Universiteit Utrecht

Master's Thesis

# Two Problems Related to the Circle Method

Jan-Willem van Ittersum
June 26, 2017

*Supervisor:* Dr. Damaris Schindler
*Second reader:* Prof. Dr. Frits Beukers

Utrecht University
Faculty of Science
Department of Mathematics

Jan-Willem van Ittersum, BSc:
Two Problems Related to the Circle Method
Master's thesis, Mathematical Sciences

Supervisor: Dr. Damaris Schindler
Second reader: Prof. Dr. Frits Beukers

Time frame: September 2016 – June 2017

## Abstract

This thesis consists of two parts. In the first part, we consider a system of polynomials $f_1, \ldots, f_R \in \mathbb{Z}[x_1, \ldots, x_n]$ of the same degree with non-singular local zeros and $n$ large compared to the Birch singular locus of these polynomials. Generalising the work of Birch [Bir62] on the circle method, we find quantitative asymptotics (in terms of the maximal coefficient of these polynomials) for the number of integer zeros of this system within a growing box. Using a quantitative version of the Nullstellensatz, we obtain an upper bound on the smallest integer zero provided the system of polynomials is non-singular.

In the second part we compute a $q$-hypergeometric generating series for overpartitions of a given rank $d$ where the difference between two successive parts may be odd only if the larger part is overlined. We show that all coefficients are divisible by 3 except for the coefficient of $q^{d^2}$.

## Acknowledgements

First and foremost, I wish to express my gratitude to Damaris Schindler for being my supervisor. I greatly benefited from her suggested research questions, guidance and helpful suggestions during our meetings. Also, I appreciated the freedom to work on the problems I was the most interested in.

I would like to thank Frits Beukers for being the second reader of this thesis.

I enjoyed to be part of the Thesis Seminar, where fellow Master's students and I presented and discussed the progress made in our theses. I would like to thank the other students who took part in this seminar.

I thank Merlijn Staps for reading a previous version of this thesis and providing me with many useful comments.

# Contents

# 1 Introduction

A famous arithmetical problem is whether every natural number can be represented as the sum of four squares. That is, does the polynomial

$$f(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - N$$

has an integer zero for every $N \in \mathbb{N}$? That this is indeed the case is Lagrange's four-square theorem. In his *Meditationes algebraicae* (1770), Edward Waring generalised this problem stating that *every number is expressible as a sum of 4 squares, or 9 cubes, or 19 biquadrates, 'and so on'*. More formally, this can be stated as: given $k \in \mathbb{N}$, there is an $n \in \mathbb{N}$ such that for all $N \in \mathbb{N}$ the polynomial

$$f(x_1, \dots, x_n) = x_1^k + \dots + x_n^k - N \tag{1.1}$$

has an integer zero $\underline{x} = (x_1, \dots, x_n)$ with $x_i \geq 0$ for all $i$. For this problem, Hardy and Littlewood developed what is nowadays called *Hardy-Littlewood circle method*. Initially, this method involved a contour integral over the unit circle, which explains the word *circle*. However, in the modern formulation, exponential sums take the role of this contour.

In this thesis, we apply the circle method to two generalisations of Waring's problem. In the first part we replace the polynomial (1.1) by a system of polynomials $\underline{f} = (f_1, \dots, f_R) \in \mathbb{Z}[x_1, \dots, x_n]$. Such a system has (infinitely many) common integer zeros provided that the system has non-singular real and $p$-adic common zeros and 'enough' variables (to be made precise later) as shown by Birch [Bir62]. We study the distribution of these zeros. In particular we provide an upper bound on the smallest integer zero. This upper bound is original work and generalises known upper bounds in case the polynomials have degree at most 3. In Section 2.2 we give a more extended introduction to this question and state our results. These results are proven in Section 2.4 and Section 2.5.

In the second part we are interested in partitions of integers. Any decomposition $N = x_1 + x_2 + \cdots$, into positive integers $x_i$ without regard to their order is called a partition of $N$. Note that for $k = 1$ a zero of $f$ in (1.1) corresponds to a partition of $N$, but that different zeros may correspond to the same partition. The number of partitions of $N$ can be given in closed form using the circle method. We study a generalisation of partitions, so-called restricted overpartitions. We prove a few congruence identities for these partitions, among which new results about the 3-divisibility of overpartitions with restricted odd differences. This is the content of Chapter 3.

# 2 Quantitative results on Diophantine equations in many variables

## 2.1 Background: the Hasse principle

A typical question in number theory is whether a given equation has integer solutions. Sometimes it is easy to show that this is not the case. For example, the equation $x^2 + y^2 = -1$ has no integer solutions $(x, y)$, because it has no real solutions. Also, the equation $x^2 - 5y^2 = 2$ has no integer solutions $(x, y)$, because it has no solutions modulo 5. These two examples illustrate two necessary conditions for an equation to have integer solutions, namely:

(i) The equation has a real solution;

(ii) The equation has a solution modulo $N$ for every positive integer $N$.

Are these two conditions sufficient? In general, the answer is 'no'. However, for some kinds of equations (i) and (ii) are sufficient. In such a case, the intuition is the following: one can find an integer solution to such an equation by using the Chinese remainder theorem to piece together solutions modulo powers of each different prime number. In the following example, only knowledge of a solution modulo $N$ for a specific positive integer $N$ is enough to find the integer solution:

**Example 2.1.** Consider the linear equation $ax = b$ with $a, b \in \mathbb{Z}$ and assume for simplicity that $a, b > 0$. Of course, we know that this equation has an integer solution if and only if $a \mid b$. In this example we show that it is enough to know $x \mod ab$ to find an integer solution $x$.

Suppose we have a solution $x_0 \in \mathbb{Z}$ of $ax_0 \equiv b \mod ab$. This implies in particular that $a \mid b$. Note that such a solution $x_0 \mod ab$ is generally not unique. We have that $x_0 \equiv \frac{b}{a} \mod b$, where $\frac{b}{a}$ is an integer as $a \mid b$. Because $\frac{b}{a} \leq b$, this implies that there exists an $x \in \mathbb{Z}$ such that $ax \equiv b \mod ab$ and $0 \leq x_0 < b$. Then, $0 \leq ax < ab$. Hence, the equation $ax = b \mod ab$ implies that $ax = b$.

Condition (i) and (ii) are called 'local' conditions. In modern number theory these are formulated in terms of local fields $\mathbb{R}$ and $\mathbb{Q}_p$. Here $\mathbb{Q}_p$ ($p$ prime) is the field of $p$-adic integers with ring of integers given by $\mathbb{Z}_p$. The second condition can be replaced by:

(ii') The equation has solutions over $\mathbb{Z}_p$ for all primes $p$.

So, by trying to answer the question whether an equation has an integer solution we encounter the $p$-adic numbers. Historically, it was for this reason that it became clear that the $p$-adic numbers have a conceptual role in mathematics. Namely, in the 1920s Hasse used the $p$-adic integers, introduced by Hensel, to express Minkowski's work on quadratic forms over the rational numbers in terms of quadratic forms over the real and $p$-adic numbers [GBGL10, p. 243] (a form is a homogeneous polynomial). This work lead to the *Hasse principle* or the *local-global principle*. The Hasse principle is not a theorem, but more a philosophy that can be formulated as:

*"studying a problem over $\mathbb{Q}$ is equivalent with studying it over $\mathbb{R}$ and all $\mathbb{Q}_p$".*

This philosophy can be stated more generally as:

*"studying a problem over a number field $k$ is equivalent with studying it over all completions $k_v$ of $k$".*

We mainly study problems over $\mathbb{Z}$ to find integer solutions. In this case we can state the Hasse principle as

*"studying a problem over $\mathbb{Z}$ is equivalent with studying it over $\mathbb{R}$ and all $\mathbb{Z}_p$".*

If this is the case, we say that the Hasse principle holds. In Example 2.1 the Hasse principle holds. By the Hasse-Minkowski theorem the Hasse principle holds for representing 0 by quadratic forms, i.e. for equations of the form $f = 0$, where $f$ is a homogeneous polynomial of degree 2.

A counterexample to the Hasse principle is given by Selmer [Sel51]. He showed that the cubic equation

$$3x^3 + 4y^3 + 5z^3 = 0$$

has non-trivial $p$-adic and real solutions, but no non-trivial rational solutions. An easier counterexample is the following:

**Example 2.2.** The equation

$$(x^2 - 2)(x^2 - 17)(x^2 - 34) = 0 \tag{2.1}$$

is a counterexample to the Hasse principle. A real solution is given by $\sqrt{2}$. The $p$-adic solutions are constructed in the following way. Let $f_a(x) = x^2 - a$, so that equation (2.1) is given by $f_2(x)f_{17}(x)f_{34}(x) = 0$. If $p$ is an odd prime number for which 2 is a quadratic residue modulo $p$, then the equation $f_2(x)$ has a non-zero solution $x_0$ modulo $p$. Moreover, $f_2'(x_0) = 2x_0 \neq 0 \mod p$. Hence, by Hensel's lemma, we obtain a zero of $f_2(x)$ in $\mathbb{Z}_p$. This zero is also solution of (2.1). Similarly, for a prime $p$ with $p \nmid a$ for which $a$ is a quadratic residue modulo $p$, we can find a zero of $f_a(x)$ in $\mathbb{Z}_p$. For $a = 2, 17$ and 34 this yields $p$-adic solutions of (2.1). Now, observe that for all primes $p$ we have the following identity of Legendre symbols:

$$\left(\frac{2}{p}\right)\left(\frac{17}{p}\right) = \left(\frac{34}{p}\right).$$

Hence, at least one of 2, 17 and 34 is a quadratic residue modulo $p$ for $p \neq 2, 17$, so (2.1) has $p$-adic solutions for $p \neq 2, 17$. Observe that for $p = 17$ we have that 2 is a quadratic residue modulo 17. For $p = 2$, observe that $f_{17}(1) = -16 \equiv 0 \mod 2^3$, whereas $f_{17}'(1) = 2 \not\equiv 0 \mod 2^2$. Hence, by Hensel's lemma, we also find a solution of (2.1) in $\mathbb{Z}_2$. It remains to show that there are no rational solutions. This follows directly because all real solutions, given by $\pm\sqrt{2}, \pm\sqrt{17}, \pm\sqrt{34}$, are irrational.

Observe that the polynomial $f(x) = (x^2-2)(x^2-17)(x^2-34)$ is reducible over $\mathbb{Q}$. Often when working with the Hasse principle, an irreducibility condition is imposed. The intuition behind such a condition is that real and $p$-adic zero's of $f$ which are in fact zeros of, say, $x^2 - 2$, do not give information about the integer zero's of, say, $x^2 - 17$. Although for $f_a(x) = x^2 - a$ with $a \in \mathbb{Z}$ the Hasse principle does hold, it does not hold in Example 2.2, because it is impossible to combine the real and $p$-adic zeros of $f_2, f_{17}$ and $f_{34}$ to find an integer zero.

As these counterexamples indicate, the Hasse principle is often too much to hope for. However, the idea of attacking a problem locally and then putting the local information together to obtain a global solution is a fundamental idea in modern mathematics. This idea is used in the proof of the modularity theorem and a central idea in the Langlands programme [GBGL10, p. 243].

## 2.2 Integer zeros of polynomials and the circle method

### 2.2.1 Polynomials in many variables

In the first part, we are interested in integer zeros of polynomials with integer coefficients. Often authors restrict theirselves to integer zeros of forms, i.e. homogeneous polynomials. Let $f$ be a rational cubic form (a homogeneous polynomial of degree 3 with rational coefficients) in $n$ variables. It is conjectured that $f$ has a rational zero as soon as $n \geq 10$. Note that multiplying such a rational zero $\underline{x}$ by the lowest common multiple of all denominators of the $x_i$ yields an integer zero of $f$, because $f$ is homogeneous. Davenport adapted the circle method to prove that $f$ has an integer zero if $n \geq 32$ [Dav59]. In later work he showed that $f$ has an integer zero if $n \geq 16$ [Dav63]. At the moment, the best result is due to Heath-Brown, who showed the same result for $n \geq 14$ [HB07]. So, trivially, the Hasse principle holds for rational cubic forms of at least 14 variables. Hooley improved on the number of variables by taking the geometry of the form $f$ into account. He showed that the Hasse principle holds for $f$ if $n \geq 9$ and the projective cubic hypersurface defined by $f$ is non-singular [Hoo88].

Birch generalises the work of Davenport to forms of arbitrary degree. In order to so, he adds geometric conditions to the work of Davenport: not only are the local zeros required to be *non-singular*, but also is the number of variables required to be large compared to degree of the forms and some *locus of singularities*. This locus is called the Birch singular locus and may be non-empty even if the projective variety associated to the system of forms is non-singular. A special case of Birch's work is the following: under the assumption that this projective variety is non-singular, the Hasse principle is satisfied as long as the number of variables is large compared to this Birch singular locus. We proceed without this assumption of non-singularity until we impose it in Section 2.5.

We now introduce some notation to be more precise about the work of Birch. Consider a system of polynomials $f_1, \ldots, f_R \in \mathbb{Z}[x_1, \ldots, x_n] = \mathbb{Z}[\underline{x}]$ of which every polynomial $f_i$ has the same degree $d \geq 2$. We write $\underline{f}(\underline{x}) = (f_1(\underline{x}), \ldots, f_R(\underline{x}))$. Let $K$ be a field of characteristic 0. For $\underline{\nu} \in K^R$, let $V_K(\underline{\nu}) = V(\underline{\nu})$ be the affine variety given by

$$V(\underline{\nu}) : \underline{f}(\underline{x}) = \underline{\nu}. \tag{2.2}$$

We write $V$ for $V(\underline{0})$. Also, denote by $\tilde{\underline{f}}$ the top degree part of $\underline{f}$ (so that $\tilde{\underline{f}}$ is homogeneous of degree $d$ and $\underline{f} - \tilde{\underline{f}}$ is a polynomial of degree at most $d-1$) and for $\underline{\nu} \in K^R$ denote by $\widetilde{V}(\underline{\nu})$ the variety given by

$$\widetilde{V} : \tilde{\underline{f}}(\underline{x}) = \underline{\nu}.$$

Again, we write $\widetilde{V}$ for $\widetilde{V}(\underline{0})$. We consider $\widetilde{V}$ as a projective variety.

**Definition 2.3.** We call a point $\underline{y}$ on an affine or projective variety $W$ given by polynomials $g_1, \ldots, g_R$ *singular* if the Jacobian matrix at $\underline{y}$ given by

$$\left(\frac{\partial g_i}{\partial x_j}(\underline{y})\right)_{i,j}, \quad 1 \leq i \leq R,\ 1 \leq j \leq n$$

does not have full rank, i.e. the rank is strictly smaller than the codimension of $W$.

Note that this definition depends on the choice of the polynomials: the polynomials $g(\underline{x}) = \prod_{i=1}^n x_i$ and $g'(\underline{x}) = \prod_{i=1}^n x_i^2$ have the same zero set, but the point $(0, 1, \ldots, 1)$ is not singular on the projective variety associated to $g$, but is singular on the projective variety associated to $g'$. Later, we see that it is natural to assume that $\widetilde{V}$ is a complete intersection, that is the ideal

$(\tilde{f}_1, \ldots, \tilde{f}_R)$ cannot be generated by fewer than $R$ elements. Under this assumption, a point $\underline{y}$ on $\widetilde{V}$ is singular if the rank of the Jacobian matrix at $\underline{y}$ is lower than $R$.

Consider the union of the loci of singularities of the $\widetilde{V}(\underline{\nu})$ over $\mathbb{C}$. In case $R = 1$ and $f_1$ is homogeneous, this union equals the locus of singularities of $\widetilde{V}$, but in general this union may be infinite. Aleksandrov and Moroz [AM02], filling a gap in the work of Birch, showed that this locus consists of all the points $\underline{x} \in \mathbb{C}^n$ for which

$$\operatorname{rk}\left(\frac{\partial \tilde{f}_i}{\partial x_j}(\underline{x})\right)_{i,j} < R$$

under the assumption that $\widetilde{V}$ is a complete intersection.

**Definition 2.4.** The *Birch singular locus* $\widetilde{V}^*$ is defined as the affine variety consisting of all points $\underline{x} \in \mathbb{C}^n$ for which

$$\operatorname{rk}\left(\frac{\partial \tilde{f}_i}{\partial x_j}(\underline{x})\right)_{i,j} < R.$$

Although Birch stated his work only for homogeneous polynomials, as Schmidt pointed out it works for polynomials as well (see [Sch85, Section 9]). The statement of Birch (for polynomials) is the following:

**Theorem 2.5** ([Bir62, Theorem 2])**.** *Suppose $\widetilde{V}$ has a non-singular real point and $V$ has a non-singular point in $\mathbb{Z}_p$ for every prime $p$. Suppose further that*

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

*Then $V$ has infinitely many integer points.*

*Remark.* Suppose that $\widetilde{V}$ is not a complete intersection. Then $\operatorname{codim} \widetilde{V} \leq R - 1$. By the definition of the Birch singular locus $\dim \widetilde{V}^* = n$, as the rank of the Jacobian matrix is at most the codimension of $\widetilde{V}$. Hence, the theorem does not hold in this case. Therefore, we can assume without loss of generality that the $\widetilde{V}$ is a complete intersection, which is normally assumed when using the circle method. Intuitively, this is because applying the circle method to some variety requires the dimension of this variety to be large relative to the degree of the corresponding polynomials, whereas varieties which are not complete intersections tend to have dimensions which are comparable with or smaller than those degrees. In [BHB17, p. 365-366] this intuition is made precise assuming conjectures on when projective varieties are complete intersections.

*Remark.* This general theorem is not applicable to Waring's problem for even $k$. Namely,

$$\widetilde{V} : x_1^k + \ldots + x_R^k = 0$$

has no non-singular real points. However, the conclusion of the theorem is also false in this case, as we know that there only finitely many integer points on

$$V : x_1^k + \ldots + x_R^k - N = 0$$

because we have to take $|x_i| \leq \sqrt[k]{N}$.

In fact, Birch's theorem also has a quantitative version, giving asymptotics for the number of integer points satisfying $|x_i| \leq P$ for all $i$ and $P \in \mathbb{R}$. This quantitative version has many advantages over Theorem 2.5, including that we can apply it to Waring's problem. We derive these asymptotics in the course of the work of this thesis, using the same ideas as Birch. We even do a little more: we make precise how this quantitative version depends on the coefficients of the polynomials. In the next sections we explain why it is interesting to obtain such a result.

### 2.2.2 Strong approximation

For now, assume that we are in the case of Theorem 2.5. Then it is natural to ask what we can deduce about these infinitely many integer points. There is no hope of deducing a general formula for these integer points, but we can investigate their distribution. We are mainly interested in integer points 'close' to a given real point on our variety $V(\underline{\nu})$. To be more precise:

(1) Is there an upper bound on the smallest integer point, i.e. an upper bound on $\min_{\underline{x}\in V(\underline{\nu})\cap\mathbb{Z}^n} |\underline{x}|$?

(2) More generally: given $\underline{x}_0 \in V_{\mathbb{R}}$ (or $\underline{x}_0 \in \mathbb{R}^n$) is there an upper bound on the distance to the closest integer point, i.e. an upper bound on $\min_{\underline{x}\in V(\underline{\nu})\cap\mathbb{Z}^n} |\underline{x} - \underline{x}_0|$?

(3) Even more generally: given $\underline{x}_0 \in V_{\mathbb{R}}$ (or $\underline{x}_0 \in \mathbb{R}^n$) is there an upper bound on the distance to the closest integer point satisfying a finite number of given modulo conditions, i.e. an upper bound on

$$\min_{\substack{\underline{x}\in V(\underline{\nu})\cap\mathbb{Z}^n \\ x_i\equiv a_i \bmod m_i}} |\underline{x} - \underline{x}_0|$$

for given $\underline{a}, \underline{m} \in \mathbb{N}^n$?

The last question can naturally be formulated in a more general framework. We now introduce this framework and the notion of quantitative strong approximation. Readers who are not familiar with the language of algebraic geometry can skip the rest of this section without problems, as these words are meant to view our question from a different perspective but are not needed to understand the rest of this thesis. Let $\mathcal{V}_{\mathbb{Z}}(\underline{\nu})$ be an integral model of $V(\underline{\nu})$. Let $\mathbb{A}$ be the adele ring of $\mathbb{Q}$ and $\mathbb{A}^{\Sigma}$ the adele ring of $\mathbb{Q}$ outside the set of places $\Sigma$, that is

$$\mathbb{A}^{\Sigma} = \prod_{v\notin\Sigma}{}'(\mathbb{Q}_v : \mathbb{Z}_v).$$

We are mostly interested in $\Sigma = \{\infty\}$ and in this case we write $\mathbb{A}^{\Sigma} = \mathbb{A}^{\infty}$. Let $\mathcal{V}_{\mathbb{A}^{\Sigma}}(\underline{\nu})$ be the base change of $\mathcal{V}_{\mathbb{Z}}(\underline{\nu})$ to $\mathbb{A}^{\Sigma}$. A basis for the opens of $V_{\mathbb{A}}(\underline{\nu})$ is given by subsets of $V_{\mathbb{A}}(\underline{\nu})$ for which the elements satisfy a finite number of given modulo conditions and lie in a given real open.

**Definition 2.6.** We say that $V(\underline{\nu})$ satisfies *strong approximation outside a set* $\Sigma$ of places if the image of the diagonal map

$$\mathcal{V}_{\mathbb{Z}}(\underline{\nu}) \to \mathcal{V}_{\mathbb{A}^{\Sigma}}(\underline{\nu})$$

is dense.

Note that if $\mathcal{V}_{\mathbb{A}^{\Sigma}}(\underline{\nu}) \neq \emptyset$ the notion of strong approximation outside $\Sigma$ implies the Hasse principle: if $\mathcal{V}_{\mathbb{Z}}(\underline{\nu})$ is dense in $\mathcal{V}_{\mathbb{A}^{\Sigma}}(\underline{\nu})$, it is in particular non-empty.

Birch's theorem implies that if $V_{\mathbb{A}}$ contains a non-singular point, then $\mathcal{V}_{\mathbb{Z}} = \mathcal{V}_{\mathbb{Z}} \cap V_{\mathbb{A}^{\infty}} \neq \emptyset$. One can even show that for all opens $U$ of $V_{\mathbb{A}^{\infty}}$, it holds that $\mathcal{V}_{\mathbb{Z}} \cap U \neq \emptyset$. Observe that this implies that for $V$ as in Birch's theorem strong approximation outside $\infty$ holds. We prove a quantitative version of this statement in Theorem 2.47, which follows directly from our main theorem. This theorem provides a quantitative answer in case $\underline{x}_0 = \underline{0}$ to the following reformulation of question (3):

(3') Let $\underline{x}_0 \in \mathbb{R}^n$ and $U$ open in $\mathcal{V}_{\mathbb{A}^{\infty}}(\underline{\nu})$. Is there an upper bound on $\min_{\underline{x}\in\mathcal{V}(\underline{\nu})\cap U} |\underline{x} - \underline{x}_0|$?

