Diophantine problems over global fields and a conjecture of Artin over function fields

Dissertation

for the award of the degree

"Doctor rerum naturalium"

of the Georg-August Universität Göttingen

within the doctoral program "Mathematical Sciences"

of the Georg-August University School of Science (GAUSS)

submitted by

Leonhard Hochfilzer

from Vienna

Göttingen, 2023

Thesis advisory committee

Prof. Dr. Damaris Schindler,
Mathematisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Jörg Brüdern,
Mathematisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Preda Mihăilescu,
Mathematisches Institut, Georg-August-Universität Göttingen

Members of the examination board

Reviewer:
Prof. Dr. Damaris Schindler
Mathematisches Institut, Georg-August-Universität Göttingen

Second reviewer:
Prof. Dr. Jörg Brüdern,
Mathematisches Institut, Georg-August-Universität Göttingen

Further members of the examination board

Prof. Dr. Preda Mihăilescu,
Mathematisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Ralf Meyer,
Mathematisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Thomas Schick,
Mathematisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Christoph Lehrenfeld,
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Date of oral examination: 23 May 2023

Declaration

Chapters 2, 3 and 5 were created in collaboration with other mathematicians. As requested per the thesis guidelines the following is a declaration of the contributions to the respective projects by the respective authors of the projects. The candidate will henceforth be referred to as LH.

Chapter 2, which deals with cubic forms over imaginary quadratic number fields was created in joint work with fellow Göttingen PhD student Christian Bernert (CB). In the spring of 2022 this topic came up in a conversation between CB and LH motivated by a previous talk by Rainer Dietmann. After a conversation with Tim Browning on this topic, LH and CB started working on the problem. The main ideas and sketches for the proof of the main result were worked out on blackboards in collaborative sessions in the mathematical institute in Göttingen. Once the authors agreed that they now together worked out a line of attack on the problem, they decided to divide the effort of writing up the content and working out the remaining details involved. LH conceived first drafts of Section 2.4, Section 2.5, Section 2.6 and Section 2.7, which is comprised of some technical preliminary lemmas, the treatment of the major arcs and a Weyl differencing estimate, which is needed for a part of the minor arcs. CB conceived first drafts of Section 2.2, Section 2.8 and Section 2.9, which includes the deduction of the applications to our result, the averaged van der Corput estimate and the estimation of the minor arcs. The introductory content of Section 2.1 and Section 2.3 were written up en passant and so both CB and LH contributed equally to the conception of these sections. All sections were proofread by both LH and CB and changes were made afterwards accordingly.

Chapter 3 stems from a collaboration with Jakob Glas (JG). The project started briefly after a conference in 2021 at the Erwin Schrödinger institute in Vienna after a suggestion by Tim Browning. Most of the work was then conceived during the course of three visits. In early 2022 LH visited JG at IST Austria where they mainly worked on understanding the previous literature on this topic. In particular it was during this visit that they both understood that the estimates provided by Browning and Vishe [17] are not in their form sufficient in order to apply them to our problem. During the second visit where JG visited LH in Göttingen the authors worked out a simple way to work around the issue of the unavailability of partial summation in this context, which paves the way to improve the estimates by Browning–Vishe. This, together with working out the remaining technical details that are needed for the first theorem was the main accomplishment during this second visit. All of this took place during collaborative blackboard sessions between JG and LH, taking turns to explain new insights or ideas to the other one on the blackboard. After this second visit the authors were equipped with most of the ideas needed in order to write up the proof of the main result.

Subsequently LH and JG divided the task of writing up the ideas and working out the remaining details involved. In particular they had not worked out the applications of weak approximation and Waring's problem in detail yet. They agreed that JG would thus first write up and think about what is now Section 3.7.1 while LH worked out and wrote up the content of Section 3.7.2. The rest of the paper was divided up as follows: JG wrote up first drafts of Section 3.4, Section 3.6 and Section 3.1 while LH wrote first drafts of Section 3.3, Section 3.5, and Section 3.8. Section 3.2 was written up by both of them en passant while writing up the remainder of the work. Everything was carefully proofread by LH and JG and changes were made accordingly afterwards.

Chapter 5 was written in collaboration with Ezra Waxman (EW). The idea for the project came after a discussion between LH and EW when they noticed the gap in the literature regarding Artin's primitive root conjecture over function fields, if the candidate element is not geometric, as defined in Chapter 5. In contrast to the collaborations described above, here many parts were worked out independently and then usually discussed together via video calls. First drafts for the proofs of some of the main results, in particular the proofs of Theorem 5.4.1, Proposition 5.5.1 and Lemma 5.3.3 were conceived by LH, as well as the ideas in Section 5.4. It was EW's suggestion to include Section 5.2 in order to make the work more accessible. This section along with Section 5.7 and Section 5.1 were conceived and written up together. First drafts of Section 5.3, Section 5.4, Section 5.5 and Section 5.6 were first conceived by LH.

Acknowledgements

Firstly I thank my supervisor Damaris Schindler with whom I have had many extremely fruitful and insightful discussions regarding every chapter of this thesis. I would like to especially thank her for enabling me to work very freely on several different projects. Further I am particularly grateful for her encouragement to travel to many different conferences and workshops, which was not only extremely inspiring but also productive in a very concrete sense – Chapters 2 and 3 for example would potentially not have been conceived if it had not been for certain conference visits. This inspiration and kindness affected me not only on an academic but also on a personal level, which made this PhD an incomparably joyful experience.

I am grateful to very insightful conversations regarding the circle method and the hyperbola method with Jörg Brüdern.

I would like to thank Tim Browning for suggesting to work on diagonal cubic forms over function fields with Jakob Glas. I am also grateful for many useful mathematical discussions and his hospitality in hosting me at ISTA for several visits, which enabled and impacted my collaboration with Jakob Glas.

I am also indebted to Efthymios Sofos, whose hospitality in Glasgow did not just come to fruition in very interesting mathematics but also affected me extremely strongly on a personal level, which he may not even be aware of. I would also like to thank him for pointing out Davenport's question to Jakob and me. On this matter, I would also like to thank Keith Matthews who kindly provided scanned copies of these letters to us.

I would like to thank my collaborators Christian Bernert, Jakob Glas and Ezra Waxman. I thoroughly enjoyed working on these projects together, and through this I have discovered not only very interesting mathematics but also three good friends.

I would like to express my gratitude towards my friends and colleagues in the mathematical institute. Especially Jonas Baltes who was always willing to answer any kind of algebraic geometry question that I would be thinking about.

I would also like to thank Chris Hall, Ofir Gorodetsky, Seoyoung Kim and Simon Rydin Myerson for helpful mathematical discussions regarding various topics of this thesis.

Finally, I would like to thank my family for their love and their supporting regarding with everything – which is why I have five instruments at home but I can only play one. Thank you Izzi for your kindness, your love and for appreciating me for who I am. And also for feeding me and keeping me sane.

Contents

Chapter 1

Introduction

The study of Diophantine problems concerns itself with finding integer solutions to polynomial equations or systems of polynomial equations in integer variables. Such problems have already been studied in ancient Greece, and while our understanding steadily improves there are still many aspects about Diophantine equations that remain elusive. One feature that makes these problems so intriguing is the simplicity with which they may be stated. For example, it is a very basic question to ask if it is possible to find all the (infintely many) right angled triangles, whose sidelengths are integer values. According to a famous theorem named after the Greek philosopher and mathematician Pythagoras, one may repackage this problem as finding all the integer triples (X, Y, Z) such that

$$X^2 + Y^2 = Z^2$$

is satisfied. Such a solution is called a *Pythagorean triple*. Encountering this question for the first time it is not simple to tell how feasible it is to solve this. Taking a geometric viewpoint there is a very elegant solution; one may yet again reformulate this problem as finding all the rational points (x, y) on the unit circle. One such point is given by $(-1, 0)$, which we call O. If we take any other point with rational entries P, say, on the unit circle then the line through O and P must have rational slope. Conversely, if we take a line through O with rational slope r, then this line intersects the unit circle in precisely one point P_r, say, other than O. It is a simple calculation that we have

$$P_r = \left(\frac{1 - r^2}{1 + r^2}, \frac{2r}{1 + r^2} \right).$$

Writing $r = a/b$, where a and b are integers, one can then deduce the solution to the original question. Namely, up to rescaling, all Pythagorean triples (X, Y, Z) are of the form

$$(b^2 - a^2, 2ab, a^2 + b^2),$$

where $a, b \in \mathbb{Z}$. It is not hard to show that a tuple of the form above satisfies $X^2 + Y^2 = Z^2$, but the argument above shows that indeed *every* solution needs to be of this shape. While such elegant solutions are often not available in more complicated situations, this should demonstrate that the geometric picture can be extremely useful in understanding and solving Diophantine problems, which is a recurring theme that appears throughout this thesis.

Now that we have understood how to characterise Pythagorean triples one might enquire about equations of a more general shape. For example, given a quadratic form $Q(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$, that is, a homogeneous degree 2 polynomial, one can ask whether there exists an integer tuple $(a_1, \dots, a_n)$ with not all entries 0 such that $Q(a_1, \dots, a_n) = 0$. We say that Q represents zero non-trivially if that is the case. Also note that due to the homogeneity of Q it makes no difference whether we obtain an integral or rational solution. If the Q is of the shape

$$Q(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2,$$

then clearly the only real solution is $a_1 = \dots = a_n = 0$, and hence there is no non-trivial rational solution. Similarly, if $n = 3$ and the equation is for example given by

$$2x_1^2 - 3x_2^2 - 4x_3^2 = 0,$$

then there are no non-trivial solutions, which can be proved by considering this equation over $\mathbb{Z}/8\mathbb{Z}$; one can easily show that $2x_1^2 - 3x_2^2 - 4x_3^2 \equiv 0 \mod 8$ implies that all of the x_i need to be even. A simple descent argument then shows that there cannot be any non-trivial rational solutions to the above quadratic form. In both examples the *local conditions* are not satisfied. In particular, for a quadratic form to represent zero non-trivially over the rationals it is necessary that there are non-trivial solutions over all $\mathbb{R}$ and $\mathbb{Q}_p$ for every p. One may wonder if the converse holds: If there are non-trivial solutions over $\mathbb{R}$ and $\mathbb{Q}_p$ for all primes p can one recover a non-trivial rational solution? If this is satisfied for a (system) of equations one says that the *Hasse Principle* holds. Such a situation is desirable since it is usually much easier to determine whether a Diophantine problem has local solutions or not.

Returning to quadratic forms, a famous theorem of Hasse and Minkowski states that the Hasse Principle is satisfied for quadratic forms. Indeed, a theorem of Meyer further shows that if $n \geq 5$ then one always obtains non-trivial p-adic solutions to a quadratic form, and therefore, unless Q is definite, one also obtains non-trivial rational solutions. A good treatment of this matter may be found in [102].

Moving to higher degree, analogous to Meyer, Demyanov [32] and Lewis [72] independently showed that the local conditions are satisfied for cubic forms as soon as $n \geq 10$. Therefore if one were to be able to show the Hasse Principle one could infer the existence of a non-trivial rational solution. Indeed using the Hasse Priniple amongst other things, Heath-Brown showed that any cubic form in $n \geq 14$ variables represents zero non-trivially [49]. When the cubic form is non-singular he showed that 10 variables suffice [43]. Later Hooley [52, 53, 54, 56] showed the Hasse Principle in the case of non-singular cubic forms whenever $n \geq 9$, and Vaughan [111] could establish the Hasse Principle for diagonal cubic forms, provided $n \geq 8$. In fact, building on Vaughan's techniques Baker establishes the existence of a non-trivial solution to diagonal cubic forms in at least 7 variables and even finds an upper bound for the size of the smallest solution when the number of variables is 7, 8 or 9. In all of these results a crucial technique used is the *circle method.*

We will henceforth often use the notation $\boldsymbol{x} = (x_1, \dots, x_n)$. Given a polynomial $F(\boldsymbol{x}) \in \mathbb{Z}[x_1, \dots, x_n]$ and a real number $P \geq 1$, one may consider the counting function

$$N(P) = \#\{\boldsymbol{x} \in \mathbb{Z}^n \colon |x_i| \leq P, \text{ for all } i,\ F(\boldsymbol{x}) = 0\}.$$

If the form has degree d, then a simple heuristic argument shows that in a generic situation one would expect the counting function to look like

$$N(P) = \sigma P^{n-d}(1 + o(1)),$$

provided n is sufficiently large. If one knows that σ is positive then such a result clearly implies the existence of a non-trivial solution to $F(\boldsymbol{x}) = 0$. The circle method is highly effective in proving such asymptotic formulas. In particular, the constant σ is usually given as the product of densities of local solutions. Therefore, one can usually show that the existence of non-trivial smooth local solutions together with Hensel's Lemma implies $\sigma > 0$. Thus in many situations such an asymptotic formula implies the *smooth Hasse Principle*, a slightly weaker version of the Hasse Principle.

We now briefly describe the basic ideas involved regarding the circle method. First, given $x \in \mathbb{R}$ we introduce the notation $e(x) = e^{2\pi i x}$. For $\alpha \in \mathbb{R}$ one then considers the exponential sum

$$S(\alpha; P) = S(\alpha) = \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^n \\ |\boldsymbol{x}| \leq P}} e(\alpha F(\boldsymbol{x})).$$

A simple Fourier analytic calculation then reveals

$$N(P) = \int_0^1 S(\alpha) d\alpha.$$

One of the key observations is that if a and q are two coprime integers then $S(a/q)$ is often roughly of size $c_{a,q}P^n$, where $c_{a,q}$ is a positive constant depending on a and q such that $c_{a,q} \to 0$ as $q \to \infty$. This might fail if certain local obstructions are present, for example, but for the sake of this heuristic the reader is encouraged to imagine that $S(a/q) \approx c_{a,q}P^n$. Thus, if α is close to a rational number with small denominator one would expect these values make a large contribution towards the counting function $N(P)$. On the other hand, if α is not close to a rational number with small denominator then the sum $S(\alpha)$ is expected to exhibit cancellation. This can be made precise and the unit interval $[0, 1]$ is thus split into *major arcs* and *minor arcs* accordingly, and integrating over these, one expects to obtain the main and error term, respectively.

We believe, but can rarely prove, that if α lies in the minor arcs, the summands of $S(\alpha)$ behave like a random variable and thus one expects to obtain square-root cancellation for $S(\alpha)$. Usually one is not able to prove such strong cancellation, except for in very special situations. In any case, even if one assumes square-root cancellation, the circle method cannot handle arbitrarily few variables. Therefore there are theoretical limitations to the range within which one may hope the circle method to be effective.

Heath-Brown's work on 14 variables [49] is a refinement of previous work by Davenport [28] who showed that any cubic form in at least 16 variables represents zero non-trivially. In both papers, when the cubic form is of a somewhat degenerate shape then one can find a solution directly, via 'geometric reasons'. In the other cases the circle method is used to successfully establish an asymptotic formula, and using the above mentioned result by Lewis and Birch one can infer the existence of a non-trivial zero. Heath-Brown's main new innovation lies in the treatment of the minor arcs, where he introduces an averaged van der Corput differencing argument in addition to a classical Weyl differencing argument.

Of course all of these problems may be considered over number fields K, and a version of the circle method may be developed analogously. As mentioned above the local conditions for cubic forms were shown to be satisfied by Lewis, and his result holds for general number fields [72]. Using this Ramanujam [89] showed that any cubic form over a number field K in at least 54 variables represents zero non-trivially, which was subsequently improved to 17 variables by Ryavec [92] and 16 variables by Pleasants [87]. In joint work with Christian Bernert we established the following.

Theorem 1.0.1 (Theorem 2.1.1). *Let $K/\mathbb{Q}$ be an quadratic imaginary extension, and let $C(\boldsymbol{x}) \in K[x_1, \ldots, x_n]$ be a cubic form. If $n \geq 14$ then $C(\boldsymbol{x})$ represents zero non-trivially over K.*

The averaged van der Corput differencing argument prevents us from easily extending this result to arbitrary number fields. In particular, Heath-Brown's approach requires a good fractional version of Dirichlet's approximation theorem, which seems to be only valid in the case when K is a quadratic imaginary extension. Nevertheless, this result has some very interesting and effective applications to cubic hypersurfaces defined over $\mathbb{Q}$.

Theorem 1.0.2 (Theorem 2.1.2). *Let $X \subset \mathbb{P}^{n-1}_{\mathbb{Q}}$ be a hypersurface defined by a cubic form. If $n \geq 33$ then X contains a rational line.*

This improves upon previous work by Wooley [117], who showed that 37 variables suffice in order to deduce the existence of rational lines on cubic hypersurfaces. In fact we would like to compare this with recent work of Brandes and Dietmann [11] who showed that if one additionally assumes X to be non-singular then $n \geq 31$ suffices to infer the result. Using an argument in forthcoming work due to Brandes and Dietmann [12], which uses the full strength of Theorem 1.0.1, one could improve Theorem 1.0.2 to 31 variables. Finally as in work by Brüdern–Dietmann–Liu–Wooley [18], Theorem 1.0.1 can also be used in conjuction with the Green-Tao theorem in order to establish the following.

Theorem 1.0.3 (Theorem 2.1.3). *Let $C(\boldsymbol{x}) \in \mathbb{Q}[x_1, \ldots, x_n]$ be a cubic form. Then, if $n \geq 33$, there are almost-prime solutions to $C(\boldsymbol{x}) = 0$ in the following sense: There are coprime integers $c_1, \ldots, c_n$ such that the equation*

$$C(c_1p_1, c_2p_2, \ldots, c_np_n) = 0$$

has infinitely many solutions in primes $p_1, \ldots, p_n$, not all equal.

Returning to $\mathbb{Q}$, as mentioned above if C is non-singular, the state of the art is due to Hooley, who showed an asymptotic formula for non-singular cubic forms in at least 9 variables. In fact under the same assumption Hooley [57, 58] could still show an asymptotic formula if $n \geq 8$ and thus establish the Hasse Principle, assuming certain analytic, Riemann Hypothesis-like properties of Hasse–Weil L-functions.

Heath-Brown [46] explored the limits of this approach by considering diagonal cubic forms. Under similar assumptions regarding certain Hasse–Weil L-functions, he showed that the counting function satisfies

$$N(P) \ll_\varepsilon P^{3+\varepsilon}, \quad \text{if } n = 6$$

and

$$N(P) = \#\{|\boldsymbol{x}| \le P\colon \boldsymbol{x} \text{ lies on a rational line contained in } \mathbb{V}(C)\} + O_\varepsilon(P^{3/2+\varepsilon}),$$

if $n = 4$. Using elementary methods Hooley [55] could show the mean value estimate

$$\#\{\boldsymbol{x} \in \mathbb{Z}^6\colon \max_i |x_i| \le P, x_1^3 + x_2^3 + x_3^3 = x_4^3 + x_5^3 + x_6^3\} \ll_\varepsilon P^{3+\varepsilon},$$

under the same assumptions regarding Hasse–Weil L-functions that Heath-Brown required for his result. Assuming an unproved conjecture concerning the growth of the rank of rational elliptic curves in terms of their conductor, Heath-Brown [47] could improve the exponent to 4/3 in the case when $n = 4$. The result established in [47] holds for all non-singular and therefore not necessarily diagonal cubic forms in 4 variables. Heath-Brown further showed in [48] that certain families of cubic forms in $n = 4, 5$ variables satisfy the Hasse Principle, assuming Selmer's conjecture on elliptic curves.

If $n = 4$ then one expects the contribution from projective rational lines on the cubic surface to dominate, whenever they exist. For example, if $C(x_1, \dots, x_4) = x_1^3 + x_2^3 + x_3^3 + x_4^3$ then one obtains a contribution of $\gg P^2$ to $N(P)$ from solutions of the form $(a, -a, b, -b)$. If one were to consider a new counting function $N_0(P)$, which ignores the contribution from lines, then according to *Manin's conjecture* one would expect

$$N_0(P) = cP(\log P)^{\rho-1}(1 + o(1)),$$

where ρ is the Picard rank of the surface defined by $C = 0$ inside $\mathbb{P}^3$ and c is a constant as predicted by Peyre [84]. We will discuss Manin's conjecture in more detail in due course.

The above mentioned results by Hooley on octonary forms and Heath-Brown on diagonal cubic forms in 4 and 6 variables rely on a variant of the circle method, sometimes referred to as the *δ-method.* One of the key inputs is a smooth decomposition of the Kronecker δ-function, according to Duke–Friedlander–Iwaniec [35], which was then further developed in the context of the circle method by Heath-Brown [45].

Let $w\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth, compactly supported weight function. One may then consider a weighted counting function $N_w(P)$, which is defined via

$$N_w(P) = \sum_{\boldsymbol{x} \in \mathbb{Z}^n} \delta(F(\boldsymbol{x}))w(\boldsymbol{x}/P),$$

and normally an asymptotic formula or an upper bound for $N_w(P)$ is sufficient to establish one for $N(P)$. If one now expresses δ with the decomposition mentioned above, after dividing $\boldsymbol{x} \in \mathbb{Z}^n$ into residue classes, then one may apply Poisson summation in order to obtain an expression of the shape

$$N_w(P) = \sum_{\boldsymbol{c} \in \mathbb{Z}^n} \sum_{q=1}^{\infty} S_q(\boldsymbol{c}) I_q(\boldsymbol{c}),$$

where $S_q(\boldsymbol{c})$ and $I_q(\boldsymbol{c})$ are certain exponential sums and integrals. The main term is expected to arise from the contribution $\boldsymbol{c} = \boldsymbol{0}$, and the remaining summands should be absorbed by the error term. While it is possible to obtain very good estimates for $I_q(\boldsymbol{c})$ the situation for the exponential sums $S_q(\boldsymbol{c})$ is a little more tricky. Via Weil's work [116] one can get very good pointwise bounds for $S_q(\boldsymbol{c})$ since the sums correspond to the coefficients of local L-functions. If one wishes to estimate averages of the form

$$\sum_{q \leq Q} S_q(\boldsymbol{c}),$$

then it would be desirable to obtain cancellation in the summands instead of estimating each summand individually. One can understand these sums in terms of the coefficients of certain Hasse–Weil L-functions, which arise from taking the Euler product of the local L-functions that provided the pointwise bounds. Therefore it is at this point where assuming a Riemann hypothesis for such Hasse–Weil L-functions yields an improvement. Obtaining a saving over averaging these exponential sums is in this context also sometimes referred to as a *double Kloosterman refinement*.

The results of Hooley and Heath-Brown are conditional but in the context of function fields the corresponding L-functions are known to satisfy the Riemann hypothesis by virtue of Deligne's seminal work [30, 31]. Therefore it seems natural to consider whether one can prove the analogous results in this context unconditionally. Before we elaborate further we briefly introduce the basic notions and definitions that are needed in order to properly state the analogous Diophantine problem. Consider the field $K = \mathbb{F}_q(t)$ of rational functions in one variable whose coefficients lie in a finite field with $q = p^k$ elements, where $p \in \mathbb{N}$ is a prime number. Given an element $f/g \in K$, where $f, g \in \mathbb{F}_q[t]$, we may define an absolute value via

$$\left| \frac{f}{g} \right| := q^{\deg(f) - \deg(g)}.$$

We may consider ϖ-adic valuations induced by irreducible polynomials $\varpi \in \mathbb{F}_q[t]$, and thus construct completions K_ϖ. Further K has a completion induced by $1/t$ (the place at infinity), which can be explicitly described as the field of Laurent series in t^{-1}. We denote this by $K_\infty = \mathbb{F}_q((t^{-1}))$ and note that this corresponds to the real numbers in the classical setting. Given a cubic form $C(\boldsymbol{x}) \in K[x_1, \dots, x_n]$ one may now ask the question whether it represents zero non-trivially over K. It is a fairly straightforward consequence of the Chevalley–Warning theorem that this is indeed the case when $n \geq 10$. For smaller values of n, the local conditions need no longer be satisfied, but one may still enquire about the Hasse Principle. Again, the circle method can be effectively applied. Especially considering the observations noted in the paragraph above one might hope to achieve very strong results. To this end given $P \in \mathbb{F}_q[t]$ define a counting function

$$N(P) := \# \left\{ \boldsymbol{x} \in \mathbb{F}_q[t]^n \colon C(\boldsymbol{x}) = 0,\ |x_i| \leq |P|,\ \text{for all } i \right\},$$

and we study its asymptotic behaviour as $|P| \to \infty$. Corresponding to Hooley's theorem regarding octonary cubic forms, Browning and Vishe [17] established an asymptotic formula of the shape

$$N(P) = \sigma |P|^{n-3}(1 + o(1)),$$

if $n = 8$ and thus deduced the Hasse Principle for $n \geq 8$, provided $\operatorname{char}(\mathbb{F}_q) > 3$. The approach followed the work of Hooley and previous authors who had developed the circle method in this context, such as Lee [70] and Kubota [66]. The setup is as follows. One may define an additive character $\psi \colon K_\infty \to \mathbb{C}^\times$ given by

$$\psi \colon \sum_{i \leq N} a_i t^i \mapsto e\left(\frac{\operatorname{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(a_{-1})}{p} \right).$$

The unit interval in this setting, is given by $\mathbb{T} = K_\infty / \mathbb{F}_q[t]$ and can be explicitly described as

$$\mathbb{T} = \{ \alpha \in K_\infty \colon |\alpha| < 1 \} = \left\{ \sum_{i \leq -1} a_i t^i \colon a_i \in \mathbb{F}_q \right\}.$$

It is then not hard to verify that given $x \in \mathbb{F}_q[t]$, we have

$$\int_{\mathbb{T}} \psi(\alpha x) d\alpha = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Therefore we may consider a weighted counting function

$$N_w(P) = \sum_{\boldsymbol{x} \in \mathbb{F}_q[t]^n} \int_{\mathbb{T}} \psi(\alpha F(\boldsymbol{x})) w(\boldsymbol{x}/P) d\alpha,$$

where $w \colon \mathbb{F}_q[t]^n \to \mathbb{R}$ is some weight function. Similarly to the procedure described above one may from here onwards apply Poisson summation and the problem reduces to estimating certain exponential sums and integrals. This is the approach used by Browning–Vishe. In joint work with Jakob Glas we refined their method and established Heath-Brown's result in this context.

Theorem 1.0.4 (Theorem 3.1.1)**.** *Let $C(\boldsymbol{x}) = \sum_{i=1}^n a_i x_i^3 \in \mathbb{F}_q[t][x_1, \dots, x_n]$ be a diagonal cubic form. We have*

$$N(P) \ll_\varepsilon |P|^{3+\varepsilon}, \quad \textit{if } n = 6 \textit{ and } \operatorname{char}(\mathbb{F}_q) \neq 3,$$

and

$$N_0(P) \ll_\varepsilon |P|^{3/2+\varepsilon}, \quad \textit{if } n = 4 \textit{ and } \operatorname{char}(\mathbb{F}_q) \neq 2, 3,$$

where $N_0(P)$ counts the number of solutions $\boldsymbol{x} \in \mathbb{F}_q[t]^n$ away from lines on the cubic hypersurface defined by $C(\boldsymbol{x}) = 0$ inside $\mathbb{P}^3_{\mathbb{F}_q}$.

We note that the restriction on the characteristic arises usually quite naturally when one wishes to apply the circle method in a function field setting, see for example the work of Kubota [66] and Lee [70]. This is an artefact from Weyl's estimate, which produces a factor of $d!$ in the exponential sum when one is dealing with polynomials of degree d. Of course, if the characteristic is then smaller than d the exponential function may only be estimated trivially when such a factor is produced. We note that in the case of Browning–Vishe the restriction on the characteristic arises in a slightly different manner. They require a point $\boldsymbol{x}_0$ at which the Hessian does not vanish. If $\operatorname{char}(\mathbb{F}_q) = 2$ or 3 such a point may not exist.

It is therefore a very nice feature that we could show the first part of the above theorem if the characteristic is 2. We also note that in the case when $\operatorname{char}(\mathbb{F}_q) = 3$ solving a diagonal cubic form over $\mathbb{F}_q[t]$ reduces to a system of linear equations, and so in a sense this is a less interesting case.

Further we note that a key difficulty in the work of Browning–Vishe was the unavailability of a suitable form of partial summation in order to take full advantage of the double Kloostermann refinement. This led to a rather complicated workaround that resulted in a slight loss in their estimates. The associated loss in the estimates was too big for our purposes. However, we managed to remedy this with a rather simple trick, which should be applicable in their setting too. We think that this idea can be applied in similar problems over function fields in the future.

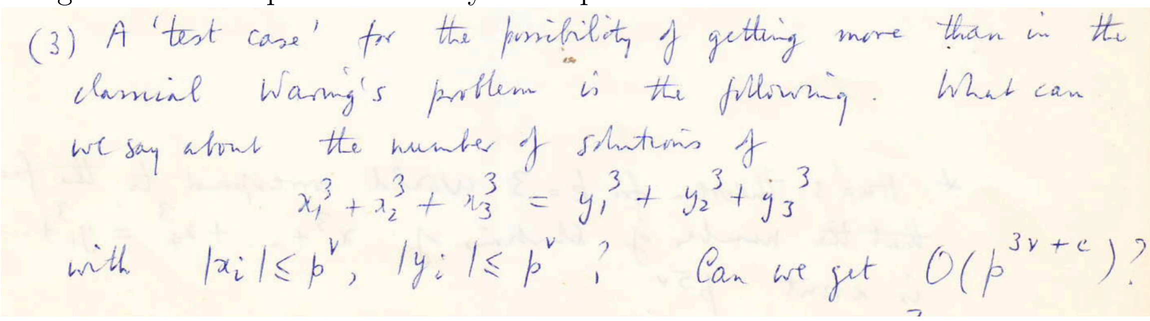
In 1964 Harold Davenport asked his PhD student at the time Keith Matthews: *"What can we say about the number of solutions of*

$$x_1^3 + x_2^3 + x_3^3 = y_1^3 + y_2^3 + y_3^3$$

with $x_i, y_i \in \mathbb{F}_q[t]$, $|x_i|, |y_i| \leq P$*? Can we get* $O(P^{3+\varepsilon})$*?"*

By considering $C(\boldsymbol{x}) = x_1^3 + x_2^3 + x_3^3 - x_4^3 - x_5^3 - x_6^3$ and applying Theorem 1.0.4 we may thus affirmatively answer Davenport's question. Davenport was enquiring about

Figure 1.1: Excerpt of a letter by Davenport written to Keith Matthews in 1964


(3) A 'test case' for the possibility of getting more than in the classical Waring's problem is the following. What can we say about the number of solutions of

$$x_1^3 + x_2^3 + x_3^3 = y_1^3 + y_2^3 + y_3^3$$

with $|x_i| \leq p^\nu$, $|y_i| \leq p^\nu$? Can we get $O(p^{3\nu+\varepsilon})$?

such an upper bound due to its relevance in connection with Waring's problem for cubes. Over the integers Waring's problem for cubes asks the following: What is the smallest positive integer s such that every sufficiently large integer may be written as the sum of at most s non-negative cubes? If we call this integer $G(3)$ then it is conjectured that $G(3) = 4$. Using a circle method approach one can tackle this problem by understanding the counting function

$$R_3(N) = \#\{\boldsymbol{x} \in \mathbb{Z}^s : x_i \geq 0, x_1^3 + \cdots + x_s^3 = N\},$$

as $N \to \infty$. The current state of the art is due to Vaughan [111, 112] who established an asymptotic formula if $s \geq 8$, and lower bounds of the correct order of magnitude for $R_3(N)$ if $s \geq 7$ from which one may then conclude $G(3) \leq 7$. Using algebraic methods Linnik [75] had already shown $G(3) \leq 7$ some years before Vaughan, but he proved no quantitative result regarding $R_3(N)$.

One may analogously consider such problems over function fields. Similarly one seeks to find the minimal integer s such that every polynomial P of sufficiently large degree may be written as the sum of s polynomials. Since there may be 'trivial' obstructions to this, one usually takes P to lie in the additive closure of the cubes $\mathbb{J}^{(3)}[t] \subset \mathbb{F}_q[t]$, in order for this question to be well defined. For example, given any positive integer n it is never possible to express t^{3n+1} as a sum of any number of cubes in $\mathbb{F}_3[t]$. We note at this point that only in the case $q = 2, 4$ or $q = 3^k$ such obstructions may occur, otherwise $\mathbb{J}^{(3)}[t] = \mathbb{F}_q[t]$, see for example [39, Lemma 5.2]. Therefore we denote by $G_q(3) = s$ the smallest integer such that every polynomial $P \in \mathbb{J}^{(3)}[t]$ of sufficiently large degree may be written as the sum of s cubes. One may also consider a counting function associated to this problem, namely

$$R_q(P) = \#\{\boldsymbol{x} \in \mathbb{F}_q[t]^n \colon |\boldsymbol{x}| \leq q^{\lceil \frac{\deg(P)}{3} \rceil}, \, x_1^3 + \cdots + x_s^3 = P\},$$

where we restricted the size of the potential summands since there may be infinitely many polynomials of large degree, which could cancel each other out – a phenomenon that does not occur over the integers, but as a result of which the counting function may otherwise not be finite for a given P. We denote by $\widetilde{G}_q(3)$ the smallest integer s such that we obtain an asymptotic formula for $R_q(P)$. Clearly $G_q(3) \leq \widetilde{G}_q(3)$. Further it is also trivial to see that $\widetilde{G}_{3^h}(3) = 1$ holds. One of the first people to study Waring's problem in this context using the circle method was Kubota [66] who showed that $\widetilde{G}_q(3) \leq 9$ holds provided that $2 \nmid q$. As before the restriction on the characteristic arises from a Weyl differencing procedure. In characteristic 2 the current best bounds available are due to Car–Cherly [20] who established $\widetilde{G}_{2^h}(3) \leq 11$. Finally, using elementary approaches Gallardo [38] and Car–Gallardo [21] showed

$$G_q(3) \leq \begin{cases} 7, & \text{if } q \notin \{7, 13, 16\} \\ 8, & \text{if } q \in \{13, 16\} \\ 9, & \text{if } q = 7. \end{cases}$$

Returning to Davenport's question, one should note that in a circle method approach to Waring's problem the count of the number of bounded solutions to

$$x_1^3 + x_2^3 + x_3^3 = y_1^3 + y_2^3 + y_3^3$$

arises very naturally and is usually handled using Hua's Lemma. In the function field setting we also note that Hua's Lemma only works when the characteristic is at least 5. Using Theorem 1.0.4 we manage to improve upon all of these results in one big sweep.

Theorem 1.0.5 (Theorem 3.1.4)**.** *We have* $\widetilde{G}_q(3) \leq 7$.

We note that if $q = 2^k$ we used a Weyl estimate, which was established by Car [19]. We may moreover use Theorem 1.0.4 in order to deduce the Hasse Principle for diagonal forms over $\mathbb{F}_q[t]$ in at least 7 variables, building on the work of Lee [70]. In fact, we can establish *weak approximation.* This is a stronger form of the Hasse principle, which states that if X is a variety over a global field k then the diagonal embedding

$$X(k) \hookrightarrow \prod_v X(k_v),$$

is dense with respect to the product topology, where the product runs over all the places of k. In more down-to-earth terms, in the case when one has a hypersurface defined by $f = 0$ with k_v-solutions for every place v, then it satisfies weak approximation precisely when given a finite set S of places of k, given $x_v \in k_v$ for all $v \in S$ with $f(x_v) = 0$ and given any $\varepsilon > 0$, then there exists a rational solution x to $f = 0$ satisfying

$$|x - x_v|_v < \varepsilon$$

for all $v \in S$. For completeness we should mention that some of the previously mentioned results regarding the Hasse Principle actually also showed weak approximation, such as the result by Browning–Vishe [46] or the work of Lee [70]. In fact, with enough care the circle method is usually capable of showing weak approximation when one can show the Hasse principle, although the technical details become more involved.

Theorem 1.0.6 (Theorem 3.1.3)**.** *Let* $C(\boldsymbol{x}) \in \mathbb{F}_q[t][x_1, \ldots, x_n]$ *be a diagonal cubic form. If* $n \geq 7$ *and if* $\operatorname{char}(\mathbb{F}_q) > 3$ *holds then weak approximation holds for the hypersurface defined by* $C(\boldsymbol{x}) = 0$.

We note that one should be able to prove this result in the case when $\operatorname{char}(\mathbb{F}_q) = 2$ by proving a suitable version of Weyl's inequality in this context using the ideas by Car [19].

So far we have hopefully demonstrated how the circle method can be effectively used in order to obtain detailed information about integer solutions to homogeneous degree 3 equations, as well as their analogues in number fields and function fields. However, the circle method is a very flexible tool that can be used in a variety of situations. One longstanding and very general result is due to Birch [9]. Consider a system of homogeneous equations $F_1, \ldots, F_R \in \mathbb{Z}[x_1, \ldots, x_n]$ which are all of degree d and assume the equations $F_1 = \cdots = F_R = 0$ define a complete intersection inside

$\mathbb{A}^n$. Let $\mathcal{B} \subset \mathbb{R}^n$ be a box whose edges are parallel to the coordinate axes and with sidelengths at most 1. Given $P \geq 1$ denote by $N(P)$ the number of integral solutions $\boldsymbol{x} \in P\mathcal{B} \cap \mathbb{Z}^n$ to this system. Birch showed if n is sufficiently large then one obtains an asymptotic of the form

$$N(P) = \sigma P^{n-dR}(1 + o(1)), \tag{1.0.1}$$

where $\sigma > 0$ provided the system has a non-singular zero over $\mathbb{R}$ inside $\mathcal{B}$ and non-singular zeroes over all $\mathbb{Q}_p$. More precisely, if we define the *Birch singular locus* to be the variety $V^* \subset \mathbb{A}^n$ defined by

$$\operatorname{rank} \left(\frac{\partial F_i}{\partial x_j} \right)_{i,j} < R,$$

then if

$$n - \dim V^* > R(R+1)(d-1)2^{d-1},$$

he showed that the asymptotic above holds. It is interesting but hardly surprising that the geometry of the equations plays a role in the number of variables needed for this result. We note also that if $\mathbb{V}(F_1, \dots, F_R) \subset \mathbb{A}^n$ is non-singular then one can show $\dim V^* \leq R-1$. Birch's result has subsequently been generalized to many other settings. To list just a few, for example Skinner [105] generalized Birch's result to number fields, Browning–Heath-Brown [14] considered a system of forms of differing degrees, Cook–Magyar [24] and Yamagishi [118] considered prime solutions to systems of forms, and Schindler [98] considered systems of bihomogeneous forms.

A recent breakthrough by Rydin Myerson [94, 95, 93] improved Birch's result significantly if the number of forms considered is large. If we denote by

$$\sigma_{\mathbb{R}} \coloneqq 1 + \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\mathbf{0}\}} \dim \operatorname{Sing} \mathbb{V} \left(\sum_{i=1}^R \beta_i F_i \right),$$

where we regard $\mathbb{V}\left(\sum_{i=1}^R \beta_i F_i\right) \subset \mathbb{P}^{n-1}$, then Rydin Myerson showed the asymptotic formula as in (1.0.1) provided

$$n - \sigma_{\mathbb{R}} > d2^d R$$

holds in the cases when $d = 2$ or $d = 3$. We note that the key novelty here is that the number of equations R appears only linearly, as opposed to quadratically, in the required number of variables. Rydin Myerson further shows that $\sigma_{\mathbb{R}} \leq \dim V^*$ always holds and therefore this also constitutes an improvement compared to Birch. In his case the quantity $\sigma_{\mathbb{R}}$ arises naturally instead of $\dim V^*$ via the methods he employs. Similar improvements, where the pencil of the system was considered instead of the Birch singular locus were proved by Schindler [99] and Dietmann [33], and very recently Yamagishi [119] replaced the Birch singular locus with a condition regarding the Hessian of the system.

Inspired by work of Müller [80, 79] on systems of quadratic inequalities Rydin Myerson's results are proved using a so-called *auxiliary inequality*. This auxiliary inequality can be used in order to exhibit a sort of repulsion behaviour for the size of the exponential sum involved for pairs of values of $\alpha \in [0,1]^R$ in the minor arcs. In order to obtain this inequality it is necessary to find a good upper bound for the number of integral solutions of bounded height to a multilinear Diophantine inequality. In the case when the degree is $d = 2$ or $d = 3$ he is able to achieve this whenever $n - \sigma_{\mathbb{R}} > d2^d R$ is satisfied, provided $\mathbb{V}(F_1, \dots, F_R)$ is a complete intersection. For higher degree he manages to establish this too for generic systems [93], in the sense that he identifies a Zariski open, nonempty subset in the coefficient space for the system of forms.

The principal objective of Chapter 4 is to apply Myerson's techniques to systems of bihomogeneous forms, and therefore improve the previous result by Schindler [98] in this direction. Consider forms $F_1(\boldsymbol{x}, \boldsymbol{y}), \dots, F_R(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{Z}[x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}]$, which are bihomogeneous of common bidegree (d_1, d_2). This means that for scalars $\lambda, \mu \in \mathbb{C}$ we have $F_i(\lambda \boldsymbol{x}, \mu \boldsymbol{y}) = \lambda^{d_1} \mu^{d_2} F_i(\boldsymbol{x}, \boldsymbol{y})$. Similar to the above one may introduce a counting function

$$N(P_1, P_2) = \# \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{Z}^{n_1+n_2} \colon F_i(\boldsymbol{x}, \boldsymbol{y}) = 0, \text{ for all } i \;\; \boldsymbol{x} \in P_1 \mathcal{B}_1, \; \boldsymbol{y} \in P_2 \mathcal{B}_2 \right\},$$

where $P_1, P_2 \geq 1$ are two real numbers and $\mathcal{B}_i \subset \mathbb{R}^{n_i}$ are two boxes whose edges are parallel to the coordinate axes with length at most 1. Similar to the Birch singular locus, Schindler defines varieties $V_1^*, V_2^* \subset \mathbb{A}^{n_1+n_2}$ given by

$$\operatorname{rank} \left(\frac{\partial F_i}{\partial x_j} \right)_{i,j} < R, \quad \text{and} \quad \operatorname{rank} \left(\frac{\partial F_i}{\partial y_j} \right)_{i,j} < R,$$

respectively. Write $b = \max \left\{ \frac{\log(P_1)}{\log(P_2)}, 1 \right\}$ and $u = \max \left\{ \frac{\log(P_2)}{\log(P_1)}, 1 \right\}$. The main result in [98] is that, provided

$$n_1 + n_2 - \dim V_i^* > 2^{d_1+d_2-2} \max\{R(R+1)(d_1+d_2-1), R(bd_1 + ud_2)\}$$

holds for $i = 1, 2$ then one obtains an asymptotic formula of the shape

$$N(P_1, P_2) = \sigma P_1^{n_1 - d_1 R} P_2^{n_2 - d_2 R} (1 + O(\min\{P_1, P_2\}^{-\delta})),$$

for some δ. As in all the previous results mentioned in this introduction, the leading constant σ can be interepreted as the product of p-adic and real zeroes of the system of bihomogeneous equations under investigation. In particular $\sigma > 0$ if the system of equations has non-singular p-adic zeroes for every p and a non-singular real zero in $\mathcal{B}_1 \times \mathcal{B}_2$.

The main result of Chapter 4 concerns systems of bidegree $(1,1)$ and $(2,1)$, which correspond to degree 2 and 3 in Myerson's case. Similar to Myerson's result, the quantity defined by Schindler, which is analogous to the Birch singular locus will be replaced with various pencils of certain varieties, which are slightly more complicated to define. We begin by stating the theorem for systems of bilinear forms. To this end, note that we may write each such form as

$$F_i(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^T A_i \boldsymbol{x},$$

where A_i is an $n_2 \times n_1$ matrix with integer coefficients. Given $\boldsymbol{\beta} \in \mathbb{R}^R$ write $A_{\boldsymbol{\beta}} = \sum_{i=1}^R \beta_i A_i$ for the linear combination of these matrices defined by $\boldsymbol{\beta}$. Define now the quantities

$$\sigma_{\mathbb{R}}^{(1)} := \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}} \dim \ker(A_{\boldsymbol{\beta}}), \quad \text{and} \quad \sigma_{\mathbb{R}}^{(2)} := \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}} \dim \ker(A_{\boldsymbol{\beta}}^T).$$

We note also that since the bidegree is $(1,1)$ the situation is completely symmetric in $\boldsymbol{x}$ and $\boldsymbol{y}$ and so we may without loss of generality state the result assuming $P_1 \geq P_2 > 1$.

Theorem 1.0.7 (Theorem 4.1.1)**.** *Let $F_1, \dots, F_R$ be bilinear forms defining a complete intersection $X \subset \mathbb{P}^{n_1-1} \times \mathbb{P}^{n_2-1}$. Let $P_1 \geq P_2 > 1$ be real numbers and write $b = \log(P_1)/\log(P_2)$. If*

$$n_i - \sigma_{\mathbb{R}}^{(i)} > (2b+2)R$$

is satisfied for $i = 1, 2$, then the asymptotic formula

$$N(P_1, P_2) = \sigma P_1^{n_1 - R} P_2^{n_2 - R}(1 + O(P_2^{-\delta})),$$

holds for some $\delta > 0$. In particular, if X is non-singular then the asymptotic formula holds provided

$$\min\{n_1, n_2\} > (2b+2)R \quad \textit{and} \quad n_1 + n_2 > (4b+5)R$$

is satisfied. The constant σ is positive if the system has a non-singular real zero in $\mathcal{B}_1 \times \mathcal{B}_2$ and non-singular p-adic zeroes for all p.

We now move on to systems of bidegree $(2,1)$. We may express a bihomogeneous polynomial of such bidegree as

$$F_i(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^T H_i(\boldsymbol{y})\boldsymbol{x},$$

where $H_i(\boldsymbol{y})$ is a symmetric $n_1 \times n_1$ matrix whose entries are linear homogeneous forms in $\boldsymbol{y}$ with coefficients in $\mathbb{Z}$. Given $\boldsymbol{\beta} \in \mathbb{R}^R$ we write

$$H_{\boldsymbol{\beta}}(\boldsymbol{y}) = \sum_{i=1}^{R} \beta_i H_i(\boldsymbol{y}).$$

For $\ell = 1,\dots,n_2$ write $\boldsymbol{e}_\ell$ for the unit standard basis vectors inside $\mathbb{R}^{n_2}$, and consider the intersection of pencils given by

$$\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_\ell \coloneqq \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_1)\boldsymbol{x},\dots,\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_{n_2})\boldsymbol{x}) \subset \mathbb{P}^{n_1-1}.$$

Define now

$$s_{\mathbb{R}}^{(1)} \coloneqq 1 + \max_{\boldsymbol{\beta}\in\mathbb{R}^R\setminus\{0\}} \dim \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_\ell.$$

Considering the system of pencils $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) = \{H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x} = \boldsymbol{0}\} \subset \mathbb{P}^{n_1-1}\times\mathbb{P}^{n_2-1}$ we define

$$s_{\mathbb{R}}^{(2)} \coloneqq 1 + \left\lfloor \frac{\max_{\boldsymbol{\beta}\in\mathbb{R}^R\setminus\{0\}} \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})}{2} \right\rfloor.$$

Theorem 1.0.8 (Theorem 4.1.2)**.** *Consider forms $F_1(\boldsymbol{x},\boldsymbol{y}),\dots,F_R(\boldsymbol{x},\boldsymbol{y})$ of bidegree $(2,1)$ defining a complete intersection $X \subset \mathbb{P}^{n_1-1}\times\mathbb{P}^{n_2-1}$. Let $P_1,P_2>1$ be real numbers and write $b = \max\left\{\frac{\log(P_1)}{\log(P_2)},1\right\}$ and $u = \max\left\{\frac{\log(P_2)}{\log(P_1)},1\right\}$. If the number of variables satisfies*

$$n_1 - s_{\mathbb{R}}^{(1)} > (8b+4u)R \quad \textit{and} \quad \frac{n_1+n_2}{2} - s_{\mathbb{R}}^{(2)} > (8b+4u)R,$$

then we have

$$N(P_1,P_2) = \sigma P_1^{n_1-2R} P_2^{n_2-R}(1+O(\min\{P_1,P_2\}^{-\delta})),$$

for some $\delta>0$. In particular, if X is non-singular then the asymptotic formula holds provided

$$n_1 > (16b+8u+1)R, \quad \textit{and} \quad n_2 > (8b+4u+1)R$$

is satisfied. The constant σ is positive if the system has a non-singular real zero in $\mathcal{B}_1\times\mathcal{B}_2$ and non-singular p-adic zeroes for all p.

Note that $N(P_1, P_2) \gg P_1^{n_1} + P_2^{n_2}$ which becomes apparent upon considering solutions of the form $(x_1, \dots, x_{n_1}, 0, \dots, 0)$ and $(0, \dots, 0, y_1, \dots, y_{n_2})$. Therefore, if an asymptotic of the above form holds then this implies

$$P_1^{n_1} + P_2^{n_2} \ll P_1^{n_1 - d_1 R} P_2^{n_2 - d_2 R}.$$

An easy calculation reveals that this forces

$$n_i > R(bd_1 + ud_2).$$

Therefore one may not, in general, hope to achieve a better result than what was stated in the previous two theorems, up to a linear factor.

One of the main motivations for considering such counting problems, aside from the fact that they are interesting in their own right in order to understand the distribution of rational points on varieties, is the potential to prove Manin's Conjecture for biprojective complete intersections, in sufficiently many variables.

Let X be a *Fano Variety* over a global field k. That is, X is a smooth projective variety such that the inverse of the canonical bundle ω_X in the Picard group is ample. Together with a choice of global sections of $(\omega_X^{-1})^m$ for some $m > 0$, this yields an *anticanonical* height function h_X, say. Manin's conjecture studies the counting function

$$N_U(P) \coloneqq \{x \in U(k) \colon h_X(x) \leq P\},$$

where $U(k) \subset X(k)$ is some subset of $X(k)$. The conjecture was first formulated by Manin [37] and Batyrev–Manin [5]. It states that if X is a Fano variety over k such that $X(k)$ is Zariski dense in X then there exists a Zariski open subset $U \subset X$ such that

$$N_U(P) \sim cP(\log P)^{\rho - 1},$$

where ρ is the Picard rank of X and c is a constant, which has received a detailed interpretation by Peyre [84]. The restriction to such an open subset is certainly necessary. For example, considering the cubic surface defined by $x_1^3 + \dots + x_4^3 = 0$ in $\mathbb{P}^3$ one obtains a contribution of $\gg P^2$ to $N_X(P)$ from rational points coming from the rational lines contained in this surface. The general idea is that there might be 'bad' accumulating subsets, which contribute disproportionately much to $N_X(P)$. The conjecture states that all of these accumulating subsets should take the form of Zariski closed subsets. The interpretation of the leading constant by Peyre along with the appearance of the Picard rank in the power of the logarithm in the formula above is a fascinating example of how the geometry of a variety (conjecturally) determines its arithmetic. Two large classes of varieties for which this conjecture is proven are for example flag varieties, which was established by Franke–Manin–Tschinkel [37] and toric varieties, which was proven by Batyrev–Tschinkel [5].

It turns out that the conjecture is false in this formulation. This was first demonstrated by Batyrev and Tschinkel [6] in the case when the field is $k = \mathbb{Q}(\sqrt{-3})$. This was subsequently generalized to arbitrary number fields by Loughran [77]. Another counterexample was found by Browing–Heath-Brown [15]. The conjecture was subsequently revised by Peyre [84], who proposes that one expects the asymptotic $N_U(P) \sim cP(\log P)^{\rho-1}$ to hold for some $U \subset X$ such that $(X \setminus U)(k)$ is a *thin* set. We recall the definition of thin sets as defined in Serre [103]. We call $A \subset X(k)$ to be of type

(C_1) if $A \subseteq Y(k)$, where $Y \subsetneq X$ is Zariski closed,

(C_2) if $A \subseteq \pi(X'(k))$, where X' is irreducible such that $\dim X = \dim X'$ and $\pi \colon X' \to X$ is a generically finite morphism of degree at least 2.

A subset of $X(k)$ is said to be *thin* if it is a finite union of sets of type (C_1) or (C_2). Since Zariski closed subsets are thin of type (C_1) this notion strictly generalises the permissible accumulating subset that one may remove in Manin's conjecture. It should be noted that Peyre has proposed two other reformulations on Manin's conjecture. One reformulation is sometimes referred to as an *all heights* approach [86], taking into account the different height functions which one may consider arising from other very ample line bundles on X. On the other hand, he also proposed a notion of *freeness*, and that one should be able to recover the desired asymptotic if one removes points of a certain freeness [85]. The latter approach was recently shown to be insufficient on its own by Sawin [97].

Returning to bihomogeneous varieties, Schindler [100] successfully verified Manin's conjecture for certain complete intersections in biprojective space defined by forms of bidegree (d_1, d_2) provided $d_1, d_2 \geq 2$ and if the number of variables is sufficiently large in terms of the bidegree and the number of forms involved. To achieve this she combined her results regarding the box count on bihomogeneous varieties [98] as described above along with a uniform count of the number of integral solutions of bounded height on certain fibres of the canonical projections $\pi_i \colon X \to \mathbb{P}^{n_i-1}$. These counting problems were then merged with a variation of the hyperbola method as developed by Blomer–Brüdern [10].

It would be interesting to explore an application of Theorem 1.0.8 to show Manin's conjecture for certain complete intersections of bihomogeneous hypersurfaces of bidegree $(2, 1)$. In particular using Theorem 1.0.8 one should be able to prove Manin's conjecture requiring fewer variables than what Schindler's approach is expected to require. In particular it would be very interesting to see whether one could obtain the desired asymptotic using the all heights approach proposed by Peyre. Biprojective varieties are in some sense one of the "simplest" class of varieties for which Peyre's all height approach is genuinely different from the other formulations of Manin's conjecture.

The final chapter of this thesis considers a multiplicative rather than an additive problem, namely *Artin's primitive root conjecture.* Recall that given a rational prime p the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic. We say that a rational number $g \in \mathbb{Q}$ is a primitive root modulo p if $v_p(g) = 0$ and if the reduction of g modulo p generates $(\mathbb{Z}/p\mathbb{Z})^\times$. We also say p is an *Artin prime* for g if this happens. One may wonder, for a given integer g, whether there are infinitely many Artin primes p for g. If g is a square then this will not be the case since $(\mathbb{Z}/p\mathbb{Z})^\times = p - 1$ and $2 \mid p - 1$ for all odd primes p. Similarly, for trivial reasons $g = \pm 1$ will not be primitive modulo infinitely many primes.

Artin conjectured in 1930 that any $g \in \mathbb{Q}^\times \setminus \{\pm 1\}$, which is not a square, is a primitive root modulo infinitely many primes. Based on a fairly simple heuristic, he conjectured a density of Artin primes for a given such g. Write h for the largest positive integer such that we can write $g = b^h$ for some rational number b. It is easy to check that p is an Artin prime for g if and only if p does not split completely in any of the splitting fields for $x^q - g$, which we denote by

$$K_q = \mathbb{Q}(\zeta_q, g^{1/q}),$$

where q is a prime. According to Chebotarev's density theorem the density of primes p splitting completely in K_q is given by $\frac{1}{[K_q \colon \mathbb{Q}]}$. Now we have

$$[K_q \colon \mathbb{Q}] = \begin{cases} \frac{1}{q(q-1)}, & \text{if } q \nmid h \\ \frac{1}{q-1}, & \text{if } q \mid h. \end{cases}$$

If we assume those splitting conditions to be independent from each other, and we ignore any kind of error terms in Chebotarev's theorem then one might expect

$$\#\{p \leq x \colon p \text{ is an Artin prime for } g\} \sim \prod_{q \nmid h} \left(1 - \frac{1}{q(q-1)}\right) \prod_{q \mid h} \left(1 - \frac{1}{q-1}\right) \frac{x}{\log x}.$$

Indeed this was the conjectured density according to Artin. Based on numerical computations by the Lehmers [71] the conjecture was revised by Artin, and first seems to have appeared in the 'correct' form in the preface of his collected works [1] edited by Lang and Tate. In particular, the assumption that the splitting conditions are independent is, in general, wrong. This causes the need for an additional *correction factor* in the density to account for such phenomena. A good overview of the history of this is given by Stevenhagen [108]. Hooley [50] proved the asymptotic, including the correction factor as stated by Lang and Tate, under the assumption of the generalised Riemann hypothesis for certain Dedekind ζ-functions.

Not so much is known unconditionally. One remarkable result in this direction is due to Heath-Brown [44], who showed that there are at most two primes, for which the conjecture fails. In particular, on member of the set $\{2, 3, 5\}$ must be a primitive root modulo infinitely many primes.

One may consider this problem over function fields. Many multiplicative number theoretic problems become significantly more approachable if one works in this setting. Instead of $\mathbb{Q}$ we now consider the field $\mathbb{F}_q(t)$, which is the function field of the projective line over $\mathbb{F}_q$. Rational primes now correspond to irreducible polynomials. Given an irreducible polynomial $p(t) \in \mathbb{F}_q[t]$ of degree n we have

$$\mathbb{F}_q[t]/(p(t)) \cong \mathbb{F}_{q^n},$$

and so the group $(\mathbb{F}_q[t]/(p(t)))^\times$ is cyclic of order $q^n - 1$. One may therefore ask again, given $g(t) \in \mathbb{F}_q(t)$, are there infinitely many irreducible polynomials $p(t)$ with $v_{p(t)}(g(t)) = 1$ such that $g(t)$ generates $(\mathbb{F}_q[t]/(p(t)))^\times$? As in the integer case, there are some obvious obstructions to this being true for certain candidate elements $g(t)$. For example, if $g(t) \in \mathbb{F}_q$ then clearly $g(t)$ is not a primitive root modulo any polynomial of degree at least 2. Further note that

$$q^n - 1 = (q - 1)(q^{n-1} + \cdots + 1),$$

and so $q - 1 \mid q^n - 1$ for any $n \geq 1$. Therefore as soon as $g(t)$ can be written as a ℓ-th power, for some prime $\ell \mid q - 1$, then similarly $g(t)$ fails to be primitive modulo any irreducible polynomials of degree at least 2. Artin's primitive root conjecture states in this context that given any $g(t) \in \mathbb{F}_q(t)^\times \setminus \mathbb{F}_q^\times$, which is not an ℓ-th power for a prime $\ell \mid q - 1$, there are infinitely many irreducible polynomials $p(t) \in \mathbb{F}_q[t]$ such that $g(t)$ is a primitive root modulo $p(t)$. One may also generalise an analogous version for the conjecture when we replace $\mathbb{F}_q(t)$ by a function field of any suitably nice variety over $\mathbb{F}_q$, and we will formulate this generalisation in due course.

Bilharz [7] considered this problem as part of his PhD thesis which supervised by Hasse at the University of Göttingen in the 1930s. Indeed he managed to prove Artin's primitive root conjecture for function fields of any non-singular curve over $\mathbb{F}_q$, under the assumption of a suitable Riemann hypothesis over function fields, which was later famously proven by Weil [116]. Bilharz' proof, however, contains a gap. In particular the proof remains only valid when the candidate elements $g(t)$ are *geometric*. In the case when the function field is $\mathbb{F}_q(t)$ this means that $g(t)$ is not of the form $g(t) = \mu h(t)^\ell$ for some rational prime ℓ different from the characteristic of $\mathbb{F}_q$ and where $\mu \in \mathbb{F}_q$. In this case, Bilharz' computations of the degrees of certain field extensions were not correct. This was already observed by Rosen [91, page 157] and a full proof of the conjecture seems not to be available in the literature. It is interesting to note the similarity of this with the history regarding the correction factor in the classical setting.

One of the principal goals of Chapter 5 is to establish a proof of Artin's primitive root conjecture in all cases. At the same time we generalise the problem to function fields of varieties of arbitrary dimension. In particular, we recover any field of finite transcendence degree over $\mathbb{F}_q$. In order to state our results we need to introduce some more language.

Let X be a geometrically integral projective variety over $\mathbb{F}_q$ of dimension r with function field denoted by K. We note that these assumptions imply that the maximal algebraic field extension of $\mathbb{F}_q$ inside K is given by $\mathbb{F}_q$. Given $g \in K$ we say that g is regular at a closed point $\mathfrak{p} \in X$ if it lies in the image of the natural embedding $\mathcal{O}_{X,\mathfrak{p}} \hookrightarrow K$. In this case we may consider the reduction of g to the residue field $\kappa_\mathfrak{p}$ at $\mathfrak{p}$. Note that $\kappa_\mathfrak{p}$ is isomorphic to a finite field extension of $\mathbb{F}_q$ and therefore $\kappa_\mathfrak{p}^\times$ is a cyclic group. We say that g is a primitive root modulo a closed point $\mathfrak{p}$ if g is regular at $\mathfrak{p}$, if its reduction to $\kappa_\mathfrak{p}$ is non-zero and generates $\kappa_\mathfrak{p}^\times$. We also call $\mathfrak{p}$ an Artin prime for g. Artin's primitive root conjecture therefore generalises in this context to say the following: Given $g \in K \setminus \mathbb{F}_q$, which is not an ℓ-th power for a prime $\ell \mid q-1$ then there exist infinitely many closed points $\mathfrak{p} \in X$ for which g is a primitive root. For example, if $X = \mathbb{P}^1_{\mathbb{F}_q}$ then this is equivalent to Artin's primitive root conjecture for $\mathbb{F}_q(t)$ as described above.

As noted above, Bilharz proved this conjecture whenever X is a curve in the cases when g is not geometric. There are a few different notions in the literature of what it means for $g \in K$ to be geometric. In Chapter 5 we prove a lemma, which shows that they are all equivalent. In particular, we can say that $g \in K$ is *not* geometric at a rational prime $\ell \neq \mathrm{char}(\mathbb{F}_q)$ if there exists some $\mu \in \mathbb{F}_q$ and $b \in K$ such that $g = \mu b^\ell$. Given $g \in K$ define

$$\mathscr{P}_g := \{\ell \in \mathbb{Z}_{\mathrm{prime}} \colon g \text{ is not geometric at } \ell\}.$$

We say that $\deg \mathfrak{p} = n$ if $[\kappa_{\mathfrak{p}} \colon \mathbb{F}_q] = n$, and write X_0 for the closed points of X. We consider the following counting function

$$N_X(g,n) = \#\{\mathfrak{p} \in X_0 \colon \deg \mathfrak{p} = n \text{ and } g \text{ is a primitive root modulo } \mathfrak{p}\}.$$

In joint work with Ezra Waxman, the main theorem of Chapter 5 is the following.

Theorem 1.0.9 (Theorem 5.4.1). *Let $X/\mathbb{F}_q$ be a geometrically integral projective variety of dimension r. Let $g \in K \setminus \mathbb{F}_q$. If g is an ℓ-th power in K for some prime $\ell \mid q-1$ then $N_X(g,n) = 0$ for all $n > 1$. Otherwise we have the asymptotic formula*

$$N_X(g,n) = \rho_g(n) \left(\frac{\varphi(q^n-1) q^{n(r-1)}}{n} + O_{X,\varepsilon}\left(q^{n(r-1/2)+\varepsilon}\right) \right),$$

where

$$\rho_g(n) = \prod_{\substack{\ell \mid q^n - 1 \\ \ell \in \mathscr{P}_g}} \left(1 - \frac{c_\ell \left(q^{n-1} + q^{n-2} + \cdots + 1\right)}{\varphi(\ell)} \right).$$

The proof of this result was inspired by a previous quantitative version of Artin's primitive root conjecture established by Pappalardi–Shparlinski [81]. They proved the above theorem in the case when X is a curve, or equivalently $r = 1$, and when g is assumed to be geometric. In particular, if g is geometric then $\rho_g(n) = 1$ for all n since the product runs over an empty set, and so we recover their result for curves.

Since $\varphi(q^n - 1) \gg_\eta q^{n(1-\eta)}$ for all $\eta \in (0,1)$ the main term indeed dominates the error term in the above asymptotic. In order to deduce the infinitude of closed points for which g is primitive it is therefore sufficient to show $\rho_g(n) \neq 0$ for infinitely many positive integers n. This is dealt with in Chapter 5. Noting that every function field over $\mathbb{F}_q$ arises as the function field of a geometrically integral projective variety over $\mathbb{F}_q$ we therefore deduce in full Artin's conjecture in this setting.

Theorem 1.0.10 (Theorem 5.4.3). *Artin's primitive root conjecture holds for any function field K of finite transcendence degree over $\mathbb{F}_q$.*

The main idea of the proof is to express the characteristic function of primitive elements for a cyclic group in terms of character sums. In particular, if G is a finite cyclic group of order M then it is not hard to show that we have

$$f_G(g) := \frac{\varphi(M)}{M}\prod_{p|M}\left(1-\frac{\sum_{\substack{\chi\in\widehat{G}\\ \mathrm{ord}\chi=p}}\chi(g)}{\varphi(p)}\right) = \begin{cases}1, & \text{if } g \text{ generates } G\\ 0, & \text{otherwise.}\end{cases}$$

Using this one may therefore transform the count $N_X(g,n)$ into an exponential sum over the $\mathbb{F}_{q^n}$-rational points of X. Certain characters will then contribute to the main term and others will only contribute to the error term. To obtain cancellation we will require the Riemann hypothesis in this context, which is a famous theorem of Weil [116]. For curves the cancellation in the character sums has been deduced from Weil's result by Perelmuter [82]. In the case of higher dimensional varieties, to the authors' surprise, no result seems to be available in the literature. In particular we required a general result of the shape

$$\sum_{\rho\in X(\mathbb{F}_q)}\chi(g(\rho)) \ll q^{r-1/2},$$

where $\chi\colon \mathbb{F}_q^\times \to \mathbb{C}^\times$ is a multiplicative character of order $\ell \notin \mathscr{P}_g$. Having the corresponding result for curves available by Perelmuter, we therefore carefully proved such an estimate using a fibration argument. This is dealt with in Proposition 5.5.1. We hope that this can be useful for future applications in many different areas since it is of such a general shape.

The study of this problem originated by attempting to count the number of Artin primes in short intervals and arithmetic progressions. The goal was to employ random matrix theory heuristics in the number field case, and an equidistribution result due to Sawin [96] in the function field case in order to compute the variance of Artin primes for a fixed candidate element g over short intervals and arithmetic progressions. However, since there appeared to be this gap in the literature it was necessary to first establish the conjecture in a quantitative form in full first.

Chapter 2

Cubic forms over quadratic imaginary number fields and rational lines on cubic hypersurfaces

2.1 Introduction

The study of integer solutions to polynomial equations is one of the most fundamental mathematical problems. Quadratic forms are very well understood but the situation already becomes much more difficult when studying cubic equations. A cubic form $C(\boldsymbol{x}) \in \mathbb{Z}[x_1, \ldots, x_s]$ is a homogeneous polynomial of degree 3. We say that C represents zero non-trivially if there is a vector $\boldsymbol{x} \in \mathbb{Z}^s \backslash \{\boldsymbol{0}\}$ such that $C(\boldsymbol{x}) = 0$. Lewis [73] and Birch [8] both independently showed that every cubic form in sufficiently many variables represents zero non-trivially.

Using the Hardy-Littlewood circle method, Davenport [26] showed that it suffices to assume $s \geq 32$ in order to show that C represents zero non-trivially, which he then improved to $s \geq 16$ in a series of papers [27, 28]. The current state of the art is due to Heath-Brown [49] who showed that 14 variables suffice.

The best one can hope for is that every cubic forms in at least 10 variables represents zero non-trivially since there exist cubic forms in 9 variables, which do not have non-trivial p-adic solutions and hence also do not represent zero non-trivially over the integers.

More is known when the cubic form is assumed to be non-singular. In this case Heath-Brown [43] showed that if $s \geq 10$ then the cubic form represents zero non-trivially, and Hooley [52] established the Hasse Principle if $s \geq 9$. That is, he showed that if a non-singular cubic form over $\mathbb{Q}$ in at least nine variables has a non-trivial

p-adic solution for every p and a non-trivial real solution then it also represents zero non-trivially over the rational numbers.

One may also consider these problems for cubic forms over a number field $K/\mathbb{Q}$. In fact the above mentioned result by Lewis was proved for any number field $K/\mathbb{Q}$. Using the circle method the number of variables required was reduced to 54 by Ramanujam [89], which was subsequently improved to 17 variables by Ryavec [92] and 16 variables by Pleasants [87]. If one assumes the cubic form to be non-singular then recent work by Browning–Vishe [16] shows that ten variables suffice in order to infer the existence of a non-trivial zero, which improves previous work by Skinner [104].

The main result of this chapter is the following.

Theorem 2.1.1. *Let $K/\mathbb{Q}$ be an imaginary quadratic number field. If $C(\boldsymbol{x})$ is a homogeneous cubic form over K in at least 14 variables then $C(\boldsymbol{x})$ represents zero nontrivially.*

It seems likely that our result should remain true for general number fields, however there are two serious obstructions in generalizing Heath-Brown's ideas to the number field setting, as we discuss in the course of our proof. We are able to remove these difficulties only in the special case of imaginary quadratic number fields.

Our result has some interesting applications to problems that do not involve, prima facie, any number fields. The first of these concerns rational lines on cubic hypersurfaces.

Theorem 2.1.2. *Let C be a cubic form in $s \geq 33$ variables with rational coefficients. Then the projective cubic hypersurface defined by $C(\boldsymbol{x}) = 0$ contains a rational line.*

This improves on work of Wooley [117] who had the same result under the assumption $s \geq 37$. We note that another two variables can be saved using ideas from forthcoming work by Brandes and Dietmann [12], thus leading to a result for $s \geq 31$ variables.

More specifically, while our argument (building on Wooley's) for the proof of Theorem 2.1.2 only requires Theorem 2.1.1 for one imaginary quadratic number field (e.g. $\mathbb{Q}(i)$), the full generality of Theorem 2.1.1 is required in the argument of Brandes and Dietmann.

It is also worth mentioning that in a different paper of the same authors [11], the result for $s \geq 31$ variables is already established under the assumption that the underlying hypersurface is nonsingular.

Based on an observation of Brüdern–Dietmann–Liu–Wooley [18], the existence of rational lines can be used in conjunction with the Green–Tao Theorem to produce almost prime solutions to cubic forms as follows:

Theorem 2.1.3. *Let C be a cubic form in $s \geq 33$ variables with rational coefficients. Then there are almost prime solutions to $C(\boldsymbol{x}) = 0$ in the following sense: There are coprime integers $c_1, \ldots, c_s$ such that the equation*

$$C(c_1p_1, c_2p_2, \ldots, c_sp_s) = 0$$

has infinitely many solutions in primes $p_1, \ldots, p_s$, not all equal.

Notation

We use $e(\alpha) = e^{2\pi i\alpha}$ and the notation $O(\ldots)$ and $\ll$ of Landau and Vinogradov, respectively. All implied constants are allowed to depend on the number field K, a choice of integral basis Ω for K, the cubic form C and a small parameter $\varepsilon > 0$ whenever it appears.

As is convenient in analytic number theory, this parameter ε may change its value finitely many times. In particular, we may write something like $M^{2\varepsilon} \ll M^{\varepsilon}$.

We often use the notation $q \sim R$ to denote the dyadic condition $R < q \leq 2R$.

2.2 Deduction of Theorems 2.1.2 and 2.1.3

In this section, we give the proofs of Theorems 2.1.2 and 2.1.3 assuming Theorem 2.1.1.

We begin with the observation that the existence of a rational line on the cubic hypersurface defined by C is equivalent to the existence of linearly independent vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ such that $C(\boldsymbol{v}+t\boldsymbol{w}) = 0$ identically in t. Expanding this formally as a cubic polynomial in t, we obtain

$$C(\boldsymbol{v}) + tQ_{\boldsymbol{w}}(\boldsymbol{v}) + t^2L_{\boldsymbol{w}}(\boldsymbol{v}) + t^3C(\boldsymbol{w}) = 0$$

for certain quadratic resp. linear forms $Q_{\boldsymbol{w}}$ and $L_{\boldsymbol{w}}$ depending on $\boldsymbol{w}$. We therefore need to find linearly independent $\boldsymbol{v}$ and $\boldsymbol{w}$ such that

$$C(\boldsymbol{v}) = Q_{\boldsymbol{w}}(\boldsymbol{v}) = L_{\boldsymbol{w}}(\boldsymbol{v}) = C(\boldsymbol{w}) = 0.$$

If we start by choosing a solution $\boldsymbol{w} \neq 0$ of $C(\boldsymbol{w}) = 0$, the linear equation $L_{\boldsymbol{w}}(\boldsymbol{v}) = 0$ and requiring $\boldsymbol{v}$ to be orthogonal to $\boldsymbol{w}$ reduce the degrees of freedom for $\boldsymbol{v}$ by two.

We are thus looking for a solution to the system $C(\boldsymbol{v}) = Q_{\boldsymbol{w}}(\boldsymbol{v}) = 0$ of one cubic and one quadratic equation in $s-2$ variables. If we knew that the signature of the quadratic form $Q_{\boldsymbol{w}}$ was sufficiently indefinite, we could infer the existence of a sufficiently large linear space on which $Q_{\boldsymbol{w}}$ vanishes, leaving us with a single cubic form in many variables, that can be dealt with by the work of Heath-Brown [49].

The crux however is that it is in general hard to control the signature of $Q_{\boldsymbol{w}}$. Instead we avoid the indefiniteness issue by passing to an imaginary quadratic number field of $\mathbb{Q}$, thus requiring our Theorem 2.1.1.

We now present the complete argument in order: We begin by choosing $\boldsymbol{w} \in \mathbb{Q}^s \backslash \{\boldsymbol{0}\}$ such that $C(\boldsymbol{w}) = 0$, which exists by the work of Heath-Brown.

Letting $K/\mathbb{Q}$ be any imaginary quadratic number field, we next show the existence of a vector $\boldsymbol{v} \in K^s$, linearly independent to $\boldsymbol{w}$ and satisfying

$$C(\boldsymbol{v}) = Q_{\boldsymbol{w}}(\boldsymbol{v}) = L_{\boldsymbol{w}}(\boldsymbol{v}) = 0.$$

To this end, we use that a hypersurface $Q(\boldsymbol{x}) = 0$ defined by a quadratic form Q in s variables contains a $\left\lfloor \frac{s-3}{2} \right\rfloor$-dimensional K-linear subspace, a fact that is easily proved by induction.

The linear space of vectors $\boldsymbol{v}$ orthogonal to $\boldsymbol{w}$ and satisfying $L_{\boldsymbol{w}}(\boldsymbol{v}) = 0$ is at least $(s-2)$-dimensional. Thus, $Q_{\boldsymbol{w}}$ vanishes on a linear subspace of dimension at least $\left\lfloor \frac{(s-2)-3}{2} \right\rfloor = \left\lfloor \frac{s-5}{2} \right\rfloor$. Note that by our assumption on s we have $\left\lfloor \frac{s-5}{2} \right\rfloor \geq 14$. We are then left to solve the equation $C(\boldsymbol{v}) = 0$ on a 14-dimensional linear space which can be done by Theorem 2.1.1.

We have thus proved that $C(\boldsymbol{v} + t\boldsymbol{w}) = 0$ identically in t for some linearly independent vectors $\boldsymbol{v} \in K^s$ and $\boldsymbol{w} \in \mathbb{Q}^s$.

By an observation of Lewis, this is enough to deduce the existence of a rational line, as we explain now, following an argument of Dietmann–Wooley [34].

Consider the K-rational spaces V spanned by $\boldsymbol{v}$ and $\boldsymbol{w}$ as well as V^* spanned by $\boldsymbol{v}^*$ and $\boldsymbol{w}$, where * denotes conjugation in K. If $\boldsymbol{v} \in \mathbb{Q}^s$ we are already done. Else, consider the three-dimensional space W spanned by $\boldsymbol{v}$, $\boldsymbol{v}^*$ and $\boldsymbol{w}$. If C vanishes on W, we are also done as W clearly contains a two-dimensional $\mathbb{Q}$-rational subspace. Else, by intersection theory the hypersurface defined by C must intersect W in a third two-dimensional K-rational subspace L. More precisely, by Theorem I.7.7 in Hartshorne [42] we have

$$i(W, C; V) + i(W, C; V^*) + \sum_j i(W, C; Z_j) \cdot \deg Z_j = (\deg W)(\deg C) = 3,$$

where $i(W, C; V)$ denotes the intersection multiplicity and Z_i are the other irreducible components of $C \cap W$. Since W is invariant under conjugation, we must have $i(W, C; V) = i(W, C; V^*)$ and thus both numbers are equal to 1, implying that there is a unique third component $L = Z_1$ which is then necessarily linear. Finally, since W and C are conjugation invariant, the three spaces V, V^* and L are permuted under conjugation and thus L itself is conjugation invariant, i.e. describes the desired rational line. □

We remark that the use of intersection theory in the previous argument can be replaced by an explicit algebraic computation, as shown in Wooley [117].

