Algebra & Number Theory

Volume 18
2024
No. 8

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2024 is US $525/year for the electronic version, and $770/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 18:8 (2024)
https://doi.org/10.2140/ant.2024.18.1403

The strong maximal rank conjecture and moduli spaces of curves

Fu Liu, Brian Osserman, Montserrat Teixidor i Bigas and Naizhen Zhang

Building on recent work of the authors, we use degenerations to chains of elliptic curves to prove two cases of the Aprodu–Farkas strong maximal rank conjecture, in genus 22 and 23. This constitutes a major step forward in Farkas' program to prove that the moduli spaces of curves of genus 22 and 23 are of general type. Our techniques involve a combination of the Eisenbud–Harris theory of limit linear series, and the notion of linked linear series developed by Osserman.

1. Introduction

The moduli space $\mathcal{M}_g$ of curves of fixed genus g is one of the most classically studied in algebraic geometry. Going back to Severi and based on examples in low genus there was a general expectation that these moduli spaces ought to be unirational. However, groundbreaking work of Harris, Mumford and Eisenbud [Harris and Mumford 1982; Harris 1984; Eisenbud and Harris 1987] in the 1980s showed that not only is $\mathcal{M}_g$ not unirational for large g, but it is in fact of general type for $g \geq 24$. Their fundamental technique was to compute the classes of certain explicit effective divisors on $\overline{\mathcal{M}}_g$ arising from Brill–Noether theory, and use this to show that the canonical class of $\overline{\mathcal{M}}_g$ can be written as the sum of an ample and an effective divisor. The particular families of divisors they considered were computable in all applicable genera, but did not suffice to prove that $\mathcal{M}_g$ is of general type for $g \leq 23$. For the last thirty

Fu Liu was partially supported by a grant from the Simons Foundation #426756. Brian Osserman was partially supported by a grant from the Simons Foundation #279151. Naizhen Zhang was supported by the Methusalem Project Pure Mathematics at KU Leuven.
MSC2010: primary 14H10; secondary 14H51, 14D06.
Keywords: maximal rank, moduli space, algebraic curve, general type.

years, no new cases have been proved of $\mathcal{M}_g$ being of general type. Over a decade ago, Farkas [2009a, §7; 2009b, §4; 2009c] proposed new families of expected divisors on $\mathcal{M}_g$ as an approach to showing that $\mathcal{M}_{22}$ and $\mathcal{M}_{23}$ are of general type. Let $\mathcal{D}_g \subseteq \mathcal{M}_g$ consist of curves X of genus g which admit a $\mathfrak{g}^6_d$ such that the resulting image of X in $\mathbb{P}^6$ lies on a quadric hypersurface. Farkas computed "virtual classes" for these expected divisors $\mathcal{D}_g$ in [Farkas 2009a] for genus 22 and in [Farkas 2018] for genus 23, and in both cases found that the classes satisfy the necessary inequalities to conclude that $\mathcal{M}_{22}$ and $\mathcal{M}_{23}$ are of general type, provided that they are indeed represented by effective divisors.

In order to conclude that $\mathcal{M}_g$ is of general type for $g = 22$ or 23, one has to check two statements: first, that $\mathcal{D}_g$ yields an effective divisor, or equivalently, that $\mathcal{D}_g \subsetneq \mathcal{M}_g$; and second, that the class induced by $\mathcal{D}_g$ agrees with the class previously computed by Farkas, or equivalently, that the subset of $\mathcal{D}_g$ consisting of curves carrying *infinitely* many $\mathfrak{g}^6_d$s whose image lie on a quadric occurs in codimension strictly higher than 1.

In this paper, we prove the first of these two statements, for both $g = 22$ and $g = 23$. An independent proof of this result has been obtained by Jensen and Payne [2018] using a tropical approach. Their tropical proof has now been merged in [Farkas et al. 2020] with the prior results of Farkas and with the missing piece that the map from the space of linear series to the moduli space of curves does not have infinite fibers over a divisorial component of the image. This completes the proof that $\mathcal{M}_{22}$, $\mathcal{M}_{23}$ are of general type.

Our main theorem is thus the following:

Theorem 1.1. *In characteristic* 0, *the loci $\mathcal{D}_{22}$ and $\mathcal{D}_{23}$ are proper subsets of $\mathcal{M}_{22}$ and $\mathcal{M}_{23}$ respectively.*

Our proof goes through unmodified for characteristic $p \geq 29$, and our techniques can in principle be applied to lower characteristics as well, but due to characteristic restrictions on the application to the geometry of $\mathcal{M}_g$, we have not pursued this. See Remark 1.3 below.

The divisor $\mathcal{D}_g \subsetneq \mathcal{M}_g$ can be presented as a particular case of a more general (conjectural) subsets of $\mathcal{M}_g$: With applications to moduli,spaces of curves in mind, Aprodu and Farkas [2011, Conjecture 5.4] proposed a "strong maximal rank conjecture", about ranks of multiplication maps of line bundles on curves. Specifically, given a linear series $(\mathscr{L}, V)$ on a curve X, we have the multiplication map

$$\operatorname{Sym}^2 V \to \Gamma(X, \mathscr{L}^{\otimes 2}). \tag{1-1}$$

Note that the source has dimension $\binom{r+2}{2}$, and assuming X is Petri-general, the target has dimension $2d + 1 - g$. The image of X under the linear series lies on a quadric if and only if (1-1) has a nonzero kernel. The classical maximal rank conjecture asserts that if $r \geq 3$, for a general X and a general $\mathfrak{g}^r_d$ on X, the map (1-1) should always be injective or surjective (and similarly for the higher-order multiplication maps). Many special cases of this were proved by various people; we omit discussion of most of these, but mention that the case of quadrics was first proved by Ballico [2012]. Subsequent proofs were given by Jensen and Payne [2016] using a tropical approach, and by the present authors [Liu et al. 2021] using a degeneration to a chain of genus-1 curves. Recently, Larson [2017; 2020] has proved the full classical maximal rank conjecture.

Since the failure of (1-1) to have maximal rank is a determinantal condition, the strong maximal rank conjecture of Aprodu and Farkas is the following:

Conjecture 1.2 [Aprodu and Farkas 2011, Conjecture 5.4]. *Set* $\rho := g-(r+1)(g+r-d)$.

On a general curve of genus g, *if* $\rho < r-2$, *the locus of* $\mathfrak{g}^r_d$*s for which* (1-1) *fails to have maximal rank is equal to the expected determinantal codimension, which is* $1+\left|\binom{r+2}{2}-(2d+1-g)\right|$. *In particular, when this expected codimension exceeds* ρ, *every linear series on* X *should have maximal rank.*[1]

The strong maximal rank conjecture remains wide open, even in the case of quadrics. The only cases solved (to our knowledge) are for $k=2,\ d\le g+1$ (see [Teixidor i Bigas 2003]) or for Brill–Noether number $\rho=0$, because it is equivalent to the (weak) maximal rank conjecture.

For the divisors $\mathcal{D}_g$, we compute that $\rho=g-21=(2d+1-g)-\binom{r+2}{2}$, so in this case Conjecture 1.2 predicts that every linear series on the generic curve should yield (1-1) of maximal rank, and more specifically, should have injective multiplication map, just as we prove in Theorem 1.1 for the cases $g=22, 23$.

Our proof builds on the ideas introduced in [Liu et al. 2021], which combine the Eisenbud–Harris theory of limit linear series with ideas from the theory of linked linear series introduced by the second author [Osserman 2006; 2014]. We start with a limit linear series on a chain X_0 of genus-1 curves, and describe a collection of global sections living in different multidegrees on X_0. We then take tensors of these sections and consider their image in a carefully chosen multidegree, showing that they have the correct-dimensional span. The first major difficulty is that while we can choose the curve, we have to consider all possible limit linear series. As a consequence, we cannot ignore degenerate limit linear series (which occur already in codimension 1). We systematically use ideas from linked linear series to prove that when $\rho=1$ or $\rho=2$ we can always produce global sections of certain prescribed forms which must lie in the specialization of the family of linear series.

The structure of the paper is as follows. In Section 2 we consider certain maps from genus-1 curves to projective spaces which arise naturally from tensor squares of linear series, and show that these are nondegenerate morphisms in cases of interest. In Section 3, we review the Eisenbud–Harris theory of limit linear series, and the related theory of linked linear series introduced by the second author. In Section 4, we describe the possible structures of linked linear series lying over a given limit linear series in the cases that appear when $\rho\le 2$ (see also Proposition 5.3). In Sections 5 and 6, we give a criterion for certain collection of sections in the tensor square of a limit linear series to be linearly independent. In Section 7, we improve and apply this criterion to a family of examples with $r=6$, which include the genus-22 and genus-23 cases of interest for the proof of Theorem 1.1. In Section 8, we focus on the behavior of degenerate sections under tensor product. For $\rho=2$ (that is, genus 23), the situation is quite a bit more complicated than for $\rho=1$ (i.e., genus 22). To handle the degenerate cases, we consider in Section 8 variant multidegrees which depend more tightly on the limit linear series in question, and (partially inspired by the earlier work of Jensen and Payne [2016] on a tropical approach to the classical

[1] In fact, Aprodu and Farkas also include higher-degree multiplication maps in their conjecture. Farkas and Ortega [2011] subsequently relax the $\rho < r-2$ hypothesis in cases such as ours, where ρ is less than the expected codimension.

maximal rank conjecture), we also consider families of curves with highly specialized directions of approach, which gives us further control over the behavior of the global sections in different multidegrees. Finally, in Section 9, we complete the proof of Theorem 1.1.

We expect that the tools we develop here will lead to proofs of specific cases of the strong maximal rank conjecture of geometric relevance. In the tropical setting, this has been carried out in the case $g = 13$ (see [Farkas et al. 2024]). We have written the different parts of the argument to be independent of r and/or ρ wherever this does not lead to unnecessary complication. In particular, Theorem 9.1 has been stated in greater generality than what we need to prove the results for $g = 22$, $g = 23$. The nature of our approach also allows for proving cases of the maximal rank conjecture where the expected codimension does not exceed ρ, so that the locus of linear series which do not have maximal rank is nonempty. Our approach should also be useful in other questions involving multiplication maps for linear series, such as the conjecture of Bakker and Farkas [2018, Remark 14], which was motivated by connections to higher-rank Brill–Noether theory. Their conjecture treats a certain specific family of cases, but with products of distinct linear series in place of symmetric squares of a fixed one. In addition, our work in Section 2 on nondegeneracy of certain morphisms from genus-1 curves to projective spaces and in Section 4 on the structure of exact linked linear series is likely to be useful in other settings as well.

In a different direction, the ideas and approach of this paper can be used in the context of vector bundles (see [Teixidor i Bigas 2023]). Brill–Noether related vector bundle problems cannot at the moment be treated with tropical techniques as there is no satisfactory theory of sections of vector bundles in the tropical settings.

Remark 1.3. We mention that although we impose characteristic-0 hypotheses in our main theorem, these do not appear to be essential. Nearly everything we do is characteristic-independent, but we use a characteristic-dependent result (Theorem 3.4 below) of Eisenbud and Harris to simplify the situation slightly by restricting our attention to "refined" limit linear series (Definition 3.1 below). In fact, the only characteristic dependence in Theorem 3.4 is the use of the Plücker inequality, which still holds in characteristic p and degree d when $p > d$; see for instance Proposition 2.4 and Corollary 2.5 of [Osserman 2006]. Thus, our proof of Theorem 1.1 extends as written to characteristic $p > 25$ for $g = 22$ and $p > 26$ for $g = 23$.

Moreover, since our key specialization result (Proposition 3.10 below) on linked linear series applies in arbitrary characteristic, there is no visible obstruction to extending our proof to lower characteristics as well. However, key portions of the argument for the implications for the geometry of $\mathcal{M}_g$ were written using characteristic 0, and as far as we are aware no one has carefully analyzed which positive characteristics they may apply to, so it seems preferable to work in characteristic zero and we will assume this is the case from now on.

2. Nondegeneracy calculations

In this section, we study maps from elliptic curves to projective space determined by comparing values of tensor products of certain tuples of sections at two points P and Q. We will need two distinct results in this direction: first, we consider the situation that we let the point Q vary. This is already considered in [Liu et al. 2021], where we showed that these maps are morphisms, described them explicitly, and gave partial criteria for nondegeneracy. Here we extend the nondegeneracy criterion to a sharp statement for the case of tensor pairs. This is used to show that if we vary the location of the nodes on individual components, we can get possible linear dependencies to vary sufficiently nontrivially. Next, we will consider a new case, where Q is fixed, but we have a separate varying parameter. This situation was not considered in [Liu et al. 2021], but will be important to us in dealing with situations where the discrete data of the limit linear series does not fix the underlying line bundle in some components.

First, given a genus-1 curve C and distinct P, Q on C, and $c, d \geq 0$, let $\mathscr{L} = \mathscr{O}_C(cP + (d-c)Q)$. Then for any $a, b \geq 0$ with $a + b = d - 1$, there is a unique section (up to scaling by $k^\times$) of $\mathscr{L}$ vanishing to order at least a at P and at least b at Q. Thus, we have a uniquely determined point R such that the divisor of the aforementioned section is $aP + bQ + R$; explicitly, R is determined by $aP + bQ + R \sim cP + (d-c)Q$, or

$$\begin{aligned} R &\sim (c-a)P + (d-c-b)Q = (c-a)P + (1+a-c)Q \\ &= P + (a+1-c)(Q-P) = Q + (a-c)(Q-P). \end{aligned} \tag{2-1}$$

We see that $R = P$ if and only if $Q - P$ is $|a + 1 - c|$-torsion, and $R = Q$ if and only if $Q - P$ is $|a-c|$-torsion. Note that (2-1) makes sense even when $Q = P$ (in which case $R = Q = P$), so we will use the formula for all P, Q, understanding that it has the initial interpretation as long as $Q \neq P$. To avoid trivial cases, we will assume that $a \neq c - 1$, and $b \neq d - c - 1$, or equivalently, $a + 1 - c \neq 0$, and $a - c \neq 0$.

Notation 2.1. Fix $P \in C$, $\ell \geq 1$, and for $j = 0, \ldots, \ell$, set numbers $a_1^j, a_2^j, b_1^j, b_2^j$ satisfying for $i = 1, 2$, $j, j' \in \{0, \ldots, \ell\}$,

$$a_i^j + b_i^j = d - 1, \quad a_i^j - c \neq 0, -1, \quad a_1^j + a_2^j = a_1^{j'} + a_2^{j'}.$$

Let U be the open subset of C consisting of all Q such that $Q - P$ is not $|a_i^j - c|$-torsion or $|a_i^j + 1 - c|$-torsion for any i, j. For $Q \in U$, choose sections s_i^j with divisors

$$a_i^j P + b_i^j Q + R_i^{j,Q}.$$

Define $s^j = s_1^j \otimes s_2^j \in \Gamma(C, \mathscr{L}^{\otimes 2})$, and normalize the s^j, uniquely up to simultaneous scalar, so that their values at P are all the same. Considering $(s^0(Q), \ldots, s^\ell(Q))$ gives a well-defined point of $\mathbb{P}^\ell$, denote by f_Q the point of $\mathbb{P}^\ell$ determined by $(s^0(Q), \ldots, s^\ell(Q))$.

In [Liu et al. 2021] we showed that the map $U \to \mathbb{P}^\ell$ given by $Q \mapsto f_Q$ extends to a morphism $f : C \to \mathbb{P}^\ell$.

Extending Corollary 2.5 of [Liu et al. 2021], we have:

Proposition 2.2. *If all the a_i^j are distinct, and $a_1^j + a_2^j \neq 2c - 1$, then f is nondegenerate.*

The proof relies on reduction to a good understanding of the $\ell = 1$ case. Indeed, we can view our map as being given by $(1, f_1, \ldots, f_\ell)$, where f_j is the rational function constructed from the quotient of sections s^j, s^0. Thus, nondegeneracy is equivalent to linear independence of the rational functions $1, f_1, \dots, f_\ell$, whose zeroes and poles are described explicitly by the following result, which combines Lemma 2.2 and Corollary 2.3 of [Liu et al. 2021]. Recall the following notation introduced in [Liu et al. 2021]. For k an integer, X a point in C and $L_1, \dots, L_{k^2}$ the line bundles in $\mathrm{Pic}^0(C)$ of order a divisor of $|k|$, $\mathcal{O}_C(X) \otimes L_i = \mathcal{O}_C(Y_i)$ for a unique $Y_i \in C$. Then, $X + T[k] := \sum_i Y_i$.

Lemma 2.3. *In the $\ell = 1$ case of Notation 2.1, the function $f : U \to k^\times$ given by $Q \mapsto (s^0/s^1)(Q)$ determines a rational function on C. We then have*

$$\operatorname{div} f = \sum_{i=1}^{2} \big((P + T[|a_i^0 - c|]) - (P + T[|a_i^1 - c|]) - (P + T[|a_i^0 + 1 - c|]) + (P + T[|a_i^1 + 1 - c|])\big),$$

Moreover, f is nonconstant if and only if $a_1^j + a_2^j \neq 2c - 1$.

Proof of Proposition 2.2. By Lemma 2.3, $f_1, \dots, f_\ell$ are all nonconstant. By reindexing the pairs we may further assume

$$a_1^\ell < a_1^{\ell-1} < \cdots < a_1^0 \le a_2^0 < a_2^1 < \cdots < a_2^\ell.$$

Let $n_i^j := |a_i^j - c + 1|$, $m_i^j := |a_i^j - c|$, and $n^j = \max\{n_1^j, n_2^j\}$.

A first observation is that $n^j > n^{j-1}$ for all j: if $a_1^{j-1} < c$ (respectively, $a_1^{j-1} > c$), then $n_1^{j-1} < n_1^j$ (respectively, $n_1^{j-1} < n_2^j$), and thus $n_1^{j-1} < n^j$; by a similar calculation, $n_2^{j-1} < n^j$; thus, $n^{j-1} < n^j$.

A second observation is that $n^j \ge \max\{m_1^0, m_2^0\}$ for all $j \ge 1$, and if equality is attained, j must be 1. Indeed, when $c < a_2^0$, we have $m_2^0 < m_2^j < n_2^j$ for all $j \ge 1$; meanwhile, either $m_1^0 \le m_2^0$ (if $c < a_1^0$) or $m_1^0 \le n_1^1 < n_1^j$ (if $c > a_1^0$) for all $j > 1$; thus, $\max\{m_1^0, m_2^0\} \le n^j$ for all $j \ge 1$. When $c > a_2^0$, $m_2^0 \le m_1^0 \le n_1^1 < n_1^j$ for all $j > 1$, and hence the same conclusion holds.

Now, we claim that f_j has poles at the strict n_j-torsion points. Recalling from Lemma 2.3 that the poles of f_j are supported among the m_i^0- and n_i^j-torsion points for $i = 1, 2$, the above two observations show that $1, \dots, f_{j-1}$ cannot have any poles at strict n^j-torsion points, which immediately implies that $1, f_1, \dots, f_\ell$ are k-linearly independent. Thus, it suffices to prove the claim. Since the potential zeroes of f_j are supported among the m_i^j- and n_i^0-torsion points, we just need to show that n^j does not divide m_i^j or n_i^0 for $i = 1, 2$ and any $j \ge 1$. Moreover, we already know that $n^j > n^0 \ge n_i^0$, so it is enough to consider the m_i^j. We consider two cases.

Case 1: $c < a_2^0$, so that also $c < a_2^j$ for all j. In this case,

$$m_2^0 < n_2^0 \le m_2^1 < n_2^1 \le \cdots \le m_2^\ell < n_2^\ell.$$

In particular, we have $n^j > m_2^j$, so it remains to compare n^j against m_1^j. If $n^j = n_1^j$, since n_1^j is always coprime to m_1^j, the claim follows instantly. If $n^j = n_2^j > n_1^j$, since $|n_1^j - m_1^j| = 1$, we have $n^j \ge m_1^j$. But equality cannot hold as it would imply that $a_1^j + a_2^j = 2c - 1$, which is ruled out by our assumption. So we conclude the claim in this case.