### 2.2.3 Results

We are mainly interested in how the answer to the above questions depends on the polynomials $\underline{f}$, given $n, d$ and $R$. Letting $C$ and $\tilde{C}$ be the (real) maximum of the absolute value of coefficients of $\underline{f}$ respectively $\underline{\tilde{f}}$, we answer the question in terms of $C$ and $\tilde{C}$. In order to do so, we make the work of Birch quantitative (in terms of $C$ and $\tilde{C}$). We use Birch's assumption on the number of variables throughout this work. That is, we let

$$K = \frac{n - \dim \widetilde{V}^*}{2^{d-1}} \tag{2.3}$$

and assume

$$K > R(R+1)(d-1). \tag{2.4}$$

Our main theorem is the following, which is proven in Section 2.5.3:

**Theorem 2.7.** *Let $f_i \in \mathbb{Z}[\underline{x}] = \mathbb{Z}[x_1, \dots, x_n]$ for $i = 1, \dots, R$ be polynomials of degree $d$ so that $K - R(R+1)(d-1) > 0$, $\underline{f}$ has zeros over $\mathbb{Z}_p$ for all primes $p$ and $\underline{\tilde{f}}$ has a real zero. Assume that the corresponding affine respectively projective varieties $V$ and $\widetilde{V}$ are smooth. Then there exists an $\underline{x} \in \mathbb{Z}^n$, polynomially bounded by the $C$ and $\tilde{C}$, such that $f(\underline{x}) = \underline{0}$, in fact*

$$|\underline{x}| \ll (C^3\tilde{C}^2)^{4n^3R(Rd)^n \cdot \frac{K+R(R+1)(d-1)}{K-R(R+1)(d-1)}}.$$

The only known upper bounds on the smallest non-trivial zero of forms in many variables are for forms of small degree. Let $\Lambda(f)$ be the smallest non-trivial integer zero of a form $f \in \mathbb{Z}[x_1, \dots, x_n]$ with coefficients bounded in absolute value by $C$. In [BDE12], the authors provide an overview of the known results if the degree $d$ equals $1, 2$ or $3$ and give improvements for the case $d = 3$ and $n \geq 17$. They show that $\Lambda(f) \leq cC^{360000}$ for some absolute constant $c$ provided $n \geq 17$, whereas by a result due to Pitman [Pit68] one has for any $\varepsilon > 0$ and sufficiently large $n$ that $\Lambda(f) \leq c_{n,\varepsilon} C^{\frac{25}{6}+\varepsilon}$, for some constant $c_{n,\varepsilon}$. In case the hypersurface corresponding to $f$ has at most isolated ordinary singularities, they provide visibly better bounds, e.g. for $n = 17$ they find $\Lambda(f) \leq cC^{1071}$. Above theorem in case $R = 1, d = 3, n = 17$ and $\dim V^* = 0$ yields

$$\Lambda(f) \leq (C^3\tilde{C}^2)^{83749461948108}.$$

So, our bounds are far worse than the known bounds and in no sense believed to be optimal. However, we provided an upper bound in many cases where it was not shown that such an upper bound exists. In Section 2.5.3 we provide a slightly better upper bound in case the polynomials are homogeneous. We also provide an upper bound on the smallest integer zero satisfying certain modulo conditions.

## 2.3 Setup

### 2.3.1 Notation

For a point $\underline{\alpha} \in \mathbb{R}^m$ we write $\underline{\alpha} = (\alpha_1, \ldots, \alpha_m)$ with $\alpha_i \in \mathbb{R}$ and introduce the supremum norm

$$|\underline{\alpha}| = \max(|\alpha_1|, \ldots, |\alpha_m|).$$

Here, $|\cdot|$ on the right-hand side is the usual absolute value, which is the real norm $|\cdot|_\infty$, but also in the $p$-adic case we write

$$|\underline{\alpha}|_p = \max(|\alpha_1|_p, \ldots, |\alpha_m|_p)$$

for $\underline{\alpha} \in \mathbb{Q}_p^m$. We denote by $\|\beta\|$ for $\beta \in \mathbb{R}$ the distance of $\beta$ to the nearest integer, i.e.

$$\|\beta\| = \min_{i \in \mathbb{Z}} |i - \beta|$$

and for a point $\underline{\alpha} \in \mathbb{R}^m$ we write

$$\|\underline{\alpha}\| = \max(\|\alpha_1\|, \ldots, \|\alpha_m\|).$$

If $\underline{a} \in \mathbb{Z}^m$ and $q \in \mathbb{Z}$, then we abbreviate $\gcd(a_1, \ldots, a_m, q)$ by $(\underline{a}, q)$. For $x \in \mathbb{R}$ we abbreviate $e^{2\pi i x}$ by $e(x)$. We do not consider 0 to be a natural number, so $\mathbb{N}$ denotes the set of positive integers.

For functions $f, g$ defined on a subset of the real numbers we use Vinogradov's notation $f \ll g$ to mean $f = O(g)$, that is there exists an $M > 0$ and $x_0 \in \mathbb{R}$ such that for all $x > x_0$ one has

$$|f(x)| \leq Mg(x). \tag{2.5}$$

Note that the statements $f \ll g$ and $|f| \ll g$ are the same. We call $M$ in (2.5) the implied constant of $f \ll g$. If we say that the implied constant only depends on variables $a, b, c$, etc. it is understood that also $x_0$ only depends on $a, b, c$, etc. Sometimes we indicate with a subscript where the implied constant depends on. For example, we use $\ll_R$ if $M$ and $x_0$ only depend on $R$ (so not on $n$ or $d$, etc.) and $\ll_1$ if $M$ and $x_0$ do not depend on one of the parameters $n, d$, etc. Without an indication the implied constant may depend on $n, R$ and $d$, but not on $C$ or $\tilde{C}$. In almost all cases we use Vinogradov's notation we consider the left- and right-hand side as functions of a variable $P$.

Denote by $\mathcal{E}$ the box $[-1, 1]^n$. Let $\mathcal{B}$ be an $n$-dimensional box contained in $\mathcal{E}$ of side-length at most 1, i.e. there are $a_j, b_j \in \mathbb{R}$ with $-1 \leq a_j < b_j \leq 1$ and $0 < b_j - a_j < 1$ such that $\mathcal{B}$ is given by $\prod_{j=1}^n [a_j, b_j]$.

### 2.3.2 Counting zeros using exponential sums

We assume throughout that $f_1, \ldots, f_R \in \mathbb{Z}[x_1, \ldots, x_n] = \mathbb{Z}[\underline{x}]$ are given with associated variety $V(\underline{\nu})$ as in Section 2.2.1. We also use the highest degree objects $\tilde{f}, \widetilde{V}, \widetilde{V}^*$ from this section.

For $\underline{\alpha} \in [0, 1)^R$, write

$$S(\underline{\alpha}) = \sum_{\underline{x} \in P\mathcal{B} \cap \mathbb{Z}^n} e(\underline{\alpha} \cdot \underline{f}(\underline{x})),$$
$$S(\underline{\alpha}, \underline{\nu}) = S(\underline{\alpha}) e(-\underline{\alpha} \cdot \underline{\nu}).$$

Denote by $M(P, \underline{\nu})$ the number of integer points on $V(\underline{\nu})$ in a box $P\mathcal{B}$. We can use the above notation to write $M(P, \underline{\nu})$ as an integral of exponential sums:

**Lemma 2.8.** *We have*

$$M(P,\underline{\nu}) = \int_{\underline{\alpha}\in[0,1]^R} S(\underline{\alpha},\underline{\nu})\ \mathrm{d}\underline{\alpha}. \tag{2.6}$$

*Proof.* If $\underline{x} \in P\mathcal{B} \cap \mathbb{Z}^n$ is such that $\underline{f}(\underline{x}) = \underline{\nu}$, then

$$\int_{[0,1]^R} S(\underline{\alpha},\underline{\nu})\ \mathrm{d}\underline{\alpha} = \int_{[0,1]^R} e\left(\underline{\alpha}\cdot\left(\underline{f}(\underline{x})-\underline{\nu}\right)\right)\ \mathrm{d}\underline{\alpha} = \int_{[0,1]^R} 1\ \mathrm{d}\underline{\alpha} = 1.$$

Conversely, if $f_i(\underline{x}) \neq \nu_i$, then

$$\int_0^1 e(\alpha_i(f_i(\underline{x})-\nu_i))\ \mathrm{d}\alpha_i = 0$$

as the exponential function has no poles. Hence, if $f(\underline{x}) \neq \underline{\nu}$, we have

$$\int_{[0,1]^R} S(\underline{\alpha},\underline{\nu})\ \mathrm{d}\underline{\alpha} = \int_{[0,1]^R} e\left(\underline{\alpha}\cdot\left(\underline{f}(\underline{x})-\underline{\nu}\right)\right)\ \mathrm{d}\underline{\alpha} = 0. \qquad \square$$

In the next chapter we provide estimates for (integrals of) exponential sums, so that in the end we can find estimates for $M(P,\underline{\nu})$.

A discrete version of above lemma is true as well. For $\underline{a} \in \mathbb{Z}^R$ and $q \in \mathbb{Z}$ such that $(\underline{a}, q) = 1$ and $1 \leq a_i \leq q$, let

$$S_{\underline{a},q} = \sum_{\underline{x} \bmod q} e(\underline{a}\cdot\underline{f}(\underline{x})/q), \tag{2.7}$$

$$S_{\underline{a},q}(\underline{\nu}) = S_{\underline{a},q}e(-\underline{a}\cdot\underline{\nu}/q). \tag{2.8}$$

Here, the summation is over $n$ complete sets of residues modulo $q$, that is over a complete set of residues modulo $q$ for every vector component of $\underline{x}$. Observe that $f_i(\underline{x}+\underline{k}q) \equiv f_i(\underline{x}) \mod q$ for all $i = 1,\ldots,n$ and $\underline{k} \in \mathbb{Z}^n$. Therefore, the above summation is well-defined.

**Lemma 2.9.** *The number of points $\underline{x} \in (\mathbb{Z}/p^N\mathbb{Z})^n$ satisfying $\underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N$ equals*

$$p^{-RN} \sum_{\underline{a} \bmod p^N} S_{\underline{a},p^N}(\underline{\nu}).$$

*Proof.* The idea of this proof is the same as the proof of Lemma 2.8. For $\underline{x} \in (\mathbb{Z}/p^N\mathbb{Z})^n$ such that $f(\underline{x}) = \underline{\nu}$ one has

$$\sum_{\underline{a} \bmod p^N} S_{\underline{a},p^N}(\underline{\nu}) = \sum_{\underline{a} \bmod p^N} e(\underline{a}\cdot(\underline{f}(\underline{x})-\underline{\nu})/p^N) = \sum_{\underline{a} \bmod p^N} 1 = p^{RN}.$$

Conversely, if $f_i(\underline{x}) \neq \nu_i$, then

$$\sum_{a_i \bmod p^N} e(a_i\cdot(f_i(\underline{x})-\nu_i)/p^N) = 0,$$

as this sum is over all $p^N$-th roots of unity. Hence, for $\underline{f}(\underline{x}) \neq \underline{\nu}$, we have

$$\sum_{\underline{a} \bmod p^N} S_{\underline{a},p^N}(\underline{\nu}) = \sum_{\underline{a} \bmod p^N} e(\underline{a}\cdot(\underline{f}(\underline{x})-\underline{\nu})/p^N) = 0.$$

Therefore,

$$p^{-RN} \sum_{\underline{x} \bmod p^N} \sum_{\underline{a} \bmod p^N} e(\underline{a}\cdot(\underline{f}(\underline{x})-\underline{\nu})/p^N) = p^{-RN} \sum_{\underline{a} \bmod p^N} S_{\underline{a},p^N}(\underline{\nu})$$

equals the number of points satisfying $\underline{f}(\underline{x}) = \underline{\nu} \mod p^N$. $\square$

We re-encounter the sum in this lemma later, after introducing the singular series.

## 2.4 Quantitative asymptotics

In this chapter we deduce asymptotics for the number of integer points on $V$ within a box $P\mathcal{B}$ for $P \to \infty$, which are quantitative in terms of the maximal coefficients $C$ and $\widetilde{C}$. We follow the work of Birch [Bir62], who found such asymptotics. The dependence of this asymptotics on $C$ and $\widetilde{C}$ is my own contribution.

### 2.4.1 Estimates of exponential sums

We obtain estimates for exponential sums of the form

$$S(\underline{\alpha}) = \sum_{x \in P\mathcal{B} \cap \mathbb{Z}^R} e(\underline{\alpha} \cdot \underline{f}(\underline{x})).$$

We use these estimates later to approximate $S(\underline{\alpha})$ in (2.6). These estimates depend on $\alpha_1, \dots, \alpha_R$ not being too well approximable by rational numbers with small denominators.

For each $i = 1, \dots, R$ write

$$\tilde{f}_i(x) = \sum_{\underline{j}=\underline{1}}^{\underline{n}} f^{(i)}_{j_0,\dots,j_{d-1}} x_{j_0} \cdots x_{j_{d-1}} \tag{2.9}$$

where the coefficients $f^{(i)}_{j_0,\dots,j_{d-1}}$ are symmetric in the suffixes $\underline{j}$ and the sum is over all $j_0, \dots, j_{d-1}$ from 1 to $n$. Note that by this condition these coefficients are not necessarily integers, but may have a denominator dividing $d!$.

**Lemma 2.10** ([Bir62, Lemma 2.1])**.** *We have*

$$|S(\underline{\alpha})|^{2^{d-1}} \ll_{n,d} P^{(2^{d-1}-d)n} \sum_{\underline{x}^{(1)},\dots,\underline{x}^{(d-1)} \in P\mathcal{E}} \left( \prod_{J=1}^{n} \min[P, \|\Phi_J(\underline{\alpha}; \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}\|^{-1}] \right), \tag{2.10}$$

*where*

$$\Phi_J\big(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}\big) = d! \sum_{i=1}^{R} \alpha_i \sum_{\underline{j}} f^{(i)}_{J,j_1,\dots,j_{d-1}} x^{(1)}_{j_1} \cdots x^{(d-1)}_{j_{d-1}}. \tag{2.11}$$

*Here,* $\sum_{\underline{j}}$ *is taken over* $j_1, \dots, j_{d-1}$ *running independently from* 1 *to* $n$ *and the* $f^{(i)}_{J,j_1,\dots,j_{d-1}}$ *are defined by* (2.9).

*Remark.* Note that the right-hand side of (2.10) does only depend on $\underline{\tilde{f}}$, whereas the left-hand side depends on $\underline{f}$.

*Proof.* We use this same ideas as in the proof of [Dav59, Lemma 3·1]. For polynomials $g_1, \dots, g_R \in \mathbb{Z}[\underline{x}]$ of degree $k$ let

$$S_k(\underline{\alpha}, \underline{g}) = \sum_{\underline{x} \in \mathcal{Q}} e\left(\underline{\alpha} \cdot \underline{g}(\underline{x})\right),$$

where $\mathcal{Q}$ is a box of side length at most $P$ (for example $P\mathcal{B}$), which we omit in the notation $S_k(\underline{\alpha}, \underline{g})$ as in every line of the proof the notation $\mathcal{Q}$ may denote a different box of side length at most $P$. Then

$$|S_k(\underline{\alpha}, \underline{g})|^2 = \sum_{\underline{z} \in \mathcal{Q}} \sum_{\underline{z}' \in \mathcal{Q}} e(\underline{\alpha} \cdot (\underline{g}(\underline{z}') - \underline{g}(\underline{z}))).$$

Let $\underline{y} = \underline{z}' - \underline{z}$. As $\underline{z}, \underline{z}' \in \mathcal{Q}$ and $\mathcal{Q}$ has side length at most $P$, we find $|\underline{y}| \le P$. Denote by $\mathcal{R}(\underline{y})$ the common part of $\mathcal{Q}$ and $\mathcal{Q} - \underline{y}$. Observe that $\mathcal{R}(\underline{y})$ is again a box of side length at most $P$. Given $g \in \mathbb{Z}[\underline{x}]$, let $\Delta_{\underline{y}} g(\underline{z}) = g(\underline{z} + \underline{y}) - g(\underline{z})$. Then

$$|S_k(\underline{\alpha}, g)|^2 \le \sum_{|\underline{y}| \le P} \sum_{\underline{z} \in \mathcal{R}(\underline{y})} e(\underline{\alpha}(g(\underline{z} + \underline{y}) - g(\underline{z}))) \le \sum_{|\underline{y}| \le P} |S_{k-1}(\underline{\alpha}, \Delta_{\underline{y}} g)|. \tag{2.12}$$

Write $\Delta_{\underline{x}^{(1)}, \dots, \underline{x}^{(k)}}$ for $\Delta_{\underline{x}^{(1)}} \cdots \Delta_{\underline{x}^{(k)}}$. Using (2.12), the Cauchy-Schwarz inequality and (2.12) again, we find that

$$\begin{aligned} |S_k(\underline{\alpha}, g)|^4 &\le \Big| \sum_{|\underline{x}^{(1)}| \le P} 1 \cdot |S_{k-1}(\underline{\alpha}, \Delta_{\underline{x}^{(1)}} g)| \Big|^2 \\ &\ll_n P^n \sum_{|\underline{x}^{(1)}| \le P} |S_{k-1}(\underline{\alpha}, \Delta_{\underline{x}^{(1)}} g)|^2 \\ &\le P^n \sum_{|\underline{x}^{(1)}| \le P} \sum_{|\underline{x}^{(2)}| \le P} |S_{k-2}(\underline{\alpha}, \Delta_{\underline{x}^{(1)}, \underline{x}^{(2)}} g)|. \end{aligned}$$

Proceeding by induction, we obtain

$$|S_d(\underline{\alpha}, \underline{f})|^{2^{d-1}} \ll_{n,d} P^{(2^{d-1}-d)n} \sum_{|\underline{x}^{(1)}| \le P} \cdots \sum_{|\underline{x}^{(d-1)}| \le P} |S_1(\underline{\alpha}, \Delta_{\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}} \underline{f})|. \tag{2.13}$$

Despite our assumption that $d \ge 2$, the following equality holds also for $d = 1$. Namely, we show by induction on $d \ge 1$ that

$$\underline{\alpha} \cdot \Delta_{\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}} \underline{f}(\underline{z}) = d! \sum_i \alpha_i {\sum_{\underline{j}}}' f^{(i)}_{j_0, \dots, j_{d-1}} z_{j_0} x^{(1)}_{j_1} \cdots x^{(d-1)}_{j_{d-1}} + \varphi(\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}) \tag{2.14}$$

$$= \sum_{J=1}^{n} \Phi_J \left( \underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)} \right) z_J + \varphi(\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}). \tag{2.15}$$

where the primed sum is over all $j_1, \dots, j_d$ running independently from 1 to $n$ and $\varphi$ is an integer polynomial independent of $\underline{z}$. The dependence of $\varphi$ on $\underline{\alpha}$ is suppressed in the notation. For $d = 1$ it is clear that (2.14) holds. Assume that (2.14) holds for $d = k$ and that all polynomials $\underline{f}$ have degree $k + 1$. Observing that $\underline{f}(\underline{z} + \underline{x}^{(k)}) - \underline{f}(\underline{z})$ has degree $k$ in $\underline{z}$, one finds

$$\begin{aligned} \underline{\alpha} \cdot \Delta_{\underline{x}^{(1)}, \dots, \underline{x}^{(k)}} \underline{f}(\underline{z}) &= \underline{\alpha} \cdot \Delta_{\underline{x}^{(1)}, \dots, \underline{x}^{(k-1)}} \left( \underline{f}(\underline{z} + \underline{x}^{(k)}) - \underline{f}(\underline{z}) \right) \\ &= k! \sum_i \alpha_i {\sum_{\underline{j}}}' (k+1) f^{(i)}_{j_0, \dots, j_{d-1}} z_{j_0} x^{(1)}_{j_1} \cdots x^{(k)}_{j_k} + \varphi(\underline{x}^{(1)}, \dots, \underline{x}^{(k)}), \end{aligned}$$

where again the primed sum is over all $j_1, \dots, j_d$ running independently from 1 to $n$ and $\varphi$ is an integer polynomial independent of $\underline{z}$. This proves equality (2.14).

For $\lambda \in \mathbb{R}$ we have

$$\left| \sum_{j=1}^{n} e(\lambda j) \right| \le \left| \frac{1 - e((n+1)\lambda)}{1 - e(\lambda)} \right| \le \frac{2}{|1 - e(\lambda)|} = \frac{1}{|\sin(\pi\lambda)|} \ll_1 \frac{1}{\|\lambda\|}.$$

Therefore,

$$\begin{aligned}|S_1(\underline{\alpha}, \Delta_{\underline{x}^{(1)},\dots,\underline{x}^{(d-1)}}\underline{f})| &= \Big|\sum_{\underline{z}\in\mathcal{Q}} e\left(\sum_{J=1}^{n}\Phi_J\left(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-1)}\right)z_J + \varphi(\underline{x}^{(1)},\dots,\underline{x}^{(d-1)})\right)\Big| \\ &= \Big|\sum_{\underline{z}\in\mathcal{Q}} e\left(\sum_{J=1}^{n}\Phi_J\left(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-1)}\right)z_J\right)\Big| \\ &\ll_1 \prod_J \min(P, \|\Phi_J\|^{-1}).\end{aligned}$$

Together with (2.13) this implies (2.10), as required. □

Denote by $N(\underline{\alpha})$ the number of distinct $(d-1)$-tuples of integer points $\underline{x}^{(1)},\dots,\underline{x}^{(d-1)}$ which satisfy

$$\begin{aligned}&|\underline{x}^{(i)}| < P && \text{for } i = 1,\dots,d-1,\\ &\|\Phi_J(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-1)})\| < P^{-1} && \text{for } J = 1,\dots,n.\end{aligned}$$

**Lemma 2.11** (Reformulation of [Bir62, Lemma 2.2])**.** *We have*

$$|S(\underline{\alpha})|^{2^{d-1}} \ll_{n,d} P^{(2^{d-1}-d+1)n}(\log P)^n N(\underline{\alpha}). \tag{2.16}$$

*Proof.* We use this same ideas as in the proof of [Dav59, Lemma 3·2]. For any $(d-2)$-tuple of integer points $\underline{x}^{(1)},\dots,\underline{x}^{(d-2)}$ and $\underline{\alpha}$ an $R$-tuple of real numbers, denote by $N(\underline{\alpha};\underline{x}^{(1)},\dots,\underline{x}^{(d-2)})$ the number of distinct points $\underline{y}$ which satisfy

$$|\underline{y}| < P, \quad \|\Phi_J(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-2)},\underline{y})\| < P^{-1}$$

for $J = 1,\dots,n$. Let $\mathrm{frac}(a)$ denote the fractional part of a real number $a$ and $\mathfrak{f}_J(\underline{y}) = \mathrm{frac}(\Phi_J(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-2)},\underline{y}))$. Then, for any integers $r_1,\dots,r_n$ with $0 \le r_j < P$, we show that there are at most $N(\underline{\alpha};\underline{x}^{(1)},\dots,\underline{x}^{(d-2)})$ integer points $\underline{y} \in P\mathcal{B}\cap\mathbb{Z}^R$ which satisfy

$$\frac{r_J}{P} \le \mathfrak{f}_J(\underline{y}) < \frac{r_J+1}{P} \tag{2.17}$$

for all $J = 1,\dots,n$. Suppose that $\underline{y}'$ and $\underline{y}$ are such points. Then

$$\|\Phi_J(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-2)},\underline{y}'-\underline{y})\| \le |\mathfrak{f}_J(\underline{y}') - \mathfrak{f}_J(\underline{y})| < P^{-1}$$

by linearity of $\Phi_J$. Also, $|\underline{y}-\underline{y}'| < P$. Hence there are at most $N(\underline{\alpha};\underline{x}^{(1)},\dots,\underline{x}^{(d-2)})$ possibilities for $\underline{y}$.

