ISOLATING SPECIAL SOLUTIONS IN THE DELTA METHOD:
THE CASE OF A DIAGONAL CUBIC EQUATION
IN EVENLY MANY VARIABLES OVER \( \mathbb{Q} \)

VICTOR Y. WANG

Abstract. Let \( F \in \mathbb{Z}[x] \) be a diagonal cubic form in \( m \in \{4, 6\} \) variables. In the “standard cubic delta method” of Duke–Friedlander–Iwaniec and Heath-Brown, we extract a weighted count of certain special solutions \( x \in \mathbb{Z}^m \) to \( F(x) = 0 \) with \( x \ll X \), asymptotically as \( X \to \infty \), by “restricting attention” to the set of vectors \( c \neq 0 \) (in the “dual lattice” \( \mathbb{Z}^m \)) for which the (projective) hyperplane section \( F(x) = c \cdot x = 0 \) is singular over \( \mathbb{C} \).

In an appendix, we explain how \( c = 0 \) instead isolates the singular series, at least for \( m \geq 5 \). We also highlight a decay property that helps to “separate” \( c = 0 \) from the aforementioned tuples \( c \neq 0 \) when \( m = 6 \), in the spirit of a conjecture of Hooley.

Our main results are unconditional. Heath-Brown’s past work actually already includes our main theorem for \( m = 4 \), but not for \( m = 6 \). Our overall strategy seems new, and also suggests a framework that may allow for non-diagonal smooth \( F \) in the future.

Table of Contents

| Section | Page |
|---------|------|
| Conventions and notes | 1 |
| 1. Introduction | 2 |
| 2. Maximal linear subvarieties under duality | 8 |
| 3. Easy sums by duality | 13 |
| 4. Bounding various sparse contributions | 16 |
| 5. A full proof outline for the theorem | 18 |
| 6. Establishing bias in exponential sums | 24 |
| Appendix A. Algebraic geometry background | 30 |
| Appendix B. Isolating the singular series | 32 |
| Acknowledgements | 33 |
| References | 33 |

Conventions and notes

On series references. This is Paper II of I–III, with I being [Wan21]. (We refer the reader to [tinyurl.com/hooley33] for any errata that may be discovered.) The present paper is logically independent of III, and essentially logically independent of I. But we will sometimes refer to I, or mention III. (Here one could “substitute” [HB98, pp. 675–684, §§2–5] for I.)

On vectors and functions. By default, norms \( \| \cdot \| \) will take place in \( \ell^\infty \) or \( L^\infty \).

We will always interpret the support of a function in the closed sense.

Date: August 10, 2021.
Given an abstract polynomial $f$ in $m$ variables, we let $\text{Hess}(f)$ denote the usual $m \times m$ Hessian matrix; $\det(\text{Hess} f)$ the Hessian determinant; $\text{hess}(f) := V_{\mathbb{A}^m}(\det(\text{Hess} f)) \subseteq \mathbb{A}^m$ the affine Hessian vanishing locus, and (when $f$ is homogeneous) $\text{hess}(V(f)) := V(\det(\text{Hess} f)) \subseteq \mathbb{P}^{m-1}$ the projective Hessian hypersurface.

Lastly, we adopt multi-index notation for multivariable calculus, especially in the context of derivatives; e.g. we let $\partial^r := \partial_{x_1}^{r_1} \cdots \partial_{x_m}^{r_m}$ (for $r \in \mathbb{Z}_{\geq 0}^m$) and $|b| := \sum_{i=1}^m b_i$ (for $b \in \mathbb{Z}^S$). For Paper III, we also let $\partial_{\log u} := u \cdot \partial u$ over $u \in \mathbb{R}$, and $\partial_{\log \|x\|} := \sum_{i} \partial_{\log x_i}$ over $x \in \mathbb{R}^m$. (Under our definitions, $\partial_{\log \|x\|} f(x) = \lambda f(\lambda x)|_{\lambda=1}$. So in any reasonable radial system of coordinates $(r, \ldots)$—with $r := \|x\|_p$ for some $p > 0$—we have $\partial_{\log \|x\|_p} = \partial_{\log \|x\|}$.)

**On inequalities and parameters.** We write $f \lesssim_S g$, or $g \gtrsim_S f$, to mean “$f \leq C g$ for some real $C = C(S) > 0$”; in practice $f, g$ will be $\geq 0$. We will often attach “size adjectives” to such *imprecise inequalities* when we really mean to describe their implied constants; e.g. “if $A \lesssim_S B$ would mean “if $C$ is small in terms of $S$, and if $A \leq CB$”.

**Remark 0.1.** Typically “$f \leq C g$” will either appear in a statement as (i) a hypothesis, in which case $C$ may be given beforehand, allowed to be arbitrary, or specified otherwise; or (ii) a result, in which case we have flexibility in choosing $C$ throughout the proof.

In this context, a phrase of the form “$P$ unless $Q$” should be read “if $\neg P$, then $Q$”.

If $f, g$ have “sizes” $|f|,|g|$, we write $f \ll g$ or $f = O(g)$, or $g \gg f$ or $g = \Omega(f)$, to mean $|f| \lesssim |g|$; in practice, the right-hand side of $\ll$, or both sides of $\gg$, will often be $\geq 0$. We write $f \asymp g$, or $f = \Theta(g)$, if $f = O(g)$ and $f = \Omega(g)$. In *exponents* (e.g. that of $X^{3-\Theta(1)}$), we enforce positivity of implied constants for $\Omega, \Theta$ (but not for $O$).

Generally, $\epsilon$ will denote an arbitrarily small loss, while $\delta$ will denote a small fixed saving. We follow the usual convention that small or large parameters like $\epsilon, \delta, \Theta(1), A$ may be modified or re-defined finitely many (i.e. a total of $O(1)$) times throughout the paper, though we will try to keep clear the order in which parameters depend on one another.

Finally (mostly only for expository purposes), we write $f = o_{S,X \to \infty}(g)$ to mean that for every $\epsilon > 0$, there exists $X(\epsilon, S)$ such that $|f| \leq \epsilon |g|$ for all $X \geq X(\epsilon, S)$.

1. Introduction

Let $F \in \mathbb{Z}[x]$ be a cubic form in $m \geq 4$ variables, smooth over $\mathbb{Q}$. (The diagonality assumption will come later.) Let $\mathcal{V}$ denote the proper scheme over $\mathbb{Z}$ cut out by $F(x) = 0$. Then $\mathcal{V} := \mathcal{V}_{\mathbb{Q}}$ is a smooth projective hypersurface in $\mathbb{P}^{m-1}_{\mathbb{Q}}$.

**Definition 1.1.** Fix $w \in C^\infty_c(\mathbb{R}^m)$ such that $\nabla F$ is bounded away from 0 on $\text{Supp} w$; equivalently (since $V_{\mathbb{Q}}$ is smooth), we require $0 \notin \text{Supp} w$. For $X \geq 1$, let $N_{F,w}(X) := \sum_{x \in \mathbb{Z}^m} w(x/X) \cdot 1_{F(x) = 0}$ denote a *smoothly weighted* count of $\mathbb{Z}$-points in homogeneously expanding regions on the affine cone $C(\mathcal{V})$ (i.e. $F(x) = 0$ sitting in $\mathbb{A}^m_{\mathbb{Z}}$).

**Remark 1.2.** From a projective standpoint, one might want to restrict to primitive $x \in \mathbb{Z}^m$. Such a restriction would likely be cosmetic for $m \geq 5$ (cf. [HB96], paragraph preceding Corollary 2, for *quadrics* in $\geq 4$ variables)—and minor, up to a factor of $\log X$, for $m = 4$ (cf. [HB96] Theorem 8 versus Corollary 2, for *ternary quadrics*).

For smooth projective $V/\mathbb{Q}$ with $m \geq 4$ and $\deg F = 3$ as above, the Hardy–Littlewood “randomness” (singular series) prediction may fail, even when $m = 6$ [Hoo86a].
Remark 1.3. The “randomness” prediction is still expected to hold uniformly for $m \geq 7$, since $V/\mathbb{Q}$ is smooth. (At least when $m \geq 8$ and $(\text{Supp } w) \cap (\text{hess } F) \neq \emptyset$, [Hoo14,Hoo15] provides a conditional affirmative proof, unconditional for $m \geq 9$.) For singular cubics, however, failure can occur even when $m \geq 7$; see [BW19] for an interesting example when $m = 8$.

But every “randomness failure” should have a good excuse! This philosophy underlies the precise conjectures that motivate the present paper.

1.1. Some conjectures of Hooley–Manin type, as motivation. A plausible version of Manin’s conjecture (in a smoothed form) for $C(V) \subseteq \mathbb{A}^m_{\mathbb{Z}}$ says that away from a certain special structured locus—namely the empty set if $m \geq 7$, and the union of all (linear) vector spaces $L \subseteq C(V)$ over $\mathbb{Q}$ of dimension $\lfloor m/2 \rfloor$ if $m \leq 6$—one should have

$$N_{F,w}(X) = \sum_{x \in \mathbb{Z}^m} w(x/X) \cdot 1_{x \in \bigcup L} = (c + o_{X \to \infty}(1)) \cdot X^{m-3}(\log X)^{r-1+1/m-4}$$

for a certain precise predicted constant $c = c_{C(MP),w} \in \mathbb{R}$ and integer rank $r \geq 1$.

Remark 1.4. If $m \geq 5$, then $r = 1$ always—so the log $X$ disappears—while $c_{C(MP),w}$, the “coned” (or “unsieved”) Manin–Peyre constant, always equals the Hardy–Littlewood constant $c_{HL,w}$. If $m = 4$, then typically, but not always, we have $(r,c_{C(MP),w}) = (1,c_{HL,w})$—if one interprets $c_{HL,w}$ generously as in [Jah14, Chapter II, Remarks 7.5–7.7].

In these cases, “Manin” differs from “Hardy–Littlewood” only in the special part. (But sometimes when $m = 4$, Brauer–Manin obstructions or other phenomena lead to more serious differences; e.g. loosely speaking, it is possible for “$c_{HL,w} \neq 0 = c_{C(MP),w}$” to hold.)

For a general overview of “Manin”, see [Bom09 §2 and references within] or [Bro09,Jah14]. In general, “Manin” is imprecise regarding the special part. However, at least when $m = 6$ and $F = x_1^3 + \cdots + x_6^3$, the specific conjecture recorded above was—in essence—stated first, and precisely, by Hooley [Hoo86a, p. 136, Conjecture 2]. We would like to “naturally detect” the expected special structured locus, in a framework that we now introduce.

1.2. The delta method in a dual form. Strictly speaking, the delta method refers to a smooth version of Kloosterman’s “averaged” circle method, in which one essentially averages over numerators to a given denominator (modulus). In practice, when studying typical quadratic or cubic counting problems, one usually begins (at least morally) in a very specific way by Poisson summation—following the pioneering works [Klo26,HB83,Hoo86b,DFI93,HB96] of Kloosterman, Heath-Brown, Hooley, and Duke–Friedlander–Iwaniec—leading to what could be more precisely called a dual form of the delta method.

1.2.1. The basic setup. There may be other good ways to apply the delta method or Poisson summation—as illustrated by the current record for quartic forms, [MV19]—but we will follow Heath-Brown’s specific cubic development [HB98]. Since [Wan21] uses the same setup, the reader may find some of [Wan21]’s exposition on [DFI93,HB96]’s delta method helpful. In any case, we now set $Y := X^{3/2}$ as on [HB98, p. 676], and recall

$$(1 + O_A(Y^{-A})) \cdot N_{F,w}(X) = Y^{-2} \sum_{n \geq 1} \sum_{c \in \mathbb{Z}^m} n^{-m} S_c(n) I_c(n)$$
Theorem 2, (1.2), up to easy manipulations from §3, where
\[
S_c(n) := \sum_{a \mod n} \sum_{x \mod n} e_n(aF(x) + c \cdot x)
\]
\[
I_c(n) = I_{c,X,Y}(n) := \int_{x \in \mathbb{R}^m} d\mathbf{x} \, w(\mathbf{x}/X) h(n/Y, F(\mathbf{x})/Y^2) e(-c \cdot \mathbf{x}/n).
\]
We will frequently work with the normalized sum \(\tilde{S}_c(n) := n^{-(1+m)/2}S_c(n)\).

Remark 1.5. For random \(c, n\), one heuristically expects \(\tilde{S}_c(n) \ll \epsilon n^\epsilon\)—which can be formalized in a way key to Papers I and III—but key to the present paper is that for certain special \(c\)'s, the truth can easily be a factor of \(\gg_n n^{1/2-\epsilon}\) larger. (Thus whereas I and III center around \(L\)-functions, the present paper centers around algebraic geometry.)

For later use, we now collect some basic facts.

Observation 1.6 (See e.g. [Wan21, §2.1]). The double sum over \(n, c\) is absolutely convergent, so e.g. we may freely switch \(n, c\). In fact, \(I_c(n)\) vanishes over some real range \(n \gtrsim_{F,w,Y} \) independent of \(c \in \mathbb{R}^m\), so we may always assume \(n \lesssim Y\). Meanwhile, up to an additive error of \(O_{F,w,c,A}(X^{-4})\), we may also assume \(\|c\| \leq X^{1/2+\epsilon}\) for any fixed \(\epsilon > 0\) of our choice.

Given \(c \in \mathbb{Z}^m\), the function \(n \mapsto S_c(n)\) is multiplicative, and the behavior of \(S_c\) (at prime powers) is roughly governed by the hyperplane section \(V_c\) of \(V\) corresponding to \(c\). Throughout the present paper, we will use the hyperplane section notation of [Wan21, Definition 1.1]. Thus for instance if \(c \in \mathbb{Z}^m \setminus \{0\}\), then \(V_c := (V_c)_\mathbb{Q}\) is a hypersurface in \((\mathbb{P}^{m-1})_\mathbb{Q} \cong \mathbb{P}^m_\mathbb{Q}\).

1.2.2. Our goal. We seek to interpret the sum
\[
Y^{-2} \sum_{n \geq 1} \sum_{c \in \mathbb{Z}^m} n^{-m} S_c(n) I_c(n) \cdot 1_{c \neq 0} \cdot 1_{V_c},
\]
in terms of special \(\mathbb{Q}\)-subvarieties (specifically, linear spaces) on \(V\). For parity reasons (see e.g. §2), we focus on the simpler case in which \(m\) is even; for several reasons, we further assume \(m \leq 6\). (The case \(m \geq 8\) could perhaps be relevant to “secondary” asymptotic terms, but this is not of immediate concern in Papers I–III.) The case \(m = 6\) is critical for us in view of Hooley’s conjecture (the focus of Paper III).

We will only fully unconditionally achieve our goal when \(F\) is diagonal and \(m = 6\) (or \(m = 4\)—to less precision). But we will still isolate explicit technical ingredients (listed in Remark 1.22) that—if true more generally—would allow for other cases.

To write succinctly, we need some notation, which we now introduce.

1.2.3. Some algebro-geometric definitions and discussion. Up to scaling, there is a unique discriminant form \(F^\vee \in \mathbb{Z}[c]\) of degree \(3 \cdot 2^{m-2}\), characterized by the property that if \(c \in \mathbb{C}^m \setminus \{0\}\), then \(F^\vee(c) = 0\) if and only if \((V_c)_\mathbb{C}\) is singular (see Remark A.3). The equation \(F^\vee(c) = 0\) cuts out the dual variety \(V^\vee\), which lies under the Gauss map \(\gamma: V \rightarrow V^\vee\) defined by \([x] \mapsto [\nabla F(x)]\). (See Appendix A for more details.)

Over any given characteristic zero field, the locus \(F^\vee = 0\) is independent of the choice of \(F^\vee\). But in the present Paper II, it is convenient to fix \(F^\vee\) once and for all (in terms of \(F\)) so that if \(F = F_1 x_1^3 + \cdots + F_m x_m^3\) is diagonal, then \(F^\vee \in (6^m)! F_1 \cdots F_m \cdot \mathbb{Z}[c]\).

It is possible, but unnecessary, to choose \(F^\vee\) to be the same throughout Papers I–II. In the present paper, we do not really use any \(F^\vee\)-dependent facts from Paper I. What is essential in the present paper is that \(F^\vee\) remain fixed in all “non-archimedean” contexts.
1.3. Our main results. Via the delta method, we can give a new and precise reformulation of §1.1’s Hooley–Manin–type conjecture, at least when \( m = 6 \) and \( F \) is diagonal.

Remark 1.7. It would be interesting to extend our analysis below to \( m = 5 \), even just for diagonal \( F \). See [Bom09 §3] for a discussion of the potentially infinite family of lines on a cubic threefold \( V \subseteq \mathbb{P}^3 \). Can one see these lines via \( V' \) (cf. Proposition 2.2)?

First, we state our main theorem, which “naturally isolates” the expected “structured” part for \( m \in \{4, 6\} \):

**Theorem 1.8.** Fix \( F \) diagonal with \( m \in \{4, 6\} \), and fix an arbitrary weight \( w \in C_{\infty}(\mathbb{R}^m) \). If we take \( Y = X^{3/2} \) in the delta method, and restrict to \( c \) with \( F^\vee(c) = 0 \), then

\[
Y^{-2} \sum_{e \in \mathbb{Z}^m} 1_{e \neq 0} \cdot \sum_{n \geq 1} n^{-m} S_e(n) I_c(n) = O_{F,w}(X^{m/2-\delta}) + \sum_{L \subseteq C(V)} \sum_{x \in L \cap \mathbb{Z}^m} w(x/X)
\]

for some absolute constant \( \delta > 0 \), where we restrict to vector spaces \( L/\mathbb{Q} \) with \( \dim(L) = m/2 \).

To avoid any confusion about the “restriction of \( c, L \)” in Theorem 1.8, we now provide an “explicit” definition thereof (which we will maintain for the rest of the present paper):

**Definition 1.9.** We let \( \sum'_{e \in * \cdots} \) denote \( \sum_{e \in * \cdots} F^\vee(e) = 0 \cdots \) (as in [Wan21 §5]), unless specified otherwise. And by default, we let \( L \) denote a vector space over \( \mathbb{Q} \), and let \( \sum_{L \subseteq *} \cdots \) and \( \{L \subseteq * \} \) denote \( \sum_{L \subseteq * \cdots} \) with \( \dim(L) = m/2 \cdots \) and \( \{L \subseteq * : \dim(L) = m/2 \} \), respectively.

**Remark 1.10.** For \( m = 4 \), Theorem 1.8 follows from [HB98, Lemmas 7.2 and 8.1], with a better error of \( O_{F,w,\epsilon}(X^{3/2+\epsilon}) \). As explained in the present §5, the case \( m = 6 \) seems to require a new strategy. But [HB98] does bound the left-hand side by \( O_{F,w,\epsilon}(X^{3+\epsilon}) \) for \( m = 6 \), so we only need “a few new ideas” (in some sense).

**Remark 1.11.** The set \( \{L \subseteq C(V)\} \) is known to be finite for general reasons (recalled in §2 below). For diagonal \( F \) as in Theorem 1.8, one can in fact compute the set \( \{L \subseteq C(V)\} \) explicitly: see Observation 2.9 (essentially classical). In any case, on the Diophantine (right-hand) side of Theorem 1.8, the sum over \( L \) is roughly equivalent to a union, since \( \#\{L \subseteq C(V)\} < \infty \) and the pairwise intersections fit in the error term.

**Remark 1.12.** The \( h \)-invariant introduced by [DL62, DL64] provides an equivalent way to think about linear subvarieties of cubic varieties. (See e.g. [Die17] Lemma 1.1 for a modern record of a more general equivalence.) In fact, the present §4 essentially relies on (a convenient choice of) “\( h \)-decompositions of \( F^\vee \)” corresponding to the \( L \)’s in Theorem 1.8.

Next, for convenience in the proof of Theorem 1.8 and the statement of Corollary 1.14, we make the following definition:

**Definition 1.13.** Given \( L/\mathbb{Q} \) of dimension \( m/2 \), let \( \Lambda := L \cap \mathbb{Z}^m \) denote the unique primitive sublattice of \( \mathbb{Z}^m \) with \( \Lambda \cdot \mathbb{Q} = L \). Then let \( \Lambda^\perp := \{c \in \mathbb{Z}^m : c \perp \Lambda\} \) denote the orthogonal complement of \( \Lambda \) (so in particular, \( \mathbb{Q} \cdot \Lambda^\perp = L^\perp \)) . Now choose bases \( \Lambda, \Lambda^\perp \) of \( \Lambda, \Lambda^\perp \), viewed as \( m \times \frac{1}{2} m \) and \( \frac{1}{2} m \times m \) integer matrices, respectively, so that \( \Lambda = \Lambda \mathbb{Z}^{m/2} \) and \( \Lambda^\perp = \mathbb{Z}^{m/2} \Lambda^\perp \) (where we view \( \Lambda \) as a “column space” and \( \Lambda^\perp \) as a “row space”).