Case 2: $c > a_1^0$, so that $c > a_1^j$ for all j. If $a_2^j > c$, $n_2^j = m_2^j + 1$ and hence $n^j > m_2^j$. Meanwhile, $n_1^j = m_1^j - 1$. Similarly to the previous case, either $n^j > m_1^j$ or n^j is coprime to m_1^j, and the claim follows. If $a_2^j < c$, $n_2^j = m_2^j - 1$. Under our assumption, $n^j = n_1^j$ so is coprime to m_1^j. But because $j \geq 1$, we have $n_1^j \geq n_2^j + 2 = m_2^j + 1$, so $n_1^j > m_2^j$ and the claim follows. □

Notation 2.4. We now consider the point Q fixed, but the line bundle $\mathscr{L}$ varies (in particular, we do not have a c). As before, for $\ell \geq 1$, and $j = 0, \ldots, \ell$, set nonnegative integers $a_1^j, a_2^j, b_1^j, b_2^j$ satisfying

$$a_i^j + b_i^j = d - 1 \quad \text{for all } i, j; \qquad a_1^j + a_2^j \text{ is independent of } j.$$

Choose a point $R = R_1^0$. Then, for every $i = 1, 2,\ j = 0, \ldots, l$, there is a well determined R_i^j such that $\mathcal{O}(a_i^j P + b_i^j Q + R_i^j) = \mathcal{O}(a_1^0 P + b_1^0 Q + R) = \mathscr{L}$. Note that, using that $a_i^j + b_i^j = d - 1$, the last equation is equivalent to

$$R_i^j = R_1^0 + (a_1^0 - a_i^j)(P - Q).$$

We have sections s_i^j of $\mathscr{L}$ with divisors $a_i^j P + b_i^j Q + R_i^j$. We can take tensor products to obtain $s^j = s_1^j \otimes s_2^j$ and obtain sections of the line bundle $\mathscr{L}^{\otimes 2}$ having divisors $(a_1^j + a_2^j)P + (b_1^j + b_2^j)Q + R_1^j + R_2^j$. Note that the condition that $a_1^j + a_2^j$ is independent of j implies that the divisors $R_1^j + R_2^j$ will all be linearly equivalent to one another. If the a_i^j, b_i^j are generic, none of the R_i^j are equal to P and we can normalize the s^j to have the same value at P. We obtain a well-defined point $(s^0(Q), \ldots, s^\ell(Q)) \in \mathbb{P}^\ell$. But because we have said that $\mathscr{L}$ is uniquely determined by R_1^0, we can view this procedure as giving a rational map from C to $\mathbb{P}^\ell$, which we will now study. The argument will be similar to that of Lemma 2.2 and Corollary 2.5 of [Liu et al. 2021], but a bit simpler.

Proposition 2.5. *Suppose that $P - Q$ is not m-torsion for any $m \leq d$. Let $U \subseteq C$ be the open subset of points $R = R_1^0$ on which the map $\varphi : U \to \mathbb{P}^\ell$ that sends $R \in U$ to $(s^0(Q), \ldots, s^\ell(Q))$ is well defined. Then φ extends to a nondegenerate morphism $C \to \mathbb{P}^\ell$.*

Proof. We first consider the case $\ell = 1$, proving that we obtain a nonconstant rational function, and showing further that the divisor of this function is equal to

$$\begin{aligned} &Q + (Q - (a_1^0 - a_2^0)(P - Q)) + (P - (a_1^0 - a_1^1)(P - Q)) + (P - (a_1^0 - a_2^1)(P - Q)) \\ &\qquad - (Q - (a_1^0 - a_1^1)(P - Q)) - (Q - (a_1^0 - a_2^1)(P - Q)) - P - (P - (a_1^0 - a_2^0)(P - Q)). \end{aligned}$$

Let D_i^j be the divisor on $C \times C$ obtained as the graph of the morphism

$$R \mapsto R + (a_1^0 - a_i^j)(P - Q)$$

so that D_1^0 is simply the diagonal, and $(R_1^0, R_i^j) \in D_i^j$. Set

$$D^j = D_1^j + D_2^j + (P - (a_1^0 - a_1^{1-j})(P - Q)) \times C + (P - (a_1^0 - a_2^{1-j})(P - Q)) \times C$$

for $j = 0, 1$. Then we claim that D^0 and D^1 are linearly equivalent. By construction, if we restrict to $\{R\} \times C$ for any R not among the $P - (a_1^0 - a_i^j)(P - Q)$, we get that D^0 and D^1 are linearly equivalent, so $D^0 - D^1 \sim D \times C$ for some divisor D on C. But if we restrict to $C \times \{P\}$, we see that $D_1^j + D_2^j$ restricts to $(P - (a_1^0 - a_1^j)(P - Q)) + (P - (a_1^0 - a_2^j)(P - Q))$, so the restrictions of D^0 and D^1 are linearly equivalent on $C \times \{P\}$, and hence on $C \times C$, as desired. Moreover, this shows that if t_0 and t_1 are sections of the resulting line bundle having D^0 and D^1 as divisors, then $t_0|_{C\times\{P\}}$ has the same divisor as $t_1|_{C\times\{P\}}$, so we can scale so that t_0 and t_1 are equal on $C \times \{P\}$. We then see that our map $U \to \mathbb{P}^1$ is given by composing $R \mapsto (R, Q)$ with the rational function induced by our normalized choice of (t_0, t_1). Thus, it is a rational function, as desired. We compute its divisor simply by looking at the restrictions of D^0 and D^1 to $C \times \{Q\}$, which gives the claimed formula.

Now, for the case of arbitrary ℓ, we can consider the map to $\mathbb{P}^\ell$ to be given by a tuple of rational functions induced from the $\ell = 1$ case, specifically by $(f_0, \dots, f_{\ell-1}, 1)$, where f_j comes from looking at s^j and s^ℓ. To show nondegeneracy, it suffices to show that the f_j are linearly independent, which we do by showing that each of them (other than $f_\ell = 1$) has a pole which none of the others have. If we order so that

$$a_1^0 < a_1^1 < \cdots < a_1^\ell \le a_2^\ell < a_2^{\ell-1} < \cdots < a_2^0,$$

we see that $P - (a_1^j - a_2^j)(P - Q)$ occurs among the poles of f_j: indeed, given our nontorsion hypothesis on $P - Q$, the only positive term in the divisor which could possibly cancel it is Q, which would require $a_1^j - a_2^j = 1$, which is not possible with our above ordering. But again using our nontorsion hypothesis, and the fact that $a_2^j - a_1^j$ strictly decreases as j increases, we see that we obtain the desired distinct poles. □

3. Background on limit linear series and linked linear series

In this section we review background on limit linear series, as introduced by Eisenbud and Harris [1986], and on linked linear series, introduced by the second author [Osserman 2006] for two-component curves and generalized to arbitrary curves of compact type in [Osserman 2014].[2] Recall that a curve of *compact type* is a projective nodal curve such that every node is disconnecting, or equivalently, the dual graph is a tree. To streamline our presentation, we will largely restrict our attention to the situation of curves of compact type together with one-parameter smoothings.

Definition 3.1. Let X_0 be a curve of compact type, with dual graph Γ. Given $r, d \ge 0$, a *limit linear series* on X_0 of dimension r and degree d is a tuple $(\mathscr{L}^v, V^v)_{v \in V(\Gamma)}$, where each $(\mathscr{L}^v, V^v)$ is a linear series of dimension r and degree d on the component Z_v of X_0 corresponding to v. Write $a_\bullet^{(v,e)} = (a_0^{(v,e)}, \dots, a_r^{(v,e)})$ with $a_0^{(v,e)} < a_1^{(v,e)} < \cdots < a_r^{(v,e)}$ for the vanishing sequence (the $r+1$ different orders of vanishing of the sections in V^v) at P_e. The following condition must be satisfied: if Z_v and $Z_{v'}$ meet at a node P_e, then

$$a_j^{(v,e)} + a_{r-j}^{(v',e)} \ge d \quad \text{for } j = 0, \dots, r.$$

A limit linear series is said to be *refined* if the above inequalities are equalities for all e and j.

[2] In [Osserman 2006], linked linear series were called "limit linear series", but the name was changed subsequently to reduce confusion.

We now consider a one-parameter smoothing of X_0, as follows.

Remark 3.2. Suppose B is the spectrum of a discrete valuation ring with algebraically closed residue field, and $\pi : X \to B$ is flat and proper, with special fiber X_0 a curve of compact type, and smooth generic fiber X_η. Suppose further that the total space X is regular, that π admits a section.

Now, suppose we have a line bundle $\mathscr{L}_\eta$ generically — more precisely, we allow for the possibility that $\mathscr{L}_\eta$ is only defined after a finite extension of the base field of X_η. We can then take a finite base change $B' \to B$ so that $\mathscr{L}_\eta$ is defined over X'_η, and then X' may not be regular, but the line bundle $\mathscr{L}_\eta$ will still extend over X_0 because X_0 is of compact type. Moreover, there is a unique extension of $\mathscr{L}_\eta$ having any specified *multidegree* (i.e., tuple of degrees one for each component) adding up to d: because X was regular each component Z_v of X_0 is a Cartier divisor in X, and twisting by the $\mathscr{O}_X(Z_v)$ (or more precisely, their pullbacks to X') will increase the degree by 1 on each component meeting Z_v, and decrease the degree on Z_v correspondingly. For a multidegree ω, we denote this unique extension by $\widetilde{\mathscr{L}}_{\boldsymbol{\omega}}$. In particular, for each Z_v, we can consider the multidegree $\boldsymbol{\omega}^v$ which concentrates degree d on Z_v, and has degree 0 elsewhere.

Proposition 3.3 [Eisenbud and Harris 1986, Proposition 2.1]. *Given a linear series $(\mathscr{L}_\eta, V^v)$ on X'_η of dimension r and degree d, if we set $\mathscr{L}^v := (\widetilde{\mathscr{L}}_{\omega^v})|_{Z_v}$, and $V^v := (V_\eta \cap \Gamma(X', \widetilde{\mathscr{L}}_{\omega^v}))|_{Z_v}$, then the resulting tuple $(\mathscr{L}^v, V^v)_v$ is a limit linear series on X_0.*

Theorem 3.4 [Eisenbud and Harris 1986, Theorem 2.6]. *In characteristic* 0, *after finite base change and blowing up nodes in the special fiber, we may assume that the specialized limit linear series constructed by Proposition 3.3 is refined.*

Note that the only effect on X_0 of the base change and blowup is that chains of genus-0 curves are introduced at the nodes. Assuming we blow up to fully resolve the singularities resulting from the base change, these chain of curves have length equal to one less than the ramification index of the base change, so in particular they are the same at every node.

We now move on to linked linear series. The first observation is that if we have two multidegrees ω and ω', then there is a unique collection of nonnegative coefficients $c_v \in \mathbb{Z}$, not all positive, such that $\widetilde{\mathscr{L}}_\omega \cong \widetilde{\mathscr{L}}_{\omega'}(-\sum_v c_v Z_v)$. In this way, we obtain an inclusion $\widetilde{\mathscr{L}}_\omega \hookrightarrow \widetilde{\mathscr{L}}_{\omega'}$ which is defined uniquely up to scaling. If we define $\mathscr{L}_\omega := \widetilde{\mathscr{L}}_\omega|_{X_0}$, we get induced maps $\mathscr{L}_\omega \to \mathscr{L}_{\omega'}$ which are no longer injective, as they vanish identically on the components Z_v with $c_v > 0$. However, they are injective on the remaining components. Passing to global sections we obtain maps

$$f_{\omega,\omega'} : \Gamma(X_0, \mathscr{L}_\omega) \to \Gamma(X_0, \mathscr{L}_{\omega'}).$$

From the construction we see that $f_{\omega,\omega'} \circ f_{\omega',\omega}$ always vanishes identically. Although the twisted line bundles $\mathscr{L}_\omega$ can be described intrinsically on the special fiber, the maps $f_{\omega,\omega'}$ depend on the smoothing of X_0 whenever the locus on which they are nonvanishing. is disconnected.

To minimize notation, we will define linked linear series only in the above specialization context.

Definition 3.5. Given $\mathscr{L}_\eta$ of degree d and the induced tuple $(\mathscr{L}_\omega)_\omega$ of line bundles, a *linked linear series* of dimension r (and degree d) on the $\mathscr{L}_\omega$ is a tuple $(V_\omega)_\omega$ for all multidegrees of total degree d where each $V_\omega \subseteq \Gamma(X_0, \mathscr{L}_\omega)$ is an $(r+1)$-dimensional space of global sections, and for every ω, ω', we have

$$f_{\omega,\omega'}(V_\omega) \subseteq V_{\omega'}.$$

From the definitions and using Remark 3.2, we have:

Proposition 3.6. *Given $(\mathscr{L}_\eta, V_\eta)$, for all ω set $V_\omega = (V_\eta \cap \Gamma(X', \widetilde{\mathscr{L}}_\omega))|_{X_0}$. We obtain a linked linear series.*

This process is compatible with the Eisenbud–Harris specialization process, and the forgetful map commutes with specialization. The definition of linked series includes a linear series for every meaningful multidegree. In particular, there are linear series for the degrees ω^v which concentrate all the degree on Z_v. Ignoring the other multidegrees, we obtain a forgetful functor:

Theorem 3.7. *If $(V_\omega)_\omega$ is a linked linear series on $\mathscr{L}_\omega$, and we set $\mathscr{L}^v = \mathscr{L}_{\omega^v}|_{Z_v}$ and $V^v = V_{\omega^v}|_{Z_v}$ for all $v \in V(\Gamma)$, then $(\mathscr{L}^v, V^v)$ is a limit linear series. We will say that the linked linear series* ***lies over*** *the limit linear series*

This is explicitly stated (in the generality of higher-rank vector bundles) as part of Theorem 4.3.4 of [Osserman 2014], but is primarily a consequence of Lemma 4.1.6 of [loc. cit.].

In [Osserman 2014], the following notion is introduced:

Definition 3.8. A linked linear series is *simple* if there exist multidegrees $\omega_0, \dots, \omega_r$ and sections $s_j \in \Gamma(X_0, \mathscr{L}_{\omega_j})$ such that for every ω, the $f_{\omega_j,\omega}(s_j)$ form a basis of V_ω.

The simple linked linear series form an open subset, and are particularly easy to understand (hence the name). However, we will be forced to consider more general linked linear series arising under specialization. We therefore introduce the following open subset, originally introduced in [Osserman 2006] in the two-component case.

Definition 3.9. A linked linear series is *exact* if for every multidegree ω, and every proper subset $S \subseteq V(\Gamma)$, if $\mathscr{L}_{\omega'} \cong \mathscr{L}_\omega(-\sum_{v\in S} Z_v)$, then

$$f_{\omega,\omega'}(V_\omega) = V_{\omega'} \cap \ker f_{\omega',\omega}.$$

An important special case in the definition, and the only one which we will use in the present paper, is that ω' is obtained from ω by decreasing the degree by 1 on a single component and increasing it correspondingly on an adjacent component.

While we cannot always ensure our linked linear series are simple, we can ensure they are exact:

Proposition 3.10. *If $(\mathscr{L}_\eta, V_\eta)$ is defined over X_η itself, then the resulting linked linear series is exact.*

The proof is exactly the same as in the two-component case, which is explained immediately before the statement of Theorem 5.2 of [Esteves and Osserman 2013]. Thus, even if $(\mathscr{L}_\eta, V_\eta)$ is not defined over X_η, we can take a finite base change to make it defined, and blow up the resulting singularities of the total space to put ourselves into position to apply Proposition 3.10.

4. Degenerate linked linear series

The purpose of this section is to analyze the structures of the possible exact linked linear series lying over limit linear series mostly when $\rho \le 2$. We will restrict our attention to the case that the reducible curve X_0 is a chain.

Definition 4.1. Let $Z_1, \dots, Z_N$ be smooth curves with distinct marked points P_i, Q_i on each Z_i. Construct X_0 by gluing Q_i to P_{i+1} for each $i = 1, \dots, N-1$. Fix a total degree which in this section, we will denote with d. Given $w = (c_2, \dots, c_N)$, define the *multidegree of a line bundle on X_0 associated to w*,

$$\mathrm{md}_d(w) = (d_1, \dots, d_N), \qquad \text{by} \qquad d_1 = c_2, \quad d_i = c_{i+1} - c_i, \quad i = 2, \dots, N-1, \quad d_N = d - c_N.$$

Note that conversely, given a multidegree ω with total degree d, there is a unique w such that $\omega = \mathrm{md}(w)$. We will assume that $0 \le c_i \le c_{i+1} \le d$ for all i.

The notation c_i is very helpful in connection with the way in which we encode the combinatorial data of a limit linear series. In order to avoid treating the end points separately, it will be convenient to use the convention that $c_1 = 0$ and $c_{N+1} = d$. To avoid notational clutter, we will frequently write simply $\mathrm{md}(w)$ when the total degree is clear, and we will abbreviate $\mathscr{L}_{\mathrm{md}(w)}$ by $\mathscr{L}_w$, $f_{\mathrm{md}(w),\mathrm{md}(w')}$ by $f_{w,w'}$, and so fourth. The assumption that $0 \le c_i \le d$ for all i guarantees that the map f_{w,w^i} is injective on the component Z_i (see Proposition 3.6 of [Liu et al. 2021]), so we can understand sections in multidegree w as being glued from the Z_i-parts of sections in the multidegrees w^i.

From Remark 3.2, given a limit linear series $(\mathscr{L}^i, \mathscr{V}^i)_{i=1,\dots,N}$ and the multidegree associated to a w, then $\mathscr{L}_w|_{Z_i}$ is obtained from $\mathscr{L}^i$ by twisting down by $c_i P_i + (d - c_{i+1}) Q_i$, leaving degree $d - c_i - (d - c_{i+1}) = c_{i+1} - c_i$. Therefore, the condition $c_i \le c_{i+1}$ is equivalent to the d_i being positive.

Notation 4.2. By construction, the components of X_0 are ordered from 1 to N. We will think of a horizontal representation of the curve, numbering the components from left to right. For example, when we talk of the curve "strictly left of i", we mean $\bigcup_{j<i} Z_j$.

We first describe the behavior of the maps $f_{w,w'}$ under the above encoding (See Proposition 3.6 of [Liu et al. 2021] for a proof):

Proposition 4.3. *Given $w = (c_2, \dots, c_N)$, $w' = (c'_2, \dots, c'_N)$ and total degree d, the map $\mathscr{L}_{w'} \to \mathscr{L}_w$ vanishes identically on the component Z_i if and only if*

$$\sum_{j=i+1}^{N} (c'_j - c_j) > \min_{1 \le i' \le N} \sum_{j=i'+1}^{N} (c'_j - c_j).$$

In particular, if $c'_i < c_i$ or $c'_{i+1} > c_{i+1}$ then the map vanishes identically on Z_i, and if $c'_i = c_i$ for $i > 1$, then the map vanishes identically on Z_i if and only if it vanishes identically on Z_{i-1}.

Proposition 4.4. *Let Z be a smooth projective curve, and $P, Q \in Z$ distinct. Let $(\mathscr{L}, V)$ be a $\mathfrak{g}^r_d$ on Z. Then there is a unique (unordered) set of pairs $(a_0, b_0), \dots, (a_r, b_r)$ with all a_j distinct and all b_j distinct such that there exists a basis $s_0, \dots, s_r$ of V with $\operatorname{ord}_P s_j = a_j$ and $\operatorname{ord}_Q s_j = b_j$ for $j = 0, \dots, r$.*

Proof. Start with a basis $s_0, \dots, s_r$ with vanishing $a_0 < \cdots < a_r$ at P. Add multiples of the s_i to s_0 to maximize vanishing at Q. Then repeat the process replacing s_1, by adding multiples of the s_j, $j \geq 2$.