Observe that for $a \in \mathbb{R}$ we have

$$\|a\| = \min(\mathrm{frac}(a), 1-\mathrm{frac}(a)).$$

Therefore, given $\underline{r} \in \mathbb{Z}^n$ with $0 \le r_j < P$ for all $j = 1,\dots,n$ satisfying (2.17) it follows that

$$\|\Phi_J(\underline{\alpha},\underline{x}^{(1)},\dots,\underline{x}^{(d-2)},\underline{y})\|^{-1} = \min\left(\mathfrak{f}_J(\underline{y}), 1-\mathfrak{f}_J(\underline{y})\right)^{-1} \le \max\left(\frac{P}{r_J}, \frac{P}{P-r_J-1}\right)$$

for all $J = 1,\dots,n$ (Davenport makes a minor mistake in the above equation by writing a minimum instead of a maximum).

Now, split the summation over $\underline{x}^{(d-1)} = \underline{y}$ in (2.10) into $2^n$ pairs, for each of which $y_1, \dots, y_n$ run through intervals of length $P$. It follows that

$$\sum_{\underline{y} \in P\mathcal{E}} \left( \prod_{J=1}^{n} \min(P, \|\Phi_J(\underline{\alpha}; \underline{x}^{(1)}, \dots, \underline{x}^{(d-2)}, \underline{y}\|^{-1}) \right) \tag{2.18}$$

$$\ll_n N(\underline{\alpha}; \underline{x}^{(1)}, \dots, \underline{x}^{(d-2)}) \sum_{\underline{r} \in \{0,\dots,P-1\}^n} \prod_{j=1}^{n} \min\left(P, \max\left(\frac{P}{r_j}, \frac{P}{P - r_j - 1}\right)\right) \tag{2.19}$$

$$\ll N(\underline{\alpha}; \underline{x}^{(1)}, \dots, \underline{x}^{(d-2)})(P \log P)^n, \tag{2.20}$$

where the last step followed as

$$\sum_{r=0}^{P-1} \min\left(P, \max\left(\frac{P}{r}, \frac{P}{P-r-1}\right)\right) = 2P + P \sum_{r=1}^{P-2} \max\left(\frac{1}{r}, \frac{1}{P-r-1}\right) \ll_1 P \log P.$$

As

$$N(\underline{\alpha}) = \sum_{\underline{x}^{(1)}, \dots, \underline{x}^{(d-2)} \in P\mathcal{E}} N(\underline{\alpha}; \underline{x}^{(1)}, \dots, \underline{x}^{(d-2)}),$$

the lemma follows from Lemma 2.10 combined with (2.18)-(2.20) summed over $\underline{x}^{(1)}, \dots, \underline{x}^{(d-2)} \in P\mathcal{E}$. □

Let $Z \in \mathbb{R}$ be a parameter to be determined later. Denote with $N(\underline{\alpha}, k, Z)$ the number of integer $(d-1)$-tuples $\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}$ satisfying

$$\begin{aligned}
&|\underline{x}^{(i)}| < ZP && \text{for all } 1 \le i \le k,\\
&|\underline{x}^{(i)}| < P && \text{for all } k < i \le d-1,\\
&\|\Phi_J(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)})\| < Z^k P^{-1} && \text{for all } J \le n.
\end{aligned}$$

**Lemma 2.12** (Reformulation of [Bir62, Lemma 2.4])**.** *For all $k = 0, \dots, d-1$ and $0 < Z < 1$ we have*

$$N(\underline{\alpha}) \ll_n Z^{-kn} N(\underline{\alpha}, k, Z).$$

*Proof.* We proceed by induction on $k$. For $k = 0$ we have $N(\underline{\alpha}, k, Z) = N(\underline{\alpha})$ and the statement is true. Assuming that for $k < d-1$ the statement holds, we apply Lemma 3.4 in [Dav59] to deduce the result for $k+1$. Let $\underline{u} = \underline{x}^{(k+1)}$ and $L(\underline{u}) = \Phi_J(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)})$, which is a symmetric linear form in $\underline{x}^{(k+1)}$. Take

$$A = PZ^{-k/2}, \quad Z_1 = Z^{k/2+1}, \quad Z_2 = Z^{k/2}.$$

Then the lemma implies that

$$\#\{\underline{x}^{(k+1)} \in \mathbb{Z}^n : |\underline{x}^{(k+1)}| < P, \|\Phi_J(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)})\| < Z^{-k} P^{-1}\}$$

is less than a constant times (depending only on $n$)

$$Z^{-n} \#\{\underline{x}^{(k+1)} \in \mathbb{Z}^n : |\underline{x}^{(k+1)}| < ZP, \|\Phi_J(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)})\| < Z^{-k-1} P^{-1}\}.$$

Summing over $\underline{x}^{(1)}, \dots, \underline{x}^{(k)}$ with $|\underline{x}^{(i)}| < ZP$ and over $\underline{x}^{(k+2)}, \dots, \underline{x}^{(d-1)}$ with $|\underline{x}^{(i)}| < P$ we obtain

$$N(\underline{\alpha}, k, Z) \ll_n Z^{-n} N(\underline{\alpha}, k+1, Z).$$

We conclude that $N(\underline{\alpha}) \ll_n Z^{-kn} N(\underline{\alpha}, k, Z)$ for all $k = 0, \dots, d-1$. □

Combining Lemma 2.10 and Lemma 2.11 with $k = d-1$ we find that for all $\varepsilon > 0$ we have

$$|S(\underline{\alpha})|^{2^{d-1}} \ll_{n,d} \frac{P^{(2^{d-1}-d+1)n}(\log P)^n}{Z^{(d-1)n}} \#\mathcal{I} \ll_{n,d} \frac{P^{(2^{d-1}-d+1)n+\varepsilon}}{Z^{(d-1)n}} \#\mathcal{I},$$

where $\mathcal{I}$ is given by

$$\mathcal{I} = \left\{ (\underline{x}^{(1)}, \dots, \underline{x}^{(d-1)}) \in \mathbb{Z}^{(d-1)n} : \begin{array}{l} |\underline{x}^{(i)}| < ZP \\ \|\Phi_J(\underline{\alpha}, \underline{x}^{(1)}, \dots, \underline{x}^{(d-1)})\| < Z^{d-1}P^{-1} \end{array} \right\}.$$

The following lemma distinguishes between $\underline{\alpha}$ for which $|S(\underline{\alpha})|$ is small and $\underline{\alpha}$ which are well approximable by fractions with small numerators. This makes [Bir62, Lemma 4.3] quantitative in terms of $\widetilde{C}$.

**Lemma 2.13.** *Let $\varepsilon > 0$ and $0 < \theta < 1$. One of the following holds:*

(i) $|S(\underline{\alpha})| \ll_R P^{n-K\theta+\varepsilon}$, *where $K$ is defined by* (2.3);

(ii) there is a rational approximation $\underline{a}/q$ to $\underline{\alpha}$ with $\underline{a} \in \mathbb{Z}^R_{\geq 0}$ and $q \in \mathbb{N}$ satisfying

$$\begin{aligned} &(\underline{a}, q) = 1, \\ &|q\underline{\alpha} - \underline{a}| \leq \widetilde{C}^{R-1} P^{-d+R(d-1)\theta}, \\ &1 \leq q \leq \widetilde{C}^R P^{R(d-1)\theta}. \end{aligned}$$

*Proof.* Letting $k = K\theta + \varepsilon$ and $Z = P^{\theta-1}$, we obtain

$$|S(\underline{\alpha})|^{2^{d-1}} \ll_{n,d} P^{(2^{d-1}-d+1)n+\varepsilon} P^{(d-1)n(\theta-1)} \#\mathcal{I}.$$

If (i) does not hold we have

$$\#\mathcal{I} \gg P^{n-k} P^{-(2^{d-1}-d+1)n-\varepsilon} P^{-(d-1)n(\theta-1)} = P^{n(d-1)\theta - 2^{d-1}k - \varepsilon}.$$

By definition of $k$ we have

$$n - 2^{d-1}k/\theta + \varepsilon = n - 2^{d-1}K + 2^{d-1}\varepsilon/\theta - \varepsilon > \dim \widetilde{V}^*.$$

Using (2.4), it follows by Lemma 2.5, 3.1 and 3.3 in [Bir62] that there exists a point in $\mathcal{I}$ with $\mathrm{rk}[\Psi_J^{(i)}] = R$, where

$$\Psi_J^{(i)}(x^{(1)}, \dots, x^{(d-1)}) = d! \sum_j{}' f^{(i)}_{J,j_1,\dots,j_{d-1}} x^{(1)}_{j_1} \cdots x^{(d-1)}_{j_{d-1}}.$$

Observe that

$$\Phi_J = \textstyle\sum_i \alpha_i \Psi_J^{(i)}$$

by (2.11) and write $\Phi_J = A_J + \delta_J$, where $A_J$ is integral and $|\delta_J| < P^{-d+(d-1)\theta}$ by definition of $\mathcal{I}$. As $\mathrm{rk}[\Psi_J^{(i)}] = R$ the matrix $[\Psi_J^{(i)}]$ has a non-vanishing $R \times R$ minor $Q$. Assume without loss of generality that this is the leading minor and write $q$ for the absolute value of its determinant. Then $q \in \mathbb{N}$. As $|\Psi_J^{(i)}| \leq \widetilde{C} P^{(d-1)\theta}$ we find that

$$q \ll_R \widetilde{C}^R P^{R(d-1)\theta}.$$

Let $a_1, \dots, a_R$ be the solutions of

$$\sum_{i=1}^{R} a_i \Psi_J^{(i)} = qA_J$$

for $J = 1, \dots, R$. Then, by Cramer's rule

$$a_j = \frac{\det(Q_k)}{q}$$

where $Q_k$ is the matrix obtained by replacing the $k$-th row of $Q$ by the vector $qA_J$. Hence, $q \mid \det(Q_k)$, so $a_1, \dots, a_R$ are integers. Also, $q\alpha_1 - a_1, \dots, q\alpha_R - a_R$ are the solutions of

$$\sum_{i=1}^{R} (q\alpha_i - a_i) \Psi_J^{(i)} = q\delta_J$$

for $J = 1, \dots, R$. Letting $Q^k$ be the matrix obtained by replacing the $k$-th row of $Q$ by the vector $q\delta_J$, we find by Cramer's rule that

$$|q\alpha_k - a_k| = \frac{|\det Q^k|}{q} \ll_R \frac{q|\delta_J| \widetilde{C}^{R-1} P^{(R-1)(d-1)\theta}}{q} < \widetilde{C}^{R-1} P^{-d+R(d-1)\theta}.$$

Finally, we can throw away a common factor of $a_1, \dots, a_R, q$ so that $(\underline{a}, q) = 1$. Hence, if (i) does not hold, then (ii) does hold. Note that by scaling $\theta$ and $k$ we can get all dependency on $R$ in the implied constant of (i). □

Recall that for $\underline{a} \in \mathbb{Z}^R$ and $q \in \mathbb{Z}$ such that $(\underline{a}, q) = 1$ and $1 \le a_i \le q$, we defined

$$S_{\underline{a},q} = \sum_{\underline{x} \bmod q} e(\underline{a} \cdot \underline{f}(\underline{x})/q), \quad S_{\underline{a},q}(\underline{\nu}) = S_{\underline{a},q} e(-\underline{a} \cdot \underline{\nu}/q).$$

We can use the previous lemma to find an upper bound on $|S_{\underline{a},q}|$, which makes [Bir62, Lemma 5.4] quantitative:

**Lemma 2.14.** *For every $\varepsilon > 0$ we have*

$$|S_{\underline{a},q}| \ll_{d,n,R,\dim \widetilde{V}^*} \widetilde{C}^{K/(d-1)} q^{n-K/R(d-1)+\varepsilon}.$$

*Proof.* Observe that $S_{\underline{a},q}$ is a particular case of $S(\underline{\alpha})$ with $P = q$, $\underline{\alpha} = \underline{a}/q$ and $\mathcal{B} : 0 \le x_j < 1$ for $j = 1, \dots, n$. In this case, the inequalities corresponding to (ii) of Lemma 2.13 are given by

$$\begin{aligned} &|q'a_i - a_i'q| \le \widetilde{C}^{R-1} q^{-(d-1)+R(d-1)\theta} \quad \text{for } i = 1, \dots, R, \\ &1 \le q' \le \widetilde{C}^R q^{R(d-1)\theta}. \end{aligned}$$

Now, take $\theta$ such that $R(d-1)\theta < 1 - \log_q(\widetilde{C}^R)$. This implies that

$$-(d-1) + R(d-1)\theta < -d + 2 - \log_q(\widetilde{C}^R) \le -\log_q(\widetilde{C}^{R-1}).$$

Then, the inequalities read

$$\begin{aligned} &|q'a_i - a_i'q| < 1 \quad \text{for } i = 1, \dots, R, \\ &1 \le q' < q. \end{aligned}$$

The first inequalitiy implies $q'a_i = a_i'q$. As $(\underline{a}, q) = 1 = (\underline{a}', q)$, it follows that $a_i = a_i'$ and $q = q'$. This contradicts $q' < q$, hence there are no solutions. Therefore, (ii) in Lemma 2.13 is not satisfied, hence (i) is satisfied. This implies that

$$\begin{aligned} S_{\underline{a},q} \ll_R q^{n-k} &< q^{n-K(1-\log_q(\widetilde{C}^R))/R(d-1)+\varepsilon} \\ &= \widetilde{C}^{K/(d-1)} q^{n-K/R(d-1)+\varepsilon}, \end{aligned}$$

as desired. □

*Remark.* Taking $R = 1, d = 3, \dim \widetilde{V}^* = 0$ we obtain the same bound as in [BDE12] for a so-called $\infty$-good form, namely

$$S_{\underline{a},q} \ll \widetilde{C}^{n/8} q^{7n/8+\varepsilon}.$$

### 2.4.2 Minor arcs

Given $\underline{a} \in \mathbb{Z}^R$, $q \in \mathbb{N}$ and $0 < \theta \leq 1$, define a *major arc* as

$$\mathfrak{M}_{\underline{a},q}(\theta) = \prod_{i=1}^{R} \left[ \frac{a_j}{q} - \frac{\widetilde{C}^{R-1} P^{-d+R(d-1)\theta}}{2q}, \frac{a_j}{q} + \frac{\widetilde{C}^{R-1} P^{-d+R(d-1)\theta}}{2q} \right].$$

Then, define *the major arcs* to be

$$\mathfrak{M}(\theta) = \bigcup_{1 \leq q \leq \widetilde{C}^R P^{R(d-1)\theta}} \bigcup_{\substack{1 \leq a_i \leq q \\ (\underline{a},q)=1}} \mathfrak{M}_{\underline{a},q}(\theta). \tag{2.21}$$

Modulo 1 we have that $\mathfrak{M}(\theta)$ consists of all $\underline{\alpha}$ satisfying (ii) in Lemma 2.13. Define the *minor arcs* by $\mathfrak{m} = [0,1] \backslash \mathfrak{M}$ modulo 1. We can find an upper bound on the volume of $\mathfrak{M}(\theta)$ as in [Bir62, Lemma 4.2]:

**Lemma 2.15.** *There exists an $\varepsilon > 0$ such that $\mathfrak{M}(\theta)$ has volume at most*

$$\widetilde{C}^{R^2} P^{-Rd+R(R+1)(d-1)\theta-\varepsilon}. \tag{2.22}$$

*Proof.* Each arc $\mathfrak{M}_{\underline{a},q}(\theta)$ has volume

$$\left( q^{-1} \widetilde{C}^{R-1} P^{-d+R(d-1)\theta} \right)^R.$$

As $\mathfrak{M}(\theta)$ is the (not necessarily disjoint) union of such boxes, an upper bound for the volume of $\mathfrak{M}(\theta)$ is given by

$$\sum_{1 \leq q \leq \widetilde{C}^R P^{R(d-1)\theta}} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a},q)=1}} \left( q^{-1} \widetilde{C}^{R-1} P^{-d+R(d-1)\theta} \right)^R.$$

As there are strictly less than $q^R$ choices for $\underline{a}$, we obtain (2.22). □

For $\theta$ small enough, the major arcs are disjoint as in [Bir62, Lemma 4.1]:

**Lemma 2.16.** *If $d > 2R(d-1)\theta + (2R-1)\log_P(\widetilde{C})$, then $\mathfrak{M}(\theta)$ is given as a disjoint union of $\mathfrak{M}_{\underline{a},q}(\theta)$ by* (2.21).

*Proof.* Suppose that $\alpha$ lies in the distinct sets $\mathfrak{M}_{\underline{b},q}(\theta)$ and $\mathfrak{M}_{\underline{b}',q'}(\theta)$. Then we have that

$$2|q\alpha_i - b_i| \le \widetilde{C}^{R-1}P^{-d+R(d-1)\theta}, \quad 2|q'\alpha_i - b_i'| \le \widetilde{C}^{R-1}P^{-d+R(d-1)\theta}, \quad 1 \le q, q' \le \widetilde{C}^R P^{R(d-1)\theta}.$$

Moreover, as the sets are distinct and $(\underline{b}, q) = 1 = (\underline{b}', q)$, it follows that there is an $i$ such that

$$b_i/q \ne b_i'/q'.$$

Then

$$1 \le |b_i'q - q'b_i| = |b_i'q - qq'\alpha_i + qq'\alpha_i - q'b_i| \le q|q'\alpha_i - b_i'| + q'|q\alpha_i - b_i| \le \widetilde{C}^{2R-1}P^{-d+2R(d-1)\theta}.$$

This contradicts $d > 2R(d-1)\theta + (2R-1)\log_P(\widetilde{C})$, which proves the lemma. □

Now, take major arcs $\mathfrak{M}(\theta_0)$, where $\eta, \theta_0, \delta$ are such that

$$\eta = R(d-1)\theta_0, \tag{2.23}$$

$$1 > \eta + R\log_P(\widetilde{C}), \tag{2.24}$$

$$\frac{K}{R(d-1)} - (R+1) > \delta\eta^{-1}. \tag{2.25}$$

Observe that assumption (2.25) is a quantative version of our main assumption (2.4). Note that (2.24) implies that the major arcs $\mathfrak{M}_{a,q}(\theta_0)$ are disjoint. In the end, we will choose $\eta, \delta$ satisfying (2.24) and (2.25).

If $\underline{\alpha}$ is not in $\mathfrak{M}(\theta_0)$ modulo 1, then by Lemma 2.13 we have that

$$S(\underline{\alpha}) \ll_1 P^{n-K\theta+\varepsilon}.$$

This estimate is stronger the larger $\theta$ is. Therefore, in order to show that $\int_{\mathfrak{m}} |S(\underline{\alpha}, \underline{\nu})| \, \mathrm{d}\underline{\alpha}$ is negligible, we use a sort of sliding scale. For most $\underline{\alpha}$, we can take $\theta$ large and have a strong estimate for $|S(\underline{\alpha})|$. When this estimate is invalid, we have to use a smaller value of $\theta$, but we have the compensation that this only happens for a set of $\alpha$ of small measure by the previous lemma. So, the worse the estimate for $|S(\underline{\alpha})|$, the smaller the set of $\underline{\alpha}$ for which it is necessary to use this estimate. Hence, we find the following generalisation of [Bir62, Lemma 4.4]:

**Lemma 2.17.**

$$\int_{\mathfrak{m}} |S(\underline{\alpha}, \underline{\nu})| \, \mathrm{d}\underline{\alpha} \ll \widetilde{C}^{R^2} P^{n-Rd-\delta}.$$

*Proof.* Observe $|S_{\underline{a},q}(\underline{\nu})| = |S_{\underline{a},q}|$ if $\underline{\nu}$ is real or $|S_{\underline{a},q}(\underline{\nu})| \ll_{\underline{\nu}} |S_{\underline{a},q}|$ if $\underline{\nu}$ is complex. Let $\varepsilon > 0$ be small. Now, define a sequence

$$\theta_T > \theta_{T-1} > \ldots > \theta_1 > \theta = \theta_0 > 0$$

such that

$$2d = (R+1)(d-1)\theta_T$$

and

$$\varepsilon\delta > R(R+1)(d-1)(\theta_{t+1} - \theta_t) \quad \text{for } 0 \le t \le T-1. \tag{2.26}$$

Then we can choose

$$T < \frac{\frac{2d}{(R+1)(d-1)}}{\frac{\varepsilon\delta}{R(R+1)(d-1)}} = \frac{2Rd}{\varepsilon\delta},$$

so $T \ll P^{\delta\varepsilon}$ (independent of $\widetilde{C}$).

By Lemma 2.13 we have that

$$\int_{\underline{\alpha}\notin\mathfrak{M}(\theta_T)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} \ll P^{n-K\theta_T+\varepsilon}.$$

As

$$-K\theta_T+\varepsilon = -\frac{K2d}{(R+1)(d-1)}+\varepsilon < -2Rd$$

by (2.4), we find

$$\int_{\underline{\alpha}\notin\mathfrak{M}(\theta_T)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} \ll P^{n-2Rd}.$$

By Lemma 2.15 and Lemma 2.13 we have

$$\begin{aligned}
\int_{\mathfrak{M}(\theta_{t+1})-\mathfrak{M}(\theta_t)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} &\ll \widetilde{C}^{R^2} P^{-Rd+R(R+1)(d-1)\theta_{t+1}} P^{n-K\theta_t-2\delta\varepsilon} \\
&\ll \widetilde{C}^{R^2} P^{n-Rd-(K-R(R+1)(d-1))\theta_t-\delta\varepsilon} && \text{(by (2.26))} \\
&\ll \widetilde{C}^{R^2} P^{n-Rd-\delta\theta_0^{-1}\theta_t-\delta\varepsilon} && \text{(by (2.25))} \\
&\ll \widetilde{C}^{R^2} P^{n-Rd-\delta-\delta\varepsilon}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\alpha\notin\mathfrak{M}(\theta_0)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} &= \int_{\alpha\notin\mathfrak{M}(\theta_T)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} + \sum_{t=0}^{T-1} \int_{\mathfrak{M}(\theta_{t+1})-\mathfrak{M}(\theta_t)} |S(\underline{\alpha},\underline{\nu})|\,\mathrm{d}\underline{\alpha} \\
&\ll P^{n-2Rd} + P^{\delta\varepsilon}\widetilde{C}^{R^2} P^{n-Rd-\delta-\delta\varepsilon} \\
&\ll \widetilde{C}^{R^2} P^{n-Rd-\delta}.
\end{aligned}$$

□

Combining (2.24) with Lemma 2.16 and Lemma 2.17, we have the following corollary:

**Corollary 2.18** (Lemma 4.5 in [Bir62])**.**

$$M(P,\underline{\nu}) = \sum_{1\le q\le \widetilde{C}^R P^\eta} \sum_{\substack{1\le a_i\le q\\(\underline{a},q)=1}} \int_{\mathfrak{M}_{\underline{a},q}(\theta_0)} S(\underline{\alpha},\underline{\nu})\,\mathrm{d}\alpha + O(\widetilde{C}^{R^2} P^{n-Rd-\delta}),$$

*where $O$ does not depend on $\widetilde{C}$.*

### 2.4.3 Approximating exponential sums by integrals

Write $\mathfrak{M}_{\underline{a},q}$ for $\mathfrak{M}_{\underline{a},q}(\theta_0)$. Let $\underline{\alpha}\in\mathfrak{M}_{\underline{a},q}$ and define $\underline{\beta} = \underline{\alpha} - \underline{a}/q$.