Here \( \Lambda = (\Lambda^\perp)^\perp \), i.e. \( \Lambda = \{x \in \mathbb{Z}^m : \Lambda^\perp x = 0\} \). So given \( L, \Lambda^\perp \), define the density

\[
\sigma_{\infty, L^\perp, w} := \lim_{\epsilon \to 0} (2\epsilon)^{-m/2} \int_{\|\Lambda^\perp x\|_\infty \leq \epsilon} d\bar{x} \, w(\bar{x});
\]
for a technically convenient alternative definition, see the second part of Lemma 3.10 in §3. Then in Theorem 1.8, we can reinterpret the sum over \(L \cap \mathbb{Z}^m = \Lambda\) as follows:

\[
\sum_{x \in \Lambda} w(x/X) = \sigma_{\infty,L+w,X^{m/2}} + O_{L,w,A}(X^{-A}),
\]

by Poisson summation over \(\Lambda\) (or, at least morally, by the circle method applied to \(\Lambda \bot x = 0\)). In particular, \(\sigma_{\infty,L+w}\) does not depend on the choice of \(\Lambda^\bot\).

We now arrive at the main result here (in the present Paper II) needed for Paper III:

**Corollary 1.14.** If \(m = 6\) in the setting of Theorem 1.8, then

\[
Y^{-2} \sum_{c \in \mathbb{Z}^m} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n) = O_{F,w}(X^{m/2-\delta}) + \sigma_{\infty,F,w,\mathcal{S}_F} X^{m-3} + \sum_{L \subseteq C(V)} \sigma_{\infty,L+w,X^{m/2}}
\]

for some absolute \(\delta > 0\), where \(\sigma_{\infty,F,w,\mathcal{S}_F}\) are as specified in Appendix 3.

**Proof.** Combine Theorem 1.8 with the routine \(c = 0\) computation in Appendix 3 (isolating the singular series). In the exponents, note that \(m/2 = m - 3 = 3\). \(\square\)

**Remark 1.15.** Theorem 1.8 and Corollary 1.14 are completely unconditional results. When \(m = 6\) and \(F\) is diagonal, these results let us reformulate 1.1’s Hooley–Manin–type conjecture as a statement purely about cancellation in the sum \(\sum_{c \in \mathbb{Z}^m} 1_{F^\vee(c) \neq 0} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n);\) see Paper III. With any luck, a similar reformulation might be possible much more generally. But at least when \(m = 4\), subtleties in the Manin–Peyre constant would likely demand a more sophisticated “delta method recipe” beyond “restriction to \(F^\vee = 0\).”

1.3.1. A proof sketch for the theorem. (In this sketch, we restrict \(L\) to \(\{L \subseteq C(V)\}'\).)

We start generally, observing that \(F^\vee|_{L^\perp} = 0\) for all \(L\) (see the first part of Proposition 2.2). Conversely, at least for diagonal \(F\), most \(c\)'s on the left-hand side of Theorem 1.8 are in fact trivial, in the sense of the following definition:

**Definition 1.16.** Call a solution \(c \in \mathbb{Z}^m\) to \(F^\vee(c) = 0\) trivial if \(c \in L^\perp\) for some \(L\).

Actually, we cannot analyze all trivial \(c\)'s uniformly, but only the “least degenerate” ones. Under (plausibly mild) hypotheses satisfied by diagonal \(F\), the second part of Proposition 2.2 establishes a vanishing baseline for the jets \(j^*F^\vee\) over \(\bigcup_L L^\perp\) which inspires the following definition:

**Definition 1.17.** Call \(F^\vee\) unsurprising if uniformly over \(C \geq 1\), the equation \(F^\vee(c) = 0\) has at most \(O((C^{m/2-1+\epsilon})\) integral solutions \(c \in [-C,C]^m\) that are either nontrivial, or else trivial with \(j^{2m/2-1} F^\vee(c) = 0\).

Indeed, in §5 we will prove that if \(F\) is diagonal, then \(F^\vee\) is unsurprising—in which case “almost all solutions to \(F^\vee = 0\) are trivial with nonzero \(2^{m/2-1}\) jet” (qualitatively speaking).

**Remark 1.18.** In particular, \(\gamma(Q) : V(Q) \rightarrow V^\vee(Q)\) can be far from surjective, even though \(\gamma : V \rightarrow V^\vee\) itself is birational and roughly log-height–preserving (up to a constant factor).

**Remark 1.19.** A weaker bound of the form \(O(C^{m/2-\delta})\) would likely suffice for our main purposes (though for \(m = 4\), a full “rigorous heuristic” delta method discussion of Manin’s conjecture would require a deeper analysis even beyond \(m/2 - 1 + \epsilon\)).
For the “least degenerate” trivial $c$’s, Lemma 5.5 isolates an explicit positive bias
\[ \tilde{S}_c(p) = [A_p(c) + O(p^{-1/2})] \cdot (1 - p^{-1}) \cdot p^{1/2} \]
for most primes $p$, with $A_p(c) \lesssim 1$ essentially a combinatorial factor measuring the $p$-adic “extent of speciality” of $c$; on average vertically, $E[A_p(c)] \approx 1$. In the dominant ranges (i.e. for $n$ large), the resulting reduction in arithmetic complexity of $S_c(n)$ lets us dramatically simplify each sum of the form $\sum_{c \in \Lambda^+} S_c(n)I_c(n)$ by “undoing” Poisson summation to $I_c(n)$ over various individual residue classes $c \equiv b \mod n_0\Lambda^\perp$ with $n_0 \lesssim X^{1/2}$ dividing $n$.

Ultimately, we get corresponding sums over $x \in \Lambda$ (as desired for Theorem 1.8). However, when $m = 6$, we must be careful to separate $c = 0$ from $c \neq 0$; Lemma B.1 (decay of the singular series over large moduli) provides the necessary input, when contrasted with Heath-Brown’s bounds giving decay of $I_c(n)$ over small moduli when $c \neq 0$.

For a full cross-referenced outline of the proof of Theorem 1.8 see §5.

Remark 1.20. We do not need horizontal cancellation over $n$: at least morally, the terms $S_c(n), I_c(n)$ are positive for trivial $c$’s, while nontrivial $c$’s are relatively sparse. This morally explains why we can prove Theorem 1.8 unconditionally. In fact, the deepest result we use (when $m = 6$) is the Weil bound for hyperelliptic curves of genus $\leq 2$.

Remark 1.21. The full proof of Theorem 1.8 requires an error analysis to “reduce consideration” to biases. Fortunately, several convenient features make our error analysis here (reducing consideration to biases over $F^\vee(c) = 0$) “half an inch” easier, or clearer, than that in Paper III (reducing to $L$-functions over $F^\vee(c) \neq 0$). Specifically, small moduli $n$ and bad primes $p$ here cause very little pain compared to those in Paper III.

Remark 1.22. We can also “axiomatize” our proof of Theorem 1.8 in the hope of generalizing Theorem 1.8 to non-diagonal $F$ (though a full diagnosis of the relevant issues would seem to require algebraic geometry beyond the author’s current expertise). Assume

1. $m \geq 4$ is even, $F$ is smooth, and $F^\vee$ is unsurprising (in the sense of Definition 1.17);
2. $w$ is supported away from the Hessian vanishing locus (hess $F$)\(\mathbb{R}\);
3. a version of Lemma 4.1 with the same hypotheses but with a modified conclusion

\[ Y^{-2} \sum_{c \in S_\{0\}} \sum_{n \geq 1} n^{(1-m)/2} |\tilde{S}_c(n)| \cdot X^{m} \max(1, X\|c\|/n)\frac{1-m/2}{1-m/2}\nu_1(\|c\|/X^{1/2}) \lesssim X^{(m-\delta)/2+\epsilon}, \]

remains true, where $\nu_1(\star)$ is a fixed function decaying as $O_A(\max(1, \star)^{-4})$; and
4. in Lemma 5.5 the formula for $S_c(p)$ and upper bound for $S_c(p^{\geq 2})$ remain true, provided $p \gtrsim_{3,m,F} 1$ exceeds some threshold (only allowed to depend on $3, m, F$).

Then the conclusion of Theorem 1.8 still holds for $F$, as does that of Corollary 1.14 if $m \geq 6$. (Depending on the precise “Definition 1.17” in (1), a weaker version of (3) may suffice.)

Remark 1.23. In Remark 1.22 we expect that (4) should be relatively routine to prove (if true), but (1) and (3) may well require substantial new ideas.

1.4. A paper outline. §2 suggests a geometric backbone for the eventual harmonic detection of $\{L \subseteq C(V)\}'$ from the locus $F^\vee(c) = 0$ in the delta method. §3 collects some easy “reverse duality” identities to be used near the end (essentially) of §5 to obtain the expected “structured” main terms. §4 provides a convenient “sparse contribution bound” (essentially due to [HB98]) to be used as an error bound at several points in §5.
Using Lemma B.1 and various technical results from HB98 and §§2–4 as a black box, §5 then reduces the proof of Theorem 1.8 to asymptotic formulas for certain exponential sums (see Lemma 5.5) proven later in §6.

Appendix A collects some classical algebraic geometry—primarily for those interested in contemplating non-diagonal $F$’s—used as a black box in parts of §2.

Finally, Appendix B analyzes the $c = 0$ contribution in the delta method—extracting the (adelic) singular series when $m \geq 5$—while also proving Lemma B.1.

Remark 1.24. Appendices A–B, §3, and possibly §6 are rather general, but §§4–5 (and maybe §6) are still “diagonally crutched” in technical but serious ways (detailed in Remark 1.22). So to obtain a satisfying geometric picture even in the “$c$-restricted” sense of Theorem 1.8, there is certainly much left to be done, even in 6 variables.

Before proceeding, we raise some further questions (about “more complicated” varieties) that one also might (at first) study in the “$c$-restricted” sense of Theorem 1.8.

1.5. Further phenomena and questions. Consider the following example (from a talk of Wooley) of a situation in which (one expects that) nonlinear special subvarieties arise.  

Example 1.25 ([Woo19]). Over boxes $[-X, X]^6$ as $X \to \infty$, one expects the 6-variable quartic $x_1^4 + x_2^4 + x_3^4 = x_4^4 + x_5^4 + x_6^4$ to have not only a “purely probabilistic” source of $\asymp X^2$ points (as $X \to \infty$), but also at least two relevant “special” sources of points:

1. the “trivial” or “diagonal-type” linear locus

$$x_1 \pm x_4 = x_2 \pm x_5 = x_3 \pm x_6 = 0$$

contributing $\asymp X^3$ points (as $X \to \infty$), as well as

2. a secondary quadratic locus

$$x_1 \pm x_2 \pm x_3 = x_4 \pm x_5 \pm x_6 = (x_1^2 + x_2^2 + x_3^2) \pm (x_4^2 + x_5^2 + x_6^2) = 0$$

contributing $\asymp X^2 \log X$ points (as $X \to \infty$).

(The underlying identity behind (2) is $a^4 + b^4 + (a + b)^4 = 2(a^2 + ab + b^2)^2$.)

Question 1.26. Can one detect (2) naturally via the delta method? Where might the quadratic aspect naturally arise?

Remark 1.27. Less formally but more generally, it is natural to ask if the delta method can serve as a “predictive telescope” into the asymptotic behavior of points on Fano hypersurfaces, even when a fully rigorous analysis lies out of reach. In particular, can one always detect “thin but special” sets naturally via the delta method?

(For general Fano varieties, the delta method could still be relevant at least sometimes, but for simplicity we have restricted attention to Fano hypersurfaces.)

2. Maximal linear subvarieties under duality

Fix a smooth cubic $V/\mathbb{Q}$ as in §1. Now use Definition 1.9 (which makes sense in general) to define the set $\{L \subseteq C(V)\}'$ in general. The reader only interested in diagonal $F$ can skim forwards to §2.2 which explicitly analyzes $\{L \subseteq C(V)\}'$ through the lens of $F^\vee$. 
2.1. A preliminary general analysis. In general, classical duality theory comes into play, leading to Proposition 2.2 below. For the definitions and basic properties of the polar and Gauss maps $[\nabla F]$, $\gamma$ associated to $V$, see the first few paragraphs of Appendix [A].

Now, if $m \geq 3$ and $L \subseteq C(V)$ is a vector space over $\mathbb{Q}$ (of arbitrary dimension), then differentiating along $L$ implies $L \perp \nabla F(x)$ for all $x \in L$. In other words, the restriction $\gamma|_{PL}$ maps into $\mathbb{P}L^\perp$. But $[\nabla F]: \mathbb{P}^{m-1} \to \mathbb{P}^{m-1}$ is regular and finite, so $\gamma|_{PL}$ is itself regular and finite (since $\mathbb{P}L$ is closed in $\mathbb{P}^{m-1}$). Thus dim($L$) $\leq \dim(L^\perp)$, i.e. dim($L$) $\leq \lfloor m/2 \rfloor$.

Remark 2.1. Although $\mathbb{P}L^\perp \subseteq (\mathbb{P}^{m-1})^\vee$ is the dual variety of $\mathbb{P}L \subseteq \mathbb{P}^{m-1}$, we prefer to write $L^\perp$ instead of $L^\vee$, to avoid confusion with the dual vector space $\text{Hom}(L, \mathbb{Q})$.

Since $\deg F \geq 3$, [Deb03, Lemma 3] (or Starr [BHB06, Appendix]) proves more: if $m$ is even, then there are at most finitely many $L$'s of dimension $m/2$. We would like to understand these “maximal” linear spaces $L$ in terms of the delta method. Proposition 2.2 suggests one plausible starting route: duality (i.e. detecting $L^\perp$ through $F^\vee$).

**Proposition 2.2.** Suppose $2 \mid m \geq 4$, and fix $L \subseteq \{L \subseteq C(V)\} \vee$. Then $\gamma|_{PL}$ is a finite flat surjective morphism $\mathbb{P}L \to \mathbb{P}L^\perp$ of degree $2^{m/2-1} \geq 2$, and $\mathbb{P}L^\perp \subseteq \text{Sing}(V^\vee)$. Furthermore, if $\mathbb{P}L^\perp \not\subseteq [\nabla F](\text{hess}(V))$ as topological spaces, then the affine jet $j^{2^{m/2-1}-1}F^\vee$ vanishes over $L^\perp$.

**Remark 2.3.** For diagonal $F$, we provide an explicit proof of Proposition 2.2 in §2.2. For non-diagonal smooth $F$, we instead rely on some classical algebraic geometry—duality theory and the ramification behavior of $[\nabla F]$—detailed in Appendix [A].

We now begin the proof of Proposition 2.2. In general, $\gamma|_{PL}$ is a finite map from $\mathbb{P}L$ to $\mathbb{P}L^\perp$. Now assume $\dim(L) = \dim(L^\perp)$. Then $\gamma|_{PL}$ is surjective, since $\mathbb{P}L^\perp$ is irreducible. But $\gamma$ maps $V$ into $V^\vee$ by definition, so $\mathbb{P}L^\perp \subseteq \text{im} \gamma|_{\mathbb{P}L} \subseteq \text{im} \gamma \subseteq V^\vee$.

**Proof of first part.** Since $\gamma|_{\mathbb{P}L}$ is finite surjective and $\mathbb{P}L, \mathbb{P}L^\perp$ are smooth, “miracle flatness” implies flatness of $\gamma|_{\mathbb{P}L}$. Also, $\gamma|_{PL}$ has degree $2^{m/2-1}$ (cf. [Dol12, top of p. 29]), since it is a morphism given by quadratic polynomials, between projective spaces of dimension $m/2 - 1$. In particular, $\gamma|_{\mathbb{P}L}: \mathbb{P}L \to \mathbb{P}L^\perp$ is nowhere birational, so the biduality theorem implies $\mathbb{P}L^\perp \subseteq \text{Sing}(V^\vee)$.

The second part of Proposition 2.2 is inspired by the factorization of $F^\vee$ over $\mathbb{Q}[c^{1/2}]$ when $F$ is diagonal (see §2.2). However, giving a rigorous “factorization” of $F^\vee$ seems to require a bit of setup, since the map $[\nabla F]$ presumably need not be Galois in general.

First, assume $\mathbb{P}L^\perp \not\subseteq [\nabla F](\text{hess}(V)) = \text{St}(V)^\vee$. Now consider the hypersurface complements $S := \mathbb{P}^{m-1} \setminus \text{St}(V)^\vee$ and $X := [\nabla F]^{-1}S \subseteq \mathbb{P}^{m-1}$. Then $S \cap \mathbb{P}L^\perp$ is a nonempty open subset of $\mathbb{P}L^\perp$, yet $[\nabla F]|_X: X \to S$ is finite étale of degree $2^{m/2-1}$. Write $\phi := [\nabla F]|_X$.

By Grothendieck’s Galois theory, there exists a finite étale Galois cover $\pi: X' \to X$ with $X'$ connected and $\phi \circ \pi: X' \to S$ (finite étale) Galois. Let $G := \text{Aut}_S(X')$ and $H := \text{Aut}_X(X')$.

2.1.1. Constructing a product “divisible” by $F^\vee$. Over every geometric point $[c] \in (S \cap V^\vee)(\overline{\mathbb{Q}})$, there exists a (geometric) point $[x] \in X_{[c]} \cap V$, i.e. $[x] \in X_{[c]}$ with $F(x) = 0$. Although $G$ may not act on $X$ itself, it acts transitively on $X'$—so after fixing a point $p \in \pi^{-1}[x]$, we can characterize the fiber $X_{[c]}$ as the set $\{\pi(gp) : g \in H\setminus G\} \subseteq X$.

Now view $F(x)$ as a section of $\mathcal{O}_X(3)$, and pull it back to $\pi^*F \in \Gamma(X', \pi^*\mathcal{O}_X(3))$. Then the product $\alpha := \prod_{g \in H\setminus G} g^*(\pi^*F)$ defines a $G$-invariant section\[^1\] of the $G$-equivariant line

\[^1\]A $G$-invariant section $\alpha \in \Gamma(X', \mathcal{L})$ is equivalent to data in a $G$-equivariant morphism $\alpha: \mathcal{O}_{X'} \to \mathcal{L}$. 
bundle $\mathcal{L} := \bigotimes_{g \in H \setminus G} g^*(\pi^* \mathcal{O}_X(3))$ on $X'$, with
\[
\alpha_{|\phi^{-1}(S\cap V')} = 0 \quad \text{(since } X \cap V \subseteq \phi^{-1}(S \cap V'), \text{ and } F|_{X \cap V} = 0) .
\]
By faithfully flat (Galois) descent\(^2\) there exist line bundles $\mathcal{F}, \mathcal{D}$ on $X, S$ with $\mathcal{L} \cong \pi^* \mathcal{F}$ and $\mathcal{F} \cong \phi^* \mathcal{D}$ and sections $\beta, \delta$ on $X, S$ vanishing along $\phi^{-1}(S \cap V'), S \cap V'$, respectively, with $\alpha = \pi^* \beta$ and $\beta = \phi^* \delta$.

But $S, X$ are hypersurface complements in $\mathbb{P}^{m-1}$, so Pic($\mathbb{P}^{m-1}$) $\to$ Pic($S$) and Pic($\mathbb{P}^{m-1}$) $\to$ Pic($X$) are surjective and we may identify $\mathcal{F}, \mathcal{D}$ with suitable powers of $\mathcal{O}_X(1), \mathcal{O}_S(1)$, respectively. Then up to a choice of nonzero constant factors, we may view $\beta, \delta$ as homogeneous rational functions (i.e. ratios of homogeneous $m$-variable polynomials) with $F^\vee(c) \mid \delta$ and $F^\vee(\nabla F(x)) = \phi^* F^\vee \mid \beta$. Here we interpret divisibility of two sections on a scheme to mean their ratio is a global section of the obvious “tensor-quotient” line bundle.

2.1.2. “Factoring” $F^\vee$. By duality theory, $F(x) \mid F^\vee(\nabla F(x))$. However, $F(x)^2 \nmid \alpha$ on $X$, since $(\pi^* F)^2 \nmid \alpha$ on $X'$. Indeed, given a geometric point $[x] \in X_{[c]}$ lying in $V$. So if $p \in \pi^{-1}[x]$ and $g \in G$, then the section $g^* \pi^* F$ evaluates to 0 at $p$ if and only if $g \in H$. Thus $\pi^* F \mid \prod_{[1]} \mathcal{L} \cap \mathcal{G}(g^* \pi^* F)$ on $X'$. It follows that $(\pi^* F)^2 \mid \alpha$, whence $F(x)^2 \mid \beta$; whence $F^\vee(c)^2 \nmid \delta$.

However, $F \mid \phi^* F^\vee$, and $\phi^* g = \phi \pi$ for all $g \in G$, so by inspection, $\alpha \mid (\pi^* \phi^* F)^\vee \pi \phi^* \delta$, i.e. $\delta \mid (F^\vee)^\pi \phi^* \delta$. By absolute irreducibility of $F^\vee$, we conclude that in fact $\delta \mid F^\vee$ and $\alpha \mid (\phi \pi)^* F^\vee$. (Divisibility also holds in the other direction, but we will not need this.)