Note that the s_j themselves are not unique, although a given s_j can be modified only by adding multiples of $s_{j'}$ which simultaneously satisfy $\operatorname{ord}_P s_{j'} > \operatorname{ord}_P s_j$ and $\operatorname{ord}_Q s_{j'} > \operatorname{ord}_Q s_j$. □

To each refined limit linear series, we can associate a table of numbers as follows:

Definition 4.5. Let $(\mathscr{L}^i, V^i)$ be a refined limit $\mathfrak{g}^r_d$ on X_0, and for each i let $(a^i_j, b^i_j)_j$ be the set of pairs given by Proposition 4.4. Construct the $(r+1) \times N$ *table* T' from left to right, with the i-th column of T' consisting of the pairs (a^i_j, b^i_j) for $j = 0, \dots, r$, and the ordering of each column determined as follows: a^1_j should be strictly increasing, and for $i > 1$ and each j, we require $a^i_j = d - b^{i-1}_j$. For fixed i, we refer to the a^i_j and the b^i_j as making up the *subcolumns* of the i-th column of T'. For each j, let $w_j = (a^2_j, \dots, a^N_j)$, and set $\omega_j = \operatorname{md}_d(w_j)$.

Example 4.6. Let X_0 be a chain of 5 elliptic curves. Construct a limit linear series on X_0 of degree 4 and dimension 1 with the following line bundles on the components:

$$L_1 = \mathcal{O}(4Q_1), \quad L_2 = \mathcal{O}(2P_2 + 2Q_2), \quad L_3 = \mathcal{O}(P_3 + 3Q_3), \quad L_4 \text{ generic}, \quad L_5 = \mathcal{O}(4P_5).$$

and sections with vanishing associated to the table

$$\begin{array}{cc|cc|cc|cc|cc} 0 & 4 & 0 & 3 & 1 & 3 & 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 2 & 2 & 1 & 3 & 0 & 4 & 0 \end{array}$$

The table has two rows corresponding to the two sections. The five columns correspond to the 5 elliptic curves with the left and right semicolumns corresponding to the vanishing at P_i and Q_i, respectively. There are two w_i one for each of the two sections and left semicolumns, starting with the second one and corresponding multidegrees ω_j as follows:

$$w_0 = (0, 1, 1, 2), \quad \omega_0 = (0, 1, 0, 1, 2), \quad w_1 = (2, 2, 3, 4), \quad \omega_1 = (2, 0, 1, 1, 0).$$

For instance $\mathscr{L}_{w_0}$ is a line bundle on the chain with restrictions to the 5 components

$$L^0_1 = \mathcal{O}, \quad L^0_2 = \mathcal{O}(2P_2 - Q_2), \quad L^0_3 = \mathcal{O}, \quad L^0_4 = L_4(-P_4 - 2Q_4), \quad L^0_5 = \mathcal{O}(2P_5),$$

while $\mathscr{L}_{w_1}$ has restrictions to the 5 components

$$L^0_1 = \mathcal{O}(2Q_1), \quad L^0_2 = \mathcal{O}, \quad L^0_3 = \mathcal{O}(2Q_3 - P_3), \quad L^0_4 = L_4(-3P_4), \quad L^0_5 = \mathcal{O}.$$

Note that the set of pairs of Proposition 4.4 is giving a relative ordering of the vanishing sequences at P and Q, so the condition that the limit linear series is refined means that we can always impose that $a^i_j = d - b^{i-1}_j$. Arranging our table ordering in this way, we can always choose sections $s^i_j \in V^i$ such that

$\operatorname{ord}_{P_i} s_j^i = a_j^i$ and $\operatorname{ord}_{Q_i} s_j^i = b_j^i$. Then in multidegree ω_j there is a unique section s_j obtained from gluing together the s_j^i (although as noted above, the choices of s_j^i are not unique in general).

Definition 4.7. We say that a *swap* occurs in column i between rows j, j' if $a_j^i < a_{j'}^i$ and $b_j^i < b_{j'}^i$ or if $a_j^i > a_{j'}^i$ and $b_j^i > b_{j'}^i$. A swap is *minimal* if further $|a_j^i - a_{j'}^i| = |b_j^i - b_{j'}^i| = 1$ and either $a_j^i + b_j^i = d$ or $a_{j'}^i + b_{j'}^i = d$.

Example 4.8. If X_0 is again a chain of 5 elliptic curves, construct a limit linear series on X_0 of degree 4 and dimension 1 with line bundles and table of vanishing

$$L_1 = \mathcal{O}(4Q_1), \quad L_2 = \mathcal{O}(2P_2 + 2Q_2), \quad L_3 = \mathcal{O}(2P_3 + 2Q_3), \quad L_4 = \mathcal{O}(2P_4 + 2Q_4), \quad L_5 = \mathcal{O}(4P_5),$$

$$\begin{array}{cc|cc|cc|cc|cc} 0 & 4 & 0 & 3 & 1 & 1 & 3 & 0 & 4 & 0 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{array}$$

A swap appears on C_3 between the only two sections on the linear series s_0, s_1. This swap is minimal as $|a_1^3 - a_0^3| = 2 - 1 = |b_1^3 - b_0^3|$ is 1 and $a_1^3 + b_1^3 = 2 + 2 = 4 = d$.

A limit linear series is *chain-adaptable* in the sense of [Osserman 2014] if there are no swaps in the table T'. For a chain-adaptable limit linear series, there is only one linked linear series lying over it that is simple, generated by the s_j described above. In the nonchain-adaptable case, the linked linear series is not necessarily unique.

A nonempty open subset of the set of possible linked linear series will always be simple, generated by sections similar to the s_j described above. However, even some exact linked linear series are not simple. We can nonetheless use exactness to obtain fairly good control over what these linked linear series look like. We address all the cases that can arise for $\rho \le 2$ below.

We will use the following observation: Fix a refined limit linear series and a choice of all the s_j^i. For any $w = (c_2, \dots, c_N)$ (assumed bounded), the linkage condition implies that the sections in the $(r+1)$-dimensional space V_w in the linked linear series are linear combinations of sections obtained by gluing, for a fixed j, the sections s_j^i to one another as i varies, where each s_j^i that appears must satisfy $a_j^i \ge c_i$ and $b_j^i \ge d - c_{i+1}$, and if the first (respectively, second) inequality is an equality, we must also have s_j^{i-1} (respectively, s_j^{i+1}) included in the gluing. Indeed, a section in V_w must be a linear combination of such s_j^i, and since the a_j^i and b_j^i are all distinct for fixed i, at most one can have equality on each side, leading to the desired form for the gluing.

Proposition 4.9. *Suppose that the j_0-th row of T' has the property that for all $j < j_0$ we have $b_j^i > b_{j_0}^i$ for $i = 1, \dots, N-1$, and for all $j > j_0$ we have $a_j^i > a_{j_0}^i$ for $i = 2, \dots, N$. Then any linked linear series lying over the given limit linear series* (*in the sense of Theorem 3.7*) *contains the expected section s_{j_0}.*

Proof. It suffices to see that the space of global sections in multidegree ω_{j_0} obtained from all possible gluings of the s_j^i has dimension exactly $r+1$, so that any linked linear series must contain the whole space, including s_{j_0}. But for $j < j_0$ since $b_j^i > b_{j_0}^i$ for $i < N$, we have $a_j^{i+1} < a_{j_0}^{i+1}$, so s_j^{i+1} cannot appear at all in multidegree ω_{j_0}. Thus, only s_j^1 can appear, glued to the zero section on every other component. Similarly, for $j > j_0$ only s_j^N can appear. And since each $s_{j_0}^i$ has precisely the desired vanishing at the nodes, s_{j_0} is the unique way to glue them together, so we obtain an $(r+1)$-dimensional space in total, as desired. □

In this paper we consider mostly spaces of linear series with Brill–Noether number $\rho = 1, 2$. We will see in Proposition 5.3 that on a generic chain of elliptic curves, the number of swaps is bounded by ρ. So, we now systematically consider all cases where the limit linear series has only one or two swaps.

Proposition 4.10. *Suppose that a limit linear series has a single swap occurring in the i_0-th column between the (j_0-1)-st and j_0-th rows, Then any linked linear series lying over that limit linear series (in the sense of Theorem 3.7) contains the expected section s_{j_0-1}. The spaces of sections of the linear series with multidegrees associated to*

$$(a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0}) \quad \textit{and} \quad (a^2_{j_0}, \dots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0-1}, \dots, a^N_{j_0-1})$$

contain the respective images of the section s_{j_0}. These images consist of 0 on the first i_0-1 components and $s^i_{j_0}$ for $i = i_0, \dots, N$, and of $s^i_{j_0}$ for $i = 1, \dots, i_0$ and 0 on the last $N-i_0$ components, respectively.

Given $w = (c_2, \dots, c_N)$. If $c_i < a^i_{j_0-1}, a^i_{j_0}$ for all i, the linked linear series contains $s^1_{j_0}$ in multidegree $\operatorname{md}(w)$, *and if $c_i > a^i_{j_0-1}, a^i_{j_0}$ for all i, the linked linear series contains $s^N_{j_0}$ in multidegree* $\operatorname{md}(w)$ (*in both cases, glued to* 0 *on the other components*).

Proof. Note that our assumptions imply that

$$b^i_j > b^i_{j_0-1}, b^i_{j_0}, \quad \text{for all } j < j_0-1,\ i = 1, \dots, N-1, \qquad a^i_j > a^i_{j_0}, a^i_{j_0-1} \quad \text{for all } j > j_0,\ i = 2, \dots, N.$$

In multidegree ω_{j_0-1}, as in the proof of Proposition 4.9, the s^i_j for $j \neq j_0-1$, j_0 can only contribute for $i = 1$ (if $j < j_0-1$) or $i = N$ (if $j > j_0$), and the $s^i_{j_0-1}$ glue uniquely to give s_{j_0-1}. Finally, the $s^i_{j_0}$ can only contribute at $i = i_0$, so we find that the space obtained from all the s^i_j is $(r+1)$-dimensional, and s_{j_0-1} must be in the linked linear series, as desired.

Next, consider $w' = (a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0})$. Note that $f_{w_{j_0}, w'}(s_{j_0})$ is equal to s_{j_0} from i_0 to N (inclusive), and 0 strictly before i_0. We claim that the space of possible sections from the s^i_j in multidegree $\operatorname{md}(w')$ is precisely $(r+1)$-dimensional, so the linked linear series is uniquely determined in this multidegree. By hypothesis, the s^i_j for $j < j_0-1$ can only contribute for $i = 1$, and the s^i_j for $j > j_0$ can only contribute for $i = N$. The $s^i_{j_0-1}$ could in principle contribute for $i < i_0$ and $i = N$, but if the $s^i_{j_0-1}$ gave rise to nonzero sections for $i < i_0$, they all would be nonvanishing at the relevant nodes, and they would have to glue to something nonvanishing in the i_0-th column. This would have to be $s^{i_0}_{j_0-1}$, which does not have enough vanishing on the right to appear in multidegree $\operatorname{md}(w')$. Thus, we conclude that the $s^i_{j_0-1}$ can only appear for $i = N$ (where it is glued to the zero section on all other columns). Finally, the $s^i_{j_0}$ can only appear for $i \geq i_0$, where they are nonzero at all interior nodes, and therefore have a unique gluing, which must yield $f_{w_{j_0}, w'}(s_{j_0})$. Thus we get the claimed dimension $r+1$, and conclude that $f_{w_{j_0}, w'}(s_{j_0})$ is contained in the linked linear series.

Similarly, if $w'' = (a^2_{j_0}, \dots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0-1}, \dots, a^N_{j_0-1})$, we find that space of possible sections is $(r+1)$-dimensional, and contains $f_{w_{j_0}, w''}(s_{j_0})$.

Now, suppose we are given w with $c_i < a^i_{j_0-1}, a^i_{j_0}$ for all i. Then Proposition 4.3 implies that $f_{w'',w}$ is nonzero precisely on the first component, so $f_{w'',w}(f_{w_{j_0},w''}(s_{j_0}))$ is equal to $s^1_{j_0}$ glued to 0, as desired. The situation with $c_i > a^i_{j_0-1}, a^i_{j_0}$ is similar, but with w' in place of w''. □

Example 4.11. In Example 4.8, $(a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0}) = (0, 1, 2, 2)$. This numbers give the required vanishing at P_i, $i = 2, \dots, 5$, while the vanishing at $Q_i = d - c_{i+1} = 4 - c_{i+1}$. This means that the required vanishing is

$$4 \text{ at } Q_1,\quad 0 \text{ at } P_2,\quad 3 \text{ at } Q_2,\quad 1 \text{ at } P_3,\quad 2 \text{ at } Q_3,\quad 2 \text{ at } P_4,\quad 2 \text{ at } Q_4,\quad 2 \text{ at } P_5.$$

The order of vanishing of s_1 at the nodes is 2 at P_3, 2 at Q_3, 2 at P_4, 2 at Q_4, 2 at P_5. As the order at P_3 is 2, strictly greater than the required 1, it can be glued to the zero section in the first two components to give rise to a section of the linked linear series. Similarly, $(a^2_{j_0}, \dots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0-1}, \dots, a^N_{j_0-1}) = (2, 2, 3, 1)$. The required vanishing is then

$$2 \text{ at } Q_1,\quad 2 \text{ at } P_2,\quad 2 \text{ at } Q_2,\quad 2 \text{ at } P_3,\quad 1 \text{ at } Q_3,\quad 3 \text{ at } P_4,\quad 0 \text{ at } Q_4,\quad 4 \text{ at } P_5.$$

The order of vanishing of s_1 at the nodes is 2 at Q_1, 2 at P_2, 2 at Q_2, 2 at P_3, 2 at Q_3. As the order at Q_3 is 1, strictly greater than the required 0, it can be glued to the zero section in the last two components to give rise to a section of the linked linear series.

When the hypotheses of Proposition 4.9 are not satisfied for every j_0, there may be linked linear series — even exact ones — which do not contain all of the s_{j_0}. They may occur as specializations of linear series on the generic fiber. This leads us to introduce the notion of mixed sections below. We will then show that there are mixed sections of rather precise forms, which can in some sense take the place of the missing s_j.

Definition 4.12. For $\ell > 1$, let $\vec{S} = (S_1, \dots, S_\ell)$ be a tuple of subsets of $\{1, \dots, N\}$ (some of which may be empty) such that for all pairs $i < i'$, every element of S_i is less than or equal to every element of $S_{i'}$ and such that every element of $\{1, \dots, g\}$ is contained in some S_i. Let $\vec{j} = (j_1, \dots, j_\ell)$ be a tuple of elements of $\{0, \dots, r\}$, possibly with repetitions. Then given a fixed limit linear series and corresponding choices of the s^i_j, a *mixed section* of type $(\vec{S}, \vec{j})$ is a w and a section s in multidegree $\mathrm{md}(w)$ which is a sum from $i = 1$ to ℓ of sections obtained by gluing $s^{i'}_{j_i}$ for all $i' \in S_i$ to the zero section on other components.

Proposition 4.13. *Suppose that a limit linear series has precisely one swap, between the j_0-th and (j_0-1)-st rows occurring in the i_0-th column. Then any linked linear series lying over the given limit linear series contains the expected sections s_j for all $j \neq j_0$. If the linked linear series is exact, then it must contain mixed sections s'_{j_0} of type $((S'_1, S'_2), (j_0 - 1, j_0))$ with S'_1 supported on $\bigcup_{i<i_0} Z_i$ and s''_{j_0} of type $((S''_1, S''_2), (j_0, j_0 - 1))$ with S''_2 supported on $\bigcup_{i>i_0} Z_i$.*

Note that this allows for the linked linear series to contain the section s_{j_0} itself when S'_1 and S''_2 are both empty.

Proof. Start with the $w' = (a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0})$ as in Proposition 4.10. Note that if $i_0 = 1$, then the proposition says that s_{j_0} itself is in our linked linear series, consistent with the stated form for s'_{j_0}.

If $i_0 > 1$, we consider iteratively changing w' by increasing the twists by 1 for $i' \leq i_0$ (starting at i_0) until they each agree with $a_{j_0}^{i'}$. We note that every such modified w' has an $(r+2)$-dimensional space of global sections obtained from the s_j^i, described explicitly as follows: s_j^1 for $j < j_0 - 1$; s_j^N for $j > j_0$; $s_{j_0-1}^N$; a section obtained by gluing the $s_{j_0-1}^i$ for i from 1 to $i'-1$ (which is the last column in which w' agrees with $a_{j_0-1}^i$); and a section obtained by gluing the $s_{j_0}^i$ from either $i'-1$ or i' to N, beginning with the last column in which w' has coefficient strictly less than $a_{j_0}^i$. For each $j \neq j_0 - 1$, since there is a unique section constructed from the s_j^i, it is necessarily equal to $f_{w_j,w'}(s_j)$. In addition, since we know s_j is in the linked linear series for $j \neq j_0$, then $f_{w_j,w'}(s_j)$ is necessarily contained in the linked linear series for $j \neq j_0 - 1, j_0$.

Now, suppose that the linked linear series contained $f_{w_{j_0},w'}(s_{j_0})$ for the old w'; we claim that it either also contains it for the new w', or contains a section of the form desired for s'_{j_0}. Indeed, increasing the twist in the i-th column corresponding to twisting once by every component from i to N., we observe that $f_{w_{j_0},w'}(s_{j_0})$ is in the kernel of the map from the old w' to the new one. By definition of exactness, the linked linear series must contain some s in the new multidegree mapping to $f_{w_{j_0},w'}(s_{j_0})$ in the old one. Using the above description of the space of global sections, this is necessarily a combination of the $f_{w_j,w'}(s_j)$ for $j < j_0 - 1$ and $j = j_0$, together with the section from the $s_{j_0-1}^i$ for $i = 1$ to $i'-1$. Moreover, since we observed above that $f_{w_j,w'}(s_j)$ is contained in the linked linear series for $j < j_0 - 1$, we can subtract these off to obtain a combination of the sections from the $j_0 - 1$ and j_0 rows. If the $j_0 - 1$ term vanishes, we have that $f_{w_{j_0},w'}(s_{j_0})$ is contained in our linked linear series for the new w', and if the $j_0 - 1$ term is nonzero, we have something of the desired form for s'_{j_0} (with the minimal element of S'_2 being either i' or $i'-1$ according to where $f_{w_{j_0},w'}(s_{j_0})$ begins), as claimed. Iterating this process, we either obtain the desired s'_{j_0}, or we eventually reach $w' = w_{j_0}$ and find that the linked linear series actually contains s_{j_0} itself.

As the situation is completely symmetric, the construction of s''_{j_0} is similar, starting from the multidegree w'' from the proof of Proposition 4.10. □

When $\rho = 2$, there are four additional types of swap (see Definition 4.7), which we consider one by one. They all involve having exactly two swaps, occurring in distinct columns. The first case is when the swaps occur in disjoint pairs of rows.