Letting $\underline{x} = \underline{z} + q\underline{y}$ we find that

$$\begin{aligned}
S(\underline{\alpha},\underline{\nu}) &= \sum_{\underline{z} \bmod q} \sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\alpha}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})) \\
&= \sum_{\underline{z} \bmod q} \sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{a}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})/q)\cdot e(\underline{\beta}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})) \\
&= \sum_{\underline{z} \bmod q} e(\underline{a}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})/q) \sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})),
\end{aligned}$$

using $\underline{f}(\underline{z}) \equiv \underline{f}(\underline{z}+q\underline{y}) \mod q$. We wish to replace the sum $\sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\underline{f}(\underline{z}+q\underline{y}))$ by the integral

$$\int_{\underline{z}+q\underline{\omega}\in P\mathcal{B}} e(\underline{\beta}\cdot\tilde{\underline{f}}(\underline{z}+q\underline{\omega}))\,\mathrm{d}\underline{\omega}.$$

We approximate the error term using the same ideas as in the Euler-Maclaurin formula. By this formula one has for some differentiable function $g:[0,m]\to\mathbb{C}$ that

$$\sum_{i=0}^{m} g(m) = \int_0^m g(x)\,\mathrm{d}x + \frac{g(0)+g(m)}{2} + R,$$

where $R \ll_1 \int_0^m |g'(x)|\,\mathrm{d}x$ (see, for example, [KC02, Chapter 25] for a proof). One can interpret $\frac{g(0)+g(m)}{2}$ as an error term coming from the boundaries and $R$ as an error term due to the variations of $g$ within $[0,m]$.

For a measurable subset $\mathcal{C}$ of $\mathcal{E}$ and $\underline{\gamma}\in\mathbb{R}^R$, we write

$$I(\mathcal{C},\underline{\gamma}) = \int_{\underline{\zeta}\in\mathcal{C}} e(\underline{\gamma}\cdot\tilde{\underline{f}}(\underline{\zeta}))\,\mathrm{d}\underline{\zeta}. \tag{2.27}$$

**Lemma 2.19.** *Given $\underline{z},\underline{\beta}\in\mathbb{R}^R$ and $q\in\mathbb{N}$, we have*

$$\sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\underline{f}(\underline{z}+q\underline{y})) = q^{-n}P^n I(\mathcal{B},P^d\underline{\beta}) + O\left(\left(C|P^d\underline{\beta}|+1\right)q^{1-n}P^{n-1}\right). \tag{2.28}$$

*Proof.* First observe that for $x\in\mathbb{R}$

$$|e(x)-1| = 2|\sin(\pi x)| \ll_1 |x|.$$

Therefore, for the system of polynomials $\underline{r}=\underline{f}-\tilde{\underline{f}}$ of degree at most $d-1$ we have

$$|e(\underline{\beta}\cdot\underline{r}(\underline{z}+q\underline{y}))-1| \ll_1 |\underline{\beta}||\underline{r}(\underline{z}+q\underline{y}))| \ll |\underline{\beta}|\cdot CP^{d-1}.$$

where we assumed that $\underline{z}+q\underline{y}\in P\mathcal{B}$. There are $O((P/q)^n)$ values of $\underline{y}$ in the sum, hence

$$\sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\underline{f}(\underline{z}+q\underline{y})) = \sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\tilde{\underline{f}}(\underline{z}+q\underline{y})) + O(|\underline{\beta}|\cdot Cq^{-n}P^{n+d-1}).$$

Next, we replace the sum in the right-hand side by the integral

$$\int_{q\underline{\omega}+\underline{z}\in P\mathcal{B}} e(\underline{\beta}\cdot\tilde{\underline{f}}(\underline{z}+q\underline{\omega}))\,\mathrm{d}\underline{\omega}. \tag{2.29}$$

The edges of the cube of summation and integration have length $P/q$. As $|e(\cdot)|\le 1$, in the replacement of the sum by the integral, we have an error of at most $\ll (P/q)^{n-1}$ coming from the boundaries. We also have to allow for variations of the integrand. The absolute value of the maximum of its derivative is $O(|\underline{\beta}|q\widetilde{C}P^{d-1})$, so the resulting error is obtained by multiplying with the volume of the region of integration: $(P/q)^n$. Hence, the total error in (2.28) is

$$\ll |\underline{\beta}|Cq^{-n}P^{n+d-1} + q^{1-n}P^{n-1} + |\underline{\beta}|\widetilde{C}q^{1-n}P^{n+d-1} \ll \left(C|P^d\underline{\beta}|+1\right)q^{1-n}P^{n-1}.$$

Applying the substitution $q\underline{\omega}+\underline{z}=P\underline{\zeta}$ to (2.29) we find that

$$\sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\underline{f}(\underline{z}+q\underline{y})) = q^{-n}P^n I(\mathcal{B},P^d\underline{\beta}) + O\left(\left(C|P^d\underline{\beta}|+1\right)q^{1-n}P^{n-1}\right)$$

as desired. □

**Corollary 2.20.** *Given $\underline{z} \in \mathbb{Z}^R$ and $\underline{\alpha} \in \mathfrak{M}_{\underline{a},q}$ so that $\underline{\beta} = \underline{\alpha} - \underline{a}/q$, we have*

$$\sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot\underline{f}(\underline{z}+q\underline{y})) = q^{-n}P^n I(\mathcal{B}, P^d\underline{\beta}) + O(C\widetilde{C}^{R-1}q^{-n}P^{n+\eta-1}). \tag{2.30}$$

*Proof.* As $\underline{\alpha} \in \mathfrak{M}_{\underline{a},q}$, we have

$$q \le \widetilde{C}^R P^{R(d-1)\theta_0} = \widetilde{C}^R P^{\eta},$$
$$|\underline{\beta}| \le \widetilde{C}^{R-1}q^{-1}P^{-d+R(d-1)\theta_0} = \widetilde{C}^{R-1}q^{-1}P^{-d+\eta}.$$

Applying these bounds to the error term in Lemma 2.19, we find

$$\left(C|P^d\underline{\beta}|+1\right)q^{1-n}P^{n-1} \ll C\widetilde{C}^{R-1}q^{-n}P^{n+\eta-1} + \widetilde{C}^R q^{-n}P^{n+\eta-1} \ll C\widetilde{C}^{R-1}q^{-n}P^{n+\eta-1},$$

as desired. □

We now have made [Bir62, Lemma 5.1] quantitative in terms of $C$ and $\widetilde{C}$:

**Corollary 2.21.** *Let $\underline{\alpha} = \underline{a}/q + \underline{\beta} \in \mathfrak{M}_{\underline{a},q}$. Then,*

$$S(\underline{\alpha},\underline{\nu}) = P^n q^{-n} S_{\underline{a},q}(\underline{\nu})\cdot I(\mathcal{B}; P^d\underline{\beta})\cdot e(-\underline{\beta}\cdot\underline{\nu}) + O(C\widetilde{C}^{R-1}P^{n+\eta-1}).$$

*Proof.* By Corollary 2.20 we have that

$$\begin{aligned}
S(\underline{\alpha},\underline{\nu}) &= \sum_{\underline{z} \bmod q} e(\underline{a}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})/q) \sum_{\underline{z}+q\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\beta}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})) \\
&= \sum_{\underline{z} \bmod q} e(\underline{a}\cdot(\underline{f}(\underline{z}+q\underline{y})-\underline{\nu})/q)\left(q^{-n}P^n I(\mathcal{B},P^d\underline{\beta}) + O(C\widetilde{C}^{R-1}q^{-n}P^{n+\eta-1})\right)e(-\underline{\beta}\cdot\underline{\nu})) \\
&= P^n q^{-n} S_{\underline{a},q}(\underline{\nu})\cdot I(\mathcal{B},P^d\underline{\beta})\cdot e(-\underline{\beta}\cdot\underline{\nu})) + O(C\widetilde{C}^{R-1}P^{n+\eta-1}).
\end{aligned}$$

□

### 2.4.4 Singular series

**Definition 2.22.** Define the *singular series* as

$$\mathfrak{S}(\underline{\nu}) = \sum_{q=1}^{\infty} q^{-n} \sum_{\substack{\underline{a} \bmod q \\ (\underline{a},q)=1}} S_{\underline{a},q}(\underline{\nu}),$$

where $S_{\underline{a},q}(\underline{\nu})$ is defined by (2.8).

The singular series converges absolutely under assumption (2.25) on $K$ as shown in [Bir62, p. 256]. To make this quantitative we first prove the following elementary lemma:

**Lemma 2.23.** *For $\alpha > 0$ and $m \in \mathbb{N}$ we have*

$$\sum_{x=m}^{\infty} x^{-1-\alpha} \ll m^{-\alpha}.$$

*Proof.* Let $k \in \mathbb{Z}_{\ge 0}$. Observe that for $2^k m \le x < 2^{k+1}m$ we have

$$x^{-1-\alpha} \le \frac{1}{2^{k(1+\alpha)}m^{1+\alpha}}.$$

Hence,

$$\sum_{x=2^k m}^{2^{k+1}m} x^{-1-\alpha} \le \frac{1}{2^{k\alpha}m^{\alpha}}$$

As the geometric series $\sum_{k=0}^{\infty}\frac{1}{2^{k\alpha}}$ converges, it follows that $\sum_{x=m}^{\infty} x^{-1-\alpha} \ll m^{-\alpha}$ as desired. □

**Lemma 2.24.** *The singular series converges absolutely. In fact, for all $\tau \geq 0$ we have*

$$\sum_{P^{\tau\eta}<q<\infty} \sum_{\substack{\underline{a} \bmod q \\ (\underline{a},q)=1}} q^{-n} |S_{\underline{a},q}(\underline{\nu})| \ll \widetilde{C}^{K/(d-1)} P^{-\tau\delta}.$$

*Proof.* Observe $|S_{\underline{a},q}(\underline{\nu})| = |S_{\underline{a},q}|$ if $\underline{\nu}$ is real or $|S_{\underline{a},q}(\underline{\nu})| \ll_{\underline{\nu}} |S_{\underline{a},q}|$ if $\underline{\nu}$ is complex. We have that

$$\begin{aligned}
\sum_{P^{\tau\eta}<q<\infty} \sum_{\substack{\underline{a} \bmod q \\ (\underline{a},q)=1}} q^{-n} |S_{\underline{a},q}(\underline{\nu})| &\ll \sum_{P^{\tau\eta}<q<\infty} \sum_{\substack{\underline{a} \bmod q \\ (\underline{a},q)=1}} q^{-n} \widetilde{C}^{K/(d-1)} q^{n-K/R(d-1)+\varepsilon} && \text{(by Lemma 2.14)} \\
&\ll \widetilde{C}^{K/(d-1)} \sum_{P^{\tau\eta}<q<\infty} q^{R-K/R(d-1)+\varepsilon} \\
&\ll \widetilde{C}^{K/(d-1)} \sum_{P^{\tau\eta}<q<\infty} q^{-1-\delta\eta^{-1}} && \text{(by (2.25))} \\
&\ll \widetilde{C}^{K/(d-1)} P^{-\tau\delta}, && \text{(by Lemma 2.23)}
\end{aligned}$$

as desired. □

**Definition 2.25.** For each prime $p$ define the *local density* at $p$ to be

$$\mathfrak{S}_p(\underline{\nu}) = \sum_{r=0}^{\infty} \sum_{\substack{\underline{a} \bmod p^r \\ (\underline{a},p)=1}} p^{-rn} S_{\underline{a},p^r}(\underline{\nu}).$$

Then, by multiplicativity of $S_{\underline{a},q}$ we can factorize the singular series as a product over the local densities, i.e.

$$\mathfrak{S}(\underline{\nu}) = \prod_{p \text{ prime}} \mathfrak{S}_p(\underline{\nu}).$$

This can be proven with the same ideas as in [Dav05, Lemma 5.2].

The next lemma provides an interpretation of the singular series. Assume that $\underline{f}(\underline{x})$ obtains every value mod $p^N$ the same number of times when $\underline{x}$ ranges over all elements of $(\mathbb{Z}/p^N\mathbb{Z})^n$ (which of course is not the case, but can be used to calculate the expected number of points on $V$ modulo $p^N$). As there are $p^{RN}$ possible outcomes of $\underline{f}(\underline{x})$ and $p^{Nn}$ possible choices for $\underline{x}$, we heuristically expect $p^{N(n-R)}$ points on $V$ modulo $p^N$. Then, we can interpret the local densities as the density of points modulo $p^N$ as $N \to \infty$:

**Proposition 2.26.**

$$\mathfrak{S}_p(\underline{\nu}) = \lim_{N\to\infty} \frac{\#\{\underline{x} \mod p^N \mid \underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N\}}{p^{N(n-R)}}.$$

*Proof.* By Lemma 2.9 we find that the number of points $\underline{x} \in (\mathbb{Z}/p^N\mathbb{Z})^n$ satisfying $\underline{f}(\underline{x}) \equiv \underline{\nu} \bmod p^N$ equals

$$p^{-NR} \sum_{\underline{a} \bmod p^N} S_{\underline{a},p^N}(\underline{\nu}).$$

Suppose $(\underline{a}, p^N) = p^s$ for some $s \geq 0$. Then

$$S_{\underline{a},p^N}(\underline{\nu}) = S_{p^{-s}\underline{a},p^{N-s}}(\underline{\nu}) p^{ns}.$$

Hence, letting $r = N - s$ for all $\underline{a} \mod p^N$, we have that

$$p^{-RN} \sum_{\underline{a} \bmod p^N} S_{a,p^N}(\underline{\nu}) = p^{N(n-R)} \sum_{r=0}^{N} \sum_{\substack{\underline{a} \bmod p^r \\ (\underline{a},p^r)=1}} p^{-rn} S_{\underline{a},p^r}(\underline{\nu})$$

is the number of points satisfying $\underline{f}(\underline{x}) = \underline{\nu} \mod p^N$. Then

$$\begin{aligned}\mathfrak{S}_p(\underline{\nu}) &= \lim_{N\to\infty} \sum_{r=0}^{N} \sum_{\substack{\underline{a} \bmod p^r \\ (\underline{a},p^r)=1}} p^{-rn} S_{\underline{a},p^r}(\underline{\nu}) \\ &= \lim_{N\to\infty} p^{N(R-n)} \#\{\underline{x} \mod p^N \mid \underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N\}. \end{aligned}$$ □

So, we can think of the singular series as a quantity giving information about the number of points on $V$ over $\mathbb{Z}_p$ for all primes $p$ at the same times.

### 2.4.5 Singular integral

In this section, we define the singular integral. This is the real analogue of the singular series. First, we prove two lemmata generalising [Bir62, Corollary on p.252 and Lemma 5.2] in order to prove convergence of the singular integral.

**Lemma 2.27.** *If $|\underline{\alpha}| < (\widetilde{C}P^d)^{-1/2}$, then*

$$S(\underline{\alpha}) \ll P^{n+\varepsilon}(\widetilde{C}^{1-R}P^d|\underline{\alpha}|)^{-K/R(d-1)}.$$

*Proof.* As $|S(\underline{\alpha})| \ll P^n$ trivially, we may suppose that $|\underline{\alpha}| > \widetilde{C}^{R-1}P^{-d}$. Define $\varphi$ by $|\underline{\alpha}| = \widetilde{C}^{R-1}P^{-d+R(d-1)\varphi}$. By the assumption $|\underline{\alpha}| < (\widetilde{C}P^d)^{-1/2}$ it follows that

$$2R(d-1)\varphi + (2R-1)\log_P(\widetilde{C}) < d.$$

Hence, by Lemma 2.16 the intervals $\mathfrak{M}_{\underline{a},q}$ are disjoint. Observe that $\underline{\alpha}$ lies on the boundary of $\mathfrak{M}_{\underline{0},1}(\varphi)$. Hence, $\underline{\alpha}$ is not in $\mathfrak{M}(\varphi - \varepsilon)$ for all $\varepsilon > 0$. By Lemma 2.13 we conclude that for all $\varepsilon > 0$ it holds that $|S(\underline{\alpha})| \ll P^{n-K\varphi+\varepsilon}$. The lemma directly follows by plugging in the definition of $\varphi$. □

**Lemma 2.28.** *For all $\underline{\gamma} \in \mathbb{R}^R$ one has*

$$\left|I(\mathcal{B};\underline{\gamma})\right| \ll \min(1, (\widetilde{C}^{1-R}|\underline{\gamma}|)^{-R-1-\delta\eta^{-1}}(\widetilde{C}|\underline{\gamma}|)^{\varepsilon}),$$

*where $I$ is defined by* (2.27).

*Proof.* $I(\mathcal{B};\underline{\gamma}) \ll 1$ follows directly as $|e(\underline{\gamma} \cdot \underline{f}(\underline{\zeta}))| \leq 1$ and $\mathcal{B}$ has volume at most 1. Therefore, in proving the second part of the inequality we may assume that

$$\widetilde{C}^{1-R}|\gamma| > 1. \tag{2.31}$$

Assume $P > (\widetilde{C}|\underline{\gamma}|^2)^{1/d}$ and let $\underline{\alpha} = P^{-d}\underline{\gamma}$. Then

$$|\underline{\alpha}| = P^{-d}|\underline{\gamma}| < (\tilde{C}P^d)^{-1/2}.$$

By Lemma 2.27 we then find that

$$|S(\underline{\alpha})| \ll P^{n+\varepsilon}(\widetilde{C}^{1-R}P^d|\underline{\alpha}|)^{-K/R(d-1)}. \tag{2.32}$$

On the other hand, by Lemma 2.19 with $\underline{z}=0, \underline{\beta}=\underline{\alpha}$ and $q=1$, we obtain

$$S(\underline{\alpha}) = \sum_{\underline{y}\in P\mathcal{B}\cap\mathbb{Z}^R} e(\underline{\alpha}\cdot\underline{f}(\underline{y})) = P^n I(\mathcal{B}, P^d\underline{\alpha}) + O\left(\left(\widetilde{C}|P^d\underline{\alpha}|+1\right)P^{n-1}\right). \tag{2.33}$$

Combining (2.32) and (2.33) we obtain

$$\left|I(\mathcal{B},\underline{\gamma})\right| \ll P^{\varepsilon}(\widetilde{C}^{1-R}|\underline{\gamma}|)^{-K/R(d-1)} + \widetilde{C}|\underline{\gamma}|P^{-1} + P^{-1}. \tag{2.34}$$

Now, take

$$P = \widetilde{C}|\underline{\gamma}|(\widetilde{C}^{1-R}|\underline{\gamma}|)^{K/R(d-1)}.$$

By (2.31) and $d\geq 2$ we find that indeed $P > (\widetilde{C}|\underline{\gamma}|^2)^{1/d}$ is satisfied. For this choice of $P$ we obtain in (2.34) that

$$\left|I(\mathcal{B},\underline{\gamma})\right| \ll (\widetilde{C}^{1-R}|\underline{\gamma}|)^{-K/R(d-1)}(\widetilde{C}|\underline{\gamma}|)^{\varepsilon}.$$

Estimating $K/R(d-1)$ by $R+1+\delta\eta^{-1}$ using (2.25) concludes the proof. □

**Definition 2.29.** For $\underline{\nu}\in\mathbb{Z}^R$ and $\Phi\in\mathbb{R}_{\geq 0}$, write

$$J(\underline{\nu},\Phi) = \int_{|\underline{\gamma}|\leq\Phi} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})\,\mathrm{d}\underline{\gamma}$$

and define the *singular integral* to be

$$J(\underline{\nu}) = \lim_{\Phi\to\infty} J(\underline{\nu},\Phi)$$

if this limit exists.

The singular integral is well-defined and this can be made quantitative as in [Bir62, Lemma 5.3]:

**Lemma 2.30.** *$J(\underline{\nu})$ exists, is continuous and for all $\Phi>0$ we have*

$$|J(\underline{\nu}) - J(\underline{\nu},\Phi)| \ll \widetilde{C}^{R^2-1+(R-1)\delta\eta^{-1}}\Phi^{-1-\delta\eta^{-1}}. \tag{2.35}$$

*Proof.* Let $\Phi_2>\Phi_1\geq 1$. Then, using Lemma 2.28 we find

$$\begin{aligned}
J(\underline{\nu},\Phi_2) - J(\underline{\nu},\Phi_1) &= \int_{\Phi_1\leq|\underline{\gamma}|\leq\Phi_2} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})\,\mathrm{d}\underline{\gamma}\\
&\ll \int_{\Phi_1\leq|\underline{\gamma}|\leq\Phi_2} (\widetilde{C}^{1-R}|\underline{\gamma}|)^{-R-1-\delta\eta^{-1}}(\widetilde{C}|\underline{\gamma}|)^{\varepsilon}\,\mathrm{d}\underline{\gamma}\\
&\ll \int_{\Phi_1}^{\Phi_2} \Gamma^{R-1}\widetilde{C}^{R^2-1+(R-1)\delta\eta^{-1}}\Gamma^{-2-\delta\eta^{-1}}\,\mathrm{d}\Gamma\\
&\ll \widetilde{C}^{R^2-1+(R-1)\delta\eta^{-1}}\Phi_1^{-1-\delta\eta^{-1}}.
\end{aligned}$$

This implies that $J(\underline{\nu})$ exists. As $J(\underline{\nu},\Phi)$ is continuous in $\underline{\nu}$ and converges uniformly to $J(\underline{\nu})$ when $\Phi\to\infty$, it follows that $J(\underline{\nu})$ is continuous. The bound (2.35) follows by taking the limit $\Phi_2\to\infty$ above. □

**Lemma 2.31.** *For all $\underline{\nu} \in \mathbb{Z}^R$ it holds that*

$$|J(\underline{\nu})| \ll \widetilde{C}^{R(R-1)}.$$

*Proof.* We have

$$|J(\underline{\nu}, \widetilde{C}^{R-1})| \leq \int_{|\underline{\gamma}| \leq \widetilde{C}^{R-1}} |I(\mathcal{B}, \underline{\gamma})| \, \mathrm{d}\underline{\gamma} \ll \widetilde{C}^{R(R-1)}$$

by the trivial bound in Lemma 2.28 and because the volume of the domain of integration is bounded by $\left(\widetilde{C}^{R-1}\right)^R$. By the previous lemma we have

$$\left| J(\underline{\nu}) - J(\underline{\nu}, \widetilde{C}^{R-1}) \right| \ll \widetilde{C}^{R^2-1+(R-1)\delta\eta^{-1}} \widetilde{C}^{(R-1)(-1-\delta\eta^{-1})} = \widetilde{C}^{R(R-1)},$$

which implies the result. □

An interpretation of the singular integral comparable to the interperation of the singular series by Proposition 2.26 will be given by Proposition 2.40.

