g}}'' = \left(\left(\mathfrak{f}_1^{(d)}, \dots, \mathfrak{f}_j^{(d)} \right), \mathfrak{g}'' \right).$$

By the induction hypothesis there is a system $\mathfrak{g}$ which is the regularization of $\tilde{\mathfrak{g}}''$ with respect to initial collection of functions F . It is clear from the construction of the system $\mathfrak{g}$ that all forms of the initial system are polynomial expressions of the ones in $\mathfrak{g}$, and the number of forms in each subsystem $\mathfrak{g}^{(i)}$ is expressible in terms of $r_1, \dots, r_d$, and d . Thus the system $\mathfrak{g}$ is the regularization of $\mathfrak{f}$ satisfying conditions (1), (2) and (3). The induction argument to go from d to $d + 1$ is simply the above argument carried out with $j = 0$. $\square$

We will need a stronger version of the above proposition relative to a partition of the variables $\mathbf{x} = (\mathbf{y}, \mathbf{z})$. To state it let us introduce a modified version of the Schmidt rank. For a single form $\mathfrak{F}(\mathbf{y}, \mathbf{z})$ of degree at least 2 defined over a field k , we define its *Schmidt rank* with respect to the variables $\mathbf{z}$, $h_k(\mathfrak{F}; \mathbf{z})$ to be the minimum value of l such that there exists a decomposition

$$\mathfrak{F}(\mathbf{y}, \mathbf{z}) = \sum_{i=1}^l U_i(\mathbf{y}, \mathbf{z}) V_i(\mathbf{y}, \mathbf{z}) + W(\mathbf{z}),$$

where U_i and V_i are forms defined over k of degree at least one. For a system $\mathfrak{f}^{(d)} = (\mathfrak{f}_1^{(d)}, \dots, \mathfrak{f}_{r_d}^{(d)})$ of forms of degree d we define $h_k(\mathfrak{f}; \mathbf{z})$ to be

$$\min \left\{ h_k \left(\lambda_1 \mathfrak{f}_1^{(d)} + \dots + \lambda_{r_d} \mathfrak{f}_{r_d}^{(d)}; \mathbf{z} \right) : \lambda_i \neq 0 \text{ for some } i \right\},$$

and set $h_k(\mathfrak{f}; \mathbf{z}) = \infty$ if $\mathfrak{f} = \emptyset$. Note that $h_k(\mathfrak{f}; \mathbf{z}) \leq h_k(\mathfrak{f})$ and $h_k(\mathfrak{f}; \mathbf{z}) = 0$ if and only if a non-trivial linear combination of the forms in $\mathfrak{f}$ depends only on the variables $\mathbf{z}$.

Proposition 1' (*Regularization of systems, parametric version*). *Let $d > 1$ be a fixed integer, and let F be any collection of non-decreasing functions $F_i(R)$ for $i = 2, \dots, d$ mapping the nonnegative integers into themselves. For a collection of non-negative integers $r_1, \dots, r_d$, there exist constants*

$$C_1(r_1, \dots, r_d, F), \dots, C_d(r_1, \dots, r_d, F)$$

such that the following holds:

Given a system of integral forms $\mathfrak{f} = (\mathfrak{f}^{(d)}, \mathfrak{f}^{(d-1)}, \dots, \mathfrak{f}^{(1)})$ where each of the $\mathfrak{f}^{(i)}$ is a system of r_i forms of degree i and a partition of the variables $\mathbf{x} = (\mathbf{y}, \mathbf{z})$, there exists a system of rational forms $\mathfrak{g} = (\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, \dots, \mathfrak{g}^{(1)})$ satisfying:

- (1) *Each form of the system $\mathfrak{f}$ can be written as a rational polynomial expression of the forms of the system $\mathfrak{g}$. In particular, the level sets of $\mathfrak{g}$ partition those of $\mathfrak{f}$.*
- (2) *The number of forms in each subsystem of $\mathfrak{g}^{(i)}$, say r'_i , is at most $C_i(r_1, \dots, r_d, F)$ for each $1 \leq i \leq d$.*
- (3) *The system $(\mathfrak{g}^{(d)}, \mathfrak{g}^{(d-1)}, \dots, \mathfrak{g}^{(2)})$ satisfies the rank condition $h_{\mathbb{Q}}(\mathfrak{g}^{(i)}) \geq F_i(R')$ for each $2 \leq i \leq d$, with $R' = r'_1 + \dots + r'_d$, moreover the system $\mathfrak{g}^{(1)}$ is a linearly independent family of linear forms.*
- (4) *The system $(\bar{\mathfrak{g}}^{(d)}, \bar{\mathfrak{g}}^{(d-1)}, \dots, \bar{\mathfrak{g}}^{(2)})$ satisfies the modified rank condition $h_{\mathbb{Q}}(\bar{\mathfrak{g}}^{(i)}; \mathbf{z}) \geq F_i(R')$ for each $2 \leq i \leq d$, where the subsystem $\bar{\mathfrak{g}}^{(i)}$ is obtained from the system $\mathfrak{g}^{(i)}$ by removing the forms depending only on the $\mathbf{z}$ variables.*

We call a system $\mathfrak{g}$ satisfying the conclusions of Proposition 1' to be a strong regularization of the system $\mathfrak{f}$ with respect to the variables $\mathbf{z}$.

Proof The argument is a slight modification of the proof of Proposition 1. To a given system $\mathfrak{f} = (\mathfrak{f}^{(d)}, \mathfrak{f}^{(d-1)}, \dots, \mathfrak{f}^{(1)})$ we assign its index which is the triple

(d, r_d, r'_d) , where $r_d = |\mathfrak{f}^{(d)}|$ resp. $\bar{r}_d = |\bar{\mathfrak{f}}^{(d)}|$ are the number of degree d forms resp. the number of degree d forms depending on at least one of the $\mathbf{y}$ variables. We will proceed via induction on the lexicographic ordering of the indexes. To be precise we deem $(d, r_d, \bar{r}_d) < (e, r_e, \bar{r}_e)$ if $d < e$, $d = e$ but $r_d < r_e$, or $d = e$, $r_d = r_e$ but $\bar{r}_d < \bar{r}_e$.

If $d = 1$ all one needs to do is to choose maximal linearly independent subsystem of $\mathfrak{f} = \mathfrak{f}^{(1)}$. So assume $d \geq 2$ (and $r_d \geq 1$), $I = (d, r_d, \bar{r}_d)$ is a fixed index and the result holds for any system with index $I' < I$. Again, let $\mathfrak{f}' = (\mathfrak{f}^{(d-1)}, \dots, \mathfrak{f}^{(1)})$ and let $\mathfrak{g}'$ be a strong regularization of $\mathfrak{f}'$ with respect to the functions $F'_i(R) = F_i(R + r_d)$ ($i = 2, \dots, d-1$) and the variables $\mathbf{z}$. Let $\tilde{\mathfrak{g}}' = (\mathfrak{f}^{(d)}, \mathfrak{g}')$. If $\tilde{\mathfrak{g}}'$ fails to be a strong regularization of $\mathfrak{f}$ with respect to the family of functions F and the variables $\mathbf{z}$, then there are two possible cases.

Either $h_{\mathbb{Q}}(\mathfrak{f}^{(d)}) < F_d(R_{\tilde{\mathfrak{g}}'} + r_d)$, where $R_{\tilde{\mathfrak{g}}'}$ is the number of forms of the system $\tilde{\mathfrak{g}}'$. As before, in this case one can replace the system $\tilde{\mathfrak{g}}'$ with a system $\tilde{\mathfrak{g}}''$ of index $I'' < I$. Then by the induction hypotheses there is a system $\mathfrak{g}$ which is a strong regularization of $\tilde{\mathfrak{g}}''$ with respect to initial collection of functions F and the variables $\mathbf{z}$. The system $\mathfrak{g}$ will satisfy the claims of the proposition.

Otherwise $\bar{r}_d \geq 1$ and we have $h_{\mathbb{Q}}(\bar{\mathfrak{f}}^{(d)}; \mathbf{z}) < F_d(R_{\tilde{\mathfrak{g}}'} + r_d)$, hence there must exist homogeneous rational polynomials U_i and V_i , $i = 1, \dots, l < F_d(R_{\tilde{\mathfrak{g}}'} + r_d)$, and a form $Q(\mathbf{z})$ of degree d such that

$$\lambda_1 \mathfrak{f}_1^{(d)} + \dots + \lambda_{\bar{r}_d} \mathfrak{f}_{\bar{r}_d}^{(d)} = \sum_{i \leq l} U_i(\mathbf{y}, \mathbf{z}) V_i(\mathbf{y}, \mathbf{z}) + Q(\mathbf{z}), \quad \left(\bar{\mathfrak{f}}^{(d)} = \left(\mathfrak{f}_1^{(d)}, \dots, \mathfrak{f}_{\bar{r}_d}^{(d)} \right) \right)$$

where without loss of generality we may assume that $\lambda_{\bar{r}_d} \neq 0$. Now let $\mathfrak{g}''$ be $\mathfrak{g}'$ adjoined with the those forms U_i and V_i which are not linear combinations of forms already in $\mathfrak{g}'$, and set

$$\tilde{\mathfrak{g}}'' = \left(\left(\mathfrak{f}^{(d)} \setminus \mathfrak{f}_{\bar{r}_d}^{(d)}, Q \right), \mathfrak{g}'' \right).$$

Note that we have replaced the form $\mathfrak{f}_{\bar{r}_d}^{(d)}(\mathbf{y}, \mathbf{z})$ with the form $Q(\mathbf{z})$ hence the index of $\tilde{\mathfrak{g}}''$ is the triple $I'' = (d, r_d, \bar{r}_d - 1) < I$. Again, there is a system $\mathfrak{g}$ which is a strong regularization of $\tilde{\mathfrak{g}}''$ with respect to initial collection of functions F and the variables $\mathbf{z}$. As explained before the system $\mathfrak{g}$ satisfies the claims of the proposition. Note that the procedure depends on the partition $\mathbf{x} = (\mathbf{y}, \mathbf{z})$, however since there only finitely many ways to partition the variables $\mathbf{x}$, the constants $C_i(r_1, \dots, r_d, F)$ can be taken independent of the partition. $\square$

Applying Proposition 1' with the functions being given by the values of the Schmidt constants $\rho_i(R, d)$ then provides the following.

Corollary 3 Let $\mathfrak{f} = (\mathfrak{f}^{(d)}, \dots, \mathfrak{f}^{(2)})$ be given system of rational forms with r_i forms of degree i composing each subsystem $\mathfrak{f}^{(i)}$ and let $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ be a partition of the independent variables. There exists a regular system of forms $\mathfrak{g}$ satisfying the conclusions of Proposition 1'.