Proposition 4.14 ("disjoint swap"). *Suppose that a limit linear series contains precisely two swaps, and these occur in disjoint pairs of rows, say $j_0 - 1$, j_0 and $j_1 - 1$, j_1. Then any linked linear series lying over the given limit linear series contains the expected sections s_j for all $j \neq j_0, j_1$. If the linked linear series is exact, then for $\ell = 0, 1$ it must contain mixed sections s'_{j_ℓ} of type $((S'_{1+2\ell}, S'_{2+2\ell}), (j_\ell - 1, j_\ell))$ with $S'_{1+2\ell}$ supported strictly left of i_ℓ and s''_{j_ℓ} of type $((S''_{1+2\ell}, S''_{2+2\ell}), (j_\ell, j_\ell - 1))$ with $S''_{2+2\ell}$ supported strictly right of i_ℓ.*

Proof. This is essentially identical to the proof of Proposition 4.13. The only new point which arises is that in constructing the sections s'_{j_0}, s''_{j_0}, we need to know that we can always subtract off any s_{j_1} part which arises in the iterative procedure, and similarly with j_0 and j_1 switched. But this follows from the last assertion of Proposition 4.10. □

Example 4.15. Assume that X_0 is a chain of 6 elliptic curves. One can construct a limit linear series on X_0 of degree 8 and dimension 3 by choosing the following line bundles on each component:

$$\mathcal{O}(8Q_1), \quad \mathcal{O}(2P_2+6Q_2), \quad \mathcal{O}(2P_3+6Q_3), \quad \mathcal{O}(6P_4+2Q_4), \quad \mathcal{O}(6P_5+2Q_5), \mathcal{O}(8P_6),$$

and sections with associated table

0	8	0	7	1	5	3	4	4	3	5	2
1	6	2	6	2	6	2	5	3	4	4	3
2	5	3	4	4	3	5	1	7	0	8	0
3	4	4	3	5	2	6	2	6	2	6	1

A swap appears on C_3 between s_0, s_1 and another at C_4 between s_2, s_3. Both swaps are minimal as required as there are the maximum number possible (two) of swaps for $\rho = 2$. In general there may be larger gaps between the rows where the swap occurs and between the columns where it occurs as well.

The next case is that a single pair of rows can undergo two swaps in different columns.

Proposition 4.16 ("repeated swap"). *Suppose that the limit linear series has precisely two swaps, both between the j_0-th and (j_0-1)-st rows, with the first occurring in the i_0-th column, and the second in the i_1-st column for some $i_1 > i_0$. Then any exact linked linear series lying over the given limit linear series contains mixed sections s'_{j_0-1} of type $((S'_1, S'_2, S'_3), (j_0-1, j_0, j_0-1))$ with S'_2 supported strictly left of i_1 and s''_{j_0-1} of type $((S''_1, S''_2), (j_0-1, j_0))$ with S''_2 supported strictly right of i_1, and it contains mixed sections s'_{j_0} of type $((S'_4, S'_5), (j_0-1, j_0))$ with S'_4 supported strictly left of i_0 and s''_{j_0} of type $((S''_3, S''_4, S''_5), (j_0, j_0-1, j_0))$ with S''_4 supported strictly right of i_0.*

Proof. The proof is similar to the proof of Proposition 4.13. For s'_{j_0-1}, we first consider

$$w' = (a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^{i_1}_{j_0}, a^{i_1+1}_{j_0-1}, \dots, a^N_{j_0-1}).$$

Note that $f_{w_{j_0-1},w'}(s_{j_0-1})$ is equal to s_{j_0-1} from i_1 to N (inclusive), and 0 elsewhere. Indeed, these are the only columns in which the $s^i_{j_0-1}$ can be supported, since they do not satisfy the correct inequalities from i_0 to i_1-1, and for $i < i_0$ they satisfy them with equality, so would have to be glued to a nonzero element in the i_0-th column. As in the proof of Proposition 4.9, we check that we have dimension exactly $r+1$ in multidegree $\mathrm{md}(w')$, with the unique contribution from the j_0 row coming from $s^N_{j_0}$. Thus, we find that $f_{w_{j_0-1},w'}(s_{j_0-1})$ is necessarily contained in multidegree $\mathrm{md}(w')$.

We then iterate changing w' by 1, increasing the twist by 1 in the i'-th column for $i' \le i_1$ to change them from $a^{i'}_{j_0}$ to $a^{i'}_{j_0-1}$. Using exactness, at each stage we either find the linked linear series still contains $f_{w_{j_0-1},w'}(s_{j_0-1})$ for the new value of w', or it contains the sum of $f_{w_{j_0-1},w'}(s_{j_0-1})$ with a section obtained by gluing the $s^i_{j_0}$ for $i = i_0, \dots, i'-1$. In the first case, we continue to iterate the process of changing w', and if we do not ever get the second case, we end up with s_{j_0-1} itself in our linked linear series. On the other hand, once the second case occurs, we begin to iteratively change w' by increasing the twist by 1 in the i'-th column for $i' \le i_0$ to change them from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$. Each time the twist increases above $a^{i'}_{j_0-1}$, we could obtain a contribution obtained from gluing $s^i_{j_0-1}$ from $i = 1$ to $i'-1$, and if this occurs, we get our desired s'_{j_0-1}. Otherwise, we keep iterating, and each time the twist at i' reaches $a^{i'}_{j_0}$, the portion of the section obtained from the $s^i_{j_0}$ extends to include $i'-1$. Again, if we never get a contribution from the $s^i_{j_0-1}$ for $i \leq i'$, we will end up with a section as required for s'_{j_0-1}, having $S'_1 = \varnothing$.

The construction of s''_{j_0-1} is similar, but simpler: we set our initial $w'' = (a^2_{j_0-1}, \dots, a^{i_1}_{j_0-1}, a^{i_1+1}_{j_0}, \dots, a^N_{j_0})$, and then we iteratively decrease the twists for $i' > i_1$ by 1 to change them from $a^{i'}_{j_0}$ to $a^{i'}_{j_0-1}$, until we obtain the desired result.

The construction of s'_{j_0} and s''_{j_0} follows the same process. For s'_{j_0}, we start with

$$w' = (a^2_{j_0-1}, \dots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0}),$$

and we iteratively increase the twists for $i' \leq i_0$ by 1 to change them from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$. Finally, for s''_{j_0}, we start with $w'' = (a^2_{j_0}, \dots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0-1}, \dots, a^{i_1}_{j_0-1}, a^{i_1+1}_{j_0}, \dots, a^N_{j_0})$, obtaining a section glued from the $s^i_{j_0}$ for $i \leq i_0$. We iteratively decrease the twists for $i' > i_0$ by 1 to change them from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$, until we obtain a contribution from the $a^i_{j_0-1}$ (necessarily ending at i_1), and then we iteratively decrease the twists for $i' > i_1$ by 1 to change them from $a^{i'}_{j_0}$ to $a^{i'}_{j_0-1}$, eventually obtaining either s_{j_0} itself, or the desired s''_{j_0}. □

Example 4.17. Assume that X_0 is a chain of 6 elliptic curves. One can construct a limit linear series on X_0 of degree 5 and dimension 1 by choosing the following line bundles on each component,

$$\mathcal{O}(5Q_1), \quad \mathcal{O}(2P_2+3Q_2), \quad \mathcal{O}(2P_3+3Q_3), \quad \mathcal{O}(3P_4+2Q_4), \quad \mathcal{O}(3P_5+2Q_5), \quad \mathcal{O}(5P_6);$$

a table associated to this limit linear series is

0	5	0	4	1	2	3	2	3	2	3	1
1	3	2	3	2	3	2	1	4	0	5	0

A swap appears on C_3 and again at C_4 between the only two sections on the linear series s_0, s_1. Both swap are minimal as required as there are the maximum number possible (two) of swaps for $\rho = 2$

Finally, three consecutive rows can undergo two swaps. This can happen in two different ways.

Proposition 4.18 ("first 3-cycle"). *Suppose that the limit linear series has one swap between the j_0-th and (j_0+1)-st rows occurring in the i_0-th column, and a second swap between the (j_0-1)-st and (j_0+1)-st rows in the i_1-st column for some $i_1 > i_0$, and no other swaps. Then any linked linear series lying over the given limit linear series contains s_{j_0-1} and s_{j_0}. If further the linked linear series is exact, then it contains mixed sections s'_{j_0+1} of type $((S'_1, S'_2, S'_3), (j_0-1, j_0, j_0+1))$ with S'_1 supported strictly left of i_1, S'_2 supported strictly left of i_0, s''_{j_0+1} of type $((S''_1, S''_2, S''_3), (j_0+1, j_0-1, j_0))$ with S''_2 supported strictly right of i_1, S''_3 supported strictly right of i_0, and s'''_{j_0+1} of type $((S'''_1, S'''_2, S'''_3), (j_0, j_0+1, j_0-1))$ with S'''_1 supported strictly left of i_0, and S'''_3 supported strictly right of i_1.*

Note that if $S'_2 = \varnothing$, then S'_1 may contain elements greater than i_0, and similarly if $S''_2 = \varnothing$, then S''_3 may contain elements less than i_1.

Proof. First, check that the multidegrees ω_{j_0-1} and ω_{j_0} both have only $(r+1)$-dimensional spaces of possible sections, so that s_{j_0-1} and s_{j_0} must both lie in any linked linear series. Indeed, for the former, the

$s^i_{j_0}$ can contribute only for $i = N$, while the $s^i_{j_0+1}$ can contribute only for $i = i_1$, while for the latter, the $s^i_{j_0-1}$ can contribute only for $i = 1$, and the $s^i_{j_0+1}$ can contribute only for $i = i_0$.

Now, to construct the sections s'_{j_0+1}, s''_{j_0+1} and s'''_{j_0+1}, we proceed as in the previous propositions. For s'_{j_0+1}, we start with $w' = (a^2_{j_0-1}, \dots, a^{i_1}_{j_0-1}, a^{i_1+1}_{j_0+1}, \dots, a^N_{j_0+1})$, and then iteratively increase the twist by 1 at a time for $i' \le i_1$, initially increasing it from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0+1}$. For $i' > i_0$, this process behaves as before, either extending the contribution from the $a^i_{j_0+1}$ iteratively to the left without introducing any other nonzero contributions, or producing a section s'_{j_0+1} as desired, having $S_2 = \varnothing$. Once $i' \le i_0$, we still iteratively increase the twist from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0+1}$, but we are required to pass $a^{i'}_{j_0}$ in the process. This introduces a third possibility: once the twist at i' is strictly greater than $a^{i'}_{j_0}$, we could obtain a contribution from $s^{i'-1}_{j_0}$. Also, for $i' < i_0$, once the twist at i' is equal to $a^{i'}_{j_0}$, we could obtain a contribution from both $s^{i'-1}_{j_0}$ and $s^{i'}_{j_0}$. If either of these occurs, we move to the next i', and for the remaining i', instead of increasing the twist from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0+1}$, we only increase to $a^{i'}_{j_0}$. Note that we may obtain contributions from the $s^i_{j_0}$ (for $i = i' - 1$ or $i = i' - 1, i'$) and $s^i_{j_0-1}$ (for $i = 1, \dots, i' - 1$) simultaneously at some point, which still gives an s'_{j_0+1} of the desired form. On the other hand, if we never obtain a contribution from the $s^i_{j_0}$, then the resulting s'_{j_0+1} simply has $S'_2 = \varnothing$.

For s''_{j_0+1}, we start with $w'' = (a^2_{j_0+1}, \dots, a^{i_0}_{j_0+1}, a^{i_0+1}_{j_0}, \dots, a^N_{j_0})$, and then follow the same procedure as for s'_{j_0+1}, iteratively decreasing the twist at $i' > i_0$ from $a^{i'}_{j_0}$ to $a^{i'}_{j_0+1}$, with the possibility of a contribution from the $s^i_{j_0-1}$ once i' passes i_1.

Finally, for s'''_{j_0+1} set $w''' = (a^2_{j_0}, \dots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0+1}, \dots, a^{i_1}_{j_0+1}, a^{i_1+1}_{j_0-1}, \dots, a^N_{j_0-1})$ initially. We then iteratively increase the twist at $i' \le i_0$ from $a^{i'}_{j_0}$ to $a^{i'}_{j_0+1}$, and iteratively decrease the twist at $i' > i_1$ from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0+1}$ to construct s'''_{j_0+1}. □

Example 4.19. Assume that X_0 is a chain of 8 elliptic curves. One can construct a limit linear series on X_0 of degree 8 and dimension 2 by choosing the following line bundles on each component

$$\mathcal{O}(8Q_1), \quad \mathcal{O}(2P_2+6Q_2), \quad \mathcal{O}(4P_3+4Q_3), \quad \mathcal{O}(4P_4+4Q_4),$$
$$\mathcal{O}(4P_5+4Q_5), \quad \mathcal{O}(4P_6+4Q_6), \quad \mathcal{O}(6P_7+2Q_7), \quad \mathcal{O}(8P_8)$$

and sections with associated table

0	8	0	7	1	6	2	5	3	3	5	2	6	2	6	1
1	6	2	6	2	5	3	3	5	2	6	1	7	0	8	0
2	5	3	4	4	4	4	4	4	4	4	4	4	3	5	2

A swap appears on C_4 between s_1, s_2 and another at C_5 between s_0, s_2. In general there may be larger gaps between the columns where the swap occurs.

Proposition 4.20 ("second 3-cycle"). *Suppose that our limit linear series has one swap between the (j_0-1)-st and j_0-th rows occurring in the i_0-th column, and a second swap between the (j_0-1)-st and (j_0+1)-st rows in the i_1-st column for some $i_1 > i_0$, and no other swaps.*

Then any linked linear series lying over the given limit linear series contains s_{j_0-1}. If further the linked linear series is exact, then it contains mixed sections s'_{j_0} and s''_{j_0} of type $((S'_1, S'_2), (j_0-1, j_0))$ and

$((S''_1, S''_2, S''_3), (j_0, j_0+1, j_0-1))$ *respectively, with* S'_1 *supported strictly left of* i_0, S''_2 *supported at or right of* i_1, *and* S''_3 *supported strictly right of* i_0. *Similarly, it contains mixed sections* s'_{j_0+1} *and* s''_{j_0+1} *of type* $((S'_3, S'_4, S'_5), (j_0-1, j_0, j_0+1))$ *and* $((S''_4, S''_5), (j_0+1, j_0-1))$ *respectively, with*S'_3 *supported strictly left of* i_1, S'_4 *supported at or left of* i_0, *and* S''_5 *supported strictly right of* i_1. *Moreover, if* $i_1 \in S''_2$ *then also* $i_1 \in S''_1$, *and if* $i_0 \in S'_4$, *then also* $i_0 \in S'_5$. *Finally, either we can have* $S'_2 = S''_4 = \{1, \ldots, N\}$, *or it also contains a mixed section* s''' *of type* $((S'''_1, S'''_2, S'''_3), (j_0, j_0-1, j_0+1))$, *where every element of* S'''_2 *is strictly between* i_0 *and* i_1.

Proof. For the most part, this is straightforward and similar to the previous propositions, but there is one new subtlety to address, and the idea for the construction of s''' is new. We first construct s'_{j_0}, starting with $w' = (a^2_{j_0-1}, \ldots, a^{i_0}_{j_0-1}, a^{i_0+1}_{j_0}, \ldots, a^N_{j_0})$. We then iteratively increase the twist from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$ for $i' \leq i_0$, and obtain s'_{j_0} as before. We then do the same procedure for s''_{j_0+1}, starting with $w'' = (a^2_{j_0+1}, \ldots, a^{i_1}_{j_0+1}, a^{i_1+1}_{j_0-1}, \ldots, a^N_{j_0-1})$.

Next, we construct s''_{j_0}, starting with $w'' = (a^2_{j_0}, \ldots, a^{i_0}_{j_0}, a^{i_0+1}_{j_0-1}, \ldots, a^N_{j_0-1})$. We then iteratively decrease the twist at $i' > i_0$ from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$. For $i' \leq i_1$, this behaves as in the previous propositions, with one new subtlety: for each intermediate value of w', the $s^i_{j_0+1}$ can contribute only in the i_1 column, but because we do not know that s_{j_0+1} is contained in the linked linear series, we also do not know *a priori* that this contribution from $s^{i_1}_{j_0+1}$ in multidegree $\mathrm{md}(w')$ is contained in our linked linear series. However, since we have already constructed s''_{j_0+1}, we can use its image in $\mathrm{md}(w')$. One checks that its only possible support in $\mathrm{md}(w')$ is in the i_1 column, so that in fact the multidegree-$\mathrm{md}(w')$ part of the linked linear series necessarily contains the section given by $s^{i_1}_{j_0+1}$, and we can subtract it off as necessary from the section we are constructing. Thus, for $i' \leq i_1$, we can iterate as before, and will either obtain an s''_{j_0} as desired (with $S''_2 = \varnothing$), or we will obtain a section made up of the $s^i_{j_0}$ for $i \leq i_1$, and vanishing identically for $i > i_1$. In the latter case, we continue to iteratively decrease the twists defining w' for $i > i_1$, but as in the construction of s'_{j_0+1} in the proof of Proposition 4.18, to get from $a^{i'}_{j_0-1}$ to $a^{i'}_{j_0}$ we need to pass $a^{i'}_{j_0+1}$, which is where the possible contribution from the j_0+1 may occur.

The construction of s'_{j_0+1} follows the same pattern as that of s''_{j_0}, but starting with

$$w' = (a^2_{j_0-1}, \ldots, a^{i_1}_{j_0-1}, a^{i_1+1}_{j_0+1}, \ldots, a^N_{j_0+1}).$$

Here we use the image of s'_{j_0} in order to subtract off any contributions of $s^{i_0}_{j_0}$ which occur.

Finally, for s''', we start with $w' = w_{j_0}$. We observe that there is an $(r+2)$-dimensional space of potential sections in multidegree ω_{j_0}, with the s^i_j for $j < j_0 - 1$ contributing only for $i = 1$, the s^i_j for $j \geq j_0+1$ contributing only for $i = N$, the $s^i_{j_0}$ contributing only with s_{j_0} itself, and the $s^i_{j_0-1}$ contributing separately for $i = 1$ and $i = N$.

We must therefore have a three-dimensional space of combinations of the four sections $s^1_{j_0-1}$, $s^N_{j_0-1}$, $s^N_{j_0+1}$, and s_{j_0}. It follows by elimination that this space must contain (at least) one of the following: s_{j_0} plus a (possibly zero) multiple of $s^1_{j_0-1}$; s_{j_0} plus a (possibly zero) multiple of $s^N_{j_0+1}$; $s^1_{j_0-1}$ and $s^N_{j_0+1}$. The first case means that we can take $S'_2 = \{1, \ldots, N\}$, while in the second we get a valid choice of s'''. In the third case, we begin with $s^N_{j_0+1}$, and iteratively twist the multidegree as before. For $i' > i_1$, we change w' from twisting down by $a^{i'}_{j_0}$ to $a^{i'}_{j_0+1}$, and at each stage, we must either obtain the desired s''', or a section made up purely of the $s^i_{j_0+1}$, in which case we continue to iterate. Note that in these multidegrees, we continue to have that the only possible contributions of the s^i_j (for $j \neq j_0$) supported strictly left of i' come for $j \le j_0 - 1$, and we can take the image of $s^1_{j_0-1}$ from multidegree ω_{j_0}, so all these can be subtracted off as necessary. When $i' \le i_1$, we will have $a^{i'}_{j_0-1}$ between $a^{i'}_{j_0}$ and $a^{i'}_{j_0+1}$; we still iteratively increase the twist, but a new possibility occurs: once we are twisting down by strictly more than $a^{i'}_{j_0-1}$, we could obtain a contribution from $a^{i'-1}_{j_0-1}$. If this occurs, we will continue to iterate, but stopping after increasing the twist from $a^{i'}_{j_0}$ to $a^{i'}_{j_0-1}$ for each smaller i'.

If we have continued with contributions from $s^{i'}_{j_0+1}$ for each i', then once we reach i_0, we will again have no other a^i_j between $a^i_{j_0}$ and $a^i_{j_0+1}$, so we will ultimately obtain an s''' of the desired form, with $S'''_2 = \varnothing$. On the other hand, if we have switched from the $s^{i'}_{j_0+1}$ to the $s^{i'}_{j_0-1}$, then we see that this must terminate (necessarily with an s''' of the desired form) before we reach $i' = i_0$, because there is no section in column i_0 which can glue to $s^{i_0+1}_{j_0-1}$.

