DIOPHANTINE PROBLEMS: INEQUALITIES AND ABELIAN VARIETIES

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr. rer. nat.)

of the Georg-August Universität Göttingen

within the doctoral program *Mathematical Sciences*

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

submitted by

Florian Maximilian Munkelt

from Wiesbaden-Dotzheim

Göttingen, 2023

Thesis committee

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

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

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

Members of the Examination board

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

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

Further members:

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

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

Prof. Dr. Russel Luke
Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen

Prof. Dr. Dominic Schuhmacher
Institut für Mathematische Stochastik, Georg-August Universität Göttingen

Date of the oral examination: 09.06.2023

Contents

Declaration of authorship

I declare that Part II has been written in colaboration with Victoria Canotral Farfan, currently employed with the University of Göttingen. Victoria Cantoral Farfan has proposed the problem and both authors have contributed equally to the proofs aswell as the writing. I have in any other places acknowledged the work of others by providing detailed references of said work.

Introduction

Diophantine problems are typically asking for the number of integer solutions to systems of polynomial equations in one or more variables with integer coefficients. We call a Diophantine equation homogenous, if it is defined by a homogenous polynomial over the rational numbers or the integers. The most famous example is the object of study in Fermat's Last Theorem

$$a^n + b^n - c^n = 0.$$

Illustrated by the three decades of active research needed to proof Fermat's Last Theorem and the negative answer to Hilbert's tenth problem, provided by Matiyasevich's theorem after 21 years of combined efforts, solutions of Diophantine equations are not easy to obtain. Many celebrated results have been achieved with analytic methods, such as the Hardy-Littlewood circle method, towards solutions to additive number theory questions such as Waring's problem and later to Diophantine equations of a more general type. We refer to [15], [22] and [68] for a comprehensive treatment of those aspects.

In order to approach the problem from a different standpoint and possibly with new and promising tools, we can reformulate our problem in a geometric fashion. Naturally, a homogenous polynomial in $n+1$ variables defines a hypersurface in n-dimensional projective space $\mathbb{P}^n$. Hence solutions of a homogenous Diophantine equation are equivalent to rational points on the corresponding projective hypersurface. Considerable effort has been made towards establishing estimations for counting functions of the shape

$$N(X;B) = \#\{x \in X \cap \mathbb{P}^n(\mathbb{Q}) \mid H(x) \leq B\},$$

where $X \subset \mathbb{P}^n$ is a projective variety, we consider $x \in \mathbb{P}^n(\mathbb{Q})$ to be an $n+1$-tuple $(x_0, ..., x_n) \in \mathbb{Z}^{n+1}$ with a suitably chosen height function H. If X is a hypersurface defined by the form $F \in \mathbb{Z}[X_0, ..., X_n]$, we shall write $N(F;B)$ instead of $N(X;B)$ and vice versa. In 1983 Heath-Brown raised the question whether for an absolutely irreducible form G of degree $d \geq 2$ in n variables one can achieve

$$N(G;B) \ll B^{n-2+\varepsilon}$$

for some $\epsilon > 0$ in [29]. This was later picked up by Serre in both [63] and [64] and appeared as the dimension growth conjecture in Browning's work [14]:

Conjecture 0.1 (Serre's dimension growth conjecture). *Let $X \subseteq \mathbb{P}^{M-1}_{\mathbb{Q}}$ be an irreducible projective varietiy of degree at least two defined over $\mathbb{Q}$. Let $N_X(B)$ be the number of rational points on X with naive height bounded by B. Then*

$$N_X(B) \ll B^{\dim X}(\log B)^c$$

for some constant $c > 0$.

The reader may be interested in the large amount of literature concerning the dimension growth conjecture and find [18] to be a nice introduction to the topic. For further reading we may refer to several examples, such as [8], [9], [10], [11], [12], [13], [30], [42], [59], [60], [61], [70].

We want to shift our attention from Diophantine equations to two distinctly different objects in diophantine geometry that illustrate the broad and rich amount of theory and tools relevant and available for the study of rational points in modern number theory.

First we want to consider a projective variety $X \subset \mathbb{P}^n_{\mathbb{R}}$ in real projective space. Assume X is given by the form $F \in \mathbb{Z}[X_0, ..., X_n]$ of degree d, then after de-homogenization (for X_0), we obtain a polynomial F_0 in n variables satisfying

$$F(X_0, ..., X_n) = X_0^d F_0\left(\frac{X_1}{X_0}, ..., \frac{X_n}{X_0}\right)$$

and a function

$$\Phi\colon \mathbb{R}^{n-1} \times \mathbb{R} \to \mathbb{R}, ((x_2, ..., x_n), x_1) \mapsto F_0(x_1, ..., x_n).$$

If the implicit function theorem is applicable, we can locally express the solutions of the equation $F_0(x_1, ..., x_n) = 0$ as the graph of a smooth function $f\colon \mathbb{R}^{n-1} \to \mathbb{R}$. Such a graph can then be naturally understood as a hypersurface in $\mathbb{R}^n$. This idea extends to a system of forms as is expected and corresponds locally to a smooth immersed manifold. Instead of Diophantine equations, we want to study Diophantine inequalities under this consideration. Robert and Sargos in [57] have given an upper bound on the integral solutions $(x_1, ..., x_4) \in [B+1, 2B]^4$ to the inequality

$$|x_1^\alpha - x_2^\alpha - x_3^\alpha - x_4^\alpha| \leq \delta B^{\alpha-1},$$

where $\alpha \in \mathbb{R}$, $\alpha \notin \{0,1\}$, $B \geq 2$ and $\delta > 0$. With the above mentioned technique, we can view these solutions equivalently as rational points within a distance of δ to the hypersurface Y parametrized by the equation $y^\alpha = x_1^\alpha - x_2^\alpha - 1$ on $[1,2]^2$. This leads to the more general question of the number of points that are close to a compact immersed submanifold $\mathscr{M} \subset \mathbb{R}^M$. Given an integer $Q \in \mathbb{N}$ and $\delta \geq 0$, we study the number

$$N(\mathscr{M}; Q, \delta) := \#\left\{(\mathbf{a}, q) \in \mathbb{Z}^M \times \mathbb{N} \,\middle|\, 1 \leq q \leq Q, \operatorname{dist}\left(\frac{\mathbf{a}}{q}, \mathscr{M}\right) \leq \frac{\delta}{q}\right\}$$

of rational points with denominators bounded by Q and L^∞-distance to $\mathscr{M}$ bounded by δ. Despite being an interesting question in its own right and the context mentioned above, there are several applications to different problems in Diophantine geometry, see for example [25], [36, §2-§5], [37], [45] or [66].

We can readily state a trivial estimate

$$N(\mathscr{M}; Q, \delta) \ll Q^{\dim \mathscr{M} + 1},$$

and a probabilistic heuristic yields

$$\delta^R Q^{\dim \mathscr{M}+1} \ll N(\mathscr{M}; Q, \delta) \ll \delta^R Q^{\dim \mathscr{M}+1},$$

where $R = M - \dim \mathscr{M}$ is the codimension of the manifold in question. It is known that this heuristic does not hold unconditionally. For example, if $\mathscr{M}$ is a rational hyperplane in $\mathbb{R}^M$ and $\delta \leq 1$, then we find that

$$Q^{\dim \mathscr{M}+1} \ll N(\mathscr{M}; Q, \delta) \ll Q^{\dim \mathscr{M}+1}.$$

We additionally see from that example, that in order to establish non-trivial bounds we may be inclined to study manifolds with a 'proper curvature condition'. Huang proposed the following conjecture in his groundbreaking work [36].

Conjecture 0.2. *Let $\mathscr{M}$ be a bounded immersed submanifold of $\mathbb{R}^M$ with boundary. Let $R = M - \dim \mathscr{M}$ and suppose $\mathscr{M}$ satisfies a 'proper curvature condition'. Then there exists a constant $c_{\mathscr{M}} > 0$ depending only on $\mathscr{M}$ such that*

$$N(\mathscr{M}; Q, \delta) \sim c_{\mathscr{M}} \delta^R Q^{\dim \mathscr{M}+1}$$

when $\delta \geq Q^{-\frac{1}{R}+\epsilon}$ for some $\epsilon > 0$ and $Q \to \infty$.

It is not made explicit what 'proper curvature conditions' means in the given context. The first non-trivial case that has been studied extensively is that of a compact curve in $\mathbb{R}^2$ with curvature bounded away from zero. In this setting, Huxley [38] was the first to obtain a notable upper bound for a $\mathcal{C}^2$ curve $\mathscr{C}$, which has later been given in the version

$$N(\mathscr{C}; Q, \delta) \ll_{\mathscr{C}} \delta^{1-\epsilon} Q^2 + Q \log Q$$

for any δ and Q in [69]. In fact, Vaughan and Velani [69] showed that

$$N(\mathscr{C}; Q, \delta) \ll_{\mathscr{C}} \delta Q^2 + Q^{1+\epsilon}$$

for a $\mathcal{C}^3$ curve $\mathscr{C}$, which is the upper bound that Conjecture 0.2 predicts.

Conversely, a sharp lower bound has been established by Beresnevich, Dickinson and Velani [3]

$$\delta Q^2 \ll_{\mathscr{C}} N(\mathscr{C}; Q, \delta)$$

with $\delta \gg Q^{-1}$ and $\delta Q \to \infty$, for a $\mathcal{C}^3$ curve $\mathscr{C}$ admitting at least one point with non-vanishing curvature. Further work by Huang [35] established an asymptotic formula for $\mathcal{C}^3$ curves. The interested reader may find more details on the case for planar curves in [35].

For the case of general manifolds Beresnevich established the sharp lower bound

$$\delta^R Q^{\dim \mathscr{M}+1} \ll_{\mathscr{M}} N(\mathscr{M}; Q, \delta)$$

for any $\delta \gg Q^{-\frac{1}{R}}$, assuming $\mathscr{M}$ is an analytic submanifold of $\mathbb{R}^M$ which admits at least one non-degenerate point, in his spectacular work [4].

Huang established Conjecture 0.2 in the case when $\mathscr{M}$ is a hypersurface with Gaussian curvature bounded away from zero in $\mathbb{R}^M$ in [36].

A recent generalization of this result is due to Schindler and Yamagishi [67], who established the Conjecture 0.2 in the case of a compact immersed submanifold of $\mathbb{R}^M$ in codimension R with a curvature condition that reduces to Huang's case for $R = 1$. In particular, if the manifold $\mathscr{M}$ is locally parametrized by the functions $f_1, ..., f_R$, the required curvature condition is as follows.

Condition 0.3. *Given any* $\mathbf{t} \in \mathbb{R}^R \setminus \{0\}$*, we have*

$$\det H_{t_1 f_1 + \cdots + t_R f_R}(\mathbf{x}_0) \neq 0,$$

where H_f *denotes the Hessian matrix of the function* f *and* $\mathbf{x}_0$ *is given as below.*

We continue by presenting the details of our main result. By the compact nature of $\mathscr{M}$, the argument reduces to a finite number of local arguments, hence we may assume without loss of generality that

$$\mathscr{M} := \{(\mathbf{x}, f_1(\mathbf{x}), ..., f_R(\mathbf{x})) \in \mathbb{R}^M \mid \mathbf{x} = (x_1, ..., x_n) \in \overline{B_{\varepsilon_0}(\mathbf{x}_0)}\}, \tag{0.1}$$

where $\mathbf{x}_0 \in \mathbb{R}^n$, $\varepsilon_0 > 0$ and $f_r \in \mathcal{C}^\ell(\mathbb{R}^n)$ for $1 \leq r \leq R$ and some $\ell \geq 2$. Note that this specifically means $\dim \mathscr{M} = n$.
Let $\omega \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ be a non-negative weight function that is compactly supported in a sufficiently small neighbourhood of $\mathbf{x}_0$ and define

$$N_\omega(Q, \delta) = \sum_{\substack{\mathbf{a} \in \mathbb{Z}^n \\ q \leq Q \\ ||q f_r(\mathbf{a}/q)|| \leq \delta \\ 1 \leq r \leq R}} \omega\left(\frac{\mathbf{a}}{q}\right),$$

where $|| \cdot ||$ denotes the distance to the closest integer. Obviously $||x|| \leq 1/2$ for any $x \in \mathbb{R}$, hence we only consider $0 \leq \delta \leq 1/2$. Let

$$N_0 := \sum_{\substack{\mathbf{a} \in \mathbb{Z}^n \\ q \leq Q}} \omega\left(\frac{\mathbf{a}}{q}\right).$$

For a given function $f \in \mathcal{C}^2(\mathbb{R}^n)$ we denote by $H_f(\mathbf{x})$ the Hessian matrix of f evaluated at $\mathbf{x}$, i.e. the $n \times n$-matrix whose entries are $\frac{\partial^2 f}{\partial x_\mu \partial x_\nu}(\mathbf{x})$ for $1 \leq \mu, \nu \leq n$.
We use the following relaxed curvature condition throughout this chapter.

Condition 0.4. *Given any* $\mathbf{t} \in \mathbb{R}^R \setminus \{0\}$ *and* $1 \leq s \leq n-2$*, we have*

$$\operatorname{rank} H_{t_1 f_1 + \cdots + t_R f_R}(\mathbf{x}_0) \geq n - s.$$

With these notations we have the following result.

Theorem 0.5. *Let* $n \geq 2$ *and* $\ell > \max\{n+1, \frac{n}{2}+4\}$*. Suppose 0.4 holds and that* $\varepsilon_0 > 0$ *is sufficiently small. Then we have*

$$|N_\omega(Q, \delta) - (2\delta)^R N_0| \ll \begin{cases} \delta^{\frac{(R-1)(n+s-2)}{n+s}} Q^{n+\frac{2s}{n+s}} \mathscr{E}_{n-s}(Q) & \text{if } \delta \geq Q^{-\frac{n-s}{n+s+2R-2}}, \\ Q^{n-\frac{(R-1)n-(R+1)s-2R+2}{n+s+2R-2}} \mathscr{E}_{n-s}(Q) & \text{if } \delta < Q^{-\frac{n-s}{n+s+2R-2}}, \end{cases}$$

where

$$\mathscr{E}_{n-s}(Q) = \begin{cases} \exp(\mathfrak{c}_1 \sqrt{\log Q}) & \text{if } n-s = 2, \\ (\log Q)^{\mathfrak{c}_2} & \text{if } n-s \geq 3, \end{cases}$$

for some positive constants $\mathfrak{c}_1$ *and* $\mathfrak{c}_2$*. These constants as well as the implicit constants only depend on* $\mathscr{M}$ *and* ω*.*

Comparing our exponents to those obtained in [67, Theorem 1.2], we have $n+\frac{2s}{n+s}$ instead of n in the first case and $n-\frac{(R-1)n-(R+1)s-2R+2}{n+s+2R-2}$ instead of $n-\frac{(n-2)(R-1)}{n+2R-2}$ in the second one. Note that the cases in [67] are distinguished by the comparison of δ against $Q^{-\frac{n}{n+2R-2}}$. As expected the bounds under the less restrictive curvature condition are worse, yet they have similar growth for large n.

By the Poisson summation formula we find that $N_0 = \sigma Q^{n+1} + O\left(Q^n\right)$ for some positive constant σ depending only on ω and n (compare [36, (6.2)]). Combining this with 0.5 yields

$$N_\omega(Q,\delta) = (2\delta)^R \sigma Q^{n+1} + O\left(\delta^{\frac{(R-1)(n+s-2)}{n+s}} Q^{n+\frac{2s}{n+s}} \mathscr{E}_{n-s}(Q)\right),$$

when $\delta \geq Q^{-\frac{n-1}{n+2R-1}+\epsilon}$ for any $\epsilon > 0$ sufficiently small. Following the arguments in [36, Section 7], we can approximate the characteristic function of $\overline{B_{\varepsilon_0}(\mathbf{x}_0)}$ by smooth weight functions and obtain:

Corollary 0.6. *Let $\mathscr{M}$ be as in (0.1), $n \geq 3$ and $\ell > \max\{n+1, \frac{n}{2}+4\}$. Suppose Condition 0.4 holds and that $\varepsilon_0 > 0$ is sufficiently small. Then there exists a constant $c_{\mathscr{M}} > 0$ depending only on $\mathscr{M}$ such that*

$$N(\mathscr{M};Q,\delta) \sim c_{\mathscr{M}} \delta^R Q^{dim \mathscr{M}+1}$$

when $\delta \geq Q^{-\frac{n-s}{n+s+2R-2}+\epsilon}$ for any $\epsilon > 0$ suffciently small and $Q \to \infty$.

Note that

$$Q^{-\frac{1}{R}} \geq Q^{-\frac{n-s}{n+s+2R-2}}$$

only holds for $R > 1$ and $n \geq \frac{(R+1)s+2R-2}{R-1}$, hence Conjecture 0.2 holds in those cases. In fact, the asymptotic formula is obtained beyond the range of δ that was conjectured in those cases.

If we let $\delta = 0$, then our (weighted) counting function gives the (weighted) number of rational points with bounded denominators that lie on the manifold $\mathscr{M}$. Applying the arguments from [36, pp. 2047] to Conjecture 0.2, we obtain

$$N(\mathscr{M};Q,0) \ll Q^{\dim \mathscr{M}+\epsilon},$$

for any $\epsilon > 0$ sufficiently small with a generally sharp upper bound. This can be interpreted as an analogue of the aforementioned dimension growth conjecture 0.1 for projective varieties in the context of smooth submanifolds of $\mathbb{R}^M$.

Corollary 0.7. *Let $\mathscr{M}$ be as in (0.1), $n-s \geq 3$ and $\ell > \max\{n+1, \frac{n}{2}+4\}$. Suppose Condition 0.4 holds and that $\varepsilon_0 > 0$ is sufficiently small. Then*

$$N(\mathscr{M};Q,0) \ll Q^{n-\frac{(R-1)n-(R+1)s-2R+2}{n+s+2R-2}} (\log Q)^c$$

for some constant $c > 0$.

Note that in contrast to the situation in [67] we do not unconditionally break the $\dim \mathscr{M}$ barrier here, only if $n > \frac{(R+1)s+2R-2}{R-1}$ and $R > 1$.

We adapt the strategy for proving Theorem 0.5 established in [67], which relies on the methods developed by Huang in [36] and fibration arguments. In particular, Schindler and Yamagishi reduced the problem to that for one function, such that the main result of [36] can be used. This is achieved by a more complicated version of the procedure developed by Huang in [36], which relates the counting problem of a function to that of its Legendre transform, for a family of functions satisfying the curvature condition 0.3 and apllying it twice. For our relaxed curvature condition 0.4 we can use a similar approach with some necessary adjustments to accomodate an additional degree of freedom.

We deal with this problem in Part I. After collecting some preliminary results in Section 1, we discuss the setup of our proof for Theorem 0.5 in Section 2. Section 3 is dedicated to establishing some auxiliary bounds, one of which depending on a result which is proven in Section 4. Lastly, we combine our findings to prove Theorem 0.5 in Section 5.

Next, we turn our attention to the Diophantine problem of writing an integer as the difference of a square and a cube. Fix an integer $k \in \mathbb{Z}$, then we are interested in integer or rational solutions to the Diophantine equation

$$y^2 - x^3 = k.$$

This is known as Bachet's equation and we refer the reader to [51, Ch. 26] for more context. An interesting property of said equation is the duplication formula discovered by Bachet in 1621, that is for any pair $(x, y) \in \mathbb{Q} \times \mathbb{Q}^\times$ that is a solution to the Bachet equation, another rational solution can be constructed as follows:

$$\left(\frac{x^2 - 8kx}{4y^2}, \frac{-x^6 - 20kx^3 + 8k^2}{8y^3} \right).$$

More general, generic equations of the form

$$y^2 = x^3 + ax + b$$

define elliptic curves, that is a smooth projective curve of genus 1 over a given field that admits a specified point $\mathcal{O}$ with coordinates in that same field. A central property of elliptic curves is that they allow for a group law on their points, extending the known duplication formula and turning the set of rational points of an elliptic curve into an abelian group. An abelian variety is a projective algebraic variety that also admits an algebraic group law on its points. In that sense, elliptic curves are exactly abelian varieties of dimension 1 and we now study the rational points of abelian varieties without an explicit mention of defining equations. The famous Mordell-Weil theorem states that for an abelian variety A defined over a number field K, the K-rational points form a finitely generated group $A(K)$, the so-called Mordell-Weil group. As a finitely generated abelian group, $A(K)$ can be decomposed as a direct sum into a free subgroup and its torsion subgroup $A(K)_{\text{tors}}$, which is in turn finite. Therefore it is a natural question to ask, whether there is a bound on the number of K-torsion points on A.

The uniform boundedness conjecture claims that for any given positive integer d all number fields K with $[K : \mathbb{Q}] = d$ and all abelian varieties A defined over K satisfy the estimation

$$|A(K)_{\text{tors}}| \ll B(d, g),$$

where $B(d, g)$ only depends on d and the dimension $g = \dim A$. Two relevant strategies have been developed in order to approach this powerful statement, both fixing certain data to achieve feasible results:

1. In the 'horizontal' approach to the problem, we fix a number field K and let the abelian varieties defined over K vary within a given class. A complete understanding of this approach exists to this date only for elliptic curves over number fields, where it has been proven (over any number field) by Merel [46] building on previous work by Mazur [44] and Kamienny [39].
2. In the 'vertical' approach to the problem, we fix an abelian variety defined over a number field K and vary over finite extensions L/K. Specifically, we ask the question: How does the order of the torsion subgroup $A(L)_{\text{tors}}$ for any finite extension L/K grow compared to the extension degree $[L:K]$?

Part II of this thesis explores the vertical approach. In a letter to Bertrand [43], Masser proved the following result in this direction:

Theorem 0.8 (Masser '86). *Let A be an abelian variety of dimension g defined over a number field K. Then there exists a constant $C(A,K)$, depending only on A and $[K:\mathbb{Q}]$, such that for every finite extension L/K, we have*

$$|A(L)_{\text{tors}}| \leq C(A,K)([L:\mathbb{Q}]\log([L:\mathbb{Q}]))^g.$$

We follow the ideas later developed by Hindry and Ratazzi to give a better exponent, introducing the invariant $\gamma(A)$.

Definition 0.9 (Invariant $\gamma(A)$). The invariant $\gamma(A)$ is defined as follows:

$$\gamma(A) = \inf\{x > 0 \,|\, \forall L/K \text{ finite}, \ |A(L)_{\text{tors}}| \ll [L:K]^x\},$$

where the notation $\ll$ means that there exists a constant $C_{A,x}$ such that

$$|A(L)_{\text{tors}}| \leq C_{A,x}[L:K]^x.$$

Note that this exponent $\gamma(A)$ is optimal in the sense that it is minimal, such that for every $\varepsilon > 0$ and every finite extension L/K we have

$$|A(L)_{\text{tors}}| \ll [L:K]^{\gamma(A)+\varepsilon}.$$

In terms of this invariant Masser's theorem 0.8 states that $\gamma(A) \leq g$. This bound on $\gamma(A)$ is optimal when A is isogenous to a power of a CM elliptic curves, but not in the general case. The work of Hindry and Ratazzi has produced an explicit formula for the invariant $\gamma(A)$ in a number of cases. First Ratazzi [58] gave an explicit formula for the case of a CM abelian variety in terms of the characters of the Mumford-Tate group $\mathrm{MT}(A)$ of A. Only shortly thereafter, Hindry and Ratazzi have worked together on an explicit formula for products of elliptic curves [31]. Later they studied abelian varieties of type GSp, meaning that they satisfy the Mumford-Tate conjecture and that their Mumford-Tate groups are isomorphic to the group of symplectic similitudes.

Any abelian variety A defined over a number field K of dimension g is isogenous, over $\overline{K}$, to a product of simple abelian varieties $\prod_{i=1}^{k} A_i^{n_i}$, where each simple factor A_i can be assigned a type (I, II, III or IV) in the sense of Albert's classification. Furthermore, we can assume that A_i is not isogenous to A_j for $i \neq j$. Within [32], Hindry and Ratazzi formulate the following conjecture:

Conjecture 0.10 ([32] Conjecture 1.1).

$$\gamma(A) = \max_{\varnothing \neq I \subseteq \{1,\ldots,k\}} \frac{2\sum_{i\in I} n_i \dim(A_i)}{\dim(\mathrm{MT}(A_i))}, \tag{0.2}$$

where for every non-empty set $I \subseteq \{1, ..., d\}$ we have $A_I = \prod_{i\in I} A_i^{n_i}$ and the Mumford–Tate group $\mathrm{MT}(A_I)$

They eventually proved this conjecture for simple abelian varieties of type I and II that are fully of Lefschetz type [33]. The conjecture has also been shown to hold for simple abelian varieties of type III that are fully of Lefschetz type, as well as abelian varieties isogenous to products of simple factors of type I, II or III, that are all fully of Lefschetz type in [17]. The remaining case to be proven is when A is a simple abelian variety fully of type IV, that is of type IV but not of CM type, and fully of Lefschetz type. Our first result is oriented towards proving the conjecture of Hindry and Ratazzi for this particular class of abelian varieties. Precisely we show the following:

Theorem 0.11. *Let A be a simple abelian variety defined over a number field K. Assume that A is fully of type IV and fully of Lefschetz type[1], then*

$$\gamma(A) = \frac{2\dim(A)}{\dim(\mathrm{MT}(A))}.$$

It should be noted that this conjecture has recently been proved by Le Fourn, Lombardo, and Zywina [27]. Nonetheless, our approach is profoundly different and the strategy developed in the proof of Theorem 0.11 allows us to obtain a lower bound for degree of the extension generated by a torsion point.

Corollary 0.12. *Under the assumptions above, there exist $c := C(A, K) > 0$ such that, for every torsion point $P \in A(\bar{K})$ of order m we have:*

$$[K(P) : K] \geq c^{\omega(m)} m^{dh},$$

where $\omega(m)$ is the number of prime divisors of m and h is the relative dimension as defined in Chapter II Section 1.

The proof of 0.11 relies on a criterion for the independance of ℓ-adic representations, introduced by Serre in [65]. This allows us to direct our attention to finite subgroups $H \subset A[\ell^\infty]$ for all primes ℓ. In particular the criterion states, that there exists a finite extension K'/K, such that for every finite subgroup $H_{\mathrm{tors}} \subset A(L)_{\mathrm{tors}}$ that we can write as

$$H_{\mathrm{tors}} = \prod_{\ell \text{ prime}} H_\ell,$$

where H_ℓ is a subgroup of $A[\ell^\infty]$, we have

$$[K'(H_{\mathrm{tors}}) : K'] \gg\ll \prod_{\ell} [K'(H_\ell) : K'].$$

We will therefore assume that K is such that the ℓ-adic representations are independent and work with the following definition of the invariant $\gamma(A)$:

$$\gamma(A) = \inf\{x > 0 \mid \forall H \subset A[\ell^\infty], |H| \ll [K(H) : K]^x\}.$$

Hence we can use the equivalence

$$|H| \ll [K(H) : K]^{\gamma(A)} \Leftrightarrow \gamma(A) \geq \frac{\log_\ell |H|}{\log_\ell [K(H) : K]} \tag{0.3}$$

[1] We refer the reader to Definition 1.13.

for every finite subgroup H of $A[\ell^\infty]$ to compute $\gamma(A)$. Note that the main contributions to the value of $\gamma(A)$ are the order of a given subgroup H and the degree of the associated extension $K(H)$ over K. Therefore, we can naturally approach the proof in two parts: calculating these two values explicitly and then using combinatorial methods to obtain the invariant $\gamma(A)$.

We discuss this problem in Part II. Beginning with the revision of preliminaries notions on abelian varieties and the relevant algebraic groups attached to them in Section 1, we follow up with a discussion of the crucial results on Unitary groups in Section 2. With these preparation we collect all necessary prerequisites on the torsion subgroup in Section 3 to then calculate the invariants and proof Theorem 0.11 and Corollary 0.12 in Section 4. We additionally give an outlook towards abelian varieties isogenous to a product with a simple factor of type IV. The remaining Sections 5 and 6 are dedicated to explicit calculations.

Part I
Inequalities:

On the number of rational points close to a compact manifold

1. Preliminaries

1.1 Notation

We denote by $\mathcal{C}^\ell(\mathcal{V})$ the set of ℓ-times continuously differentiable functions defined on an open set $\mathcal{V} \subseteq \mathbb{R}^n$. Analogously $\mathcal{C}^\infty(\mathcal{V})$ denotes the set of smooth functions defined on $\mathcal{V}$ and $\mathcal{C}_c^\infty(\mathcal{V})$ the set of smooth functions defined on $\mathcal{V}$ that have a compact support. Given any $f \in \mathcal{C}^1(\mathbb{R}^n)$ we let $\nabla f = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n})$ be the gradient of f. For a subset $X \subseteq \mathbb{R}^n$ we denote the boundary of X by $\partial X = \overline{X} \setminus X^\circ$, where $\overline{X}$ denotes the closure of X and X° denotes the interior of X. For any $z \in \mathbb{R}$ we let $e(z) = e^{2\pi \mathrm{i} z}$ and $||z||$ denotes the distance to the closest integer. For $\mathbf{z} = (z_1, ..., z_n) \in \mathbb{R}^n$ let $|\mathbf{z}| = \max_{1 \leq i \leq n} |z_i|$ denote the L^∞-Norm and given any $\varepsilon > 0$ we let

$$B_\varepsilon^n(\mathbf{z}) = \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x} - \mathbf{z}| < \varepsilon\} = (z_1 - \varepsilon, z_1 + \varepsilon) \times \cdots \times (z_n - \varepsilon, z_n + \varepsilon).$$

We may write $B_\varepsilon(\mathbf{z})$ instead of $B_\varepsilon^n(\mathbf{z})$ if the dimension is clear from context. For natural numbers $k \leq m$ let $[m] = \{1, ..., m\}$ be the set of natural numbers smaller or equal to m and

$$[m]^k = \{\boldsymbol{\nu} \subseteq [m] \mid \#\boldsymbol{\nu} = k\}$$

the set of k-subsets of $[m]$. By the notation $f(\mathbf{x}) \ll g(\mathbf{x})$ or $f = O\left(g(\mathbf{x})\right)$ we mean that there exists a constant $C > 0$ such that $|f(\mathbf{x})| \leq Cg(\mathbf{x})$ for all $\mathbf{x}$ in consideration.

Definition 1.1. Let $F \in \mathcal{C}^\ell(\mathbb{R}^n)$ and let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open subset such that ∇F is invertible on $\mathcal{U}$. We define the Legendre transform $F^* \colon \nabla F(\mathcal{U}) \to \mathbb{R}$ of F via

$$F^*(\mathbf{z}) = \mathbf{z} \cdot (\nabla F)^{-1}(\mathbf{z}) - (F \circ (\nabla F)^{-1})(\mathbf{z}).$$

It can be verified that F^* is ℓ-times continuously differentiable, $F^{**} = F$ and $\nabla F^* = (\nabla F)^{-1}$. If $\mathbf{x} = \nabla F(\mathbf{z})$ we obviously have

$$F^*(\mathbf{x}) = \mathbf{x} \cdot \mathbf{z} - F(\mathbf{z}) \tag{1.1}$$

and furthermore

$$H_{F^*}(\mathbf{x}) = H_F(\mathbf{z})^{-1}. \tag{1.2}$$

1.2 Oscillatory integrals

The proofs presented in this part rely fundamentally on the estimation of oscillatory integrals. We discuss the theory briefly, following [34] (compare specifically Theorem 7.7.1 and 7.7.5).

An oscillatory integral is given in the form

$$\int_{\mathbb{R}^d} \omega(\mathbf{x}) e(\lambda\varphi(\mathbf{x}))\mathrm{d}\mathbf{x},$$

where $\lambda > 0$, φ is called the phase function and ω is a smooth weight function. The easiest method to estimate such an integral is to use iterated partial integration, however this is only possible if φ has no stationary points in the support of ω. For simplicity consider the case $d = 1$ and assume $\varphi'(x) \neq 0$ for all $x \in \operatorname{supp}\omega = [a, b]$, $a \leq b$. Then

$$\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x))\mathrm{d}x = \int_a^b \omega(x) e(\lambda\varphi(x))\mathrm{d}x.$$

Since $\varphi'(x) \neq 0$ for all $x \in [a, b]$ we can introduce a complicated 1

$$\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x))\mathrm{d}x = \int_a^b \omega(x)\frac{2\pi\mathrm{i}\lambda\varphi'(x)}{2\pi\mathrm{i}\lambda\varphi'(x)} e(\lambda\varphi(x))\mathrm{d}x = \int_a^b \frac{\omega(x)}{2\pi\mathrm{i}\lambda\varphi'(x)}\frac{\mathrm{d}}{\mathrm{d}x} e(\lambda\varphi(x))\mathrm{d}x.$$

Utilizing integration by parts for the functions $\frac{\omega(x)}{\varphi'(x)}$ and $e(\lambda\varphi(x))$ we find

$$\int_{\mathbb{R}} \omega(x) e(\mathrm{i}\lambda\varphi(x))\mathrm{d}x = \frac{1}{2\pi\mathrm{i}\lambda}\left(\left.\frac{e(\lambda\varphi(x))\omega(x)}{\varphi'(x)}\right|_a^b - \int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x))\mathrm{d}x\right).$$

Since ω is smooth with compact support, we have $\omega(a) = \omega(b) = 0$, hence we can simplify to

$$\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x))\mathrm{d}x = -\frac{1}{2\pi\mathrm{i}\lambda}\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x))\mathrm{d}x.$$

Using integration by parts again for $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\omega(x)}{\varphi'(x)}\right)$ and $e(\lambda\varphi(x))$ and the quotient rule for derivatives the right hand side is

$$-\left(\frac{1}{2\pi\mathrm{i}\lambda}\right)^2\left(\left.\frac{e(\lambda\varphi(x))(\omega(x)\varphi''(x) - \omega'(x)\varphi(x))}{\varphi'(x)^2}\right|_a^b - \int_a^b \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x))\mathrm{d}x\right).$$

Just like ω, all it's derivatives vanish at a and b, hence we obtain

$$\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x))\mathrm{d}x = \left(\frac{-1}{2\pi\mathrm{i}\lambda}\right)^2 \int_a^b \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x))\mathrm{d}x.$$

Since ω and φ are smooth, we can repeat this process arbitrarily often and in every step the non-integral contribution of the integration by parts clearly vanishes because every additive term contains a factor of $\omega^{(k)}(a) = 0$ or $\omega^{(k)}(b) = 0$ for a suitable $k \in \mathbb{N}$. We conclude that for every $\ell \in \mathbb{N}$

$$\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x)) \mathrm{d}x = \left(\frac{-1}{2\pi \mathrm{i}\lambda}\right)^{\ell} \int_a^b \frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x)) \mathrm{d}x.$$

Now taking absolute values we see that

$$\begin{aligned}\left|\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x)) \mathrm{d}x\right| &= \frac{1}{(2\pi\lambda)^\ell}\left|\int_a^b \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\omega(x)}{\varphi'(x)}\right) e(\lambda\varphi(x)) \mathrm{d}x\right| \\ &\leq \frac{1}{(2\pi\lambda)^\ell}\int_a^b \left|\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\left(\frac{\omega(x)}{\varphi'(x)}\right)\right| \mathrm{d}x \\ &\leq \frac{M(b-a)}{\lambda^\ell},\end{aligned}$$

where $M = (2\pi)^{-\ell} \max_{[a,b]}\left(\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\left(\frac{\omega(x)}{\varphi'(x)}\right)\right)$. In essence

$$\left|\int_{\mathbb{R}} \omega(x) e(\lambda\varphi(x)) \mathrm{d}x\right| \leq M(\omega, \varphi, \ell)\lambda^{-\ell},$$

where $M(\omega, \varphi, \ell)$ is a constant depending on ω, φ and ℓ, i.e. the integral is $O(\lambda^{-\ell})$ for every $\ell \in \mathbb{N}$. For the higher dimensional case we argue similarly, integrating over one variable after the other. We will use the following precise result in full generality:

Lemma 1.2 (non-stationary phase). *Let $\ell \in \mathbb{N}$ and $U_+ \subseteq \mathbb{R}^d$ a bounded open set. Let $\omega \in \mathcal{C}_c^{\ell-1}(\mathbb{R}^d)$ with $\operatorname{supp}\omega \subseteq U_+$ and $\varphi \in \mathcal{C}^\ell(U_+)$ with $\nabla\varphi(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in \operatorname{supp}\omega$. Then for any $\lambda > 0$*

$$\left|\int_{\mathbb{R}^d} \omega(\mathbf{x}) e(\lambda\varphi(\mathbf{x})) \mathrm{d}\mathbf{x}\right| \leq c_\ell \lambda^{-\ell+1},$$

where the constant c_ℓ only depends on ℓ, d, upper bounds for the absolute values of finitely many derivatives of ω and φ on U_+ and a lower bound for $|\nabla\varphi|$ on $\operatorname{supp}\omega$.