To deduce Theorem 2.1.3, we follow the strategy in [18]. In particular, we show that the existence of a rational line implies the existence of almost prime solutions, regardless of the number of variables. We thus assume that for some linearly independent vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{Z}^s$, we have $C(\boldsymbol{a}t + \boldsymbol{b}u) = 0$ identically in t and u. If $a_i = b_i = 0$ for some i, then we can set $c_i = 1$ and continue to work with the other variables. By taking a suitable linear combination, we can then assume that indeed all a_i and b_i are different from 0. Rescaling u by a factor of $a_1 a_2 \ldots a_s$ and then rescaling the variables by a factor of a_i (thereby changing c_i by a factor of a_i), we may even assume that all the a_i are equal to 1, i.e.

$$C(t + b_1 u, t + b_2 u, \ldots, t + b_n u) = 0$$

identically in t and u. By the Green–Tao Theorem [40], the primes contain infinitely many arithmetic progressions of length $2M + 1$ where $M = 2 \max_i |b_i| + 1$, i.e. there are infinitely many pairs (ℓ, d) such that $\ell + kd$ is prime for all $|k| \leq M$. Choosing $t = \ell$ and $u = k$ then yields the desired result with $c_i = 1$. □

2.3 Algebraic preliminaries

While our main result is proved only for imaginary quadratic number fields we will introduce the matter in a general fashion and not restrict ourselves to these fields for now. We will aim to highlight whenever phenomena occur that set apart the situation for imaginary quadratic number fields from a general setting. In particular, even when $K/\mathbb{Q}$ is an imaginary quadratic number field we still sometimes prefer to write $n = [K : \mathbb{Q}]$.

Let K be a number field of degree n over $\mathbb{Q}$ and denote by $\mathcal{O}$ its ring of integers.

Define the $\mathbb{R}$-vector space $K_{\mathbb{R}} := K \otimes_{\mathbb{Q}} \mathbb{R}$ and note that we have natural embeddings $\mathcal{O} \subset K \subset K_{\mathbb{R}}$. The space $K_{\mathbb{R}}$ is sometimes referred to as the *Minkowski space* of K. Note that there exist integers n_1 and n_2 with $n_1 + 2n_2 = n$ such that K admits n_1 real embeddings $\sigma_1, \dots, \sigma_{n_1}$ and $2n_2$ complex embeddings $\sigma_{n_1+1}, \overline{\sigma}_{n_1+1}, \dots, \sigma_{n_1+n_2}, \overline{\sigma}_{n_1+n_2}$ so that $K_{\mathbb{R}} \cong \mathbb{R}^{n_1} \times \mathbb{C}^{n_2}$.

Denote by π_i the projection from $K_{\mathbb{R}} \cong \mathbb{R}^{n_1} \times \mathbb{C}^{n_2}$ to the i-th coordinate, which may take real or complex values. We define the trace map $\mathrm{tr}\colon K_{\mathbb{R}} \to \mathbb{R}$ and norm map $\mathrm{Norm}\colon K_{\mathbb{R}} \to \mathbb{R}$ as

$$\mathrm{tr}(\alpha) = \sum_{i=1}^{n_1} \pi_i(\alpha) + \sum_{i=n_1+1}^{n_2} \mathrm{Re}(\pi_i(\alpha)),$$

and

$$\mathrm{Norm}(\alpha) = \prod_{i=1}^{n_1} |\pi_i(\alpha)| \prod_{i=n_1+1}^{n_2} |\pi_i(\alpha)|^2,$$

respectively. If $\alpha \in K$ then these are just the usual norm and trace function from algebraic number theory.

Pick a basis $\Omega = \{\omega_1, \dots, \omega_n\}$ of $\mathcal{O}$. Any element $\alpha \in K_{\mathbb{R}}$ may be expressed in the form $\alpha = \sum_{j=1}^{n} \alpha_j \omega_j$ for some $\alpha_j \in \mathbb{R}$. For such α we define a height

$$|\alpha| := \max_j |\alpha_j|.$$

Note that this depends on the choice of basis Ω for $\mathcal{O}$. Given a vector $\boldsymbol{\alpha} = (\alpha^{(1)}, \dots, \alpha^{(s)}) \in K_{\mathbb{R}}^s$ we further denote

$$|\boldsymbol{\alpha}| := \max_k |\alpha^{(k)}|.$$

We may alternatively define another height on $K_{\mathbb{R}}$ given by

$$|\alpha|_K := \max_p |\pi_p(\alpha)|.$$

As noted by Pleasants [87, Section 2] we have

$$|\alpha| \asymp |\alpha|_K,$$

for all $\alpha \in K_{\mathbb{R}}$. If $\alpha, \beta \in K_{\mathbb{R}}$ then it is easy to see that this height satisfies

$$|\alpha\beta|_K \leq |\alpha|_K |\beta|_K,$$
$$|\alpha + \beta|_K \leq |\alpha|_K + |\beta|_K$$
$$|\alpha^{-1}|_K \leq \frac{|\alpha|_K^{n-1}}{\mathrm{Norm}(\alpha)}.$$

The same inequalities therefore hold for $|\cdot|$ if we replace the symbols $\leq$ by $\ll_K$. It would be desirable to have the last inequality in the form $|\alpha^{-1}| \asymp |\alpha|^{-1}$ which would result if $\mathrm{Norm}(\alpha) \asymp |\alpha|^n$. However, if α is a unit in $\mathcal{O}$ then $\mathrm{Norm}(\alpha) = 1$ while the height $|\alpha|$ may be unbounded, at least whenever K is not an imaginary quadratic number field. This is one of the points where our argument crucially depends on the latter assumption.

If $K = \mathbb{Q}(\sqrt{-d})$ is an imaginary quadratic number field then, depending on the value of the residue class of d mod 4, we can choose $\{1, \sqrt{-d}\}$ or $\{1, (1+\sqrt{-d})/2\}$ as an integral basis for $\mathcal{O}$. We thus find that

$$\mathrm{Norm}(\alpha) \asymp |\alpha|^2.$$

In particular we find

$$|\alpha^{-1}| \asymp |\alpha|^{-1}.$$

Given an ideal $J \subset \mathcal{O}$ we recall that $\mathcal{O}/J$ is finite and we define as usual the norm of the ideal to be

$$N(J) := \#\left(\mathcal{O}/J\right).$$

For a fractional ideal of K this norm is, as usual, extended multiplicatively using the unique factorization into prime ideals inside K. Given $\gamma \in K$ we further define the *denominator ideal* of γ as

$$\mathfrak{a}_\gamma := \{x \in \mathcal{O} \colon x\gamma \in \mathcal{O}\}.$$

As the name suggests, and this is not very difficult to verify, $\mathfrak{a}_\gamma$ is an ideal inside $\mathcal{O}$, contained in the fractional ideal $(\gamma)^{-1}$. We will need the following fact several times.

Lemma 2.3.1. *Let $J \subset \mathcal{O}$ be an ideal. Then there are at most $N(J)$ different elements $\gamma \in K/\mathcal{O}$ such that $\mathfrak{a}_\gamma = J$.*

Proof. To see this, note first that for any two fractional ideals $\mathfrak{b}, \mathfrak{c} \subset K$ with $\mathfrak{b} \supset \mathfrak{c}$ there exists some $d \in \mathcal{O}$ such that $d\mathfrak{b}, d\mathfrak{c} \subset \mathcal{O}$. Thus

$$[\mathfrak{b} \colon \mathfrak{c}] = [d\mathfrak{b} \colon d\mathfrak{c}] = \frac{[\mathcal{O} \colon d\mathfrak{c}]}{[\mathcal{O} \colon d\mathfrak{b}]} = N(d\mathfrak{c})/N(d\mathfrak{b}) = N(\mathfrak{c})/N(\mathfrak{b}).$$

Now note that if $\mathfrak{a}_\gamma = J$ we must have $\gamma \in J^{-1}\mathcal{O}$, where

$$J^{-1} = \{x \in K \colon xJ \subset \mathcal{O}\}.$$

But now $[J^{-1}\mathcal{O} \colon \mathcal{O}] = N(J)$ and so the result follows. □

We shall further require a version of Dirichlet's approximation theorem.

Lemma 2.3.2. *Let $K/\mathbb{Q}$ be a number field of degree n. Let $\alpha \in K_{\mathbb{R}}$ and let $Q \geq 1$. Then there exist some $a, q \in \mathcal{O}$ with $1 \leq |q| \leq Q$ such that*

$$|q\alpha - a| \leq \frac{1}{Q}.$$

Proof. Consider the set $\mathcal{Q}$ of algebraic integers given by

$$\mathcal{Q} = \left\{ \sum_j q_j \omega_j \in \mathcal{O} \colon 0 \leq q_j \leq Q \right\}.$$

For any $q \in \mathcal{Q}$ we may express $q\alpha$ as

$$q\alpha = a_q + x_q,$$

where $a_q \in \mathcal{O}$ and $x_q = \sum_j x_{q,j}\omega_j$ such that $0 \leq x_{q,j} < 1$ for $j = 1, \ldots, n$. By considering $\lceil Q \rceil^n$ boxes centered around x_q inside $K_{\mathbb{R}}/\mathcal{O} = \left\{ \sum_j x_j \omega_j \colon 0 \leq x_j < 1 \right\}$ whose edges have side lengths $1/Q$, we find that two such boxes must necessarily intersect. Hence there must be $q_1, q_2 \in \mathcal{Q}$ with $q_1 \neq q_2$ such that x_{q_1} and x_{q_2} lie in the same box according to the partition above. Therefore we find

$$|(q_1 - q_2)\alpha - (a_{q_1} - a_{q_2})| = |x_{q_1} - x_{q_2}| \leq 1/Q.$$

Taking $q = q_1 - q_2$ and $a = a_{q_1} - a_{q_2}$ delivers the result. □

For the application to the mean-square averaging method introduced by Heath-Brown, we need a fractional form of Dirichlet's theorem. We are only able to obtain a satisfactory version for imaginary quadratic number fields, this being the first of the obstructions regarding possible generalizations mentioned in the introduction. Note that this is special to Heath-Brown's method and hence was not an issue in the work of Ramanujam, Ryavec and Pleasants.

Lemma 2.3.3. *Let $K/\mathbb{Q}$ be an imaginary quadratic number field (in particular $n = 2$). Let $\alpha \in K_{\mathbb{R}}$ and let $Q \geq 1$. Then there exists some $\gamma \in K$ with $N(\mathfrak{a}_\gamma) \leq Q^n$ such that*

$$|\alpha - \gamma| \ll \frac{1}{N(\mathfrak{a}_\gamma)^{\frac{1}{n}} Q}.$$

Proof. From Lemma 2.3.3 we find that there exist $a, q \in \mathcal{O}$ with $q \neq 0$ and $|q| \leq Q$ such that

$$|q\alpha - a| \leq 1/Q.$$

Set $\gamma = a/q \in K$ and note that $(q) \subseteq \mathfrak{a}_\gamma$. In particular from this it follows that

$$N(\mathfrak{a}_\gamma) \leq N((q)) = \mathrm{Norm}(q) \asymp |q|^n,$$

where the last estimate is true since K is an imaginary quadratic number field. Thus

$$|q|^{-1} \ll N(\mathfrak{a}_\gamma)^{-1/n},$$

and so we obtain

$$|\alpha - \gamma| \ll |q|^{-1}|q\alpha - a| \ll \frac{1}{N(\mathfrak{a}_\gamma)^{\frac{1}{n}} Q},$$

as desired. □

We shall sometimes require the following easy lemma.

Lemma 2.3.4. *Let $J \subset \mathcal{O}$ be an ideal. Then there exist constants c_1, c_2 only depending on K such that for any non-zero $g \in J$ we have*

$$c_1 N(J)^{1/n} \leq |g|,$$

and we may always find a non-zero element $g \in J$ such that

$$|g| \leq c_2 N(J)^{1/n}.$$

Proof. First note that if $g \in J$ then $(g) \subset J$ and therefore

$$N(J) \leq N((g)) = \mathrm{Norm}(g) \ll |g|^n.$$

For the second inequality note that there are at least $N(J) + 1$ algebraic integers whose height does not exceed $N(J)^{1/n}$. By definition $N(J) = \#(\mathcal{O}/J)$ and hence at least two of these integers must lie in the same residue class modulo J. Their difference is therefore an algebraic integer $g \in J$ with $|g| \leq 2N(J)^{1/n}$. □

Finally we will also need the following.

Lemma 2.3.5. *Let $K/\mathbb{Q}$ be a number field and let Δ be the discriminant of this extension. Let $\alpha \in K_\mathbb{R}$ and assume that $\{\omega_i\}_i$ is an integral basis for $\mathcal{O}$. If*

$$\Delta^{-1}\mathrm{tr}(\alpha\omega_i) \in \mathbb{Z}$$

holds for all $i = 1, \ldots, n$ then $\alpha \in \mathcal{O}$.

Proof. Write $\alpha = \sum_{j=1}^{n} \alpha_j \omega_j$, where $\alpha_j \in \mathbb{R}$. Due to the additivity of the trace we have

$$\mathrm{tr}(\alpha\omega_i) = \sum_{j=1}^{n} \alpha_j \mathrm{tr}(\omega_i\omega_j).$$

Denote by $\mathbf{T}$ the trace form, that is, the $n \times n$ matrix with entries $\mathrm{tr}(\omega_i\omega_j)$. Then if we identify $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n$, the assumption of the lemma is equivalent to

$$\Delta^{-1}\mathbf{T}(\boldsymbol{\alpha}) \in \mathbb{Z}^n.$$

By definition $\det \mathbf{T} = \Delta$. Hence $\mathbf{T}' := \Delta \mathbf{T}^{-1}$ has integer entries. Combining this with our previous observation yields

$$\boldsymbol{\alpha} = \mathbf{T}^{-1}\mathbf{T}(\boldsymbol{\alpha}) = \mathbf{T}'(\Delta^{-1}\mathbf{T}(\boldsymbol{\alpha})) \in \mathbb{Z}^n.$$

Hence $\alpha \in \mathcal{O}$ as required. □

2.4 The dichotomy

Let $C \in \mathcal{O}[x_1, \ldots, x_s]$ be a homogeneous cubic form. Our goal is to show that there always exists a non-trivial solution to $C = 0$ over K provided $s \geq 14$ and K is an imaginary quadratic number field. We follow the strategy of Davenport that was later refined by Heath-Brown [49]: Either C represents zero non-trivially for 'geometric reasons', or we can establish an asymptotic formula for the number of solutions of bounded height, using the circle method.

2.4.1 Davenport's geometric condition

We may express $C(\boldsymbol{x})$ as

$$C(\boldsymbol{x}) = \sum_{i,j,k} c_{ijk} x_i x_j x_k,$$

where the coefficients c_{ijk} are fully symmetric in the indices and lie in $\mathcal{O}$, after replacing $C(\boldsymbol{x})$ by $6C(\boldsymbol{x})$ if required. For $i = 1, \ldots, s$ further define the bilinear forms $B_i(\boldsymbol{x}, \boldsymbol{y})$ by

$$B_i(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j,k} c_{ijk} x_j y_k.$$

Finally, we also consider an $s \times s$ matrix $M(\boldsymbol{x})$, the *Hessian* of $C(\boldsymbol{x})$, whose entries are defined by

$$M(\boldsymbol{x})_{jk} = \sum_{i} c_{ijk} x_i,$$

so that

$$(M(\boldsymbol{x})\boldsymbol{y})_i = B_i(\boldsymbol{x}, \boldsymbol{y}).$$

We note that the entries of $M(\boldsymbol{x})$ are linear forms in the variables $\boldsymbol{x}$. Denote the rank of the matrix by

$$r(\boldsymbol{x}) = \operatorname{rank}(M(\boldsymbol{x})).$$

As in Davenport's and Heath-Brown's work we obtain a dichotomy.

Lemma 2.4.1. *One of the following two alternatives holds.*

(i) Davenport's Geometric Condition: For every integer $0 \le r \le s$ *we have*

$$\#\{\boldsymbol{x} \in \mathcal{O}^s \colon |\boldsymbol{x}| < H,\ r(\boldsymbol{x}) = r\} \ll H^{nr}. \tag{2.4.1}$$

(ii) The cubic form $C(\boldsymbol{x})$ *has a non-trivial zero in* $\mathcal{O}$*.*

Proof. Consider the least integer $h = h(C)$ such that the cubic form may be written as

$$C(\boldsymbol{x}) = \sum_{i=1}^{h} L_i(\boldsymbol{x}) Q_i(\boldsymbol{x}),$$

where L_i are linear and Q_i are quadratic forms defined over K. This is the h-invariant of C. It is easy to see that $1 \le h \le s$ holds, and that $C(\boldsymbol{x}) = 0$ has a non-trivial solution over K if and only if $h < s$.

We will show that if $h = s$ then alternative (1) holds. In fact, Pleasants [87, Lemma 3.5] showed that the number of points $\boldsymbol{x} \in \mathcal{O}^s$ such that $|\boldsymbol{x}| < H$ holds, for which the equations $B_i(\boldsymbol{x}, \boldsymbol{y}) = 0$, $i = 1, \ldots, s$ have exactly $s - r$ linearly independent solutions $\boldsymbol{y}$ is bounded by $O(H^{n(s-h+r)})$. Hence taking $h = s$ delivers the desired bound (2.4.1). □

We will henceforth assume that Davenport's Geometric Condition (2.4.1) is satisfied and apply the circle method. In particular as in [49] this condition implies that we have

$$\#\{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{O}^s \colon |\boldsymbol{x}|, |\boldsymbol{y}| < H, B_i(\boldsymbol{x}, \boldsymbol{y}) = 0, \forall i\} \ll H^{ns}, \tag{2.4.2}$$

for any $H \ge 1$.

2.4.2 The circle method

Let $\mathcal{B} \subset K_{\mathbb{R}}^s \cong \mathbb{R}^{ns}$ be a box of the form

$$\mathcal{B} = \left\{ \left(\sum_j \alpha_{ij}\omega_j \right)_i \in K_{\mathbb{R}}^s \colon b_{ij}^- \leq \alpha_{ij} \leq b_{ij}^+ \right\},$$

where $b_{ij}^- < b_{ij}^+$ are some real numbers and we will throughout assume $b_{ij}^+ - b_{ij}^- \leq 1$. For $P \geq 1$ consider the counting function

$$N(P;\mathcal{B}) = N(P) = \{ \boldsymbol{x} \in P\mathcal{B} \cap \mathcal{O}^s \colon C(\boldsymbol{x}) = 0 \}.$$

For $\alpha \in K_{\mathbb{R}}$ and $P \geq 1$ we define the exponential sum

$$S(\alpha) = S(\alpha; P) = \sum_{\boldsymbol{x} \in P\mathcal{B} \cap \mathcal{O}^s} e\left(\operatorname{tr}(\alpha C(\boldsymbol{x}))\right).$$

Denote by $I \subset K_{\mathbb{R}}$ the set given by

$$I = \left\{ \sum_{j=1}^n \alpha_j \omega_j \colon 0 \leq \alpha_j \leq 1 \right\},$$

which may also be regarded as $K_{\mathbb{R}}/\mathcal{O}$. Due to orthogonality of characters we obtain

$$N(P) = \int_{\alpha \in I} S(\alpha) d\alpha.$$

We are now able to state the main technical theorem of this chapter.

Theorem 2.4.2. *Let $K/\mathbb{Q}$ be an imaginary quadratic number field and let $C(\boldsymbol{x})$ be a cubic form in $s \geq 14$ variables over K. Suppose that $C(\boldsymbol{x})$ is irreducible over K and that Davenport's Geometric Condition* (2.4.1) *is satisfied. Then we have the asymptotic formula*

$$N(P) = \sigma P^{n(s-3)} + o\left(P^{n(s-3)}\right), \quad \text{as} \quad P \to \infty,$$

where $\sigma > 0$ is the product of the usual singular integral and singular series.

Therefore Theorem 2.1.1 follows directly from Lemma 2.4.1 and Theorem 2.4.2 where we also note that a reducible cubic form always contains a linear factor over K and therefore has a non-trivial solution for obvious reasons.

2.4.3 The major arcs

For this section we do not need to assume that K is an imaginary quadratic number field of $\mathbb{Q}$. As in Pleasants, we choose as center of our box $\mathcal{B} = \mathcal{B}(\boldsymbol{z})$ a solution $\boldsymbol{z} \in K_{\mathbb{R}}$ of $C(\boldsymbol{z}) = 0$ satisfying $\frac{\partial C}{\partial x_1}(\boldsymbol{z}) \neq 0$ and $z_1, \ldots, z_n \neq 0$. Such a vector $\boldsymbol{z}$ always exists by [87, Lemma 7.2] provided C is irreducible.

Let $\gamma \in K/\mathcal{O}$ and define

$$\mathfrak{M}_\gamma := \left\{\alpha \in I \colon |\alpha - \gamma| < P^{-3+\nu}\right\},$$

where we regard $I = K_{\mathbb{R}}/\mathcal{O}$. We define the *major arcs* as

$$\mathfrak{M} = \bigcup_{\substack{\gamma \in K/\mathcal{O} \\ N(\mathfrak{a}_\gamma) \leq P^\nu}} \mathfrak{M}_\gamma,$$

and the *minor arcs* as

$$\mathfrak{m} = I \setminus \mathfrak{M}.$$

Further, define the sum S_γ via

$$S_\gamma = \sum_{\boldsymbol{x} \bmod N(\mathfrak{a}_\gamma)} e(\operatorname{tr}(\gamma C(\boldsymbol{x}))).$$

Given a parameter $R \geq 1$ we define the *truncated singular series* to be

$$\mathfrak{S}(R) := \sum_{\substack{\gamma \in K/\mathcal{O} \\ N(\mathfrak{a}_\gamma) \leq R}} N(\mathfrak{a}_\gamma)^{-ns} S_\gamma,$$

and the *truncated singular integral* to be

$$\mathfrak{I}(R) := \int_{|\zeta| < R^\nu} \int_{\mathcal{B}} e(\operatorname{tr}(\zeta R^{-3} C(R\boldsymbol{\xi}))) d\boldsymbol{\xi} d\zeta.$$

Pleasants [87, Lemma 7.1] shows that if $\nu < \frac{1}{n+4}$ is satisfied then we have

$$\int_{\mathfrak{M}} S(\alpha) d\alpha = \mathfrak{S}(P^\nu)\mathfrak{I}(P)P^{n(s-3)} + o(P^{n(s-3)}).$$

Moreover, if $\mathcal{B} = \mathcal{B}(\boldsymbol{z})$ is the box as in the beginning of the section, provided that the sidelengths of the boxes are sufficiently small, and if $C(\boldsymbol{x})$ is irreducible over K then Pleasants [87, Lemma 7.2] further shows that $\mathfrak{I}(R)$ converges absolutely to a positive number $\mathfrak{I}$.

We remark that Lemma 7.2 in [87] was originally stated under the weaker assumption that $C(\boldsymbol{x})$ is not a rational multiple of a cube of a linear form. His proof relies on a result by Davenport [26, Lemma 6.2], which assumes the existence of a non-singular, real solution $\boldsymbol{\xi} \in \mathbb{R}^n$ of a rational cubic form G such that

$$\frac{\partial G}{\partial x_i}(\boldsymbol{\xi}) \neq 0, \quad \xi_i \neq 0,$$

holds for some i. In particular Pleasants writes that *"this hypothesis is not used in the proof of the lemma, however, and in any case the argument that follows could easily be adapted to provide it"*. While one can always find $\boldsymbol{\xi} \in \mathbb{R}^s$ with $\frac{\partial G}{\partial x_i}(\boldsymbol{\xi}) \neq 0$ unless G is a rational multiple of a cube of a linear form, one can not necessarily ensure that $\xi_i \neq 0$ for the same index i. Consider for example $G(x_1, \ldots, x_n) = x_1(x_2^2 + \cdots + x_n^2)$. It is possible that Davenport's result [26, Lemma 6.2] holds nevertheless in this generality but at least the standard method of establishing bounded variation of the auxiliary function involved in the proof by showing the existence of right and left derivatives, see for example [29, Lemma 16.1], fails in general.

The singular series $\mathfrak{S}(R)$ may or may not converge absolutely as $R \to \infty$. If it does converge, then provided non-singular $\mathfrak{p}$-adic solutions of $C(\boldsymbol{x}) = 0$ exist for all primes $\mathfrak{p}$, by standard arguments it follows that $\mathfrak{S} > 0$. See for example the proof of Lemma 7.4 in [87], where this argumentation is carried out in our setting. Finally, Lewis [72] showed that these non-singular $\mathfrak{p}$-adic solutions always exist whenever $s \geq 10$. Therefore we obtain the following.

Theorem 2.4.3. *Let $C \in \mathcal{O}[x_1, \ldots, x_s]$ be an irreducible cubic form. Assume that $s \geq 10$. If the singular series $\mathfrak{S}(R)$ converges absolutely as $R \to \infty$ then*

$$\int_{\mathfrak{M}} S(\alpha) d\alpha = \sigma P^{n(s-3)} + o(P^{n(s-3)}),$$

for some $\sigma > 0$ as $P \to \infty$.

In particular, in Section 2.7 we will establish the following.

Theorem 2.4.4. *Assume that $s \geq 13$ and that Davenport's Geometric Condition* (2.4.1) *is satisfied then the singular series converges absolutely. Therefore if $C(\boldsymbol{x})$ is irreducible we have*

$$\int_{\mathfrak{M}} S(\alpha) d\alpha = \sigma P^{n(s-3)} + o\left(P^{n(s-3)}\right),$$

for some $\sigma > 0$ as $P \to \infty$.

We remark that we show this result for any number field K.

2.5 Auxiliary Diophantine inequalities

To bound the Weyl sum $S(\alpha)$ of a general cubic form, classical Weyl differencing leaves us with the task of examining the number of solutions to certain auxiliary Diophantine inequalities. Davenport's crucial idea was to bootstrap these inequalities using his *Shrinking Lemma*, combined with the observation that sufficiently strong Diophantine inequalities already imply divisibility or even equality.

In this section, we prepare these arguments by providing a version of this observation adapted to our setting. We are only able to show a satisfactory version of this lemma if $K/\mathbb{Q}$ is an imaginary quadratic number field, this being the second of the obstructions mentioned in the introduction.

Lemma 2.5.1. *Assume that $K/\mathbb{Q}$ is a number field and denote by Δ the discriminant of this extension. There exists a real positive constant $A > 0$ depending only on K and the choice of integral basis Ω for K such that the following statement holds.*

Let $M \geq 0$ be a real number and let $\alpha \in K_{\mathbb{R}}$. Suppose that $\alpha = \gamma + \theta$ with $\gamma \in K$ and $M|\theta|N(\mathfrak{a}_\gamma)^{1/n} \leq A$. If $m \in \mathcal{O}$ is such that $|m| \leq M$ and $\|\Delta^{-1}\mathrm{tr}(\alpha m \omega_j)\| < P_0^{-1}$ holds for all $j = 1, \ldots, n$ where $AP_0 \geq N(\mathfrak{a}_\gamma)^{1/n}$ then $m \in \mathfrak{a}_\gamma$. In particular if either of the conditions

(i) $M \leq AN(\mathfrak{a}_\gamma)^{1/n}$, or

(ii) K is an imaginary quadratic number field and $A|\theta| \geq N(\mathfrak{a}_\gamma)^{-1/n}P_0^{-1}$

is satisfied, then we must have $m = 0$.

Proof. Note first that

$$\left\|\Delta^{-1}\mathrm{tr}(\gamma m \omega_j)\right\| \leq \left\|\Delta^{-1}\mathrm{tr}(\alpha m \omega_j)\right\| + \left\|\Delta^{-1}\mathrm{tr}(\theta m \omega_j)\right\|.$$

Now due to our assumption we have $\|\Delta^{-1}\mathrm{tr}(\alpha m \omega_j)\| < P_0^{-1}$. Further it is easy to see that

$$\Delta^{-1}|\mathrm{tr}(\theta m \omega_j)| \ll |\theta| M.$$

Therefore choosing A sufficiently small we find

$$\left\|\Delta^{-1}\mathrm{tr}(\gamma m \omega_j)\right\| < \frac{A^{1/2}}{N(\mathfrak{a}_\gamma)^{1/n}}, \tag{2.5.1}$$

for all $j = 1, \ldots, n$. As before write $\mathbf{T} = (\mathrm{tr}(\omega_i\omega_j))_{i,j}$ for the trace form. Write $\boldsymbol{x} \in \mathbb{R}^n$ for the real vector obtained from γm under the isomorphism $K_{\mathbb{R}} \cong \mathbb{R}^n$ with respect to the integral basis Ω. Then (2.5.1) is equivalent to saying that there exist $\boldsymbol{a} \in \mathbb{Z}^n$ and $\boldsymbol{r} \in \mathbb{R}^n$ with $|\boldsymbol{r}| < \frac{A^{1/2}}{N(\mathfrak{a}_\gamma)^{1/n}}$ such that

$$\mathbf{T}(\Delta^{-1}\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{r}.$$

Recall that $\Delta\mathbf{T}^{-1}$ is an integral matrix whose entries are bounded in terms of K. Therefore

$$\boldsymbol{x} = \Delta\mathbf{T}^{-1}(\boldsymbol{a}) + \Delta\mathbf{T}^{-1}(\boldsymbol{r}).$$

Now $\Delta\mathbf{T}^{-1}(\boldsymbol{a}) \in \mathbb{Z}^n$ and

$$|\Delta\mathbf{T}^{-1}(\boldsymbol{r})| < \frac{A^{1/3}}{N(\mathfrak{a}_\gamma)^{1/n}},$$

after decreasing A if necessary. We thus find that

$$\gamma m = a + \rho,$$

where $a \in \mathcal{O}$ and $|\rho| < \frac{A^{1/3}}{N(\mathfrak{a}_\gamma)^{1/n}}$. By Lemma 2.3.4 there exists $g \in \mathfrak{a}_\gamma$ with $|g| \asymp N(\mathfrak{a}_\gamma)^{1/n}$. From the above equation we see that $g\rho \in \mathcal{O}$, and so, unless $\rho = 0$ we have

$$1 \leq |g\rho| < A^{1/4},$$

after decreasing A if necessary. Choosing A suitably small therefore leads to a contradiction whence we must have $\rho = 0$, and so $m \in \mathfrak{a}_\gamma$. This finishes the first part of the proof.

If we now assume that $M \leq AN(\mathfrak{a}_\gamma)^{1/n}$ is satisfied then by choosing A suitably small this implies that $m = 0$ via Lemma 2.3.4.

Finally, assume that $A|\theta| > (N(\mathfrak{a}_\gamma)^{1/n}P_0)^{-1}$ is satisfied and that K is an imaginary quadratic number field. Upon choosing A even smaller if necessary, we find that

$$\Delta^{-1}|\mathrm{tr}(\theta m\omega_j)| \leq \frac{1}{2},$$

for all $j = 1, \ldots, n$ and thus

$$\Delta^{-1}|\mathrm{tr}(\theta m\omega_j)| = \left\|\Delta^{-1}\mathrm{tr}(\theta m\omega_j)\right\| \leq \left\|\Delta^{-1}\mathrm{tr}(\gamma m\omega_j)\right\| + \left\|\Delta^{-1}\mathrm{tr}(\alpha m\omega_j)\right\| < P_0^{-1},$$

for all $j = 1, \ldots, n$. Write $\boldsymbol{y} = (y_1, \ldots, y_n)$ for the image of θm under the isomorphism $K_\mathbb{R} \cong \mathbb{R}^n$ and let $\mathbf{T}$ be the trace form as above. The above inequality is equivalent to saying that there exists some $\boldsymbol{t} \in \mathbb{R}^n$ with $|\boldsymbol{t}| < P_0^{-1}$ such that

$$\mathbf{T}(\Delta^{-1}\boldsymbol{y}) = \boldsymbol{t}.$$

As before the inverse of $\mathbf{T}$ is a matrix with rational entries, whose absolute value is bounded by $O(1)$. Hence

$$|\boldsymbol{y}| = \Delta|\mathbf{T}^{-1}(\boldsymbol{t})| \ll |\boldsymbol{t}| < P_0^{-1}.$$

Further $|\boldsymbol{y}| = |\theta m|$, and since K is an imaginary quadratic number field we have $|\theta^{-1}| \asymp |\theta|^{-1}$ and so

$$|m| \ll (P_0|\theta|)^{-1}.$$

Hence for sufficiently small A we obtain

$$|m| < A^{1/2} N(\mathfrak{a}_\gamma)^{1/n}.$$

Choosing A to be suitably small implies $m = 0$ by Lemma 2.3.4. □

We now recall Davenport's shrinking lemma [29, Lemma 12.6].

Lemma 2.5.2. *Let $L\colon \mathbb{R}^m \to \mathbb{R}^m$ be a linear map. Let $a > 0$ be a real number and for a real number $Z > 0$ consider*

$$N(Z) = \left\{\boldsymbol{u} \in \mathbb{Z}^m \colon |\boldsymbol{u}| < aZ,\ \|(L(\boldsymbol{u}))_i\| < a^{-1}Z, \text{for all } i\right\}.$$

Then if $0 < Z \leq 1$ we have

$$N(1) \ll_m Z^{-m} N(Z).$$

As noted in [49] the lemma was originally only stated when $a \geq 1$ but we may extend the range of a to all positive real numbers since the result holds trivially if $0 < a < 1$.

2.6 Weyl differencing

One of the main innovations in [49] is to introduce an averaged van der Corput differencing approach in order to bound the contribution from the minor arcs. Since this cannot handle the entire range of minor arcs we need to supplement it with an estimate coming from conventional Weyl differencing.

Let $\alpha \in K_\mathbb{R}$. Throughout this section we will write

$$\alpha = \gamma + \theta,$$

where $\gamma \in K$ and $\theta \in K_\mathbb{R}$. Note as in [87, Lemma 2.1] we find

$$|S(\alpha)|^4 \ll P^{ns} \sum_{|\boldsymbol{x}|,|\boldsymbol{y}|<P} \prod_{i=1}^{s} \prod_{j=1}^{n} \min\left(P, \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x},\boldsymbol{y}))\|^{-1}\right). \tag{2.6.1}$$

This estimate is proved using a classical Weyl differencing procedure adjusted to this context. Following standard arguments as in Davenport [29, Chapter 13] we now transform this into a counting problem.

Given $\alpha \in \mathbb{R}$ and $P \geq 1$ define

$$N(\alpha, P) \coloneqq \#\left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{O}^{2s} \colon |\boldsymbol{x}| < P,\, |\boldsymbol{y}| < P,\, \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\| < P^{-1},\, \forall i, j\right\}.$$

For a fixed $\boldsymbol{x} \in \mathcal{O}^s$ write further

$$N(\boldsymbol{x}) \coloneqq \#\left\{\boldsymbol{y} \in \mathcal{O}^s \colon |\boldsymbol{y}| < P,\, \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\| < P^{-1},\, \forall i, j\right\},$$

so that

$$N(\alpha, P) = \sum_{|\boldsymbol{x}|<P} N(\boldsymbol{x}).$$

Let r_{ij} be integers such that $0 \leq r_{ij} < P$ for $i = 1, \ldots, s$, $j = 1, \ldots, n$. We claim that there exist no more than $N(\boldsymbol{x})$ integer tuples $\boldsymbol{y} \in \mathcal{O}^s$, which lie in a box whose edges have sidelengths at most P such that

$$\frac{r_{ij}}{P} \leq \{\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\} < \frac{r_{ij}+1}{P}$$

is satisfied for all $i = 1, \ldots, s$ and $j = 1, \ldots, n$, where $\{x\}$ denotes the fractional part of a real number x. Indeed, if $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are two such integer tuples satisfying the above system of inequalities then $|\boldsymbol{y}_1 - \boldsymbol{y}_2| < P$ and

$$\|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}_1 - \boldsymbol{y}_2))\| < P^{-1}$$

holds for all i, j. Hence, since $\boldsymbol{y} = \boldsymbol{0}$ is a possible solution, there are no more than $N(\boldsymbol{x})$ possible solutions to the system of inequalities above. Dividing the box $P\mathcal{B}$ into 2^{ns} boxes whose edges have sidelength at most P we find

$$\begin{aligned}\sum_{|\boldsymbol{y}|<P} \prod_{i=1}^{s} \prod_{j=1}^{i} \min\left(P, \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\|^{-1}\right) &\ll N(\boldsymbol{x}) \prod_{i,j} \sum_{r_{ij}=0}^{P} \min\left(P, \frac{P}{r_{ij}}, \frac{P}{P - r_{ij} - 1}\right) \\ &\ll N(\boldsymbol{x})(P \log P)^{ns}.\end{aligned}$$

Upon summing this estimate over $|\boldsymbol{x}| < P$ and using (2.6.1) we obtain

$$|S(\alpha)|^4 \ll P^{2ns}(\log P)^{ns} N(\alpha, P). \tag{2.6.2}$$

We now proceed to estimate $N(\alpha, P)$ using the results from the previous section.

For fixed $\boldsymbol{x} \in \mathcal{O}^s$ identifying $\mathcal{O}^s \cong \mathbb{R}^{ns}$ and given $\boldsymbol{y} \in \mathcal{O}^s$ one may view the map

$$\boldsymbol{y} \mapsto (\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y})))_{i,j}$$

as a linear map $\mathbb{R}^{ns} \to \mathbb{R}^{ns}$. Hence we can apply Lemma 2.5.2 where $N(\boldsymbol{x}) = N(1)$ in the notation of the lemma where Z is to be determined in due course. Summing over the $|\boldsymbol{x}| < P$ then yields

$$N(\alpha, P) \ll Z^{-ns} \# \big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{O}^{2s} \colon |\boldsymbol{x}| < P,\ |\boldsymbol{y}| < ZP, \\ \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\| < ZP^{-1},\ \forall i, j \big\}. \quad (2.6.3)$$

If we apply the same procedure to the quantity on the right hand side of (2.6.3), but now with the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$ reversed we obtain

$$N(\alpha, P) \ll Z^{-2ns} \# \big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{O}^{2s} \colon |\boldsymbol{x}| < ZP,\ |\boldsymbol{y}| < ZP, \\ \|\mathrm{tr}(6\alpha\omega_j B_i(\boldsymbol{x}, \boldsymbol{y}))\| < Z^2P^{-1},\ \forall i, j \big\}. \quad (2.6.4)$$

At this point we will employ Lemma 2.5.1. We wish to choose Z such that the bilinear forms appearing in the right hand side of (2.6.4) are forced to vanish. To this end, in the notation of the lemma we take $m = 6\Delta B_i(\boldsymbol{x}, \boldsymbol{y})$, $M \asymp 6Z^2P^2$ and $P_0^{-1} = Z^2P^{-1}$. Choose the parameter Z so that it satisfies

$$0 < Z < 1, \quad Z^2 \ll (P^2|\theta|N(\mathfrak{a}_\gamma)^{1/n})^{-1}, \quad Z^2 \ll \frac{P}{N(\mathfrak{a}_\gamma)^{1/n}},$$

as well as

$$Z^2 \ll \max\left(\frac{N(\mathfrak{a}_\gamma)^{1/n}}{P^2}, N(\mathfrak{a}_\gamma)^{1/n}|\theta|P \right),$$

where the implicit constants involved are sufficiently small such that the assumptions of Lemma 2.5.1 are satisfied. Provided K is an imaginary quadratic number field, Lemma 2.5.1 and (2.6.4) give

$$N(\alpha, P) \ll Z^{-2ns} \big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{O}^{2s} \colon |\boldsymbol{x}| < ZP,\ |\boldsymbol{y}| < ZP,\ B_i(\boldsymbol{x}, \boldsymbol{y}) = 0,\ i = 1, \ldots, s \big\},$$

where we note that clearly $6\Delta B_i(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $B_i(\boldsymbol{x}, \boldsymbol{y}) = 0$.

Since we assume that Davenport's Geometric Condition (2.4.1) is satisfied it follows from the simple observation (2.4.2) that

$$N(\alpha, P) \ll Z^{-2ns}(ZP)^{ns}.$$

From (2.6.2) for permissible Z as described above we therefore have

$$|S(\alpha)|^4 \ll P^{3ns+\varepsilon} Z^{-ns}.$$

The estimate is optimised when Z is as large as possible. Hence if we take

$$Z^2 \asymp \min\left\{ 1, (P^2|\theta|N(\mathfrak{a}_\gamma)^{1/n})^{-1}, \frac{P}{N(\mathfrak{a}_\gamma)^{1/n}}, \max\left(\frac{N(\mathfrak{a}_\gamma)^{1/n}}{P^2}, N(\mathfrak{a}_\gamma)^{1/n}|\theta|P \right) \right\}$$

then Z is clearly in the permissible range, and we deduce

$$|S(\alpha)|^4 \ll P^{3ns+\varepsilon}\left(1+P^2|\theta|N(\mathfrak{a}_\gamma)^{1/n}+P^{-1}N(\mathfrak{a}_\gamma)^{1/n}\right.$$
$$\left.+\min\left(P^2N(\mathfrak{a}_\gamma)^{-1/n},\left(N(\mathfrak{a}_\gamma)^{1/n}|\theta|P\right)^{-1}\right)\right)^{\frac{ns}{2}}.$$

In particular, if $N(\mathfrak{a}_\gamma)^{1/n}\leq P^{3/2}$ then $P^{-1}N(\mathfrak{a}_\gamma)^{1/n}\leq P^{1/2}$ and so we find

$$|S(\alpha)| \ll P^{ns+\varepsilon}\left(N(\mathfrak{a}_\gamma)^{1/n}|\theta|+(N(\mathfrak{a}_\gamma)^{1/n}|\theta|P^3)^{-1}+P^{-3/2}\right)^{\frac{ns}{8}}$$

in this case. Finally since $X^{1/2}\leq X/Y+Y$ for any two positive real numbers X and Y we see that the last term of the right hand side above is dominated by the other two summands. We summarise the main result of this section.

Lemma 2.6.1. *Let $K/\mathbb{Q}$ be an imaginary quadratic number field. Let $\alpha\in K_{\mathbb{R}}$ and write $\alpha=\gamma+\theta$ where $\gamma\in K$ and $\theta\in K_{\mathbb{R}}$. If $N(\mathfrak{a}_\gamma)^{1/n}\leq P^{3/2}$ then we have*

$$S(\alpha) \ll P^{ns+\varepsilon}\left(N(\mathfrak{a}_\gamma)^{1/n}|\theta|+(N(\mathfrak{a}_\gamma)^{1/n}|\theta|P^3)^{-1}\right)^{\frac{ns}{8}}.$$

This bound will be useful for the range in the minor arcs when the parameter θ is small.

2.7 Pointwise van der Corput differencing and the singular series

In this section we will perform a pointwise van der Corput differencing argument, in order to show that the singular series converges absolutely. This argument works over a general number field. We start by considering the exponential sum $S(\gamma)$, where $\gamma\in K$ and we set $P=N(\mathfrak{a}_\gamma)$. Further in this section we take the box $\mathcal{B}=\mathcal{B}_{\mathfrak{S}}=\{(\sum_j x_{ij}\omega_j)_i\in K_{\mathbb{R}}^s\colon 0\leq x_{ij}<1\}$ so that the goal of this section is to study the sum S_γ as it was defined in Section 2.4.3. To be completely explicit with our choice of box we then have

$$S_\gamma=S(\gamma)=\sum_{0\leq \boldsymbol{x}<N(\mathfrak{a}_\gamma)} e\left(\operatorname{tr}(\gamma C(\boldsymbol{x}))\right),$$

where the condition $0\leq\boldsymbol{x}<N(\mathfrak{a}_\gamma)$ denotes the sum over elements $\boldsymbol{x}=\left(\sum_j x_{ij}\omega_j\right)_i\in\mathcal{O}^s$ such that $0\leq x_{ij}<N(\mathfrak{a}_\gamma)$ holds. The main goal of this section is to establish the bound

$$S_\gamma \ll N(\mathfrak{a}_\gamma)^{s(n-1/6)+\varepsilon}. \tag{2.7.1}$$

Let H be a positive integer that satisfies $H \leq N(\mathfrak{a}_\gamma)$. Clearly we have

$$H^{ns} S(\gamma) = \sum_{0 \leq \boldsymbol{h} < H} \sum_{\substack{0 \leq \boldsymbol{x} < N(\mathfrak{a}_\gamma) \\ 0 \leq \boldsymbol{x}+\boldsymbol{h} < N(\mathfrak{a}_\gamma)}} e\left(\operatorname{tr}(\gamma C(\boldsymbol{x}+\boldsymbol{h}))\right).$$

Interchanging the order of summation gives

$$H^{ns} S(\gamma) = \sum_{0 \leq \boldsymbol{x} < N(\mathfrak{a}_\gamma)} \sum_{\substack{0 \leq \boldsymbol{h} < H \\ 0 \leq \boldsymbol{x}+\boldsymbol{h} < N(\mathfrak{a}_\gamma)}} e\left(\operatorname{tr}(\gamma C(\boldsymbol{x}+\boldsymbol{h}))\right).$$

An application of Cauchy-Schwarz yields

$$H^{2ns} |S(\gamma)|^2 \ll N(\mathfrak{a}_\gamma)^{ns} \sum_{0 \leq \boldsymbol{x} < N(\mathfrak{a}_\gamma)} \left| \sum_{\substack{0 \leq \boldsymbol{h} < H \\ 0 \leq \boldsymbol{x}+\boldsymbol{h} < N(\mathfrak{a}_\gamma)}} e\left(\operatorname{tr}(\gamma C(\boldsymbol{x}+\boldsymbol{h}))\right) \right|^2.$$

Expanding the square one obtains

$$H^{2ns} |S(\gamma)|^2 \ll N(\mathfrak{a}_\gamma)^{ns} \sum_{0 \leq \boldsymbol{x} < N(\mathfrak{a}_\gamma)} \sum_{\substack{0 \leq \boldsymbol{h}_1, \boldsymbol{h}_2 < H \\ 0 \leq \boldsymbol{x}+\boldsymbol{h}_1, \boldsymbol{x}+\boldsymbol{h}_2 < N(\mathfrak{a}_\gamma)}} e\left(\operatorname{tr}(\gamma C(\boldsymbol{x}+\boldsymbol{h}_1) - C(\boldsymbol{x}+\boldsymbol{h}_2))\right).$$

Set $\boldsymbol{y} = \boldsymbol{x}+\boldsymbol{h}_2$ and $\boldsymbol{h} = \boldsymbol{h}_1 - \boldsymbol{h}_2$. Note that after this change of coordinates each value of $\boldsymbol{h}$ in the sum above appears at most H^{ns} times. Therefore the previous display gives

$$H^{ns} |S(\gamma)|^2 \ll N(\mathfrak{a}_\gamma)^{ns} \sum_{|\boldsymbol{h}| \leq H} |T(\boldsymbol{h}, \gamma)|, \tag{2.7.2}$$

where

$$T(\boldsymbol{h}, \gamma) = \sum_{\boldsymbol{y} \in \mathcal{R}(\boldsymbol{h})} e\left(\operatorname{tr}(\gamma(C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})))\right),$$

and where $\mathcal{R}(\boldsymbol{h})$ is a box whose sidelengths are $O(N(\mathfrak{a}_\gamma))$. We take the square of the absolute value of this expression, and expand the resulting sum in order to obtain

$$|T(\boldsymbol{h}, \gamma)|^2 = \sum_{\boldsymbol{y}, \boldsymbol{z} \in \mathcal{R}(\boldsymbol{h})} e\left(\operatorname{tr}(\gamma(C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y}) - C(\boldsymbol{z}+\boldsymbol{h}) + C(\boldsymbol{z})))\right).$$

Making the change of variables $\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{w}$ we find

$$|T(\boldsymbol{h}, \gamma)|^2 = \sum_{|\boldsymbol{w}| < N(\mathfrak{a}_\gamma)} \sum_{\boldsymbol{z}} e\left(\operatorname{tr}(\gamma C(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{z}))\right),$$

where the inner sum ranges over a (potentially empty) box $\mathcal{S}(\boldsymbol{h}, \boldsymbol{w})$ whose sidelengths are $O(N(\mathfrak{a}_\gamma))$ and where we write $C(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{z})$ for the multilinear form given by

$$C(\boldsymbol{w}, \boldsymbol{h}, \boldsymbol{z}) = C(\boldsymbol{w}+\boldsymbol{h}+\boldsymbol{z}) - C(\boldsymbol{w}+\boldsymbol{z}) - C(\boldsymbol{h}+\boldsymbol{z}) + C(\boldsymbol{z}).$$

In particular we have

$$C(\boldsymbol{w},\boldsymbol{h},\boldsymbol{z}) = 6\sum_{i=1}^{s} z_i B_i(\boldsymbol{w},\boldsymbol{h}) + \Psi(\boldsymbol{w},\boldsymbol{h}),$$

where B_i are the bilinear forms associated to C, and where Ψ is a certain polynomial whose precise shape is of no importance to us. Therefore we find

$$|T(\boldsymbol{h},\gamma)|^2 = \sum_{\boldsymbol{w}}\sum_{\boldsymbol{z}} e\left(\mathrm{tr}\left(6\gamma\sum_{i=1}^{s} z_i B_i(\boldsymbol{w},\boldsymbol{h}) + \gamma\Psi(\boldsymbol{w},\boldsymbol{h})\right)\right).$$

Writing $z_i = \sum_j z_{ij}\omega_j$ we may regard the inner sum as an exponential sum over integer variables z_{ij}. This is a linear exponential sum and the coefficient of z_{ij} is given by $6\mathrm{tr}(\gamma\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))$. A standard argument regarding geometric sums now yields

$$|T(\boldsymbol{h},\gamma)|^2 \ll \sum_{\boldsymbol{w}}\prod_{i=1}^{s}\prod_{j=1}^{n} \min\left(N(\mathfrak{a}_\gamma), \|6\mathrm{tr}(\gamma\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\|^{-1}\right).$$

In particular the same argument that led to (2.6.2) shows that

$$|T(\boldsymbol{h},\gamma)|^2 \ll N(\mathfrak{a}_\gamma)^{ns+\varepsilon} N(\gamma, N(\mathfrak{a}_\gamma), \boldsymbol{h}), \tag{2.7.3}$$

where

$$N(\gamma, N(\mathfrak{a}_\gamma), \boldsymbol{h}) = \#\left\{\boldsymbol{w}\in\mathcal{O}^s \colon |\boldsymbol{w}| < N(\mathfrak{a}_\gamma),\ \|6\mathrm{tr}(\gamma\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| < N(\mathfrak{a}_\gamma)^{-1},\ \forall i,j\right\}.$$

Note that the condition in the sum already implies that $6\Delta B_i(\boldsymbol{x},\boldsymbol{y}) \in \mathfrak{a}_\gamma$ holds for all i, but we prefer to write it in the above shape in order to highlight the similarities with the argument in the previous section.

As in Section 2.6 we may regard $\boldsymbol{w} \mapsto \mathrm{tr}(\gamma\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))$ as a linear map $\mathbb{R}^{ns} \to \mathbb{R}^{ns}$. Hence we can apply Lemma 2.5.2 so that for any $Z \in (0,1]$ we have

$$N(\gamma, N(\mathfrak{a}_\gamma), \boldsymbol{h}) \ll Z^{-ns}\#\left\{\boldsymbol{w}\in\mathcal{O}^s \colon |\boldsymbol{w}| < ZN(\mathfrak{a}_\gamma),\ \|6\mathrm{tr}(\gamma\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| < ZN(\mathfrak{a}_\gamma)^{-1}\right\}.$$

We now wish to choose Z in such a way that we can apply Lemma 2.5.1. In the notation of this lemma we have $m = \Delta\omega_j B_i(\boldsymbol{w},\boldsymbol{h})$ and $\theta = 0$. We take $Z \in (0,1]$ such that $Z \asymp H^{-1}N(\mathfrak{a}_\gamma)^{\frac{1}{n}-1}$ for a suitable implied constant. Then Lemma 2.5.1 implies

$$N(\gamma, P, \boldsymbol{h}) \ll H^{ns}N(\mathfrak{a}_\gamma)^{ns-s}\#\left\{\boldsymbol{w}\in\mathcal{O}^s \colon |\boldsymbol{w}| < H^{-1}N(\mathfrak{a}_\gamma)^{1/n},\ B_i(\boldsymbol{w},\boldsymbol{h}) = 0,\ \forall i,j\right\}.$$

Recalling that $r(\boldsymbol{h})$ is the rank of $B_i(\boldsymbol{h},\cdot)\colon K_{\mathbb{R}}^s \to K_{\mathbb{R}}^s$, using (2.7.3) we find

$$T(\boldsymbol{h},\gamma) \ll N(\mathfrak{a}_\gamma)^{ns-\frac{r(\boldsymbol{h})}{2}+\varepsilon} H^{\frac{nr(\boldsymbol{h})}{2}}.$$

Hence (2.7.2) delivers

$$|S(\gamma)|^2 \ll H^{-ns} N(\mathfrak{a}_\gamma)^{2ns+\varepsilon} \sum_{|\boldsymbol{h}| \le H} \left(H^n N(\mathfrak{a}_\gamma)^{-1} \right)^{\frac{r(\boldsymbol{h})}{2}} .$$

By (2.4.1), for any r the number of $\boldsymbol{h}$ with $r(\boldsymbol{h}) = r$ is $O(H^{nr})$. Therefore we find

$$|S(\gamma)|^2 \ll H^{-ns} N(\mathfrak{a}_\gamma)^{2ns+\varepsilon} \sum_{r=0}^{s} \left(H^{3n} N(\mathfrak{a}_\gamma)^{-1} \right)^{\frac{r}{2}} .$$

The sum is maximal either when $r = 0$ or when $r = s$, and thus

$$|S(\gamma)|^2 \ll H^{-ns} N(\mathfrak{a}_\gamma)^{2ns+\varepsilon} \left(1 + H^{3ns/2} N(\mathfrak{a}_\gamma)^{-s/2} \right) .$$

Choosing $H = \lfloor N(\mathfrak{a}_\gamma) \rfloor^{1/3n}$ this finally yields

$$S(\gamma) \ll N(\mathfrak{a}_\gamma)^{s(n-1/6)+\varepsilon}.$$

2.7.1 Proof of Theorem 2.4.4

By Theorem 2.4.3 it suffices to show that $\mathfrak{S}(R)$ converges absolutely as $R \to \infty$.

Given a positive integer k the number of ideals of $\mathcal{O}$ of norm k is $O(k^\varepsilon)$ using the divisor bound. Hence together with Lemma 2.3.1 we obtain that the number of $\gamma \in K/\mathcal{O}$ such that $N(\mathfrak{a}_\gamma) = k$ is bounded by $O(k^{1+\varepsilon})$. Thus, using (2.7.1) we find

$$\mathfrak{S}(R) \ll \sum_{k=0}^{R} k^{-ns+1+\varepsilon} k^{ns-s/6} = \sum_{k=0}^{R} k^{1-s/6+\varepsilon}.$$

Therefore $\mathfrak{S}(R)$ converges absolutely to some real number $\mathfrak{S}$ as $R \to \infty$ provided $s \ge 13$. □

We remark that using the ideas of Heath-Brown [49, Section 7] it would be possible to establish the absolute convergence of $\mathfrak{S}(R)$ already for $s \ge 11$.

2.8 Van der Corput on average

In this section, we work towards a bound for the Weyl sum $S(\alpha)$ on the minor arcs. As observed by Heath-Brown, the simple pointwise van der Corput differencing is not sufficient to improve on Davenport's result for $s \ge 16$.

It is however possible to exploit the fact that we are averaging both over the modulus $\mathfrak{a}_\gamma$ as well as the integration variable β in the minor arcs, thus leading to a version of van der Corput differencing on average.

From now on we continue to work with the box $\mathcal{B} = \mathcal{B}(\boldsymbol{z})$ as defined in the beginning of Section 2.4.3. Instead of a pointwise bound for $S(\alpha)$, we will seek to bound the mean-square average

$$M(\alpha,\kappa) = \int_{|\beta-\alpha|<\kappa} |S(\beta)|^2 d\beta$$

for $\alpha \in K_{\mathbb{R}}$ and a small parameter $\kappa \in (0,1)$, where we remind the reader that the integration is over a region of $K_{\mathbb{R}}$.

In conjunction with the Cauchy-Schwarz inequality and an appropriate dyadic dissection of the minor arcs, a satisfactory bound for $M(\alpha,\kappa)$ will be sufficient to control the total minor arc contribution.

The idea now is that the mean square integral automatically shortens all the n coordinates of h_1 in the van der Corput differencing, allowing us to effectively save a factor $\frac{H^n}{P^n}$ over the pointwise bound. Here and throughout we denote $\boldsymbol{h} = (h_i)_i = \left(\sum_j h_{ij}\omega_j\right)_i \in \mathcal{O}^s$.

To this end, we initiate the van der Corput differencing with parameters $1 \leq H_{ij} \leq P$ to be determined, obtaining

$$\prod_{i,j} H_{ij} S(\beta) = \sum_{0\leq h_{ij}<H_{ij}} \sum_{\boldsymbol{x}+\boldsymbol{h}\in P\mathcal{B}} e\left(\mathrm{tr}(\beta C(\boldsymbol{x}+\boldsymbol{h}))\right) = \sum_{\boldsymbol{x}\in\mathcal{O}^s} \sum_{\boldsymbol{x}+\boldsymbol{h}\in P\mathcal{B}} e\left(\mathrm{tr}(\beta C(\boldsymbol{x}+\boldsymbol{h}))\right),$$

where implicitly we still restrict to $\boldsymbol{h}$ such that $0 \leq h_{ij} < H_{ij}$ is satisfied. Note that the condition $H_{ij} \leq P$ ensures that the sum over $\boldsymbol{x}$ is restricted to $O(P^{ns})$ many summands. An application of Cauchy-Schwarz thus yields

$$\prod_{i,j} H_{ij}^2 |S(\beta)|^2 \ll P^{ns} \sum_{\boldsymbol{x}\in\mathcal{O}^s} \left| \sum_{\boldsymbol{x}+\boldsymbol{h}\in P\mathcal{B}} e\left(\mathrm{tr}(\beta C(\boldsymbol{x}+\boldsymbol{h}))\right)\right|^2.$$

Opening the square on the RHS, we obtain

$$\prod_{i,j} H_{ij}^2 |S(\beta)|^2 \ll P^{ns} \sum_{\boldsymbol{x}\in\mathcal{O}^s} \sum_{\boldsymbol{x}+\boldsymbol{h_1},\boldsymbol{x}+\boldsymbol{h_2}\in P\mathcal{B}} e\left(\mathrm{tr}(\beta\left[C(\boldsymbol{x}+\boldsymbol{h_1}) - C(\boldsymbol{x}+\boldsymbol{h_2})\right])\right)$$

On substituting $\boldsymbol{y} = \boldsymbol{x}+\boldsymbol{h_2}$ and $\boldsymbol{h} = \boldsymbol{h_1}-\boldsymbol{h_2}$, this becomes

$$\prod_{i,j} H_{ij}^2 |S(\beta)|^2 \ll P^{ns} \sum_{|h_{ij}|\leq H_{ij}} w(\boldsymbol{h}) \sum_{\boldsymbol{y}\in\mathcal{R}(\boldsymbol{h})} e\left(\mathrm{tr}(\beta\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right])\right)$$

where $w(\boldsymbol{h}) = \#\{\boldsymbol{h_1},\boldsymbol{h_2} : \boldsymbol{h} = \boldsymbol{h_1}-\boldsymbol{h_2}\} \leq \prod_{i,j} H_{ij}$ and $\mathcal{R}(\boldsymbol{h})$ is a certain box depending only on $\boldsymbol{h}$.

Instead of taking absolute values, we now first integrate over $\beta = \sum_j \beta_j \omega_j$ with a smooth cutoff function to obtain

$$\begin{aligned} M(\alpha,\kappa) &\leq e^n \int_{K_{\mathbb{R}}} \exp\left(-\frac{\sum_j (\beta_j - \alpha_j)^2}{\kappa^2}\right) \cdot |S(\beta)|^2 d\beta \\ &\ll \frac{P^{ns}}{\prod_{i,j} H_{ij}^2} \sum_{|h_{ij}| \leq H_{ij}} w(\boldsymbol{h}) \sum_{\boldsymbol{y} \in \mathcal{R}(\boldsymbol{h})} I(\boldsymbol{h},\boldsymbol{y}) \\ &\ll \frac{P^{ns}}{\prod_{i,j} H_{ij}} \sum_{|h_{ij}| \leq H_{ij}} \left| \sum_{\boldsymbol{y} \in \mathcal{R}(\boldsymbol{h})} I(\boldsymbol{h},\boldsymbol{y}) \right|, \end{aligned}$$

where

$$\begin{aligned} I(\boldsymbol{h},\boldsymbol{y}) &= \int_{K_{\mathbb{R}}} \exp\left(-\frac{\sum_j (\beta_j - \alpha_j)^2}{\kappa^2}\right) \cdot e\left(\operatorname{tr}(\beta\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right])\right) d\beta \\ &= \pi^{n/2}\kappa^n \prod_{j=1}^n \exp\big(-\pi^2\kappa^2 \operatorname{tr}(\omega_j\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right])^2\big) \cdot e(\operatorname{tr}(\alpha\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right])). \end{aligned}$$

Heuristically, for large $h_1 \in \mathcal{O}$, we should have $C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y}) \approx h_1 \cdot \frac{\partial C(\boldsymbol{y})}{\partial x_1}$ so that by our choice of the box $\mathcal{B}(\boldsymbol{z})$, this difference is large. But then for some j, the trace of this number multiplied with ω_j must be large as well, leading to a negligible contribution to $M(\alpha,\kappa)$ from those terms, thus effectively cutting down the range to small h_1.

We now fix the choice $H_{ij} = H$ for $i \neq 1$ and $H_{1j} = cP$ for a sufficiently small constant c and make the above heuristic discussion precise. For $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{h})$ we have

$$C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y}) = h_1 \cdot \frac{\partial C(\boldsymbol{y})}{\partial x_1} + O(HP^2 + |h_1|^2|\boldsymbol{y}|).$$

If the width of the box $\mathcal{B}(\boldsymbol{z})$ and c are sufficiently small, the fact that $\frac{\partial C(\boldsymbol{z})}{\partial x_1} \neq 0$ then implies that

$$|C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})| \gg |h_1| \cdot P^2$$

unless $|h_1| \ll H$. Additionally, unless $|h_1| \ll \frac{(\log P)^2}{\kappa P^2}$, we even have that

$$|C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})| \gg \frac{(\log P)^2}{\kappa}$$

so that for some j we must have

$$|\operatorname{tr}\left(\omega_j\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right]\right)| \gg \frac{(\log P)^2}{\kappa}$$

and we infer from our previous calculations that the contribution of such $\boldsymbol{h}$ to $M(\alpha,\kappa)$ is $O(1)$. Hence,

$$M(\alpha,\kappa) \ll 1 + \frac{P^{ns-n}}{H^{ns-n}} \sum_{|h_i| \ll H} \left| \sum_{\boldsymbol{y}} I(\boldsymbol{h},\boldsymbol{y}) \right|$$

if we choose $\kappa \asymp \frac{(\log P)^2}{HP^2}$.

Moreover, the range $|\beta-\alpha| \geq \kappa \log P$ in the definition of $I(\boldsymbol{h},\boldsymbol{y})$ clearly gives a total contribution of $O(1)$ to $M(\alpha,\kappa)$ so that we end up with the estimate

$$M(\alpha,\kappa) \ll 1 + \frac{P^{ns-n}}{H^{ns-n}} \sum_{|h_i| \ll H} \int_{|\beta-\alpha|<\kappa\log P} |T(\boldsymbol{h},\beta)| d\beta$$

with

$$T(\boldsymbol{h},\beta) = \sum_{\boldsymbol{y}\in\mathcal{R}(\boldsymbol{h})} e\left(\operatorname{tr}(\beta\left[C(\boldsymbol{y}+\boldsymbol{h}) - C(\boldsymbol{y})\right])\right).$$

As in Section 2.7, we obtain

$$|T(\boldsymbol{h},\beta)|^2 \ll P^{ns+\varepsilon} N(\beta,P,\boldsymbol{h})$$

where

$$N(\beta,P,\boldsymbol{h}) = \#\{\boldsymbol{w}\in\mathcal{O}^s : |\boldsymbol{w}| < P, \|6\operatorname{tr}(\beta\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| < P^{-1}, \forall i,j\}$$

so that

$$M(\alpha,\kappa) \ll 1 + \frac{\kappa^n P^{\frac{3ns}{2}-n+\varepsilon}}{H^{ns-n}} \sum_{|h_i|\ll H} \max_{\beta\in\mathcal{I}} N(\beta,P,\boldsymbol{h})^{\frac{1}{2}} \tag{2.8.1}$$

for $\mathcal{I} = \{\beta : |\beta-\alpha| \leq \kappa\log P\}$.

We next claim that

$$\max_{\beta\in\mathcal{I}} N(\beta,P,\boldsymbol{h}) \ll P^{\varepsilon} N(\alpha,P,\boldsymbol{h}).$$

Indeed, consider a vector $\boldsymbol{w}$ counted by $N(\beta,P,\boldsymbol{h})$. It thus satisfies $|\boldsymbol{w}| < P$ as well as $\|6\operatorname{tr}(\beta\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| < P^{-1}$ so that

$$\|6\operatorname{tr}(\alpha\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| \ll \frac{1}{P} + |\beta-\alpha|\cdot|B_i(\boldsymbol{w},\boldsymbol{h})| \ll \frac{1}{P} + \kappa\log P\cdot HP \ll \frac{(\log P)^3}{P}.$$

We thus obtain

$$\begin{aligned} N(\beta,P,\boldsymbol{h}) &\ll \#\{\boldsymbol{w}\in\mathcal{O}^s : |\boldsymbol{w}| < P, \|6\operatorname{tr}(\alpha\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| \ll \frac{(\log P)^3}{P}, \forall i,j\} \\ &\ll P^{\varepsilon} N(\alpha,P,\boldsymbol{h}) \end{aligned}$$

where the last estimate is a consequence of Lemma 2.5.2 upon choosing suitable $Z \asymp (\log P)^{-3}$.

We conclude that

$$M(\alpha,\kappa) \ll 1 + \frac{\kappa^n P^{\frac{3ns}{2}-n+\varepsilon}}{H^{ns-n}} \sum_{|h_i| \ll H} N(\alpha, P, \boldsymbol{h})^{\frac{1}{2}}.$$

Let $\alpha = \gamma + \theta$ with $\gamma \in K$ and $\theta \in K_{\mathbb{R}}$ (which we think of as being small). We are now prepared for an application of Lemmas 2.5.2 and 2.5.1. Indeed, Lemma 2.5.2 implies that

$$N(\alpha, P, \boldsymbol{h}) \ll Z^{-ns} \#\{\boldsymbol{w} \in \mathcal{O}^s : |\boldsymbol{w}| < ZP, \|6\operatorname{tr}(\alpha\omega_j B_i(\boldsymbol{w},\boldsymbol{h}))\| < ZP^{-1}, \forall i,j\}.$$

Following Heath-Brown, we will make two different choices of Z: In the first one, we will choose $Z = Z_1$ sufficiently small so that Lemma 2.5.1 implies that $B_i(\boldsymbol{w},\boldsymbol{h}) = 0$. In the second choice $Z = Z_2$, we will only force $6\Delta B_i(\boldsymbol{w},\boldsymbol{h}) \in \mathfrak{a}_\gamma$, a consequence followed by a study of how often such a divisibility property can occur, crucially using an average over γ.

By Lemma 2.5.1, if we choose $Z \leq 1$ satisfying

$$Z \ll \frac{P}{N(\mathfrak{a}_\gamma)^{1/n}}$$

and

$$Z \ll \frac{1}{PH|\theta|N(\mathfrak{a}_\gamma)^{1/n}}$$

we can conclude that $6\Delta B_i(\boldsymbol{w},\boldsymbol{h}) \in \mathfrak{a}_\gamma$. If, moreover

$$Z \ll \frac{N(\mathfrak{a}_\gamma)^{1/n}}{PH}$$

or

$$Z \ll |\theta| P N(\mathfrak{a}_\gamma)^{1/n}$$

we obtain that $B_i(\boldsymbol{w},\boldsymbol{h}) = 0$. Here, all the implicit constants need to be sufficiently small in order to satisfy the conditions in Lemma 2.5.1.

Writing

$$\eta = |\theta| + \frac{1}{P^2 H} \tag{2.8.2}$$

we should therefore choose

$$Z_1 \asymp \min\left(N(\mathfrak{a}_\gamma)^{1/n} P\eta, \frac{1}{PH\eta N(\mathfrak{a}_\gamma)^{1/n}}\right),$$

noting that this automatically implies that $Z_1 \leq 1$. Similarly we should choose

$$Z_2 \asymp \min\left(1, \frac{1}{PH\eta N(\mathfrak{a}_\gamma)^{1/n}}\right).$$

In the application with $Z = Z_1$ we thus obtain

$$\begin{aligned} N(\alpha, P, \boldsymbol{h}) &\ll Z_1^{-ns} \#\{\boldsymbol{w} \in \mathcal{O}^s : |\boldsymbol{w}| < Z_1 P, B_i(\boldsymbol{w}, \boldsymbol{h}) = 0, \forall i\} \\ &\ll Z_1^{-ns} \cdot (Z_1 P)^{n(s-r)} \\ &\ll P^{ns} \cdot \left(\frac{1}{N(\mathfrak{a}_\gamma)^{1/n} P^2 \eta} + H\eta N(\mathfrak{a}_\gamma)^{1/n}\right)^{nr} \end{aligned}$$

with $r = r(\boldsymbol{h})$. Instead, in the application with $Z = Z_2$, we end up with the bound

$$N(\alpha, P, \boldsymbol{h}) \ll Z_2^{-ns} \#\{\boldsymbol{w} \in \mathcal{O}^s : |\boldsymbol{w}| < Z_2 P, 6\Delta B_i(\boldsymbol{w}, \boldsymbol{h}) \in \mathfrak{a}_\gamma, \forall i\}. \tag{2.8.3}$$

We thus need to count vectors $\boldsymbol{w}$ with $6\Delta B_i(\boldsymbol{w}, \boldsymbol{h}) \in \mathfrak{a}_\gamma$. For any prime ideal $\mathfrak{p}$, let $r_\mathfrak{p}(\boldsymbol{h})$ be the rank of $M(\boldsymbol{h})$ modulo $\mathfrak{p}$. Clearly, $r_\mathfrak{p}(\boldsymbol{h}) \leq r(\boldsymbol{h}) = r$ with strict inequality if and only if $\mathfrak{p}$ divides all $r \times r$ minors of $M(\boldsymbol{h})$. This means that there are only relatively few such 'bad' primes, which we will exploit later.

We now decompose $\mathfrak{a}_\gamma = \mathfrak{q}_1 \cdot \mathfrak{q}_2$ where $\mathfrak{q}_1$ contains all the primes $\mathfrak{p}$ dividing $\mathfrak{a}_\gamma$ with $r_\mathfrak{p}(\boldsymbol{h}) < r$ and $\mathfrak{q}_2$ consists of those with $r_\mathfrak{p}(\boldsymbol{h}) = r$.

As we are looking for an upper bound, we can replace $\mathfrak{a}_\gamma$ by the larger $\mathfrak{q}_2$ in (2.8.3).

For fixed $\boldsymbol{h}$ with $r(\boldsymbol{h}) = r$, the condition $6\Delta B_i(\boldsymbol{h}, \boldsymbol{w}) \in \mathfrak{q}_2, \forall i$ defines a lattice $\Lambda(\boldsymbol{h})$ for $\boldsymbol{w} \in \mathcal{O}^s$ which we view as a lattice in $\mathbb{R}^{ns}$.

To estimate the number of integer points in such a lattice we use [49, Lemma 5.1] implying that

$$\#\{\boldsymbol{x} \in \Lambda(\boldsymbol{h}) : |\boldsymbol{x}| \leq B\} \ll \prod_i \left(1 + \frac{B}{\lambda_i}\right) \tag{2.8.4}$$

where $\lambda_1, \ldots, \lambda_{ns}$ are the successive minima of $\Lambda(\boldsymbol{h})$.

In order to make this estimate useful, we need a bound on the determinant $d(\Lambda)$ which is proportional to $\prod_i \lambda_i$ as well as a bound on the skewness of the measure, i.e. upper and lower bounds for the λ_i.

For the determinant, we note that for $\mathfrak{p}^e \mid \mathfrak{q}_2$, the matrix $M(\boldsymbol{h})$ has rank r modulo $\mathfrak{p}$ (hence also modulo $\mathfrak{p}^e$) and therefore $B_i(\boldsymbol{h}, \boldsymbol{w})$ has $N(\mathfrak{p}^e)^{s-r}$ solutions modulo $\mathfrak{p}^e$ so that $N(\mathfrak{p}^e)^r$ divides $d(\Lambda)$. It thus follows that $N(\mathfrak{q}_2)^r \mid d(\Lambda)$ and hence $d(\Lambda) \geq N(\mathfrak{q}_2)^r$.

Regarding the skewness, we clearly have $\lambda_i \gg 1$ for all i, while in the other direction we have $\mathfrak{q}_2 \mathcal{O}^s \subset \Lambda(\boldsymbol{h})$ so that Lemma 2.3.4 implies $\lambda_i \ll N(\mathfrak{q}_2)^{1/n}$.

Optimizing the RHS of (2.8.4) with these constraints shows that the maximum is obtained when rn of the λ_i are of order $N(\mathfrak{q}_2)^{1/n}$ while the others are of order 1.

This shows that

$$N(\alpha,P,\boldsymbol{h}) \ll Z_2^{-ns}\left(1+\frac{Z_2P}{N(\mathfrak{q}_2)^{1/n}}\right)^{rn}\cdot (Z_2P)^{(s-r)n} = P^{ns}\left(\frac{1}{Z_2P}+\frac{1}{N(\mathfrak{q}_2)^{1/n}}\right)^{rn}$$

if $Z_2P \gg 1$ but we note that the bound is trivially true for $Z_2P \ll 1$.

Recalling our choice of Z_2, this bound becomes

$$N(\alpha,P,\boldsymbol{h}) \ll P^{ns}\left(\frac{1}{P}+\frac{1}{N(\mathfrak{q}_2)^{1/n}}+H\eta N(\mathfrak{a}_\gamma)^{1/n}\right)^{rn}.$$

Combining our two estimates, we obtain

$$N(\alpha,P,\boldsymbol{h}) \ll P^{ns}\left(\frac{1}{P}+H\eta N(\mathfrak{a}_\gamma)^{1/n}+\min\left(\frac{1}{N(\mathfrak{a}_\gamma)^{1/n}P^2\eta},\frac{1}{N(\mathfrak{q}_2)^{1/n}}\right)\right)^{rn}.$$

We now need to insert this into our expression for $M(\alpha,\kappa)$ which already involves the average over $\boldsymbol{h}$. Additionally, we want to average over $\mathfrak{a}_\gamma$ allowing us to use that $N(\mathfrak{q}_2)$ is almost as large as $N(\mathfrak{a}_\gamma)$ most of the time.

Our object of study thus becomes

$$A(\theta,R,H,P) := \sum_{\gamma:N(\mathfrak{a}_\gamma)^{1/n}\sim R}\ \sum_{|h_i|\ll H} N(\alpha,P,\boldsymbol{h})^{1/2} \tag{2.8.5}$$

where we continue to write $\alpha=\gamma+\theta$ and we remind the reader of the notation $q\sim R$ for the dyadic condition $R<q\leq 2R$. We then obtain

$$A(\theta,R,H,P) \ll R^nP^{ns/2}\sum_{|h_i|\ll H}\ \sum_{N(\mathfrak{a})^{1/n}\sim R}\left(\frac{1}{P}+H\eta R+\min\left(\frac{1}{RP^2\eta},\frac{1}{N(\mathfrak{q}_2)^{1/n}}\right)\right)^{\frac{r(\boldsymbol{h})n}{2}}$$

where we used that there are at most $N(\mathfrak{a})$ choices of γ with $\mathfrak{a}_\gamma=\mathfrak{a}$ by Lemma 2.3.1 and we remind the reader that $\mathfrak{q}_2$ depends on $\mathfrak{a}$ and $\boldsymbol{h}$.