Now, if the above construction did not give s''' because we had $S'_2 = \{1, \dots, N\}$, we apply precisely the same process starting in multidegree ω_{j_0+1}, and we find that unless we also have $S''_4 = \{1, \dots, N\}$, we end up with the desired s'''. □

Example 4.21. Assume that X_0 is a chain of 8 elliptic curves. One can construct a limit linear series on X_0 of degree 8 and dimension 2 by choosing the following line bundles on each component

$$\mathcal{O}(8Q_1),\quad \mathcal{O}(2P_2+6Q_2),\quad \mathcal{O}(2P_3+6Q_3),\quad \mathcal{O}(5P_4+3Q_4),$$
$$\mathcal{O}(5P_5+3Q_5),\quad \mathcal{O}(4P_6+4Q_6),\quad \mathcal{O}(6P_7+2Q_7),\quad \mathcal{O}(8P_8)$$

and sections with associated table

0	8	0	7	1	5	3	4	4	2	6	1	7	0	8	0
1	6	2	6	2	6	2	5	3	4	4	4	4	3	5	2
2	5	3	4	4	3	5	3	5	3	5	2	6	2	6	1

A swap appears on C_3 between s_0, s_1 and another at C_5 between s_0, s_2. In general there may be larger gaps between the columns where the swap occurs.

Up until now, everything we have done has been insensitive to insertion of genus-0 components. However, to handle the genus-23 case, we will need to impose restrictions on direction of approach; more precisely, we will require that the genus-1 components be separated by exponentially increasing numbers of genus-0 components (going from right to left). The reason for doing this is that, if our limit linear series has all changes to the λ_i occurring in the genus-1 components, the pattern of the genus-0 components will force the support of every s_j in every multidegree to be precisely the leftmost segment of potential support (see Proposition 8.8). So, we obtain better control over the situation when the potential support is disconnected. That this sort of restriction could potentially be useful is already pointed out in

Remark 4.12 of [Liu et al. 2021], and is also influenced by the earlier work of Jensen and Payne [2016] on their tropical approach to the classical maximal rank conjecture.

Definition 4.22. Assume that we have a chain of curves of genus zero and one with the first and last components of the chain being of genus 1. We denote by $\boldsymbol{\ell}_i$ the number of nodes between the i-th and $(i+1)$-st components of genus-1. We say that X_0 is *left-weighted* if

$$\ell_i \geq 4d \sum_{i'=i+1}^{g-1} \ell_{i'}.$$

In our notation, $\ell_i - 1$ is the number of genus-0 components between two genus one components. If we take a ramified base change with ramification index e, and then blow up to resolve the resulting singularities, we will insert e new genus-0 components at every node, which has the effect of multiplying all the ℓ_i by e. Thus, the ratios of the ℓ_i (and therefore the left-weightedness) are invariant under this operation.

Definition 4.23. Given $\vec{S} = (S_1, \ldots, S_\ell)$ and $\vec{j} = (j_1, \ldots, j_\ell)$, a mixed section of type $(\vec{S}, \vec{j})$ is said to be *controlled* if for every $i = 2, \ldots, \ell$ with $S_i \neq \varnothing$, the minimal element of S_i is either a genus-1 component or strictly closer to the next genus-1 component to the right than to the previous one on the left.

Proposition 4.24. *Suppose that X_0 is left-weighted. Then*:

(1) *In the situation of Proposition 4.14, if we assume further that i_0 and i_1 have genus* 1, *then we may require that s'_{j_0} and s'_{j_1} are controlled, that S'_2 does not contain any $i < i_0$ which has genus* 1, *and that S'_4 does not contain any $i < i_1$ which has genus* 1.

(2) *In the situation of Proposition 4.20, if we assume further that i_0 and i_1 have genus* 1, *then we may require that s'_{j_0} is controlled, and that S'_2 does not contain any $i < i_0$ which has genus* 1.

Proof. (1) With the notations of the proof of Proposition 4.14, for $s = 0, 1$ and every value of w' arising in the iterative procedure, the potential support of the $s^i_{j_s-1}$ in multidegree $\mathrm{md}(w')$ has two connected components: one extending from $i = 1$ to $i = i' - 1$, and the other supported at $i = N$. We cannot continue our iterative procedure indefinitely if $f_{w_{j_s-1},w'}(s_{j_s-1})$ is supported (partially or entirely) at $i = N$. If we write $w' = (c'_2, \ldots, c'_N)$, then $a^i_{j_s-1} - c'_i > 0$ for $i > i_s$, $a^i_{j_s-1} - c'_i < 0$ for $i' \leq i \leq i_s$, and $a^i_{j_s-1} - c'_i = 0$ for $i < i'$. Suppose i_s is the m_s-th genus-1 component. Note that $a^i_{j_s-1} - c'_i \leq a^i_{j_s-1} \leq d$. Then in the notation used in Definition 4.22 above, we can say (extremely conservatively) that

$$\sum_{i=i_s+1}^{N} (a^i_{j_s-1} - c'_i) \leq d \sum_{i=m_s}^{g-1} \ell_i \leq \tfrac{1}{4}\ell_{m_s-1}. \tag{4-1}$$

Thus, if $i_s - i' > \frac{1}{4}\ell_{m_s-1}$, then $\sum_{i=i'}^{N}(a^i_{j_s-1} - c'_i) < 0$, so $f_{w_{j_s-1},w'}(s_{j_s-1})$ is supported entirely on the left. We can then subtract off any contribution from the $s^i_{j_s-1}$ and continue our iterative procedure. The desired conditions on s'_{j_s} follow.

(2) This is essentially the same as (1) (the analogous statement for s'_{j_0+1} is a bit more complicated, but we don't need it). ☐

row	col = 1	2	3	4	5	6	7	8	9	10	11
0	0 25	0 24	1 23	2 22	3 21	4 20	5 19	6 19	6 18	7 17	8 16
1	1 23	2 23	2 22	3 21	4 20	5 19	6 18	7 17	8 16	9 16	9 15
2	2 22	3 21	4 21	4 20	5 19	6 18	7 17	8 16	9 14	11 13	12 12
3	3 21	4 20	5 19	6 19	6 18	7 17	8 16	9 15	10 15	10 14	11 14
4	4 20	5 19	6 18	7 17	8 17	8 16	9 15	10 14	11 13	12 12	13 11
5	5 19	6 18	7 17	8 16	9 15	10 15	10 14	11 13	12 12	13 11	14 10
6	6 18	7 17	8 16	9 15	10 14	11 13	12 13	12 12	13 11	14 10	15 9

row	col = 12	13	14	15	16	17	18	19	20	21	22
0	9 15	10 14	11 13	12 12	13 12	13 11	14 10	15 9	16 8	17 7	18 6
1	10 14	11 13	12 12	13 11	14 10	15 10	15 9	16 8	17 7	18 6	19 5
2	13 12	13 11	14 10	15 9	16 8	17 7	18 6	19 6	19 5	20 4	21 3
3	11 13	12 12	13 11	14 10	15 9	16 8	17 8	17 7	18 6	19 5	20 4
4	14 10	15 10	15 9	16 8	17 7	18 6	19 5	20 4	21 4	21 3	22 2
5	15 9	16 8	17 8	17 7	18 6	19 5	20 4	21 3	22 2	23 2	23 1
6	16 8	17 7	18 6	19 6	19 5	20 4	21 3	22 2	23 1	24 0	25 0

Table 1. Example 4.25. A possible table T' associated to a limit linear series for $r = 6$, $g = 22$, $d = 25$ with all component of genus 1. The table is split in two. Top: left part. Bottom: right part.

An example in a case we are ultimately interested in is presented now.

Example 4.25. Table 1 shows a possible table T' associated to a limit linear series for $r = 6$, $g = 22$, $d = 25$ with all component of genus 1.

Since there is no ramification at P_1, the first entries of the table agree with the row labels. Note that we have a single swap, occurring in the ninth column between the $j = 2$ and $j = 3$ rows. This leads to having an extra dimension of possibilities in the multidegree obtained from the $j = 3$ row, as the $j = 2$ row can appear either in the first or last columns. Consequently, it is possible that an exact linked linear series lying over this limit linear series might not contain s_3, but might only contain mixed sections s_3' and s_3'', as in Proposition 4.13, with s_3' agreeing with s_3 for $i \geq 9$, but switching to s_2 at some $i < 9$, and s_3'' agreeing with s_3 for $i \leq 9$, but switching to s_2 at some $i > 9$. In both cases, the switch occurs in a column mixing s_2^i and s_3^i unless, the column in question has a gap of at least 2 between the $j = 2$ and $j = 3$ rows. Since this doesn't occur for $i < 9$, we see that s_3' always has a mixed column, while s_3'' may not.

5. General setup

We are working with chains of N curves. While we imagine starting from a chain of genus-1 curves, we allow for inserting any number of rational components at nodes so that all components Z_i are of genus at most one, and exactly g have genus 1 (including the first and last components). Given i between 1 and N, we denote by $g(i)$ the number of genus-1 components between 1 and i, inclusive. In particular, $g(0) = 0$ by convention.

We will suppose also that all the P_i and Q_i are general (and in particular each $P_i - Q_i$ is not ℓ-torsion for any $\ell \leq d$). We need generality conditions that go beyond what can be imposed component by component, as it also involves interaction between components. This will be needed in the proof of Lemma 6.3.

No matter the genus of Z_i, for all j, $a^i_j + b^i_j \leq d$. If Z_i has genus 0, there are no further restrictions. If the genus of Z_i is 1, the genericity hypothesis implies that there is at most one value of j such that $a^i_j + b^i_j = d$. In this case the underlying line bundle is uniquely determined as $\mathcal{O}(a^i_j P_i + b^i_j Q_i)$. The generic situation is that $a^i_j + b^i_j = d - 1$ for all other j, but in positive codimension we can have strictly smaller sums as well

Definition 5.1. We say that the j-th row is *exceptional* in column i if $a^i_j + b^i_j < d - 1$ when Z_i has genus 1, or if $a^i_j + b^i_j < d$ when Z_i has genus 0. For $i = 1, \dots, N$, we write $\boldsymbol{g}(i)$ for the number of genus one components between 1 and i inclusive. For $j = 0, \dots, r$ and $i = 0, \dots, N$, define $\boldsymbol{\lambda}_{i,j}$ by $a^{i+1}_j = g(i) + j - \lambda_{i,j}$. For a given i, if there is a j such that $\lambda_{i,j} > \lambda_{i-1,j}$, we say that *there is a* $\boldsymbol{\delta}_i$. and more specifically, that $\boldsymbol{\delta}_i = j$. Otherwise, we say that *there is no* δ_i.

In this way, we obtain a sequence $\lambda_i = (\lambda_{i,0}, \dots, \lambda_{i,r})$. For a fixed i, we think of the $\lambda_{i,j}$, $j = 0, \dots, r$ as rows with $\lambda_{i,j}$, squares, where a negative number of squares appears as "holes" to the left of the level 0 line. In this ways, we get a collection of "shapes", $i = 0, \dots, N$ (not necessarily skew, or connected) generalizing the Young Tableaux usually associated to limit linear series on chains of elliptic curves. They behave as follows: $\lambda_{0,j} \leq 0$ for all j, λ_0 is the empty shape if $a^1_j = j$ and in general the λ_0 shape is entirely made of holes. Going from i to $i + 1$, any number of "squares" can be removed from the right of any row (and then the row is exceptional). At most one "square" is added (and then $\delta_i = j$), with the possibility of adding a "square" only in the genus-1 components.

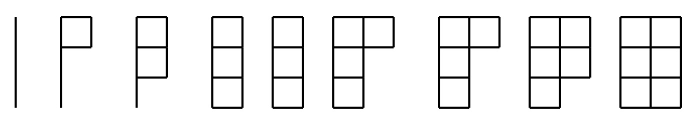
Example 5.2. Let X_0 be a generic chain of 8 elliptic curves. Construct a limit linear series on X_0 of degree 8 and dimension 2 by choosing generic line bundles L_4, L_6 on C_4, C_6 and the rest as follows:

$$\mathcal{O}(8Q_1),\quad \mathcal{O}(2P_2+6Q_2),\quad \mathcal{O}(4P_3+4Q_3),\quad L_4,\quad \mathcal{O}(3P_5+5Q_5),\quad L_6,\quad \mathcal{O}(6P_7+2Q_7),\quad \mathcal{O}(8P_8)$$

and sections with associated table

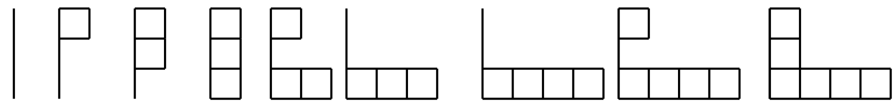
0	8	0	7	1	6	2	5	3	5	3	4	4	3	5	2
1	6	2	6	2	5	3	4	4	3	5	2	6	2	6	1
2	5	3	4	4	4	4	3	5	2	6	1	7	0	8	0

The corresponding λ shapes are



λ_0 λ_1 λ_2 λ_3 λ_4 λ_5 λ_6 λ_7 λ_8

By contrast, the λ shapes corresponding to Example 4.19 which has a swap on C_4 between rows 1 and 2 and another swap on C_5 between rows 0 and 2 are



λ_0 λ_1 λ_2 λ_3 λ_4 λ_5 λ_6 λ_7 λ_8

Proposition 5.3. *Assume that the curve X_0 is a chain of elliptic and rational curves. Choose a limit linear series and a linked series lying over it, then the number of swaps is bounded by ρ. Moreover, if there are ρ swaps, then the swaps must all be minimal, and occur in genus-1 columns, and there cannot be any exceptional behavior (as in Definition 5.1) other than what is needed for the swaps.*

Proof. The b_j^N are nonnegative (and distinct) for all j. Equivalently the a_j^{N+1} are bounded by

$$d-r, d-r+1, \dots, d.$$

In particular, $\sum_{j=0}^r a_j^{N+1} \le (r+1)d - \binom{r+1}{2}$, so

$$\sum_{j=0}^{r} \lambda_{N,j} \ge (r+1)g + \binom{r+1}{2} - (r+1)d + \binom{r+1}{2} = (r+1)(g+r-d) = g-\rho.$$

Since $\sum_j \lambda_{i,j}$ can increase by at most 1 as i increases (and only on genus-1 components), and $\lambda_{0,j} \le 0$ for all j, we see that for $\rho = 0$, we must have $\lambda_{0,j} = 0$ for all j (i.e., minimal initial vanishing sequence at P_1), no places where $\lambda_{i,j}$ decreases (i.e., no exceptional columns for any row), and a δ_i for every genus-1 column i. When $\rho > 0$, the total amount of initial ramification, exceptional columns, and genus-1 columns without δ_i is bounded by ρ.

A swap occurs when the vanishings of two of the sections at P_i, Q_i are of the form $(a, d-a-l)$ and $(a+k, d-a-k-l')$, respectively, with $k > 0$, $k + l' < l$. Hence, a swap is necessarily a case of an exceptional column, and can contribute exactly 1 to ρ precisely when it is minimal and occurs in a genus-1 column. □

We now describe the tensor square of a limit linear series considering images in a fixed multidegree of total degree $2d$. Essentially the discrete data from the base limit linear series is extended to its tensor square.

Notation 5.4. In the situation of Definition 4.5, let T be the $\binom{r+2}{2} \times N$ table with rows indexed by unordered pairs (j, j') with $j, j' \in \{0, \dots, r\}$, having entries $(a^i_{(j,j')}, b^i_{(j,j')})$ defined by

$$a^i_{(j,j')} = a^i_j + a^i_{j'}, \quad \text{and} \quad b^i_{(j,j')} = b^i_j + b^i_{j'}.$$

We update a definition from [Liu et al. 2021] to allow for genus-0 components:

Definition 5.5. We say a multidegree of total degree $2d$ is *unimaginative* if it assigns degree 0 to every genus-0 component, and degree 2 or 3 to every genus-1 component. By extension, we will say that w is unimaginative if $\mathrm{md}_{2d}(w)$ is. Given a fixed unimaginative multidegree, we will let $\boldsymbol{\gamma}_i$ be the number of 3s in the first i columns.

We will work throughout only with unimaginative multidegrees. Thus, the multidegree is encoded by twisting down by $2d-2g(i)-\gamma_i$ on the right-hand of the i-th column, and by twisting down by $2g(i)+\gamma_i$ on the left-hand side of the $(i+1)$-st column, for all $i<N$. We introduce some notation that we will use:

Definition 5.6. In the situation of Notation 5.4, fix total degree $2d$, and $w=(c_2,\dots,c_N)$. We say that the (j,j') row is *potentially present* (respectively *potentially starting*, respectively *potentially ending*) in column i and multidegree $\mathrm{md}_{2d}(w)$ if $a^i_{(j,j')}\geq c_i$ and $b^i_{(j,j')}\geq 2d-c_{i+1}$ (respectively $a^i_{(j,j')}>c_i$ and $b^i_{(j,j')}\geq 2d-c_{i+1}$, respectively $a^i_{(j,j')}\geq c_i$ and $b^i_{(j,j')}>2d-c_{i+1}$).

The next proposition is an immediate consequence of the definitions.

Proposition 5.7. *If a row (j_1,j_2) is potentially present in the i-th column, then*

$$j_1+j_2-\lambda_{i-1,j_1}-\lambda_{i-1,j_2}\geq\gamma_{i-1}\quad\textit{and}\quad j_1+j_2-\lambda_{i,j_1}-\lambda_{i,j_2}\leq\gamma_i.$$

If a row (j_1,j_2) is potentially starting (respectively ending) in the i-th column, then the first (respectively second) inequality is strict.

If a row (j_1,j_2) is potentially present in the i-th and $(i+1)$-st columns, then

$$j_1+j_2-\lambda_{i,j_1}-\lambda_{i,j_2}=\gamma_i.$$

Note that the sequence $j_1+j_2-\lambda_{i,j_1}-\lambda_{i,j_2}$ decreases by at most 1 each time i increases, unless $j_1=j_2=\delta_i$, when it can decrease by 2. Similarly, from our assumptions on degrees and the definition of γ_i, γ_i is nondecreasing, and increases by at most 1 each time i increases.

Corollary 5.8. *Assume the multidegree is 2 on the i-th column. There can be a row potentially starting in the i-th column only if δ_i exists and either there exists j such that*

$$\delta_i+j-\lambda_{i,\delta_i}-\lambda_{i,j}=\gamma_i\quad\textit{or}\quad 2\delta_i-2\lambda_{i,\delta_i}=\gamma_i-1.$$

In these cases, the potentially starting rows are (δ_i,j) or (δ_i,δ_i), respectively.

There can be a row potentially ending in the i-th column only if δ_i exists and either there exists j such that

$$\delta_i+j-\lambda_{i,\delta_i}-\lambda_{i,j}=\gamma_i-1\quad\textit{or}\quad 2\delta_i-2\lambda_{i,\delta_i}=\gamma_i-2.$$

In these cases, the potentially ending rows are (δ_i,j) or (δ_i,δ_i), respectively.

In any case, there can be at most one row potentially starting on the i-th column, and at most one row potentially ending in it.

Proof. Since in this case $\gamma_i=\gamma_{i-1}$, Proposition 5.7 implies that the (j_1,j_2) row can be potentially starting in the i-th column only if $\lambda_{i,j_1}>\lambda_{i-1,j_1}$ or $\lambda_{i,j_2}>\lambda_{i-1,j_2}$, which is to say if δ_i exists and j_1 or j_2 is equal to δ_i. Moreover, in this case $\lambda_{i,\delta_i} = \lambda_{i-1,\delta_i} + 1$, so we conclude that the two stated cases are the only possibilities for having

$$j_1 + j_2 - \lambda_{i-1,j_1} - \lambda_{i-1,j_2} > \gamma_{i-1} = \gamma_i \geq j_1 + j_2 - \lambda_{i,j_1} - \lambda_{i,j_2},$$

and that moreover in the first case we must also have $\lambda_{i-1,j} = \lambda_{i,j}$ unless $j = \delta_i$.

Now, there is at most one j satisfying the first identity of the corollary, since the $j - \lambda_{i,j}$ are all distinct. Moreover, if there is some j satisfying the first, then the second one cannot hold, since this would force

$$\delta_i - \lambda_{i-1,\delta_i} = \delta_i - \lambda_{i,\delta_i} + 1 = j - \lambda_{i,j} = j - \lambda_{i-1,j},$$

which is not allowed. This completes the proof of the assertions on rows potentially starting in the i-th column. The assertion on rows potentially ending in the i-th column is proved similarly. □

Next corollary has a similar proof, which we omit.