Proof For fixed d , let $\mathfrak{g}$ be a strong regularization of the system $\mathfrak{f}$ with respect to the functions $F_i(R) := \rho_i(R, d) + R$ and the variables $\mathbf{z}$. We show that $|V_{\mathfrak{g}, \mathbf{0}} \cap [N]^n| = O(N^{n-R})$.

Let $H := V_{\mathfrak{g}^{(1)}, \mathbf{0}}$ be the nullset of the system $\mathfrak{g}^{(1)}$, and let $A \in GL_n(\mathbb{Q})$ be a linear transformation such that $M := A(H)$ is coordinate subspace of codimension r_1 . It is easy to see that $A([N]^n) \subseteq K_1^{-1} \cdot [-K_2 N, K_2 N]^n$ for some constants K_1 and K_2 , where by $\lambda \cdot \mathbf{x}$ we mean a dilation of the point $\mathbf{x}$ by a factor λ . Let $\mathfrak{g}' = \mathfrak{g} \circ A^{-1}$, then $V_{\mathfrak{g}', \mathbf{0}} = A(V_{\mathfrak{g}, \mathbf{0}})$, thus by homogeneity we have that

$$|V_{\mathfrak{g}, \mathbf{0}} \cap [N]^n| \leq |V_{\mathfrak{g}', \mathbf{0}} \cap [-KN, KN]^n|, \quad (3.1)$$

with $K = K_1 K_2$. We have that $h_{\mathbb{Q}}(\mathfrak{g}'^{(i)}) = h_{\mathbb{Q}}(\mathfrak{g}^{(i)}) \geq F_i(R)$, and hence $h_{\mathbb{Q}}(\mathfrak{g}'^{(i)}|_M) \geq \rho_i(R, d)$, for $i = 2, \dots, d$, where R denotes the total number forms of $\mathfrak{g}'$. The quantity on the right side of (3.1) is the number of integral points $\mathbf{x}$ in $M \cap [-KN, KN]^n = [-KN, KN]^{n-r_1}$ such that $\mathfrak{g}'^{(i)}(\mathbf{x}) = 0$ for $i = 2, \dots, d$. Since the system $(\mathfrak{g}'^{(d)}, \dots, \mathfrak{g}'^{(2)})$ satisfies the conditions of Theorem A, we have that $|V_{\mathfrak{g}', \mathbf{0}} \cap [-KN, KN]^n| = O(N^{n-R})$. This proves the corollary. $\square$

The somewhat technical last conclusion in Proposition 1' is utilized in the following lemma which will play an important role in our minor arcs estimates.

Lemma 2 Let $\mathfrak{g}^{(k)}(\mathbf{y}, \mathbf{z}) = (\mathfrak{g}_1^{(k)}(\mathbf{y}, \mathbf{z}), \dots, \mathfrak{g}_s^{(k)}(\mathbf{y}, \mathbf{z}))$ be system of homogeneous forms of degree k . Then for the system $\tilde{\mathfrak{g}}^{(k)}(\mathbf{y}, \mathbf{y}', \mathbf{z}) = (\mathfrak{g}^{(k)}(\mathbf{y}, \mathbf{z}), \mathfrak{g}^{(k)}(\mathbf{y}', \mathbf{z}))$ we have that

$$h_{\mathbb{Q}}(\tilde{\mathfrak{g}}^{(k)}; \mathbf{z}) = h_{\mathbb{Q}}(\mathfrak{g}^{(k)}; \mathbf{z}).$$

Here $\mathbf{y}$, $\mathbf{y}'$ and $\mathbf{z}$ represent distinct sets of variables.

Proof Since $\mathfrak{g}^{(k)}$ is a subsystem of $\tilde{\mathfrak{g}}^{(k)}$ it is clear that $h_{\mathbb{Q}}(\tilde{\mathfrak{g}}^{(k)}; \mathbf{z}) \leq h_{\mathbb{Q}}(\mathfrak{g}^{(k)}; \mathbf{z})$. Now let λ, μ be s -tuples of rational numbers not all 0. Using the short hand notation $\lambda \cdot \mathfrak{g}^{(k)} = \sum_i \lambda_i \mathfrak{g}_i^{(k)}$ assume that we have a decomposition

$$\lambda \cdot \mathfrak{g}^{(k)}(\mathbf{y}, \mathbf{z}) + \mu \cdot \mathfrak{g}^{(k)}(\mathbf{y}', \mathbf{z}) = \sum_{i=1}^h U_i(\mathbf{y}, \mathbf{y}', \mathbf{z}) V_i(\mathbf{y}, \mathbf{y}', \mathbf{z}) + Q(\mathbf{z}), \quad (3.2)$$

where all forms U_i, V_i have positive degree. We argue that $h \geq h_{\mathbb{Q}}(\mathfrak{g}^{(k)})$. Assuming without loss of generality that $\lambda \neq \mathbf{0}$ and substituting $\mathbf{y}' = \mathbf{0}$ into (3.2) we have

$$\lambda \cdot \mathfrak{g}^{(k)}(\mathbf{y}, \mathbf{z}) = \sum_{i=1}^h U_i(\mathbf{y}, \mathbf{0}, \mathbf{z}) V_i(\mathbf{y}, \mathbf{0}, \mathbf{z}) + Q'(\mathbf{z}),$$

with $Q'(\mathbf{z}) = Q(\mathbf{z}) - \mu \cdot \mathfrak{g}^{(k)}(\mathbf{0}, \mathbf{z})$. For each $1 \leq i \leq h$ the term $U_i(\mathbf{y}, \mathbf{0}, \mathbf{z}) V_i(\mathbf{y}, \mathbf{0}, \mathbf{z})$ vanishes identically or both forms $U_i(\mathbf{y}, \mathbf{0}, \mathbf{z})$ and $V_i(\mathbf{y}, \mathbf{0}, \mathbf{z})$ have positive degrees, thus by the definition of the modified Schmidt rank we have that $h \geq h_{\mathbb{Q}}(\mathfrak{g}^{(k)}; \mathbf{z})$. $\square$

4 A decomposition of forms

For $I \subset [n]$, let $\mathbf{y} = (y_i)_{i \in I}$ be the vector with components $y_i = x_i$ for $i \in I$. Also let $\mathbf{z}$ be the vector defined similarly for the set $[n] \setminus I$. Note that $(\mathbf{y}, \mathbf{z}) = \mathbf{x}$. To such a partition of the variables there corresponds a unique decomposition of a form $\mathfrak{f}$

$$\mathfrak{f}(\mathbf{x}) = \mathfrak{f}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z}) + \mathfrak{f}_2(\mathbf{z}),$$

where $\mathfrak{f}_1(\mathbf{y})$ is the sum of all monomials in $\mathfrak{f}$ which depends only on $\mathbf{y}$ variables, $\mathfrak{g}(\mathbf{y}, \mathbf{z})$ is the sum of the monomials depending both on $\mathbf{y}$ and $\mathbf{z}$, and $\mathfrak{f}_2(\mathbf{z})$ is the sum of the remaining terms. The decomposition extends to a system $\mathfrak{f}$ in the obvious manner. The aim of this section is to prove the following result.

Proposition 2 *Let positive integers C_1 and C_2 be given. Let $\mathfrak{f}$ be given system of r rational forms with $\mathcal{B}_d(\mathfrak{f})$ sufficiently large with respect to C_1, C_2, r and $d \geq 1$. There exists an $I \subset [n]$ such that $|I| \leq C_1 r$ and the associated decomposition*

$$\mathfrak{f}(\mathbf{x}) = \mathfrak{f}_1(\mathbf{y}) + \mathfrak{f}_2(\mathbf{z}) + \mathfrak{g}(\mathbf{y}, \mathbf{z})$$

satisfies $\mathcal{B}_d(\mathfrak{f}_1 + \mathfrak{g}) \geq C_1$ and $\mathcal{B}_d(\mathfrak{f}_2) \geq C_2$.

The proof of this result is carried out at the end of this section, and is done so by dealing directly with the Jacobian matrices. Some notation is helpful. Let $M = M(\mathbf{x})$ be a $i \times j$ matrix whose entries depend on $\mathbf{x}$. The notation $M \preceq M'$ is used to imply that M is a submatrix of M' obtained by the deletion of columns, so that M is an $i \times j'$ matrix with $j' \leq j$. Let V_M^* be the collection of $\mathbf{x}$ where $M = M(\mathbf{x})$ has rank strictly less than i . Clearly one has that if $M \preceq M'$, then $V_{M'}^* \subseteq V_M^*$.

Lemma 3 *If $\mathfrak{f}$ is a system of r integral forms of degree $d > 1$ in n variables, then the restriction of $\mathfrak{f}$ to the hyperplane defined by $x_n = 0$ has rank at least $\mathcal{B}_d(\mathfrak{f}) - r$. In the case $d = 1$ we have the improved lower bound $\mathcal{B}_1(\mathfrak{f}) - 1$.*

Proof First we consider the case $d > 1$. Denote the restriction of $\mathfrak{f}$ to the subspace defined by $x_n = 0$ as $\mathcal{F}$. The matrix $Jac_{\mathcal{F}}$ is then the matrix $Jac_{\mathfrak{f}}$ with the last column deleted and restricted to the space $x_n = 0$. It follows that $V_{\mathcal{F}}^* \cap H \subseteq V_{\mathfrak{f}}^* \cap \{x_n = 0\}$, where H denotes the variety where the last column of $Jac_{\mathfrak{f}}$ has all entries equal to zero. As H is defined by at most r equations, it has co-dimension at most r ([10], Ch. 7). In turn it follows from the sub-additivity of the dimension of intersection of varieties that $\mathcal{B}_d(\mathcal{F}) + r \geq \mathcal{B}_d(\mathfrak{f})$.

Now let $\mathfrak{f} = (f_1, \dots, f_r)$ be a system of r linear forms, and $\mathcal{F} = (\mathcal{F}_1, \dots, \mathcal{F}_r)$ be the system of forms obtained by setting $x_n = 0$. We have for any $\lambda_1, \dots, \lambda_r$

$$\lambda_1 f(\mathbf{x}) + \dots + \lambda_r f_r(\mathbf{x}) = \lambda_1 \mathcal{F}(\mathbf{x}) + \dots + \lambda_r \mathcal{F}_r(\mathbf{x}) + (\lambda_1 a_{1,n} + \dots + \lambda_r a_{r,n}) x_n,$$

where $a_{i,n}$ is the x_n coefficient in $f_i(\mathbf{x})$. From this the result is clear from the definition of $\mathcal{B}_1$. $\square$

Note that the Birch rank of a system of forms $\mathcal{F}$ is invariant under invertible linear transformations via the multivariate chain rule, and this fact remains true for a non-homogeneous system, as its rank is defined to be the rank of its homogeneous part. Thus we have more generally

Corollary 4 *If H is an affine linear space of co-dimension m , then the restriction of $\mathfrak{f}$ to H has rank at least $\mathcal{B}_d(\mathfrak{f}) - mr$ when $d > 1$. In the case $d = 1$ we have the improved lower bound $\mathcal{B}_1(\mathfrak{f}) - m$.*

Now define $\mathcal{C}_{\mathfrak{f}}(k)$ to be the minimal value of m such that there exists a matrix valued function $M \preceq Jac_{\mathfrak{f}}$ of size $r \times m$ such that V_M^* has dimension at most $n - k$. This is defined to be infinite if no such value exists. Thus $\mathcal{C}_{\mathfrak{f}}(k) \leq l$ if one can select l columns of the Jacobian matrix $Jac_{\mathfrak{f}}$ forming a matrix M so that V_M^* has co-dimension at least k .