We thus proceed to estimate

$$V(\boldsymbol{h},R,\eta) := \sum_{N(\mathfrak{a})^{1/n}\sim R}\min\left(\frac{1}{RP^2\eta},\frac{1}{N(\mathfrak{q}_2)^{1/n}}\right)^{\frac{rn}{2}}$$

for $r=r(\boldsymbol{h})$ via a dyadic decomposition as follows:

$$\begin{aligned}V(\boldsymbol{h},R,\eta) &\ll P^\varepsilon\max_{S\leq R}\sum_{N(\mathfrak{q}_1)^{1/n}\sim S}\ \sum_{N(\mathfrak{q}_2)^{1/n}\sim\frac{R}{S}}\min\left(\frac{1}{RP^2\eta},\frac{S}{R}\right)^{\frac{rn}{2}}\\ &\ll P^\varepsilon\max_{S\leq R}\frac{R^n}{S^n}\min\left(\frac{1}{RP^2\eta},\frac{S}{R}\right)^{\frac{rn}{2}}\#\{\mathfrak{q}_1 : N(\mathfrak{q}_1)^{1/n}\leq 2S\}.\end{aligned}$$

Now recall that $\mathfrak{q}_1$ only contains prime ideals dividing a certain non-zero $r \times r$ determinant M_0 of $M(\boldsymbol{h})$. In particular, we have $M_0 \ll H^r$. Applying Rankin's trick, we then obtain

$$\#\{\mathfrak{q}_1 : N(\mathfrak{q}_1)^{1/n} \leq 2S\} \ll S^{\varepsilon} \sum_{\mathfrak{q}_1} N(\mathfrak{q}_1)^{-\varepsilon} = S^{\varepsilon} \prod_{\mathfrak{p}|M_0} \frac{1}{1 - N(\mathfrak{p})^{-\varepsilon}} \ll S^{\varepsilon} M_0^{\varepsilon} \ll P^{\varepsilon}$$

and thus

$$V(\boldsymbol{h}, R, \eta) \ll P^{\varepsilon} \max_{S \leq R} \frac{R^n}{S^n} \min\left(\frac{1}{RP^2\eta}, \frac{S}{R}\right)^{\frac{rn}{2}}$$

Maximizing for S we find that

$$V(\boldsymbol{h}, R, \eta) \ll P^{\varepsilon} \frac{R^n}{(RP^2\eta)^{rn/2}} \min(1, P^2\eta)^{ne(r)}$$

with $e(0) = 0$, $e(1) = \frac{1}{2}$ and $e(r) = 1$ for $r \geq 2$.

Putting everything together, we obtain the estimate

$$\begin{aligned}
A(\theta, R, H, P) &\ll R^{2n} P^{\frac{ns}{2}} \sum_{|h_i| \ll H} \left[\left(\frac{1}{P} + H\eta R \right)^{\frac{nr(\boldsymbol{h})}{2}} + \frac{1}{R^n} V(\boldsymbol{h}, R, \eta) \right] \\
&\ll R^{2n} P^{\frac{ns}{2}+\varepsilon} \sum_{|h_i| \ll H} \left[\left(\frac{1}{P} + H\eta R \right)^{\frac{nr(\boldsymbol{h})}{2}} + \frac{1}{(RP^2\eta)^{\frac{r(\boldsymbol{h})n}{2}}} \min(1, P^2\eta)^{ne(r(\boldsymbol{h}))} \right] \\
&\ll R^{2n} P^{\frac{ns}{2}+\varepsilon} \sum_{r=0}^{s} H^{nr} \left[\left(\frac{1}{P} + H\eta R \right)^{\frac{nr}{2}} + \frac{1}{(RP^2\eta)^{\frac{rn}{2}}} \min(1, P^2\eta)^{ne(r)} \right] \\
&\ll \left[R^2 P^{s/2+\varepsilon} \left(1 + (RH^3\eta)^{s/2} + \frac{H^s}{P^{s/2}} + \frac{H^s}{(RP^2\eta)^{s/2}} \min(1, P^2\eta) \right) \right]^n .
\end{aligned}$$

Finally, we argue that the third term $\frac{H^s}{P^{s/2}}$ is negligible.

Indeed, if $HRP\eta \geq 1$, then it is smaller than the second term. Otherwise, if $HRP\eta \leq 1$, we have $(RP\eta)^{s/2} \leq RP\eta \leq \frac{1}{H} \leq \min(1, \eta P^2)$ on recalling that $\eta \geq \frac{1}{P^2 H}$ and hence the term $\frac{H^s}{P^{s/2}}$ is dominated by the fourth term in that case.

In any case, it now follows that

$$A(\theta, R, H, P) \ll \left[R^2 P^{s/2+\varepsilon} \left(1 + (RH^3\eta)^{s/2} + \frac{H^s}{(RP^2\eta)^{s/2}} \min(1, P^2\eta) \right) \right]^n . \quad (2.8.6)$$

2.9 The minor arcs

Finally, we synthesize the bounds obtained by Weyl and van der Corput differencing to estimate the total minor arc contribution $\int_{\mathfrak{m}} S(\alpha) d\alpha$.

We dissect $\mathfrak{m}$ with the help of the version of Dirichlet's Approximation Theorem provided by Lemma 2.3.3, applied for some parameter $1 \leq Q \leq P^{3/2}$ to be determined. Thus, every $\alpha \in K_{\mathbb{R}}$ has an approximation $\alpha = \gamma + \theta$ with $\gamma \in K$ and $N(\mathfrak{a}_\gamma) \leq Q^n$ as well as $|\theta| \ll \frac{1}{N(\mathfrak{a}_\gamma)^{1/n} Q}$.

The assumption $\alpha \in \mathfrak{m}$ then implies that $N(\mathfrak{a}_\gamma) > P^\nu$ or $|\theta| > P^{-3+\nu}$. Note that as the contribution to the minor arcs coming from $|\theta| \leq \frac{1}{P^s}$ is $O(Q^{n+1})$, we may assume that $|\theta| \geq P^{-s}$.

By a double dyadic decomposition with respect to $|\theta|$ and $N(\mathfrak{a}_\gamma)^{1/n}$, we then obtain that

$$\int_{\mathfrak{m}} S(\alpha) d\alpha \ll Q^{n+1} + P^\varepsilon \max_{R \leq Q, \phi \leq \frac{1}{RQ}} \Sigma(R, \phi)$$

where

$$\Sigma(R, \phi) := \sum_{\gamma : N(\mathfrak{a}_\gamma)^{1/n} \sim R} \int_{|\theta| \sim \phi} |S(\gamma + \theta)| \, d\theta$$

and we note that the region of integration is given by a rectangular annulus.

To establish Theorem 2.4.2, it thus suffices to prove that $\Sigma(R, \phi) \ll P^{n(s-3)-\varepsilon}$. To employ the mean-value estimates from the previous section, we use Cauchy-Schwarz to obtain

$$\Sigma(R, \phi) \ll R^n \phi^{n/2} \left(\sum_{\gamma : N(\mathfrak{a}_\gamma)^{1/n} \sim R} \int_{|\theta| \sim \phi} |S(\gamma + \theta)|^2 \, d\theta \right)^{1/2}.$$

We next cover the annulus $|\theta| \sim \phi$ with $O\left(\left(1 + \frac{\phi}{\kappa}\right)^n \right)$ boxes of size κ, all centered at values of $\alpha = \gamma + \theta$ with $|\theta| \sim \phi$, so that we obtain

$$\Sigma(R, \phi) \ll R^n \phi^{n/2} \left(1 + \frac{\phi}{\kappa}\right)^{n/2} \max_{|\theta| \sim \phi} \left(\sum_{\gamma : N(\mathfrak{a}_\gamma)^{1/n} \sim R} M(\gamma + \theta, \kappa) \right)^{1/2}$$

and using (2.8.1) and (2.8.5) we obtain

$$\Sigma(R, \phi) \ll R^n \phi^{n/2} \left(1 + \frac{\phi}{\kappa}\right)^{n/2} \max_{|\theta| \sim \phi} \left(R^{2n+\varepsilon} + \frac{\kappa^n P^{\frac{3ns}{2} - n + \varepsilon}}{H^{ns-n}} A(\theta, R, H, P) \right)^{1/2}$$

so that (2.8.6) implies that

$$\Sigma(R, \phi) \ll \left[P^\varepsilon R^2 \phi^{1/2} \left(1 + \frac{\phi}{\kappa}\right)^{1/2} \left(1 + \frac{\kappa P^{2s-1}}{H^{s-1}} E\right)^{1/2} \right]^n \tag{2.9.1}$$

where $E = 1 + (RH^3\eta)^{s/2} + \frac{H^s}{(RP^2\eta)^{s/2}} P^2 \eta$. Here we simply estimated $\min(1, P^2\eta) \leq P^2\eta$ which turns out to be sufficient.

Suppose we can show that $E \ll 1$. Recall that $\kappa \asymp \frac{(\log P)^2}{HP^2}$ so that

$$1 + \frac{\phi}{\kappa} \ll \frac{P^\varepsilon \eta}{\kappa}$$

from the definition (2.8.2) of η.

Since $\kappa \gg \frac{1}{P^s}$, both summands in the last bracket of (2.9.1) are bounded by $\frac{\kappa P^{2s-1}}{H^{s-1}}$. Still assuming $E \ll 1$, we then obtain

$$\Sigma(R, \phi) \ll \left[P^\varepsilon R^2 \phi^{1/2} \eta^{1/2} \frac{P^{s-\frac{1}{2}}}{H^{\frac{s-1}{2}}} \right]^n .$$

Recalling our desired bound $\Sigma(R, \phi) \ll P^{n(s-3)-\varepsilon}$, it now suffices to prove that

$$H^{s-1} \gg R^4 \phi \eta P^{5+\varepsilon},$$

still under the assumption $E \ll 1$. Putting $s = 14$ for convenience (as we may without loss of generality) and recalling the definition (2.8.2) of η, it suffices to have

$$H^{13} \gg R^4 \phi^2 P^{5+\varepsilon}$$

as well as

$$H^{14} \gg R^4 \phi P^{3+\varepsilon}.$$

We thus choose

$$H \asymp P^\varepsilon \max \left\{ \left(R^4 \phi^2 P^5\right)^{1/13}, \left(R^4 \phi P^3\right)^{1/14}, 1 \right\}.$$

In order for this choice to satisfy $H \leq P$, we require $R^4 \phi^2 \ll P^{8-\varepsilon}$ as well as $R^4 \phi \ll P^{11-\varepsilon}$.

Recalling $\phi \leq \frac{1}{QR} \leq \frac{1}{R^2}$, both conditions are satisfied for any $Q \leq P^{3/2}$.

We have thus found an admissible choice of H, leading to a satisfactory estimate for $\Sigma(R, \phi)$ under the assumption of $E \ll 1$.

We now enquire under which circumstances this assumption is justified.

For convenience, denote $\phi_0 = (R^4 P^{31})^{-\frac{1}{15}}$. The relevance of this parameter comes from the observation that for $\phi \leq \phi_0$, one has

$$H \asymp P^\varepsilon \max \left\{ \left(R^4 \phi P^3\right)^{1/14}, 1 \right\}$$

and $\eta \asymp \frac{1}{HP^2}$ whereas for $\phi \geq \phi_0$, one has

$$H \asymp P^\varepsilon \max \left\{ \left(R^4 \phi^2 P^5\right)^{1/13}, 1 \right\}$$

and $\eta \asymp \phi$.

To prove $E \ll 1$, we need to check that $RH^3\eta \ll P^{-\varepsilon}$ as well as $\left(\frac{H^2}{RP^2\eta}\right)^7 P^2\eta \ll P^{-\varepsilon}$.

We begin by checking that $RH^3\eta \ll P^{-\varepsilon}$. First, if $\phi \le \phi_0$, we have

$$\begin{aligned} RH^3\eta &\ll P^{\varepsilon}\frac{QH^2}{P^2} \\ &\ll P^{\varepsilon}\frac{Q}{P^2}\left(1 + (R^4\phi P^3)^{1/14}\right) \\ &\ll P^{\varepsilon}\cdot\left(\frac{Q}{P^2} + \frac{Q^{9/7}}{P^{11/7}}\right). \end{aligned}$$

This bound is satisfactory if $Q \ll P^{11/9-\varepsilon}$.

Next, if $\phi \ge \phi_0$, we have

$$\begin{aligned} RH^3\eta &\ll P^{\varepsilon}RH^3\phi \\ &\ll P^{\varepsilon}\cdot\frac{1}{Q}\cdot\left(1 + \left(R^4\phi^2P^5\right)^{3/13}\right) \\ &\ll P^{\varepsilon}\cdot\frac{P^{15/13}}{Q} \end{aligned}$$

which is satisfactory if $Q \gg P^{15/13+\varepsilon}$.

We thus choose $Q = P^{13/11}$, ensuring that $RH^3\eta \ll P^{-\varepsilon}$ in both cases, and noting that this also satisfies our earlier rough assumption $Q \le P^{3/2}$.

Finally, we need to enquire whether $\left(\frac{H^2}{RP^2\eta}\right)^7 P^2\eta \ll P^{-\varepsilon}$.

For $\phi \le \phi_0$, we have $\eta \asymp \frac{1}{HP^2}$ so that

$$\left(\frac{H^2}{RP^2\eta}\right)^7 P^2\eta \ll P^{\varepsilon}\frac{H^{20}}{R^7}$$

so that it suffices to have $H \ll R^{7/20-\varepsilon}$.

Recalling our choice of H in this case, it is thus sufficient to have $R \gg P^{\varepsilon}$ as well as additionally $\phi \le \phi_1$ where

$$\phi_1 = R^{9/10}P^{-3-\varepsilon}.$$

Similarly, if $\phi \ge \phi_0$ we have $\eta \asymp \phi$ so that

$$\left(\frac{H^2}{RP^2\eta}\right)^7 P^2\eta \ll \frac{H^{14}}{R^7P^{12}\phi^6}$$

and hence by our definition of H, it suffices to have $R \gg P^{\varepsilon}$ as well as additionally $\phi \ge \phi_2$ where

$$\phi_2 = \frac{1}{P^{\frac{43}{25}-\varepsilon}R^{7/10}}.$$

Summarizing, we have obtained a satisfactory bound for $\Sigma(R, \phi)$ if $R \gg P^{\varepsilon}$ and $\phi \leq \min(\phi_0, \phi_1)$ or $\phi \geq \max(\phi_0, \phi_2)$.

Letting $R_0 = P^{4/5+\varepsilon}$, a quick computation shows that $\phi_2 \leq \phi_0 \leq \phi_1$ if $R \geq R_0$ whereas $P^{-\varepsilon}\phi_1 \leq \phi_0 \leq \phi_2 P^{\varepsilon}$ if $R \leq R_0$.

In the first case, our argument already covers all possible values of ϕ. We are thus left with the case where $R \leq R_0$ and $P^{-\varepsilon}\phi_1 \leq \phi \leq \phi_2 P^{\varepsilon}$ or $R \leq P^{\varepsilon}$.

It is here that we require the bound obtained by Weyl differencing. Indeed, applying Lemma (2.6.1) with $s = 14$ and noting that the assumption $Q \leq P^{3/2}$ is satisfied, we obtain

$$\Sigma(R, \phi) \ll P^{\varepsilon} \left[R^2 \phi P^{14} \left(R\phi + \frac{1}{R\phi P^3} \right)^{7/4} \right]^n.$$

Recalling our goal $\Sigma(R, \phi) \ll P^{11n-\varepsilon}$, it then suffices to have

$$R^2 \phi P^3 \left(R\phi + \frac{1}{R\phi P^3} \right)^{7/4} \ll P^{-\varepsilon}.$$

But this will be satisfied if

$$\frac{R^{1/3}}{P^{3-\varepsilon}} \ll \phi \ll \frac{1}{P^{12/11+\varepsilon} R^{15/11}}. \tag{2.9.2}$$

Under the assumption $R \leq R_0$ and $P^{-\varepsilon}\phi_1 \leq \phi \leq \phi_2 P^{\varepsilon}$, this will thus be true as soon as

$$\phi_1 \gg \frac{R^{1/3+\varepsilon}}{P^3}$$

as well as

$$\phi_2 \ll \frac{1}{P^{12/11+\varepsilon} R^{15/11}}.$$

The first condition is always satisfied for $R \gg P^{\varepsilon}$ while the second one is satisfied for $R \ll P^{\frac{346}{365}-\varepsilon}$ which is indeed true under the assumption $R \leq R_0$.

Finally, we need to treat the cases where $R \leq P^{\varepsilon}$. Here of course, we need to use that we are on the minor arcs so that $\phi \geq P^{-3+\nu}$. But it is easy to see that in that case (2.9.2) is also satisfied, thus finishing our proof of Theorem 2.4.2. □

Chapter 3

Diagonal cubic forms over $\mathbb{F}_q[t]$

3.1 Introduction

Given a non-singular cubic form $F \in K[x_1, \ldots, x_n]$ with coefficients in a global field K, it is natural to study the distribution of rational points on the hypersurface $X \subset \mathbb{P}^{n-1}$ defined by F. In a quantitative sense, this entails understanding the counting function

$$N(P) = \#\{\boldsymbol{x} \in \mathcal{O}^n \colon \max_i |x_i| < |P|, F(\boldsymbol{x}) = 0\}, \tag{3.1.1}$$

where $\mathcal{O} \subset K$ is the ring of integers, $P \in \mathcal{O}$ and $|\cdot|$ is a suitable absolute value on K. For $n \geq 5$, one generally expects an asymptotic formula of the form

$$N(P) \sim c|P|^{n-3} \tag{3.1.2}$$

as $|P| \to \infty$ for some constant $c \geq 0$. For large values of n, this has been successfully achieved using the Hardy–Littlewood circle method. For $K = \mathbb{Q}$, the current state of the art is due to Hooley [52], who showed that $n \geq 9$ suffices for (3.1.2) to hold. In fact, conditional on unproved hypotheses about certain Hasse–Weil L-functions, in [57] he pushed his approach further with the outcome that $n \geq 8$ is enough. For $K = \mathbb{F}_q(t)$, using the fact that the analogous hypotheses are in fact theorems by virtue of Deligne's work [31], Browning–Vishe [17] proved unconditionally the asymptotic formula (3.1.2) for $n \geq 8$ and $\mathrm{char}(K) > 3$. However, for small values of n, an asymptotic remains largely out of reach. Assuming F to be non-singular and diagonal, which means

$$F(\boldsymbol{x}) = \sum_{i=1}^n F_i x_i^3, \quad F_i \in \mathcal{O} \setminus \{0\}, \tag{3.1.3}$$

Heath-Brown [46] provided an upper bound of the form $N(P) \ll |P|^{3+\varepsilon}$ for $n = 6$ and $K = \mathbb{Q}$, matching the predicted asymptotic up to a factor of $|P|^{\varepsilon}$. However, his work relies on deep unproven conjectures about certain Hasse–Weil L-functions.

Our first goal of this work is to prove the analogous result unconditionally for $K = \mathbb{F}_q(t)$. One of the main novelties of our work is that we also obtain results when $\mathrm{char}(K) = 2$. Usually the circle method breaks down in small characteristic due to a Weyl differencing process. We manage to bypass this issue by applying Poisson summation instead, along with a recursion argument regarding the density of solutions of the dual form F^* of F.

From now on we write $\mathcal{O} = \mathbb{F}_q[t]$ and we work with the absolute value given by $|P| = q^{\deg P}$ for $P \in \mathcal{O}$. By abuse of notation we also write $|\boldsymbol{x}| := \max_i |x_i|$ for $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{O}^n$.

Theorem 3.1.1. *Let $K = \mathbb{F}_q(t)$ with $\mathrm{char}(K) \neq 3$. Suppose F is given by (3.1.3). Then for $n = 6$ we have*

$$N(P) \ll |P|^{3+\varepsilon}.$$

In applications of the circle method one frequently uses upper bounds for the counting function

$$M(P) = \#\left\{\boldsymbol{x} \in \mathcal{O}^6 \colon x_1^3 + x_2^3 + x_3^3 = x_4^3 + x_5^3 + x_6^3 \colon |\boldsymbol{x}| < |P|\right\}$$

to estimate the contribution from the minor arcs. Until now the strongest estimate followed from Hua's lemma, which gives $M(P) \ll |P|^{7/2+\varepsilon}$. Heath-Brown's results mentioned above show $M(P) \ll_\varepsilon P^{3+\varepsilon}$, if we take $\mathcal{O} = \mathbb{Z}$. The same was established by Hooley [55] using different methods but his results are also conditional on some hypotheses regarding certain Hasse–Weil L-functions.

We now return to the case when $\mathcal{O} = \mathbb{F}_q[t]$. In a 1964 letter to Keith Matthews [25] Davenport asked whether one could achieve the bound $M(P) \ll |P|^{3+\varepsilon}$. Theorem 3.1.1 provides an affirmative answer to his question.

For $n = 4$ the situation is more complicated and one does not expect (3.1.2) to hold in general. The cubic surface $X \subset \mathbb{P}^3$ might contain rational lines and any such will contribute $\gg |P|^2$ rational points to the counting function $N(P)$. According to Manin's conjecture [37], one expects

$$N^\circ(P) \sim c|P|(\log|P|)^{\rho-1},$$

where $N^\circ(P)$ only counts rational points that do not lie on any rational line contained in X and ρ is the rank of the Picard group of X.

Over $K = \mathbb{Q}$, partial progress was made by Heath-Brown [46], who showed how to isolate the contribution to $N(P)$ coming from points on rational lines when F is diagonal. He also managed to give an upper bound of the form $N^\circ(P) \ll |P|^{3/2+\varepsilon}$, again only conditionally on certain conjectures about Hasse–Weil L-functions. As for $n = 6$, working over $K = \mathbb{F}_q(t)$ allows us to establish the estimates unconditionally and we also succeed in isolating the contribution coming from points on rational lines under certain restrictions on the characteristic of K.

Theorem 3.1.2. *Suppose F is given by (3.1.3). If $\mathrm{char}(K) > 3$, then for $n = 4$, we have*

$$N^\circ(P) \ll |P|^{3/2+\varepsilon},$$

where $N^\circ(P)$ is defined as $N(P)$ with the extra condition that $\boldsymbol{x}$ does not lie on any rational line contained in the surface $F = 0$. These lines, if they exist, are of the form

$$b_i x_i + b_j x_j = b_k x_k + b_l x_l = 0,$$

for some $b_i, b_j, b_k, b_l \in K$ such that

$$\left(\frac{b_i}{b_j}\right)^3 = \frac{F_i}{F_j}, \quad \text{and} \quad \left(\frac{b_k}{b_l}\right)^3 = \frac{F_k}{F_l},$$

where $\{i, j, k, l\} = \{1, 2, 3, 4\}$.

While if $\mathrm{char}(K) = 2$, then for $n = 4$ we have

$$N(P) \ll |P|^{2+\varepsilon}.$$

In characteristic 2 the shape of the dual form of F prevents us from isolating the contribution coming from rational points on rational lines to $N(P)$. However, we still manage to give a non-trivial upper bound for the counting function $N(P)$, thereby providing evidence that the main contribution to $N(P)$ comes from points on rational lines.

In fact, assuming certain unproved conjectures regarding the growth of the rank of rational elliptic curves, Heath-Brown [47] showed $N^\circ(P) \ll_\varepsilon P^{4/3+\varepsilon}$ for any non-singular cubic form in 4 variables defined over $\mathbb{Q}$. He further showed in [48] that certain families of cubic forms in 4 and 5 variables satisfy the Hasse principle, assuming a conjecture of Selmer on elliptic curves.

Our work also shares some similarity with the recent findings of Wang. In [113] he established an asymptotic formula for $N(P)$ for diagonal cubic forms over $\mathbb{Q}$ when $n = 6$ conditional on conjectures about mean values of ratios of L-functions and the large sieve. His approach required to isolate the contribution coming from rational points on rational linear subspaces, which he achieved in [115], similar to Heath-Brown's [46] treatment when $n = 4$. It would be interesting to see to what extent his work can be made unconditional over $\mathbb{F}_q(t)$.

So far we have ignored the constant c appearing in the asymptotic formula (3.1.2), despite its arithmetic significance. It encapsulates information about the existence of rational points on X and has received a conjectural interpretation as an adelic volume by Peyre [84]. For $n \geq 6$ it is expected to be positive as soon as $X(K_\nu) \neq \emptyset$ for all completions K_ν of K, or in other words, it reflects that X is expected to satisfy the Hasse principle. A key feature of the circle method is that when it provides an asymptotic formula, it automatically confirms the Hasse principle. So in particular, thanks to Hooley [52], we know that the Hasse principle holds for non-singular cubic forms in $n \geq 9$ variables over $\mathbb{Q}$ and the work of Browning–Vishe establishes the Hasse principle for non-singular cubic forms over $\mathbb{F}_q(t)$ in at least 8 variables.

In fact, by imposing further congruence conditions on $\boldsymbol{x}$ in the definition of $N(P)$ in (3.1.1) Browning–Vishe show that X satisfies weak approximation, which means that under the diagonal embedding

$$X(K) \longrightarrow \prod_\nu X(K_\nu)$$

the image of $X(K)$ is dense with respect to the product topology. Using Theorem 3.1.1 as a mean value estimate for the minor arc contribution, we can apply a classical version of the circle method to draw the same conclusions for diagonal cubic forms in $n \geq 7$ variables.

Theorem 3.1.3. *Let $K = \mathbb{F}_q(t)$ with $\operatorname{char}(K) > 3$ and F be a diagonal cubic form in $n \geq 7$ variables. Then the hypersurface $X \subset \mathbb{P}^{n-1}$ cut out by F satisfies the Hasse principle and weak approximation.*

One reason for being able to deal with fewer variables than Browning–Vishe is that when F is diagonal we have better control over the exponential sums involved and that we get stronger estimates for the density of solutions of bounded height of the dual form F^* of F. However, this alone along with the estimates by Browning–Vishe on averages of exponential sums would not be sufficient to prove Theorem 3.1.1–3.1.3. We additionally make use of slightly better estimates through an argument that enables us to bypass the lack of a convenient form of partial summation over K.

It should be noted that the Hasse principle over $K = \mathbb{F}_q(t)$ is an easy consequence of the Lang–Tsen theory of C_i fields for $n \geq 10$, which in fact establishes that $X(K) \neq \emptyset$ in this case. For smaller values of n, only little is known about the Hasse principle or weak approximation over $\mathbb{F}_q(t)$. Colliot-Thélène [23] has established the Hasse principle for diagonal cubic forms in $n \geq 5$ variables when $q \equiv 2 \mod 3$ and for $n = 4$ for the same range of q under some additional combinatorial constraints on the coefficients of F. Furthermore, for arbitrary non-singular cubic hypersurfaces $X \subset \mathbb{P}^{n-1}$ Tian [109] has shown that the Hasse principle holds when $\mathrm{char}(K) > 5$ and $n \geq 6$. Assuming the existence of a rational point, Tian–Zhang [110] have also verified that X satisfies weak approximation at places of good reduction whose residue fields have at least 11 elements as soon as $n \geq 4$. In fact, the results by Colliot-Thélène, Tian and Tian–Zhang were all shown to hold for any global function field K of a smooth curve over a finite field.

As a further application of Theorem 3.1.1, we are able to improve Waring's problem over $\mathbb{F}_q(t)$ for cubes. Waring's problem in degree d in this context is concerned with finding the smallest value of n such that

$$P = x_1^d + \cdots + x_n^d$$

has a solution in $\boldsymbol{x} \in \mathcal{O}^n$ for every $P \in \mathcal{O}$ with sufficiently large degree. Over $\mathbb{F}_q(t)$, in contrast to the integer setting, there might be global obstructions for P to be representable as a sum of d-th powers, for example if its leading coefficient is not a sum of n d-th powers in $\mathbb{F}_q$. Therefore, one usually restricts to $P \in \mathbb{J}_q^d[t]$, which is defined as the additive closure of d-th powers in $\mathbb{F}_q[t]$. In order to avoid cancellation in the x_i variables coming from the terms of degree larger than $\deg P$, it is more natural to consider the *strict Waring problem.* There, one is concerned with finding the minimal number $G_q(d) = n$ such that every sufficiently large polynomial $P \in \mathbb{J}_q^d[t]$ can be written as

$$P = x_1^d + \cdots + x_n^d,$$

where $\deg x_i \leq \lceil \frac{\deg P}{d} \rceil$. In order to study a more refined version of Waring's problem, we introduce the quantity $\widetilde{G}_q(d)$, which is the smallest number n such that we obtain an asymptotic formula for

$$R_n(P) = \#\{\boldsymbol{x} \in \mathcal{O}^n \colon |\boldsymbol{x}| \leq q^{\lceil \frac{\deg(P)}{d} \rceil},\, x_1^d + \cdots + x_n^d = P\},$$

for $P \in \mathbb{J}_q^d[t]$ as $\deg(P) \to \infty$. In his PhD thesis [66] Kubota tackled the asymptotic strict Waring problem over $\mathbb{F}_q(t)$ and showed $\widetilde{G}_q(d) \leq 2^d + 1$ whenever $\mathrm{char}(\mathbb{F}_q) > d$. The restriction in Kubota's work on the characteristic comes from Weyl differencing, producing a factor of $d!$ and hence rendering trivial bounds when estimating exponential sums if $\mathrm{char}(\mathbb{F}_q) \leq d$. For degrees $d \geq 4$ this was improved by Liu–Wooley [76] by replacing Weyl differencing with an application of the large sieve to also obtain results for $\mathrm{char}(\mathbb{F}_q) \leq d$.

Returning to the case of cubes, in characteristic 2 the current state of the art is due to Car–Cherly [20] who showed $\widetilde{G}_{2^h}(3) \leq 11$. They managed to avoid Weyl differencing with an application of Poisson summation along with a version of Weyl's inequality in characteristic 2 developed in [19].

Further, work by Gallardo [38] and Car–Gallardo [21] shows

$$G_q(3) \leq \begin{cases} 7, & \text{if } q \notin \{7, 13, 16\} \\ 8, & \text{if } q \in \{13, 16\} \\ 9, & \text{if } q = 7. \end{cases}$$

Rather than using a circle method approach, the last set of bounds are achieved using elementary arguments. As a result these methods do not produce an asymptotic formula, hence do not yield new bounds for $\widetilde{G}_q(3)$.

We can again use Theorem 3.1.1 as a minor arc mean value estimate in order to improve the current best known bound for $\widetilde{G}_q(3)$ for any q not divisible by 3 as well as for $G_7(3)$, $G_{13}(3)$ and $G_{16}(3)$. Our work on Waring's problem for cubes constitutes a significant improvement on the current state of the art. In particular, our result improves the previously best known upper bound of $\widetilde{G}_q(3)$ by 4 variables if q is even and by 2 variables if q is odd.

Theorem 3.1.4. *If* $\mathrm{char}(\mathbb{F}_q) \neq 3$*, then we have* $\widetilde{G}_q(3) \leq 7$ *and thus also* $G_q(3) \leq 7$.

This theorem is the function field counterpart of a result by Hooley [51], who proved the asymptotic Waring problem for cubes over integers in $n \geq 7$ variables conditional on hypotheses on certain Hasse–Weil L-functions. We also obtain a power saving error term in the asymptotic formula for $R_n(P)$. The best unconditional result in the integer setting is due to Vaughan [111], who resolved the asymptotic Waring problem for cubes in 8 variables, although he obtained only log savings in the error term. It should further be mentioned that in subsequent work, Vaughan [112] established lower bounds of the expected order of magnitude in the case when $n = 7$. Building on Vaughan's techniques, Baker [2, 3, 4] established the existence of a non-trivial zero to a diagonal cubic form in $n = 7, 8, 9$ variables, and even finds impressive upper bounds for the smallest such non-trivial solution, depending on the size of the coefficients of the form.

To deduce Theorem 3.1.4 from Theorem 3.1.1, we require a power saving when estimating a certain Weyl sum. For Waring's problem this has been carried out by Car [19], which allows us to establish Theorem 3.1.4 in characteristic 2. Although it would be possible to adapt the work of Car adequately to handle the Weyl sums appearing in the treatment of weak approximation and thus extend Theorem 3.1.3 to the case $\mathrm{char}(K) = 2$, we have decided against including such an adaption here given the length of this chapter .

While the techniques used to prove Theorems 3.1.1 – 3.1.4 are not applicable when $\mathrm{char}(K) = 3$, one can almost trivially deal with the problems directly. In fact, studying the solutions to the diagonal cubic equation (3.1.3) reduces to solving a system of linear equations. In particular, the Hasse principle and weak approximation hold trivially. Further it is easy to see that $\widetilde{G}_q(3) = 1$ holds when $\mathrm{char}(K) = 3$.

Outline

To prove Theorem 3.1.1 and Theorem 3.1.2 we employ a technique known as the *delta method* over $\mathbb{F}_q(t)$ developed by Browning–Vishe [17], but which is much simpler than the version of Heath-Brown [46] invoked over the integers. The starting point of the delta method is a smooth decomposition of the Kronecker delta function, a technique that goes back to Duke–Friedlander–Iwaniec [35]. Over $\mathbb{F}_q(t)$, indicator functions of intervals are smooth in an appropriate sense and so this decomposition is essentially rendered trivial.

In Section 3.2, we begin by reviewing some essential facts that are required to perform the analysis and arrive at an expression of the form

$$N(w, P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \sum_{\boldsymbol{c} \in \mathcal{O}^n} S_r(\boldsymbol{c}) I_r(\boldsymbol{c}),$$

for a weighted version of the main counting function, involving certain exponential sums $S_r(\boldsymbol{c})$ and oscillatory integrals $I_r(\boldsymbol{c})$.

In Sections 3.3 and 3.4, we estimate the integrals $I_r(\boldsymbol{c})$ and the exponential sums $S_r(\boldsymbol{c})$, respectively. More precisely, we obtain cancellations when averaging $S_r(\boldsymbol{c})$ over r giving essentially optimal bounds. These estimates are possible due to work by Deligne [31] and the required analysis of the relevant L-functions has been carried out in [17, Section 3]. The quality of the estimates of the exponential sums is connected to the dual form of the cubic form. This prompts us to study its rational solutions in Section 3.5.

Classically, to combine these estimates one would use partial summation, a tool that is not available in a useful form to us in the function field setting. In [17] this causes significant difficulty, and in fact the approach by Browning–Vishe comes with a slight loss in the estimates rendering them insufficient for our purposes. We can resolve this issue with Lemma 3.3.6, where we show that $I_r(\boldsymbol{c})$ only depends on the absolute value of r and so via q-adic summation we can separate the quantities without any loss.

In Section 3.6, we combine the estimates using this new approach and finish our treatment in the case $n = 6$, thereby proving Theorem 3.1.1. In the case $\mathrm{char}(K) = 2$, it turns out that the dual form F^* of F is again a non-singular cubic form. For this reason, in Section 3.6.3, we can introduce a self-improving process in the proof of Theorem 3.1.1 and the second part of Theorem 3.1.2 that turns any saving into the desired upper bound. Finally, we use Theorem 3.1.1 as a mean value estimate in an application of the classical circle method to deal with the asymptotic Waring's problem for cubes and weak approximation for diagonal cubic hypersurfaces in $n \geq 7$ variables in Section 3.7.

If $n = 4$ and $\mathrm{char}(K) > 3$ we need to deal separately with the terms coming from *special solutions* of the dual form. This is the content of Section 3.8, where we show that these terms correspond to points coming from rational lines on X.

Conventions

Given $a_1, a_2 \in \mathcal{O}$ we denote by (a_1, a_2) their highest common factor. The letter ε will always denote an arbitrarily small positive real number, whose value might change from one line to the next. All of the implied constants throughout the chapter are allowed to depend on ε, the cardinality of the constant field q and on the form F.

3.2 Function field background

In this section we collect some basic facts concerning analysis over function fields. A more detailed summary can be found in [13, Chapter 5]. Let $K = \mathbb{F}_q(t)$ with ring of integers $\mathcal{O} = \mathbb{F}_q[t]$ and $K_\infty = \mathbb{F}_q((t^{-1}))$ be the field of Laurent series in t^{-1}. For $M \in \mathbb{R}$, we shall write $\widehat{M} := q^M$. Any $\alpha \in K_\infty \setminus \{0\}$ can be written uniquely as

$$\alpha = \sum_{i \leq M} \alpha_i t^i, \quad \alpha_M \neq 0, \tag{3.2.1}$$

for some $M \in \mathbb{Z}$. If we set $|\alpha| := \widehat{M}$, then $|\cdot|$ naturally extends the absolute value induced by t^{-1} on K to K_∞. We also note that K_∞ is the completion of K with respect to this absolute value. The analogue of the unit interval in K_∞ is given by

$$\mathbb{T} := \{\alpha \in K_\infty \colon |\alpha| < 1\}.$$

In fact, K_∞ is a local field and thus can be endowed with a unique Haar measure $\mathrm{d}\alpha$ such that $\int_{\mathbb{T}} \mathrm{d}\alpha = 1$. We can extend the absolute value to K_∞^n by $|\boldsymbol{\alpha}| = \max_{i=1,\dots,n} |\alpha_i|$ and the Haar measure by $\mathrm{d}\boldsymbol{\alpha} = \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n$ for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in K_\infty^n$.

Just like over the rational numbers, Dirichlet's approximation theorem holds. That is, for any $\alpha \in \mathbb{T}$ and $Q \in \mathbb{N}$ there exist polynomials $a, r \in \mathcal{O}$ with r monic such that $(a, r) = 1$ and $|a| < |r| \leq \widehat{Q}$ satisfying

$$\left|\alpha - \frac{a}{r}\right| < \frac{1}{|r|\widehat{Q}}.$$

In fact, from the ultrametric property it follows that Dirichlet's approximation Theorem is already enough to obtain for any $Q \geq 1$ an analogue of a *Farey dissection* of the unit interval:

$$\mathbb{T} = \bigsqcup_{\substack{|r| \leq \widehat{Q} \\ r \text{ monic}}} \bigsqcup_{\substack{|a| < |r| \\ (a,r)=1}} \{\alpha \in \mathbb{T} \colon |r\alpha - a| < \widehat{Q}^{-1}\}, \tag{3.2.2}$$

where $a, r \in \mathcal{O}$.

Characters. For $\alpha \in K_\infty$ given by (3.2.1), we define

$$\psi \colon K_\infty \to \mathbb{C}^\times, \quad \psi(\alpha) = e\left(\frac{\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(\alpha_{-1})}{p}\right),$$

and set $\psi(0) = 1$, where as usual we write $e(x) = \exp(2\pi i x)$ for $x \in \mathbb{R}$. It is easy to see that ψ is a non-trivial additive character of K_∞ that satisfies for $x \in K_\infty$ and $N \in \mathbb{Z}_{\geq 0}$,

$$\int_{|\alpha| < \widehat{N}^{-1}} \psi(\alpha x) \mathrm{d}\alpha = \begin{cases} \widehat{N}^{-1} & \text{if } |x| < \widehat{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.3}$$

In particular, if $x \in \mathcal{O}$ then this implies

$$\int_{\mathbb{T}} \psi(\alpha x) \mathrm{d}\alpha = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Further, we will make frequent use of the following formulae for exponential sums. If $r, a \in \mathcal{O}$ are such that $r \neq 0$, then

$$\frac{1}{|r|} \sum_{|x| < |r|} \psi\left(\frac{ax}{r}\right) = \begin{cases} 1 & \text{if } r \mid a, \\ 0 & \text{otherwise.} \end{cases}$$

We also obtain the expected outcome for Ramanujan sums of prime powers. Let $a, \varpi \in \mathcal{O}$ be such that ϖ is prime and let $k \geq 1$ be a natural number. Then we have

$$\sideset{}{'}\sum_{|x|<|\varpi|^k} \psi\left(\frac{ax}{\varpi^k}\right) = \begin{cases} 0 & \text{if } \varpi^{k-1} \nmid a, \\ -|\varpi|^{k-1} & \text{if } \varpi^{k-1} \parallel a, \\ |\varpi|^{k-1}(|\varpi|-1) & \text{if } \varpi^k \mid a, \end{cases}$$

where the notation $\sideset{}{'}\sum_{|x|<|\varpi|^k}$ indicates that the sum runs over x which are coprime to ϖ.

Poisson Summation. We call a function $w\colon K_\infty^n \to \mathbb{C}$ *smooth* if it is locally constant. Denote by $S(K_\infty^n)$ the space of all smooth functions $w\colon K_\infty^n \to \mathbb{C}$ with compact support. If $w \in S(K_\infty^n)$ then we call w a *Schwarz-Bruhat function.* For such functions the Poisson summation formula [17, Lemma 2.1] holds.

Lemma 3.2.1. *Let $f \in K_\infty[x_1, \ldots, x_n]$ and let $w \in S(K_\infty^n)$. Then we have*

$$\sum_{\boldsymbol{z}\in\mathcal{O}^n} w(\boldsymbol{z})\psi(f(\boldsymbol{z})) = \sum_{\boldsymbol{c}\in\mathcal{O}^n} \int_{K_\infty^n} w(\boldsymbol{u})\psi(f(\boldsymbol{u}) + \boldsymbol{c}\cdot\boldsymbol{u})\mathrm{d}\boldsymbol{u}. \tag{3.2.4}$$

Delta method. Given a polynomial $F \in \mathcal{O}[x_1, \ldots, x_n]$ and $w \in S(K_\infty^n)$, we are interested in the counting function

$$N(w, P) = \sum_{\substack{\boldsymbol{x}\in\mathcal{O}^n \\ F(\boldsymbol{x})=0}} w\left(\frac{\boldsymbol{x}}{P}\right).$$

For estimating the integrals appearing in our work, it is necessary to work with such a weighted counting function, since we require ∇F to be bounded away from 0 on $\mathrm{supp}(w)$. To estimate our original counting function defined in (3.1.1), it suffices to take w to be the characteristic function of the set $\{\boldsymbol{x} \in \mathbb{T}\colon |\boldsymbol{x}| = q^{-1}\}$. Indeed, it follows that

$$N(w, P) = \#\{\boldsymbol{x} \in \mathcal{O}^n\colon F(\boldsymbol{x}) = 0, |\boldsymbol{x}| = q^{-1}|P|\},$$

so that an upper bound of the shape $N(w, P) \ll |P|^k$ yields $N(P) \ll |P|^{k+\varepsilon}$ for any $\varepsilon > 0$ by summing over q-adic ranges for $|P|$.

For a fixed parameter $Q \geq 1$ to be specified later, we deduce from (3.2.2) and (3.2.3) the identity

$$N(w, P) = \sum_{\substack{r \text{ monic} \\ |r|\leq\widehat{Q}}} \sideset{}{'}\sum_{|a|<|r|} \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} S(a/r + \theta)\mathrm{d}\theta,$$

where $\sum'_{|a|<|r|}$ means that we sum over $a \in \mathcal{O}$ with $(a, r) = 1$ only and

$$S(\alpha) = \sum_{\boldsymbol{x} \in \mathcal{O}^n} \psi(\alpha F(\boldsymbol{x})) w(\boldsymbol{x}/P)$$

for $\alpha \in \mathbb{T}$. As explained in [17, Chapter 4], since w is a Schwartz-Bruhat function we can evaluate $S(\theta + a/r)$ using Poisson summation (3.2.4) to obtain

$$N(w, P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \sum_{\boldsymbol{c} \in \mathcal{O}^n} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta, \tag{3.2.5}$$

where

$$S_r(\boldsymbol{c}) = \sum_{|a|<|r|}' \sum_{|\boldsymbol{x}|<|r|} \psi\left(\frac{aF(\boldsymbol{x}) + \boldsymbol{c} \cdot \boldsymbol{x}}{r}\right) \tag{3.2.6}$$

and

$$I_r(\theta, \boldsymbol{c}) = \int_{K_\infty^n} w(\boldsymbol{x}) \psi\left(\theta P^3 F(\boldsymbol{x}) + \frac{P\boldsymbol{c} \cdot \boldsymbol{x}}{r}\right) \mathrm{d}\boldsymbol{x}. \tag{3.2.7}$$

The expression (3.2.5) is the starting point for our work and from now on we will mostly be concerned about estimating the integrals $I_r(\theta, \boldsymbol{c})$ and the sums $S_r(\boldsymbol{c})$.

3.3 Integral estimates

As a preliminary lemma we note the following result on a linear change of variables, the proof of which is completely analogous to the proof of Lemma 7.4.2 in [62].

Lemma 3.3.1. *Let $R_1, \ldots, R_n \in \mathbb{R}$ and let $\Gamma \subset K_\infty^n$ be the region given by*

$$\Gamma = \{\boldsymbol{x} \in K_\infty^n \colon |x_i| \leq \widehat{R_i}\}.$$

Let $g \colon \Gamma \to \mathbb{C}$ be a continuous function and let $M \in \mathrm{GL}_n(K_\infty)$. Then we have

$$\int_\Gamma g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = |\det M| \int_{M\boldsymbol{\beta} \in \Gamma} g(M\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}.$$

For $f \in K_\infty[x_1, \ldots, x_n]$, we denote by H_f its height, that is, the maximum of the absolute values of its coefficients. Given $\gamma \in K_\infty$, $\boldsymbol{w} \in K_\infty^n$ and $f \in K_\infty[x_1, \ldots, x_n]$, integrals of the form

$$J_f(\gamma, \boldsymbol{w}) \coloneqq \int_{K_\infty^n} w(\boldsymbol{x}) \psi(\gamma f(\boldsymbol{x}) + \boldsymbol{w} \cdot \boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

appear quite frequently in our work. We shall now collect the required estimates for them. Upon noting that $w(\boldsymbol{x}) = \chi_{\mathbb{T}}(\boldsymbol{x}) - \chi_{t^{-1}\mathbb{T}}(\boldsymbol{x})$, the next lemma follows directly from [17, Lemma 2.4].

Lemma 3.3.2. *Let $\gamma \in K_\infty$ and $\boldsymbol{w} \in K_\infty^n$ be such that $|\boldsymbol{w}| > q$ and $|\boldsymbol{w}| \geq H_f|\gamma|$. Then $J_f(\gamma, \boldsymbol{w}) = 0$.*

The next result [17, Lemma 2.7] is the main ingredient for estimating the integrals $J_f(\gamma, \boldsymbol{w})$.

Lemma 3.3.3. *We have*

$$\int_{\mathbb{T}^n \setminus \Omega} \psi(\gamma f(\boldsymbol{x}) + \boldsymbol{w} \cdot \boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 0,$$

where $\Omega \subset \mathbb{T}^n$ is given by

$$\Omega = \left\{ \boldsymbol{x} \in \mathbb{T}^n \colon |\gamma \nabla f(\boldsymbol{x}) + \boldsymbol{w}| \leq H_f \max\left\{1, |\gamma|^{1/2}\right\}\right\}.$$

In our setting, this leads to the following estimate.

Lemma 3.3.4. *Suppose $F \in K_\infty[x_1, \ldots, x_n]$ is a non-singular cubic form. Let $\gamma \in K_\infty$ and $\boldsymbol{w} \in K_\infty^n \setminus \{\boldsymbol{0}\}$ be such that $|\boldsymbol{w}| \gg 1$. Then $J_F(\gamma, \boldsymbol{w}) = 0$, unless*

$$|\boldsymbol{w}| \ll |\gamma| \ll |\boldsymbol{w}|,$$

in which case

$$J_F(\gamma, \boldsymbol{w}) \ll \mathrm{meas}(\{\boldsymbol{x} \in \mathrm{supp}(w) \colon |\gamma \nabla F(\boldsymbol{x}) + \boldsymbol{w}| \ll |\boldsymbol{w}|^{1/2}\}).$$

Proof. First note $J_F(\gamma, \boldsymbol{w}) = 0$ if $|\boldsymbol{w}| > \max\{q, H_F|\gamma|\}$ by Lemma 3.3.2. Since by assumption $1 \ll |\boldsymbol{w}|$, we may thus assume $1 \ll |\boldsymbol{w}| \ll |\gamma|$. For $\boldsymbol{a} \in \mathbb{F}_q^n \setminus \{\boldsymbol{0}\}$, let

$$w_{\boldsymbol{a}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } |\boldsymbol{x} - \boldsymbol{a}t^{-1}| < |t|^{-1}, \\ 0 & \text{else.} \end{cases}$$

We can then write $w(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in \mathbb{F}_q^n \setminus \{\boldsymbol{0}\}} w_{\boldsymbol{a}}(\boldsymbol{x})$, so that

$$\begin{aligned} J_F(\gamma, \boldsymbol{w}) &= \sum_{\boldsymbol{a} \in \mathbb{F}_q^n \setminus \{\boldsymbol{0}\}} \int_{\mathbb{T}^n} w_{\boldsymbol{a}}(\boldsymbol{x}) \psi(\gamma F(\boldsymbol{x}) + \boldsymbol{w} \cdot \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \sum_{\boldsymbol{a} \in \mathbb{F}_q^n \setminus \{\boldsymbol{0}\}} q^{-n} \psi(t^{-1}\boldsymbol{w} \cdot \boldsymbol{a}) \int_{\mathbb{T}^n} \psi(\gamma G_{\boldsymbol{a}}(\boldsymbol{y}) + t^{-1}\boldsymbol{w} \cdot \boldsymbol{y}) \mathrm{d}\boldsymbol{y}, \end{aligned} \tag{3.3.1}$$

where we performed the change of variables $\boldsymbol{y} = t\boldsymbol{x} - \boldsymbol{a}$ and wrote $G_{\boldsymbol{a}}(\boldsymbol{y}) = F((\boldsymbol{y} + \boldsymbol{a})t^{-1})$. From Lemma 3.3.3 we deduce that each inner integral is bounded by

$$\mathrm{meas}(\{\boldsymbol{y} \in \mathbb{T}^n \colon |\gamma \nabla G_{\boldsymbol{a}}(\boldsymbol{y}) + t^{-1}\boldsymbol{w}| \ll H_{G_{\boldsymbol{a}}}|\gamma|^{1/2}\}),$$

which in turn may be bounded from above by

$$\operatorname{meas}(\{\boldsymbol{x} \in \operatorname{supp}(w_{\boldsymbol{a}}) \colon |\gamma \nabla F(\boldsymbol{x}) + \boldsymbol{w}| \ll H_F |\gamma|^{1/2}\}), \tag{3.3.2}$$

since $H_{G_{\boldsymbol{a}}} \leq H_F$. Denote the set in (3.3.2) by $\Omega_{\boldsymbol{a}}$. Note that since F is assumed to be non-singular, we have $\nabla F(\boldsymbol{x}) \neq 0$ for all $\boldsymbol{x} \in \Omega_{\boldsymbol{a}}$. Since $\operatorname{supp}(w_{\boldsymbol{a}})$ is compact for every $\boldsymbol{a}$, this implies $\nabla F(\boldsymbol{x}) \gg_w 1$ for all $\boldsymbol{x} \in \Omega_{\boldsymbol{a}}$. In particular, unless $|\boldsymbol{w}| \gg |\gamma \nabla F(\boldsymbol{x})| \gg |\gamma|$ the sets $\Omega_{\boldsymbol{a}}$ are all empty and the integral vanishes. Finally the Lemma follows upon noting

$$\operatorname{meas}(\Omega_{\boldsymbol{a}}) \ll \operatorname{meas}(\{\boldsymbol{x} \in \operatorname{supp}(w) \colon |\gamma \nabla F(\boldsymbol{x}) + \boldsymbol{w}| \ll |\boldsymbol{w}|^{1/2}\}),$$

for any $\boldsymbol{a} \in \mathbb{F}_q^n \setminus \{\boldsymbol{0}\}$ and substituting this into (3.3.1). ☐

Since we work with a diagonal cubic form $F(\boldsymbol{x}) = \sum_{i=1}^n F_i x_i^3$ with $F_i \in \mathcal{O} \setminus \{0\}$, we have $\nabla F(\boldsymbol{x}) = (3F_1 x_1^2, \ldots, 3F_n x_n^2)$. Therefore in order to find an upper bound for $J_F(\gamma, \boldsymbol{w})$ the following lemma will be useful.

Lemma 3.3.5. *Let $a, b \in K_\infty$ and consider the set*

$$P_{a,b} = \{x \in \mathbb{T} \colon |x^2 - a| < |b|\}.$$

Then we have

$$\operatorname{meas}(P_{a,b}) \ll \min\{|b|^{1/2}, |b||a|^{-1/2}\}.$$

Proof. Note first that the result is trivial if $a = 0$ or $b = 0$. Hence we may write

$$a = \sum_{i \leq K} a_i t^i, \quad \text{and} \quad b = \sum_{j \leq M} b_j t^j,$$

where $a_K, b_M \neq 0$. We will proceed in two cases.

Case 1: $|a| < |b|$. Then via the ultrametric triangle inequality we note

$$|x^2 - a| < |b| \iff |x|^2 < |b|,$$

for any $x \in \mathbb{T}$. Thus $\operatorname{meas}(P_{a,b}) \ll |b|^{1/2} = \min\{|b|^{1/2}, |b||a|^{-1/2}\}$.

Case 2: $|a| \geq |b|$. Let $x = \sum_{i \leq -1} x_i t^i \in \mathbb{T}$. Then $|x^2 - a| < |b|$ can only hold if $|x|^2 = |a|$. In particular K must be even, $K \leq -1$ must hold and $x_{K/2+1} = \cdots = x_{-1} = 0$. Write

$$x^2 = \sum_{\ell \leq K} X_\ell t^\ell,$$

where $X_\ell = \sum_{i+j=\ell} x_i x_j$. Then, requiring

$$|x^2 - a| < |b| = q^M$$

implies $X_\ell = a_\ell$ for $\ell = M, \dots, K$. Now $X_K = x_{K/2}^2$, so the condition $X_K = a_K$ yields at most two possible solutions for $x_{K/2}$. Further, since

$$X_{K-r} = 2x_{K/2}x_{K/2-r} + \sum_{\substack{i+j=K-r \\ K/2-r<i,j<K/2}} x_i x_j,$$

we find inductively that a solution to $x_{K/2}^2 = a_K$ uniquely determines $x_{K/2-r}$ for $r = 1, \dots, M+K$. To summarise, in this case, there are at most two possibilities for the values of the coefficients $x_{-1}, \dots, x_{M-K/2}$. Therefore we obtain

$$\operatorname{meas}(P_{a,b}) \ll \operatorname{meas}\left(t^{M-K/2}\mathbb{T}\right) = q^{M-K/2} = |b||a|^{-1/2}.$$

Finally, noticing that $|b||a|^{-1/2} \leq |b|^{1/2}$ if $|a| \geq |b|$ finishes the proof of this lemma. $\square$

In light of Lemma 3.3.5 we thus find

$$\operatorname{meas}(\{\boldsymbol{x} \in \operatorname{supp}(w) \colon |\gamma \nabla F(\boldsymbol{x}) + \boldsymbol{w}| \ll |\boldsymbol{w}|^{1/2}\}) \ll \prod_{i=1}^n \min\{|\boldsymbol{w}|^{-1/4}, |w_i|^{-1/2}\}$$

if F is a diagonal cubic form. Noting that the expression on the right hand side is $\gg_q 1$ if $|\boldsymbol{w}| \ll 1$ we infer from Lemma 3.3.4

$$J_F(\gamma, \boldsymbol{w}) \ll \prod_{i=1}^n \min\{|\boldsymbol{w}|^{-1/4}, |w_i|^{-1/2}\}, \tag{3.3.3}$$

for all $\gamma \in K_\infty$ and all $\boldsymbol{w} \in K_\infty^n \setminus \{\mathbf{0}\}$.

We will also have to deal with averages of $I_r(\theta, \boldsymbol{c})$ over θ, which are of the form

$$I_r(\boldsymbol{c}) \coloneqq \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta.$$

While we do not have a convenient form of partial summation available in the function field setting, the next lemma will be crucial in replacing this tool.

Lemma 3.3.6. *Assume that f is a cubic form. Let $r_1, r_2 \in \mathcal{O}$ be such that $|r_1| = |r_2|$. Then $I_{r_1}(\boldsymbol{c}) = I_{r_2}(\boldsymbol{c})$.*

Proof. Write $r = r_1$ for brevity. We shall show that $I_r(\boldsymbol{c})$ only depends on the absolute value of r. Indeed, recalling (3.2.7), for $\boldsymbol{c}$ fixed we have

$$\begin{aligned} I_r(\boldsymbol{c}) &= \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \int_{K_\infty^n} w(\boldsymbol{x})\psi\left(\theta P^3 f(\boldsymbol{x}) + \frac{P\boldsymbol{c}\cdot\boldsymbol{x}}{r}\right)\mathrm{d}\boldsymbol{x}\mathrm{d}\theta \\ &= |r|^n \int_{K_\infty^n} w(r\boldsymbol{y})\psi(P\boldsymbol{c}\cdot\boldsymbol{y})\int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \psi(\theta P^3 r^3 f(\boldsymbol{y}))\mathrm{d}\theta\mathrm{d}\boldsymbol{y}, \end{aligned} \tag{3.3.4}$$

where we used Fubini's theorem and applied the change of variables $\boldsymbol{y} = \boldsymbol{x}r^{-1}$. It follows from (3.2.3) that

$$\int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \psi(\theta P^3 r^3 f(\boldsymbol{y}))\mathrm{d}\theta = \begin{cases} (|r|\widehat{Q})^{-1} & \text{if } |P^3 f(\boldsymbol{y})| < |r|^{-2}\widehat{Q}, \\ 0 & \text{else.} \end{cases}$$

We conclude that the value of the inner integral in (3.3.4) only depends on $|r|$ for $\boldsymbol{y}$ and $\boldsymbol{c}$ fixed. The claim now follows, since w only depends on the absolute value of its argument. □

To highlight this dependence, we shall write $I_{\widehat{Y}}(\boldsymbol{c}) = I_r(\boldsymbol{c})$ if $|r| = \widehat{Y}$ from now on. In the notation above, for $r \in \mathcal{O}\setminus\{0\}$, $\boldsymbol{c} \in \mathcal{O}^n$, $\theta \in \mathbb{T}$ and $P \in \mathcal{O}$ we have

$$I_r(\theta, \boldsymbol{c}) = J_F\left(P^3\theta, \frac{P}{r}\boldsymbol{c}\right).$$

Since $I_r(\theta, \boldsymbol{c})$ vanishes unless $\frac{|P||\boldsymbol{c}|}{|r|} \ll |\theta||P|^3 \ll \frac{|P||\boldsymbol{c}|}{|r|}$, we deduce from (3.3.3) the following integral estimate.

Lemma 3.3.7. *Let $Y \geq 0$, $\boldsymbol{c} \in \mathcal{O}^n \setminus \{\boldsymbol{0}\}$, and $P \in \mathcal{O}$. Then*

$$I_{\widehat{Y}}(\boldsymbol{c}) \ll \min\left\{\frac{|\boldsymbol{c}|}{\widehat{Y}|P|^2}, \widehat{Y}^{-1}\widehat{Q}^{-1}\right\}\prod_{i=1}^n \min\left\{\left(\frac{|P||\boldsymbol{c}|}{\widehat{Y}}\right)^{-1/4}, \left(\frac{|P||c_i|}{\widehat{Y}}\right)^{-1/2}\right\}.$$

So far we have not yet achieved any non-trivial estimates for $I_{\widehat{Y}}(\boldsymbol{0})$ and in fact we will have to do slightly better than the trivial bound for our treatment.

Lemma 3.3.8. *Assume $n \geq 4$. Let $P \in \mathcal{O}\setminus\{\boldsymbol{0}\}$. Then for any $Y \geq 1$ we have*

$$I_{\widehat{Y}}(\boldsymbol{0}) \ll |P|^{-3+\varepsilon}.$$

Proof. For $r \in \mathcal{O}\setminus\{0\}$ such that $|r| = \widehat{Y}$, Lemma 3.3.3 gives

$$\tilde{I}_r(\theta, \boldsymbol{0}) := \int_{\mathbb{T}^n} \psi\left(\theta P^3 F(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x} \ll \mathrm{meas}(\{\boldsymbol{x} \in \mathbb{T}^n \colon |\nabla F(\boldsymbol{x})| \leq \max\{1, |\theta||P|^3\}^{-1/2}\}).$$

Now it is not hard to see that $I_r(\theta, \mathbf{0}) = \tilde{I}_r(\theta, \mathbf{0}) - q^{-n}\tilde{I}_r(q^{-3}\theta, \mathbf{0})$. From Lemma 3.3.4 we deduce

$$I_r(\theta, \mathbf{0}) \ll \operatorname{meas}(\{\boldsymbol{x} \in \mathbb{T}^n : |\nabla F(\boldsymbol{x})| \ll \max\{1, |\theta||P|^3\}^{-1/2}\}).$$

Since F is diagonal we have $|\nabla F(\boldsymbol{x})| \geq |\boldsymbol{x}|^2$ whence

$$I_r(\theta, \mathbf{0}) \ll \max\{1, |\theta||P|^3\}^{-n/4}.$$

By definition of $I_{\widehat{Y}}(\mathbf{0})$ we may divide the area of integration up as follows

$$I_{\widehat{Y}}(\mathbf{0}) = \int_{|\theta| \ll |P|^{-3}} I_r(\theta, \mathbf{0}) \mathrm{d}\theta + \int_{|P|^{-3} \ll |\theta| < \widehat{Q}^{-1}\widehat{Y}^{-1}} I_r(\theta, \mathbf{0}) \mathrm{d}\theta.$$

The first term is trivially $O(|P|^{-3})$. For the second term note

$$\int_{|P|^{-3} \ll |\theta| < \widehat{Q}^{-1}\widehat{Y}^{-1}} I_r(\theta, \mathbf{0}) \mathrm{d}\theta \ll \int_{|P|^{-3} \ll |\theta| < \widehat{Q}^{-1}\widehat{Y}^{-1}} |P|^{-3n/4} |\theta|^{-n/4} \mathrm{d}\theta \ll |P|^{-3+\varepsilon}.$$

The result now follows. □

3.4 Exponential sum estimates

We want to estimate the sum

$$\begin{aligned} S_r(\boldsymbol{c}) &= \sideset{}{'}\sum_{|a|<|r|} \sum_{|\boldsymbol{x}|<|r|} \psi\left(\frac{aF(\boldsymbol{x}) + \boldsymbol{c} \cdot \boldsymbol{x}}{r}\right) \\ &= \sideset{}{'}\sum_{|a|<|r|} \prod_{i=1}^{n} \sum_{|x|<|r|} \psi\left(\frac{aF_i x^3 + c_i x}{r}\right), \end{aligned} \tag{3.4.1}$$

where $F(\boldsymbol{x}) = \sum_{i=1}^n F_i x_i^3$. The corresponding sum over the integers has already been subject to thorough investigation by Heath-Brown [46] and Hooley [51]. Browning–Vishe [17] have translated many of the properties to the function field setting, some of which we shall record here.

The quality of our estimates is intimately connected to the dual form F^* of F, which is an absolutely irreducible polynomial of degree $2^{n-2} \times 3$ whose zero locus parameterises hyperplanes that have a singular intersection with the projective hypersurface cut out by F. As explained by Wang [114, Appendix D], if F is diagonal and $\operatorname{char}(K) > 3$, we can take

$$F^*(\boldsymbol{c}) = \left(\prod_{i=1}^{n} F_i\right)^{2^{n-2}} \prod \left((F_1^{-1}c_1^3)^{1/2} \pm \cdots \pm (F_n^{-1}c_n^3)^{1/2}\right), \tag{3.4.2}$$

where the inner product runs through all possible combinations of $\pm$. In fact, in [114] this is only shown for $K = \mathbb{Q}$, but one can check that the requirement $\mathrm{char}(K) > 3$ is sufficient for (3.4.2) to hold. In characteristic 2, we have the following result.

Lemma 3.4.1. *Let K be a field of characteristic 2 and let $F(\boldsymbol{x}) = \sum_{i=1}^n F_i x_i^3 \in K[x_1, \ldots, x_n]$ be a non-singular cubic form. Then the dual form of F is given by*

$$F^*(\boldsymbol{c}) = \left(\prod_{i=1}^n F_i\right) \sum_{i=1}^n F_i^{-1} c_i^3.$$

Proof. By definition the zero locus $V(F^*) \subset \mathbb{P}^{n-1}$ parameterises points $\boldsymbol{c} \in \mathbb{P}^{n-1}$ such that the hyperplane $\boldsymbol{c} \cdot \boldsymbol{x} = 0$ has a singular intersection with $V(F^*)$. This means, that there exists $\boldsymbol{x} \in \mathbb{P}^{n-1}(\overline{K})$ such that

$$\mathrm{rank} \begin{pmatrix} \nabla F(\boldsymbol{x}) \\ \boldsymbol{c} \end{pmatrix} = 1, \quad \boldsymbol{c} \cdot \boldsymbol{x} = 0 \quad \text{and} \quad F(\boldsymbol{x}) = 0. \tag{3.4.3}$$

Since we assume F to be non-singular, the rank condition implies that $\boldsymbol{c}$ is proportional to $\nabla F(\boldsymbol{x})$, that is, $x_i^2 = \lambda F_i^{-1} c_i$ for some $\lambda \in \overline{K}^\times$ and $i = 1, \ldots, n$. Any pair $(\boldsymbol{x}, \boldsymbol{c})$ having this property then satisfies $F(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{c} \cdot \boldsymbol{x} = 0$. Moreover, the third condition in (3.4.3) is equivalent to

$$\sum_{i=1}^n F_i^{-1/2} c_i^{3/2} = 0,$$

where we used that every element of $\overline{K}$ has a unique square-root as $\mathrm{char}(K) = 2$. However, again since we are in characteristic 2, this is is equivalent to

$$\sum_{i=1}^n F_i^{-1} c_i^3 = 0.$$

The result now follows after clearing denominators. $\square$

Note that if $r_1, r_2 \in \mathcal{O}$ are coprime, then

$$S_{r_1 r_2}(\boldsymbol{c}) = S_{r_1}(\boldsymbol{c}) S_{r_2}(\boldsymbol{c}), \tag{3.4.4}$$

which follows readily from the Chinese remainder theorem. This essentially reduces the task of estimating $S_r(\boldsymbol{c})$ to prime power moduli. Indeed, suppose $S_{\varpi^k}(\boldsymbol{c}) \leq C|\varpi|^{k\alpha}$ for some $\alpha > 0$ and some absolute constant C. Let $\Omega(r)$ be the number of prime divisors of r. Then by multiplicativity of $S_r(\boldsymbol{c})$ we have

$$S_r(\boldsymbol{c}) = \prod_{\varpi^k \| r} S_{\varpi^k}(\boldsymbol{c}) \leq \prod_{\varpi^k \| r} C|\varpi|^{k\alpha} = C^{\Omega(r)} |r|^\alpha \ll \tau(r)|r|^\alpha \ll |r|^{\alpha+\varepsilon}$$

by the usual estimate for the divisor function $\tau(r)$, see [13, Lemma 5.9].

Further, if ϖ is irreducible such that $\varpi \nmid F^*(\boldsymbol{c})$, then Browning–Vishe [17, Section 5] show

$$S_{\varpi^k}(\boldsymbol{c}) = 0 \quad \text{for } k \geq 2. \tag{3.4.5}$$

3.4.1 Square-free moduli contribution

Deligne's resolution of the Weil conjectures [30] shows that we get square-root cancellation for the sums $S_\varpi(\boldsymbol{c})$ whenever ϖ is suitably generic:

$$S_\varpi(\boldsymbol{c}) \ll |\varpi|^{(n+1)/2} |(\varpi, \nabla F^*(\boldsymbol{c}))|^{1/2}. \tag{3.4.6}$$

However, this is not sufficient for our purposes. In the integer setting Hooley [51] was the first to achieve an extra saving when averaging the sums $S_r(\boldsymbol{c})$ over r by appealing to certain hypotheses about Hasse–Weil L-functions associated to cubic threefolds. By virtue of Deligne's proof of the Weil conjectures [31] these hypotheses are in fact theorems in the function field setting. This enabled Browning–Vishe [17, Lemma 8.5] to establish the following result unconditionally.

Lemma 3.4.2. *Suppose n is even and $F^*(\boldsymbol{c}) \neq 0$. Then for any $Z \geq 0$ and $\varepsilon > 0$, we have*

$$\sum_{\substack{|r| \leq \widehat{Z} \\ (r, \Delta_F F^*(\boldsymbol{c}))=1}} \frac{S_r(\boldsymbol{c})}{|r|^{(n+1)/2}} \ll |\boldsymbol{c}|^\varepsilon \widehat{Z}^{1/2+\varepsilon},$$

where Δ_F is the discriminant of F and by virtue of (3.4.5) *r ranges over square-free values only.*

Remark 3.4.3. In fact Browning–Vishe have to consider averages of $S_r(\boldsymbol{c})$ twisted by a Dirichlet character of K_∞ since they were unable to separate the integral $I_r(\theta, \boldsymbol{c})$ from summation. However, we can resolve this issue with Lemma 3.3.6 allowing us to combine Lemma 3.4.2 with the integral bounds from Lemma 3.3.7 more efficiently.

3.4.2 Pointwise estimates

For $B \in \mathcal{O}$ fixed and $a, r \in \mathcal{O} \setminus \{0\}$ with $(a, r) = 1$, let

$$S_r(a, c) = \sum_{|x| < |r|} \psi\left(\frac{aBx^3 + cx}{r}\right).$$

In view of (3.4.1) upper bounds for $S_r(a, c)$ directly transform into estimates for $S_r(\boldsymbol{c})$. Moreover, by (3.4.4) it suffices to consider the case $r = \varpi^k$, where ϖ is irreducible.

Hooley [51] has provided upper bounds for the integer-analogue of the sum $S_{\varpi^k}(a,c)$ whenever $\varpi \nmid B$. As explained by Heath-Brown [46], these estimates also hold if $\varpi \mid B$ when we allow the implied constant to depend on B. Hooley's and Heath-Brown's proofs of these results go through almost verbatim in the function field setting and so we spare the reader from the tedious exercise of reproducing them here. To state the final outcome, we need some notation. First, we set $\{\varpi^k, c\} = (\varpi^k, c)$ for $k = 2$ and for $k \geq 3$, we define $\{\varpi^k, c\} = |\varpi|^{-1}$ if $\varpi \parallel c$ and $\{\varpi^k, c\} = (\varpi^k, c)$ else. For later use, we generalise this to square-full r by setting

$$\{r, c\} \coloneqq \prod_{\varpi^k \| r} \{\varpi^k, c\}.$$

We then have

$$S_{\varpi^k}(a,c) \ll |\varpi|^{k/2} |\{\varpi^k, c\}|^{1/4} \quad \text{for } k \geq 2. \tag{3.4.7}$$

We shall also use an estimate of Hua [61, Lemma 1.1], whose proof, again, readily translates to the function field setting. If $g(x) = \sum_{i=0}^d g_i x^i \in \mathcal{O}[x]$, then for any $\varpi \in \mathcal{O}$ irreducible we have

$$\sum_{|x|<|\varpi|^k} \psi\left(\frac{g(x)}{\varpi^k}\right) \ll |\varpi|^{k(1-1/d)} |(\varpi^k, g_0, \ldots, g_d)|^{1/d}, \tag{3.4.8}$$

where the constant depends only on ε and d. Originally this was stated in the case when $\varpi \nmid (g_0, \ldots, g_d)$, but the factor $|(\varpi^k, g_0, \ldots, g_d)|^{1/d}$ in the estimate accounts for the possibility of $\varpi \mid (g_0, \ldots, g_d)$. Therefore we obtain

$$S_{\varpi^k}(a,c) \ll |\varpi|^{2k/3},$$

where the implied constant depends on ε but crucially not on a since we assumed $\varpi \nmid a$. Using (3.4.1), we can immediately deduce the following lemma from (3.4.7) and (3.4.8), which is the analogue of [46, Lemma 5.1.].

Lemma 3.4.4. *It holds that*

$$S_{\varpi^2}(\boldsymbol{c}) \ll |\varpi|^{2+n}.$$

In addition, if $(\varpi^k, \boldsymbol{c}) = H_\varpi$ and there are at least m indices i such that $(\varpi^k, c_i) = H_\varpi$, then

$$S_{\varpi^k}(\boldsymbol{c}) \ll |\varpi|^{k+2(n-m)k/3+mk/2} |H_\varpi|^{m/4}.$$

3.4.3 Averages over square-full moduli

Suppose we are given a set of t indices $\mathcal{T} \subset \{1,\dots,n\}$ and positive integers C_i for $i \in \mathcal{T}$. For $\boldsymbol{C} := (C_i)_{i\in\mathcal{T}}$ we define $\mathcal{R}(\boldsymbol{C}) \subset \mathcal{O}^n$ to be the set of tuples $\boldsymbol{c} = (c_1,\dots,c_n)$ such that $|c_i| = \widehat{C}_i$ if $i \in \mathcal{T}$ and $c_j = 0$ whenever $j \notin \mathcal{T}$. Given $Y \in \mathbb{Z}_{>0}$, we are interested in averages of the form

$$\mathcal{A}(\mathcal{R}(\boldsymbol{C}), \widehat{Y}) := \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C}) \\ F^*(\boldsymbol{c})\neq 0}} \sum_{\substack{r\in\mathcal{O} \\ |r|=\widehat{Y}}} |S_r(\boldsymbol{c})|,$$

where r is restricted to square-full polynomials.

Lemma 3.4.5. *With the notation from above, we have*

$$\mathcal{A}(\mathcal{R}(\boldsymbol{C}), \widehat{Y}) \ll_\varepsilon \widehat{Y}^{1+n/2+(n-t)/6}(\widehat{Y}\widehat{C})^\varepsilon \#\mathcal{R}(\boldsymbol{C}),$$

where $\widehat{C} = \max_{i\in\mathcal{T}} \widehat{C}_i$.

The proof of Lemma 3.4.5 is along the same lines as that of [46, Lemma 5.2], and so we shall be brief.

Proof. First of all, we introduce some notation. Fix $\boldsymbol{c} \in \mathcal{R}(\boldsymbol{C})$. For $r \in \mathcal{O}$ monic square-full, we write

$$r = r_* \prod_{i\in\mathcal{T}} r_i, \tag{3.4.9}$$

where the various coprime factors r_*, r_i are defined as follows. We let r_* be the product of those monic prime powers ϖ^k such that $\varpi^k \parallel r$ and $k = 2$ or $\varpi \nmid c_i$ for $i \in \mathcal{T}$. Moreover, for $i \in \mathcal{T}$, we define r_i to be the product of monic prime powers $\varpi^k \parallel r$ such that $\varpi \mid c_i$, but $\varpi \nmid c_j$ for any $j \in \mathcal{T}$ with $j < i$. In particular, any r_i is cube-full. Since all the factors in (3.4.9) are coprime, it follows from (3.4.4) that

$$S_r(\boldsymbol{c}) = S_{r_*}(\boldsymbol{c}) \prod_{i\in\mathcal{T}} S_{r_i}(\boldsymbol{c}).$$

Using the fact that $S_{\varpi^k}(\boldsymbol{c}) = 0$ if $\varpi \nmid F^*(\boldsymbol{c})$ for $k \geq 2$ and the estimates (3.4.7) and (3.4.8), we deduce that

$$S_r(\boldsymbol{c}) \ll \eta(r, \boldsymbol{c})|r|^{1+n/2+(n-t)/6+\varepsilon} \prod_{i,j\in\mathcal{T}} |\{r_i, c_j\}|^{1/4},$$

where $\eta(r, \boldsymbol{c}) = 1$ if $\varpi \mid F^*(\boldsymbol{c})$ for all primes $\varpi \mid r_*$ and $\eta(r, \boldsymbol{c}) = 0$ else. Let us now fix the absolute values of r_* and of the various r_i's, say $|r_*| = \widehat{Y}_*$ and $|r_i| = \widehat{Y}_i$, and denote their contribution to $\mathcal{A}(\mathcal{R}(\boldsymbol{C}), \widehat{Y})$ by $\mathcal{A}(Y_*, \boldsymbol{Y})$, where $\boldsymbol{Y} = (Y_i)_{i\in\mathcal{T}}$. We then have

$$\mathcal{A}(Y_*, \boldsymbol{Y}) \ll \widehat{Y}^{1+n/2+(n-t)/6+\varepsilon} \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C}) \\ F^*(\boldsymbol{c})\neq 0}} \sum_{\substack{|r_i|=\widehat{Y}_i \\ i\in\mathcal{T}}} \prod_{i,j\in\mathcal{T}} |\{r_i, c_j\}|^{1/4} S_{\boldsymbol{c}},$$

where we have suppressed the dependence of r_* and of the r_i's on $\boldsymbol{c}$ in the notation and where

$$S_{\boldsymbol{c}} = \sum_{|r_*|=\widehat{Y}_*} \eta(r, \boldsymbol{c}).$$

Heath-Brown's argument for estimating $S_{\boldsymbol{c}}$ goes through almost verbatim in our setting and gives $S_{\boldsymbol{c}} \ll (\widehat{Y}\widehat{C})^{\varepsilon}$. Therefore, we have

$$\mathcal{A}(Y_*, \boldsymbol{Y}) \ll \widehat{Y}^{1+n/2+(n-t)/6+\varepsilon} (\widehat{Y}\widehat{C})^{\varepsilon} \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C}) \\ F^*(\boldsymbol{c})\neq 0}} \sum_{\substack{|r_i|=\widehat{Y}_i \\ i\in\mathcal{T}}} \prod_{i,j\in\mathcal{T}} |\{r_i, c_j\}|^{1/4}.$$

To achieve the desired upper bound, we shall now only require that each r_i is cube-full and that $\varpi \mid c_i$ whenever $\varpi \mid r_i$, so that in particular the r_i's do not depend on $\boldsymbol{c}$ anymore. Thus, after setting

$$S(j) = \sum_{|c_j|=\widehat{C}_j} \prod_{i\in\mathcal{T}} |\{r_i, c_j\}|^{1/4},$$

we obtain

$$\mathcal{A}(Y_*, \boldsymbol{Y}) \ll \widehat{Y}^{1+n/2+(n-t)/6+\varepsilon} (\widehat{Y}\widehat{C})^{\varepsilon} \sum_{\substack{|r_i|=\widehat{C}_i \\ i\in\mathcal{T}}} \prod_{j\in\mathcal{T}} S(j).$$

It is again straightforward to verify that Heath-Brown's argument continues to hold in our setting, yielding

$$\sum_{\substack{|r_i|=\widehat{C}_i \\ i\in\mathcal{T}}} \prod_{j\in\mathcal{T}} S(j) \ll \widehat{Y}^{(n+1)\varepsilon} \#\mathcal{R}(\boldsymbol{C}).$$

With a new choice of ε, we conclude

$$\mathcal{A}(Y_*, \boldsymbol{Y}) \ll \widehat{Y}^{1+n/2+(n-t)/6} (\widehat{Y}\widehat{C})^{\varepsilon} \#\mathcal{R}(\boldsymbol{C}),$$

so that the statement of the lemma follows from the fact that there are only $\widehat{Y}^{\varepsilon}$ possibilities for admissible tuples $(Y_*, \boldsymbol{Y})$. □

3.5 Rational points on the dual hypersurface

In this section we study roots of the dual form F^* of F that was defined in (3.4.2). Our first goal is to find an upper bound for the number of solutions $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| \leq \widehat{C}$ when $\mathrm{char}(K) > 3$. In order to achieve this we closely follow the strategy of Heath-Brown [46, Section 7]. The result of Lemma 3.5.2 is standard over the rational numbers, however we could not find a proof in the literature for our setting and so we included a proof here.