Corollary 5.9. *If the multidegree has a* 3 *in the i-th column, then there can be at most one row potentially starting and ending in the i-th column, and this occurs only if δ_i exists and either there exists j such that*

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} = \gamma_i - 1 \quad \text{or} \quad 2\delta_i - 2\lambda_{i,\delta_i} = \gamma_i - 2.$$

In addition, for a fixed $j \neq \delta_i$, there is at most one value of j' such that the (j, j') row is potentially starting in column i, and at most one value of j' such that the (j, j') row is potentially ending in column i.

Proposition 5.10. *For a fixed limit linear series, w, and column i, if $a^i_{(j,j')} > c_i$, then (j, j') has a component of potential support $\bigcup_{j \geq i} Z_j$. If $a^i_{(j,j')} < c_i$, then (j, j') has a component of potential support $\bigcup_{j<i} Z_j$.*

Conversely, suppose further that w is unimaginative. If (j, j') has a component of potential support $\bigcup_{j \geq i} Z_j$, and if neither j nor j' is exceptional in any column strictly right of $i - 1$, then $a^i_{(j,j')} > c_i$. Similarly, if (j, j') has a component of potential support $\bigcup_{j<i} Z_j$., and if neither j nor j' is exceptional in any column $j < i$, then $a^i_{(j,j')} < c_i$.

In particular, in the unimaginative case, the potential support of (j, j') is connected unless the sum of the number of swaps for which the j-th row is exceptional and the number of swaps for which the j'-th row is exceptional is at least 2.

Proof. The first part is straightforward, and we omit the proof. For the second part, the point is that the unimaginative hypothesis together with the nonexceptional hypothesis imply that the relevant portion of the sequence $a^{i'}_{(j,j')} - c_{i'}$ is nondecreasing in the relevant range as i' decreases, so in the first case if its positivity for some $i' \geq i$ implies it remains positive at $i' = i$, while in the second case its negativity for some $i' \leq i$ implies it remains negative at $i' = i$.

For the last assertion, we can have disconnected potential support in the (j, j') row only if the sequence $a^i_{(j,j')} - c_i$ goes from positive to negative as i decreases, possibly over multiple columns. But we observe that if only one of j and j' are exceptional at a swap, which is moreover minimal and in a genus-1 column, then the sequence $a^i_{(j,j')} - c_i$ can decrease only by 1 as i decreases. Thus, if this occurs only once, it cannot go from positive to negative, and it cannot have disconnected potential support. □

Definition 5.11. Given a refined limit linear series, we construct a second table $\overline{\boldsymbol{T}}$ of vanishing numbers obtained from the first by reordering each subcolumn into strictly increasing (respectively, decreasing) order. We denote the λ sequence obtained from $\overline{T}$ by $\bar{\boldsymbol{\lambda}}_i$, and the entries of the table $\overline{T}$ by $(\bar{\boldsymbol{a}}^i_j, \bar{\boldsymbol{b}}^i_j)$. For $\ell \geq 1$, we denote by $\bar{\boldsymbol{\lambda}}^\ell_i$ the number of j such that $\bar{\lambda}_{i,j} \geq \ell$.

Put differently, $\overline{T}$ is obtained from the limit linear series simply by taking vanishing sequences at each point, and ignoring the interplay between the pair of points. If we picture skewing the rows of the $\bar{\lambda}_i$ according to the initial ramification sequence $a^1_j - j$, the sequence $\bar{\lambda}_i$ will give a genuine sequence of skew shapes, terminating with a skew shape containing the one obtained by starting from the usual $(r+1) \times (r+g-d)$ center rectangle, and adding squares on the left determined by the initial ramification sequence.

The following lemma is the key to our analysis, showing in particular that if we place multidegree 3 in the correct places, we can obtain fine control over what happens with the rows involving δ_{i+1}.

Lemma 5.12. *Given* $1 \leq \ell_1 < \ell_2$ *and* $n > 0$, *suppose that* $\bar{\lambda}^{\ell_1}_i + \bar{\lambda}^{\ell_2}_i = n$ *and* $\bar{\lambda}^{\ell_1}_{i-1} + \bar{\lambda}^{\ell_2}_{i-1} < n$. *Then*

$$\text{for all } j, \quad \delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} \neq n - 1 - \ell_1 - \ell_2, \tag{5-1}$$

$$2\delta_i - 2\lambda_{i,\delta_i} \neq n - 2 - \ell_1 - \ell_2. \tag{5-2}$$

Moreover, *if for some* j, $\lambda_{i,j} < \lambda_{i-1,j}$, *then*

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} \neq n - \ell_1 - \ell_2, \tag{5-3}$$

$$\delta_i + j - \lambda_{i-1,\delta_i} - \lambda_{i-1,j} \neq n - 1 - \ell_1 - \ell_2. \tag{5-4}$$

Proof. We first prove the case that $\bar{\lambda}_{i'} = \lambda_{i'}$ for all i'. Note that the assumption that $\bar{\lambda}^{\ell_1}_i + \bar{\lambda}^{\ell_2}_i > \bar{\lambda}^{\ell_1}_{i-1} + \bar{\lambda}^{\ell_2}_{i-1}$ implies that there is a δ_i and it is the row of the last square in either the ℓ_1-th or ℓ_2-th columns of λ_i.

Case when $\lambda^{\ell_1}_i, \lambda^{\ell_2}_i$ *are distinct and positive*. Set $j_s = \lambda^{\ell_s}_i - 1$, $s = 1, 2$. In particular, $\delta_i = j_1$ or j_2 and $(j_1 + 1) + (j_2 + 1) = n$. For $s = 1, 2$ write $m_s = \lambda_{i,j_s} - \ell_s$, so that $m_s \geq 0$ for $s = 1, 2$, with equality for at least one s. Then,

$$\begin{aligned} j_1 + j_2 - \lambda_{i,j_1} - \lambda_{i,j_2} &= (j_1 + 1) + (j_2 + 1) - 2 - (m_1 + \ell_1) - (m_2 + \ell_2) \\ &= n - 2 - \ell_1 - \ell_2 - m_1 - m_2 \\ &< n - 1 - \ell_1 - \ell_2. \end{aligned}$$

Thus, the only way to get equality in (5-1) would be to set j to be strictly greater than whichever j_s is not equal to δ_i. Now, because j_s was determined as the lowest row with a square in the ℓ_s-th column, we have

$$\lambda_{i,j_s+1} < \ell_s = \lambda_{i,j_s} - m_s,$$

so if we use $j > j_s$ in place of j_s, the value of the above expression jumps by at least $2 + m_s$. Moreover, we can only use j in place of j_1 if $\delta_i = j_2$, in which case we must have $m_2 = 0$, and similarly if we use j in place of j_2. We conclude that (5-1) is satisfied.

Similarly, if we had equality in (5-3), then necessarily $j = j_s + 1$ and $\lambda_{i,j} = \ell_s - 1$. On the other hand, the assumption that $\bar{\lambda}_i^{\ell_1} + \bar{\lambda}_i^{\ell_2} > \bar{\lambda}_{i-1}^{\ell_1} + \bar{\lambda}_{i-1}^{\ell_2}$ implies that if $\lambda_{i,j} < \lambda_{i-1,j}$ for some j, then $\lambda_{i,j} \neq \ell_1 - 1, \ell_2 - 1$. Therefore, (5-3) is an inequality as stated.

By the same reasoning, if $\delta_i = j_1 > j_2$, then equality in (5-2) is also impossible: because $m_1 = 0$ replacing j_2 by δ_i increases the left-hand side by at least $2 + m_2$. On the other hand, if $\delta_i = j_2$ then replacing j_1 by δ_i decreases the left-hand side, making it too small to satisfy (5-2).

Finally, suppose we have some j such that $\lambda_{i,j} < \lambda_{i-1,j}$; say $\lambda_{i,j} = \lambda_{i-1,j} - p$ for some $p > 0$. Then equality in (5-4) is equivalent to

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} = n - 2 - \ell_1 - \ell_2 + p.$$

If as above $j_s \neq \delta_i$, then necessarily $j > j_s$. Then, by definition of j_s we must have $\lambda_{i,j} < \ell_s$. On the other hand, since we have assumed that $\lambda_{i-1}^{\ell_1} + \lambda_{i-1}^{\ell_2} < n$, we must have that $\lambda_{i,j}, \ldots, \lambda_{i,j} + p$ does not contain ℓ_s. It follows that $\lambda_{i,j} + p < \ell_s = \lambda_{i,j_s} - m_s$. We conclude that $j - \lambda_{i,j} > 1 + j_s - \lambda_{i,j_s} + p + m_s$, so

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} > (n - 2 - \ell_1 - \ell_2 - m_s) + 1 + p + m_s = n - 1 - \ell_1 - \ell_2 + p,$$

proving (5-4).

Case when λ_i has no entries in the ℓ_2-th column. Then, $\bar{\lambda}_i^{\ell_2} = 0$, δ_i is the row of the last square in the ℓ_1-th column and $\lambda_{i,\delta_i} = \ell_1$. As the rows are numbered starting at 0, the condition $\bar{\lambda}_i^{\ell_1} + \bar{\lambda}_i^{\ell_2} = n$ becomes $\delta_i + 1 = n$. Hence,

$$\delta_i - \lambda_{i,\delta_i} = (\delta_i + 1) - 1 - \ell_1 = n - 1 - \ell_1.$$

But since the ℓ_2-th column is empty for all j, we have $\lambda_{i,j} < \ell_2$, so we find that

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} > n - 1 - \ell_1 + j - \ell_2 \geq n - 1 - \ell_1 - \ell_2,$$

showing that (5-1) holds, and that equality in (5-3) occurs only if $\lambda_{i,j} = \ell_2 - 1$. But if we assume $\lambda_{i,j} < \lambda_{i-1,j}$, then $\lambda_{i-1,j} \geq \ell_2$, contradicting the assumption that $\bar{\lambda}_{i-1}^{\ell_1} + \bar{\lambda}_{i-1}^{\ell_2} < n$. Therefore, (5-3) holds.

We also see that

$$2\delta_i - 2\lambda_{i,\delta_i} = 2n - 2 - 2\ell_1 > 2n - 2 - \ell_1 - \ell_2 > n - 2 - \ell_1 - \ell_2,$$

so (5-2) holds. Finally, as $\lambda_{i-1}^{\ell_1} + \lambda_{i-1}^{\ell_2} < n$, then $\lambda_{i-1,j} < \ell_2$, so

$$\delta_i + j - \lambda_{i-1,\delta_i} - \lambda_{i-1,j} = \delta_i + j - \lambda_{i,\delta_i} - \lambda_{i-1,j} + 1 > n - 1 - \ell_1 - \ell_2 + 1 = n - \ell_1 - \ell_2,$$

proving (5-4).

Case when λ_i has the same number of entries in the ℓ_1-th and ℓ_2-th columns. So we must have n even, with $\delta_i + 1 = n/2$, and also $\lambda_{i,\delta_i} = \ell_2$. In this case, we have

$$\delta_i - \lambda_{i,\delta_i} = (\delta_i + 1) - 1 - \ell_2 = n/2 - 1 - \ell_2.$$

Therefore

$$2\delta_i - 2\lambda_{i,\delta_i} = n-2-2\ell_2 < n-2-\ell_1-\ell_2,$$

proving (5-2).

For $j_1 < \delta_i,\ \lambda_{i,j_1} \geq \lambda_{i,\delta_i}$. Therefore

$$j_1 + \delta_i - \lambda_{i,j_1} - \lambda_{i,\delta_i} < 2\delta_i - 2\lambda_{i,\delta_i} < n-2-\ell_1-\ell_2,$$

proving (5-1) in this case. For $j_1 > \delta_i$, as the ℓ_1-th column has exactly $\delta_i + 1$ entries,

$$\lambda_{i,j_1} \leq \ell_1 - 1 = \ell_2 - (\ell_2-\ell_1) - 1 = \lambda_{i,\delta_i} - (\ell_2-\ell_1) - 1,$$
$$j_1 + \delta_i - \lambda_{i,j_1} - \lambda_{i,\delta_i} \geq \delta_i + 1 + \delta_i - 2\lambda_{i,\delta_i} + \ell_2 - \ell_1 + 1 = n - 2\ell_2 + \ell_2 - \ell_1 = n - \ell_2 - \ell_1,$$

completing the proof of (5-1). Moreover, we can have equality in (5-3) only if $j = \delta_i + 1$ and $\lambda_{i,j} = \ell_1 - 1$, so (5-3) holds under the stated condition $\lambda_{i,j} < \lambda_{i-1,j}$. Finally, if $\lambda_{i,j} = \lambda_{i-1,j} - p$ for $p > 0$, then in order to have equality in (5-4) we would need to have $j > \delta_i$, which implies that $\lambda_{i,j} + p < \ell_1 = \lambda_{i,\delta_i} - \ell_2 + \ell_1$, so

$$\delta_i + j - \lambda_{i,\delta_i} - \lambda_{i,j} > (n-2-2\ell_2) + 1 + (\ell_2 - \ell_1 + p) = n - 1 - \ell_1 - \ell_2 + p,$$

again yielding (5-4).

This completes the proof of the lemma in the case that $\bar{\lambda}_{i'} = \lambda_{i'}$ for all i'. We will see that the general case follows.

General case ($\bar{\lambda}_{i'}$ not necessarily equal to $\lambda_{i'}$). The main observation is the following: if $\bar{\lambda}_{i,j} = \bar{\lambda}_{i-1,j} + 1$, and we let j' be such that $\bar{a}_j^{i+1} = a_{j'}^{i+1}$, then we necessarily have $\lambda_{i,j'} = \lambda_{i-1,j'} + 1$, and we cannot have any swaps in the i-th column involving the j'-th row. Indeed, the identity $\bar{\lambda}_{i,j} = \bar{\lambda}_{i-1,j} + 1$ means that we have $\bar{a}_j^i = \bar{a}_j^{i+1}$, which means that $a_{j'}^{i+1} = a_{j''}^i$ for the j'' such that exactly j values of a_m^i are less than $a_{j''}^i$. We also have exactly j values of a_m^{i+1} less than $a_{j'}^{i+1}$. It then follows that we must have $j'' = j'$: we cannot have $a_{j'}^i > a_{j''}^i$, since then we would have $a_{j'}^i > a_{j'}^{i+1}$. But if $a_{j'}^i < a_{j''}^i$, then j' occurs among the values of m with $a_m^i < a_{j''}^i$, so there is necessarily some m with $a_m^{i+1} < a_{j'}^{i+1}$ but $a_m^i \geq a_{j''}^i$, again leading to a contradiction. This proves the observation, noting that the fact that $j' = j''$ rules out any swaps involving the j'-th row.

Note that equations (5-1) and (5-2) can be phrased in terms of the values of $j - \lambda_{i',j} = a_j^{i'+1} - g(i')$. Using $\bar{\delta}_i$ to denote the values of δ coming from $\overline{T}$, our above observation implies that we have $a_{\delta_i}^i = a_{\delta_i}^{i+1} = \bar{a}_{\bar{\delta}_i}^{i+1} = \bar{a}_{\bar{\delta}_i}^i$. Therefore, the proof of these two equations follows from the case $\bar{\lambda}_{i'} = \lambda_{i'}$.

Next, suppose that we have some j with $\lambda_{i,j} < \lambda_{i-1,j}$. We claim that if j' is such that $a_j^{i+1} = \bar{a}_{j'}^{i+1}$, and j'' is such that $a_j^i = \bar{a}_{j''}^i$, then we necessarily also have that $\bar{\lambda}_{i,j'} < \bar{\lambda}_{i-1,j'}$ and $\bar{\lambda}_{i,j''} < \bar{\lambda}_{i-1,j''}$. Given this claim, (5-3) and (5-4) follow from the case that $\bar{\lambda}_{i'} = \lambda_{i'}$ for all i'. By our above observations on the case $\bar{\lambda}_{i,j} = \bar{\lambda}_{i-1,j} + 1$, it suffices to prove that $\bar{\lambda}_{i,j'} \neq \bar{\lambda}_{i-1,j'}$ and $\bar{\lambda}_{i,j''} \neq \bar{\lambda}_{i-1,j''}$, Equivalently, $\bar{a}_{j'}^{i+1} \neq \bar{a}_{j'}^i + 1$ and $\bar{a}_{j''}^{i+1} \neq \bar{a}_{j''}^i + 1$. In order to have $a_j^{i+1} = \bar{a}_{j'}^{i+1} = \bar{a}_{j'}^i + 1$, we would need to have $a_j^{i+1} - 1$ occurring among the $a_\bullet^i$, with precisely j' strictly smaller values also occurring. By definition we have j' values strictly smaller than a_j^{i+1} occurring in $a_\bullet^{i+1}$, and using our observation on lack of swaps when $\bar{\lambda}_{i,j} = \bar{\lambda}_{i-1,j} + 1$, we see that every one of these also must yield a value of $a^i_\bullet$ strictly smaller than $a^{i+1}_j - 1$. But we have in addition that $a^i_j < a^{i+1}_j - 1$, so we conclude that there are at least $j' + 1$ values in $a^i_\bullet$ strictly less than $a^{i+1}_j - 1$, proving the desired inequality by contradiction.

Similarly, in order to have $\bar{a}^{i+1}_{j''} = \bar{a}^i_{j''} + 1 = a^i_j + 1$, we would need to have $a^i_j + 1$ occurring among the $a^{i+1}_\bullet$, with precisely j'' strictly smaller values also occurring. By definition, we have only j'' values among the $a^i_\bullet$ strictly smaller than a^i_j, and every value of $a^{i+1}_\bullet$ which is strictly smaller than $a^i_j + 1$ must come from one of these. But again using our observation on the lack of swaps when $\bar{\lambda}_{i,j} = \bar{\lambda}_{i-1,j} + 1$, we see that the value $a^i_j + 1$ in $a^{i+1}_\bullet$ must itself come from a row in $a^i_\bullet$ with value strictly smaller than a^i_j, so we conclude that if $a^i_j + 1$ occurs in $a^{i+1}_\bullet$, there must be strictly fewer than j'' entries in $a^{i+1}_\bullet$ which are strictly smaller than it. This proves the claim, and the lemma. □

6. An independence criterion

Suppose we have a limit linear series, and fix choices of sections s^i_j matching the vanishing orders in our table. We make the following definition:

Definition 6.1. Given an unimaginative multidegree ω, for all (j_1, j_2), let $n_{(j_1,j_2)}$ be the number of places (i.e., collections of contiguous columns) where the (j_1, j_2) row is potentially present in the multidegree ω. Let $s_{(j_1,j_2),i}$ for $i = 1, \dots, n_{(j_1,j_2)}$ be the induced sections in multidegree ω with precisely the given support. Then the full collection of $s_{(j_1,j_2),i}$ are the *potentially present* sections in multidegree ω, and their span in $\Gamma(X_0, (\mathscr{L}^{\otimes 2})_\omega)$ is the *potential ambient space*.

Note that in the above, we require that each $s_{(j_1,j_2),i}$ be potentially starting in its first column of support and potentially ending in its last column of support. Thus, there may be individual columns in which the (j, j') row satisfies the inequalities to be potentially present in that column, but which does not occur in any of the $s_{(j_1,j_2),i}$ because it fails inequalities in other columns.

The $s_{(j_1,j_2),i}$ are each unique up to scaling given a choice of the s^i_j. The s^i_j are not unique, but they can differ only by multiples of $s^i_{j'}$ with strictly higher vanishing at both points. Then if s^i_j has potential support (in the i-th column), necessarily $s^i_{j'}$ has a connected component of potential support consisting precisely of the i-th column. We conclude that the potential ambient space is independent of the choice of the s^i_j. Consequently, the dimension of the span — and in particular the linear independence — of the potentially appearing sections is likewise independent of choices.