Lemma 4 *For a system $\mathfrak{f}$ of r forms of degree $d > 1$, one has that $\mathcal{C}_{\mathfrak{f}}(k) \leq kr$ as long as $(k - 1)((d - 1)r)^{k-1} < \mathcal{B}(\mathfrak{f})$.*

Proof Write the singular variety $V_{\mathfrak{f}}^*$ as an intersection of varieties V_I , where V_I is the zero set of the determinant of the $r \times r$ minor coming from the selecting the columns $I = \{i_1, \dots, i_r\} \subset [n]$. Let $\mathfrak{I}_{n,r}$ denote the family of r -element subsets of $[n]$.

Proceeding inductively, we show that one can select index sets $I_1, \dots, I_l$, $I_j \in \mathfrak{I}_{n,r}$, so that $\text{codim} \left(\bigcap_{j=1}^l V_{I_j} \right) = l$, as long as $\mathcal{B}_d(\mathfrak{f})$ is sufficiently

large with respect to l, d and r . Indeed, assume that this is true for a given l and let $V^{(l)} := \bigcap_{j=1}^l V_{I_j}$. The degree of each hypersurface V_I is at most $r(d-1)$, and hence by the basic properties of the degree, see [10], Ch.7, $V^{(l)}$ has at most $((d-1)r)^l$ irreducible components. Label the components with dimension precisely $n-l$ as $Y_1, \dots, Y_J$, where $J \leq ((d-1)r)^l$.

For each Y_j , set $N(j)$ to be the set of $I \subseteq \mathcal{I}_{n,r}$ such that $Y_j \subseteq V_I$. If there is an index set I such that $I \notin \bigcup_{j=1}^J N(j)$, then $\dim(V_I \cap V^{(l)}) = n-l-1$, and one can choose $I_{l+1} = I$. Otherwise it follows that $\bigcup_{j=1}^J N(j) = \mathcal{I}_{n,r}$. In this case, we have $V_f^* = \bigcap_{j=1}^J (\bigcap_{I \in N(j)} V_I)$, and in turn

$$\operatorname{codim}(V_f^*) \leq \sum_{j=1}^J \operatorname{codim} \left(\bigcap_{I \in N(j)} V_I \right) \leq \sum_{j=1}^J \operatorname{codim}(Y_j) \leq l((d-1)r)^l,$$

which cannot happen if $l((d-1)r)^l < \mathcal{B}_d(f)$. Thus one can choose a I_{l+1} such that $V^{(l+1)}$ has co-dimension $l+1$. Now let $I' := \bigcup_{j=1}^l I_j$ and let $M = M_{I'}$ be the associated submatrix of Jac_f obtained by selecting the columns belonging to I' . It is clear that $V_M^* \subseteq \bigcap_{j=1}^l V_{I_j}$ and hence $\operatorname{codim}(V_M^*) \geq l$ while $|I'| \leq lr$. $\square$

Proof of Proposition 2 The proof is carried out separately for the cases $d = 1$ and $d > 1$, starting with the latter. We have $\mathcal{B}_d(f) = \mathcal{B}_d(f^{(d)})$ so we may assume that f consists of forms of degree d . Start by applying Lemma 4 with $k = C_1$, valid by assuming that $\mathcal{B}_d(f)$ is sufficiently large. Then there are at most $C_1 r$ columns of Jac_f providing a sub-matrix $M \preccurlyeq Jac_f$ with $\operatorname{codim}(V_M^*) \geq C_1$. Let I denote the collection of the indices of these columns, noting that $m := |I| \leq C_1 r$. Let $\mathbf{y} = (x_i)_{i \in I}$ and $\mathbf{z} = (x_i)_{i \notin I}$ and $f(\mathbf{x}) = f(\mathbf{y}, \mathbf{z}) = f_1(\mathbf{y}) + g(\mathbf{y}, \mathbf{z}) + f_2(\mathbf{z})$ be the associated decomposition of the system f . It is easily seen that $M \preccurlyeq Jac_{f_1+g}$, and it follows that $V_{f_1+g}^* \subseteq V_M^*$, and hence $\mathcal{B}(f_1 + g) \geq C_1$.

Now look at the the matrix $W = W(\mathbf{y}, \mathbf{z})$ obtained by deleting the columns of M from Jac_f . One now has $Jac_{f_2}(\mathbf{z}) = W(\mathbf{0}, \mathbf{z})$ as $\partial f / \partial \mathbf{z}(\mathbf{0}, \mathbf{z}) = \partial f_2 / \partial \mathbf{z}(\mathbf{z})$. Let H denote the variety of points $\mathbf{z}$ so that $M(\mathbf{0}, \mathbf{z}) = \mathbf{0}$, then we have that $V_{f_2}^* \cap H \subseteq \left\{ \mathbf{z}; (\mathbf{0}, \mathbf{z}) \in V_f^* \right\}$ hence

$$\operatorname{codim}(V_{f_2}^*) + C_1 r^2 \geq \mathcal{B}_d(f) - C_1 r,$$

as $\operatorname{codim}(H) \leq C_1 r^2$ and $m \leq C_1 r$. Thus assuming $\mathcal{B}_d(f) \geq C_2 + C_1(r^2 + r)$ provides the bound $\mathcal{B}_d(f_2) \geq C_2$ as required.

When $d = 1$ write the linear system f as $A\mathbf{x}$ for some integral matrix A of size $r \times n$. Note that the term g does not appear in this case, and we are simply

looking for a block decomposition $A = [A_1|A_2]$, where each A_i is an $r \times n_i$ matrix with $n_1 + n_2 = n$. If the linear system has $\mathcal{B}_1(\mathfrak{f})$ at least $C_1r + C_2$ then there is an $r \times r$ submatrix B_1 which has full rank. Now write $A = [B_1|A']$. By Corollary 4 we have that the linear system defined by A' in the variables $(x_{r+1}, \dots, x_n)$ has rank at least $(C_1 - 1)r + C_2$. Repeating this process gives a block decomposition $A = [B_1|B_2|\dots|B_{C_1}|A_2]$ where each B_i is $r \times r$ full rank matrix. We now set $A_1 := [B_1|\dots|B_{C_1}]$. As each B_i has full rank, it follows the rank of the linear system defined by A_1 in the variables $x_1, \dots, x_{rC_1}$ has rank at least C_1 because any non-trivial linear combination of the associated forms must contain at least one non-zero coefficient for one variable from each of the sets $\{x_{jr+1}, \dots, x_{(j+1)r}\}$ for $0 \leq j \leq C_1 - 1$. Applying Corollary 4 shows that the system defined by A_2 in the remaining variables has rank at least C_2 . $\square$

5 The minor arcs

Assume now throughout this section that we have a fixed system of integral polynomials $\mathfrak{p} = (\mathfrak{p}_1, \dots, \mathfrak{p}_r)$, where each $\mathfrak{p}_i$ is of degree d . The system $\mathfrak{f}$ is again the highest degree homogenous parts of $\mathfrak{p}$.

For a given value of $C > 0$ and an integer $q \leq (\log N)^C$, define a major arc

$$\mathfrak{M}_{\mathbf{a},q}(C) = \left\{ \alpha \in [0, 1]^r : \max_{1 \leq i \leq r} |\alpha_i - a_i/q| \leq N^{-d}(\log N)^C \right\}$$

for each $\mathbf{a} = (a_1, \dots, a_r) \in U_q^r$. When $q = 1$ it is to be understood that $U_1 = \{0\}$. These arcs are disjoint, and the union

$$\bigcup_{q \leq (\log N)^C} \bigcup_{\mathbf{a} \in U_q^r} \mathfrak{M}_{\mathbf{a},q}(C)$$

defines the major arcs $\mathfrak{M}(C)$. The minor arcs are then given by

$$\mathfrak{m}(C) = [0, 1]^r \setminus \mathfrak{M}(C).$$

The main result in this section is to deal with the integral representation on the minor arcs.

Proposition 3 *There exists constant $\chi(r, d)$ such that if we have $\mathcal{B}_d(\mathfrak{p}) \geq \chi(r, d)$, then the following holds. To any given value $c > 0$ there exists a $C > 0$ such that*

$$\int_{\mathfrak{m}(C)} e(-\mathbf{s} \cdot \alpha) \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) e(\mathbf{p}(\mathbf{x}) \cdot \alpha) d\alpha = O\left(N^{n-D} (\log N)^{-c}\right), \quad (5.1)$$

with an implied constant independent of $\mathbf{s}$.

Another set of minor arcs is also required for an exponential sum estimate. For each $1 \leq i \leq d$, define for $a^{(i)} \in U_q$ the major arc

$$\mathfrak{N}_{a^{(i)},q}^{(i)}(C) = \left\{ \mathbf{x} i^{(i)} \in \mathbb{T} : \left| \mathbf{x} i^{(i)} - \frac{a^{(i)}}{q} \right| \leq N^{-i} (\log N)^C \right\}.$$

Set

$$\mathfrak{N}_{\mathbf{a},q}(C) = \mathfrak{N}_{a^{(d)},q}^{(d)}(C) \times \dots \times \mathfrak{N}_{a^{(1)},q}^{(1)}(C),$$

where $\mathbf{a} = a^{(d)} \times \dots \times a^{(1)}$. The major arcs are now

$$\mathfrak{N}(C) = \bigcup_{q \leq (\log N)^C} \bigcup_{\mathbf{a} \in U_q^d} \mathfrak{N}_{\mathbf{a},q}(C).$$

Let $\mathfrak{n}(C)$ denote a set of minor arcs $\mathfrak{n}(C) = [0, 1]^d \setminus \mathfrak{N}(C)$.

Define the exponential sum

$$S_0(\beta) = \sum_{x \in [N]} \Lambda(x) e\left(\beta_d x^d + \dots + \beta_1 x\right)$$

for $(\beta_d, \dots, \beta_1) \in \mathbb{T}^d$.

Lemma 5 (Hua) *Given $c > 0$, there exists a C such that $\|S_0\|_{\infty(\mathfrak{n}(C))} = O(N(\log N)^{-c})$.*

For a proof the reader is referred to [11] (Ch. 10, §5, Lemma 10.8).