**Remark 2.4.** Our proof of $\delta \mid F^\vee$ requires $(S \cap V') \setminus \text{Sing}(V') \neq \emptyset$, i.e. that $S \cap V'$ and $V' \setminus \text{Sing}(V')$ are nonempty (open) subsets of $V'$; since $\mathbb{P}L^\perp \subseteq V'$, the latter nonemptiness follows conveniently from our assumption $S \cap \mathbb{P}L^\perp \neq \emptyset$ (but see Question A.9), while the latter follows (unconditionally) from “generic smoothness” in characteristic 0.

2.1.3. Differentiating the product. Using $S \cap \mathbb{P}L^\perp \neq \emptyset$ one last time (more seriously than before), we will now complete the proof of the second part of Proposition 2.2.

**Completion of proof.** By assumption, $S \cap \mathbb{P}L^\perp \neq \emptyset$. Yet $\phi|_{X \cap \mathbb{P}L} = \gamma|_{X \cap \mathbb{P}L} : X \cap \mathbb{P}L \to S \cap \mathbb{P}L^\perp$ is finite étale of degree $2^m/2 - 1$. Say $[c] \in (S \cap \mathbb{P}L^\perp)(\overline{Q})$ is a geometric point, and fix $p \in X_{[c]}$. Then there exist at least $2^m/2 - 1$ cosets $g \in H \setminus G$ with $\pi gp \in X \cap \mathbb{P}L \subseteq V$. In a $G$-invariant affine open neighborhood of $p$ in $X'$, the Leibniz rule—applied after locally trivializing $g^* \pi^* \mathcal{O}_X(3)$ for all $g \in G$—thus implies that $j_p^\beta \alpha(p) = 0$ for $r := 2^m/2 - 1$, where $j^r : \mathcal{L} \to j^r \mathcal{L}$ denotes the $r$th-order jet map “along” $\mathcal{L}$ (from $\mathcal{L}$ to its $r$th jet bundle $j^r \mathcal{L}$).

Since $\alpha \mid (\phi \pi)^* F^\vee$, Leibniz then implies $j_p^\beta (\phi \pi)^* F^\vee(p) = 0$ “along” the pullback line bundle $(\phi \pi)^* \mathcal{O}_S(\deg F^\vee)$. But $\phi : X' \to S$ is étale at $p \in X'$, so $j_p^\beta (\phi \pi)^* F^\vee([c]) = 0$ “along” $\mathcal{O}_S(\deg F^\vee)$ itself, over all points $[c] \in (S \cap \mathbb{P}L^\perp)(\overline{Q})$. Finally, $S \cap \mathbb{P}L^\perp$ is dense in $\mathbb{P}L^\perp$, so the vanishing of the $r$th-order jet section $j^r \mathcal{F}^\vee$ extends to all points $[c] \in \mathbb{P}L^\perp$, as desired. \hfill $\Box$

**Remark 2.5.** In the friendly setting above, our étale morphisms (such as $\phi \pi : X' \to S$), after base change to an algebraically closed field, always induce isomorphisms on completed local rings. So e.g. at regular points we can do calculus purely in terms of formal power series (in general by the Cohen structure theorem, but for us $S$ is already given as a piece of $\mathbb{P}^{m-1}$).

---

\(^1\)In the form of an equivalence of categories

\(^2\)It suffices to find $\mathcal{D}$ with $\mathcal{L} \cong (\phi \pi)^* \mathcal{D}$, and then define $\mathcal{F} := \phi^* \mathcal{D}$. 
Question 2.6. Is $PL^1 \subseteq \{\nabla F|\hess(V)\}$ possible (assuming $V/Q$ is smooth and $\dim(L) = m/2$)? If so, then in such situations, does the conclusion of Proposition 2.2 still hold?

2.2. The diagonal case. Say $F$ is diagonal, and write $F = F_1x_1^3 + \cdots + F_mx_m^3$. Then we can explicitly verify all the theory above. Here $[\nabla F]: [x] \mapsto [3F_1x_1^2, \ldots, 3F_mx_m^2]$ is Galois with abelian Galois group $\mu^m_2/\mu_2 \cong (\Z/2)^{m-1}$, and $\hess(V)$ is cut out by $(6F_1x_1) \cdots (6F_mx_m) = 0$.

2.2.1. Describing $\{L \subseteq C(V)\}'$. It is known (see e.g. a claim of Starr [BHB06, top of p. 302] on Fermat hypersurfaces of degree $d \geq 3$ in $s \in \{4, 6, 8, \ldots\}$ variables) that since $F$ is diagonal, the number of $m/2$-dimensional vector spaces $L \subseteq C(V)_C$ over $C$ is $(\deg F)^{m/2} = 3^{m/2}$ times $C_{m/2-1} := (m-1)!!$, the number of pairings of $[m]$. So we make a combinatorial definition:

Definition 2.7. Let $J = (J(k))_{k \geq 1}$ denote an ordered partition of $[m]$. We identify $J, J'$ as the same if $\{J(k) : k\} = \{J'(k) : k\}$ as sets. For convenience, we always assume $|J(1)| \geq |J(2)| \geq \cdots$, and restrict attention to indices $k$ with $J(k) \neq \emptyset$.

Call $J$ a pairing if $\forall k, |J(k)| = 2$. Call $J$ permissible if $i, j \in J(k) \implies F_j/F_i \in (Q^\times)^3$.

For a permissible $J$, let $\mathcal{R}_J := \{c \in \Z^m : i, j \in J(k) \implies c_i/F_i^{1/3} = c_j/F_j^{1/3}\}$—and given $c \in \mathcal{R}_J$, define $c: \{k\} \to \R$ so that $\forall k, \forall i \in J(k)$, $c_i/F_i^{1/3} = c(k)$.

Knowing the size of $\{L \subseteq C(V)_C\}'$, we can then give an exhaustive construction: each pairing $J$ yields $3^{m/2}$ distinct $L/C$, obtained by setting $F_ix_i^3 + F_jx_j^3 = 0$ for each part $J(k) = \{i, j\}$. (Over $Q$, we must set $x_i + (F_j/F_i)^{1/3}x_j = 0$—which is valid when $F_i \equiv F_j$ mod $(Q^\times)^3$.)

In other words, the only $m/2$-dimensional $L$’s on $C(V)_C, C(V)$ (over $C, Q$, respectively) are the “obvious” ones.

Remark 2.8. Since we do not know of a reference proving the aforementioned claim of Starr, we should mention that given $d, s$, the claim easily follows from Gaussian elimination, symmetry, and the fact that over $C$, the only linear automorphisms of the “halved” (i.e. $s/2$-variable degree-$d$) Fermat hypersurface are the “obvious” ones (see e.g. [Shi88] or [Kon02] proofs of Proposition 3.1 and Example 1]).

Therefore we obtain the following (essentially classical) result:

Observation 2.9. There is a canonical bijection, between $\{L \subseteq C(V)\}'$ and the set of permissible pairings $J$, characterized by $L \cap \Z^m = \mathcal{R}_J^\perp$ (an equality of sublattices of $\Z^m$).

2.2.2. Analyzing the discriminant. For convenience, fix square roots $F_i^{1/2} \in \overline{Q}^\times$. Up to scaling (which matters in the non-archimedean analysis of Lemma 5.5 but not here), the discriminant form $F^\vee(c)$ factors in $\overline{Q}[c^{1/2}]$ as

$$
\prod_{\epsilon} (\epsilon_1 F_1^{-1/2} c_1^{3/2} + \epsilon_2 F_2^{-1/2} c_2^{3/2} + \cdots + \epsilon_m F_m^{-1/2} c_m^{3/2}) \in Q[c],
$$

with the product taken over $\epsilon$ with $\epsilon_1 = 1$ and $\epsilon_2, \ldots, \epsilon_m = \pm 1$. (This formula is classical; see [Wan21] §1.2, proof of Proposition-Definition 1.8 for diagonal $F$.)

Now fix a tuple $c \neq 0$ with $F^\vee(c) = 0$, and fix square roots $c_i^{1/2} \in \overline{Q}$. Then using formal power series calculus over $c_i \neq 0$ (by Remark 2.5 applied to $A_1^1 \to A_1^1, t \mapsto t^2$ away from the origin), we will prove the following result, which precisely characterizes the order of vanishing of $F^\vee$ at $c$:

Proposition 2.10. Assume $F$ is diagonal. Fix $r \geq 0$. Then the affine jet $j^r F^\vee$ vanishes at a given point $c \neq 0$ if and only if there exist at least $r + 1$ distinct $\epsilon$ with $\langle \cdots \rangle = 0$. 

Remark 2.11. A short computation yields a geometric interpretation of the number of $\epsilon$'s:
\[
\# \{\text{such distinct } \epsilon \text{'s} \} = \sum_{[x] \in \gamma(\mathbb{Q})^{-1}([\epsilon])} 2^{\# \{i \in [m] : x_i = 0 \}},
\]
where $\gamma(\mathbb{Q})^{-1}([\epsilon]) := \{[x] \in V(\mathbb{Q}) : [\nabla F(x)] = [\epsilon] \} = \{\text{singular } \mathbb{Q}\text{-points of } V_\epsilon \}$. Here $x_i$ corresponds to $\epsilon_j F_i^{-1/2} C_i^{1/2}$, with some ambiguity (multiplicity) in $\epsilon_i$ when $x_i = 0$.

Thus we can formulate Proposition 2.10 more geometrically, without reference to $\epsilon$'s. Does such a geometric formulation generalize somehow to the case of non-diagonal smooth $F$?

Proof. Induct on $r \geq 0$. The base case $r = 0$ follows directly from the factorization of $F^\vee$. Now suppose $r \geq 1$, and assume the result for $r - 1$.

First, we prove the forwards implication for $r$. Here it suffices to work with “pure” derivatives $\partial_{c_i}^r$, for just a single index $i$ with $c_i \neq 0$. For example, if $c_i \neq 0$, and there exist exactly $r$ distinct $\epsilon_1, \ldots, \epsilon_r$ with $(\cdots) = 0$, then $r \leq 2^{m-1}$, and the product rule implies
\[
\partial_{c_i}^r F^\vee(c) \propto_{3,r,F} (c_1^{1/2})^r \prod_{\epsilon \neq \epsilon_1, \ldots, \epsilon_r} (\cdots) \neq 0.
\]
But if $j^r F^\vee(c) = 0$, then by the inductive hypothesis, there must exist at least $r$ distinct $\epsilon$’s with $(\cdots) = 0$, and thus at least $r + 1$. This proves the forwards implication for $r$.

It remains to prove the backwards implication, i.e. that if there exist at least $r + 1$ distinct $\epsilon$’s with $(\cdots) = 0$, then $j^r F^\vee(c) = 0$. We must take extra care if $c_1 \cdots c_m = 0$. Say $c_i = 0$ for $i \in I$. Then the following “formal analytic functions”—indexed by certain triples $(a, b, E)$—span a $\mathbb{Q}$-vector space closed under $\mathbf{c}$-differentiation:
\[
\prod_{i \in I} c_i^{a_i} \prod_{i \notin I} c_i^{b_i/2} \prod_{\epsilon \in E} (\epsilon_1 F_1^{-1/2} C_1^{3/2} + \cdots + \epsilon_m F_m^{-1/2} C_m^{3/2}) \in \mathbb{Q}^{|I| + |\epsilon|} [c_i, c_i^{-1}]
\]
for $(a, b) \in \mathbb{Z}_{\geq 0}^{|I|} \times \mathbb{Z}^{[m]|I|}$ and $E \subseteq \{\epsilon \in \{-1, 1\}^m : \epsilon_1 = 1\}$ with $E$ mod $\pm 1$ invariant under flipping $\epsilon_i$ for $i \in I$.

Specifically, differentiating in $c_i$ leads to terms with $a_i \mapsto a_i - 1$ or $(a_i, |E|) \mapsto (a_i + 2, |E| - 2)$ if $i \in I$, and to terms with $b_i \mapsto b_i - 2$ or $(b_i, |E|) \mapsto (b_i + 1, |E| - 1)$ if $i \notin I$. In each case, applying $\partial_{c_i}$ decreases $\min_{a,b,E} (|a| + |E|)$ by at most 1.

Now fix $r \in \mathbb{Z}_{\geq 0}^m$ with $|r| \leq r$. Then $\partial_{c}^r F^\vee$ is a $\mathbb{Q}$-linear combination of functions indexed by triples $(a, b, E)$ with $|a| + |E| \geq 2^{m-1} - r$ (and thus $|E| \geq 2^{m-1} - r$ or $|a| \geq 1$). Each such function must vanish at our original given point $c$, so $\partial_{c}^r F^\vee = 0$, as desired.

2.2.3. Evaluating $F^\vee$ on $L^\perp$ for special $L$’s. Fix $L$ in $\{L \subset C(V)\}'$. Then first, using Proposition 2.10, we can explicitly verify the conclusion of Proposition 2.2.

Corollary 2.12. For $L$ as above, we have $(j^{2^{m-2} - 1} F^\vee)|_{L^\perp} = 0$.

Proof. $L$ corresponds to some pairing $J$. For each part $J(k) = \{i, j\}$, there are exactly two choices of signs $(\epsilon_i, \epsilon_j) \in \{ \pm 1 \}^2$—or only one choice if 1 $\in J(k)$—such that $\epsilon_i F_i^{-1/2} C_i^{3/2} + \epsilon_j F_j^{-1/2} C_j^{3/2}$ vanishes over all $c \in L^\perp \cap \mathbb{Z}^m = R_{J}$ lying in a given orthant of $\mathbb{R}^m$. So given $c \in L^\perp \setminus \{0\}$, one can apply Proposition 2.10 “backwards” with $r := 2^{m/2} - 1$. (Of course, $L^\perp \setminus \{0\}$ is dense in $L^\perp$, so the vanishing then extends to all of $L^\perp$.)

The next result shows that in fact, $F^\vee$ generally does not vanish to higher order along $L^\perp$. 

Observation 2.13. Given $L$ as above, fix $c \in L^\perp \cap \mathbb{Z}^m = \mathcal{R}_\mathcal{J}$. For each $k$, choose a square root $c(k)^{3/2} := F_1^{-1/2} c_i^{3/2}$ in $\mathcal{Q}$ (say). Then $j^{2m/2-1} F^\vee(c) = 0$ if and only if there exist $l \geq 1$ distinct indices $k_1, \ldots, k_l$ such that $c(k_1)^{3/2} \pm \cdots \pm c(k_l)^{3/2} = 0$ for some choice of signs. Consequently, if $j^{2m/2-1} F^\vee(c) = 0$, then $c(k_j)^3 c(k_j)^3 \in (\mathbb{Q}^\times)^2 \cup \{0\}$ for some distinct $k_1, k_2$.

Proof. For the equivalence, apply Proposition 2.10 “forwards” with $r := 2^{m/2-1}$ (and then “simplify” the resulting conclusion using the fact that $\mathcal{J}$ is a pairing). To obtain the final conclusion, note that the condition $c(k_1)^{3/2} \pm \cdots \pm c(k_l)^{3/2} = 0$ implies the following, provided $l$ is minimal among all possible $l$’s (as we may certainly assume):

1. If $l = 1$, then $c(k)^3 = 0$ for some $k$.
2. If $l = 2$ is minimal, then $c(k_1)^3 = c(k_2)^3 \in \mathbb{Q}^\times$ for some distinct $k_1, k_2$.
3. If $l \geq 3$ is minimal, then $c(k_j)^3 \in \mathbb{Q}^\times$ for all $t \in [l]$, and by multi-quadratic field theory in characteristic 0, the square classes $c(k_1)^3, \ldots, c(k_l)^3 \mod (\mathbb{Q}^\times)^2$ must all coincide. (More precisely, given indices $i_t \in \mathcal{J}(k_t)$ for $t \in [l]$, we must have $c(k_i)^3 = F_i^2 x_i^3 d^3 \in d \cdot (\mathbb{Q}^\times)^2$ for some $d, x_i \in \mathbb{Q}^\times$ such that $F_i x_i^3 + \cdots + F_i x_i^3 = 0$. If $\mathcal{J}(k_1), \ldots, \mathcal{J}(k_l)$ cover $[m]$, as must be the case if $m = 6$, then this would imply that $[c] \in \mathbb{P}L^\perp$ actually lies in the image of $\gamma(\mathbb{Q})$—unlike most points of $\mathbb{P}L^\perp$.)

In each case, the claimed multiplicative relationship exists for some distinct $k_1, k_2$. \hfill \square

Remark 2.14. (2.2) is written in characteristic 0, but since $F^\vee \in \mathbb{Z}[c]$ (by §1.2.3’s conventions), the results, with their algebraic proofs, carry over to arbitrary fields of characteristic $p \not| (6^m)!F_1 \cdots F_m$. Over $\mathbb{F}_p$, such extensions of Proposition 2.10 (in its geometric formulation), Corollary 2.12, and (the equivalence part of) Observation 2.13 will prove useful in §6.

(Though unimportant for us over $\mathbb{F}_p$, the other results of §2.2 also carry over. For instance, regarding the field theory behind the last part of Observation 2.13 if $K$ is a field of characteristic $p \not| 2$, and $d_1, \ldots, d_l \in K^\times$ are pairwise incongruent modulo $(K^\times)^2$, then $\sqrt{d_1}, \ldots, \sqrt{d_l} \in K(\sqrt{d_1}, \ldots, \sqrt{d_l})$ are linearly independent over $K$.)

3. Easy sums by duality

Fix a vector space $L/\mathbb{Q}$ in the set $\{L \subseteq C(V)\}'$ (see Definition 1.9). Now define $\Lambda, \Lambda^\perp$, and choose $\Lambda, \Lambda^\perp$, following Definition 1.13. Then $F_i|_L = 0$, and $\Lambda = \mathbb{Z}^m/2\Lambda$ and $\Lambda^\perp = \mathbb{Z}^{m/2}\Lambda^\perp$ are primitive rank-$m/2$ sublattices of $\mathbb{Z}^m$. We seek to prove Lemma 3.10 and Proposition 3.11 below (essentially for the “endgame” of §5). Before proceeding, we first remark on certain congruences appearing in the statements of these results.

Remark 3.1. By primitivity, $\Lambda^\perp \cap n_0\mathbb{Z}^m = n_0\Lambda^\perp$. So given $b \in \Lambda^\perp$, we have

$$\{ c \in \Lambda^\perp : c \equiv b \mod n_0 \} = \{ c \in \Lambda^\perp : c \equiv b \mod n_0\Lambda^\perp \},$$

i.e. the two notions of “congruence modulo $n_0$” on $\Lambda^\perp$ coincide. For us, the second notion ($\Lambda^\perp/n_0\Lambda^\perp$) is more fundamental; we mainly use primitivity to simplify local calculations.

For the rest of §3, we will let $\mathbf{x}, \mathbf{h}, \mathbf{x}'$ denote column vectors and $\mathbf{c}, \mathbf{b}, \mathbf{v}$ row vectors. In particular, the dot product $\mathbf{c} \cdot \mathbf{x}$ will coincide with standard matrix multiplication.

3.1. Preliminaries. For calculations to come, it will help to extend $\Lambda^\perp$ to a basis of $\mathbb{Z}^m$.

Definition 3.2. By primitivity of $\Lambda^\perp$, choose $\Gamma$ (itself primitive) such that $\mathbb{Z}^m = \Lambda^\perp \oplus \Gamma$. Then choose a $\frac{1}{2}m \times m$ basis matrix $\Gamma$ so that $\Gamma = \mathbb{Z}^{m/2}\Gamma$. 
Remark 3.3. Here det $[\Lambda^+]_1 = \pm 1$, since the rows of $[\Lambda^+]_1$ form a basis of $Z^m$. (Alternatively, note that $y \mapsto y \, [\Lambda^+]_1$ defines a surjective, hence bijective, $Z$-linear map $Z^m \to Z^m$. When proving the observation below, we will also use the bijectivity of $x \mapsto [\Lambda^+]_1 \, x$.)

Remark 3.4. $Z^m$ is unimodular, so $Z^m = \text{Hom}(Z^m, Z) = \text{Hom}(\Lambda^+, Z) \oplus \text{Hom}(\Gamma, Z) = \text{Hom}(Z^m/\Gamma, Z) \oplus \text{Hom}(Z^m/\Lambda^+, Z) = \Gamma^+ \oplus \Lambda^+$—but we do not need this.

When evaluating Fourier transforms at $c \in \Lambda^+$, it will help to rewrite $c \cdot x$. We also need to compute with $\Lambda, \Lambda^+$ over various rings. Thus we make a slightly tedious observation:

Observation 3.5. Let $R$ denote a ring, e.g. $\mathbb{R}$ or $Z/nZ$. Given a $Z$-module $M$, let $M_R := M \otimes R$; let $\Lambda^+_R := (\Lambda^+)_R$. Let $R \cdot \Lambda := R \cdot \Lambda^+ \subseteq R^{m \times 1}$ and $\Lambda \cdot \Lambda^+ := R^{m/2} \Lambda^+ \subseteq R^{1 \times m}$.