Fix a multidegree ω and consider the images of $s_j \otimes s_{j'}$ focusing attention on potentially present sections. Assume we have a linear combination of images of sections equal to 0. As in [Liu et al. 2021],, we prove successively that the coefficient of particular sections must be zero. When this happens, we say that we "*drop*" that section and talk about *"remaining" sections* (i.e., those which have not yet been dropped).

Definition 6.2. We say that the i-th column of T is *semicritical* in multidegree ω if it satisfies the following conditions:

- it has a value of δ_i (see Definition 5.1), in particular, it has genus 1;

- on the potentially present sections remaining, the two subcolumns of column i exhibiting the minimal values add to at least $2d-2$;
- if the (j, δ_i) row remains in the i-th column for some $j \neq \delta_i$, then the j-th row is not exceptional.

If further the minimal values among the remaining potentially present sections are not both one less than the values in the (δ_i, δ_i) row, we say that the i-th column is *critical*.

We start with the following independence criterion:

Lemma 6.3. *For a given limit linear series, and given unimaginative multidegree ω, we can drop potentially present sections if they satisfy the following rules*:

(i) *If in some column i, there is a unique remaining potentially present section supported in that column having minimal $a^i_{(j,j')}$ value, or a unique one having minimal $b^i_{(j,j')}$, then the one achieving the minimum may be dropped.*

(ii) *If there are at most two remaining potentially present sections with support in some genus-*1 *column i, and neither of them involves an exceptional row,* (*see Definition 5.1*), *then they can both be dropped.*

(iii) *If there are $i < i'$ such that the block of columns from i to i' has the following properties, then all the remaining potentially present sections supported in this block can be dropped*:

- *There are at most* 3 *remaining potentially present sections supported in each of the i-th and i'-th columns.*
- *Within the block, there are at most three potentially present sections continuing from any column to the next.*
- *Every column strictly between i and i' has degree* 2.
- *Both the i-th column and the i'-th column are semicritical and either i is critical with no remaining potentially present section ending in the i-th column, or i' is critical with no remaining potentially present section starting in the i'-th column.*

Proof. Suppose we had a hypothetical linear dependence among the potentially present sections. We claim that in each case (i), (ii), (iii), the coefficients of the relevant potentially present sections would be forced to vanish.

In case (i), the uniqueness of the minimal value of $a^i_{(j,j')}$ (or of $b^i_{(j,j')}$) means that $s^i_{(j,j')}$ vanishes to strictly smaller order at P_i than any other remaining potentially present section, which forces us to drop that section.

In case (ii), we need to see that for a fixed column i, any two $s^i_{(j,j')}$ have to be linearly independent provided that they do not involve any exceptional rows. If either of them involves δ_i, this is automatic, since either the $a^i_{(j,j')}$ or $b^i_{(j,j')}$ values are forced to be distinct. On the other hand, if neither involves δ_i, we claim that the sections in question must have distinct zeroes on Z_i away from P_i and Q_i. Indeed, if we have a, b, a', b' with $a+b=d-1=a'+b'$, then the unique sections s, s' of the given line bundle vanishing to order at least a at P_i and b at Q_i (respectively, a' at P_i and b' at Q_i) have

$$\operatorname{div} s = aP_i + bQ_i + R \quad \text{and} \quad \operatorname{div} s' = a'P_i + b'Q_i + R'$$

for some R, R'. We see that we have a linear equivalence $R - R' \sim (a'-a)P_i + (b'-b)Q_i$, and if $0 \le a, a' \le d$, we see that $R \neq R'$ because of the generality hypothesis on P_i, Q_i. Thus, tensors of different sections of this form always have zeroes in distinct places on Z_i, and must be linearly independent.

For case (iii), note that the condition that the degree is 2 on every column between i and i' means by Corollary 5.8 that there is at most one potentially starting and at most one potentially ending section in each of these columns. Noting that the situation is fully symmetric, suppose without loss of generality that i' is critical, with no remaining potentially present sections starting in it. If i or i' has fewer than three remaining potentially present sections, we may use (ii) to drop these, and then move iteratively through the rest of the block, using that at most one section starts or ends in each column together with (i) to drop the remaining sections. Thus, suppose that i and i' both have three remaining potentially present sections. Note also that if any column i'' has only one potentially present section spanning i'' and $i''+1$, then the minimal value in the right subcolumn of i'' is necessarily unique, so we can use case (i) to drop the section in question. Moreover, there can be at most one other potentially present section supported in column i'' (the one ending there), so we can drop this one as well, and then we can move iteratively left and right to drop the entire block. Thus, we may further suppose that every column in the block has at least two potentially present sections spanning it and the next column.

Next, normalize the sections as follows: scale all sections spanning the $i'-1$ and i' column so they agree at $Q_{i'-1}$, and then go back one column at a time, scaling any newly appearing section so that its value at the previous node agrees with the value of a section which has already been fixed. In this way, we will fix a normalization of all the sections except for those which are supported in only one column. Although the normalization depends on some choices, they are of a discrete nature, and can be fixed based purely on the discrete data of the limit linear series.

Now, consider a hypothetical nonzero linear dependence involving the rows in our block. First, the coefficients of the linear dependence cannot vanish identically in the remaining potential sections of any column, since otherwise the condition that at most one potentially present section ends in each column would imply that there was a column with exactly one nonzero coefficient among its remaining sections. Next, we see that the coefficients are unique up to simultaneous scaling for the three potentially present sections in column i. Indeed, since we have assumed that i is semicritical, its three potentially present sections must be pairwise independent.

Since we have at most one new potentially present section in each column, we find that the coefficient for any new one is always uniquely determined by the previous ones. Since there are no new potentially present sections in column i', we find that even before considering this column, we have already uniquely determined all of the coefficients (up to simultaneous scaling) of all of the potentially present sections remaining in the block. Moreover, we claim that these coefficients (excluding the ones for potentially present sections supported only in a single column) are uniquely determined up to finite indeterminacy by the marked curves $Z_i, \dots, Z_{i'-1}$ together with the discrete data of the limit linear series. Indeed, there are only two ways in which nontrivial moduli can enter the picture: if there are columns i'' between i and $i'-1$ either having no $\delta_{i''}$, or having some sections $s_j^{i''}$ which are not uniquely determined up to scalar. This becomes slightly delicate, since in both these cases, varying the moduli could affect both the normalization we have chosen and the linear dependence. However, we will show that in both cases, there will in fact be only finitely many possibilities which still preserve the linear dependence. Note that by hypothesis, neither of these nontrivial moduli occurs in the i-th column. Note also that we cannot have both occurring at once, as the $s_j^{i''}$ can only fail to be determined up to scalar if they involve an exceptional row, and since we have assumed we have degree 2 between i and i', these can only appear if paired with the $\delta_{i''}$ row.

First consider the case that we have no $\delta_{i''}$. Then since we have degree 2, every potentially present section in column i'' must extend to both the preceding and subsequent columns. By assumption, there are then at most three such sections. If there are fewer than three, they cannot be independent, leading to an immediate contradiction. As a line bundle of degree two on an elliptic curve has at most two independent sections, if there are three, say $s_0^{i''}, s_1^{i''}, s_2^{i''}$, they are necessarily dependent, with a unique dependence $c_0 s_0^{i''} + c_1 s_1^{i''} + c_2 s_2^{i''} = 0$ which can be determined by requiring that it holds at both $P_{i''}$ and $Q_{i''}$. We claim that for any fixed choice of c_0, c_1, c_2 (not all zero), there are only finitely many choices of the line bundle $\mathscr{L}^{i''}$ such that the resulting cancellation holds at both points. For this claim, we can renormalize our sections so that the values of the $s_j^{i''}$ agree at $P_{i''}$. We want to see that the values at $Q_{i''}$ move nondegenerately in $\mathbb{P}^2$ as $\mathscr{L}^{i''}$ varies. But this is precisely the content of Proposition 2.5.

Next, suppose that we have an exceptional row j involved in column i'', necessarily paired with the $\delta_{i''}$ row. As before, a linear dependence in the i'' necessarily has to give cancellation at both $P_{i''}$ and $Q_{i''}$. Suppose that the j-th row and the $\delta_{i''}$-th row have entries a, b and a', b' respectively, so that $a+b=d-2$ and $a'+b'=d$. There are two cases: if $a=a'-1$, so that also $b=b'-1$ (and i'' has a swap in it), then the moduli for the section $s_j^{i''}$ consists simply of adding multiples of the section $s_{\delta_{i''}}^{i''}$, which doesn't affect the value at either $P_{i''}$ or $Q_{i''}$, and only affects the coefficient of the $(\delta_{i''}, \delta_{i''})$ row, which in this case is supported purely in the i'' column. On the other hand, if $a \neq a'-1$, observe that since the degree is 2 in this column, we cannot have any other sections involving $\delta_{i''}$ starting or ending in the column, and therefore we have no sections starting or ending in the column. Thus, there are at most three potentially present sections in column i'', and the other ones can't involve any exceptional row and must therefore be linearly independent. It follows that in our linear dependence, the coefficient of $s_{(j,\delta_{i''})}$ must be nonzero. Now, varying s_j will change the relationship between the values at $P_{i''}$ and $Q_{i''}$ (we can view the moduli for s_j as adding multiples of a section vanishing to order $a+1$ at $P_{i''}$ and order b at $Q_{i''}$). Since this variation of moduli affects only a single potentially present section, and we know it must have nonzero coefficient in our linear dependence, there is only one choice of $s_j^{i''}$ compatible with the previously determined linear dependence, and we have no nontrivial moduli in this case.

Finally, note that although our normalization was not determined for potentially present sections supported in a single column, scaling these does not affect the coefficients of any of the sections spanning the $i'-1$ and i' column, so we have that the possible coefficients of these sections are determined up to finitely many possibilities. It thus suffices to show that if we vary the gluing points on the component corresponding to the final column, the (unique, if it exists) linear independence on the three potentially present sections varies nontrivially.

As there are no remaining potentially present section starting in the i'-th column, the three rows in its left subcolumn necessarily have the same a value. Let b be the minimal value for the right subcolumn. By criticality, $a+b=2d-2$. Using (i), there are two cases to consider, either b is attained twice, or in all three rows. The last condition in the definition of criticality implies that none of the (a,b) rows are obtained by adding the $\delta_{i'}$ row to an exceptional row. Now, if all three rows are (a,b) rows, we can directly apply Proposition 2.2 to conclude that the linear dependence in the i'-th column varies nontrivially with $P_{i'}$, $Q_{i'}$, as desired. On the other hand, if two rows are (a,b) rows, we again apply Proposition 2.2 to these two rows, and since we have normalized all three rows so that the values at $P_{i'}$ agree, we again see that the linear dependence among the three has to vary nontrivially with $P_{i'}$, $Q_{i'}$, as desired. □

7. The case $r=6$

We now specialize to $r=6$, $g=21+\epsilon$ and $d=24+\epsilon$ for some $\epsilon\geq 0$ (so that $\rho=\epsilon$). As the total degree is $2d=2g+6=3\times 6+2\times(g-6)$, a multidegree can be determined by placing threes in six columns, and twos in the rest.

Definition 7.1. For a limit linear series and with the $\bar{\lambda}_i$ of Definition 5.11, the *default multidegree* ω_{def} is determined by placing a 3

(1) in the first column;

(2) in the first column with $\bar{\lambda}_i^1+\bar{\lambda}_i^2=5$;

(3) in the first column with $\bar{\lambda}_i^1+\bar{\lambda}_i^3=7$;

(4) in the column immediately after the last column with $\bar{\lambda}_i^1+\bar{\lambda}_i^3=7$;

(5) in the column immediately after the last column with $\bar{\lambda}_i^2+\bar{\lambda}_i^3=9$;

(6) in the last column.

As we are assuming that the first and last component in the chain have genus one and $\bar{\lambda}_i^\ell$ can only increase in a genus-1 column, the default multidegree is unimaginative. Also, as the goal is to fill an $(r+1)(g-d+r)=7\times 3$ rectangle, conditions (2)–(5), (3)–(4) and of course (1)–(6) are symmetric with respect to flipping the start and end of the chain. Recall that an unimaginative multidegree has degrees 2 or 3 on every elliptic component and that γ_i denotes the number of 3s in the first i components(see Definition 5.5).

Proposition 7.2. *Fix an unimaginative multidegree. Then for a column i, there can be at most three rows spanning columns i and $i+1$ except in the following circumstances*:

(i) $\gamma_i=0$ *and* $\bar{\lambda}_i^1+\bar{\lambda}_i^3\geq 8$;

(ii) $\gamma_i = 2$ *and* $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 \geq 7$;

(iii) $\gamma_i = 4$ *and* $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 \leq 7$;

(iv) $\gamma_i = 6$ *and* $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 \leq 6$.

In particular, in the default multidegree there are never more than three rows spanning a given pair of columns.

Proof. We will use the criterion from Proposition 5.7. Since this only involves the values of $j - \lambda_{i,j} = a_j^{i+1} - g(i)$, the general case reduces immediately to the notationally simpler situation that $\bar{\lambda}_i = \lambda_i$ for all i. We thus assume that we are in this situation. Then, because the sequence $j - \lambda_{i,j}$ is strictly increasing in j, we see that pairs (j_1, j_2) satisfying the identity for appearing in the i-th and $(i+1)$-st columns from Proposition 5.7 must be strictly nested, so we can have at most $r/2+1 = 4$ of them, and we only have all of these if $\lambda_{i,j} + \lambda_{i,r-j}$ is constant for all j, in particular $2\lambda_{i,r/2} = \lambda_{i,0} + \lambda_{i,r}$. Moreover,

$$\gamma_i = r - 2\lambda_{i,r/2} = 6 - 2\lambda_{i,r/2}, \qquad \gamma_i = r - \lambda_{i,j} - \lambda_{i,r-j}, \quad j = 0, \dots, r;$$

in particular γ_i is even. Adding these identities, we find that

$$\sum_{j=0}^{r} \lambda_{i,j} = \frac{(r+1)(r-\gamma_i)}{2} = 7\big(3 - \tfrac{1}{2}\gamma_i\big),$$

so $\lambda_{i,r/2} = 3 - \gamma_i/2$.

If $\gamma_i = 0$, then $\lambda_{i,r/2} = 3$. As $r/2 = 3$ and we start numbering the sections at 0, there are at least 4 values of λ that contribute to $\bar{\lambda}_i^3$ (and therefore to $\bar{\lambda}_i^1$). We conclude that $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 \geq 8$.

If $\gamma_i = 2$, then $\lambda_{i,r/2} = 2$. Let n be the number of values of j with $\lambda_{i,j} \leq 0$. Then also $\lambda_{i,r-j} \geq 4$ for the same n values of j. So $\lambda_i^1 + \lambda_i^3 \geq (r+1-n) + n = 7$, as desired.

Similarly, if $\gamma_i = 4$ then $\lambda_{i,r/2} = 1$. If there are n values of j with $\lambda_{i,j} \geq 3$, then also $\lambda_{i,r-j} \leq -1$. As before we find $\lambda_i^1 + \lambda_i^3 \leq (r+1-n) + n = 7$.

Finally, if $\gamma_i = 6$ then $\lambda_{i,r/2} = 0$. Therefore, $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 \leq 6$, as claimed. □

We can now prove the following theorem, which will in particular prove the desired maximal rank statement in all sufficiently nondegenerate cases for all ϵ in our family of cases. It will also suffice to prove the genus-22 case of our main theorem.

Theorem 7.3. *In the default multidegree, we can always drop all potentially present sections using the rules from Lemma 6.3, so the potentially present sections are all linearly independent.*

Proof. The vanishing of sections of a line bundle of degree d at Q_1 is at most either d or $d-1$(but not both), $d-2, d-3\dots$. So, at most the rows (0, 0), (0, 1) and (0, 2) can be among the potentially present if there is no swap and at most the rows (0, 1) and (1, 1) can be potentially present if there is a swap. In both cases, these sections have distinct orders of vanishing at Q_1, so they can be dropped.

According to Corollary 5.8, we will have at most one new row with a potentially present section in each column until we get to the next column of degree 3, so these can all be dropped.

Now, suppose that i is minimal such that $\bar{\lambda}_i^1 + \bar{\lambda}_i^2 = 5$. Take $\ell_1 = 1$ and $\ell_2 = 2$ in Lemma 5.12, so $\gamma_i - 1 = 1 = 5 - 1 - \ell_1 - \ell_2$. From Lemma 5.12 and Corollary 5.9, we have no potentially present sections supported entirely in the i-th column. Any other new potentially present sections would have to be supported in the i-th and $(i+1)$-st columns, so by Proposition 7.2, we have at most three of these. Choose i' minimal so that $\bar{\lambda}_{i'}^1 + \bar{\lambda}_{i'}^2 = 6$, then with $\ell_1 = 1$ and $\ell_2 = 2$, $\gamma_{i'} = 2 = 6 - 1 - \ell_1 - \ell_2$. According to Corollary 5.8 and Lemma 5.12, there is no row starting in the i'-th column. Then the i-th (respectively, i'-th) columns are critical: if a, b are the minimum values in the subcolumns, they have to add to at least $2d-2$ or the rows would not be potentially starting in the i-th column (respectively, potentially supported in the i'-th column). The last condition of semicriticality and the condition for criticality then follow from the second and first parts of Lemma 5.12, respectively. It follows that the hypotheses of Lemma 6.3(iii) are satisfied, so we can drop all rows occurring in this block. We can then again handle any additional columns before the next degree-3 one.

The setup being symmetric, we can also start at the end of the chain and in the same manner, eliminating all potentially present sections occurring in any columns outside the middle two degree-3 columns. For these columns, we are considering $\ell_1 = 1$ and $\ell_2 = 3$, so we have

$$\gamma_{i+1} - 1 = 2 = 7 - 1 - \ell_1 - \ell_2 \quad \text{and} \quad \gamma_{i+1} - 1 = 3 = 8 - 1 - \ell_1 - \ell_2,$$

respectively, and according to Corollary 5.9 and Lemma 5.12, neither column has any potentially present section supported entirely in it. As before, we find we must have a block satisfying the hypotheses of Lemma 6.3(iii), which we can then eliminate. □

If the specialization of our linear series contains the "expected" sections s_j for every $j = 0, \dots, r$ in the expected multidegrees ω_j (as in Proposition 4.9), then Theorem 7.3 implies that the images of each $s_j \otimes s_{j'}$ in the default multidegree are linearly independent, so the multiplication map has the desired rank $\binom{r+2}{2} = 28$. However, some linear series may have more degenerate specializations. The remainder of the paper will be devoted to applying Theorem 7.3 (and variants thereof) to handle these situations as well. For this, the statement in terms of potentially present sections (as opposed to the separate rows considered in [Liu et al. 2021]) is crucial. In interesting cases, we can have strictly more than 28 potentially present sections. This does not contradict the fact that the multiplication map can have rank at most 28, because these do not occur separately in the linked linear series coming as the specialization of any fixed family of linear series on the smooth fibers. In most limits, for every (j_1, j_2) we will have a unique linear combination of the potentially present sections in the (j_1, j_2) row which actually arise in the specialization. What makes the degenerate cases more interesting is that in these cases, we may have more than one linear combination occurring from a given row, precisely in situations where the specialization fails to contain any potentially present sections from some other row — see Example 8.3.

Ultimately, the default multidegree used in Theorem 7.3 will be sufficient to handle the genus-22 case, and most of the genus-23 cases. However, for certain degenerate cases we will need to consider other multidegrees instead. The following results allows for some flexibility in the choice of multidegree while maintaining linear independence.