Proof of Proposition 3 Our first goal is to pick an appropriate splitting of the variables of the form $\mathbf{x} = (x_1, \dots, x_K, \mathbf{y}, \mathbf{z})$ which then induces a decomposition of the form

$$\mathbf{p}(\mathbf{x}) = \mathbf{p}_0(x_1, \dots, x_K, \mathbf{y}, \mathbf{z}) + \mathbf{p}_1(\mathbf{y}) + \mathbf{g}(\mathbf{y}, \mathbf{z}) + \mathbf{p}_2(\mathbf{z})$$

such that three things hold: First we need $x_1, \dots, x_K$ selected which are useful for applying Lemma 5. Secondly we need $\mathbf{y}$ consisting of m variables chosen so that $\mathbf{p}_1 + \mathbf{g}$ has large rank with respect to K . Lastly we need the rank of

p_2 to be large with respect to K and m . The last two conditions are going to be achieved by an application of Proposition 2 assuming that the rank of $\mathfrak{p}$ is sufficiently large.

We select the variables $x_1, \dots, x_K$ first; we consider the associated system of forms $\mathfrak{f}$ of degree d . We collect the r coefficients of each term $x_{i_1} \dots x_{i_d}$ into a vector $\mathbf{b}_{i_1, \dots, i_d}$. We select r of these which are linearly independent, this is possible as the system $\mathfrak{f}$ is linearly independent. The total number of indices involved is our value of K , in this choice is at most dr , and we assume that the corresponding variables to be the first $1 \leq i \leq K$. The variables $x_1, \dots, x_K$ have now been selected.

For any choice of $\mathbf{y}$ and $\mathbf{z}$, i.e. any generic splitting of the variables $(x_{K+1}, \dots, x_n)$, we have a decomposition of the shape

$$\begin{aligned} \mathfrak{p}(x_1, \dots, x_K, \mathbf{y}, \mathbf{z}) &= \mathfrak{p}(x_1, \dots, x_K, 0, \dots, 0) \\ &+ \sum_{j=1}^{d-1} \sum_{1 \leq i_1 < \dots < i_j \leq K} \left(\sum_{k=1}^{d-j} \mathfrak{G}_{i_1, \dots, i_j}^{(k)}(\mathbf{y}, \mathbf{z}) \right) x_{i_1} \dots x_{i_j} \\ &+ \sum_{\kappa=1}^{d-1} \sum_{1 \leq \iota_1 < \dots < \iota_\kappa \leq m} \left(\sum_{k=1}^{d-\kappa} \mathfrak{H}_{0; \iota_1, \dots, \iota_\kappa}^{(k)}(\mathbf{z}) \right) y_{\iota_1} \dots y_{\iota_\kappa} \\ &+ \mathfrak{p}_1(0, \dots, 0, \mathbf{y}, 0) + \mathfrak{p}_2(0, \dots, 0, 0, \mathbf{z}), \end{aligned} \quad (5.2)$$

where for each appropriate set of indices the $\mathfrak{G}_{i_1, \dots, i_j}^{(k)}$ and the $\mathfrak{H}_{0; \iota_1, \dots, \iota_\kappa}^{(k)}$ are systems of at most r integral forms of degree k in the appropriate variables. We write $\mathfrak{p}_1(\mathbf{y})$ for $\mathfrak{p}_1(0, \dots, 0, \mathbf{y}, 0)$, and similarly $\mathfrak{p}_2(\mathbf{z})$ for $\mathfrak{p}_2(0, \dots, 0, 0, \mathbf{z})$.

Let $\mathfrak{G}$ be the collection of all the forms $\mathfrak{G}_{i_1, \dots, i_j}^{(k)}$. These are independent of the choice of the $(\mathbf{y}, \mathbf{z})$ partition and there are crudely at most $R_{\mathfrak{G}} \leq d^2 K^d r \leq d^{d+2} r^{d+1}$ of them. By Proposition 1', there is a system $\mathfrak{g}_{\mathfrak{G}}$ which is a strong regularization of $\mathfrak{G}$ with respect to the functions $F_i(R) = \rho_i(2R + 2r_d, d)$ ($i = 1, \dots, d-1$) and the variables $\mathbf{z}$ so that the number of forms of degree i in $\mathfrak{g}_{\mathfrak{G}}$ is bounded in terms of r and d independently of the choice of the partition of the variables. Correspondingly for any choice of the variables $(\mathbf{y}, \mathbf{z})$ we have $R_{\mathfrak{g}_{\mathfrak{G}}} \leq R_0$ for an appropriate constant $R_0 = R_0(r, d)$.

The variables $\mathbf{y}$, and thus $\mathbf{z}$ as well, are now chosen by Proposition 2 with the choice of $C_1 = 2^{d-1} \rho_d(2R_0 + 2r_d, d)$ so that the forms

$$\mathfrak{f}(0, \dots, 0, \mathbf{y}, \mathbf{z}) = \mathfrak{f}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z}) + \mathfrak{f}_2(\mathbf{z})$$

have $\mathcal{B}_d(\mathfrak{f}_1(\mathbf{y}) + \mathfrak{g}(\mathbf{y}, \mathbf{z})) \geq C_1$ with the number of $\mathbf{y}$ variables, m , being at most $C_1 r$. With this choice we have that the system obtained by adjoining either of the systems $\mathfrak{f}_1 + \mathfrak{g}$ or $\mathfrak{p}_1 + \mathfrak{p}_3$ to the $\mathfrak{g}_{\mathfrak{G}}$ is a regular system by Theorem

A, where p_3 is defined to be $p(0, \dots, 0, \mathbf{y}, \mathbf{z}) - p_1(\mathbf{y}) - p_2(\mathbf{z})$. Moreover $h_{\mathbb{Q}}(\bar{\mathbf{g}}_{\mathfrak{G}}^{(i)}; \mathbf{z}) \geq \rho_i(2R_{\mathfrak{g}_{\mathfrak{G}}} + 2r_d, d)$ for $i = 1, \dots, d-1$ by Proposition 1', with $\bar{\mathbf{g}}_{\mathfrak{G}}$ denoting the subsystem obtained by removing the forms of $\mathbf{g}_{\mathfrak{G}}$ depending only on the $\mathbf{z}$ variables.

We break down further the forms of $\mathbf{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z})$ by separating the $\mathbf{y}$ and $\mathbf{z}$ parts:

$$(\mathbf{g}_{\mathfrak{G}})_i^{(l)}(\mathbf{y}, \mathbf{z}) = \sum_{\kappa=0}^l \sum_{1 \leq \iota_1 < \dots < \iota_{\kappa} \leq m} \mathfrak{H}_{i; \iota_1, \dots, \iota_{\kappa}}^{(k; l)}(\mathbf{z}) y_{\iota_1} \dots y_{\iota_{\kappa}}. \quad (5.3)$$

Note that the right hand side introduces at most $lm^l \leq dm^d$ forms in $\mathbf{z}$ for the i th form of degree l in $\mathbf{g}_{\mathfrak{G}}$. We collect the forms $\mathfrak{H}_{0; \iota_1, \dots, \iota_{\kappa}}^{(k)}$ and $\mathfrak{H}_{i; \iota_1, \dots, \iota_{\kappa}}^{(k; l)}$ into a system $\mathfrak{H}$. Then the number of forms $R_{\mathfrak{H}}$ of $\mathfrak{H}$ is at most $R_{\mathfrak{g}_{\mathfrak{G}}} dm^d + r d^2 m^d$. Now we regularize the system $\mathfrak{H}$ with respect to the functions $F_i(R) = \rho_i(2R + 2r, d)$ for $1 \leq i \leq d-1$ and call the resulting system $\mathbf{g}_{\mathfrak{H}}$. Note that $|\mathbf{g}_{\mathfrak{H}}| \leq R_1$ where R_1 is a constant depending only on the original parameters R and d .

If the system $\mathbf{f}_2(\mathbf{z})$ has rank at least as large as $C_2 = \rho_d(2R_1 + 2r_d, d)$, then system $\mathbf{f}_2$ adjoined to the system $\mathbf{g}_{\mathfrak{H}}$ is a regular system by Theorem A. This can be guaranteed as long the rank of $\mathbf{f}(0, \dots, 0, \mathbf{y}, \mathbf{z})$ is sufficiently large with respect to our choices of C_1 and C_2 by appealing to Proposition 2. As losing the first K variables can drop the rank by at most $K(r+1) \leq r(r+1)d$, and C_1 and C_2 are dependent only on d and r , this is our choice of $\chi(r, d)$.

We now define the following sets:

$$W_z(H) = \{\mathbf{z} \in [N]^{n-K-m} : \mathbf{g}_{\mathfrak{H}}(\mathbf{z}) = H\},$$

$$W_y(G; H) = \{\mathbf{y} \in [N]^m : \bar{\mathbf{g}}_{\mathfrak{G}}(\mathbf{y}, W_z(H)) = G\}.$$

Note that for $\mathbf{z} \in W_z(H)$ the system $\mathbf{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z})$ are of the form $q(\mathbf{y}, H)$ as all of its coefficient forms are constants depending only on the parameter H . The number of H required is $N^{D_{\mathfrak{g}_{\mathfrak{H}}}}$. The image of $[N]^{n-K}$ under $\bar{\mathbf{g}}_{\mathfrak{G}}$ is $O(N^{D_{\bar{\mathbf{g}}_{\mathfrak{G}}}})$, and this is an upper bound of the number of G 's for any fixed H , where the implied constant does not depend on H .