1. The canonical maps $\kappa: \Lambda_R \to \Lambda \cdot R$ and $\kappa: \Lambda^+_R \to R \cdot \Lambda^+$ are isomorphisms.
2. If $c \in \Lambda^+_R$, then there exists a unique row vector $c^* \in R^{m/2}$ with $\kappa c = c^* \Lambda$.
3. The $Z$-bilinear map $c \cdot x: \Lambda^+ \times Z^m \to Z$ induces an $R$-bilinear map $c \cdot x: \Lambda^+_R \times R^m \to R$.

Furthermore, $c \cdot x = (\kappa c) \cdot x = c^* \cdot \Lambda^+ \cdot x$ holds for all $(c, x) \in \Lambda^+_R \times R^m$.
4. If $x \in R^m$, then $\Lambda^+ \cdot x = 0$ if and only if $x \in \kappa \Lambda R$.

Proof of (1). The maps $\kappa$ are induced by inclusions $i: \Lambda \to Z^m$ and $i: \Lambda^+ \to Z^m$. But $\Lambda, \Lambda^+$ are primitive, so the tensored $i \otimes R$—with images $\Lambda \cdot R, R \cdot \Lambda^+$—remain injective.

Proof of (2). Write $c^* \Lambda^+ = [c^* \, 0] \, [\Lambda^+_1]$, and note that $[\Lambda^+_1]_1$ is invertible.

Proof of (3). The extension-of-scalars map $\text{Hom}(Z^m, Z) \otimes R \to \text{Hom}_R(R^m, R)$ induces a map $\text{Hom}(\Lambda^+, \text{Hom}(Z^m, Z)) \to \text{Hom}_R(\Lambda^+_R, \text{Hom}_R(R^m, R))$, by extension of scalars. This implies the first part of (3). Furthermore, the $Z$-linear map $c \cdot \star: \Lambda^+ \to \text{Hom}(Z^m, Z)$ factors through $\iota: \Lambda^+ \to Z^m$, so the induced $R$-linear map $c \cdot \star: \Lambda^+_R \to \text{Hom}(R^m, R)$ factors through $\kappa: \Lambda^+_R \to R^m$. The second part of (3) now easily follows from (1)-(2).

Proof of (4). Since $\det [\Lambda^+_1] = \pm 1$, the map $Z^m \ni x \mapsto \Lambda^+ \cdot x \in Z^{m/2}$ is surjective, and thus defines an exact sequence $\Lambda \to Z^m \to Z^{m/2} \to 0$. The right exact functor $\otimes R$ therefore gives an exact sequence $\Lambda_R \to R^m \to R^{m/2} \to 0$. So $\ker(\Lambda^+_1 \mid R^m) = \im(\Lambda_R \to R^m) = \kappa \Lambda_R$.

Remark 3.6. In fact $\Lambda \to Z^m \to Z^{m/2}$ and $\Lambda_R \to R^m \to R^{m/2}$ define (split) short exact sequences. But there are no canonical splittings—which is perhaps confusing, at least in our notation (which conflates $(Z^*)^t = \text{Hom}(Z^*, Z)$ with $Z^*$).

In view of the observation, we now make the following definition, for any given $x \in R^m$:

Definition 3.7. Let $h := \Lambda^+ \cdot x \in R^{m/2}$ and $x' := \Gamma x \in R^{m/2}$, i.e. $[h^*] := [\Lambda^+_1]_1 \, x$.

Remark 3.8. The observation implies that if $x \in R^m$, then $h = 0 \iff x \in \Lambda \cdot R$.

Example 3.9. If $F = x_1^3 + \cdots + x_m^3$ and $\Lambda^+ = \mathcal{R}$ (in the notation of Observation 2.9), then we can choose $\Lambda^+$ so that $h_k := \sum_{i \in \mathcal{J}(k)} x_i$ for all $k$.

3.2. Averaging the oscillatory integrals.

Lemma 3.10. If $n = n_0 n_1$ and $b \in \Lambda^+$, then

$$n_1^{-m/2} \sum_{c \in b + n_0 \Lambda^+} I_c(n) = \sum_{h \in n_1 Z^{m/2}} e_n(-b^* \cdot h) \int_{\mathbb{R}^m} d\mathbf{x}' w(\mathbf{x}/X) h(n/Y, F(\mathbf{x})/Y^2).$$

Here $h = 0$ summand simplifies to $\sigma_{\infty, L^+, w} X^{m/2} \cdot h(n/Y, 0)$. Furthermore, if $n_1 \geq_{\Lambda^+, w} X$ with a sufficiently large implied constant, then each $h \neq 0$ summand vanishes.
Proof of third part. If \( n_1 \gtrsim_{\Lambda, \omega} X \) is sufficiently large, then the conditions \( n_1 \mid h = \Lambda \perp x \) and \( w(x/X) \neq 0 \) cannot simultaneously hold unless \( h = 0 \).

Proof of second part. Simplify using \( F|_{\Lambda, \omega} = 0 \), normalize via \( \hat{x} := x/X \), and observe that
\[
\int_{(h, \hat{x}) \in \{0\} \times \mathbb{R}^{m/2}} d\hat{x}' w(\hat{x}) = \lim_{\epsilon \to 0} (2\epsilon)^{-m/2} \int_{(h, \hat{x}) \in \mathbb{R}^m} \frac{d(\hat{h}, \hat{x})}{d\hat{x}} w(\hat{x}) \cdot 1_{\|\hat{h}\|_\infty \leq \epsilon} = \sigma_{\infty, L^\perp, w}
\]
by an \( \epsilon \)-thickening in \( \hat{h} \) followed by pullback under the unimodular map \( \hat{x} \mapsto (\hat{h}, \hat{x}') \).

Proof of first part. Write \( c = b + n_0 v \), with \( c^*, b^*, v^* \in \mathbb{Z}^{m/2} \) defined as earlier. Then
\[
c \cdot x/n = c^* \cdot h/n = b^* \cdot h/n + v^* \cdot h/n_1
\]
is independent of \( x' \); the \( x \)-dependence lies in \( h \). Since \( x \mapsto (h, x') \) is unimodular,
the integral
\[
I_c(n) = \int_{(h, x')} dx \, w(x/X) h(n/Y, F(x)/Y^2) e(-c \cdot x/n)
\]
is therefore the Fourier transform at \( v^*/n_1 \in \mathbb{R}^{m/2} \) of the \( (h \text{-dependent}) \) function
\[
\mathbb{R}^{m/2} \rightarrow \mathbb{R}, \quad h \mapsto e(-b^* \cdot h/n) \int_{x'} dx' \, w(x/X) h(n/Y, F(x)/Y^2).
\]
So Poisson summation over \( h \in n_1 \mathbb{Z}^{m/2} \), or “in reverse” over \( v^*/n_1 \in n_1^{-1} \mathbb{Z}^{m/2} \), yields
\[
n_1^{-m/2} \sum_{v^* \in \mathbb{Z}^{m/2}} I_{b + n_0 v}(n) = \sum_{h \in n_1 \mathbb{Z}^{m/2}} e(-b^* \cdot h/n) \int_{x'} dx' \, w(x/X) h(n/Y, F(x)/Y^2),
\]
as desired. \( \square \)

3.3. Vertically averaging the exponential sums.

Proposition 3.11. \( S_c(n) \) is a function of \( n \) and \( c \mod n \). If \( j \in \mathbb{Z}^{m/2} \), then
\[
\mathbb{E}_{c \in \Lambda^\perp / n \Lambda^\perp} [S_c(n)e_n(-c^* \cdot j)] = \sum_{a \mod n} \sum_{x \mod n} e_n(aF(x)) \cdot 1_{n|h-j}.
\]
In particular, if \( j = 0 \), then \( \mathbb{E}_{c \in \Lambda^\perp / n \Lambda^\perp} [S_c(n)] = \phi(n) n^{m/2} \).

Remark 3.12. We only need the \( j = 0 \) case, but in general, the right-hand side equals \( T_n(j) \) as defined (for diagonal \( F \)) on [Hö98] p. 692.

Proof. The first part is clear by definition: \( S_c(n) := \sum_{a \mod n} \sum_{x \mod n} e_n(aF(x) + c \cdot x) \).

For the second part, fix \( a, x \) and let \( R := \mathbb{Z} / n \mathbb{Z} \). Now restrict to \( c \in \Lambda^\perp / n \Lambda^\perp = \Lambda^\perp \), corresponding isomorphically to \( c^* \in R^{m/2} = (\mathbb{Z} / n \mathbb{Z})^{m/2} \) as specified earlier—using primitivity of \( \Lambda^\perp \)—so that \( c \cdot x = c^* \cdot h \). Then note that the average of \( e_n(c^* \cdot h - c^* \cdot j) \) over \( c \), or equivalently over \( c^* \), is precisely \( 1_{n|h-j} \). Summing over \( a, x \) gives the desired result.

Finally, for the third part, recall that \( h = 0 \) in \( R \) if and only if \( x \in \Lambda \cdot R \)—in which case \( F(x) = 0 \) in \( R \), and the “right-hand side from the second part” simplifies to \( \phi(n) n^{m/2} \). \( \square \)
4. Bounding various sparse contributions

Let $S \subseteq \{c \in \mathbb{Z}^m : F^v(c) = 0\}$ be a homogeneous (i.e. invariant under scaling, so $c \in S$ implies $\mathbb{Z} \cdot c \subseteq S$) subset of $c$’s with $F^v(c) = 0$. At several technical points in §3 the following lemma will let us cleanly discard various contributions from sparse homogeneous sets $S$—when restricted to $c \neq 0$, at least.

**Lemma 4.1.** If $S \cap [-C,C]^m$ has size $O(C^{m/2-\delta+\epsilon})$ for all $C \gtrsim 1$, then
\[
Y^{-2} \sum_{c \in S \setminus \{0\}} \sum_{n \geq 1} |I_c(n)| \cdot n^{1-m/2} \cdot \max_{n_i | n} n_i^{-1/2} |\tilde{S}_c(n_i)| \lesssim \epsilon X^{(m-\delta)/2+\epsilon}
\]
provided $4 \leq m \leq 6$ and $\delta \leq \min((m+2)/4, (m-1)/2)$.

In the absence of a deeper algebro-geometric understanding of $S_c$, we will rely heavily on the bound in the following remark, valid for diagonal $F$.

**Remark 4.2.** For all integers $n \geq 1$ and tuples $c \in \mathbb{Z}^m$, we have
\[
n^{-1/2} |\tilde{S}_c(n)| \lesssim_F O(1)^{\omega(n)} \prod_{j \in [m]} \gcd(\text{cub}(n)^{1/6}, \gcd(\text{cub}(n), \text{sq}(c_j))^{1/4}),
\]
where $\text{cub}(\ast)$ and $\text{sq}(\ast)$ denote the cube-full and square-full parts of $\ast$, respectively. (See [Wan21, §2.3, Proposition 2.8] for details and precise references.)

[Hb98] proves a simplified version of Lemma 4.1 with $(m, \delta) \in \{(4, 1), (6, 0)\}$, and with $1_{n_i = n}$ instead of $\max_{n_i | n}$. The argument of [Hb98] generalizes without significant modification, so we focus on the key ideas in the proof of Lemma 4.1.

**Proof.** We may restrict to $c \in S$ with $||c|| \asymp C$. Since $F$ is diagonal, the homogeneous polynomial $F^v(c) \in \mathbb{Z}[c]$ (of degree $D$, say) contains the monomials $c_1^D, \ldots, c_m^D$ (each with nonzero coefficient). So $|c_i| \gtrsim C$ for at least two indices $i \in [m]$.

In particular, $|I_c(n)| \lesssim \epsilon X^{m+\epsilon} (XC/N)^{(2-m)/4}$ [Hb98, p. 688, (7.3)].

On the other hand, if $c = gc'$ with $g > 0$ and $c'$ primitive, then for all primes $p \nmid 3 \prod F_i = O_F(1)$, the equation $F^v(c') = 0$ forces $p \nmid c'_i$ for at least two indices $i \in [m]$. The bound
\[
n^{-1/2} |\tilde{S}_c(n_i)| \lesssim O(1)^{\omega(n_i)} \prod_{i \in [m]} \gcd(\text{cub}(n_i)^{1/6}, \gcd(\text{cub}(n_i), \text{sq}(c_i))^{1/4})
\]
for $n_i | n$ now implies (since $\text{cub}(n_i) \mid \text{cub}(n)$, and $p \nmid c'_i$ implies $v_p(\text{sq}(c_i)) = v_p(\text{sq}(g)))$
\[
\max_{n_i | n} n_i^{-1/2} |\tilde{S}_c(n_i)| \lesssim O(1)^{\omega(n)} \text{cub}(n)^{(m-2)/6} \gcd(\text{cub}(n), O_F(1)^{\infty})^{2/6} \gcd(\text{cub}(n), g^F)^{2/4},
\]
where $g^F$ is the prime-to-$O_F(1)$ part of $g$.

Now write $n = n_2n_3$, with $n_3 := \text{cub}(n)$ the cube-full part and $n_2 := n/n_3$ the cube-free part. Then for fixed $c$, the dyadic intervals $n_2 \asymp N_2$ and $n_3 \asymp N_3$ contribute
\[
\sum_{n_2, n_3} |I_c(n)| \cdot n^{1-m/2} \cdot \max_{n_i | n} n_i^{-1/2} |\tilde{S}_c(n_i)| \lesssim \epsilon X^{m+\epsilon} (XC/N)^{(2-m)/4} \cdot N_1^{1-m/2} \cdot N_3^{(m-2)/6} (N_2 \cdot \text{cub}(g)^{1/6} N_3^{1/3})
\]
under the preceding bounds on $S_c, I_c$, upon the sum estimates $\sum_{n_2} 1 \lesssim N_2$ and
\[
\sum_{n_3} \gcd(n_3, O_F(1)^{\infty})^{1/3} \gcd(n_3, \text{cub}(g^F))^{1/2} \lesssim \epsilon X^{\epsilon} \text{cub}(g^F)^{1/6} N_3^{1/3}.
\]
To prove the last estimate over $n_3$, let $d_F := \gcd(n_3, O_F(1)^\infty)$ and $d_i := \gcd(n_3, g_i^F)$ for $i = 2, 3$, where $g_i^F := \text{cub}(g_i^F)$ and $g_2^F := \text{sq}(g_2^F)/g_3^F$. Then $d_F, d_3$ are cube-full and $d_2$ is exactly square-full. So $D := d_Fd_2^{3/2}d_3$ divides $n_3$. Let $n_3'$ be the prime-to-$D$ part of $n_3$, so that $n_3'$ is cube-full of size $O(N_3/D)$, while $n_3/n_3' \mid D^\infty$. Then given cube-full $d_F \mid O_F(1)^\infty$ and $d_3 \mid g_3^F$, and square-full $d_2 \mid g_2^F$, there are at most $O_\epsilon(X^\epsilon(N_3/D)^{1/3})$ possibilities for $n_3 \sim N_3$. Summing $d_F^{1/2}(d_2d_3)^{1/2} \cdot X^{-1}(N_3/D)^{1/3}$ over the (very sparse!) variables $d_F \leq N_3$ and $d_2, d_3$ gives roughly $\text{cub}(g_3^F)^{1/6}N_3^{1/3}$, as desired.

The bound in the penultimate display simplifies to $\text{cub}(\gcd(c))^{1/6}X^{m+\epsilon}(XC)^{(2-m)/4}$ times 
\[ N^{1/2-m/4}N_2N_3^{m/6} \lesssim N^{3/2-m/4}, \]

since $m \leq 6$. Finally, summing over $c$ restricted to lie in $S \setminus \{0\}$, we obtain 
\[ Y^{-2} \sum_{|c| \leq Cn_2 \cdot n_3} \sum' |I_c(n)| \cdot n^{1-m/2} \cdot \max_{n_* | n} |S_c(n_*)| \lesssim Y^{-2}X^{m+\epsilon}(XC)^{(2-m)/4}N^{3/2-m/4} \sum_{g \geq 1} \sum' \text{cub}(g)^{1/6} \cdot 1_{\gcd(c) = g}. \]

If $g_3 := \text{cub}(\gcd(c))$, then $g_3 \mid c$, so the sum over $g, c$ is at most 
\[ \sum' \sum' g_3^{1/6} \cdot 1_{g_3c} \lesssim \sum' g_3^{1/6} (C/g_3)^{m/2-\delta} = C^{m/2-\delta} \sum_{g_3 \leq C} g_3^{1/6+\delta-m/2}, \]

with $g_3$ only restricted to be cube-full. There are $O(G_3^{1/3})$ cube-full numbers $g_3 \sim G_3$, so the original sum over $g, c$ simplifies to 
\[ O(C^{m/2-\delta} \max(1, C^{1/2+\delta-m/2}) \log C), \]

for a final $C$-exponent of roughly (keeping the condition $\delta \leq (m + 2)/4$ in mind) 
\[ (2-m)/4 + m/2 - \delta + \max(0, 1/2 + \delta - m/2) = \max((m+2)/4 - \delta, (4-m)/4) \geq 0. \]

The $N$-exponent $3/2 - m/4 = (6-m)/4$ is also nonnegative (for $m \leq 6$), so we may “replace” $C$ with $Z \lesssim X^{1/2} \log^{O(1)}(X)$, and $N$ with $Y = X^{3/2}$, for a final bound of roughly 
\[ Y^{-2}X^{m+\epsilon}Y^{(2-m)/4}X^{3/2-m/4}(X/Y)^{m/2-\delta + \max(0, \delta - (m-1)/2)} \approx X^{m/2}(Y/X)^{-\delta + \max(0, \delta - (m-1)/2)}. \]

Now plug in $Y/X = X^{1/2}$, and recall the condition $\delta \leq (m - 1)/2$, to finish. \(\square\)

Remark 4.3. “...each of which divides $c$. It follows that there is some such $H$ for which...” [HB98] p. 689] suggests a clever way to address [HB98]’s factor $H^{1/2} \approx \gcd(n_3, g_F^{1/2})$: use the clean $\ell^1$ bound $\sum_{H} \left( \sum' |n_3| \cdot 1_{H|n_3} \right) \lesssim \sum' \tau(n_3)$ to write 
\[ \sum_{H \geq 1} \sum' 1_{H|n_3} 1_{H|c} H^{1/2} N_3^{(m-2)/6} \cdot \sup_{H \geq 1} \left( \sum' \tau(n_3) \right) \cdot \prod_{i \in [m]} (X|e_i|/n)^{-1/2}. \]

Remark 4.4. It would be nice to give a more conceptual treatment avoiding numerical coincidences like $1/3 + (m - 2)/6 \leq 1$ (for $m \leq 6$). [HB98]’s most convenient bound 
\[ |I_c(n)| \lesssim X^{m+\epsilon}(X \|c\|/n) \prod_{i \in [m]} (X|e_i|/n)^{-1/2} \]
happens to present some technical difficulties when the coordinates $c_i$ have rather different sizes, but Heath-Brown’s techniques suggest a plausibly true improved bound of

$$X^{m+\epsilon}(X\|c\|/n)^{1-m/4} \prod_{i \in [m]} (X|c_i|/n)^{-1/4} \quad \text{(with proof for certain “standard” weights $w$)},$$

which would likely help to reveal the dominant $c$-ranges. Alternatively, if we restrict to weights $w$ supported away from the Hessian of $F$ (i.e. $w$ supported away from all coordinate hyperplanes $x_i = 0$), then a clean uniform bound of the form $X^{m+\epsilon}(X\|c\|/n)^{1-m/2}$ should hold [Hoo14 p. 252], which would likely simplify the argument even further.

5. A FULL PROOF OUTLINE FOR THE THEOREM

Fix $F, w$ in Theorem 1.8. Restrict $L$ to $\{L \subseteq C(V)\}'$. The first part of Proposition 2.2 implies that $F^y|_{L^\perp} = 0$ holds (for general reasons) for each $L$ at hand. So in terms of the sublattices $\Lambda, \Lambda^\perp$ from Definition 1.13 (see the beginning of §3), $F^y(c) = 0$ has certain trivial integral solutions, namely $c \in \bigcup_L \Lambda^\perp$ (agreeing with the earlier Definition 1.16).

Since $F$ is diagonal, Observation 2.9 characterizes the trivial $c$'s via certain pairings introduced in Definition 2.7. More precisely, the identification $\Lambda^\perp = \mathcal{R}_J$ defines a bijection between $\{L \subseteq C(V)\}'$ and the the set of permissible pairings $\mathcal{J}$ of $[m]$.

We now begin to use the diagonality assumption on $F$ in various serious but technical ways (as outlined in Remark 1.22). Consider the sum over $\{c \neq 0 : F^y(c) = 0\}$ in Theorem 1.8. We begin by reducing consideration to trivial $c$'s.