In order to handle a phase function that admits a stationary point, we first consider the case where φ is a (one dimensional) quadratic form $\varphi\colon [-a, b] \to \mathbb{R}, x \mapsto \alpha x^2$ with $\alpha, a, b > 0$. Then

$$\int_{-a}^b e(\varphi(x)) \mathrm{d}x = \int_0^a e(\alpha x^2) \mathrm{d}x + \int_0^b e(\alpha x^2) \mathrm{d}x,$$

so we shall consider only one of those integrals. By the substitution $x \mapsto \sqrt{\alpha}x$ we obtain

$$\int_0^a e(\alpha x^2) \mathrm{d}x = \frac{1}{\sqrt{\alpha}} \int_0^{\sqrt{\alpha}a} e(z^2) \mathrm{d}z.$$

Let $\gamma_1\colon [0, 1] \to \mathbb{C}, t \mapsto e(1/8)tw$ and $\gamma_2\colon [0, 1/8] \to \mathbb{C}, t \mapsto e(t)w$ for some $w > 0$. We can consider the complex line integrals over the curve following γ_1 and then the inverse of γ_2 and by Cauchy's theorem obtain

$$\int_0^w e(z^2) \mathrm{d}z = \int_{\gamma_1} e(z^2) \mathrm{d}z + \int_{-\gamma_2} e(z^2) \mathrm{d}z.$$

A straight forward calculation shows that

$$\int_{\gamma_1} e(z^2)\mathrm{d}z = \frac{e(1/8)}{2\sqrt{2}} + O\left(\frac{1}{w}\right)$$

and

$$\int_{-\gamma_2} e(z^2)\mathrm{d}z = O\left(\frac{1}{w}\right).$$

Resubstitution now yields

$$\int_{-a}^{b} e(\alpha x^2)\mathrm{d}x = \frac{e(1/8)}{\sqrt{2A}} + O\left(\frac{1}{a\alpha} + \frac{1}{b\alpha}\right),$$

which we can extend to a quadratic function of the shape $\varphi\colon [a,b] \to \mathbb{R}, x \mapsto \alpha(x-x_0)^2+\beta$ by a linear substitution. For a more general phase function φ, that is sufficiently smooth, with a stationary point at x_0 we use Taylor's formula to write

$$\varphi(x) = \varphi(x_0) + \frac{1}{2}\varphi''(x_0)(x-x_0)^2 + R$$

with an error term R. Now we shall use the estimation for quadratic functions to obtain an estimate for the oscillatory integral in question. The precise result in full generality which we will make use of is as follows:

Lemma 1.3 (stationary phase). *Let $\ell > \frac{d}{2} + 4$ and $\mathscr{D}, \mathscr{D}_+ \subseteq \mathbb{R}^d$ bounded open sets such that $\overline{\mathscr{D}} \subseteq \mathscr{D}_+$. Let $\omega \in \mathcal{C}_c^{\ell-1}(\mathbb{R}^d)$ with $\operatorname{supp}\omega \subseteq \mathscr{D}$ and $\varphi \in \mathcal{C}^{\ell}(\mathscr{D})$. Suppose $\nabla\varphi(\mathbf{v}_0) = \mathbf{0}$ and $H_\varphi(\mathbf{v}_0) \neq 0$ for some $\mathbf{v}_0 \in \mathscr{D}$. Let σ be the signature of $H_\varphi(\mathbf{v}_0)$ and $\Delta = |\det H_\varphi(\mathbf{v}_0)|$. Suppose further that $\nabla\varphi(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in \overline{\mathscr{D}} \setminus \{\mathbf{v}_0\}$. Then for any $\lambda > 0$*

$$\int_{\mathbb{R}^d} \omega(\mathbf{x})e(\lambda\varphi(\mathbf{x}))\mathrm{d}\mathbf{x} = e\left(\lambda\varphi(\mathbf{v}_0) + \frac{\sigma}{8}\right)\Delta^{-\frac{1}{2}}\lambda^{-\frac{d}{2}}(\omega(\mathbf{v}_0) + O\left(\lambda^{-1}\right)),$$

where the implicit constant only depends on ℓ, d, upper bounds for the absolute values of finitely many derivatives of ω and φ on $\mathscr{D}_+$, an upper bound for $|\mathbf{x} - \mathbf{v}_0|/|\nabla\varphi(\mathbf{v}_0)|$ on $\mathscr{D}_+$, and a lower bound for Δ.

Recall given a symmetric matrix we define its signature to be the number of positive eigenvalues minus the number of negative eigenvalues and note that this is a simplified version of [34, Theorem 7.7.5], where the assumption on ℓ can be deduced from [34, pp. 222, Remark].

1.3 Compactly parametrized functions

The relevant phase function for our oscillatory integrals happen to be smooth functions, that are parametrized over a compact set, hence we collect two fundamental properties of such functions. Let $m \in \mathbb{N}_0$ and $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2 \subseteq \mathbb{R}^{n+m}$, where $\mathcal{G}_1 \subseteq \mathbb{R}^n$ and $\mathcal{G}_2 \subseteq \mathbb{R}^m$ are bounded connected open sets. Let $\mathbf{x}_0$ be a fixed point in $\mathcal{G}_1$.

Lemma 1.4. *Let $G \in \mathcal{C}^\ell(\mathcal{G})$, $\ell \geq 2$ and assume $H_{G_\mathbf{t}}(\mathbf{x}) \neq 0$ for every $\mathbf{t} \in \mathcal{G}_2$, where $G_\mathbf{t}$ is the real-valued map on $\mathcal{G}_1$ given by $\mathbf{x} \mapsto G(\mathbf{x}, \mathbf{t})$. Let $\mathcal{F}_2$ be a compact set contained in $\mathcal{G}_2$, then there exist a real number $\tau > 0$ and constants $c_1, c_2 > 0$ such that*

$$c_1 \leq |\det H_{G_\mathbf{t}}(\mathbf{x})| \leq c_2$$

for all $\mathbf{x} \in \overline{B_\tau(\mathbf{x}_0)}$ and $\mathbf{t} \in \mathcal{F}_2$. Moreover the map $\mathbf{x} \mapsto \nabla G_\mathbf{t}(\mathbf{x})$ is a $\mathcal{C}^{\ell-1}$-diffeomorphism on $\overline{B_\tau(\mathbf{x}_0)}$ for all $\mathbf{t} \in \mathcal{F}_2$.

Lemma 1.5. *Let $G \in \mathcal{C}^\ell(\mathcal{G})$, $\ell \geq 2$ and assume $H_{G_\mathbf{t}}(\mathbf{x}) \neq 0$ for every $\mathbf{t} \in \mathcal{G}_2$. Let $\mathcal{F}_2$ and τ be as in Lemma 1.4. Then for any $0 < \kappa < \tau$ sufficiently small, there exists $\rho > 0$ such that*

$$\operatorname{dist}(\partial(\nabla G_\mathbf{t}(B_\tau(\mathbf{x}_0))), \partial(\nabla G_\mathbf{t}(B_\kappa(\mathbf{x}_0)))) \geq 2\rho$$

for all $\mathbf{t} \in \mathcal{F}_2$.

For proofs of Lemma 1.4 and Lemma 1.5 we refer the reader to [67, Lemmas 3.4 and 3.5].

2. Setting up the proof of Theorem 0.5

By virtue of the characteristic functions

$$\chi_\delta(\theta) = \begin{cases} 1 & \text{if } ||\theta|| \leq \delta, \\ 0 & \text{else,} \end{cases} \tag{2.1}$$

for $0 < \delta \leq 1/2$ we can rewrite

$$N_\omega(Q, \delta) = \sum_{\substack{\mathbf{a} \in \mathbb{Z}^n \\ q \leq Q}} \omega\left(\frac{\mathbf{a}}{q}\right) \prod_{r=1}^{R} \chi_\delta\left(q f_r\left(\frac{\mathbf{a}}{q}\right)\right).$$

Consider the Selberg magic functions as described in [49] for the interval $[-\delta, \delta] \subseteq \mathbb{R}/\mathbb{Z}$ and a parameter $J \in \mathbb{N}$

$$S_J^{\pm}(x) = \sum_{|j| \leq J} \widehat{S}_J^{\pm}(j) e(jx).$$

They obey the properties

$$S_J^-(y) \leq \chi_\delta(y) \leq S_J^+(y)$$

and

$$\widehat{S}_J^{\pm}(0) = 2\delta \pm \frac{1}{J+1}$$

and are bounded by

$$|\widehat{S}_J^{\pm}(j)| \leq \frac{1}{J+1} + \min\left(2\delta, \frac{1}{\pi|j|}\right) \tag{2.2}$$

for all $y \in \mathbb{R}/\mathbb{Z}$ and $0 \le |j| \le J$. Hence we can bound the characteristic functions from above by the Selberg magic functions and obtain

$$\begin{aligned} N_\omega(Q,\delta) &\le \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n \\ q\le Q}} \omega\left(\frac{\mathbf{a}}{q}\right) \prod_{r=1}^{R} S_J^+\left(qf_r\left(\frac{\mathbf{a}}{q}\right)\right) \\ &= \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n \\ q\le Q}} \omega\left(\frac{\mathbf{a}}{q}\right) \prod_{r=1}^{R} \left(\sum_{j_r=-J}^{J} \widehat{S}_J^+(j_r) e\left(j_r q f_r\left(\frac{\mathbf{a}}{q}\right)\right)\right) \\ &= \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n \\ q\le Q}} \omega\left(\frac{\mathbf{a}}{q}\right) \sum_{\substack{0\le|j_i|\le J \\ 1\le i\le R}} \left(\prod_{r=1}^{R} \widehat{S}_J^+(j_r)\right) e\left(\sum_{r=1}^{R} j_r q f_r\left(\frac{\mathbf{a}}{q}\right)\right). \end{aligned}$$

The terms with $j_i = 0$ for all $i = 1, ..., R$ contribute

$$\left(2\delta + \frac{1}{J+1}\right)^R N_0 = (2\delta)^R N_0 + O\left(\delta^{R-1}\frac{Q^{n+1}}{J} + \frac{Q^{n+1}}{J^R}\right),$$

with the implicit constant possibly depending on R and an upper bound for the diameter of $\operatorname{supp} w$. Bounding the characteristic function from below by the Selberg magic functions yields a similar result, such that we conclude

$$\begin{aligned} |N_\omega(Q,\delta) - (2\delta)^R N_0| &\ll \delta^{R-1}\frac{Q^{n+1}}{J} + \frac{Q^{n+1}}{J^R} \\ &\quad + \sum_{\substack{1\le|j_i|\le J \\ 1\le i\le R \\ \mathbf{j}\neq\mathbf{0}}} \left(\prod_{r=1}^{R} b_{j_r}\right) \left|\sum_{\substack{\mathbf{a}\in\mathbb{Z}^n \\ q\le Q}} \omega\left(\frac{\mathbf{a}}{q}\right) e\left(\sum_{r=1}^{R} j_r q f_r\left(\frac{\mathbf{a}}{q}\right)\right)\right|, \end{aligned} \tag{2.3}$$

where

$$b_{j_r} := \frac{1}{J+1} + \min\left(2\delta, \frac{1}{\pi|j_r|}\right)$$

is the bound for the Selberg magic functions given in (2.2). Via the Poisson summation formula we can rewrite

$$\begin{aligned} &\sum_{\mathbf{a}\in\mathbb{Z}^n} \omega\left(\frac{\mathbf{a}}{q}\right) e\left(\sum_{r=1}^{R} j_r q f_r\left(\frac{\mathbf{a}}{q}\right)\right) \\ &= \sum_{\mathbf{k}\in\mathbb{Z}^n} \int_{\mathbb{R}^n} \omega\left(\frac{\mathbf{x}}{q}\right) e\left(\sum_{r=1}^{R} j_r q f_r\left(\frac{\mathbf{x}}{q}\right) - \mathbf{k}\cdot\mathbf{x}\right) \mathrm{d}\mathbf{x} \\ &= q^n \sum_{\mathbf{k}\in\mathbb{Z}^n} I(q;\mathbf{j};\mathbf{k}) \end{aligned} \tag{2.4}$$

with

$$I(q;\mathbf{j};\mathbf{k}) = \int_{\mathbb{R}^n} \omega(\mathbf{x}) e\left(\sum_{r=1}^{R} q j_r f_r(\mathbf{x}) - q\mathbf{k}\cdot\mathbf{x}\right) \mathrm{d}\mathbf{x}.$$

We now make use of Condition 0.4. By asssumption the Hessian matrix $H_{t_1 f_1 + \cdots + t_R f R}(\mathbf{x}_0)$ has a non-vanishing minor of size $n-s$, i.e. we can find s columns $i_1, ..., i_s$ and rows $j_1, ..., j_s$ such that after deleting them, the resulting matrix is invertible. Since the Hessian matrix is symmetric, we can choose $\{i_1, ..., i_s\} = \{j_1, ..., j_s\}$ in this case. Consider the functions

$$\vartheta_{\boldsymbol{\nu}} \colon \mathbb{R}^R \setminus \{\mathbf{0}\} \to \mathbb{R}, \mathbf{t} \mapsto \det \left(\frac{\partial^2 (t_1 f_1 + \cdots + t_R f_R)}{\partial x_\nu \partial x_\mu} \right)_{\substack{1 \le \nu, \mu \le n \\ \nu, \mu \notin \boldsymbol{\nu}}} (\mathbf{x}_0)$$

for $\boldsymbol{\nu} \in [n]^s$, where $[n]^s$ denotes the set of s-subsets of $\{1, ..., n\}$. Given Condition 0.4 the preimages $\{\vartheta_{\boldsymbol{\nu}}^{-1}(\mathbb{R} \setminus \{0\}) \mid \boldsymbol{\nu} \in [n]^s\}$ form an open cover of $\mathbb{R}^R \setminus \{\mathbf{0}\}$. For any fixed $\mathbf{t}$ the rank condition is invariant under scaling with a linear factor $a \in \mathbb{R} \setminus \{0\}$, i.e. if $\mathbf{t}$ belongs to $\vartheta_{\boldsymbol{\nu}}^{-1}(\mathbb{R} \setminus \{0\})$ so does $a\mathbf{t} = (at_1, ..., at_R)$. Therefore we can assume $\mathbf{t}$ to be normalized in the sense $|\mathbf{t}| = 1$. For every $\boldsymbol{\nu} \in [n]^s$ let $\tilde{T}_{\boldsymbol{\nu}} = \vartheta_{\boldsymbol{\nu}}^{-1}(\mathbb{R} \setminus \{0\}) \cap B_1^R(\mathbf{0})$, then $\{\tilde{T}_{\boldsymbol{\nu}} \mid \boldsymbol{\nu} \in [n]^s\}$ is an open cover of $B_1^R(\mathbf{0})$. Since $B_1^R(\mathbf{0})$ is compact and Hausdroff, it is also normal, hence the open cover $\{\tilde{T}_{\boldsymbol{\nu}} \mid \boldsymbol{\nu} \in [n]^s\}$ admits a shrinkage. That is an open cover $\{T'_{\boldsymbol{\nu}} \mid \nu \in [n]^s\}$ such that $T_{\boldsymbol{\nu}} := \overline{T'_{\boldsymbol{\nu}}} \subseteq \tilde{T}_i$ for $\boldsymbol{\nu} \in [n]^s$.

To find a bound for the last term in (2.3) it suffices to find an upper bound for

$$N^{(r;\boldsymbol{\epsilon};\boldsymbol{\nu})}(Q, \delta) = \sum_{\substack{1 \le j_r \le J \\ 0 \le j_s \le j_r \\ (\mathbf{j}/j_r) \in T_{\boldsymbol{\nu}}}} \left(\prod_{r=1}^R b_{j_r} \right) \left| \sum_{q \le Q} q^n \sum_{\mathbf{k} \in \mathbb{Z}^n} I(q; (\epsilon_1 j_1, ..., \epsilon_R j_R); \mathbf{k}) \right| \tag{2.5}$$

for each $1 \le r \le R$, $\boldsymbol{\epsilon} \in \{-1, 1\}^R$ and $\boldsymbol{\nu} \in [n]^s$. The arguments turn out to be identical for all $(r; \boldsymbol{\epsilon}; \boldsymbol{\nu})$, since different choices of r or ϵ admit only to relabeling and the choice of $\boldsymbol{\nu}$ is merely an exercise in notation. Therefore we only present the details for $N^{(1;(1,...,1);\mathcal{I}_s^n)}(Q, \delta)$, where $\mathcal{I}_s^n = [n] - [n-s] = \{n-s+1, ..., n\}$. Note that the same upper bound in fact holds for all $N^{(r;\boldsymbol{\epsilon};\boldsymbol{\nu})}$.

To $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ let $\widehat{\mathbf{x}} = (x_1, ..., x_{n-s}) \in \mathbb{R}^{n-s}$ and define functions

$$\widehat{f}_{j,\mathbf{y}} \colon \mathbb{R}^{n-s} \to \mathbb{R}, \widehat{\mathbf{x}} \mapsto f_j(\widehat{\mathbf{x}}, \mathbf{y})$$

for $\mathbf{y}$ in a sufficiently small neighborhood $U(x_{0,n-s+1}, ...x_{0,n})$ of $(x_{0,n-s+1}, ..., x_{0,n}) \in \mathbb{R}^s$. Define further

$$\widehat{G}_{\mathbf{y}}(\widehat{\mathbf{x}}, \mathbf{t}) = \widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{x}}) + \sum_{r=2}^R t_r \widehat{f}_{r,\mathbf{y}}(\widehat{\mathbf{x}})$$

and consider the continuous function

$$\psi \colon U(x_{0,n-s+1}, ..., x_{0,n}) \to \mathbb{R}, \mathbf{y} \mapsto \det \left(\frac{\partial^2 (f_1 + t_2 f_{2+} \cdots + t_R f_R)}{\partial x_\nu \partial x_\mu} \right)_{1 \le \nu, \mu \le n-s} (\widehat{\mathbf{x}}, \mathbf{y}).$$

For a suitable $\varepsilon' > 0$ we have $\psi(y) \neq 0$ for $\mathbf{y} \in B_\varepsilon(x_{0,n-s+1}, ..., x_{0,n})$, since $\psi(x_{0,n-s+1}, ..., x_{0,n})$ is non-zero by construction. Take $0 < \varepsilon_1 < \varepsilon'$ sufficiently small, then on the compact set $\overline{\mathscr{Y}}$ with $\mathscr{Y} = B_{\varepsilon_1}(x_{0,n-s}, ..., x_{0,n})$ we have

$$c'_1 \le \left| \det \left(\frac{\partial^2 (f_1 + t_2 f_2 + \cdots + t_R f_R)}{\partial x_\nu \partial x_\mu} \right)_{1 \le \nu, \mu \le n-s} (\widehat{\mathbf{x}}_0, \mathbf{y}) \right| \le c'_2$$

with constants $0 < c_1', c_2'$ for all $\mathbf{t} \in T_{\mathcal{I}_s^n}$. Now $\widehat{G}_{\mathbf{y}}$ satisfies the conditions of Lemmas 1.4 and 1.5 for $\mathcal{F}_2 = \overline{\mathscr{Y}} \times T_{\mathcal{I}_s^n}$, i.e. there are constants $\tau_{(1;(1,...,1);\mathcal{I}_s^n)} > 0$ and $c_1, c_2 > 0$ such that

$$c_1 \leq \left| \det \left(\frac{\partial^2 (f_1 + t_2 f_2 \cdots + t_R f_R)}{\partial x_\nu \partial x_\mu} \right)_{1 \leq \nu, \mu \leq n-s} (\widehat{\mathbf{x}}, \mathbf{y}) \right| \leq c_2 \tag{2.6}$$

for all $\mathbf{t} \in T_{\mathcal{I}_s^n}$, $\mathbf{y} \in \overline{\mathscr{Y}}$ and $\widehat{\mathbf{x}} \in B_{2\tau_{(1;(1,...,1);\mathcal{I}_s^n)}}(\widehat{\mathbf{x}}_0)$. Moreover, the map

$$\widehat{\mathbf{x}} \mapsto \left(\widehat{f}_{1,\mathbf{y}} + \sum_{r=2}^{R} t_r \widehat{f}_{r,\mathbf{y}} \right) (\widehat{\mathbf{x}})$$

is a $\mathcal{C}^{\ell-1}$ diffeomorphism on $\overline{B_{2\tau_{(1;(1,...,1);\mathcal{I}_s^n)}}(\widehat{\mathbf{x}}_0)}$ for all $\mathbf{t} \in T_{\mathcal{I}_s^n}$ and $\mathbf{y} \in \overline{\mathscr{Y}}$. Define $\tau_{(r;\epsilon;\boldsymbol{\nu})}$ in the same way for $1 \leq r \leq R$, $\epsilon \in \{\pm 1\}^R$ and $\boldsymbol{\nu} \in [n]^s$ and let

$$0 < \tau \leq \min_{\substack{1 \leq r \leq R \\ \epsilon \in \{\pm 1\}^R \\ \boldsymbol{\nu} \in [n]^s}} \tau_{(r;\epsilon;\boldsymbol{\nu})}$$

be sufficiently small (such that Lemma 3.3 is going to be applicable). For this choice of τ with Lemma 1.5 we find constants $0 < \kappa < \tau$ and ρ such that

$$\operatorname{dist} \left(\partial \left(\nabla \left(\widehat{f}_{1,\mathbf{y}} + \sum_{r=2}^{R} t_r \widehat{f}_{r,\mathbf{y}} \right) (B_\tau(\widehat{\mathbf{x}}_0)) \right), \partial \left(\nabla \left(\widehat{f}_{1,\mathbf{y}} + \sum_{r=2}^{R} t_r \widehat{f}_{r,\mathbf{y}} \right) (B_\kappa(\widehat{\mathbf{x}}_0)) \right) \right) \geq 2\rho \tag{2.7}$$

for all $\mathbf{t} \in T_{\mathcal{I}_s^n}$ and $\mathbf{y} \in \overline{\mathscr{Y}}$. Note that $\varepsilon_0 < 2\tau$ is a sufficient choice in Theorem 0.5.

Let $\mathscr{D} = B_\tau(\widehat{\mathbf{x}}_0)$ and let $\omega \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ be a non-negative weight function such that for any $\mathbf{y} \in \mathscr{Y}$ the closure of

$$U_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{R}^{n-s} \mid \omega(\mathbf{x}, \mathbf{y}) \neq 0\}$$

is contained in $B_\kappa(\widehat{\mathbf{x}}_0)$. Define the function $\widehat{F}_{\mathbf{y},\mathbf{j}} = \widehat{f}_{1,\mathbf{y}} + (j_2/j_1)\widehat{f}_{2,\mathbf{y}} + \cdots + (j_R/j_1)\widehat{f}_{R,\mathbf{y}}$ and $V_{\mathbf{y},\mathbf{j}} = \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(U_{\mathbf{y}})$. Since $0 \leq j_r/j_1 \leq 1$ for $2 \leq r \leq R$ we know that $\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}$ is a diffeomorphism on $U_{\mathbf{y}}$ and $\mathscr{D}$.

Lemma 2.1. *The functions $\widehat{f}_{r,\mathbf{y}}$ for $1 \leq r \leq R$ are bounded on $\overline{B_{2\tau}(\widehat{\mathbf{x}}_0)}$ for all $\mathbf{y} \in \overline{\mathscr{Y}}$ and the bounds are independend of $\mathbf{y}$. Additionally, there is $L \in \mathbb{N}$ such that for all $\mathbf{t} \in T_{\mathcal{I}_s^n}$ and all $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$ we have*

$$\left| \frac{\partial^{i_1 + \cdots + i_{n-s}} (t_1 f_1 + \cdots t_R f_R)}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}} (\mathbf{x}) \right| \leq L$$

on $\overline{B_{2\tau}(\widehat{\mathbf{x}}_0)} \times \overline{\mathscr{Y}}$ and $\rho \leq L$ for ρ in (2.7).

Proof. By assumption $f_r(\mathbf{x})$ is smooth, hence on the compact domain $\overline{B_{2\tau}(\widehat{\mathbf{x}}_0)} \times \overline{\mathscr{Y}}$ it attains a maximum M_r. Now by definition $\widehat{f}_{r,\mathbf{y}}(\widehat{\mathbf{x}}) = f_r(\widehat{\mathbf{x}}, \mathbf{y})$, hence for any $\mathbf{y} \in \overline{\mathscr{Y}}$

$$|\widehat{f}_{r,\mathbf{y}}(\widehat{\mathbf{x}})| < M_r$$

on $\overline{B_{2\tau}(\widehat{\mathbf{x}}_0)}$. For the derivatives note that all domains of definition, i.e. $\overline{B_{2\tau}(\widehat{\mathbf{x}}_0)}$, $\overline{\mathscr{Y}}$ and $T_{\mathcal{I}_s^n}$ are compact and all the relevant functions depend at least continuously on $\mathbf{x} = (\widehat{\mathbf{x}}, \mathbf{y})$ and $\mathbf{t}$, hence for any given suitable $(i_1, ..., i_{n-s})$ there exists a maximum

$$M_{(i_1,...,i_{n-s})} = \max_{\mathbf{x},\mathbf{t}} \left| \frac{\partial^{i_1+\cdots+i_{n-s}}(t_1 f_1 + \cdots t_R f_R)}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\mathbf{x}) \right|.$$

Since there are only finitely many suitable choices for $(i_1, ..., i_{n-s})$ we can also take the maximum over them and define L to be the smallest natural number such that

$$L \geq \max_{(i_1,...,i_{n-s})} M_{(i_1,...,i_{n-s})}$$

and

$$L \geq \rho.$$

□

Specifically we have that $V_{\mathbf{y},\mathbf{j}} \subseteq [-L, L]^{n-s}$ independently of $\mathbf{y}$ and $\mathbf{j}$.

We split the set of $\mathbf{k} \in \mathbb{Z}^n$ into three disjoint subsets as follows. Let $\widehat{\mathbf{k}} = (k_1, ..., k_{n-s})$ and $\mathbf{k}^* = (k_{n-s+1}, ..., k_n)$. Let

$$D(\widehat{\mathbf{k}}, \mathbf{j}) = \min_{\mathbf{y} \in \overline{\mathscr{Y}}} \operatorname{dist}\left(\frac{\widehat{\mathbf{k}}}{j_1}, V_{\mathbf{y},\mathbf{j}}\right).$$

Now define

$$\mathscr{K}_{\mathbf{j};1} = \left\{ \mathbf{k} \in \mathbb{Z}^n \,\middle|\, \frac{\widehat{\mathbf{k}}}{j_1} \in \bigcup_{y \in \overline{\mathscr{Y}}} V_{y,\mathbf{j}}, |\mathbf{k}^*| \leq 2j_1 L \right\},$$

$$\mathscr{K}_{\mathbf{j};2} = \left\{ \mathbf{k} \in \mathbb{Z}^n \,\middle|\, D(\widehat{\mathbf{k}}, \mathbf{j}) \geq \rho \right\} \cup \{\mathbf{k} \in \mathbb{Z}^n \mid |\mathbf{k}^*| > 2j_1 L\}$$

and

$$\mathscr{K}_{\mathbf{j};3} = \left\{ \mathbf{k} \in \mathbb{Z}^n \,\middle|\, D(\widehat{\mathbf{k}}, \mathbf{j}) < \rho, \frac{\widehat{\mathbf{k}}}{j_1} \notin \bigcup_{y \in \overline{\mathscr{Y}}} V_{y,\mathbf{j}}, |\mathbf{k}^*| \leq 2j_1 L \right\}.$$

Remark. This decomposition is inspired by Huang's original work and is designed to reflect the cases where the phase function of the oscillatory integral $I(q; \mathbf{j}; \mathbf{k})$ has stationary ($\mathscr{K}_{\mathbf{j};1}$ and $\mathscr{K}_{\mathbf{j};3}$) or non-stationary phase ($\mathscr{K}_{\mathbf{j};2}$). Note that an important difference lies with the distinction of variables. Huang's original conditions are expressed for $\widehat{\mathbf{k}}$ in the integral over $\widehat{\mathbf{x}}$, since by choice of the variables the stronger curvature condition holds there. When the remaining $\mathbf{k}^*$ are big enough, the integral over $\mathbf{y}$ will show rapid decay, hence the two components of $\mathscr{K}_{\mathbf{j};2}$.

For each $1 \leq i \leq 3$, we let

$$N_i = \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_2, ..., j_R \leq j_1 \\ (\mathbf{j}/j_1) \in T_{\mathcal{I}_s^n}}} \left(\prod_{r=1}^{R} b_{j_r} \right) \left| \sum_{q \leq Q} q^n \sum_{\mathbf{k} \in \mathscr{K}_{\mathbf{j};i}} I(q; \mathbf{j}; \mathbf{k}) \right| \tag{2.8}$$

such that

$$N^{(1;(1,\dots,1);\mathcal{I}_s^n)}(Q,\delta) \ll N_1 + N_2 + N_3 \tag{2.9}$$

and proceed to bound each N_i seperately.

3. Bounds for N_1, N_2 and N_3

Lemma 3.1. *For any $K > 0$ we have that*

$$\left\{ \frac{\widehat{\mathbf{k}}}{j_1} \mid \widehat{\mathbf{k}} \in \mathbb{Z}^{n-s}, D(\widehat{\mathbf{k}}, \mathbf{j}) < K \right\} \subseteq [-L-K, L+K]^{n-s},$$

where L is defined as in Lemma 2.1.

Proof. For $(\widehat{\mathbf{k}}/j_1) \in [-L, L]^{n-s}$ the inclusion is obvious, so let $(\widehat{\mathbf{k}}/j_1) \notin [-L, L]^{n-s}$. We have

$$\begin{aligned} K > D(\widehat{\mathbf{k}}, \mathbf{j}) &= \min_{\mathbf{y} \in \overline{\mathscr{Y}}} \inf_{\mathbf{z} \in V_{\mathbf{y},\mathbf{j}}} \left| \frac{\widehat{\mathbf{k}}}{j_1} - \mathbf{z} \right| \geq \min_{\mathbf{z} \in [-L,L]^{n-s}} \left| \frac{\widehat{\mathbf{k}}}{j_1} - \mathbf{z} \right| \\ &\geq \min_{\mathbf{z} \in [-L,L]^{n-s}} \left| \frac{|\widehat{\mathbf{k}}|}{j_1} - |\mathbf{z}| \right| = \frac{|\widehat{\mathbf{k}}|}{j_1} - L, \end{aligned}$$

hence $K + L > |\widehat{\mathbf{k}}/j_1|$ as desired.

□

Case $\mathbf{k} \in \mathscr{K}_{\mathbf{j};2}$.

Let

$$D_1(\widehat{\mathbf{k}}, \mathbf{j}) = j_1 D(\widehat{\mathbf{k}}, \mathbf{j}) = \min_{\mathbf{y} \in \overline{\mathscr{Y}}} \operatorname{dist}(\widehat{\mathbf{k}}, j_1 V_{\mathbf{y},\mathbf{j}}).$$

For a fixed $\mathbf{k}^* \in \mathbb{Z}^s$, consider the integral

$$\int_{\mathbb{R}^{n-s}} \omega(\widehat{\mathbf{x}}, \mathbf{y}) e\left(q j_1 \left(\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}} + \mathbf{k}^* \cdot \mathbf{y}}{j_1} \right) \right) \mathrm{d}\widehat{\mathbf{x}}.$$

with $(\widehat{\mathbf{k}}, \mathbf{k}^*) \in \mathscr{K}_{\mathbf{j};2}$ and $D(\widehat{\mathbf{k}}, \mathbf{j}) \geq \rho$. Let

$$\varphi_{\mathbf{y},1}(\widehat{\mathbf{x}}) = \frac{j_1 \widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{x}}) + \cdots + j_R \widehat{f}_{R,\mathbf{y}}(\widehat{\mathbf{x}}) - \widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}} - \mathbf{k}^* \cdot \mathbf{y}}{D_1(\widehat{\mathbf{k}}, \mathbf{j})}$$

and $\lambda_1 = q D_1(\widehat{\mathbf{k}}, \mathbf{j})$. Then by definition of $V_{\mathbf{y},\mathbf{j}}$

$$|\nabla \varphi_{\mathbf{y},1}(\widehat{\mathbf{x}})| = \frac{|j_1 \nabla \widehat{f}_{1,\mathbf{y}} + \cdots + j_R \nabla \widehat{f}_{R,\mathbf{y}}(\widehat{\mathbf{x}}) - \widehat{\mathbf{k}}|}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \geq 1$$

for $\widehat{\mathbf{x}} \in U_{\mathbf{y}}$. Let $U_{\mathbf{y}}^+ \subseteq \mathbb{R}^{n-s}$ be an open set such that $\overline{U_{\mathbf{y}}} \subseteq U_{\mathbf{y}}^+ \subseteq \mathscr{D}$, $V_{\mathbf{y},\mathbf{j}}^+ = \nabla F_{\mathbf{j}}\left(U_{\mathbf{y}}^+\right) \subseteq [-2L, 2L]^{n-s}$,

$$\min_{\mathbf{z} \in \overline{U_{\mathbf{y}}}} |\widehat{\mathbf{x}} - \mathbf{z}| < \frac{\operatorname{dist}(\partial \mathscr{D}, \partial U_{\mathbf{y}})}{4} \tag{3.1}$$

for any $\widehat{\mathbf{x}} \in U_y^+$ and

$$\min_{\mathbf{z}\in\overline{U_{\mathbf{y}}}} \max_{\mathbf{t}\in T_{\mathcal{I}_s^n}} |\nabla(t_1 f_{1,\mathbf{y}} + \cdots + t_R f_{R,\mathbf{y}})(\widehat{\mathbf{x}}) - \nabla(t_1 f_{1,\mathbf{y}} + \cdots + t_R f_{R,\mathbf{y}})(\mathbf{z})| < \frac{\rho}{2}$$

for any $\widehat{\mathbf{x}} \in U_{\mathbf{y}}^+$. Then we have

$$\nabla\varphi_{\mathbf{y},1}(\widehat{\mathbf{x}}) \geq \frac{1}{2}$$

for all $\widehat{\mathbf{x}} \in U_{\mathbf{y}}^+$.

Now assume that $|\mathbf{k}^*| \leq D_1(\widehat{\mathbf{k}}, \mathbf{j})$ or $|\mathbf{k}^*| \leq 2j_1 L$.

Lemma 3.2. *Let $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$. Then for all $\widehat{\mathbf{x}} \in U_{\mathbf{y}}^+$ we have*

$$\left| \frac{\partial^{i_1+\cdots+i_{n-s}}\varphi_{\mathbf{y},1}}{\partial_{x_1}^{i_1}\cdots\partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| \ll 1,$$

where the implicit constant depends only on $(i_1, ..., i_{n-s})$, ρ, τ, ε_1 and upper bounds for (the absolute values of) finitely many derivatives of f_r on $\mathscr{D} \times \mathscr{Y}$ for $1 \leq r \leq R$.