For any choice of $\mathbf{z} \in W_z(H)$ and $\mathbf{y} \in W_y(G; H)$, the polynomials $\mathbf{p}$ now take the shape given in (2.1)

$$\begin{aligned} \mathbf{p}(x_1, \dots, x_K, \mathbf{y}, \mathbf{z}) &= \mathbf{p}(x_1, \dots, x_K, 0, \dots, 0) \\ &\quad + \sum_{j=1}^{d-1} \sum_{1 \leq i_1 < \dots < i_j \leq K} \mathbf{c}_{i_1, \dots, i_j}^{(d)}(G, H) x_{i_1} \dots x_{i_j} \end{aligned}$$

$$\begin{aligned}
 & + \sum_{\kappa=1}^{d-1} \sum_{1 \leq l_1 < \dots < l_\kappa \leq m} \mathbf{c}_{0; i_1, \dots, i_j}^{(d)}(H) y_{l_1} \dots y_{l_\kappa} \\
 & + \mathfrak{p}_1(0, \dots, 0, \mathbf{y}, 0) + \mathfrak{p}_2(0, \dots, 0, 0, \mathbf{z}) \\
 & := \mathfrak{P}_0(x_1, \dots, x_K, G, H) + \mathfrak{P}_1(\mathbf{y}, H) + \mathfrak{p}_2(\mathbf{z}). \quad (5.4)
 \end{aligned}$$

Indeed on the level set $W_z(H)$ the values of the forms $\mathfrak{g}_{\mathfrak{H}}$ and in turn those of the forms $\mathfrak{H}$ are fixed, including the forms of $\mathfrak{g}_{\mathfrak{G}}$ depending only on the $\mathbf{z}$ variables. Then for $\mathbf{y} \in W_y(G; H)$ the values of the forms $\bar{\mathfrak{g}}_{\mathfrak{G}}$ and hence the values of the all forms in $\mathfrak{g}_{\mathfrak{G}}$ are fixed, which in turn fix the values of coefficient forms $\mathfrak{G}$. Define the exponential sums

$$S_0(\alpha, G, H) = \sum_{x_1, \dots, x_K \in [N]} \Lambda(x_1) \dots \Lambda(x_K) e(\alpha \cdot \mathfrak{P}_0(x_1, \dots, x_K, G, H)).$$

$$S_1(\alpha, G, H) = \sum_{\mathbf{y} \in W_y(G; H)} \Lambda(\mathbf{y}) e(\alpha \cdot \mathfrak{P}_1(\mathbf{y}, H)).$$

$$S_2(\alpha, H) = \sum_{\mathbf{z} \in W_z(H)} \Lambda(\mathbf{z}) e(\alpha \cdot \mathfrak{p}_2(\mathbf{z})).$$

Now we have to bound the expression

$$E_C(N) = \sum_H \sum_{G \in \mathfrak{m}(C)} \int S_0(\alpha, G, H) S_1(\alpha, G, H) S_2(\alpha, H) e(-\mathbf{s} \cdot \alpha) d\alpha.$$

Proceeding as in Sect. 2.1, we obtain

$$\begin{aligned}
 (E_C(N))^2 & \lesssim \left((\log N)^{2n} N^{D_{\bar{\mathfrak{g}}_{\mathfrak{G}}} + D_{\mathfrak{g}_{\mathfrak{H}}}} \sup_{H, G} \|S_0(\cdot, G, H)\|_{\infty(\mathfrak{m}(C))}^2 \right) \\
 & \times \sum_H \sum_G \|S_1(\cdot, G, H)\|_2^2 \|S_2(\cdot, H)\|_2^2 \quad (5.5)
 \end{aligned}$$

The summands on the right hand side can be expressed as the number of solutions of $\mathfrak{P}_1(\mathbf{y}, H) = \mathfrak{P}_1(\mathbf{y}', H)$ for $\mathbf{y}, \mathbf{y}' \in W_y(G; H)$ times the number of solutions to $\mathfrak{p}_2(\mathbf{z}) = \mathfrak{p}_2(\mathbf{z}')$ for $\mathbf{z}, \mathbf{z}' \in W_z(H)$. The conditions $\mathbf{z}, \mathbf{z}' \in W_z(H)$ may be replaced by the conditions $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$ and $\mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) = \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H$. The conditions $\mathbf{y}, \mathbf{y}' \in W_y(G; H)$ may be replaced by the conditions $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) = \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) = G$ for any $\mathbf{z} \in W_z(H)$.

In short, we are summing over all G and H the number of solutions to the system

$$\begin{aligned}\mathfrak{P}_1(\mathbf{y}, H) &= \mathfrak{P}_1(\mathbf{y}', H) \\ \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) &= \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) = G \\ \mathfrak{p}_2(\mathbf{z}) &= \mathfrak{p}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H\end{aligned}$$

for $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$. With a little rearrangement this becomes

$$\begin{aligned}\mathfrak{P}_1(\mathbf{y}, \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) &= \mathfrak{P}_1(\mathbf{y}', \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) \\ \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}, \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) &= \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}', \mathfrak{g}_{\mathfrak{H}}(\mathbf{z})) = G \\ \mathfrak{p}_2(\mathbf{z}) &= \mathfrak{p}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}') = H,\end{aligned}$$

and summing over G and H now simply removes two of the equalities. And after doing so, by removing $\mathfrak{g}_{\mathfrak{H}}$ as an argument puts us in the final form

$$\begin{aligned}\mathfrak{p}_1(\mathbf{y}) + \mathfrak{p}_3(\mathbf{y}, \mathbf{z}) &= \mathfrak{p}_1(\mathbf{y}') + \mathfrak{p}_3(\mathbf{y}', \mathbf{z}) \\ \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) &= \bar{\mathfrak{g}}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) \\ \mathfrak{p}_2(\mathbf{z}) &= \mathfrak{p}_2(\mathbf{z}') \\ \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}) &= \mathfrak{g}_{\mathfrak{H}}(\mathbf{z}'),\end{aligned}\tag{5.6}$$

for $\mathbf{y}, \mathbf{y}' \in [N]^m$ and $\mathbf{z}, \mathbf{z}' \in [N]^{(n-K-m)}$. Let us call the number of solutions to the system (5.6) $\mathcal{W}$. Then (5.5) takes the form

$$(E_C(N))^2 \leq O\left((\log N)^{2n} \mathcal{W} N^{D_{\bar{\mathfrak{g}}_{\mathfrak{G}}} + D_{\mathfrak{g}_{\mathfrak{H}}}} \sup_{H, G} \|S_0(\cdot, G, H)\|_{\infty(\mathfrak{m}(C))}^2\right).\tag{5.7}$$

Proposition 3 follows immediately from the following two claims.

Claim 1: *Given $c > 0$, there is a C such that the bound*

$$\|S_0(\cdot, G, H)\|_{\infty(\mathfrak{m}(C))} = O((\log N)^{-c} N^K)\tag{5.8}$$

holds uniformly in G and H .

Claim 2: *With the rank of $\mathfrak{p}_1(\mathbf{y}) + \mathfrak{p}_3(\mathbf{y}, \mathbf{z})$ and $\mathfrak{p}_2(\mathbf{z})$ sufficiently large, the bound*

$$\mathcal{W} = O\left(N^{2(n-K) - 2D_{\mathfrak{p}} - D_{\mathfrak{g}_{\mathfrak{H}}} - D_{\bar{\mathfrak{g}}_{\mathfrak{G}}}}\right)\tag{5.9}$$

holds.

Let us start with claim 1. We look at $\alpha \cdot \mathfrak{P}_0(x_1, \dots, x_K, G, H)$, focusing on the coefficients of terms of the form $x_{i_1} \dots x_{i_d}$ for $1 \leq i_j \leq K$. From our choice of $x_1, \dots, x_K$, there is a collection of indices, say $(i_1^\kappa, \dots, i_d^\kappa)$ for each $1 \leq \kappa \leq r$, such that the collection $\{\mathbf{b}_{i_1^\kappa, \dots, i_d^\kappa}\}$ is linearly independent. Let M denote the $r \times r$ matrix of these coefficient vectors as rows. The coefficient of $x_{i_1^\kappa} \dots x_{i_d^\kappa}$ in $\alpha \cdot \mathfrak{P}_0(x_1, \dots, x_K, G, H)$ is $(M\alpha)_\kappa$. Because M has full rank, there is some term of the form $x_{i_1^\kappa} \dots x_{i_d^\kappa}$ with a coefficient β where $\beta \in \mathfrak{m}(C')$ for some slightly larger C' .

If it happens to be the case that the indices $i_1^\kappa, \dots, i_d^\kappa$ are equal, say all 1, then the bound follows directly from the bound in Lemma 5 for the x_1 summation, and claim 1 follows by treating the other sums trivially. Otherwise we assume that $x_{i_1^\kappa} \dots x_{i_d^\kappa} = x_1^{\gamma_1} \dots x_l^{\gamma_l}$ where $\sum \gamma_i = d$ and $l < d$. Now look at the sum S_0 in the form

$$\begin{aligned} & \sum_{x_{l+1}, \dots, x_K \in [N]} \Lambda(x_{l+1}) \dots \Lambda(x_K) \\ & \times \sum_{x_1, \dots, x_l \in [N]} \Lambda(x_1) \dots \Lambda(x_l) e(\beta x_1^{\gamma_1} \dots x_l^{\gamma_l} + Q(x_1, \dots, x_K, G, H)) \end{aligned}$$

where $Q(x_1, \dots, x_K, G, H)$ is viewed as a polynomial in $x_1, \dots, x_l$ of degree less than d with coefficients in the other x_i and the G and the H . We proceed by using the Weyl differencing method, that is applying the Cauchy–Schwarz inequality γ_i times to the inner sum T_0 for each of the variables x_i . This gives the upper bound

$$\begin{aligned} |T_0|^{2^d} & \leq (\log N)^{d2^d} N^{2^d l - d - l} \sum_{x_1, \dots, x_l} \sum_{w_1^1, \dots, w_{\gamma_1}^1} \dots \sum_{w_1^l \dots w_{\gamma_l}^l} \\ & \times \left(\prod_{i=1}^l \Delta_{w_1^i, \dots, w_{\gamma_i}^i} \mathbf{1}_{x_i \in [N]} \right) e\left(\beta w_1^1 \dots w_{\gamma_l}^l\right), \end{aligned}$$

where $\Delta_w f(x) = f(x + w)f(x)$ is the multiplicative differencing operator, and $\Delta_{w_1, w_2} = \Delta_{w_2}(\Delta_{w_1})$, and so on. For a fixed value of x -variables, the inner sum is taken over a convex set $\mathcal{K}(x) \subseteq [-N, N]^d$. Renaming the w -variables as $\mathbf{w} = (w_1, \dots, w_d)$ and fixing $w_1, \dots, w_{d-1}$, the sum in w_d is taken over an interval of length at most $2N$ of a phase which is linear in w_d and is estimated by $\min\left(2N, \frac{1}{\|\beta w_1 \dots w_{d-1}\|}\right)$. Here $\|\gamma\|$ denotes the distance of a real number

γ to the nearest integer. Thus we have that

$$|T_0|^{2^d} \leq (\log N)^{d2^d} N^{2^d l - d} \\ \times \sum_{(w_1, \dots, w_{d-1}) \in [-N, N]^{d-1}} \min \left(2N, \frac{1}{\|\beta w_1 \dots w_{d-1}\|} \right).$$

At this point we proceed as in [2] Sec.2, see Lemmas 2.1–2.4, however as our context is slightly different we briefly describe the argument below. First note that by the multi-linearity of the expression in the denominator, we have that the sum in the w -variables is bounded by $2N \log N$ -times the cardinality of the set

$$\mathcal{A}_N := \left\{ \mathbf{w} = (w_1, \dots, w_{d-1}) \in [-N, N]^{d-1}; \|\beta w_1 \dots w_{d-1}\| \leq \frac{1}{N} \right\}.$$

For given $1 \leq M < N$ define the sets

$$\mathcal{A}_{N,M} := \left\{ \mathbf{w} = (w_1, \dots, w_{d-1}) \in \left[-\frac{N}{M}, \frac{N}{M} \right]^{d-1}; \right. \\ \left. \|\beta w_1 \dots w_{d-1}\| \leq \frac{1}{NM^{d-1}} \right\}.$$

Applying Lemma 2.3 in [2]—a result from the geometry of numbers - successively in the variables w_i , we have that

$$|\mathcal{A}_N| \lesssim M^{d-1} |\mathcal{A}_{N,M}|.$$

Choose $M := N(\log N)^{-C''}$ with $C'' = C'/d$, and notice if there is a vector $\mathbf{w} = (w_1, \dots, w_{d-1}) \in \mathcal{A}_{N,M}$ so that $q := |w_1 \dots w_{d-1}| \geq 1$, then $\|q\beta\| \leq N^{-d} (\log N)^{C'}$ and hence β would be in the major arcs $\mathcal{M}(C')$ contradicting our assumption. Thus all vectors $\mathbf{w} \in \mathcal{A}_{N,M}$ have at least one zero coordinate which implies that $|\mathcal{A}_{N,M}| \lesssim (\log N)^{(d-2)C''}$, which gives the upper bound

$$|T_0|^{2^d} \lesssim N^{2^d l} (\log N)^{d2^d + 1 - C''} \leq N^{2^d l} (\log N)^{-C''/2},$$

if C , and hence C' and C'' , is chosen sufficiently large with respect to d . Claim 1 follows by taking the 2^d th root and summing trivially in $x_{l+1}, \dots, x_K$.