### 2.4.6 Major arcs

We are now ready to give an asymptotic for the number of integer points in a box $P\mathcal{B}$, generalising [Bir62, Lemma 5.5]:

**Lemma 2.32.**

$$M(P; \underline{\nu}) = P^{n-Rd} \mathfrak{S}(\underline{\nu}) J(P^{-d}\underline{\nu}) + O\left( \widetilde{C}^{R^2-R} P^{n-Rd} \left( C\widetilde{C}^{R^2+2R-1} P^{-1+2(R+1)\eta} + \widetilde{C}^{K/(d-1)} P^{-\delta} \right) \right).$$

*Proof.* By Corollary 2.18 we have that

$$\begin{aligned} M(P; \underline{\nu}) &= \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a}, q) = 1}} \int_{\mathfrak{M}_{\underline{a},q}} S(\underline{\alpha}, \underline{\nu}) \, \mathrm{d}\underline{\alpha} + O(\widetilde{C}^{R^2} P^{n-Rd-\delta}) \\ &= \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a}, q) = 1}} \int_{|\underline{\beta}| \leq \widetilde{C}^{R-1} P^{-d+\eta}} S(\underline{\alpha}, \underline{\nu}) \, \mathrm{d}\underline{\beta} + O(\widetilde{C}^{R^2} P^{n-Rd-\delta}), \end{aligned}$$

where in the second integral it is understood that $\underline{\alpha} = \underline{a}/q + \underline{\beta}$. As $S_{\underline{a},q}(\underline{\nu}) \leq q^n$ and there are at most $(\widetilde{C}^R P^\eta)^{R+1}$ choices for $\underline{a}$ and $q$, we find using Corollary 2.21 that

$$\begin{aligned} M(P, \underline{\nu}) &= P^n \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a}, q) = 1}} S_{\underline{a},q}(\underline{\nu}) \int_{|\underline{\beta}| \ll \widetilde{C}^{R-1} P^{-d+\eta}} I(\mathcal{B}; P^d \underline{\beta}) \cdot e(-\underline{\beta} \cdot \underline{\nu}) \, \mathrm{d}\underline{\beta} \\ &\qquad + O((\widetilde{C}^R P^\eta)^{R+1} (\widetilde{C}^{R-1} P^{-d+\eta})^R C\widetilde{C}^{R-1} P^{n+\eta-1}) + O(\widetilde{C}^{R^2} P^{n-Rd-\delta}) \\ &= P^{n-Rd} \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a}, q) = 1}} S_{\underline{a},q}(\underline{\nu}) \int_{|\underline{\gamma}| \ll \widetilde{C}^{R-1} P^\eta} I(\mathcal{B}; \underline{\gamma}) \cdot e(-\underline{\gamma} \cdot P^{-d}\underline{\nu}) \, \mathrm{d}\underline{\gamma} + \mathfrak{O} \\ &= P^{n-Rd} \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a}, q) = 1}} S_{\underline{a},q}(\underline{\nu}) J(P^{-d}\underline{\nu}, \widetilde{C}^{R-1} P^\eta) + \mathfrak{O}, \end{aligned}$$

where

$$\mathfrak{O} = O(C\widetilde{C}^{2R^2+R-1} P^{n-Rd-1+2(R+1)\eta}) + O(\widetilde{C}^{K/(d-1)+R^2-R} P^{n-Rd-\delta}).$$

Here we multiplied the second error term with $\widetilde{C}^{\frac{K}{d-1}-R} \geq 1$ so that all subsequent error terms are at most $\mathfrak{O}$.

Using Lemma 2.30 and Lemma 2.24 for $\tau = 0$, we can plug in the singular integral to find that $M(P, \underline{\nu})$ equals

$$\begin{aligned}
&P^{n-Rd} \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a},q)=1}} S_{\underline{a},q}(\underline{\nu}) \left( J(P^{-d}\underline{\nu}) + O\left( \widetilde{C}^{R^2-1+(R-1)\delta\eta^{-1}} (\widetilde{C}^{R-1} P^\eta)^{-1-\delta\eta^{-1}} \right) \right) + \mathfrak{O} \\
&= P^{n-Rd} \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a},q)=1}} S_{\underline{a},q}(\underline{\nu}) J(P^{-d}\underline{\nu}) + O\left( P^{n-Rd} \widetilde{C}^{K/(d-1)} \widetilde{C}^{R^2-R} P^{-\eta-\delta} \right) + \mathfrak{O} \\
&= P^{n-Rd} \sum_{1 \leq q \leq \widetilde{C}^R P^\eta} q^{-n} \sum_{\substack{1 \leq a_i \leq q \\ (\underline{a},q)=1}} S_{\underline{a},q}(\underline{\nu}) J(P^{-d}\underline{\nu}) + \mathfrak{O}.
\end{aligned}$$

By Lemma 2.24 for $\tau = 1$ and Lemma 2.31 we can plug in the singular series and obtain

$$\begin{aligned}
M(P, \underline{\nu}) &= P^{n-Rd} \left( \mathfrak{S}(\underline{\nu}) + O(\widetilde{C}^{K/(d-1)} P^{-\delta}) \right) J(P^{-d}\underline{\nu}) + \mathfrak{O} \\
&= P^{n-Rd} \mathfrak{S}(\underline{\nu}) J(P^{-d}\underline{\nu}) + O(P^{n-Rd} \widetilde{C}^{K/(d-1)} P^{-\delta} \widetilde{C}^{R(R-1)}) + \mathfrak{O} \\
&= P^{n-Rd} \mathfrak{S}(\underline{\nu}) J(P^{-d}\underline{\nu}) + \mathfrak{O}.
\end{aligned}$$

□

**Theorem 2.33.**

$$M(P, \underline{\nu}) = P^{n-Rd} \mathfrak{S}(\underline{\nu}) J(P^{-d}\underline{\nu}) + O(C \widetilde{C}^{K/(d-1)+R^2-1} P^{n-Rd-\delta}),$$

*where*

$$\delta < \frac{K - R(R+1)(d-1)}{K + R(R+1)(d-1)}.$$

*Proof.* Let $\varepsilon > 0$ be given and take

$$\delta = \left( \frac{K}{R(d-1)} - (R+1) \right) \eta - \varepsilon, \quad \eta = \frac{1}{K/R(d-1) + R + 1}.$$

It follows directly that (2.24) and (2.25) are satisfied. We have that

$$\begin{aligned}
\delta = \left( \frac{K}{R(d-1)} + (R+1) \right) \eta - 2(R+1)\eta - \varepsilon &< 1 - 2(R+1)\eta \\
&= \frac{K - R(R+1)(d-1)}{K + R(R+1)(d-1)}.
\end{aligned}$$

Hence, $P^{-1+2(R+1)\eta} < P^{-\delta}$. The statement now follows directly from Lemma 2.32. □

We see that if $\mathfrak{S}(\underline{0})J(\underline{0}) > 0$, we have that $M(P, \underline{0}) \to \infty$ as $P \to \infty$. Birch already showed that $\mathfrak{S}(\underline{0}) > 0$ if $\underline{f}$ has non-singular zeros over $\mathbb{Z}_p$ for all primes $p$ and that $J(\underline{0}) > 0$ if $\underline{\tilde{f}}$ has a non-singular zero over $\mathbb{R}$. In the next chapter we make this quantitative, given lower bounds for the singular series and the singular integral. This is used in Theorem 2.7 to give an upper bound on the smallest integer solution of $\underline{f}$.

## 2.5 Quantitative strong approximation

### 2.5.1 Lower bound for the singular series

Suppose $\mathfrak{S}(\underline{\nu}) = 0$. By uniform convergence of the product of local densities, this implies that $\mathfrak{S}_p(\underline{\nu}) = 0$ for some prime $p$. We will see in Lemma 2.36 that this implies that $\underline{f}(\underline{x}) = \underline{\nu}$ has no non-singular solution over $\mathbb{Z}_p$. As we proceed with the assumption that $\underline{f}$ has non-singular zeros over $\mathbb{Z}_p$ and $\mathbb{R}$ for all primes $p$, we conclude that $\mathfrak{S}(\underline{0})$ is strictly positive. In this section we make this statement quantitative.

**Definition 2.34.** Let $p$ be prime and $e \in \mathbb{Z}_{\geq 0}$. A solution of $\underline{f}(\underline{x}) = \underline{\nu} \mod p^{2e+1}$ is called *non-singular* if the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}\right)_{i,j}$ has a minor with determinant non-zero modulo $p^{e+1}$.

Observe that a solution $\underline{x}$ of $\underline{f}(\underline{x}) = \underline{\nu}$ in $\mathbb{Z}_p$ is non-singular if and only if there is an $e$ such that $\underline{x} \mod p^{2e+1}$ is a non-singular solution of $\underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^{2e+1}$. Namely, for a non-singular solution $\underline{x}$ of $\underline{f}(\underline{x}) = \underline{\nu}$ in $\mathbb{Z}_p$ the Jacobian matrix of $\underline{f}$ in $\underline{x}$ has full rank.

**Lemma 2.35.** *Let $E > 2e+1$ with $E, e \in \mathbb{Z}_{\geq 0}$. A non-singular solution $\underline{a}$ of $\underline{f}(\underline{a}) = \underline{\nu} \mod p^{2e+1}$ lifts to $p^{(n-R)(E-2e-1)}$ non-singular solutions $\underline{x}$ of $\underline{f}(\underline{x}) = \underline{\nu} \mod p^E$.*

*Proof.* This is a consequence of the Multidimensional Hensel's lemma (see, for example, [Gre69, Proposition 5.20]). Suppose without loss of generality that the minor with determinant non-zero modulo $p^{e+1}$ is the leading minor $\Delta$. Now, choose $b_{n-R}, \ldots, b_n \mod p^E$ such that $b_i = a_i \mod p^{2e+1}$ for all $i = R+1, \ldots, n$. There are $p^{(n-R)(E-2e-1)}$ such choices. Given such $b_{R+1}, \ldots, b_n \mod p^E$ we have that if also $b_i \equiv a_i \mod p^{2e+1}$ for $i = 1, \ldots, R$, then

$$\underline{f}(\underline{b}) \equiv \underline{f}(\underline{a}) \equiv \underline{\nu} \mod p^{2e+1} \quad \text{and} \quad \Delta(\underline{b}) \equiv \Delta(\underline{a}) \not\equiv 0 \mod p^E.$$

Hence, Hensel's lemma indicates that there is a solution $(x_1, \ldots, x_R) \in \mathbb{Z}_p^n$ such that $x_i \equiv b_i \mod p^{2e+1}$. Then $(x_1, \ldots, x_R, b_{R+1}, \ldots, b_R) \mod p^E$ is a solution of $\underline{f}(\underline{x}) = \underline{\nu} \mod p^E$. □

Let $I$ be a subset of $[n] := \{1, \ldots, n\}$ of size $R$ and let $\Delta_I(\underline{x})$ be the $R \times R$-minor of the Jacobian matrix of $\underline{f}$ with rows given by the elements of $I$. Here, the Jacobian matrix of $\underline{f}$ is given by

$$\left(\frac{\partial f_i}{\partial x_j}(\underline{x})\right)_{i,j}.$$

Similarly, let $\tilde{\Delta}_I(\underline{x})$ be the $R \times R$-minor of the Jacobian matrix of $\underline{\tilde{f}}$ with rows given by the elements of $I$.

**Lemma 2.36.** *If there exists a non-singular solution $\underline{x}_0 \in \mathbb{Z}_p^n$ to $\underline{f}(\underline{x}_0) = \underline{\nu}$, then*

$$\mathfrak{S}_p(\underline{\nu}) \geq \left(p^{-1} \max_I |\Delta_I(\underline{x}_0)|_p^2\right)^{n-R}.$$

*Proof.* Take $e \in \mathbb{Z}$ such that $p^{-e} = \max_I |\Delta_I(\underline{x}_0)|_p$ and assume that $N > 2e+1$. The non-singular solution $\underline{x}_0 \in \mathbb{Z}_p^n$ gives a non-singular solution modulo $p^{2e+1}$. By Lemma 2.35 we find that there are at least $p^{(n-R)(N-2e-1)}$ solutions of $\underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N$. Hence, by Proposition 2.26 we find

$$\mathfrak{S}_p(\underline{\nu}) \geq \lim_{N \to \infty} p^{N(R-n)} p^{(n-R)(N-2e-1)} = p^{-(n-R)(2e+1)} = (p^{-1} \max_I |\Delta_I(\underline{x}_0)|_p^2)^{n-R}. \qquad \square$$

Note that if we apply the above lemma to all primes $p$, we cannot deduce more than $\mathfrak{S}(\underline{\nu}) \geq 0$. Hence, we need stronger estimates. Therefore, from now on assume that $V$ and $\widetilde{V}$ are non-singular over $\overline{\mathbb{Q}}$ as affine respectively projective varieties. This depends on the choice of our system $\underline{f}$, because the definition of a singular point (Definition 2.3) depends not only on the zero set of the system, but also on the choice of the polynomials.

Consider the polynomials $f_1, \ldots, f_R$ together with the polynomials $\Delta_I$. As $V$ is non-singular, these polynomials have no common zero over $\overline{\mathbb{Q}}$. Hence, by the Nullstellensatz, the ideal generated by these polynomials equals $\overline{\mathbb{Q}}[\underline{x}]$. This is made quantitative in Theorem 1 of [KPS01]: there exist an $N \in \mathbb{N}$ and polynomials $g_1, \ldots, g_R$ and $g_I$ in $\mathbb{Z}[\underline{x}]$ for all $I \subset [n]$ with $|I| = R$ such that

$$\sum_{i=1}^{R} f_i(\underline{x}) g_i(\underline{x}) + \sum_{I} \Delta_I(\underline{x}) g_I(\underline{x}) = N, \tag{2.36}$$

satisfying the estimate

$$\log(N) \ll 4n(n+1)D^n \log(C^R).$$

Here, $D$ is such that $\deg f_i \leq D$ and $\deg \Delta_I \leq D$. Taking $D = \max(R(d-1), d)$, we find

$$N \ll C^{4n(n+1)R\max(R(d-1),d)^n} = \mathfrak{C}, \tag{2.37}$$

where the above equation defines $\mathfrak{C}$.

For the projective variety $\widetilde{V}$ we have to be slightly more careful. Namely, the polynomials $\tilde{f}_1, \cdots, \tilde{f}_R$ have the common zero $(0, 0, \cdots, 0)$. Therefore, we do the same as above on every affine patch obtained by setting one of the coordinates $x_j = 1$. Let $1 \leq j \leq n$ be given. Because $\widetilde{V}$ is non-singular over $\overline{\mathbb{Q}}$, we can find $\tilde{N}_j \in \mathbb{N}$ and polynomials $\tilde{g}_{1,j}, \ldots, \tilde{g}_{R,j}$ and $\tilde{g}_{I,j}$ in $\mathbb{Z}[x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$ for all $I \subset [n]$ with $|I| = R$ such that

$$\sum_{i=1}^{R} \tilde{f}_i(\underline{x}) \tilde{g}_{i,j}(\underline{x}) + \sum_{I} \tilde{\Delta}_I(\underline{x}) \tilde{g}_{I,j}(\underline{x}) = \tilde{N}_j \tag{2.38}$$

for all $\underline{x}$ with $x_j = 1$. Denote with $\|g\|_\infty$ the height of a polynomial $g$, that is, $\|g\|_\infty$ is the maximum of the absolute values of the coefficients of $g$. Then, by Theorem 1 in [KPS01] equation (2.38) satisfies the following estimate:

$$\log \|\tilde{g}_{I,j}\|_\infty \ll 4n(n-1)D^{n-1} \log(\widetilde{C}^R),$$

for all $I \subset [n]$ with $|I| = R$. Here, we can take $D = \max(R(d-1), d)$ so that

$$\|\tilde{g}_{I,j}\|_\infty \ll \widetilde{C}^{4n(n-1)R\max(R(d-1),d)^{n-1}} = \tilde{\mathfrak{C}}, \tag{2.39}$$

where the above equation defines $\tilde{\mathfrak{C}}$.

**Lemma 2.37.** *For all primes $p$ for which there exists a solution $\underline{x}_0 \in \mathbb{Z}_p^n$ of $\underline{f}(\underline{x}_0) = \underline{0}$ we have*

$$\max_I |\Delta_I(\underline{x}_0)|_p \geq |N|_p.$$

*Proof.* Let $p$ be a prime such that there exists an $\underline{x}_0 \in \mathbb{Z}_p^n$ with $\underline{f}(\underline{x}_0) = \underline{0}$, so that the first set of terms on the left-hand side of (2.36) vanishes for $\underline{x} = \underline{x}_0$. Then taking $p$-adic absolute values in (2.36) shows that

$$\max_I |\Delta_I(\underline{x}_0)|_p \max_I |g_I(\underline{x}_0)|_p \geq |N|_p.$$

As $g_I \in \mathbb{Z}[\underline{x}]$ we have $|g_I(\underline{x}_0)|_p \leq 1$, so we obtain $\max_I |\Delta_I(\underline{x}_0)|_p \geq |N|_p$. □

**Lemma 2.38.** *If $p$ is prime such that $p \nmid d$ and $p \nmid N$, then*

$$\mathfrak{S}_p(\underline{0}) - 1 \ll p^{-n/2+R+\varepsilon}. \tag{2.40}$$

*Proof.* Suppose $V(\underline{0})$ is singular over $\mathbb{F}_p$. Then, there exists an $\underline{x} \in \mathbb{F}_p^n$ such that $\underline{f}(\underline{x}) = \underline{0}$ and $\Delta_I(\underline{x}) = 0$ over $\mathbb{F}_p$ for all $I \subset [n]$ with $|I| = R$. Considering (2.36) over $\mathbb{F}_p$, it follows that $N \equiv 0 \mod p$. This contradicts our assumption, so $V$ is non-singular over $\mathbb{F}_p$.

As pointed out by Schmidt [Sch84], a result of Deligne, worked out in the appendix of [Ser77], then shows that

$$\#V_{\mathbb{F}_p}(\underline{0}) = p^{n-R} + O(p^{n/2+\varepsilon})$$

if $p \nmid d$, where the implied constant depends at most on $n$ and $d$. Observe that if $\underline{x} \in \mathbb{Z}^n$ is a solution of $\underline{f}(\underline{x}) = \underline{0} \mod p^e$ for some $e \in \mathbb{N}$, then $\underline{x}$ reduces to a non-singular point on $V_{\mathbb{F}_p}(\underline{0})$, as $V_{\mathbb{F}_p}$ is non-singular over $\mathbb{F}_p$. Hence, $\underline{x} \mod p^e$ can be obtained by lifting a point of $V_{\mathbb{F}_p}(\underline{0})$ as in Lemma 2.35. We conclude that

$$\#\{\underline{x} \mod p^N \mid \underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N\} = p^{N(n-R)} + (p^{(n-R)(N-1)+n/2+\varepsilon}).$$

Similar to the proof of Lemma 2.36, we find by Proposition 2.26 that

$$\begin{aligned}
\mathfrak{S}_p(\underline{0}) &= \lim_{N\to\infty} p^{N(R-n)} \#\{\underline{x} \mod p^N \mid \underline{f}(\underline{x}) \equiv \underline{\nu} \mod p^N\} \\
&= \lim_{N\to\infty} p^{N(R-n)} \left(p^{N(n-R)} + O(p^{(N-1)(n-R)+n/2+\varepsilon})\right) \\
&= 1 + O(p^{-n/2+R+\varepsilon}).
\end{aligned}$$

□

The uniform convergence of the product $\prod_p \mathfrak{S}_p$ implies that $\mathfrak{S}_p > 0$ for $p$ sufficiently large. Above lemma can be interpreted as a quantitative version of that statement. Namely, for $p$ large, so that $p > Nd$ and $p^{-n/2+R+\varepsilon}$ is smaller than the implied constant in (2.40) above lemma implies that that $\mathfrak{S}_p(\underline{\nu}) > 0$. Hence, the existence of infinitely many $p$-adic point on $V(\underline{\nu})$ is automatic for $p$ sufficiently large.

**Proposition 2.39.** *Suppose that for all primes $p$ there exists a non-singular solution $\underline{x}_0 \in \mathbb{Z}_p^n$ to $\underline{f}(\underline{x}_0) = \underline{0}$. Then*

$$\mathfrak{S}(\underline{0}) \gg N^{-3(n-R)}.$$

*Proof.* In this proof $p$ will always denote a prime. Let $S$ be the finite set of 'bad' primes $p$, i.e. primes such that $p \mid d$ or $p \mid N$. Applying Lemma 2.36 and Lemma 2.37 we obtain

$$\prod_{p\in S} \mathfrak{S}_p(\underline{0}) \geq \prod_{p\in S} \left(p^{-1}|N|_p^2\right)^{n-R}. \tag{2.41}$$

As the product over all bad primes is at most $Nd$, one has

$$\prod_{p\in S} p^{-1} \gg_d N^{-1}. \tag{2.42}$$

Using the fact that $N$ is an integer and then the product formula for $|\cdot|_p$, we find

$$\prod_{p\in S} |N|_p \geq \prod_p |N|_p = N^{-1}. \tag{2.43}$$

Applying (2.42) and (2.43) to (2.41) we obtain

$$\prod_{p\in S} \mathfrak{S}_p(\underline{0}) \gg N^{3(R-n)}.$$

For $p_0$ large enough (not depending on $C$) we have that

$$\prod_{p\notin S, p\geq p_0} 1 + O(p^{-n/2+R+\varepsilon})$$

converges (here $O$ does not depend on $C$). Moreover,

$$\prod_{p\leq p_0} \mathfrak{S}_p(\underline{0}) \gg 1$$

by Lemma 2.36 as $p_0$ does not depend on $C$ and $\max_I |g_I(\underline{x}_0)|_p = 1$ for $p \notin S$ . Hence, it follows by Lemma 2.38 that

$$\prod_{p\notin S} \mathfrak{S}_p(\underline{0}) \gg 1.$$

Therefore, we conclude that

$$\mathfrak{S}(\underline{0}) = \prod_p \mathfrak{S}_p(\underline{0}) = \prod_{p\in S} \mathfrak{S}_p(\underline{0}) \prod_{p\notin S} \mathfrak{S}_p(\underline{0}) \gg N^{-3(n-R)}. \qquad \square$$

### 2.5.2 Lower bound for the singular integral

In this section we find a lower bound for the singular integral $J(\underline{0})$. In order to do so, we make again use of the quantitative version of the Nullstellensatz. First, we rewrite the singular integral:

**Proposition 2.40** (Paragraph 11 of [Sch82])**.** *Let $\underline{\nu} \in \mathbb{R}^R$. Then*

$$J(\underline{\nu}) = \lim_{t\to\infty} t^R \int_{|\underline{\tilde{f}}(\underline{x})-\underline{\nu}|\leq t^{-1}} \prod_{i=1}^{R} (1 - t|\tilde{f}_i(\underline{x}) - \mu_i|)\, \mathrm{d}\underline{x}.$$

*Proof.* For $t > 0$ and $y \in \mathbb{R}$ let

$$\chi_t(y) = \begin{cases} t(1-t|y|) & \text{if } |y| \leq t^{-1} \\ 0 & \text{else.} \end{cases}$$

As the Fourier transform of $\chi_1(y)$ is given by $\left(\frac{\sin \pi y}{\pi y}\right)^2$, it follows that

$$\chi_1(y) = \int_{-\infty}^{\infty} e(\gamma y) \left(\frac{\sin \pi\gamma}{\pi\gamma}\right)^2 \mathrm{d}\gamma.$$

Observe that $\chi_t(y) = t\chi_1(ty)$. Hence, by the basic properties of the Fourier transform we have

$$\chi_t(y) = \int_{-\infty}^{\infty} e(\gamma y) \left(\frac{\sin \pi\gamma t^{-1}}{\pi\gamma t^{-1}}\right)^2 \mathrm{d}\gamma. \tag{2.44}$$

Now, for $\underline{z} \in \mathbb{R}^R$ put

$$\chi_t(\underline{z}) = \prod_{i=1}^{n} \chi_t(z_i).$$

and for $\underline{\gamma} \in \mathbb{R}^R$ let

$$K_t(\underline{\gamma}) = \prod_{i=1}^{R} \left(\frac{\sin \pi\gamma_i t^{-1}}{\pi\gamma_i t^{-1}}\right)^2.$$

Then, by (2.44) we obtain

$$\chi_t(\underline{z}) = \int_{\mathbb{R}^R} e(\underline{\gamma}\cdot\underline{z})K_t(\underline{\gamma})\,\mathrm{d}\underline{\gamma}. \tag{2.45}$$

Let

$$J_t(\underline{\nu}) = \int_{\mathcal{B}} \chi_t(\tilde{\underline{f}}(\underline{\xi}) - \underline{\nu})\,\mathrm{d}\underline{\xi}.$$

Replacing $\chi_t(\tilde{\underline{f}}(\underline{\xi})-\underline{\nu})$ by the absolute convergent integral (2.45), we may interchange the order of integration to obtain

$$J_t(\underline{\nu}) = \int_{\mathbb{R}^R} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})K_t(\underline{\gamma})\,\mathrm{d}\underline{\gamma},$$

where $I$ is defined by (2.27). We thus have

$$J_t(\underline{\nu}) - J(\underline{\nu}) = \int_{\mathbb{R}^R} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})(1-K_t(\underline{\gamma}))\,\mathrm{d}\underline{\gamma}.$$

We split this integral in integrals over the domains $|\underline{\gamma}|<t$ and $|\underline{\gamma}|\geq t$. For $|\underline{\gamma}|<t$ we have

$$\sin\pi\gamma_i t^{-1} = \pi\gamma_i t^{-1} + O((\gamma_i t^{-1})^3) = \pi\gamma_i t^{-1}(1+O(|\underline{\gamma}|^2 t^{-2})),$$

whence

$$K_t(\underline{\gamma}) = 1 + O(|\underline{\gamma}|^2 t^{-2}).$$

By Lemma 2.28, we obtain

$$\int_{|\underline{\gamma}|<t} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})(1-K_t(\underline{\gamma}))\,\mathrm{d}\underline{\gamma} \ll t^{-2}\int_{|\underline{\gamma}|<t}|\underline{\gamma}|^{2-R-1}\,\mathrm{d}\underline{\gamma} \ll t^{-2}\int_{\Gamma=0}^{t}\Gamma^{R-1}\Gamma^{1-R}\,\mathrm{d}\Gamma = t^{-1},$$

where the implied constant depends on $C$. In the last inequality we used that the integrand only depends on $\Gamma := |\underline{\gamma}|$.