Proof. Choose $C > 0$ such that

$$\frac{1}{C} \max_{\substack{\mathbf{t}\in T_{\mathcal{I}_s^n} \\ \mathbf{y}\in\overline{\mathscr{Y}} \\ \widehat{\mathbf{x}}\in\overline{U_{\mathbf{y}}}}} |\nabla(t_1 \widehat{f}_{1,\mathbf{y}} + \cdots + t_R \widehat{f}_{R,\mathbf{y}})(\widehat{\mathbf{x}})| < \frac{1}{2}.$$

and assume $j_1 C \leq |\widehat{\mathbf{k}}|$. Then we have

$$\left| \frac{\widehat{\mathbf{k}}}{|\widehat{\mathbf{k}}|} - \frac{j_1\mathbf{z}}{|\widehat{\mathbf{k}}|} \right| \geq 1 - \frac{j_1\mathbf{z}}{|\widehat{\mathbf{k}}|} > \frac{1}{2}$$

for all $\mathbf{z} \in \bigcup_{\mathbf{y}\in\mathscr{Y}} V_{\mathbf{y},\mathbf{j}}$ and hence

$$\frac{1}{|\widehat{\mathbf{k}}|} D_1(\widehat{\mathbf{k}}, \mathbf{j}) = \min_{y\in\overline{\mathscr{Y}}} \operatorname{dist}\left(\frac{\widehat{\mathbf{k}}}{|\widehat{\mathbf{k}}|}, \frac{j_1 V_{\mathbf{y},\mathbf{j}}}{|\widehat{\mathbf{k}}|}\right) \geq \frac{1}{2}.$$

Therefore

$$\begin{aligned} |\varphi_{\mathbf{y},1}(\widehat{\mathbf{x}})| &\leq \left| \frac{\frac{j_1}{|\widehat{\mathbf{k}}|}\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{x}}) + \cdots + \frac{j_R}{|\widehat{\mathbf{k}}|}\widehat{f}_{R,\mathbf{y}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{|\widehat{\mathbf{k}}|}\cdot\widehat{\mathbf{x}}}{\frac{1}{|\widehat{\mathbf{k}}|}D_1(\widehat{\mathbf{k}}, \mathbf{j})} \right| + \left| \frac{\mathbf{k}^*\cdot\mathbf{y}}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \right| \\ &\leq 2\left(\frac{j_1}{|\widehat{\mathbf{k}}|}|\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{x}})| + \cdots + \frac{j_R}{|\widehat{\mathbf{k}}|}|\widehat{f}_{R,\mathbf{y}}(\widehat{\mathbf{x}})| + \left|\frac{\widehat{\mathbf{k}}}{|\widehat{\mathbf{k}}|}\right| \cdot |\widehat{\mathbf{x}}| \right) + \frac{|\mathbf{k}^*|}{D_1(\widehat{\mathbf{k}}, \mathbf{j})}|\mathbf{y}|. \end{aligned}$$

Since $|\widehat{\mathbf{k}}| > j_1 C$, we have $\frac{j_r}{|\widehat{\mathbf{k}}|} \leq 1$ for $1 \leq r \leq R$. Let $M_r > 0$ be the bound for $|\widehat{f}_{r,\mathbf{y}}|$ established in 2.1 on $U_{\mathbf{y}}^+$ for $1 \leq r \leq R$ and $S = M_1 + \cdots + M_R$. Now by assumption we either have

$$\frac{|\mathbf{k}^*|}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \leq 1 \qquad \text{or} \qquad \frac{|\mathbf{k}^*|}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \leq \frac{2L}{\rho},$$

hence the final term is also bounded independently of $\mathbf{j}$. Then

$$|\varphi_{\mathbf{y},1}(\widehat{\mathbf{x}})| \leq 2(S+\tau)+\varepsilon_1 \qquad \text{or} \qquad |\varphi_{y,1}(\widehat{\mathbf{x}})| \leq 2(S+\tau)+\frac{2L}{\rho}\varepsilon_1$$

as desired. If $|\widehat{\mathbf{k}}| < Cj_1$ we immediately conclude

$$|\varphi_{\mathbf{y},1}(\widehat{\mathbf{x}})| \leq \left| \frac{\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{x}}) + \frac{j_2}{j_1}\widehat{f}_{2,\mathbf{y}}(\widehat{\mathbf{x}}) \cdots + \frac{j_R}{j_1}\widehat{f}_{R,\mathbf{y}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{j_1}\widehat{\mathbf{x}}}{\rho} \right| + \left| \frac{\mathbf{k}^* \cdot \mathbf{y}}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \right| \ll 1$$

for all $\widehat{\mathbf{x}} \in U_{\mathbf{y}}^+$ with the same argument for the final term as above. For the first partial derivatives consider

$$\left| \frac{\partial \varphi_{\mathbf{y},1}}{\partial x_i}(\widehat{\mathbf{x}}) \right| = \left| \frac{j_1 \frac{\partial \widehat{f}_{1,\mathbf{y}}}{\partial x_i}(\widehat{\mathbf{x}}) + \cdots + j_R \frac{\partial \widehat{f}_{R,\mathbf{y}}}{\partial x_i}(\widehat{\mathbf{x}}) - k_i}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \right|$$

which can be treated with a similar argument. For higher partial derivatives the terms $\widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}}$ vanish and the desired result follows easily with

$$\begin{aligned} \left| \frac{\partial^{i_1+\cdots+i_{n-s}} \varphi_{\mathbf{y},1}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-1}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| &= \left| \frac{j_1 \frac{\partial^{i_1+\cdots+i_{n-s}} \widehat{f}_{1,\mathbf{y}}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) + \cdots + j_R \frac{\partial^{i_1+\cdots+i_{n-s}} \widehat{f}_{R,\mathbf{y}}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}})}{D_1(\widehat{\mathbf{k}}, \mathbf{j})} \right| \\ &\leq \frac{1}{\rho} \left(\left| \frac{\partial^{i_1+\cdots+i_{n-s}} \widehat{f}_{1,\mathbf{y}}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| + \sum_{r=2}^{R} \frac{j_r}{j_1} \left| \frac{\partial^{i_1+\cdots+i_{n-s}} \widehat{f}_{r,\mathbf{y}}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| \right). \end{aligned}$$

□

Therefore with Lemma 1.2 for $\varphi = \varphi_{\mathbf{y},1}$ and $\lambda = \lambda_1$ as chosen above we have

$$\int_{\mathbb{R}^{n-s}} \omega(\widehat{\mathbf{x}}, \mathbf{y}) e\left(qj_1 \left(\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}} + \mathbf{k}^* \cdot \mathbf{y}}{j_1} \right) \right) \mathrm{d}\widehat{\mathbf{x}} \ll (qD_1(\widehat{\mathbf{k}}, \mathbf{j}))^{-\ell+1} \tag{3.2}$$

and hence

$$I(q;\mathbf{j};\mathbf{k}) \ll \int_{\mathscr{Y}} (qD_1(\widehat{\mathbf{k}}, \mathbf{j}))^{-\ell+1} \mathrm{d}y \ll_{\varepsilon_1} (qD_1(\widehat{\mathbf{k}}, \mathbf{j}))^{-\ell+1}. \tag{3.3}$$

Now assume that $|\mathbf{k}^*| > D_1(\widehat{\mathbf{k}}, \mathbf{j})$ and $|\mathbf{k}^*| > 2j_1 L$. Consider

$$\int_{\mathbb{R}^s} \omega(\widehat{\mathbf{x}}, \mathbf{y}) e\left(qj_1 \left(\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}} + \mathbf{k}^* \cdot \mathbf{y}}{j_1} \right) \right) \mathrm{d}\mathbf{y}.$$

Let $\lambda_2 = q|\mathbf{k}^*|$ and

$$\varphi_{\widehat{\mathbf{x}},2}(\mathbf{y}) = \frac{j_1 f_1(\widehat{\mathbf{x}}, y) + \cdots + j_R f_R(\widehat{\mathbf{x}}, y) - \widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}} - \mathbf{k}^* \cdot \mathbf{y}}{|\mathbf{k}^*|}.$$

Observe that

$$\begin{aligned} |\nabla\varphi_{\widehat{\mathbf{x}},2}(\mathbf{y})| &= \left|\frac{j_1\nabla f_1(\widehat{\mathbf{x}},\mathbf{y})+\cdots+j_R\nabla f_R(\widehat{\mathbf{x}},\mathbf{y})-\mathbf{k}^*}{|\mathbf{k}^*|}\right| \\ &\geq \left|\frac{\mathbf{k}^*}{|\mathbf{k}^*|}-\frac{\nabla f_1(\widehat{\mathbf{x}},\mathbf{y})+\frac{j_2}{j_1}\nabla f_2(\widehat{\mathbf{x}},\mathbf{y})+\cdots+\frac{j_R}{j_1}\nabla f_R(\widehat{\mathbf{x}},\mathbf{y})}{2L}\right| \\ &\geq \frac{1}{2}. \end{aligned}$$

Consider further that

$$\begin{aligned} |\varphi_{\widehat{\mathbf{x}},2}(\mathbf{y})| &= \left|\frac{j_1 f_1(\widehat{\mathbf{x}},\mathbf{y})+\cdots j_R f_R(\widehat{\mathbf{x}},\mathbf{y})-\widehat{\mathbf{k}}\cdot\widehat{\mathbf{x}}-\mathbf{k}^*\cdot\mathbf{y}}{|\mathbf{k}^*|}\right| \\ &\leq \frac{|f_1(\widehat{\mathbf{x}},\mathbf{y})|+\frac{j_2}{j_1}|f_2(\widehat{\mathbf{x}},\mathbf{y})|\cdots\frac{j_R}{j_1}|f_R(\widehat{\mathbf{x}},\mathbf{y})|+\frac{|\widehat{\mathbf{k}}|}{j_1}\cdot|\widehat{\mathbf{x}}|}{\frac{1}{j_1}|\mathbf{k}^*|}+|\mathbf{y}| \\ &\leq \frac{|f_1(\widehat{\mathbf{x}},\mathbf{y})|+\frac{j_2}{j_1}|f_2(\widehat{\mathbf{x}},\mathbf{y})|\cdots\frac{j_R}{j_1}|f_R(\widehat{\mathbf{x}},\mathbf{y})|+(L+\frac{1}{j_1}|\mathbf{k}^*|)\cdot|\widehat{\mathbf{x}}|}{\frac{1}{j_1}|\mathbf{k}^*|}+|\mathbf{y}| \\ &\leq \frac{|f_1(\widehat{\mathbf{x}},\mathbf{y})|+\frac{j_2}{j_1}|f_2(\widehat{\mathbf{x}},y)|\cdots\frac{j_R}{j_1}|f_R(\widehat{\mathbf{x}},y)|+L|\widehat{\mathbf{x}}|}{2L}+|\widehat{\mathbf{x}}|+|\mathbf{y}|, \end{aligned}$$

and

$$\begin{aligned} |\nabla\varphi_{\widehat{\mathbf{x}},2}(\mathbf{y})| &= \left|\frac{j_1\nabla f_1(\widehat{\mathbf{x}},\mathbf{y})+\cdots+j_R\nabla f_R(\widehat{\mathbf{x}},\mathbf{y})-\mathbf{k}^*}{|\mathbf{k}^*|}\right| \\ &\leq \frac{j_1}{|\mathbf{k}^*|}\left|\nabla f_1(\widehat{\mathbf{x}},\mathbf{y})\right|+\cdots+\frac{j_R}{|\mathbf{k}^*|}\left|\nabla f_R(\widehat{\mathbf{x}},\mathbf{y})\right|+1, \end{aligned}$$

hence with analogous arguments as above we conclude that

$$\left|\frac{\partial^{i_1+\cdots+i_{n-s}}\varphi_{\widehat{\mathbf{x}},2}}{\partial_{x_1}^{i_1}\cdots\partial_{x_{n-1}}^{i_{n-s}}}(\mathbf{y})\right| \ll 1,$$

where the implicit constants only depend on $i,\rho,\tau,\varepsilon_1$ and upper bounds for (the absolute values of) finitely many derivatives of f_r on $\mathscr{D}\times\mathscr{Y}$. So Lemma 1.2 applies in the s-dimensional case, hence

$$\int_{\mathbb{R}^s}\omega(\widehat{\mathbf{x}},\mathbf{y})e\left(qj_1\left(\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}})-\frac{\widehat{\mathbf{k}}\cdot\widehat{\mathbf{x}}+\mathbf{k}^*\cdot\mathbf{y}}{j_1}\right)\right)\mathrm{d}\mathbf{y} \ll (q|\mathbf{k}^*|)^{-\ell+1} \tag{3.4}$$

and therefore

$$I(q;\mathbf{j};\mathbf{k}) \ll \int_{U_{\mathbf{y}}}(q|\mathbf{k}^*|)^{-\ell+1}\mathrm{d}\widehat{\mathbf{x}} \ll_\kappa (q|\mathbf{k}^*|)^{-\ell+1}. \tag{3.5}$$

Given these estimates we can split up the sum

$$\sum_{\mathbf{k}\in\mathscr{K}(\mathbf{j};2)} I(q;\mathbf{j};\mathbf{k}) = \sum_{\mathbf{k}\in\mathscr{K}(\mathbf{j};2)}\int_{\mathbb{R}^n}\omega(\widehat{\mathbf{x}},\mathbf{y})e\left(qj_1\left(\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}})-\frac{\widehat{\mathbf{k}}\cdot\widehat{\mathbf{x}}+\mathbf{k}^*\cdot\mathbf{y}}{j_1}\right)\right)\mathrm{d}\widehat{\mathbf{x}}\mathrm{d}\mathbf{y}$$

as follows

$$\sum_{(\widehat{\mathbf{k}},\mathbf{k}^*)\in\mathscr{K}(\mathbf{j};2)} I(q;\mathbf{j};(\widehat{\mathbf{k}},\mathbf{k}^*)) = \sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ D_1(\widehat{\mathbf{k}},\mathbf{j})\geq|\mathbf{k}^*|}} I(q;\mathbf{j},(\widehat{\mathbf{k}},\mathbf{k}^*)) + \sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ 2j_1L\geq|\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j})}} I(q;\mathbf{j},(\widehat{\mathbf{k}},\mathbf{k}^*))$$
$$+ \sum_{\substack{|\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j}) \\ |\mathbf{k}^*|>2j_1L}} I(q;\mathbf{j};(\widehat{\mathbf{k}},\mathbf{k}^*))$$

With (3.3) and (3.5) we obtain

$$\begin{aligned}
\sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ D_1(\widehat{\mathbf{k}},\mathbf{j})\geq|\mathbf{k}^*|}} I(q;\mathbf{j},(\widehat{\mathbf{k}},\mathbf{k}^*)) &\ll q^{-\ell+1} \sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ D_1(\widehat{\mathbf{k}},\mathbf{j})\geq|\mathbf{k}^*|}} D_1(\widehat{\mathbf{k}},\mathbf{j})^{-\ell+1} \\
&= q^{-\ell+1}\sum_{d=0}^{\infty} \sum_{\substack{D_1(\widehat{\mathbf{k}},\mathbf{j})\geq|\mathbf{k}^*| \\ 2^d j_1\rho\leq D_1(\widehat{\mathbf{k}},\mathbf{j})<2^{d+1}j_1\rho}} D_1(\widehat{\mathbf{k}},\mathbf{j})^{-\ell+1} \\
&\leq q^{-\ell+1}\sum_{d=0}^{\infty} \sum_{\substack{D_1(\widehat{\mathbf{k}},\mathbf{j})\geq|\mathbf{k}^*| \\ 2^d j_1\rho\leq D_1(\widehat{\mathbf{k}},\mathbf{j})<2^{d+1}j_1\rho}} \frac{1}{(2^d j_1\rho)^{\ell-1}} \\
&\ll q^{-\ell+1}\sum_{d=0}^{\infty} \frac{(j_1L+2^{d+1}j_1\rho)^n}{(2^d j_1\rho)^{\ell-1}} \\
&\ll_{L,n} q^{-\ell+1}
\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}
\sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ 2j_1L\geq|\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j})}} I(q;\mathbf{j},(\widehat{\mathbf{k}},\mathbf{k}^*)) &\ll q^{-\ell+1} \sum_{\substack{D(\widehat{\mathbf{k}},\mathbf{j})\geq\rho \\ 2j_1L\geq|\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j})}} D_1(\widehat{\mathbf{k}},\mathbf{j})^{-\ell+1} \\
&\leq q^{-\ell+1}\sum_{d=0}^{\infty} \sum_{\substack{2j_1L\geq|\mathbf{k}^*| \\ 2^d j_1\rho\leq D_1(\widehat{\mathbf{k}},\mathbf{j})<2^{d+1}j_1\rho}} D_1(\widehat{\mathbf{k}},\mathbf{j})^{-\ell+1} \\
&\leq q^{-\ell+1}\sum_{d=0}^{\infty} \sum_{\substack{2j_1L\geq|\mathbf{k}^*| \\ 2^d j_1\rho\leq D_1(\widehat{\mathbf{k}},\mathbf{j})<2^{d+1}j_1\rho}} \frac{1}{(2^d j_1\rho)^{\ell-1}} \\
&\ll q^{-\ell+1}\sum_{d=0}^{\infty} (2j_1L)^s \frac{(j_1L+2^{d+1}j_1\rho)^{n-s}}{(2^d j_1\rho)^{\ell-1}} \\
&\ll_{L,n} q^{-\ell+1}
\end{aligned} \tag{3.7}$$

and

$$\sum_{\substack{|\mathbf{k}^*|>2j_1L\\ |\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j})}} I(q;\mathbf{k};\mathbf{j}) \ll q^{-\ell+1} \sum_{\substack{|\mathbf{k}^*|>2j_1L\\ |\mathbf{k}^*|>D_1(\widehat{\mathbf{k}},\mathbf{j})}} |\mathbf{k}^*|^{-\ell+1} \tag{3.8}$$
$$\ll q^{-\ell+1} \sum_{|\mathbf{k}^*|>2j_1L} \frac{(j_1L+|\mathbf{k}^*|)^{n-s}}{|\mathbf{k}^*|^{\ell-1}}$$
$$\ll_{n,s} q^{-\ell+1} \sum_{|\mathbf{k}^*|>2j_1L} |\mathbf{k}^*|^{n-s-\ell+1}$$
$$\ll q^{-\ell+1} \int_{2j_1L}^{\infty} \frac{1}{t^{\ell-n}} \mathrm{d}t$$
$$\ll q^{-\ell+1} j_1^{n-\ell+1}.$$

Note that by assumption ℓ is sufficiently large. Here the implicit constants only depend on $L, \rho, n, \ell, s, \varepsilon_1, \kappa$ and upper bounds for (the absolute values of) finitely many derivatives of ω and f_r for $1 \leq r \leq R$ on $\mathscr{D} \times \mathscr{Y}$. Consequently we obtain

$$N_2 \ll \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_R\leq j_1\\ \mathbf{j}/j_1\in T_{\mathcal{I}_s^n}}} \left(\prod_{r=1}^{R}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)\right) j_1^{n-s-\ell} \sum_{q\leq Q} q^{n-\ell+1} \tag{3.9}$$
$$\ll \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_R\leq j_1}} \left(\prod_{r=1}^{R}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)\right) \int_1^Q q^{n-\ell+1}\mathrm{d}q$$
$$\ll \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_R\leq j_1}} \left(\prod_{r=1}^{R}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)\right) \int_1^Q \frac{1}{q}\mathrm{d}q$$
$$= \log Q \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_{R-1}\leq j_1}} \left(\left(\prod_{r=1}^{R-1}\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right) \sum_{0\leq j_R\leq j_1}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)\right)$$
$$= \log Q \sum_{1\leq j_1\leq J} \left(\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_1}\right)\right)\prod_{r=2}^{R}\sum_{0\leq j_r\leq j_1}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)\right)$$
$$\leq \log Q \prod_{r=1}^{R}\sum_{0\leq j_r\leq J}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j_r}\right)\right)$$
$$= \log Q \left(\sum_{0\leq j\leq J}\left(\frac{1}{J}+\min\left(\delta,\frac{1}{j}\right)\right)\right)^R$$
$$\ll \log Q(1+\log J)^R.$$

Case $\mathbf{k} \in \mathscr{K}_{\mathbf{j};3}$.

Let $\lambda = qj_1$ and

$$\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}) = \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{j_1} \cdot \widehat{\mathbf{x}} - \frac{\mathbf{k}^*}{j_1} \cdot \mathbf{y}.$$

For each $\mathbf{y}$ we know that $\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}$ is a diffeomorphism on $\mathscr{D}$, hence for any fixed $\mathbf{j}$ and any $\widehat{\mathbf{k}}$ with $(\widehat{\mathbf{k}}/j_1) \in \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\mathscr{D})$ we have a unique preimage

$$\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}} = (\nabla \widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}(\widehat{\mathbf{k}}/j_1).$$

This defines a critical point for $\varphi_{\mathbf{y}}$, since

$$\nabla \varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) = \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) - \frac{\widehat{\mathbf{k}}}{j_1} = \mathbf{0}.$$

Let $\mathscr{D}_+$ be an open set such that $\overline{\mathscr{D}} \subseteq \mathscr{D}_+ \subseteq B_{3\tau/2}(\widehat{\mathbf{x}}_0)$.

Lemma 3.3. *Let $\widehat{\mathbf{x}} \in \mathscr{D}_+ \setminus \{\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}\}$. Then*

$$\frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla \varphi_{\mathbf{y}}(\widehat{\mathbf{x}})|} \ll 1$$

where the implicit constant is independant of $\mathbf{y}$, $\mathbf{j}$ and $\widehat{\mathbf{k}}$.

Proof. By definition of $\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}$ as a preimage of $\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}$ we have

$$\frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla \varphi_{y}(\widehat{\mathbf{x}})|} = \frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla \widehat{F}_{y,\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{j_1}|} = \frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|}.$$

Now for any distinct $\widehat{\mathbf{x}}, \widehat{\mathbf{z}} \in \overline{\mathscr{D}_+}$ we know by Taylor's theorem that

$$\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}}) = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}})(\widehat{\mathbf{x}} - \widehat{\mathbf{z}}) + O(|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}|^2),$$

where the implicit constant does not depend on $\mathbf{j}$. Considering the eigenvalues of the invertible real symmetric matrix $H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}$ we have that

$$\lambda_{\min}|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}| \ll |H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}})(\widehat{\mathbf{x}} - \widehat{\mathbf{z}})|,$$

where $\lambda_{\min}$ is the minimum of the absolute values of the eigenvalues and the implicit constant depends only on n. We already showed that $|\det H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}|$ is bounded away from zero on $\overline{\mathscr{D}_+}$ hence by virtue of the eigenvalues being continuous in the coefficients of the matrix we find constants $C, \eta > 0$ such that

$$|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}| \leq C|\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}})|$$

for all $|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}| \leq \eta$. Here C, η are independend of $\mathbf{j}$. Now in particular choosing $\widehat{\mathbf{z}} = \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}$ yields the desired result. □

Following similar arguments as in the proof of Lemma 3.2 we find that for $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$. Then for all $\widehat{\mathbf{x}} \in \mathscr{D}_+$ we have

$$\left| \frac{\partial^{i_1+\cdots+i_{n-s}} \varphi_{\mathbf{y}}}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| \ll 1,$$

where the implicit constant depends only on $(i_1, ..., i_{n-s})$, ρ, τ, ε_1 and upper bounds for (the absolute values of) finitely many derivatives of f_r on $\mathscr{D}_+ \times \mathscr{Y}$ for $1 \leq r \leq R$. Notice that scaling with $D_1(\widehat{\mathbf{k}}, \mathbf{j})$ is unnecessary here since for $\mathbf{k} \in \mathscr{K}_{\mathbf{j};3}$ this distance is bounded from above by ρ and that the bounds are all independant of $\mathbf{j}$ and $\mathbf{k}$. By definition $H_{\varphi_{\mathbf{y}}} = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}$, hence we can apply Lemma 1.3 for $\varphi = \varphi_{\mathbf{y}}$, and λ as above together with (2.6) to obtain

$$\begin{aligned} I(q; \mathbf{j}, \mathbf{k}) &\ll \int_{\mathscr{Y}} \left| \det H_{\widehat{F}_{y,\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) \right|^{-\frac{1}{2}} (qj_1)^{-\frac{n-s}{2}-1} \mathrm{d}\mathbf{y} \\ &\ll_{c_1,\varepsilon_1} (qj_1)^{-\frac{n-s+2}{2}}, \end{aligned}$$

where the implicit constant only depends on ℓ, n, ε_1, upper bounds for the absolute values of finitely many derivatives of ω and $\varphi_{\mathbf{y}}$ on $\mathscr{D}_+$, an upper bound for $|\mathbf{x} - \mathbf{v}_0|/|\nabla \varphi_{\mathbf{y}}(\mathbf{v}_0)|$ on $\mathscr{D}_+$, and a lower bound for $\det H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})$, all of which are independant of $\mathbf{j}$ and $\mathbf{k}$. We obtain

$$\begin{aligned} \sum_{(\widehat{\mathbf{k}}, \mathbf{k}^*) \in \mathscr{K}(\mathbf{j},3)} I(q; \mathbf{j}, (\widehat{\mathbf{k}}, \mathbf{k}^*)) &\ll \sum_{(\widehat{\mathbf{k}}, k_n) \in \mathscr{K}(\mathbf{j},3)} (qj_1)^{-\frac{n-s+2}{2}} \ll (2j_1 L)^s \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ D(\widehat{\mathbf{k}}, \mathbf{j}) < \rho}} (qj_1)^{-\frac{n-s+2}{2}} \\ &\ll j_1^s \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ (\widehat{\mathbf{k}}/j_1) \in [-L-\rho, L+\rho]^{n-s}}} (qj_1)^{-\frac{n-s+2}{2}} \\ &\ll q^{-\frac{n-s+2}{2}} j_1^{\frac{n+s-2}{2}}, \end{aligned}$$

where the implicit constant only depends additionally on L, ρ, n and s. Arguing similarly to (3.9) we obtain

$$\begin{aligned} N_3 &\ll \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_2, \ldots, j_R \leq j_1 \\ \mathbf{j}/j_1 \in T_{\mathcal{I}_s^n}}} \left(\prod_{r=1}^{R} \left(\frac{1}{J} + \min\left(\delta, \frac{1}{j_r} \right) \right) \right) j_1^{\frac{n+s-2}{2}} \sum_{q \leq Q} q^{\frac{n+s-2}{2}} \\ &\ll \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_2, \ldots, j_R \leq j_1}} \left(\prod_{r=1}^{R} \left(\frac{1}{J} + \min\left(\delta, \frac{1}{j_r} \right) \right) \right) J^{\frac{n+s}{2}-1} Q^{\frac{n+s}{2}} \\ &\ll J^{\frac{n+s}{2}-1} Q^{\frac{n+s}{2}} (1 + \log J)^R. \end{aligned} \tag{3.10}$$

Case $\mathbf{k} \in \mathscr{K}_{\mathbf{j};1}$.

Choose $\varphi_{\mathbf{y}}$ and λ as in the previous case, such that we still have

$$\nabla \varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) = \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) - \frac{\widehat{\mathbf{k}}}{j_1}$$

and 3.3 still applies. The signature σ of the matrix $H_{\varphi_{\mathbf{y}}}(\widehat{\mathbf{x}}_{\mathbf{j};\mathbf{k}}) = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\mathbf{k}})$ is constant for all $\mathbf{j}, \mathbf{k}$ and $\mathbf{y}$ in consideration, since the determinant is bounded away from zero, the eigenvalues of a matrix depend continuously on its coefficients and the coefficients depend continously on $\mathbf{y}$. Applying Lemma 1.3 again yields

$$I(q;\mathbf{j};\mathbf{k}) = \int_{\mathbb{R}^s} (qj_1)^{-\frac{n-s}{2}} \frac{\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}, \mathbf{y})}{|\det H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} e\left(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) + \frac{\sigma}{8}\right) \mathrm{d}\mathbf{y} + O\left((qj_1)^{-\frac{n-s}{2}-1}\right). \tag{3.11}$$

Since all of $\omega, \varphi_{\mathbf{y}}$ and $\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}$ depend on $\mathbf{y}$, we delay evaluating the integral for now and consider the terms

$$\begin{aligned}
&N_{1,\mathbf{y},\mathbf{j}} \\
&= \sum_{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}} \sum_{q\le Q} q^n \left((qj_1)^{-\frac{n-s}{2}} \frac{\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}, \mathbf{y})}{|\det H_{\widehat{F}_{y,\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} e\left(qj_1\varphi_y(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) + \frac{\sigma}{8}\right)\right) \\
&+ O\left(\sum_{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}} \sum_{q\le Q} q^n (qj_1)^{-\frac{n-s}{2}-1}\right) \\
&= \sum_{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}} \sum_{q\le Q} q^n \left((qj_1)^{-\frac{n-s}{2}} \frac{\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}, \mathbf{y})}{|\det H_{\widehat{F}_{y,\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} e\left(qj_1\varphi_y(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) + \frac{\sigma}{8}\right)\right) \\
&+ O\left(j_1^{\frac{n+s}{2}-1} Q^{\frac{n+s}{2}}\right)
\end{aligned}$$

and

$$N_{1,\mathbf{y}} = \sum_{\substack{1\le j_1\le J \\ 0\le j_2,\ldots,j_R\le j_1 \\ (\mathbf{j}/j_1)\in T_{\mathcal{I}_s^n}}} \left(\prod_{r=1}^R b_{j_r}\right) N_{1,\mathbf{y},\mathbf{j}}.$$

We start with the inner most sum

$$\begin{aligned}
&\left|\sum_{q\le Q} q^n \left((qj_1)^{-\frac{n-s}{2}} \frac{\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}, \mathbf{y})}{|\det H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} e\left(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) + \frac{\sigma}{8}\right)\right)\right| \\
&\quad\ll \frac{\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}, \mathbf{y})}{|\det H_{\widehat{F}_{y,\mathbf{j}}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} j_1^{-\frac{n-s}{2}} \left|\sum_{q\le Q}\left(q^{\frac{n+s}{2}} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}))\right)\right|.
\end{aligned} \tag{3.12}$$

The remaining sum over q can be dealt with by means of partial summation

$$\begin{aligned}
\sum_{q\le Q} q^{\frac{n+s}{2}} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})) &= Q^{\frac{n+s}{2}} \sum_{q\le Q} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})) \\
&\quad - \int_1^Q \sum_{q\le \xi} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})) \frac{n+s}{2} \xi^{\frac{n+s}{2}-1} \mathrm{d}\xi
\end{aligned}$$

Note that we have the bound

$$\left|\sum_{q\le\xi} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}))\right| \ll \min\{\xi, ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||^{-1}\},$$

so we distinguish two cases. First if $||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|| \geq Q^{-1}$ we obtain

$$\sum_{q\leq Q} q^{\frac{n+s}{2}} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})) \ll \frac{Q^{\frac{n+s}{2}}}{||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||} + \frac{1}{||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||}\int_1^Q \frac{n+s}{2}\xi^{\frac{n+s}{2}-1}\mathrm{d}\xi \ll \frac{Q^{\frac{n+s}{2}}}{||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||}. \tag{3.13}$$

On the other hand if $||j_1\varphi_y(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})|| < Q^{-1}$we have

$$\sum_{q\leq Q} q^{\frac{n+s}{2}} e(qj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})) \ll Q^{\frac{n+s}{2}+1} + \frac{n+s}{2}\int_1^Q \xi^{\frac{n+s}{2}}\mathrm{d}\xi \ll Q^{\frac{n+s}{2}+1}. \tag{3.14}$$

Hence we obtain

$$\begin{aligned} N_{1,\mathbf{y}} \ll & \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||\geq Q^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}} Q^{\frac{n+s}{2}} ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||^{-1} \\ & + \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<Q^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}} Q^{\frac{n+s}{2}+1} \\ & + \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) j_1^{\frac{n+s}{2}-1} Q^{\frac{n+s}{2}} \\ \ll_L & \; Q^{\frac{n+s}{2}} \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||\geq Q^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}} ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||^{-1} \\ & + Q^{\frac{n+s}{2}+1} \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<Q^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}} \\ & + \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\dots,j_R\leq j_1}} \left(\prod_{r=1}^R b_{j_r}\right) j_1^{\frac{n+s}{2}-1} Q^{\frac{n+s}{2}} \end{aligned} \tag{3.15}$$

The last term can be bounded similarly to (3.9) and (3.10). We have the following essential result to be proven in Section 4.

Proposition 3.4. *Let $T > 0$ and $J_2, ..., J_R \in [1, J]$. Then with the notations of this section and for all $y \in \overline{\mathscr{Y}}$, we have*

$$\sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq \min\{J_r,j_1\}\\ 2\leq r\leq R}} \sum_{\substack{\mathbf{k}\in\mathbb{Z}^n\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) \ll \left(\prod_{r=2}^R J_r\right)(J^{n+1}T^{-1} + J^n\mathscr{E}_{n-s}(J)),$$

where

$$\mathscr{E}_{n-s}(J)=\mathscr{E}_{n-s}^{(\mathfrak{c}_1';\mathfrak{c}_2')}=\begin{cases}\exp(\mathfrak{c}_1'\sqrt{\log J}) & \text{if } n=2+s\\ (\log J)^{\mathfrak{c}_2'} & \text{if } n\geq 2+s+1\end{cases}$$

for some positive constants $\mathfrak{c}_1'$ and $\mathfrak{c}_2'$. Here the implicit constant as well as $\mathfrak{c}_1'$ and $\mathfrak{c}_2'$ only depend on n, R, c_1 and c_2 in (2.6), ρ in (2.7), ρ' in (4.10) and upper bounds for (the absolute value) of finitely many derivatives of ω and f_r for $2\leq r\leq R$ on $\mathscr{D}_+\times\mathscr{Y}$. In particular, they are independant of T, $J_2,...,J_R$ and $\mathbf{y}$.

Recall that $b_j\ll\frac{1}{j}$ for $1\leq j\leq J$. Hence with $\mathfrak{I}_0=\{0\}$ and $\mathfrak{I}_s=[2^{s-1},2^s]$ it follows that

$$\begin{aligned}&\sum_{\substack{1\leq j_1\leq J\\0\leq j_2,...,j_R\leq j_1}}\left(\prod_{r=1}^R b_{j_r}\right)j_1^{-\frac{n-s}{2}}\sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}}\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) \qquad (3.16)\\
&\ll\sum_{0\leq s_2,...,s_R\leq\frac{\log J}{\log 2}+1}\left(\prod_{r=2}^R 2^{-s_r}\right)\sum_{\substack{1\leq j_1\leq J\\ j_r\in\mathfrak{I}_{s_r}\cap[0,j_1]\\2\leq r\leq R}}j_1^{-\frac{n-s}{2}-1}\sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}}\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y})\\
&\ll\sum_{0\leq s_2,...,s_R\leq\frac{\log J}{\log 2}+1}\left(\prod_{r=2}^R 2^{-s_r}\right)\sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq\min\{2^{s_r},j_1\}\\2\leq r\leq R}}j_1^{-\frac{n-s}{2}-1}\sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}}\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}).\end{aligned}$$

Now using partial summation and Proposition 3.4, we find that for all s_r in consideration

$$\begin{aligned}&\sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq\min\{2^{s_r},j_1\}\\2\leq r\leq R}}j_1^{-\frac{n-s}{2}-1}\sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}}\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) \qquad (3.17)\\
&\ll J^{-\frac{n-s}{2}-1}\left(\prod_{r=2}^R 2^{s_r}\right)(J^{n+1}T^{-1}+J^n\mathscr{E}_{n-s}(J)).\end{aligned}$$

Therefore

$$\begin{aligned}&\sum_{\substack{1\leq j_1\leq J\\0\leq j_2,...,j_R\leq j_1}}\left(\prod_{r=1}^R b_{j_r}\right)j_1^{-\frac{n-s}{2}}\sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}}\omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) \qquad (3.18)\\
&\ll(1+\log J)^R J^{-\frac{n-s}{2}-1}(J^{n+1}T^{-1}+J^n\mathscr{E}_{n-s}(J)).\end{aligned}$$

Now the second term in (3.15) can be estimated by taking $T=Q$ in (3.18). For the first sum in (3.15) we split the interval $[Q^{-1},1/2]$ into dyadic intervals. We may assume

$Q \geq 2$, i.e. $Q^{-1} \leq 1/2$, and conclude

$$\sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_R\leq j_1}} \left(\prod_{r=1}^{R} b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||\geq Q^{-1}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}} ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||^{-1} \tag{3.19}$$

$$\leq \sum_{1\leq i\leq \frac{\log Q}{\log 2}+1} Q2^{1-i} \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_2,\ldots,j_R\leq j_1}} \left(\prod_{r=1}^{R} b_{j_r}\right) \sum_{\substack{\mathbf{k}\in\mathscr{K}_{\mathbf{j};1}\\ \frac{2^{i-1}}{Q}\leq ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||\leq \frac{2^i}{Q}}} \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y}) j_1^{-\frac{n-s}{2}}$$

$$\ll \sum_{1\leq i\leq \frac{\log Q}{\log 2}+1} Q2^{1-i}(1+\log J)^R J^{-\frac{n-s}{2}-1}(2^i J^{n+1}Q^{-1} + J^n \mathscr{E}_{n-s}(J))$$

$$\ll (1+\log J)^R((\log Q)J^{\frac{n+s}{2}} + QJ^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J))$$

using (3.18) again. Combining (3.15), (3.18) and (3.19), we obtain

$$\begin{aligned} N_{1,\mathbf{y}} &\ll Q^{\frac{n+s}{2}}(1+\log J)^R((\log Q)J^{\frac{n+s}{2}} + QJ^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J)) \\ &\quad + Q^{\frac{n+s}{2}+1}(1+\log J)^R J^{-\frac{n-s}{2}-1}(J^{n+1}Q^{-1} + J^n\mathscr{E}_{n-s}(J)) \\ &\quad + (1+\log J)^R Q^{\frac{n+s}{2}} J^{\frac{n+s}{2}-1} \\ &\ll (1+\log J)^R((\log Q)Q^{\frac{n+s}{2}} J^{\frac{n+s}{2}} + Q^{\frac{n+s}{2}+1} J^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J)). \end{aligned} \tag{3.20}$$

Consequently we have

$$N_1 \ll_{\varepsilon_1} (1+\log J)^R((\log Q)Q^{\frac{n+s}{2}} J^{\frac{n+s}{2}} + Q^{\frac{n+s}{2}+1} J^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J)). \tag{3.21}$$

4. Proof of Proposition 3.4

Recall that we defined functions

$$\widehat{f}_{j,\mathbf{y}} \colon \mathbb{R}^{n-s} \to \mathbb{R}, \widehat{\mathbf{x}} \mapsto f_j(\widehat{\mathbf{x}},\mathbf{y})$$

for $\mathbf{y} \in \overline{\mathscr{Y}}$, ε_1 as in (2.6), and

$$\widehat{F}_{\mathbf{y},\mathbf{j}} = \widehat{f}_{1,\mathbf{y}} + \frac{j_2}{j_1}\widehat{f}_{2,\mathbf{y}} + \cdots + \frac{j_R}{j_1}\widehat{f}_{R,\mathbf{y}}$$

for $\mathbf{j} \in \mathbb{R}^R \setminus \{0\}$. For a non-negative weight function $\omega \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ we defined

$$U_{\mathbf{y}} = \{\widehat{\mathbf{x}} \in \mathbb{R}^{n-s} \mid \omega(\widehat{\mathbf{x}}, y) \neq 0\}$$

and $V_{\mathbf{y},\mathbf{j}} = \nabla\widehat{F}_{y,\mathbf{j}}$. Note that $\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}$ is a diffeomorphism on $U_{\mathbf{y}}$. Now let $\omega^*_{\mathbf{j}} = \omega\circ(\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}$, and for $T \geq 2$ and $J_2, ..., J_R \in [1, J]$, define

$$\begin{aligned} \mathscr{M}(J,T^{-1}) &= \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq \min\{J_r,j_1\}\\ 2\leq r\leq R}} \sum_{\substack{\mathbf{a}\in\mathscr{K}_{\mathbf{j};1}\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}})||\leq T^{-1}}} \omega^*_{\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) \\ &= \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq \min\{J_r,j_1\}\\ 2\leq r\leq R}} \sum_{\substack{\mathbf{a}\in\mathscr{K}_{\mathbf{j};1}\\ |\mathbf{a}^*|\leq 2j_1L\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}})||\leq T^{-1}}} \omega^*_{\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right). \end{aligned} \tag{4.1}$$

Note that Proposition 3.4 for $0 < T < 2$ immediately follows from the case $T = 2$. We consider the Fejér kernel

$$\mathscr{F}_D(\theta) = D^{-2}\left|\sum_{d=1}^{D} e(d\theta)\right|^2 = \left(\frac{\sin(\pi D\theta)}{D\sin(\pi\theta)}\right)^2 = \sum_{d=-D}^{D} \frac{D-|d|}{D^2} e(d\theta) \tag{4.2}$$

for $D = \lfloor T/2 \rfloor$. Let $\theta \in \mathbb{R}$ with $0 < ||\theta|| \le T^{-1}$, then by the concave property of the sine function on $[0, \pi/2]$ we have

$$\left(\frac{\sin(D\pi\theta)}{D\sin(\pi\theta)}\right)^2 \ge \left(\frac{2\pi^{-1}D\pi||\theta||}{D\pi||\theta||}\right) \ge \frac{4}{\pi^2}.$$

Therefore, it follows that

$$\chi_{T^{-1}}(\theta) \le \frac{\pi^2}{4}\mathscr{F}_D(\theta), \tag{4.3}$$

with $\chi_{T^{-1}}$ as in (2.1). Combining (4.1), (4.2) and (4.3) we obtain

$$\mathscr{M}(J, T^{-1}) \le \frac{\pi^2}{4} \sum_{\substack{1\le j_1\le J\\ 0\le j_r\le \min\{J_r, j_1\}\\ 2\le r\le R}} \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n\\ |\mathbf{a}^*|\le 2j_1L}} \sum_{d=-D}^{D} \frac{D-|d|}{D^2}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) e(dj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}})). \tag{4.4}$$