5.1. Reducing to trivial $c$'s. In fact, for $m = 4$, [HB98] pp. 688–689 already eliminates all but the trivial (“special”) $c$'s, with a satisfactory error of $O_\epsilon(X^{3/2+\epsilon})$.

**Remark 5.1.** For $m = 4$ in the Fermat case, $|\mathcal{I}(k)| = 3$ (in [HB98]'s notation) is essentially impossible, so each nontrivial solution $c \in \mathbb{Z}^m$ to $F^y(c) = 0$ must factor as $ge^2$, with $g \in \mathbb{Z}$ and $F(e) = 0$. It would be interesting to compute $S_c(n)$ explicitly for such $c$, and to decide whether such $c$ really contribute $X^{3/2+\epsilon}$ in the delta method (or if the $\epsilon$ can be removed).

Heath-Brown’s proof easily generalizes: suppose we restrict to nontrivial $c$’s for either $m = 4$ or $m = 6$. In boxes $\|c\| \lesssim C$, one can prove that such $c$’s are relatively sparse—with a count of $O_\epsilon(C^{m/2-1+\epsilon})$—by analyzing the combinatorics of [HB98] p. 687. For $m = 6$, we can then replace $O_\epsilon(C^{5+\epsilon})$ in [HB98] p. 688, Lemma 7.1 with $O_\epsilon(C^{2+\epsilon})$, leading to an improvement of the final bound on [HB98] p. 689 by a factor of $X^{1/2}$ (see Lemma 4.1).

Therefore, to prove Theorem 1.8, it remains to evaluate

$$Y^{-2} \sum_{c \in \bigcup_L \Lambda^\perp \setminus \{0\}} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n) \quad \text{(where $L$ is restricted to $\{L \subseteq C(V)\}'$).}$$

In fact, it will suffice to consider a single $L$ at a time, as the following reduction shows.

**Reduction of Theorem 1.8 to Theorem 5.2 (below).** The $O_\epsilon(C^{m/2-1+\epsilon})$ count above, together with Lemma 1.11 (a “sparse contribution bound”), reduces consideration to $c \in \bigcup_L \Lambda^\perp \setminus \{0\}$. Each pairwise intersection $\Lambda^\perp \cap \Lambda^\perp \subseteq L^\perp \cap L^\perp$ is also “relatively sparse” with an $O_\epsilon(C^{m/2-1+\epsilon})$ count, so it is harmless (again by Lemma 4.1) to replace the union over $L$ with a sum.

For each individual $L$, Theorem 5.2 (below) applies. By the discussion preceding Corollary 1.14, we can rewrite $\sigma_{\infty,L^\perp,w} X^{m/2}$ in terms of $\sum_{x \in L^\perp} w(x/X)$, up to a negligible error. Now sum over $L$ to finish.

□
**Theorem 5.2.** In the setting of Theorem 1.8, fix $F, w, L$ (with $L$ in \{L $\subseteq C(V)\}'). Then
\[
Y^{-2} \sum_{c \in \Lambda \setminus \{0\}} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n) = O(X^{m/2-\delta}) + \sigma_{\infty,L,w} X^{m/2}
\]
for some absolute constant $\delta > 0$.

The remainder of §5 is devoted to the proof of Theorem 5.2.

5.2. The difficulty in Heath-Brown’s approach. For $m = 4$, [HB98, pp. 690–692] writes
\[
X^{-3} \sum_{c \in R_J} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n) = X^{-3} \sum_{n \geq 1} n^{-m} \cdot n^{m/2} \sum_{j \in \mathbb{Z}^{m/2}} T_q(j) J_q(j)
\]
for each permissible pairing $J$. Here $j = 0$ captures the “$J$-diagonal” contribution, which—if $w$ is “positive” (say)—strictly dominates the singular series if $m = 4$. (In total over $j \neq 0$, Heath-Brown obtains a final error of $O_e(X^{3/2+\varepsilon})$ [HB98, p. 690, Lemma 8.1].)

The same equality (properly interpreted) holds for all $m \in \{4, 6, 8, \ldots\}$. However, if $m = 6$ (and $w$ is “positive”), then the singular series roughly matches each $J$-diagonal contribution in magnitude, so that for any given $J$, the locus $j \neq 0$ likely carries part of the main term. It may well be that for typical $j \neq 0$, the (multiplicative!) sums $T_q(j)$—related to the local analysis of certain $m/2$-variable quadratic equations $F(\ast)(j, \ast) = 0$ in $\ast$ [HB98, p. 691]—can be analyzed in terms of $L$-functions, but it is not a priori clear where the singular series would arise for $m = 6$.

A deeper analysis—perhaps considering small and large $q$ separately—would likely be enlightening for any given value of $m$. But at least for the relatively qualitative purpose of proving Theorem 5.2, a more direct approach will suffice, as we now explain.

5.3. Our alternative approach. First, rewrite the left-hand side of Theorem 5.2 in terms of $\tilde{S}_c(n) := n^{-(1+m)/2} S_c(n)$:
\[
Y^{-2} \sum_{c \in \Lambda \setminus \{0\}} \sum_{n \geq 1} n^{-m} S_c(n) I_c(n) = Y^{-2} \sum_{c \in \Lambda \setminus \{0\}} \sum_{n \geq 1} n^{(1-m)/2} \tilde{S}_c(n) I_c(n).
\]
Recall the running assumptions $n \lesssim Y$ and $\|c\| \leq X^{1/2+\varepsilon}$ from Observation 1.6.

**Remark 5.3** (Building intuition). Summing “crudely”—using known pointwise bounds on $|S_c(n)|, |I_c(n)|$ from [HB98]—already proves $Y^{-2} \sum_{c,n} \cdots \ll_{\varepsilon} X^{m/2+\varepsilon}$. (See e.g. [Wan21] §5.) Furthermore, on dyadic ranges, the precise bound thus obtained is strictly increasing in the modulus size range $n \asymp N$. So over $n \lesssim Y^{1-\delta}$, one already has a power saving.

**Definition 5.4.** Let $P = P_3$ be a small power of $Y$ to be determined.

Now fix a modulus $n \lesssim Y$, with $n \gtrsim Y^{1-\delta}$ being the most interesting.

5.3.1. Recognizing local biases. Before summing over $c \in \Lambda$, we will first decompose $S_c(n)$ into simpler pieces, motivated by the key lemma below exposing the bias of $S_c(p^\delta)$. Recall that by §1.2.3’s conventions, $F^\vee \in ((6^m)!F_1 \cdots F_m) \cdot \mathbb{Z}[c]$ (since $F$ is diagonal).

**Lemma 5.5.** Assume $F$ is diagonal, with $m \in \{4, 6\}$. Suppose $c \in \Lambda = R_J$ is trivial, and $p \nmid (j^{2m/2-1} F^\vee)(c)$ is a prime. Then
\[
\tilde{S}_c(p) = \phi(p)p^{-1/2} + O(1).
\]
Also, in the notation of Definition 2.7, \(c(k)^3 \in \mathbb{Q}\) is a well-defined \(p\)-adic unit for all \(k\), and
\[
\tilde{S}_c(p') = \phi(p')p^{-1/2} \cdot \prod_{k \in [m/2-1]} [1 + \chi(c(k)^3(c(k+1)^3))] \ll \phi(p')p^{-1/2}
\]
for all integers \(l \geq 2\), if we let \(\chi(r) := \left(\frac{r}{p}\right)\). Both implied constants depend only on \(m\).

**Remark 5.6.** In general, \(E_{c \in \Lambda^\perp /n\Lambda^\perp} [\tilde{S}_c(n)] = \phi(n)n^{-1/2}\) by Proposition 3.11. Hence we have chosen to formulate Lemma 5.5 using \(\phi(n)n^{-1/2}\), not \(n^{1/2}\). Even though the choice does not matter to first (or even second) order, it will prove convenient near the end of §5.

**Proof.** We will prove the first part immediately after Lemma 6.6. For the claim \(c(k)^3 \in \mathbb{Z}_p^\times\), see list items (1)–(2) near the beginning of §6. Finally, for \(l \geq 2\), see Lemma 6.10. \(\square\)

**Remark 5.7.** Analogously, it might be interesting to try computing \(I_c(n)\) to greater precision, as a way to reduce our reliance on Lemma 5.5’s pointwise understanding of \(S_c(n)\).

### 5.3.2. Separating local bias and error

Fix \(c \in \Lambda^\perp\) with \((J^{2m/2-1}F^\vee)(c) \neq 0\). As suggested above, it is natural to compare \(\tilde{S}_c(n)\) with \(\phi(n)n^{-1/2}\). More precisely, by Lemma 5.5, the Dirichlet series \(\Phi(c,s) := \sum_{n \geq 1} n^{-s}\tilde{S}_c(n)\) should resemble (to leading order) the series
\[
\sum_{n \geq 1} n^{-s}\phi(n)n^{-1/2} = \sum_{n \geq 1} n^{-s-1/2} (\mu \ast \text{Id})(n) = \frac{1}{\zeta(s + 1/2)} \cdot \zeta(s - 1/2).
\]

Thus we divide the former by the latter to define the “multiplicative error”
\[
\sum_{n \geq 1} n^{-s}\tilde{S}_c'(n) := \frac{\Phi(c,s)}{\zeta(s - 1/2)/\zeta(s + 1/2)} = \frac{1}{\zeta(s + 1/2)} \cdot \zeta(s - 1/2) \cdot \Phi(c,s).
\]

The expansions \(1/\zeta(s) = \sum_{n \geq 1} n^{-s}\mu(n)\) and \(\zeta(s) = \sum_{n \geq 1} n^{-s}\) lead to the “error expansion”
\[
\tilde{S}'_c(n) := \sum_{d_0d_1d_2 = n} \mu(d_0)d_0^{1/2} \cdot d_1^{-1/2} \cdot \tilde{S}_c(d_2),
\]

from which we obtain rigorous pointwise “error bounds”

1. \(\tilde{S}'_c(p) = -p^{1/2} + O(1) + \phi(p)p^{-1/2} = O(1)\) if \(p \mid j^{2m/2-1}F^\vee(c)\), by Lemma 5.5 for \(S_c(p)\);
2. \(\tilde{S}'_c(p') \ll \tau_3(p')p^{1/2}\) if \(p \mid j^{2m/2-1}F^\vee(c)\) and \(l \geq 2\), by Lemma 5.5 for \(S_c(p^{2l})\); and
3. \(\tilde{S}'_c(n) \ll \tau_3(n)n^{1/2} \cdot \max_{n_1|n} n_*^{-1/2} |\tilde{S}_c(n_*)|\) in general, by the triangle inequality, even if \(\gcd(n,j^{2m/2-1}F^\vee(c)) > 1\).

(Here \(\tau_3(n) := \sum_{d_0d_1d_2 = n} 1\).)

**Remark 5.8.** For convenience, write \(S'_c(n) := n^{(1+m)/2}\tilde{S}'_c(n)\). Given \(c \in \Lambda^\perp\), the multiplicativity of \(S_c(n)\), \(\phi(n)\) in \(n\) leads to Euler products for \(\Phi, \zeta\), and then to multiplicativity of \(S'_c(n)\). So in particular, we can always factor \(S'_c(n)\) before applying the bounds above.

Given \(n\), we may now decompose \(\tilde{S}_c(n)\) as the following Dirichlet convolution:
\[
\tilde{S}_c(n) = \sum_{n_0n_1 = n} \tilde{S}'_c(n_0) \cdot \phi(n_1)n_1^{-1/2}.
\]

Before proceeding, we make a small observation that will later justify our definition of \(\tilde{S}_c\).

**Observation 5.9.** \(\tilde{S}_c'(n)\) is a function of \(n\) and \(c \mod n\). On average, \(E_{c \in \Lambda^\perp /n\Lambda^\perp} [\tilde{S}_c'(n)] = 1_{n = 1}\).
Proof. By the first part of Proposition \ref{prop:decomp}, \( S_c(n) \) depends at most on \( n \) and \( c \mod n \). But \( S'_c(n) \) depends at most on the list of values \( (S_c(d_2))_{d_2|n} \), hence at most on \( n \) and \( c \mod n \). From the pointwise Dirichlet series identity
\[
\sum_{n \geq 1} n^{-s} \tilde{S}_c(n) = \left( \sum_{n \geq 1} n^{-s} \tilde{S}'_c(n) \right) \left( \sum_{n \geq 1} n^{-s} \phi(n)n^{-1/2} \right)
\]
(equivalent to the decomposition of \( S_c(n) \) in terms of \( S'_c(n_0) \), \( \phi(n) n_1^{-1/2} \)), we thus obtain
\[
\sum_{n \geq 1} n^{-s} \mathbb{E}_{c \in \Lambda^+ \cap n\Lambda^+} [\tilde{S}_c(n)] = \left( \sum_{n \geq 1} n^{-s} \mathbb{E}_{c \in \Lambda^+ \cap n\Lambda^+} [\tilde{S}'_c(n)] \right) \left( \sum_{n \geq 1} n^{-s} \phi(n)n^{-1/2} \right)
\]
by averaging formally (coefficient-wise) over \( c \in \Lambda^+ \). However, the second part of Proposition \ref{prop:decomp} implies \( \mathbb{E}_{c \in \Lambda^+ \cap n\Lambda^+} [\tilde{S}_c(n)] = \phi(n)n^{-1/2} \) for all \( n \). So \( \mathbb{E}_{c \in \Lambda^+ \cap n\Lambda^+} [\tilde{S}'_c(n)] = 1_{n=1} \) follows formally by division.

Remark 5.10. In general, if \( A_c(n) \) is a function of \( n \) and \( c \mod n \), multiplicative in \( n \), then by the Chinese remainder theorem, \( \mathbb{E}_{c \in \Lambda^+ \cap n\Lambda^+} [A_c(n)] \) is also multiplicative in \( n \). Thus when proving Observation 5.9, it would have sufficed to apply Proposition 3.11 to \( n = p^j \) only.

5.3.3 Proving Theorem 5.2. Recall the running assumptions \( n \lesssim Y \) and \( \|c\| \lesssim X^{1/2+\epsilon} \) from Observation 1.6. Returning to our Dirichlet decomposition of \( \tilde{S}_c(n) \), we now separate the pieces \( n_1 \geq Y/P \) and \( n_1 \lesssim Y/P \), bounding the latter sum in absolute value:
\[
\tilde{S}_c(n) = \sum_{n_0 n_1 = n} \tilde{S}'_c(n_0) \cdot \phi(n_1) n_1^{-1/2} \cdot 1_{n_1 \geq Y/P} + O \left( \sum_{n_0 n_1 = n} |\tilde{S}'_c(n_0)| \cdot \phi(n_1) n_1^{-1/2} \cdot 1_{n_1 \lesssim Y/P} \right).
\]
We may then factor \( n_0 = n_c n_2 n'_1 \) to get \( S'_c(n_0) = S'_c(n_c) S'_c(n_2) S'_c(n'_1) \) with
- \( n_c : = \prod_{p | j^{2m/2-1} F^v(c)} n_p^{v_p(n_0)} \) the “bad factor”;
- \( n_2 : = \text{sq}(n_0/n_c) \) the largest square-full divisor of \( n_0/n_c \); and
- \( n'_1 : = n_0/(n_c n_2) \) the leftover part, implicitly square-free—

to which (E3), (E2), and (E1) apply, respectively.

Now, our strategy to prove Theorem 5.2 is as follows:

(S1) Eliminate \( c \in \Lambda^+ \setminus \{0\} \) with \( j^{2m/2-1} F^v(c) = 0 \) by summing crudely over \( c, n \) using the sparsity of \( c \).

(S2) Over \( c \in \Lambda^+ \setminus \{0\} \) with \( j^{2m/2-1} F^v(c) \neq 0 \), eliminate \( n_1 \lesssim Y/P \) by summing crudely over \( c, n \) using the sparsity of \( n_c, n_2 \) and smallness of \( S'_c(n'_1) \), whose contributions are too small to “compensate the enemy” for the fact that \( Y/n_1 \gtrsim P \).

(S3) Over \( c \in \Lambda^+ \setminus \{0\} \) with \( j^{2m/2-1} F^v(c) \neq 0 \), the key range \( n_1 \gtrsim Y/P \) remains. But here \( n_0 \lesssim P \), so fix \( n_0 \) and use Poisson summation over each residue class \( c \equiv b \mod n_0 \Lambda^+ \) to evaluate sums exactly (after over-extending to include \( c = 0 \) and \( j^{2m/2-1} F^v(c) = 0 \) in an artificial way), eventually leading to the main term of Theorem 5.2.

Remark 5.11. For \( m = 6 \), when over-extending in the last step, it is a priori safe to include \( c = 0 \), because small moduli \( n \lesssim P \) dominate the singular series. (For \( m = 4 \) the singular series is more mysterious, but the \( c = 0 \) contribution is dominated by the diagonal.) So for our basic purposes, we do not need a complicated inclusion-exclusion argument anywhere.

We will assume familiarity with the following crude bounds on \( |S_c(n)|, |I_c(n)| \) from [HB98]:
• For all integers \( n \geq 1 \) and tuples \( \mathbf{c} \in \mathbb{Z}^m \) with \( c_1 \cdots c_m \neq 0 \), we have
  \[
  n^{-1/2} |\tilde{S}_c(n)| \lesssim_F O(1)^{\omega(n)} \prod_{j \in [m]} \text{sq}(c_j)^{1/4}.
  \]

  (This is a special case of the bound stated, with precise references, in Remark 4.2)

• For all reals \( n \geq 1 \) and tuples \( \mathbf{c} \in \mathbb{R}^m \) with \( c_1 \cdots c_m \neq 0 \), we have
  \[
  |I_c(n)| \lesssim_{F,w,\epsilon} X^{m+\epsilon} n^{-1/2} \prod_{i \in [m]} \text{sq}(c_i)^{1/4}.
  \]

  by [HB98, p. 678, Lemma 3.2]. (One could likely weaken \( n \geq 1 \) to \( n > 0 \), and remove
  the \( X^c \); see Paper III for an affirmative proof when \((\text{Supp } w) \cap (\text{hess } F) \neq \emptyset\).

We now prove Theorem 5.2 in three steps, following the plan above.

First step. If \( \mathbf{c} \in \Lambda^+ \) satisfies \( j^{2m/2-1} F^\vee(\mathbf{c}) = 0 \), then by Observation 2.13 there must exist
  distinct \( k_1, k_2 \) with \( \|c(k_1)^3 c(k_2)^3 \| \in (\mathbb{Q} \times)^2 \cup \{0\} \). That leaves at most \( O(\epsilon(C^{m/2-1+c}) \) possible \( \mathbf{c} \in \Lambda^+ \) with \( \|\mathbf{c}\| \lesssim C \): given \( i \in J(k_1) \) and \( j \in J(k_2) \), there are at most \( O_{F_i,F_j}(C^{1+c}) \) pairs of integers \( c_i, c_j \leq C \) with \( (c_i^k F_i)(c_j^k F_j) \in \mathbb{Q}^2 \), i.e. \( (F_i c_i)(F_j c_j) \in \mathbb{Z}^2 \). Now finish by
  Lemma 4.1 just as in our earlier elimination of nontrivial \( \mathbf{c}' \)’s.

\[ \square \]

Remark 5.12. By now, it is clear that \( F^\vee \) is unsurprising (in the sense of Definition 1.17).

The precise goal of the second step, (S2), is to prove, under the restriction \( j^{2m/2-1} F^\vee(\mathbf{c}) \neq 0 \), a power saving of the form
  \[
  Y^{-2} \sum_{n \geq 1} n^{(1-m)/2} \left( \sum_{n_0 n_1 = n} |\tilde{S}_c(n_0)| \cdot \phi(n_1)^{1/2} \cdot 1_{n_0 \leq Y/P} \right) |I_c(n)| \lesssim X^{m/2-\delta}.
  \]

Second step. Fix a restricted \( \mathbf{c} \in \Lambda^+ = \mathcal{R}_J \). Then \( c_1 \cdots c_m \neq 0 \) by Observation 2.13 so:

• If \( |c(k)| \approx C(k) \) and \( \|\mathbf{c}\| \approx C \), then
  \[
  |I_c(n)| \lesssim X^{k+\epsilon} n^{-1/2} \prod_{i \in [m]} \text{sq}(c_i)^{1/4}.
  \]

• Fix \( n, n_0, n_1 \). Then \( n c_{e_{n,1}/2} |\tilde{S}_c(n_0)| \lesssim O(1)^{\omega(n,e)} \prod_{i \in [m]} \text{sq}(c_i)^{1/4} \) for all \( n_{e,1} \) so
  \[
  \phi(n_1)^{1/2} |\tilde{S}_c(n_0)| \lesssim n_1^{1/2} \cdot O(1)^{\omega(n)} \prod_{i \in [m]} \text{sq}(c_i)^{1/4}
  \]

by the factorization \( n_0 = n c_{n_2, n_1} \) and the bounds (E1)–(E3).

To proceed, we merely plug in these pointwise bounds—relaxed dyadically for convenience—and then sum dyadically over \( n_1, (n_c, n_2, n_1), (c(k))_{k \geq 1} \) in that order.