On the other hand, we have that $0\leq K_t(\underline{\gamma})\leq 1$. Therefore, we deduce that in the case $|\underline{\gamma}|\geq t$ we have

$$\int_{|\underline{\gamma}|\geq t} I(\mathcal{B},\underline{\gamma})e(-\underline{\gamma}\cdot\underline{\nu})(1-K_t(\underline{\gamma}))\,\mathrm{d}\underline{\gamma} \ll \int_{|\underline{\gamma}|\geq t}|\underline{\gamma}|^{-R-1}\,\mathrm{d}\underline{\gamma} \ll \int_{\Gamma=t}^{\infty}\Gamma^{R-1}\Gamma^{-1-R}\,\mathrm{d}\Gamma = t^{-1}.$$

Here, again the implied constant may depend on $C$. Combining these two estimates we find that $\lim_{t\to\infty} J_t(\mu) = J(\mu)$. □

We now use ideas of [PSW16, Lemma 9.3] to provide an analogue of Lemma 2.35:

**Lemma 2.41** (Quantitative version of the inverse function theorem)**.** *Given $\underline{x}_0\in\mathbb{R}^n$ with $|\underline{x}_0|\leq\Lambda$ with $\Lambda\geq 1$, assume that $M := \max_{I\subset[n],|I|=R}|\Delta_I(\underline{x}_0)| = |\Delta(\underline{x}_0)| > 0$. Let $g:\mathbb{R}^n\to\mathbb{R}^n$ be given by*

$$g:\underline{x}\mapsto(\tilde{f}_1(\underline{x}),\dots,\tilde{f}_R(\underline{x}),x_{R+1},\dots,x_n).$$

*Then there are open subsets $W\subset\mathbb{R}^n$ and $W'\subset\mathbb{R}^n$ with $\underline{x}_0\in W$ and $g(\underline{x}_0)\in W'$ such that $g$ is a bijection from $W$ to $W'$ and has differentiable inverse $g^{-1}$ on $W'$ with $\det((g^{-1})')\geq M^{-1}$ on $W'$. Furthermore, one may choose*

$$W' = \left\{\underline{y}\in\mathbb{R}^n : |g(\underline{x}_0)-\underline{y}| \ll_{n,d,\Lambda} \frac{M^2}{\widetilde{C}^{2R-1}\Lambda^{R(d-1)-1}}\right\}.$$

Before we are going to prove this lemma, we recall a lemma in multivariable analysis, which shows how a multivariable function can be approximated by its derivatives.

**Lemma 2.42.** *Let $A \subset \mathbb{R}^n$ be a rectangle and $\varphi : A \to \mathbb{R}^r$ with $r \leq n$ be continuously differentiable. Suppose $B \in \mathbb{R}$ is such that $\left|\frac{\partial \varphi_i(\underline{x})}{\partial x_j}\right| \leq B$ for all $\underline{x}$ in the interior of $A$ and all $i = 1, \dots, r$ and $j = 1, \dots, n$. Then for all $\underline{x}, \underline{y} \in A$ we have*

$$|\varphi(\underline{x}) - \varphi(\underline{y})| \ll_n B|\underline{x} - \underline{y}|.$$

*Proof.* The case $n = r$ is Lemma 2.10 in [Spi65]. If $r < n$, extend $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ by $\varphi_i(\underline{x}) = 0$ for $r < i \leq n$. Then $\varphi$ remains continuously differentiable and $\frac{\partial \varphi_i(\underline{x})}{\partial x_j} = 0$ for $r < i \leq n$. Therefore, the statement follows from the case $n = r$. □

*Proof of Lemma 2.41.* We explicitly find a small open neighbourhood of $\underline{x}_0$ in which the implicit function theorem is applicable, following the proof of Theorem 2.11 in [Spi65] or the proof of Lemma 9.3 in [PSW16]. From $\left|\frac{\partial \tilde{f}_i}{\partial x_j}(\underline{x}_0)\right| \ll \widetilde{C}\Lambda^{d-1}$ as $\Lambda \geq 1$, it follows that

$$M \ll \widetilde{C}^R \Lambda^{R(d-1)}. \tag{2.46}$$

Let $U$ be the closed rectangle given by

$$U = \left\{ x \in \mathbb{R}^n : |x - \underline{x}_0| \leq a \frac{M}{\widetilde{C}^R \Lambda^{R(d-1)-1}} \right\}, \tag{2.47}$$

for a sufficiently small constant $a \in \mathbb{R}$ depending only on $d, n$ and $R$. Then for $\underline{x} \in U$ we have by (2.46) that

$$|\underline{x}| \leq |\underline{x} - \underline{x}_0| + |\underline{x}_0| \leq a \frac{M}{\widetilde{C}^R \Lambda^{R(d-1)-1}} + \Lambda \ll a\Lambda + \Lambda \ll \Lambda.$$

We claim that if $a$ is sufficiently small, the following three properties hold for all $\underline{x}, \underline{x}_1, \underline{x}_2 \in U$:

(1) $\left|\frac{\partial g_i}{\partial x_j}(\underline{x}) - \frac{\partial g_i}{\partial x_j}(\underline{x}_0)\right| \ll a\widetilde{C}^{-R+1}M$, for all $1 \leq i, j \leq n$;

(2) $|\Delta(\underline{x})| \gg M$;

(3) $|g(\underline{x}_1) - g(\underline{x}_2)| \geq \frac{1}{2}M|\underline{x}_1 - \underline{x}_2|$.

For the first property, observe that $\frac{\partial g_i}{\partial x_j x_k}(\underline{x})$ for all $1 \leq i, j, k \leq n$ is a polynomial with maximal coefficient $\ll \widetilde{C}$ of degree at most $d - 2$. Hence,

$$\frac{\partial g_i}{\partial x_j x_k}(\underline{x}) \ll \widetilde{C}\Lambda^{d-2}.$$

Applying Lemma 2.42 with $\varphi = \frac{\partial g_i}{\partial x_j}$ we find that

$$\left|\frac{\partial g_i}{\partial x_j}(\underline{x}) - \frac{\partial g_i}{\partial x_j}(\underline{x}_0)\right| \ll \widetilde{C}\Lambda^{d-2}|\underline{x} - \underline{x}_0| \ll a\frac{\widetilde{C}\Lambda^{d-2}M}{\widetilde{C}^R\Lambda^{R(d-1)-1}} \ll a\widetilde{C}^{-R+1}M.$$

For the second property, note that $\Delta(\underline{x})$ is a polynomial of degree at most $R(d-1)$. Hence, for all $\underline{x} \in U$ we have

$$\left|\frac{\partial \Delta(\underline{x})}{\partial x_j}\right| \ll \widetilde{C}^R \Lambda^{R(d-1)-1}.$$

Therefore, applying Lemma 2.42 with $\varphi = \Delta(\underline{x})$ we find that

$$|\Delta(\underline{x}) - \Delta(\underline{x}_0)| \leq \widetilde{C}^R \Lambda^{R(d-1)-1} |\underline{x} - \underline{x}_0| \leq a \frac{\widetilde{C}^R \Lambda^{R(d-1)-1} M}{\widetilde{C}^R \Lambda^{R(d-1)-1}} = aM.$$

Assuming $a$ is small enough, we have for all $\underline{x} \in U$ that

$$|\Delta(\underline{x})| \geq |\Delta(\underline{x}_0)| - |\Delta(\underline{x}) - \Delta(\underline{x}_0)| \geq (1-a)M \gg M.$$

For the third property, we apply Lemma 2.42 to $h(\underline{x}) = g(\underline{x}) - Dg(\underline{x}_0) \cdot \underline{x}$, where $Dg$ is the Jacobian matrix of $g$. By the first property we have that

$$\left| \frac{\partial h_i(\underline{x})}{\partial x_j} \right| = \left| \frac{\partial g_i}{\partial x_j}(\underline{x}) - \frac{\partial g_i}{\partial x_j}(\underline{x}_0) \right| \ll a \widetilde{C}^{-R+1} M,$$

whence for $\underline{x}_1, \underline{x}_2 \in U$ we have

$$|g(\underline{x}_1) - Dg(\underline{x}_0) \cdot \underline{x}_1 - g(\underline{x}_2) + Dg(\underline{x}_0) \cdot \underline{x}_2| \ll a \widetilde{C}^{-R+1} M |\underline{x}_1 - \underline{x}_2|. \tag{2.48}$$

Let $A$ be an invertible $n \times n$-matrix, denote with $|A| = \max_{i,j} |A_{i,j}|$ the maximum norm and assume that $|A| \ll 1$. For all $\underline{h} \in \mathbb{R}^n$ one has

$$|\underline{h}| = |A^{-1} A \underline{h}| \leq |A^{-1}||A\underline{h}| = \frac{1}{\det A} |\mathrm{adj}(A)||A\underline{h}| \ll \frac{1}{\det A} |A\underline{h}|,$$

where $\mathrm{adj}(A)_{i,j} = (-1)^{i+j} A_{j,i}$ so that $|\mathrm{adj}(A)| = |A| \ll 1$. Now, let $A = \widetilde{C}^{-1} Dg(\underline{x}_0)$. Then indeed $|A| \ll 1$, where the implied constant does not depend on $C$ or $\widetilde{C}$, but may depend on $\Lambda, n$ and $d$. Since $M = |\Delta(\underline{x}_0)| = |Dg(\underline{x}_0)|$, we find that for $\underline{x}_1, \underline{x}_2 \in U$ we have

$$|\widetilde{C}^{-1} Dg(\underline{x}_0)(\underline{x}_1 - \underline{x}_2)| \gg \det(\widetilde{C}^{-1} Dg(\underline{x}_0)) |\underline{x}_1 - \underline{x}_2| = \widetilde{C}^{-R} M |\underline{x}_1 - \underline{x}_2|.$$

Hence,

$$\begin{aligned} |g(\underline{x}_1) - Dg(\underline{x}_0) \cdot \underline{x}_1 - g(\underline{x}_2) + Dg(\underline{x}_0) \cdot \underline{x}_2| + |g(\underline{x}_1) - g(\underline{x}_2)| &\geq |Dg(\underline{x}_0)(\underline{x}_1 - \underline{x}_2)| \\ &\gg \widetilde{C}^{-R+1} M |\underline{x}_1 - \underline{x}_2|. \end{aligned}$$

Therefore, using (2.48) for $a$ small enough, we find for all $\underline{x}_1, \underline{x}_2 \in U$ that

$$|g(\underline{x}_1) - g(\underline{x}_2)| \gg \widetilde{C}^{-R+1} M |\underline{x}_1 - \underline{x}_2|.$$

This implies that if $\underline{x}$ is on the boundary of $U$ we have

$$|g(\underline{x}) - g(\underline{x}_0)| \gg \widetilde{C}^{-R+1} M |\underline{x} - \underline{x}_0| = a \frac{M^2}{\widetilde{C}^{2R-1} \Lambda^{R(d-1)-1}} \tag{2.49}$$

Set $b \gg a \frac{M^2}{\widetilde{C}^{2R-1} \Lambda^{R(d-1)-1}}$ so that for $\underline{x}$ and the boundary of $U$ it holds that $|g(\underline{x}) - g(\underline{x}_0)| \gg b$ and define

$$W' = \{\underline{y} \in \mathbb{R}^n : |\underline{y} - g(\underline{x}_0)| < \tfrac{1}{2} b\}.$$

Then by (2.49) we have that if $\underline{y} \in W'$ and $\underline{x}$ is on the boundary of $U$

$$|\underline{y} - g(\underline{x}_0)| < \tfrac{1}{2} b < |g(\underline{x}) - g(\underline{x}_0)| - |\underline{y} - g(\underline{x}_0)| \leq |\underline{y} - g(\underline{x})|. \tag{2.50}$$

We show that for all $\underline{y} \in W'$ there is a unique $\underline{x}$ in the interior of $U$ such that $g(\underline{x}) = \underline{y}$. For this, consider the function $h : U \to \mathbb{R}$ defined by

$$h(\underline{x}) = |\underline{y} - g(\underline{x})|^2 = \sum_{i=1}^{n} (y_i - g_i(\underline{x}))^2.$$

As $h$ is continuous, it attains a minimum on $U$ and from (2.50) it follows that this minimum does not occur on the boundary of $U$. Hence, a minimum $\underline{x}$ is obtained in the interior of $U$ and as all partial derivatives of $h$ exist this implies that $\frac{\partial h}{\partial x_j}(\underline{x}) = 0$ for all $1 \leq j \leq n$. That is, for all $1 \leq j \leq n$ we have

$$\sum_{i=1}^{n} 2(y_i - g_i(\underline{x})) \frac{\partial g_i}{\partial x_j}(\underline{x}) = 0.$$

Property (2) implies that $|Dg(\underline{x})| = |\Delta(\underline{x})| \neq 0$. Therefore, we find that $y_i = \tilde{f}_i(\underline{x})$ for all $1 \leq i \leq n$. By property (3) uniqueness of $\underline{x}$ follows. Let $W = \operatorname{int}(U) \cap g^{-1}(W')$. Similar as in the proof of Theorem 2.11 in [Spi65] or Lemma 9.3 in [PSW16] we conclude that $g : W \to W'$ has a differentiable inverse with $\det((g^{-1})') \geq M^{-1}$. □

**Theorem 2.43.** *Suppose that $\underline{x}_0 \in \mathbb{R}^n$ with $|\underline{x}_0| \leq \Lambda$ satisfies $\underline{\tilde{f}}(\underline{x}_0) = \underline{0}$ and $\Lambda \geq 1$ such that $M = \max_{I \subset [n], |I| = R} |\Delta_I(\underline{x}_0)| > 0$. Then, we have*

$$J(\underline{0}) \gg M^{-1} \left( \frac{M^2}{\widetilde{C}^{2R-1} \Lambda^{R(d-1)-1}} \right)^{n-R}.$$

*Proof.* Let $\mathbf{1}_{1/2t} : \mathbb{R} \to \{0, 1\}$ be the characteristic function of the interval $[-\frac{1}{2t}, \frac{1}{2t}]$. Let $W, W'$ as in Lemma 2.41. Then by Proposition 2.40 we have that

$$J(\underline{0}) \geq \lim_{t \to \infty} \left( \frac{t}{2} \right)^R \int_W \prod_{i=1}^{R} \mathbf{1}_{1/2t} \tilde{f}_i(\underline{x}) \, \mathrm{d}\underline{x}.$$

Applying the change of variables as in Proposition 2.40 we obtain

$$\int_W \prod_{i=1}^{R} \mathbf{1}_{1/2t} \tilde{f}_i(\underline{x}) \, \mathrm{d}\underline{x} = \int_{W'} |\det(((\tilde{f})^{-1})')| \prod_{i=1}^{R} \mathbf{1}_{1/2t}(y_i) \, \mathrm{d}\underline{y} \geq \int_{W'} M^{-1} \prod_{i=1}^{R} \mathbf{1}_{1/2t}(y_i) \, \mathrm{d}\underline{y}.$$

For $t$ sufficiently large, so that $\mathbf{1}_{1/2t} \equiv 1$ on $W'$, this is

$$\gg M^{-1} \frac{1}{t^R} \left( \frac{M^2}{\widetilde{C}^{2R-1} \Lambda^{R(d-1)-1}} \right)^{n-R}.$$

Therefore,

$$J(\underline{0}) \gg M^{-1} \left( \frac{M^2}{\widetilde{C}^{2R-1} \Lambda^{R(d-1)-1}} \right)^{n-R},$$

concluding the theorem. □

**Lemma 2.44.** *Let $\underline{x}_0 \in \mathbb{R}^n$ be such that $|\underline{x}_0| = 1$ and $\underline{\tilde{f}}(\underline{x}_0) = \underline{0}$. Take $j \in [n]$ such that $(x_0)_j = |\underline{x}_0| = 1$. Then, one has*

$$\max_I |\Delta_I(\underline{x}_0)| \gg \tilde{\mathfrak{C}}^{-1} \tilde{N}_j.$$

*Proof.* This is essentially the same proof as the proof of Lemma 2.37. Substitute $\underline{x} = \underline{x}_0$ in (2.38) where $j$ is such that $(x_0)_j = |\underline{x}_0| = 1$. Then the first sum vanishes and we find that

$$\max_I |\tilde{\Delta}_I(\underline{x}_0)| \max_I |\tilde{g}_I(\underline{x}_0)| \gg |\tilde{N}_j|.$$

As $|\underline{x}_0| = 1$, this implies that $g_I(\underline{x}_0) \ll \widetilde{\mathfrak{C}}$. This implies that

$$\max_I |\tilde{\Delta}_I(\underline{x}_0)| \gg \widetilde{\mathfrak{C}}^{-1} \tilde{N}_j. \qquad \square$$

**Corollary 2.45.** *Suppose $\widetilde{V}$ is non-singular and $\underline{\tilde{f}}$ has a non-singular real zero. Then*

$$J(\underline{0}) \gg \tilde{\mathfrak{C}}^{-(2(n-R)-1)} \widetilde{C}^{-2R+1} \tilde{N}^{2(n-R)-1}.$$

*Proof.* Observe that by homogeneity of $\underline{\tilde{f}}$ we can assume that the non-singular real zero $\underline{x}_0$ satisfies $|\underline{x}_0| = 1$. The corollary then follows directly from Theorem 2.43 and Lemma 2.44. $\square$

### 2.5.3 Main theorems

*Proof of Theorem 2.7.* From Theorem 2.33, it follows that for $P$ satisfying

$$P \gg \left( \frac{C \widetilde{C}^{K/(d-1)+R^2-1}}{\mathfrak{S}(\underline{0}) J(\underline{0})} \right)^{1/\delta}$$

we have that $M(P, \underline{0}) > 0$ (by letting the implied constant above large enough). By Proposition 2.39, Corollary 2.45, (2.37) and (2.39) it follows that

$$\begin{aligned} \mathfrak{S}(\underline{0}) J(\underline{0}) &\gg \tilde{\mathfrak{C}}^{-(2(n-R)-1)} \widetilde{C}^{-2R+1} \left( \frac{\tilde{N}^2}{N^3} \right)^{n-R} \tilde{N}^{-1} \\ &\gg \mathfrak{C}^{-3(n-R)} \tilde{\mathfrak{C}}^{-(2(n-R)-1)} \widetilde{C}^{-2R+1}. \end{aligned}$$

Using that $(n+1)(n-R) < n^2$, one finds that one can take

$$P = c(C^3 \widetilde{C}^2)^{4n^3 R (Rd)^n \cdot \frac{K+R(R+1)(d-1)}{K-R(R+1)(d-1)}}.$$

where $c$ is a constant not depending on $C$ and $\widetilde{C}$. Hence, for the above choice of $P$ there exists an integer zero $\underline{x}$ of $\underline{f}$ with $|\underline{x}| \leq P$. $\square$

*Remark.* The bound given above is believed to be far from optimal. One can find a slight improvement by considering more carefully which value of $P$ one can take in the above proof. However, due to the the large bounds one gets for $N$ and $\tilde{N}$ by applying the quantitative version of the Nullstellensatz, this bound cannot significantly be improved with the techniques used in this work.

We can do slightly better in case we add the assumption that the polynomials $\underline{f}$ are homogeneous:

**Theorem 2.46.** *Suppose $\tilde{f}_i \in \mathbb{Z}[\underline{x}]$ for $i = 1, \ldots, R$ are homogeneous polynomials of degree $d$ so that $K - R(R+1)(d-1) > 0$, $\underline{\tilde{f}}$ has a zeros over $\mathbb{Z}_p$ for all primes $p$ and a non-singular real zero. Assume that the corresponding projective variety $\widetilde{V}$ is non-singular. Then there exists an $\underline{x} \in \mathbb{Z}^n \backslash \{\underline{0}\}$, polynomially bounded by $C$ and $\widetilde{C}$, such that $\tilde{f}(\underline{x}) = \underline{0}$, namely*

$$|\underline{x}| \ll \widetilde{C}^{12n^3 R (Rd)^n \cdot \frac{K+R(R+1)(d-1)}{K-R(R+1)(d-1)}}.$$

*Proof.* As in the proof of Theorem 2.7 (with $C = \widetilde{C}$) we use that for $P$ satisfying

$$P \gg \left( \frac{\widetilde{C}^{K/(d-1)+R^2}}{\mathfrak{S}(\underline{0})J(\underline{0})} \right)^{1/\delta}$$

we have that $M(P, \underline{0}) > 0$ (by letting the implied constant above large enough). Moreover, the quantitative version of the Nullstellensatz for $\underline{\tilde{f}}$ given in (2.39) does still hold. Take $j$ in this equation such that for the integer zero $\underline{x}$ we have $x_j \neq 0$ and write $\tilde{N} = \tilde{N}_j$.