If $n = 4$ and $\mathrm{char}(K) > 3$ we call a solution $\boldsymbol{c}$ to $F^*(\boldsymbol{c}) = 0$ *special* if $c_1, \ldots, c_4 \neq 0$ and there are indices i, j, k, l such that $\{i, j, k, l\} = \{1, 2, 3, 4\}$ and

$$(F_i^{-1}c_i^3)^{1/2} + (F_j^{-1}c_j^3)^{1/2} = (F_k^{-1}c_k^3)^{1/2} + (F_l^{-1}c_l^3)^{1/2} = 0$$

holds for a suitable choice of square roots. We call a solution $\boldsymbol{c}$ to $F^*(\boldsymbol{c}) = 0$ *ordinary* if it is not special. In particular, if $\mathrm{char}(K) = 2$ every solution is ordinary.

Lemma 3.5.1. *Assume* $\mathrm{char}(K) > 3$. *If $n = 6$, then the number of solutions to $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| \leq \widehat{C}$ is bounded by $O(\widehat{C}^{3+\varepsilon})$. Moreover, if $n = 4$, then the number of ordinary solutions to $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| \leq \widehat{C}$ is bounded by $O(\widehat{C}^{1+\varepsilon})$.*

Before we can begin with the proof of this lemma, we need an auxiliary result. In the following we fix $\zeta \in \mathbb{F}_q^\times$ to be a representative of a non-trivial element in $\mathbb{F}_q^\times/\mathbb{F}_q^{\times,2}$. If $\mathrm{char}(\mathbb{F}_q) > 2$ this certainly exists — we may for example pick ζ to be a primitive root of $\mathbb{F}_q^\times$.

Lemma 3.5.2. *Suppose* $\mathrm{char}(K) > 3$. *Let $m_1, \ldots, m_n \in \mathcal{O}$ be a collection of distinct square-free polynomials such that each m_i is either monic or has leading coefficient ζ. Then $\{\sqrt{m_1}, \ldots, \sqrt{m_n}\}$ is a linearly independent set over K.*

Proof. We will prove the result by induction on n. The cases $1 \leq n \leq 3$ can easily be verified directly, so suppose $n \geq 4$. Assume for a contradiction that $\lambda_1, \ldots, \lambda_n \in K$ not all zero are such that

$$\sum_{k=1}^{n} \lambda_k \sqrt{m_k} = 0.$$

Note that we may assume $\lambda_i \neq 0$ for all $i = 1, \ldots, n$ since otherwise the result would follow immediately from the induction hypothesis. In particular it is sufficient to show that there exists some index k with $\lambda_k = 0$. Since $n \geq 3$ there exist two distinct indices i, j such that $m_i/m_j \notin \mathbb{F}_q^\times$. From the $n = 3$ case it follows that $K_{i,j} \coloneqq K(\sqrt{m_i}, \sqrt{m_j})$ is a Galois extension of degree 2 or 4 over K. In either case there exists $\sigma \in \mathrm{Gal}(K_{i,j}/K)$ such that $\sigma(\sqrt{m_i}) = -\sqrt{m_i}$ and $\sigma(\sqrt{m_j}) = \sqrt{m_j}$. We may lift this to an element $\tilde{\sigma} \in \mathrm{Gal}(K^s/K)$ where K^s is the separable closure of K. Then we find

$$0 = \tilde{\sigma}\left(\sum_{k=1}^{n} \lambda_k \sqrt{m_k}\right) + \sum_{k=1}^{n} \lambda_k \sqrt{m_k} = 2\lambda_j \sqrt{m_j} + \sum_{k \neq i,j} \widetilde{\lambda}_k \sqrt{m_k},$$

where $\widetilde{\lambda}_k \in \{0, 2\lambda_k\}$. From the induction hypothesis we get $\lambda_j = 0$, which yields the desired result as remarked above. □

Proof of Lemma 3.5.1. First note that $F^*(\boldsymbol{c}) = 0$ if and only if

$$(F_1^{-1} c_1^3)^{1/2} + \cdots + (F_n^{-1} c_n^3)^{1/2} = 0, \tag{3.5.1}$$

for a suitable choice of square roots. Let $m_k \in \mathcal{O}$ be a square-free polynomial, which is either monic or has leading coefficient ζ. Say $i \in \mathcal{I}(k)$ if there exists some $d_i \in \mathcal{O}$ such that $F_i c_i^3 = m_k d_i^2$. From Lemma 3.5.2 we find that (3.5.1) implies

$$\sum_{i \in \mathcal{I}(k)} F_i^{-1} d_i = 0.$$

We have $c_i^2 \mid m_k d_i^2$ and consequently $c_i \mid d_i$ since m_k is square-free. Thus there exists $e_i \in \mathcal{O}$ such that $d_i = c_i e_i$. Substituting this into the relation $F_i c_i^3 = m_k d_i^2$ we find $c_i = m_k F_i^{-1} e_i^2$ and hence $d_i = c_i e_i = m_k F_i^{-1} e_i^3$. Therefore $F_i^{-1} d_i = m_k F_i \left(\frac{e_i}{F_i}\right)^3$ and the preceding display gives

$$\sum_{i \in \mathcal{I}(k)} F_i \left(\frac{e_i}{F_i}\right)^3 = 0. \tag{3.5.2}$$

We will now estimate the number of solutions $\boldsymbol{e}$ to (3.5.2) such that $|\boldsymbol{e}| \leq \widehat{E} = \sqrt{\widehat{C}/|m_k|}$. This will then enable us to estimate the number of solutions of (3.5.1). Note that if $\#\mathcal{I}(k) = 1$ then the only solution is given by $e_i = 0$. Using this, Hölder's inequality and Hua's Lemma in this context (cf. [13, Lemma 5.12]) we find

$$\#\left\{ |\boldsymbol{e}| \leq \widehat{E} \colon \sum_{i \in \mathcal{I}(k)} F_i \left(\frac{e_i}{F_i}\right)^3 = 0 \right\} \ll \begin{cases} 1 & \text{if } \#\mathcal{I}(k) = 1, \\ \widehat{E}^{2+\varepsilon} & \text{if } 2 \leq \#\mathcal{I}(k) \leq 4, \\ \widehat{E}^{\#\mathcal{I}(k) - 2 + \varepsilon} & \text{if } 5 \leq \#\mathcal{I}(k) \leq 6. \end{cases}$$

Note that at this point it is crucial to assume $\mathrm{char}(K) > 3$, because the Weyl differencing argument in the proof of Hua's lemma breaks down otherwise. Therefore for a fixed partition $\bigsqcup_j \mathcal{I}(k_j) = \{1, \ldots, n\}$ corresponding to $\{m_{k_j}\}$ the number of $|\boldsymbol{c}| \leq \widehat{C}$ satisfying (3.5.1) is bounded above by

$$\prod_j \left(\frac{\widehat{C}}{|m_{k_j}|} \right)^{e_{k_j}/2+\varepsilon},$$

where

$$e_{k_j} = \begin{cases} 0, & \text{if } \#\mathcal{I}(k_j) = 1 \\ 2, & \text{if } 2 \leq \#\mathcal{I}(k_j) \leq 4 \\ 3, & \text{if } \#\mathcal{I}(k_j) = 5 \\ 4, & \text{if } \#\mathcal{I}(k_j) = 6. \end{cases}$$

By considering all possible square-free elements $|m_{k_j}| \ll \widehat{C}$, we see that the total number of solutions of (3.5.1) corresponding to a fixed partition is bounded above by

$$\sum_{|m_{k_j}| \leq \widehat{C}} \prod_j \left(\frac{\widehat{C}}{|m_{k_j}|} \right)^{e_{k_j}/2+\varepsilon} \ll \prod_j \widehat{C}^{e_{k_j}/2+\varepsilon},$$

where we note that $m_{k_j} = 0$ is the only permissible value in the sum above if $e_{k_j} = 0$. It is easily checked that for any possible partition this is bounded above by $O(\widehat{C}^{3+\varepsilon})$ if $n = 6$. Therefore the total number of solutions to $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| \leq \widehat{C}$ has the same upper bound. In the case $n = 4$ one can similarly obtain $O(\widehat{C}^{1+\varepsilon})$ for the number of solutions corresponding to any partition, except in the case where $\#\mathcal{I}(k_1) = \#\mathcal{I}(k_2) = 2$. But solutions arising from such partitions are precisely the special solutions. This finishes the proof of the lemma. □

3.6 Circle method

As explained in the introduction, we are considering a diagonal cubic form $F \in \mathcal{O}[x_1, \ldots, x_n]$ of the shape

$$F(\boldsymbol{x}) = \sum_{i=1}^{n} F_i x_i^3, \quad F_i \in \mathcal{O} \setminus \{0\}.$$

Recall from (3.2.5) that the associated counting function can be written as

$$N(w, P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \int_{|\theta| < |r|^{-1}\widehat{Q}^{-1}} \sum_{\boldsymbol{c} \in \mathcal{O}^n} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta.$$

Throughout the parameter Q is chosen in such a way that

$$|P|^{3/2} \leq \widehat{Q} \leq q|P|^{3/2} \tag{3.6.1}$$

ensuring that the measure of the set $\{|\theta| < |r|^{-1}\widehat{Q}^{-1}\}$ is $O(|P|^{-3})$ when $|r| = \widehat{Q}$. It follows from Lemma 3.3.2 that $I_r(\theta, \boldsymbol{c})$ vanishes unless $|\boldsymbol{c}| < |r||P|^{-1}\max\{q, H_F|P|^3\theta\}$. Since $H_F|P|^3|\theta| \leq H_F|P|^3\widehat{Q}^{-1}|r|^{-1}$ and $|P|^3\widehat{Q}^{-1}|r|^{-1} \gg 1$, we can truncate the sum over $\boldsymbol{c}$ in (3.2.5) at $|\boldsymbol{c}| \ll \widehat{C}$, where $\widehat{C} := |P|^2\widehat{Q}^{-1}$.

We now split up $N(w, P)$ according to the quality of our available estimates into

$$N(w, P) = N_0(P) + E_1(P) + E_2(P),$$

where

$$N_0(P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} S_r(\mathbf{0}) I_r(\theta, \mathbf{0}) \mathrm{d}\theta, \tag{3.6.2}$$

$$E_1(P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \sum_{\substack{\boldsymbol{c} \in \mathcal{O}^n \\ F^*(\boldsymbol{c}) \neq 0}} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta, \tag{3.6.3}$$

$$E_2(P) = |P|^n \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-n} \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \sum_{\substack{\boldsymbol{c} \in \mathcal{O}^n \setminus \{\mathbf{0}\} \\ F^*(\boldsymbol{c}) = 0}} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta. \tag{3.6.4}$$

For $n = 4$ we will later divide the term $E_2(P)$ into special and ordinary solutions of $F^*(\boldsymbol{c}) = 0$ as defined in Section 3.5. Usually one expects that the main term in an asymptotic formula for $N(w, P)$ should come from $N_0(P)$. As we are only interested in an upper bound for $N(w, P)$, the contribution from $N_0(P)$ will be rather straightforward to deal with. Handling the terms $E_1(P)$, $E_2(P)$ turns out to be a more challenging task and will occupy most of the remainder of our work. For $E_1(P)$ we can make use of the full power of our exponential sum estimates, in particular we gain an extra saving when averaging $S_r(\boldsymbol{c})$ over r. This is not possible for $E_2(P)$, but we shall benefit from the sparsity of $\boldsymbol{c}$'s such that $F^*(\boldsymbol{c}) = 0$, at least for ordinary solutions when $n = 4$.

3.6.1 Contribution from $N_0(P)$

For this we write again $r = r_1 r_2$, where r_1 is cube-free and r_2 is cube-full. It thus follows from (3.4.6) and Lemma 3.4.4 with $m = 0$ that

$$S_r(\boldsymbol{c}) \ll |r_1|^{1+n/2+\varepsilon} |r_2|^{1+2n/3+\varepsilon}.$$

From Lemma 3.3.8 we obtain the estimate $I_r(\mathbf{0}) \ll |P|^{-3+\varepsilon}$. We thus get

$$\begin{aligned} N_0(P) &\ll |P|^{n-3+\varepsilon} \sum_{|r_1|\le\widehat{Q}} |r_1|^{-n} S_{r_1}(\boldsymbol{c}) \sum_{|r_2|\le\widehat{Q}/|r_1|} |r_2|^{-n} S_{r_2}(\boldsymbol{c}) \\ &\ll |P|^{n-3+\varepsilon} \sum_{|r_1|\le\widehat{Q}} |r_1|^{1-n/2} \sum_{|r_2|\le\widehat{Q}/|r_1|} |r_2|^{1-n/3} \\ &\ll |P|^{n-3+\varepsilon}, \end{aligned}$$

since there are $O(\widehat{Y}^{1/3})$ cube-full r_2 with $|r_2| = \widehat{Y}$.

3.6.2 Contribution from $E_1(P)$

We begin with some preparations for the term $E_1(P)$. Let $0 \le Y \le Q$ and fix the absolute value of r to be $\widehat{Y}$. As in Section 3.4.3, we will also fix a set of indices $\mathcal{T} \subset \{1,\dots,n\}$ of cardinality t, as well as a tuple $\boldsymbol{C} = (C_i)_{i\in\mathcal{T}}$, where $1 \le C_i \le C$ and denote by $\mathcal{R}(\boldsymbol{C})$ the set of vectors $\boldsymbol{c} = (c_1,\dots,c_n) \in \mathcal{O}^n$ such that $|c_i| = \widehat{C}_i$ if $i \in \mathcal{T}$ and $c_j = 0$ if $j \notin \mathcal{T}$. Let us put $\mathcal{C} = \max_{i\in\mathcal{T}} C_i$, so that $|\boldsymbol{c}| = \widehat{\mathcal{C}}$ whenever $\boldsymbol{c} \in \mathcal{R}(\boldsymbol{C})$. We then define $E_1(\mathcal{R}(\boldsymbol{C}), \widehat{Y})$ to be the contribution coming from $\boldsymbol{c} \in \mathcal{R}(\boldsymbol{C})$ and $|r| = \widehat{Y}$ in the definition of $E_1(P)$ given in (3.6.3). Explicitly, this means

$$E_1(\mathcal{R}(\boldsymbol{C}), \widehat{Y}) = \frac{|P|^n}{\widehat{Y}^n} \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C}) \\ F^*(\boldsymbol{c})\neq 0}} \sum_{\substack{r \text{ monic} \\ |r|=\widehat{Y}}} S_r(\boldsymbol{c}) I_{\widehat{Y}}(\boldsymbol{c}),$$

where

$$I_{\widehat{Y}}(\boldsymbol{c}) = \int_{|\theta|<\widehat{Y}^{-1}\widehat{Q}^{-1}} I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta.$$

The definition of $I_{\widehat{Y}}(\boldsymbol{c})$ makes sense by Lemma 3.3.6, which shows that the value of the double integral in the definition of $I_{\widehat{Y}}(\boldsymbol{c})$ only depends on the absolute value of r for $\boldsymbol{c}$ fixed.

Note that there are $Q+1 \ll |P|^\varepsilon$ possibilities for Y and $O(C^n) = O(|P|^\varepsilon)$ choices for $\boldsymbol{C}$. In particular, if we can show that $E_1(\mathcal{R}(\boldsymbol{C}), \widehat{Y}) \ll |P|^{3n/4-3/2+\varepsilon}$ holds, then the same estimate will be true for $E_1(P)$ with a new value of $\varepsilon > 0$. Next we tansform $E_1(P)$ in such a way that Lemma 3.4.2 and Lemma 3.4.5 are applicable. For this we write $r = b_1' b_1 r_2$, where r_2 is the square-full part of r and $b_1' b_1$ is the square-free part of r. Moreover, if we let S be the set of prime divisors of $\Delta_F F^*(\boldsymbol{c})$, then we further require that $(b_1, S) = 1$ and each prime $\varpi \mid b_1'$ satisfies $\varpi \in S$. It then follows from (3.4.4) that

$$E_1(\mathcal{R}(\boldsymbol{C}),\widehat{Y}) = \frac{|P|^n}{\widehat{Y}^{(n-1)/2}} \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C})\\ F^*(\boldsymbol{c})\neq 0}} I_{\widehat{Y}}(\boldsymbol{c}) \sum_{|r_2|\leq \widehat{Y}} \frac{S_{r_2}(\boldsymbol{c})}{|r_2|^{(n+1)/2}} \sum_{|b_1'|\leq \frac{\widehat{Y}}{|r_2|}} \frac{S_{b_1'}(\boldsymbol{c})}{|b_1'|^{(n+1)/2}} \sum_{\substack{|b_1|=\frac{\widehat{Y}}{|r_2 b_1'|}\\ (b_1,S)=1}} \frac{S_{b_1}(\boldsymbol{c})}{|b_1|^{(n+1)/2}}. \quad (3.6.5)$$

We can now apply Lemma 3.4.2 to the innermost sum to obtain

$$\sum_{\substack{|b_1|=\frac{\widehat{Y}}{|r_2 b_1'|}\\ (b_1,S)=1}} \frac{S_{b_1}(\boldsymbol{c})}{|b_1|^{(n+1)/2}} \ll \widehat{C}^{\varepsilon}(\widehat{Y}|r_2 b_1'|^{-1})^{1/2+\varepsilon}. \quad (3.6.6)$$

Moreover, by (3.4.6) and (3.4.4) we also have

$$\sum_{|b_1'|\leq \frac{\widehat{Y}}{|r_2|}} \frac{|S_{b_1'}(\boldsymbol{c})|}{|b_1'|^{n/2+1}} \ll |P|^{\varepsilon} \sum_{|b_1'|\leq \widehat{Y}/|r_2|} \frac{|(b_1',\nabla F^*(\boldsymbol{c}))|^{1/2}}{|b_1'|^{1/2}} \ll |P|^{\varepsilon}, \quad (3.6.7)$$

where we used that there at most $O((\widehat{Y}|r_2|^{-1}|F^*(\boldsymbol{c})|)^{\varepsilon}) = O(|P|^{\varepsilon})$ possibilities for square-free b_1' whose prime divisors are restricted to S with $|b_1'| \leq \widehat{Y}|r_2|^{-1}$. After inserting (3.6.6) and (3.6.7) into (3.6.5), we see that

$$E_1(\mathcal{R}(\boldsymbol{C}),\widehat{Y}) \ll \frac{|P|^{n+\varepsilon}}{\widehat{Y}^{n/2-1}} \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C})\\ F^*(\boldsymbol{c})\neq 0}} |I_{\widehat{Y}}(\boldsymbol{c})| \sum_{|r_2|\leq \widehat{Y}} \frac{|S_{r_2}(\boldsymbol{c})|}{|r_2|^{n/2+1}}.$$

We can now estimate $I_{\widehat{Y}}(\boldsymbol{c})$ with Lemma 3.3.7:

$$\begin{aligned} I_{\widehat{Y}}(\boldsymbol{c}) &\ll \widehat{Y}^{-1}\widehat{Q}^{-1} \prod_{i=1}^{n} \min\left\{ \left(\frac{|P||\boldsymbol{c}|}{\widehat{Y}}\right)^{-1/4}, \left(\frac{|P||c_i|}{\widehat{Y}}\right)^{-1/2} \right\} \\ &= \widehat{Y}^{-1}\widehat{Q}^{-1} \left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{(n-t)/4} \prod_{i\in\mathcal{T}} \min\left\{ \left(\frac{|P|\widehat{\mathcal{C}}}{\widehat{Y}}\right)^{-1/4}, \left(\frac{|P|\widehat{C}_i}{\widehat{Y}}\right)^{-1/2} \right\}, \end{aligned}$$

where we used that $\min\left\{ \left(\frac{|P|\widehat{\mathcal{C}}}{\widehat{Y}}\right)^{-1/4}, \left(\frac{|P||c_i|}{\widehat{Y}}\right)^{-1/2} \right\} = (|P|\widehat{\mathcal{C}}\widehat{Y}^{-1})^{-1/4}$ if $i \notin \mathcal{T}$. Denote the last product above by Π. Then after dividing r_2 into q-adic ranges, Lemma 3.4.5 implies

$$\begin{aligned} E_1(\mathcal{R}(\boldsymbol{C}),\widehat{Y}) &\ll \frac{|P|^{n+\varepsilon}}{\widehat{Y}^{n/2}\widehat{Q}} \left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{(n-t)/4} \Pi \sum_{\substack{\boldsymbol{c}\in\mathcal{R}(\boldsymbol{C})\\ F^*(\boldsymbol{c})\neq 0}} \sum_{|r_2|\leq \widehat{Y}} \frac{|S_{r_2}(\boldsymbol{c})|}{|r_2|^{n/2+1}} \\ &\ll \frac{|P|^{n+\varepsilon}}{\widehat{Y}^{n/2}\widehat{Q}} \left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{(n-t)/4} \widehat{Y}^{(n-t)/6}\Pi\#\mathcal{R}(\boldsymbol{C}). \end{aligned}$$

From the fact that $\#\mathcal{R}(\boldsymbol{C}) \ll \prod_{i\in\mathcal{T}} \widehat{C}_i$ we deduce that

$$\begin{aligned}\#\mathcal{R}(\boldsymbol{C})\Pi &\ll \prod_{i\in\mathcal{T}} \min\left\{\widehat{C}_i\left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{1/4}, \left(\frac{\widehat{C}_i\widehat{Y}}{|P|}\right)^{1/2}\right\}\\ &\ll \widehat{\mathcal{C}}^{\,t}\left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{t/4}\min\left\{1, \frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right\}^{t/4},\end{aligned}$$

where we used that $\widehat{C}_i \leq \widehat{\mathcal{C}}$. Recalling (3.6.1), we therefore have

$$E_1(\mathcal{R}(\boldsymbol{C}), \widehat{Y}) \ll \frac{|P|^{n-3/2+\varepsilon}}{\widehat{Y}^{n/2}}\left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{n/4}\widehat{Y}^{(n-t)/6}\widehat{\mathcal{C}}^{\,t}\min\left\{1, \frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right\}^{t/4}.$$

One easily sees that the expression above is maximal either at $t = 0$ or $t = n$. For $t = 0$, we get

$$\begin{aligned}\frac{|P|^{n-3/2+\varepsilon}}{\widehat{Y}^{n/2}}\left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{n/4}\widehat{Y}^{n/6} &= |P|^{3n/4-3/2+\varepsilon}\widehat{Y}^{-n/12}\widehat{\mathcal{C}}^{-n/4}\\ &\ll |P|^{3n/4-3/2+\varepsilon}\end{aligned}$$

as desired. For $t = n$, we have

$$\begin{aligned}\frac{|P|^{n-3/2+\varepsilon}}{\widehat{Y}^{n/2}}\left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}}\right)^{n/4}\widehat{\mathcal{C}}^{n}\min\left\{1, \frac{\widehat{Y}}{\widehat{\mathcal{C}}|P|}\right\}^{n/4} &\ll |P|^{n/2-3/2+\varepsilon}\widehat{\mathcal{C}}^{n/2}\\ &\ll |P|^{3n/4-3/2+\varepsilon}\end{aligned}$$

since $\widehat{\mathcal{C}} \leq \widehat{C} \ll |P|^{1/2}$. This finishes our treatment of $E_1(P)$.

3.6.3 Contribution from $E_2(P)$ for ordinary solutions

Now we turn our attention to the term $E_2(P)$. For $n = 4$ we further divide it into $E_2(P) = E_2^{\mathrm{ord}}(P) + E_2^{\mathrm{spec}}(P)$, where $E_2^{\mathrm{spec}}(P)$ is restricted to special solutions of $F^*(\boldsymbol{c}) = 0$ in the sense of Section 3.5 and $E_2^{\mathrm{ord}}(P)$ to ordinary solutions of $F^*(\boldsymbol{c}) = 0$. In this section we deal with $E_2(P)$ for $n = 6$ and $E_2^{\mathrm{ord}}(P)$ for $n = 4$.

We shall again fix the absolute value of r to be $\widehat{Y}$ for some $0 \leq Y \leq Q$ and the absolute value of $\boldsymbol{c}$ to be $\widehat{\mathcal{C}}$ for some $0 < \mathcal{C} \leq C$. We will then consider the sum

$$E_2(Y, \mathcal{C}) := \frac{|P|^n}{\widehat{Y}^n}\sum_{\substack{|\boldsymbol{c}|=\widehat{\mathcal{C}}\\ F^*(\boldsymbol{c})=0}}\sum_{\substack{r \text{ monic}\\ |r|=\widehat{Y}}} S_r(\boldsymbol{c})I_{\widehat{Y}}(\boldsymbol{c}),$$

where the sum over $\boldsymbol{c}$ is restricted to ordinary solutions of $F^*(\boldsymbol{c}) = 0$ for $n = 4$. Once we have shown $E_2(Y, \mathcal{C}) \ll |P|^{3n/4-3/2+\varepsilon}$ the same estimate will follow for $E_2(P)$ for $n = 6$ and for $E_2^{\mathrm{ord}}(P)$ for $n = 4$, because there are only $O(|P|^\varepsilon)$ possible pairs of Y's and $\mathcal{C}$'s.

Lemma 3.6.1. *Let F be a non-singular cubic form in 4 or 6 variables, and let F^* be its dual form. Suppose there exists some $\eta > 0$ such that for any $\widehat{\mathcal{C}} \geq 1$ the following bound holds*

$$\#\{\boldsymbol{x} \in \mathcal{O}^n : \boldsymbol{x} \text{ is an ordinary solution to } F^*(\boldsymbol{x}) = 0, |\boldsymbol{x}| \leq \widehat{\mathcal{C}}\} \ll \widehat{\mathcal{C}}^{n-3+\eta}.$$

Then we have

$$E_2(P) \ll |P|^{3n/4-3/2+\eta/2+\varepsilon}.$$

Proof. If $D = \deg F^*$, then we see from (3.4.2) and Lemma 3.4.1 that F^* has non-zero monomials of the form $G_i x_i^D$ for every $i = 1, \dots, n$. In particular, if $|\boldsymbol{c}| = \widehat{\mathcal{C}}$ and $F^*(\boldsymbol{c}) = 0$, then there must be at least two indices $i \neq j$ such that $\widehat{\mathcal{C}} \ll |c_i| \ll |c_j| \ll \widehat{\mathcal{C}}$. Therefore, from Lemma 3.3.7 we deduce

$$I_{\widehat{Y}}(\boldsymbol{c}) \ll \frac{\widehat{\mathcal{C}}}{|P|^2 \widehat{Y}} \prod_{i=1}^{n} \min\left\{ \left(\frac{\widehat{Y}}{|P||c_i|} \right)^{1/2}, \left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}} \right)^{1/4} \right\} \ll \left(\frac{\widehat{Y}}{|P|\widehat{\mathcal{C}}} \right)^{(n-2)/4} |P|^{-3}. \quad (3.6.8)$$

Next we deal with the sum $S_r(\boldsymbol{c})$. Write $r = r_1 r_2 r_3$ into coprime monic factors r_i, where r_1 is cube-free, r_2 is cube-full and each prime divisor of r_3 divides $\prod F_i$.

Let us begin with $S_{r_2}(\boldsymbol{c})$. Suppose $\varpi^k \parallel r_2$ and write $H_\varpi = (\varpi^k, \boldsymbol{c})$. It follows that $\boldsymbol{c} = H_\varpi \boldsymbol{c}'$ for some $\boldsymbol{c}' \in \mathcal{O}^n$ with $(\varpi, \boldsymbol{c}') = 1$. It is again easy to see that any prime divisor of the coefficients G_i of the top-degree monomials x_i^D of F^* divides $\prod F_i$. In particular, if $H_\varpi \neq \varpi^k$, then $F^*(\boldsymbol{c}') = 0$ implies that at least two entries of $\boldsymbol{c}'$ are coprime to ϖ. On the other hand, if $H_\varpi = \varpi^k$, then $(\varpi^k, c_i) = \varpi^k$ for every $i = 1, \dots, n$, so that in any case there are always least two distinct indices $i \neq j$ such that $(\varpi^k, c_i) = (\varpi^k, c_j) = H_\varpi$. Consequently it follows from Lemma 3.4.4 with $m = 2$ that

$$S_{r_2}(\boldsymbol{c}) \ll |r_2|^{2/3+2n/3+\varepsilon} |H|^{1/2},$$

where $H = \prod_{\varpi | r_2} H_\varpi$ divides each entry of $\boldsymbol{c}$.

In addition, (3.4.6) and Lemma 3.4.4 give us $S_{r_1}(\boldsymbol{c}) \ll |r_1|^{1+n/2+\varepsilon}$ and (3.4.8) tells us that $S_{r_3}(\boldsymbol{c}) \ll |r_3|^{1+2n/3+\varepsilon}$. To sum up, we have

$$S_r(\boldsymbol{c}) \ll |r|^{\varepsilon}|r_1|^{1+n/2}|r_2|^{2/3+2n/3}|r_3|^{1+2n/3}|H|^{1/2}.$$

Let us fix $|r_i| = \widehat{Y}_i$, where $0 \le Y_i \le Y$ and $Y_1 + Y_2 + Y_3 = Y$. We want to give an upper bound for

$$\mathcal{S} := \sum_{|r_i|=\widehat{Y}_i, i=1,2,3} \sum_{\substack{|\boldsymbol{c}|=\widehat{C} \\ F^*(\boldsymbol{c})=0}} |S_r(\boldsymbol{c})|.$$

Taking into account that the number of available r_1 and r_3 is $O(\widehat{Y}_1)$ and $O(|P|^{\varepsilon})$ respectively, we see that

$$\begin{aligned}\mathcal{S} &\ll |P|^{\varepsilon}\widehat{Y}_1^{2+n/2}\widehat{Y}_2^{2/3+2n/3}\widehat{Y}_3^{1+2n/3} \sum_{|r_2|=\widehat{Y}_2} \sum_{H|r_2} |H|^{1/2} \sum_{\substack{|\boldsymbol{c}|=\widehat{C}/|H| \\ F^*(\boldsymbol{c})=0}} 1 \\ &\ll |P|^{\varepsilon}\widehat{C}^{n-3+\eta}\widehat{Y}_1^{2+n/2}\widehat{Y}_2^{2/3+2n/3}\widehat{Y}_3^{1+2n/3} \sum_{|r_2|=\widehat{Y}_2} \sum_{H|r_2} |H|^{7/2-n-\eta},\end{aligned}$$

where we used the main assumption of the lemma in order to bound the number of ordinary solutions of $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| = \widehat{C}/|H|$ for the second inequality. Since $n \ge 4$ clearly $7/2 - n - \eta \le 0$ holds and since the number of available r_2 is $O(\widehat{Y}_2^{1/3})$, it follows that

$$\mathcal{S} \ll |P|^{\varepsilon}\widehat{C}^{n-3+\eta}\widehat{Y}_1^{2+n/2}\widehat{Y}_2^{1+2n/3}\widehat{Y}_3^{1+2n/3} \ll |P|^{\varepsilon}\widehat{C}^{n-3+\eta}\widehat{Y}^{2+n/2}, \tag{3.6.9}$$

because $2 + n/2 \ge 1 + 2n/3$ for $n \le 6$. As there are only $O(|P|^{\varepsilon})$ possibilities for permissible triples (Y_1, Y_2, Y_3), we deduce from (3.6.8) and (3.6.9) that

$$E_2(Y, \widehat{C}) \ll |P|^{3n/4-5/2+\varepsilon}\widehat{Y}^{3/2-n/4}\widehat{\mathcal{C}}^{3n/4-5/2+\eta}.$$

In particular, since $\widehat{\mathcal{C}} \ll |P|^{1/2}$ and $\widehat{Y} \ll |P|^{3/2}$, we thus obtain

$$\begin{aligned}E_2(Y, \mathcal{C}) &\ll |P|^{3n/4-5/2+\varepsilon}|P|^{9/4-3n/8}|P|^{3n/8-5/4+\eta/2} \\ &\ll |P|^{3n/4-3/2+\eta/2+\varepsilon},\end{aligned}$$

which completes the proof. □

At this point our treatment of $E_2(P)$ differs depending on the characteristic of K.

If $\mathrm{char}(K) > 3$, then by virtue of Lemma 3.5.1 we know that the number of ordinary solutions of the dual form $F^*(\boldsymbol{c}) = 0$ such that $|\boldsymbol{c}| \le \widehat{\mathcal{C}}$ is bounded by $O(\widehat{\mathcal{C}}^{n-3+\varepsilon})$. Therefore Lemma 3.6.1 implies

$$E_2^{\mathrm{ord}}(P) \ll |P|^{3n/4-3/2+\varepsilon} \quad \text{and} \quad E_2(P) \ll |P|^{3n/4-3/2+\varepsilon},$$

for $n = 4$ and $n = 6$, respectively. This finishes our treatment of $E_2(P)$ in this case.

If $\mathrm{char}(K) = 2$, then we need to argue differently. We begin by considering the case when $n = 6$. According to Lemma 3.4.1 the dual form takes the shape of a non-singular diagonal cubic form. In particular, we can trivially bound the number of solutions to $F^*(\boldsymbol{c}) = 0$ such that $|\boldsymbol{c}| \leq \widehat{\mathcal{C}}$ by $O(\widehat{\mathcal{C}}^6) = O(\widehat{\mathcal{C}}^{n-3+\eta})$, where $\eta = 3$. Therefore, Lemma 3.6.1 gives

$$E_2(P) \ll |P|^{3n/4-3/2+\eta/2+\varepsilon} = |P|^{n-3+\eta/2+\varepsilon}.$$

This, together with our bounds for $N_0(P)$ and $E_1(P)$ established earlier in this section, shows that

$$N(P) \ll |P|^{n-3+\eta/2+\varepsilon}.$$

This holds for any non-singular, diagonal cubic form over K when $\mathrm{char}(K) = 2$. In particular, as a result we can bound the number of solutions to $F^*(\boldsymbol{c}) = 0$ with $|\boldsymbol{c}| \leq \widehat{\mathcal{C}}$ by $O(\widehat{\mathcal{C}}^{n-3+\eta/2+\varepsilon})$. Another application of Lemma 3.6.1 yields

$$E_2(P) \ll |P|^{3n/4-3/2+\eta/4+\varepsilon}$$

and we may argue as above to deduce

$$N(P) \ll |P|^{n-3+\eta/4+\varepsilon}.$$

If we repeat this process k-times, where $2^{-k+1} \leq \varepsilon$ we find

$$E_2(P) \ll |P|^{3n/4-3/2+2\varepsilon},$$

which concludes our treatment for $E_2(P)$ in this case.

On the other hand, if $n = 4$ we can trivially estimate the number of solutions to $F^*(\boldsymbol{c}) = 0$ of bounded height $\widehat{\mathcal{C}}$ by $O(\widehat{\mathcal{C}}^4) = O(\widehat{\mathcal{C}}^{n-3+\eta})$, where $\eta = 3$. Lemma 3.6.1 then yields

$$E_2(P) \ll |P|^{3n/4-3/2+\eta/2+\varepsilon} = |P|^{n-3+1/2+\eta/2+\varepsilon},$$

which in turn implies

$$N(P) \ll |P|^{n-3+1/2+\eta/2+\varepsilon}.$$

Repeating this process k-times, where $k > 1/\varepsilon$ we thus find

$$E_2(P) \ll |P|^{3n/4-3/2+1/2+2\varepsilon} = |P|^{2+2\varepsilon}.$$

3.7 Waring's problem and weak approximation

Having completed our task for $n = 6$, we will now apply it to Waring's problem and weak approximation for diagonal cubic hypersurfaces of dimension at least 5.

3.7.1 Waring's problem for $n \geq 7$

Recall that $\mathbb{J}_q^3[t]$ is the additive closure of all cubes in $\mathcal{O}$. Given $P \in \mathbb{J}_q^3[t]$, we define $B := \left\lceil \frac{\deg(P)}{3} \right\rceil + 1$ and the counting function

$$R_n(P) := \#\{\boldsymbol{x} \in \mathcal{O}^n \colon |\boldsymbol{x}| < \widehat{B}, x_1^3 + \cdots + x_n^3 = P\}.$$

Our next goal is to deduce Theorem 3.1.4 from our findings. We shall accomplish this goal with a classical version of the circle method. For $\alpha \in \mathbb{T}$, we define

$$T(\alpha) := \sum_{\substack{x \in \mathcal{O} \\ |x| < \widehat{B}}} \psi(\alpha x^3).$$

It then follows from (3.2.3) that we can write our counting function as

$$R_n(P) = \int_{\mathbb{T}} T(\alpha)^n \psi(-\alpha P) \mathrm{d}\alpha.$$

We then define our set of major arcs to be

$$\mathfrak{M} := \bigcup_{\substack{|r| \leq \widehat{B} \\ r \text{ monic}}} \bigcup_{\substack{|a| < |r| \\ (a,r)=1}} \{\alpha \in \mathbb{T} \colon |r\alpha - a| < \widehat{B}^{-2}\}$$

and $\mathfrak{m} := \mathbb{T} \setminus \mathfrak{M}$ constitutes our set of minor arcs. The following lemma is a consequence of [66, Theorem 30].

Lemma 3.7.1. *Suppose* $\mathrm{char}(K) \nmid 3$ *and* $n \geq 7$. *Then there exists* $\delta > 0$ *such that for all* $P \in \mathbb{J}_q^3[t]$ *we have*

$$\int_{\mathfrak{M}} T(\alpha)^n \psi(-\alpha P) \mathrm{d}\alpha = \mathfrak{S}(P) \sigma_\infty(P) \widehat{B}^{n-3} + O\left(\widehat{B}^{n-3-\delta}\right),$$

where $\mathfrak{S}(P)$ *and* $\sigma_\infty(P)$ *are the singular series and singular integral associated to* P. *Furthermore, they satisfy*

$$1 \ll \mathfrak{S}(P) \sigma_\infty(P) \ll 1.$$

Remark 3.7.2. In fact, Kubota states Lemma 3.7.1 only for $n \geq 10$. However, as explained by Liu–Wooley in [76, Lemma 5.2], this is a result of an oversight and Kubota's argument already works for $n \geq 7$.

We now have

$$\left|\int_{\mathfrak{m}} T(\alpha)^n \psi(-\alpha P)\mathrm{d}\alpha\right| \leq \sup_{\alpha\in\mathfrak{m}} |T(\alpha)|^{n-6} \int_{\mathbb{T}} |T(\alpha)|^6 \mathrm{d}\alpha. \tag{3.7.1}$$

If $\alpha \in \mathfrak{m}$, then (3.2.2) with $\widehat{Q} = \widehat{B}$ implies the existence of $a, r \in \mathcal{O}$ with r monic such that $|a| < |r| \leq \widehat{B}$, $(a, r) = 1$ and $|r\alpha - a| < \widehat{B}^{-1}$. As $\alpha \in \mathfrak{m}$, we must have $|\alpha - a/r| \geq \widehat{B}^{-2}|r|^{-1}$. Under these circumstances Weyl's inequality, see [13, Lemma 5.10] for $\mathrm{char}(K) > 3$ and [19, Proposition IV.4] for $\mathrm{char}(K) = 2$, guarantees the existence of $\delta > 0$ such that

$$\sup_{\alpha\in\mathfrak{m}} |T(\alpha)|^{n-6} \ll \widehat{B}^{(n-6)(1-\delta)}. \tag{3.7.2}$$

Since

$$\int_{\mathbb{T}} |T(\alpha)|^6 \mathrm{d}\alpha = \#\{\boldsymbol{x} \in \mathcal{O}^6 \colon |\boldsymbol{x}| < \widehat{B}, x_1^3 + x_2^3 + x_3^3 = x_4^3 + x_5^3 + x_6^3\},$$

Theorem 3.1.1 implies

$$\int_{\mathbb{T}} |T(\alpha)|^6 \mathrm{d}\alpha \ll \widehat{B}^{3+\varepsilon}. \tag{3.7.3}$$

Plugging (3.7.2) and (3.7.3) into (3.7.1) yields

$$\begin{aligned}\int_{\mathfrak{m}} T(\alpha)^n \psi(-\alpha P)\mathrm{d}\alpha &\ll \widehat{B}^{(n-6)(1-\delta)+3+\varepsilon} \\ &= \widehat{B}^{n-3-\delta(n-6)+\varepsilon}.\end{aligned}$$

After choosing $\varepsilon = \delta(n-6)/2$, we see that the contribution of the minor arcs is

$$\int_{\mathfrak{m}} T(\alpha)^n \psi(-\alpha P)\mathrm{d}\alpha \ll \widehat{B}^{n-3-\delta(n-6)/2}.$$

Since $n \geq 7$, combining this with Lemma 3.7.1 therefore completes the proof of Theorem 3.1.4.

3.7.2 Weak approximation for cubic diagonal hypersurfaces

We will show that weak approximation holds for the diagonal cubic hypersurface defined by $F(\boldsymbol{x}) = \sum_{i=1}^n F_i x_i^3$ if $n \geq 7$. Fix $\boldsymbol{x}_0 \in \mathbb{T}^n$, $M \in \mathcal{O}$, $\boldsymbol{b} \in \mathcal{O}^n$ and $N \in \mathbb{Z}_{\geq 0}$ such that $|\boldsymbol{b}| < |M|$ and such that N is bounded in terms of M. Define the weight function $\widetilde{w} \colon K_\infty^n \to \mathbb{R}$ via

$$\widetilde{w}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } |\boldsymbol{x} - \boldsymbol{x}_0| < \widehat{N}^{-1}, \\ 0 & \text{otherwise.} \end{cases}$$

Further for $P \in \mathcal{O}$ we introduce the counting function

$$N(\widetilde{w}, P) := \sum_{\substack{\boldsymbol{x} \in \mathcal{O}^n \\ F(M\boldsymbol{x}+\boldsymbol{b})=0}} \widetilde{w}\left(\frac{M\boldsymbol{x}+\boldsymbol{b}}{P}\right).$$

As usual, we can write this as an integral over an exponential sum

$$N(\widetilde{w}, P) = \int_{\mathbb{T}} \widetilde{S}(\alpha)\mathrm{d}\alpha,$$

where

$$\widetilde{S}(\alpha) = \sum_{\boldsymbol{x} \in \mathcal{O}^n} \psi\left(\alpha F(M\boldsymbol{x}+\boldsymbol{b})\right) \widetilde{w}\left(\frac{M\boldsymbol{x}+\boldsymbol{b}}{P}\right).$$

Since F is diagonal we may factorise $\widetilde{S}(\alpha)$ as

$$\widetilde{S}(\alpha) = \prod_{i=1}^{n} \widetilde{T}_i(\alpha),$$

where

$$\widetilde{T}_i(\alpha) = \sum_{\substack{x \in \mathcal{O} \\ |Mx+b_i-x_{0,i}|<|P|\widehat{N}^{-1}}} \psi(\alpha F_i(Mx+b_i)^3).$$

Note that our counting function $N(\widetilde{w}, P)$ agrees with the function $\rho_{M,\boldsymbol{b}}(n)$ and $\widetilde{S}(\alpha)$ agrees with $T(\alpha)$ in [70, Chapter 4]. In order to show weak approximation for the variety $X = \mathbb{V}(F) \subset \mathbb{P}^{n-1}$, by the same argument as the one provided in Section 4.9 of [70], it is enough to show the following result.

Theorem 3.7.3. *Suppose* $\mathrm{char}(K) > 3$. *Then there exists some* $\delta > 0$ *such that*

$$N(\widetilde{w}, P) = |M|^{-3}\mathfrak{S}\mathfrak{I}|P|^{n-3} + O(|P|^{n-3-\delta}),$$

where $\mathfrak{S}$ *and* $\mathfrak{I}$ *are the singular series and the singular integral respectively as defined in* (3.7.5) *and* (3.7.6).

We tackle this using a traditional circle method argument.

We define the major arcs to be the set $\mathcal{M} \subset \mathbb{T}$ given by

$$\mathcal{M} = \bigcup_{\substack{r \in \mathcal{O} \\ |r|<|P|^{1/2} \\ r \text{ monic}}} \bigcup_{\substack{a \in \mathcal{O} \\ |a|<|r| \\ (a,q)=1}} \left\{\alpha \in \mathbb{T} \colon |r\alpha - a| < H_F^{-1}|M|^{-3}|r||P|^{-5/2}\right\},$$

and we take the minor arcs to be the complement $\mathfrak{m} = \mathbb{T} \setminus \mathcal{M}$.

In this context, provided $\mathrm{char}(K) > 3$, Weyl's inequality [70, Lemma 4.3.6] tells us that

$$|\widetilde{T}_i(\alpha)| \ll |P|^{1+\varepsilon} \left(\frac{|P| + |r| + |P|^3|r\alpha - a|}{|P|^3} + \frac{1}{|r| + |P|^3|r\alpha - a|} \right)^{1/4}$$

for $i = 1, \ldots, n$ if $a, r \in \mathcal{O}$ are such that $|a| < |r|$, r monic and $(a, r) = 1$. Using (3.2.2) and the definition of the minor arcs, a similar argument that handed us (3.7.2) gives

$$\sup_{\alpha \in \mathfrak{m}} \left| \widetilde{T}_i(\alpha) \right| \ll |P|^{7/8+\varepsilon}, \tag{3.7.4}$$

for any $\varepsilon > 0$. We are now ready to finish our treatment of the minor arcs. If $n \geq 7$ we obtain

$$\int_{\mathfrak{m}} |\widetilde{S}(\alpha)| \mathrm{d}\alpha = \int_{\mathfrak{m}} \left| \prod_{i=1}^{n} \widetilde{T}_i(\alpha) \right| \mathrm{d}\alpha \ll \sup_{\alpha \in \mathfrak{m}} \left| \widetilde{T}_7(\alpha) \cdots \widetilde{T}_n(\alpha) \right| \int_{\mathbb{T}} \left| \prod_{i=1}^{6} \widetilde{T}_i(\alpha) \right| \mathrm{d}\alpha.$$

The integral can be dealt with as follows. By Hölder's inequality we find

$$\int_{\mathfrak{m}} \left| \prod_{i=1}^{6} \widetilde{T}_i(\alpha) \right| \mathrm{d}\alpha \leq \prod_{i=1}^{6} \left(\int_{\mathbb{T}} |\widetilde{T}_i(\alpha)|^6 \mathrm{d}\alpha \right)^{1/6}.$$

Now the last quantity is equal to

$$\prod_{i=1}^{6} \# \left\{ \boldsymbol{x} \in \mathcal{O}^6 \colon x_j \equiv b_i \bmod M, |x_j/P - x_{0,i}| < \widehat{N}^{-1}, \text{ for all } j, \sum_{j=1}^{3} x_j^3 = \sum_{j=4}^{6} x_j^3 \right\}^{1/6},$$

which in turn is bounded by

$$\prod_{i=1}^{6} \#\{\boldsymbol{x} \in \mathcal{O}^6 \colon |\boldsymbol{x}| < |\boldsymbol{x}_0||P|,\, x_1^3 + x_2^3 + x_3^3 = x_4^3 + x_5^3 + x_6^3\}^{1/6},$$

if $|P|$ is sufficiently large. An application of Theorem 3.1.1 therefore yields

$$\int_{\mathbb{T}} \left| \prod_{i=1}^{6} \widetilde{T}_i(\alpha) \right| \mathrm{d}\alpha \ll |P|^{3+\varepsilon}.$$

Once combined with (3.7.4) we thus obtain

$$\int_{\mathfrak{m}} |\widetilde{S}(\alpha)| \mathrm{d}\alpha \ll |P|^{n-3-(n-6)/8+\varepsilon}$$

for any $\varepsilon > 0$, which is satisfactory if $n \geq 7$. We now turn to the major arcs. Given $a, r \in \mathcal{O}$ write

$$\widetilde{S}_r(a) := \sum_{|\boldsymbol{x}| < |r|} \psi \left(\frac{aF(M\boldsymbol{x} + \boldsymbol{b})}{r} \right).$$

For any $Y \in \mathbb{R}$ we define the truncated singular series

$$\mathfrak{S}(\widehat{Y}) := \sum_{\substack{|r|<\widehat{Y} \\ r \text{ monic}}} \sum_{\substack{|a|<|r| \\ (a,r)=1}} |r|^{-n} \widetilde{S}_r(a),$$

and the truncated singular integral to be

$$\mathfrak{I}(\widehat{Y}) = \int_{|\gamma|<H_F^{-1}\widehat{Y}} I(\gamma) \mathrm{d}\gamma,$$

where

$$I(\gamma) = \int_{\mathbb{T}^n} \psi(\gamma F(\boldsymbol{x})) \widetilde{w}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

Then from (4.6.30) in [70] it follows that we have

$$\int_{\mathcal{M}} \widetilde{S}(\alpha) \mathrm{d}\alpha = |M|^{-3} \mathfrak{S}(|P|^{1/2}) \mathfrak{I}(|P|^{1/2}) |P|^{n-3}.$$

It remains to study the convergence of the singular integral and singular series. In order to handle the singular series we will need upper bounds for $\widetilde{S}_r(a)$. First, we record the following multiplicative property, which is shown in [70, Lemma 4.7.2]. If $r_1, r_2 \in \mathcal{O}$ are coprime then

$$\widetilde{S}_{r_1 r_2}(a) = \widetilde{S}_{r_1}(a_1) \widetilde{S}_{r_2}(a_2),$$

where $a_i \in \mathcal{O}$ are such that $a_1 \equiv a\tilde{r}_2 \mod r_1$ and $a_2 \equiv a\tilde{r}_1 \mod r_2$, where $\tilde{r}_1, \tilde{r}_2$ denote the multiplicative inverses modulo r_2, r_1, respectively. Thus, from (3.4.8) in combination with the divisor estimate, it follows that we have

$$\widetilde{S}_r(a) \ll |r|^{2n/3+\varepsilon},$$

where the constant may depend on M, b and ε.

Using this we see that

$$\sum_{\substack{|r|=\widehat{Y} \\ r \text{ monic}}} \sum_{\substack{|a|<|r| \\ (a,r)=1}} |r|^{-n} \left| \widetilde{S}_r(a) \right| \ll \widehat{Y}^{(2-n/3+\varepsilon)}.$$

Since $n \geq 7$ we deduce absolute convergence of the series

$$\mathfrak{S} = \sum_{r \text{ monic}} \sum_{\substack{|a|<|r| \\ (a,r)=1}} |r|^{-n} \widetilde{S}_r(a), \tag{3.7.5}$$

which is the singular series. Moreover choosing positive $\varepsilon < (n-6)/6$ we find

$$\mathfrak{S} - \mathfrak{S}(|P|^{1/2}) \ll |P|^{1-n/6+\varepsilon},$$

if $n \geq 7$ upon redefining ε. We turn to the singular integral. Let $\boldsymbol{x}_0 \in K_\infty$ be a non-singular point of $X \subset \mathbb{P}^{n-1}$. In [17] it is shown in Lemma 7.5 and the paragraphs preceding it that

$$\mathfrak{I}(\widehat{Y}) = \mathfrak{I}(\widehat{N}/|\nabla F(\boldsymbol{x}_0)|) = \frac{1}{|\nabla F(\boldsymbol{x}_0)|\widehat{N}^{n-1}}$$

whenever $\widehat{Y} \geq \widehat{N}/|\nabla F(\boldsymbol{x}_0)|$. Thus clearly $\lim_{\widehat{Y}\to\infty} \mathfrak{I}(\widehat{Y})$ exists and is equal to

$$\mathfrak{I} := \lim_{\widehat{Y}\to\infty} \mathfrak{I}(\widehat{Y}) = \frac{1}{|\nabla F(\boldsymbol{x}_0)|\widehat{N}^{n-1}}. \tag{3.7.6}$$

We conclude that

$$N(\widetilde{w}, P) = |M|^{-3}\mathfrak{S}\mathfrak{I}|P|^{n-3} + O(|P|^{n-3-1/8+\varepsilon}),$$

as desired.

3.8 Special solutions and the case $n = 4$

In this section we will concern ourselves with understanding how the special solutions of $F^*(\boldsymbol{c}) = 0$ in the case $n = 4$ relate to the solutions of $F(\boldsymbol{x}) = 0$ on rational lines. The goal of this section is to prove the following lemma, from which Theorem 3.1.2 immediately follows.

Lemma 3.8.1. *For any $\varepsilon > 0$ the following holds*

$$|P|^4 \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-4} \int_{|\theta| < |r|^{-1}\widehat{Q}^{-1}} \sum_{\boldsymbol{c}}^{\text{spec}} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) \mathrm{d}\theta = \sum_{\boldsymbol{x}}^{\text{line}} w(P^{-1}\boldsymbol{x}) + O(|P|^{3/2+\varepsilon}), \tag{3.8.1}$$

where $\sum_{\boldsymbol{c}}^{\text{spec}}$ denotes the sum over the special solutions *$\boldsymbol{c} \in \mathcal{O}^4 \setminus \{\mathbf{0}\}$ of $F^*(\boldsymbol{c}) = 0$ such that*

$$(F_1^{-1}c_1^3)^{1/2} \pm (F_2^{-1}c_2^3)^{1/2} = (F_3^{-1}c_3^3)^{1/2} \pm (F_4^{-1}c_4^3)^{1/2} = 0 \tag{3.8.2}$$

and $\sum_{\boldsymbol{x}}^{\text{line}}$ denotes the sum over points $\boldsymbol{x} \in \mathcal{O}^4$ satisfying

$$F_1x_1^3 + F_2x_2^3 = F_3x_3^3 + F_4x_4^3 = 0. \tag{3.8.3}$$

For notational convenience, this lemma only considers the case of lines such that $(i, j, k, l) = (1, 2, 3, 4)$ in the language of Theorem 3.1.2. By the symmetry of the situation at hand it is clear that the result follows for any permutation of indices.

3.8.1 Analysis of special solutions

We begin by noting that with an error of $O(|P|^{3/2+\varepsilon})$ we may include tuples $\boldsymbol{c} \in \mathcal{O}^4 \setminus \{\boldsymbol{0}\}$ satisfying (3.8.2) such that $c_i = 0$ for at least one i in the sum appearing in the left hand side of (3.8.1). Write $\sum_{\boldsymbol{c}}^{\widetilde{\text{spec}}}$ for the sum over such tuples $\boldsymbol{c}$. Note for such $\boldsymbol{c}$ Lemma 3.3.7 gives

$$I_r(\boldsymbol{c}) \ll |P|^{-5/2}|\boldsymbol{c}|^{-1},$$

for any $r \in \mathcal{O}$. Also note that $I_r(\theta, \boldsymbol{c}) = 0$ if $|\boldsymbol{c}| \gg |P|^{1/2}$. From (3.4.6) and Lemma 3.4.5, where we apply the second part with $m = 0$, we obtain

$$S_r(\boldsymbol{c}) \ll |r|^{\varepsilon}|r_1|^3|r_2|^{4-1/3},$$

where r_1 denotes the cube-free and r_2 the cube-full part of r. Hence

$$\sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-4}S_r(\boldsymbol{c}) \ll |P|^{\varepsilon} \left(\sum_{|r_1| \leq \widehat{Q}} |r_1|^{-1} \right) \left(\sum_{|r_2| \leq \widehat{Q}} |r_2|^{-1/3} \right) \ll |P|^{\varepsilon},$$

since the number of cube-full r_2 of a fixed absolute value of $\widehat{Y}$, say, is at most $P(\widehat{Y}^{1/3})$. To summarise, we found that the contribution to the left hand side of (3.8.1) is at most

$$|P|^4 \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-4} \sum_{\boldsymbol{c}}^{\widetilde{\text{spec}}} S_r(\boldsymbol{c})I_r(\boldsymbol{c}) \ll |P|^{3/2+\varepsilon} \sum_{0<|\boldsymbol{c}|\leq |P|^{1/2}}^{\widetilde{\text{spec}}} |\boldsymbol{c}|^{-1} \ll |P|^{3/2+\varepsilon},$$

where the last estimate follows since there are only $O(\widehat{C})$ vectors $\boldsymbol{c}$ of absolute value $\widehat{C}$, say, appearing in $\sum_{\boldsymbol{c}}^{\widetilde{\text{spec}}}$.

We may assume that both F_1/F_2 and F_3/F_4 are cubes in K. Otherwise the conclusion of the lemma is easily seen to be true, since there are no special solutions and $O(|P|)$ points $\boldsymbol{x}$ satisfying (3.8.3). Therefore there exist at most $O(1)$ many different possible $\rho_i \in \mathcal{O}$ with $(\rho_1, \rho_2) = (\rho_3, \rho_4) = 1$ and $\lambda, \mu \in \mathcal{O}$ such that

$$F_1 = \lambda\rho_1^3, \quad F_2 = \lambda\rho_2^3, \quad F_3 = \mu\rho_3^3, \quad F_4 = \mu\rho_4^3.$$

The different possibilites for ρ_i come from the potential existence of non-trivial third roots of unity in K. For a choice of $\rho_i \in \mathcal{O}$ if we write

$$c_1 = \rho_1 d_1, \quad c_2 = \rho_2 d_1, \quad c_3 = \rho_3 d_2, \quad c_4 = \rho_4 d_2,$$

then as we run through the possible choices of ρ_i and as $\boldsymbol{d}$ runs through $\mathcal{O}^2$, then $\boldsymbol{c}$ runs through solutions of $F^*(\boldsymbol{c}) = 0$ satisfying (3.8.2) . Given a choice of ρ_i there exist $\rho_i' \in \mathcal{O}$ such that

$$\rho_1\rho_2' - \rho_2\rho_1' = \rho_3\rho_4' - \rho_4\rho_3' = 1.$$

Then the change of variables $(x_1, x_2, x_3, x_4) \mapsto (y_1, y_2, z_1, z_2)$ given by

$$\begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 & 0 & 0 \\ \rho_1' & \rho_2' & 0 & 0 \\ 0 & 0 & \rho_3 & \rho_4 \\ 0 & 0 & \rho_3' & \rho_4' \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

is unimodular. Moreover the inverse of this is easily seen to be

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \rho_2' & -\rho_2 & 0 & 0 \\ -\rho_1' & \rho_1 & 0 & 0 \\ 0 & 0 & \rho_4' & -\rho_4 \\ 0 & 0 & -\rho_3' & \rho_3 \end{pmatrix} \begin{pmatrix} y_1 \\ z_1 \\ y_2 \\ z_2 \end{pmatrix}.$$

We will write $\boldsymbol{x}(\boldsymbol{y}, \boldsymbol{z})$ for $\boldsymbol{x}$ arising from this linear transformation. An easy calculation reveals

$$F(\boldsymbol{x}(\boldsymbol{y}, \boldsymbol{z})) = y_1 Q_1(y_1, z_1) + y_2 Q_2(y_2, z_2) =: \widetilde{F}(\boldsymbol{y}, \boldsymbol{z}),$$

where Q_i are the quadratic forms given by

$$Q_1(y, z) = \frac{\lambda}{4}\left(y^2 + 3\{2\rho_1\rho_2 z - (\rho_1\rho_2' + \rho_1'\rho_2)y\}^2\right),$$

and

$$Q_2(y, z) = \frac{\mu}{4}\left(y^2 + 3\{2\rho_3\rho_4 z - (\rho_3\rho_4' + \rho_3'\rho_4)y\}^2\right).$$

With this notation we then find

$$S_r(\boldsymbol{c}) = \sideset{}{'}\sum_{|a|<|r|} \sum_{|\boldsymbol{g}|,|\boldsymbol{h}|<|r|} \psi\left(\frac{a\widetilde{F}(\boldsymbol{g}, \boldsymbol{h}) + \boldsymbol{g}\cdot\boldsymbol{d}}{r}\right),$$

and

$$I_r(\theta, \boldsymbol{c}) = \int_{K_\infty^2} \int_{K_\infty^2} w(\boldsymbol{x}(\boldsymbol{y}, \boldsymbol{z}))\psi\left(\theta P^3 \widetilde{F}(\boldsymbol{y}, \boldsymbol{z}) + P\frac{\boldsymbol{y}\cdot\boldsymbol{d}}{r}\right) \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}.$$

We make the change of variables $\boldsymbol{y} = P^{-1}(\boldsymbol{g} + r\boldsymbol{v})$ in the integral to obtain

$$I_r(\theta, \boldsymbol{c}) = |r|^2|P|^{-2} \int_{K_\infty^2} \int_{K_\infty^2} w(\boldsymbol{x}(P^{-1}(\boldsymbol{g} + r\boldsymbol{v}), \boldsymbol{z})) \times \psi\left(\theta P^3 \widetilde{F}(P^{-1}(\boldsymbol{g} + r\boldsymbol{v}), \boldsymbol{z}) + \frac{\boldsymbol{g}\cdot\boldsymbol{d}}{r}\right) \psi(\boldsymbol{v}\cdot\boldsymbol{d})\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{z}.$$

Hence we find

$$\sum_{\boldsymbol{c}}{}^{\text{spec}} S_r(\boldsymbol{c}) I_r(\theta, \boldsymbol{c}) = |r|^2 |P|^{-2} \sum_{\rho_i} \sum_{|\boldsymbol{g}|<|r|} \int_{K_\infty^2} \sum_{\boldsymbol{d}\in\mathcal{O}^2} \int_{K_\infty^2} f_{\boldsymbol{g},\boldsymbol{z}}(\theta, \boldsymbol{v}) \psi(\boldsymbol{v}\cdot\boldsymbol{d}) \mathrm{d}\boldsymbol{v} \mathrm{d}\boldsymbol{z},$$

where $\sum_{\rho_i}$ sums over the finitely many possible choices for $\rho_i \in \mathcal{O}$ as above and where

$$f_{\boldsymbol{g},\boldsymbol{z}}(\theta, \boldsymbol{v}) = \sum_{|a|<|r|}{}' \sum_{|\boldsymbol{h}|<|r|} w(\boldsymbol{x}(P^{-1}(\boldsymbol{g}+r\boldsymbol{v}), \boldsymbol{z})) \psi\left(\theta P^3 \widetilde{F}(P^{-1}(\boldsymbol{g}+r\boldsymbol{v}), \boldsymbol{z}) + \frac{a\widetilde{F}(\boldsymbol{g},\boldsymbol{h})}{r}\right).$$

Poisson summation (3.2.4) yields

$$\sum_{\boldsymbol{d}\in\mathcal{O}^2} \int_{K_\infty^2} f_{\boldsymbol{g},\boldsymbol{z}}(\theta, \boldsymbol{v}) \psi(\boldsymbol{v}\cdot\boldsymbol{d}) \mathrm{d}\boldsymbol{v} = \sum_{\boldsymbol{s}\in\mathcal{O}^2} f_{\boldsymbol{g},\boldsymbol{z}}(\theta, \boldsymbol{s}).$$

We make the change of variables $\boldsymbol{j} = \boldsymbol{g} + r\boldsymbol{s}$ and the substitution $\boldsymbol{z} = P^{-1}\boldsymbol{t}$ in order to obtain

$$\sum_{\boldsymbol{c}}{}^{\text{spec}} S_r(\boldsymbol{c}) I_r(\boldsymbol{c}) = |r|^2 |P|^{-4} \sum_{\rho_i} \sum_{\boldsymbol{j}\in\mathcal{O}^2} T_r(\boldsymbol{j}) J_r(\boldsymbol{j}, \theta),$$

where

$$T_r(\boldsymbol{j}) = \sum_{|a|<|r|}{}' \sum_{|\boldsymbol{h}|<|r|} \psi\left(\frac{a\widetilde{F}(\boldsymbol{j},\boldsymbol{h})}{r}\right),$$

and

$$J_r(\boldsymbol{j}, \theta) = \int_{K_\infty^2} w(P^{-1}\boldsymbol{x}(\boldsymbol{j},\boldsymbol{t})) \psi(\theta \widetilde{F}(\boldsymbol{j},\boldsymbol{t})) \mathrm{d}\boldsymbol{t}.$$

Further we will write

$$J_r(\boldsymbol{j}) := \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} J_r(\boldsymbol{j}, \theta) \mathrm{d}\theta.$$

We can summarise our findings until now as follows.

Lemma 3.8.2. *We have*

$$|P|^4 \sum_{\substack{r \text{ monic} \\ |r|\le\widehat{Q}}} |r|^{-4} \sum_{\boldsymbol{c}}{}^{\text{spec}} S_r(\boldsymbol{c}) I_r(\boldsymbol{c}) = \sum_{\rho_i} \sum_{\substack{r \text{ monic} \\ |r|\le\widehat{Q}}} |r|^{-2} \sum_{\boldsymbol{j}\in\mathcal{O}^2} T_r(\boldsymbol{j}) J_r(\boldsymbol{j}) + O(|P|^{3/2+\varepsilon}). \tag{3.8.4}$$

We now follow a strategy that is very similar to the usual delta method. The main term will come from $\boldsymbol{j} = \boldsymbol{0}$ and it then remains to estimate $T_r(\boldsymbol{j})$ and $J_r(\boldsymbol{j}, \theta)$ for $\boldsymbol{j} \neq \boldsymbol{0}$.

3.8.2 The main term

Lemma 3.8.3. *For all $P \in \mathcal{O} \setminus \{0\}$ we have*

$$\sum_{\rho_i} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-2} T_r(\mathbf{0}) J_r(\mathbf{0}) = \sum_{\boldsymbol{x}}^{\text{line}} w(P^{-1}\boldsymbol{x}) + O(1).$$

Proof. Since $\widetilde{F}(\mathbf{0}, \boldsymbol{z}) = 0$ for all $\boldsymbol{z} \in K_\infty^2$ we have

$$T_r(\mathbf{0}) = \sum_{|a|<|r|}{}' |r|^2,$$

and

$$J_r(\mathbf{0}, \theta) = \int_{K_\infty^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{t})) \mathrm{d}\boldsymbol{t}.$$

Therefore, the term arising from $\boldsymbol{j} = \mathbf{0}$ on the right hand side of (3.8.4) is equal to

$$\sum_{\rho_i} \int_{K_\infty^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{t})) \mathrm{d}\boldsymbol{t} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} \sum_{|a|<|r|}{}' \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \mathrm{d}\theta.$$

But from Dirichlet's approximation theorem (3.2.2) we see

$$\sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} \sum_{|a|<|r|}{}' \int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \mathrm{d}\theta = \mu(\mathbb{T}) = 1.$$

Further, it is easily seen that

$$\sum_{\boldsymbol{x}}^{\text{line}} w(P^{-1}\boldsymbol{x}) = \sum_{\rho_i} \sum_{\boldsymbol{z} \in \mathcal{O}^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{z})).$$

But since $K_\infty^2 = \bigsqcup_{\boldsymbol{z} \in \mathcal{O}^2} (\boldsymbol{z} + \mathbb{T})$ we have

$$\int_{K_\infty^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{t})) \mathrm{d}\boldsymbol{t} = \sum_{\boldsymbol{z} \in \mathcal{O}^2} \int_{\mathbb{T}^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{z} + \boldsymbol{\alpha})) \mathrm{d}\boldsymbol{\alpha}.$$

If $\boldsymbol{z} \in \mathcal{O} \setminus \{\mathbf{0}\}$ then $|\boldsymbol{x}(\mathbf{0}, \boldsymbol{z} + \boldsymbol{\alpha})| = |\boldsymbol{x}(\mathbf{0}, \boldsymbol{z})|$ for all $\boldsymbol{\alpha} \in \mathbb{T}^2$ and so

$$\int_{\mathbb{T}^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{z} + \boldsymbol{\alpha})) \mathrm{d}\boldsymbol{\alpha} = w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{z}))$$

for such $\boldsymbol{z}$. We also clearly have $\int_{\mathbb{T}^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{\alpha})) \mathrm{d}\boldsymbol{\alpha} \ll 1$ and so

$$\int_{K_\infty^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{t})) \mathrm{d}\boldsymbol{t} = \sum_{\boldsymbol{z} \in \mathcal{O}^2} w(P^{-1}\boldsymbol{x}(\mathbf{0}, \boldsymbol{z})) + O(1),$$

whence the Lemma follows. □

3.8.3 Estimating the error term

In this section we make a choice of $\rho_1, \dots, \rho_4$ and bound the contribution made from terms such that $\boldsymbol{j} \neq \mathbf{0}$. Once we showed the desired bound for a particular choice, Lemma 3.8.1 will follow since there are only $O(1)$ different possibilities for ρ_i.

We begin by bounding $J_r(\boldsymbol{j})$ where $\boldsymbol{j} \neq \mathbf{0}$. Note first that $w(P^{-1}(\boldsymbol{x}(\boldsymbol{j}, \boldsymbol{t}))) = 0$ if $\boldsymbol{j} \gg |P|$ and so $J_r(\boldsymbol{j}) = 0$ if $\boldsymbol{j} \gg |P|$. Further this allows us to exchange the integral over θ with the sum over $\boldsymbol{j}$ in (3.8.4). Note further from (3.2.3) that we have

$$\int_{|\theta|<|r|^{-1}\widehat{Q}^{-1}} \psi(\theta \widetilde{F}(\boldsymbol{j}, \boldsymbol{t})) \mathrm{d}\theta = \begin{cases} |r|^{-1}\widehat{Q}^{-1}, & \text{if } |\widetilde{F}(\boldsymbol{j}, \boldsymbol{t})| < |r|\widehat{Q} \\ 0, & \text{otherwise.} \end{cases}$$

Thus we find

$$J_r(\boldsymbol{j}) \ll \mu(\boldsymbol{j}, r)|r|^{-1}\widehat{Q}^{-1},$$

where

$$\mu(\boldsymbol{j}, r) = \operatorname{meas}\left(\left\{\boldsymbol{t} \in K_\infty^2 \colon |\boldsymbol{t}| \ll |P|,\ |\widetilde{F}(\boldsymbol{j}, \boldsymbol{t})| < |r|\widehat{Q}\right\}\right).$$

To estimate this measure we simplify the expressions involved by making the substitution

$$u_1 = 2\rho_1\rho_2 t_1 - (\rho_1\rho_2' + \rho_1'\rho_2)j_1, \quad u_2 = 2\rho_3\rho_4 t_2 - (\rho_3\rho_4' + \rho_3'\rho_4)j_2.$$

After this linear change of variables $\widetilde{F}$ takes the form

$$\widetilde{G}(\boldsymbol{j}, \boldsymbol{u}) = \lambda j_1(3u_1^2 + j_1^2) + \mu j_2(3u_2^2 + j_2^2).$$

Since the change of variables is linear of constant, non-vanishing Jacobian it is sufficient to consider

$$\mu_{\widetilde{G}}(\boldsymbol{j}, r) := \operatorname{meas}\left(\left\{\boldsymbol{u} \in K_\infty^2 \colon |\boldsymbol{u}| \ll |P|,\ |\widetilde{G}(\boldsymbol{j}, \boldsymbol{u})| < |r|\widehat{Q}\right\}\right).$$

If $j_2 = 0$ then using Lemma 3.3.5 it is easily seen that

$$\mu_{\widetilde{G}}(\boldsymbol{j}, r) \ll |P| \left(\frac{|r|\widehat{Q}}{|j_1|}\right)^{1/2},$$

and similarly if $j_1 = 0$. So assume $j_1 j_2 \neq 0$. In this case, note that we have

$$\mu_{\widetilde{G}}(\boldsymbol{u}, r) \ll \sum_{k,m=-\infty}^{\log_q |P|} \sum_{\substack{U_1 = q^k \\ U_2 = q^m}} \mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2),$$

where

$$\mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2) = \mathrm{meas}\left(\left\{\boldsymbol{u} \in K_\infty^2 \colon |u_1| = U_1,\ |u_2| = U_2,\ \left|\widetilde{G}(\boldsymbol{j}, \boldsymbol{u})\right| < |r|\widehat{Q}\right\}\right).$$

In the case where U_1 or $U_2 < |P|^{-1}$ we can use the trivial bound $O(U_1U_2)$ for $\mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2)$ to deduce that the total contribution arising from such U_1, U_2 is bounded by $O(1)$. For the remaining contribution note if $\boldsymbol{u}$ satisfies $\widetilde{G}(\boldsymbol{j}, \boldsymbol{u}) = 0$ then $u_1^2 = A + O(|r|\widehat{Q}/|j_1|)$ for some function $A(j_1, j_2, u_2)$ and thus u_1 lies in a subset of measure $O(|r|\widehat{Q}/(U_1|j_1|))$. Therefore $\mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2) \ll U_2|r|\widehat{Q}/(U_1|j_1|)$. Similarly, $\mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2) \ll U_1|r|\widehat{Q}/(U_2|j_2|)$. Putting this together yields

$$\mu_{\widetilde{G}}(\boldsymbol{j}, r, U_1, U_2) \ll |r|\widehat{Q}|j_1j_2|^{-1/2}.$$

Since there are $|P|^\varepsilon$ pairs U_1, U_2 such that $|P|^{-1} \le U_1, U_2 \le |P|$ we deduce

$$\mu(\boldsymbol{j}, r) \ll 1 + |P|^\varepsilon |r|\widehat{Q}|j_1j_2|^{-1/2}.$$

We summarise our observations in the following lemma.