Specifically, in the expression \( Y^{-2} \sum_{n_{e,n} \cdots} \) of interest, fix dyadic intervals \( n_1 \approx N_1 \approx Y/P \); \( n_0 \approx N_0 \) \( |c(k)| \approx C(k) \); etc., with \( N \approx N_0 N_1 \) and \( N_0 \approx N_c N_2 N_1’ \). Now fix \( \mathbf{c} \), and use
  \[
  Y^{-2} n^{(1-m)/2} |I_c(n)| \lesssim_Y Y^{-2} N^{(1-m)/2} X^{m+\epsilon} (XC/N) \prod_{k \geq 1} (XC(k)/N)^{-1}
  \]

as an \( \ell^\infty \) bound over \( \mathbf{c} \). Over \( n_1, (n_c, n_2, n_1) \), it then remains to bound the “factor”
  \[
  \sum_{n_1, (n_c, n_2, n_1)} |\tilde{S}_c(n_0)| \cdot \phi(n_1)^{1/2}.
  \]
Each summand is at most \( O_\epsilon(N^\epsilon) \cdot (N_1N_2N_\epsilon)^{1/2} \prod_{i \in [m]} sq(c_i)^{1/4} \), for a total of at most
\[
(1 \cdot N_2^{1/2} \cdot N_1' \cdot N_1) \cdot (N_1N_2N_\epsilon)^{1/2} \prod_{i \in [m]} sq(c_i)^{1/4} = N_1^{3/2}N_1'N_2N_\epsilon^{1/2} \prod_{i \in [m]} sq(c_i)^{1/4}
\]
up to \( O_\epsilon(N^\epsilon\|c\|^{\epsilon}) \), since there are at most
- \( O_\epsilon(N^\epsilon\|c\|^{\epsilon}) \) divisors \( n_\epsilon \simeq N_\epsilon \) of \( j^{2m/2-1}F^\vee(c) \);
- \( O(N_1^{1/2}) \) square-full numbers \( n_2 \simeq N_2 \); and of course
- \( O(N_1'N_1) \) choices for \( n_1', n_1 \).

Note that if \( J(k) = \{i, j\} \), then \( sq(c_i)^{1/4} \cdot sq(c_j)^{1/4} \lesssim_F sq(c_{\min,J(k)})^{1/2} \).

Upon multiplying the two “factors” above, and summing over \( c \in \Lambda^\perp \), we obtain roughly
\[
Y^{-2}N^{(1-m)/2}X^m(XC/N) \cdot N_1^{1/2}N \prod_{k \geq 1} (XC(k)/N)^{-1} \sum_{|c(k)| \geq C(k)} sq(c_{\min,J(k)})^{1/2}.
\]

The sum over \( c(k) \) is roughly \( C(k) \) (up to \( \log C(k) \)), so we simplify to get roughly
\[
Y^{-2}N^{(1-m)/2}X^m(XC/N) \cdot N_1^{1/2}N(X/N)^{-m/2} = Y^{-2}N^{3/2}X^m/2(XC/N) \cdot N_1^{1/2}.
\]

The final exponents for \( C \) and \( N \) are positive; plugging in \( C \lesssim X^{1/2+\epsilon} \) and \( N \lesssim Y \) gives a final bound of roughly \( X^{m/2}(N_1/Y)^{1/2} \lesssim X^{m/2}P^{-1/2} \), as desired. \( \square \)

Having covered \( n_1 \lesssim Y/P \) via (S2), it remains to address the complementary sum
\[
Y^{-2} \sum_{c \in \Lambda^\perp} \sum'_{n \geq 1} n^{(1-m)/2} \left( \sum_{n_0/n_1 = n} \tilde{S}_c(n_0) \cdot \phi(n_1)n_1^{-1/2} \cdot 1_{n_1 \geq Y/P} \right) I_c(n),
\]
still under the restriction \( j^{2m/2-1}F^\vee(c) \neq 0 \). As suggested in (S3), we will first over-extend to all \( c \in \Lambda^\perp \), and then separately (later) cover the “loose end” \( j^{2m/2-1}F^\vee(c) = 0 \).

First part of third step. Fix \( n, n_0, n_1 \) with \( n \lesssim Y \) and \( n_1 \gtrsim Y/P \). Then \( n_0 \lesssim P \). But by the first part of Observation \( 5.9 \), \( \tilde{S}_c(n_0) \) only depends on \( c \mod n_0 \). “Reverse Poisson summation” (Lemma 3.10), applied to each individual residue class \( b \mod n_0\Lambda^\perp \), therefore implies
\[
\sum_{c \in \Lambda^\perp} \tilde{S}_c(n_0) \cdot I_c(n) = \sum_{b \in \Lambda^\perp/n_0\Lambda^\perp} \tilde{S}_b(n_0) \cdot n_1^{m/2} \cdot \sigma_{\infty,L^\perp,w}X^{m/2}h(n/Y, 0),
\]
provided \( Y/P \gtrsim_{F,w} X \) (which we may assume). But \( \sum_{b \in \Lambda^\perp/n_0\Lambda^\perp} \tilde{S}_b(n_0) = |\Lambda^\perp/n_0\Lambda^\perp| \cdot 1_{n_0=1} \) by the second part of Observation 5.9. Here \( |\Lambda^\perp/n_0\Lambda^\perp| = n_0^{m/2} \), so ultimately
\[
\sum_{c \in \Lambda^\perp} \tilde{S}_c(n_0) \cdot I_c(n) = n_0^{m/2} \cdot \sigma_{\infty,L^\perp,w}X^{m/2}h(n/Y, 0) \cdot 1_{n_0=1}.
\]

Now for each individual \( n \), the sum
\[
Y^{-2} \sum_{c \in \Lambda^\perp} n^{(1-m)/2} \left( \sum_{n_0/n_1 = n} \tilde{S}_c(n_0) \cdot \phi(n_1)n_1^{-1/2} \cdot 1_{n_1 \geq Y/P} \right) I_c(n),
\]
simplifies to
\[
Y^{-2}n^{(1-m)/2} \cdot \sum_{n_0/n_1 = n} n_0^{m/2} \cdot \sigma_{\infty,L^\perp,w}X^{m/2}h(n/Y, 0) \cdot 1_{n_0=1} \cdot \phi(n_1)n_1^{-1/2}1_{n_1 \geq Y/P},
\]
i.e. \( Y^{-2} \cdot \sigma_{\infty,L^+} X^{m/2} h(n/Y,0) \cdot \phi(n) 1_{n \geq Y/P} \). By Proposition \([5.13]\) (below), summing over \( n \geq 1 \) yields \( \sigma_{\infty,L^+} X^{m/2} \), up to a satisfactory error of \( O(X^{m/2P-1}) + O_A(Y^{-A}) \). \( \square \)

**Proposition 5.13.** \( \sum_{n \geq 1} Y^{-2} \phi(n) h(n/Y,0) \cdot 1_{n \geq Y/P} = 1 + O_A(Y^{-A}) + O(P^{-1}) \).

**Proof.** Up to a question of endpoints, write \( 1_{n \geq Y/P} = 1 - 1_{n \leq Y/P} \). Then we have (rigorously)

\[
\sum_{n \geq 1} Y^{-2} \phi(n) h(n/Y,0) \cdot 1_{n \geq Y/P} = \sum_{n \geq 1} Y^{-2} \phi(n) h(n/Y,0) + O \left( \sum_{n \leq Y/P} Y^{-2} \phi(n) \cdot |h(n/Y,0)| \right).
\]

Being unrestricted, the first sum on the right evaluates to \( Y^{-1} \sum_{n \geq 1} \omega(n/Y) = 1 + O_A(Y^{-A}) \) \([HB98, p. 692]\). Meanwhile, \( h(x,0) = \sum_{j \geq 1} \omega(xj) / (xj) \ll x^{-1} \) \([HB96, p. 168, Lemma 4]\) (since \( \omega \) is supported away from the origin by definition), so the second sum (i.e. the “absolute error” over \( n \)) is bounded by \( \sum_{n \leq Y/P} Y^{-2} \phi(n) \cdot Y/n \leq \sum_{n \leq Y/P} Y^{-1} \ll P^{-1} \). \( \square \)

In the first part of the third step, we “artificially over-extended” over \( c \) to include

\[
Y^{-2} \sum_{c \in \Lambda^+} 1_{j^2 \geq m/2 - 1} F^\vee(c) = 0 \sum_{n \geq 1} n^{(1-m)/2} \left( \sum_{n_0 = n} \tilde{S}_c(n_0) \phi(n_1) n_1^{-1/2} \cdot 1_{n_1 \geq Y/P} \right) \cdot I_c(n),
\]

a piece that we must now bound by \( O(X^{m/2-\delta}) \). This final “loose end” will complete the proof of Theorem \([5.2]\) and hence the proof of Theorem \([1.8]\) as well.

**Loose ends.** The main subtlety here is that we must treat \( c \neq 0 \) and \( c = 0 \) separately.

First, given \( n \), the bound (E3), applied directly to \( S_c(n_0) \) for each \( n_0 \mid n \), implies

\[
\sum_{n_0 = n} |S_c(n_0)| \cdot \phi(n_1) n_1^{-1/2} \cdot 1_{n_1 \geq Y/P} \leq \tau(n) \cdot \tau_3(n) n^{1/2} \cdot 1_{n \geq Y/P} \cdot \max_{n_0, \star} n_0^{-1/2} \left| \tilde{S}_c(n_0, \star) \right|.
\]

Even without the \( 1_{n \geq Y/P} \) factor, Lemma \([4.1]\) (an entirely “positive” statement) provides a satisfactory bound \( O(X^{(m-1)/2+\epsilon}) \) in total over \( \{ c \in \Lambda^+ \setminus \{0\} : j^{2m/2-1} F^\vee(c) = 0 \} \).

For \( c = 0 \), the factor \( 1_{n \geq Y/P} \) is crucial. Lemma \([B.1]\) together with \( I_0(n) \ll X^m \), yields

\[
Y^{-2} \sum_{n \geq 1} n^{1-m/2} \cdot 1_{n \geq Y/P} \cdot \max_{n_0, \star} n_0^{-1/2} |\tilde{S}_c(n_0, \star)| \cdot |I_0(n)| \leq Y^{-2} X^m \sup_{Y \geq N \geq Y/P} N^{(4-m)/3+\epsilon},
\]

i.e. roughly \( X^{m-3} (Y/P)^{(4-m)/3} \) (since \( m \geq 4 \)), which is certainly \( O(X^{m/2-\delta}) \) for \( m \leq 6 \). \( \square \)

**Remark 5.14.** If \( n \leq Y \) and \( n_1 \geq Y/P \), then \( n_0, \star \mid n_0 \leq P \). So up to a loss of \( P^{O(1)} \), we could simplify the “loose ends” via the trivial bound \( S_c(-) \ll P^{O(1)} \) — thus reducing to weaker versions of Lemmas \([4.1]\) and \([B.1]\) in which “\( \max_{n, \star} \)” is replaced by “\( 1 \)” (as if \( n_\star = 1 \)).

6. **Estimating Bias in Exponential Sums**

Fix \( F \) diagonal with \( m \in \{4,6\} \). In the setting of \([So]\) fix \( L/\mathbb{Q} \). We seek to prove Lemma \([5.5]\) regarding \( \tilde{S}_c(g^l) := p^{-l(1+m)/2} S_c(g^l) \) for \( l \geq 1 \), assuming \( c \in \Lambda^+ = \mathcal{R}_d \) and \( p \nmid j^{2m/2-1} F^\vee(c) \) is prime. In particular, by the choice of \( F^\vee \) in \([1.2.3]\) we may assume \( p \nmid (6^m)!F_1 \cdots F_m \).

Certain hyperplane sections govern the behavior of \( S_c(g^l) \). Throughout the present \([So]\) only, we adopt the following pair of definitions (and the convenient notation within).
Definition 6.1. Fix a prime \( p \nmid (6^m)! F_1 \cdots F_m \), and let \( V \) and \( V_c \) denote the proper schemes over \( \mathbb{Z} \), defined by the equations \( F(x) = 0 \) and \( F(x) = c \cdot x = 0 \), respectively. Then let \( V := V_{F_p} \) and \( V_c := (V_c)_{F_p} \) denote the corresponding fibers over \( \mathbb{F}_p \).

Definition 6.2 (Cf. [Wan21, §1.2, Definition 1.6]). For \( q = p^r \), let \( \rho(q) := |V(\mathbb{F}_q)| \) and \( \rho_c(q) := |V_c(\mathbb{F}_q)| \). Optimize \( E(q) := \rho(q) - (q^{m-1} - 1)/(q - 1) \) and \( E_c(q) := \rho_c(q) - (q^{m-2} - 1)/(q - 1) \) to get \( E_c(q) := q^m - E_c(q) \) and \( E(q) := q^{m-1} E_c(q) \), where \( m := m - 3 \).

The assumptions \( p \nmid (6^m)! F_1 \cdots F_m \) and \( p \nmid j^{2m^2-1} F^c(c) \) imply the following:

1. Each \( c(k)^3 \) is a well-defined rational number in \( \mathbb{Z}(p) = \mathbb{Q} \cap \mathbb{Z}_p \).
2. If \( k_1 < \cdots < k_t \), then \( p \nmid \prod (c(k_i)^3) \pm \cdots \pm c(k_i)^3/2 \). In particular, \( p \nmid c(k)^3 \), hence \( c(k)^3 \in \mathbb{Z}_p^* \), for each \( k \)—thus improving (1). Also, \( p \nmid c(i)^3 - c(j)^3 \) when \( i \neq j \).
3. \( V_c \) has exactly \( 2^{m^2}/2 \) singular \( \mathbb{F}_p \)-points.

Proof of (1). By definition, \( c(k)^3 = c_i^3/F_i \in \mathbb{Q} \) for all \( i \in J(k) \). Here \( c_i \in \mathbb{Z} \) and \( p \nmid F_i \). □

Proof of (2). Use the equivalence part of Observation 2.13 carried over to \( \mathbb{Q} \to \mathbb{F}_p \) via Remark 2.14.

Proof of (3). \( p \nmid j^{2m^2-1} F^c(c) \) by Corollary 2.12 carried over to \( \mathbb{F}_p \). But \( p \nmid j^{2m^2-1} F^c(c) \) by assumption. So by the geometric formulation of Proposition 2.10—carried over to \( \mathbb{F}_p \)—the hyperplane section \( V_c \) has at least, but also at most, \( 2^{m^2}/2 \) singular \( \mathbb{F}_p \)-points. □

6.1. A change of variables. As suggested in Remark 1.22, Lemma 5.5 might generalize to non-diagonal smooth \( F \). But to minimize unforeseen technical issues, and to maximize the accessibility of §6 we have chosen not to pursue such generalizations here.

Instead, we will use an ad hoc change of coordinates, highlighting specific features (of diagonal forms) that may be of independent interest.

Observation 6.3. There exist invertible \( \mathbb{Z}_p \)-linear maps \( x \mapsto x' \) and \( c \mapsto c' \), and a diagonal cubic form \( F' \) with \( F'_i = F'_i \) for each \( J(k) = \{ i, j \} \)—all depending on \( F, L \)—such that

- \( F(x), c \cdot x \) are preserved, i.e. we have \( F(x') = F'(x') \) and \( c \cdot x = c' \cdot x' \);
- \( c(k)^3 \) is preserved for each \( k \), i.e. we have \( c(k)^3 := c_i^3/F_i = (c_i^3/F_i)' =: c'(k)^3 \);
- the earlier list items (1)–(3) carry over from \( F, c, V_c \) to \( F', c', V'_c \); and
- if \( a \in \mathbb{Z}_p^* \) and \( x, c \in \mathbb{Z}_p^m \), then \( e_p^i(aF(x) + c \cdot x) = e_p^i(aF'(x') + c' \cdot x') \) for all \( l \geq 1 \).

Proof. By Observation 2.9 \( \Lambda^* = \mathcal{R}_J \) is cut out by the system of equations that includes, for each \( J(k) = \{ i, j \} \), the equation \( c_i/F_i^{1/3} = c_j/F_j^{1/3} \), where \( F_i, F_j \in F(k) \cdot (\mathbb{Q}_p)^3 \) for some unique cube-free integer \( F(k) \).

Since \( p \nmid F_1 \cdots F_m \) by assumption, we can define \( F', x', c' \) by specifying \( (F'_i, F'_j) := (F(k), F(k)) \) first, then \( F'_i(x_i)^3 := F_i x_i^3 \) and \( c'_i x_i := c_i x_i \), where we insist linearity of \( x_i \mapsto x'_i \). (There are three possible \( \mathbb{Z}_p \)-linear branches if \( p \equiv 1 \mod 3 \.) With the first bullet point thus satisfied, the second requirement, \( c_i^3/F_i = (c'_i)^3/F'_i \), automatically follows. The third and fourth then easily follow from the first two.

Finally, by inspection, the \( \mathbb{Z}_p \)-linear maps \( x \mapsto x' \) and \( c \mapsto c' \) are indeed invertible. □

Remark 6.4. In general, such a simple correspondence between \( (F, x, c), (F', x', c') \) is only possible after localizing away from finitely many primes—which we do do throughout §6.
For the remainder of §6, we may thus assume $F_i = F_j$ for each $J(k) = \{i, j\}$, with $\Lambda^+ = R_J$ cut out by the equations $c_i = c_j$. Then define $F(k) := F_i = F_j$ and $c_k^* := c_i = c_j$, so that $c(k)^3 = (c_k^*)^3/F(k)$.

Now consider the equations $F(x) = 0$ and $c \cdot x = 0$ defining $V_c$. After a linear change of variables over $\mathbb{Z}[1/6]$, they become

$$\sum_{k \in [m/2]} F(k) \cdot h[k]y[k]^2 = -3 \sum_{k \in [m/2]} F(k) \cdot h[k]^3 \quad \text{and} \quad \sum_{k \in [m/2]} c_k^* \cdot h[k] = 0.$$ 

Explicitly, if $J(k) = \{i, j\}$ with $i < j$, then we may take $h[k] := x_i + x_j$ and $y[k] := 3(x_i - x_j)$, so that the equation $h[1] = \cdots = h[m/2] = 0$ cuts out $\Lambda = R_J$.

**Remark 6.5.** We use the letter “$h$” in analogy with van der Corput or Weyl differencing. The definition of $h := (h[k])_{k \in [m/2]}$ is compatible with that in §3.

Geometrically speaking over $K := \mathbb{F}_p$, the space $\{[h] \in \mathbb{P}^{m/2-1} : c^* \cdot h = 0\} \cong \mathbb{P}^{m/2-2}$ parameterizes projective $m/2$-planes $\mathbb{P}H \subseteq \mathbb{P}c^* \cong \mathbb{P}^{m-2}$ containing the fixed $(m/2-1)$-plane $\mathbb{P}\Lambda_K$. So the explicit “quadratic fibration”

$$V_c \setminus \mathbb{P}\Lambda_K \rightarrow \mathbb{P}[c^*] \cong \mathbb{P}^{m/2-2}, \quad [x] \mapsto [h]$$

above is closely related to the blow-up of $V_c$ along $\mathbb{P}\Lambda_K$. Concretely, each slice $V_c \cap \mathbb{P}H$ consists of $\mathbb{P}\Lambda_K$ and a (possibly reducible or singular) quadric hypersurface $Q_H \subseteq \mathbb{P}H$ of dimension $m/2 - 1$, where $Q_H \setminus \mathbb{P}\Lambda_K$ is the fiber of $V_c \setminus \mathbb{P}\Lambda_K$ over $\mathbb{P}H$.

**6.2. Analyzing sums to prime moduli.** We begin as in [Wan21, §2.2’s discussion of $S_c(p)$]. First, $p \nmid c$ implies $S_c(p) = p^2 E_c(p) - pE(p)$, i.e., $S_c(p) = E_c(p) - p^{-1/2} E(p)$. Here $E(p) \ll 1$.

On the other hand, we now have the following lemma.

**Lemma 6.6.** In the setting of Lemma 5.5, $\tilde{E}_c(p) = p^{1/2} + O(1)$.

**Proof of first part of Lemma 5.5.** Plug Lemma 6.6 and $\tilde{E}(p) \ll 1$ into the proposition. □

**Proof of Lemma 6.6.** We must show that $E_c(p) = p^{(m-2)/2} + O(p^{(m-3)/2})$, or equivalently—in terms of the affine cone $C(V_c)$—that $1 + (p - 1)|V_c(\mathbb{F}_p)| = p^{m-2} + p^{m/2} + O(p^{(m-1)/2})$.

First, reduce to simpler $F, L$ as detailed in §6.1. We can then count solutions to $F(x) = c^* \cdot x = 0$ using the $(h, y)$ coordinates described near the end of §6.1. The locus $h = 0$ contributes $|\Lambda/p\Lambda| = p^{m/2}$ solutions.