Considering $\underline{x}$ as a zero of $\underline{\tilde{f}}$ in $\mathbb{Z}_p$ for all primes $p$, one deduces with a proof similar to that of Proposition 2.39 that $\mathfrak{S}(\underline{0}) \geq \tilde{N}^{-3(n-R)}$. Together with Corollary 2.45 (where a scalar multiple of $\underline{x}$ is considered as the real zero), (2.37) and (2.39) it follows that

$$\begin{aligned} \mathfrak{S}(\underline{0})J(\underline{0}) &\gg \tilde{\mathfrak{C}}^{-(2(n-R)-1)}\widetilde{C}^{-2R+1}\tilde{N}^{-n+R-1} \\ &\gg \tilde{\mathfrak{C}}^{-(3(n-R)-2)}\widetilde{C}^{-2R+1}. \end{aligned}$$

One finds that one can take

$$P = c\widetilde{C}^{12n^3R(Rd)^n \cdot \frac{K+R(R+1)(d-1)}{K-R(R+1)(d-1)}}.$$

where $c$ is a constant not depending on $C$ and $\widetilde{C}$. □

*Remark.* In case of one homogeneous polynomial of degree $d = 3$ with $\dim \widetilde{V}^* = 0$, so that $K = n/4$, one has

$$|\underline{x}| \ll \widetilde{C}^{12n^3 3^n \cdot \frac{n+8}{n-8}}.$$

in the above theorem. This is visibly worse than the bound found in Theorem 1 of [BDE12].

As already indicated in Section 2.2.2, we are not only interested in integer zeros of $\underline{f}$, but also in integer zeros of $\underline{f}$ satisfying certain modulo conditions. This provides an answer to question (3) for $\underline{x}_0 = \underline{0}$.

**Theorem 2.47.** *Let $\underline{m}, \underline{M} \in \mathbb{N}^n$. Suppose $f_i \in \mathbb{Z}[\underline{x}]$ for $i = 1, \ldots, R$ are polynomials of degree $d$ so that $K - R(R+1)(d-1) > 0$ and the corresponding varieties $V$ and $\widetilde{V}$ are non-singular affine respectively projective varieties. Suppose that a zero $\underline{y} \in \mathbb{Z}_p$ of $\underline{f}$ satisfying $y_i \equiv m_i \mod M_i$ exists for every prime $p$ and suppose $\underline{\tilde{f}}$ has a real zero. Then, there exists an $\underline{x} \in \mathbb{Z}^n$, polynomially bounded by $C$ and $\widetilde{C}$, such that*

$$f(\underline{x}) = \underline{0} \quad \textit{and} \quad x_i \equiv m_i \mod M_i,$$

*and*

$$|\underline{x}| \ll (|\underline{M}|^{5d} C^3 \widetilde{C}^2)^{4n^3 R(Rd)^n \cdot \frac{K+R(R+1)(d-1)}{K-R(R+1)(d-1)}}.$$

*Proof.* Let

$$\underline{g}(\underline{y}) = \underline{f}(\underline{My} + \underline{m}) \quad \text{and} \quad \underline{\tilde{g}}(\underline{y}) = \underline{f}(\widetilde{\underline{My} + \underline{m}}) = \underline{\tilde{f}}(\underline{My}),$$

where the $i$th component of $\underline{My}$ is given by $(My)_i = M_i y_i$. Observe that over $\overline{\mathbb{Q}}$ we have that $\underline{f}$ or $\underline{\tilde{f}}$ is non-singular if and only if $\underline{g}$ respectively $\underline{\tilde{g}}$ is non-singular. Moreover, the condition on the existence of zeros of $\underline{f}$ ensures that $\underline{g}$ has zeros over $\mathbb{Z}_p$ for all primes $p$ and over $\mathbb{R}$. Hence, we can apply Theorem 2.7 to $\underline{g}$. As the maximal coefficient of $\underline{g}$ and $\underline{\tilde{g}}$ is $\ll |\underline{M}|^d C$, respectively $|\underline{M}|^d \widetilde{C}$, the theorem follows. □

# 3 The circle method and families of partitions

## 3.1 Partitions

The initial idea of the circle method is attributed to Hardy and Ramanujan in their work on partitions [HR18].

**Definition 3.1.** A *partition* $\lambda$ of a non-negative integer $n$ is a non-increasing sequence of non-negative integers $(\lambda_i)_{i\in\mathbb{N}}$ such that $\sum_{i=1}^{\infty}\lambda_i = n$. The non-zero $\lambda_i$ are called the *parts* of the partition.

For a partition of $r$ parts we write $\lambda = (\lambda_1, \ldots, \lambda_r)$ or simply $\lambda_1 + \ldots + \lambda_r$ instead of $(\lambda_1, \ldots, \lambda_r, 0, 0, \ldots)$. So, $(4, 2, 1, 1)$ and $4 + 2 + 1 + 1$ denote the same partition of 8. Let $p(n)$ be the number of partitions of $n$. For example, $p(4) = 5$, as the partitions of 4 are given by

$$4, \quad 3+1, \quad 2+2, \quad 2+1+1, \quad 1+1+1+1.$$

The number of partitions $p(n)$ grows rapidly with $n$, for example $p(10) = 42$, $p(100) = 190\ 569\ 292$ and $p(1000) = 24\ 061\ 467\ 864\ 032\ 622\ 473\ 692\ 149\ 727\ 991$. This is captured in the asymptotic formula of Hardy and Rumanujan

$$p(n) \sim \frac{\sqrt{3}}{12n} e^{\frac{\sqrt{6}\pi}{3}\sqrt{n}}, \quad n \to \infty. \tag{3.1}$$

Rademacher refined the use of circle method and provided an exact formula for $p(n)$, namely

**Theorem 3.2** ([Rad37, IK04])**.** *For $n \geq 1$ it holds that*

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{c=1}^{\infty} \sqrt{c} A_c(n) \frac{d}{dn} \frac{1}{\lambda_n} \sinh\left(\frac{B}{c}\lambda_n\right),$$

*where $B = \frac{1}{3}\sqrt{6}\pi, \lambda_n = (n - \frac{1}{24})^{\frac{1}{2}}$ and*

$$A_c(n) = \sum_{\substack{a \bmod c \\ (a,c)=1}} \omega_{a,c} e(-an/c),$$

*with $\omega_{a,c}$ a certain 24th root of unity whose exact value can be found in* [IK04, p. 450-451].

As often with series of integers, it is useful to consider its generating series

$$P(z) = \sum_{n=0}^{\infty} p(n) z^n = \prod_{n=1}^{\infty} \frac{1}{1 - z^n},$$

which was already studied by Euler. This generating series is related to the Dedekind eta function

$$\eta(z) = e(z/24) \prod_{n=1}^{\infty} (1 - e(nz)),$$

by $P(e(z)) = e(z/24)/\eta(z)$. One of the main ingredients in the proof of Rademacher's exact formula is the modularity of $\eta(z)$. To be more precise, it is used that $z \mapsto \frac{1}{\eta(z)}$ is a nearly holomorphic form of weight $-\frac{1}{2}$. A *nearly holomorphic modular form* satisfies the same properties as a modular form, except that the function is allowed to have poles at the cusps.

Ramanujan discovered the following congruences for the partition function:

$$\begin{aligned} p(5k+4) &\equiv 0 \mod 5, \\ p(7k+5) &\equiv 0 \mod 7, \\ p(11k+6) &\equiv 0 \mod 11, \end{aligned}$$

where $k \in \mathbb{Z}_{\geq 0}$. Hardy extracted proofs for these identities from an unpublished manuscript of Ramanujan employing Eisenstein series [Ram21]. Later, a combinatorial explanation for these identities was given by Andrews and Garvan [AG88]. They defined the *crank* of a partition (which was already conjectured to exist by Dyson) in the following way. For a partition $\lambda$, let $\omega(\lambda)$ denote the number of parts equal to 1 and $\nu(\lambda)$ be the number of parts larger than $\omega(\lambda)$. Then the crank $c(\lambda)$ of a partition $\lambda$ is given by

$$c(\lambda) = \begin{cases} \lambda_1 & \text{if } \omega(\lambda) = 0, \\ \nu(\lambda) - \omega(\lambda) & \text{else.} \end{cases}$$

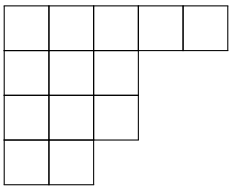
Letting $\mathcal{M}(m, q, n)$ be the number of partitions of $n$ with crank equal to $m$ modulo $q$, they showed that

$$\begin{aligned} \mathcal{M}(m, 5, 5k+4) &= \tfrac{1}{5}p(5k+4), \\ \mathcal{M}(m, 7, 7k+5) &= \tfrac{1}{7}p(7k+5), \\ \mathcal{M}(m, 11, 11k+6) &= \tfrac{1}{11}p(11k+6) \end{aligned}$$

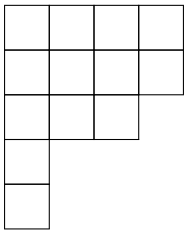
for all $m \in \mathbb{Z}$. Congruence identities for partitions have continued to attract much attention, see [AO05] for a survey. We will encounter congruence identities for so-called restricted overpartitions in Section 3.3.

## 3.2 Counting partitions

In this section we introduce and provide examples of some elementary notations in the theory of partitions, following [And98]. A useful tool in studying partitions is the graphical representation in a *Young tableau* or *Young diagram*. Given a partition $\lambda = (\lambda_1, \ldots, \lambda_n)$, such a diagram consists of $n$ rows with $\lambda_i$ boxes in row $i$. For example, the diagram



corresponds to the partition $5 + 3 + 3 + 2$ of 13. Note that the transpose of such a diagram (obtained by reflecting in the main diagonal) corresponds to a partition as well. For example, the transpose of the above diagram



corresponds to the partition $4 + 4 + 3 + 1 + 1$ of 13.

**Definition 3.3.** The *conjugate* $\lambda' = (\lambda'_i)_{i\in\mathbb{N}}$ of a partition $\lambda = (\lambda_i)_{i\in\mathbb{N}}$ is defined by letting $\lambda'_i$ be the number of parts of $\lambda$ that are $\geq i$.

Note that this definition corresponds to taking the transpose of the Young diagram. We can use this construction to prove the following lemma:

**Lemma 3.4.** *The number of partitions of $n$ with at most $m$ parts equals the number of partitions of $n$ in which no part exceeds $m$.*

*Proof.* The map from the set of all partitions of $n$ to itself given by conjugation is a bijection. The image of all partitions of $n$ with a most $m$ parts under this mapping consists of precisely all partitions in which no part exceeds $m$ and vice versa. Hence, we found a bijection between partitions of $n$ 'with at most $m$ parts' and 'in which no part exceeds $m$'. □

Denote with $p(\leq m, n)$ the number of partitions of $n$ with at most $m$ parts or, equivalently, the number of partitions of $n$ in which no part exceeds $m$. We calculate its generating series $\sum_{n=0}^{\infty} p(\leq m, n)q^n$. We introduce the $q$-Pochhammer symbol

$$(a, q)_n = \prod_{i=1}^{n}(1 - aq^{i-1}),$$

for $n \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ and write $(q)_n$ for $(q, q)_n$. Here, $(a, q)_0$ is considered to be 1. Note that in this notation

$$\sum_{n=0}^{\infty} p(n)q^n = \frac{1}{(q)_\infty}.$$

We consider above and other identities involving infinite sums of powers of $q$ as formal sums with $q$ (and later also $x$) a formal parameter. However, often it is not hard to show that such identities hold as well as if we consider $q$ (or $x$) to be a complex number with $|q| < 1$.

**Lemma 3.5.** *For all $m \in \mathbb{Z}_{\geq 0}$ we have*

$$\sum_{n=0}^{\infty} p(\leq m, n)q^n = \frac{1}{(q)_m}.$$

*Proof.* We have

$$\begin{aligned}\sum_{n=0}^{\infty} p(\leq m, n)q^n &= \sum_{a_1 \geq 0, \ldots, a_m \geq 0} q^{a_1 + a_2 \cdot 2 + \ldots + a_m \cdot m} \\ &= (1 + q + q^2 + \ldots) \cdots (1 + q^m + q^{2m} + \ldots) \\ &= \prod_{i=1}^{m} \frac{1}{1 - q^i} = \frac{1}{(q)_m}.\end{aligned}$$

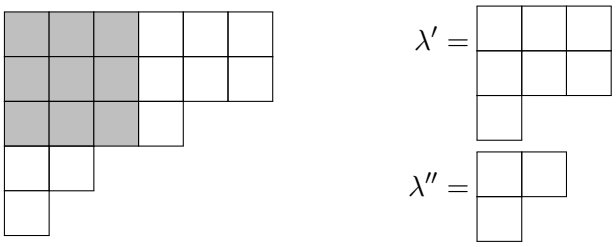
□

We conclude this section with an introduction to the rank of partition.

**Definition 3.6.** For a partition $\lambda$ we define the *rank* $d(\lambda)$ as the number of $\lambda_j$ such that $\lambda_j \geq j$.

This definition should not be confused with the rank of a partition defined by Dyson. He defined the rank of a partition as the largest part minus the number of parts. Both definitions appear in the literature. In this thesis the rank of a partition is always as in Definition 3.6.

In the Young diagram the rank is related to the largest square contained in the partition. For example, for the partition $(6, 6, 4, 2, 1)$ the rank equals 3 and the largest square has size $3 \times 3$.




The largest square in the Young diagram (placed in the top left) is called the *Durfee square* and $d(\lambda)$ is the length of the diagonal (or side) of this square. Observe that a partition $\lambda$ is uniquely determined by its rank, the partition $\lambda'$ to the right of the Durfee square and the partition $\lambda''$ below the Durfee square. Denote with $p(n)_d$ the number of partitions $\lambda$ with $d(\lambda) = d$.

**Lemma 3.7.** *For all $d \geq 0$ we have*

$$\sum_{n=0}^{\infty} p(n)_d = \frac{q^{d^2}}{(q)_d^2}. \tag{3.2}$$

*Proof.* By the above observation the number of partitions $\lambda$ of rank $d$ equals the product of the number of partitions $\lambda'$ and $\lambda''$, where $\lambda'$ is a partition in at most $d$ parts (corresponding to the squares next to the Durfee square) and $\lambda''$ is a partition whose parts are $\leq d$ (corresponding to the squares below the Durfee square). Since the generating series of both $\lambda'$ and $\lambda''$ are given by $\frac{1}{(q)_d}$ by Lemma 3.5, we find that

$$\sum_{n=0}^{\infty} p(n)_d = \frac{q^{d^2}}{(q)_d^2}.$$

Here, the factor $q^{d^2}$ corresponds to the $d^2$ squares of the Durfee square. □

Summing (3.2) over all possible ranks $d$, we obtain the generating series of all partitions. Hence, we have the following identity:

**Corollary 3.8.**

$$\sum_{d\geq 0} \frac{q^{d^2}}{(q)_d^2} = \frac{1}{(q)_\infty}.$$

## 3.3 Restricted overpartitions

We now study overpartitions with restricted odd differences, which were introduced in [BDLM15], mostly for the study of so-called mixed mock modular forms. Overpartitions appear in combinatorial proofs of $q$-series identities and in the study of hypergeometric series. A short history where overpartitions appear and a discussion about the structure of overpartitions can be found in [CL04].

**Definition 3.9.** An *overpartition* of a non-negative integer $n$ is a pair $(\lambda, \mu)$ where $\lambda$ is a partition of $n$ and $\mu \in \{0,1\}^{\mathbb{N}}$ such that $\mu_i = 0$ if $\lambda_i = \lambda_{i+1}$. An *overpartition with restricted odd differences* is an overpartition of $n$ where $\lambda_i - \lambda_{i+1}$ is odd only if $\mu_i = 1$.

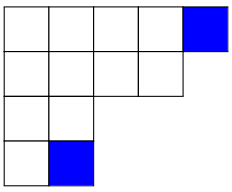
Given an overpartition $(\lambda, \mu)$ we say that $\lambda_i$ is overlined if $\mu_i = 1$. We denote such a part overlined. Note that $\lambda_i$ can only be overlined if $\lambda_i$ is non-zero. One can think of overpartitions as partitions in which the final occurrence of a number may be overlined. For example, the eight overpartitions of 3 are given by

$$3,\ \overline{3},\ 2+1,\ \overline{2}+1,\ 2+\overline{1},\ \overline{2}+\overline{1},\ 1+1+1,\ 1+1+\overline{1}.$$

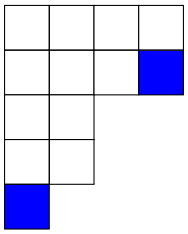
Of those, the overpartitions

$$\overline{3},\ \overline{2}+\overline{1},\ 1+1+\overline{1}$$

have restricted odd differences. To an overpartition we associate a Young diagram where the last cell of each row corresponding to an overlined part is coloured. For example, the diagram



corresponds to the overpartition $\overline{5}+4+2+\overline{2}$. Note that this is an overpartition with restricted odd differences. The transpose of this diagram is the overpartition $4+\overline{4}+2+2+\overline{1}$, which is not an overpartition with restricted odd differences:



This leads to the definition of the conjugate of an overpartition, which generalises Definition 3.3.

**Definition 3.10.** The *conjugate* $(\lambda', \mu')$ of an overpartition $(\lambda, \mu)$ is defined by letting $\lambda'$ be the conjugate of $\lambda$ and $\mu'_i = 1$ precisely if there is a $j$ such that $\lambda_j = i$ and $\mu_j = 1$ (and else $\mu'_i = 0$).

Note that this definition corresponds to taking the transpose of the diagram associated to an overpartition, as illustrated above.

Let $t(n)$ be the number of overpartitions with restricted odd differences of a non-negative integer $n$. Observe that the conjugate of a partition counted by $t(n)$ is an overpartition such that if $\overline{m}$ does not occur as a part, then $m$ occurs an even number of times. Hence,

$$\sum_{n=0}^{\infty} t(n)q^n = \prod_{m\geq 1}\left(\frac{q^m}{1-q^m} + \frac{1}{1-q^{2m}}\right) = \frac{(q^3,q^3)_\infty}{(q)_\infty (q^2,q^2)_\infty}.$$

Here the term $\frac{q^m}{1-q^m}$ corresponds to the parts of size $m$ in case $m$ occurs overlined in the overpartition and $\frac{1}{1-q^{2m}}$ corresponds to the parts of size $m$ where $m$ does not occur overlined. By a generalisation of Rademacher's refinement of the circle method [RZ38], for $n \to \infty$ one has

$$t(n) \sim \frac{\sqrt{21}}{36n} e^{\frac{\sqrt{7}\pi}{3}\sqrt{n}},$$

as mentioned in [BDLM15, p. 4]. Note that this is a faster growth than the growth of $p(n)$ in (3.1) as partitions correspond to a strict subset of restricted overpartitions. Namely, overlining the last occurence of every number in a partition gives an overpartition with restricted odd differences. Specifically, one has

$$\frac{t(n)}{p(n)} \sim \frac{\sqrt{7}}{3} e^{\frac{\pi(\sqrt{7}-\sqrt{6})}{3}\sqrt{n}}$$

as $n \to \infty$.

It turns out that $t(n)$ satisfies a congruence identity modulo 3.

**Lemma 3.11** ([BDLM15, Corollary 3])**.** *For $n \geq 1$ we have*

$$t(n) \equiv \begin{cases} (-1)^{k+1} \mod 3 & \text{if } n = k^2 \text{ for some } k \in \mathbb{N}, \\ 0 \mod 3 & \text{else.} \end{cases}$$

*Proof.* Note that

$$\frac{1-q^{3n}}{(1-q^n)(1-q^{2n})} = \frac{1+q^n+q^{2n}}{1-q^{2n}} \equiv \frac{1-2q^n+q^{2n}}{1-q^{2n}} = \frac{1-q^n}{1+q^n} \mod 3.$$

Hence,

$$\sum_{n=0}^{\infty} t(n)q^n = \frac{(q^3,q^3)}{(q)_\infty (q^2,q^2)_\infty} \equiv \frac{(q)_\infty}{(-q,q)_\infty} = 1 + 2\sum_{n\geq 1} (-1)^n q^{n^2} \mod 3.$$

The last equality is Gauss' identity [And98, Equation (2.2.12)]. □

In the definition of overpartitions with restricted odd differences there is nothing special about *odd differences*. More generally, one can define restricted overpartitions:

**Definition 3.12.** For a positive integer $N \in \mathbb{N}$ and a set $C \subseteq \{1, 2, \ldots, N-1\}$ let $t_N^C(n)$ count the number of overpartitions $\lambda$ of $n$ where for all $c \in C$ it holds that $\lambda_i - \lambda_{i+1} \equiv c \mod N$ only if $\mu_i = 1$. Define *restricted overpartitions* with parameters $C$ and $N$ as overpartitions counted by $t_N^C(n)$.

Note that $t_2^{\{1\}}(n) = t(n)$ and that $t_N^{\emptyset}(n)$ (for arbitrary $N \in \mathbb{N}$) counts all overpartitions of $n$. Denote with $C'$ the union of $\{0\}$ and the complement of $C$ in $\{1, 2 \ldots, N-1\}$. Then by a similar argument as above we have

$$\sum_{n=0}^{\infty} t_N^C(n)q^n = \prod_{m\geq 1} \left( \frac{q^m}{1-q^m} + \sum_{c\in C'} \frac{q^{cm}}{1-q^{Nm}} \right).$$

For two choices of $(C, N)$ we found congruences identities similar to Lemma 3.11.

**Conjecture 3.13.** For $n \geq 1$ we have

$$t_3^{\{1,2\}}(n) \equiv \begin{cases} 1 \mod 2 & \text{if } n = k^2 \text{ for some } k \not\equiv 0 \mod 3, \\ 0 \mod 2 & \text{else.} \end{cases}$$

This conjecture has been checked for all $n \leq 1000$.

**Lemma 3.14.** *For $n \geq 1$ we have*

$$t_1^{\emptyset}(n) \equiv \begin{cases} 2 \mod 4 & \text{if } n = k^2 \text{ for some } k \in \mathbb{N}, \\ 0 \mod 4 & \text{else.} \end{cases}$$

*Proof.* Note that

$$\sum_{n=0}^{\infty} t_1^{\emptyset}(n) q^n = \prod_{m \geq 1} \left( \frac{q^m}{1-q^m} + \frac{1}{1-q^m} \right) = \frac{(-q,q)_\infty}{(q)_\infty} = \frac{1}{1 + 2\sum_{n=1}^{\infty} (-1)^n q^{n^2}}.$$

Here, the last equality is Gauss' identity [And98, Equation (2.2.12)]. As already worked out in [Mah04], we have that

$$\begin{aligned} \frac{1}{1 + 2\sum_{n=1}^{\infty} (-1)^n q^{n^2}} &= 1 + \sum_{m=1}^{\infty} \left( -2 \sum_{n=1}^{\infty} (-1)^n q^{n^2} \right)^m \\ &= 1 + \sum_{m=1}^{\infty} (-2)^m \sum_{n_1, \ldots, n_m = 1}^{\infty} (-1)^{n_1 + \ldots + n_m} q^{n_1^2 + \ldots + n_m^2} \\ &= 1 + \sum_{m=1}^{\infty} 2^m \sum_{n=1}^{\infty} (-1)^{n+m} c_m(n) q^n, \end{aligned}$$

where $c_m(n)$ denotes the number of representations of $n$ as a sum of $m$ squares of positive integers. Note that

$$c_1(n) = \begin{cases} 1 & n = k^2 \\ 0 & \text{else.} \end{cases}$$

Hence,

$$\sum_{n=0}^{\infty} t_1^{\emptyset}(n) q^n \equiv 1 + 2 \sum_{n=1}^{\infty} q^{n^2} \mod 4. \qquad \square$$

One can wonder whether these congruence identities in above two lemmata and conjecture can be deduced from an explicit $3,2$- respectively 4-fold symmetry for the overpartitions counted by $t_N^C(n)$. As mentioned before, for the Ramanujan congruences this is indeed the case as can be shown using the crank of a partition. The authors of [BDLM15] raise this question for Lemma 3.11. This is still an open question. We prove a stronger result than Lemma 3.11. Let $t(n)_d$ count the number of overpartions with restricted odd differences and rank $d$. Here the rank of an overpartition $(\lambda, \mu)$ is the rank of $\lambda$.