Lemma 3.8.4. *Let $\boldsymbol{j} \in \mathcal{O}^2 \setminus \{\boldsymbol{0}\}$ be such that $|\boldsymbol{j}| \ll |P|$. If $j_1j_2 \neq 0$, then we have*

$$J_r(\boldsymbol{j}) \ll |P|^\varepsilon |j_1j_2|^{-1/2}.$$

If $j_2 = 0$, then we have

$$J_r(\boldsymbol{j}) \ll \frac{|P|^{1/4}}{(|j_1||r|)^{1/2}}.$$

Next, we turn to estimating the exponential sums $T_r(\boldsymbol{j})$. Via the Chinese remainder theorem we have for all $r_1, r_2 \in \mathcal{O}$ such that $(r_1, r_2) = 1$ that

$$T_{r_1r_2}(\boldsymbol{j}) = T_{r_1}(\boldsymbol{j})T_{r_2}(\boldsymbol{j}). \tag{3.8.5}$$

Thus we may put our focus on $T_r(\boldsymbol{j})$ where $r = \varpi^k$ for irreducible $\varpi \in \mathcal{O}$. Note that

$$\left|\sum_{|\boldsymbol{h}|<|r|} \psi\left(\frac{a\widetilde{F}(\boldsymbol{j}, \boldsymbol{h})}{r}\right)\right| \le \left|\sum_{|h_1|<|r|} \psi\left(\frac{aj_1Q_1(j_1, h_1)}{r}\right)\right| \left|\sum_{|h_2|<|r|} \psi\left(\frac{aj_2Q_2(j_2, h_2)}{r}\right)\right|.$$

A simple Weyl differencing type of argument further yields

$$\begin{aligned}\left|\sum_{|h_1|<|r|} \psi\left(\frac{aj_1Q_1(j_1, h_1)}{r}\right)\right|^2 &= \sum_{|h|,|h_1|<|r|} \psi\left(\frac{aj_1(Q_1(j_1, h+h_1) - Q_1(j_1, h_1))}{r}\right)\\ &\ll \sum_{|h|<|r|}\left|\sum_{|h_1|<|r|} \psi\left(\frac{6a\lambda j_1\rho_1^2\rho_2^2 j_1h_1h}{r}\right)\right|\\ &= |r|\,\#\{h \in \mathcal{O}\colon |h| < |r|, r \mid 6a\lambda j_1\rho_1^2\rho_2^2 h\}\\ &\ll |r|\,|(r, 6a\lambda j_1\rho_1^2\rho_2^2 h)|\\ &\ll |r|\,|(r, j_1)|.\end{aligned}$$

We can find a similar estimate for the sum over h_2, which gives

$$T_r(\boldsymbol{j}) \ll |r|^2 |(r, j_1)|^{1/2} |(r, j_2)|^{1/2}.$$

This will be sufficient for our purposes if r is cube-full. However, for $r = \varpi$ or $r = \varpi^2$ we can do better. We begin by considering the case when $r = \varpi$ and we will further assume $\varpi \nmid (j_1, j_2)$. Note first that

$$\sum_{|a|<|\varpi|}{}' \psi\left(\frac{a\widetilde{F}(\boldsymbol{j}, \boldsymbol{h})}{\varpi}\right) = \sum_{\substack{|a|<|\varpi| \\ a\neq 0}} \psi\left(\frac{a\widetilde{F}(\boldsymbol{j}, \boldsymbol{h})}{\varpi}\right) = \begin{cases} |\varpi| - 1, & \text{if } \varpi \mid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}), \\ -1, & \text{otherwise.} \end{cases}$$

Therefore we get

$$\begin{aligned} T_\varpi(\boldsymbol{j}) &= (|\varpi| - 1)\# \left\{ |\boldsymbol{h}| < |\varpi| \colon \varpi \mid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \right\} - \# \left\{ |\boldsymbol{h}| < |\varpi| \colon \varpi \nmid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \right\} \\ &= |\varpi| \# \left\{ |\boldsymbol{h}| < |\varpi| \colon \varpi \mid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \right\} - |\varpi|^2. \end{aligned}$$

The equation $\widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \equiv 0 \mod \varpi$ may be regarded as $Q(h_1, h_2, 1)$ for a ternary quadratic form $Q(x, y, z)$. The quadratic form Q is non-singular in $\mathcal{O}/\varpi$ if $\varpi \nmid j_1 j_2 F_0(\boldsymbol{j})$, where $F_0(\boldsymbol{j}) = \lambda j_1^3 + \mu j_2^3$. Since ϖ is irreducible we have $\mathcal{O}/\varpi \cong \mathbb{F}_{|\varpi|}$ and so if $\varpi \nmid j_1 j_2 F_0(\boldsymbol{j})$ then Theorem 6.26 in [74] gives

$$\# \left\{ |\boldsymbol{h}| < |\varpi| \colon \varpi \mid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \right\} = |\varpi| + O(1).$$

We deduce $T_\varpi(\boldsymbol{j}) \ll |\varpi|$ in this case. Since $\varpi \nmid (j_1, j_2)$ the form Q does not vanish identically in $\mathcal{O}/\varpi$ and so we have

$$\# \left\{ |\boldsymbol{h}| < |\varpi| \colon \varpi \mid \widetilde{F}(\boldsymbol{j}, \boldsymbol{h}) \right\} \ll |\varpi|,$$

whence $T_\varpi(\boldsymbol{j}) \ll |\varpi|^2$ if $\varpi \mid j_1 j_2 F_0(\boldsymbol{j})$.

We now turn to analysing $T_{\varpi^2}(\boldsymbol{j})$. We assume $\varpi \nmid \lambda\mu \prod_{i=1}^5 \rho_i$. This condition affects only finitely many primes ϖ and so the estimates that we obtain under this condition hold in general by adjusting the resulting constant. Put

$$k_1 = 2\rho_1\rho_2 h_1 - (\rho_1\rho_2' + \rho_1'\rho_2) j_1, \quad \text{and} \quad k_2 = 2\rho_3\rho_4 h_2 - (\rho_3\rho_4' + \rho_3'\rho_4) j_2,$$

so that after this change of variables we have

$$\widetilde{F}(\boldsymbol{j}, \boldsymbol{k}(\boldsymbol{h})) = \frac{1}{4} F_0(\boldsymbol{j}) + \frac{3}{4}(\lambda j_1 k_1^2 + \mu j_2 k_2^2).$$

By our assumption on ϖ, as $\boldsymbol{h}$ ranges through values $|\boldsymbol{h}| < |\varpi^2|$ we also have that $\boldsymbol{k}$ ranges through $|\boldsymbol{k}| < |\varpi^2|$ under this change of variables. Hence we obtain

$$T_{\varpi^2}(\boldsymbol{j}) = \sideset{}{'}\sum_{|a|<|\varpi|^2} \psi\left(\frac{aF_0(\boldsymbol{j})}{4\varpi^2}\right) \sum_{|\boldsymbol{k}|<|\varpi|^2} \psi\left(\frac{3a(\lambda j_1 k_1^2 + \mu j_2 k_2^2)}{4\varpi^2}\right).$$

We can write $\boldsymbol{k} = \boldsymbol{u} + \varpi\boldsymbol{v}$ where $|\boldsymbol{u}|, |\boldsymbol{v}| < |\varpi|$. Then

$$\begin{aligned}\sum_{|k_i|<|\varpi|^2} \psi\left(\frac{3a\lambda j_i k_i^2}{4\varpi^2}\right) &= \sum_{|u_i|<|\varpi|} \psi\left(\frac{3a\lambda j_i u_i^2}{4\varpi^2}\right) \sum_{|v_i|<|\varpi|} \psi\left(\frac{3a\lambda j_i u_i v_i}{4\varpi^2}\right) \\ &= |\varpi| \sum_{\substack{|u_i|<|\varpi| \\ \varpi | j_i u_i}} \psi\left(\frac{3a\lambda j_i u_i^2}{2\varpi}\right),\end{aligned}$$

for $i = 1, 2$ since $\varpi \nmid a\lambda$. If $\varpi \nmid j_1 j_2$ the above expression is just $|\varpi|$ and so we get in this case

$$T_{\varpi^2}(\boldsymbol{j}) = |\varpi|^2 \sideset{}{'}\sum_{|a|<|\varpi|^2} \psi\left(\frac{aF_0(\boldsymbol{j})}{4\varpi^2}\right) = \begin{cases} 0, & \text{if } \varpi \nmid F_0(\boldsymbol{j}), \\ -|\varpi|^3 & \text{if } \varpi \parallel F_0(\boldsymbol{j}), \\ |\varpi|^4 - |\varpi|^3 & \text{if } \varpi^2 \mid F_0(\boldsymbol{j}), \end{cases}$$

and so in particular

$$T_{\varpi^2}(\boldsymbol{j}) \ll |\varpi|^2 |(\varpi^2, F_0(\boldsymbol{j}))|.$$

If, on the other hand, $\varpi \mid j_1$ we claim that $T_{\varpi^2}(\boldsymbol{j}) = 0$. Due to the standing assumption $\varpi \nmid (j_1, j_2)$ it follows that $\varpi \nmid j_2$ and thus the above gives

$$T_{\varpi^2}(\boldsymbol{j}) = |\varpi|^2 \sum_{|u_1|<|\varpi|} \sideset{}{'}\sum_{|a|<|\varpi|^2} \psi\left(\frac{a(F_0(\boldsymbol{j}) + 3\lambda j_1 u_1^2)}{4\varpi^2}\right).$$

This vanishes unless $\varpi \mid F_0(\boldsymbol{j}) + 3\lambda j_1 u_1^2$. But since $\varpi \mid j_1$ this would imply $\varpi \mid \mu j_2^3$ and hence $\varpi \mid j_2$. As we excluded this case by assumption the claim follows. We summarise our analysis of $T_r(\boldsymbol{j})$ in a lemma.

Lemma 3.8.5. *Let $\boldsymbol{j} \in \mathcal{O}^2 \setminus \{\boldsymbol{0}\}$. Then we have*

$$T_r(\boldsymbol{j}) \ll |r|^2 |(r, j_1)|^{1/2} |(r, j_2)|^{1/2}$$

for any $r \in \mathcal{O} \setminus \{0\}$. Further, if $r = \varpi$ or $r = \varpi^2$ for some irreducible $\varpi \in \mathcal{O}$ and if $\varpi \nmid (j_1, j_2)$ then we get

$$T_r(\boldsymbol{j}) \ll |r||(r, j_1 j_2 F_0(\boldsymbol{j}))|.$$

We are now finally in a position to give a sufficiently good upper bound for the right hand side of (3.8.4) and thus complete the proof of Theorem 3.1.2. For this we fix a choice of ρ_i and estimate the sum

$$\mathcal{S} := \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-2} \sum_{\substack{\boldsymbol{j} \in \mathcal{O}^2 \\ |\boldsymbol{j}| \ll |P|}} T_r(\boldsymbol{j}) J_r(\boldsymbol{j}).$$

Since there are $O(1)$ possibilities for the ρ_i's, this will be enough to show $\mathcal{S} \ll |P|^{3/2+\varepsilon}$.

We begin with the case when $j_1 j_2 F_0(\boldsymbol{j}) \neq 0$. In this situation Lemma 3.8.4 yields

$$\mathcal{S} \ll |P|^{\varepsilon} \sum_{\boldsymbol{j}} |j_1 j_2|^{-1/2} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |r|^{-2} |T_r(\boldsymbol{j})|.$$

Next we write $r = r_1 r_2$ where r_1, r_2 monic are coprime, and where r_1 is cube-free and $\varpi \mid r_1$ implies $\varpi \nmid (j_1, j_2)$. We can then factor $T_r(\boldsymbol{j})$ by (3.8.5) to obtain

$$\begin{aligned} \mathcal{S} &\ll |P|^{\varepsilon} \sum_{\boldsymbol{j}} |j_1 j_2|^{-1/2} \sum_{r_2} |r_2|^{-2} |T_{r_2}(\boldsymbol{j})| \sum_{r_1} |r_1|^{-2} |T_{r_1}(\boldsymbol{j})| \\ &\ll |P|^{\varepsilon} \sum_{\boldsymbol{j}} |j_1 j_2|^{-1/2} \sum_{r_2} |r_2|^{-2} |T_{r_2}(\boldsymbol{j})| \sum_{r_1} \frac{|(r_1, j_1 j_2 F_0(\boldsymbol{j}))|}{|r_1|}, \end{aligned}$$

where we used Lemma 3.8.5 to estimate $T_{r_1}(\boldsymbol{j})$. For the inner sum we have

$$\sum_{r_1} \frac{|(r_1, j_1 j_2 F_0(\boldsymbol{j}))|}{|r_1|} \ll |P|^{\varepsilon} |j_1 j_2 F_0(\boldsymbol{j})|^{\varepsilon} \ll |P|^{2\varepsilon},$$

since we assume $j_1 j_2 F_0(\boldsymbol{j}) \neq 0$ and in general it holds $\widehat{Y}^{-1} \sum_{|r| = \widehat{Y}} |(G, r)| \ll (|G| \widehat{Y})^{\varepsilon}$ for any $Y \in \mathbb{Z}_{\geq 0}$ and $G \in \mathcal{O}$.
Note that if $\varpi \parallel r_2$ or $\varpi^2 \parallel r_2$, then $\varpi \mid (j_1, j_2)$. In particular, if we put $\eta(r_2) = \prod \varpi$, where the product is over all $\varpi \mid r_2$ such that $\varpi \parallel r_2$ or $\varpi^2 \parallel r_2$, then we have $\boldsymbol{j} = \eta(r_2) \boldsymbol{k}$ for some $|\boldsymbol{k}| \ll |P| / |\eta(r_2)|$. It follows that

$$\begin{aligned} \mathcal{S} &\ll |P|^{\varepsilon} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} |\eta(r)|^{-1} \sum_{\substack{|\boldsymbol{k}| \ll |P| / |\eta(r)| \\ k_1 k_2 \neq 0}} \frac{|(r, \eta(r) k_1)|^{1/2} |(r, \eta(r) k_2)|^{1/2}}{|k_1 k_2|^{1/2}} \\ &\ll |P|^{\varepsilon} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} \sum_{\substack{|\boldsymbol{k}| \ll |P| / |\eta(r)| \\ k_1 k_2 \neq 0}} \frac{|(r, k_1)|^{1/2} |(r, k_2)|^{1/2}}{|k_1 k_2|^{1/2}}. \end{aligned}$$

The sum over $\boldsymbol{k}$ above factors into $(\sum_k |(r,k)|^{1/2}|k|^{-1/2})^2$, which we can estimate as

$$\sum_{\substack{|k|\ll |P|/|\eta(r)| \\ k\neq 0}} \frac{|(r,k)|^{1/2}}{|k|^{1/2}} \ll \sum_{d|r} |d|^{1/2} \sum_{\substack{|k'|\ll |P|/|\eta(r)d| \\ (r,k')=1}} |k'd|^{-1/2}$$
$$\ll \sum_{d|r} |P|^{1/2}|\eta(r)|^{-1/2}.$$

Since $\sum_{d|r} 1 \ll |r|^{\varepsilon} \ll |P|^{\varepsilon}$, we thus arrive at

$$\mathcal{S} \ll |P|^{1+\varepsilon} \sum_{|r|\leq \widehat{Q}} |\eta(r)|^{-1}.$$

Next we write $r = st_1^2t_3$, where s, t_1, t_3 are pairwise coprime and monic, t_3 is cube-full and s is square-free. With this notation we clearly have $\eta(r) = st_1$ and there are at most $O(\widehat{Q}^{1/3}) = O(|P|^{1/2})$ available t_3, so that

$$\begin{aligned}\mathcal{S} &\ll |P|^{3/2+\varepsilon} \sum_{|s|\leq \widehat{Q}} |s|^{-1} \sum_{|t_1|\leq (\widehat{Q}/|s|)^{1/2}} |t_1|^{-1} \\ &\ll |P|^{3/2+\varepsilon} \sum_{|s|\leq \widehat{Q}} |s|^{-1}(\widehat{Q}/|s|)^{\varepsilon/2} \\ &\ll |P|^{3/2+\varepsilon}\widehat{Q}^{3\varepsilon/2}.\end{aligned}$$

With a new choice of ε this estimate suffices for our purpose.

Next we consider the case when $j_1j_2F_0(\boldsymbol{j}) = 0$. If $j_1j_2 \neq 0$ but $F_0(\boldsymbol{j}) = 0$, then there exist some $j, \nu_i \in \mathcal{O}$ such that $j_i = \nu_i j$. The number of possible ν_i can be estimated by $O(1)$. In this case Lemma 3.8.4 and Lemma 3.8.5 yield

$$J_r(\boldsymbol{j}) \ll |P|^{\varepsilon}|j|^{-1}, \quad \text{and} \quad T_r(\boldsymbol{j}) \ll |r|^2|(r,j)|.$$

The total contribution to $\mathcal{S}$ of such $\boldsymbol{j}$ is therefore bounded by

$$|P|^{\varepsilon} \sum_{\substack{r \text{ monic} \\ |r|\leq \widehat{Q}}} \sum_{\substack{j\ll P \\ j\neq 0}} |j|^{-1}|(r,j)| \ll |P|^{3/2+\varepsilon},$$

which is sufficient.

Finally we need to consider the case when one of $j_i = 0$. We may assume $j_2 = 0$ since the other case is analogous. Write $j_1 = j$, then the second part of Lemma 3.8.4 gives

$$J_r(\boldsymbol{j}) \ll \frac{|P|^{1/4}}{(|j||r|)^{1/2}}.$$

Combining the estimates in Lemma 3.8.5 also gives

$$T_r(\boldsymbol{j}) \ll |r|^{5/2+\varepsilon}|(j,r)|m(r)^{-1/2},$$

where $m(r) = \prod_{\varpi \| r} \varpi$. The contribution to $\mathcal{S}$ of $\boldsymbol{j}$ under consideration is therefore bounded by

$$|P|^{1/4} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} \sum_{\substack{j \ll P \\ j \neq 0}} |(j,r)||j|^{-1/2} m(r)^{-1/2}.$$

Since $\sum_{0<j\ll P}|(j,r)||j|^{-1/2} \ll q^{\varepsilon}|P|^{1/2+\varepsilon}$ we get an overall bound

$$|P|^{3/4+\varepsilon} \sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} m(r)^{-1/2}.$$

Write $r = r_1 r_2$ where r_1 is square-free and r_2 is square-full. Note that then $m(r) = r_1$ and there are at most $O\left(\left(\widehat{Q}/|r_1|\right)^{1/2}\right)$ available r_2. Hence

$$\sum_{\substack{r \text{ monic} \\ |r| \leq \widehat{Q}}} m(r)^{-1/2} \ll \widehat{Q}^{1/2} \sum_{\substack{r_1 \text{ monic} \\ |r_1| \leq \widehat{Q}}} |r_1|^{-1} \ll |P|^{3/4+\varepsilon},$$

and so the desired bound of $O(|P|^{3/2+\varepsilon})$ contributed from $\boldsymbol{j}$'s such that either $j_1 = 0$ or $j_2 = 0$ follows. Altogether, we have shown

$$\mathcal{S} \ll |P|^{3/2+\varepsilon},$$

which completes the proof of Lemma 3.8.1.

Chapter 4

Systems of forms of small bidegree

4.1 Introduction

Studying the number of rational solutions of bounded height on a system of equations is a fundamental tool in order to understand the distribution of rational points on varieties. A longstanding result by Birch [9] establishes an asymptotic formula for the number of integer points of bounded height that are solutions to a system of homogeneous forms of the same degree in a general setting, provided the number of variables is sufficiently big relative to the singular locus of the variety defined by the system of equations. This was recently improved upon by Rydin Myerson [94, 95] whenever the degree is 2 or 3. These results may be used in order to prove Manin's conjecture for certain Fano varieties, which arise as complete intersections in projective space.

Analogous to Birch's result, Schindler studied systems of bihomogeneous forms [98]. Using the hyperbola method, Schindler established Manin's conjecture for certain bihomogeneous varieties as a result [100]. The aim of this chapter is to improve Schindler's result by applying the ideas of Rydin Myerson to the bihomogeneous setting.

Consider a system of bihomogeneous forms $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = (F_1(\boldsymbol{x}, \boldsymbol{y}), \dots, F_R(\boldsymbol{x}, \boldsymbol{y}))$ with integer coefficients in variables $\boldsymbol{x} = (x_1, \dots, x_{n_1})$ and $\boldsymbol{y} = (y_1, \dots, y_{n_2})$. We assume that all of the forms have the same bidegree, which we denote by (d_1, d_2) for nonnegative integers d_1, d_2. By this we mean that for any scalars $\lambda, \mu \in \mathbb{C}$ we have

$$F_i(\lambda \boldsymbol{x}, \mu \boldsymbol{y}) = \lambda^{d_1} \mu^{d_2} F_i(\boldsymbol{x}, \boldsymbol{y}),$$

for all $i = 1, \dots, R$. This system defines a biprojective variety $V \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$. One can also interpret the system in the affine variables $(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2})$ and thus $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y})$ also defines an affine variety which we will denote by $V_0 \subset \mathbb{A}_{\mathbb{Q}}^{n_1+n_2}$.

We are interested in studying the set of integer solutions to this system of bihomogeneous equations. Consider two boxes $\mathcal{B}_i \subset [-1,1]^{n_i}$ where each edge is of side length at most one and they are all parallel to the coordinate axes. In order to study the questions from an analytic point of view, for $P_1, P_2 > 1$ we define the following counting function

$$N(P_1,P_2) = \#\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{Z}^{n_1} \times \mathbb{Z}^{n_2} \mid \boldsymbol{x}/P_1 \in \mathcal{B}_1,\ \boldsymbol{y}/P_2 \in \mathcal{B}_2,\ \boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}\}.$$

Generalising the work of Birch [9], Schindler [98] used the circle method to achieve an asymptotic formula for $N(P_1,P_2)$ as $P_1, P_2 \to \infty$ provided certain conditions on the number of variables are satisfied, as we shall describe below. Before we can state Schindler's result, consider the varieties V_1^* and V_2^* in $\mathbb{A}_{\mathbb{Q}}^{n_1+n_2}$ to be defined by the equations

$$\operatorname{rank}\left(\frac{\partial F_i}{\partial x_j}\right)_{i,j} < R, \quad \text{and} \quad \operatorname{rank}\left(\frac{\partial F_i}{\partial y_j}\right)_{i,j} < R$$

respectively. Assume that V_0 is a complete intersection, which means that $\dim V_0 = n_1+n_2-R$. Write $b = \max\left\{\frac{\log(P_1)}{\log(P_2)}, 1\right\}$ and $u = \max\left\{\frac{\log(P_2)}{\log(P_1)}, 1\right\}$. If $n_i > R$ and

$$n_1+n_2-\dim V_i^* > 2^{d_1+d_2-2}\max\{R(R+1)(d_1+d_2-1), R(bd_1+ud_2)\}, \qquad (4.1.1)$$

is satisfied, for $i=1,2$ then Schindler showed the asymptotic formula

$$N(P_1,P_2) = \sigma P_1^{n_1-Rd_1}P_2^{n_2-Rd_2} + O\left(P_1^{n_1-Rd_1}P_2^{n_2-Rd_2}\min\{P_1,P_2\}^{-\delta}\right), \qquad (4.1.2)$$

for some $\delta > 0$ and where σ is positive if the system $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}$ has a smooth p-adic zero for all primes p, and the variety V_0 has a smooth real zero in $\mathcal{B}_1 \times \mathcal{B}_2$.

In the case when the equations $F_1(\boldsymbol{x},\boldsymbol{y}),\dots,F_R(\boldsymbol{x},\boldsymbol{y})$ define a smooth complete intersection V, and where the bidegree is $(1,1)$ or $(2,1)$ the goal of this chapter is to improve the restriction on the number of variables (4.1.1) and still show (4.1.2).

The result by Schindler generalises a well-known result by Birch [9], which deals with systems of homogeneous equations; Let $\mathcal{B} \subset [-1,1]^n$ be a box containing the origin with side lengths at most 1 and edges parallel to the coordinate axes. Given homogeneous equations $G_1(\boldsymbol{x}),\dots,G_R(\boldsymbol{x})$ with rational coefficients of common degree d define the counting function

$$N(P) = \#\{\boldsymbol{x} \in \mathbb{Z}^n \colon \boldsymbol{x}/P \in \mathcal{B},\ G_1(\boldsymbol{x}) = \cdots = G_R(\boldsymbol{x}) = 0\}.$$

Write $V^* \subset \mathbb{A}_{\mathbb{Q}}^n$ for the variety defined by

$$\operatorname{rank}\left(\frac{\partial G_i}{\partial x_j}\right)_{i,j} < R,$$

commonly referred to as the *Birch singular locus*. Assuming that $G_1, \dots, G_R$ define a complete intersection $X \subset \mathbb{P}_{\mathbb{Q}}^{n-1}$ and that the number of variables satisfies

$$n - \dim V^* > R(R+1)(d-1)2^{d-1}, \tag{4.1.3}$$

then Birch showed

$$N(P) = \tilde{\sigma} P^{n-dR} + O(P^{n-dR-\varepsilon}), \tag{4.1.4}$$

where $\tilde{\sigma} > 0$ if the system $\boldsymbol{G}(\boldsymbol{x})$ has a smooth p-adic zero for all primes p and the variety X has a smooth real zero in $\mathcal{B}$.

Building on ideas of Müller [79, 80] on quadratic Diophantine inequalities, Rydin Myerson improved Birch's theorem. He weakened the assumption on the number of variables in the cases $d = 2, 3$ [94, 95] whenever R is reasonably large. Assuming that $X \subset \mathbb{P}_{\mathbb{Q}}^{n-1}$ defines a complete intersection, he was able to replace the condition in (4.1.3) by

$$n - \sigma_{\mathbb{R}} > d2^d R, \tag{4.1.5}$$

where

$$\sigma_{\mathbb{R}} = 1 + \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}} \dim \operatorname{Sing} \mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{G}),$$

and where $\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{G})$ is the pencil defined by $\sum_{i=1}^R \beta_i G(\boldsymbol{x})$ in $\mathbb{P}_{\mathbb{Q}}^{n-1}$. We note at this point that several other authors have replaced the Birch singular locus condition with weaker assumptions, such as Schindler [99] and Dietmann [33] who also considered dimensions of pencils, and very recently Yamagishi [119] who replaced the Birch singular locus with a condition regarding the Hessian of the system. Returning to Rydin Myerson's result if X is non-singular then one can show

$$\sigma_{\mathbb{R}} \leq R - 1$$

and in this case if $n \geq (d2^d + 1)R$ then one obtains the desired asymptotic. Notably, the work of Rydin Myerson showed the number of variables n thus only has to grow linearly in the number of equations R, whereas R appeared quadratically in Birch's work. If $d \geq 4$ he showed that for *generic* systems of forms it suffices to assume (4.1.5) for the asymptotic (4.1.4) to hold. Generic here means that the set of coefficients is required to lie in some non-empty Zariski open subset of the parameter space of coefficients of the equations.

Our goal in this chapter is to generalise the results obtained by Rydin Myerson to the case of bihomogeneous varieties whenever the bidegree of the forms is $(1, 1)$ or $(2, 1)$. Those two cases correspond to degrees 2 and 3 in the homogeneous case, respectively. We call a bihomogeneous form *bilinear* if the bidegree is $(1,1)$. Given a bilinear form $F_i(\boldsymbol{x},\boldsymbol{y})$ we may write it as

$$F_i(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{y}^T A_i \boldsymbol{x},$$

for some $n_2 \times n_1$-dimensional matrices A_i with rational entries. Given $\boldsymbol{\beta} \in \mathbb{R}^R$ write

$$A_{\boldsymbol{\beta}} = \sum_{i=1}^{R} \beta_i A_i.$$

Regarding $A_{\boldsymbol{\beta}}$ as a map $\mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ and and $A_{\boldsymbol{\beta}}^T$ as a map $\mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$ we define the quantities

$$\sigma_{\mathbb{R}}^{(1)} := \max_{\boldsymbol{\beta}\in\mathbb{R}^R\setminus\{\boldsymbol{0}\}} \dim\ker(A_{\boldsymbol{\beta}}), \quad \text{and} \quad \sigma_{\mathbb{R}}^{(2)} := \max_{\boldsymbol{\beta}\in\mathbb{R}^R\setminus\{\boldsymbol{0}\}} \dim\ker(A_{\boldsymbol{\beta}}^T).$$

We state our first theorem for systems of bilinear forms. Since the situation is completely symmetric with respect to the $\boldsymbol{x}$ and $\boldsymbol{y}$ variables if the forms are bilinear, we may without loss of generality assume $P_1 \geq P_2$ in the counting function, and still obtain the full result.

Theorem 4.1.1. *Let $F_1(\boldsymbol{x},\boldsymbol{y}),\dots,F_R(\boldsymbol{x},\boldsymbol{y})$ be bilinear forms with integer coefficients such that the biprojective variety $\mathbb{V}(F_1,\dots,F_R) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ is a complete intersection. Let $P_1 \geq P_2 > 1$, write $b = \frac{\log(P_1)}{\log(P_2)}$ and assume further that*

$$n_i - \sigma_{\mathbb{R}}^{(i)} > (2b+2)R \tag{4.1.6}$$

holds for $i = 1,2$. Then there exists some $\delta > 0$ depending at most on b, $\boldsymbol{F}$, R and n_i such that

$$N(P_1,P_2) = \sigma P_1^{n_1-R} P_2^{n_2-R} + O(P_1^{n_1-R} P_2^{n_2-R-\delta})$$

holds, where $\sigma > 0$ if the system $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}$ has a smooth p-adic zero for all primes p and if the variety V_0 has a smooth real zero in $\mathcal{B}_1 \times \mathcal{B}_2$.

Moreover, if we assume $\mathbb{V}(F_1,\dots,F_R) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ to be smooth the same conclusions hold if we assume

$$\min\{n_1,n_2\} > (2b+2)R \quad \textit{and} \quad n_1 + n_2 > (4b+5)R$$

instead of (4.1.6).

We now move on to systems of forms $F_1(\boldsymbol{x},\boldsymbol{y}),\dots,F_R(\boldsymbol{x},\boldsymbol{y})$ of bidegree $(2,1)$. We may write such a form $F_i(\boldsymbol{x},\boldsymbol{y})$ as

$$F_i(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^T H_i(\boldsymbol{y}) \boldsymbol{x},$$

where $H_i(\boldsymbol{y})$ is a symmetric $n_1 \times n_1$ matrix whose entries are linear forms in the variables $\boldsymbol{y} = (y_1, \dots, y_{n_2})$. Similarly to above, given $\boldsymbol{\beta} \in \mathbb{R}^R$ we write

$$H_{\boldsymbol{\beta}}(\boldsymbol{y}) = \sum_{i=1}^{R} \beta_i H_i(\boldsymbol{y}).$$

Given $\ell \in \{1, \dots, n_2\}$ write $\boldsymbol{e}_\ell \in \mathbb{R}^{n_2}$ for the standard unit basis vectors. Write

$$\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} = \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_1)\boldsymbol{x}, \dots, \boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_{n_2})\boldsymbol{x}) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1}$$

for this intersection of pencils, and define

$$s_{\mathbb{R}}^{(1)} := 1 + \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}} \dim \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2}. \tag{4.1.7}$$

Further write $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})$ for the biprojective variety defined by the system of equations

$$\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) = \mathbb{V}((H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})_1, \dots, (H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})_{n_1}) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$$

and define

$$s_{\mathbb{R}}^{(2)} := \left\lfloor \frac{\max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})}{2} \right\rfloor + 1, \tag{4.1.8}$$

where $\lfloor x \rfloor$ denotes the largest integer m such that $m \leq x$.

Theorem 4.1.2. *Let $F_1(\boldsymbol{x}, \boldsymbol{y}), \dots, F_R(\boldsymbol{x}, \boldsymbol{y})$ be bihomogeneous forms with integer coefficients of bidegree $(2,1)$ such that the biprojective variety $\mathbb{V}(F_1, \dots, F_R) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ is a complete intersection. Let $P_1, P_2 > 1$ be real numbers. Write $b = \max\left\{\frac{\log(P_1)}{\log(P_2)}, 1\right\}$ and $u = \max\left\{\frac{\log(P_2)}{\log(P_1)}, 1\right\}$. Assume further that*

$$n_1 - s_{\mathbb{R}}^{(1)} > (8b + 4u)R \quad \text{and} \quad \frac{n_1 + n_2}{2} - s_{\mathbb{R}}^{(2)} > (8b + 4u)R \tag{4.1.9}$$

is satisfied. Then there exists some $\delta > 0$ depending at most on b, u, R, n_i and $\boldsymbol{F}$ such that

$$N(P_1, P_2) = \sigma P_1^{n_1-2R} P_2^{n_2-R} + O(P_1^{n_1-2R} P_2^{n_2-R} \min\{P_1, P_2\}^{-\delta}) \tag{4.1.10}$$

holds, where $\sigma > 0$ if the system $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ has a smooth p-adic zero for all primes p, and if the variety V_0 has a smooth real zero in $\mathcal{B}_1 \times \mathcal{B}_2$.

If we assume that $\mathbb{V}(F_1, \dots, F_R) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ is smooth, then the same conclusions hold if we assume

$$n_1 > (16b + 8u + 1)R, \quad \text{and} \quad n_2 > (8b + 4u + 1)R \tag{4.1.11}$$

instead of (4.1.9).

We remark that we preferred to give conditions in terms of the geometry of the variety regarded as a biprojective variety, as opposed to an affine variety. The reason for this is the potential application of this result to proving Manin's conjecture for this variety, which will be addressed in due course.

Compared to the result by Schindler we thus basically remove the assumption that the number of variables needs to grow at least quadratically in R. In particular, if the complete intersection defined by the system is assumed to be smooth, then our results requires fewer variables than Schindler's provided

$$d_1 b + d_2 u < \frac{R+1}{2}$$

is satisfied, in the cases $(d_1, d_2) = (1,1)$ or $(2,1)$. In particular, if R is large this means our result provides significantly more flexibility in the choice of u and b.

One cannot hope to achieve the asymptotic formula (4.1.2) in general where a condition of the shape $n_i > R(bd_1 + ud_2)$ is not present. To see this note that the counting function satisfies

$$N(P_1, P_2) \gg P_1^{n_1} + P_2^{n_2},$$

coming from the solutions when $x_1 = \cdots = x_{n_1} = 0$ and $y_1 = \cdots = y_{n_2} = 0$. The asymptotic formula (4.1.2) thus implies

$$P_i^{n_i} \ll P_1^{n_1 - d_1 R} P_2^{n_2 - d_2 R},$$

for $i = 1, 2$. Noting that $P_1^u = P_2$ if $u > 1$ and $P_2^b = P_1$ if $b > 1$ and comparing the exponents one necessarily finds $n_i > R(bd_1 + ud_2)$.

If the forms are diagonal then one can take boxes $\mathcal{B}_i$, which avoid the coordinate axes in order to remedy this obstruction. In fact this is the approach taken by Blomer and Brüdern [10] and they proved an asymptotic formula of a system of multihomogeneous equations without a restriction on the number of variables similar to the type described above.

If the forms are not diagonal the problem still persists, even if one were to take boxes avoiding the coordinate axes. In general there may be 'bad' vectors $\boldsymbol{y}$ away from the coordinate axes such that

$$\#\{\boldsymbol{x} \in \mathbb{Z}^{n_1} \colon \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}, |\boldsymbol{x}| \leq P_1\} \gg P_1^{n_1 - a},$$

where $a < d_1 R$ for example. This is in contrast to the diagonal case, where the only vectors $\boldsymbol{y}$ where this occurs lie on at least one coordinate axis. It would be interesting to consider a modified counting function where one excludes such vectors $\boldsymbol{y}$, and analogously 'bad' vectors $\boldsymbol{x}$. In a general setting it seems difficult to control the set of such vectors. In particular, it is not clear how one would deal with the Weyl differencing step if one were to consider such a counting function.

4.1.1 Manin's conjecture

Let $V \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ be a non-singular complete intersection defined by a system of forms $F_i(\boldsymbol{x}, \boldsymbol{y})$, $i = 1, \dots, R$ of common bidegree (d_1, d_2). Assume $n_i > d_i R$ so that V is a Fano variety, which means that the inverse of the canonical bundle in the Picard group, the *anticanonical bundle*, is very ample. For a field K, write $V(K)$ for the set of K-rational points of V. In the context of Manin's conjecture we define this to be the set of K-morphisms

$$\mathrm{Spec}(K) \to V_K,$$

where V_K denotes the base change of V to the field K. For a subset $U(\mathbb{Q}) \subset V(\mathbb{Q})$ and $P \geq 1$ consider the counting function

$$N_U(P) = \#\left\{(\boldsymbol{x}, \boldsymbol{y}) \in U(\mathbb{Q}) \colon H(\boldsymbol{x}, \boldsymbol{y}) \leq P\right\},$$

where $H(\cdot, \cdot)$ is the *anticanonical height* induced by the anticanonical bundle and a choice of global sections. In our case one such height may be explicitly given as follows. If $(\boldsymbol{x}, \boldsymbol{y}) \in U(\mathbb{Q})$ we may pick representatives $\boldsymbol{x} \in \mathbb{Z}^{n_1}$, and $\boldsymbol{y} \in \mathbb{Z}^{n_2}$ such that $(x_1, \dots, x_{n_1}) = (y_1, \dots, y_{n_2}) = 1$ and we define

$$H(\boldsymbol{x}, \boldsymbol{y}) = \left(\max_i |x_i|\right)^{n_1 - Rd_1} \left(\max_i |y_i|\right)^{n_2 - Rd_2}.$$

Manin's Conjecture in this context states that, provided V is a Fano variety such that $V(\mathbb{Q}) \subset V$ is Zariski dense, there exists a subset $U(\mathbb{Q}) \subset V(\mathbb{Q})$ where $(V \setminus U)(\mathbb{Q})$ is a *thin* set such that

$$N_U(P) \sim cP(\log P)^{\rho - 1},$$

where ρ is the Picard rank of the variety V and c is a constant as predicted and interpreted by Peyre [84]. We briefly recall the definition of a thin set, according to Serre [103]. First recall a set $A \subset V(K)$ is of type

(C_1) if $A \subseteq W(K)$, where $W \subsetneq V$ is Zariski closed,

(C_2) if $A \subseteq \pi(V'(K))$, where V' is irreducible such that $\dim V = \dim V'$, where $\pi \colon V' \to V$ is a generically finite morphism of degree at least 2.

Now a subset of the K-rational points of V is *thin* if it is a finite union of sets of type (C_1) or (C_2). Originally Batyrev–Manin [5] conjectured that it suffices to assume that $(V \setminus U)$ is Zariski closed, but there have been found various counterexamples to this, the first one being due to Batyrev–Tschinkel [6].

In [100] Schindler showed an asymptotic formula of the shape above, if V is smooth and $d_1, d_2 \geq 2$ and

$$n_i > 3 \cdot 2^{d_1+d_2} d_1 d_2 R^3 + R$$

is satisfied for $i = 1, 2$. If $R = 1$ she moreover verified that the constant obtained agrees with the one predicted by Peyre, and thus proved Manin's conjecture for bihomogeneous hypersurfaces when the conditions above are met. The proof uses the asymptotic (4.1.2) established in [98] along with uniform counting results on fibres. That is, for a vector $\boldsymbol{y} \in \mathbb{Z}^{n_2}$ one may consider the counting function

$$N_{\boldsymbol{y}}(P) = \# \left\{ \boldsymbol{x} \in \mathbb{Z}^{n_1} \colon \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}, |\boldsymbol{x}| \leq P \right\},$$

and to understand its asymptotic behaviour uniformly means to understand the dependence of $\boldsymbol{y}$ on the constant in the error term. Similarly she considered $N_{\boldsymbol{x}}(P)$ for 'good' $\boldsymbol{x}$ and combined the three resulting estimates to obtain an asymptotic formula for the number of solutions $\widetilde{N}(P_1, P_2)$ to the system $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$, where $|\boldsymbol{x}| \leq P_1$, $|\boldsymbol{y}| \leq P_2$ and $\boldsymbol{x}, \boldsymbol{y}$ are 'good'. Considering only 'good' tuples essentially removes a closed subset from V, and thus, after an application of a slight modification of the hyperbola method developed as in [10] she obtained an asymptotic formula for $N_U(P)$ of the desired shape.

One can be hopeful that the result established in Theorem 4.1.2 is useful in verifying Manin's Conjecture for V, when $(d_1, d_2) = (2, 1)$ in fewer variables than would be expected using Schindler's method as described above. Further, since the Picard rank of V is strictly greater than 1, it would be interesting to consider the *all heights approach* as suggested by Peyre [86, Question V.4.8]. As noted by Peyre himself, in the case when a variety has Picard rank 1, the answer to his Question 4.8 follows provided one can prove Manin's conjecture with respect to the height function induced by the anticanonical bundle.

Schindler's results have been improved upon in a few special cases. Browning and Hu showed Manin's conjecture in the case of smooth biquadratic hypersurfaces in $\mathbb{P}_{\mathbb{Q}}^{n-1} \times \mathbb{P}_{\mathbb{Q}}^{n-1}$ if the number of variables satisfies $n > 35$. If the bidegree is $(2, 1)$ then Hu showed that $n > 25$ suffices in order to obtain Manin's conjecture. Systems of bilinear varieties are flag varieties and thus Manin's conjecture follows from the result for flag varieties, which was proven by Franke, Manin and Tschinkel [37] using the theory of

Eisenstein series. In the special case when the variety is defined by $\sum_{i=0}^{s} x_i y_i = 0$ then Robbiani [90] showed how one may use the circle method to establish Manin's conjecture if $s \geq 3$, which was later improved to $s \geq 2$ by Spencer [106].

Conventions

The symbol $\varepsilon > 0$ is an arbitrarily small value, which we may redefine whenever convenient, as is usual in analytic number theory. Given forms g_ℓ, $\ell = 1, \ldots, k$ we write $\mathbb{V}(g_\ell)_{\ell=1,\ldots,k}$ or sometimes just $\mathbb{V}(g_\ell)_\ell$ for the intersection $\mathbb{V}(g_1, \ldots, g_k)$. Further, we may sometimes consider a vector of forms $\boldsymbol{h} = (h_1, \ldots, h_k)$ and we similarly write $\mathbb{V}(\boldsymbol{h})$ for the intersection $\mathbb{V}(h_1, \ldots, h_k)$.

For a real number $x \in \mathbb{R}$ we will write $e(x) = e^{2\pi i x}$. We will use Vinogradov's notation $O(\cdot)$ and $\ll$.

We shall repeatedly use the convention that the dimension of the empty set -1.

4.2 Multilinear forms

Both Theorem 4.1.1 and Theorem 4.1.2 follow from a more general result. If we have control over the number of 'small' solutions to the associated linearised forms then we can show that the asymptotic (4.1.2) holds. More explicitly, given a bihomogeneous form $F(\boldsymbol{x}, \boldsymbol{y})$ with integer coefficients of bidegree (d_1, d_2) for positive integers d_1, d_2, we may write it as

$$F(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} F_{\boldsymbol{j},\boldsymbol{k}} x_{j_1} \cdots x_{j_{d_1}} y_{k_1} \cdots y_{k_{d_2}},$$

where the coefficients $F_{\boldsymbol{j},\boldsymbol{k}} \in \mathbb{Q}$ are symmetric in $\boldsymbol{j}$ and $\boldsymbol{k}$. We define the associated multilinear form

$$\Gamma_F(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) := d_1! d_2! \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} F_{\boldsymbol{j},\boldsymbol{k}} x_{j_1}^{(1)} \cdots x_{j_{d_1}}^{(d_1)} y_{k_1}^{(1)} \cdots y_{k_{d_2}}^{(d_2)},$$

where $\widetilde{\boldsymbol{x}} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d_1)})$ and $\widetilde{\boldsymbol{y}} = (\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(d_2)})$ for vectors $\boldsymbol{x}^{(i)}$ of n_1 variables and vectors $\boldsymbol{y}^{(i)}$ of n_2 variables. Write further $\widehat{\boldsymbol{x}} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d_1-1)})$ and $\widehat{\boldsymbol{y}} = (\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(d_2-1)})$. Given $\boldsymbol{\beta} \in \mathbb{R}^R$ we define the auxiliary counting function $N_1^{\text{aux}}(\boldsymbol{\beta}; B)$ to be the number of integer vectors satisfying $\widehat{\boldsymbol{x}} \in (-B, B)^{(d_1-1)n_1}$ and $\widetilde{\boldsymbol{y}} \in (-B, B)^{d_2 n_2}$ such that

$$|\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widetilde{\boldsymbol{y}})| < \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty B^{d_1+d_2-2},$$

for $\ell = 1, \ldots, n_1$ where $\|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty := \frac{1}{d_1! d_2!} \max_{\boldsymbol{j},\boldsymbol{k}} \left| \frac{\partial^{d_1+d_2}(\boldsymbol{\beta}\cdot\boldsymbol{F})}{\partial x_{j_1} \cdots \partial x_{j_{d_1}} \partial y_{k_1} \cdots \partial y_{k_{d_2}}} \right|$. We define $N_2^{\text{aux}}(\boldsymbol{\beta}; B)$ analogously.

The technical core of this chapter is the following theorem.

Theorem 4.2.1. *Assume $n_1, n_2 > (d_1+d_2)R$ and let $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = (F_1(\boldsymbol{x}, \boldsymbol{y}), \ldots, F_R(\boldsymbol{x}, \boldsymbol{y}))$ be a system of bihomogeneous forms with integer coefficients of common bidegree (d_1, d_2) such that the variety $\mathbb{V}(\boldsymbol{F}) \subset \mathbb{P}_{\mathbb{Q}}^{n_1-1} \times \mathbb{P}_{\mathbb{Q}}^{n_2-1}$ is a complete intersection. Write $b = \max\{\log(P_1)/\log(P_2), 1\}$ and $u = \max\{\log(P_2)/\log(P_1), 1\}$.*

Assume there exist $C_0 \geq 1$ and $\mathscr{C} > (bd_1 + ud_2)R$ such that for all $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}$ and all $B > 0$ we have

$$N_i^{\text{aux}}(\boldsymbol{\beta}; B) \leq C_0 B^{d_1 n_1 + d_2 n_2 - n_i - 2^{d_1+d_2-1}\mathscr{C}} \tag{4.2.1}$$

for $i = 1, 2$. There exists some $\delta > 0$ depending on b, u, C_0, R, d_i and n_i such that

$$N(P_1, P_2) = \sigma P_1^{n_1 - d_1 R} P_2^{n_2 - d_2 R} + O\left(P_1^{n_1-d_1R} P_2^{n_2-d_2R} \min\{P_1, P_2\}^{-\delta}\right).$$

The factor $\sigma = \mathfrak{I}\mathfrak{S}$ is the product of the singular integral $\mathfrak{I}$ and the singular series $\mathfrak{S}$, as defined in (4.5.26) *and* (4.5.23)*, respectively. Moreover, if the system $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ has a non-singular real zero in $\mathcal{B}_1 \times \mathcal{B}_2$ and a non-singular p-adic zero for every prime p, then $\sigma > 0$.*

While showing that (4.2.1) holds is rather straightforward when the bidegree is $(1, 1)$ it becomes significantly more difficult when the bidegree increases. In fact, in Rydin Myerson's work a similar upper bound on a similar auxiliary counting function needs to be shown. He is successful in doing so when the degree is 2 or 3 and the system defines a complete intersection, but for higher degrees he was only able to show this upper bound for generic systems.

Our strategy is as follows. We will establish Theorem 4.2.1 in Section 4.4 and Section 4.5 and then use this to show Theorem 4.1.1 and Theorem 4.1.2 in Section 4.6 and in Section 4.7.

4.3 Geometric preliminaries

The following Lemma is taken from [100].

Lemma 4.3.1 (Lemma 2.2 in [100])**.** *Let W be a smooth variety that is complete over some algebraically closed field and consider a closed irreducible subvariety $Z \subseteq W$ such that* $\dim Z \geq 1$. *Given an effective divisor D on W then the dimension of every* *irreducible component of* $D \cap Z$ *is at least* $\dim Z - 1$*. If* D *is moreover ample we have in addition that* $D \cap Z$ *is nonempty.*

In particular the following corollary will be very useful.

Corollary 4.3.2. *Let* $V \subseteq \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ *be a closed variety such that* $\dim V \geq 1$*. Consider* $H = \mathbb{V}(f)$ *where* $f(\boldsymbol{x}, \boldsymbol{y})$ *is a polynomial of bidegree at least* $(1,1)$ *in the variables* $(\boldsymbol{x}, \boldsymbol{y}) = (x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2})$*. Then*

$$\dim(V \cap H) \geq \dim V - 1,$$

in particular $V \cap H$ *is non-empty.*

Proof. Since the bidegree of f is at least $(1,1)$ we have that H defines an effective and ample divisor on $\mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$. We apply Lemma 4.3.1 with $W = \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$, $D = H$ and Z any irreducible component of V. □

Lemma 4.3.3. *Let* $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y})$ *be a system of* R *bihomogeneous equations of the same bidegree* (d_1, d_2) *with* $d_1, d_2 \geq 1$*. Assume that* $\mathbb{V}(\boldsymbol{F}) \subset \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ *is a smooth complete intersection. Given* $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}$ *we have*

$$\dim \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \leq R - 2,$$

where we write $\boldsymbol{\beta} \cdot \boldsymbol{F} = \sum_i \beta_i F_i$*.*

Proof. The singular locus of $\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F})$ is given by

$$\mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) = \mathbb{V}\left(\frac{\partial(\boldsymbol{\beta} \cdot \boldsymbol{F})}{\partial x_j}\right)_{j=1,\dots,n_1} \cap \mathbb{V}\left(\frac{\partial(\boldsymbol{\beta} \cdot \boldsymbol{F})}{\partial y_j}\right)_{j=1,\dots,n_2}.$$

Assume without loss of generality $\beta_R \neq 0$ so that $\mathbb{V}(\boldsymbol{F}) = \mathbb{V}(F_1, \dots, F_{R-1}, \boldsymbol{\beta} \cdot \boldsymbol{F})$. We claim that we have the following inclusion

$$\mathbb{V}(F_1, \dots, F_{R-1}) \cap \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \subseteq \mathrm{Sing}\mathbb{V}(\boldsymbol{F}). \tag{4.3.1}$$

To see this note first that $\mathbb{V}(F_1, \dots, F_{R-1}) \cap \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \subseteq \mathbb{V}(\boldsymbol{F})$. Further, the Jacobian matrix $J(\boldsymbol{F})$ of $\boldsymbol{F}$ is given by

$$J(\boldsymbol{F}) = \left(\frac{\partial F_i}{\partial z_j}\right)_{ij},$$

where $i = 1, \dots, R$ and z_j ranges through $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$. Now if the equations

$$\frac{\partial(\boldsymbol{\beta} \cdot \boldsymbol{F})}{\partial x_j} = \frac{\partial(\boldsymbol{\beta} \cdot \boldsymbol{F})}{\partial y_j} = 0,$$

are satisfied then this implies that the rows of $J(\boldsymbol{F})$ are linearly dependent. Since $\mathbb{V}(\boldsymbol{F})$ is a complete intersection we deduce the claim.

Assume now for a contradiction that $\dim \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) \geq R-1$ holds. Applying Corollary 4.3.2 $(R-1)$-times with $V = \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F})$, noting that the bidegree of F_i is at least $(1,1)$, we find

$$\mathbb{V}(F_1,\dots,F_{R-1}) \cap \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) \neq \emptyset.$$

This contradicts (4.3.1) since $\mathrm{Sing}\mathbb{V}(\boldsymbol{F}) = \emptyset$ by assumption. □

Lemma 4.3.4. *Let $n_1 \leq n_2$ be two positive integers. For $i = 1,\dots,n_2$ let $A_i \in \mathrm{M}_{n_1\times n_1}(\mathbb{C})$ be symmetric matrices. Consider the varieties $V_1 \subset \mathbb{P}_{\mathbb{C}}^{n_1-1}$ and $V_2 \subset \mathbb{P}_{\mathbb{C}}^{n_1-1}\times\mathbb{P}_{\mathbb{C}}^{n_2-1}$ defined by*

$$V_1 = \mathbb{V}(\boldsymbol{t}^T A_i \boldsymbol{t})_{i=1,\dots,n_2}$$
$$V_2 = \mathbb{V}\left(\sum_{i=1}^{n_2} y_i A_i \boldsymbol{x}\right).$$

Then we have

$$\dim V_2 \leq \dim V_1 + n_2 - 1.$$

In particular, if $V_1 = \emptyset$ then $\dim V_2 \leq n_2 - 2$.

Proof. Consider the variety $V_3 \subset \mathbb{P}_{\mathbb{C}}^{n_1-1}\times\mathbb{P}_{\mathbb{C}}^{n_1-1}$ defined by

$$V_3 = \mathbb{V}(\boldsymbol{z}^T A_i \boldsymbol{x})_{i=1,\dots,n_2}.$$

Further for $\boldsymbol{x} = (x_1,\dots,x_{n_1})^T$ consider

$$A(\boldsymbol{x}) = (A_1\boldsymbol{x}\cdots A_{n_2}\boldsymbol{x}) \in \mathrm{M}_{n_1\times n_2}(\mathbb{C})[x_1,\dots,x_{n_1}].$$

We may write $V_2 = \mathbb{V}(A(\boldsymbol{x})\boldsymbol{y})$ and $V_3 = \mathbb{V}(\boldsymbol{z}^T A(\boldsymbol{x}))$. Our first goal is to relate the dimensions of the varieties above as follows

$$\dim V_2 \leq \dim V_3 + n_2 - n_1. \tag{4.3.2}$$

For $r = 0,\dots,n_1$ define the quasi-projective varieties $D_r \subset \mathbb{P}_{\mathbb{C}}^{n_1-1}$ given by

$$D_r = \{\boldsymbol{x} \in \mathbb{P}_{\mathbb{C}}^{n_1-1} \colon \mathrm{rank}(A(\boldsymbol{x})) = r\}.$$

These are quasiprojective since they may be written as the intersection of the vanishing of all $(r+1)\times(r+1)$ minors of $A(\boldsymbol{x})$ with the complement of the vanishing of all $r\times r$ minors. For each r let

$$D_r = \bigcup_{i\in I_r} D_r^{(i)}$$

be a decomposition into finitely many irreducible components. Since $\bigcup_r D_r = \mathbb{P}_{\mathbb{C}}^{n_1-1}$ we have

$$\dim V_2 = \max_{\substack{0\leq r<n_2 \\ i\in I_r}} \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2).$$

Note that $r = n_2$ doesn't play a role here, since the intersection $(D_{n_2}^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2$ is empty. Similarly we get

$$\dim V_3 = \max_{\substack{0\leq r<n_2 \\ i\in I_r}} \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}) \cap V_3).$$

For $0 \leq r < n_2$ and $i \in I_r$ consider now the surjective projection maps

$$\pi_{2,r,i}\colon (D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2 \to D_r^{(i)},\ (\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x},$$

and

$$\pi_{3,r,i}\colon (D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}) \cap V_3 \to D_r^{(i)},\ (\boldsymbol{x}, \boldsymbol{z}) \mapsto \boldsymbol{x},$$

We note that by the way $D_r^{(i)}$ was constructed here, the fibres of both of these projection morphisms have constant dimension for fixed r. By the rank-nullity theorem we find that the dimensions of the fibres are related as follows

$$\dim \pi_{2,r,i}^{-1}(\boldsymbol{x}) = \dim \pi_{3,r,i}^{-1}(\boldsymbol{x}) + n_2 - n_1. \tag{4.3.3}$$

We claim that the morphism $\pi_{2,r,i}$ is proper. For this note that the structure morphism $\mathbb{P}_{\mathbb{C}}^{n_1-1} \to \operatorname{Spec} \mathbb{C}$ is proper whence $D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1} \to D_r^{(i)}$ must be proper too, as properness is preserved under base change. As $(D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2$ is closed inside $D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}$ the restriction $\pi_{2,r,i}$ must also be proper. By an analogous argument it follows $\pi_{3,r,i}$ is also proper.

Further note that the fibres of $\pi_{2,r,i}$ are irreducible since they define linear subspaces of $(D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2$, and similarly the fibres of $\pi_{3,r,i}$ are irreducible. Since $D_r^{(i)}$ is irreducible by construction and all the fibres have constant dimension, it follows that $(D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2$ is irreducible. Similarly $(D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}) \cap V_3$ is irreducible.

Hence all the conditions of Chevalley's upper semicontinuity theorem are satisfied [41, Théorème 13.1.3], so that for any $\boldsymbol{x} \in D_r^{(i)}$ we obtain

$$\dim \pi_{2,r,i}^{-1}(\boldsymbol{x}) = \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2) - \dim D_r^{(i)}, \tag{4.3.4}$$

and

$$\dim \pi_{3,r,i}^{-1}(\boldsymbol{x}) = \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}) \cap V_3) - \dim D_r^{(i)}. \tag{4.3.5}$$

Hence (4.3.4) and (4.3.5) together with (4.3.3) yield

$$\dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2) = \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_1-1}) \cap V_3) + n_2 - n_1.$$

Choosing r and i such that $\dim V_2 = \dim((D_r^{(i)} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}) \cap V_2)$ the claim (4.3.2) now follows.

Thus it is enough to find an upper bound for $\dim V_3$. To this end, consider the affine cones $\widetilde{V}_1 = \mathbb{V}(\boldsymbol{u}^T A_i \boldsymbol{u})_{i=1,\dots,n_2} \subset \mathbb{A}_{\mathbb{C}}^{n_1}$ and $\widetilde{V}_3 = \mathbb{V}(\boldsymbol{x}^T A(\boldsymbol{z})) \subset \mathbb{A}_{\mathbb{C}}^{n_1} \times \mathbb{A}_{\mathbb{C}}^{n_1}$. Note in particular, that $\widetilde{V}_1 \neq \emptyset$ even if $V_1 = \emptyset$.

Write $\widetilde{\Delta} \subset \mathbb{A}_{\mathbb{C}}^{n_1} \times \mathbb{A}_{\mathbb{C}}^{n_1}$ for the diagonal given by $\mathbb{V}(x_i = z_i)_i$. Then $\widetilde{V}_3 \cap \widetilde{\Delta} \cong \widetilde{V}_1 \neq \emptyset$. Thus, the affine dimension theorem [42, Proposition 7.1] yields

$$\dim \widetilde{V}_1 \geq \dim \widetilde{V}_3 - n_1.$$

Noting $\dim V_1 + 1 \geq \dim \widetilde{V}_1$ and $\dim \widetilde{V}_3 \geq \dim V_3 + 2$ now gives the desired result. We remind the reader at this point that this is compatible with the convention $\dim \emptyset = -1$. $\square$

4.4 The auxiliary inequality

We remind the reader of the notation $e(x) = e^{2\pi i x}$. For $\boldsymbol{\alpha} \in [0,1]^R$ define

$$S(\boldsymbol{\alpha}, P_1, P_2) = S(\boldsymbol{\alpha}) \coloneqq \sum_{\boldsymbol{x} \in P_1 \mathcal{B}_1} \sum_{\boldsymbol{y} \in P_2 \mathcal{B}_2} e\left(\boldsymbol{\alpha} \cdot \boldsymbol{F}\left(\boldsymbol{x}, \boldsymbol{y}\right)\right),$$

where the sum ranges over $\boldsymbol{x} \in \mathbb{Z}^{n_1}$ such that $\boldsymbol{x}/P_1 \in \mathcal{B}_1$ and similarly for $\boldsymbol{y}$. Throughout this section we will assume $P_1 \geq P_2$. Note crucially that we have

$$N(P_1, P_2) = \int_{[0,1]^R} S(\boldsymbol{\alpha}) d\boldsymbol{\alpha}.$$

As noted in the introduction we can rewrite the forms as

$$F_i(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} F_{\boldsymbol{j},\boldsymbol{k}}^{(i)} x_{j_1} \cdots x_{j_{d_1}} y_{k_1} \cdots y_{k_{d_2}},$$

and given $\boldsymbol{\alpha} \in \mathbb{R}^R$, as in [98], we consider the multilinear forms

$$\Gamma_{\boldsymbol{\alpha} \cdot \boldsymbol{F}}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \coloneqq d_1! d_2! \sum_i \alpha_i \sum_{\boldsymbol{j}} \sum_{\boldsymbol{k}} F_{\boldsymbol{j},\boldsymbol{k}}^{(i)} x_{j_1}^{(1)} \cdots x_{j_{d_1}}^{(d_1)} y_{k_1}^{(1)} \cdots y_{k_{d_2}}^{(d_2)}.$$

Further we write $\widehat{\boldsymbol{x}} = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(d_1-1)})$ and similarly for $\widehat{\boldsymbol{y}}$. For any real number λ we write $\|\lambda\| = \min_{k \in \mathbb{Z}} |\lambda - k|$. We now define $M_1(\boldsymbol{\alpha} \cdot \boldsymbol{F}; P_1, P_2, P^{-1})$ to be the number of integral $\widehat{\boldsymbol{x}} \in (-P_1, P_1)^{(d_1-1)n_1}$ and $\widetilde{\boldsymbol{y}} \in (-P_2, P_2)^{d_2 n_2}$ such that for all $\ell = 1, \ldots, n_1$ we have

$$\|\Gamma_{\boldsymbol{\alpha}\cdot\boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widetilde{\boldsymbol{y}})\| < P^{-1}.$$

Similarly, we define $M_2(\boldsymbol{\alpha}\cdot\boldsymbol{F}; P_1, P_2, P^{-1})$ to be the number of integral $\widetilde{\boldsymbol{x}} \in (-P_1, P_1)^{d_1 n_1}$ and $\widehat{\boldsymbol{y}} \in (-P_2, P_2)^{(d_2-1)n_2}$ such that for all $\ell = 1, \ldots, n_2$ we have

$$\|\Gamma_{\boldsymbol{\alpha}\cdot\boldsymbol{F}}(\widetilde{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{e}_\ell,)\| < P^{-1}.$$

For our purposes we will need a slight generalization of Lemma 2.1 in [98] that deals with a polynomial $G(\boldsymbol{x}, \boldsymbol{y})$, which is not necessarily bihomogeneous. If $G(\boldsymbol{x}, \boldsymbol{y})$ has bidegree (d_1, d_2) write

$$G(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{0 \leq r \leq d_1 \\ 0 \leq l \leq d_2}} G^{(r,l)}(\boldsymbol{x}, \boldsymbol{y}),$$

where $G^{(r,l)}(\boldsymbol{x}, \boldsymbol{y})$ is homogeneous of bidegree (r, l). Using notation as above we first show the following preliminary Lemma, which is a version of Weyl's inequality for our context.

From now on we will often use the notation $\tilde{d} = d_1 + d_2 - 2$.

Lemma 4.4.1. *Let $\varepsilon > 0$. Let $G(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}[x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2}]$ be a polynomial of bidegree (d_1, d_2) with $d_1, d_2 \geq 1$. For the exponential sum*

$$S_G(P_1, P_2) = \sum_{\boldsymbol{x} \in P_1 \mathcal{B}_1} \sum_{\boldsymbol{x} \in P_2 \mathcal{B}_2} e\left(G(\boldsymbol{x}, \boldsymbol{y})\right)$$

we have the following bound

$$|S_G(P_1, P_2)|^{2^{\tilde{d}}} \ll P_1^{n_1(2^{\tilde{d}} - d_1 + 1) + \varepsilon} P_2^{n_2(2^{\tilde{d}} - d_2)} M_1\left(G^{(d_1, d_2)}, P_1, P_2, P_1^{-1}\right).$$

Proof. The proof is quite involved but follows closely the proof of Lemma 2.1 in [98], which in turn is based on idas of Schmidt [101, Section 11] and Davenport [26, Section 3].

Our first goal is to apply a Weyl differencing process $d_2 - 1$-times to the $\boldsymbol{y}$ part of G and then $d_1 - 1$-times to the $\boldsymbol{x}$ part of the resulting polynomial. Clearly this is trivial if $d_2 = 1$ or $d_1 = 1$, respectively. Therefore assume for now that $d_2 \geq 2$. We start by applying the Cauchy-Schwarz inequality and the triangle inequality to find

$$|S_G(P_1, P_2)|^{2^{d_2-1}} \ll P_1^{n_1(2^{d_2-1}-1)} \sum_{\boldsymbol{x} \in P_1 \mathcal{B}_1} |S_{\boldsymbol{x}}(P_1, P_2)|^{2^{d_2-1}}, \tag{4.4.1}$$

where we define

$$S_{\boldsymbol{x}}(P_1,P_2)=\sum_{\boldsymbol{y}\in P_2\mathcal{B}_2} e(G(\boldsymbol{x},\boldsymbol{y})).$$

Now write $\mathcal{U}=P_2\mathcal{B}_2$, write $\mathcal{U}^D=\mathcal{U}-\mathcal{U}$ for the difference set and define

$$\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(t)})=\bigcap_{\varepsilon_1=0,1}\cdots\bigcap_{\varepsilon_t=0,1}\left(\mathcal{U}-\varepsilon_1\boldsymbol{y}^{(1)}-\ldots-\varepsilon_t\boldsymbol{y}^{(t)}\right).$$

Write $\mathcal{F}(\boldsymbol{y})=G(\boldsymbol{x},\boldsymbol{y})$ and set

$$\mathcal{F}_d(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d)})=\sum_{\varepsilon_1=0,1}\cdots\sum_{\varepsilon_d=0,1}(-1)^{\varepsilon_1+\ldots+\varepsilon_d}\mathcal{F}(\varepsilon_1\boldsymbol{y}^{(1)}+\ldots+\varepsilon_d\boldsymbol{y}^{(d)}).$$

Equation (11.2) in [101] applied to our situation gives

$$|S_{\boldsymbol{x}}(P_1,P_2)|^{2^{d_2-1}}\ll\left|\mathcal{U}^D\right|^{2^{d_2-1}-d_2}\sum_{\boldsymbol{y}^{(1)}\in\mathcal{U}^D}\cdots\sum_{\boldsymbol{y}^{(d_2-2)}\in\mathcal{U}^D}\left|\sum_{\boldsymbol{y}^{(d_2-1)}\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-2)})}e\left(\mathcal{F}_{d_2-1}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)}\right)\right)\right|^2,$$

and we note that this did not require $\mathcal{F}(\boldsymbol{y})$ to be homogeneous in Schmidt's work. It is not hard to see that for $\boldsymbol{z},\boldsymbol{z}'\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-2)})$ we have

$$\mathcal{F}_{d_2-1}(\boldsymbol{y}^{(1)},\cdots,\boldsymbol{z})-\mathcal{F}_{d_2-1}(\boldsymbol{y}^{(1)},\cdots,\boldsymbol{z}')=\mathcal{F}_{d_2}(\boldsymbol{y}^{(1)},\cdots,\boldsymbol{y}^{(d_2-1)},\boldsymbol{y}^{(d_2)})-\mathcal{F}_{d_2-1}(\boldsymbol{y}^{(1)},\cdots,\boldsymbol{y}^{(d_2-1)}),$$

for some $\boldsymbol{y}^{(d_2-1)}\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-2)})^D$ and $\boldsymbol{y}^{(d_2)}\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)})$. Thus we find

$$|S_{\boldsymbol{x}}(P_1,P_2)|^{2^{d_2-1}}\ll\left|\mathcal{U}^D\right|^{2^{d_2-1}-d_2}\sum_{\boldsymbol{y}^{(1)}\in\mathcal{U}^D}\cdots\sum_{\boldsymbol{y}^{(d_2-2)}\in\mathcal{U}^D}\sum_{\boldsymbol{y}^{(d_2-1)}\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-2)})^D}\sum_{\boldsymbol{y}^{(d_2)}\in\mathcal{U}(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)})}e\left(\mathcal{F}_{d_2}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2)}\right)-\mathcal{F}_{d_2-1}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)}\right)\right). \quad (4.4.2)$$

We may write the polynomial $G(\boldsymbol{x},\boldsymbol{y})$ as follows

$$G(\boldsymbol{x},\boldsymbol{y})=\sum_{\substack{0\leq r\leq d_1\\0\leq l\leq d_2}}\sum_{\boldsymbol{j}_r,\boldsymbol{k}_l}G^{(r,l)}_{\boldsymbol{j}_r,\boldsymbol{k}_l}\boldsymbol{x}_{\boldsymbol{j}_r}\boldsymbol{y}_{\boldsymbol{k}_l},$$

for some real $G^{(r,l)}_{\boldsymbol{j}_r,\boldsymbol{k}_l}$. Further write $\mathcal{F}(\boldsymbol{y})=\mathcal{F}^{(0)}(\boldsymbol{y})+\ldots+\mathcal{F}^{(d_2)}(\boldsymbol{y})$, where $\mathcal{F}^{(d)}(\boldsymbol{y})$ denotes the degree d homogeneous part of $\mathcal{F}(\boldsymbol{y})$. Lemma 11.4 (A) in [101] states that $\mathcal{F}_{d_2}$ transpires to be the multilinear form associated to $\mathcal{F}^{(d_2)}(\boldsymbol{y})$. From this we see

$$\mathcal{F}_{d_2}-\mathcal{F}_{d_2-1}=\sum_{\substack{0\leq r\leq d_1\\0\leq l\leq d_2}}\sum_{\boldsymbol{j}_r,\boldsymbol{k}_l}G^{(r,l)}_{\boldsymbol{j}_r,\boldsymbol{k}_l}x_{j_r(1)}\cdots x_{j_r(r)}h_{\boldsymbol{k}_l}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2)}\right), \quad (4.4.3)$$

where

$$h_{\boldsymbol{k}_{d_2}}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2)}\right) = d_2! y^{(1)}_{k_{d_2}(1)}\cdots y^{(d_2)}_{k_{d_2}(d_2)} + \tilde{h}_{\boldsymbol{k}_{d_2}}\left(\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)}\right),$$

for some polynomials $\tilde{h}_{\boldsymbol{k}_{d_2}}$ of degree d_2 that are independent of $\boldsymbol{y}^{(d_2)}$ and further $h_{\boldsymbol{k}_l}$ are polynomials of degree l that are always independent of $\boldsymbol{y}^{(d_2)}$ whenever $l \leq d_2 - 1$. Write $\widetilde{\boldsymbol{y}} = (\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2)})$. Now set

$$S_{\widetilde{\boldsymbol{y}}} = \sum_{\boldsymbol{x}\in P_1\mathcal{B}_1} e\left(\sum_{\substack{0\leq r\leq d_1\\ 0\leq l\leq d_2}}\sum_{\boldsymbol{j}_r,\boldsymbol{k}_l} G^{(r,l)}_{\boldsymbol{j}_r,\boldsymbol{k}_l} x_{j_r(1)}\cdots x_{j_r(r)} h_{\boldsymbol{k}_l}(\widetilde{\boldsymbol{y}})\right).$$

Now we swap the order of summation of $\sum_{\boldsymbol{x}}$ in (4.4.1) with the sums over $\boldsymbol{y}^{(i)}$ in (4.4.2). Using the Cauchy-Schwarz inequality and (4.4.3) we thus obtain

$$|S_G(P_1,P_2)|^{2^{\tilde{d}}} \ll P_1^{n_1(2^{\tilde{d}}-2^{d_1-1})} P_2^{n_2(2^{\tilde{d}}-d_2)} \sum_{\boldsymbol{y}^{(1)}}\cdots\sum_{\boldsymbol{y}^{(d_2)}} |S_{\widetilde{\boldsymbol{y}}}|^{2^{d_1-1}}.$$

The above still holds if $d_2 = 1$, which can be seen directly. Applying the same differencing process to $S_{\widetilde{\boldsymbol{y}}}$ gives

$$|S_G(P_1,P_2)|^{2^{\tilde{d}}} \ll P_1^{n_1(2^{\tilde{d}}-d_1)} P_2^{n_2(2^{\tilde{d}}-d_2)} \sum_{\boldsymbol{y}^{(1)}}\cdots\sum_{\boldsymbol{y}^{(d_2)}}\sum_{\boldsymbol{x}^{(1)}}\cdots\left|\sum_{\boldsymbol{x}^{(d_1)}} e\left(\gamma(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\right)\right|, \quad (4.4.4)$$

where

$$\gamma(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}) = \sum_{\substack{0\leq r\leq d_1\\ 0\leq l\leq d_2}}\sum_{\boldsymbol{j}_r,\boldsymbol{k}_l} G^{(r,l)}_{\boldsymbol{j}_r,\boldsymbol{k}_l} g_{\boldsymbol{j}_r}(\widetilde{\boldsymbol{x}}) h_{\boldsymbol{k}_l}(\widetilde{\boldsymbol{y}}),$$

and where similar to before we have

$$g_{\boldsymbol{j}_{d_1}}(\widetilde{\boldsymbol{x}}) = d_1! x^{(1)}_{j_{d_1}(1)}\cdots x^{(d_1)}_{j_{d_1}(d_1)} + \tilde{g}_{\boldsymbol{j}_{d_1}}(\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(d_1-1)}),$$

with $\tilde{g}_{\boldsymbol{j}_{d_1}}$ and $g_{\boldsymbol{j}_r}$ for $r < d_1$ not depending on $\boldsymbol{x}^{(d_1)}$. We note that (4.4.4) holds for all $d_1, d_2 \geq 1$ and all the summations $\sum_{\boldsymbol{x}^{(i)}}$ and $\sum_{\boldsymbol{y}^{(j)}}$ in (4.4.4) are over boxes contained in $[-P_1,P_1]^{n_1}$ and $[-P_2,P_2]^{n_2}$, respectively. Write $\widehat{\boldsymbol{x}} = (\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(d_1-1)})$ and $\widehat{\boldsymbol{y}} = (\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(d_2-1)})$. We now wish to estimate the quantity

$$\sum(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}) := \sum_{\boldsymbol{y}^{(d_2)}}\left|\sum_{\boldsymbol{x}^{(d_1)}} e\left(\gamma(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}})\right)\right|. \quad (4.4.5)$$

Viewing $\sum_{a<x\leq b} e(\beta x)$ for $b-a \geq 1$ as a geometric series we recall the following elementary estimate

$$\left|\sum_{a<x\leq b} e(\beta x)\right| \ll \min\{b-a, \|\beta\|^{-1}\}.$$

This yields

$$\left|\sum_{\boldsymbol{x}^{(d_1)}} e\left(\gamma(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})\right)\right| \ll \prod_{\ell=1}^{n_1} \min\left\{P_1, \|\widetilde{\gamma}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widetilde{\boldsymbol{y}})\|^{-1}\right\},$$

where $\boldsymbol{e}_\ell$ denotes the ℓ-th unit vector and where

$$\widetilde{\gamma}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) = d_1! \sum_{0 \le l \le d_2} \sum_{\boldsymbol{j}_{d_1}, \boldsymbol{k}_l} G^{(d_1,l)}_{\boldsymbol{j}_{d_1}, \boldsymbol{k}_l} x^{(1)}_{j_{d_1}(1)} \cdots x^{(d_1)}_{j_{d_1}(d_1)} h_{\boldsymbol{k}_l}(\widetilde{\boldsymbol{y}}).$$

We now apply a standard argument in order to estimate this product, as in Davenport [29, Chapter 13]. For a real number z write $\{z\}$ for its fractional part. Let $\boldsymbol{r} = (r_1, \ldots, r_{n_1}) \in \mathbb{Z}^{n_1}$ be such that $0 \le r_\ell < P_1$ holds for $\ell = 1, \ldots, n_1$. Define $\mathcal{A}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r})$ to be the set of $\boldsymbol{y}^{(d_2)}$ in the sum in (4.4.5) such that

$$r_\ell P_1^{-1} \le \left\{\widetilde{\gamma}\left(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{y}^{(d_2)}\right)\right\} < (r_\ell + 1) P_1^{-1},$$

holds for all $\ell = 1, \ldots, n_1$ and write $A(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r})$ for its cardinality. We obtain the estimate

$$\sum(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \ll \sum_{\boldsymbol{r}} A(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r}) \prod_{\ell=1}^{n_1} \min\left\{P_1, \max\left\{\frac{P_1}{r_\ell}, \frac{P_1}{P_1 - r_\ell - 1}\right\}\right\},$$

where the sum $\sum_{\boldsymbol{r}}$ is over integral $\boldsymbol{r}$ with $0 \le r_\ell < P_1$ for all $\ell = 1, \ldots, n_1$. Our next aim is to find a bound for $A(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r})$ that is independent of $\boldsymbol{r}$. Given $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{A}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r})$ then

$$\|\widetilde{\gamma}\left(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{u}\right) - \widetilde{\gamma}\left(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{v}\right)\| < P_1^{-1},$$

for $\ell = 1, \ldots, n_1$. Similar as before we now define the multilinear forms

$$\Gamma_G(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) := d_1! d_2! \sum_{\boldsymbol{j}_{d_1}, \boldsymbol{k}_{d_2}} G^{(d_1,d_2)}_{\boldsymbol{j}_{d_1}, \boldsymbol{k}_{d_2}} x^{(1)}_{j_{d_1}(1)} \cdots x^{(d_1)}_{j_{d_1}(d_1)} y^{(1)}_{k_{d_2}(1)} \cdots y^{(d_2)}_{k_{d_2}(d_2)},$$

which only depend on the (d_1, d_2)-degree part of G. For fixed $\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}$ let $N(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$ be the number of $\boldsymbol{y} \in (-P_2, P_2)^{n_2}$ such that

$$\|\Gamma_G(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{y})\| < P_1^{-1},$$

for al $\ell = 1, \ldots, n_1$. Observe now crucially

$$\widetilde{\gamma}\left(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{u}\right) - \widetilde{\gamma}\left(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{v}\right) = \Gamma_G(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widehat{\boldsymbol{y}}, \boldsymbol{u} - \boldsymbol{v}).$$

Thus we find $A(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}, \boldsymbol{r}) \le N(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$ for all $\boldsymbol{r}$ as specified above. Using this we get

$$\sum_{\boldsymbol{y}^{(d_2)}} \left|\sum_{\boldsymbol{x}^{(d_1)}} e\left(\gamma(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})\right)\right| \ll N(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) (P_1 \log P_1)^{n_1}.$$

Finally, summing over $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ we obtain

$$|S_G(P_1, P_2)|^{2^{\tilde{d}}} \ll P_1^{n_1(2^{\tilde{d}} - d_1 + 1) + \varepsilon} P_2^{n_2(2^{\tilde{d}} - d_2)} M_1\left(G^{(d_1,d_2)}, P_1, P_2, P_1^{-1}\right). \qquad \square$$

Inspecting the proof of Lemma 4.1 in [98] we find that for a polynomial $G(\boldsymbol{x}, \boldsymbol{y})$ as above given $\theta \in (0,1]$ the following holds

$$M_1(G^{(d_1,d_2)}, P_1, P_2, P_1^{-1}) \ll P_1^{n_1(d_1-1)} P_2^{n_2 d_2} P_2^{-\theta(n_1 d_1 + n_2 d_2)} \times \max_{i=1,2} \left\{ P_2^{n_i \theta} M_i \left(G^{(d_1,d_2)}; P_2^{\theta}, P_2^{\theta}, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)} \right) \right\}$$

Using this and Lemma 4.4.1 we deduce the next Lemma.