By definition $\omega_{\mathbf{j}}^*$ vanishes outside of $\bigcup_{\mathbf{y}\in\overline{\mathscr{Y}}} V_{\mathbf{y},\mathbf{j}} \subseteq [-L, L]^n$, hence the contribution of terms with $d = 0$ in (4.4) is

$$\frac{\pi^2}{4D} \sum_{\substack{1\le j_1\le J\\ 0\le j_r\le \min\{J_r, j_1\}\\ 2\le r\le R}} \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n\\ |\mathbf{a}^*|\le 2j_1L}} \omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) \ll \frac{1}{D}\left(\prod_{r=2}^{R} J_r\right) \sum_{1\le j_1\le J} j_1^n \ll \frac{J^{n+1}}{D}\left(\prod_{r=2}^{R} J_r\right), \tag{4.5}$$

where the implicit constants only depend on n and L. Let $\widehat{F}_{\mathbf{y},\mathbf{j}}^*$ be the Legendre transform of $\widehat{F}_{\mathbf{y},\mathbf{j}}$. Then with (1.1) and since $\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}} = (\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}(\widehat{\mathbf{a}}/j_1)$ we have

$$\widehat{F}_{\mathbf{y},\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) = \widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}} \cdot \frac{\widehat{\mathbf{a}}}{j_1} - \widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}}) = -\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}}) - \frac{\mathbf{a}^*\cdot\mathbf{y}}{j_1}.$$

Now we can rewrite

$$\begin{aligned}
&\sum_{\substack{\mathbf{a}\in\mathbb{Z}^n\\ |\mathbf{a}^*|\le 2j_1L}} \sum_{d=-D}^{D} \frac{D-|d|}{D^2}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) e(dj_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{a}}})) \qquad (4.6)\\
&= \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n\\ |\mathbf{a}^*|\le 2j_1L}} \sum_{d=-D}^{D} \frac{D-|d|}{D^2}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) e\left(-dj_1\left(\frac{\mathbf{a}^*\cdot\mathbf{y}}{j_1} + \widehat{F}_{\mathbf{y},\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)\right)\\
&= \sum_{\substack{\mathbf{a}\in\mathbb{Z}^n\\ |\mathbf{a}^*|\le 2j_1L}} \sum_{d=-D}^{D} \frac{D-|d|}{D^2}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) e(-d\mathbf{a}^*\cdot\mathbf{y}) e\left(-dj_1\widehat{F}_{\mathbf{y},\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)\\
&= \sum_{|\mathbf{a}^*|\le 2j_1L} \sum_{d=-D}^{D} \frac{D-|d|}{D^2} e(-d\mathbf{a}^*\cdot\mathbf{y}) \sum_{\widehat{\mathbf{a}}\in\mathbb{Z}^{n-s}} \omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right) e\left(-dj_1\widehat{F}_{\mathbf{y},\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right),
\end{aligned}$$

and since

$$e\left(-dj_1\widehat{F}_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)=\overline{e\left(dj_1\widehat{F}_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)},\quad e(-d\mathbf{a}^*\cdot\mathbf{y})=\overline{e(d\mathbf{a}^*\cdot\mathbf{y})}$$

we have

$$\begin{aligned}&\left|\sum_{|\mathbf{a}^*|\le 2j_1L}\sum_{1\le|d|\le D}\frac{D-|d|}{D^2}e(d\mathbf{a}^*\cdot\mathbf{y})\sum_{\widehat{\mathbf{a}}\in\mathbb{Z}^{n-s}}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)e\left(dj_1\widehat{F}^*_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)\right| \qquad (4.7)\\&\le 2\left|\sum_{|\mathbf{a}^*|\le 2j_1L}\sum_{d=1}^{D}\frac{D-d}{D^2}e(d\mathbf{a}^*\cdot\mathbf{y})\sum_{\widehat{\mathbf{a}}\in\mathbb{Z}^{n-s}}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)e\left(dj_1\widehat{F}^*_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right)\right|.\end{aligned}$$

Applying the n-dimensional Poisson summation formula to the inner most sum yields

$$\begin{aligned}&\sum_{\widehat{\mathbf{a}}\in\mathbb{Z}^{n-s}}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)e\left(dj_1\widehat{F}^*_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{a}}}{j_1}\right)\right) \qquad (4.8)\\&=\sum_{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s}}\int_{\mathbb{R}^{n-s}}\omega_{\mathbf{j}}^*\left(\frac{\widehat{\mathbf{z}}}{j_1}\right)e\left(dj_1\widehat{F}^*_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{z}}}{j_1}\right)-\widehat{\mathbf{k}}\cdot\widehat{\mathbf{z}}\right)\mathrm{d}\widehat{\mathbf{z}}\\&=j_1^{n-s}\sum_{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s}}I_0(d;\mathbf{j};\widehat{\mathbf{k}})\end{aligned}$$

where

$$I_0(d;\mathbf{j};\widehat{\mathbf{k}})=\int_{\mathbb{R}^{n-s}}\omega_{\mathbf{j}}^*(\widehat{\mathbf{x}})e\left(dj_1\widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}})-j_1\widehat{\mathbf{k}}\cdot\widehat{\mathbf{x}}\right)\mathrm{d}\widehat{\mathbf{x}}.$$

Therefore to obtain a bound for $\mathscr{M}(J,T^{-1})$ it is sufficient to provide a bound for

$$\left|\sum_{\substack{1\le j_1\le J\\0\le j_r\le\min\{J_r,j_1\}\\2\le r\le R}}\sum_{|\mathbf{a}^*|\le 2j_1L}\sum_{d=1}^{D}\frac{D-d}{D^2}e(d\mathbf{a}^*\cdot\mathbf{y})j_1^{n-s}\sum_{\widehat{\mathbf{a}}\in\mathbb{Z}^{n-s}}I_0(d;\mathbf{j};\widehat{\mathbf{k}})\right|. \qquad (4.9)$$

Since $\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}=(\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}$ and $\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}$ is a diffeomorphism on $\overline{\mathscr{D}}$ we have that $\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}$ is a diffeomorphism on $\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\mathscr{D})$ and $\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}(V^+_{\mathbf{y},\mathbf{j}})=U^+_{\mathbf{y}}$. Let

$$\rho'=\frac{\mathrm{dist}(\partial\mathscr{D},\partial U_{\mathbf{y}})}{2}. \qquad (4.10)$$

We repeat the technique of Section 3 and split the set $\mathbb{Z}^{n-s}$ into three disjoint subsets. Let

$$\mathscr{K}_1=\left\{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s}\,\middle|\,\frac{\widehat{\mathbf{k}}}{d}\in U_{\mathbf{y}}\right\},$$

$$\mathscr{K}_2=\left\{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s}\,\middle|\,\mathrm{dist}\left(\frac{\widehat{\mathbf{k}}}{d},U_{\mathbf{y}}\right)\ge\rho'\right\}$$

and

$$\mathscr{K}_3 = \mathbb{Z}^{n-s} \setminus (\mathscr{K}_1 \cup \mathscr{K}_2).$$

For each $1 \leq i \leq 3$ we define

$$M_i = \sum_{d=1}^{D} \frac{D-d}{D^2} \left| \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_r \leq \min\{J_r, j_1 \\ 2 \leq r \leq R}} \sum_{|\mathbf{a}^*| \leq 2 j_1 L} e(d\mathbf{a}^* \cdot \mathbf{y}) j_1^{n-s} \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_i} I_0(d; \mathbf{j}; \widehat{\mathbf{k}}) \right| \tag{4.11}$$

$$\ll_{L,s} \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_r \leq \min\{J_r, j_1\} \\ 2 \leq r \leq R}} \sum_{d=1}^{D} \frac{D-d}{D^2} \left| j_1^n \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_i} I_0(d; \mathbf{j}; \widehat{\mathbf{k}}) \right|,$$

such that

$$\mathscr{M}(J, T^{-1}) \ll \left(\prod_{r=2}^{R} J_r \right) \frac{J^{n+1}}{D} + M_1 + M_2 + M_3, \tag{4.12}$$

and seek to bound each M_i seperately.

Case $\widehat{\mathbf{k}} \in \mathscr{K}_2$.

Define

$$\varphi_1(\widehat{\mathbf{x}}) = \frac{d\widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \widehat{\mathbf{k}} \cdot \widehat{\mathbf{x}}}{\operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}})}$$

and

$$\lambda_1 = j_1 \operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}}).$$

Then for all $\widehat{\mathbf{x}} \in V_{y,\mathbf{j}}$

$$|\nabla \varphi_1(\widehat{\mathbf{x}})| = \frac{|d\nabla \widehat{F}^*_{\mathbf{y},\mathbf{j}}(\mathbf{x}) - \widehat{\mathbf{k}}|}{\operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}})} \geq 1$$

and like in (3.1) we conclude

$$|\nabla \varphi_1(\widehat{\mathbf{x}})| \geq \frac{1}{2}$$

for $\widehat{\mathbf{x}} \in V^+_{\mathbf{y},\mathbf{j}}$. Next we establish upper bounds for the derivatives of φ_1 and $\omega^*_{\mathbf{j}}$. In order to do so, we estbalish bounds for the derivatives of $\widehat{F}^*_{\mathbf{y},\mathbf{j}}$ first.

Lemma 4.1. *Let $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$. Then for all $\widehat{\mathbf{x}} \in U^+_y$ we have*

$$\left| \frac{\partial^{i_1 + \cdots + i_{n-s}} \widehat{F}^*_{\mathbf{y},\mathbf{j}}}{\partial^{i_1}_{x_1} \cdots \partial^{i_{n-s}}_{x_{n-s}}}(\widehat{\mathbf{x}}) \right| \ll 1,$$

where the implicit constant depends only on $(i_1, ..., i_{n-s})$, ρ', τ and upper bounds for (the absolute values of) finitely many derivatives of f_r on $U^+_y \times \mathscr{Y}$ for $1 \leq r \leq R$.

Proof. For $\widehat{\mathbf{x}} \in V^{+}_{\mathbf{y},\mathbf{j}}$ and $\widehat{\mathbf{z}} \in U^{+}_{\mathbf{y}}$ we have

$$|\widehat{\mathbf{x}} \cdot \widehat{\mathbf{z}}| + |\widehat{F}_{\mathbf{y},\mathbf{j}}| \ll 1,$$

hence we easily deduce $|\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}| \ll 1$ with (1.1). Recall that $\nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}} = (\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}$ and $U^{+}_{\mathbf{y}} = (\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}(V^{+}_{\mathbf{y},\mathbf{j}})$. Hence $|\nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}})| \ll 1$ for $\widehat{\mathbf{x}} \in V^{+}_{\mathbf{y},\mathbf{j}}$. For $\widehat{\mathbf{x}} = \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}})$ with $\widehat{\mathbf{z}} \in U^{+}_{\mathbf{y}}$ we have

$$\mathrm{Jac}_{\nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}) = \mathrm{Jac}_{(\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}}(\widehat{\mathbf{x}}) = (\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}}))^{-1}, \tag{4.13}$$

where Jac_f denotes the Jacobian matrix of the function f and we used the chain rule. Consequently every second partial derivative of $\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}$, i.e. the entries of the Jacobian matrix, can be written as

$$\frac{P}{\det(\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}}))}, \tag{4.14}$$

where P is a polynomial expression in the terms of the entries of $\mathrm{Jac}_{\nabla\widehat{F}_{y,\mathbf{j}}}(\widehat{\mathbf{z}})$. Note that P has degree $(n-s)$ and each coefficient can only be ± 1 or 0. Since

$$\left|\frac{\partial^{i_1+\cdots+i_{n-s}}\widehat{F}_{\mathbf{y},\mathbf{j}}}{\partial^{i_1}_{x_1}\cdots\partial^{i_{n-s}}_{x_{n-s}}}(\widehat{\mathbf{z}})\right| \ll 1$$

is obvious for any $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$ and $\widehat{\mathbf{z}} \in U^{+}_{\mathbf{y}}$, and $\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}} = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}$, the desired bounds for the second derivatives follows directly with (4.14) and (2.6). Essentially the same idea will be used to argue for higher partial derivatives. Note that for any $k \in \mathbb{N}$ we can express the k-th partial derivative with respect to the $\widehat{\mathbf{x}}$-variables of an entry in $\mathrm{Jac}_{\nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}})$ as a real polynomial with coefficients independant of $\mathbf{j}$ in terms of:

(i) $(\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}}))^{-m}$, where $m \leq k+1$;
(ii) entries of $\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}})$;
(iii) m-th partial derivatives with respect to the $\widehat{\mathbf{z}}$-variables of the entries in $\mathrm{Jac}_{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}}$, where $m \geq k$;
(iv) m-th partial derivatives with respect to the $\widehat{\mathbf{x}}$-variables of the entries in $\nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}} = (\nabla\widehat{F}_{\mathbf{y},\mathbf{j}})^{-1}(\widehat{\mathbf{x}})$, where $m \geq k$.

Now, using (2.6) again, the desired result follows inductively. □

Now with similar arguments as in the proof of Lemma 3.2 we can deduce the following.

Corollary 4.2. *Let $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$. Then for all $\widehat{\mathbf{x}} \in U^{+}_{\mathbf{y}}$ we have*

$$\left|\frac{\partial^{i_1+\cdots+i_{n-1}}\varphi_1}{\partial^{i_1}_{x_1}\cdots\partial^{i_{n-s}}_{x_{n-s}}}(\widehat{\mathbf{x}})\right| \ll 1,$$

where the implicit constant depends only on $(i_1, ..., i_{n-s})$, ρ', τ and upper bounds for (the absolute values of) finitely many derivatives of f_r on $U^{+}_{y} \times \mathscr{Y}$ for $1 \leq r \leq R$.

Recall that $\omega^{*}_{\mathbf{j}} = \omega \circ \nabla\widehat{F}^{*}_{\mathbf{y},\mathbf{j}}$, hence we also obtain the following.

Corollary 4.3. *Let $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell - 1$. Then for all $\widehat{\mathbf{x}} \in U_{\mathbf{y}}^+$ we have*

$$\left| \frac{\partial^{i_1+\cdots+i_{n-s}} \omega_{\mathbf{j}}^*}{\partial_{x_1}^{i_1} \cdots \partial_{x_{n-s}}^{i_{n-s}}}(\widehat{\mathbf{x}}) \right| \ll 1,$$

where the implicit constant depends only on $(i_1, ..., i_{n-s})$, ρ', τ and upper bounds for (the absolute values of) finitely many derivatives of f_r on $U_y^+ \times \mathscr{Y}$ for $1 \leq r \leq R$.

Apllying 1.3 with $\varphi = \varphi_1$ and $\lambda = \lambda_1$ as defined above now yields

$$I_0(d; \mathbf{j}; \widehat{\mathbf{k}}) \ll \lambda^{-\ell+1} = (j_1 \operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}}))^{-\ell+1},$$

where the implicit constant is independant of $\mathbf{y}$, $\mathbf{j}$ and $\widehat{\mathbf{k}}$. Now since $\ell - (n-s) > 1$ we find similarly to (3.6)

$$\begin{aligned} \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_2} I_0(d; \mathbf{j}; \widehat{\mathbf{k}}) &\ll j_1^{-\ell+1} \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_2} \operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}})^{-\ell+1} \\ &\ll j_1^{-\ell+1} \sum_{m=0}^{\infty} \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ 2^m d\rho' \leq \operatorname{dist}(\widehat{\mathbf{k}}, dU_{\mathbf{y}}) < 2^{m+1} d\rho'}} \frac{1}{(2^m d\rho')^{\ell-1}}, \\ &\ll j_1^{-\ell+1} \sum_{m=0}^{\infty} \frac{(Ld + 2^{m+1} d\rho')^{n-s}}{(2^m d\rho')^{\ell-1}}, \\ &\ll j_1^{-\ell+1}, \end{aligned} \tag{4.15}$$

where the implicit constant is independant of $\mathbf{y}$ and d. Consequently we obtain

$$\begin{aligned} M_2 &\leq \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_r \leq \min\{J_r, j_1\} \\ 2 \leq r \leq R}} \sum_{d=1}^{D} \frac{D-d}{D^2} j_1^n \left| \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_2} I_0(d; \mathbf{j}; \widehat{\mathbf{k}}) \right| \\ &\ll \frac{D-1}{2D} \sum_{\substack{1 \leq j_1 \leq J \\ 0 \leq j_r \leq \min\{J_r, j_1\} \\ 2 \leq r \leq R}} j_1^{n-\ell+1} \\ &\ll \left(\prod_{r=2}^{R} J_r \right) \log J. \end{aligned} \tag{4.16}$$

Case $\widehat{\mathbf{k}} \in \mathscr{K}_3$.

Let $\lambda = j_1 d$ and

$$\varphi(\widehat{\mathbf{x}}) = \widehat{F}_{\mathbf{y},\mathbf{j}}^*(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{d} \cdot \widehat{\mathbf{x}}.$$

By definition, for each fixed d we have that $\widehat{\mathbf{k}} \in d\mathscr{D}$ determines a unique preimage

$$\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}} = (\nabla \widehat{F}_{\mathbf{y},\mathbf{j}}^*)^{-1} \left(\frac{\widehat{\mathbf{k}}}{d} \right) = \nabla \widehat{F}_{\mathbf{y},\mathbf{j}} \left(\frac{\widehat{\mathbf{k}}}{d} \right)$$

that is also a critical point for φ in the sense that

$$\nabla\varphi(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) = \nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) - \frac{\widehat{\mathbf{k}}}{d} = \mathbf{0}.$$

Lemma 4.4. *Let* $\widehat{\mathbf{x}} \in \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\mathscr{D}_+) \setminus \{\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}\}$*. Then*

$$\frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla\varphi(\widehat{\mathbf{x}})|} \ll 1$$

where the implicit constant is independant of d*,* $\mathbf{j}$ *and* $\widehat{\mathbf{k}}$*.*

Proof. Observe that for $\widehat{\mathbf{k}}/d \in \mathscr{D}$ we have

$$\frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}|}{|\nabla\varphi(\widehat{\mathbf{x}})|} = \frac{|\widehat{\mathbf{x}} - (\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}})^{-1}(\widehat{\mathbf{k}}/d)|}{|\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \frac{\widehat{\mathbf{k}}}{d}|},$$

hence it is sufficient to prove

$$\frac{|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}|}{|\nabla\widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}) - \nabla\widehat{F}^*_{y,\mathbf{j}}(\widehat{\mathbf{z}})|} \ll 1$$

for $\widehat{\mathbf{x}}, \widehat{\mathbf{z}} \in \overline{\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\mathscr{D}_+)}$ and $\widehat{\mathbf{x}} \neq \widehat{\mathbf{z}}$. Taking $\widehat{\mathbf{x}}', \widehat{\mathbf{z}}' \in \overline{\mathscr{D}_+}$ with $\widehat{\mathbf{x}}' \neq \widehat{\mathbf{z}}'$ such that $\widehat{\mathbf{x}} = \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}')$ (and the same for $\widehat{\mathbf{z}}$) the inequality is equivalent to

$$\frac{|\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}') - \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}}')|}{|\widehat{\mathbf{x}}' - \widehat{\mathbf{z}}'|} \ll 1.$$

or alternatively

$$1 \ll \frac{|\widehat{\mathbf{x}}' - \widehat{\mathbf{z}}'|}{|\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}') - \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}}')|}.$$

We have already established in the proof of 3.3 that

$$\nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}') - \nabla\widehat{F}_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{z}}') = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{z}}')(\widehat{\mathbf{x}}' - \widehat{\mathbf{z}}') + O(|\widehat{\mathbf{x}} - \widehat{\mathbf{z}}'|^2),$$

which yields the desired lower bound immediately. □

Note that because of Lemma 4.1 we can deduce the same result from Corollary 4.2 for φ in this case, i.e. for given $i_1, ..., i_{n-s} \in \mathbb{Z}_{\geq 0}$ with $\sum_{\mu=1}^{n-s} i_\mu \leq \ell$ and $\widehat{\mathbf{x}} \in V^+_{\mathbf{y},\mathbf{j}}$ we have

$$\left| \frac{\partial^{i_1+\cdots+i_{n-s}}\varphi}{\partial^{i_1}_{x_1}\cdots\partial^{i_{n-s}}_{x_{n-s}}}(\widehat{\mathbf{x}}) \right| \ll 1. \tag{4.17}$$

The implicit constant is again independent of $\mathbf{y}$, d, $\mathbf{j}$ and such $\widehat{\mathbf{k}}$ that satisfy this case. By construction we have $H_\varphi = H_{\widehat{F}^*_{\mathbf{y},\mathbf{j}}}$, so with (1.2) and (2.6) we have $H_\varphi(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) \neq 0$ and consequently Lemma 1.3 yields

$$I_0(d;\mathbf{j};\widehat{\mathbf{k}}) \ll \lambda^{-\frac{n-s}{2}-1} = j_1^{-\frac{n-s}{2}-1} d^{-\frac{n-s}{2}-1}.$$

Note that we chose $\widehat{\mathbf{k}}/d \notin U_{\mathbf{y}}$, hence $\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}} \notin \nabla \widehat{F}_{\mathbf{y},\mathbf{j}}(U_y)$, i.e. $\omega^*_{\mathbf{j}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) = 0$. With a similar argument as in the proof of Lemma 3.1 we obtain

$$\sum_{\widehat{\mathbf{k}}\in\mathscr{K}_3} I_0(d;\mathbf{j};\widehat{\mathbf{k}}) \ll d^{n-s} j_1^{-\frac{n-s}{2}-1} d^{-\frac{n-s}{2}-1} = j_1^{-\frac{n-s}{2}-1} d^{\frac{n-s}{2}-1}.$$

Hence we have

$$\begin{aligned} M_3 &\leq \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq \min\{J_r,j_1\}\\ 2\leq r\leq R}} \sum_{d=1}^{D} \frac{D-d}{D^2} \left| j_1^n \sum_{\widehat{\mathbf{k}}\in\mathscr{K}_3} I_0(d;\mathbf{j};\widehat{\mathbf{k}})\right| \\ &\ll \frac{1}{D} \sum_{\substack{1\leq j_1\leq J\\ 0\leq j_r\leq \min\{J_r,j_1\}\\ 2\leq r\leq R}} j_1^{\frac{n+s}{2}-1} \sum_{d=1}^{D} d^{\frac{n-s}{2}-1} \\ &\ll \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}-1}. \end{aligned} \tag{4.18}$$

Case $\widehat{\mathbf{k}} \in \mathscr{K}_1$.

Let λ and φ be as in the previous case, specifically maintaining Lemma 4.4 and (4.17). Then we have

$$\varphi(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) = \widehat{F}^*_{\mathbf{y},\mathbf{j}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) - \frac{\widehat{\mathbf{k}}}{d} \cdot \widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}} = -\widehat{F}_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{k}}}{d}\right) \tag{4.19}$$

and by (1.2)

$$H_{\widehat{F}^*_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}}) = H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}\left(\frac{\widehat{\mathbf{k}}}{d}\right)^{-1}.$$

Similar to the arguments right before (3.11) we find that the signature σ of $H_\varphi(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}}) = H_{\widehat{F}^*_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}})$ is constant for all d, $\mathbf{j}$ and $\widehat{\mathbf{k}}$ in consideration, hence with Lemma 1.3, (2.6) and (4.19) we obtain

$$\begin{aligned} &I_0(d;\mathbf{j};\widehat{\mathbf{k}}) \\ &\quad= \frac{\omega^*_{\mathbf{j}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}})}{|\det H_{\widehat{F}^*_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{x}}_{d;\mathbf{j};\widehat{\mathbf{k}}})|^{\frac{1}{2}}} (j_1 d)^{-\frac{n-s}{2}} e\left(-j_1 d \widehat{F}_{\mathbf{y},\mathbf{j}}\left(\frac{\widehat{\mathbf{k}}}{d}\right) + \frac{\sigma}{8}\right) + O\left((j_1 d)^{-\frac{n-s}{2}-1}\right) \\ &\quad= \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) |\det H_{\widehat{F}_{\mathbf{y},\mathbf{j}}}(\widehat{\mathbf{k}}/d)|^{\frac{1}{2}} (j_1 d)^{-\frac{n-s}{2}} e\left(-d(j_1\widehat{f}_{1,\mathbf{y}} + \cdots + j_R \widehat{f}_{R,\mathbf{y}})\left(\frac{\widehat{\mathbf{k}}}{d}\right) + \frac{\sigma}{8}\right) \\ &\quad+ O\left((j_1 d)^{-\frac{n-s}{2}-1}\right), \end{aligned} \tag{4.20}$$

where the implicit constant is independant of $\mathbf{y}$, d, $\mathbf{j}$ and $\widehat{\mathbf{k}}$. For $(u_1, ..., u_R) \in \mathbb{R}_{>0} \times \mathbb{R}^{R-1}_{\geq 0}$ we consider the function

$$\Psi_{\widehat{\mathbf{k}};d}(u_1, ..., u_R) = u_1^{\frac{n+s}{2}} |\det H_{\widehat{f}_{1,\mathbf{y}} + \frac{u_2}{u_1}\widehat{f}_{2,\mathbf{y}} + \cdots + \frac{u_R}{u_1} f_{R,\mathbf{y}}}(\widehat{\mathbf{k}}/d)|^{\frac{1}{2}}$$

and obtain

$$\left| \sum_{\substack{1\le j_1\le J \\ 0\le j_r\le \min\{J_r,j_1\} \\ 2\le r\le R}} j_1^n I_0(d;\mathbf{j};\widehat{\mathbf{k}}) \right| \tag{4.21}$$
$$\ll \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) d^{-\frac{n-s}{2}} \sum_{\substack{0\le j_r\le J_r \\ 2\le r\le R}} \left| \sum_{\max\{1,j_2,...,j_R\}\le j_1\le J} \Psi_{\widehat{\mathbf{k}};d}(j_1,...,j_R) e(-dj_1\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)) \right|$$
$$+ \left(\prod_{r=2}^R J_r\right) J^{\frac{n+s}{2}} d^{-\frac{n-s}{2}-1}.$$

Given any fixed $u_2,...,u_R\in\mathbb{R}_{\ge 0}$ we find that $\Psi(\cdot,u_2,...,u_R)$ is a smooth function on the set $\{u_1\in\mathbb{R}_{>0}\mid u_1\ge u_r, 2\le r\le R\}$. Let $\Psi^{(1)}_{\widehat{\mathbf{k}};d}$ denote the partial derivative of $\Psi_{\widehat{\mathbf{k}};d}$ in u_1-direction. For the following argument let

$$\mu=\max\{1,j_2,...,j_R\}.$$

Then by partial summation we have for the innermost sum

$$\sum_{\mu\le j_1\le J} \Psi_{\widehat{\mathbf{k}};d}(j_1,....,j_R) e(-dj_1\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)) \tag{4.22}$$
$$\le \Psi_{\widehat{\mathbf{k}};d}(J,j_2,...,j_R)\sum_{j_1=\mu}^{J} e(-j_1d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)) - \int_\mu^J \sum_{j_1=\mu}^{\xi} e(-j_1d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d))\Psi^{(1)}_{\widehat{\mathbf{k}};d}(\xi,j_2,...,j_R)\mathrm{d}\xi.$$

We distinguish two cases. First if $||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||\ge J^{-1}$ we obtain

$$\sum_{\mu\le j_1\le J} \Psi_{\widehat{\mathbf{k}};d}(j_1,....,j_R) e(-dj_1\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)) \tag{4.23}$$
$$\ll \frac{\Psi_{\widehat{\mathbf{k}};d}(J,j_2,...,j_R)}{||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||} + \frac{1}{||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||}\int_1^J \Psi^{(1)}_{\widehat{\mathbf{k}};d}(\xi,j_2,...,j_R)\mathrm{d}\xi$$
$$\ll \frac{\Psi_{\widehat{\mathbf{k}};d}(J,j_2,...,j_R)}{||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||}.$$

On the other hand if $||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||< J^{-1}$ we have

$$\sum_{\mu\le j_1\le J} \Psi_{\widehat{\mathbf{k}};d}(j_1,....,j_R) e(-dj_1\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)) \tag{4.24}$$
$$\ll \Psi_{\widehat{\mathbf{k}};d}(J,j_2,...,j_R)J + \int_1^J \xi\Psi^{(1)}_{\widehat{\mathbf{k}};d}(\xi,j_2,...,j_R)\mathrm{d}\xi.$$

To simplify further, we need estimates for $\Psi_{\widehat{\mathbf{k}};d}$ and $\Psi^{(1)}_{\widehat{\mathbf{k}};d}$ respectively.

Lemma 4.5. *Let $(u_1, ..., u_R) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}^{R-1}$ be such that $u_r \leq u_1$ for $2 \leq r \leq R$. Then for any $\widehat{\mathbf{k}} \in \mathscr{K}_1$ we have*

$$|\Psi_{\widehat{\mathbf{k}};d}(u_1, ..., u_R)| \ll u_1^{\frac{n+s}{2}}$$

and

$$|\Psi^{(1)}_{\widehat{\mathbf{k}};d}(u_1, ..., u_R)| \ll u_1^{\frac{n+s}{2}-1},$$

where the respective implicit constants are independant of $\mathbf{y}$, d and $\widehat{\mathbf{k}}$.

Proof. The first estimate is an obvious consequence of (2.6). Write

$$\det H_{\widehat{f}_{1,\mathbf{y}}+\alpha_2 \widehat{f}_{2,\mathbf{y}}+\cdots+\alpha_R \widehat{f}_{R,\mathbf{y}}} = \sum_{\substack{0\leq \nu_2+\cdots+\nu_R \leq n-s \\ 0 \leq \nu_2,...,\nu_R}} A_{\nu_2,...,\nu_R} \alpha_2^{\nu_2} \cdots \alpha_R^{\nu_R}.$$

Then since $\widehat{\mathbf{k}}/d \in U_{\mathbf{y}}$ we have $|A_{\nu_2,...,\nu_R}| \ll 1$, where the implicit constant is independant of $\mathbf{y}$, d and $\widehat{\mathbf{k}}$. Now by product and chain rule

$$\begin{aligned}
&|\Psi^{(1)}_{\widehat{\mathbf{k}};d}(u_1, ..., u_R)| \\
&\ll \frac{n+s}{2} u_1^{\frac{n+s}{2}-1} |\det H_{\widehat{f}_{1,\mathbf{y}}+\frac{u_2}{u_1}\widehat{f}_{2,\mathbf{y}}+\cdots+\frac{u_R}{u_1} f_{R,\mathbf{y}}}(\widehat{\mathbf{k}}/d)|^{\frac{1}{2}} \\
&+ \frac{u_1^{\frac{n+s}{2}-1}}{2|\det H_{\widehat{f}_{1,\mathbf{y}}+\frac{u_2}{u_1}\widehat{f}_{2,\mathbf{y}}+\cdots+\frac{u_R}{u_1} f_{R,\mathbf{y}}}(\widehat{\mathbf{k}}/d)|^{\frac{1}{2}}} \sum_{\substack{0\leq \nu_2+\cdots+\nu_R \leq n-s \\ 0 \leq \nu_2,...,\nu_R}} |A_{\nu_2,...,\nu_R}| \left(\frac{u_2}{u_1}\right)^{\nu_2} \cdots \left(\frac{u_R}{u_1}\right)^{\nu_R} \\
&\ll u_1^{\frac{n+s}{2}-1},
\end{aligned}$$

since $0 \leq u_2, ..., u_R \leq u_1$ and $u_1 > 0$ by assumption. ☐

Therefore we obtain

$$\begin{aligned}
M_1 \ll& \frac{1}{D} \sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ ||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)|| \leq J^{-1}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d}, \mathbf{y}\right) d^{-\frac{n-s}{2}} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}+1} \qquad (4.25) \\
&+ \frac{1}{D} \sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ J^{-1} < ||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||}} \omega\left(\frac{\widehat{\mathbf{k}}}{d}\right) d^{-\frac{n-s}{2}} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} ||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||^{-1} \\
&+ \frac{1}{D} \sum_{d=1}^{D} \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_1} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} d^{-\frac{n-s}{2}-1}.
\end{aligned}$$

With a similar argument as in the proof of Lemma 3.1 we can bound the last term in (4.25) by

$$\begin{aligned}
\frac{1}{D} \sum_{d=1}^{D} \sum_{\widehat{\mathbf{k}} \in \mathscr{K}_1} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} d^{-\frac{n-s}{2}-1} &\ll \left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}}}{D} \sum_{d=1}^{D} d^{\frac{n-s}{2}-1} \qquad (4.26) \\
&\ll \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}-1}.
\end{aligned}$$

In order to estimate the second term in (4.25) we want to sum dyadically. Note that since $||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)|| \leq \frac{1}{2}$ we can assume $J^{-1} \leq \frac{1}{2}$ and obtain

$$\frac{1}{D}\sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s} \\ J^{-1}<||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||}} \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) d^{-\frac{n-s}{2}} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} ||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||^{-1} \tag{4.27}$$

$$\leq \left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}}}{D} \sum_{i=1}^{\frac{\log J}{\log 2}+1} J2^{1-i} \sum_{d=1}^{D} d^{-\frac{n-s}{2}} \sum_{\substack{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s} \\ \frac{2^{i-1}}{J}<||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)||\leq\frac{2^i}{J}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right).$$

Now for both the first term in (4.25) and (4.27) we utilize the following result from [36, Theorem 2]: For any $X > 0$ we have

$$\sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s} \\ ||df_1(\widehat{\mathbf{k}}/d)||\leq X^{-1}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) \ll X^{-1}D^{n-s+1} + D^{n-s}\mathscr{E}_{n-s}(D), \tag{4.28}$$

where

$$\mathscr{E}_m(D) = \mathscr{E}_m^{(\mathfrak{c}_3;\mathfrak{c}_4)}(D) = \begin{cases} \exp(\mathfrak{c}_3\sqrt{\log D}) & \text{if } m = 2 \\ (\log D)^{\mathfrak{c}_4} & \text{if } m \geq 3 \end{cases}.$$

for some positive constans $\mathfrak{c}_3$ and $\mathfrak{c}_4$. Here the implicit constants as well as $\mathfrak{c}_3$ and $\mathfrak{c}_4$ only depend on n, s, c_1 and c_2 in (2.6), ρ in (2.7), ρ' in (4.10) and upper bounds for (the absolute values) of finitely many derivatives of ω and f_1 on $\mathscr{D}_+ \times \mathscr{Y}$. In particular, they are independent of $\mathbf{y}$.
By partial summation we find that

$$\sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s} \\ ||d\widehat{f}_{1,y}(\widehat{\mathbf{k}}/d)||\leq X^{-1}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) d^{-\frac{n-s}{2}} \ll D^{-\frac{n-s}{2}}(X^{-1}D^{n-s+1} + D^{n-s}\mathscr{E}_{n-s}(D)). \tag{4.29}$$

Therefore we can estimate the first term in (4.25) by

$$\frac{1}{D}\sum_{d=1}^{D} \sum_{\substack{\widehat{\mathbf{k}}\in\mathbb{Z}^{n-s} \\ ||d\widehat{f}_{1,y}(\widehat{\mathbf{k}}/d)||\leq J^{-1}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d},\mathbf{y}\right) d^{-\frac{n-s}{2}} \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}+1} \tag{4.30}$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}+1}}{D} D^{-\frac{n-s}{2}}(J^{-1}D^{n-s+1} + D^{n-s}\mathscr{E}_{n-s}(D))$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) (J^{\frac{n+s}{2}}D^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1}D^{\frac{n-s}{2}-1}\mathscr{E}_{n-s}(D))$$

and (4.27) by

$$\left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}}}{D} \sum_{i=1}^{\frac{\log J}{\log 2}+1} J2^{1-i} \sum_{d=1}^{D} d^{-\frac{n-s}{2}} \sum_{\substack{\widehat{\mathbf{k}} \in \mathbb{Z}^{n-s} \\ \frac{2^{i-1}}{J} < ||d\widehat{f}_{1,\mathbf{y}}(\widehat{\mathbf{k}}/d)|| \leq \frac{2^i}{J}}} \omega\left(\frac{\widehat{\mathbf{k}}}{d}, \mathbf{y}\right) \tag{4.31}$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}}}{D} \sum_{i=1}^{\frac{\log J}{\log 2}+1} J2^{1-i} D^{-\frac{n-s}{2}} (2^i J^{-1} D^{n-s+1} + D^{n-s} \mathscr{E}_{n-s}(D))$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) \frac{J^{\frac{n+s}{2}}}{D} ((\log J) D^{\frac{n-s}{2}+1} + J D^{\frac{n-s}{2}} \mathscr{E}_{n-s}(D))$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) ((\log J) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} D^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(D)).$$

Putting together (4.26), (4.30) and (4.31) yields

$$M_1 \ll \left(\prod_{r=2}^{R} J_r\right) (J^{\frac{n+s}{2}} D^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} D^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(D)) \tag{4.32}$$

$$+ \left(\prod_{r=2}^{R} J_r\right) ((\log J) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} D^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(D))$$

$$+ \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}-1}$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) ((\log J) J^{\frac{n+s}{2}} D^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} D^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(D)).$$

Final estimate. Recall $D = \lfloor T/2 \rfloor$ and $T \geq 2$. By combining (4.12), (4.16), (4.18) and (4.32) we obtain

$$\mathscr{M}(J, T^{-1}) \ll \left(\prod_{r=2}^{R} J_r\right) \frac{J^{n+1}}{T} + \left(\prod_{r=2}^{R} J_r\right) ((\log J) J^{\frac{n+s}{2}} T^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} T^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(T)) \tag{4.33}$$

$$+ \left(\prod_{r=2}^{R} J_r\right) \log J + \left(\prod_{r=2}^{R} J_r\right) J^{\frac{n+s}{2}} T^{\frac{n-s}{2}-1}$$

$$\ll \left(\prod_{r=2}^{R} J_r\right) (J^{n+1} T^{-1} + (\log J) J^{\frac{n+s}{2}} T^{\frac{n-s}{2}} + J^{\frac{n+s}{2}+1} T^{\frac{n-s}{2}-1} \mathscr{E}_{n-s}(T)).$$

We distinguish two cases. First if $T^{-1} \leq J^{-1}$ we have

$$\mathscr{M}(J, T^{-1}) \leq \mathscr{M}(J, J^{-1}) \ll \left(\prod_{r=2}^{R} J_r\right) J^n ((\log J) + \mathscr{E}_{n-s}(J)). \tag{4.34}$$

On the other hand, if $T^{-1} > J^{-1}$, i.e $J > T$, then

$$\mathscr{M}(J,T^{-1}) \ll \left(\prod_{r=2}^{R} J_r\right)(J^{n+1}T^{-1} + J^n((\log J) + \mathscr{E}_{n-s}(J))). \tag{4.35}$$

Therefore we conclude

$$\sum_{\substack{1\le j_1\le J\\ 0\le j_r\le \min\{J_r,j_1\}\\ 2\le r\le R}} \sum_{\substack{\mathbf{k}\in\mathbb{Z}^n\\ ||j_1\varphi_{\mathbf{y}}(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}})||<T^{-1}}} \omega^*_{\mathbf{j}}\left(\frac{\widehat{\mathbf{k}}}{j_1}\right) = \mathscr{M}(J,T^{-1}) \ll \left(\prod_{r=2}^{R} J_r\right)(J^{n+1}T^{-1}+J^n\mathscr{E}_{n-s}(J)).$$

Recall that $\omega^*_{\mathbf{j}}(\widehat{\mathbf{k}}/j_1) = \omega\circ(\nabla \widehat{F}_{y,\mathbf{j}})^{-1}(\widehat{\mathbf{k}}/j_1) = \omega(\widehat{\mathbf{x}}_{\mathbf{j};\widehat{\mathbf{k}}},\mathbf{y})$, hence Proposition 3.4 follows.