We turn to the proof of claim 2. One may write $\mathcal{W} = \sum_{\mathbf{z}} T_1(\mathbf{z}) T_2(\mathbf{z})$ where, for fixed $\mathbf{z}$, $T_1(\mathbf{z})$ is the number of solutions $\mathbf{y}, \mathbf{y}' \in [N]^m$ to the system

$$\mathbf{p}_1(\mathbf{y}) + \mathbf{p}_3(\mathbf{y}, \mathbf{z}) = \mathbf{p}_1(\mathbf{y}') + \mathbf{p}_3(\mathbf{y}', \mathbf{z}) \\ \bar{\mathbf{g}}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) = \bar{\mathbf{g}}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}),$$

while $T_2(\mathbf{z})$ is the number of solutions $\mathbf{z}' \in [N]^{n-K-m}$ to

$$\begin{aligned} p_2(\mathbf{z}) &= p_2(\mathbf{z}') \\ g_{\mathfrak{H}}(\mathbf{z}) &= g_{\mathfrak{H}}(\mathbf{z}'), \end{aligned}$$

By the Cauchy–Schwarz inequality we have that $\mathcal{W}^2 \leq (\sum_{\mathbf{z}} T_1(\mathbf{z})^2) (\sum_{\mathbf{z}} T_2(\mathbf{z})^2) =: \mathcal{W}_1 \mathcal{W}_2$.

Now, $\mathcal{W}_1$ is the number of $\mathbf{y}, \mathbf{y}', \mathbf{u}, \mathbf{u}' \in [N]^m$ and $\mathbf{z} \in [N]^{n-K-m}$ satisfying the equations

$$\begin{aligned} p_1(\mathbf{y}) + p_3(\mathbf{y}, \mathbf{z}) - p_1(\mathbf{y}') - p_3(\mathbf{y}', \mathbf{z}) &= 0 \\ p_1(\mathbf{u}) + p_3(\mathbf{u}, \mathbf{z}) - p_1(\mathbf{u}') - p_3(\mathbf{u}', \mathbf{z}) &= 0 \\ \bar{g}_{\mathfrak{G}}(\mathbf{y}, \mathbf{z}) - \bar{g}_{\mathfrak{G}}(\mathbf{y}', \mathbf{z}) &= 0, \\ \bar{g}_{\mathfrak{G}}(\mathbf{u}, \mathbf{z}) - \bar{g}_{\mathfrak{G}}(\mathbf{u}', \mathbf{z}) &= 0, \end{aligned} \quad (5.10)$$

and $\mathcal{W}_2$ is the number of $\mathbf{z}, \mathbf{z}', \mathbf{v}' \in [N]^{n-K-m}$ satisfying

$$p_2(\mathbf{z}) - p_2(\mathbf{z}') = 0, \quad p_2(\mathbf{z}) - p_2(\mathbf{v}') = 0, \quad g_{\mathfrak{H}}(\mathbf{z}) - g_{\mathfrak{H}}(\mathbf{z}') = 0, \quad g_{\mathfrak{H}}(\mathbf{z}) - g_{\mathfrak{H}}(\mathbf{v}') = 0 \quad (5.11)$$

First we consider the system in (5.10), and estimate the rank of the family of degree d forms. For given $\lambda, \mu \in \mathbb{Q}^r$ assume we have the decomposition

$$\begin{aligned} \lambda \cdot (f_1(\mathbf{y}) + g(\mathbf{y}, \mathbf{z}) - f_1(\mathbf{y}') - g(\mathbf{y}', \mathbf{z})) + \mu \cdot (f_1(\mathbf{u}) \\ + g(\mathbf{u}, \mathbf{z}) - f_1(\mathbf{u}') - g(\mathbf{u}', \mathbf{z})) = \sum_{i=1}^h U_i V_i, \end{aligned}$$

where $U_i(\mathbf{y}, \mathbf{y}', \mathbf{u}, \mathbf{u}', \mathbf{z})$ and $V_i(\mathbf{y}, \mathbf{y}', \mathbf{u}, \mathbf{u}', \mathbf{z})$ are homogeneous forms of positive degree. Using the facts that $f_1(\mathbf{0}) = \mathbf{0}$ and $g(\mathbf{0}, \mathbf{z})$ vanishes identically, substituting $\mathbf{y}' = \mathbf{u} = \mathbf{u}' = \mathbf{0}$ gives

$$\lambda \cdot (f_1(\mathbf{y}) + g(\mathbf{y}, \mathbf{z})) = \sum_{i=1}^h U'_i(\mathbf{y}, \mathbf{z}) V'_i(\mathbf{y}, \mathbf{z}),$$

thus $h \geq h_{\mathbb{Q}}(f_1 + g) \geq 2^{1-d} \mathcal{B}_d(f_1 + g) \geq \rho_d(2R_{\bar{g}_{\mathfrak{G}}} + 2r_d)$. Here we used the fact that $h_{\mathbb{Q}}(f) \geq h_{\mathbb{C}}(f) \geq 2^{1-d} \mathcal{B}_d(f)$ for a homogeneous form f of degree d , see [17], Lemma 16.1. To estimate the rank of the degree i forms for given

$2 \leq i \leq d-1$ we invoke Lemma 2. We have that

$$\begin{aligned} & h_{\mathbb{Q}} \left(\left(\bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}, \mathbf{z}) - \bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}', \mathbf{z}), \bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{u}, \mathbf{z}) - \bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{u}', \mathbf{z}) \right); \mathbf{z} \right) \\ &= h_{\mathbb{Q}} \left(\bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}, \mathbf{z}) - \bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}', \mathbf{z}); \mathbf{z} \right) \geq h_{\mathbb{Q}} \left(\bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}, \mathbf{z}), \bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}', \mathbf{z}); \mathbf{z} \right) \\ &= h_{\mathbb{Q}} \left(\bar{\mathfrak{g}}_{\mathfrak{S}}^{(i)}(\mathbf{y}, \mathbf{z}); \mathbf{z} \right) \geq \rho_i (2R_{\bar{\mathfrak{g}}_{\mathfrak{S}}} + 2r_d) \end{aligned}$$

Since $h_{\mathbb{Q}}(f(\mathbf{y}, \mathbf{z})) \geq h_{\mathbb{Q}}(f(\mathbf{y}, \mathbf{z}); \mathbf{z})$ for any system of forms f and the total number of forms of system (5.10) is $2R_{\bar{\mathfrak{g}}_{\mathfrak{S}}} + 2r_d$, one has by Theorem A

$$\mathcal{W}_1 = O \left(N^{4m+n-K-m-2D_p-2D_{\bar{\mathfrak{g}}_{\mathfrak{S}}}} \right).$$

For system (5.11) the estimates are simpler. Similarly as above it is easy to see directly from the definition that forms of this system of degree i have Schmidt rank at least $\rho_i(2R_{\mathfrak{g}_{\mathfrak{S}}} + 2r_d, d)$. Indeed, suppose for $\lambda, \mu \in \mathbb{Q}^r$ with say $\lambda \neq \mathbf{0}$

$$\lambda \cdot (f_2(\mathbf{z}) - f_2(\mathbf{z}')) + \mu \cdot (f_2(\mathbf{z}) - f_2(\mathbf{v}')) = \sum_{i=1}^h U_i(\mathbf{z}, \mathbf{z}'\mathbf{v},) V_i(\mathbf{z}, \mathbf{z}'\mathbf{v},).$$

Substituting $\mathbf{z}' = \mathbf{v}' = \mathbf{0}$ we have $\lambda \cdot f_2(\mathbf{z}) = \sum_{i=1}^h U'_i(\mathbf{z}) V'_i(\mathbf{z})$ and as not all terms in the sum can be zero and for each non-zero term $U'_i(\mathbf{z}) V'_i(\mathbf{z})$ both forms have positive degrees we have that $h \geq h_{\mathbb{Q}}(f_2)$. The rank of the systems $\mathfrak{g}_{\mathfrak{S}}^{(i)}$ are estimated the same way. The system (5.11) hence

$$\mathcal{W}_2 = O \left(N^{3(n-K-m)-2D_p-2D_{\mathfrak{g}_{\mathfrak{S}}}} \right).$$

Putting these estimates together

$$\mathcal{W} \leq \sqrt{\mathcal{W}_1 \mathcal{W}_2} = O \left(N^{2(n-K)-2D_p-D_{\bar{\mathfrak{g}}_{\mathfrak{S}}}-D_{\mathfrak{g}_{\mathfrak{S}}}} \right).$$

This proves claim 2 and Proposition 3 follows. $\square$

6 The major arcs

The treatment of the major arcs is fairly standard and differs little from the scenario with diagonal equations (see Hua [11]). Recall that the major arcs are a union of boxes of the form $\mathfrak{M}_{\mathbf{a},q}(C)$, where $q \leq (\log N)^C$ and C is now a fixed constant chosen large enough so that Lemma 3 holds with $c = 1$.

For a fixed $\mathbf{a}$ and q , the small size of the associated major arc means that the exponential sum

$$T_{\mathbf{p}}(\alpha) = \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) e(\mathbf{p}(\mathbf{x}) \cdot \alpha)$$

can be replaced by any approximation that has a sufficiently large logarithmic power gain in the error.