**Theorem 3.15.** *For all $n \geq 1, d \geq 1$ we have*

$$t(n)_d \equiv \begin{cases} (-1)^{d+1} \mod 3 & \text{if } n = d^2 \\ 0 \mod 3 & \text{else.} \end{cases}$$

We give a proof at the end of this section, after proving a few useful lemmata.

Let $r(m,n)$ be the number of conjugate restricted overpartitions of $n$ in precisely $m$ parts. That is, $r(m,n)$ counts the overpartitions of $n$ in $m$ parts where if $\bar{l}$ does not occur as a part, then $l$ occurs an even number of times. First we express $t(n)_d$ in terms of $r(l,n)$ for $l \leq d$:

**Lemma 3.16.** *For all $d \geq 1$ we have*

$$\sum_{n=0}^{\infty} t(n)_d q^n = \frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}} \sum_{n=0}^{\infty} \left( \frac{1-q^d}{q^d} r(d,n) + \frac{3q^d}{1-q^{2d}} \sum_{l=0}^{d} r(l,n) \right) q^n.$$

*Proof.* Let $t(m,n,l)$ be the number of restricted overpartitions of $n$ consisting of $m$ parts and with largest part of size $l$ (here $m$ is either an integer or of the form $< i$ for some integer $i$ in which case there are $< i$ parts). We use a dot if we do not want to specify one of these values, e.g. $t(m,n,\cdot)$ counts the number of restricted overpartitions of $n$ consisting of $m$ parts and arbitrary largest part. We use an upper/lower $+$ and $-$ to denote that the largest/smallest part is even respectivically odd. For example, $t^+(m,n,l)$ counts the overpartitions counted by $t(m,n,l)$ with largest part even and $t_-(m,n,l)$ counts the overpartitions counted by $t(m,n,l)$ with smallest part odd.

The rough idea of this proof is the same as the proof of Lemma 3.7. Given $n, d$ we count the number of possible restricted overpartitions $\lambda'$ next to the Durfee square and $\lambda''$ below this square. However, we have to distinguish a few cases because the overlining of $\lambda_d = d + \lambda'_d$ in $\lambda$ depends on $\lambda_{d+1} = \lambda''_1$. For example, assume $d$ and $\lambda'_d$ are odd. Then $\lambda'_d$ is overlined in $\lambda'$. Note that $\lambda_d = d + \lambda'_d$ is even. Hence, if $\lambda''_1$ is even, $\lambda_d$ may or may not be overlined. However, if $\lambda''_1$ is odd, then $\lambda_d$ should be overlined.

Hence, assuming $d$ is odd, we find that $t(n)_d$ equals

$$\begin{aligned}
&\sum_{m=0}^{\infty} t_-(d,m,\cdot) \cdot \left(2t^+(\cdot, n-m-d^2, \leq d) + t^-(\cdot, n-m-d^2, \leq d)\right) + \\
&\sum_{m=0}^{\infty} \tfrac{1}{2} t_+(d,m,\cdot) \cdot \left(t^+(\cdot, n-m-d^2, \leq d) + 2t^-(\cdot, n-m-d^2, \leq d)\right) + \\
&\sum_{m=0}^{\infty} t(<d,m,\cdot) \cdot \left(t^+(\cdot, n-m-d^2, \leq d) + 2t^-(\cdot, n-m-d^2, < d) + t^-(\cdot, n-m-d^2, d)\right).
\end{aligned}$$

If $d$ is even and positive, then $t(n)_d$ equals

$$\begin{aligned}
&\sum_{m=0}^{\infty} t_-(d,m,\cdot) \cdot \left(2t^-(\cdot, n-m-d^2, \leq d) + t^+(\cdot, n-m-d^2, \leq d)\right) + \\
&\sum_{m=0}^{\infty} \tfrac{1}{2} t_+(d,m,\cdot) \cdot \left(t^-(\cdot, n-m-d^2, \leq d) + 2t^+(\cdot, n-m-d^2, \leq d)\right) + \\
&\sum_{m=0}^{\infty} t(<d,m,\cdot) \cdot \left(t^-(\cdot, n-m-d^2, \leq d) + 2t^+(\cdot, n-m-d^2, < d) + t^+(\cdot, n-m-d^2, d)\right).
\end{aligned}$$

Note that an overpartition counted by $t_-(d,m,\cdot)$ corresponds under conjugation to an overpartition with the largest part equal to $d$ such that there are an odd number of parts of size $d$ and if $\overline{m}$ does not occur as a part, then $m$ occurs on even number of times. Hence,

$$\sum_{m=0}^{\infty} t_-(d,m,\cdot) q^m = \frac{q^d}{1-q^{2d}} \cdot \prod_{i=1}^{d-1} \left( \frac{q^i}{1-q^i} + \frac{1}{1-q^{2i}} \right) = \frac{q^d}{1-q^{2d}} \frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}.$$

Similarly, an overpartition counted by $t_-(d,m,\cdot)$ corresponds under conjugation to an overpartition with the largest part equal to $d$ such that there are an *even* number of parts of size $d$ and

if $\overline{m}$ does not occur as a part, then $m$ occurs on even number of times. Hence,

$$\sum_{m=0}^{\infty} t_+(d,m,\cdot)q^m = \frac{2q^{2d}}{1-q^{2d}} \cdot \prod_{i=1}^{d-1}\left(\frac{q^i}{1-q^i}+\frac{1}{1-q^{2i}}\right) = \frac{2q^{2d}}{1-q^{2d}}\frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}.$$

Counting overpartitions with the largest part at most $d-1$ such that if $\overline{m}$ does not occur as a part, then $m$ occurs on even number of times, we find

$$\sum_{m=0}^{\infty} t(<d,m,\cdot)q^m = \prod_{i=1}^{d-1}\left(\frac{q^i}{1-q^i}+\frac{1}{1-q^{2i}}\right) = \frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}.$$

Replacing overpartitions by their conjugate we find that

$$\sum_{m=0}^{\infty} t^+(\cdot,m,\leq d)q^m = \sum_{m=0}^{\infty}\sum_{\substack{l\geq 0\\ \text{even}}}^{d} r(l,m)q^m, \qquad \sum_{m=0}^{\infty} t^-(\cdot,m,\leq d)q^m = \sum_{m=0}^{\infty}\sum_{\substack{l\geq 0\\ \text{odd}}}^{d} r(l,m)q^m,$$

$$\sum_{m=0}^{\infty} t^-(\cdot,m,< d)q^m = \sum_{m=0}^{\infty}\sum_{\substack{l\geq 0\\ \text{odd}}}^{d-1} r(l,m)q^m, \qquad \sum_{m=0}^{\infty} t^-(\cdot,m,d)q^m = \sum_{m=0}^{\infty} r(d,m)q^m.$$

Hence, for all $d>0$ we have

$$\begin{aligned}\sum_{n=0}^{\infty} t(n)_d q^n =& q^{d^2}\sum_{n=0}^{\infty}\frac{q^d}{1-q^{2d}}\frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\left(2\sum_{\substack{l\geq 0\\ l\not\equiv d(2)}}^{d-1} r(l,n)q^n + \sum_{\substack{l\geq 0\\ l\equiv d(2)}}^{d} r(l,n)q^n\right)+\\ & q^{d^2}\sum_{n=0}^{\infty}\frac{q^{2d}}{1-q^{2d}}\frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\left(\sum_{\substack{l\geq 0\\ l\not\equiv d(2)}}^{d-1} r(l,n)q^n + 2\sum_{\substack{l\geq 0\\ l\equiv d(2)}}^{d} r(l,n)q^n\right)+\\ & q^{d^2}\sum_{n=0}^{\infty}\frac{(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\left(\sum_{\substack{l\geq 0\\ l\not\equiv d(2)}}^{d-1} r(l,n)q^n + 2\sum_{\substack{l\geq 0\\ l\equiv d(2)}}^{d-2} r(l,n)q^n + r(d,n)q^n\right).\end{aligned}$$

Let $a$ be the number of 1's occurring in a partition $\lambda$ counted by $r(d,n)$. Subtracting 1 from each part $\lambda_i$, we obtain a partition counted by $r(d-a,n-d)$. Note that if $a$ is even and positive, the 1's in $\lambda$ may be both overlined or not overlined, whereas in the other cases the 1's are necessarily overlined or do no occur as parts. Hence, we find a $2:1$ map from partitions counted by $r(d,n)$ to partitions counted by $r(d-a,n-d)$ if $a$ is even and positive and a $1:1$ map from partitions counted by $r(d,n)$ to partitions counted by $r(d-a,n-d)$ else. Hence,

$$r(d,n) = r(d,n-d) + 2\sum_{a>0 \text{ even}} r(d-a,n-d) + \sum_{a \text{ odd}} r(d-a,n-d).$$

So, for $d>0$ we have

$$\begin{aligned}\sum_{n=0}^{\infty} r(d,n)q^n &= \sum_{n=0}^{\infty}\left(r(d,n)q^{n+d} + 2\sum_{\substack{a>0\\ \text{even}}} r(d-a,n)q^{n+d} + \sum_{\substack{a>0\\ \text{odd}}} r(d-a,n)q^{n+d}\right)\\ &= \sum_{n=0}^{\infty}\left(r(d,n)q^{n+d} + 2\sum_{\substack{l=1\\ l\equiv d(2)}}^{d-2} r(l,n)q^{n+d} + \sum_{\substack{l=0\\ l\not\equiv d(2)}}^{d-1} r(l,n)q^{n+d}\right).\end{aligned}$$

We conclude that

$$\sum_{n=0}^{\infty} r(d,n)q^n = \frac{q^d}{1+q^d}\left(\sum_{\substack{l=0\\ l\not\equiv d(2)}}^{d-1} r(l,n)q^n + 2\sum_{\substack{l=1\\ l\equiv d(2)}}^{d} r(l,n)q^n\right).$$

Hence, we have for all $d>0$ that $\sum_{n=0}^{\infty} t(n)_d q^n$ equals

$$\begin{aligned}
&\frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\left(-\frac{q^d}{1-q^{2d}}\frac{1+q^d}{q^d}+\frac{q^{2d}}{1-q^{2d}}\frac{1+q^d}{q^d}+\frac{1+q^d}{q^d}-1\right)\sum_{n=0}^{\infty} r(d,n)q^n + \\
&\quad + 3\frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\frac{q^d}{1-q^{2d}}\sum_{n=0}^{\infty}\sum_{l=0}^{d} r(l,n)q^n \\
&= \frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\left(\left(\frac{-1}{1-q^d}+\frac{q^d}{1-q^d}+q^{-d}\right)\sum_{n=0}^{\infty} r(d,n)q^n + \frac{3q^d}{1-q^{2d}}\sum_{n=0}^{\infty}\sum_{l=0}^{d} r(l,n)q^n\right) \\
&= \frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\sum_{n=0}^{\infty}\left(\frac{1-q^d}{q^d}r(d,n)+\frac{3q^d}{1-q^{2d}}\sum_{l=0}^{d} r(l,n)\right)q^n. \qquad \square
\end{aligned}$$

**Corollary 3.17.**

$$\sum_{n=0}^{\infty} t(n)_1 q^n = q+3q^2+3q^3+6q^4+6q^5+9q^6+9q^7+\ldots+3\left\lfloor\frac{n}{2}\right\rfloor q^n+\ldots$$

*Proof.* Observe that $r(0,n)=\delta_{0n}$ and $r(1,n)=1-\delta_{0n}$. Hence,

$$\begin{aligned}
\sum_{n=0}^{\infty} t(n)_1 q^n &= q\frac{3q}{1-q^2} + q\sum_{n\geq 1}\left(\frac{1-q}{q}+\frac{3q}{1-q^2}\right)q^n \\
&= \frac{3q^2}{1-q^2} + \frac{q^3+2q^2-q+1}{1-q^2}\frac{q}{1-q} \\
&= \frac{q\left(q^3-q^2+2q+1\right)}{(1-q)\left(1-q^2\right)} \\
&= \frac{3}{2}\left(q+q^2\right)\frac{d}{dq}\frac{q^2}{1-q^2} \\
&= q+3q^2+3q^3+6q^4+6q^5+9q^6+9q^7+\ldots+3\left\lfloor\frac{n}{2}\right\rfloor q^n+\ldots. \qquad \square
\end{aligned}$$

**Lemma 3.18.** *We have the following identities*

$$\begin{aligned}
&\sum_{m,n\geq 0} r(m,n)x^m q^n = \frac{(x^3q^3;q^3)_\infty}{(xq;q)_\infty}, \\
&\sum_{n=0}^{\infty} r(d,n)q^n \equiv (-1)^{d-1}\frac{q^d}{1-q^d}\frac{(q^2;q^2)_{d-1}}{(q)_{d-1}^2} \mod 3.
\end{aligned}$$

*Proof.* For overpartitions counted by $r(m,n)$ we have that if $\overline{k}$ does not occur, $k$ occurs an even number of times. Hence,

$$\sum_{m,n\geq 0} r(m,n)x^m q^n = \prod_{k\geq 1}\left(\frac{xq^k}{1-xq^k}+\frac{1}{1-x^2q^{2k}}\right) = \frac{(x^3q^3;q^3)_\infty}{(xq;q)_\infty(x^2q^2;q^2)_\infty}.$$

Note that

$$\frac{(x^3q^3;q^3)_\infty}{(xq;q)_\infty(x^2q^2;q^2)_\infty} \equiv \frac{(xq,q)_\infty}{(-xq,q)_\infty} \mod 3$$

by a similar argument as in the proof of Lemma 3.11. Now,

$$\frac{(xq;q)_\infty}{(-xq;q)_\infty} = 1 + \sum_{m=1}^{\infty} \frac{(-1;q)_m}{(q)_m}(-xq)^m = 1 + 2\sum_{m=1}^{\infty}(-1)^m \frac{q^m}{1-q^m}\frac{(-q;q)_{m-1}}{(q)_{m-1}}x^m$$

by Theorem 2.1 in [And98] with $t = -xq, a = -1$, so that for $d > 0$ we have

$$\sum_{n=0}^{\infty} r(d,n)q^n \equiv 2(-1)^d \frac{q^d}{1-q^d}\frac{(-q;q)_{d-1}}{(q)_{d-1}} \equiv (-1)^{d-1}\frac{q^d}{1-q^d}\frac{(q^2;q^2)_{d-1}}{(q)_{d-1}^2} \mod 3. \qquad \square$$

*Proof of Theorem 3.15.* Reducing the equality of Lemma 3.16 modulo 3 we find that

$$\sum_{n=0}^{\infty} t(n)_d q^n \equiv \frac{q^{d^2}(q^3;q^3)_{d-1}}{(q)_{d-1}(q^2;q^2)_{d-1}}\frac{1-q^d}{q^d}\sum_{n=0}^{\infty} r(d,n)q^n \equiv \frac{q^{d^2}(q)_{d-1}^2}{(q^2;q^2)_{d-1}}\frac{1-q^d}{q^d}\sum_{n=0}^{\infty} r(d,n)q^n \mod 3.$$

Substituting the second identity of Lemma 3.18 we find for $d > 0$ that

$$\sum_{n=0}^{\infty} t(n)_d q^n \equiv (-1)^{d-1}q^{d^2} \mod 3. \qquad \square$$

In further research one could study the double generating series

$$T(q,x) := \sum_{n,d\geq 0} t(n)_d q^n x^{d^2}.$$

Generating series comparable to this one are studied in [BDLM15]. Note that $T(q,1)$ is the generating series for the number of overpartitions with restricted odd differences, whereas $T(q,-1)$ is the generating series for overpartitions with restricted odd differences with even rank minus the number of such overpartitions with odd rank. By Theorem 3.15 one has

$$T(q,x) \equiv 1 + \sum_{d=1}^{\infty}(-1)^{d-1}(xq)^{d^2} \equiv 1 + 2\sum_{d=1}^{\infty}(-1)^d(xq)^{d^2} = \frac{(xq,xq)_\infty}{(-xq,xq)_\infty} \mod 3.$$

# 4 Concluding remarks

In this thesis we discussed two problems related to the circle method. In Chapter 2 we explained how the circle method can be used to prove the existence of integer zeros of a system of integral polynomials. The original contributions are bounds on the smallest integer zero in Theorem 2.7. These bounds hold for all systems of polynomials of all degrees provided the same conditions as in the work of Birch and that the corresponding varieties are non-singular. In order to establish these bounds asymptotics for the number of such zeros in Theorem 2.33, which are explicit in terms of the coefficients of the polynomials, have been found. It would be interesting to provide these bounds also in cases where the corresponding varieties are singular. However, our approach falls short in this case, because we cannot use the Nullstellensatz to control the singular series and -integral. We have not answered all questions raised in Section 2.2.2, in particular we have not answered question (2) for $\underline{x} \neq \underline{0}$: is it possible to find an upper bound on $\min_{\underline{x} \in V(\underline{\nu}) \cap \mathbb{Z}^n} |\underline{x} - \underline{x}_0|$ which is stronger than the one obtained by using the triangle equality on $|\underline{x} - \underline{x}_0|$?

In Chapter 3 we introduced the counting function $t(n)$ for overpartitions with restricted odd differences. For such partitions and related partitions counted by $t_N^C(n)$ we found a few congruence identities. The original contribution is a congruence identity modulo 3 for overpartitions with restricted odd differences of a given rank, counted by $t(n)_d$. See Theorem 3.15 for a more precise statement. It would be interesting to investigate whether a similar result holds for the other two congruence identities. Moreover, the proofs of these congruence identities depend on manipulating generating series. It would be interesting to find a combinatorial explanation of these results.

# Bibliography


[AG88] G. E. Andrews and F. G. Garvan. Dyson's crank of a partition. *Bull. Amer. Math. Soc. (N.S.)*, 18(2):167–171, 1988.

[AM02] A. G. Aleksandrov and B. Z. Moroz. Complete intersections in relation to a paper of B. J. Birch. *Bull. London Math. Soc.*, 34(2):149–154, 2002.

[And98] G. E. Andrews. *The theory of partitions*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1998. Reprint of the 1976 original.

[AO05] G. E. Andrews and K. Ono. Ramanujan's congruences and Dyson's crank. *Proc. Natl. Acad. Sci. USA*, 102(43):15277, 2005.

[BDE12] T. D. Browning, R. Dietmann, and P. D. T. A. Elliott. Least zero of a cubic form. *Math. Ann.*, 352(3):745–778, 2012.

[BDLM15] K. Bringmann, J. Dousse, J. Lovejoy, and K. Mahlburg. Overpartitions with restricted odd differences. *Electron. J. Combin.*, 22(3), 2015.

[BHB17] T. Browning and R. Heath-Brown. Forms in many variables and differing degrees. *J. Eur. Math. Soc. (JEMS)*, 19(2):357–394, 2017.

[Bir62] B. J. Birch. Forms in many variables. *Proc. Roy. Soc. Ser. A*, 265:245–263, 1961/1962.

[CL04] S. Corteel and J. Lovejoy. Overpartitions. *Trans. Amer. Math. Soc.*, 356(4):1623–1635, 2004.

[Dav59] H. Davenport. Cubic forms in thirty-two variables. *Philos. Trans. Roy. Soc. London. Ser. A*, 251:193–232, 1959.

[Dav63] H. Davenport. Cubic forms in sixteen variables. *Proc. Roy. Soc. Ser. A*, 272:285–303, 1963.

[Dav05] H. Davenport. *Analytic methods for Diophantine equations and Diophantine inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2005.

[GBGL10] T. Gowers, J. Barrow-Green, and I. Leader. *The Princeton Companion to Mathematics*. Princeton University Press, 2010.

[Gre69] M. J. Greenberg. *Lectures on forms in many variables*. W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[HB07] D. R. Heath-Brown. Cubic forms in 14 variables. *Invent. Math.*, 170(1):199–230, 2007.

[Hoo88] C. Hooley. On nonary cubic forms. *J. Reine Angew. Math.*, 386:32–98, 1988.

[HR18] G. H. Hardy and S. Ramanujan. Asymptotic Formulaae in Combinatory Analysis. *Proc. London Math. Soc.*, S2-17(1):75, 1918.

[IK04] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[KC02] V. Kac and P. Cheung. *Quantum calculus*. Universitext. Springer-Verlag, New York, 2002.

[KPS01] T. Krick, L. M. Pardo, and M. Sombra. Sharp estimates for the arithmetic Nullstellensatz. *Duke Math. J.*, 109(3):521–598, 2001.

[Mah04] K. Mahlburg. The overpartition function modulo small powers of 2. *Discrete Math.*, 286(3):263–267, 2004.

[Pit68] J. Pitman. Cubic inequalities. *J. London Math. Soc.*, 43:119–126, 1968.

[PSW16] L. B. Pierce, D. Schindler, and M. M. Wood. Representations of integers by systems of three quadratic forms. *Proc. Lond. Math. Soc. (3)*, 113(3):289–344, 2016.

[Rad37] H. A. Rademacher. Convergent Series for the Partition Function $p(n)$. *Proc. Natl. Acad. Sci. USA*, 23(2):78–84, 1937.

[Ram21] S. Ramanujan. Congruence properties of partitions. *Math. Z.*, 9(1-2):147–153, 1921.

[RZ38] H. Rademacher and H. S. Zuckerman. On the Fourier coefficients of certain modular forms of positive dimension. *Ann. of Math. (2)*, 39(2):433–462, 1938.

[Sch82] W. M. Schmidt. Simultaneous rational zeros of quadratic forms. In *Seminar on Number Theory, Paris 1980-81 (Paris, 1980/1981)*, volume 22 of *Progr. Math.*, pages 281–307. Birkhäuser, Boston, Mass., 1982.

[Sch84] W. M. Schmidt. Bounds for exponential sums. *Acta Arith.*, 44(3):281–297, 1984.

[Sch85] W. M. Schmidt. The density of integer points on homogeneous varieties. *Acta Math.*, 154(3-4):243–296, 1985.

[Sel51] E. S. Selmer. The Diophantine equation $ax^3+by^3+cz^3 = 0$. *Acta Math.*, 85:203–362, 1951.

[Ser77] J.-P. Serre. Majorations de sommes exponentielles. In *Journées Arithmétiques de Caen (Univ. Caen, Caen, 1976)*, pages 111–126. Astérisque No. 41–42. Soc. Math. France, Paris, 1977.

[Spi65] M. Spivak. *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*. W. A. Benjamin, Inc., New York-Amsterdam, 1965.