5. Proof of Theorem 0.5

Recall (2.9), hence with the bounds obtained for N_1, N_2 and N_3 in Section 3, i.e. (3.21), (3.9) and (3.10), we have

$$\begin{aligned} N^{(1;(1,\ldots,1),\mathcal{I}^n_s)}(Q,\delta) &\ll N_1+N_2+N_3 \\ &\ll (1+\log J)^R((\log Q)Q^{\frac{n+s}{2}}J^{\frac{n+s}{2}} + Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J)) \\ &+ \log Q(1+\log J)^R + J^{\frac{n+s}{2}-1}Q^{\frac{n+s}{2}}(1+\log J)^R. \\ &\ll (1+\log J)^R((\log Q)Q^{\frac{n+s}{2}}J^{\frac{n+s}{2}} + Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J)) \end{aligned} \tag{5.1}$$

Now with (2.3), (2.5) and the remark made right after (2.5) we obtain

$$\begin{aligned} &|N_\omega(Q,\delta) - (2\delta)^R N_0| \\ &\ll \delta^{R-1}\frac{Q^{n+1}}{J} + \frac{Q^{n+1}}{J^R} + (1+\log J)^R(\log Q)Q^{\frac{n+s}{2}}J^{\frac{n+s}{2}} \\ &+ (1+\log J)^R Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1}\mathscr{E}_{n-s}(J). \end{aligned} \tag{5.2}$$

Note that the constants $\mathfrak{c}'_1$ and $\mathfrak{c}'_2$ in $\mathscr{E}_{n-s}(J) = \mathscr{E}^{(\mathfrak{c}'_1;\mathfrak{c}'_2)}_{n-s}(J)$ as well as the implicit constants only depend on n, s, R, c_1 and c_2 in (2.6), ρ in (2.7), ρ' in (4.10) for each choice of $(r;\boldsymbol{\epsilon};\boldsymbol{\nu}$ $(1\le r\le R, \boldsymbol{\epsilon}\in\{\pm1\}^R, \boldsymbol{\nu}\in[n]^s)$ and upper bounds for (the absolute values) of finitely many derivatives of ω and f_1 on $\mathscr{D}_+\times\mathscr{Y}$. Recall that while we only adressed the case $(1;(1,...,1);\mathcal{I}^n_s)$ but the bounds are identical in each case. Since we can still choose the parameter $J\ge 1$, consider the equivalences

$$\begin{aligned} Q^{\frac{n+s}{2}}J^{\frac{n+s}{2}} < Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1} \quad &\Leftrightarrow \quad J<Q \\ \delta^{R-1}\frac{Q^{n+1}}{J} < \frac{Q^{n+1}}{J^R} \quad &\Leftrightarrow \quad J<\delta^{-1} \\ \frac{Q^{n+1}}{J^R} \le Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1} \quad &\Leftrightarrow \quad Q^{\frac{n-s}{n+s+2R-2}} \le J \\ \delta^{R-1}\frac{Q^{n+1}}{J} \le Q^{\frac{n+s}{2}+1}J^{\frac{n+s}{2}-1} \quad &\Leftrightarrow \quad \delta^{\frac{2(R-1)}{n+s}}Q^{\frac{n-s}{n+s}} \le J \end{aligned} \tag{5.3}$$

and distinguish two cases. If $\delta^{-1} > Q^{\frac{n-s}{n+s+2R-2}}$ then let $J = Q^{\frac{n-s}{n+s+2R-2}}$, so by the first, second and third equivalence in (5.3) we have

$$|N_\omega(Q,\delta) - (2\delta)^R N_0| \ll (\log Q)^R Q^{\frac{n^2+(s+R-1)n+(R+1)s+2R-2}{n+s+2R-2}} \mathscr{E}_{n-s}(Q).$$

If $\delta^{-1} \le Q^{\frac{n-s}{n+s+2R-2}}$ then let $J = \delta^{\frac{2(R-1)}{n+s}} Q^{\frac{n-s}{n+s}} \ge \delta^{-1}$, so by the second, third and fourth equivalence in (5.3) we have

$$|N_\omega(Q,\delta) - (2\delta)^R N_0| \ll \delta^{\frac{(R-1)(n+s-2)}{n+s}} (\log Q)^R Q^{\frac{n^2+sn+2s}{n+s}} \mathscr{E}_{n-s}(Q).$$

Note that

$$(\log Q)^R \mathscr{E}_{n-s}(Q) = (\log Q)^R \mathscr{E}_{n-s}^{(\mathfrak{c}_1';\mathfrak{c}_2')}(Q) = \begin{cases} \exp(\mathfrak{c}_1' \sqrt{\log Q} + R \log\log Q) & \text{if } n-s=2, \\ (\log Q)^{\mathfrak{c}_2'+R} & \text{if } n-s \ge 3, \end{cases}$$

hence for some absolute constant c_0 we can choose $\mathfrak{c}_1 = \mathfrak{c}_1' + c_0 R$ and $\mathfrak{c}_2 = \mathfrak{c}_2' + R$ and obtain

$$(\log Q)^R \mathscr{E}_{n-s}^{(\mathfrak{c}_1';\mathfrak{c}_2')}(Q) \ll \mathscr{E}_{n-s}^{(\mathfrak{c}_1;\mathfrak{c}_2)}(Q).$$

This completes the proof of Theorem 0.5.

Part II
Abelian varieties :

Bounds on the torsion for abelian varieties of type IV

1. Preliminaries

1.1 Abelian varieties

An abelian variety A over a number field K can be defined by several means, for example as a complete algebraic variety over K that carries a group law defined by regular functions or alternatively, in the language of schemes, as a smooth, connected, proper K-group scheme. Before working over a number field K, we give a preliminary discussion of abelian varieties over the field $\mathbb{C}$ of complex numbers and refer to [52] and [48] for further details.

Let $\mathscr{V}$ be a finite dimensional $\mathbb{C}$-vector space of dimension g and $\Lambda \subset \mathscr{V}$ a lattice of (maximal) rank $2g$. We call the quotient $\mathscr{V}/\lambda$ a complex torus of dimension g. A positive definite hermitian form $\mathcal{H}\colon \mathscr{V} \times \mathscr{V} \to \mathbb{C}$ is called a Riemann form on $\mathscr{V}/\Lambda$, if the imaginary part $\operatorname{Im}\mathcal{H}\colon \mathscr{V} \times \mathscr{V} \to \mathbb{R}$ only takes integer values on Λ.

Definition 1.1. An abelian variety A over $\mathbb{C}$ is a complex torus that admits a Riemann form $\mathcal{H}$.

Let $\varphi\colon A \to A'$ be a morphism between complex abelian varieties. We introduce the notion of an isogeny:

Definition 1.2. An isogeny is a surjective morphism $\varphi\colon A \to A'$ between complex abelian varieties that has a finite kernel. If such an isogeny exists between A and A' we call them isogenous (and non-isogenous otherwise).

A crucial result for the structure of abelian varieties is the following theorem:

Theorem 1.3 (Poincaré Reducibility Theorem). *Let A be a complex abelian variety. For any abelian subvariety $A' \subset A$ there exists another abelian subvariety $A'' \subset A$ and an isogeny $A \to A' \times A''$.*

We call an abelian variety simple, if it does not contain any proper, non-zero abelian subvarieties. Theorem 1.3 now implies that any abelian variety admits a decomposition into simple factors. Precisely there exist finitely many simple subvarieties $A_1, ..., A_k \subset A$, that are pairwise non-isogenous, positive integers $n_1, ..., n_k$ and an isogeny

$$A \longrightarrow \prod_{i=1}^{k} A_i^{n_i}.$$

It is standard to work in the category of abelain varieties up to isogeny, that is the category with abelian varieties as objects and the set of morphism between abelian varieties A, B defined to be the vectorspace $\mathrm{Hom}(A,B)\otimes_{\mathbb{Z}}\mathbb{Q}$. We write $\mathrm{End}^\circ(A) := \mathrm{End}_{\overline{K}}(A)\otimes_{\mathbb{Z}}\mathbb{Q}$ and obtain from the isogeny decomposition above the decomposition

$$\mathrm{End}^\circ(A) = \prod_{i=1}^{k} \mathrm{Mat}_{n_i}(\mathrm{End}^\circ(A_i)).$$

In the following we will assume that all endomorphisms of A are already defined over K, that is $\mathrm{End}_K(A) = \mathrm{End}_{\overline{K}}(A)$. If A is a simple abelian variety and we are not at risk of ambiguity, we usually write $D = \mathrm{End}^\circ(A)$.

Proposition 1.4. *Let A be a simple abelian variety, then D is a division algebra. If A is any (not necessarily simple) abelian variety, then* $\mathrm{End}^\circ(A)$ *is a semi-simple* $\mathbb{Q}$*-algebra.*

In algebraic geometric terms the notion of a Riemann form corresponds to a polarization, that is an isogeny $\phi\colon A \to A^\vee$ from A into the dual variety $A^\vee$. The degree of ϕ as a polarization is just its degree as an isogeny, that is the degree of its finite kernel. We can naturally associate an involution

$$^\dagger\colon \mathrm{End}^\circ(A) \longrightarrow \mathrm{End}^\circ(A), \alpha \mapsto \alpha^\dagger$$

to ϕ via

$$\phi \circ \alpha^\dagger = \alpha^\vee \circ \phi,$$

where $\alpha^\vee \in \mathrm{End}^\circ(A^\vee)$ is the element in the opposite algebra corresponding to $\alpha \in \mathrm{End}^\circ(A)$. One can show that for $\alpha \in \mathrm{End}^\circ(A)$ there exists $\beta \in \mathrm{End}^\circ(A)$ and a positive integer m such that

$$\phi \circ b = ma^\vee \circ \phi.$$

Hence we define the positive Rosati involution by

$$a^\dagger = \frac{1}{m} \times b.$$

For the following let A be a simple abelian variety defined over a number field K with a fixed polarization $\phi\colon A \to A^\vee$ and the corresponding Rosati involution $^\dagger\colon D \to D$ as described above. Denote by $E := Z(D)$ the center of D, making D into a central E-algebra, and denote further $E_0 := \{\varphi \in E \mid \varphi^\dagger = \varphi\}$ the subset of elements fixed by the involution. We therefore have a tower

$$D \supset E \supset E_0 \supset \mathbb{Q},$$

where $E/\mathbb{Q}$ and $E_0/\mathbb{Q}$ are field extensions with degrees $e := [E : \mathbb{Q}]$ and $e_0 := [E_0 : \mathbb{Q}]$ respectively. Note that D as a central simple algebra has a square degree over E, hence we denote $d^2 := [D : E]$.

Due to Albert [52, Chapter 4], simple abelian varieties can be classified according to the type of their endomorphism algebra. We distinguish four types:

- Type I: $D = E = E_0$ is a totally real field and $^\dagger = \mathrm{id}_D$.
- Type II: $E = E_0$ is a totally real field and D is an indefinite quaternion algebra over $\mathbb{Q}$, such that $D \otimes_{E_0} \mathbb{R} = \mathrm{Mat}_2(\mathbb{R})$ for any embedding $E_0 \hookrightarrow \mathbb{R}$. $^\dagger = (x \mapsto a^{-1}(Tr^0_{D/E}(x) - x)a)$ where $Tr^0_{D/E}$ is the reduced trace of D/E and $a \in D \setminus \{0\}$ with $a^2 \in E$ totally negative.
- Type III: $E = E_0$ is a totally real field and D is a definite quaternion algebra over $\mathbb{Q}$, such that $D \otimes_{E_0} \mathbb{R} = \mathbb{H}$ for any embedding $E_0 \hookrightarrow \mathbb{R}$. $^\dagger = (x \mapsto Tr^0_{D/E}(x) - x)$ with $Tr^0_{D/E}$ as before.
- Type IV: E_0 is a totally real field, E is totally imaginary quadratic extension of E_0 and D is a division algebra over $\mathbb{Q}$, such that $D \otimes_{E_0} \mathbb{R} = \mathrm{Mat}_d(\mathbb{C})$. $^\dagger{}_{|E} = (x \mapsto \bar{x})$.

Remark 1.5. A totally imaginary quadratic field extension E over a totally real field E_0 as present in the type IV case is called a CM field. This distinction leads us to the following definition.

For a simple abelian variety A, we say that A is of type I, II, III or IV, if the endomorphism algebra (and the respective involution) are of the corresponding type. If A is of type I, we necessarily have $d = 1$. For types II and III respectively we have $d = 2$. For A of type IV $d \geq 1$ can be arbitrary and moreover, $e = 2e_0$ in that case. If A is of type IV and $d = 1$ we say that A is an abelian variety with complex multiplication or A is of CM type.

Definition 1.6. A simple abelian variety A is called fully of type IV, if it is of type IV but not of CM type.

Definition 1.7. Let A be a simple abelian variety of dimension g. With the notation from above we define the relative dimension of A as

$$h := \dim_{rel}(A) := \begin{cases} \frac{g}{e} & \text{if } A \text{ is of type I,} \\ \frac{g}{2e} & \text{if } A \text{ is of type II or III,} \\ \frac{g}{d^2 e_0} & \text{if } A \text{ is of type IV.} \end{cases}$$

1.2 Tate Module

Let ℓ be a prime number, such that ℓ does not divide $\deg \phi$. We define the Tate module of A to be

$$T_\ell(A) := \varprojlim_n A[\ell^n],$$

where $A[\ell^n]$ is the kernel of the multiplication by ℓ^n map. Furthermore let

$$V_\ell(A) = T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell.$$

Then we have isomorphisms of topological groups:

$$T_\ell(A) \simeq \mathbb{Z}_\ell^{2g} \quad \text{and} \quad V_\ell(A) \simeq \mathbb{Q}_\ell^{2g}.$$

There exists a non-degenerate bilinear form

$$< \cdot, \cdot >: T_\ell(A) \times T_\ell(A^\vee) \to \mathbb{Z}_\ell(1) = \varprojlim \mu_{\ell^n},$$

which is Galois equivariant, called the Weil pairing. The polarization $\phi\colon A \to A^\vee$ therefore induces a non-degenerate, alternating, bilinear pairing

$$\phi_{\ell^\infty}: T_\ell(A) \times T_\ell(A) \xrightarrow{\mathrm{id}\times\phi} T_\ell(A) \times T_\ell(A^\vee) \xrightarrow{<\cdot,\cdot>} \mathbb{Z}_\ell(1) = \varprojlim \mu_{\ell^n}.$$

Since ℓ does not divide the degree of the polarization ϕ, the ℓ-adic Weil pairing is non-degenerate on $A[\ell^n]$ for all $n \geq 1$. Let $\mathcal{O}^*_{E_\ell}$ denote the dual of $\mathcal{O}_{E_\ell}$, the ring of integers of $E_\ell = E \otimes \mathbb{Q}_\ell$, induced by the trace map

$$\mathrm{Tr}_{E_\ell/\mathbb{Q}_\ell}\colon \mathrm{Hom}_{\mathcal{O}_{E_\ell}}(T_\ell(A) \otimes_{\mathcal{O}_{E_\ell}} T_\ell(A), \mathcal{O}^*_{E_\ell}) \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(A) \otimes_{\mathcal{O}_{E_\ell}} T_\ell(A), \mathbb{Z}_\ell).$$

By [33, Lemme 3.3] there exists a unique $\mathcal{O}_{E_\ell}$-linear pairing

$$\phi^*_{\ell^\infty}: T_\ell(A) \times T_\ell(A) \to \mathcal{O}^*_{E_\ell}(1)$$

such that $\mathrm{Tr}_{E_\ell/\mathbb{Q}_\ell}(\phi^*_{\ell^\infty}) = \phi_{\ell^\infty}$. If ℓ is unramified in $\mathcal{O}_E$, we have $\mathcal{O}_{E_\ell} = \mathcal{O}^*_{E_\ell}$, hence we have

$$\phi^*_{\ell^\infty}: T_\ell(A) \times T_\ell(A) \to \mathcal{O}_{E_\ell}(1).$$

Let $\mathcal{O}^0_E = \mathrm{End}(A) \cap \mathcal{O}_E$ and for each prime ideal λ in $\mathcal{O}_E$ dividing ℓ let $\mathcal{O}_\lambda$ be the completion of $\mathcal{O}_E$ at λ and E_λ its fraction field. Then

$$E_\ell = \prod_{\lambda|\ell} E_\lambda \quad \text{and} \quad \mathcal{O}_{E_\ell} = \prod_{\lambda|\ell} \mathcal{O}_\lambda$$

for all ℓ. If ℓ does not divide $(\mathcal{O}_E : \mathcal{O}^0_E)$ we have $\mathcal{O}_{E_\ell} = \mathcal{O}_E \otimes_\mathbb{Z} \mathbb{Z}_\ell = \mathcal{O}^0_E \otimes_\mathbb{Z} \mathbb{Z}_\ell$ and $\mathcal{O}_{E_\ell}$ acts on $T_\ell(A)$. Therefore we obtain a decomposition

$$T_\ell(A) = \prod_{\lambda|\ell} \mathcal{T}_\lambda,$$

where $\mathcal{T}_\lambda = T_\ell(A) \otimes_{\mathcal{O}_{E_\ell}} \mathcal{O}_\lambda$, and an $\mathcal{O}_\lambda$-linear pairing

$$\phi_{\lambda^\infty}: \mathcal{T}_\lambda \times \mathcal{T}_\lambda \to \mathcal{O}_\lambda(1).$$

The explicit decomposition of $\mathcal{T}_\lambda$ is going to be a key ingredient for the proof of our theorem and will be further discussed in Section 3.2. Note that all the ℓ-adic and λ-adic pairings defined above are Galois equivariant and we may write $\phi^0_{\ell^\infty}$ and $\phi^0_{\lambda^\infty}$ respectively for the pairings defined on V_ℓ and V_λ with values in E_ℓ and E_λ.

In order that the above definitions are well-defined, we have restricted the set of primes ℓ that we consider. The following definition makes this precise:

Definition 1.8. Denote $\mathcal{O}^0_E = \mathrm{End}(A) \cap \mathcal{O}_E$ and $\mathcal{O}^0_{E_0} = \mathrm{End}(A) \cap \mathcal{O}_{E_0}$. Let L be a Galois extension over E_0 containing E such that

$$D \otimes_E L \cong \mathrm{Mat}_d(L)$$

as an L-algebra (compare [2, Lemma 2.1]). Let $\mathcal{P}$ be the set of prime numbers ℓ such that:

1. $\ell \nmid (\mathcal{O}_E : \mathcal{O}^0_E), (\mathcal{O}_{E_0} : \mathcal{O}^0_{E_0})$;

2. ℓ is unramified in $\mathcal{O}_L$;

3. $\ell \nmid \deg(\phi)$, where ϕ is the fixed polarization of A;

4. $\exists \lambda_0 \in \mathcal{O}_{E_0}, \lambda \in \mathcal{O}_E$, $\lambda_0 | \ell$; λ is inert over λ_0 and splits completely in $\mathcal{O}_L$.

We sometimes want to address the complementary set of primes:

Definition 1.9. Let $\mathcal{S} = \{\ell \text{ prime} \mid \ell \notin \mathcal{P}\}$.

Remark 1.10. Note that from Corollary 2.3 and Lemma 2.1 in [2] we obtain that $\mathcal{S}$ is a finite set. For a prime $\ell \in \mathcal{S}$ at least one of the following conditions holds:

1. ℓ is ramified in $\mathcal{O}_L$,

2. $\ell \mid \deg(\phi)$,

3. D does not split over E_λ for at least one $\lambda | \ell$.

For the rest of the paper we may assume that any prime ℓ is not contained in $\mathcal{S}$. We refer the reader to Section 6 for a treatment of the the case $\ell \in \mathcal{S}$.

1.3 Algebraic groups attached to abelian varieties

In this section, we discuss several algebraic groups that arise naturally within the theory of abelian varieties and are particularly important in the study of the invariant

$$\gamma(A) = \inf\{x > 0 \,|\, \forall L/K \text{ finite},\ |A(L)_{\mathrm{tors}}| \ll [L:K]^x\}.$$

Definition 1.11 (Lefschetz group). The Lefschetz group $L(A) := C_{\mathrm{GSp}(V,\psi)}(D)$ of A is the centralizer of D in $\mathrm{GSp}(V,\psi)$. We write $L(A)^\circ$ for the (connected) identity component of $L(A)$.

Remark 1.12. Note that $L(A)^\circ$ is a connected, reductive algebraic group subgroup of $\mathrm{GSp}(V,\psi)$ over $\mathbb{Q}$. For a definition of the symplective, orthogonal and unitary groups and spaces, that will appear throughout this section, we refer to Section 2.

From the work of Deligne [23], Piatetski-Shapiro [54] and Borovoi [7, Lemma 2] it is known that the Lefschetz, Hodge and ℓ-adic monodromy groups of A can be compared to one another in the following way for every prime number ℓ:

$$G_{\ell,1} \subseteq \mathrm{Hg}(A)_{\mathbb{Q}_\ell} \subseteq L(A)^\circ_{\mathbb{Q}_\ell}. \tag{1.1}$$

Definition 1.13 (Fully of Lefschetz type). An abelian variety A is said to be fully of Lefschetz type when the Mumford-Tate conjecture holds for A, and $\mathrm{MT}(A) = L(A)$.

The following classification of the Lefschetz groups for simple abelian varieties A of dimension g depending on their type is essentially due to Milne [47]. Note his definition of the group $S(A)$ ([47, p. 8]) and the table ([47, p. 14]) which is the summary of section 2 in his paper:

Type of A	Group
I	$\mathrm{Sp}_{\frac{2g}{e_0}}$
II	$\mathrm{Sp}_{\frac{g}{e_0}}$
III	$O_{\frac{g}{e_0}}$
IV	$\mathrm{GL}_{\frac{g}{de_0}}$

Now the group $S(A)_{\overline{K}}$ is isomorphic over $\overline{K}$ to e_0 copies of the groups listed above. A paper of Murty [53] following the definition of Milne gives a classification of possible factors of $L(A)$. For that, consider a maximal commutative subfield $\mathcal{K}$ of D and its maximally totally real subfield $\mathcal{K}_0$. This induces a decomposition

$$V_{\mathbb{R}} = \prod_{\sigma\colon \mathcal{K}_0 \to \mathbb{R}} V_\sigma$$

indexed by all possible real embeddings of $\mathcal{K}_0$. Except for type I (where $\mathcal{K}$ is automatically totally real as well), $\mathcal{K} \otimes \mathbb{R}$ acts on $V_{\mathbb{R}}$ and therefore induces a complex structure on V_σ. Write $V_{\sigma,\mathbb{C}}$ for the corresponding $\mathbb{C}$-vector space and note that this is in general different from $V_\sigma \otimes \mathbb{C}$. Now we have

$$L(A)_{\mathbb{R}} \simeq \prod_\sigma L_\sigma,$$

where L_σ is the projection of $L(A)_{\mathbb{R}}$ onto $\mathrm{GL}(V_\sigma)$ and σ runs through a set of representatives for the equivalence $\sigma_1 \sim \sigma_2$ if and only if ${\sigma_1}_{|_{E_0}} = {\sigma_2}_{|_{E_0}}$. The following factors occur:

Type of A	L_σ
I	$\mathrm{Sp}(V_\sigma)$
II	$\mathrm{U}(V_{\sigma,\mathbb{C}}) \cap \mathrm{Sp}(V_{\sigma,\mathbb{C}})$
III	$\mathrm{U}(V_{\sigma,\mathbb{C}}) \cap O(V_{\sigma,\mathbb{C}})$
IV	$\mathrm{U}(V_{\sigma,\mathbb{C}})$

Murty further notes that after complexification one obtains

Type of A	$L_\tau \otimes \mathbb{C}$
I	$\mathrm{Sp}(V_\tau \otimes \mathbb{C})$
II	$\mathrm{Sp}(V_\tau \otimes \mathbb{C})$
III	$O(V_\tau \otimes \mathbb{C})$
IV	$\mathrm{GL}(V_\sigma)$

where $\tau\colon E_0 \hookrightarrow \mathbb{R}$ is any embedding of the (totally real) fixed field of the Rosati involution and σ is any extension of τ to $\mathcal{K}$. We deduce the following properties of $L(A)$:
- $L(A)$ is contained in the group of symplectic similitudes GSp_{2g} if A is of type I or II,
- $L(A)$ is contained in the group of orthogonal similitudes GO_{2g} if A is of type III ,
- $L(A)$ is contained in the group of unitary similitudes GU_{2g} if A is of type IV.

Now let V be a finite dimensional $\mathbb{Q}$-vector space. By extension of scalars, we consider the complex vector space $V_{\mathbb{C}} := V \otimes_{\mathbb{Q}} \mathbb{C}$ together with a complex conjugation defined via

$$\overline{v \otimes z} = v \otimes \overline{z}$$

for $v \in V$ and $z \in \mathbb{C}$. A decomposition

$$V_{\mathbb{C}} = \bigoplus_{p+q=n} V^{p,q}$$

of V into complex vector spaces with $\overline{V^{p,q}} = V^{q,p}$ is called a pure Hodge structure of weight $n \in \mathbb{Z}$. A Hodge class is a vector $v \in V$ that belongs to $V^{0,0}$ and we call the collection of pairs (p,q), such that $V^{p,q}$ is non-trivial, the type of the Hodge structure. Equivalently, a Hodge structure is given by a Hodge cocharacter

$$h \colon \mathbb{S} \to \mathrm{GL}(V)_{\mathbb{R}}.$$

That is a representation of the Deligne torus $\mathbb{S} := \mathrm{Res}_{\mathbb{C}/\mathbb{R}}(\mathbb{G}_{m,\mathbb{C}})$, which arises from Weil restriction of scalars with the properties $\mathbb{S}(\mathbb{R}) = \mathbb{C}^\times$ and $\mathbb{S}(\mathbb{C}) = \mathbb{C}^\times \times \mathbb{C}^\times$. The subspace $V_{\mathbb{C}}^{p,q}$ is precisely the subspace of elements $v \in V_{\mathbb{C}}$ on which $z = (z_1, z_2) \in \mathbb{S}(\mathbb{C})$ acts by multiplication of $z_1^{-p} z_2^{-q}$.

Now let A be an abelian variety defined over the number field K and fix an embedding $\sigma \colon K \hookrightarrow \mathbb{C}$ of the number field into the complex numbers. Then the vector space $V = H_1(A_\sigma(\mathbb{C}), \mathbb{Q})$ carries a natural Hodge structure of weight -1, where A_σ denotes the complex abelian variety induced by the embedding. In the following we will by abuse of notation suppress the subscript σ for clarity.

Definition 1.14 (Mumford-Tate group). The Mumford-Tate group of A, denoted $\mathrm{MT}(A)$, is the smallest algebraic subgroup of $\mathrm{GL}(V)$ over $\mathbb{Q}$, such that the Hodge cocharacter h factors through $\mathrm{MT}(A)_{\mathbb{R}}$.

We introduce the unit circle group $\mathcal{U}_1 \subset \mathbb{S}$, whose real valued points correspond to the unit circle $U_1 = \{z \in \mathbb{C}^\times \mid z\overline{z} = 1\}$.

Definition 1.15 (Hodge group). The Hodge group of A, denoted $\mathrm{Hg}(A)$, is the smallest algebraic subgroup of $\mathrm{GL}(V)$ over $\mathbb{Q}$, such that $h\big|_{\mathcal{U}_1}$ factors through $\mathrm{Hg}(A)_{\mathbb{R}}$.

Remark 1.16. Note that, by construction, $\mathrm{MT}(A)$ is the almost-direct product inside $\mathrm{GL}(V)$ of $\mathrm{Hg}(A)$ with the central torus of homotheties $\mathbb{G}_{m,\mathbb{Q}}$.

Another algebraic group naturally attached to the abelian variety A can be defined via the ℓ-adic representation for the absolute Galois group of K, denoted $G_K := \mathrm{Gal}(\overline{K}/K)$,

$$\rho_\ell \colon G_K \to \mathrm{GL}(T_\ell(A))$$

for any prime ℓ as follows:

Definition 1.17 (ℓ-adic monodromy group). The ℓ-adic monodromy group of A, denoted by G_ℓ, is the Zariski closure $\overline{\rho_\ell(G_K)}^{\mathrm{Zar}}$ of the image of the ℓ-adic representation ρ_ℓ. Similarly, we define the special ℓ-adic monodromy group $G_{\ell,1} := G_\ell \cap \mathrm{SL}_{V_\ell}$.

The relation between the Mumford-Tate and Hodge groups is immediate from their definitions, the relation of the ℓ-adic monodromy group to them is famously stated as the Mumford-Tate conjecture:

Conjecture 1.18 (Mumford–Tate conjecture). *For every prime number ℓ we have $G_\ell^\circ = \mathrm{MT}(A)_{\mathbb{Q}_\ell}$, or equivalently $G_{\ell,1}^\circ = \mathrm{Hg}(A)_{\mathbb{Q}_\ell}$.*

Many results have been produced towards identifying whether certain classes of abelian varieties satisfy the Mumford-Tate conjecture. Already in 1972, Serre showed that all elliptic curves satisfy the Mumford-Tate conjecture. Ribet managed to prove (without the powerful result of Faltings existing) that the Mumford-Tate conjecture holds for all abelian varieties of dimension g with endomorphism algebra a totally real field of dimension g and that admit a semi-stable place of bad reduction [56, Paragraph V]. Following Faltings theorem, many more cases of the Mumford-Tate conjecture have been verified. In 1985 Serre proved that all abelian varieties of odd dimension whose endomorphism algebra is $\mathbb{Q}$ satisfy the Mumford-Tate conjecture [62]. In 1991 all simple abelian varieties of prime dimension have been proved to satisfy the Mumford-Tate conjecture by Chi [19]. One year later, Chi showed some additional cases:

Theorem 1.19 (Chi 1992, [20]**).** *The Mumford-Tate conjecture holds in the following four cases:*
If A is a simple abelian variety of type I and

(i) g is an odd integer or 2 and further $\mathrm{End}^\circ(A) = \mathbb{Q}$;
(ii) $g = 2d$ where d is an odd integer and further $\mathrm{End}^\circ(A) = E$ is a real quadratic field.

If A is a simple abelian variety of type II and

(iii) $g = 2d$ where d is an odd integer or 2 and further D is an indefinite quaternion algebra over $\mathbb{Q}$,
(iv) $g = 4d$ where d is an odd integer and further D is an indefinite quaternion algebra over a real quadratic field.

Pink defined the following set

$$\Sigma := \left\{g \geq 1 \mid \exists\, k \geq 3, \exists\, a \geq 1, g = 2^{k-1}a^k\right\} \cup \left\{g \geq 1 \mid \exists\, k \geq 3, 2g = \binom{2k}{k}\right\},$$

where k is an odd integer, and showed that all abelian varieties with endomorphism algebra $\mathbb{Q}$ of dimension $g \notin \Sigma$ satisfy the Mumford-Tate conjecture [55]. This result has then been used by Banszak, Gajda and Krasoń to generalize the findings of Chi for abelian varieties of type I and II:

Theorem 1.20 (Banaszak, Gajda, Krasoń, [1]**).** *Lat A be a simple abelian variety of type I or II with relative dimension either equal to 2 or odd. Then $\mathrm{MT}(A) = \mathrm{GSp}_{2g}$ and $G_\ell^\circ = \mathrm{MT}(A)_{\mathbb{Q}_\ell}$, hence A satisfies the Mumford-Tate conjecture.*

A theorem by Hall from 2011 covers the following cases:

Theorem 1.21 (Hall, [28]**).** *Let A be an abelian variety with $\mathrm{End}^\circ(A) = \mathbb{Q}$ and such that the Néron model of A over $\mathcal{O}_K$ has a semi-stable fibre of toric dimension equal to 1. Then $G_\ell^\circ = \mathrm{GSp}_{2g,\mathbb{Q}_\ell}$, hence A satisfies the Mumford-Tate conjecture.*

Hindry and Ratazzi managed to generalize the results of Hall and Pink:

Theorem 1.22 (Hindry, Ratazzi, [33]**).** *Let A be a geometrically simple abelian variety of type I or II, such that the center of $\mathrm{End}^\circ(A)$ is a totally real field of degree e. Then A is fully of Lefschetz type (see Definition 1.13) and in particular satisfies the Mumford-Tate conjecture if one of the following conditions holds:*

(i) The relative dimension h of A is either equal to 2 or odd.
(ii) We have $e = 1$ and $h \notin \Sigma$.

(iii) A has a semi-stable place of bad reduction with toric dimension equal to e if A is of type I or 2e if A is of type II.

Note that an analogous result exists for a product of abelian varieties [33, Corollaire 1.15]. Additionally, it has been shown by Moonen and Zharin [50] and by Lombardo [40] that any abelian variety (not necessarily simple) of dimsion 5 or lower satisfies the Mumford-Tate conjecture.

We want to work with abelian varieties A that not only satisfy the Mumford-Tate conjecture, additionally we would like A to have the biggest possible Mumford-Tate group. In order to define this notion precisely, let $D = \mathrm{End}^\circ(A)$ as before and note that the polarization $\phi\colon A \to A^\vee$ induces a symplectic bilinear form ψ on V.

1.4 Notations

We shall make us of the following notations:

Let L_1, L_2 be number fields that are contained in one common field L and that additionally depend on A/K and a finite set of other parameters. We write $L_1 \asymp L_2$ if there exists a constant $c(A/K)$ that only depends on A/K, such that the inequalities

$$[L_1 : L_1 \cap L_2] \leq c(A/K)$$

and

$$[L_2 : L_1 \cap L_2] \leq c(A/K)$$

hold for all the other parameters given. For groups G_1, G_2 that are subgroups in one common group G and depend on a finite set of parameters, we similarly write $G_1 \asymp G_2$ if there exists a constant $C(A/K)$ that only depends on A/K, such that the inequalities

$$(G_1 : G_1 \cap G_2) \leq C(A/K)$$

and

$$(G_2 : G_1 \cap G_2) \leq C(A/K)$$

hold for all the other parameters given.

In both cases we say that the equality $L_1 = L_2$ (resp. $G_1 = G_2$) holds up to a bounded index.

2. Group lemmas

This section is mainly devoted to establish the index $(G(\mathbb{Z}_\ell) : G(H))$, for an abelian variety A, a finite subgroup $H \subset A[\ell^\infty]$ and an algebraic subgroup $G \subset \mathrm{GL}_{2g}$ over $\mathbb{Z}$. The group $G(H)$ is defined as follows

$$G(H) = \{\sigma \in G(\mathbb{Z}_\ell) \mid \sigma_{|H} = \mathrm{id}_H\}$$

and can be interpreted as a stabilizer of a group action of G on $T_\ell(A)$. A precise definition of this notion is given in Section 2.3. These results will be useful for the proof of the main theorem in Section 4 for $G = \mathrm{Hg}(A)$ as is described in Section 3.1.

2.1 Hermitian spaces and unitary group

From [2, Section 3] we know that we have non-degenerate hermitian pairings $\phi^{\circ}_{\ell^\infty}, \phi^{\circ}_{\lambda^\infty}, \phi_{\ell^\infty}, \phi_{\lambda^\infty}$ on the spaces $V_\ell, V_\lambda, T_\ell, T_\lambda$ respectively. Let (V, ψ) be a hermitian space of dimension n, where we may assume that $V \in \{V_\ell, V_\lambda, T_\ell, T_\lambda\}$ and $\psi \in \{\phi^{\circ}_{\ell^\infty}, \phi^{\circ}_{\lambda^\infty}, \phi_{\ell^\infty}, \phi_{\lambda^\infty}\}$.

Definition 2.1. We define the unitary group $U(V, \psi)$ and the special unitary group $SU(V, \psi)$ as follows:

(i) $\mathrm{U}(V, \psi) = \{u \in \mathrm{GL}(V) \mid \psi(ux, uy) = \psi(x, y), \forall\, x, y \in V\}$,

(ii) $\mathrm{SU}(V, \psi) = U \cap \mathrm{SL}(V)$.

If there is no ambiguity at risk concerning the form ψ, we shall write U_n and SU_n respectively.

Let F be the field or ring over which V is defined.

Definition 2.2. We say that a linear transformation u on V is a unitary similitude, if there is a constant $a \in F^\times$, such that for all $x, y \in V$

$$\psi(ux, uy) = a\psi(x, y).$$

We denote the group of unitary similitudes as $\mathrm{GU}(V, \psi)$ (resp. GU_n as before).

For a unitary similitude u we denote the unique associated factor a defined above as mult_u. Consequently we have a map

$$\mathrm{mult}\colon\ \mathrm{GU}_n \longrightarrow \mathbb{G}_m, u \mapsto \mathrm{mult}_u$$

and conclude that $u \in U_n$ if $\mathrm{mult}_u = 1$.

Definition 2.3. A unitary similitude $u \in \mathrm{GU}_n$ is called a (unitary) isometry, if $\mathrm{mult}_u = 1$ or equivalently if $u \in U_n$.