Now restrict attention to $h \neq 0$ with $c^* \cdot h = 0$, which means at most one of the $h[k]$ can be zero—since $m/2 \leq 3$ and $p \nmid c_k^*$ for all $k \in [m/2]$. For convenience, let $\chi(r) := (\frac{r}{p})$.

1. If $m = 4$, then since $p \nmid c_1^* c_2^*$, we can fix an affine chart by setting $h[2] = c_1^*$, say—so that $h[1] = -c_2^*$, and the equation $F(x) = 0$ becomes

$$-F(1) \cdot c_2^* \cdot y[1]^2 + F(2) \cdot c_1^* \cdot y[2]^2 = 3[F(1) \cdot (c^*_2)^3 - F(2) \cdot (c^*_1)^3].$$

Since $p \nmid c(2)^3 - c(1)^3$, we may compare with $\mathbb{P}^1$ to get a point count of $p - \chi(c(1)^3 c(2)^3)$ for the affine chart; we then multiply by $p - 1$ to get the affine cone contribution.

2. If $m = 6$ and $h[k] = 0$ for some $k \in [m/2]$, then the affine cone contribution reduces to the product of $\mathbb{A}^1$ (with coordinate $y[k]$) with an affine cone for $m = 4$—for a total of $p^3 + O(p^2)$ for an individual $k$, or $3p^3 + O(p^2)$ as $k \in [m/2]$ varies.

3. If $m = 6$ and $h[k] \neq 0$ for all $k \in [m/2]$, then for $h$ fixed (with $c^* \cdot h = 0$), there are

$$p^2 + p \cdot \chi (F(1) h[1] \cdots F(3) h[3] \cdot 3(F(1) h[1]^3 + \cdots + F(3) h[3]^3))$$
solutions \( y \in \mathbb{F}_p^{m/2} \) to \( F(x) = 0 \) (a ternary quadratic in \( y \)). Indeed, if \( \sum F(k)h[k]^3 = 0 \), then the space \( \{ y : F(x) = 0 \} \) forms an affine cone over a conic \( Q \cong \mathbb{P}^1 \); otherwise, \( \{ y : F(x) = 0 \} \) forms a non-degenerate affine quadric surface in \( \mathbb{A}^3 \).

Now fix an affine chart by setting \( h[3] = 1 \), say; then \( h[2] = -(c_3^*)^{-1}(c_1^* \cdot h[1] + c_3^*) \).

The sum \( \Sigma \) of the previous display over \( t := h[1] \in \mathbb{F}_p^\times \setminus \{-c_3^*/c_1^*\} \)—where we restrict \( t \) so that \( h[1], h[2] \neq 0 \)—is given by

\[
-\sum_{t \in \{0, -c_3^*/c_1^*\}} + \sum_{t \in \mathbb{F}_p} = -2p^2 + p^3 + p \cdot \left( \# \{(z, t) \in \mathbb{F}_p^2 : z^2 = P_c(t) - p \} \right),
\]

where

\[
P_c(t) := -3F(1)F(2)F(3)(c_3^*)^{-1} \cdot t \cdot (c_1^* \cdot t + c_3^*) \cdot [F(1)t^3 - F(2)(c_2^*)^{-3}(c_1^* \cdot t + c_3^*)^3 + F(3)].
\]

Here \( \deg P_c = 5 \), since \( p \nmid c(1)^3 - c(2)^3 \). By a routine computation, the discriminant of \( P_c(t) \) simplifies—up to a harmless “unit monomial” in \( 3, \pm F(1/k), (c_3^*)^{1 \pm 1} \)—to

\[
[c(1)^3 - c(3)^3] \cdot [c(2)^3 - c(3)^3] \cdot \prod [c(1)^{3/2} \pm c(2)^{3/2} \pm c(3)^{3/2}],
\]

which is an integer coprime to \( p \) (since \( p \nmid f^{2m/2-1}F^\vee(c) \)). Thus \( z^2 = P_c(t) \) defines a smooth, affine curve over \( \mathbb{F}_p \) of genus 2.

Finally, to get the affine cone contribution, we multiply \(-2p^2 + p^3 + \cdots \) by \( p - 1 \).

All in all, for \( m = 4 \) we have

\[
|C(V_c)(\mathbb{F}_p)| = |\Lambda/p\Lambda| + (p^2 - p) - \chi(c(1)^3 c(2)^3) \cdot (p - 1) = p^{m-2} + p^{m/2} + O(p),
\]

while for \( m = 6 \) we have (by the Weil bound for \( z^2 = P_c(t) \))

\[
|C(V_c)(\mathbb{F}_p)| = |\Lambda/p\Lambda| + 3(p^3 - p^2) - \sum_{i < j} \chi(c(i)^3 c(j)^3) \cdot (p^2 - p) - 2(p^3 - p^2) + (p^2 - p) + O(p^{1/2}),
\]

which simplifies to \( |\Lambda/p\Lambda| + p^4 + O(p^{5/2}) = p^{m-2} + p^{m/2} + O(p^{(m-1)/2}) \), as desired. \( \square \)

**Remark 6.7.** By Lang-Weil for curves, we only need \( z^2 = P_c(t) \) to be absolutely irreducible over \( \mathbb{F}_p \)—not necessarily smooth. However, \( p \nmid c(k)^3 \) and \( p \nmid c(i)^3 - c(j)^3 \) remain essential throughout the proof of Lemma 6.6, without them, the bias could increase.

### 6.3. Analyzing sums to prime-power moduli

Prime powers \( p^l \), with \( l \geq 2 \), induce arithmetico-geometric behavior of a somewhat different flavor.

**Definition 6.8.** Given \( l \geq a, b \), let \( S = S_{a,b}(l) \) be the set of \( x \) modulo \( p^l \) with \( p^a \mid F(x) \) and \( p^b \mid c \cdot x \). Let \( \mathcal{Z} \) be the subset of \( S \) with \( p \mid \nabla F(x) \); let \( \mathcal{U} := S \setminus \mathcal{Z} \). Let \( \mathcal{B} \) be the subset of \( \mathcal{U} \) with \( \nabla F(x) \) and \( c \) linearly dependent over \( \mathbb{F}_p \) (with no additional condition modulo \( p^{2 \geq} \)); let \( \mathcal{G} := \mathcal{U} \setminus \mathcal{B} \). Let \( \mathcal{S}(l) := S_{a,b}(l) \), and similarly define \( \mathcal{Z}(l), \ldots \).

**Proposition 6.9 (Cf. Hoo86b, pp. 65–66, Lemma 7) and Hoo14, p. 255, (39)).** Fix \( l \geq 2 \).

If \( p \nmid F^\vee(c) \), then \( S_c(p^l) = 0 \). In general, if \( p \nmid c \), then

\[
\phi(p^l)S_c(p^l) = p^{2l} |\mathcal{B}(l)| - p^{2l+m-2} |\mathcal{B}(l - 1)|.
\]

**Proof.** Write

\[
S_c(p^l) = \sum_{a \in \mathbb{Z}/p^l} \sum_{x \in \mathbb{Z}/p^l} e_{p^l}(aF(x) + c \cdot x) = \sum_{x \in \mathbb{Z}/p^l} e_{p^l}(c \cdot x) [-p^{l-1} \cdot 1_{p^{l-1} | F(x)} + p^l \cdot 1_{p^l | F(x)}].
\]
A scalar symmetry argument gives
\[
\phi(p^l)S_c(p^l) = \sum_{\mathbf{x} \in \mathbb{Z}_p^{lm}} [-p^{l-1} \cdot 1_{p^{l-1} \cdot \mathbf{x}} + p^l \cdot 1_{p^l \cdot \mathbf{x}}] [-p^{l-1} \cdot 1_{p^{l-1} \cdot F(\mathbf{x})} + p^l \cdot 1_{p^l \cdot F(\mathbf{x})}] \\
= p^{2t-2} |S_{l-1,l-1}(l)| - p^{2t-1} [S_{l-1,l}(l)] + |S_{l,l-1}(l)| + p^{2t} |S(l)|.
\]
To think about this, write \( \mathbf{x} = \mathbf{x}_0 + p^{l-1} \mathbf{r} \) — where \( \mathbf{x}_0 \in \mathbb{Z}^m \) — so \( F(\mathbf{x}) \equiv F(\mathbf{x}_0) + p^{l-1} \nabla F(\mathbf{x}_0) \cdot \mathbf{r} \mod p^l \) and \( \mathbf{c} \cdot \mathbf{x}_0 \equiv \mathbf{c} \cdot \mathbf{x}_0 + p^{l-1} \mathbf{c} \cdot \mathbf{r} \mod p^l \).

1. Each fiber of \( S_{l-1,l-1}(l) \to \mathbb{S}(l-1) \) is “full,” of size \( p^m \): here “all lifts are valid.” So for instance, \( |Z_{l-1,l-1}(l)| = \mathbb{P}^m |Z(l-1)| \) and \( |U_{l-1,l-1}(l)| = \mathbb{P}^m |U(l-1)| \).

2. If \( p \nmid \mathbf{c} \), then each fiber of \( S_{l-1,l}(l) \to \mathbb{S}(l-1) \) is of size \( p^{m-1} \), cut out by a single affine-linear equation \( \mathbf{c} \cdot \mathbf{r} \equiv \mathbf{x} \mod p \). So \( |S_{l-1,l}(l)| = p^{m-1} |\mathbb{S}(l-1)| \).

3. If \( p | \mathbf{c} \), then each fiber of \( Z_{l,l-1}(l) \to \mathbb{Z}(l-1) \) is \( p \) times the corresponding fiber of \( Z(l) \to \mathbb{Z}(l-1) \). More precisely, either (i) both fibers are empty (i.e. \( F(\mathbf{x}_0 + p^{l-1} \mathbf{r}) \) for all \( \mathbf{r} \)); or (ii) the first fiber is “full,” of size \( p^m \), while the second fiber is of size \( p^{m-1} \) (cut out by a single equation \( \mathbf{c} \cdot \mathbf{r} \equiv \mathbf{x} \mod p \)).

For proof, note that deg \( F \geq 2 \) — so for all \( (\mathbf{x}_0, \mathbf{r}) \in \mathbb{Z}(l-1) \times \mathbb{Z}^m \), we have
\[
F(\mathbf{x}_0 + p^{l-1} \mathbf{r}) = F(\mathbf{x}_0) + p^{l-1} \nabla F(\mathbf{x}_0) \cdot p^{l-1} \mathbf{r} + O(p^{2t-2}) = F(\mathbf{x}_0) \mod p^l.
\]

4. By definition, \( p \nmid \nabla F(\mathbf{x}_0) \) for \( \mathbf{x}_0 \in \mathbb{U} \). So each fiber of \( U_{l-1,l}(l) \to \mathbb{U}(l-1) \) is of size \( p^{m-1} \), cut out by a single equation \( \nabla F(\mathbf{x}_0) \cdot \mathbf{r} \equiv \mathbf{x} \mod p \).

5. If \( p \nmid \mathbf{c} \), then each fiber of \( G(l) \to \mathbb{G}(l-1) \) is of size \( p^{m-2} \), cut out by independent equations \( \mathbf{c} \cdot \mathbf{r} \equiv \mathbf{x} \mod p \) and \( \nabla F(\mathbf{x}_0) \cdot \mathbf{r} \equiv \mathbf{x} \mod p \).

6. If \( p | \mathbf{c} \), then each fiber of \( \mathbb{B}(l) \to \mathbb{B}(l-1) \) is either empty or of size \( p^{m-1} \).

Now it easily follows — by “eliminating” \( \mathbb{Z}, \mathbb{G} \) to leave \( \mathbb{B} \subseteq \mathbb{U} \) — that \( \phi(p^l)S_c(p^l) = p^2 |\mathbb{B}(l)| - p^{2t-1} |\mathbb{B}_{l-1}(l)| = p^2 |\mathbb{B}(l)| - p^{2t+m^{-2}} |\mathbb{B}(l-1)| \), as desired. \( \Box \)

**Lemma 6.10.** In the setting of Lemma 5.5, fix an integer \( l \geq 2 \). Let \( \chi(\mathbf{r}) := (\mathbf{r}^{\mathbf{p}}) \). Then
\[
S_c(p^l) = 1_{\chi(\mathbf{c}(1)^3) \cdots \chi(\mathbf{c}(m/2)^3)} \cdot 2^{m/2-1} (p-1)p^{(m+2)/2-1},
\]
or equivalently,
\[
\tilde{S}_c(p^l) = 1_{\chi(\mathbf{c}(1)^3) \cdots \chi(\mathbf{c}(m/2)^3)} \cdot 2^{m/2-1} \phi(p^l)^{p^{-1}/2}.
\]

**Proof.** To go from the first statement to the second, write \( (p-1)p^{(m+2)/2-1} = \phi(p^l)^{p^{-m/2}} \) and then divide by \( p^{-t(1+m)/2} \). We will prove the first statement, in the equivalent form \( p^{-2}S_c(p^l) = 1_{\chi(\mathbf{c}(1)^3) \cdots \chi(\mathbf{c}(m/2)^3)} \cdot 2^{m/2-1} (p-1)p^{(m+2)/2-1} \).

To start, we reduce to simpler \( F, L \) as detailed in §6.1. We can then rewrite \( F(\mathbf{x}), \mathbf{c} \cdot \mathbf{x} \) in terms of the \( (\mathbf{h}, \mathbf{y}) \) coordinates described near the end of §6.1.

Now, recall that \( p \nmid 6F_1 \cdots F_m \) and \( p \nmid c(k)^3 := (c_k^3)/F(k) \) for all \( k \in [m/2] \). If \( \chi(\mathbf{c}(i)^3) \neq \chi(\mathbf{c}(j)^3) \), i.e. \( \chi(c_i^3/F(i)) \chi(c_j^3/F(j)) = -1 \), for some \( i, j \); then \( \mathbb{B}(1) = 0 \); in fact, there are no \( \mathbf{x} \in \mathbb{F}_p^m \setminus \{0\} \) with \( \nabla F(\mathbf{x}), \mathbf{c} \) linearly dependent over \( \mathbb{F}_p \). Thus \( S_c(p^2) = 0 \) unless \( \chi(\mathbf{c}(i)^3) = \chi(\mathbf{c}(j)^3) \) for all \( i, j \in [m/2] \) — which we assume from now on.

Since \( p \neq 2 \), there exists \( \lambda \in \mathbb{Z}_p^\times \) such that \( \lambda \cdot c_k^3/F(k) \in (\mathbb{Z}_p^\times)^2 \) for all \( k \in [m/2] \). Say \( \lambda \cdot c_k^3 = F(k)d(k)^2 \) for some choices \( d(k) \in \mathbb{Z}_p^\times \); write \( d_i = d(k) \) when \( i \in J(k) \). Then \( \mathbb{B}(1)/\mathbb{F}_p^\times \) consists of the singular \( \mathbb{F}_p \)-points of the \( m \)-dimensional projective variety \( V_c \), i.e. \( \mathbf{x} = [\pm d_i]_{i \in [m]} \) with \( F(\mathbf{x}) = 0 \). But \( p \nmid 2^{m/2-1} F^{\vee}(\mathbf{c}) \), so \( V_c \) has exactly \( 2^{m/2-1} \) singular
points \([x] \in V_c(\mathbb{F}_p)\), which—since \(\pm \bar{d}_i \in \mathbb{F}_p\) for all \(i \in [m]\)—must all lie in \(V_c(\mathbb{F}_p)\). Explicitly, these \(2^{m/2-1}\) points \([x]\) arise from the sign choices for which \(h = 0\).

We now seek to count \(\mathcal{B}(l)\) for \(l \geq 2\)—i.e. \(p\)-adic lifts (for the equation \(F(x) = c \cdot x = 0\)) of the \(2^{m/2-1}\) singular points \(x \in \mathcal{B}(1)/\mathbb{F}^\times_p\). But \(p \nmid x\) for all \(x \in \mathcal{B}(l)\), so \(\mathbb{Z}_p^\times\) acts freely on \(\mathcal{B}(l)\). Thus it suffices to count \(\mathcal{B}'(l) := \mathcal{B}(l)/\mathbb{Z}_p^\times\) instead. Fix a point \(x \in \mathcal{B}(1)/\mathbb{F}^\times_p\), say given in \((h, y)\) coordinates by \(h \equiv 0 \mod p\) and \(y[k] \equiv d(k) \mod p\). Replacing \(c_k^*\) with \(F(k)d(k)/\lambda \in \mathbb{Z}_p^\times\), we want to count lifts to modulus \(p^l\) for the system

\[
\sum_{k \in [m/2]} F(k)h[k]y[k]^2 \equiv \rho \quad -3\sum_{k \in [m/2]} F(k)h[k]^3 \quad \land \quad \sum_{k \in [m/2]} F(k)d(k)^2h[k] \equiv \rho \quad 0.
\]

Denote by \(\mathcal{B}_d(l)\) and \(\mathcal{B}'_d(l) := \mathcal{B}_d(l)/\mathbb{Z}_p^\times\) the sets of such lifts (over our fixed \(x \in \mathcal{B}(1)/\mathbb{F}^\times_p\)).

Fix an affine chart (i.e. representatives in \(\mathcal{B}'_d(l)\)) by setting \(y[m/2] = d(m/2)\) identically over \(\mathbb{Z}_p\). Write \(h[k] = p^s h_s[k]\) and \(y[k] = d(k) + p^s y_s[k]\) (with \(s = 1\) for now, but all \(s \geq 1\) to be relevant below), so \(y_s[m/2] = 0\). Then our system becomes

\[
\sum_{k \neq m/2} F(k)h_s[k](2d(k)y_s[k] + p^s y_s[k]^2) \equiv \rho \quad -3p^s\sum_{k} F(k)h_s[k]^3 \quad \land \quad \sum_{k} F(k)d(k)^2h_s[k] \equiv \rho \quad 0.
\]

So \(|\mathcal{B}'_d(l)| = p^{m-2}|\mathcal{A}_s(l-2)|\), where \(\mathcal{A}_s(l)\) is the (non-homogeneous, affine) system

\[
\sum_{k \neq m/2} F(k)h_s[k](2d(k)y_s[k] + p^s y_s[k]^2) \equiv \rho \quad -3p^s\sum_{k} F(k)h_s[k]^3 \quad \land \quad \sum_{k} F(k)d(k)^2h_s[k] \equiv \rho \quad 0.
\]

Fix \(s \geq 1\). Clearly \(|\mathcal{A}_s(0)| = 1\), while \(\mathcal{A}_s(1)\) is isomorphic to a cone over a smooth\(^4\) quadric in \(m - 2\) variables (i.e. in \(\mathbb{P}^m_\ast\), of even dimension \(m_\ast - 1\)) with discriminant in \((-1)^{m/2-1}(\mathbb{F}_p)^\times\), so \(|\mathcal{A}_s(1)| = p^{m_\ast} + (p-1)p^{(m_\ast-1)/2}\). For \(l \geq 2\), the origin of the cone \(\mathcal{A}_s(1)\) contributes \(p^{m_\ast-2}|\mathcal{A}_{s+1}(l-2)|\) points to \(\mathcal{A}_s(l)\), while points away from the origin (i.e. smooth points!) lift uniformly to a total of \((|\mathcal{A}_s(1)| - 1) \cdot p^{(l-1)m_\ast}\) points of \(\mathcal{A}_s(l)\). Thus

\[
|\mathcal{A}_s(l)| = p^{m_\ast+1}|\mathcal{A}_{s+1}(l-2)| + (|\mathcal{A}_s(1)| - 1) \cdot p^{(l-1)m_\ast}
\]

for \(s \geq 1\) and \(l \geq 2\). (The same holds for \(l = 1\), provided we interpret \(|\mathcal{A}_s(-1)| := p^{-(m_\ast+1)}\).)

By induction on \(l \geq 0\) (with base cases \(l = 0, 1\)), we immediately find that \(|\mathcal{A}_s(l)|\) is independent of the choice of \(\lambda\) and the \(p\)-adic square roots \(d(k)\); furthermore, \(|\mathcal{A}_s(l)| = |\mathcal{A}_s(1)|\) for all \(s \geq 1\), i.e. there is no \(d\)-dependence or \(s\)-dependence!