Upon the actual fixing of a $q \leq (\log N)^C$ and an $\mathbf{a} \in U_q^n$, one has²

$$\begin{aligned} T_{\mathbf{p}}(\alpha) &= \sum_{\mathbf{x} \in [N]^n} \Lambda(\mathbf{x}) e(\mathbf{p}(\mathbf{x}) \cdot \alpha) \\ &= \sum_{\mathbf{g} \in Z_q^n} \sum_{\mathbf{x} \in [N]^n} \mathbf{1}_{\mathbf{x} \equiv \mathbf{g} (q)} \Lambda(\mathbf{x}) e(\mathbf{a} \cdot \mathbf{p}(\mathbf{g})/q) e(\mathbf{p}(\mathbf{x}) \cdot \tau) \\ &= \sum_{\mathbf{g} \in Z_q^n} e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}), \end{aligned} \quad (6.1)$$

where the notations introduced here are $\tau_i = \alpha_i - a_i/q$, and $\psi_{\mathbf{g}}(\mathbf{X}) = \psi_{g_1}(X_1) \dots \psi_{g_n}(X_n)$ with

$$\psi_l(v) = \sum_{t \leq v, t \equiv l (q)} \Lambda(t),$$

and $\mathcal{J}$ is the unit cube $[0, 1]^n \subset \mathbb{R}^n$.

Lemma 6 *For any given a constant c , the estimate*

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}) = \mathbf{1}_{\mathbf{g} \in U_q^n} \phi(q)^{-n} \int_{\mathbf{z} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{z}) \cdot \tau) d\mathbf{z} + O((\log N)^{-c} N^n) \quad (6.2)$$

holds on each major arc $\mathfrak{M}_{a,q}(C)$ provided that C is sufficiently large.

Proof Define for a fixed l the one dimensional signed measure $dv_l = d\psi_l - d\omega_l$, where $d\omega_l$ is the Lebesgue measure multiplied by the reciprocal of the totient of q if $l \in U_q$, and is zero otherwise. For a continuous function f one then has

$$\int_0^N f(X) dv_l(X) = \sum_{x \in [N], x \equiv l (q)} \Lambda(x) f(x) - \phi(q)^{-1} \int_0^N f(z) dz.$$

² There is some ambiguity in the case where N is a prime power, however, there is no harm in assuming that this is not so due to the fact that the prime powers are sparse.

Now set $d|v_l| = d\omega_l + d\psi_l$, so that

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\psi_{\mathbf{g}}(\mathbf{X}) = \int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) \prod_{i=1}^n (dv_{g_i}(X_i) + d\omega_{g_i}(X_i)).$$

Expanding out the product in the last integral gives the form

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\omega_{\mathbf{g}}(\mathbf{X}) + \sum_{i=1}^{2^n-1} \int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\mu_{i,\mathbf{g}}(\mathbf{X}),$$

where $d\mu_{i,\mathbf{g}}$ runs over all of the corresponding product measures, barring the $d\omega_{\mathbf{g}}(\mathbf{X})$ term.

Consider

$$\int_{\mathbf{X} \in N\mathcal{J}} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\mu_{i,\mathbf{g}}(\mathbf{X})$$

for some fixed i . Assume without loss of generality that $d\mu_{i,\mathbf{y}}$ is of the form

$$dv_{g_1}(X_1) d\sigma_{\mathbf{g}}(X_2, \dots, X_n),$$

where $d\sigma_{\mathbf{g}}$ may be signed in some variables and is independent of g_1 . The range of integration for the X_1 variable is a copy of the continuous interval $[0, N]$, and is to be split into smaller disjoint intervals of size $N^1(\log N)^{-c'}$. Here c' is chosen to be between $(c+C)$ and $2(c+C)$ such that $(\log N)^{c'}$ is an integer, say B . The equality $[0, N] = \bigcup_{j=1}^B I_j$ follows. Also set $\mathcal{J}'_j = I_j \times [0, N]^{n-1}$, which absorbs the factor of N .

Now, for a fixed interval I_j , select some $t \in I_j$. Then write

$$\begin{aligned} \int_{\mathbf{X} \in \mathcal{J}'_j} e(\mathbf{p}(\mathbf{X}) \cdot \tau) d\mu_{i,\mathbf{g}} &= \int_{\mathbf{X} \in \mathcal{J}'_j} e(\mathbf{p}(t, X_2, \dots, X_n) \cdot \tau) dv_{g_1}(X_1) d\sigma_{\mathbf{g}}(X_2, \dots, X_n) \\ &\quad + \int_{\mathbf{X} \in \mathcal{B}_j} (e(\mathbf{p}(X_1, \dots, X_n) \cdot \tau) - e(\mathbf{p}(t, X_2, \dots, X_n) \cdot \tau)) \\ &\quad \times dv_{g_1}(X_1) d\sigma_{\mathbf{g}}(X_2, \dots, X_n) \\ &:= E_1 + E_2 \end{aligned}$$

The first error term satisfies

$$|E_1| \leq \int_{X_2, \dots, X_n \in [0, N]} \left| \int_{I_j} dv_{g_1}(X_1) \right| d|\sigma_{\mathbf{g}}|(X_2, \dots, X_n) = O\left(N^n e^{-c_0 \sqrt{\log N}}\right)$$

for some positive constant c_0 by the Siegel–Walfisz theorem, as $q \leq (\log N)^C$. To bound E_2 , note that on I_j the integrand is

$$O(|\mathfrak{p}(X_1, \dots, X_n) - \mathfrak{p}(t, X_2, \dots, X_n) \cdot \tau|) = O\left((\log N)^{C-c'}\right).$$

In turn,

$$\begin{aligned} |E_2| &= O((\log N)^{C-c'}) \int_{\mathbf{X} \in \mathcal{J}'_j} d|v_{g_1}|(X_1) d|\sigma_{\mathbf{g}}|(X_2, \dots, X_n) \\ &= O(N^n (\log N)^{C-2c'}). \end{aligned}$$

There are $2^n - 1$ error terms on each interval, so summing over the $(\log N)^{c'}$ intervals completes the proof. $\square$

The integral appearing in the last result has a quick reduction, namely

$$\int_{N\mathcal{J}} e(\mathfrak{p}(\mathbf{X})\tau) d\mathbf{X} = \int_{N\mathcal{J}} e(\mathfrak{f}(\mathbf{X})\tau) d\mathbf{X} + O(N^{n-1+\epsilon}),$$

recalling that $\mathfrak{f}$ is the highest degree part of $\mathfrak{p}$. Following along with the work of Birch,

$$\int_{N\mathcal{J}} e(\mathfrak{f}(\mathbf{X})\tau) d\mathbf{X} = N^n \int_{\zeta \in \mathcal{J}} e(\mathfrak{f}(\zeta) \cdot N^2 \tau) d\zeta,$$

is denoted by $N^n \mathcal{I}(\mathcal{J}, N^d \tau)$ in [2]. This function is independent of $\mathbf{a}$ and q . Thus the integral over any major arc yields the common integral

$$\int_{|\tau| \leq (\log N)^C} \mathcal{I}(\mathcal{J}, N^d \tau) e(-\mathbf{s} \cdot \tau) d\tau.$$

With $\mu = N^{-d} \mathbf{s}$, set

$$J(\mu; \Phi) = \int_{|\tau| \leq \Phi} \mathcal{I}(\mathcal{J}, \tau) e(-\mu \cdot \tau) d\tau,$$

and

$$J(\mu) = \lim_{\Phi \rightarrow \infty} J(\mu; \Phi).$$

The following is Lemma 5.3 in [2].

Lemma 7 *The function $J(\mu)$ is continuous and uniformly bounded in μ . Moreover,*

$$|J(\mu) - J(\mu, \Phi)| \lesssim \Phi^{-\frac{1}{2}}$$

holds uniformly in μ .

By defining

$$W_{\mathbf{a},q} = \sum_{\mathbf{g} \in U_q^n} e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q),$$

one then has

Lemma 8 *For any given $c > 0$, the estimate*

$$\begin{aligned} \int_{\mathfrak{M}_{\mathbf{a},q}(C)} T_{\mathbf{p}}(\alpha) e(-\mathbf{s} \cdot \alpha) d\alpha &= N^{n-dr} \phi(q)^{-n} W_{\mathbf{a},q} e(-\mathbf{s} \cdot \mathbf{a}/q) J(\mu) \\ &\quad + O(N^{n-dr} (\log N)^{-c}), \end{aligned}$$

where $\mu = N^{-2}\mathbf{s}$, holds on each major arc $\mathfrak{M}_{\mathbf{a},q}(C)$.

The measure of the major arcs is easily at most $N^{-dr} (\log N)^K$ for some constant K . By defining

$$\begin{aligned} B(\mathbf{s}, q) &= \sum_{\mathbf{a} \in U_q^r} \phi(q)^{-n} W_{\mathbf{a},q} e(-\mathbf{s} \cdot \mathbf{a}/q) \\ \mathfrak{S}(\mathbf{s}, N) &= \sum_{q \leq (\log N)^C} B(\mathbf{s}, q), \end{aligned}$$

it then follows that

Lemma 9 *The estimate*

$$\mathcal{M}_{\mathbf{p},\mathbf{s}}(N) = \mathfrak{S}(\mathbf{s}, N) J(\mu) N^{n-dr} + O((\log N)^{-c} N^{n-dr}) \quad (6.3)$$

holds for any chosen value of c .

7 The singular series

Following the outline of Hua ([11], chapter VIII, §2, Lemma 8.1), one can show that $B(\mathbf{s}, q)$ is multiplicative as a function of q . This leads to consideration of the formal identity

$$\mathfrak{S}(\mathbf{s}) := \lim_{N \rightarrow \infty} \mathfrak{S}(\mathbf{s}, N) = \prod_{p < \infty} \left(1 + \sum_{t=1}^{\infty} B(\mathbf{s}, p^t) \right). \quad (7.1)$$

Lemma 10 *If $q = p^t$ is a prime power, then*

$$B(\mathbf{s}, q) = O\left(q^{r-\mathcal{B}(\mathbf{p})/((d-1)2^d r)+\epsilon}\right) \quad (7.2)$$

holds uniformly in $\mathbf{s}$ as $t \rightarrow \infty$. The implied constants can be made independent of p .

Proof It is shown here that

$$W_{\mathbf{a},q} = O\left(q^{n-\mathcal{B}(\mathbf{p})/((d-1)2^d r)+\epsilon}\right),$$

uniformly for $\mathbf{a} \in U_q^n$, which clearly implies the result by the definition of $B(\mathbf{s}, q)$ and the fact that $q^n/\phi(q)^n \leq 2^n$ independent of p .