Remark 2.4. The following notions of orthogonality, isotropy and the results on the stabilizers are essentially equivalent, if we replace the hermitian space (V, ψ) by a quadratic space (resp. a symplectic space) and the unitary group U and group of similitudes GU by the orthogonal groups O and GO (resp. by the symplectic groups Sp and GSp), which are defined in the exact same way for a symmetric bilinear form (resp. an antisymmetric bilniear form) ψ. These appear for abelian varieties of type III (resp. type I and II). The respective dimensions of these groups are known:

$$\dim \mathrm{U}_n = n^2; \quad \dim O_n = \frac{n(n-1)}{2}; \quad \dim \mathrm{Sp}_n = \frac{n(n+1)}{2},$$

which can be deduced with a careful study of [6, 23.9 p. 260ff].

Let G be a group acting on V. Naturally we can have $G \in \{\mathrm{GU}, \mathrm{U}, \mathrm{GO}, O, \mathrm{GSp}, \mathrm{Sp}\}$. Given a vector subspace $W \subset V$ we are interested in the stabilizer of W for the action of G, which is defined as

$$G_W := \{\tau \in G \mid \tau_{|W} = \mathrm{id}_W\}.$$

In order to study G_W we employ properties of the subspace W. The following definitions as well as the theorem of Witt can be formulated equivalently for (V, ψ) being a hermitian, quadratic or symplectic space.

Definition 2.5. Two vectors $x, y \in V$ are said to be orthogonal, if $\psi(x,y) = 0$. If $W \subset V$ is a vector subspace, the set of all vectors that are orthogonal to every vector in W is called the orthogonal complement of W and denoted by

$$W^\perp = \{x \in V \mid \psi(x,w) = 0 \quad \forall\, w \in W\}.$$

We say ψ is non-degenerate if $V^\perp = \{0\}$.

The relation between W and $W^\perp$ is useful to classify the properties of W and its stabilizer.

Definition 2.6. A non-trivial vector $x \in V \setminus \{0\}$ is called isotropic, if $\psi(x,x) = 0$. A vector subspace $W \subset V$ is called isotropic if $W \cap W^\perp \neq \{0\}$ and totally isotropic if $W \subset W^\perp$.

Definition 2.7. The index ν of the form ψ is the biggest dimension amongst the totally isotropic subspaces of V; any totally isotropic subspace of dimension ν is called a maximal isotropic subspace.

Totally isotropic subspaces usually appear in complementary pairs in the following sense:

Proposition 2.8. [24]*(§11, p. 21, 1)) Let $W \subset V$ be a totally isotropic subspace, then there exists another totally isotropic subspace $P \subset V$ with $\dim P = \dim W$ and $W \cap P = \{0\}$. Furthermore we can find a basis $\{e_1, ..., e_r\}$ of W and a basis $\{e_{r+1}, ..., e_{2r}\}$ of P such that $\psi(e_i, e_{r+j}) = \delta_{i,j}$.*

We easily deduce the following corollary:

Corollary 2.9. *Let W be a totally isotropic subspace of dimension r, then there exists a totally isotropic subspace W' of dimension r such that $W \oplus W' = P_1 \oplus \cdots \oplus P_r$, where the P_i are hyperbolic planes.*

Proof. Let $\{e_1, ..., e_r\}$ be a basis of W, then we can find a vector $f_1' \in V$ with $\psi(f_1', f_1') = 0$ and $\psi(e_1, f_1') = 1$, such that $< e_1, f_1' >$ is a hyperbolic plane. Consider the vector

$$f_1 := f_1' + \sum_{i=2}^{r} \lambda_i e_i,$$

where λ_i are chosen such that $\psi(e_i, f_1) = 0$ for all $i = 2, ..., r$. Hence we obtain

$$< e_1, ..., e_r, f_1 >=< e_1, f_1 > \perp < e_2, ..., e_r > .$$

Repeating this construction in a similar fashion yields isotropic vectors $f_2, ..., f_r$ such that

$$< e_1, ..., e_r, f_1, ..., f_r >=< e_{1,f_1} > \perp \cdots \perp < e_r, f_r > .$$

Now $W' =< f_1, ..., f_r >$ has all desired properties. □

Therefore a maximally isotropic subspace can at most have dimension $n/2$, hence $2\nu \leq \dim V$. The discussion in [24](§11, p.21-22, 2)) yields that isometries can be extended from a subspace W to V. Precisely we have the following:

Theorem 2.10 (Witt). *Let (V, ψ) be a finite dimensional, non-degenerate hermitian space and $W \subset V$ a vector-subspace. For every isometry $u \in U(W, \psi_{|W})$ of W we can find an extension to V, i.e. an isometry $a \in U(V, \psi)$ such that $a_{|W} = u$.*

2.2 Dimension of the stabilizer

We now compute $\dim G_W$.

Theorem 2.11. *Let (V, ψ) be a non-degenerate hermitian space of dimension n and $W \subset V$ a vector subspace of codimension d. For the stabilizer of W, defined as follows*

$$G_W = \{\tau \in U(V, \psi), \tau_{|W} = \mathrm{id}_W\} \subset \mathrm{GL}_n,$$

we find that

$$\dim G_W \leq d^2.$$

Proof. We present the proof in two steps.

First assume that W is non-isotropic, i.e. $W \cap W^\perp = \{0\}$. Then we have the decomposition $V = W \oplus W^\perp$ with $\dim W^\perp = d$ and $\dim W = r = n - d$. Let $\{e_1, ..., e_r\}$ be a basis of W, then we can extend it to a basis $\{e_1, ..., e_n\}$ of V. By construction we have $W^\perp =< e_{r+1}, ..., e_n >$ and hence we can write the stabilizer within the basis as

$$G_W = \left\{ \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & \mathrm{GU}_d \end{array} \right) \right\}.$$

Therefore $G_W \cong \mathrm{GU}(W^\perp, \psi_{|_{W^\perp}})$ and

$$\dim G_W = d^2.$$

Now assume that W is isotropic. We present the case where W is a maximally isotropic subspace.

Assume further that $2r = n$. By Proposition 2.8 we can find another maximally isotropic subspace $P \subset V$ with $W \oplus P = V$ and a basis $\{e_1, ..., e_{2r}\}$ of V such that $< e_1, ..., e_r >= W$, $< e_{r+1}, ..., e_{2r} >= P$ and $\phi(e_i, e_{r+j}) = \delta_{ij}$. For $u \in G_W$ write $v(x) = u(x) - x$ and note that v (as a transformation of V) is identically 0 on W and for any $x \in W, y \in P$ we find that

$$\psi(x, v(y)) = \psi(u(x), u(y)) - \psi(x, y) = 0.$$

If both $x, y \in P$ we have

$$\begin{aligned} \psi(x, v(y)) + \psi(v(x), y) &= \psi(x, u(y)) - \psi(x, y) + \psi(u(x), y) - \psi(x, y) \\ &= \psi(x, u(y)) - \psi(u(x), u(y)) + \psi(u(x), y) - \psi(u(x), u(y)) \\ &= \psi(x - u(x), 0) + \psi(0, y - u(y)) \\ &= 0 \end{aligned}$$

and therefore

$$\psi(x, v(y)) = -\overline{\psi(y, v(x))}.$$

Hence expressing u within the given basis yields

$$u = \left(\begin{array}{c|c} I_r & S \\ \hline 0 & I_r \end{array} \right)$$

with a matrix S satisfying ${}^tS = -\overline{S}$, i.e. S is an antihermitian matrix. Hence

$$G_W \cong (\{S \text{ antihermitian of order } r\}, +)$$

where the right side is of dimension r^2. Note that in this case $r = d$.

The case where $2r < n$ works similarly. Again by Proposition 2.8 we find $P \subset V$ totally isotropic of dimension r and $P \cap W = \{0\}$. Then $W \oplus P$ is non-isotropic and similar to the case before $u \in G_W$ can be written as

$$u = \left(\begin{array}{cc|c} I_r & S & \\ 0 & I_r & \\ \hline & & \mathrm{GU}_{n-2r} \end{array}\right).$$

Now since $r = n - d$ we easily find that

$$\dim G_W = r^2 + (n-2r)^2.$$

Coparing this expression to d^2 yields

$$\begin{aligned} r^2 + (n-2r)^2 \leq d^2 &\Leftrightarrow (n-d)^2 + (n-2r)^2 \leq d^2 \\ &\Leftrightarrow n^2 - 2nd + d^2 + n^2 - 4nr + 4r^2 \leq d^2 \\ &\Leftrightarrow 2n^2 - 2(nd + 2nr) + 4r^2 \leq 0 \\ &\Leftrightarrow n^2 - n(d+2r) + 2r^2 \leq 0 \\ &\Leftrightarrow n^2 - n(n+r) + 2r^2 \leq 0 \\ &\Leftrightarrow 2r^2 \leq nr \\ &\Leftrightarrow 2r \leq n \end{aligned}$$

which is true. □

We can formulate this result in view of our situation, where $V = V_\ell(A)$ and $\psi = \phi^\circ_{\ell^\infty}$.

Theorem 2.12. *Let A be a simple abelian variety fully of type IV and dimension g. We consider the hermitian space $(V_\ell, \phi^\circ_{\ell^\infty})$ with the hermitian form $\phi^\circ_{\ell^\infty} : V_\ell \times V_\ell \to E_\ell$. Let W be a $\mathbb{Q}_\ell$-subspace of $V_\ell(A)$ of codimension d and consider the stabilizer*

$$G_W = \{g \in \mathrm{U}(V_\ell(A)) \mid g_{|W} = \mathrm{id}_W\} \subset \mathrm{GL}_{2g}(\mathbb{Q}_\ell).$$

Then

$$\dim G_W \leq d^2.$$

2.3 Calculating the index

The rest of the section is devoted to determining the index $(G(\mathbb{Z}_\ell) : G(H))$. Using results of Serre, Oesterle and Robba, we can estimate this index directly, given that all of the stabilizers are smooth. This is fulfilled at least for all primes ℓ that are sufficiently large, due to a resultsof Lombardo.

Lemma 2.13. [41, Lem 2.13] *There exists a prime number $\ell_0 = \ell_0(A, K)$, such that for all primes $\ell \geq \ell_0$ and all submodules $\widehat{H} \subset T_\ell(A)$ the Zariski-closure of the stabilizer $G_{\widehat{H}}$ is smooth over $\mathbb{Z}_\ell$.*

Additionally, we need the following lemma due to Hindry and Ratazzi:

Lemma 2.14. [32, Lemme 2.1] *Let $G/\mathbb{Z}_\ell$ be an algebraic subgroup of GL_n of dimension d, such that the reduction over $\mathbb{F}_\ell$ is smooth. Then for all $m \geq 1$ we have*

$$|G(\mathbb{Z}/\ell^m\mathbb{Z})| = \ell^{d(m-1)}|G(\mathbb{Z}/\ell\mathbb{Z})|.$$

For the abelian variety A of dimension g recall the set $\mathcal{S}$ defined in 1.9. Let $\ell \notin \mathcal{S}$ be a prime number and let $H \subset A[\ell^\infty]$ be a finite subgroup. We can choose a positive integer n such that $H \subset A[\ell^n]$ and a suitable basis of $T_\ell(A)$, such that we can write

$$H = \prod_{i=1}^{t} (\mathbb{Z}/\ell^{m_{t-(i-1)}}\mathbb{Z})^{\alpha_i},$$

where the m_i form a strictly increasing sequence. Denote $H_i = (\mathbb{Z}/\ell^{m_i}\mathbb{Z})^{\alpha_{t-(i-1)}}$, then there is a non-canonical choice of a submodule $\widehat{H}_i \subset T_\ell(A)$ that lifts H_i and hence $\mathrm{rk}_{\mathbb{Z}/\ell^{m_i}\mathbb{Z}} H_i = \mathrm{rk}_{\mathbb{Z}_\ell} \widehat{H}_i$.

In the simple case that $H = (\mathbb{Z}/\ell^{m_1})^{\alpha_1} \subset A[\ell^{m_1}]$ we can choose a $(\mathbb{Z}/\ell^{m_1}\mathbb{Z})$-basis $\{e_1, ..., e_r\}$ of H and a $\mathbb{Z}_\ell$-basis $\{\widehat{e}_1, ..., \widehat{e}_r\}$ of $\widehat{H}$, such that $\widehat{e}_i \equiv e_i \mod \ell^{m_1}$ for all $1 \leq i \leq r$. Considering [32, Lemme 3.7] yields the following:

Lemma 2.15. *Let $H = (\mathbb{Z}/\ell^{m_1}\mathbb{Z})^{\alpha_1} \subset A[\ell^{m_1}]$ be a totally isotropic subgroup and $\pi_{m_1} : T_\ell(A) \to A[\ell^{m_1}]$ the canonical projection. Then there exists a totally isotropic submodule $H_\infty \subset T_\ell(A)$, such that $\pi_{m_1}(H_\infty) = H$.*

Note that if H happens to be an $\mathrm{End}(A)$-module, we can choose $\widehat{H}$ to be an $\mathrm{End}(A)$-module as well.

Now for a general $H \subset A[\ell^n]$ with a decomposition as described above $H = \prod_{i=1}^t H_i$, we can choose submodules $\widehat{H}_i \subset T_\ell(A)$, such that for all $1 \leq i \leq t$ we have $\pi_{m_{t-(i-1)}}(\widehat{H}_i) = H_i$ under the canonical projections. Let $W_1 = \widehat{H}_1 + \cdots + \widehat{H}_t$ and

$$W_i = \widehat{H}_1 + \cdots + \widehat{H}_{t-(i-1)},$$

such that $\pi_{m_1}(W_1) = H[\ell^{m_1}] \simeq (\mathbb{Z}/\ell^{m_1}\mathbb{Z})^{\alpha_1+\cdots+\alpha_t}$ and $\pi_{m_t}(W_t) = (\mathbb{Z}/\ell^{m_1}\mathbb{Z})^{\alpha_t}$. Then we have a filtration $W_t \subset \cdots \subset W_1$ of submodules of $T_\ell(A)$. Given a group action by a group G on $T_\ell(A)$ we denote the stabilizer of W_i as G_{W_i}. Now we can describe the stabilizer $G(H)$ of H as follows:

$$G(H) = \{M \in G(\mathbb{Z}_\ell) \mid M \in G_{W_i} \mod \ell^{m_i}, 1 \leq i \leq t\}.$$

An alternative construction to consider are the subgroups

$$V_i := \prod_{k=1}^{t-(i-1)} (\mathbb{Z}/\ell^{t-(k-1)}\mathbb{Z})^{\alpha_k},$$

such that $V_t \subset \cdots \subset V_1 = H$ and $V_t = (\mathbb{Z}/\ell^{m_t}\mathbb{Z})^{\alpha_1}$. Let $\widehat{V}_i$ be an associated submodule of $T_\ell(A)$ and $G_{\widehat{V}_i}$ its stabilizer (note that this association is not canonic). With these submodules we can describe the stabilizer $G(H)$ as follows:

$$G(H) = \{M \in G(\mathbb{Z}_\ell) \mid M \in G_{\widehat{V}_i} \mod \ell^{m_i}, 1 \leq i \leq t\}.$$

We denote $g_i := \dim G_{\widehat{V}_i}$, $d_i := \operatorname{codim} G_{\widehat{V}_i}$ and depending on the $0 < m_1 < \cdots < m_t$ and all $1 \leq t \leq 2g$

$$H(m_1, ..., m_t) := \{M \in G(\mathbb{Z}_\ell) \mid M \in G_{\widehat{V}_i} \mod \ell^{m_i}, 1 \leq i \leq t\}.$$

The index of $H(m_1, ..., m_t)$ has been calculated in [17] using the following results:

Lemma 2.16. [17, Lemma 2.3] *Let G be an algebraic subgroup over $\mathbb{Z}$ of GL_{2g}, t a non-zero integer and $G_{\widehat{V}_1} \subset \cdots \subset G_{\widehat{V}_t}$ algebraic subgroups over $\mathbb{Z}_\ell$ of $G_{\mathbb{Z}_\ell}$ defined as above of codimension d_i. Then*

$$(G(\mathbb{Z}_\ell) : G(H)) \gg\ll_A \ell^{\sum_{i=1}^{t} d_i(m_i - m_{i-1})}$$

for all primes ℓ.

Let H_λ be a finite subgroup of $A[\lambda^\infty]$ for prime λ in $\mathcal{O}_E$ dividing ℓ. The following lemma is a reformulation of Lemma 2.16 in this situation:

Lemma 2.17. *Let G be an algebraic subgroup over $\mathbb{Z}$ of GL_{2g}. For each $\lambda|\ell$ let t_λ be a non-zero integer, $G_{\widehat{V}_1} \subset \cdots \subset G_{\widehat{V}_{t_\lambda}}$ algebraic subgroups over $\mathbb{Z}_\ell$ of $G_{\mathbb{Z}_\ell}$ of codimension d_i and $1 \leq m_{\lambda,t_\lambda} < ... < m_{\lambda,1}$ a sequence of decreasing integers. Furthermore the stabilizer $G(H_\lambda)$ is defined for each $\lambda|\ell$ by*

$$G(H_\lambda) = \{M \in G(\mathcal{O}_\lambda) \mid M \in G_{\widehat{V}_i} \mod \lambda^{m_{\lambda,t_\lambda-(i-1)}}, 1 \leq i \leq t_\lambda\}.$$

Then

$$(G(\mathcal{O}_\lambda) : G(H_\lambda)) \gg\ll_A (\#\mathbb{F}_\lambda)^{\sum_{i=1}^{t_\lambda} d_i(m_{\lambda,i} - m_{\lambda,i-1})}.$$

3. Properties of the torsion subgroup

3.1 Main idea

Let A be a simple abelian variety of type IV, defined over a number field K, and dimension g. Recall the definition of the invariant $\gamma(A)$:

$$\gamma(A) = \inf\{x > 0 \mid \forall L/K, |A(L)_{\mathrm{tors}}| \ll [L : K]^x\}.$$

We have been working with the connected component of the ℓ-adic monodromy group G_ℓ (resp. $G_{\ell,1}$). Due to the following theorem by Serre, after replacing the number field K by a suitable extension, we can assume G_ℓ to be connected.

Theorem 3.1 (Serre). *There exists a finite extension K'/K, such that for all primes ℓ the algebraic group G_ℓ is connected.*

Furthermore, by a theorem of Faltings [26, Satz 3], we know that G_ℓ is reductive. A crucial ingredient for our proof is the following theorem about the independance of ℓ-adic representations:

Theorem 3.2 (Serre). *There exists a finite extension K'/K, such that all ℓ-adic representations are independent. In other terms, the morphism*

$$\mathrm{Gal}(K'(A(K')_{\mathrm{tors}})/K) \to \prod_{\ell \text{ prime}} \mathrm{Gal}(K'(A[\ell^\infty])/K')$$

is a bijection.

After replacing K by an extension K' such that G_ℓ is connected and all ℓ-adic representations are independent we can make use of the following result:

Corollary 3.3. *Let $H \subset A(\overline{K})_{\text{tors}}$ such that $H = \prod_\ell H_\ell$ with $H_\ell \subset A[\ell^\infty]$. Then*

$$[K(H) : K] = \prod_\ell [K(H_\ell) : K].$$

This is an immediate consequence of Theorem 3.2 after replacing K by K' and abuse of notation. Following Corollary 3.3 we can reformulate the invariant $\gamma(A)$ as follows:

$$\gamma(A) = \inf\{x > 0 \mid \forall H \subset A[\ell^\infty], |H| \ll [K(H) : K]^x\}.$$

This yields the equivalence

$$|H| \ll [K(H) : K]^{\gamma(A)} \;\Leftrightarrow\; \gamma(A) \geq \frac{\log_\ell |H|}{\log_\ell [K(H) : K]} \tag{3.1}$$

for every finite subgroup $H \subset A[\ell^\infty]$ and we will therefore determine $\gamma(A)$ as follows:

Step 1: Calculate the order of H.

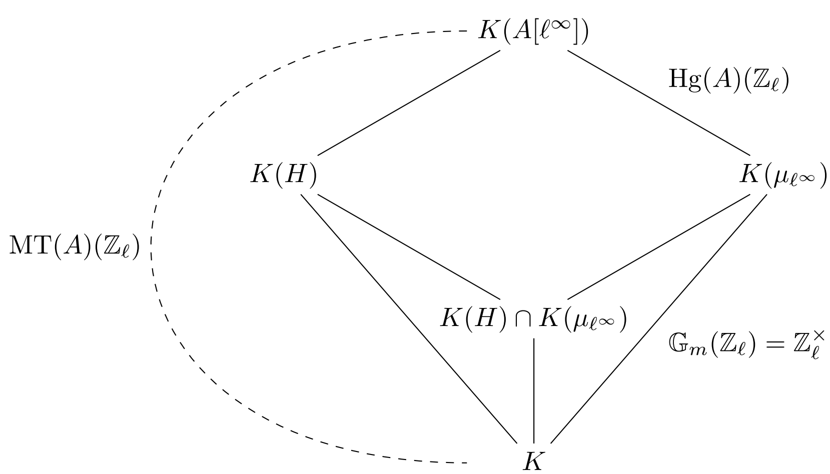
Step 2: Calculate the degree of the extension $K(H)/K$.

Step 3: Use combinatorial methods to determine $\gamma(A)$ with these two values.

Calculating the order of a subgroup $H \subset A[\ell^\infty]$ is straightforward and mainly relies on the structure of the Tate module $T_\ell(A)$. To determine the degree of the extension $K(H)$ over K we consider the Galois extensions $K(A[\ell^\infty])$ and $K(\mu_{\ell^\infty})$ over K. Recall that assuming A satisfies the Mumford-Tate conjecture, we have an isomorphism

$$G_\ell \cong \mathrm{MT}(A)_{\mathbb{Q}_\ell}.$$

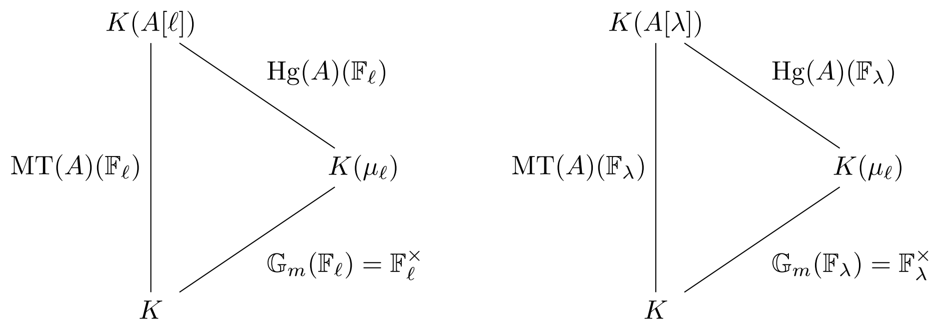
Moreover from Theorem 3.2 we know that we can restrict our attention to any prime ℓ, therefore we can compare the Galois groups of the following diagram with the $\mathbb{Z}_\ell$ points of the Mumford-Tate group, the Hodge group and the torus of homotheties $\mathbb{G}_m$.



Therefore we have the following identifications:

$$\mathrm{Gal}(K(A[\ell^\infty]/K)) \asymp \mathrm{MT}(A)(\mathbb{Z}_\ell),$$
$$\mathrm{Gal}(K(A[\ell^\infty]/K(\mu_{\ell^\infty}))) \asymp \mathrm{Hg}(A)(\mathbb{Z}_\ell),$$
$$\mathrm{Gal}(K(\mu_{\ell^\infty})/K) \asymp \mathbb{G}_m(\mathbb{Z}_\ell).$$

Note that we have analogue diagrams for the cases $H \subset A[\ell]$ and $H_\lambda \subset A[\lambda]$ for $\lambda|\ell$ stemming from the triangles



Note that for the case of $H_\lambda \subset A[\lambda]$ a precise treatment has been done for simple abelian varieties of tyoes I and II (see [33, Section 3]). Combining their methods with the decomposition of $T_\ell(A)$ in Theorem 3.5 yields similar results for type IV. Therefore we have the following identifications:

$$\mathrm{Gal}(K(A[\lambda]/K)) \asymp \mathrm{MT}(A)(\mathbb{F}_\lambda),$$
$$\mathrm{Gal}(K(A[\lambda]/K(\mu_\ell))) \asymp \mathrm{Hg}(A)(\mathbb{F}_\lambda),$$
$$\mathrm{Gal}(K(\mu_\ell)/K) \asymp \mathbb{G}_m(\mathbb{F}_\lambda).$$

In order to calculate $[K(H) : K]$ we consider the groups

$$G_0(H) := \{\sigma \in \mathrm{MT}(A)(\mathbb{Z}_\ell) \mid \sigma_{|H} = \mathrm{id}_H\} \asymp \mathrm{Gal}(K(A[\ell^\infty]/K(H)))$$

and

$$G(H) := \{\sigma \in \mathrm{Hg}(A)(\mathbb{Z}_\ell) \mid \sigma_{|H} = \mathrm{id}_H\} \asymp G_0(H) \cap \mathrm{Hg}(A)(\mathbb{Z}_\ell).$$

Let $\delta(H) := [K(H) \cap K(\mu_{\ell^\infty}) : K]$. Then the following lemma establishes how to determine $[K(H) : K]$.

Lemma 3.4. *Let $H \subset A[\ell^\infty]$. Then we have up to a finite index*

$$\delta(H) = (\mathbb{Z}_\ell^\times : \mathrm{mult}(G_0(H))(\mathbb{Z}_\ell))$$

and additionally

$$[K(H) : K] = (\mathrm{Hg}(A)(\mathbb{Z}_\ell) : G(H)) \cdot \delta(H).$$

We refer the reader to [33, Proposition 5.5] for a treatment of a similar case. There are analogous results for the cases $H \subset A[\ell]$ and $H_\lambda \subset A[\lambda]$ respectively.

As we have established in Lemma 2.16 for $H \subset A[\ell^\infty]$ the index $(\mathrm{Hg}(A)(\mathcal{O}_\lambda) : G(H_\lambda))$ can be calculated for every $\lambda|\ell$ with the codimension of $G(H_\lambda)$. Establishing said codimension is therefore essential to this section. Lastly we will discuss the value $\delta(H)$.

3.2 Structure of the Tate module

Recall the Tate module $T_\ell(A) = \varprojlim A[\ell^n]$ and the set $\mathcal{P}$ of primes (Definition 1.8), such that we have decompositions

$$T_\ell(A) = \prod_{\lambda|\ell} \mathcal{T}_\lambda, \tag{3.2}$$

$$V_\ell(A) = \prod_{\lambda|\ell} \mathcal{V}_\lambda.$$

Here for all $\lambda|\ell$ we have $\mathcal{T}_\lambda = T_\ell(A) \otimes_{\mathcal{O}_{E_\ell}} \mathcal{O}_\lambda$ and an E_ℓ-vector space $\mathcal{V}_\lambda = V_\ell(A) \otimes_{E_\ell} E_\lambda$. Furthermore we have unique non-degenerate, Galois equivariant bilinear forms over $\mathcal{O}_{E_\ell}$ and over E_ℓ respectively

$$\phi_{\ell^\infty} : T_\ell(A) \times T_\ell(A) \to \mathcal{O}_{E_\ell},$$

$$\phi^\circ_{\ell^\infty} : V_\ell(A) \times V_\ell(A) \to E_\ell.$$

For all primes in $\mathcal{P}$, or equivalently all $\ell \notin \mathcal{S}$, we have the following result of Banaszak and Kaim-Garnek, that will be helpful with the calculation of the order of a subgroup $H \subset A[\ell^\infty]$.

Theorem 3.5. [2, Theorem 1.1] *Let A be a simple abelian variety of type IV. Let ℓ be a prime outside of the finite set $\mathcal{S}$. Then the $\mathcal{O}_\lambda[G_K]$-module $\mathcal{T}_\lambda(A)$ has a decomposition*

$$\mathcal{T}_\lambda(A) \cong T_\lambda(A)^d,$$

where $T_\lambda(A)$ is a free $\mathcal{O}_\lambda$ module of rank $2g/ed$ with a non-degenerate, hermitian, G_K-equivariant form

$$\phi_{\lambda^\infty} : T_\lambda(A) \times T_\lambda(A) \to \mathcal{O}_\lambda,$$

such that $V_\lambda(A) := T_\lambda(A) \otimes_{\mathcal{O}_\lambda} E_\lambda$ is an absolutely irreducible G_K-module with a non-degenate, hermitian, G_K-equivariant form $\phi^\circ_{\lambda^\infty} := \phi_{\lambda^\infty} \otimes_{\mathcal{O}_\lambda} E_\lambda$ and respectively $\overline{T_\lambda}(A) := T_\lambda \otimes_{\mathcal{O}_\lambda} (\mathcal{O}_\lambda/\lambda)$ is an absolutely irreducible G_K-module with a non-degenerate, hermitian, G_K-equivariant form $\overline{\phi}_{\lambda^\infty} := \phi_{\lambda^\infty} \otimes_{\mathcal{O}_\lambda} (\mathcal{O}_\lambda/\lambda)$.

3.3 Order of the subgroup H

From now on assume all primes ℓ to be in the set $\mathcal{P}$ defined in Definition 1.9. For an explicit treatment of the remaining set of primes $\mathcal{S}$ see Section 6.

Lemma 3.6. *Let $H \subset A[\ell^\infty]$, then*

$$\log_\ell |H| = \sum_{\lambda|\ell} \sum_{i=1}^{t_\lambda} df(\lambda)\alpha_{\lambda,i} m_{\lambda,i},$$

where $t_\lambda \leq dh, \alpha_{\lambda,i}$ and $m_{\lambda,i}$, $1 \leq i \leq t_\lambda$ only depend on H and λ respectively. Here $f(\lambda)$ is the residual degree of λ over ℓ and h is the relative dimension of A as in Definition 1.7.

Proof. First assume $H \subseteq A[\ell]$. By definition $A[\ell] = T_\ell(A)/\ell T_\ell(A)$, hence with the decomposition Eq. (3.2) and Theorem 3.5 we can write

$$A[\ell] \cong \prod_{\lambda|\ell} (\mathcal{T}_\lambda(A)/\ell\mathcal{T}_\lambda(A)) \otimes_{\mathcal{O}_\lambda/\ell\mathcal{O}_\lambda} \mathcal{O}_{E_\ell}/\ell\mathcal{O}_{E_\ell} \cong \prod_{\lambda|\ell} (T_\lambda(A)/\ell T_\lambda(A))^d$$

and denote $T_\lambda[\ell] = T_\lambda(A)/\ell T_\lambda(A)$. Therefore we can write

$$H = \prod_{\lambda|\ell} \mathcal{H}_\lambda,$$

where

$$\mathcal{H}_\lambda = \underbrace{H_\lambda \oplus \cdots \oplus H_\lambda}_{d\text{-times}}$$

and $H_\lambda \subset T_\lambda[\ell]$. Remember that $T_\ell(A) \cong \mathbb{Z}_\ell^{2g}$ as a $\mathbb{Z}_\ell$-module and with the relative dimension $h = g/d^2 e_0$ we must have $T_\lambda(A) \cong \mathbb{Z}_\ell^{dh}$. Let $r_\lambda = \operatorname{rk}_{\mathbb{F}_\lambda} H_\lambda$, then by what we just established $r_\lambda \in [0, dh]$ and further

$$\operatorname{rk}_{\mathbb{F}_\ell} H = d \sum_{\lambda|\ell} f(\lambda) r_\lambda.$$

Then for the order of H we must have

$$|H| = \ell^{d \sum_{\lambda|\ell} f(\lambda) r_\lambda},$$

where $t_\lambda = 1 = m_{\lambda,1}$ and $\alpha_{\lambda,1} = r_\lambda$.

Now let $H \subset A[\ell^n]$ for some $n > 1$. Note that $A[\ell^n] = T_\ell(A)/\ell^n T_\ell(A)$ and denote $T_\lambda[\ell^n] = T_\lambda(A)/\ell^n T_\lambda(A)$. In similar fashion as before we find that $H = \prod_{\lambda|\ell} \mathcal{H}_\lambda$ with $\mathcal{H}_\lambda = H_\lambda \oplus \cdots \oplus H_\lambda$, consisting of d copies of $H_\lambda \subset T_\lambda[\ell^n]$. For each $\lambda|\ell$ choose a basis of T_λ, such that

$$H_\lambda = \prod_{i=1}^{t_\lambda} (\mathbb{Z}/\ell^{m_{\lambda,i}}\mathbb{Z})^{\alpha_{\lambda,i} f(\lambda)}, \tag{3.3}$$

where $t_\lambda \in [1, dh]$ and the $m_{\lambda,i}$ form a strictly decreasing sequence, $1 \leq m_{\lambda,t_\lambda} < \cdots < m_{\lambda,1}$. We have

$$\operatorname{rk}_{\mathbb{F}_\lambda} H_\lambda = \sum_{i=1}^{t_\lambda} \alpha_{\lambda,i}$$

and

$$\operatorname{rk}_{\mathbb{F}_\ell} H_\lambda = f(\lambda) \cdot \operatorname{rk}_{\mathbb{F}_\lambda} H_\lambda.$$

Therefore, since $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda$ and there are d-copies of H_λ for every $\lambda|\ell$, we find

$$\operatorname{rk}_{\mathbb{F}_\ell} H = d \cdot \operatorname{rk}_{\mathbb{F}_\ell} H_\lambda$$

and hence

$$|H| = \ell^{d \sum_{\lambda|\ell} f(\lambda) \sum_{i=1}^{t_\lambda} \alpha_{\lambda,i} m_{\lambda,i}}.$$

□

3.4 Codimension and stabilizers

Recall that by Lemma 3.4 the calculation of the degree $[K(H) : K]$ relies essentially on the index $(\mathrm{Hg}(A)(\mathbb{Z}_\ell) : G(H))$ and by Lemma 2.16 said index relies on the codimension of the stabilizer $G(H)$. We will therefore calculate these codimensions explicitly, first for the case $H \subset A[\ell]$ and then in general for $H \subset A[\ell^\infty]$.

- Consider $H \subset A[\ell]$ and for $\lambda|\ell$ the stabilizer $G(H_\lambda)$ can be expressed by the corresponding subspace of $T_\lambda(A)$ of dimension r_λ. Since $T_\lambda(A) \cong \mathbb{Z}_\ell^{dh}$ and Theorem 2.12 we have:

  Proposition 3.7. *Let $H \subset A[\ell]$, $\lambda|\ell$, H_λ the subset of H corresponding to the decomposition $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda$ and $G(H_\lambda)$ the stabilizer of H_λ in $\mathrm{Hg}(A)$. Then*

$$\operatorname{codim}(G(H_\lambda)) = (dh)^2 - \dim G(H_\lambda) \geq (dh)^2 - (dh - r_\lambda)^2 = r_\lambda(2dh - r_\lambda).$$

- Let $H \subset A[\ell^\infty]$. We use the decomposition of H_λ presented in (3.3) and consider to each $1 \leq i \leq t_\lambda$ the natural projection map

$$\pi_{m_i} : T_\lambda(A) \to T_\lambda(A)/\ell^{m_i} T_\lambda(A).$$

  Hence we obtain a filtration of submodules $W_{t_\lambda} \subset \cdots \subset W_1$ of $T_\lambda(A)$ such that

$$\pi_{m_i}(W_i) = H_\lambda[\ell^{m_i}] = (\mathbb{Z}/\ell^{m_{\lambda,i}})^{\alpha_{\lambda,i} f(\lambda)}.$$

  Note that $\dim W_i = \alpha_{\lambda,i} f(\lambda)$. We give another decomposition of T_λ:

$$T_\lambda = \left(\bigoplus_{i=1}^{t_\lambda} W_i \right) \oplus W',$$

  where W' is a suitable complementary subspace. Define subgroups

$$V_i := \prod_{k=1}^{t_\lambda - (i-1)} (\mathbb{Z}/\ell^{m_{\lambda,k}} \mathbb{Z})^{\alpha_{\lambda,k} f(\lambda)}$$

  and note that $V_{t_\lambda} \subset \cdots \subset V_1$ and

$$V_i \cong \prod_{k=1}^{t_\lambda - (i-1)} (\mathcal{O}_\lambda/\ell^{m_{\lambda,k}} \mathcal{O}_\lambda)^{\alpha_{\lambda,k} f(\lambda)}.$$

  Let $\widehat{V}_i$ denote the corresponding submodules of T_λ and $r_i := \mathrm{rk}_{\mathcal{O}_\lambda} \widehat{V}_i$ and let $G_{\widehat{V}_i}$ be the stabilizer of $\widehat{V}_i$, hence we can write

$$G(H_\lambda) = \{M \in \mathrm{MT}(A)(\mathcal{O}_\lambda) \mid M \in G_{\widehat{V}_i} \mod \ell^{m_{\lambda, t_\lambda - (i-1)}}, 1 \leq i \leq t_\lambda\}. \tag{3.4}$$

  Using Theorem 2.12 again we find that:

  Proposition 3.8. *Let $H \subset A[\ell^\infty]$, $\lambda|\ell$, H_λ the subset of H corresponding to the decomposition $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda$ and $G(H_\lambda)$ and $G(H_\lambda)$ defined as in (3.4). Then*

$$d_i := \operatorname{codim} G_{\widehat{V}_i} = (dh)^2 - \dim G_{\widehat{V}_i} \geq (dh)^2 - (dh - r_i)^2 = r_i(2dh - r_i).$$

3.5 Property μ

In order to calculate $[K(H) : K]$ we not only need the codimension of the stabilizer, but also the value $\delta(H) = [K(H) \cap K(\mu_{\ell^\infty}) : K]$. In order to understand $\delta(H)$ we introduce the property μ as described by Hindry and Ratazzi.

Definition 3.9 ([31](Définition 6.3)). Let A be an abelian variety defined over a number field K. We say that A satisfies the property μ, if for all primes ℓ and all finite subgroups $H \subset A[\ell^\infty]$ there exists an integer $m = m(H)$, such that up to a finite index (independent of ℓ) we have

$$K(H) \cap K(\mu_{\ell^\infty}) \asymp K(\mu_{\ell^m}).$$

As before we can make a similar definition for $H \subset A[\ell]$ and $H_\lambda \subset A[\lambda]$.