Finally, by symmetry, \(|\mathcal{B}'(l)| = 2^{m/2-1}|\mathcal{B}'_d(l)| = 2^{m/2-1}p^{m_\ast-2}|\mathcal{A}_s(l-2)|\) for all \(l \geq 1\). (For \(l = 1\), recall \(|\mathcal{A}_1(-1)| := p^{-(m_\ast+1)} = p^{-(m-2)}\).) Thus

\[
p^{-2l}S_c(p^l) = \phi(p^l)^{-1}|\mathcal{B}(l)| - p^{-2l}\phi(p^l)^{-1}|\mathcal{B}(l-1)|
\]

\[
= |\mathcal{B}'(l)| - p^{m_\ast}|\mathcal{B}'(l-1)| = 2^{m/2-1}p^{m_\ast-2}(|\mathcal{A}_s(l-2)| - p^{m_\ast}|\mathcal{A}_s(l-3)|)
\]

for \(l \geq 2\); here we have used the fact that \(\mathbb{Z}_p^\times\) acts freely on \(\mathcal{B}(l)\). To prove \(p^{-2l}S_c(p^l) = 2^{m/2-1}(p-1)p^{(m-2)/2-1}\) (the desired answer), it remains to show that

\[
|\mathcal{A}_s(l)| - p^{m_\ast}|\mathcal{A}_s(l-1)| = (p-1)p^{(m-2)/2-1} = (p-1)p^{(m_\ast+1)/2-1}
\]

for \(l \geq 0\). To this end, we compute \(|\mathcal{A}_1(l)| - p^{m_\ast}|\mathcal{A}_1(l-1)|\) (recursively if \(l \geq 2\)) to get

\begin{align*}
(1) & \quad 1 - p^{m_\ast} \cdot p^{-(m_\ast+1)} = 1 - p^{-1}, \text{ i.e. } (p-1)p^{-1}, \text{ for } l = 0; \\
(2) & \quad \left[p^{m_\ast} + (p-1)p^{(m_\ast-1)/2}\right] - p^{m_\ast} \cdot 1 = (p-1)p^{(m_\ast-1)/2}, \text{ i.e. } (p-1)p^{(m_\ast+1)/2-1}, \text{ for } l = 1;
\end{align*}

\(^4h_s[m/2]\) is determined by the remaining \(h_s[k]\), and \(\sum_{k \neq m/2} F(k)d(k) \cdot h_s[k]y_s[k] = 0\) is smooth.
(3) \( p^{m+1}(|A_1(l-2)| - p^{m}|A_1(l-3)|) \) for \( l \geq 2 \), since \( p^{l-1|m|} = p^{m} \cdot p^{(l-2)m} \).

(Equivalently in terms of \( S_c \), this means that \( p^{-2l}S_c(p^l) = p^{m+1}p^{-2(l-2)}S_c(p^{l-2}) \), i.e. \( S_c(p^l) = p^{m+2}S_c(p^{l-2}) \), for \( l \geq 4 \).)

By induction on \( l \geq 0 \), we are done, since \( p^{m+1} \cdot p^{(l-2)(m+1)/2-1} = p^{l(m+1)/2-1} \). \( \square \)

Remark 6.11. By induction on \( l \geq 0 \) (with base cases \( l = 0, 1 \)), we can explicitly compute

\[
|A_2(l)| = p^{m} + (p-1)p^{(l+1)/2-1}p^{(m-1)/2-1} - 1 \]

(also valid for \( l = -1 \)). Furthermore,

\[
|E_d'(l)| = p^{m+1}|A_1(l-2)| = p^{(l-1)m+1} + (p-1)p^{(l+1)/2-1}p^{(l-2)(m-1)/2-1} - 1 \]

for \( l \geq 1 \), since \( p^{m+1}p^{(l-2)m} = p^{(l-1)m+1} \) and \( p^{m+1}p^{(l-2)(m+1)/2} - 1 = p^{(l+1)/2-1} \).

APPENDIX A. Algebraic geometry background

For non-diagonal \( F \), we need some classical results on the gradient map \( \nabla F \), its image, and its ramification. For diagonal \( F \), a more explicit analysis is possible (see \( \S 2.2 \).

A.1. The dual hypersurface and the discriminant form. Since \( V \) is smooth, the polar map \( [\nabla F]: \mathbb{P}^{m-1} \to \mathbb{P}^{m-1} \) is regular, i.e. defined everywhere. (It would be better to write \( \mathbb{P}^{m-1} \to (\mathbb{P}^{m-1})^\vee \), but no confusion should arise.) The map \( [\nabla F] \) is finite of degree \( 2^{m-1} \) between irreducible equidimensional projective varieties, hence surjective. Since \( \mathbb{P}^{m-1} \) is smooth, \( [\nabla F] \) must then be flat (by “miracle flatness”).

Upon restricting \( [\nabla F] \) to \( V \), we get the finite surjective Gauss map \( \gamma: V \to V^\vee \), where \( V^\vee \subseteq (\mathbb{P}^{m-1})^\vee \) denotes the dual variety of \( V \), i.e. the closure of the union of hyperplanes tangent to the smooth locus \( V_{sm} \) of \( V \). (For us, \( V \) is pure of codimension \( 1 \), and \( V_{sm} = V \). So \( V^\vee = [\nabla F]V_{sm} = [\nabla F]V = [\nabla F]V \). Hence \( \gamma \) is indeed well-defined and surjective; and \( \gamma \) is finite because it is a quasi-finite map between projective varieties.)

Here \( V/Q \) is irreducible over \( \mathbb{C} \), so \( V^\vee \) must be too, by [GKZ94] p. 15, Proposition 1.3. Thus \( V^\vee/Q \) is a projective hypersurface, cut out by an absolutely irreducible form \( F^\vee \in Q[c] \).

(At least for diagonal \( F \), one can easily explicitly compute \( F^\vee \); see \( \S 2.2 \).

Remark A.1. The notation \( F^\vee \) is convenient for us, but likely not standard.

**Proposition A.2** (Classical). Here \( \deg F^\vee = 3 \cdot 2^m \). Given \( c \in \mathbb{C}^m \setminus \{0\} \), we have \( F^\vee(c) = 0 \) if and only if \( (V_c)^\vee \) is singular. Furthermore, we may choose \( F^\vee \in \mathbb{Z}[c] \) so that for all \( c \in \mathbb{Z}^m \) and primes \( p \nmid F^\vee(c) \), the special fiber \( (V_c)_{\mathbb{F}_p} \) is smooth of dimension \( m - 3 \).

**Remark A.3.** Proposition A.2 together with the absolute irreducibility of \( F^\vee \), implies that [Wan21, §1.2, Proposition-Definition 1.8] is valid for all \( F \) smooth over \( \mathbb{Q} \)—even though only provides a proof when \( F \) is diagonal.

**Proof sketch for Proposition A.2.** For the degree computation, see e.g. [Dol12, p. 33, (1.47)].

For the “singular hyperplane section” interpretation over \( K = \mathbb{C} \) or over \( K = \mathbb{F}_p \) with \( p > p_0 \), see e.g. [Hoo14, pp. 244–245]. (The key is that if \( V_K \) is smooth and \( c \in K^m \setminus \{0\} \), then \( (V_K)^\vee \) is singular if and only if there exists \( [x] \in (V_K)^\vee(K) \) such that \( [c] = [\nabla F](|x|) \).

We can then “absorb” the primes \( p \leq p_0 \) (once and for all) into the polynomial \( F^\vee \).

Alternatively, for a convenient reference touching briefly on both matters above (at least in characteristic 0), see [Bro09, pp. 115–116, §7.1]. \( \square \)
It is known that \((V^\vee)^\vee = V\) \cite[Reflexivity Theorem]{Dol12}. Furthermore, the definition of \(V^\vee\) (together with the fact that \(V, V^\vee\) are hypersurfaces with \(V\) irreducible) implies the divisibility \(F(x) \mid F^\vee(\nabla F(x))\), and reflexivity (together with the fact that \(V^\vee, V\) are hypersurfaces with \(V^\vee\) irreducible) similarly implies the divisibility \(F^\vee(c) \mid F(\nabla F^\vee(c))\). These (symmetric!) divisibilities capture much of the basic duality theory for \(V\).

The apparent symmetry between \(V, V^\vee\) thus far is deceptive, however. What complicates matters is that \(V^\vee\), or equivalently \(F^\vee\), must be singular if \(\deg F \geq 3\). (Otherwise, \((V^\vee)^\vee\) would be a hypersurface of degree larger than \(\deg F\), contradicting reflexivity.) Hence the polar map \([\nabla F^\vee] : \mathbb{P}^{m-1} \to \mathbb{P}^{m-1}\) is only a rational map, defined away from \(\text{Sing}(V^\vee)\) (a proper closed subset of \(V^\vee\)).

Nonetheless, given smooth points \([x] \in V\) and \([c] \in V^\vee\), the biduality theorem says that \([\nabla F(x)] = [c] \iff [\nabla F^\vee(c)] = [x]\). (See e.g. \cite[Proposition 2.2]{Dol12}.) For \(V_{\text{sm}} = V\), so biduality implies that \([\nabla F], [\nabla F^\vee]\) restrict to inverse morphisms between \(V \setminus [\nabla F]^{-1}(\text{Sing}(V^\vee))\) and \(V^\vee \setminus \text{Sing}(V^\vee)\).

**Remark A.4.** The map \(\gamma : V \to V^\vee\) is finite surjective (and birational, i.e. of degree 1), but not necessarily flat (or equivalently, locally free). In fact, the finite \(\mathcal{O}_{V^\vee}\)-algebra \(\gamma_*\mathcal{O}_V\) is isomorphic to \(\mathcal{O}_{V^\vee}\) generically over \(V^\vee\), but not necessarily everywhere—for instance, the geometric fiber \(V \times_k k(p) = \text{Spec}(\mathcal{O}_V p \otimes_{\mathcal{O}_{V^\vee}, p} k(p))\), and in fact the analogous set-theoretic geometric fiber, may have size \(\geq 2\) at some point \(p \in \text{Sing}(V^\vee)\).

**Question A.5.** What is known about the fiber of \(\gamma\) over a singular point \([c] \in V^\vee\)?

Presumably the singular structure of \(V_c\) should play a role. Perhaps works of Aluffi and Cukierman (such as \cite{AC93, Alu95}) can help to give a precise statement.

### A.2. Ramification and the Hessian.

\([\nabla F]\) is a finite surjective morphism of smooth varieties, so its ramification theory is well-behaved. By \cite[Proposition 1.2.1]{Dol12},

1. the ramification divisor of \([\nabla F] : \mathbb{P}^{m-1} \to \mathbb{P}^{m-1}\) is the Hessian hypersurface \(\text{hess}(V) \subseteq \mathbb{P}^{m-1}\); and

2. its image, the branch divisor, is \(\text{St}(V)^\vee\), the dual of the Steinerian hypersurface \(\text{St}(V)\) (with \(\text{St}(V)\) defined scheme-theoretically as in \cite[§1.1.6]{Dol12}).

In particular, both \(\text{hess}(V), \text{St}(V)^\vee\) are (possibly reducible or non-reduced) projective hypersurfaces in \(\mathbb{P}^{m-1}\).

When studying the Hasse principle (or even weak approximation), one can certainly localize before counting points. Thus a useful fact worth mentioning now is the following:

**Proposition A.6** \cite[Lemma 1]{Hoo88}. If \(V\) is a smooth cubic, then \(V \not\subseteq \text{hess}(V)\).

**Remark A.7.** Thus on an arbitrary smooth cubic \(V\), as in \cite{Hoo88} or \cite{Hoo14} for instance, it is always non-vacuous to count points on \(V\) with nonvanishing Hessian determinant.

**Remark A.8** \cite[Remarks in paragraph before Lemma 1]{Hoo88}. The intersection \(V \cap \text{hess}(V)\) consists of inflection points if \(m = 3\), and of parabolic points if \(m \geq 4\) \cite[Theorem 1.1.20]{Dol12}. But it does not seem easy to find a general reference proving the existence of non-inflection or non-parabolic points on \(V\) (according as \(m = 3\) or \(m \geq 4\)).

The following stronger technical question comes up in Proposition \[2.2\] although we happen to be able to sidestep it there (as remarked after the proof of Proposition \[2.2\]).
Question A.9. Is it always true (for smooth cubic \( V \)) that \( V^\vee \not\subseteq \text{St}(V)^\vee \)?

Remark A.10. If \( V \subseteq \text{hess}(V) \), then applying \([\nabla F]\) would imply \( V^\vee \subseteq \text{St}(V)^\vee \). Thus an affirmative answer to the question would give another proof of Proposition A.6.

APPENDIX B. ISOLATING THE SINGULAR SERIES

Lemma B.1. If \( m \geq 4 \), then \( \sum_{n \geq N} n^{1-m/2} \max_{n_* | n} n_*^{1/2} |\tilde{S}_0(n_*)| \lesssim \epsilon \cdot N^{(m-4)/3} \).

Remark B.2. The \( \max_{n_* | n} \) version is not really used in the present Appendix (which only needs \( 1_{n_* = n} \)), but rather in §5 (to help separate \( c = 0 \) from \( c \neq 0 \)).

Proof. We have \( \tilde{S}_0(n_*) \ll n_*^{1/2 + \epsilon} \text{cub}(n_*)^{m/6} \) by [Hoo88, p. 95, (170)]. Clearly \( \text{cub}(n_*) | \text{cub}(n) \) for \( n_* | n \), so factoring out the cube-full part \( n_3 \) of \( n \asymp N \), we get a bound of

\[
\sum_{n \geq N} n^{1-m/2+\epsilon} \text{cub}(n)^{m/6} \ll \epsilon \cdot N^{1-m/2} \approx \epsilon \cdot N \sum_{n_3 \leq N} n_3^{1-m/3} \cdot (N/n_3)^{2-m/2} \\
\approx \epsilon \sum_{N_3 \leq N} N_3^{1/3} \cdot N_3^{m/6-1} \cdot N_3^{2-m/2} 
\]

since \( 1/3 + (m/6 - 1) \geq 0 \). (Here \( N_3 \) is a dyadic parameter, ranging over powers of 2.) \( \square \)

Remark B.3. The statement above should still hold if \( 0 \) is replaced with any fixed \( c \in \mathbb{Z}^m \). On the other hand, \( I_0(n) \) behaves rather differently than \( I_c(n) \) for \( c \neq 0 \).

The lemma guarantees absolute convergence of the singular series

\[
\mathcal{G}_F = \sum_{n \geq 1} n^{-m} S_0(n) = \sum_{n \geq 1} n^{(1-m)/2} \tilde{S}_0(n)
\]

for \( m \geq 5 \). But as noted in [Wan21, §4], certain results from [HB96] imply that

\[
\sigma_{\infty,F,w} = \lim_{\epsilon \to 0} (2\epsilon)^{-1} \int_{|F(x)| \leq \epsilon} dx \ w(x) \ll_{F,w} 1,
\]

and—if we let \( \tilde{I}_c(n) := X^{-m} I_c(n) \) as in [Wan21, §2.1]—that \( \tilde{I}_0(n) = \sigma_{\infty,F,w} + O_A((n/Y)^A) \) for all \( n \geq 1 \). Here \( \tilde{I}_0(n) = 0 \) for \( n \gtrsim Y \), so

\[
\sum_{n \geq 1} n^{-m} S_0(n) \cdot \tilde{I}_0(n) = \mathcal{G}_F \cdot \sigma_{\infty,F,w} + O_{A,\epsilon} \left( \sum_{N \leq Y} N^{(m-4)/3+\epsilon} \cdot (N/Y)^A + \sum_{N \geq Y} N^{(m-4)/3+\epsilon} \right),
\]

with \( N \in 2^\mathbb{N} \). If we choose \( A = (m-4)/3 \), then both \( N \)-sums are geometric series peaking at \( N \asymp Y \), giving a final error term of \( O_\epsilon (Y^{(m-4)/3+\epsilon}) = O_\epsilon (X^{(m-4)/3+\epsilon}) \).

Thus (since \( Y^2 = X^{\deg F} \) and \( I_0(n) = X^m \tilde{I}_0(n) \))

\[
Y^{-2} \sum_{n \geq 1} n^{-m} S_0(n) I_0(n) = X^{m-3} \cdot [\sigma_{\infty,F,w} \mathcal{G}_F + O_\epsilon (X^{(m-4)/2+\epsilon})].
\]

If \( m \geq 5 \), then \( (4-m)/2 < 0 \), so we have isolated the expected singular series from \( c = 0 \).
Remark B.4. One may well do better by analyzing the Dirichlet series $\sum_{n \geq 1} n^{-s} S_0(n)$, using the derivative bound $\partial^k I_0(n) \ll_k n^{-k} X^m$ (stated in [HB96, p. 183, Lemma 16] for $k = 0, 1$, but valid for $k \geq 0$ with little change in proof). It would be interesting to extend the above analysis to the case $m = 4$ (with powers of $\log X$ expected to occur for certain $F$’s).

Acknowledgements

This paper complements Paper III; many of my acknowledgements there apply here as well. I also thank my advisor, Peter Sarnak, for many helpful suggestions and questions on the exposition, references, assumptions, and scope of (various drafts of) the present work.

References

[AC93] P. Aluffi and F. Cukierman, Multiplicities of discriminants, Manuscripta Math. 78 (1993), no. 3, 245–258. MR1206155

[Alu95] P. Aluffi, Singular schemes of hypersurfaces, Duke Math. J. 80 (1995), no. 2, 325–351. MR1369396

[BHB06] T. D. Browning and D. R. Heath-Brown, The density of rational points on non-singular hypersurfaces. II, Proc. London Math. Soc. (3) 93 (2006), no. 2, 273–303. With an appendix by J. M. Starr. MR2251154

[Bom09] E. Bombieri, Problems and results on the distribution of algebraic points on algebraic varieties, J. Théor. Nombres Bordeaux 21 (2009), no. 1, 41–57. MR2537702

[Bro09] T. D. Browning, Quantitative arithmetic of projective varieties, Progress in Mathematics, vol. 277, Birkhäuser Verlag, Basel, 2009. MR2559866

[BW19] J. Brüdern and T. D. Wooley, An instance where the major and minor arc integrals meet, Bull. Lond. Math. Soc. 51 (2019), no. 6, 1113–1128. MR4041016

[Deb03] O. Debarre, Lines on smooth hypersurfaces (November 2003), available at https://www.math.ens.fr/~debarre/Lines_hypersurfaces.pdf

[DFI93] W. Duke, J. B. Friedlander, and H. Iwaniec, Bounds for automorphic $L$-functions, Invent. Math. 112 (1993), no. 1, 1–8. MR1207474

[Die17] R. Dietmann, On the $h$-invariant of cubic forms, and systems of cubic forms, Q. J. Math. 68 (2017), no. 2, 485–501. MR3667211

[DL62] H. Davenport and D. J. Lewis, Exponential sums in many variables, Amer. J. Math. 84 (1962), 649–665. MR144862

[DL64] _____, Non-homogeneous cubic equations, J. London Math. Soc. 39 (1964), 657–671. MR167458

[Dol12] I. V. Dolgachev, Classical algebraic geometry, Cambridge University Press, Cambridge, 2012. A modern view. MR2964027

[GKZ94] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, Discriminants, resultants, and multidimensional determinants, Mathematics: Theory & Applications, Birkhäuser Boston, Inc., Boston, MA, 1994. MR1264417

[HB83] D. R. Heath-Brown, Cubic forms in ten variables, Proc. London Math. Soc. (3) 47 (1983), no. 2, 225–257. MR703978

[HB96] _____, A new form of the circle method, and its application to quadratic forms, J. Reine Angew. Math. 481 (1996), 149–206. MR1421949

[HB98] _____, The circle method and diagonal cubic forms, R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci. 356 (1998), no. 1738, 673–699. MR1620820

[Hoo14] C. Hooley, On octonary cubic forms, Proc. Lond. Math. Soc. (3) 109 (2014), no. 1, 241–281. MR3237742

[Hoo15] _____, On octonary cubic forms: II, Bull. Lond. Math. Soc. 47 (2015), no. 1, 85–94. MR3312967

This work was partially supported by NSF grant DMS-1802211.
[Hoo86a] ______, On some topics connected with Waring’s problem, J. Reine Angew. Math. 369 (1986), 110–153. MR850631
[Hoo86b] ______, On Waring’s problem, Acta Math. 157 (1986), no. 1-2, 49–97. MR857679
[Hoo88] ______, On nonary cubic forms, J. Reine Angew. Math. 386 (1988), 32–98. MR936992
[Jah14] J. Jahnel, Brauer groups, Tamagawa measures, and rational points on algebraic varieties, Mathematical Surveys and Monographs, vol. 198, American Mathematical Society, Providence, RI, 2014. MR3242964
[Klo26] H. D. Kloosterman, On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$, Acta Math. 49 (1926), no. 3-4, 407–464. MR1555249
[Kon02] A. Kontogeorgis, Automorphisms of Fermat-like varieties, Manuscripta Math. 107 (2002), no. 2, 187–205. MR1894739
[MV19] O. Marmon and P. Vishe, On the Hasse principle for quartic hypersurfaces, Duke Math. J. 168 (2019), no. 14, 2727–2799. MR4012347
[Shi88] T. Shioda, Arithmetic and geometry of Fermat curves, Algebraic Geometry Seminar (Singapore, 1987), 1988, pp. 95–102. MR966448
[Wan21] V. Y. Wang, Diagonal cubic forms and the large sieve (August 7, 2021).
[Woo19] T. D. Wooley, A slice or two of a diagonal cubic: arithmetic stratification via the circle method, Joint IAS/Princeton University Number Theory Seminar, 2019. URL: https://youtu.be/r1nzdaFCMjA? t=1453 (accessed 2020-11-08).