Lemma 4.4.2. *Let $P_1, P_2 > 1$, $\theta \in (0,1]$ and $\boldsymbol{\alpha} \in \mathbb{R}^R$. Write $S_G = S_G(P_1, P_2)$. Using the same notation as above for $i = 1$ or $i = 2$ we have*

$$|S_G|^{2^{\tilde{d}}} \ll_{d_i, n_i, \varepsilon} P_1^{n_1 2^{\tilde{d}} + \varepsilon} P_2^{n_2 2^{\tilde{d}}} P_2^{\theta n_i - \theta(n_1 d_1 + n_2 d_2)} \times M_i \left(G^{(d_1,d_2)}; P_2^{\theta}, P_2^{\theta}, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)} \right).$$

Using the preceding Lemma and adapting the proof of [94, Lemma 3.1] to our setting we can now show the following.

Lemma 4.4.3. *Let $\varepsilon > 0$, $\theta \in (0,1]$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^R$. Then for $i = 1$ or $i = 2$ we have*

$$\min \left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right| \right\}^{2^{\tilde{d}+1}} \ll_{d_i, n_i, \varepsilon} \frac{M_i \left(\boldsymbol{\beta} \cdot \boldsymbol{F}; P_2^{\theta}, P_2^{\theta}, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)} \right)}{P_2^{\theta(n_1 d_1 + n_2 d_2) - \theta n_i}} \tag{4.4.6}$$

Proof. Note first that for two real numbers $\lambda, \mu > 0$ we have

$$\min\{\lambda, \mu\} \leq \sqrt{\lambda \mu}.$$

Therefore it suffices to show

$$\left| \frac{S(\boldsymbol{\alpha}) S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{2n_1+2\varepsilon} P_2^{2n_2}} \right|^{2^{\tilde{d}}} \ll_{d_i, n_i, \varepsilon} \frac{M_i \left(\boldsymbol{\beta}; P_2^{\theta}, P_2^{\theta}, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)} \right)}{P_2^{\theta(n_1 d_1 + n_2 d_2) - \theta n_i}}.$$

holds for $i = 1$ or $i = 2$. Note first that

$$\left| S(\boldsymbol{\alpha}+\boldsymbol{\beta}) \overline{S}(\boldsymbol{\alpha}) \right| = \left| \sum_{\substack{\boldsymbol{x} \in P_1 \mathcal{B}_1 \\ \boldsymbol{y} \in P_2 \mathcal{B}_2}} \sum_{\substack{\boldsymbol{x}+\boldsymbol{z} \in P_1 \mathcal{B}_1 \\ \boldsymbol{y}+\boldsymbol{w} \in P_2 \mathcal{B}_2}} e\left((\boldsymbol{\alpha}+\boldsymbol{\beta}) \cdot \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{\alpha} \cdot \boldsymbol{F}(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+\boldsymbol{w}) \right) \right|,$$

so by the triangle inequality we get

$$\left| S(\boldsymbol{\alpha}+\boldsymbol{\beta}) \overline{S}(\boldsymbol{\alpha}) \right| \leq \sum_{\substack{\|\boldsymbol{z}\|_\infty \leq P_1 \\ \|\boldsymbol{w}\|_\infty \leq P_2}} \left| \sum_{\substack{\boldsymbol{x} \in P_1 \mathcal{B}_{\boldsymbol{z}} \\ \boldsymbol{y} \in P_2 \mathcal{B}_{\boldsymbol{w}}}} e\left(\boldsymbol{\beta} \cdot \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) - g_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{y}) \right) \right|,$$

where $g_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{x},\boldsymbol{y})$ is of degree at most d_1+d_2-1 in $(\boldsymbol{x},\boldsymbol{y})$ and we have some boxes $\mathcal{B}_{\boldsymbol{z}} \subset \mathcal{B}_1$ and $\mathcal{B}_{\boldsymbol{w}} \subset \mathcal{B}_2$. Applying Cauchy's inequality $\tilde{d}$-times we deduce

$$\left|S(\boldsymbol{\alpha}+\boldsymbol{\beta})\overline{S}(\boldsymbol{\alpha})\right|^{2^{\tilde{d}}} \leq P_1^{n_1(2^{\tilde{d}}-1)} P_2^{n_2(2^{\tilde{d}}-1)} \sum_{\substack{\|\boldsymbol{z}\|_\infty \leq P_1 \\ \|\boldsymbol{w}\|_\infty \leq P_2}} \left| \sum_{\substack{\boldsymbol{x}\in P_1\mathcal{B}_{\boldsymbol{z}} \\ \boldsymbol{y}\in P_2\mathcal{B}_{\boldsymbol{w}}}} e\left(\boldsymbol{\beta}\cdot\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) - g_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{x},\boldsymbol{y})\right)\right|^{2^{\tilde{d}}}.$$

If we write $G(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{\beta}\cdot\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) - g_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{x},\boldsymbol{y})$ then note that $G^{(d_1,d_2)} = \boldsymbol{\beta}\cdot\boldsymbol{F}$. Using Lemma 4.4.2 we therefore obtain

$$\left|S(\boldsymbol{\alpha}+\boldsymbol{\beta})\overline{S}(\boldsymbol{\alpha})\right|^{2^{\tilde{d}}} \ll P_1^{2^{\tilde{d}+1}n_1+\varepsilon} P_2^{2^{\tilde{d}+1}n_2} P_2^{-\theta(n_1d_1+n_2d_2)+\theta n_i} \times M_i(\boldsymbol{\beta}\cdot\boldsymbol{F}, P_2^\theta, P_2^\theta, P_1^{-d_1}P_2^{-d_2}P_2^{\theta(\tilde{d}+1)}),$$

for $i=1$ or $i=2$, which readily delivers the result. □

As in the introduction, for $\boldsymbol{\beta}\in\mathbb{R}^R$ we define the auxiliary counting function $N_1^{\mathrm{aux}}(\boldsymbol{\beta};B)$ to be the number of integer vectors $\widehat{\boldsymbol{x}} \in (-B,B)^{(d_1-1)n_1}$ and $\widetilde{\boldsymbol{y}} \in (-B,B)^{d_2n_2}$ such that

$$|\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widetilde{\boldsymbol{y}})| < \|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty B^{\tilde{d}},$$

for $\ell = 1,\dots,n_1$ where $\|f\|_\infty := \frac{1}{d_1!d_2!}\max_{\boldsymbol{j},\boldsymbol{k}}\left|\frac{\partial^{d_1+d_2}f}{\partial x_{j_1}\cdots\partial x_{j_{d_1}}\partial y_{k_1}\cdots\partial y_{k_{d_2}}}\right|$. We also analogously define $N_2^{\mathrm{aux}}(\boldsymbol{\beta};B)$. We now formulate an analogue for [94, Proposition 3.1].

Proposition 4.4.4. *Let $C_0 \geq 1$ and $\mathscr{C} > 0$ such that for all $\boldsymbol{\beta}\in\mathbb{R}^R$ and $B>0$ we have for $i=1,2$ that*

$$N_i^{\mathrm{aux}}(\boldsymbol{\beta};B) \leq C_0 B^{d_1n_1+d_2n_2-n_i-2^{\tilde{d}+1}\mathscr{C}}. \tag{4.4.7}$$

Assume further that the forms F_i are linearly independent, so that there exist $M > \mu > 0$ such that

$$\mu\|\boldsymbol{\beta}\|_\infty \leq \|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty \leq M\|\boldsymbol{\beta}\|_\infty. \tag{4.4.8}$$

Then there exists a constant $C>0$ depending on C_0, d_i, n_i, μ and M such that the following auxiliary inequality

$$\min\left\{\left|\frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon}P_2^{n_2}}\right|, \left|\frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon}P_2^{n_2}}\right|\right\} \leq C\max\left\{P_2^{-1}, P_1^{-d_1}P_2^{-d_2}\|\boldsymbol{\beta}\|_\infty^{-1}, \|\boldsymbol{\beta}\|_\infty^{\frac{1}{\tilde{d}+1}}\right\}^{\mathscr{C}}$$

holds for all real numbers $P_1, P_2 > 1$.

Proof. The strategy of this proof will closely follow the proof of [94, Proposition 3.1]. By Lemma 4.4.3 we know that (4.4.6) holds for $i = 1$ or $i = 2$. Assume that there is some $\theta \in (0, 1]$ such that for the same i we have

$$N_i^{\mathrm{aux}}(\boldsymbol{\beta}; P_2^\theta) < M_i(\boldsymbol{\beta} \cdot \boldsymbol{F}, P_2^\theta, P_2^\theta, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)}), \tag{4.4.9}$$

Going forward with the case $i = 1$, noting that the case $i = 2$ can be proven completely analogously, this means that there exists a $(d_1 - 1)$-tuple $\widehat{\boldsymbol{x}}$ and a d_2-tuple $\widetilde{\boldsymbol{y}}$ which is counted by $M_1(\boldsymbol{\beta} \cdot \boldsymbol{F}, P_2^\theta, P_2^\theta, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)})$ but not by $N_1^{\mathrm{aux}}(\boldsymbol{\beta}; P_2^\theta)$. Therefore this pair of tuples satisfies

$$\left\|\widehat{\boldsymbol{x}}^{(i)}\right\|_\infty, \left\|\widetilde{\boldsymbol{y}}^{(j)}\right\|_\infty \leq P_2^\theta, \text{ for } i = 1, \ldots, d_1 - 1 \text{ and } j = 1, \ldots, d_2, \tag{4.4.10}$$

and

$$\|\Gamma_{\boldsymbol{\beta} \cdot \boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_\ell, \widetilde{\boldsymbol{y}})\| < P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)}, \text{ for } \ell = 1, \ldots, n_1, \tag{4.4.11}$$

since it is counted by $M_1(\boldsymbol{\beta} \cdot \boldsymbol{F}, P_2^\theta, P_2^\theta, P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)})$. On the other hand, since it is not counted by $N_1^{\mathrm{aux}}(\boldsymbol{\beta}; P_2^\theta)$ there exists $\ell_0 \in \{1, \ldots, n_1\}$ such that

$$|\Gamma_{\boldsymbol{\beta} \cdot \boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_{\ell_0}, \widetilde{\boldsymbol{y}})| \geq \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty P_2^{\tilde{d}\theta}. \tag{4.4.12}$$

From (4.4.11) we get that for ℓ_0 we must have either

$$|\Gamma_{\boldsymbol{\beta} \cdot \boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_{\ell_0}, \widetilde{\boldsymbol{y}})| < P_1^{-d_1} P_2^{-d_2} P_2^{\theta(\tilde{d}+1)} \tag{4.4.13}$$

or

$$|\Gamma_{\boldsymbol{\beta} \cdot \boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_{\ell_0}, \widetilde{\boldsymbol{y}})| \geq \frac{1}{2}. \tag{4.4.14}$$

If (4.4.13) holds then (4.4.12) implies

$$\|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty < \frac{P_1^{-d_1} P_2^{-d_2} P_2^{(\tilde{d}+1)\theta}}{P_2^{\tilde{d}\theta}} = P_2^\theta P_1^{-d_1} P_2^{-d_2} \tag{4.4.15}$$

If on the other hand (4.4.14) holds then (4.4.10) gives

$$\frac{1}{2} \leq |\Gamma_{\boldsymbol{\beta} \cdot \boldsymbol{F}}(\widehat{\boldsymbol{x}}, \boldsymbol{e}_{\ell_0}, \widetilde{\boldsymbol{y}})| \ll \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty P_2^{(\tilde{d}+1)\theta}. \tag{4.4.16}$$

Since either (4.4.15) or (4.4.16) holds then via (4.4.8) we deduce

$$P_2^{-\theta} \ll_{\mu, M} \max\left\{ P_1^{-d_1} P_2^{-d_2} \|\boldsymbol{\beta}\|_\infty^{-1}, \|\boldsymbol{\beta}\|_\infty^{\frac{1}{\tilde{d}+1}} \right\}. \tag{4.4.17}$$

Since (4.4.6) holds for $i = 1$ and due to the assumption (4.4.7) we see that (4.4.9) holds if there exists some $C_1 > 0$ such that

$$P_2^{-\theta 2^{\tilde{d}+1}\mathscr{C}} \leq C_1 \min\left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right| \right\}^{2^{\tilde{d}+1}}. \tag{4.4.18}$$

Now *define* θ such that we have equality in the equation above, i.e. such that we have

$$P_2^{\theta} = C_1^{\frac{1}{2^{\tilde{d}+1}\mathscr{C}}} \min\left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right| \right\}^{-\frac{1}{\mathscr{C}}}. \tag{4.4.19}$$

If $\theta \in (0, 1]$ then (4.4.18) holds and so together with the assumption (4.4.7) as argued above this implies (4.4.17) holds, which gives the result in this case. But θ will always be positive; for if $\theta \leq 0$ then (4.4.19) implies

$$\min\left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right| \right\} \geq C_1^{-\frac{1}{2^{\tilde{d}+1}}}.$$

However, note that clearly $|S(\boldsymbol{\alpha})| \leq (P_1+1)^{n_1}(P_2+1)^{n_2}$. Without loss of generality we may take P_i large enough, depending on ε, so that this clearly leads to a contradiction. Finally, if $\theta \geq 1$ then we find $P_2^{-\mathscr{C}\theta} \leq P_2^{-\mathscr{C}}$, and so from (4.4.19) we obtain.

$$\min\left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1+\varepsilon} P_2^{n_2}} \right| \right\} \ll P_2^{-\mathscr{C}}.$$

This gives the result. □

4.5 The circle method

The aim of this section is to use the auxiliary inequality

$$P_1^{-\varepsilon} \min\left\{ \left| \frac{S(\boldsymbol{\alpha})}{P_1^{n_1} P_2^{n_2}} \right|, \left| \frac{S(\boldsymbol{\alpha}+\boldsymbol{\beta})}{P_1^{n_1} P_2^{n_2}} \right| \right\} \leq C \max\left\{ P_2^{-1}, P_1^{-d_1} P_2^{-d_2} \|\boldsymbol{\beta}\|_\infty^{-1}, \|\boldsymbol{\beta}\|_\infty^{\frac{1}{\tilde{d}+1}} \right\}^{\mathscr{C}}, \tag{4.5.1}$$

where $C \geq 1$ and apply the circle method in order to deduce an estimate for $N(P_1, P_2)$. In this section we will use the notation $P = P_1^{d_1} P_2^{d_2}$. Write $b = \max\{1, \log P_1/\log P_2\}$ and $u = \max\{1, \log P_2/\log P_1\}$. If $P_1 \geq P_2$ then $b = \log P_1/\log P_2$ and thus $P_2^{bd_1+d_2} = P$ holds. The main result will be the following.

Proposition 4.5.1. *Let $\mathscr{C} > (bd_1 + ud_2)R$, $C \geq 1$ and $\varepsilon > 0$ such that the auxiliary inequality* (4.5.1) *holds for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^R$, all $P_1, P_2 > 1$ and all boxes $\mathcal{B}_i \subset [-1,1]^{n_i}$*

with side lengths at most 1 *and edges parallel to the coordinate axes. There exists some* $\delta > 0$ *depending on* b, u, R, d_i *and* n_i *such that*

$$N(P_1, P_2) = \sigma P_1^{n_1-d_1R} P_2^{n_2-d_2R} + O\left(P_1^{n_1-d_1R} P_2^{n_2-d_2R} P^{-\delta}\right).$$

The factor $\sigma = \mathfrak{I}\mathfrak{S}$ *is the product of the singular integral* $\mathfrak{I}$ *and the singular series* $\mathfrak{S}$*, as defined in* (4.5.26) *and* (4.5.23)*, respectively.*

Note that this result holds for general bidegree, and therefore in the proof one may assume $P_1 \geq P_2$ throughout. For instance if one wishes to show the above proposition for bidegree $(2,1)$, the result follows from the asymmetric results of bidegree $(2,1)$ and bidegree $(1,2)$.

4.5.1 The minor arcs

First we will show that the contributions from the minor arcs do not affect the main term. For this we will prove a Lemma similar to Lemma 2.1 in [94].

Lemma 4.5.2. *Let* $r_1, r_2\colon (0,\infty) \to (0,\infty)$ *be strictly decreasing and increasing bijections, respectively, and let* $A > 0$ *be a real number. For any* $\nu > 0$ *let* $E_0 \subset \mathbb{R}^R$ *be a hypercube of side lengths* ν *whose edges are parallel to the coordinate axes. Let* $E \subseteq E_0$ *be a measurable set and let* $\varphi\colon E \to [0,\infty)$ *be a measurable function.*

Assume that for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^R$ *such that* $\boldsymbol{\alpha}, \boldsymbol{\alpha}+\boldsymbol{\beta} \in E$ *we have*

$$\min\{\varphi(\boldsymbol{\alpha}), \varphi(\boldsymbol{\alpha}+\boldsymbol{\beta})\} \leq \max\left\{A, r_1^{-1}\left(\|\boldsymbol{\beta}\|_\infty\right), r_2^{-1}\left(\|\boldsymbol{\beta}\|_\infty\right)\right\}. \tag{4.5.2}$$

Then for all integers $k \leq \ell$ *such that* $A < 2^k$ *we get*

$$\int_E \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \ll_R \nu^R 2^k + \sum_{i=k}^{\ell-1} 2^i \left(\frac{\nu r_1(2^i)}{\min\{r_2(2^i), \nu\}}\right)^R + \left(\frac{\nu r_1(2^\ell)}{\min\{r_2(2^\ell), \nu\}}\right)^R \sup_{\boldsymbol{\alpha}\in E} \varphi(\boldsymbol{\alpha}). \tag{4.5.3}$$

Note that if we take

$$\varphi(\boldsymbol{\alpha}) = C^{-1} P_1^{-n_1-\varepsilon} P_2^{-n_2} |S(\boldsymbol{\alpha})|, \quad r_1(t) = P_1^{-d_1} P_2^{-d_2} t^{-\frac{1}{\mathscr{C}}}, \quad r_2(t) = t^{\frac{\tilde{d}+1}{\mathscr{C}}}, \quad A = P_2^{-\mathscr{C}}$$

where C is the constant in (4.5.1), then the assumption (4.5.2) is just the auxiliary inequality (4.5.1).

Proof. Given $t \geq 0$ define the set

$$D(t) = \{\boldsymbol{\alpha} \in E \colon \varphi(\boldsymbol{\alpha}) \geq t\}.$$

If $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}+\boldsymbol{\beta}$ are both contained in $D(t)$ then by (4.5.2) one of the following must hold

$$A \geq t, \quad \|\boldsymbol{\beta}\|_\infty \leq r_1(t), \quad \text{or} \quad \|\boldsymbol{\beta}\|_\infty \geq r_2(t).$$

In particular, if $t > A$ then either $\|\boldsymbol{\beta}\|_\infty \leq r_1(t)$ or $\|\boldsymbol{\beta}\|_\infty \geq r_2(t)$. Assuming that $t > A$ is satisfied consider a box $\mathfrak{b} \subset \mathbb{R}^R$ with sidelengths $r_2(t)/2$ whose edges are parallel to the coordinate axes. Given $\boldsymbol{\alpha} \in \mathfrak{b} \cap D(t)$ set

$$\mathfrak{B}(\boldsymbol{\alpha}) = \left\{\boldsymbol{\alpha}+\boldsymbol{\beta} \colon \boldsymbol{\beta} \in \mathbb{R}^R, \|\boldsymbol{\beta}\|_\infty \leq r_1(t)\right\}.$$

If $\boldsymbol{\alpha}+\boldsymbol{\beta} \in \mathfrak{b}\cap D(t)$ then by construction $\|\boldsymbol{\beta}\|_\infty \leq r_2(t)/2 < r_2(t)$ whence $\|\boldsymbol{\beta}\|_\infty \leq r_1(t)$. Therefore we have $\mathfrak{b}\cap D(t) \subset \mathfrak{B}(\boldsymbol{\alpha})$, which in turn implies that the measure of $\mathfrak{b}\cap D(t)$ is bounded by $(2r_1(t))^R$. Since $D(t)$ is contained in E_0 one can cover $D(t)$ with at most

$$\ll_R \frac{\nu^R}{\min\{r_2(t), \nu\}^R}$$

boxes $\mathfrak{b}$ whose sidelenghts are $r_2(t)/2$. Therefore we find

$$\mu(D(t)) \ll_R \left(\frac{\nu r_1(t)}{\min\{r_2(t), \nu\}}\right)^R,$$

where we write $\mu(D(t))$ for the Lebesgue measure of $D(t)$. If $k < \ell$ are two integers then

$$\int_E \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} = \int_{E\setminus D(2^k)} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} + \sum_{i=k}^{\ell} \int_{D(2^i)\setminus D(2^{i+1})} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} + \int_{D(2^\ell)} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}.$$

We can trivially bound $\int_{E\setminus D(2^k)} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \leq \nu^R 2^k$, and further we can bound

$$\int_{D(2^i)\setminus D(2^{i+1})} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \leq 2^{i+1}\mu(D(2^i)), \quad \text{and} \quad \int_{D(2^\ell)} \varphi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \leq \mu(D(2^\ell)) \sup_{\boldsymbol{\alpha}\in E} \varphi(\boldsymbol{\alpha}).$$

If $2^k > A$ then for any $i \geq k$ by our discussion above we find

$$\mu(D(2^i)) \ll_R \left(\frac{\nu r_1(2^i)}{\min\{r_2(2^i), \nu\}}\right)^R.$$

Therefore the result follows. □

Recall the notation $P = P_1^{d_1} P_2^{d_2}$. From now on we will assume $P_1 \geq P_2$. Note that the assumption in Proposition 4.4.4 that $\mathscr{C} > R(bd_1 + ud_2)$ holds, is equivalent to $\mathscr{C} > R(bd_1 + d_2)$ if $P_1 \geq P_2$.

Lemma 4.5.3. *Let $T\colon \mathbb{R}^R \to \mathbb{C}$ be a measurable function. With notation as in Lemma 4.5.2 assume that for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^R$ and for all $P_1 \geq P_2 > 1$, and $\mathscr{C} > 0$ we have*

$$\min \left\{ \left| \frac{T(\boldsymbol{\alpha})}{P_1^{n_1} P_2^{n_2}} \right|, \left| \frac{T(\boldsymbol{\alpha} + \boldsymbol{\beta})}{P_1^{n_1} P_2^{n_2}} \right| \right\} \leq \max \left\{ P_2^{-1}, P_1^{-d_1} P_2^{-d_2} \left\|\boldsymbol{\beta}\right\|_\infty^{-1}, \left\|\boldsymbol{\beta}\right\|_\infty^{\frac{1}{\tilde{d}+1}} \right\}^{\mathscr{C}}. \quad (4.5.4)$$

Write $P = P_1^{d_1} P_2^{d_2}$ and assume that that we have

$$\sup_{\boldsymbol{\alpha} \in E} |T(\boldsymbol{\alpha})| \leq P_1^{n_1} P_2^{n_2} P^{-\delta}, \quad (4.5.5)$$

for some $\delta > 0$. Then we have

$$\int_E \frac{T(\boldsymbol{\alpha})}{P_1^{n_1} P_2^{n_2}} d\boldsymbol{\alpha} \ll_{\mathscr{C}, d_i, R}$$
$$\begin{cases} \nu^R P^{-R} P_2^{(\tilde{d}+2)R - \mathscr{C}} + P_2^{-\mathscr{C}} & \text{if } \mathscr{C} < R \\ \nu^R P^{-R} P_2^{(\tilde{d}+2)R - \mathscr{C}} + P^{-R} \log P_2 + P_2^{-\mathscr{C}} & \text{if } \mathscr{C} = R \\ \nu^R P^{-R} P_2^{(\tilde{d}+2)R - \mathscr{C}} + P^{-R - \delta(1 - R/\mathscr{C})} + P_2^{-\mathscr{C}} & \text{if } R < \mathscr{C} < (d_1 + d_2) R \\ \nu^R P^{-R} \log P_2 + P^{-R - \delta(1 - R/\mathscr{C})} + P_2^{-\mathscr{C}} & \text{if } \mathscr{C} = (d_1 + d_2) R \\ \nu^R P^{-R - \delta(1 - (d_1 + d_2) R/\mathscr{C})} + P^{-R - \delta(1 - R/\mathscr{C})} + P_2^{-\mathscr{C}} & \text{if } \mathscr{C} > (d_1 + d_2) R. \end{cases} \quad (4.5.6)$$

We expect the main term of $N(P_1, P_2)$ to be of order $P_1^{n_1 - Rd_1} P_2^{n_2 - Rd_2} = P_1^{n_1} P_2^{n_2} P^{-R}$. Thus the Lemma indicates why it is necessary for us to assume $\mathscr{C} > R(bd_1 + d_2)$, using this method of proof at least.

Proof. We apply Lemma 4.4.3 by taking

$$\varphi(\boldsymbol{\alpha}) = \frac{|T(\boldsymbol{\alpha})|}{P_1^{n_1} P_2^{n_2}}, \quad r_1(t) = P_1^{-d_1} P_2^{-d_2} t^{-\frac{1}{\mathscr{C}}}, \quad r_2(t) = t^{\frac{\tilde{d}+1}{\mathscr{C}}}, \text{ and } A = P_2^{-\mathscr{C}}. \quad (4.5.7)$$

Then our assumption (4.5.4) is just (4.5.2). We will choose our parameters k and ℓ such that the $\sum_{i=k}^{\ell-1}$ term dominates the right hand side of (4.5.3). Let

$$k = \left\lceil \log_2 P_2^{-\mathscr{C}} \right\rceil, \quad \text{and} \quad \ell = \left\lceil \log_2 P^{-\delta} \right\rceil, \quad (4.5.8)$$

so that we have

$$P_2^{-\mathscr{C}} < 2^k \leq 2P_2^{-\mathscr{C}}, \quad \text{and} \quad P^{-\delta} \leq 2^\ell < 2P^{-\delta}.$$

Without loss of generality we assume $k < \ell$ since otherwise the bound in the assumption (4.5.5) would be sharper than any of those listed in (4.5.6). Substituting our choices (4.5.7) into (4.5.3) we get

$$\int_E \frac{|T(\boldsymbol{\alpha})|}{P_1^{n_1}P_2^{n_2}} \ll_R \nu^R 2^k + \sum_{i=k}^{\ell-1} 2^i \left(\frac{\nu P_1^{-d_1} P_2^{-d_2} 2^{-i/\mathscr{C}}}{\min\left\{\nu, 2^{i(\tilde{d}+1)/\mathscr{C}}\right\}} \right)^R + \left(\frac{\nu P_1^{-d_1} P_2^{-d_2} 2^{-\ell/\mathscr{C}}}{\min\left\{\nu, 2^{\ell(\tilde{d}+1)/\mathscr{C}}\right\}} \right)^R \sup_{\boldsymbol{\alpha}\in E} \frac{|T(\boldsymbol{\alpha})|}{P_1^{n_1}P_2^{n_2}}. \quad (4.5.9)$$

From (4.5.5) and (4.5.8) we see that

$$\sup_{\boldsymbol{\alpha}\in E} \frac{|T(\boldsymbol{\alpha})|}{P_1^{n_1}P_2^{n_2}} \leq P^{-\delta} \leq 2^\ell. \quad (4.5.10)$$

Further, we clearly have

$$\frac{P_1^{-d_1} P_2^{-d_2} 2^{-i/\mathscr{C}}}{\min\left\{\nu, 2^{i(\tilde{d}+1)/\mathscr{C}}\right\}} \leq \nu^{-1} P_1^{-d_1} P_2^{-d_2} 2^{-i/\mathscr{C}} + 2^{-i(\tilde{d}+2)/\mathscr{C}} P_1^{-d_1} P_2^{-d_2}. \quad (4.5.11)$$

Substituting the estimates (4.5.10) and (4.5.11) into (4.5.9) we obtain

$$\int_E \frac{|T(\boldsymbol{\alpha})|}{P_1^{n_1}P_2^{n_2}} \ll_R \nu^R 2^k + \sum_{i=k}^{\ell} \nu^R P_1^{-d_1 R} P_2^{-d_2 R} 2^{i(1-(\tilde{d}+2)R/\mathscr{C})} + \sum_{i=k}^{\ell} P_1^{-d_1 R} P_2^{-d_2 R} 2^{i(1-R/\mathscr{C})}. \quad (4.5.12)$$

Note now that

$$\sum_{i=k}^{\ell} 2^{i(1-R(\tilde{d}+2)/\mathscr{C})} \ll_{\mathscr{C},d_i,R} \begin{cases} 2^{k(1-R(\tilde{d}+2)/\mathscr{C})} & \text{if } \mathscr{C} < (\tilde{d}+2)R \\ \ell - k & \text{if } \mathscr{C} = (\tilde{d}+2)R \\ 2^{\ell(1-R(\tilde{d}+2)/\mathscr{C})} & \text{if } \mathscr{C} > (\tilde{d}+2)R, \end{cases} \quad (4.5.13)$$

where we used $k < \ell$ for the second alternative. Recall from (4.5.8) that we have

$$2^k \geq P_2^{-\mathscr{C}} \quad \text{and} \quad 2^\ell \leq 2P^{-\delta},$$

so using this in (4.5.13) we get

$$\sum_{i=k}^{\ell} 2^{i(1-(\tilde{d}+2)/\mathscr{C})} \ll_{\mathscr{C},d_i,R} \begin{cases} P_2^{(\tilde{d}+2)R-\mathscr{C}} & \text{if } \mathscr{C} < (\tilde{d}+2)R \\ \log P_2 & \text{if } \mathscr{C} = (\tilde{d}+2)R \\ P^{-\delta(1-(\tilde{d}+2)R/\mathscr{C})} & \text{if } \mathscr{C} > (\tilde{d}+2)R. \end{cases} \quad (4.5.14)$$

Arguing similarly for $\sum_{i=k}^{\ell} 2^{i(1-R/\mathscr{C})}$ we find

$$\sum_{i=k}^{\ell} 2^{i(1-R/\mathscr{C})} \ll_{\mathscr{C},d_i,R} \begin{cases} P_2^{R-\mathscr{C}} & \text{if } \mathscr{C} < R \\ \log P_2 & \text{if } \mathscr{C} = R \\ P^{-\delta(1-R/\mathscr{C})} & \text{if } \mathscr{C} > R. \end{cases} \quad (4.5.15)$$

Finally we note that by our choice of k we have $2^k \leq 2P_2^{-\mathscr{C}}$ and we recall that $\tilde{d}+2 = d_1 + d_2$. Using this, as well as (4.5.14) and (4.5.15) in (4.5.12) we deduce the result. $\square$

We will finish this section by defining the major and minor arcs and showing that the minor arcs do not contribute to the main term. For $\Delta > 0$ we define the *major arcs* to be the set given by

$$\mathfrak{M}(\Delta) := \bigcup_{\substack{q\in\mathbb{N} \\ q \leq P^\Delta}} \bigcup_{\substack{0\leq a_i \leq q \\ (a_1,\dots,a_R,q)=1}} \left\{ \boldsymbol{\alpha} \in [0,1]^R \colon 2\left\| q\boldsymbol{\alpha} - \boldsymbol{a} \right\|_\infty < P_1^{-d_1} P_2^{-d_2} P^\Delta \right\},$$

and the *minor arcs* to be the given by

$$\mathfrak{m}(\Delta) := [0,1]^R \setminus \mathfrak{M}(\Delta).$$

Write further

$$\delta_0 = \frac{\min_{i=1,2} \{ n_1 + n_2 - \dim V_i^* \}}{(\tilde{d}+1)2^{\tilde{d}}R}. \tag{4.5.16}$$

Note that if the forms F_i are linearly independent, then V_i^* are proper subvarieties of $\mathbb{A}_{\mathbb{C}}^{n_1+n_2}$ so that $\dim V_i^* \leq n_1 + n_2 - 1$ whence $\delta_0 \geq \frac{1}{(\tilde{d}+1)2^{\tilde{d}}R}$. To see this for V_1^* note that requiring

$$\operatorname{rank}\left(\frac{\partial F_i}{\partial x_j}\right)_{i,j} < R$$

is equivalent to requiring all the $R \times R$ minors of $\left(\frac{\partial F_i}{\partial x_j}\right)_{i,j}$ vanish. This defines a system of polynomials of degree $R(d_1 + d_2 - 1)$ in $(\boldsymbol{x}, \boldsymbol{y})$, which are not all zero unless there exists $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\boldsymbol{0}\}$ such that

$$\sum_{i=1}^R \beta_i \left(\frac{\partial F_i}{\partial x_j}\right) = 0 \quad \text{for } j = 1, \dots, n_1$$

holds identically in $(\boldsymbol{x}, \boldsymbol{y})$. This is the same as saying that

$$\nabla_{\boldsymbol{x}} \left(\sum_{i=1}^R \beta_i F_i \right) = 0$$

holds identically. From this we find that $\sum_{i=1}^R \beta_i F_i$ must be a form entirely in the $\boldsymbol{y}$ variables. But this is a linear combination of homogeneous bidegree (d_1, d_2) forms with $d_1 \geq 1$ and thus we must in fact have $\sum_{i=1}^R \beta_i F_i = 0$ identically, contradicting linear independence. The argument works analogously for V_2^*.

The next Lemma shows that the assumption (4.5.5) holds with $E = \mathfrak{m}(\Delta)$ and $T(\boldsymbol{\alpha}) = C^{-1} P_1^{-\varepsilon} S(\boldsymbol{\alpha})$.

Lemma 4.5.4. *Let $0 < \Delta \leq R(\tilde{d}+1)(bd_1+d_2)^{-1}$ and let $\varepsilon > 0$. Then we have the upper bound*

$$\sup_{\boldsymbol{\alpha} \in \mathfrak{m}(\Delta)} |S(\boldsymbol{\alpha})| \ll P_1^{n_1} P_2^{n_2} P^{-\Delta\delta_0+\varepsilon}. \tag{4.5.17}$$

Proof. The result follows straightforward from [98, Lemma 4.3] by setting the parameter θ to be

$$\theta = \frac{\Delta}{(\tilde{d}+1)R}.$$

If we have $0 < \Delta \leq R(\tilde{d}+1)(bd_1+d_2)^{-1}$ this ensures that the assumption $0 < \theta \leq (bd_1+d_2)^{-1}$ in [98, Lemma 4.3] is satisfied. □

Before we state the next proposition, recall that we assume $P_1 \geq P_2$ throughout, as was mentioned at the beginning of this section.

Proposition 4.5.5. *Let $\varepsilon > 0$ and let $0 < \Delta \leq R(\tilde{d}+1)(bd_1+d_2)^{-1}$. Under the assumptions of Proposition 4.5.1 we have*

$$\int_{\mathfrak{m}(\Delta)} S(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \ll P_1^{n_1-d_1R} P_2^{n_2-d_2R} P^{-\Delta\delta_0(1-(d_1+d_2)R/\mathscr{C})+\varepsilon}.$$

Proof. We apply Lemma 4.5.2 with

$$T(\boldsymbol{\alpha}) = C^{-1}P^{-\varepsilon}S(\boldsymbol{\alpha}), \quad E_0 = [0,1]^R, \quad E = \mathfrak{m}(\Delta), \quad \text{and} \quad \delta = \Delta\delta_0,$$

where $C > 0$ is some real number. With these choices (4.5.4) follows from the auxiliary inequality (4.5.1) since for any $\varepsilon > 0$ we have $P^{-\varepsilon} \leq P_1^{-\varepsilon}$. From Lemma 4.5.4 we have the bound

$$\sup_{\boldsymbol{\alpha} \in E} CT(\boldsymbol{\alpha}) \ll P_1^{n_1} P_2^{n_2} P^{-\delta}.$$

We may increase C if necessary so that we recover (4.5.5). Therefore the hypotheses of Lemma 4.5.3. Since we assume $\mathscr{C} > (bd_1+d_2)R$, we also note

$$P_2^{-\mathscr{C}} = P^{-R} P^{R-\mathscr{C}(bd_1+d_2)^{-1}} \ll_{\mathscr{C}} P^{-R-\tilde{\delta}},$$

for some $\tilde{\delta} > 0$. Therefore if we assume $\mathscr{C} > (bd_1+d_2)R$ then Lemma 4.5.3 gives

$$\int_{\mathfrak{m}(\Delta)} S(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \ll P_1^{n_1-d_1R} P_2^{n_2-d_2R} P^{-\Delta\delta_0(1-(d_1+d_2)R/\mathscr{C})+\varepsilon},$$

as desired. □

4.5.2 The major arcs

The aim of this section is to identify the main term via integrating the exponential sum $S(\boldsymbol{\alpha})$ over the major arcs, and analyse the singular integral and singular series appropriately. For $\boldsymbol{a} \in \mathbb{Z}^R$ and $q \in \mathbb{N}$ consider the complete exponential sum

$$S_{\boldsymbol{a},q} := q^{-n_1-n_2} \sum_{\boldsymbol{x},\boldsymbol{y}} e\left(\frac{\boldsymbol{a}}{q} \cdot \boldsymbol{F}(\boldsymbol{x},\boldsymbol{y})\right),$$

where the sum $\sum_{\boldsymbol{x},\boldsymbol{y}}$ runs through a complete set of residues modulo q. Further, for $P \geq 1$ and $\Delta > 0$ we define the truncated singular series

$$\mathfrak{S}(P) := \sum_{q \leq P^\Delta} \sum_{\boldsymbol{a}} S_{\boldsymbol{a},q},$$

where the sum $\sum_{\boldsymbol{a}}$ runs over $\boldsymbol{a} \in \mathbb{Z}^R$ such that $0 \leq a_i < q$ for $i = 1, \ldots, R$ and $(a_1, \ldots, a_R, q) = 1$. For $\boldsymbol{\gamma} \in \mathbb{R}^R$ we further define

$$S_\infty(\boldsymbol{\gamma}) := \int_{\mathcal{B}_1 \times \mathcal{B}_2} e\left(\boldsymbol{\gamma} \cdot \boldsymbol{F}(\boldsymbol{u},\boldsymbol{v})\right) d\boldsymbol{u} d\boldsymbol{v},$$

and we define the truncated singular integral for $P \geq 1$, $\Delta > 0$ as follows

$$\mathfrak{I}(P) := \int_{\|\boldsymbol{\gamma}\|_\infty \leq P^\Delta} S_\infty(\boldsymbol{\gamma}) d\boldsymbol{\gamma}.$$

From now on we assume that our parameter $\Delta > 0$ satisfies

$$(bd_1 + d_2)^{-1} > \Delta(2R+3) + \delta \tag{4.5.18}$$

for some $\delta > 0$. Since $\mathscr{C} > R(bd_1 + d_2)$ we are always able to choose such Δ in terms of $\mathscr{C}$. Further as in [98] we now define some slightly modified major arcs $\mathfrak{M}'(\Delta)$ as follows

$$\mathfrak{M}'(\Delta) := \bigcup_{1 \leq q \leq P^\Delta} \bigcup_{\substack{0 \leq a_i < q \\ (a_1, \ldots, a_R, q) = 1}} \mathfrak{M}'_{\boldsymbol{a},q}(\Delta),$$

where $\mathfrak{M}'_{\boldsymbol{a},q}(\Delta) = \left\{\boldsymbol{\alpha} \in [0,1]^R \colon \left\|\boldsymbol{\alpha} - \frac{\boldsymbol{a}}{q}\right\|_\infty < P_1^{-d_1} P_2^{-d_2} P^\Delta\right\}$. The sets $\mathfrak{M}'_{\boldsymbol{a},q}$ are disjoint for our choice of Δ; for if there is some

$$\boldsymbol{\alpha} \in \mathfrak{M}'_{\boldsymbol{a},q}(\Delta) \cap \mathfrak{M}'_{\tilde{\mathbf{a}},\tilde{q}}(\Delta),$$

where $\mathfrak{M}'_{\tilde{\mathbf{a}},\tilde{q}}(\Delta) \neq \mathfrak{M}'_{\boldsymbol{a},q}(\Delta)$ then there is some $i \in \{1, \ldots, R\}$ such that

$$P^{-2\Delta} \leq \frac{1}{q\tilde{q}} \leq \left|\frac{a_i}{q} - \frac{\tilde{a}_i}{\tilde{q}}\right| \leq 2P^{\Delta-1},$$

which is impossible for large P, since by our assumption (4.5.18) we have $3\Delta - 1 < 0$. Further we note that clearly $\mathfrak{M}'(\Delta) \supseteq \mathfrak{M}(\Delta)$ whence $\mathfrak{m}'(\Delta) \subseteq \mathfrak{m}(\Delta)$ and so the conclusions of Proposition 4.5.5 hold with $\mathfrak{m}(\Delta)$ replaced by $\mathfrak{m}'(\Delta)$.

The next result expands the exponential sum $S(\boldsymbol{\alpha})$ when $\boldsymbol{\alpha}$ can be well-approximated by a rational number. In particular for our applications it is important to obtain an error term in which the constant does not depend on $\boldsymbol{\beta}$, whence we cannot just use Lemma 5.3 in [98] as it is stated there.

Lemma 4.5.6. *Let $\Delta > 0$ satisfy (4.5.18), let $\boldsymbol{\alpha} \in \mathfrak{M}'_{\boldsymbol{a},q}(\Delta)$ where $q \leq P^{\Delta}$, and write $\boldsymbol{\alpha} = \boldsymbol{a}/q + \boldsymbol{\beta}$ such that $1 \leq a_i < q$ and $(a_1, \ldots, a_R, q) = 1$. If $P_1 \geq P_2 > 1$ then*

$$S(\boldsymbol{\alpha}) = P_1^{n_1} P_2^{n_2} S_{\boldsymbol{a},q} S_\infty(P\boldsymbol{\beta}) + O\left(q P_1^{n_1} P_2^{n_2-1} \left(1 + P \left\|\boldsymbol{\beta}\right\|_\infty\right)\right), \tag{4.5.19}$$

where the implied constant in the error term does not depend on q or on $\boldsymbol{\beta}$.

Proof. In the sum for $S(\boldsymbol{\alpha})$ we begin by writing $\boldsymbol{x} = \boldsymbol{z}^{(1)} + q\boldsymbol{x}'$ and $\boldsymbol{y} = \boldsymbol{z}^{(2)} + q\boldsymbol{y}'$ where $0 \leq z_i^{(1)} < q$ and $0 \leq z_j^{(2)} < q$ for all $1 \leq i \leq n_1$ and for all $1 \leq j \leq n_2$. A simple calculation now shows

$$\begin{aligned} S(\boldsymbol{\alpha}) &= \sum_{\boldsymbol{x} \in P_1 \mathcal{B}_1} \sum_{\boldsymbol{y} \in P_2 \mathcal{B}_2} e\left(\boldsymbol{\alpha} \cdot \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y})\right) \\ &= \sum_{\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)} \bmod q} e\left(\frac{\boldsymbol{a}}{q} \cdot \boldsymbol{F}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)})\right) \tilde{S}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}) \end{aligned} \tag{4.5.20}$$

where

$$\tilde{S}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}) = \sum_{\boldsymbol{x}', \boldsymbol{y}'} e\left(\boldsymbol{\beta} \cdot \boldsymbol{F}(q\boldsymbol{x}' + \boldsymbol{z}^{(1)}, q\boldsymbol{y}' + \boldsymbol{z}^{(2)})\right),$$

where $\boldsymbol{x}', \boldsymbol{y}'$ in the sum runs through integer tuples such that $q\boldsymbol{x}' + \boldsymbol{z}^{(1)} \in P_1\mathcal{B}_1$ and $q\boldsymbol{y}' + \boldsymbol{z}^{(2)} \in P_2\mathcal{B}_2$ is satisfied. Now consider $\boldsymbol{x}', \boldsymbol{x}''$ and $\boldsymbol{y}', \boldsymbol{y}''$ such that

$$\left\|\boldsymbol{x}' - \boldsymbol{x}''\right\|_\infty, \left\|\boldsymbol{y}' - \boldsymbol{y}''\right\|_\infty \leq 2.$$

Then for all $i = 1, \ldots, R$ we have

$$\begin{aligned} &\left|F_i(q\boldsymbol{x}' + \boldsymbol{z}^{(1)}, q\boldsymbol{y}' + \boldsymbol{z}^{(2)}) - F_i(q\boldsymbol{x}'' + \boldsymbol{z}^{(1)}, q\boldsymbol{y}'' + \boldsymbol{z}^{(2)})\right| \\ &\qquad \ll q P_1^{d_1-1} P_2^{d_2} + q P_1^{d_1} P_2^{d_2-1} \ll q P_1^{d_1} P_2^{d_2-1}, \end{aligned}$$

where we used $P_1 \geq P_2 > 1$ for the last estimate. We note that the implied constant here does not depend on q. We now use this to replace the sum in $\tilde{S}$ by an integral to obtain

$$\tilde{S}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}) = \int_{q\tilde{\boldsymbol{v}} \in P_1\mathcal{B}_1} \int_{q\tilde{\boldsymbol{w}} \in P_2\mathcal{B}_2} e\left(\sum_{i=1}^{R} \beta_i F_i(q\tilde{\boldsymbol{v}}, q\tilde{\boldsymbol{w}}) \right) d\tilde{\boldsymbol{v}} d\tilde{\boldsymbol{w}} \\ + O\left(\|\boldsymbol{\beta}\|_\infty q P_1^{d_1} P_2^{d_2-1} \left(\frac{P_1}{q}\right)^{n_1} \left(\frac{P_2}{q}\right)^{n_2} + \left(\frac{P_1}{q}\right)^{n_1} \left(\frac{P_2}{q}\right)^{n_2-1} \right),$$

where we used that $q \leq P_2$, which is implied by our assumptions, but we mention here for the convenience of the reader. In the integral above we perform a substitution $\boldsymbol{v} = qP_1^{-1}\tilde{\boldsymbol{v}}$ and $\boldsymbol{w} = qP_2^{-1}\tilde{\boldsymbol{w}}$ to get

$$\tilde{S}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}) = P_1^{n_1} P_2^{n_2} q^{-n_1-n_2} \mathfrak{I}(P\boldsymbol{\beta}) + q^{-n_1-n_2} O\left(q P_1^{n_1} P_2^{n_2-1} \left(1 + P\|\boldsymbol{\beta}\|_\infty\right)\right),$$

where the implied constant does not depend on $\boldsymbol{\beta}$ or q. Substituting this into (4.5.20) gives the result. ☐

From the Lemma und using that the sets $\mathfrak{M}'_{\boldsymbol{a},q}$ are disjoint we deduce

$$\int_{\mathfrak{M}'(\Delta)} S(\boldsymbol{\alpha}) d\boldsymbol{\alpha} = P_1^{n_1} P_2^{n_2} \sum_{1 \leq q \leq P^\Delta} \sum_{\boldsymbol{a}} S_{\boldsymbol{a},q} \int_{|\boldsymbol{\beta}|} S_\infty(P\boldsymbol{\beta}) d\boldsymbol{\beta} \\ + O\left(P_1^{n_1} P_2^{n_2} P^{2\Delta} P_2^{-1} \mathrm{meas}\left(\mathfrak{M}'(\Delta)\right)\right), \quad (4.5.21)$$

where we used $q \leq P^\Delta$ and $P\|\boldsymbol{\beta}\|_\infty \leq P^\Delta$ for the error term. Now we can bound the measure of the major arcs by

$$\mathrm{meas}(\mathfrak{M}'(\Delta)) \ll \sum_{q \leq P^\Delta} q^R P^{-R+\Delta R} \ll P^{-R+\Delta(2R+1)}.$$

Using this and making the substitution $\boldsymbol{\gamma} = P\boldsymbol{\beta}$ in the integral in (4.5.21) we find

$$\int_{\mathfrak{M}'(\Delta)} S(\boldsymbol{\alpha}) d\boldsymbol{\alpha} = P_1^{n_1} P_2^{n_2} P^{-R} \mathfrak{S}(P) \mathfrak{I}(P) \\ + O\left(P_1^{n_1} P_2^{n_2} P^{-R+\Delta(2R+3)-1/(bd_1+d_2)}\right). \quad (4.5.22)$$

It becomes transparent why the assumption (4.5.18) is in place, because then the error term in (4.5.22) is bounded by $O(P_1^{n_1} P_2^{n_2} P^{-R-\delta})$ and thus is of smaller order than the main term.

We now focus on the singular series $\mathfrak{S}(P)$ and the singular integral $\mathfrak{I}(P)$ in the next two Lemmas.

Lemma 4.5.7. *Let $\varepsilon > 0$ and assume that the bound* (4.5.1) *holds for some $C \geq 1$, $\mathscr{C} > 1 + b\varepsilon$, for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^R$ and all real $P_1 \geq P_2 > 1$. Then we have the following:*

(i) *For all* $\varepsilon' > 0$ *such that* $\varepsilon' = O_{\mathscr{C}}(\varepsilon)$ *we have*

$$\min\{|S_{\boldsymbol{a},q}|, |S_{\boldsymbol{a}',q'}|\} \ll_C (q'+q)^{\varepsilon} \left\| \frac{\boldsymbol{a}}{q} - \frac{\boldsymbol{a}'}{q'} \right\|_{\infty}^{\frac{\mathscr{C}-\varepsilon'}{\tilde{d}+1}}$$

for all $q, q' \in \mathbb{N}$ *and all* $\boldsymbol{a} \in \{1, \dots, q\}^R$ *and* $\boldsymbol{a}' \in \{1, \dots, q'\}^R$ *with* $\frac{\boldsymbol{a}}{q} \neq \frac{\boldsymbol{a}'}{q'}$.

(ii) *If* $\mathscr{C} > \varepsilon'$ *then for all* $t \in \mathbb{R}_{>0}$ *and* $q_0 \in \mathbb{N}$ *we have*

$$\#\left\{ \frac{\boldsymbol{a}}{q} \in [0,1]^R \cap \mathbb{Q}^R \colon q \leq q_0, |S_{\boldsymbol{a},q}| \geq t \right\} \ll_C (q_0^{-\varepsilon} t)^{-\frac{(\tilde{d}+1)R}{\mathscr{C}-\varepsilon'}},$$

where the fractions in the set above are in lowest terms.

(iii) *Assume that the forms* $F_i(\boldsymbol{x}, \boldsymbol{y})$ *are linearly independent. Then for all* $q \in \mathbb{N}$ *and* $\boldsymbol{a} \in \mathbb{Z}^R$ *with* $(a_1, \dots, a_R, q) = 1$ *there exists some* $\nu > 0$ *depending at most on* d_i *and* R *such that*

$$|S_{\boldsymbol{a},q}| \ll q^{-\nu}.$$

(iv) *Assume* $\mathscr{C} > (\tilde{d}+1)R$ *and assume the forms* $F_i(\boldsymbol{x}, \boldsymbol{y})$ *are linearly independent. Then the* singular series

$$\mathfrak{S} = \sum_{q=1}^{\infty} \sum_{\boldsymbol{a} \bmod q} S_{\boldsymbol{a},q} \tag{4.5.23}$$

exists and converges absolutely, with

$$|\mathfrak{S}(P) - \mathfrak{S}| \ll_{C,\mathscr{C}} P^{-\Delta\delta_1},$$

for some $\delta_1 > 0$ *depending only on* $\mathscr{C}, d_i$ *and* R.

Proof of (i). Take $\mathcal{B}_i = [0,1]^{n_i}$ so that $S_\infty(\mathbf{0}) = 1$. Therefore (4.5.19) implies that

$$\frac{S\left(\frac{\boldsymbol{a}}{q}\right)}{P_1^{n_1} P_2^{n_2}} = S_{\boldsymbol{a},q} + O\left(q P_2^{-1}\right) \quad \text{and} \quad \frac{S\left(\frac{\boldsymbol{a}'}{q'}\right)}{P_1^{n_1} P_2^{n_2}} = S_{\boldsymbol{a}',q'} + O\left(q' P_2^{-1}\right).$$

Using this and the bound (4.5.1) we obtain

$$\min\{|S_{\boldsymbol{a},q}|, |S_{\boldsymbol{a}',q'}|\} \leq C P_1^{\varepsilon} P^{-\mathscr{C}} \left\| \frac{\boldsymbol{a}}{q} - \frac{\boldsymbol{a}'}{q'} \right\|_{\infty}^{-\mathscr{C}} + C P_1^{\varepsilon} \left\| \frac{\boldsymbol{a}}{q} - \frac{\boldsymbol{a}'}{q'} \right\|_{\infty}^{\frac{\mathscr{C}}{\tilde{d}+1}} + O\left((q'+q) P_2^{-1}\right), \tag{4.5.24}$$

where we note that $P_1^{\varepsilon} P_2^{-\mathscr{C}} = O(P_2^{-1})$ due to our assumptions on $\mathscr{C}$. Now set

$$P_1 = P_2 = (q+q') \left\| \frac{\boldsymbol{a}}{q} - \frac{\boldsymbol{a}'}{q'} \right\|_{\infty}^{-\frac{1+\mathscr{C}}{\tilde{d}+1}}.$$

Note $(q+q')\geq 1$ and $\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty \leq 1$ so that this gives $P_i\geq 1$. Substituting these choices into (4.5.24) we get

$$\min\left\{|S_{\boldsymbol{a},q}|,|S_{\boldsymbol{a}',q'}|\right\}\leq P_1^\varepsilon (q+q')^{-\mathscr{C}(d_1+d_2)}\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty^{\frac{\mathscr{C}^2+\mathscr{C}}{\tilde{d}+1}(d_1+d_2)-\mathscr{C}}+$$
$$CP_1^\varepsilon\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty^{\frac{\mathscr{C}}{\tilde{d}+1}}+O\left(\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty^{\frac{1+\mathscr{C}}{(\tilde{d}+1)}}\right).$$

Noting again that $(q+q')\geq 1$, $\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty\leq 1$ and also that $\frac{\mathscr{C}^2+\mathscr{C}}{\tilde{d}+1}(d_1+d_2)-\mathscr{C}\geq\frac{\mathscr{C}}{\tilde{d}+1}$ we see that the second term on the right hand side above dominates the expression. Hence we finally obtain

$$\min\left\{|S_{\boldsymbol{a},q}|,|S_{\boldsymbol{a}',q'}|\right\}\ll_C P_1^\varepsilon\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty^{\frac{\mathscr{C}}{\tilde{d}+1}}=(q'+q)^\varepsilon\left\|\frac{\boldsymbol{a}}{q}-\frac{\boldsymbol{a}'}{q'}\right\|_\infty^{\frac{\mathscr{C}-\varepsilon'}{\tilde{d}+1}},$$

for some $\varepsilon'=O_{\mathscr{C}}(\varepsilon)$. □

Proof of (ii). This now follows almost directly from (i). The points in the set

$$\left\{\frac{\boldsymbol{a}}{q}\in[0,1]^R\cap\mathbb{Q}^R\colon q\leq q_0, |S_{\boldsymbol{a},q}|\geq t\right\}$$

are separated by gaps of size $\gg_C (q_0^{-\varepsilon}t)^{\frac{\tilde{d}+1}{\mathscr{C}-\varepsilon'}}$. Hence at most $O_C((q_0^{-\varepsilon}t)^{-\frac{\tilde{d}+1}{\mathscr{C}-\varepsilon'}})$ fit in the box $[0,1]^R$ so the result follows. □

Proof of (iii). Setting $P_1=P_2=q$ and $\boldsymbol{\alpha}=\boldsymbol{a}/q$ we find $S_{\boldsymbol{a},q}=q^{-n_1-n_2}S(\boldsymbol{\alpha})$. Let δ_0 be defined as in (4.5.16). We can define Δ by $(d_1+d_2)\Delta=1-\varepsilon''$ for some $\varepsilon''\in(0,1)$. We claim that $\boldsymbol{a}/q$ does not lie in the major arcs $\mathfrak{M}(\Delta)$ if $(a_1,\dots,a_r,q)=1$. For if, then there exist $q',\boldsymbol{a}'$ such that

$$1\leq q'\leq q^{(d_1+d_2)\Delta},$$

and

$$2\left|q'a_i-qa_i'\right|\leq q^{1-d_1-d_2}q^{(d_1+d_2)\Delta}<1,$$

which is clearly impossible. The bound (4.5.17) applied to our situation gives

$$|S_{\boldsymbol{a},q}|\ll q^{-R\delta_0(1-\varepsilon'')+\varepsilon}.$$

As the forms F_i are linearly independent we know that $\delta_0\geq\frac{1}{(\tilde{d}+1)2^{\tilde{d}}R}$. Thus, choosing some small enough ε delivers the result. □

Proof of (iv). For $Q > 0$ let

$$s(Q) = \sum_{\substack{\boldsymbol{a}/q \in [0,1)^R \\ Q < q \leq 2Q}} |S_{\boldsymbol{a},q}|,$$

where $\sum_{\boldsymbol{a}/q \in [0,1)^R}$ is shorthand for the sum running over $\sum_{q=1}^{\infty} \sum_{\|\boldsymbol{a}\|_\infty \leq q}$ such that $(a_1, \ldots, a_R, q) = 1$. We claim that $s(Q) \ll_{C,\mathscr{C}} Q^{-\delta_1}$ for some $\delta_1 > 0$. To see this, let $\ell \in \mathbb{Z}$. Then

$$\begin{aligned} s(Q) = &\sum_{\substack{\boldsymbol{a}/q \in [0,1)^R \\ Q < q \leq 2Q \\ |S_{\boldsymbol{a},q}| \geq 2^{-\ell}}} |S_{\boldsymbol{a},q}| + \sum_{i=\ell}^{\infty} \sum_{\substack{\boldsymbol{a}/q \in [0,1)^R \\ Q < q \leq 2Q \\ 2^{-i} > |S_{\boldsymbol{a},q}| \geq 2^{-i-1}}} |S_{\boldsymbol{a},q}| \\ &\leq \# \left\{ \frac{\boldsymbol{a}}{q} \in [0,1)^R \cap \mathbb{Q}^R \colon q \leq 2Q, |S_{\boldsymbol{a},q}| \geq 2^{-\ell} \right\} \cdot \sup_{q > Q} |S_{\boldsymbol{a},q}| \\ &\quad + \sum_{i=\ell}^{\infty} \# \left\{ \frac{\boldsymbol{a}}{q} \in [0,1)^R \cap \mathbb{Q}^R \colon q \leq 2Q, |S_{\boldsymbol{a},q}| \geq 2^{-i-1} \right\} \cdot 2^{-i}. \end{aligned} \tag{4.5.25}$$

Now from (ii) we know

$$\# \left\{ \frac{\boldsymbol{a}}{q} \in [0,1)^R \cap \mathbb{Q}^R \colon q \leq 2Q, |S_{\boldsymbol{a},q}| \geq t \right\} \ll_C (Q^{-\varepsilon} t)^{-\frac{(\tilde{d}+1)R}{\mathscr{C} - \varepsilon'}},$$

and from (iii) we know, since F_i are linearly independent there is some $\nu > 0$ such that

$$\sup_{q > Q} |S_{\boldsymbol{a},q}| \ll Q^{-\nu}.$$

Using these estimates in (4.5.25) we get

$$s(Q) \ll_C Q^{O_{\mathscr{C}}(\varepsilon) - \nu} 2^{\ell \frac{(\tilde{d}+1)R}{\mathscr{C} - \varepsilon'}} + Q^{O_{\mathscr{C}}(\varepsilon)} \sum_{i=\ell}^{\infty} 2^{(i+1)\left(\frac{(\tilde{d}+1)R}{\mathscr{C} - \varepsilon''}\right) - i}.$$

Since we assumed $\mathscr{C} > (\tilde{d}+1)R$ and since ε' is small in terms of $\mathscr{C}$ we may also assume $\mathscr{C} > (\tilde{d}+1)R + \varepsilon'$. Therefore, summing the geometric expression gives

$$s(Q) \ll_{C,\mathscr{C}} Q^{O_{\mathscr{C}}(\varepsilon)} 2^{\ell \frac{(\tilde{d}+1)R}{\mathscr{C} - \varepsilon'}} \left(Q^{-\nu} + 2^{-\ell} \right).$$

Now choose $\ell = \lfloor \log_2 Q^{\nu} \rfloor$ to get

$$s(Q) \ll Q^{\nu \frac{(\tilde{d}+1)R - \mathscr{C}}{\mathscr{C}} + O_{\mathscr{C}}(\varepsilon)}.$$

Letting ε be small enough in terms of $\mathscr{C}, d_i$, R we get some $\delta_1 > 0$ depending on $\mathscr{C}, d_i$ and R such that

$$s(Q) \ll Q^{-\delta_1},$$

which proves the claim. Finally using this and splitting the sum into dyadic intervals we find

$$|\mathfrak{S}(P)-\mathfrak{S}| \leq \sum_{\substack{\boldsymbol{a}/q\in[0,1)^R\\ q>P^\Delta}} |S_{\boldsymbol{a},q}| = \sum_{k=0}^{\infty} \sum_{Q=2^kP^\Delta} s(Q) \ll \sum_{k=0}^{\infty} \left(2^kP^\Delta\right)^{-\delta_1},$$

which proves (iv). □

The next Lemma handles the singular integral.

Lemma 4.5.8. *Let $\varepsilon>0$ and assume that the bound* (4.5.1) *holds for some $C\geq 1$, $\mathscr{C}>1+b\varepsilon$ and for all $\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{R}^R$ and all real $P_1\geq P_2>1$. Then:*

(i) For all $\boldsymbol{\gamma}\in\mathbb{R}^R$ we have

$$S_\infty(\boldsymbol{\gamma}) \ll_C \|\boldsymbol{\gamma}\|_\infty^{-\mathscr{C}+\varepsilon'},$$

for some $\varepsilon>0$ such that $\varepsilon'=O_{\mathscr{C}}(\varepsilon)$.

(ii) Assume that $\mathscr{C}-\varepsilon'>R$. Then for all $P_1,P_2>1$ we have

$$|\mathfrak{I}(P)-\mathfrak{I}| \ll_{\mathscr{C},C,\varepsilon'} P^{-\Delta(\mathscr{C}-\varepsilon'-R)},$$

where $\mathfrak{I}$ is the singular integral

$$\mathfrak{I} = \int_{\boldsymbol{\gamma}\in\mathbb{R}^R} S_\infty(\boldsymbol{\gamma})d\boldsymbol{\gamma}. \tag{4.5.26}$$

In particular we see that $\mathfrak{I}$ exists and converges absolutely.

Proof of (i). It is easy to see that for all $\boldsymbol{\beta}\in\mathbb{R}^R$ we have $|S(\boldsymbol{\beta})|\leq|S(\mathbf{0})|$. Thus applying (4.5.1) with $\boldsymbol{\alpha}=\mathbf{0}$ and $\boldsymbol{\beta}=P^{-1}\boldsymbol{\gamma}$ we get

$$\left|S(P^{-1}\boldsymbol{\gamma})\right| \leq CP_1^{n_1}P_2^{n_2}P_1^{\varepsilon}\max\left\{P_2^{-1}\|\boldsymbol{\gamma}\|_\infty^{-1}, P^{-\frac{1}{\tilde{d}+1}}\|\boldsymbol{\gamma}\|^{\frac{1}{\tilde{d}+1}}\right\}^{\mathscr{C}}. \tag{4.5.27}$$

Now from (4.5.19) with $\boldsymbol{a}=\mathbf{0}$ and $\boldsymbol{\beta}=P^{-1}\boldsymbol{\gamma}$ we have

$$S(P^{-1}\boldsymbol{\gamma}) = P_1^{n_1}P_2^{n_2}S_\infty(\boldsymbol{\gamma}) + O\left(P_1^{n_1}P_2^{n_2-1}(1+\|\boldsymbol{\gamma}\|_\infty)\right), \tag{4.5.28}$$

where we used as in the proof of part (i) Lemma 4.5.7 that $P_1^\varepsilon P_2^{-\mathscr{C}}\leq P_2^{-1}$ due to our assumptions on $\mathscr{C}$. Combining (4.5.27) and (4.5.28) we obtain

$$S_\infty(\boldsymbol{\gamma}) \ll_C P_1^\varepsilon\max\left\{\|\boldsymbol{\gamma}\|_\infty^{-1}, P^{-\frac{1}{\tilde{d}+1}}\|\boldsymbol{\gamma}\|_\infty^{\frac{1}{\tilde{d}+1}}\right\}^{\mathscr{C}} + P_2^{-1} + \|\boldsymbol{\gamma}\|_\infty P_2^{-1}.$$

Taking $P_1=P_2=\max\{1,\|\boldsymbol{\gamma}\|_\infty^{1+\mathscr{C}}\}$ gives the result. □

Proof of (ii). For this simply note that by part (i) we get

$$|\mathfrak{I}(P) - \mathfrak{I}| = \int_{\|\boldsymbol{\gamma}\|_\infty \geq P^\Delta} S_\infty(\boldsymbol{\gamma}) d\boldsymbol{\gamma} \ll_{\mathscr{C},C,\varepsilon'} \int_{\|\boldsymbol{\gamma}\|_\infty \geq P^\Delta} \|\boldsymbol{\gamma}\|_\infty^{-\mathscr{C}-\varepsilon'} d\boldsymbol{\gamma} \ll P^{-\Delta(\mathscr{C}-\varepsilon'-R)},$$

where the last estimate follows since we assumed $\mathscr{C} - \varepsilon' > R$. □

Before we finish the proof of the main result we state two different expressions for the singular series and the singular integral that will be useful later on. If $\mathscr{C} > R(d_1 + d_2)$ then $\mathfrak{I}$ and $\mathfrak{S}$ converge absolutely, as was shown in the previous two Lemmas. Therefore, as in §7 of [9], by regarding the bihomogeneous forms under investigation simply as homogeneous forms we may express the singular series as an absolutely convergent product

$$\mathfrak{S} = \prod_p \mathfrak{S}_p, \tag{4.5.29}$$

where

$$\mathfrak{S}_p = \lim_{k\to\infty} \frac{1}{p^{k(n_1+n_2-R)}} \# \left\{ (\boldsymbol{u}, \boldsymbol{v}) \in \{1, \dots, p^k\}^{n_1+n_2} : F_i(\boldsymbol{u}, \boldsymbol{v}) \equiv 0 \ (\mathrm{mod}\, p), i = 1, \dots, R \right\}.$$

Lemma 2.6 in [94] further shows that we can write the singular integral as

$$\mathfrak{I} = \lim_{P\to\infty} \frac{1}{P^{n_1+n_2-(d_1+d_2)R}} \mu\big\{(\boldsymbol{t}_1, \boldsymbol{t}_2)/P \in \mathcal{B}_1 \times \mathcal{B}_2 : \\ |F_i(\boldsymbol{t}_1, \boldsymbol{t}_2)| \leq 1/2,\ i = 1, \dots, R\big\}, \tag{4.5.30}$$

where $\mu(\cdot)$ denotes the Lebesgue measure. We may therefore interpret the quantities $\mathfrak{I}$ and $\mathfrak{S}_p$ as the real and p-adic *densities*, respectively, of the system of equations $F_1(\boldsymbol{x}, \boldsymbol{y}) = \cdots = F_R(\boldsymbol{x}, \boldsymbol{y}) = 0$.

4.5.3 Proofs of Proposition 4.5.1 and Theorem 4.2.1

Proof of Proposition 4.5.1. From Proposition 4.5.5, the estimate (4.5.22), Lemma 4.5.7 and Lemma 4.5.8, for any $\varepsilon > 0$ we find

$$\frac{N(P_1, P_2)}{P_1^{n_1} P_2^{n_2} P^{-R}} - \mathfrak{S}\mathfrak{I} \ll \\ O\left(P^{-\Delta\delta_1} + P^{-\Delta\delta_0(1-(d_1+d_2)R/\mathscr{C})+\varepsilon} + P^{(2R+3)\Delta - 1/(bd_1+d_2)} + P^{-\Delta(\mathscr{C}-\varepsilon'-R)}\right).$$

for some $\delta_1 > 0$ and some $1 > \varepsilon' > 0$. Recall we assumed $\mathscr{C} > (bd_1 + d_2)R$ and assuming the forms F_i are linearly independent we also have $\delta_0 \geq \frac{1}{(\tilde{d}+1)2^{\tilde{d}}R}$. Therefore choosing suitably small $\Delta > 0$ there exists some $\delta > 0$ such that

$$\frac{N(P_1, P_2)}{P_1^{n_1} P_2^{n_2} P^{-R}} - \mathfrak{S}\mathfrak{I} \ll P^{-\delta}$$

as desired. Finally, since we assume that the equations F_i define a complete intersection, it is a standard fact to see that $\mathfrak{S}$ is positive if there exists a non-singular p-adic zero for all primes P, and similarly $\mathfrak{J}$ is positive if there exists a non-singular real zero within $\mathcal{B}_1 \times \mathcal{B}_2$. A detailed argument of this fact using a version of Hensel's Lemma for $\mathfrak{S}$ and the implicit function theorem for $\mathfrak{J}$ can be found for example in §4 of [94]. □

We finish this section by deducing the technical main theorem, namely Theorem 4.2.1.

Proof of Theorem 4.2.1. Assume the estimate in (4.2.1) holds for some constant $C_0 > 0$. From Proposition 4.4.4 it thus follows that the auxiliary inequality (4.5.1) holds with a constant $C > 0$ depending on C_0, d_i, n_i, μ and M, where all of these quantities follow the same notation as in Section 4.4. Therefore the assumptions of Proposition 4.5.1 so we can apply it to obtain the desired conclusions. □

4.6 Systems of bilinear forms – the proof of Theorem 4.1.1

In this section we assume $d_1 = d_2 = 1$. Then we can write our system as

$$F_i(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^T A_i \boldsymbol{x},$$

where A_i are $n_2 \times n_1$-dimensional matrices with integer entries. For $\boldsymbol{\beta} \in \mathbb{R}^R$ we now have

$$\boldsymbol{\beta} \cdot \boldsymbol{F} = \boldsymbol{y}^T A_{\boldsymbol{\beta}} \boldsymbol{x},$$

where $A_{\boldsymbol{\beta}} = \sum_i \beta_i A_i$. Recall that we put

$$\sigma_{\mathbb{R}}^{(1)} = \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \dim \ker(A_{\boldsymbol{\beta}}) \quad \text{and} \quad \sigma_{\mathbb{R}}^{(2)} = \max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \dim \ker(A_{\boldsymbol{\beta}}^T).$$

Since the row rank of a matrix is equal to its column rank we can also define

$$\rho_{\mathbb{R}} := \min_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \operatorname{rank}(A_{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \operatorname{rank}(A_{\boldsymbol{\beta}}^T).$$

Due to the rank-nullity theorem the conditions

$$n_i - \sigma_{\mathbb{R}}^{(i)} > (2b+2)R$$

for $i = 1, 2$ are equivalent to

$$\rho_{\mathbb{R}} > (2b+2)R.$$

Lemma 4.6.1. *Assume that $\mathbb{V}(F_1, \dots, F_R) \subset \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ is a smooth complete intersection. Let $b \geq 1$ be a real number. Assume further*

$$\min\{n_1, n_2\} > (2b+2)R, \quad \textit{and} \quad n_1 + n_2 > (4b+5)R. \tag{4.6.1}$$

Then we have

$$n_i - \sigma_{\mathbb{R}}^{(i)} > (2b+2)R \tag{4.6.2}$$

for $i = 1, 2$.

Proof. Without loss of generality assume $n_1 \geq n_2$. Pick $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\mathbf{0}\}$ such that $\operatorname{rank}(A_{\boldsymbol{\beta}}) = \rho_{\mathbb{R}}$. In particular then

$$\dim \ker(A_{\boldsymbol{\beta}}) = \sigma_{\mathbb{R}}^{(1)}, \quad \text{and} \quad \dim \ker(A_{\boldsymbol{\beta}}^T) = \sigma_{\mathbb{R}}^{(2)}.$$

We proceed in distinguishing two cases. Firstly, if $\sigma_{\mathbb{R}}^{(2)} = 0$ then (4.6.2) follows for $i = 2$ by the assumption (4.6.1). Further by comparing row rank and column rank of $A_{\boldsymbol{\beta}}$ in this case we must then have $\sigma_{\mathbb{R}}^{(1)} \leq n_1 - n_2$, and therefore

$$n_1 - \sigma_{\mathbb{R}}^{(1)} \geq n_2 > (2b+2)R,$$

so (4.6.2) follows for $i = 1$.

Now we turn to the case $\sigma_{\mathbb{R}}^{(2)} > 0$. Then also $\sigma_{\mathbb{R}}^{(1)} > 0$. The singular locus of the variety $\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \subset \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ is given by

$$\operatorname{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) = \mathbb{V}(\boldsymbol{y}^T A_{\boldsymbol{\beta}}) \cap \mathbb{V}(A_{\boldsymbol{\beta}} \boldsymbol{x}).$$

Therefore we have

$$\dim \operatorname{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) = \sigma_{\mathbb{R}}^{(1)} + \sigma_{\mathbb{R}}^{(2)} - 2.$$

Since we assumed $\mathbb{V}(\boldsymbol{F})$ to be a smooth complete intersection we can apply Lemma 4.3.3 to get $\dim \operatorname{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \leq R - 2$. Therefore we find

$$\sigma_{\mathbb{R}}^{(1)} + \sigma_{\mathbb{R}}^{(2)} \leq R.$$

From our previous remarks we know that showing (4.6.2) is equivalent to showing $\rho_{\mathbb{R}} > (2b+2)R$. But now

$$\rho_{\mathbb{R}} = \frac{1}{2}\left(n_1 + n_2 - \sigma_{\mathbb{R}}^{(1)} - \sigma_{\mathbb{R}}^{(2)}\right) \geq \frac{1}{2}(n_1 + n_2 - R) > (2b+2)R,$$

where the last inequality followed from the assumption (4.6.1). Therefore (4.6.2) follows as desired. □

Proof of Theorem 4.1.1. Recall the notation $b = \frac{\log P_1}{\log P_2}$. By virtue of Theorem 4.2.1 it suffices to show that assuming

$$n_i - \sigma_{\mathbb{R}}^{(i)} > (2b+2)R$$

for $i = 1, 2$ implies (4.2.1). We will show (4.2.1) for $i = 1$, the other case follows analogously. Let $\mathscr{C} = \frac{n_2 - \sigma_{\mathbb{R}}^{(2)}}{2}$ and we note that we have $\mathscr{C} > (bd_1 + d_2)R = (b+1)R$ precisely when $n_2 - \sigma_{\mathbb{R}}^{(2)} > (2b+2)R$ holds. Therefore it suffices to show that

$$N_1^{\mathrm{aux}}(\boldsymbol{\beta}, B) \ll B^{\sigma_{\mathbb{R}}^{(2)}}. \tag{4.6.3}$$

for all $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\mathbf{0}\}$ with the implied constant not depending on $\boldsymbol{\beta}$. In our case we have

$$\boldsymbol{\Gamma}(\boldsymbol{u}) = \boldsymbol{u}^T A(\boldsymbol{\beta}),$$

where $\boldsymbol{u} \in \mathbb{Z}^{n_2}$. Therefore $N_1^{\mathrm{aux}}(\boldsymbol{\beta}, B)$ counts vectors $\boldsymbol{u} \in \mathbb{Z}^{n_2}$ such that

$$\|\boldsymbol{u}\|_\infty \leq B \quad \text{and} \quad \left\|\boldsymbol{u}^T A(\boldsymbol{\beta})\right\|_\infty \leq \|A(\boldsymbol{\beta})\|_\infty = \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty.$$

In particular, all of the vectors $\boldsymbol{u} \in \mathbb{Z}^{n_2}$, which are counted by $N_1^{\mathrm{aux}}(\boldsymbol{\beta}, B)$ are contained in the ellipsoid

$$E_{\boldsymbol{\beta}} \coloneqq \left\{\boldsymbol{t} \in \mathbb{R}^{n_2} \colon \boldsymbol{t}^T A_{\boldsymbol{\beta}} A_{\boldsymbol{\beta}}^T \boldsymbol{t} < n_2 \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty^2\right\}.$$

The principal radii of $E_{\boldsymbol{\beta}}$ are given by $|\lambda_i|^{-1} n_2^{1/2} \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty$ for $i = 1, \ldots, n_2$, where λ_i run through the n_2 singular values of $A_{\boldsymbol{\beta}}$ and are listed in increasing order of absolute value. Thus we find

$$N_1^{\mathrm{aux}}(\boldsymbol{\beta}, B) \ll \prod_{i=1}^{n_2} \min\left\{|\lambda_i|^{-1} \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty + 1, B\right\}.$$

If $\left|\lambda_{\sigma_{\mathbb{R}}^{(2)}+1}\right| \gg \|\boldsymbol{\beta} \cdot \boldsymbol{F}\|_\infty$ holds then (4.6.3) would follow. So suppose for a contradiction that there exists a sequence $(\boldsymbol{\beta}^{(i)})$ such that $\left|\lambda_{\sigma_{\mathbb{R}}^{(2)}+1}\right| = o\left(\left\|\boldsymbol{\beta}^{(i)} \cdot \boldsymbol{F}\right\|_\infty\right)$. Let $\boldsymbol{\beta}$ be the limit of a subsequence of $\boldsymbol{\beta}^{(i)} / \left\|\boldsymbol{\beta}^{(i)}\right\|$, which must exist by the Bolzano–Weierstrass theorem. For this $\boldsymbol{\beta}$ we must then have $\lambda_{\sigma_{\mathbb{R}}^{(2)}+1} = 0$. Since the singular values were listed in order of increasing absolute value it follows that

$$\lambda_1 = \cdots = \lambda_{\sigma_{\mathbb{R}}^{(2)}+1} = 0,$$

and so $\dim \ker A_{\boldsymbol{\beta}}^T = \sigma_{\mathbb{R}}^{(2)} + 1$. This contradicts the maximality of $\sigma_{\mathbb{R}}^{(2)} + 1$.