Similarly as in the proofs of [33](Proposition 5.5 & Proposition 7.3), we can establish that property μ holds (see Proposition 3.10 below) for a simple abelian varietiy A defined over a number field K that is fully of type IV and fully of Lefschetz type. In that case we have

$$T_\ell(A) = \prod_{\lambda|\ell} \underbrace{T_\lambda \oplus \cdots \oplus T_\lambda}_{d\text{-times}}.$$

Write $A[\lambda] = T_\lambda/\ell T_\lambda$, $A[\lambda^n] = T_\lambda/\ell^n T_\lambda$ and $A[\lambda^\infty] = \bigcup_n A[\lambda^n]$.

Proposition 3.10. *Under the above notations, we have the following results:*

1. *for every $n \geq 1$ and for all $\lambda|\ell$ and all $H_\lambda \subset A[\lambda^n]$, there exists an integer m_λ such that*
$$K(H_\lambda) \cap K(\mu_{\ell^\infty}) \asymp K(\mu_{\ell^{m_\lambda}});$$
2. *we have*
$$\mathrm{Gal}(K(A[\ell])/K(\mu_{\ell^\infty})) \simeq \prod_{\lambda|\ell} \mathrm{Gal}(K(A[\lambda])/K(\mu_{\ell^\infty}));$$
and
$$\mathrm{Gal}(K(A[\ell^\infty])/K(\mu_{\ell^\infty})) \simeq \prod_{\lambda|\ell} \mathrm{Gal}(K(A[\lambda^\infty])/K(\mu_{\ell^\infty}));$$
3. *with $m := \max_\lambda m_\lambda$, for all finite subgroups $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda \subset A[\ell]$ or $H \subset A[\ell^\infty]$ we have $K(H) \cap K(\mu_{\ell^\infty}) \asymp K(\mu_{\ell^m})$ and*
$$[K(H) : K(\mu_{\ell^m})] \gg\ll \prod_{\lambda|\ell} [K(H_\lambda) : K(\mu_{\ell^{m_\lambda}})].$$

Recall that we defined the multiplier

$$\mathrm{mult}\colon \mathrm{GU}_{2g} \to \mathbb{G}_m$$

and we have an embedding $\mathrm{MT}(A) \hookrightarrow \mathrm{GU}_{2g}$ given a fixed polarisation. Therefore, we can restrict mult to a map $\mathrm{MT}(A) \to \mathbb{G}_m$ and identify the Hodge group $\mathrm{Hg}(A)$ as the connected component of the kernel of this map. Given that A is simple and fully of type IV, dimension g and fully of Lefschetz type, i.e. $\mathrm{MT}(A) = \mathcal{L}(A)$, we have that

$$\mathrm{MT}(A)(\mathbb{Z}_\ell) = \{(\sigma_\lambda)_\lambda \in \prod_{\lambda|\ell} \mathrm{GU}_{\mathcal{O}_\lambda} \mid \mathrm{mult}(\sigma_\lambda) \in \mathbb{Z}_\ell^\times\}.$$

Proposition 3.11. *Let $\widehat{H}$ be a $\mathbb{Z}_\ell$-submodule of $T_\ell(A)$. If $\widehat{H}$ is a maximal isotropic submodule of dimension at most g, then the multipliier restricted to the stabilizer $G_{\widehat{H}}$ of $\widehat{H}$ in $\mathrm{MT}(A)(\mathbb{Z}_\ell)$ is surjective.*

Note that the preceding proposition implies that for every submodule $\widehat{H}$ that is contained in a maximally isotropic submodule $\mathrm{mult}(G_{\widehat{H}})(\mathbb{Z}_\ell) = \mathbb{G}_m(\mathbb{Z}_\ell) = \mathbb{Z}_\ell^\times$ and hence

$$\delta(\widehat{H}) = (\mathbb{Z}_\ell^\times : \mathrm{mult}(G_{\widehat{H}})(\mathbb{Z}_\ell)) = 1.$$

Note that in particular for $H_\lambda \subset A[\lambda]$ we have necessarily that

$$\delta(H_\lambda) = \begin{cases} 1 & H_\lambda \text{ is contained in a maximal isotropic submodule,} \\ \ell & \text{else.} \end{cases}$$

Let us introduce the notation $\delta := \log_\ell \delta(H_\lambda)$, i.e. $\delta(H_\lambda) = \ell^\delta$, depending on H, such that

$$\delta = \begin{cases} 0 & H_\lambda \text{ is contained in a maximal isotropic submodule,} \\ 1 & \text{else.} \end{cases}$$

Proof of Proposition 3.11. Let W be the maximal isotropic subspace associated to $\widehat{H}$ in V_ℓ and $< e_1, ..., e_h >$ a basis. Considering the stabilizer of W in $\mathrm{GL}_{2g}(\mathbb{Q}_\ell)$, we can express it as

$$G_W = \{M \in \mathrm{GU}_g \mid Me_i = e_i, i = 1, ..., h\}.$$

Specifically, $G_W \subset \mathrm{MT}(A)(\mathbb{Q}_\ell)$, hence it is sufficient to show that $\mathrm{mult}\colon G_W \to \mathbb{G}_m$ is surjective for any maximal isotropic subspace W.

First assume that $W =< e_1 >$, where e_1 is an isotropic vector. Then we can find an isotropic vector $e_2 \in V_\ell$ with $\phi^\circ_{\ell^\infty}(e_1, e_2) = 1$ and a hyperbolic plane $\Pi =< e_1, e_2 >$ with $\Pi \cap \Pi^\perp = \{0\}$ and $\Pi \oplus \Pi^\perp = V_\ell$. For any matrix $M \in \mathrm{GU}_g$ we have $\phi^\circ_{\ell^\infty}(Me_1, Me_2) = \mathrm{mult}(M)\phi^\circ_{\ell^\infty}(e_1, e_2)$, hence for any $M \in G_W$ we have $Me_1 = e_1$ and $Me_2 = \mathrm{mult}(M)e_2$. Therefore M also stabilizes Π and consequently $\Pi^\perp$. We can write M with diagonal blocks

$$M = \left(\begin{array}{cc|c} 1 & & \\ & \mathrm{mult}(M) & \\ \hline & & M_1 \end{array}\right)$$

where $M_1 \in \mathrm{GU}(\phi^\circ_{\ell^\infty \,|\Pi^\perp})$. Let

$$G_2 = \{M \in \mathrm{GU}_g \mid M =< 1, \mathrm{mult}(M) > \oplus M_1, M_1 \in \mathrm{GU}(\phi^\circ_{\ell^\infty|\Pi^\perp})\}.$$

Since the restriction of mult to $\mathrm{GU}(\phi^\circ_{\ell^\infty|\Pi^\perp})$ is surjective, the decomposition

$$\mathrm{mult}_{|\mathrm{GU}(\phi^\circ_{\ell^\infty|\Pi^\perp})}\colon\ \mathrm{GU}(\phi^\circ_{\ell^\infty|\Pi^\perp}) \twoheadrightarrow G_2 \stackrel{\mathrm{mult}_{|G_2}}{\longrightarrow} \mathbb{G}_m$$

yields that $\mathrm{mult}\colon G_W \to \mathbb{G}_m$ is surjective. For the general case let $W =< e_1, ..., e_h >$ with $h \le g$. By Corollary 2.9 we find h hyperbolic planes $\Pi_1, ..., \Pi_h$, such that

$$V_\ell = (\Pi_1 \oplus \cdots \oplus \Pi_h) \oplus (\Pi_1 \oplus \cdots \oplus \Pi_h)^\perp.$$

Write $\Pi = \Pi_1 \oplus \cdots \oplus \Pi_h$ and let $M_1 \in \mathrm{GU}(\phi^\circ_{\ell^\infty|\Pi^\perp})$. Similar to before, every matrix M in $G_W = \{M \in \mathrm{GU}_g \mid Me_i = e_i, 1 \leq i \leq h\}$ can be written in the form

$$M = \left(\begin{array}{cccccc|c} 1 & & & & & & \\ & \mathrm{mult}(M) & & & & & \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \mathrm{mult}(M) & \\ \hline & & & & & & M_1 \end{array}\right).$$

If W is not totally isotropic, we have that $\phi^\circ_{\ell^\infty}(e_1, e_2) \neq 0$. Consequently every matrix M in G_W satisfies both of the following equations:

$$\phi^\circ_{\ell^\infty}(Me_1, Me_2) = \phi^\circ_{\ell^\infty}(e_1, e_2); \quad \phi^\circ_{\ell^\infty}(Me_1, Me_2) = \mathrm{mult}(M)\phi^\circ_{\ell^\infty}(e_1, e_2).$$

Obviously it follows that $\mathrm{mult}(M) = 1$. □

4. Main result

4.1 Simple abelian varieties of type IV

Theorem 4.1. *Let A be a simple abelian variety defined over a number field K. Assume that A is fully of type IV and fully of Lefschetz type. Then*

$$\gamma(A) = \frac{2\dim(A)}{\dim(\mathrm{MT}(A))}.$$

Recall that we have established the order of the subgroup $H \subset A[\ell^\infty]$ in Lemma 3.6, the codimension of the stabilizers of H_λ for each $\lambda|\ell$ in Section 3.4 and the values $\delta(H_\lambda)$ in Proposition 3.11.

Proof of Theorem 4.1. First we assume that $H \subset A[\ell]$ to illustrate the strategy of proof. We will make use the symbol $\gamma(A)$ by abuse of notation even though we only consider subgroups of $A[\ell]$ for simplicity. By Lemma 3.4 we have that

$$[K(H) : K] = [K(H) : K(\mu_{\ell^m})][K(\mu_{\ell^m}) : K] = (\mathrm{Hg}(A)(\mathbb{F}_\ell) : G(H)) \cdot \delta(H),$$

and with Proposition 3.10 and the decomposition $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda$ we have

$$(\mathrm{Hg}(A)(\mathbb{F}_\ell) : G(H)) = \prod_{\lambda|\ell} (\mathrm{Hg}(A)(\mathbb{F}_\lambda) : G(H_\lambda)).$$

Therefore we can resort to determine $(\mathrm{Hg}(A)(\mathbb{F}_\lambda) : G(H_\lambda))$ for every $\lambda|\ell$. Using Lemma 2.17 we find that

$$(\mathrm{Hg}(A)(\mathbb{F}_\lambda) : G(H_\lambda)) \gg\ll_A (\#\mathbb{F}_\lambda)^{\mathrm{codim}\, G(H_\lambda)},$$

where by Proposition 3.7

$$\mathrm{codim}\, G(H_\lambda) = (dh)^2 - (dh - r_\lambda)^2 \geq r_\lambda(2dh - r_\lambda).$$

Moreover, $\delta(H) = \ell^{\delta}$, where $\delta = 0$ or $\delta = 1$, hence

$$[K(H):K] \gg\ll \ell^{\sum_{\lambda|\ell} f(\lambda)\,\mathrm{codim}(G(H_\lambda))+\delta}.$$

Using the equivalence (3.1) and considering the d copies of H_λ in H we obtain

$$\gamma(A) = d \cdot \max_{\lambda|\ell} \psi(\underline{r}), \tag{4.1}$$

where $\underline{r} = (r_\lambda)_{\lambda|\ell}$ and

$$\psi(\underline{r}) := \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{\delta + \sum_{\lambda|\ell} f(\lambda)\,\mathrm{codim}(G(H_\lambda))}.$$

We want to study the maximum of the function $\psi(\underline{r})$ over all possible values of r_λ and distinguish two cases:

First consider the case where $\delta = 0$. By Proposition 3.11 we know that H_λ is therefore contained in a maximally isotropic subspace and $0 \le r_\lambda \le \frac{dh}{2}$. Hence we find that

$$\psi(\underline{r}) \ge \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{\sum_{\lambda|\ell} f(\lambda) r_\lambda (2dh - r_\lambda)}$$

and a study of the extremal values[1] of the function

$$f(\underline{r}) = \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{\sum_{\lambda|\ell} f(\lambda) r_\lambda (2dh - r_\lambda)}$$

for $r_\lambda \in [0, dh/2]$ yields that the maximum is attained for $r_\lambda = dh/2$ for all $\lambda|\ell$. Since the case $r_\lambda = dh/2$ is attained when H_λ is a maximally isotropic subspace of biggest possible dimension, the equality $\mathrm{codim}\, G(H_\lambda) = r_\lambda(2dh - r_\lambda)$ holds and f and ψ have the same maximum. In particular we have

$$\max_{r_\lambda \in [0, dh/2]} \psi(\underline{r}) = \frac{2}{3dh}.$$

Secondly, if $\delta = 1$, H_λ is not contained in a maximally isotropic subspace and we study the extremal values of the function

$$g(\underline{r}) = \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{1 + \sum_{\lambda|\ell} f(\lambda) r_\lambda (2dh - r_\lambda)}$$

for $r_\lambda \in [0, dh]$. Again the function attains its maximum in the boundary case that $r_\lambda = dh$ for all $\lambda|\ell$. This means essentially, that H_λ is anisotropic and hence the equality $\mathrm{codim}\, G(H_\lambda) = r_\lambda(2dh - r_\lambda)$ holds again and we have

$$\max_{r_\lambda \in [0, dh]} \psi(\underline{r}) = \frac{edh}{1 + e(dh)^2}.$$

Finally, we compare the two maxima obtained above to find the unconditional maximum of ψ for $r_\lambda \in [0, dh]$:

$$\max_{r_\lambda} \psi(\underline{r}) = \left\{ \frac{2}{3dh}, \frac{edh}{1 + e(dh)^2} \right\} = \frac{edh}{1 + e(dh)^2}.$$

[1] For details on the study of extremal values we refer the reader to Section 5.

Therefore Eq. (4.1) gives

$$\gamma(A) = \frac{ed^2h}{1+e(dh)^2} = \frac{2g}{1+\mathrm{Res}_{E/\mathbb{Q}}\,\mathrm{SU}_{dh}} = \frac{2\dim A}{\dim \mathrm{MT}(A)}.$$

Now assume that $H \subset A[\ell^\infty]$. Using Lemma 3.4 again we have

$$[K(H):K] = [K(H):K(\mu_{\ell^m})][K(\mu_{\ell^m}):K] = (\mathrm{Hg}(A)(\mathbb{Z}_\ell):G(H))\cdot\delta(H),$$

together with Proposition 3.10 and the decomposition $H = \prod_{\lambda|\ell} H_\lambda \oplus \cdots \oplus H_\lambda$ we have

$$(\mathrm{Hg}(A)(\mathbb{Z}_\ell):G(H)) = \prod_{\lambda|\ell} \mathrm{Hg}(A))(\mathcal{O}_\lambda):G(H_\lambda).$$

Therefore we can resort to determine $(\mathrm{Hg}(A)(\mathcal{O}_\lambda):G(H_\lambda))$ for every $\lambda|\ell$. By Lemma 2.17 we know that for every ℓ (up to some constants)

$$(\mathrm{Hg}(A)(\mathcal{O}_\lambda):G(H_\lambda)) \gg\ll_A \ell^{f(\lambda)\sum_{i=1}^{t_\lambda} d_i(m_{\lambda,t_\lambda-(i-1)}-m_{\lambda,t_\lambda-(i-1)+1})}.$$

Consequently we obtain

$$(\mathrm{Hg}(A)(\mathbb{Z}_\ell):G(H)) \gg\ll_A \ell^{\sum_{\lambda|\ell} f(\lambda)\sum_{i=1}^{t_\lambda} d_i(m_{\lambda,t_\lambda-(i-1)}-m_{\lambda,t_\lambda-(i-1)+1})},$$

where we conventionally let $m_{\lambda,t_\lambda+1}=0$. For the degree of the extension in question we therefore have

$$[K(H):K] \gg\ll \ell^{\sum_{\lambda|\ell} f(\lambda)\sum_{i=1}^{t_\lambda} d_i(m_{\lambda,t_\lambda-(i-1)}-m_{\lambda,t_\lambda-(i-1)+1})}\cdot\delta(H).$$

To estimate $\delta(H)$ we introduce two integers in the following way:

1. Let $m_H \geq 1$ be the maximal integer, such that elements $P, Q \in H$ of order ℓ^{m_H} exist and additionally $e_\ell(\ell^{m_H-1}P, \ell^{m_H-1}Q) \in \mu_\ell$. If no such elements exist, we let $m_H = 0$.
2. Let $h_H \in [1, t_\lambda]$ be minimal such that $m^{h_H} \leq m_H$. If $m_H = 0$, we let $h_H = t_\lambda + 1$, where m is given as in Proposition 3.10.

Since $\widehat{V}_{t_\lambda} \subset \cdots \subset \widehat{V}_1$ we have $G_{\widehat{V}_1} \subset \cdots \subset G_{\widehat{V}_{t_\lambda}}$. To each stabilizer we associate $\delta_{\lambda,i}$, which is either 0 or 1 depending whether $\widehat{V}_i$ is contained in a maximally isotropic subspace or not. Obviously, if $\delta_{\lambda,i}=1$ for some i, we necessarily have $\delta_{j,\lambda}=1$ for all $1\leq j\leq i$. By construction we have $\delta(H) \gg \ell^{m^{h_H}}$ and hence

$$\delta_{\lambda,1} = \cdots = \delta_{\lambda,t_\lambda+1-h_H} = 1; \delta_{\lambda,t_\lambda+1-(h_H-1)} = \cdots = \delta_{\lambda,t_\lambda} = 0.$$

In particular we have

$$m^{h_H} = \sum_{i=1}^{t_\lambda} \delta_{\lambda,i}(m_{\lambda,t_\lambda-(i-1)} - m_{\lambda,t_\lambda-(i-1)+1})$$

and with Lemma 3.4 up to a finite index

$$\begin{aligned}[K(H):K] &\gg \ell^{m^{h_H}+\sum_{\lambda|\ell} f(\lambda)\sum_{i=1}^{t_\lambda} d_i(m_{\lambda,t_\lambda-(i-1)}-m_{\lambda,t_\lambda-(i-1)+1})} \\ &= \ell^{\sum_{\lambda|\ell} f(\lambda)\sum_{i=1}^{t_\lambda} (d_i+\delta_{\lambda,i})(m_{\lambda,t_\lambda-(i-1)}-m_{\lambda,t_\lambda-(i-1)+1})}.\end{aligned} \tag{4.2}$$

As before, we want to employ a combinatoric argument, using the equivalence (3.1):

$$|H| \ll [K(H):K]^{\gamma(A)} \Leftrightarrow \gamma(A) \gg \max_{m_{\lambda,i}} \left\{ \frac{\sum_{\lambda|\ell} \sum_{i=1}^{t_\lambda} f(\lambda) d\alpha_{\lambda,i} m_{\lambda,i}}{\sum_{\lambda|\ell} f(\lambda) \sum_{i=1}^{t_\lambda} (d_i + \delta_{\lambda,i})(m_{\lambda,t_\lambda-(i-1)} - m_{\lambda,t_\lambda-(i-1)+1})} \right\}.$$

Reformulating the denominator in terms of $m_{\lambda,i}$ yields

$$\gamma(A) \gg \max_{m_{\lambda,i}} \left\{ \frac{\sum_{\lambda|\ell} \sum_{i=1}^{t_\lambda} f(\lambda) d\alpha_{\lambda,i} m_{\lambda,i}}{\sum_{\lambda|\ell} f(\lambda) \sum_{i=1}^{t_\lambda} m_{\lambda,i}(d_{t_\lambda+1-i} + \delta_{\lambda,t_\lambda+1-i} - d_{t_\lambda+2-i} - \delta_{\lambda,t_\lambda+2-i})} \right\}.$$

Let M denote the maximum given above, then following [32, Lemme 2.8] we have

$$M = \max_{1 \le k \le t_\lambda} \left\{ \frac{\sum_{\lambda|\ell} \sum_{i=1}^{t_k} f(\lambda) d\alpha_{\lambda,i}}{\sum_{\lambda|\ell} \sum_{i=1}^{k} f(\lambda)(d_{t_\lambda+1-i} + \delta_{\lambda,t_\lambda+1-i} - d_{t_\lambda+2-i} - \delta_{\lambda,t_\lambda+2-i})} \right\},$$

with the convention $d_{t_\lambda+1} = 0 = \delta_{\lambda,t_\lambda+1}$. Evaluating the telescope sum in the denominator and noting that $r_{t_\lambda+1-k} = \sum_{i=1}^{k} \alpha_{\lambda,i}$, we find that

$$M = \max_{1 \le k \le t_\lambda} \left\{ \frac{\sum_{\lambda|\ell} f(\lambda) d r_{t_\lambda+1-i}}{\sum_{\lambda|\ell} f(\lambda)(d_{t_\lambda+1-k} + \delta_{\lambda,t_\lambda+1-k})} \right\}.$$

We can assume without loss of generality, that $\delta(H) = \delta(H_{\lambda'})$ for some fixed place λ' over ℓ. Necessarily for all $\lambda \neq \lambda'$ with $\lambda|\ell$ we have $\delta_{t_\lambda+1-k} = 0$ for $1 \le k \le t_\lambda$. Remember that $\delta_{\lambda',t_{\lambda'}+1-k}$ is either 0 or 1 depending, whether $k < h_H$ or $k \ge h_H$. Hence

$$M = \max \left\{ \max_{1 \le k < h_H} \frac{\sum_{\lambda|\ell} f(\lambda) d r_{t_\lambda+1-i}}{\sum_{\lambda|\ell} f(\lambda) d_{t_\lambda+1-k}}, \max_{h_H \le k \le t_\lambda} \frac{\sum_{\lambda|\ell} f(\lambda) d r_{t_\lambda+1-i}}{1 + \sum_{\lambda|\ell} f(\lambda) d_{t_\lambda+1-k}} \right\}.$$

By Proposition 3.8 we have $d_{t_\lambda+1-k} \ge r_{t_\lambda+1-k}(2dh - r_{t_\lambda+1-k})$ and write

$$f_1(r_{t_\lambda+1-k}) = \frac{\sum_{\lambda|\ell} f(\lambda) d r_{t_\lambda+1-i}}{\sum_{\lambda|\ell} f(\lambda) r_{t_\lambda+1-k}(2dh - r_{t_\lambda+1-k})}$$

as well as

$$f_2(r_{t_\lambda+1-k}) = \frac{\sum_{\lambda|\ell} f(\lambda) d r_{t_\lambda+1-i}}{1 + \sum_{\lambda|\ell} f(\lambda) r_{t_\lambda+1-k}(2dh - r_{t_\lambda+1-k})}.$$

With essentially the same analysis of extremal values for f_1, f_2 as we did in the previous case we find that

$$M = \max \left\{ \frac{2d}{3dh}, \frac{ed^2h}{1 + e(dh)^2} \right\} = \frac{ed^2h}{1 + e(dh)^2}.$$

□

4.2 Products of abelian varieties with a factor of type IV

Recall that an abelian variety A can be written as the product of simple abelian varieties up to isogeny. Together with the treatment of simple abelian varieties of type I and II (compare [33]) or type III (compare [17]) previous work has also established the invariant

$\gamma(A)$ for abelian varieties that are isogenous to a product of simple factors of type I and II or type III. Specifically, let $A_1, ..., A_k$ be simple abelian varieties, pair-wise non-isogenous, all fully of Lefschetz type and of type I, II or III . Assume that A is isogenous to the product

$$\left(\prod_{i=1}^{k} A_i^{n_i}\right)$$

with positive integers $n_1, ..., n_k$. Denote for all $1 \leq i \leq k$ the dimension $g_i = \dim A_i$, the relative dimension $h_i = \dim_{rel} A_i$, E_i the center of $\mathrm{End}^\circ(A_i)$ with its degree $e_i = [E_i : \mathbb{Q}]$ and the degree

$$d_i = [\mathrm{End}^\circ(A_i) : E_i] = \begin{cases} 1 & A_i \text{ is of type I,} \\ 2 & A_i \text{ is of type II or III.} \end{cases}$$

Then [17, Theorem 1.7] establishes

$$\gamma(A) = \max_{I \subset \{1,...,k\}} \frac{2\sum_{i \in I} \dim A_i}{1 + \dim \mathrm{Hg}\left(\prod_{i \in I} A_i\right)} = \max_{I \subset \{1,...,k\}} \frac{2\sum_{i \in I} n_i d_i e_i h_i}{1 + e_i(2h_i^2 + \eta_i h_i)},$$

where

$$\eta_i = \begin{cases} +1 & A_i \text{ is of type I or II,} \\ -1 & A_i \text{ is of type III,} \end{cases}$$

In principle we should be able to extend this theorem for an abelian variety that is isogenous to a product of simple factors of type I, II, III or IV that are fully of Lefschetz type. However, we shall note the behaviour of the (ℓ-adic) Hodge group in that case, as it is a major ingredient of $\gamma(A)$. Let A_i for $1 \leq i \leq k$ be as above and A'_j (for $1 \leq j \leq k'$) be simple abelian varieties of type IV that are pairwise non-isogenous. Then due to a result of Lombardo [40, Theorem 4.7] we have

$$\mathrm{Hg}\left(\prod_{i=1}^{k} A_i \times \prod_{j=1}^{k'} A'_j\right)_{\mathbb{Q}_\ell} = \prod_{i=1}^{k} \mathrm{Hg}(A_i)_{\mathbb{Q}_\ell} \times \mathrm{Hg}\left(\prod_{j=1}^{k'} A'_j\right)_{\mathbb{Q}_\ell}.$$

Furthermore for any positive integers $n_1, ..., n_k, m_1, ..., m_{k'}$ we have [40, Proposition 2.8]

$$\mathrm{Hg}(A_1^{n_1} \times \cdots \times {A'}_k^{n_k} \times {A'}_1^{m_1} \times \cdots \times A_{k'}^{m_{k'}})_{\mathbb{Q}_\ell} = \mathrm{Hg}(A_1 \times \cdots \times A_k \times A'_1 \times \cdots \times A'_{k'})_{\mathbb{Q}_\ell}.$$

Hence we can not expect a result for any abelian variety with any amount of simple factors of type IV. However, if we restrict ourselves to abelian varieties with exactly one simple factor of type IV, an extension of [17, Theorem 1.7] seems feasible.

Let therefore A_i for $1 \leq i \leq k$ be as above and additionally let A_{k+1} be a simple abelian variety of type IV and fully of Lefschetz type. Denote $g = \dim A_{k+1}$ and $h = \dim_{rel} A$. Let E be the center of $\mathrm{End}^\circ(A_{k+1})$ with its degree e and $d = [\mathrm{End}^\circ(A_{k+1}) : E]$. Assume A is isogenous to the product

$$\left(\prod_{i=1}^{k} A_i^{n_i}\right) \times A_{k+1},$$

then we should be able to establish

$$\gamma(A) = \max_{I \subset \{1,...,k+1\}} \left\{\frac{\eta_{k+1} e d^2 h + 2\sum_{i \in J} n_i d_i e_i h_i}{1 + \eta_{k+1} e (dh)^2 + \sum_{i \in J} 2 e_i h_i^2 + \eta_i e_i h_i}\right\}, \tag{4.3}$$

where $J = I \setminus \{k+1\}$ and

$$\eta_{k+1} = \begin{cases} 1 & k+1 \in I, \\ 0 & k+1 \notin I. \end{cases},$$

following essentially the same strategy as in Section 4.1. Precisely we work with the following setup:

Let $H \subset A[\ell^\infty]$ be a finite subgroup and write

$$H = \prod_{i=1}^{k+1} H_i^{n_i},$$

where for all $1 \le i \le k+1$ we have $H_i \subset A_i[\ell^\infty]$. Remember that in order to determine $\gamma(A)$ we need to determine the order of H and the degree $[K(H) : K]$. We find a natural number n such that for each $1 \le i \le k+1$ we have $H_i \subset A_i[\ell^n]$ and further

$$H_i = \prod_{\lambda|\ell} \mathcal{H}_{\lambda,i},$$

where for all $1 \le i \le k+1$ and all $\lambda|\ell$

$$\mathcal{H}_{\lambda,i} = \begin{cases} H_{\lambda,i} \text{ if } A_i \text{ is of type I,} \\ H_{\lambda,i} \oplus H_{\lambda,i} \text{ if } A_i \text{ is of type II or III,} \\ \underbrace{H_{\lambda,i} \oplus \cdots \oplus H_{\lambda,i}}_{d\text{-times}} \text{ if } i = k+1. \end{cases}$$

Note that respectively we have

$$T_\ell(A_i) = \begin{cases} \prod_{\lambda|\ell} T_\lambda(A_i) \text{ if } A_i \text{ is of type I,} \\ \prod_{\lambda|\ell} (T_\lambda(A_i) \oplus T_\lambda(A_i)) \text{ if } A_i \text{ is of type II or III,} \\ \prod_{\lambda|\ell} (\underbrace{T_\lambda(A_i) \oplus \cdots \oplus T_\lambda(A_i)}_{d\text{-times}}) \text{ if } i = k+1. \end{cases}$$

Let

$$I_\ell := \{(\lambda, i) \mid 1 \le i \le k+1 \text{ and } \lambda \text{ is a place in } \mathrm{End}^\circ(A_i) \text{ over } \ell\},$$

then we can write

$$H = \prod_{(\lambda,i)\in I_\ell} {\mathcal{H}_{\lambda,i}}^{n_i}.$$

As seen before, choosing a suitable basis of $T_\lambda(A_i)$ for each $(\lambda, i) \in I_\ell$ we have

$$H_{\lambda,i} = \prod_{j=1}^{t_{\lambda,i}} (\mathbb{Z}/\ell^{m_j}\mathbb{Z})^{\alpha_j f(\lambda)} \simeq \prod_{j=1}^{t_{\lambda,i}} (\mathcal{O}_{\lambda,i}/\ell^{m_j}\mathcal{O}_{\lambda,i})^{\alpha_j},$$

where $1 \le t_{\lambda,i} \le d_i h_i$, $f(\lambda)$ is the residual degree of λ over ℓ and $1 \le m_{t_{\lambda,i}} \le \cdots \le m_1$. Note that both α_j and m_j depend (λ, i), which we will suppress in the notation for clarity.

With these notions established we have three steps to complete in order to proof (4.3):

Step 1: Calculate the order of H.

Step 2: Calculate the degree of the extension $K(H)/K$.

Step 3: Use combinatorial methods to determine $\gamma(A)$ with these two values.

We can calculate the order of H with the observation that

$$\operatorname{rk}_{\mathbb{F}_\lambda} H_{\lambda,i} = \sum_{j=1}^{t_{\lambda,i}} \alpha_j \quad \text{and} \quad \operatorname{rk}_{\mathbb{F}_\ell} H_{\lambda,i} = f(\lambda) \sum_{j=1}^{t_{\lambda,i}} \alpha_j$$

and hence

$$|H_{\lambda,i}| = \ell^{f(\lambda) \sum_{j=1}^{t_{\lambda,i}} \alpha_j m_j}.$$

Since

$$H_i = \prod_{\lambda | \ell} \bigoplus_{\nu=1}^{d_i} H_{\lambda,i}$$

we have

$$\operatorname{rk}_{\mathbb{F}_\lambda} H_i = d_i \operatorname{rk}_{\mathbb{F}_\ell} H_{\lambda,i} \quad \text{and therefore} \quad |H_i| = \ell^{d_i \sum_{\lambda|\ell} f(\lambda) \sum_{j=1}^{t_{\lambda,i}} \alpha_j m_j}.$$

Consequently we obtain

$$|H| = \prod_{i=1}^{k+1} |H_i|^{n_i} = \ell^{\sum_{i=1}^{k+1} n_i d_i \sum_{\lambda|\ell} f(\lambda) \sum_{j=1}^{t_{\lambda,i}} \alpha_j m_j}.$$

To write this expression for the order of H with the introduced set I_ℓ we denote $a_{ij} := d_i n_i \alpha_j$ and hence

$$|H| = \ell^{\sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} a_{ij} f(\lambda) m_j}.$$

Remember that α_j and hence a_{ij} depends on (λ, i), which we again suppress in the notation for simplicity. For the final expression of $\gamma(A)$ we take the ℓ-log of the order of H:

$$\log_\ell |H| = \sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} a_{ij} f(\lambda) m_j.$$

In order to calculate the degree $[K(H) : K]$ we remember the property μ to obtain a natural number m, such that $K(H) \cap K(\mu_{\ell^\infty}) = K(\mu_{\ell^m})$ and write

$$[K(H) : K] = [K(H) : K(\mu_{\ell^m})] \cdot [K(\mu_{\ell^m}) : K] = (\operatorname{Hg}(A)(\mathbb{Z}_\ell) : G(H)) \cdot \delta(H)$$

as before. By [40] we know that

$$\operatorname{Hg}(A) = \prod_{i=1}^{k+1} \operatorname{Hg}(A_i)$$

and hence obtain

$$\begin{aligned}(\operatorname{Hg}(A)(\mathbb{Z}_\ell) : G(H)) &= \prod_{i=1}^{k+1} (\operatorname{Hg}(A_i)(\mathbb{Z}_\ell) : G(H_i)) = \prod_{i=1}^{k+1} \prod_{\lambda|\ell} (\operatorname{Hg}(A_i)(\mathcal{O}_{\lambda,i}) : G(H_{\lambda,i})) \\ &= \prod_{(\lambda,i)\in I_\ell} (\operatorname{Hg}(A_i)(\mathcal{O}_{\lambda,i}) : G(H_{\lambda,i})).\end{aligned}$$

Applying Lemma 2.17 for each $\lambda|\ell$ to each $H_{\lambda,i}$ will yield the desired result. In order to achieve that, let $W_{i,j}$ denote the subspace of $T_\lambda(A_i)$ with

$$W_{i,j} \twoheadrightarrow (\mathbb{Z}/\ell^{m_j}\mathbb{Z})^{f(\lambda)\alpha_j}$$

for each $(\lambda, i) \in I_\ell$ and each $1 \leq j \leq t_{\lambda,i}$. Note that $\dim W_{i,j} = f(\lambda)\alpha_j$ and we have a decomposition

$$T_\lambda(A_i) = \left(\bigoplus_{j=1}^{t_{\lambda,i}} W_{i,j}\right) \oplus W'.$$

Let further

$$V_{i,j} := \prod_{\nu=1}^{t_{\lambda,i}-(j-1)} (\mathbb{Z}/\ell^{m_\nu}\mathbb{Z})^{f(\lambda)\alpha_\nu} \simeq \prod_{\nu=1}^{t_{\lambda,i}-(j-1)} (\mathcal{O}_{\lambda,i}/\ell^{m_\nu}\mathcal{O}_{\lambda,i})^{\alpha_\nu}$$

be subgroups such that $V_{i,t_{\lambda,i}} \subset \cdots \subset V_{i,1}$. Denote $\widehat{V}_{i,j}$ the corresponding submodule of $T_\lambda(A_i)$ and define

$$r_{i,j} = \mathrm{rk}_{\mathcal{O}_{\lambda,i}} \widehat{V}_{i,j} = \sum_{\nu=1}^{t_{\lambda,i}-(j-1)} \alpha_\nu.$$

As before let $G_{\widehat{V}_{i,j}}$ be the stabilizer of $V_{i,j}$, then we can write

$$G(H_{\lambda,i}) = \{M \in G(\mathcal{O}_{\lambda,i}) \mid M \in G_{\widehat{V}_{i,j}} \mod \ell^{m_{t_{\lambda,i}-(j-1)}}, 1 \leq j \leq t_{\lambda,i}\}.$$

For all $(\lambda, i) \in I_\ell$ and $1 \leq j \leq t_{\lambda,i}$ let $d_{i,j}$ denote the codimension of $G_{\widehat{V}_{i,j}}$. The analogue to Theorem 2.12 for the type I, II and III is

Theorem 4.2. *Let A be a simple abelian variety of type I or II (resp. type III) and dimension g. We consider the symplectic space (resp. hermitian space) $(V_\ell, \phi^\circ_{\ell^\infty})$ with the antisymmetric (resp. symmetric) bilinear form $\phi^\circ_{\ell^\infty} \colon V_\ell \times V_\ell \to E_\ell$. Let W be a $\mathbb{Q}_\ell$-subspace of $V_\ell(A)$ of codimension d and consider the stabilizer*

$$G_W = \{g \in G(V_\ell(A)) \mid g_{|W} = \mathrm{id}_W\} \subset \mathrm{GL}_{2g}(\mathbb{Q}_\ell)$$

where $G(V_\ell(A)) = \mathrm{Sp}(V_\ell(A)$ (resp. $G(V_\ell(A)) = O(V_\ell(A))$). Then

$$\dim G_W = \frac{d(d+\epsilon)}{2},$$

where

$$\epsilon = \begin{cases} 1 & \text{type I or II,} \\ -1 & \text{type III.} \end{cases}$$

This follows essentially from [17, Thm 2.1]. We therefore distinguish three cases:

1. If A_i is of type I or II, we have

$$d_{i,j} = \frac{2h_i(h_i+1)}{2} - \dim G_{\widehat{V}_{i,j}} = \frac{r_{i,j}(4h_i + 1 - r_{i,j})}{2},$$

2. If A_i is of type III, we have

$$d_{i,j} = \frac{2h_i(h_i-1)}{2} - \dim G_{\widehat{V}_{i,j}} = \frac{r_{i,j}(4h_i - 1 - r_{i,j})}{2},$$

3. if $i = k+1$, we have

$$d_{k+1,j} = (dh)^2 - \dim G_{\widehat{V}_{k+1,j}} \geq r_{k+1,j}(2dh - r_{k+1,j}).$$

With these values established we can use Lemma 2.17 to obtain

$$(\mathrm{Hg}(A_i)(\mathcal{O}_{\lambda,i}) : G(H_{\lambda,i})) \gg\ll \ell^{f(\lambda)\sum_{j=1}^{t_{\lambda,i}} d_{i,j}(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1})}$$

up to a multiplicative constant. Consequently

$$(\mathrm{Hg}(A)(\mathbb{Z}_\ell) : G(H)) \gg\ll \ell^{\sum_{(\lambda,i)\in I_\ell} f(\lambda)\sum_{j=1}^{t_{\lambda,i}} d_{i,j}(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1})}$$

and hence

$$\log_\ell(\mathrm{Hg}(A)(\mathbb{Z}_\ell) : G(H)) = \sum_{(\lambda,i)\in I_\ell} f(\lambda)\sum_{j=1}^{t_{\lambda,i}} d_{i,j}(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1}) + O(1).$$

Finally, we need to estimate $\delta(H)$. As before, we choose one of the $H_{\lambda,i}$ that shares the same value for δ as H. Fix a $(\lambda_1, 1) \in I_\ell$, such that after possibly reordering the factors A_i we have $H_1 \subset A_1[\ell^\infty]$ with a decomposition

$$H_1 = \prod_{\lambda|\ell} \bigoplus_{\mu=1}^{d_1} H_{\lambda,1},$$

a fixed place λ_1 in $\mathrm{End}^\circ(A_1)$ over ℓ with $H_{\lambda_1,1} \subset A_1[\lambda_1^\infty]$ and the property $\delta(H_{\lambda_1,1}) = \delta(H)$. We introduce two integers:

1. Let $m_{\lambda_1,1}$ be the maximal integer, such that elements $P, Q \in H_{\lambda_1,1}$ of order $\ell^{m_{\lambda_1,1}}$ exist and additionally $e_{\lambda^n}(P,Q)$ is of order $\ell^{m_{\lambda_1,1}}$. If no such elements exist, we let $m_{\lambda_1,1} = 0$.