The inclusion-exclusion principle is used to bound $W_{\mathbf{a},q}$ when $q = p^t$ when $t \leq d$. Let such a t be fixed, and note that the characteristic function of U_{p^t} decomposes as

$$\mathbf{1}_{U_{p^t}}(x) = 1 - \sum_{h \in Z_{p^{t-1}}} \mathbf{1}_{x=hp}.$$

Applying this in the definition gives

$$\begin{aligned} W_{\mathbf{a},q} &= \sum_{\mathbf{g} \in U_q^n} e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \\ &= \sum_{\mathbf{g} \in Z_q^n} \prod_{i=1}^n \left(1 - \sum_{h_i \in Z_{p^{t-1}}} \mathbf{1}_{g_i=h_i p}\right) e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \\ &= \sum_{I \subseteq [n]} (-1)^{|I|} \sum_{\mathbf{h} \in Z_{p^{t-1}}^{|I|}} \sum_{\mathbf{g} \in Z_q^n} F_I(\mathbf{g}; \mathbf{h}) e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q), \end{aligned} \quad (7.3)$$

where

$$F_I(\mathbf{g}; \mathbf{h}) = \prod_{i \in I} \mathbf{1}_{g_i=ph_i}$$

for $\mathbf{h} \in Z_p^{|I|}$. In other words, F_I is the characteristic function of the set $H_{I,\mathbf{h}} = \{\mathbf{g} : g_i = ph_i \forall i \in I\}$.

The sets $I \subseteq [n]$ divided into two categories according to whether $|I| \leq \mathcal{B}(\mathbf{p})/(r+1)$ or not. If I is a set fitting into the latter category, then the trivial

estimate is

$$\left| \sum_{\mathbf{h} \in Z_{p^{t-1}}^{|I|}} \sum_{\mathbf{g} \in Z_q^n} F_I(\mathbf{g}; \mathbf{h}) e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \right| = p^{(t-1)|I|} (p^t)^{n-|I|} \\ = (p^t)^{n-|I|/t} \leq q^{n-\mathcal{B}(\mathbf{p})/(tr+t)} \leq q^{n-\mathcal{B}(\mathbf{p})/((d-1)2^d r)}.$$

Now let I be a fixed subset of $[n]$ with $|I| \leq \mathcal{B}(\mathbf{p})/(r+1)$. For each $\mathbf{h}$ the restriction of $\mathbf{p}$ to the set $H_{I,\mathbf{h}}$ has Birch rank at least $\mathcal{B}(\mathbf{p}) - |I|(r+1)$ by corollary 4. By the work of Birch ([2], Lemma 5.4) it follows that

$$\sum_{\mathbf{h} \in Z_p^{(t-1)|I|}} \left| \sum_{\mathbf{g} \in Z_q^n} F_I(\mathbf{g}; \mathbf{h}) e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \right| \\ \lesssim q^{(t-1)|I|/t} q^{n-|I|-(\mathcal{B}(\mathbf{p})-|I|(r+1))/((d-1)2^{d-1}r)+\epsilon} \\ \lesssim q^{n-\mathcal{B}(\mathbf{p})/((d-1)2^d r)+\epsilon}.$$

Summing over all I yields the bound.

Now let $q = p^t$ for $t > d$. Going back to the definition gives

$$W_{\mathbf{a},q} = \sum_{\mathbf{g} \in U_q^n} e(\mathbf{p}(\mathbf{g}) \cdot \mathbf{a}/q) \\ = \sum_{\mathbf{g} \in U_p^n} \sum_{\mathbf{h} \in Z_{p^{t-1}}^n} e(\mathbf{p}(\mathbf{g} + p\mathbf{h}) \cdot \mathbf{a}/q).$$

The system of forms in the exponent can be expanded for each fixed $\mathbf{g}$ as

$$\mathbf{p}(\mathbf{g} + p\mathbf{h}) = p^d \mathbf{p}(\mathbf{h}) + f_{\mathbf{g}}(\mathbf{h})$$

for some polynomial $f_{\mathbf{g}}$ of degree at most $d-1$. Then it follows that

$$|W_{\mathbf{a},q}| \leq \sum_{\mathbf{g} \in U_p^n} \left| \sum_{\mathbf{h} \in Z_{p^{t-1}}^n} e(f_{\mathbf{g}}(\mathbf{h}) + p^d \mathbf{p}(\mathbf{h})) \cdot \mathbf{a}/q \right|. \quad (7.4)$$

The inner sum is now bounded uniformly in $\mathbf{g}$ by an application of the exponential sum estimates in [2] as follows.

Set $P = p^{t-1}$ and $q_1 = p^{t-d}$. Then, for each $i = 1, \dots, r$,

$$2|q'_i a_i - a'_i q_1| \leq P^{-(d-1)+(d-1)r\theta}$$

and

$$1 \leq q' \leq P^{(d-1)r\theta}$$

cannot be satisfied if $\theta < 1/(d-1)r$. Then, by Lemma 4.3 of [2],

$$\sum_{\mathbf{h} \in Z_{p^{t-1}}^n} e((p^d \mathbf{p}(\mathbf{h})) \cdot \mathbf{a}/q + f(\mathbf{h})) = O(P^{n-B(\mathbf{p})/((d-1)2^d r)+\epsilon})$$

for any polynomial $f(\mathbf{h})$ of degree strictly less than d . In turn,

$$|W_{\mathbf{a},q}| \leq \sum_{\mathbf{g} \in U_p^n} O(P^{n-B(\mathbf{p})/((d-1)2^d r)+\epsilon}) = O(q^{n-B(\mathbf{p})/((d-1)2^d r)+\epsilon}),$$

which is what is needed to complete the proof in this last and final case. $\square$

Now define the local factor for a finite prime p as

$$\mu_p = 1 + \sum_{t=1}^{\infty} B(\mathbf{s}, p^t), \quad (7.5)$$

which is well defined as the series is absolutely convergent provided that $B(\mathbf{p}) > (d-1)2^d r(r+1)$. The following result is again an straight forward extension of the results for a single form.

Lemma 11 *For each finite prime p , the local factor may be represented as*

$$\mu_p = \lim_{t \rightarrow \infty} \frac{(p^t)^R M(p^t)}{\phi^n(p^t)}, \quad (7.6)$$

where $M(p^t)$ represents the number of solutions to Eq. 2.2 in the multiplicative group U_{p^t} .

At our disposal now is the fact that the μ_p are positive, which then easily gives the following.

Lemma 12 *If $B(\mathbf{p}) > (d-1)2^d r(r+1)$ then the local factor for each finite prime satisfies the estimate*

$$\mu_p = 1 + O(p^{-(1+\delta)})$$

for some positive δ , and therefore the product in Eq. 7.1 is absolutely convergent and thusly is in fact well defined.

The observation that

$$|\mathfrak{S}(\mathbf{s}, N) - \mathfrak{S}(\mathbf{s})| = o(1)$$

gives the final form of the asymptotic

$$\mathcal{M}_{\mathbf{p},\mathbf{s}}(N) = \mathfrak{S}(\mathbf{s})J(\mu)N^{n-dr} + O((\log N)^{-c}N^{n-dr}).$$

This proves our main result, Theorem 1. The non-vanishing of local factors follows from the exact same lifting argument as when dealing with integer solutions [2].

8 Further remarks

There are further possible refinements of the main result of this paper. It is expected, similarly to the case of integer solutions [17], that Theorem 1 holds assuming only the largeness of the rational Schmidt rank of the system. To prove this one needs to find a suitable analogue of Proposition 2 for the rational Schmidt rank instead of the Birch rank. The tower type bounds on the ranks are due to the regularization process expressed in Proposition 1. It is expected that exponential type lower bounds on the rank of the system are sufficient. It might be possible that there are “transfer principles” for higher degree forms, to allow a more direct transition to find solutions in the primes.

The methods of this paper may extend to other special sequences not just the primes. For example for a translation invariant system of forms one might modify our arguments to show the existence of solutions chosen from a set A of upper positive density. Indeed, assuming that A is sufficiently uniform, one expects that an analogue of Lemma 5 and hence our main estimate Proposition 3 holds, with the minor arcs replaced by $[0, 1]^r$. Otherwise the balanced function of A should correlate with a polynomial phase function which naturally leads to a standard density increment argument. Note that the existence of solutions in the set A already follows from known results, Szemerédi’s theorem [8] together with the main result of Birch [2], providing though weaker quantitative bounds on the density. We do not pursue these matters here.

References

1. Balog, A.: Linear equations in primes. *Mathematika* **39**(2), 367–378 (1992)
2. Birch, B.: Forms in many variables. *Proc. Royal Soc. Lond. Ser. A.* **265**(1321), 245–263 (1962)
3. Bourgain, J., Gamburd, A., Sarnak, P.: Afne linear sieve, expanders, and sum-product. *Invent. Math.* **179**, 559644 (2010)

4. Brdern, J., Dietmann, J., Liu, J., Wooley, T.: A Birch–Goldbach theorem. *Archiv der Mathematik* **94**(1), 53–58 (2010)
5. Goldston, D., Pintz, J.: Yildirim, Primes in tuples I. *Ann. Math.* **170**, 819–862 (2009)
6. Golsefidy, A.S., Sarnak, P.: The affine sieve. *J. Am. Math. Soc.* **26**(4), 1085–1105 (2013)
7. Green, B., Tao, T.: Linear equations in the primes. *Ann. Math. (2)* **171**(3), 1753–1850 (2010)
8. Green, B., Tao, T.: Yet another proof of Szemerédi’s theorem. *An irregular mind*, pp. 335–342. Springer, Berlin Heidelberg, (2010)
9. Green, B., Tao, T.: The distribution of polynomials over finite fields, with applications to the Gowers Norms. *Contrib. Discret. Math.* **4**, 1–36 (2009)
10. Hartshorne, R.: *Algebraic geometry*, Graduate Texts in Mathematics, vol. 52. Springer, Verlag (1977)
11. Hua, L.K.: *Additive theory of prime numbers*, Translations of Mathematical Monographs, vol. 13”. Am. Math. Soc., Providence (1965)
12. Kaufman, T., Lovett, S.: Worst case to average case reductions for polynomials. In: 49th Annual IEEE Symposium on Foundations of Computer, Science, pp. 166–175, (2008)
13. Keil, E.: Translation invariant quadratic forms in dense variables, preprint
14. Liu, J., Sarnak, P.: Integral points on quadrics in three variables whose coordinates have few prime factors. *Israel J. Math.* **178**, 393–426 (2010)
15. Liu, J.: Integral points on quadrics with prime coordinates. *Monats. Math.* **164**(4), 439–465 (2011)
16. Duke, W., Rudnick, Z., Sarnak, P.: Density of integer points on affine homogeneous varieties. *Duke Math. J.* **71**(1), 143–179 (1993)
17. Schmidt, W.: The density of integral points on homogeneous varieties. *Acta. Math.* **154**(3), 243–296 (1985)
18. Vinogradov, I.M.: Representations of an odd integer as a sum of three primes Goldbach Conjecture **4**, 61 (2002)
19. Zhang, Y.: Bounded gaps between primes. *Ann. Math.*, to appear