The second part of the theorem is now a direct consequence of Lemma 4.6.1. □

4.7 Systems of forms of bidegree $(2,1)$

We consider a system $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y})$ of homogeneous equations of bidegree $(2,1)$, where $\boldsymbol{x}=(x_1,\ldots,x_{n_1})$ and $\boldsymbol{y}=(y_1,\ldots,y_{n_2})$. We will first assume $n_1=n_2=n$, say, and then deduce Theorem 4.1.2 afterwards. Therefore the initial main goal is to establish the following.

Proposition 4.7.1. *Let $F_1(\boldsymbol{x},\boldsymbol{y}),\ldots,F_R(\boldsymbol{x},\boldsymbol{y})$ be bihomogeneous forms of bidegree $(2,1)$ such that the biprojective variety $\mathbb{V}(F_1,\ldots,F_R)\subset\mathbb{P}_{\mathbb{Q}}^{n-1}\times\mathbb{P}_{\mathbb{Q}}^{n-1}$ is a complete intersection. Write $b=\max\{\log P_1/\log P_2,1\}$ and $u=\max\{\log P_2/\log P_1,1\}$ Assume that*

$$n-s_{\mathbb{R}}^{(i)}>(8b+4u)R \tag{4.7.1}$$

holds for $i=1,2$, where $s_{\mathbb{R}}^{(i)}$ are as defined in (4.1.7) and (4.1.8). Then there exists some $\delta>0$ depending at most on $\boldsymbol{F}$, R, n, b and u such that we have

$$N(P_1,P_2)=\sigma P_1^{n-2R}P_2^{n-R}+O(P_1^{n-2R}P_2^{n-R}\min\{P_1,P_2\}^{-\delta})$$

where $\sigma>0$ if the system $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y})=\mathbf{0}$ has a smooth p-adic zero for all primes p and a smooth real zero in $\mathcal{B}_1\times\mathcal{B}_2$.

If we assume that $\mathbb{V}(F_1,\ldots,F_R)\subset\mathbb{P}_{\mathbb{Q}}^{n-1}\times\mathbb{P}_{\mathbb{Q}}^{n-1}$ is smooth, then the same conclusions hold if we assume

$$n>(16b+8u+1)R$$

instead of (4.7.1).

For $r=1,\ldots,R$ we can write each form $F_r(\boldsymbol{x},\boldsymbol{y})$ as

$$F_r(\boldsymbol{x},\boldsymbol{y})=\sum_{i,j,k}F_{ijk}^{(r)}x_ix_jy_k,$$

where the coefficients $F_{ijk}^{(r)}$ are symmetric in i and j. In particular, for any $r=1,\ldots,R$ we have an $n\times n$ matrix given by $H_r(\boldsymbol{y})=(\sum_k F_{ijk}^{(r)}y_k)_{ij}$ whose entries are linear homogeneous polynomials in $\boldsymbol{y}$. We may thus also write each equation in the form

$$F_r(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}^TH_r(\boldsymbol{y})\boldsymbol{x}.$$

The strategy of the proof of Proposition 4.7.1 is the same as in the bilinear case, however this time more techincal arguments are required. We need to obtain a good upper bound for the counting functions $N_i^{\mathrm{aux}}(\boldsymbol{\beta};B)$ so that we can apply Theorem 4.2.1. For $\boldsymbol{\beta}\in\mathbb{R}^R$ we consider $\boldsymbol{\beta}\cdot\boldsymbol{F}$, which we can rewrite in our case as

$$\boldsymbol{\beta}\cdot\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}^TH_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}$$

where $H_{\boldsymbol{\beta}}(\boldsymbol{y}) = \sum_{i=1}^{R} \beta_i H_i(\boldsymbol{y})$ is a symmetric $n \times n$ matrix whose entries are linear and homogeneous in $\boldsymbol{y}$. The associated multilinear form $\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{y})$ is thus given by

$$\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{y}) = 2\left(\boldsymbol{x}^{(1)}\right)^T H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}^{(2)}.$$

Recall $N_1^{\mathrm{aux}}(\boldsymbol{\beta}, B)$ counts integral tuples $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^n$ satisfying $\|\boldsymbol{x}\|_\infty, \|\boldsymbol{y}\|_\infty \leq B$ and

$$\left\|(\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}, \boldsymbol{e}_1, \boldsymbol{y}), \ldots, \Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}, \boldsymbol{e}_n, \boldsymbol{y}))^T\right\|_\infty = 2\|H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}\|_\infty \leq \|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty B.$$

Now $N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)$ counts integral tuples $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$ with $\left\|\boldsymbol{x}^{(1)}\right\|_\infty, \left\|\boldsymbol{x}^{(2)}\right\|_\infty \leq B$ and

$$\left\|\left(\Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{e}_1), \ldots, \Gamma_{\boldsymbol{\beta}\cdot\boldsymbol{F}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{e}_n)\right)^T\right\|_\infty \leq \|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty B.$$

We may rewrite this as saying that

$$\left\|\boldsymbol{x}^{(1)} H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x}^{(2)}\right\| \leq \|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty B$$

is satisfied for $\ell = 1, \ldots, n$.

As in the proof of Theorem 4.1.1 using Proposition 4.4.4 and Proposition 4.5.1 we find that for the proof of Theorem 4.7.1 it is enough to show that there exists a positive constant C_0 such that for all $B \geq 1$ and all $\boldsymbol{\beta} \in \mathbb{R}^r \setminus \{0\}$ we have

$$N_i^{\mathrm{aux}}(\boldsymbol{\beta}; B) \leq C_0 B^{2n-4\mathscr{C}}$$

for $i = 1, 2$, where $\mathscr{C} > (2b+u)R$. The remainder of this section establishes these upper bounds.

4.7.1 The first auxiliary counting function

This is the easier case and the problem of finding a suitable upper bound for $N_1^{\mathrm{aux}}(\boldsymbol{\beta}; B)$ is essentially handled in [95].

Lemma 4.7.2 (Corollary 5.2 of [95])**.** *Let $H_{\boldsymbol{\beta}}(\boldsymbol{y})$ and $N_1^{\mathrm{aux}}(\boldsymbol{\beta}; B)$ be as above. Let $B, C \geq 1$, let $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ and let $\sigma \in \{0, \ldots, n-1\}$. Then we either obtain the bound*

$$N_1^{\mathrm{aux}}(\boldsymbol{\beta}; B) \ll_{C,n} B^{n+\sigma}(\log B)^n$$

or there exist non-trivial linear subspaces $U, V \subseteq \mathbb{R}^n$ with $\dim U + \dim V = n + \sigma + 1$ such that for all $\boldsymbol{v} \in V$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$ we have

$$\frac{\left|\boldsymbol{u}_1^T H_{\boldsymbol{\beta}}(\boldsymbol{v})\boldsymbol{u}_2\right|}{\|\boldsymbol{\beta}\cdot\boldsymbol{F}\|_\infty} \ll_n C^{-1}\|\boldsymbol{u}_1\|_\infty \|\boldsymbol{v}\|_\infty \|\boldsymbol{u}_2\|_\infty.$$

Recall the quantity

$$s_{\mathbb{R}}^{(1)} := 1 + \max_{\boldsymbol{\beta}\in\mathbb{R}^R\setminus\{0\}} \dim \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2},$$

where we regard $\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \subset \mathbb{P}_{\mathbb{C}}^{n_1-1}$ as a projective variety. Note that for this definition we do not necessarily require $n_1 = n_2$.

Proposition 4.7.3. *Let $\varepsilon > 0$. For all $B \geq 1$, $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ we have*

$$N_1^{\mathrm{aux}}(\boldsymbol{\beta}; B) \ll_\varepsilon B^{n+s_{\mathbb{R}}^{(1)}+\varepsilon}. \tag{4.7.2}$$

Proof. Assume for a contradiction that the estimate in (4.7.2) does not hold. In this case Lemma 4.7.2 gives that for each $N \in \mathbb{N}$ there exist $\boldsymbol{\beta}_N \in \mathbb{R}^R$ and there are non-trivial linear subspaces $U_N, V_N \subseteq \mathbb{R}^n$ with $\dim U_N + \dim V_N = n + s_{\mathbb{R}}^{(1)} + 1$ such that for all $\boldsymbol{v} \in V_N$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U_N$ we have

$$\frac{\left|\boldsymbol{u}_1^T H_{\boldsymbol{\beta}_N}(\boldsymbol{v})\boldsymbol{u}_2\right|}{\|\boldsymbol{\beta}_N \cdot \boldsymbol{F}\|_\infty} \ll_n N^{-1} \|\boldsymbol{u}_1\|_\infty \|\boldsymbol{v}\|_\infty \|\boldsymbol{u}_2\|_\infty .$$

If we change $\boldsymbol{\beta}_N$ by a scalar then $2\frac{\left|H_{\boldsymbol{\beta}_N}(\boldsymbol{y})\right|}{\|\boldsymbol{\beta}_N\cdot\boldsymbol{F}\|_\infty}$ remains unchanged for any $\boldsymbol{y} \in \mathbb{R}^n$. Therefore we may without loss of generality assume $\|\boldsymbol{\beta}_N\|_\infty = 1$. Thus there exists a convergent subsequence of $(\boldsymbol{\beta}_N)$ whose limit we will denote by $\boldsymbol{\beta}$. Hence we find subspaces $U, V \subseteq \mathbb{R}^n$ with $\dim U + \dim V = n + s_{\mathbb{R}}^{(1)} + 1$ such that for all $\boldsymbol{v} \in V$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U$ we have

$$\boldsymbol{u}_1^T H_{\boldsymbol{\beta}}(\boldsymbol{v})\boldsymbol{u}_2 = 0.$$

Let k denote the nonnegative integer such that

$$\dim V = n - k, \quad \text{and} \quad \dim U = s_{\mathbb{R}}^{(1)} + k + 1$$

holds. Consider now a basis $\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n$ of V that we extend to a basis $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ of $\mathbb{R}^n$. Write also $[U] \subseteq \mathbb{P}_{\mathbb{C}}^{n-1}$ for the projectivisation of U. Define $W \subseteq [U]$ to be the projective variety defined by the equations

$$\boldsymbol{u}^T H_{\boldsymbol{\beta}}(\boldsymbol{v}_i)\boldsymbol{u} = 0, \quad \text{for } i = 1, \dots, k$$

We find $\dim W \geq \dim[U] - k = s_{\mathbb{R}}^{(1)}$. Since $W \subseteq [U]$ and by the definition of W, noting that the entries of $H_{\boldsymbol{\beta}}(\boldsymbol{y})$ are linear in $\boldsymbol{y}$ we get that if $\boldsymbol{u} \in W$ then

$$\boldsymbol{u}^T H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{u} = 0 \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^n.$$

In particular it follows that $W \subseteq \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n} \subset \mathbb{P}_{\mathbb{C}}^{n-1}$ and thus

$$s_{\mathbb{R}}^{(1)} - 1 \geq \dim W \geq s_{\mathbb{R}}^{(1)},$$

which is clearly a contradiction. □

Now that we found an upper bound in terms of the geometry of $\mathbb{V}(\boldsymbol{F})$ the next Lemma shows that if $\boldsymbol{F}$ defines a non-singular variety then $s_{\mathbb{R}}^{(1)}$ is not too large. For the next Lemma we will not assume $n_1 = n_2$ as we will require it later in the slightly more general context when this assumption is not necessarily satisfied.

Lemma 4.7.4. *Let $s_{\mathbb{R}}^{(1)}$ be defined as above and assume that $\boldsymbol{F}$ is a system of bihomogenous equations of bidegree $(2,1)$ that defines a smooth complete intersection $\mathbb{V}(\boldsymbol{F}) \subset \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$. Then*

$$s_{\mathbb{R}}^{(1)} \leq \max\{0, R + n_1 - n_2\}.$$

Proof. Let $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ be such that $\dim \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} = s_{\mathbb{R}}^{(1)} - 1$. If $\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} = \emptyset$ then the statement in the lemma is trivially true. Hence we may assume that this is not the case. The singular locus of $\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) \subseteq \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ is given by

$$\mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) = \left(\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}\right) \cap \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}).$$

From Lemma 4.3.3 we obtain

$$\dim \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) \leq R - 2.$$

Further, since $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})$ is a system of n_1 bilinear equations, Lemma 4.3.1 gives

$$\dim \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta}\cdot\boldsymbol{F}) \geq s_{\mathbb{R}}^{(1)} - 1 + n_2 - 1 - n_1.$$

Combining the previous two inequalities yields

$$s_{\mathbb{R}}^{(1)} \leq R + n_1 - n_2,$$

as desired. □

We remark here that the proof of Lemma 4.7.4 shows that if $\mathbb{V}(\boldsymbol{F})$ defines a smooth complete intersection and if $s_{\mathbb{R}}^{(1)} > 0$ then $n_2 < n_1 + R$.

4.7.2 The second auxiliary counting function

Define $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(1)})$ to be the $n\times n$ matrix with the rows given by $(\boldsymbol{x}^{(1)})^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)/\left\|\boldsymbol{\beta}\cdot\boldsymbol{F}\right\|_\infty$ for $\ell = 1,\dots,n$. Using this notation $N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)$ counts the number of integer tuples $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$ such that $\left\|\boldsymbol{x}^{(1)}\right\|_\infty, \left\|\boldsymbol{x}^{(2)}\right\|_\infty \leq B$ and

$$\left\|\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(1)})\boldsymbol{x}^{(2)}\right\|_\infty \leq B,$$

is satisfied. The entries of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(1)})$ are homogeneous linear polynomials in $\boldsymbol{x}^{(1)}$ whose coefficients do not exceed absolute value 1.

Let A be a real $m \times n$ matrix. Then $A^T A$ is a symmetric and positive definite $n \times n$ matrix, with eigenvalues $\lambda_1^2, \ldots, \lambda_n^2$. The nonnegative real numbers $\{\lambda_i\}$ are the *singular values* of A.

Notation. Given a matrix $M = (m_{ij})$ we define $\|M\|_\infty := \max_{i,j} |m_{ij}|$. For simplicity we will from now on write $\boldsymbol{x}$ instead of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{y}$ instead of $\boldsymbol{x}^{(2)}$. For $\boldsymbol{x} \in \mathbb{R}^n$ let $\lambda_{\boldsymbol{\beta},1}(\boldsymbol{x}), \ldots, \lambda_{\boldsymbol{\beta},n}(\boldsymbol{x})$ denote the singular values of the real $n \times n$ matrix $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$ in descending order, counted with multiplicity. Note that $\lambda_{\boldsymbol{\beta},i}(\boldsymbol{x})$ are real and nonnegative. Also note

$$\lambda_{\boldsymbol{\beta},1}^2(\boldsymbol{x}) \leq n \left\| \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) \right\|_\infty \leq n^2 \left\| \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) \right\|_\infty^2 \leq n^4 \|\boldsymbol{x}\|_\infty^2 .$$

Taking square roots we find the following useful estimates

$$\lambda_{\boldsymbol{\beta},1}(\boldsymbol{x}) \leq n \left\| \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) \right\|_\infty \leq n^2 \|\boldsymbol{x}\|_\infty \tag{4.7.3}$$

Let $i \in \{1, \ldots, n\}$ and write $\boldsymbol{D}^{(\boldsymbol{\beta},i)}(\boldsymbol{x})$ for the vector with $\binom{n}{i}^2$ entries being the $i \times i$ minors of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$. Note that the entries are homogeneous polynomials in $\boldsymbol{x}$ of degree i.

Finally write $J_{\boldsymbol{D}^{(\boldsymbol{\beta},i)}}(\boldsymbol{x})$ for the Jacobian matrix of $\boldsymbol{D}^{(\boldsymbol{\beta},i)}(\boldsymbol{x})$. That is, $J_{\boldsymbol{D}^{(\boldsymbol{\beta},i)}}(\boldsymbol{x})$ is the $\binom{n}{i}^2 \times n$ matrix given by

$$(J_{\boldsymbol{D}^{(\boldsymbol{\beta},i)}}(\boldsymbol{x}))_{jk} = \frac{\partial D_j^{(\boldsymbol{\beta},i)}}{\partial x_k}.$$

Definition 4.7.5. Let $k \in \{0, \ldots, n\}$ and let $E_1, \ldots, E_{k+1} \in \mathbb{R}$ be such that $E_1 \geq \ldots \geq E_{k+1} \geq 1$ holds. We define $K_k(E_1, \ldots, E_{k+1}) \subseteq \mathbb{R}^n$ to be the set containing $\boldsymbol{x} \in \mathbb{R}^n$ such that the following three conditions are satisfied:

(i) $\|\boldsymbol{x}\|_\infty \leq B$,

(ii) $\frac{1}{2} E_i < \lambda_{\boldsymbol{\beta},i}(\boldsymbol{x}) \leq E_i$ if $1 \leq i \leq k$, and

(iii) $\lambda_{\boldsymbol{\beta},i}(\boldsymbol{x}) \leq E_{k+1}$ if $k+1 \leq i \leq n$.

Lemma 4.7.6. *Let $\widetilde{H}$ be an $n \times n$ matrix with real entries, and denote its singular values in descending order by $\lambda_1, \ldots, \lambda_n$. Let $C, B \geq 1$ and assume $\lambda_1 \leq CB$. Write $N_{\widetilde{H}}(B)$ for the number of integral vectors $\boldsymbol{y} \in \mathbb{Z}^n$ such that*

$$\|\boldsymbol{y}\|_\infty \leq B, \quad \textit{and} \quad \left\| \widetilde{H} \boldsymbol{y} \right\|_\infty \leq B$$

holds. Then

$$N_{\widetilde{H}}(B) \ll_{C,n} \min_{1 \leq i \leq n} \frac{B^n}{1 + \lambda_1 \cdots \lambda_i}.$$

Proof. Consider the ellipsoid

$$\mathcal{E} := \{\boldsymbol{t} \in \mathbb{R}^n : \boldsymbol{t}^T \widetilde{H}^T \widetilde{H} \boldsymbol{t} \leq nB^2\}.$$

Note that any $\boldsymbol{y} \in \mathbb{Z}^n$ counted by $N_{\widetilde{H}}(B)$ is contained in $\mathcal{E} \cap [-B, B]^n$. Now recall that $\widetilde{H}^T \widetilde{H}$ is a symmetric matrix with eigenvalues $\lambda_1^2, \dots, \lambda_n^2$. Therefore the principal radii of the ellipsoid $\mathcal{E}$ are given by $\lambda_i^{-1}\sqrt{n}B$. Hence we find

$$N_{\widetilde{H}}(B) \ll_n \prod_{i=1}^{n} \min\{1 + \lambda_i^{-1}\sqrt{n}B, B\} \tag{4.7.4}$$

By assumption we have $\lambda_i \leq CB$ and so the quantity on the right hand side of (4.7.4) is bounded above by

$$\prod_{i=1}^{n} \min\{2C\lambda_i^{-1}\sqrt{n}B, B\},$$

and thus

$$N_{\widetilde{H}}(B) \ll_{C,n} B^n \prod_{i=1}^{n} \min\{\lambda_i^{-1}, 1\}.$$

Since $\lambda_1 \geq \cdots \geq \lambda_n$ the result now follows. □

Lemma 4.7.7. *Given $B \geq 1$ one of the following three possibilities must be true. Either we have*

$$\frac{N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_0(1)), \tag{4.7.5}$$

or there exist nonnegative integers $e_1, \dots, e_k$ for some $k \in \{1, \dots, n-1\}$ such that $\log B \gg_n e_1 \geq \ldots \geq e_k$ *and*

$$\frac{2^{e_1 + \cdots + e_k} N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_k(2^{e_1}, \dots, 2^{e_k}, 1)), \tag{4.7.6}$$

or there exist nonnegative integers $e_1, \dots, e_n$ such that $\log B \gg_n e_1 \geq \ldots \geq e_n$ *and*

$$\frac{2^{e_1 + \cdots + e_n} N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_{n-1}(2^{e_1}, \dots, 2^{e_n})). \tag{4.7.7}$$

Proof. If $k = n$ then condition (iii) in Definition 4.7.5 is always trivially satisfied and thus

$$K_n(2^{e_1}, \dots, 2^{e_n}, 1) \subseteq K_{n-1}(2^{e_1}, \dots, 2^{e_n}).$$

In particular, (4.7.7) follows from (4.7.6) with $k = n$. We are left showing that either (4.7.5) holds or there exist nonnegative integers $e_1, \dots, e_k$ for some $k \in \{1, \dots, n\}$ such that $\log B \gg_n e_1 \geq \ldots \geq e_k$ and (4.7.6) holds.

Note that the box $[-B,B]^n$ is the disjoint union of $K_0(1)$ and $K_k(2^{e_1},\dots,2^{e_k},1)$ where k runs over $1,\dots,n$ and e_i run over integers $\log B \gg_n e_1 \geq \ldots \geq e_k$. Given $\boldsymbol{x}\in\mathbb{Z}^n$ write

$$N_{\boldsymbol{x}}(B) = \#\left\{\boldsymbol{y}\in\mathbb{Z}^n\colon\ \|\boldsymbol{y}\|_\infty \leq B,\ \left\|\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})\boldsymbol{y}\right\|_\infty \leq B\right\}.$$

We thus obtain

$$N_2^{\mathrm{aux}}(\boldsymbol{\beta},B) = \sum_{\substack{\boldsymbol{x}\in\mathbb{Z}^n\\ \boldsymbol{x}\in K_0(1)}} N_{\boldsymbol{x}}(B) + \sum_{\substack{1\leq k\leq n\\ 1\leq e_k\leq\ldots\leq e_1\\ e_1\ll_n \log B}} \sum_{\substack{\boldsymbol{x}\in\mathbb{Z}^n\\ \boldsymbol{x}\in K_k(2^{e_1},\ldots,2^{e_k},1)}} N_{\boldsymbol{x}}(B). \tag{4.7.8}$$

Note that the number of terms of the outer sum of the second term of the right hand side of (4.7.8) is bounded by $\ll_n (\log B)^n$. From this it follows that we either have

$$\sum_{\substack{\boldsymbol{x}\in\mathbb{Z}^n\\ \boldsymbol{x}\in K_0(1)}} N_{\boldsymbol{x}}(B) \gg_n \frac{N_2^{\mathrm{aux}}(\boldsymbol{\beta},B)}{(\log B)^n} \tag{4.7.9}$$

or there exists an integer $k\in\{1,\ldots,n\}$ and integers $e_1\geq\ldots\geq e_k\geq 1$ such that

$$\sum_{\substack{\boldsymbol{x}\in\mathbb{Z}^n\\ \boldsymbol{x}\in K_k(2^{e_1},\ldots,2^{e_k},1)}} N_{\boldsymbol{x}}(B) \gg_n \frac{N_2^{\mathrm{aux}}(\boldsymbol{\beta},B)}{(\log B)^n}. \tag{4.7.10}$$

If (4.7.9) holds then (4.7.5) follows from the trivial bound $N_{\boldsymbol{x}}(B)\ll_n B^n$. Assume now (4.7.10) holds. From (4.7.3), for each $\boldsymbol{x}$ appearing in the sum of (4.7.10) we have the bound

$$\lambda_{\boldsymbol{\beta},1}(\boldsymbol{x}) \leq n^2 B.$$

Applying Lemma 4.7.6 with $C=n^2$ and $\widetilde{H}=\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$ we find

$$N_{\boldsymbol{x}}(B) \ll_n \frac{B^n}{2^{e_1+\ldots+e_k}}. \tag{4.7.11}$$

Substituting (4.7.11) into (4.7.10) delivers (4.7.6). □

We now recall two Lemmas from [95] that are conveniently stated in a form so that they apply to our setting.

Lemma 4.7.8 (Lemma 3.2 in [95])**.** *Let M be a real $m\times n$ matrix with singular values $\lambda_1,\ldots,\lambda_n$ listed with multiplicity in descending order. For $k\leq\min\{m,n\}$ denote by $\boldsymbol{D}^{(k)}$ the vector of $k\times k$ minors of M. Given such k, the following statements are true:*

(i) *We have*

$$\left\| \boldsymbol{D}^{(k)} \right\|_\infty \asymp \lambda_1 \cdots \lambda_k$$

(ii) *There is a k-dimensional subspace $V \subset \mathbb{R}^n$, which can be taken to be a span of standard basis vectors $\boldsymbol{e}_i$, such that for all $\boldsymbol{v} \in V$ the following holds*

$$\|M\boldsymbol{v}\|_\infty \gg_{m,n} \|\boldsymbol{v}\|_\infty \lambda_k$$

(iii) *Given $C \geq 1$ one of the following alternatives holds. Either there exists a $(n-k+1)$-dimensional subspace $X \subset \mathbb{R}^n$ such that*

$$\|M\boldsymbol{X}\|_\infty \leq C^{-1} \|\boldsymbol{X}\|_\infty \quad \textit{for all } \boldsymbol{X} \in X,$$

or there is a k-dimensional subspace $V \subset \mathbb{R}^n$ spanned by standard basis vectors such that

$$\|M\boldsymbol{v}\|_\infty \gg_{m,n} C^{-1} \|\boldsymbol{v}\|_\infty \quad \textit{for all } \boldsymbol{v} \in V.$$

Next, we are interested in counting the number of integer tuples contained in the sets $K_k(E_1, \ldots, E_{k+1})$. The next Lemma is taken from [95].

Lemma 4.7.9 (Lemma 4.1 in [95])**.** *Let $B, C \geq 1$, $\sigma \in \{0, \ldots, n-1\}$ and $k \in \{0, \ldots, n-\sigma-1\}$. Assume further $CB \geq E_1 \geq \ldots \geq E_{k+1} \geq 1$. Then one of the following alternatives must hold.*

(I)$_k$ *We have the estimate*

$$\#(\mathbb{Z}^n \cap K_k(E_1, \ldots, E_{k+1})) \ll_{C,n} B^\sigma (E_1 \cdots E_{k+1}) E_{k+1}^{n-\sigma-k-1}.$$

(II)$_k$ *For some integer $b \in \{1, \ldots, k\}$ there exists a $(\sigma+b+1)$-dimensional subspace $X \subset \mathbb{R}^n$ and there exists $\boldsymbol{x}^{(0)} \in K_b(E_1, \ldots, E_{b+1})$ such that $E_{b+1} < C^{-1} E_b$ and*

$$\left\| J_{\boldsymbol{D}^{(\boldsymbol{\beta}, b+1)}}(\boldsymbol{x}^{(0)}) \boldsymbol{X} \right\|_\infty \leq C^{-1} \left\| \boldsymbol{D}^{(\boldsymbol{\beta}, b)}(\boldsymbol{x}^{(0)}) \right\|_\infty \|\boldsymbol{X}\|_\infty \quad \textit{for all } \boldsymbol{X} \in X.$$

(III) *There exists a $(\sigma+1)$-dimensional subspace $X \subset \mathbb{R}^n$ such that*

$$\left\| \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{X}) \right\|_\infty \leq C^{-1} \|\boldsymbol{X}\|_\infty \quad \textit{for all } \boldsymbol{X} \in X. \tag{4.7.12}$$

Remark 4.7.10. In [95], Lemma 4.7.9 was stated for $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$ being a symmetric matrix, and $\lambda_{\boldsymbol{\beta},i}(\boldsymbol{x})$ were taken to be the eigenvalues of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$ whose absolute values coincide with its singular values. However, an inspection of the proof shows that only the estimates in Lemma 4.7.8 as well as (4.7.3) were used, which are valid for singular values as well as the (absolute values) of the eigenvalues. Therefore the proof remains valid in our setting.

The next Lemma is similar to Lemma 5.1 in [95], however we need to account for the fact that $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})$ is not necessarily a symmetric matrix.

Lemma 4.7.11. *Let $b \in \{1, \ldots, n-1\}$ and $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ be such that $\boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \neq 0$. Then there exist subspaces $Y_1, Y_2 \subseteq \mathbb{R}^n$ with $\dim Y_1 = \dim Y_2 = n-b$ such that for all $\boldsymbol{Y}_1 \in Y_1$, $\boldsymbol{Y}_2 \in Y_2$ and $\boldsymbol{t} \in \mathbb{R}^n$ we have*

$$\boldsymbol{Y}_1^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t}) \boldsymbol{Y}_2 \ll_n \left(\frac{\left\| J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}^{(0)}) \boldsymbol{t} \right\|_\infty}{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty} + \frac{\lambda_{\boldsymbol{\beta},b+1}(\boldsymbol{x}^{(0)}) \cdot \|\boldsymbol{t}\|_\infty}{\lambda_{\boldsymbol{\beta},b}(\boldsymbol{x}^{(0)})} \right) \|\boldsymbol{Y}_1\|_\infty \|\boldsymbol{Y}_2\|_\infty \tag{4.7.13}$$

where the implied constant only depends on n but is otherwise independent from $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t})$

Proof. Given $\boldsymbol{x} \in \mathbb{R}^n$ define $\boldsymbol{y}_1^{(1)}(\boldsymbol{x}), \ldots, \boldsymbol{y}_1^{(n-b)}(\boldsymbol{x})$ in the following way. The j-th entries are given by

$$(y_1^{(i)}(\boldsymbol{x}))_j = \begin{cases} (-1)^{n-b} \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=n-b+1,\ldots,n \\ \ell=n-b+1,\ldots,n}} \right) & \text{if } j = i, \\ (-1)^j \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=i,n-b+1,\ldots,n;\ k\neq j \\ \ell=n-b+1,\ldots,n}} \right) & \text{if } j > n-b, \\ 0 & \text{otherwise,} \end{cases} \tag{4.7.14}$$

where $k = i, n-b+1, \ldots, n;\ k \neq j$ denotes that we let the index k run over the values $i, n-b+1, \ldots, n$ with $k = j$ omitted. Similarly we define $\boldsymbol{y}_2^{(1)}(\boldsymbol{x}), \ldots, \boldsymbol{y}_2^{(n-b)}(\boldsymbol{x})$ by

$$(y_2^{(i)}(\boldsymbol{x}))_j = \begin{cases} (-1)^{n-b} \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=n-b+1,\ldots,n \\ \ell=n-b+1,\ldots,n}} \right) & \text{if } j = i, \\ (-1)^j \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=n-b+1,\ldots,n \\ \ell=i,n-b+1,\ldots,n;\ \ell\neq j}} \right) & \text{if } j > n-b, \\ 0 & \text{otherwise.} \end{cases}$$

Using the Laplace expansion of a determinant along columns and rows we thus obtain

$$(\boldsymbol{y}_1^{(i)}(\boldsymbol{x})^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}))_j = \begin{cases} (-1)^{n-b} \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=i,n-b+1,\ldots,n \\ \ell=j,n-b+1,\ldots,n}} \right) & \text{if } j \leq n-b, \\ 0 & \text{otherwise,} \end{cases} \tag{4.7.15}$$

and

$$(\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) \boldsymbol{y}_2^{(i)}(\boldsymbol{x}))_j = \begin{cases} (-1)^{n-b} \det \left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})_{k\ell})_{\substack{k=j,n-b+1,\ldots,n \\ \ell=i,n-b+1,\ldots,n}} \right) & \text{if } j \leq n-b, \\ 0 & \text{otherwise,} \end{cases} \tag{4.7.16}$$

respectively. It follows from (4.7.14) — (4.7.16) that there exist matrices $L_1^{(i)}$, $L_2^{(i)}$, $M_1^{(i)}$ and $M_2^{(i)}$ for $i = 1, \ldots, n-b$ with entries only in $\{0, \pm 1\}$ such that we obtain

$$\boldsymbol{y}_1^{(i)}(\boldsymbol{x}) = L_1^{(i)} \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}), \tag{4.7.17}$$

$$\boldsymbol{y}_2^{(i)}(\boldsymbol{x}) = L_2^{(i)} \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}), \tag{4.7.18}$$

$$(\boldsymbol{y}_1^{(i)}(\boldsymbol{x}))^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) = [M_1^{(i)} \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x})]^T, \quad \text{and} \tag{4.7.19}$$

$$\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) \boldsymbol{y}_2^{(i)}(\boldsymbol{x}) = M_2^{(i)} \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x}). \tag{4.7.20}$$

Given $\boldsymbol{t} \in \mathbb{R}^n$ we write $\partial_{\boldsymbol{t}}$ for the directional derivative given by $\sum t_i \frac{\partial}{\partial x_i}$. Applying $\partial_{\boldsymbol{t}}$ to both sides of (4.7.20) we obtain

$$[\partial_{\boldsymbol{t}} \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x})] \boldsymbol{y}_2^{(i)}(\boldsymbol{x}) + \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) [\partial_{\boldsymbol{t}} \boldsymbol{y}_2^{(i)}(\boldsymbol{x})] = M_2^{(i)} [\partial_{\boldsymbol{t}} \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x})]. \tag{4.7.21}$$

Now note

$$\partial_{\boldsymbol{t}} \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x}) = J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}) \boldsymbol{t}, \quad \text{and} \quad \partial_{\boldsymbol{t}} \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) = \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t}). \tag{4.7.22}$$

Substituting (4.7.22) and (4.7.18) into (4.7.21) yields

$$\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t}) \boldsymbol{y}_2^{(i)}(\boldsymbol{x}) = M_2^{(i)} J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}) \boldsymbol{t} - \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}) L_2^{(i)} \partial_{\boldsymbol{t}} \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}).$$

If we premultiply this by $\boldsymbol{y}_1^{(j)}(\boldsymbol{x})^T$ and use (4.7.19) then we obtain

$$\begin{aligned} \boldsymbol{y}_1^{(j)}(\boldsymbol{x})^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t}) \boldsymbol{y}_2^{(i)}(\boldsymbol{x}) = \boldsymbol{y}_1^{(j)}(\boldsymbol{x})^T M_2^{(i)} J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}) \boldsymbol{t} \\ - [M_1^{(j)} \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x})]^T [L_2^{(i)} \partial_{\boldsymbol{t}} \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x})]. \end{aligned} \tag{4.7.23}$$

Lemma 4.7.8 (i) yields the bounds

$$\frac{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b+1)}(\boldsymbol{x}) \right\|_\infty}{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}) \right\|_\infty} \ll_n \lambda_{\boldsymbol{\beta},b+1}(\boldsymbol{x}), \tag{4.7.24}$$

and

$$\frac{\left\| \partial_{\boldsymbol{t}} \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}) \right\|_\infty}{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}) \right\|_\infty} \ll_n \frac{\|\boldsymbol{t}\|_\infty}{\lambda_{\boldsymbol{\beta},b}(\boldsymbol{x})}. \tag{4.7.25}$$

Now we specify $\boldsymbol{x} = \boldsymbol{x}^{(0)}$ so by assumption we have $\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty > 0$. Thus define

$$\boldsymbol{Y}_k^{(i)} = \frac{\boldsymbol{y}_k^{(i)}(\boldsymbol{x}^{(0)})}{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty}, \quad \text{for } i = 1, \ldots, n-b \text{ and } k = 1, 2. \tag{4.7.26}$$

Dividing (4.7.23) by $1/ \left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty^2$ and using (4.7.26) as well as the bounds (4.7.24) and (4.7.25) gives

$$\left| \boldsymbol{Y}_1^{(j)} \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{t}) \boldsymbol{Y}_2^{(i)} \right| \ll_n \frac{\left\| J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}^{(0)}) \boldsymbol{t} \right\|_\infty}{\left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty} + \frac{\lambda_{\boldsymbol{\beta},b+1}(\boldsymbol{x}^{(0)}) \left\| \boldsymbol{t} \right\|_\infty}{\lambda_{\boldsymbol{\beta},b}(\boldsymbol{x}^{(0)})}.$$

We claim now that we can take the subspaces $Y_k \subseteq \mathbb{R}^n$ to be defined as the span of $\boldsymbol{Y}_k^{(1)}, \dots, \boldsymbol{Y}_k^{(n-b)}$ for $k = 1, 2$ respectively, so that the Lemma holds. For this we need to show that (4.7.13) holds, and also that $\dim Y_1 = \dim Y_2 = n - b$. Therefore it suffices to show the following claim: Given $\boldsymbol{\gamma} \in \mathbb{R}^{n-b}$ if we take $\boldsymbol{Y}_k = \sum \gamma_i Y_k^{(i)}$ then $\|\boldsymbol{\gamma}\|_\infty \ll_n \|\boldsymbol{Y}_k\|_\infty$, for $k = 1, 2$ respectively.

Assume that the $b \times b$ minor of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})$ of largest absolute value lies in the bottom right corner of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})$. In other words, we assume

$$\left\|\boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)})\right\|_\infty = \left|\det\left((\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})_{k\ell})_{\substack{k=n-b+1,\dots,n \\ \ell=n-b+1,\dots,n}} \right)\right|. \tag{4.7.27}$$

After permuting the rows and columns of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})$ the identity (4.7.27) will always be true. The vectors $\boldsymbol{Y}_k^{(i)}$ depend on minors of $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})$. Thus we can apply the same permutations to $\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(0)})$ that ensure that (4.7.27) holds to the definition of these vectors. From this we see that we can always reduce the general case to the case where (4.7.27) holds.

Now for $k = 1, 2$ we define matrices

$$Q_k = \left(\boldsymbol{Y}_k^{(1)} \middle| \cdots \middle| \boldsymbol{Y}_k^{(n-b)} \middle| \boldsymbol{e}_{n-b+1} \middle| \cdots \middle| \boldsymbol{e}_n \right).$$

By the definition of $\boldsymbol{Y}_k^{(i)}$ we see that Q_k must be of the following form

$$Q_k = \begin{pmatrix} I_{n-b} & 0 \\ \widetilde{Q}_k & I_b \end{pmatrix},$$

for some matrix $\widetilde{Q}_k$. In particular we find $\det Q_k = 1$ and so $\left\|Q_k^{-1}\right\|_\infty \ll_n 1$. Given $\boldsymbol{Y}_k = \sum \gamma_i Y_k^{(i)}$ we thus find

$$\|\boldsymbol{\gamma}\|_\infty = \left\|Q_k^{-1}\boldsymbol{Y}_k\right\|_\infty \ll_n \|\boldsymbol{Y}_k\|_\infty,$$

and so the Lemma follows. □

The next Corollary is the main technical result from this section, which will allow us to deduce that either $N_2^{\text{aux}}(\boldsymbol{\beta}, B)$ is small or a suitable singular locus is large.

Corollary 4.7.12. *Let $B, C \geq 1$ and let $\sigma \in \{0, \dots, n-1\}$. Then one of the following alternatives is true. Either we have the bound*

$$N_2^{\text{aux}}(\boldsymbol{\beta}, B) \ll_{C,n} B^{n+\sigma}(\log B)^n, \tag{4.7.28}$$

or there exist subspaces $X, Y_1, Y_2 \subseteq \mathbb{R}^n$ with $\dim X + \dim Y_1 = \dim X + \dim Y_2 = n + \sigma + 1$, such that

$$\left|\boldsymbol{Y}_1^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{X})\boldsymbol{Y}_2\right| \ll_n C^{-1} \|\boldsymbol{Y}_1\|_\infty \|\boldsymbol{X}\|_\infty \|\boldsymbol{Y}_2\|_\infty \tag{4.7.29}$$

holds for all $\boldsymbol{X} \in X, \boldsymbol{Y}_1 \in Y_1, \boldsymbol{Y}_2 \in Y_2$.

Proof. Let $k \in \{0, \ldots, n-\sigma-1\}$ and $E_1, \ldots, E_{k+1} \in \mathbb{R}$ be such that

$$CB \geq E_1 \geq \ldots \geq E_{k+1} \geq 1.$$

We know that one of the alternatives $(\mathrm{I})_k$, $(\mathrm{II})_k$ or (III) in Lemma 4.7.9 holds. Assume first that $(\mathrm{I})_k$ always holds so that the estimate

$$\#(\mathbb{Z}^n \cap K_k(E_1, \ldots, E_{k+1})) \ll_{C,n} B^\sigma (E_1 \cdots E_{k+1}) E_{k+1}^{n-\sigma-k-1}. \tag{4.7.30}$$

holds for every $k \in \{0, \ldots, n-\sigma-1\}$ and $E_1, \ldots, E_{k+1} \in \mathbb{R}$ such that $CB \geq E_1 \geq \ldots \geq E_{k+1} \geq 1$. From Lemma 4.7.7 we find that either we have

$$\frac{N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_0(1)), \tag{4.7.31}$$

or there exist nonnegative integers $e_1, \ldots, e_k$ for some $k \in \{1, \ldots, n-1\}$ such that $\log B \gg_n e_1 \geq \ldots \geq e_k$ and

$$\frac{2^{e_1+\cdots+e_k} N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_k(2^{e_1}, \ldots, 2^{e_k}, 1)), \tag{4.7.32}$$

or there exist nonnegative integers $e_1, \ldots, e_n$ such that $\log B \gg_n e_1 \geq \ldots \geq e_n$ and

$$\frac{2^{e_1+\cdots+e_n} N_2^{\mathrm{aux}}(\boldsymbol{\beta}, B)}{B^n (\log B)^n} \ll_n \#(\mathbb{Z}^n \cap K_{n-1}(2^{e_1}, \ldots, 2^{e_n})). \tag{4.7.33}$$

We may take C to be large enough depending on n such that $CB \geq 2^{e_1}$ is satisfied. Then upon sbustituting the bound (4.7.30) into either of (4.7.31), (4.7.32) or (4.7.33) gives (4.7.28).

If (III) holds in Lemma 4.7.9 we can take $Y_1 = Y_2 = \mathbb{R}^n$ so that (4.7.29) follows from (4.7.12).

Finally, assume there exist $k \in \{0, \ldots, n-\sigma-1\}$ and $E_1, \ldots, E_{k+1} \in \mathbb{R}$ with $CB \geq E_1 \geq \ldots \geq E_{k+1} \geq 1$ such that $(\mathrm{II})_k$ in Lemma 4.7.9 holds. Recall this means there exists some integer $b \in \{1, \ldots, k\}$, a $(\sigma+b+1)$-dimensional subspace $X \subset \mathbb{R}^n$ and $\boldsymbol{x}^{(0)} \in K_b(E_1, \ldots, E_{b+1})$ such that $E_{b+1} < C^{-1} E_b$ and

$$\left\| J_{\boldsymbol{D}^{(\boldsymbol{\beta},b+1)}}(\boldsymbol{x}^{(0)}) \boldsymbol{X} \right\|_\infty \leq C^{-1} \left\| \boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \right\|_\infty \|\boldsymbol{X}\|_\infty \quad \text{for all } \boldsymbol{X} \in X. \tag{4.7.34}$$

As $\boldsymbol{x}^{(0)} \in K_b(E_1, \ldots, E_{b+1})$ we have $E_i/2 < \lambda_{\boldsymbol{\beta},i}(\boldsymbol{x}^{(0)}) \leq E_i$ for $i = 1, \ldots, k$ and $\lambda_{\boldsymbol{\beta},b+1}(\boldsymbol{x}^{(0)}) \leq E_{b+1}$. This, together with the fact that $E_{b+1} < C^{-1} E_b$ implies

$$\lambda_{\boldsymbol{\beta},b+1}(\boldsymbol{x}^{(0)}) < 2C^{-1} \lambda_{\boldsymbol{\beta},b}(\boldsymbol{x}^{(0)}). \tag{4.7.35}$$

Also we find $\lambda_{\boldsymbol{\beta},b}(\boldsymbol{x}^{(0)}) \neq 0$, from which in turn it follows from Lemma 4.7.8 (i) that $\boldsymbol{D}^{(\boldsymbol{\beta},b)}(\boldsymbol{x}^{(0)}) \neq 0$. Thus we may apply Lemma 4.7.11 to obtain spaces $Y_1, Y_2 \subseteq \mathbb{R}^n$ with $\dim Y_1 = \dim Y_2 = n - b$ such that the estimate (4.7.13) holds. Now taking $\boldsymbol{t} = \boldsymbol{X}$ in (4.7.13) and using (4.7.34) and (4.7.35) then (4.7.29) follows. Since $\dim X = \sigma + b + 1$ we also have $\dim X + \dim Y_1 = \dim X + \dim Y_2 = n + \sigma + 1$ as desired. $\square$

Recall the definition of the quantity

$$s_{\mathbb{R}}^{(2)} := \left\lfloor \frac{\max_{\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}} \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})}{2} \right\rfloor + 1,$$

where $\lfloor x \rfloor$ denotes the largest integer m such that $m \leq x$. Although we have been assuming $n_1 = n_2$ throughout the definition of this quantity remains valid if $n_1 \neq n_2$. Note that we have $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \subsetneq \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ for all $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$. For if not, then the matrix $H_{\boldsymbol{\beta}}(\boldsymbol{y})$ is identically zero for some $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ contradicting the fact that $\mathbb{V}(\boldsymbol{F})$ is a complete intersection. In particular this yields $s_{\mathbb{R}}^{(2)} \leq \frac{n_1+n_2}{2} - 1$.

Before we prove the main result of this section we require another small Lemma.

Lemma 4.7.13. *Let $\boldsymbol{\beta} \in \mathbb{R} \setminus \{0\}$. The system of equations*

$$\boldsymbol{y}^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x} = 0, \text{ for } \ell = 1, \dots, n \quad \text{and} \quad H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x} = \boldsymbol{0}$$

define the same variety in $\mathbb{P}_{\mathbb{C}}^{n-1} \times \mathbb{P}_{\mathbb{C}}^{n-1}$.

Proof. Recall that by definition we have

$$\widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{z}) = \begin{pmatrix} \boldsymbol{z}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_1) \\ \vdots \\ \boldsymbol{z}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_n). \end{pmatrix}$$

For $\ell \in \{1, \dots, n\}$ we get

$$\boldsymbol{y}^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x} = \boldsymbol{y}^T \begin{pmatrix} \boldsymbol{e}_\ell^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_1)\boldsymbol{x} \\ \vdots \\ \boldsymbol{e}_\ell^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_n)\boldsymbol{x} \end{pmatrix} = \sum_{i=1}^n y_i \boldsymbol{e}_\ell^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_i)\boldsymbol{x} = \boldsymbol{e}_\ell^T H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x},$$

where the last line follows since the entries of $H_{\boldsymbol{\beta}}(\boldsymbol{y})$ are linear homogeneous in $\boldsymbol{y}$. The result is now immediate. $\square$

Proposition 4.7.14. *Let $s_{\mathbb{R}}^{(2)}$ be defined as above and let $B \geq 1$. Then for all $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ the following holds*

$$N_2^{\text{aux}}(\boldsymbol{\beta}, B) \ll_n B^{n+s_{\mathbb{R}}^{(2)}} (\log B)^n.$$

Proof. Suppose for a contradiction the result were false. Then for each positive integer N there exists some $\boldsymbol{\beta}_N$ such that

$$N_2^{\mathrm{aux}}(\boldsymbol{\beta}_N, B) \geq N B^{n+s_{\mathbb{R}}^{(2)}} (\log B)^n.$$

From Corollary 4.7.12 it follows that there are linear subspaces $X^{(N)}, Y_1^{(N)}, Y_2^{(N)} \subset \mathbb{R}^n$ with

$$\dim X^{(N)} + \dim Y_i^{(N)} = n + s_{\mathbb{R}}^{(2)} + 1, \quad i = 1, 2,$$

such that for all $\boldsymbol{X} \in X^{(N)}$, $\boldsymbol{Y}_i \in Y_i^{(N)}$ we get

$$\left| \boldsymbol{Y}_1^T \widetilde{H}_{\boldsymbol{\beta}_N}(\boldsymbol{X}) \boldsymbol{Y}_2 \right| \leq N^{-1} \left\| \boldsymbol{Y}_1 \right\|_\infty \left\| \boldsymbol{X} \right\|_\infty \left\| \boldsymbol{Y}_2 \right\|_\infty.$$

Note that $\widetilde{H}_{\boldsymbol{\beta}_N}(\boldsymbol{\beta})$ is unchanged when $\boldsymbol{\beta}_N$ is multiplied by a constant. Thus we may assume $\|\boldsymbol{\beta}_N\|_\infty = 1$ and consider a converging subsequence of $\boldsymbol{\beta}_{N_r}$ converging to $\boldsymbol{\beta}$, say, as $N \to \infty$. This delivers subspaces $X, Y_1, Y_2 \subset \mathbb{R}^n$ with $\dim X + \dim Y_i = n + s_{\mathbb{R}}^{(2)} + 1$ for $i = 1, 2$ such that

$$\boldsymbol{Y}_1^T \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{X}) \boldsymbol{Y}_2 = 0 \quad \text{for all } \boldsymbol{X} \in X, \boldsymbol{Y}_1 \in Y_1, \boldsymbol{Y}_2 \in Y_2.$$

There exists some $b \in \{0, \dots, n - s_{\mathbb{R}}^{(2)} - 1\}$ such that $\dim X = n - b$ and $\dim Y_i = s_{\mathbb{R}}^{(2)} + b + 1$. Now let $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)}$ be a basis for $\mathbb{R}^n$ such that $\boldsymbol{x}^{(b+1)}, \dots, \boldsymbol{x}^{(n)}$ is a basis for X. Write $[Y_i] \subset \mathbb{P}_{\mathbb{C}}^{n-1}$ for the linear subspace of $\mathbb{P}_{\mathbb{C}}^{n-1}$ associated to Y_i for $i = 1, 2$.

Define the biprojective variety $W \subset [Y_1] \times [Y_2]$ in the variables $(\boldsymbol{y}_1, \boldsymbol{y}_2)$ by

$$W = \mathbb{V}(\boldsymbol{y}_1 \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(i)}) \boldsymbol{y}_2)_{i=1,\dots,b}.$$

Since the non-trivial equations defining W have bidegree $(1, 1)$ we can apply Corollary 4.3.2 to find

$$\dim W \geq \dim [Y_1] \times [Y_2] - b = 2 s_{\mathbb{R}}^{(2)} + b. \tag{4.7.36}$$

Given $(\boldsymbol{y}_1, \boldsymbol{y}_2) \in W$ we have in particular $(\boldsymbol{y}_1, \boldsymbol{y}_2) \in [Y_1] \times [Y_2]$ and so

$$\boldsymbol{y}_1 \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{x}^{(i)}) \boldsymbol{y}_2 = 0, \quad \text{for } i = b+1, \dots, n,$$

and hence $\boldsymbol{y}_1 \widetilde{H}_{\boldsymbol{\beta}}(\boldsymbol{z}) \boldsymbol{y}_2 = 0$ for all $\boldsymbol{z} \in \mathbb{R}^n$. From Lemma 4.7.13 we thus see $H_{\boldsymbol{\beta}}(\boldsymbol{y}_1) \boldsymbol{y}_2 = 0$ for all $(\boldsymbol{y}_1, \boldsymbol{y}_2) \in W$. Hence in particular

$$\dim W \leq \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y}) \boldsymbol{x}) \leq 2 s_{\mathbb{R}}^{(2)} - 1,$$

where we regard $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y}) \boldsymbol{x})$ as a variety in $\mathbb{P}_{\mathbb{C}}^{n-1} \times \mathbb{P}_{\mathbb{C}}^{n-1}$ in the variables $(\boldsymbol{x}, \boldsymbol{y})$. This together with (4.7.36) implies $b \leq -1$, which is clearly a contradiction. $\square$

In the next Lemma we show that $s_{\mathbb{R}}^{(2)}$ is small if $\mathbb{V}(\boldsymbol{F})$ defines a smooth complete intersection. For this we no longer assume $n_1 = n_2$.

Lemma 4.7.15. *Let $s_{\mathbb{R}}^{(2)}$ be defined as above. If $\mathbb{V}(\boldsymbol{F})$ is a smooth complete intersection in $\mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ then we have the bound*

$$\frac{n_2-1}{2} \leq s_{\mathbb{R}}^{(2)} \leq \frac{n_2+R}{2}. \tag{4.7.37}$$

Proof. Let $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{\mathbf{0}\}$ be such that

$$s_{\mathbb{R}}^{(2)} = \left\lfloor \frac{\dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})}{2} \right\rfloor + 1.$$

Note that then

$$2s_{\mathbb{R}}^{(2)} - 2 \leq \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \leq 2s_{\mathbb{R}}^{(2)} - 1. \tag{4.7.38}$$

The variety $\mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \subset \mathbb{P}_{\mathbb{C}}^{n_1-1} \times \mathbb{P}_{\mathbb{C}}^{n_2-1}$ is defined by n_1 bilinear polynomials. Using Corollary 4.3.2 we thus find

$$\dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \geq n_2 - 2$$

so the lower bound in (4.7.37) follows. We proceed by considering two cases.

Case 1: $\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} = \emptyset$**.** Note that this can only happen if $n_2 \geq n_1$. We can therefore apply Lemma 4.3.4 with $V_1 = \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2}$, $V_2 = \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})$ and $A_i = H_{\boldsymbol{\beta}}(\boldsymbol{e}_i)$ to find

$$\dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \leq n_2 - 1 + \dim \mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} = n_2 - 2.$$

From this and (4.7.38) the upper bound in (4.7.37) follows for this case.

Case 2: $\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \neq \emptyset$**:** By assumption there exists $\boldsymbol{x} \in \mathbb{C}^{n_1} \setminus \{\mathbf{0}\}$ such that

$$\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x} = 0, \quad \text{for all } \ell = 1, \dots, n_2.$$

We claim that there exists $\boldsymbol{y} \in \mathbb{C}^{n_2} \setminus \{\mathbf{0}\}$ such that $H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x} = \mathbf{0}$. For this define the vectors

$$\boldsymbol{u}_\ell = H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x}, \quad \ell = 1, \dots, n_2.$$

Note that $\boldsymbol{x} \in \langle \boldsymbol{u}_1, \dots, \boldsymbol{u}_{n_2} \rangle^\perp$ so these vectors must be linearly dependent. Thus there exist $y_1, \dots, y_{n_2} \in \mathbb{C}$ not all zero, such that

$$H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x} = \sum_{\ell=1}^{n_2} y_\ell H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x} = \mathbf{0},$$

where the first equality followed since the entries of $H_{\boldsymbol{\beta}}(\boldsymbol{y})$ are linear homogeneous in $\boldsymbol{y}$. The claim follows. In particular it follows from this that

$$\left(\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \times \mathbb{P}^{n_2-1}\right) \cap \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) \neq \emptyset.$$

Using Lemma 4.3.1 and (4.7.38) we therefore find

$$\begin{aligned}\dim\left[\left(\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \times \mathbb{P}^{n_2-1}\right) \cap \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x})\right] \geq \\ \dim \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}) - n_2 \geq 2s_{\mathbb{R}}^{(2)} - n_2 - 2. \quad (4.7.39)\end{aligned}$$

Recall $\boldsymbol{\beta} \cdot \boldsymbol{F} = \boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}$ so that

$$\mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) = \left(\mathbb{V}(\boldsymbol{x}^T H_{\boldsymbol{\beta}}(\boldsymbol{e}_\ell)\boldsymbol{x})_{\ell=1,\dots,n_2} \times \mathbb{P}^{n_2-1}\right) \cap \mathbb{V}(H_{\boldsymbol{\beta}}(\boldsymbol{y})\boldsymbol{x}).$$

Under our assumptions we can apply Lemma 4.3.3 to find $\dim \mathrm{Sing}\mathbb{V}(\boldsymbol{\beta} \cdot \boldsymbol{F}) \leq R - 2$. The result follows from this and (4.7.39). □

Proof of Theorem 4.7.1. Applying Theorem 4.2.1 it suffices to show

$$N_i^{\mathrm{aux}}(\boldsymbol{\beta}; B) \leq C_0 B^{2n-4\mathscr{C}}, \quad (4.7.40)$$

holds for all $\boldsymbol{\beta} \in \mathbb{R}^R \setminus \{0\}$ and $i = 1, 2$, where $\mathscr{C} > (2b + u)R$. Let

$$s = \max\{s_{\mathbb{R}}^{(1)}, s_{\mathbb{R}}^{(2)}\},$$

where $s_{\mathbb{R}}^{(1)}$ and $s_{\mathbb{R}}^{(2)}$ are as defined in (4.1.7) and (4.1.8), respectively. From Proposition 4.7.3 and Proposition 4.7.14 for any $\varepsilon > 0$ we get

$$N_i^{\mathrm{aux}}(\boldsymbol{\beta}; B) \ll_\varepsilon B^{n+s+\varepsilon},$$

with the implied constant not depending on $\boldsymbol{\beta}$. Choose $\varepsilon = \frac{n-s-(8b+4u)R}{2}$, which is a positive real number by our assumption (4.7.1). Taking

$$\mathscr{C} = \frac{n - s - \varepsilon}{4},$$

we see that from the assumption $n - s_{\mathbb{R}}^{(i)} > (8b + 4u)R$ for $i = 1, 2$ we must have $\mathscr{C} > (2b + u)R$ for this choice. Therefore (4.7.40) holds and the first part of the theorem follows upon applying Theorem 4.2.1.

For the second part recall we assume $n > (16b + 8u + 1)R$ and that the forms $F_i(\boldsymbol{x}, \boldsymbol{y})$ define a smooth complete intersection in $\mathbb{P}_{\mathbb{C}}^{n-1} \times \mathbb{P}_{\mathbb{C}}^{n-1}$. By Lemma 4.7.4 in this case we obtain

$$s_{\mathbb{R}}^{(1)} \leq R,$$

and from Lemma 4.7.15 we find

$$s_{\mathbb{R}}^{(2)} \leq \frac{n+R}{2}.$$

Therefore it is easily seen that assuming $n > (16b + 8u + 1)R$ implies that

$$n - s_{\mathbb{R}}^{(i)} > (8b + 4u)R$$

holds for $i = 1, 2$, which is what we wanted to show. □

4.7.3 Proof of Theorem 4.1.2

Proof of Theorem 4.1.2. If $n_1 = n_2$ then the result follows immediately from Proposition 4.7.1. We have two cases to consider and although their strategies are very similar they are not entirely symmetric. Therefore it is necessary to consider them individually.

Case 1: $n_1 > n_2$. We consider a new system of equations $\widetilde{F}_i(\boldsymbol{x}, \tilde{\boldsymbol{y}})$ in the variables $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\tilde{\boldsymbol{y}} = (y_1, \ldots, y_{n_2}, y_{n_2+1}, \ldots, y_{n_1})$ where the forms $\widetilde{F}_i(\boldsymbol{x}, \tilde{\boldsymbol{y}})$ satisfy

$$\widetilde{F}_i(\boldsymbol{x}, \tilde{\boldsymbol{y}}) = F(\boldsymbol{x}, \boldsymbol{y}),$$

where $\boldsymbol{y} = (y_1, \ldots, y_{n_2})$. Write $\widetilde{N}(P_1, P_2)$ for the counting function associated to the system $\widetilde{\boldsymbol{F}} = \boldsymbol{0}$ and the boxes $\mathcal{B}_1 \times (\mathcal{B}_2 \times [0, 1]^{n_1 - n_2})$. Note in particular, that if we replace F by $\widetilde{F}$ in (4.5.30) and (4.5.29) then the expressions for the singular series and the singular integral remain unchanged. Further denote by $\tilde{s}_{\mathbb{R}}^{(i)}$ the quantities defined in (4.1.7) and (4.1.8) but with F replaced by $\widetilde{F}$. Note that we have $\tilde{s}_{\mathbb{R}}^{(1)} = s_{\mathbb{R}}^{(1)}$ and $\tilde{s}_{\mathbb{R}}^{(2)} \leq s_{\mathbb{R}}^{(2)} + \frac{n_1 - n_2}{2}$. Therefore the assumptions (4.1.9) imply

$$n_1 - \tilde{s}_{\mathbb{R}}^{(i)} > (8b + 4u)R$$

for $i = 1, 2$. Hence we may apply Proposition 4.7.1 in order to obtain

$$\widetilde{N}(P_1, P_2) = \mathfrak{I}\mathfrak{S} P_1^{n_1 - 2R} P_2^{n_1 - R} + O(P_1^{n_1 - 2R} P_2^{n_1 - R} \min\{P_1, P_2\}^{-\delta}),$$

for some $\delta > 0$. Finally it is easy to see that

$$\begin{aligned}\widetilde{N}(P_1, P_2) &= N(P_1, P_2) \# \left\{ \boldsymbol{t} \in \mathbb{Z}^{n_1 - n_2} \cap [0, P_2]^{n_1 - n_2} \right\} \\ &= N(P_1, P_2)(P_2^{n_1 - n_2} + O(P_2^{n_1 - n_2 - 1})),\end{aligned}$$

and so (4.1.10) follows.

Case 2: $n_2 > n_1$ We deal with this very similarly as in the first case; we define a new system of forms $\widetilde{F}_i(\tilde{\boldsymbol{x}}, \boldsymbol{y})$ in the variables $\tilde{\boldsymbol{x}} = (x_1, \ldots, x_{n_2})$ and $\boldsymbol{y} = (y_1, \ldots, y_{n_2})$ such that

$$\widetilde{F}_i(\boldsymbol{x}, \tilde{\boldsymbol{y}}) = F_i(\boldsymbol{x}, \boldsymbol{y})$$

holds. As before we define a new counting function $\widetilde{N}(P_1, P_2)$ with respect to the new product of boxes $(\mathcal{B}_1 \times [0,1]^{n_2-n_1}) \times \mathcal{B}_2$, and we define $\tilde{s}_{\mathbb{R}}^{(i)}$ similarly to the previous case. Note that $\tilde{s}_{\mathbb{R}}^{(1)} = s_{\mathbb{R}}^{(1)} + n_2 - n_1$ and $\tilde{s}_{\mathbb{R}}^{(2)} \leq s_{\mathbb{R}}^{(2)} + \frac{n_2-n_1}{2}$ so that (4.1.9) gives

$$n_2 - \tilde{s}_{\mathbb{R}}^{(i)} > (8b + 4u)R,$$

for $i = 1, 2$. Therefore Proposition 4.7.1 applies and we deduce again that (4.1.10) holds as desired.

Finally we turn to the case when $\mathbb{V}(\boldsymbol{F})$ defines a smooth complete intersection. Note first that by Lemma 4.7.15 we have

$$s_{\mathbb{R}}^{(2)} \leq \frac{n_2 + R}{2},$$

and therefore the condition

$$\frac{n_1 + n_2}{2} - s_{\mathbb{R}}^{(2)} > (8b + 4u)R$$

is satisfied if we assume $n_1 > (16b + 8u + 1)R$. Further, by Lemma 4.7.4 we have

$$s_{\mathbb{R}}^{(1)} \leq \max\{0, n_1 + R - n_2\},$$

and so we may replace the condition $n_1 - s_{\mathbb{R}}^{(1)} > (8b + 4u)R$ by

$$n_1 - \max\{0, n_1 + R - n_2\} > (8b + 4u)R.$$

If $n_2 \geq n_1 + R$ then this reduces to assuming $n_1 > (8b + 4u + 1)R$, which follows immediately since we assumed $n_1 > (16b + 8u + 1)R$. If $n_2 \leq n_1 + R$ on the other hand, then this is equivalent to assuming

$$n_2 > (8b + 4u + 1)R.$$

In any case, the assumptions (4.1.11) imply the assumptions (4.1.9) as desired. $\square$

Chapter 5

Artin's primitive root conjecture over function fields

5.1 Introduction

5.1.1 Primitive roots over $\mathbb{Z}$

For an odd prime $p \in \mathbb{N}$ recall that the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ is a finite cyclic group of order $p - 1$. We say that $g \in \mathbb{Z}$ is a *primitive root* mod p (denoted $ord_p(g) = p - 1$) if $p \nmid g$ and if the reduction g mod p generates the group $(\mathbb{Z}/p\mathbb{Z})^\times$. We call a prime number p an *Artin prime* for g, if g is a primitive root mod p. Note that if g is ± 1 or a perfect square, then it is easy to see that there are at most finitely many Artin primes for g. *Artin's primitive root conjecture* states that if g is neither a perfect square nor ± 1, then there are infinitely many Artin primes for g. This conjecture was proven by Hooley [50] conditionally on the generalised Riemann Hypothesis for Dedekind ζ-functions.

5.1.2 Primitive roots over $\mathbb{F}_q[t]$

Artin's primitive root conjecture may analogousely be formulated in the function field setting. This problem was first proposed by Hasse to his PhD student, Herbert Bilharz [7]. The simplest instance of the problem in this setting is as follows. Let $\mathbb{F}_q$ denote a finite field of q elements and $\mathbb{F}_q[t]$ the ring of polynomials with coefficients in $\mathbb{F}_q$. We moreover let $\mathcal{P}_n \subset \mathbb{F}_q[t]$ denote the subset of prime monic polynomials of degree $n \in \mathbb{N}$. For a polynomial $g(t) \in \mathbb{F}_q[t]$, we let $\mathrm{ord}_P(g(t))$ denote the order of $g(t)$ in the multiplicative group $(\mathbb{F}_q[t]/(P))^\times$, where $(P) \subseteq \mathbb{F}_q[t]$ denotes the prime ideal generated by some $P \in \mathcal{P}_n$. In particular, $g(t)$ generates $(\mathbb{F}_q[t]/(P))^\times$ if and only if $\mathrm{ord}_P(g) = q^n - 1$, in which case we say that g is a *primitive root* mod P.

Two immediate obstructions prevent $g(t)$ from being a primitive root modulo infinitely many P. First, note that if $g(t) \in \mathbb{F}_q^\times$ is a unit in $\mathbb{F}_q[t]$, then $\mathrm{ord}_P(g(t)) \leq q-1$, and therefore $g(t)$ cannot be a primitive root modulo $P \in \mathcal{P}_n$, whenever $n > 1$. Second, suppose $g(t) = h(t)^\ell$ where $\ell > 1$, and $\ell \mid q-1$. Since

$$q^n - 1 = (q-1)(q^{n-1} + \cdots + q + 1),$$

we then find that $\ell \mid q^n - 1$ for any $n \in \mathbb{N}$. Thus $\mathrm{ord}_P(g) \leq \frac{q^n-1}{\ell} < q^n - 1$, from which it follows that $g(t)$ cannot be a primitive root modulo any prime $P \in \mathbb{F}_q[t]$.

We therefore assume that $g(t) \notin \mathbb{F}_q^\times$, and moreover that $g(t) \in \mathbb{F}_q[t]$ is not an ℓ^{th} power, for any $\ell > 1$ such that $\ell \mid q-1$. In this setting, *Artin's primitive root conjecture* is the claim that there exist infinitely many prime polynomials $P \in \mathbb{F}_q[t]$ such that $g(t)$ is a primitive root mod P.

5.1.3 Primitive roots over function fields

To formulate Artin's primitive root conjecture over more general function fields, it seems appropriate to take a geometric viewpoint. (A rather self-contained overview of this geometric set-up is provided in Section 5.2.) Let X be a geometrically irreducible projective variety over $\mathbb{F}_q$ of dimension $r > 0$, and write $K = \mathbb{F}_q(X)$ for its function field. Given a closed point $\mathfrak{p}$ of X, denote by $\mathcal{O}_\mathfrak{p} = \mathcal{O}_{X,\mathfrak{p}}$ the stalk of the structure sheaf $\mathcal{O}_X$ at $\mathfrak{p}$. Abusing notation, write $\mathfrak{p} \subset \mathcal{O}_\mathfrak{p}$ for the unique maximal ideal and let $\kappa_\mathfrak{p} = \mathcal{O}_\mathfrak{p}/\mathfrak{p}$ denote the corresponding residue field. It is an elementary result that in this situation $\kappa_\mathfrak{p}$ is a finite field extension of the base field of X, i.e. of $\mathbb{F}_q$. We write $\deg \mathfrak{p} = [\kappa_\mathfrak{p} \colon \mathbb{F}_q]$ for the degree of $\mathfrak{p}$, which is a finite number.

Let $g \in K$ and let $\mathfrak{p}$ be a closed point of X. We say g is *regular* at $\mathfrak{p}$ if g lies in the image of the embedding $\mathcal{O}_\mathfrak{p} \hookrightarrow K$. By pulling g back under this embedding we may then consider $g \in \kappa_\mathfrak{p}$. We say that $g \in K$ is a *primitive root modulo* $\mathfrak{p}$ if g is regular at $\mathfrak{p}$ and if g generates the multiplicative group $\kappa_\mathfrak{p}^\times$. In such a case, we moreover refer to $\mathfrak{p}$ as an *Artin prime* for g.

Suppose $g \in K \setminus \mathbb{F}_q$ is not an ℓ^{th} power for any rational prime $\ell \mid q-1$. *Artin's primitive root conjecture* over K then states that there exist infinitely many Artin primes for g.

Note that every function field over $\mathbb{F}_q$, i.e. every field extension $K/\mathbb{F}_q$ of positive finite transcendence degree, may be recovered as $K = \mathbb{F}_q(X)$, where X is a geometrically integral projective variety over $\mathbb{F}_q$. Our main result is then the following theorem:

Theorem 5.1.1 (Artin's primitive root conjecture over function fields). *Artin's primitive root conjecture holds for any function field K over $\mathbb{F}_q$.*

As an example, consider the case $X = \mathbb{P}^1_{\mathbb{F}_q}$. Then $K = \mathbb{F}_q(t)$ and the closed points correspond to irreducible monic polynomials in $\mathbb{F}_q[t]$ in addition to the point at infinity. Theorem 5.1.1 then reduces to the setting described in Section 5.1.2.

Bilharz [7] addressed the particular case of Theorem 5.1.1 in which X is a geometrically irreducible projective curve over $\mathbb{F}_q$ (i.e. the case in which K is a global function field). In particular, he provided a proof conditional on the Riemann hypothesis for finite fields − a result which was later established by André Weil [116]. Bilharz's proof fails, however, in particular instances; namely cases in which $g \in K$ is not a *geometric element* (see Definition 5.3.2). Though this mistake has previously been noted, no corrected proof of Conjecture 5.1.1 for the case in which $g \in K$ has thus far appeared anywhere in the literature (see [91, pp. 155] for a more detailed discussion). In this work, we therefore remove the assumption that $g \in K$ be a geometric element. We moreover generalize to projective varieties of arbitrary dimension; thereby completing a proof of Conjecture 5.1.1 (see Theorem 5.4.1).

For the special case $g(t) = t^m + c$, an elementary proof of Artin's primitive root conjecture over $\mathbb{F}_q[t]$, which uses only the theory of Gauss sums is given in [63]. For irreducible $g(t)$, a proof which relies only on establishing a zero-free region of relevant L-functions, instead of the results of Weil, is given in [65, 64]. Several further variations of Artin's primitive root conjecture over function fields have also been studied, for example, over Carlitz modules [36, 59], rank one Drinfeld modules [60, 120, 67], and one dimensional tori over function fields [22].

5.1.4 Main results

Bilharz demonstrated that the Dirichlet density of Artin primes for geometric g is positive, from which he then deduced the infinitude of Artin primes for g. For a more quantitative description, let $N_X(g, n)$, denote the number of Artin primes for g of fixed degree n. In the particular case that $X = \mathscr{C}$ is a non-singular algebraic curve, and $g \in K$ is geometric, Pappalardi and Shparlinski demonstrated that for any $\varepsilon > 0$,

$$N_{\mathscr{C}}(g, n) = \frac{\varphi(q^n - 1)}{n} + O_{\varepsilon, g, \mathscr{C}}\left(q^{n/2(1+\varepsilon)}\right). \tag{5.1.1}$$

In this work, we generalize the above result by providing an asymptotic count for $N_X(g, n)$ where X is *any* geometrically irreducible projective variety of dimension $r \geq 1$. As a further highlight the assumption that $g \in \mathbb{F}_q(X)$ is geometric has been removed. Recall that if $g \in \mathbb{F}_q^\times$, or if g is an ℓ^{th} power for some rational prime $\ell \mid q-1$, then $N_X(g,n) = 0$ for all $n > 1$. Otherwise, we find that

$$N_X(g,n) = \rho_g(n)\left(\frac{\varphi(q^n-1)q^{n(r-1)}}{n} + O_{\varepsilon,g,X}\left(q^{n(r-1/2+\varepsilon)}\right)\right), \tag{5.1.2}$$

where $\rho_g(n)$ is as in (5.4.3). In particular, when $g \in K$ is geometric, we find that $\rho_g(n) = 1$, thereby recovering (5.1.1) in the case that $X = \mathscr{C}$ is a curve (i.e. $r = 1$). For non-geometric g, it is possible that $\rho_g(n) = 0$ for certain values of n. Nonetheless, in an argument provided in the proof of Corollary 5.4.3, we show that $\rho_g(n) \geq 1$ for infinitely many $n \in \mathbb{N}$, thereby confirming Conjecture 5.1.1.

5.1.5 Comparison to classical setting

When $g \in \mathbb{N}$ is not an exact power, Artin conjectured that the natural density of Artin primes for g, denoted $\mathcal{P}_g \subseteq \mathcal{P}$, is equal to

$$A := \prod_{p \text{ prime}}\left(1 - \frac{1}{p(p-1)}\right) \approx .3739558,$$

known as *Artin's constant.* Due to careful numerical observations pioneered by Derrick and Emma Lehmer, it later emerged that, for certain g, an additional correction factor is needed. Slightly more generally, the natural density of $\mathcal{P}_g \subseteq \mathcal{P}$ is conjectured to equal $c_g A_h$, where A_h is an explicit Euler product, and $c_g \in \mathbb{Q}$. More specifically, A_h is a *linear* factor, which depends on the value of the largest integer h such that g is an h^{th} power in $\mathbb{Z}$, while c_g is an additional quadratic correction factor, which takes into account *entanglements* between number fields of the form $\mathbb{Q}(\zeta_\ell, g^{1/\ell})$. This modified conjecture was eventually proven correct by Hooley [50] under the assumption of the generalised Riemann Hypothesis.

Going back to the function field setting let P_n denote the closed points of X of fixed degree n, and let $P_n(g) \subseteq P_n$ denote the subset of Artin primes for g, so that $\#P_n(g) = N_X(g,n)$. When X is a non-singular curve and g is geometric, it follows from (5.1.1) that the density of $P_n(g) \subseteq P_n$ is asymptotic to $A(n) := \varphi(q^n-1)q^{-n}$, in the limit as $q^n \to \infty$. Note that $A(n)$ does not converge. In fact, even the natural density of Artin primes, namely the limit

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} N_X(g,n)}{\sum_{n=1}^{N} \#P_n},$$

does not, in general, exist. This was demonstrated by Bilharz [7] and expanded upon by Perng [83].

More generally, from (5.1.2) we find that the density of $P_n(g) \subseteq P_n$ is asymptotic to $A_g(n) := \rho_g(n)\varphi(q^n - 1)q^{-n}$, where $\rho_g(n)$ depends on the factorization behaviour of g in $K \otimes_{\mathbb{F}_q} \overline{\mathbb{F}}_q$.

Outline

The remainder of this chapter is structured as follows. In Section 5.2 we provide an overview of the relevant geometric set-up for Artin's primitive root conjecture for varieties of arbitrary dimension over $\mathbb{F}_q$. In Section 5.3 we then discuss geometric extensions and geometric elements, and in Section 5.4 we state our quantitative results (Theorem 5.4.1), from which a proof of Theorem 5.1.1 follows (Corollary 5.4.3). Section 5.5 uses Weil's theorem to establish a very general estimate for exponential sums over a variety. This step is crucial for extending our results from curves to varieties. Section 5.6 then establishes a proof of Theorem 5.4.1, and finally Section 5.7 provides a heuristic interpretation of the counting function, $N_X(g, n)$, in order to draw a conceptual comparison between our correction factor, $\rho_g(n)$, and the classical correction factor, c_g.