2. Let $h_{H_{\lambda_1,1}} \in [1, t_{\lambda_1,1}]$ be minimal such that $m_{h_{H_{\lambda_1,1}}} \leq m_{\lambda_1,1}$. If $m_{\lambda_1,1} = 0$, we let $h_{H_{\lambda_1,1}} = t_{\lambda_1,1} + 1$.

Then

$$\ell^{m_{h_{H_{\lambda_1,1}}}} \ll \delta(H_{\lambda_1,1}) = \delta(H).$$

Now define for all $1 \leq j \leq t_{\lambda_1,1}$ the integer $\delta_{1,j}$ such that

$$\delta_{1,t_{\lambda_1,1}} = \cdots = \delta_{1,t_{\lambda_1,1}-(h_{H_{\lambda_1,1}}-1)+1} = 0 \quad \text{and} \quad \delta_{1,t_{\lambda_1,1}-h_{H_{\lambda_1,1}}+1} = \cdots = \delta_{1,1} = 1.$$

With the convention $m_{t_{\lambda_1,1}+1} = 0$ we can now write

$$m_{h_{H_{\lambda_1,1}}} = \sum_{j=1}^{t_{\lambda_1,1}} \delta_{1,j}(m_{t_{\lambda_1,1}-(j-1)} - m_{t_{\lambda_1,1}} - (j-1)+1).$$

Let $m_{\lambda,i}$ be such that $K(H_{\lambda,i}) \cap K(\mu_{\ell^\infty}) = K(\mu_{\ell^{m_{\lambda,i}}})$ for all $(\lambda, i) \in I_\ell$ in accordance with property μ, then by construction $m_{h_{H_{\lambda_1,1}}} \leq m_{\lambda_1,1} = \max_{(\lambda,i)} m_{\lambda,i}$. Hence

$$K(\mu_{\ell^{m_{\lambda,i}}}) \subset K(\mu_{\ell^{m_{h_{H_{\lambda_1,1}}}}})$$

for all $(\lambda, i) \in I_\ell$ and we can assume that for all $(\lambda, i) \neq (\lambda_1, 1)$ that $\delta_{i,j} = 0$ for all $1 \leq j \leq t_{\lambda,i}$. Therefore

$$m_{h_{H_{\lambda_1,1}}} = \sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} \delta_{i,j}(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1})$$

and consequently

$$\ell^{\sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} \delta_{i,j}(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1})} \ll \delta(H).$$

Combining the results above we find that

$$[K(H) : K] \gg \ell^{\sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} (f(\lambda)d_{i,j}+\delta_{i,j})(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1})}$$

and hence

$$\log_\ell[K(H) : H] \geq \sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} (f(\lambda)d_{i,j} + \delta_{i,j})(m_{t_{\lambda,i}-(j-1)} - m_{t_{\lambda,i}-(j-1)+1}) + O(1).$$

We introduce the notation

$$b_{i,j} := f(\lambda)(d_{i,t_{\lambda,i}-(j-1)} - d_{i,t_{\lambda,i}-(j-1)+1}) + \delta_{i,t_{\lambda,i}-(j-1)} - \delta_{i,t_{\lambda,1}-(j-1)+1}$$

and after a change of indices write

$$\log_\ell[K(H) : K] \geq \sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} m_j b_{i,j} + O(1).$$

Remember the equivalence

$$|H| \ll [K(H) : K]^{\gamma(A)} \Leftrightarrow \frac{\log_\ell |H|}{\log_\ell[K(H) : K]} \leq \gamma(A).$$

Combining the arguments above therefore yields

$$|H| \ll [K(H) : K]^{\gamma(A)} \Leftrightarrow \max_{m_1 > \cdots > m_{t_{\lambda,i}}} \left\{ \frac{\sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} f(\lambda)a_{i,j}m_j}{\sum_{(\lambda,i)\in I_\ell} \sum_{j=1}^{t_{\lambda,i}} b_{i,j}m_j} \right\} \leq \gamma(A).$$

This completes steps 1 and 2. The remaining step, being essentially an effort in calculation, has not been established yet.

4.3 Order of the extension generated by a torsion point

As a consequence of the methods used to prove Theorem 4.1 we can conclude a lower bound for the degree of the extension that is generated by a torsion point. Despite being interesting in its own right, there are various applications to Diophantine geometry.

Corollary 4.3. *Under the assumptions above, there exist $c := C(A, K) > 0$ such that, for every torsion point $P \in A(\bar{K})$ of order m we have:*

$$[K(P) : K] \geq c^{\omega(m)} m^{dh},$$

where $\omega(m)$ is the number of prime divisors of m.

Proof. Let P be a torsion point of order m. Denote H_P the End(A)-module generated by P, then $K(P) = K(H_P)$. We follow the argument of [33, Paragraph 9]. Let $m = \prod_{i=1}^{r} \ell_i^{n_i}$ be the decomposition of m into prime factors. Since we assume K to be a field over which all ℓ-adic representations are independant (compare the explanation in Section 3.1), up to multiplicative constant we obtain

$$[K(P : K)] = [K(P_1, ..., P_r)] \gg \prod_{i=1}^{r} [K(P_i) : K],$$

uniformly in m and P, where each point P_i is a torsion point of order $\ell_i^{n_i}$. Furthermore, up to multiplicative constant, we have uniformly in ℓ_i and P_i that

$$[K(P_i) : K] \gg \ell_i^{dhn_i},$$

since P generates a module of codimension dh. Now let $\omega(m)$ denote the number of prime factors of m, then there exists a constant $c = C(A, K)$ such that

$$[K(P) : K] = [K(P_1, ..., P_r) : K] \gg \prod_{i=1}^{r} c\ell_i^{dhn_i} \geq c^{\omega(m)} m^{dh}.$$

□

5. Computation of the maximum in 4.1

In this part we present the details for the calculation of the extremal values of the function

$$f(\underline{r}) = \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{\sum_{\lambda|\ell} f(\lambda) r_\lambda (2dh - r_\lambda)},$$

where $\underline{r} = (r_\lambda)_{\lambda|\ell}$ and $r_\lambda \in [0, dh/2]$ for all $\lambda|\ell$. Recall that ℓ is assumed to be unramified, hence $\sum_{\lambda|\ell} f(\lambda) = e$ is the degree of the extension $E/\mathbb{Q}$. In fact, we will show that it is sufficient to study the extremal values of the function

$$h(\underline{x}) = \frac{\sum_{i=1}^{e} x_i}{\sum_{i=1}^{e} x_i (2dh - x_i)}$$

on the cube $[0, dh/2]^e$. Let $\lambda_1, ..., \lambda_s$ be the (finitely many) places over ℓ, then we claim that

$$\max_{[0,dh/2]^s} f(\underline{r}) = \max_{[0,dh/2]^s} \frac{\sum_{\lambda=\lambda_1}^{\lambda_s} f(\lambda) r_\lambda}{\sum_{\lambda=\lambda_1}^{\lambda_s} f(\lambda) r_\lambda (2dh - r_\lambda)} = \max_{[0,dh/2]^e} \frac{\sum_{\lambda=1}^{e} x_i}{\sum_{\lambda=1}^{e} x_i(2dh - x_i)} = \max_{[0,dh/2]^e} h(\underline{x}).$$

Note that $f(\lambda_i)$ is a positive integer for all $1 \le i \le s$ and recall that $f(\lambda_1) + \cdots + f(\lambda_s) = e$. We can therefore write

$$f(\underline{r}) = \frac{\overbrace{r_{\lambda_1} + \cdots + r_{\lambda_1}}^{f(\lambda_1)\text{-times}} + \cdots + \overbrace{r_{\lambda_s} + \cdots + r_{\lambda_s}}^{f(\lambda_s)\text{-times}}}{\underbrace{r_{\lambda_1}(2dh - r_{\lambda_1}) + \cdots + r_{\lambda_1}(2dh - r_{\lambda_1})}_{f(\lambda_1)\text{-times}} + \cdots + \underbrace{r_{\lambda_s}(2dh - r_{\lambda_s}) + \cdots + r_{\lambda_s}(2dh - r_{\lambda_s})}_{f(\lambda_s)\text{-times}}}$$

Numerator and denominator are sums with e terms, hence if a point $(x_1, ..., x_e) \in [0, dh/2]^e$ obeys the properties

$$\begin{aligned} x_{\sigma(1)} &= \cdots = x_{\sigma(f(\lambda_1))} \\ x_{\sigma(f(\lambda_1)+1)} &= \cdots = x_{\sigma(f(\lambda_1)+f(\lambda_2))} \\ &\vdots \\ x_{\sigma(f(\lambda_1)+\cdots+f(\lambda_{s-1})+1)} &= \cdots = x_{\sigma(f(\lambda_1)+\cdots+f(\lambda_s))} \end{aligned}$$

for some permutation $\sigma \in S_e$ in the symmetric group over e elements, we have

$$h(x_1, ..., x_e) = f((x_{\sigma(f(\lambda_1))}, x_{\sigma(f(\lambda_1)+f(\lambda_2))}, ..., x_{\sigma(f(\lambda_1)+\cdots+f(\lambda_s))})).$$

Write $\Lambda \subset [0, dh/2]^e$ for the subset of all points that obey this property, then the image in $\mathbb{R}$ of f on $[0, dh/2]^s$ and the image in $\mathbb{R}$ of $h_{|\Lambda}$ on Λ are equal and so are their maxima. This immediately yields

$$\max_{[0,dh/2]^s} f(\underline{r}) \le \max_{[0,dh/2]^e} h(\underline{x}).$$

It is therefore sufficient to show that the point at which h maximizes is in Λ. Let $I \subseteq \{1, ..., e\}$. We consider the function

$$h\colon \mathbb{R}^{|I|} \supset [0, dh/2]^{|I|} \to \mathbb{R}, (x_i)_{i \in I} \mapsto \frac{\sum_I x_i}{\sum_I x_i(2dh - x_i)}.$$

defined on the compact set $[0, dh/2]^{|I|}$ and let x^* denote its maximum. Assume $x^* \in (0, dh/2)^{|I|}$, i.e. h attains its maximum on the interior of its domain, then x^* is a critical point of h. Hence $\nabla h(x^*) = 0$, that is

$$\partial_j h(x^*) = \frac{\partial h}{\partial x_j}(x^*) = 0$$

for all $j \in I$. We have

$$\partial_j h(\underline{x}) = \frac{\sum_{i \in I} x_i(2dh - x_i) - (2dh - 2x_j) \sum_{i \in I} x_i}{\left(\sum_{i \in I} x_i(2dh - x_i)\right)^2},$$

therefore

$$
\begin{aligned}
0 = \partial_j h(\underline{x}) \Leftrightarrow 0 &= \sum_{i\in I} (2dhx_i - x_i^2 - 2dhx_i + 2x_i x_j) = \sum_{i\in I} (2x_i x_j - x_i^2) \\
&\Leftrightarrow x_j^2 + 2\sum_{i\neq j} x_i x_j - \sum_{i\neq j} x_i^2 = 0 \\
&\Leftrightarrow x_j = \pm\sqrt{\left(\sum_{i\neq j} x_i\right)^2 + \sum_{i\neq j} x_i^2} - \sum_{i\neq j} x_i \qquad (*)
\end{aligned}
$$

Since the sums could be empty if there are less than two elements in I, we distinguish two cases:

Case 1: $|I| = 1$.

In this case

$$
h(x) = \frac{x}{x(2dh - x)} \qquad \text{and} \qquad h'(x) = \frac{1}{(2dh-x)^2} \neq 0.
$$

So obviously there are no critical points and hence the maximum has to be attained on the boundary. We have

$$
h(0) = \frac{1}{2dh} < \frac{2}{3dh} = h(dh/2).
$$

Case 2: $|I| > 1$.

Let $k \in I$ be such that $x_k^* \geq x_i^*$ for all $i \in I$. Then

$$
\begin{aligned}
x_k^* = \pm&\sqrt{\left(\sum_{i\neq k} x_i^*\right)^2 + \sum_{i\neq k} (x_i^*)^2} - \sum_{i\neq k} x_i^* \\
&\Leftrightarrow \sum_{i\in I} x_i^* = \pm\sqrt{\left(\sum_{i\neq k} x_i^*\right)^2 + \sum_{i\neq k} (x_i^*)^2} \\
&\Rightarrow \left(\sum_{i\in I} x_i^*\right)^2 = \left(\sum_{i\neq k} x_i^*\right)^2 + \sum_{i\neq k} (x_i^*)^2 \\
&\Leftrightarrow \sum_{i\in I} (x_i^*)^2 + \sum_{i\neq k} x_i^* x_j^* = 2\sum_{i\neq k} (x_i^*)^2 + \sum_{\substack{i\neq j\\ i,j\neq k}} x_i^* x_j^* \\
&\Leftrightarrow (x_k^*)^2 = \sum_{i\neq k} ((x_i^*)^2 - x_i^* x_k^*).
\end{aligned}
$$

By our choice of k we know that $(x_i^*)^2 - x_i^* x_k^* \leq 0$ for all $i \in I$. Hence we reached an obvious contradiction and h has no critical points. Therefore h attains its maximum on the boundary of its domain, i.e. $x_j^* = 0$ or $x_j^* = dh/2$ for at least one $j \in I$. Fix the corresponding entry $j \in I$.

If $x_j^* = 0$, then

$$
h(x^*) = \frac{\sum_{i\neq j} x_j^*}{\sum_{i\neq j} x_j^*(2dh - x_j^*)}
$$

and we can conclude by induction, replacing I with $I \setminus \{j\}$.

If $x^* = dh/2$, then

$$h(x^*) = \frac{\frac{dh}{2} + \sum_{i \neq j} x_i^*}{\frac{3}{4}(dh)^2 + \sum_{i \neq j} x_i^*(2dh - x_i^*)}$$

and for all $k \neq j$

$$\partial_k h(x^*) = \frac{\frac{3}{4}(dh)^2 + \sum_{i \neq j} x_i^*(2dh - x_i^*) - (2dh - 2x_k^*)\left(\frac{dh}{2} + \sum_{i \neq j} x_i^*\right)}{\left(\frac{3}{4}(dh)^2 + \sum_{i \neq j} x_i^*(2dh - x_i^*)\right)^2}.$$

Repeating a similar argument, we show that h considered as a function on $[0, dh/2]^{|I \setminus \{j\}|}$ has no critical points in the interior. Consider

$$\begin{aligned}
0 = \partial_k h(\underline{x}) \Leftrightarrow -\frac{3}{4}(dh)^2 &= \sum_{i \neq j} (2x_k x_i - x_i^2) - (dh)^2 + dh x_k \\
\Leftrightarrow 0 &= x_k^2 + \left(dh + 2 \sum_{i \neq k,j} x_i\right) x_k - \left(\sum_{i \neq k,j} x_i^2 + \frac{(dh)^2}{4}\right) \\
\Leftrightarrow x_k &= \pm \sqrt{\left(\sum_{i \neq k,j} x_i + \frac{dh}{2}\right)^2 + \sum_{i \neq k,j} x_i^2 + \frac{(dh)^2}{4}} - \sum_{i \neq k,j} x_i - \frac{dh}{2}
\end{aligned}$$

and let $m \in I \setminus \{j\}$ be such that $x_m^* \geq x_i$ for all $i \in I \setminus \{j\}$. Then

$$\begin{aligned}
&\left(\sum_{i \neq j} x_i^* + \frac{dh}{2}\right)^2 = \left(\sum_{i \neq j,m} x_i^* + \frac{dh}{2}\right)^2 + \sum_{i \neq j,m} (x_i^*)^2 + \frac{(dh)^2}{4} \\
&\Leftrightarrow \left(\sum_{i \neq j} x_i^*\right)^2 + dh \sum_{i \neq j} x_i^* = \left(\sum_{i \neq j} x_i^*\right)^2 + dh \sum_{i \neq j,m} x_i^* + \sum_{i \neq j,m} (x_i^*)^2 + \frac{(dh)^2}{4} \\
&\Leftrightarrow \sum_{i \neq j} (x_i^*)^2 + \sum_{\substack{i \neq k \\ i,k \neq j}} x_i^* x_k + dh x_m^* = \sum_{i \neq j,m} (x_i^*)^2 + \sum_{\substack{i \neq k \\ i,k \neq j,m}} x_i^* x_k^* + \sum_{i \neq j,m} (x_i^*)^2 + \frac{(dh)^2}{4} \\
&\Leftrightarrow (x_m^*)^2 + dh x_m^* = \sum_{i \neq j,m} \left((x_i^*)^2 - x_i^* x_m^*\right) + \frac{(dh)^2}{4} \\
&\Leftrightarrow \left(x_m^* - \frac{dh}{2}\right)^2 = \sum_{i \neq j,m} \left((x_i^*)^2 - x_i^* x_m^*\right)
\end{aligned}$$

which yields the same contradiction. We inductively conclude that for each $j \in I$ we have $x_j^* = 0$ or $x_j^* = dh/2$. Let $J \subset I$ be such that $x_j^* = 0$ for all $j \in J$ and $x_j^* = dh/2$ else. Then

$$h(x^*) = \frac{\sum_{I \setminus J} \frac{dh}{2}}{\sum_{I \setminus J} \frac{dh}{2}(2dh - dh/2)} = \frac{|I \setminus J|}{|I \setminus J| 3/2 dh} = \frac{2}{3dh}.$$

A similar calculation with essentially the same arguments for

$$g(\underline{r}) = \frac{\sum_{\lambda|\ell} f(\lambda) r_\lambda}{1 + \sum_{\lambda|\ell} f(\lambda) r_\lambda (2dh - r_\lambda)}$$

yields that a maximum is achieved at $x^* = (dh, ..., dh)$ with

$$g(x^*) = \frac{edh}{1 + e(dh)^2}.$$

6. SMALL PRIMES

In this section we discuss the cases of finite subgroups $H \subset A[\ell^\infty]$ where $\ell \in \mathcal{S}$ (compare Definition 1.9). We use a strategy similar to [33, Section 8]. Recall that every prime in $\mathcal{S}$ satisfies at least one of the following conditions:

1. ℓ is ramified in $\mathcal{O}_L$, $\ell \nmid \deg \phi$ and D decomposes over E_λ for all $\lambda|\ell$;
2. $\ell \mid \deg \phi$ and D decomposes over E_λ for all $\lambda|\ell$;
3. D does not decompose (and consequently ℓ is ramified in $\mathcal{O}_E$).

Case (1): ℓ ramified Note that if ℓ is ramified in $\mathcal{O}_L$ we have $\ell\mathcal{O}_E = \prod_{\lambda|\ell} \lambda^{e(\lambda)}$. We shall assume $e(\lambda) > 1$ for all $\lambda|\ell$ and define the pairing

$$\phi^*_{\ell^\infty} : T_\ell(A) \times T_\ell(A) \to \mathcal{O}^*_{E_\ell}(1)$$

and note that $\mathcal{O}_{E_\ell} \subset \mathcal{O}^*_{E_\ell}$.

Consider first that the image of $\phi^*_{\ell^\infty}$ is in $\mathcal{O}_{E_\ell}$. Then we obtain the following pairing

$$\phi_{\lambda^n} : (T_\lambda/\lambda^n T_\lambda) \times (T_\lambda/\lambda^n T_\lambda) \to \mathcal{O}_\lambda/\lambda(1)$$

by reduction modulo λ^n (for each $n \geq 1$). Since $\ell\mathcal{O}_\lambda = \lambda^{e(\lambda)}$ we have the decomposition

$$A[\ell^n] = T_\ell(A)/\ell^n T_\ell(A) = \prod_{\lambda|\ell} (T_\lambda/\lambda^{e(\lambda)n} T_\lambda)^d.$$

Since $\mathrm{rk}_{\mathbb{Z}_\ell}(\mathcal{O}_\lambda) = f(\lambda)e(\lambda)$ we have

$$(\mathcal{O}_\lambda/\ell^n\mathcal{O}_\lambda) \cong (\mathbb{Z}/\ell^n\mathbb{Z})^{f(\lambda)e(\lambda)}.$$

We can therefore employ the same methods as in Section 3 with the following modifications:

For a finite subgroup $H \subset A[\ell^n]$ we have a decomposition $\prod_{\lambda|\ell} H^d_\lambda$, where $H_\lambda \subset T_\lambda[\lambda^{e(\lambda)n}]$. For the property μ we use

$$m_1(H_\lambda) = \max\{k \in \mathbb{N} \mid \exists n \geq 0, \exists P, Q \in H_\lambda \text{ of order } \lambda^{e(\lambda)n},\ \phi_{\lambda^{e(\lambda)n}(P,Q)} \text{ is of order } \lambda^{e(\lambda)k}\}.$$

Furthermore we replace $f(\lambda)$ by $f(\lambda)e(\lambda)$ and use the pairing

$$\phi_{\lambda^{e(\lambda)n}} : T_\lambda[\lambda^{e(\lambda)n}] \times T_\lambda[\lambda^{e(\lambda)n}] \to \mathcal{O}_\lambda/\lambda^{e(\lambda)n}(1) = \mathcal{O}_\lambda/\ell^n\mathcal{O}_\lambda(1).$$

for the calculations.

Next assume the image of $\phi^*_{\ell^\infty}$ is in $\mathcal{O}^*_{E_\ell}$. Denote E_λ as the completion of E at a place λ, then $\mathcal{O}^*_\lambda$ is a fractional ideal in E_λ. Hence there is an integer m_λ and a unit π_λ such that

$$\mathcal{O}^*_\lambda = \pi_\lambda^{-m_\lambda}\mathcal{O}_\lambda.$$

Let

$$m_0 = \gcd\{m_\lambda \mid \lambda|\ell,\ \ell \text{ ramified in } \mathcal{O}_L\},$$

then $\ell^{m_0}\mathcal{O}^*_{E_\ell} \subset \mathcal{O}_{E_\ell}$. It follows that

$$\mathcal{O}^*_{E_\ell} = \prod_{\lambda|\ell}\mathcal{O}^*_\lambda = \prod_{\lambda|\ell}\pi_\lambda^{-m_\lambda}\mathcal{O}_\lambda \subset \ell^{-m_0}\prod_{\lambda|\ell}\mathcal{O}_\lambda = \ell^{-m_0}\mathcal{O}_{E_\ell}.$$

Replacing $T_\ell(A)$ with $T'_\ell = \ell^{m_0}T_\ell(A)$ and a finite subgroup $H \subset A[\ell^n]$ with $H' = \ell^{m_0}H \subset T'_\ell/\ell^n T'_\ell$, we can employ the methods from Section 3 to obtain

$$|H'| \ll [K(H') : K]^{\gamma(A)} \leq [K(H) : K]^{\gamma}$$

and

$$|H| \leq |H'| + |A[\ell^{m_0}]|.$$

Since $\mathcal{S}$ is a finite set $|A[\ell^{m_0}]|$ is necessarily bounded and we obtain

$$|H| \ll [K(H) : K]^{\gamma(A)}$$

for all $H \subset A[\ell^\infty]$.

Case (2): $\ell|\deg(\phi)$ Let

$$m_0 := \max_{k\in\mathbb{N}}\{\ell^k|\deg(\phi)\}.$$

Since $\ell|\deg(\phi)$ the associated pairing is not going to be non-degenerate modulo ℓ^n for all $n \geq 1$. Hence we make the modifications

$$T'_\ell := \ell^{m_0}T_\ell(A)$$

and

$$H' := \ell^{m_0}H \subset T'_\ell/\ell^n T'_\ell$$

for all $n \geq 1$ in order to obtain a non-degenerate pairing. Define the pairing as follows:

$$T'_\ell \times T'_\ell \to \mathbb{Z}_\ell(1), (x, y) \mapsto \phi_{\ell^\infty}(x, \phi(y)).$$

We employ the methods of Section 3 again to obtain

$$|H'| \ll [K(H') : K]^{\gamma(A)} \leq [K(H) : K]^{\gamma(A)}$$

and

$$|H| \leq |H'| + |A[\ell^{m_0}]|$$

uniformly in H. With the same arguments as for Case (1) above we conclude

$$|H| \ll [K(H) : K]^{\gamma(A)}$$

for all $H \subset A[\ell^\infty]$.

Case (3): D does not decompose If ℓ is ramified in $\mathcal{O}_E$ and the division algebra D does not decompose we can find a finite extension L/E such that

$$D \otimes_E L \cong \mathrm{Mat}_d(L)$$

and for any ℓ in consideration

$$D_\ell \otimes_{E_\ell} L_\ell = \prod_{\lambda|\ell} \mathrm{Mat}_d(L_\lambda),$$

where $L_\lambda = E_\lambda \otimes_E L$ (compare [2, Lemma 2.1]). Denote

$$T_{\ell,L}(A) := T_\ell(A) \otimes_{\mathcal{O}_{E_\ell}} \mathcal{O}_{L_\ell},$$

then we have a decomposition

$$T_{\ell,L} = \prod_{\lambda|\ell} \mathcal{T}_{\lambda,L}$$

with

$$\mathcal{T}_{\lambda,L} = \underbrace{T_{\lambda,L} \oplus \cdots \oplus T_{\lambda,L}}_{d\text{-times}}.$$

For $H \subset A[\ell^n]$ with

$$H = \prod_{\lambda|\ell} \mathcal{H}_\lambda$$

where $\mathcal{H}_\lambda \subset \mathcal{T}_{\lambda,L}/\ell^n \mathcal{T}_{\lambda,L}$ we get a decomposition

$$\mathcal{H}_\lambda = \bigoplus_{\sigma \in \mathrm{Gal}(L/E)} H_\lambda^\sigma$$

with $H_\lambda \subset T_{\lambda,L}/\ell^n T_{\lambda,L}$. Since we don't have d copies of H_λ as before we instead calculate the order of $\mathcal{H}_\lambda$ and the degree $[K(\mathcal{H}_\lambda) : K]$ to obtain the invariant $\gamma(A)$ in a similar fashion as before.

Bibliography

[1] G. Banaszak, W. Gajda, P. Krasoń *On the image of ℓ–adic Galois representations for abelian varieties of type I and II*, Doc. Math., Extra volume Coates, 55-95, 2006.

[2] G. Banaszak, A. Kaim–Garnek, *The Tate module of a simple abelian variety of type IV*, New York J. Math. 27, 1240-1257, 2021.

[3] V. Beresnevich, D. Dickinson and S. Velani, *Diophantine approximation on planar curves and the distribution of rational points*, With an Appendix II by R.C. Vaughan, Ann. of Math.(2) 166, no. 2, 367-426, 2007.

[4] V. Beresnevich, *Rational points near manifolds and metric Diophantine approximation*, Ann. of Math.(2) 175, no. 1, 187-235, 2012.

[5] V. Beresnevich, R.C. Vaughan, S. Velani and E. Zorin, *Diophantine Approximation on Manifolds and the Distribution of Rational Points: Contributions to the Convergence Theory*, Int. Math. Res. Not. IMRN, no. 10, 2885-2908, 2017.

[6] A. Borel, *Linear algebraic groups*, 1991.

[7] M. V. Borovoi, *On the Galois action on the rational cohomology classes of type (p,p) of abelian varieties*, Math. USSR Sb., no. 23(4), 613-616 1974.

[8] N. Broberg, *A note on a paper by R. Heath–Brown: "The density of rational points on curves and surfaces"*, J. reine angew. Math. 571, 159-178, 2004.

[9] N. Broberg and P. Salberger, *Counting rational points on threefolds*, Arithmetic of higher–dimensional algebraic varieties, Prog. Math. 226, Birkhaeuser, 105-120, 2003.

[10] T.D. Browning and D.R. Heath-Brown, *Counting rational points on hypersurfaces*, J. Reine Angew. Math. 584, 83-115, 2005.

[11] T.D. Browning and D.R. Heath-Brown, *The density of rational points on non-singular hypersurfaces, I*, Bull. London Math. Soc. 93, 401-410, 2006.

[12] T.D. Browning and D.R. Heath-Brown, *The density of rational points on non-singular hypersurfaces, II*, Proc. London Math. Soc. 38, 273-303, 2006.

[13] T.D. Browning, D.R. Heath-Brown and P. Salberger, *Counting rational points on algebraic varieties*, Duke Math. J. 132, 545-578, 2006.

[14] T.D. Browning, *Quantitive arithmetic of projective varieties*, Prog. Math., vol. 277, 2009.

[15] T.D. Browning, *Cubic Forms and the circle method*, Prog. Math., vol. 343, 2021.

[16] V. Cantoral-Farfán, *Points de torsion pour les variétés abéliennes de type III.*, PhD Thesis, DOI 10.13140/RG.2.2.17695.05283, 2017.

[17] V. Cantoral-Farfán, *Torsion for abelian varieties of type III*, J. Number Theory 198, 346-380, 2019.

[18] W. Castryck, R. Cluckers, P. Dittman and K. H. Nguyen, *The dimension growth conjecture, polynomial in the degree and without logarithmic factors*, Algebra Number Theory 12, no. 8, 2261-2294, 2020.

[19] W.Chi, *On the ℓ-adic representations attached to simple abelian varieties of type iv*, Bull. Austral. Math. Soc., no. 444: 71 - 78, 1991.

[20] W.Chi, *ℓ-adic and λ-adic representations associated to abelian varieties defined over number fields*, Am. J. Math., no. 114(2):315 - 353, 1992.

[21] J. Commelin, *The Mumford-Tate conjecture for products of abelian varieties*, Algebr. Geom. 6, no. 6, 650-677, 2019.

[22] H. Davenport, *Analytic Methods for Diophantine Equations and Diophantine Inequalities,*, Cambridge Mathematical Library, prepared for publication by T. D. Browning, 2005.

[23] P. Deligne, J.S. Milne, A. Ogus and K.-Y. Shih *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math., vol. 900 1982.

[24] N. Dieudonné, *La géométrie des groupes classiques*, 1963.

[25] N.D. Elkies, *Rational points near curves and small nonzero* $|x^3 - y^2|$ *via lattice reduction*, Algorithmic number theory, (Leiden, 2000), 33-63, Lecture Notes in Comput. Sci., 1838 Springer, Berlin 2000.

[26] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent Math, no. 73, 349-366, 1983.

[27] S. Le Fourn, D. Lombardo, D. Zywina *Torsion bounds for a fixed abelian variety and varying number field*, arXiv:2208.02345, 2022.

[28] C. Hall, *An open-image theorem for a general class of abelian varieties with an appendix by Emmanuel Kowalski.*, Bull. Lond. Math. Soc., no. 43(4), 703-711, 2011.

[29] D.R. Heath-Brown, *Cubic forms in Ten Variables*, Proc. London Math. Soc., s3-47, 225-257, 1983.

[30] D.R. Heath-Brown, *The density of rational points on curves and surfaces*, Ann. of Math. (2) 155, no. 2, 553-595, 2002.

[31] M. Hindry, N. Ratazzi *Torsion dans un produit de courbes elliptiques*, J. Ramanujan Math. Soc. 25, 1-31, 2010.

[32] M. Hindry, N. Ratazzi *Points de torsion sur les variétés abéliennes de type* GSp, Journal of the Institute of Mathematics of Jussieu 11 (01), 27-65, 2012.

[33] M. Hindry, N. Ratazzi *Torsion pour les variétés abéliennes de type I et II*, Algebra Number Theory 10, no. 9, 1845-1891, 2016.

[34] L. Hörmander, *The analysis of linear partial differential operators. I. Distribution theory and Fourier analysis*, Second edition. Grundlehren der Mathematischen Wissenschaften, 256. Springer-Verlag, Berlin, 1990.

[35] J.-J. Huang, *Rational points near planar curves and Diophantine approximation*, Adv. Math. 274, 490-515. 2015.

[36] J.-J. Huang, *The density of rational points near hypersurfaces*, Duke Math. J. 169, 2045-2077. 2020.

[37] C.P. Hughes and Z. Rudnick, *On the distribution of lattice points in thin annuli*, Int. Math. Res. Not. 2004, no. 13, 637-658, 2004.

[38] M. N. Huxley, *The rational points close to a curve*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 21, no. 3, 357-375, 1994.

[39] S. Kamienny, *Torsion points on elliptic curves over fields of higher degree*, Internat. Math. Res. Notices , no.6, 129-133, 1992.

[40] D. Lombardo, *On the* ℓ*-adic Hodge group of nonsimple abelian varieties*, Ann. de l'Institut Fourier 66 , no.3, 1217-1245, 2016.

[41] D. Lombardo, *Roots of unity and torsion points of abelian varieties*, Ramanujan J 43, 383-403, 2017.

[42] O. Marmon, *The density of integral points on hypersurfaces of degree at least four*, Acta. Arith. 141, no. 3, 211-240, 2010.

[43] D. Masser, *Lettre à Daniel Bertrand du 10 novembre 1986*, 1986.

[44] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math, no. 44(2), 129-162, 1978.

[45] B. Mazur, *Perturbations, deformations, and variations (and "near-misses") in geometry, physics, and number theory*, Bull. Amer. Math. Soc. (N.S.) 41, no. 3, 307-336, 2004.

[46] L. Merel, *Bornes pour la torsion des courbes elliptiques sur les corps de nombres*, Invent. Math., no. 124(1-3), 437-339, 1996.

[47] J. S. Milne, *Lefschetz classes on abelian varieties*, Duke Math. J., no. 96(3), 639-670, 1999.

[48] J. S. Milne, *Abelian Varieties*, http ://www.jmilne.org/math/CourseNotes/AV.pdf, Version 2.0.

[49] H. Montgomery, *Ten lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics (Book 84), American Mathematical Society, 1994.

[50] B.J.J. Moonen et Yu.G. Zarhin, *Hodge classes on abelian varieties of low dimension*, Math. Ann., no. 315, 711–733, 1999.

[51] L. J. Mordell, *Diophanitne equations*, Academic Press, New York and London, 1969.

[52] D. Mumford, *Abelian varieties*, Oxford Univ. Press, 1974.

[53] V. K. Murty, *Hodge and Weil classes on abelian varieties*, The Arithmetic and Geometry of Algebraic Cycles, 83-115 2000.

[54] I.I. Piatetski-Shapiro, *Interrelations betwen the Tate and Hodge conjectures for abelian varieties*, Math. USSR Sb., no. 14(4), 615-624 1971.

[55] R. Pink, *ℓ-adic algebraic monodromy groups, cocharacters, and the Mumford-Tate conjecture*, J. reine angew. Math, 495, 1998.

[56] K. Ribet, *Galois Action on Division Points of Abelian Varieties with Real Multiplications*, Amer. J. Math. 98(3), 751-804 1976.

[57] O. Robert, P. Sargos *Three-dimensional exponential sums with monomials*, J. Reine Angew. Math. 591, 1-20 2006.

[58] N. Ratazzi, *Borne sur la torsion dans les variétés abéliennes de type CM*, Ann. Sci. de l'École Norm. Supérieur 40, no. 6, 951-983 2007.

[59] P. Salberger, *On the density of rational and integral points on algebraic varieties*, J. Reine u. Angew. Math. 606, 123-147, 2007.

[60] P. Salberger, *Rational points of bounded height on projective surfaces*, Math. Z. 258, no. 4, 805-826, 2008.

[61] P. Salberger, *Uniform bounds for rational points on cubic hypersurfaces*, in Arithmetic and geometry, London Math. Soc. Lecture Note Ser., vol. 420, Cambridge Univ. Press, Cambridge, 401-421, 2015.

[62] J.-P. Serre, *Résumé des cours au Collège de France 1985-1986*, Springer Collected Works in Mathematics. 2003.

[63] J.-P. Serre, *Lectures on the Mordell-Weil theorem*, Research Notes in Mathematics, vol.1, Lecture notes prepared by Henri Darmon [Henri Darmon], With a foreword by Darmon and the author 1989.

[64] J.-P. Serre, *Topics in Galois theory*, Aspects of Mathematics, E15, Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt, 1989.

[65] J.-P. Serre, *Un critère d'indépendance pour une famille de représentations ℓ-adiques*, Comment. Math. Helv 88, no. 3, 541-554, 2013.

[66] D. Schindler, *Diophantine Inequalities for generic ternary diagonal forms*, Int. Math. Res. Not. IMRN, no. 11, 3396-3416, 2020.

[67] D. Schindler, S. Yamagishi, *Density of rational points near/on compact manifolds with certain curvature conditions*, arXiv:2103.05281, 2021.

[68] R.C. Vaughan, *The Hardy-Littlewood Method*, Cambridge Tracts in Math. 125, 1997.

[69] R.C. Vaughan and S. Velani, *Diophantine approximation on planar curves: the convergence theory*, Int. Math. 166, no. 1, 103-124, 2006.

[70] M.N. Walsh, *Bounded rational points of curves*, Int. Math. Res. Not. IMRN 2015, no. 14, 5644-5658, 2015.

Mathematisches Institut, Georg-August Universität Göttingen, Bunsenstrasse 3-5, 37073 Göttingen, Germany

Email address: `florian.munkelt@mathematik.uni-goettingen.de`