5.2 Background on projective schemes

Projective Schemes

A *graded ring* is a ring S endowed with a direct sum decomposition $S = \oplus_{d\geq 0} S_d$ of the underlying additive group, such that $S_d S_e \subset S_{d+e}$. We say that a non-zero element $a \in S$ is *homogeneous* of degree d, denoted $\deg a = d$, if $a \in S_d$. A *homogenous ideal* is an ideal $I \subset S$ that is generated by a set of homogenous elements. The ideal consisting of elements of positive degree, namely $S_+ := \oplus_{d>0} S_d$, is referred to as the *irrelevant ideal.* If $S = \oplus_{d\geq 0} S_d$ is a graded ring, and $I \triangleleft S$ a homogenous ideal, then the quotient ring $R = S/I$ is itself a graded ring, with $R_d = S_d/(I \cap S_d)$.

Consider the set

$$\mathrm{Proj}(S) := \{\mathfrak{p} \subseteq S : \mathfrak{p} \text{ homogenous prime ideal, } S_+ \not\subseteq \mathfrak{p}\}.$$

We define a topology on $X = \mathrm{Proj}(S)$ (called the *Zariski topology*) by designating the closed sets of $\mathrm{Proj}(S)$ to be of the form

$$Z(I) := \{\mathfrak{p} \in \mathrm{Proj}(S) : I \subseteq \mathfrak{p}\},$$

where $I \subset S$ denotes a homogenous ideal. A point $\mathfrak{p} \in X$ is said to be a *closed point* if $\overline{\{\mathfrak{p}\}} = \{\mathfrak{p}\}$, equivalently, if there is no $\mathfrak{q} \in X$ such that $\mathfrak{p} \subsetneq \mathfrak{q}$.

The *distinguished open set* associated to any homogenous element $f \in S_+$ is then given by

$$X_f := \mathrm{Proj}(S) \setminus Z(\langle f \rangle) = \{\mathfrak{p} \in \mathrm{Proj}(S) : f \notin \mathfrak{p}\},$$

and the collection of such sets, namely $\{X_f : f \in S_+\}$, forms a basis for the topology on $\mathrm{Proj}(S)$. The space $\mathrm{Proj}(S)$, together with its Zariski topology, is referred to as a *projective scheme.*

The *structure sheaf*, denoted $\mathcal{O}_X$, is a *sheaf* on $\mathrm{Proj}(S)$, defined on the distinguished open sets X_f, $f \in S_+$ homogeneous, as

$$\mathcal{O}_X(X_f) := S_{(f)} = \left\{ \frac{a}{f^n} : a \in S \text{ is homogenous}, n \in \mathbb{Z}_{\geq 0},\ \deg\, a = n \cdot \deg\, f \right\},$$

i.e. as the zero-degree component of the localization $\{1, f, f^2, \dots\}^{-1}S$. The projective scheme $X = \mathrm{Proj}(S)$ is called *integral* if $S_{(f)}$ is an integral domain for any homogeneous $f \in S_+$. An integral projective scheme is, in particular, irreducible as a topological space. It is an elementary fact that any integral scheme X has a *generic point* η. That is, an element $\eta \in X$ such that $\overline{\{\eta\}} = X$.

Function Fields, Stalks, and Residue Fields

Let X denote an integral projective scheme. We then find that for any homogeneous $f, g \in S_+$, $\mathrm{Frac}(S_{(f)}) \cong \mathrm{Frac}(S_{(g)})$, where $\mathrm{Frac}(R)$ denotes the fraction field of an integral domain R. We define $\mathcal{K}(X) := \mathrm{Frac}(S_{(f)})$ for any homogeneous $f \in S_+$ to be the *function field* of X, which can be expressed explicitly as

$$\mathcal{K}(X) := \left\{ \frac{a}{b} : a, b \in S_d \text{ for some } d \in \mathbb{Z}_{\geq 0}, b \neq 0 \right\}.$$

The *stalk* at a point $\mathfrak{p} \in X$ refers to the local ring

$$\mathcal{O}_{X,\mathfrak{p}} := \left\{ \frac{a}{b} \in \mathcal{K}(X) : a, b \in S_d \text{ for some } d \in \mathbb{Z}_{\geq 0}, b \notin \mathfrak{p} \right\},$$

whose unique maximal ideal is given explicitly by

$$\mathfrak{p}\mathcal{O}_{X,\mathfrak{p}} := \left\{ \frac{a}{b} \in \mathcal{K}(X) : a, b \in S_d \text{ for some } d \in \mathbb{Z}_{\geq 0}, a \in \mathfrak{p}, b \notin \mathfrak{p} \right\}.$$

The intersection of all such stalks, namely

$$\mathcal{O}_X(X) := \bigcap_{\mathfrak{p} \in X} \mathcal{O}_{X,\mathfrak{p}},$$

is referred to as the *global sections* of X. We say that X is *normal* if $\mathcal{O}_{X,\mathfrak{p}}$ is an integrally closed domain inside $\mathcal{K}(X)$ for every $\mathfrak{p} \in X$. Finally, we refer to $\kappa_{\mathfrak{p}} := \mathcal{O}_{X,\mathfrak{p}}/\mathfrak{p}\mathcal{O}_{X,\mathfrak{p}}$ as the *residue field* of $\mathfrak{p}$.

Two noteworthy properties of the function field $\mathcal{K}(X)$ are as follows: if $\eta \in X$ is the generic point of X, then $\mathcal{K}(X)$ is isomorphic to the stalk $\mathcal{O}_{X,\eta}$. Moreover, if $\mathrm{Spec}(R) \subset X$ is an open affine then R must be an integral domain and $\mathcal{K}(X)$ is again given by the fraction field of R.

Projective Varieties

A *projective variety* X over the field k is a projective integral scheme of the form $X = \mathrm{Proj}(S)$, where $S = k[x_0, \ldots, x_n]/I$ is a finitely generated k-algebra, and where $I \subseteq k[x_0, \ldots, x_n]$ is a homogenous ideal. Under these assumptions, we note that X is both *Noetherian* and *separated*. We denote its function field by $k(X)$ and we note that the *dimension* of X, denoted $\dim(X)$, is equal to the transcendence degree of $k(X)$ over k. A projective variety of dimension one is referred to as a *projective curve*.

Let $S = k[x_0, \ldots, x_n]/I$ be as above. $\mathrm{Proj}(S)$ is said to be *geometrically integral* if $\mathrm{Proj}(\overline{S})$ is integral, where $\overline{S} = (k[x_0, \ldots, x_n]/I) \otimes_k \overline{k}$. For example, if $f \in k[x_0, x_1, x_2]$ is homogeneous of positive degree and *absolutely* irreducible (i.e. irreducible over $\overline{k}$), then $\mathrm{Proj}(k[x_0, x_1, x_2]/\langle f \rangle)$ is a geometrically integral projective curve. We moreover let $\overline{k}(X)$ refer to the function field of $\mathrm{Proj}(\overline{S})$.

Let $X = \mathrm{Proj}(S)$ be a geometrically integral projective variety over $\mathbb{F}_q$, and let $K = \mathbb{F}_q(X)$ denote the function field of X. Note that when $\mathfrak{p} \in X$ is closed, $\kappa_{\mathfrak{p}}$ is then a finite algebraic extension of $\mathbb{F}_q$. We moreover define $\deg \mathfrak{p} := [\kappa_{\mathfrak{p}} \colon \mathbb{F}_q]$ to be the *degree* of $\mathfrak{p} \in X$.

Fix $g \in K$ and let $\mathfrak{p} \in X$ be closed. We say that g is *regular* at $\mathfrak{p}$ if $g \in \mathcal{O}_{X,\mathfrak{p}} \subset K$. We say $g \in K$ is a *primitive root modulo* $\mathfrak{p}$ if g is regular at $\mathfrak{p}$ and if $g \bmod \mathfrak{p}\mathcal{O}_{X,\mathfrak{p}}$ generates the multiplicative group $\kappa_{\mathfrak{p}}^{\times}$.

Divisors and Valuations

Let $X = \mathrm{Proj}(S)$ be a normal, geometrically integral, projective variety over $\mathbb{F}_q$. In particular, X is a normal Noetherian integral separated scheme. A *prime divisor* Y of X is a closed integral subscheme $Y \subset X$ of codimension one, i.e. such that $\dim(Y) = \dim(X) - 1$. It then follows that if $\eta_Y \in Y$ is the generic point of Y, the stalk $\mathcal{O}_{X,\eta_Y}$ is in fact a discrete valuation ring inducing a valuation $v_Y : \mathcal{O}_{X,\eta_Y} \to \mathbb{Z}$. Furthermore, since the fraction field of $\mathcal{O}_{X,\eta_Y}$ is the function field K, we find that for any prime divisor Y, the valuation v_Y may be extended to a function $v_Y : K \to \mathbb{Z}$.

Given $g \in K^\times$, it follows from [42, Lemma II.6.1] that $v_Y(g) = 0$ for all but finitely many prime divisors $Y \subset X$. We may thus define the *degree* of g to be

$$\deg(g) := \sum_{Y \subset X} |v_Y(g)|,$$

where the sum runs over all prime divisors Y of X. Note that $\deg(g)$ may be viewed as the number of poles and zeros on X, counted with multiplicity.

In the particular case in which X is a curve, we note that the set of prime divisors of X is precisely given by the set of closed points in X.

Rational Points

Let R be a finitely generated $\mathbb{F}_q$-algebra, and we denote by $\mathrm{Spec}(R)$ the *affine* $\mathbb{F}_q$-*scheme*, which is the affine scheme whose underlying set is the collection of prime ideals in R coming with a morphism $\mathrm{Spec}(R) \to \mathrm{Spec}(\mathbb{F}_q)$ induced by the $\mathbb{F}_q$-algebra structure. Note that the closed points of $\mathrm{Spec}(R)$ are given by the maximal ideals of R. For finitely generated $\mathbb{F}_q$-algebras R and S, we further recall that morphisms ρ: $\mathrm{Spec}(S) \to \mathrm{Spec}(R)$ between $\mathbb{F}_q$-schemes are in one-to-one correspondence with the $\mathbb{F}_q$-algebra homomorphisms $\rho^\# : R \to S$.

An $\mathbb{F}_{q^n}$-*rational point* of $\mathrm{Spec}(R)$ is an $\mathbb{F}_q$-scheme morphism

$$\rho : \mathrm{Spec}(\mathbb{F}_{q^n}) \to \mathrm{Spec}(R),$$

which then corresponds to a homomorphism of $\mathbb{F}_q$-algebras

$$\rho^\# : R = \mathbb{F}_q[x_1, \ldots, x_m]/I \to \mathbb{F}_{q^n}.$$

The image $\mathrm{Im}(\rho^\#)$ is a subring of $\mathbb{F}_{q^n}$ containing $\mathbb{F}_q$, and hence must be a field between $\mathbb{F}_q$ and $\mathbb{F}_{q^n}$. To any $\mathbb{F}_{q^n}$-rational point ρ, one may associate an ideal $\mathfrak{m} = \ker(\rho^\#) \subset R$, which, by the first isomorphism theorem, is maximal.

Conversely, suppose $\mathfrak{m} \subset R$ is a maximal ideal. Then $R/\mathfrak{m} \cong \mathbb{F}_{q^m}$ for some positive integer m, where m is the degree of $\mathfrak{m}$. If $m \leq n$ one may then associate precisely m different $\mathbb{F}_{q^n}$-rational points to the closed point $\mathfrak{m} \in \mathrm{Spec}(R)$ as follows. Note that there are precisely m different $\mathbb{F}_q$-invariant inclusions $\varphi \colon \mathbb{F}_{q^m} \hookrightarrow \mathbb{F}_{q^n}$ coming from the elements in $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q) \cong \mathbb{Z}/m\mathbb{Z}$.

Let $\pi : R \to R/\mathfrak{m}$ denote the projection map. Then for any φ as above the $\mathbb{F}_q$-algebra homomorphism $\rho^\#_\varphi : R \to \mathbb{F}_{q^m}$ given by $\rho^\#_\varphi = \varphi \circ \pi$ corresponds to a unique $\mathbb{F}_{q^n}$-rational point. Thus each closed point $\mathfrak{m} \in \mathrm{Spec}(R)$ of degree m gives rise to precisely m distinct $\mathbb{F}_{q^n}$-rational points, $\rho^\#_\varphi$.

Let X be a geometrically integral projective variety over $\mathbb{F}_q$ with function field $K = \mathbb{F}_q(X)$. We let $X(\mathbb{F}_{q^n})$ denote the set of $\mathbb{F}_{q^n}$-rational points of X, i.e. the set of $\mathbb{F}_q$-scheme morphisms

$$\rho\colon \operatorname{Spec}(\mathbb{F}_{q^n}) \to X.$$

We now describe how to evaluate an element $g \in \mathbb{F}_q(X)$ at a rational point $\rho \in X(\mathbb{F}_{q^n})$. Note that ρ must factor through some open affine subscheme $\operatorname{Spec}(R) \subset X$. Thus considering the restriction of this morphism whose image is contained in $\operatorname{Spec}(R)$ this induces an $\mathbb{F}_q$-algebra homomorphism

$$\rho^{\#}\colon R \to \mathbb{F}_{q^n}.$$

Since X is integral, R is an integral domain with field of fractions K. We therefore may write $g = a/b$ for some $a, b \in R$ with $b \neq 0$. If $\rho^{\#}(b) = 0$ we say that g has a *pole* at ρ. Otherwise we may evaluate g at ρ as follows

$$g(\rho) := \frac{\rho^{\#}(a)}{\rho^{\#}(b)} \in \mathbb{F}_{q^n}. \tag{5.2.1}$$

Recall that the closed point $\mathfrak{p}$ corresponding to ρ is given by $\mathfrak{p} = \operatorname{Ker}(\rho^{\#}) \subset R$. Clearly $\rho^{\#}(b) = 0$ precisely when $b \in \operatorname{Ker}(\rho^{\#}) = \mathfrak{p}$. Hence $g = a/b$ is regular at $\mathfrak{p}$ if and only if g does not have a pole at ρ.

Number field analogue

We conclude this section by noting that the classical version of Artin's primitive root conjecture (i.e. the case over number fields) may also be phrased in a geometric language. Let $K/\mathbb{Q}$ be a number field with ring of integers $\mathcal{O}_K$, and recall that the closed points of $\operatorname{Spec}(\mathcal{O}_K)$ is the set of non-zero prime ideals of $\mathcal{O}_K$. The residue field of a closed point $P \in \operatorname{Spec}(\mathcal{O}_K)$, i.e. of a non-zero prime ideal of $\mathcal{O}_K$, is given by $\kappa_P = \mathcal{O}_K/P$. Given $g \in K$ we say that g is a primitive root modulo a non-zero prime ideal $P \in \operatorname{Spec}(\mathcal{O}_K)$ if $v_P(g) \geq 0$ and if g generates the multiplicative group $\kappa_P^\times = (\mathcal{O}_K/P)^\times$. Artin's primitive root conjecture over K states that the number of prime ideals P for which g is a primitive root is infinite.

5.3 Geometric extensions

Let $X = \operatorname{Proj}(S)$ denote a geometrically integral projective variety over $\mathbb{F}_q$, where $\operatorname{char}(\mathbb{F}_q) = p$. Let $K = \mathbb{F}_q(X)$ denote its function field. Recall that we write $\overline{\mathbb{F}}_q(X)$

to refer to the function field of the base change of X to $\overline{\mathbb{F}}_q$, namely to the compositum of fields $K\overline{\mathbb{F}}_q$. We moreover note that this is isomorphic to $\mathbb{F}_q(X) \otimes \overline{\mathbb{F}}_q$.

Let $\overline{K}$ denote a fixed algebraic closure of K, and consider the algebraic field extensions L/K and $M/\mathbb{F}_q$. Note that L, M, as well as the compositum LM, may then all be embedded inside $\overline{K}$. Given an algebraic field extension L/K we thus write $\overline{\mathbb{F}}_q \cap L \subset \overline{K}$ to be the maximal algebraic subextension of $\mathbb{F}_q$ inside L. Using this notation we note by [88, Proposition 2.2.22] that $K \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$.

Definition 5.3.1. Let $L_2/L_1/K$ be a tower of algebraic field extensions. We say that L_2/L_1 is a **geometric field extension** if $L_2 \cap \overline{\mathbb{F}}_q = L_1 \cap \overline{\mathbb{F}}_q$. In particular, if L/K is an algebraic field extension, we say that L/K is geometric if $L \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$.

Definition 5.3.2. Let $a \in K$. We say that a is **geometric at a rational prime** $\ell \neq p$ if, for all roots $\alpha \in \overline{K}$ of the polynomial $X^\ell - a$, the extension $K(\alpha)/K$ is a proper geometric extension of fields. If $a \in K$ is geometric at all primes $\ell \neq p$, we refer to $a \in K$ as a **geometric element**.

Previous work has only considered geometric elements and the aim of this chapter is to prove Artin's primitive root conjecture for elements, which are not necessarily geometric. To this end we will prove a lemma providing equivalent characterisations of elements, which are not geometric.

Lemma 5.3.3. *Let $K = \mathbb{F}_q(X)$, let $a \in K$ and let $\ell \neq p$ be a rational prime. The following are equivalent:*

(i) a is not geometric at ℓ.

(ii) There exists $\mu \in \mathbb{F}_q$ and $b \in K$ such that $a = \mu b^\ell$.

(iii) There exists $\tilde{a} \in K\overline{\mathbb{F}}_q$ such that $a = \tilde{a}^\ell$.

Proof. (i) $\implies$ (ii): Since a is not geometric at ℓ, by definition there exists a root $\alpha \in \overline{K}$ of $X^\ell - a$ such that $K(\alpha)/K$ is either not a proper field extension or such that $K(\alpha) \cap \overline{\mathbb{F}}_q \neq \mathbb{F}_q$. In the former case, we may write $a = \mu b^\ell$ where $b = \alpha$ and $\mu = 1$, and we're done.

We therefore assume that $K(\alpha)/K$ is a proper field extension which is not geometric, i.e. $M := K(\alpha) \cap \overline{\mathbb{F}}_q \neq \mathbb{F}_q$. Note that since a is not an ℓ^{th} power the polynomial $X^\ell - a$ is irreducible over K (cf. [68, VI §9]) and therefore $[K(\alpha) : K] = \ell$. Since $M \supsetneq \mathbb{F}_q$ we further note that $K(\alpha) \supseteq MK \supsetneq K$. Since $[K(\alpha) : K] = \ell$ is prime, it follows that $MK = K(\alpha)$. Next, note that $M/\mathbb{F}_q$ is a finite extension of finite fields and hence Galois. It then further follows that the extension of composita $MK/\mathbb{F}_qK$ is also Galois. In other words, $K(\alpha)/K$ is Galois, and we conclude that $K(\alpha)$ is the splitting field of the polynomial $X^\ell - a$. In particular, it follows that $\zeta_\ell \in K(\alpha)$, where ζ_ℓ denotes a fixed primitive ℓ^{th} root of unity. Note further that $K(\alpha) \supseteq K(\zeta_\ell) \supseteq K$, and moreover that $K(\alpha) \neq K(\zeta_\ell)$ since $[K(\zeta_\ell) : K] \leq \ell - 1 < \ell$. Since ℓ is prime, it follows that $K(\zeta_\ell) = K$. Since elements in K of finite order lie in $K \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$, it follows that $\zeta_\ell \in \mathbb{F}_q^\times$. Noting that $\zeta_\ell \in \mathbb{F}_q^\times$ if and only if $\ell \mid q-1$, we further conclude that $\ell \mid q-1$.

By [91, Proposition 8.1], we find that $[M : \mathbb{F}_q] = [K(\alpha) : K] = \ell$. Since $\ell \mid q-1$, we moreover find that $\#\mathbb{F}_q^{\times,\ell} = \frac{q-1}{\ell}$, and in particular, that there exists an element $\mu \in \mathbb{F}_q^\times$, which is *not* an ℓ^{th} power. By [68, VI §9], we find that the polynomial $X^\ell - \mu$ is irreducible over $\mathbb{F}_q$, and thus by the uniqueness of finite field extensions, it follows that $M = \mathbb{F}_q(\beta)$, where $\beta^\ell = \mu \in \mathbb{F}_q^\times$. From [91, Proposition 8.1] it then also follows that $\{1, \beta, \ldots, \beta^{\ell-1}\}$ form a basis for $K(\alpha)/K$, and therefore $K(\alpha) = K(\beta)$. Hence there exist $b_i \in K$, $i = 0, 1, \ldots, \ell-1$ such that

$$\alpha = \sum_{i=0}^{\ell-1} b_i \beta^i.$$

Let σ be a non-trivial element of $\mathrm{Gal}(K(\alpha)/K)$, then $\sigma(\alpha) = \zeta_\ell^n \alpha$ and $\sigma(\beta) = \zeta_\ell^m \beta$ for two integers $n, m \in \{1, \ldots, \ell-1\}$. Thus we find

$$\sum_{i=0}^{\ell-1} b_i \zeta_\ell^n \beta^i = \zeta_\ell^n \alpha = \sigma(\alpha) = \sigma\left(\sum_{i=0}^{\ell-1} b_i \beta^i\right) = \sum_{i=0}^{\ell-1} b_i \zeta_\ell^{mi} \beta^i.$$

Since $\{1, \beta, \ldots, \beta^{\ell-1}\}$ are linearly independent over K, it follows that

$$b_i \zeta_\ell^n = b_i \zeta_\ell^{mi} \quad \text{for all} \quad 0 \leq i \leq \ell - 1,$$

and therefore whenever $n \not\equiv mi \pmod{\ell}$ this implies $b_i = 0$. Since there exists a unique $0 \leq i_0 \leq \ell - 1$ such that $n \equiv mi_0 \pmod{\ell}$, it follows that $\alpha = b_{i_0} \beta^{i_0}$, and therefore

$$a = \alpha^\ell = \tilde{\mu} b_{i_0}^\ell$$

where $\tilde{\mu} = \mu^{i_0} \in \mathbb{F}_q$. This shows the desired claim.

(ii) $\implies$ (iii): Set $\tilde{a} = \sqrt[\ell]{\mu} b$, where $\sqrt[\ell]{\mu}$ denotes any root of $X^\ell - \mu$ in $\overline{\mathbb{F}}_q$.

(iii) $\implies$ (i): If $\tilde{a} \in K$ then a is clearly not geometric at ℓ. So, assume $\tilde{a} \notin K$. Then $K \subsetneq K(\tilde{a}) \subset K\overline{\mathbb{F}}_q$. Since $K(\tilde{a})/K$ is a proper finite extension, in fact $K(\tilde{a}) \subseteq K\mathbb{F}_{q^n}$, for some $n > 1$. Moreover, since $\mathrm{Gal}(K\mathbb{F}_{q^n}/K) \cong \mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$, it follows that

$K\mathbb{F}_{q^n}/K$ is a cyclic extension. Thus, there exists a unique subgroup of $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$, of any given order dividing $n = |\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)|$. By the fundamental theorem of Galois theory, we then find that there exists a unique subextension of $K\mathbb{F}_{q^n}/K$ of degree $\ell = [K(\tilde{a}) : K]$, and thus may conclude that $K(\tilde{a}) = K\mathbb{F}_{q^\ell}$. By [91, Proposition 8.3], it follows that

$$K(\tilde{a}) \cap \overline{\mathbb{F}}_q = K\mathbb{F}_{q^\ell} \cap \overline{\mathbb{F}}_q = \mathbb{F}_{q^\ell} \supsetneq \mathbb{F}_q,$$

i.e. a is not geometric, as desired. □

Remark 5.3.4. Note that the first part of the proof of Lemma 5.3.3 shows that when a is not geometric at a prime ℓ such that $\ell \nmid q-1$ then a is already a full ℓ^{th} power in K. In particular, if $\ell \mid q^n - 1$ then by the same argument provided in Section 5.1.2, a is not be a primitive root modulo any closed point of degree n.

5.4 Quantitative results and Artin's primitive root conjecture

For positive integers m, k, consider the Ramanujan sum

$$c_m(k) := \sum_{\substack{1\le a\le m \\ (a,m)=1}} e\left(\frac{ak}{m}\right),$$

and recall the following elementary property

$$c_\ell(n) = \begin{cases} -1 & \text{if } \ell \nmid n, \\ \varphi(\ell) & \text{if } \ell \mid n, \end{cases} \tag{5.4.1}$$

for a rational prime ℓ. Our main result is the following:

Theorem 5.4.1. *Let $X/\mathbb{F}_q$ be a geometrically integral projective variety of dimension $r \ge 1$ with function field $K = \mathbb{F}_q(X)$. Let $g \in K \setminus \mathbb{F}_q$ be a rational function. Let $\mathscr{P}_g$ denote the set of primes $\ell \ne p$ at which g is not geometric. If g is a full ℓ^{th} power in K for some rational prime $\ell \mid q-1$ then $N_X(g,n) = 0$ for all $n \ge 1$. Otherwise we have*

$$N_X(g,n) = \rho_g(n)\left(\frac{\varphi(q^n-1)q^{n(r-1)}}{n} + O_{\varepsilon,X,g}\left(q^{n(r-1/2+\varepsilon)}\right)\right), \tag{5.4.2}$$

where

$$\rho_g(n) := \prod_{\substack{\ell \mid q^n-1 \\ \ell\in\mathscr{P}_g}} \left(1 - \frac{c_\ell\left(q^{n-1}+q^{n-2}+\cdots+1\right)}{\varphi(\ell)}\right). \tag{5.4.3}$$

Moreover, $\rho_g(n) > 0$ if and only if for all primes $\ell \in \mathscr{P}_g$ such that $\ell \mid q^n-1$ we have $\ell \mid q-1$ and $\ell \nmid n$.

Note that if g is geometric at every prime $\ell \mid q^n - 1$, then $\rho_g(n) = 1$. As noted in the introduction, we then recover equation (5.1.1) in the case when $r = 1$. By [78, Theorem 2.9], we moreover note that $\varphi(q^n - 1) \gg_\nu q^{n(1-\nu)}$ for any $\nu \in (0, 1)$. Thus (5.4.2) indeed yields a main term, in the limit as $n \to \infty$.

To finish this section we deduce Artin's primitive root conjecture over function fields in full generality from Theorem 5.4.1 by demonstrating that $N_X(g, n) > 0$ for infinitely many $n \in \mathbb{N}$.

Lemma 5.4.2. *Let $g \in K \setminus \mathbb{F}_q$. Then the set $\mathscr{P}_g$ of primes $\ell \neq p$ at which g is not geometric, is finite.*

Proof. By [107, Lemma 035Q] and [107, Lemma 0GK4], there exists a geometrically integral normal projective variety X_ν over $\mathbb{F}_q$, such that $\mathbb{F}_q(X_\nu) \cong \mathbb{F}_q(X)$. In particular, since X_ν is a geometrically integral projective variety, by [88, Proposition 2.2.22] we find that the global sections are given by $\mathcal{O}_{X_\nu}(X_\nu) = \mathbb{F}_q$.

Suppose $\mathscr{P}_g$ is infinite. Since $\mathcal{O}_{X_\nu}(X_\nu) \subset \overline{\mathbb{F}}_q$ it suffices to show that g lies in the global sections $\mathcal{O}_{X_\nu}(X_\nu)$, since then $g \in \mathbb{F}_q(X) \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$.

If $\mathscr{P}_g$ is infinite, then by Lemma 5.3.3 there exists an arbitrarily large $\ell \in \mathbb{N}$ such that $g = \mu b^\ell$, where $\mu \in \mathbb{F}_q$ and $b \in K$. Note that as Y ranges over prime divisors of X_ν, the maximum value of $|v_Y(g)|$ is bounded by $\deg(g)$. Let $\ell > \deg(g)$ such that $g = \mu b^\ell$. Then for any prime divisor $Y \subset X_\nu$, we find that

$$v_Y(g) = v_Y(\mu) + \ell \cdot v_Y(b) = \ell \cdot v_Y(b),$$

Thus $v_Y(g) = 0$ for any prime divisor $Y \subset X_\nu$, and in particular $v_Y(g) \geq 0$ for all prime divisors Y. It follows from [42, Proposition 6.3A] that $g \in \mathcal{O}_{X_\nu}(X_\nu)$, as desired. □

Corollary 5.4.3. *Artin's primitive root conjecture holds for any function field K over $\mathbb{F}_q$.*

Proof. Firstly note that any such field K is the function field of a geometrically integral, projective variety $X/\mathbb{F}_q$, i.e. $K = \mathbb{F}_q(X)$.

Let $g \in K \setminus \mathbb{F}_q$ and assume g is not a full ℓ^{th} power for any prime $\ell \mid q - 1$, so that (5.4.2) holds. We wish to show that there exist infinitely many closed points $\mathfrak{p}$ of X such that g is a primitive root modulo $\mathfrak{p}$, i.e. that there exist infinitely many $n \in \mathbb{N}$ such that $N_X(g, n) \neq 0$.

Note that $\rho_g(n) \geq 1$ whenever $\rho_g(n) \neq 0$. To show that $N_X(g, n) \neq 0$ infinitely often, it therefore suffices to show that there exist infinitely many $n \in \mathbb{N}$ such that

$\rho_g(n) > 0$. Let $\mathscr{P}_g = \{\ell_s : s \in S\}$ denote the set of primes $\ell \neq p$ at which g is not geometric. Since $g \notin \mathbb{F}_q$ we note that $\mathscr{P}_g$ is a finite set, by Lemma 5.4.2. Let $I \subset S$ be such that $i \in I$ whenever $\ell_i \mid q-1$, and let $J = S \setminus I$ be such that $j \in J$ whenever $\ell_j \nmid q-1$. Given $m \in \mathbb{N}$ we then consider

$$n = 1 + m \prod_{i \in I} \ell_i \prod_{j \in J} (\ell_j - 1). \tag{5.4.4}$$

We claim that the set of primes in $\mathscr{P}_g$, which divide $q^n - 1$, is precisely given by $\{\ell_i \colon i \in I\}$. Note first that $\ell_i \mid q^n - 1$ for all $i \in I$ since, in fact for any $n \in \mathbb{N}$, we have that $(q-1) \mid q^n - 1$. On the other hand, let $j_0 \in J$. Then $q \not\equiv 1 \mod \ell_{j_0}$ and thus

$$q^n \equiv q^{1+m\prod_{i\in I}\ell_i \prod_{j\in J}(\ell_j - 1)} \equiv q \not\equiv 1 \mod \ell_{j_0}$$

since $q^{\ell_{j_0}-1} \equiv 1 \mod \ell_{j_0}$ by Fermat's little theorem. Hence $\ell_{j_0} \nmid q^n - 1$.

Finally note that $n \not\equiv 0 \mod \ell_i$ for all $i \in I$. From the last part of Theorem 5.4.1 it then follows that $\rho_g(n) > 0$. The result now follows upon noting that there are infinitely many $n \in \mathbb{N}$ of the form in (5.4.4). □

5.5 A bound on exponential sums

One of the key ingredients of the proof of Theorem 5.4.1 is the following estimate for exponential sums, which is of independent interest. As we were unable to find a suitable result of our desired form in the existing literature, we present a proof here.

Proposition 5.5.1. *Let X be a geometrically integral projective variety of dimension r. Let $\chi \in \widehat{\mathbb{F}_q^\times}$ be a non-trivial character of order $\delta > 1$. Let $g \in K$ and assume that there exists a prime $\ell \mid \delta$ such that g is not of the form $g = b^\ell$ for some $b \in \overline{\mathbb{F}}_q(X)$. Write $\mathcal{R}_g \subset X(\mathbb{F}_q)$ for the set of the $\mathbb{F}_q$-rational points on X that are neither zeroes nor poles of g. Then*

$$\sum_{\rho \in \mathcal{R}_g} \chi(g(\rho)) \ll_X q^{r-1/2}.$$

Proof. Let $U \subset X$ be an affine open subset of X on which g is invertible, i.e. it has neither poles nor zeroes on U. It suffices to show the estimate for the sum over $U(\mathbb{F}_q)$ since $X \setminus U$ is a proper closed subset of X and thus by irreducibility of X has codimension at least 1. Therefore by the Lang–Weil bounds [69] the number of rational points we do not consider is bounded by $O(q^{r-1})$.

By Noether's normalization theorem there exists a finite surjective morphism $U \to \mathbb{A}^r$. Obviously there also exists a surjective map $\mathbb{A}^r \to \mathbb{A}^{r-1}$ by projecting on the first

$r-1$ coordinates, say. The composition of these maps yields a surjective morphism of locally finite type $\varphi\colon U \to \mathbb{A}^{r-1}$.

From Chevalley's upper semicontinuity theorem (cf. [41, Théorème 13.1.3]) it follows that the elements $x \in U$ such that $\dim \varphi^{-1}(\varphi(x)) > 1$ holds lie in a proper closed subset, which has dimension at most $r-1$ since U is irreducible. The number of rational points in this subset is bounded by $O(q^{r-1})$ via Lang–Weil and hence we may bound the contribution arising from these rational points trivially.

It therefore remains to estimate

$$\sum_{\substack{y \in \varphi(U)(\mathbb{F}_q) \\ \dim \varphi^{-1}(y)=1}} \sum_{\rho \in \varphi^{-1}(y)(\mathbb{F}_q)} \chi(g(\rho)),$$

where with an abuse of notation we write $\varphi^{-1}(y)$ for the fibre of the closed point in $\varphi(U)$ corresponding to y. On the fibres where $\dim \varphi^{-1}(y) = 1$ we apply a theorem of Perelmuter [82, Theorem 2] according to which we have

$$\sum_{\rho \in \varphi^{-1}(y)(\mathbb{F}_q)} \chi(g(\rho)) \ll_X q^{1/2}$$

uniformly in y, as long as g restricted to an irreducible component of $\varphi^{-1}(y)$ after changing the base field to $\overline{\mathbb{F}}_q$ is not a δ^{th} power of some element in $\overline{\mathbb{F}}_q(X)$. The remainder of the proof is concerned with showing that for generic $y \in \varphi(U)$, the element g is not an ℓ^{th} power restricted to an irreducible component of $\varphi^{-1}(y)_{\overline{\mathbb{F}}_q}$, where ℓ is as in the statement of the proposition, and hence Perelmuter's theorem is applicable.

Call $y \in \mathbb{A}^{r-1}$ *bad* if this occurs and *good* otherwise. For the sake of easing notation, but without loss of generality, in the following we will assume that φ is surjective onto $\mathbb{A}^{r-1}$ and that $\dim \varphi^{-1}(y) = 1$ for all $y \in \mathbb{A}^{r-1}$, since we took care of the other potential cases already. We claim that there exists a constructible set $C \subset \mathbb{A}^{r-1}$ that is contained in the set of good points with $\eta \in C$ where η is the generic point. Deferring the proof of this claim for now, by [107, Lemma 005K] we deduce that C contains an open dense subset in $\mathbb{A}^{r-1}$ and so $\mathbb{A}^{r-1} \setminus C$ is contained in a proper closed subset of $\mathbb{A}^{r-1}$. Since $\mathbb{A}^{r-1} \setminus C$ contains the set of bad points, by Lang–Weil the number of bad $\mathbb{F}_q$-rational points is bounded by $O(q^{r-2})$. Therefore trivially bounding the character sums for the fibres coming from $\mathbb{A}^{r-1} \setminus C$ the overall contribution is $O(q^{r-1})$.

To show the claim made above we will employ [107, Lemma 055B]. This states that if $h\colon Z \to Y$ is a morphism of finite presentation, and if $n_h\colon Y \to \{0, 1, \ldots\}$ is the number of irreducible components of the fibre $h^{-1}(y)$ after base change to $\overline{\mathbb{F}}_q$ then for any positive integer n the set

$$E_n = \{y \in Y : n_h(y) = n\}$$

is constructible. Recall that U is an affine open subset of X and hence is of the form $U = \mathrm{Spec}(R)$, where $R = \mathbb{F}_q[x_1, \ldots, x_n]/I$ for some ideal I. We may consider g restricted to U as an element in R since g is invertible on U. Consider $U_g = \mathrm{Spec}(R[z]/(z^\ell - g))$, and note that we have a natural map $\psi : U_g \to U$ induced by the inclusion map $R \hookrightarrow R[z]/(z^\ell - g)$. Write f for the composition $f = \varphi \circ \psi : U_g \to \mathbb{A}^{r-1}$. Note f is locally of finite type since all the schemes involved are Noetherian and so it follows from [107, Lemma 01TX] that f and φ are of finite presentation.

Note that the generic fibre $\varphi^{-1}(\eta)$ is integral with function field isomorphic to K and in particular it is also integral after changing base to $\overline{\mathbb{F}}_q$. Further η is good since g was assumed not to be an ℓ^{th} power in $\overline{\mathbb{F}}_q(X)$. Now the set

$$C = \{y \in \mathbb{A}^{r-1} : n_f(y) = 1\}$$

is constructible as mentioned above and clearly $\eta \in C$. Further note that if $n_f(y) = n_\varphi(n)$ then y is good. Otherwise, if y is bad then essentially by construction we have $n_\varphi(y) < n_f(y)$. Thus C is a constructible set contained in the set of good points y and also $\eta \in C$, as desired. $\square$

5.6 The proof of Theorem 5.4.1

Consider a finite cyclic group G of order M, and let $\widehat{G} = \mathrm{Hom}(G, \mathbb{C}^\times)$ denote its group of characters. Let

$$f_G(g) := \frac{\varphi(M)}{M} \prod_{p \mid M} \left(1 - \frac{\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = p}} \chi(g)}{\varphi(p)} \right).$$

We begin by noting the following general formula:

Lemma 5.6.1. *For $g \in G$, we have that*

$$f_G(g) = \frac{\varphi(M)}{M} \sum_{d \mid M} \frac{\mu(d)}{\varphi(d)} \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = d}} \chi(g) = \begin{cases} 1 & \text{if } g \text{ generates } G \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Suppose $g \in G$ does not generate G. Then we may write $g = h^p$ for some $h \in G$, where $p \mid M$ is prime. In such a case, we moreover find that

$$\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = p}} \chi(g) = \varphi(p),$$

and therefore $f_G(g) = 0$. Alternatively, suppose $g \in G$ generates G. Then

$$\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = p}} \chi(g) = -1,$$

Now it is easy to check that

$$\frac{M}{\varphi(M)} = \prod_{p \mid M} \left(1 + \frac{1}{p-1}\right),$$

and so we conclude that

$$f_G(g) = \begin{cases} 1 & \text{if } g \text{ generates } G \\ 0 & \text{otherwise,} \end{cases}$$

as desired. Finally, we note that for $(d_1, d_2) = 1$ and $g \in G$,

$$\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = d_1 d_2}} \chi(g) = \Bigg(\sum_{\substack{\psi \in \widehat{G} \\ \mathrm{ord}\psi = d_1}} \psi(g)\Bigg)\Bigg(\sum_{\substack{\phi \in \widehat{G} \\ \mathrm{ord}\phi = d_2}} \phi(g)\Bigg).$$

By multiplicativity, we thus conclude that

$$\begin{aligned} f_G(g) &= \frac{\varphi(M)}{M} \prod_{p \mid M} \left(1 - \frac{\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = p}} \chi(g)}{\varphi(p)}\right) \\ &= \frac{\varphi(M)}{M} \sum_{d \mid M} \frac{\mu(d)}{\varphi(d)} \sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = d}} \chi(g). \end{aligned}$$

□

Proof of Theorem 5.4.1. Let $X/\mathbb{F}_q$ be a geometrically integral projective variety. We write K for the function field of X and we fix algebraic closures $\overline{\mathbb{F}}_q$ and $\overline{K}$. Let $\mathfrak{p} \in X$ be a closed point of degree n, i.e. $[\kappa_{\mathfrak{p}} : \mathbb{F}_q] = n$. There then exists an isomorphism

$$\kappa_{\mathfrak{p}}^\times \cong \mathbb{F}_{q^n}^\times \subset \overline{\mathbb{F}}_q,$$

which may be explicitly described as follows. Let $[f] \in \kappa_{\mathfrak{p}}^\times$ denote a residue class represented by some $f \in K$ that is regular at $\mathfrak{p}$. We then map $[f]$ to $f(\rho)$, where ρ

is an $\mathbb{F}_{q^n}$-rational point of X corresponding to $\mathfrak{p}$. Note that since $f \in K$ is neither regular at $\mathfrak{p}$ nor vanishes at $\mathfrak{p}$ and f does not have a pole at ρ, and thus evaluating f at ρ, as in (5.2.1), is well-defined. For a fixed $g \in K$, we thus find that g is a primitive root mod $\mathfrak{p}$ if and only if g is regular at $\mathfrak{p}$ and its image $g(\rho)$ under this isomorphism generates $\mathbb{F}_{q^n}^\times$. Note that for any closed point of degree n there exist n different $\mathbb{F}_{q^n}$-rational points on X corresponding to it obtained by the action of the Frobenius element, see [42, Lemma II.4.4]. For an affine open $U \subset X$ this was explained in Section 5.2. It would also be sufficient to only consider these rational points on U, since the number of $\mathbb{F}_{q^n}$-rational points on $X \setminus U$ is $O(q^{r(n-1)})$ by Lang–Weil.

Let $\mathcal{R}_g^{(n)} \subset X(\mathbb{F}_{q^n})$ denote the set of $\mathbb{F}_{q^n}$-rational points on X that are neither zeroes nor poles of g, so that for any $\rho \in \mathcal{R}_g^{(n)}$, we have that $g(\rho) \in \mathbb{F}_{q^n}^\times$. If we then consider any $\rho \in \mathcal{R}_g^{(n)}$ corresponding to a closed point $\mathfrak{q}$ with $\deg(\mathfrak{q}) < n$ then $g(\rho)$ is contained in a proper subfield of $\mathbb{F}_{q^n}$ and therefore will not generate $\mathbb{F}_{q^n}^\times$. Here we may consider this subfield as a subset of $\mathbb{F}_{q^n}$ since we fixed an algebraic closure $\overline{\mathbb{F}}_q$. Further note that if $\rho \notin \mathcal{R}_g^{(n)}$ then it corresponds to a closed point for which g is either not regular or vanishes. Clearly g is not a primitive root for such a closed point since all powers of $g(\rho)$ will be contained in a proper subfield of $\mathbb{F}_{q^n}$.

Thus we find

$$N_X(g,n) = \frac{1}{n}\#\{\rho \in \mathcal{R}_g^{(n)} \colon \langle g(\rho)\rangle = \mathbb{F}_{q^n}^\times\}.$$

Combining this observation with Lemma 5.6.1 leads to

$$N_X(g,n) = \frac{\varphi(q^n-1)}{n(q^n-1)} \sum_{\rho \in \mathcal{R}_g^{(n)}} \prod_{\ell \mid q^n-1} \left(1 - \frac{\sum_{\substack{\chi \in \widehat{G} \\ \mathrm{ord}\chi = \ell}} \chi(g(\rho))}{\varphi(\ell)}\right).$$

Note that if g is an ℓ^{th} power, where $\ell \mid q-1$, then $N_X(g,n) = 0$ for all $n \geq 1$. This follows directly from the above formula but can also be seen from elementary group theoretic considerations, as noted in Section 5.1.2. Henceforth, we may thus assume that $g \in K$ is not an ℓ^{th} power for any prime $\ell \mid q-1$.

Let $\ell \mid q^n - 1$ be a prime such that g is not geometric, that is, $\ell \in \mathscr{P}_g$. Then by virtue of Lemma 5.3.3 we may write $g = \mu_\ell b_\ell^\ell$ for some $\mu_\ell \in \mathbb{F}_q^\times$ and some $b_\ell \in K$. Therefore, if $\mathrm{ord}(\chi) = \ell$ we have $\chi(g(\rho)) = \chi(\mu_\ell)$ for any $\rho \in \mathcal{R}_g^{(n)}$. Let r denote the order of $\mu_\ell \in \mathbb{F}_q^\times$. Since we assumed that g is not a full ℓ^{th} power for any prime $\ell \mid q-1$ we find that $\ell \nmid \frac{q-1}{r}$ or, equivalently, that $(\ell, \frac{q-1}{r}) = 1$.

For $x \in \mathbb{R}$, write $e(x) := e^{2\pi i x}$. Consider an embedding

$$\psi \colon \mathbb{F}_{q^n}^\times \hookrightarrow \mathbb{C}^\times$$

such that for $\mu_\ell \in \mathbb{F}_q^\times \subset \mathbb{F}_{q^n}^\times$ as above, we have

$$\psi(\mu_\ell) = e\left(\frac{1}{r}\right).$$

A character $\chi\colon \mathbb{F}_{q^n}^\times \to \mathbb{C}^\times$ of order ℓ then acts on an element $\alpha \in \mathbb{F}_{q^n}^\times$ via

$$\chi(\alpha) = \psi(\alpha)^{\frac{(q^n-1)a}{\ell}},$$

for some $a \in \{1, \ldots, \ell-1\}$. It follows that for any given $\rho \in \mathcal{R}_g^{(n)}$,

$$\begin{aligned}
\frac{1}{\varphi(\ell)} \sum_{\mathrm{ord}\chi=\ell} \chi(g(\rho)) &= \frac{1}{\varphi(\ell)} \sum_{1\le a\le \ell-1} e\left(\frac{1}{r}\right)^{\frac{(q^n-1)a}{\ell}} \\
&= \frac{1}{\varphi(\ell)} \sum_{1\le a\le \ell-1} e\left(\frac{a(q^n-1)}{\ell r}\right) \\
&= \frac{1}{\varphi(\ell)} c_\ell\left(\frac{q^n-1}{r}\right),
\end{aligned}$$

where $c_\ell\left(\frac{q^n-1}{r}\right)$ denotes a Ramanujan sum. Since $(\ell, \frac{q-1}{r}) = 1$ we have $\ell \mid \frac{q^n-1}{r}$ if and only if $\ell \mid q^{n-1} + \cdots + 1$, and so

$$c_\ell\left(\frac{q^n-1}{r}\right) = c_\ell\left(q^{n-1}+\cdots+1\right) = \begin{cases} -1 & \text{if } \ell \nmid q^{n-1}+\cdots+1 \\ \varphi(\ell) & \text{otherwise.} \end{cases}$$

Upon setting

$$\rho_g(n) := \prod_{\substack{\ell \mid q^n-1 \\ \ell \in \mathscr{P}_g}} \left(1 - \frac{c_\ell(q^{n-1}+\cdots+1)}{\varphi(\ell)}\right),$$

we therefore find that

$$N_X(g,n) = \rho_g(n) \frac{\varphi(q^n-1)}{n(q^n-1)} \sum_{\rho \in \mathcal{R}_g^{(n)}} \prod_{\substack{\ell \mid q^n-1 \\ \ell \notin \mathscr{P}_g}} \left(1 - \frac{\sum_{\substack{\chi\in\widehat{G} \\ \mathrm{ord}\chi=\ell}} \chi(g(\rho))}{\varphi(\ell)}\right). \tag{5.6.1}$$

For an integer $\delta \in \mathbb{N}$ write $(\delta, \mathscr{P}_g) = 1$ if $(\delta, \ell) = 1$ for every prime $\ell \in \mathscr{P}_g$. As in the proof of Lemma 5.6.1 we may then expand (5.6.1) to obtain

$$N_X(g,n) = \rho_g(n) \frac{\varphi(q^n-1)}{n(q^n-1)} \sum_{\substack{\delta \mid q^n-1 \\ (\delta, \mathscr{P}_g)=1}} \frac{\mu(\delta)}{\varphi(\delta)} \sum_{\mathrm{ord}\chi=\delta} \sum_{\rho \in \mathcal{R}_g^{(n)}} \chi(g(\rho)).$$

By the Lang–Weil bounds [69], the number of $\mathbb{F}_{q^n}$−rational points on X, denoted $\#X(\mathbb{F}_{q^n})$, is given by

$$|\#X(\mathbb{F}_{q^n}) - q^{nr}| \ll_X q^{n(r-1/2)}.$$

Noting, moreover, that g has at most m poles and zeroes, it follows that for fixed g,

$$|\#\mathcal{R}_g^{(n)} - q^{nr}| \leq m + O_X\left(q^{n(r-1/2)}\right) = O_{g,X}\left(q^{n(r-1/2)}\right),$$

and thus the contribution from the trivial character χ_0 is given by

$$\sum_{\rho \in \mathcal{R}_g^{(n)}} \chi_0(g(\rho)) = \#\mathcal{R}_g^{(n)} = q^{nr} + O\left(q^{n(r-1/2)}\right).$$

If $\delta \mid q^n - 1$ is such that $(\delta, \mathscr{P}_g) = 1$ and $\delta > 1$, then by Proposition 5.5.1 we moreover find that

$$\sum_{\rho \in \mathcal{R}_g^{(n)}} \chi(g(\rho)) = O\left(q^{n(r-1/2)}\right).$$

Combining the above observations, and applying the divisor bound $\sum_{\delta | q^n - 1} |\mu(\delta)| = O_\varepsilon(q^{n\varepsilon})$, we obtain

$$N_X(g,n) = \rho_g(n)\left(\frac{\varphi(q^n - 1)q^{n(r-1)}}{n} + O_\varepsilon\left(q^{n(r-1/2+\varepsilon)}\right)\right),$$

thereby yielding the first part of the theorem.

Finally, we turn to studying the product

$$\rho_g(n) = \prod_{\substack{\ell | q^n - 1 \\ \ell \in \mathscr{P}_g}} \left(1 - \frac{c_\ell\left(q^{n-1} + \cdots + 1\right)}{\varphi(\ell)}\right).$$

Let ℓ be a prime dividing $q^n - 1$ such that g is not geometric at ℓ. We proceed in two cases. First, suppose $\ell \mid q - 1$. Then $q \equiv 1 \mod \ell$, and therefore

$$q^{n-1} + \cdots + 1 \equiv 0 \mod \ell \iff n \equiv 0 \mod \ell.$$

By (5.4.1), we then find that $c_\ell(q^{n-1} + \cdots + 1) = -1$ if and only if $\ell \nmid n$. Otherwise, $c_\ell(q^{n-1} + \cdots + 1) = \varphi(\ell)$, in which case $\rho_n(g) = 0$.

Next, suppose $\ell \nmid q - 1$. Since $\ell \mid q^n - 1$ by assumption, we then find that $\ell \mid q^{n-1} + \cdots + 1$. By (5.4.1) it then follows that $c_\ell(q^{n-1} + \cdots + 1) = \varphi(\ell)$, and therefore $\rho_n(g) = 0$. In conclusion, we find that $\rho_g(n) > 0$ if and only if for all primes $\ell \in \mathscr{P}_g$ such that $\ell \mid q^n - 1$ we have $\ell \mid q - 1$ and $\ell \nmid n$, as desired. □

5.7 A heuristic interpretation

Artin arrived at the quantitative version of his primitive root conjecture using a well-known heuristic concerning the splitting properties of primes across the fields

$\mathbb{Q}(\zeta_\ell, g^{1/\ell})$, for varying primes $\ell \in \mathbb{N}$. In this section we suggest an analogous heuristic, which interprets the constant $\rho_g(n)$ in terms of splitting properties of primes in K. In contrast to the classical setting, we may obtain the correct density while maintaining the assumption that the various splitting conditions are independent, across the primes $\ell \in \mathbb{N}$.

In what follows, we restrict ourselves to the case in which K is a global function field, or equivalently, we assume that X is a normal geometrically irreducible projective curve over $\mathbb{F}_q$. A *prime* P of K is, by definition, a discrete valuation ring O_P with maximal ideal P whose field of fractions is K. Since K is a global function field, then the stalk $\mathcal{O}_{X,\mathfrak{p}}$ at each closed point $\mathfrak{p} \in X$ is a discrete valuation ring, and such stalks are in fact in $1:1$ correspondence with the primes in K. If L/K is a field extension then a prime $\mathfrak{P}$ of L is said to *lie above* P if $O_{\mathfrak{P}} \cap K = O_P$. We moreover say that P *splits completely* in L if the number of primes $\mathfrak{P}$ in L lying above P is equal to $[L : K]$.

Let $g \in K \setminus \mathbb{F}_q$. Let $\mathfrak{p} \in X$ be a closed point such that g is regular at $\mathfrak{p}$. Such $g \in K$ then fails to be primitive modulo $\mathfrak{p}$ if and only if the prime P corresponding to $\mathfrak{p}$ splits completely in $K_\ell := K(\sqrt[\ell]{g}, \zeta_\ell)$, the splitting field of $X^\ell - g$, for some prime $\ell \in \mathbb{N}$, where $\ell \nmid q$ [91, Lemma 10.1 and Proposition 10.6]. We may therefore formulate a heuristic for the density of primes P of degree n for which g is a primitive root by understanding the density of primes P in K which split completely in K_ℓ, for each prime $\ell \in \mathbb{N}$.

Suppose g is not a full ℓ^{th} power, for any prime $\ell \mid q-1$. Otherwise, $N_X(g, n) = 0$, trivially for all $n \geq 1$. Given a prime $\ell \in \mathbb{N}$, let

$$\mathbb{P}_\ell := \mathbb{P}(P \text{ splits completely in } K_\ell \mid \deg(P) = n).$$

Then under the heuristic assumption that the splitting conditions of P in K_ℓ are independent for the various fields K_ℓ, we can expect the desired density to be given by

$$A = \prod_{\ell} (1 - \mathbb{P}_\ell)\,.$$

Note that $\ell \mid q^n - 1$ if and only if P splits in $K(\zeta_\ell)$ (cf. [91, Proposition 10.2]). Thus $\ell \nmid q^n - 1$ implies that P does not split in K_ℓ, i.e. $\mathbb{P}_\ell = 0$ for all $\ell \nmid q^n - 1$, and thus

$$A = \prod_{\ell \mid q^n - 1} (1 - \mathbb{P}_\ell)\,.$$

Note that when $\ell \mid q^n - 1$ then again by [91, Proposition 10.2] we find that

$$\mathbb{P}_\ell = \mathbb{P}(P \text{ splits completely in } K_\ell \mid \deg(P) = n, P \text{ splits completely in } K(\zeta_\ell)).$$

Note that since K_ℓ is the splitting field of $X^\ell - g$, the field extension K_ℓ/K is Galois, and thus $K_\ell/K(\zeta_\ell)$ is also Galois, by the fundamental theorem of Galois theory. Moreover, since $K(\zeta_\ell)$ is the splitting field of $X^\ell - 1$, it follows that $K(\zeta_\ell)/K$ is also Galois. Now suppose P splits completely in $K(\zeta_\ell)$. Since $K_\ell/K(\zeta_\ell)$ and $K(\zeta_\ell)/K$ are both Galois extensions, we find that a given prime $\mathfrak{q}$ in $K(\zeta_\ell)$ lying above P splits in K_ℓ if and only if *all* such primes $\mathfrak{q}$ in $K(\zeta_\ell)$ lying above P split in K_ℓ (cf. [91, Proposition 9.3]).

Recall that since $\ell \mid q^n - 1$, every prime $\mathfrak{q}$ in $K(\zeta_\ell)$ lies above some prime $P \in K$ that splits completely in $K(\zeta_\ell)$. By the above remarks, we thus find that $\mathbb{P}_\ell$ is equal to the density of primes $\mathfrak{q}$ in $K(\zeta_\ell)$ which split completely in K_ℓ. Note that if g is geometric at ℓ then $K_\ell/K(\zeta_\ell)$ is a geometric extension. In such a case, we may apply *Chebotarev's density theorem* [91, Theorem 9.13B] to establish that the desired density is given by

$$\mathbb{P}_\ell = \frac{1}{[K_\ell : K(\zeta_\ell)]} = \frac{1}{\ell}.$$

If g is not geometric at ℓ, then we may no longer apply Chebatarev's density theorem. In such a case, however, we have sufficient information to compute $\mathbb{P}_\ell$ precisely. By Lemma 5.3.3, we write $g = \mu b^\ell$ where $\mu \in \mathbb{F}_q^\times$. Let r denote the order of μ in $\mathbb{F}_q^\times$, and let ζ denote a generator of $\mathbb{F}_q^\times$ such that $\mu = \zeta^{\frac{q-1}{r}}$. By [91, Proposition 10.6] we find that a prime P which splits completely in $K(\zeta_\ell)$ also splits completely in K_ℓ, if and only if

$$g^{\frac{q^n-1}{\ell}} \equiv \zeta^{\frac{q-1}{r}\cdot\frac{q^n-1}{\ell}} b^{\frac{q^n-1}{\ell}\cdot \ell} \equiv \zeta^{\frac{q-1}{r}\cdot\frac{q^n-1}{\ell}} \equiv 1 \mod P.$$

Note that this in turn occurs if and only if $q-1 \mid \frac{q-1}{r}\cdot\frac{q^n-1}{\ell}$, enabling us to conclude that

$$\mathbb{P}_\ell = \begin{cases} 1 & \text{if } \ell \mid \frac{q-1}{r}(q^{n-1}+\cdots+1) \\ 0 & \text{otherwise.} \end{cases}$$

In particular, since $\ell \mid q^n - 1 = (q-1)(q^{n-1}+\cdots+1)$, we find that $\mathbb{P}_\ell = 1$ whenever $\ell \nmid q-1$.

So suppose $\ell \mid q-1$. If $\ell \mid \frac{(q-1)}{r}$, then $\mu = \zeta^{\frac{q-1}{r}}$ is an ℓ^{th} power, in which case g is also an ℓ^{th} power, contradicting our initial assumption. We may therefore assume that $\ell \nmid \frac{(q-1)}{r}$. In this case, $\mathbb{P}_\ell = 1$ if and only if $\ell \mid (q^{n-1}+\cdots+1)$. Since $q \equiv 1 \mod \ell$, we find that

$$q^{n-1}+\cdots+1 \equiv n \mod \ell,$$

so that

$$\mathbb{P}_\ell = \begin{cases} 1 & \text{if } n \equiv 0 \mod \ell \\ 0 & \text{if } n \not\equiv 0 \mod \ell. \end{cases}$$

We thus conclude as follows. Suppose g is not a full ℓ^{th} power for any prime $\ell \mid q-1$. If, for all primes $\ell \in \mathscr{P}_g$ such that $\ell \mid q^n - 1$, we have $\ell \mid q-1$ and $n \not\equiv 0 \mod \ell$, then the density is given by

$$A = \prod_{\substack{\ell \mid q^n - 1 \\ \ell \notin \mathscr{P}_g}} \left(1 - \frac{1}{\ell}\right) \prod_{\substack{\ell \mid q^n - 1 \\ \ell \in \mathscr{P}_g}} 1.$$

Otherwise $A = 0$. In all cases, A is then given by

$$A = \prod_{\ell \mid q^n - 1} \left(1 - \frac{1}{\ell}\right) \prod_{\substack{\ell \mid q^n - 1 \\ \ell \in \mathscr{P}_g}} \left(1 - \frac{c_\ell \left(q^{n-1} + q^{n-2} + \cdots + 1\right)}{\varphi(\ell)}\right) = \frac{\varphi(q^n - 1)}{q^n - 1} \rho_g(n),$$

as expected.

Bibliography

[1] E. Artin. *The collected papers of Emil Artin.* Addison-Wesley Publishing Co., Inc., Reading, Mass.-London, 1965. Edited by Serge Lang and John T. Tate.

[2] R. C. Baker. Diagonal cubic equations. I. In *Théorie des nombres (Quebec, PQ, 1987)*, pages 15–28. de Gruyter, Berlin, 1989.

[3] R. C. Baker. Diagonal cubic equations. II. *Acta Arith.*, 53(3):217–250, 1989.

[4] R. C. Baker. Diagonal cubic equations. III. *Proc. London Math. Soc. (3)*, 58(3):495–518, 1989.

[5] V. V. Batyrev and Y. I. Manin. Sur le nombre des points rationnels de hauteur borné des variétés algébriques. *Math. Ann.*, 286(1-3):27–43, 1990.

[6] V. V. Batyrev and Y. Tschinkel. Rational points on some Fano cubic bundles. *C. R. Acad. Sci. Paris Sér. I Math.*, 323(1):41–46, 1996.

[7] H. Bilharz. Primdivisoren mit vorgegebener Primitivwurzel. *Math. Ann.*, 114(1):476–492, 1937.

[8] B. J. Birch. Homogeneous forms of odd degree in a large number of variables. *Mathematika*, 4:102–105, 1957.

[9] B. J. Birch. Forms in many variables. *Proc. Roy. Soc. London Ser. A*, 265:245–263, 1962.

[10] V. Blomer and J. Brüdern. Counting in hyperbolic spikes: the Diophantine analysis of multihomogeneous diagonal equations. *J. Reine Angew. Math.*, 737:255–300, 2018.

[11] J. Brandes and R. Dietmann. Rational lines on cubic hypersurfaces. *Math. Proc. Cambridge Philos. Soc.*, 171(1):99–112, 2021.

[12] J. Brandes and R. Dietmann. Rational lines on cubic hypersurfaces II. *In preparation*, 2023.

[13] T. D. Browning. *Cubic Forms and the Circle Method*. Progress in Mathematics. Birkhäuser, 2021.

[14] T. D. Browning and D. R. Heath-Brown. Forms in many variables and differing degrees. *J. Eur. Math. Soc. (JEMS)*, 19(2):357–394, 2017.

[15] T. D. Browning and D. R. Heath-Brown. Density of rational points on a quadric bundle in $\mathbb{P}^3 \times \mathbb{P}^3$. *Duke Math. J.*, 169(16):3099–3165, 2020.

[16] T. D. Browning and P. Vishe. Cubic hypersurfaces and a version of the circle method for number fields. *Duke Math. J.*, 163(10):1825–1883, 2014.

[17] T. D. Browning and P. Vishe. Rational points on cubic hypersurfaces over $\mathbb{F}_q(t)$. *Geometric and functional analysis.*, 25(3):671–732, 2015.

[18] J. Brüdern, R. Dietmann, J. Y. Liu, and T. D. Wooley. A Birch–Goldbach theorem. *Arch. Math. (Basel)*, 94(1):53–58, 2010.

[19] M. Car. Sommes d'exponentielles dans $\mathbb{F}_{2^h}((x^{-1}))$. *Acta Arithmetica*, 62:303–328, 1992.

[20] M. Car and J. Cherly. Sommes de cubes dans l'anneau $\mathbb{F}_{2^h}[X]$. *Acta Arithmetica*, 65(3):227–241, 1993.

[21] M. Car and L. Gallardo. Sums of cubes of polynomials. *Acta Arithmetica*, 112:41–50, 2004.

[22] Y.-M. J. Chen, Y. Kitaoka, and J. Yu. On primitive roots of tori: the case of function fields. *Mathematische Zeitschrift*, 243(2):201–215, 2003.

[23] J.-L. Colliot-Thélène. Points rationnels sur les fibrations. In *Higher dimensional varieties and rational points*, pages 171–221. Springer, 2003.

[24] B. Cook and Á. Magyar. Diophantine equations in the primes. *Invent. Math.*, 198(3):701–737, 2014.

[25] H. Davenport. Private letter to Keith Matthews. 1964. Online summary of the letter at http://www.numbertheory.org/thesis/remarks_by_davenport.pdf .

[26] H. Davenport. Cubic forms in thirty-two variables. *Philos. Trans. Roy. Soc. London Ser. A*, 251:193–232, 1959.

[27] H. Davenport. Cubic forms in 29 variables. *Proc. Roy. Soc. London Ser. A*, 266:287–298, 1962.

[28] H. Davenport. Cubic forms in sixteen variables. *Proc. Roy. Soc. London Ser. A*, 272:285–303, 1963.

[29] H. Davenport. *Analytic methods for Diophantine equations and Diophantine inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2005.

[30] P. Deligne. La conjecture de Weil: I. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 43(1):273–307, 1974.

[31] P. Deligne. La conjecture de Weil: II. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 52:137–252, 1980.

[32] V. B. Demyanov. On cubic forms in discretely normed fields. *Doklady Akad. Nauk SSSR (N.S.)*, 74:889–891, 1950.

[33] R. Dietmann. Weyl's inequality and systems of forms. *Q. J. Math.*, 66(1):97–110, 2015.

[34] R. Dietmann and T. D. Wooley. Pairs of cubic forms in many variables. *Acta Arith.*, 110(2):125–140, 2003.

[35] W. Duke, J. Friedlander, and H. Iwaniec. Bounds for automorphic L-functions. *Inventiones mathematicae*, 112(1):1–8, 1993.

[36] E. Eisenstein, L. K. Jain, and W. Kuo. On extending Artin's conjecture to composite moduli in function fields. *J. Number Theory*, 213:285–318, 2020.

[37] J. Franke, Y. I. Manin, and Y. Tschinkel. Rational points of bounded height on Fano varieties. *Invent. Math.*, 95(2):421–435, 1989.

[38] L. Gallardo. On the restricted Waring problem over $\mathbb{F}_{2^n}[t]$. *Acta Arithmetica*, 92(2):109–113, 2000.

[39] L. Gallardo. Waring's problem for cubes and squares over a finite field of even characteristic. *Bull. Belg. Math. Soc. Simon Stevin*, 12(3):349–362, 2005.

[40] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[41] A. Grothendieck. Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas IV. *Inst. Hautes Études Sci. Publ. Math.*, (32):361, 1967.

[42] R. Hartshorne. *Algebraic geometry*, volume 52. Springer Science & Business Media, 2013.

[43] D. R. Heath-Brown. Cubic forms in ten variables. *Proc. London Math. Soc. (3)*, 47(2):225–257, 1983.

[44] D. R. Heath-Brown. Artin's conjecture for primitive roots. *Quart. J. Math. Oxford Ser. (2)*, 37(145):27–38, 1986.

[45] D. R. Heath-Brown. A new form of the circle method, and its application to quadratic forms. *J. Reine Angew. Math.*, 481:149–206, 1996.

[46] D. R. Heath-Brown. The circle method and diagonal cubic forms. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 356(1738):673–699, 1998.

[47] D. R. Heath-Brown. Counting rational points on cubic surfaces. Number 251, pages 13–30. 1998. Nombre et répartition de points de hauteur bornée (Paris, 1996).

[48] D. R. Heath-Brown. The solubility of diagonal cubic Diophantine equations. *Proc. London Math. Soc. (3)*, 79(2):241–259, 1999.

[49] D. R. Heath-Brown. Cubic forms in 14 variables. *Invent. Math.*, 170(1):199–230, 2007.

[50] C. Hooley. On Artin's conjecture. *J. Reine Angew. Math.*, 225:209–220, 1967.

[51] C. Hooley. On Waring's problem. *Acta Mathematica*, 157:49 – 97, 1986.

[52] C. Hooley. On nonary cubic forms. *J. Reine Angew. Math.*, 386:32–98, 1988.

[53] C. Hooley. On nonary cubic forms. II. *J. Reine Angew. Math.*, 415:95–165, 1991.

[54] C. Hooley. On nonary cubic forms. III. *J. Reine Angew. Math.*, 456:53–63, 1994.

[55] C. Hooley. On Hypothesis K^* in Waring's problem. In *Sieve methods, exponential sums, and their applications in number theory (Cardiff, 1995)*, volume 237 of *London Math. Soc. Lecture Note Ser.*, pages 175–185. Cambridge Univ. Press, Cambridge, 1997.

[56] C. Hooley. On nonary cubic forms. IV. *J. Reine Angew. Math.*, 680:23–39, 2013.

[57] C. Hooley. On octonary cubic forms. *Proc. London Math. Soc.*, 109(1):241–281, 2014.

[58] C. Hooley. On octonary cubic forms: II. *Bull. Lond. Math. Soc.*, 47(1):85–94, 2015.

[59] C.-H. Hsu. On Artin's conjecture for the Carlitz module. *Compositio Mathematica*, 106(3):247266, 1997.

[60] C.-N. Hsu and J. Yu. On Artin's conjecture for rank one Drinfeld modules. *J. Number Theory*, 88(1):157–174, 2001.

[61] L. Hua. *Additive theory of prime numbers*, volume 13 of *Translations of mathematical monographs 13.* American Mathematical Society, Providence, Rhode Island, 1965. XIII, 190 S.

[62] J. Igusa. *An introduction to the theory of local zeta functions*, volume 14. American Mathematical Soc., 2007.

[63] E. Jensen and M. Ram Murty. Artin's conjecture for polynomials over finite fields. In *Number theory*, pages 167–181. Basel: Birkhäuser, 2000.

[64] S. Kim and M. Ram Murty. Artin's primitive root conjecture for function fields revisited. *Finite Fields Appl.*, 67:15, 2020. Id/No 101713.

[65] S. Kim and M. Ram Murty. Corrigendum to: "Artin's primitive root conjecture for function fields revisited". *Finite Fields Appl.*, 78:4, 2022. Id/No 101963.

[66] R. M. Kubota. Waring's problem for $\mathbf{F}_q[x]$. *Dissertationes Math. (Rozprawy Mat.)*, 117:60, 1974.

[67] W. Kuo and D. Tweedle. Artin's conjecture for Drinfeld modules. *Algebra Number Theory*, 16(5):1025–1070, 2022.

[68] S. Lang. *Algebra*, volume 211. Springer Science & Business Media, 2012.

[69] S. Lang and A. Weil. Number of points of varieties in finite fields. *Amer. J. Math.*, 76:819–827, 1954.

[70] S. A. Lee. *On the applications of the circle method to function fields, and related topics*. PhD thesis, University of Bristol, 2013.

[71] D. H. Lehmer and E. Lehmer. Heuristics, anyone? In *Studies in mathematical analysis and related topics*, pages 202–210. Stanford Univ. Press, Stanford, Calif., 1962.

[72] D. J. Lewis. Cubic homogeneous polynomials over p-adic number fields. *Ann. of Math. (2)*, 56:473–478, 1952.

[73] D. J. Lewis. Cubic forms over algebraic number fields. *Mathematika*, 4:97–101, 1957.

[74] R. Lidl and H. Niederreiter. *Finite Fields*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1996.

[75] U. V. Linnik. On the representation of large numbers as sums of seven cubes. *Rec. Math. [Mat. Sbornik] N. S.*, 12(54):218–224, 1943.

[76] Y.-R. Liu and T. D. Wooley. Waring's problem in function fields. *J. Reine Angew. Math.*, 2010(638):1–67, 2010.

[77] D. Loughran. Rational points of bounded height and the Weil restriction. *Israel J. Math.*, 210(1):47–79, 2015.

[78] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.

[79] W. Müller. Systems of quadratic Diophantine inequalities. *J. Théor. Nombres Bordeaux*, 17(1):217–236, 2005.

[80] W. Müller. Systems of quadratic Diophantine inequalities and the value distribution of quadratic forms. *Monatsh. Math.*, 153(3):233–250, 2008.

[81] F. Pappalardi and I. Shparlinski. On artins conjecture over function fields. *Finite Fields Appl*, 1(4):399–404, 1995.

[82] G. I. Perelmuter. Estimation of a sum along an algebraic curve. *Mathematical notes of the Academy of Sciences of the USSR*, 5(3):223–227, 1969.

[83] C. Perng. A note on Bilharz's example regarding nonexistence of natural density. *Int. Math. Forum*, 7(37-40):1877–1884, 2012.

[84] E. Peyre. Hauteurs et mesures de Tamagawa sur les variétés de Fano. *Duke Math. J.*, 79(1):101–218, 1995.

[85] E. Peyre. Liberté et accumulation. *Doc. Math.*, 22:1615–1659, 2017.

[86] E. Peyre and G. Rémond, editors. *Arakelov geometry and Diophantine applications*, volume 2276 of *Lecture Notes in Mathematics*. Springer, 2021.

[87] P. A. B. Pleasants. Cubic polynomials over algebraic number fields. *J. Number Theory*, 7(3):310–344, 1975.

[88] B. Poonen. *Rational points on varieties*, volume 186. American Mathematical Soc., 2017.

[89] C. P. Ramanujam. Cubic forms over algebraic number fields. *Proc. Cambridge Philos. Soc.*, 59:683–705, 1963.

[90] M. Robbiani. On the number of rational points of bounded height on smooth bilinear hypersurfaces in biprojective space. *J. London Math. Soc. (2)*, 63(1):33–51, 2001.

[91] M. Rosen. *Number theory in function fields*, volume 210. Springer Science & Business Media, 2002.

[92] C. Ryavec. Cubic forms over algebraic number fields. *Proc. Cambridge Philos. Soc.*, 66:323–333, 1969.

[93] S. L. Rydin Myerson. Systems of forms in many variables. *arXiv preprint arXiv:1709.08917*, 2017.

[94] S. L. Rydin Myerson. Quadratic forms and systems of forms in many variables. *Invent. Math.*, 213(1):205–235, 2018.

[95] S. L. Rydin Myerson. Systems of cubic forms in many variables. *J. Reine Angew. Math.*, 757:309–328, 2019.

[96] W. Sawin. The equidistribution of L-functions of twists by Witt vector Dirichlet characters over function fields. *arXiv preprint arXiv:1805.04330*, 2018.

[97] W. Sawin. Freeness alone is insufficient for Manin-Peyre. *arXiv preprint arXiv:2001.06078*, 2020.

[98] D. Schindler. Bihomogeneous forms in many variables. *J. Théor. Nombres Bordx.*, 26(2):483–506, 2014.

[99] D. Schindler. A variant of Weyl's inequality for systems of forms and applications. In *Advances in the theory of numbers*, volume 77 of *Fields Inst. Commun.*, pages 207–218. Fields Inst. Res. Math. Sci., Toronto, ON, 2015.

[100] D. Schindler. Manin's conjecture for certain biprojective hypersurfaces. *J. Reine Angew. Math.*, 714:209–250, 2016.

[101] W. M. Schmidt. The density of integer points on homogeneous varieties. *Acta Math.*, 154:243–296, 1985.

[102] J.-P. Serre. *Cours d'arithmétique*, volume 2 of *Collection SUP: "Le Mathématicien"*. Presses Universitaires de France, Paris, 1970.

[103] J.-P. Serre. *Topics in Galois theory*, volume 1 of *Research Notes in Mathematics*. A K Peters, Ltd., Wellesley, MA, second edition, 2008. With notes by Henri Darmon.

[104] C. M. Skinner. Rational points on nonsingular cubic hypersurfaces. *Duke Math. J.*, 75(2):409–466, 1994.

[105] C. M. Skinner. Forms over number fields and weak approximation. *Compositio Math.*, 106(1):11–29, 1997.

[106] C. V. Spencer. The Manin conjecture for $x_0y_0 + \cdots + x_sy_s = 0$. *J. Number Theory*, 129(6):1505–1521, 2009.

[107] The Stacks Project Authors. *Stacks Project*. `https://stacks.math.columbia.edu`, 2018.

[108] P. Stevenhagen. The correction factor in Artin's primitive root conjecture. volume 15, pages 383–391. 2003. Les XXIIèmes Journées Arithmetiques (Lille, 2001).

[109] Z. Tian. Hasse principle for three classes of varieties over global function fields. *Duke Math. J.*, 166(17):3349–3424, 2017.

[110] Z. Tian and L. Zhang. Weak approximation for cubic hypersurfaces and degree 4 del Pezzo surfaces. *Int. Math. Res. Not.*, 2018(3):762–784, 2018.

[111] R. C. Vaughan. On Waring's problem for cubes. *J. Reine Angew. Math.*, 1986(365):122–170, 1986.

[112] R. C. Vaughan. On Waring's problem for cubes. II. *J. London Math. Soc. (2)*, 39(2):205–218, 1989.

[113] V. Y. Wang. Approaching cubic Diophantine statistics via mean-value L-function conjectures of Random Matrix Theory type. *arXiv preprint arXiv:2108.03398*, 2021.

[114] V. Y. Wang. Diagonal cubic forms and the large sieve. *arXiv preprint arXiv:2108.03395*, 2021.

[115] V. Y. Wang. Isolating special solutions in the delta method: The case of a diagonal cubic equation in evenly many variables over $\mathbb{Q}$. *arXiv preprint arXiv:2108.03396*, 2021.

[116] A. Weil. *Sur les courbes algébriques et les variétés qui s'en déduisent*. Number 1041. Hermann, 1948.

[117] T. D. Wooley. Linear spaces on cubic hypersurfaces, and pairs of homogeneous cubic equations. *Bull. London Math. Soc.*, 29(5):556–562, 1997.

[118] S. Yamagishi. Prime solutions to polynomial equations in many variables and differing degrees. *Forum Math. Sigma*, 6:Paper No. e19, 89, 2018.

[119] S. Yamagishi. Birch's theorem on forms in many variables with a Hessian condition. *arXiv preprint arXiv:2304.02620*, 2023.

[120] W.-C. Yao and J. Yu. On primitive roots for rank one Drinfeld modules. *J. Number Theory*, 130(2):370–385, 2010.