Proposition 7.4. *Suppose ω is an unimaginative multidegree determined by placing degree 3 in genus-1 columns as follows*:

(1) *In one column which is either the first, or a column with no exceptional rows and satisfying $\bar{\lambda}_i^1+\bar{\lambda}_i^2 \le 4$ and $\bar{\lambda}_{i,0} \le 2$.*

(2) *In one column with $\bar{\lambda}_i^1+\bar{\lambda}_i^2 = 5$ but $\bar{\lambda}_{i-1}^1+\bar{\lambda}_{i-1}^2 = 4$.*

(3) *In one column between the first column with $\bar{\lambda}_i^1+\bar{\lambda}_i^2 = 6$ and the first column with $\bar{\lambda}_i^1+\bar{\lambda}_i^3 = 7$ (inclusive).*

(4) *In one column between the column immediately after the last column with $\bar{\lambda}_i^1+\bar{\lambda}_i^3 = 7$ and the column immediately after the last column with $\bar{\lambda}_i^2+\bar{\lambda}_i^3 = 8$ (inclusive).*

(5) *In one column with $\bar{\lambda}_i^2+\bar{\lambda}_i^3 = 10$ but $\bar{\lambda}_{i-1}^2+\bar{\lambda}_{i-1}^3 = 9$.*

(6) *In one column which is either the last, or a column with no exceptional rows and satisfying*

$$\bar{\lambda}_{i-1}^2+\bar{\lambda}_{i-1}^3 \ge 10 \quad \textit{and} \quad \bar{\lambda}_{i-1,6} \ge 1.$$

Then the potentially present sections in multidegree ω are still linearly independent.

Proof. The main new ingredient is verifying that if we place the first degree 3 in a (genus-1) column after the first, but still satisfying $\bar{\lambda}_i^1+\bar{\lambda}_i^2 \le 4$ and $\bar{\lambda}_{i,0} \le 2$, then provided we also have no exceptional rows (and therefore no swaps), we will in fact obtain at most two potentially present sections starting in the i-th column. By Proposition 5.7, for the (j, j') row to have a potentially present section starting in the i-th column, we will need $j+j'-\lambda_{i-1,j}-\lambda_{i-1,j'} > \gamma_{i-1} = 0$ and $j+j'-\lambda_{i,j}-\lambda_{i,j'} \le \gamma_i = 1$, or equivalently

$$\lambda_{i-1,j}+\lambda_{i-1,j'} < j+j' \le 1+\lambda_{i,j}+\lambda_{i,j'}. \tag{7-1}$$

As there are no swaps in the i-th column, it suffices to check this assertion with $\bar{\lambda}_i = \lambda_i$ for all i. Then, $\bar{\lambda}_i^1+\bar{\lambda}_i^2 \le 4$ implies $\lambda_{i,j} \le 0$ for $j \ge 4$ and $\lambda_{i,j} \le 1$ for $j = 2, 3$. It follows that to satisfy the right-hand inequality above, we must have at least one of j, j' equal to 0 or 1. Moreover, by Corollary 5.9 for $j = 0, 1$, if $j \ne \delta_i$, then there is at most one value of j' satisfying the above inequalities. In particular, we conclude that if $\delta_i \ne 0, 1$, there are at most two potentially present sections, as claimed.

Assume $\delta_i = 0$, and show that at most two rows of the form $(0, j')$ are present in the i-th column, and if two are present, then none of the form $(1, j')$ is for $j' > 0$. Suppose first $(0, 0)$ is potentially starting in the i-th column. By (7-1) this could only happen if $\lambda_{i-1,0} < 0$, so $\lambda_{i,j'} = \lambda_{i-1,j'} < 0$ for all $j' > 0$, and then $(0, j')$ cannot satisfy the right-hand side of (7-1) for any $j' > 0$. On the other hand, if $j'' > j' > 0$ are such that $(0, j')$ and $(0, j'')$ are both present, then

$$\lambda_{i-1,0}+\lambda_{i-1,j'} < j' < j'' \le 1+\lambda_{i,0}+\lambda_{i,j''} \le 2+\lambda_{i-1,0}+\lambda_{i-1,j'},$$

so the only possibility is that $j' = 1+\lambda_{i-1,0}+\lambda_{i-1,j'}$ and $j'' = j'+1$, with $\lambda_{i,j''} = \lambda_{i,j'}$. It follows that no other $(0, j''')$ is present for $j''' \ne 0, j', j''$. Moreover, $(1, j''')$ cannot be potentially present for any $j''' > 0$ in this situation: If the $(1, 1)$ row were present, (7-1) implies $\lambda_{i-1,1} \le 0$, $\lambda_{i,1} \ge 1$ against the assumption $\delta_i = 0$. As $1 + j'''$ will be too large if $j''' \geq j'$, in order to have $(1, j''')$, $j''' \geq 2$ present we would need $j'' \geq 4$. But the original assumptions imply $1 + \lambda_{i,0} + \lambda_{i,j''} \leq 3$, contradicting (7-1).

Finally, consider the case that $\delta_i = 1$. If the $(1, 1)$ row is potentially starting in the i-th column, by parity we have $1 = \lambda_{i,1}$. So for all $j > 1$ $\lambda_{i,j} = \lambda_{i-1,j} \leq \lambda_{i-1,1} = 0$. Then we cannot have $(1, j')$ potentially starting for any $j' > 1$, so we have at most two rows potentially starting. On the other hand, if we have $j'' > j' > 1$ potentially starting in the i-th column, we are just as above forced to have $j' = \lambda_{i-1,1} + \lambda_{i-1,j'}$ and $j'' = j' + 1$, with $\lambda_{i,j''} = \lambda_{i,j'}$, and we claim we cannot have $(0, j''')$ potentially starting for any j'''. Indeed, if $j''' \leq j''$, then we have $j''' - \lambda_{i-1,j'''} \leq j'' - \lambda_{i-1,j''}$, so

$$\begin{aligned} j''' &\leq j'' - \lambda_{i-1,j''} + \lambda_{i-1,j'''} \\ &= 1 + \lambda_{i-1,1} + \lambda_{i-1,j''} - \lambda_{i-1,j''} + \lambda_{i-1,j'''} \\ &\leq \lambda_{i-1,0} + \lambda_{i-1,j'''}, \end{aligned}$$

violating (7-1). But $j'' \geq 3$, so if $j''' \geq 4$ we cannot satisfy (7-1) without violating our hypothesis that $\lambda_{i-1,0} \leq 2$. We thus conclude the desired statement on the number of potentially present sections starting in column i.

Now, since we have assumed that our first column with degree 3 has no exceptional rows, the fact that it has at most two potentially present sections starting in it means that we can still eliminate sections starting at the beginning of the chain until we reach the second column of degree 3, just as in the proof of Theorem 7.3 and the second column of degree 3 will still be critical, with at most three potentially present sections starting in it. The next step depends on the location of the third column of degree 3. If the first column with $\bar{\lambda}_i^1 + \bar{\lambda}_i^2 = 6$ still has degree 2, we will eliminate this block in increasing order, as before. On the other hand, if the first column with $\bar{\lambda}_i^1 + \bar{\lambda}_i^3 = 7$ has degree 2, we do not need to have eliminated everything in components with smaller indices in order to eliminate the central block, since the potentially supported rows in multidegree ω will be precisely the same as the potentially starting rows in ω_{def}. Thus, if the third column of degree 3 is strictly between these, we can eliminate both adjacent blocks first, and then eliminate all potentially present sections one by one from both sides until we reach this final column, which can have at most one remaining potentially present section by Corollary 5.9. However, if the third column of degree 3 is the first column with $\bar{\lambda}_i^1 + \bar{\lambda}_i^2 = 6$, we see that this will be critical with at most three potentially present sections ending in it, and we will instead eliminate the central block first, and then eliminate the block between the second and third columns of degree 3 last.

The situation is symmetric on the right, so we see that in all cases we will be able to eliminate all potentially present sections in a suitable order. □

It will also be important to consider moving degree 3 into a column with a swap, which we analyze below:

Lemma 7.5. *Choose a multidegree* $\mathrm{md}_{2d}(w)$ *for* $w = (c_2, \ldots, c_g)$ *that assigns degree* 3 *to the* i*-th column having a* δ_i. *If* $a^i_{\delta_i} = a^i_{j_0} + 1,\ a^{i+1}_{\delta_i} = a^{i+1}_{j_0} - 1$ *for some* j_0, *then there are at most four potentially starting* *sections on the i-th column. Moreover, if there are actually four either*

$$2\bar{a}^i_3 = c_i + 1, \tag{7-2}$$

or (δ_i, δ_i) *is potentially starting, and one of the following three possibilities holds*:

(1) $a^i_{(\delta_i,\delta_i)} = c_i + 1$;

(2) $a^i_{(\delta_i,\delta_i)} = c_i + 2$, *and the* (j_0, δ_i) *row is potentially starting*;

(3) $a^i_{(\delta_i,\delta_i)} = c_i + 3$, *with* $a^i_{\delta_i} = \bar{a}^i_4$.

At most four potentially present sections end in the i-th column, with four of them ending only if either

$$2\bar{a}^i_3 = c_i, \tag{7-3}$$

or if (δ_i, δ_i) *is potentially ending, and one the following three possibilities holds*:

(1) $a^i_{(\delta_i,\delta_i)} = c_i$, *with* $a^i_{\delta_i} = \bar{a}^i_3$;

(2) $a^i_{(\delta_i,\delta_i)} = c_i + 1$, *and the* (j_0, δ_i) *row is potentially ending*;

(3) $a^i_{(\delta_i,\delta_i)} = c_i + 2$;

We have written the above to allow for swaps having occurred prior to the i-th column. If no swaps have occurred, the j_0 in the lemma statement is necessarily $\delta_i - 1$, and the third exceptional case would require $\delta_i = 4$ (respectively, $\delta_i = 2$) in the statement on potential support starting (respectively, ending).

Proof. In order to have a potentially starting section in the (j, j') row, one needs $a^i_{(j,j')} > c_i$ and $a^{i+1}_{(j,j')} \leq c_{i+1} = c_i + 3$. It follows that if $j, j' \neq \delta_i$, then $a^i_{(j,j')} = c_i + 1$, and neither j nor j' equal to j_0. If $j' = \delta_i,\ j \neq \delta_i$, it is possible that $a^i_{(j,\delta_i)} = c_i + 2$, provided that $j \neq j_0$. And (j_0, δ_i) is potentially starting only if $a^i_{(j_0,\delta_i)} = c_i + 1$, or equivalently if $a^i_{(\delta_i,\delta_i)} = c_i + 2$. Recall that Corollary 5.9 says that if (j, δ_i) is potentially starting for some $j \neq \delta_i$, then there is no $j' \neq \delta_i$ with (j, j') also potentially starting. Next, we note that we can have at most two rows of the form (j, δ_i) potentially starting. Indeed, if $a^i_{(\delta_i,\delta_i)} = c_i + 3$, then $a^i_{(j,\delta_i)} = c_i + 2$ only for $j = j_0$, so the (j_0, δ_i) row does not occur, and we can have at most one additional row, having $a^i_{(j,\delta_i)} = c_i + 1$. On the other hand, if $a^i_{(\delta_i,\delta_i)} \neq c_i + 3$, then we have at most two rows, because they have to satisfy $a^i_{(j,\delta_i)} = c_i + 1$ or $c_i + 2$. We also observe that we can have a row of the form (j, j) for $j \neq \delta_i$ only for a unique choice of j, necessarily with $2a^i_j = c_i + 1$ and $j \neq j_0$, and then we cannot have (δ_i, δ_i) occurring, since $a^i_j \neq a^i_{\delta_i} - 1$ for $j \neq j_0$.

If no (j, δ_i) is potentially starting, then in particular (j_0, δ_i) also cannot be potentially starting. We can obtain at most three pairs (allowing one of them to have repeated entries) with fixed sum of vanishing. Similarly, if exactly one (j, δ_i) is potentially starting, then necessarily $j \neq j_0$, or we would be in the 2nd exceptional case with also (δ_i, δ_i) potentially starting, so for the remaining pairs we must choose from values not equal to δ_i, j_0, j, leaving four values, and at most two pairs. We therefore see that in order to have four rows potentially starting, two of them need to involve δ_i.

If (j_1, δ_i) and (j_2, δ_i) are potentially starting, with neither j_1, j_2 equal to δ_i (and hence also neither equal to j_0), then any remaining rows have to be chosen as distinct pairs from the remaining $(r+1) - 4 = 3$

indices, with at most one pair having repeated value. We thus obtain at most four rows, with four occurring only if $2a^i_j = a^i_{j_3} + a^i_{j_4} = c_i + 1$ for some j, j_3, $j_4 \neq \delta_i$, j_0, j_1, j_2. Moreover, we see that there must be exactly three values of j' with $a^i_{j'} < a^i_j$ in this case: if $a^i_{\delta_i} < a^i_j$, then these are δ_i, j_0, and exactly one of j_3, j_4, with necessarily j_1, j_2 and the other of j_3, j_4 having $a^i_{j'} > a^i_j$. If $a^i_{\delta_i} > a^i_j$, then $a^i_{j_0}$ must also be greater than a^i_j, so we similarly find exactly three values are smaller. Thus (7-2) must hold.

It remains to consider the case the (δ_i, δ_i) row is potentially starting, and the only thing left to prove is the description of case (3), where $a^i_{(\delta_i,\delta_i)} = c_i + 3$. Here, we must also have a j with $a^i_{(j,\delta_i)} = c_i + 1$, and if we have two additional rows appearing, these must come from two additional pairs nested around $a^i_{(j,\delta_i)}$, so since $a^i_j < a^i_{j_0} < a^i_{\delta_i}$ in this case, we obtain the desired statement.

The statement on rows ending is symmetric. □

We are now ready to deal with the two possible cases of 3-cycles in the next two corollaries.

Corollary 7.6. *Suppose that $\rho = 2$ and $r = 6$ and we are in the "first 3-cycle" situation of Proposition 4.18. Then, there exists an unimaginative multidegree ω' such that the $(j_0 - 1, j_0)$ row has a unique potentially present section in multidegree ω', whose support does not contain i_0 or i_1, and such that all the potentially present sections are linearly independent.*

Proof. Consider the default multidegree ω_{def}. If all (j_1, j_2) have connected potential support, we are done. With the notation of Proposition 4.18, the only rows that have a semicolumn adding to $d-2$ are j_0, $j_0 - 1$. From Proposition 5.10, the only row which could have disconnected potential support in some unimaginative multidegree is $(j_0 - 1, j_0)$. More specifically, the potential support of $(j_0 - 1, j_0)$ can be disconnected only if $a^{i_0}_{j_0-1} + a^{i_0}_{j_0} = c_{i_0} - 1$, $a^{i_1+1}_{j_0-1} + a^{i_1+1}_{j_0} = c_{i_1+1} + 1$, there is degree 2 in every column from i_0 to i_1 inclusive, and no δ_i equals $j_0 - 1$ or j_0 for any i between i_0 and i_1. It then follows in particular that $a^{i_0}_{(j_0+1,j_0+1)} \geq c_{i_0} + 2$ and $a^{i_1+1}_{(j_0+1,j_0+1)} \leq c_{i_1+1} - 2$, or equivalently, $a^{i_1}_{(j_0+1,j_0+1)} \leq c_{i_1}$. If the $(j_0 - 1, j_0)$ row has disconnected potential support, then we will use Lemma 7.5 to verify that we can move one degree 3 into either the i_0 or i_1 column to achieve connected potential support while maintaining the independence conclusion of Theorem 7.3. If the 3 was moved to the i_0 column, then the $(j_0 - 1, j_0)$ still cannot have any potential support at i_1. If the 3 was moved from the right, we still have $a^{i_0}_{j_0-1} + a^{i_0}_{j_0} = c_{i_0} - 1$, ruling out potential support at i_0. But if it was moved from the left, then this will decrease c_{i_0} by 1, and we will then have $a^i_{j_0-1} + a^i_{j_0} = c_i$ for $i_0 \leq i \leq i_1$, meaning that any potential support at i_0 would have to continue right to i_1, but we will still have $a^{i_1+1}_{j_0-1} + a^{i_1+1}_{j_0} = c_{i_1} + 1$, so there cannot be any potential support at i_1. A similar analysis holds if we moved the 3 to i_1, proving the desired result.

To prove that we can always move a 3 as desired, we first make some general observations regarding when we will be able to move degree 3 from the left or right onto i_0 or i_1. Recall that, from the assumption of having a 3-cycle, $\delta_{i_0} = \delta_{i_1} = j_0 + 1$. Since $a^{i_1}_{(j_0+1,j_0+1)} \leq c_{i_1}$, moving a degree 3 to i_1 from the right will always lead to at most 3 rows starting in the i_1 column, unless $2\bar{a}^{i_1}_3 = c_{i_1} + 1$, or equivalently,

$$5 - \gamma_{i_1-1} = 2\bar{\lambda}_{i_1-1,3}. \tag{7-4}$$

In addition, $a^{i_1}_{(j_0-1,j_0+1)} < c_{i_1}$, so the $(j_0 - 1, j_0 + 1)$ row will not be among the potentially present rows.

We next consider what happens if we move a degree 3 to i_0 from the left. This will decrease c_{i_0} by 1, so we have to rule out that in multidegree ω_{def} we have $2\bar{a}_3^{i_0} = c_{i_0}$, or equivalently,

$$6 - \gamma_{i_0 - 1} = 2\bar{\lambda}_{i_0-1,3}. \tag{7-5}$$

Additionally, if $a^{i_0}_{(j_0+1, j_0+1)} \geq c_{i_0} + 3$ in ω_{def}, then after moving the degree 3 to i_0, none of the other exceptional cases of Lemma 7.5 can occur, so as long as we do not have (7-5), we will have at most three rows with potential support starting at i_0. The only other possibility is that $a^{i_0}_{(j_0+1, j_0+1)} = c_{i_0} + 2$, which is equivalent to $2j_0 - \gamma_{i_0-1} = 2\lambda_{i_0-1, j_0+1}$; moreover, after moving a 3 from the left to i_0 we will have $a^{i_0}_{(j_0+1, j_0+1)} = c_{i_0} + 3$, so we could potentially be only in the third exceptional case in Lemma 7.5. Thus, the only case for concern is that $j_0 + 1 = 4$, so we simply need to check that in cases where we wish to move a 3 from the left, we never have

$$6 - \gamma_{i_0 - 1} = 2\lambda_{i_0-1,4}. \tag{7-6}$$

Finally, in either case after the move we will have $a^{i_0}_{(j_0, j_0+1)} = a^{i_0}_{(j_0+1, j_0+1)} - 1 \geq c_{i_0} + 2$, so the (j_0, j_0+1) row cannot be among the rows starting at i_0.

We now describe how to modify our default multidegree, depending on the location of i_0 and i_1. If we have $\gamma_{i_0} = \gamma_{i_1} = 1$, then we will move the next 3 from the right to column i_1, and we will obtain at most three rows with potential support starting in i_1: by the above observation, it suffices to rule out (7-4), but we have $5 - \gamma_{i_1 - 1} = 4$. To have equality we would need $\bar{\lambda}_{i_1-1,3} = 2$, which would imply $\bar{\lambda}^1_{i_1-1} + \bar{\lambda}^2_{i_1-1} \geq 8$, in which case we would not have had $\gamma_{i_1} = 1$ in ω_{def}.

Next, suppose $\gamma_{i_0} = \gamma_{i_1} = 2$, and we have $\bar{\lambda}^1_{i_1} + \bar{\lambda}^2_{i_1} < 6$. In this case, we will move the 3 to i_0 from the left, and $\gamma_{i_0-1} = 2$ in ω_{def}, so if either (7-6) or (7-5) is satisfied, we must have $\lambda_{i_0-1,3} \geq 2$. But this would force

$$\bar{\lambda}^1_{i_1} + \bar{\lambda}^2_{i_1} \geq \bar{\lambda}^1_{i_0-1} + \bar{\lambda}^2_{i_0-1} \geq 8,$$

contradicting the hypothesis for the case in question. We again conclude that there are at most 3 rows starting, and again the (j_0, j_0+1) row is not among them.

On the other hand, if $\gamma_{i_0} = \gamma_{i_1} = 2$, and $\bar{\lambda}^1_{i_1} + \bar{\lambda}^2_{i_1} \geq 6$, then we will move a 3 to i_1 from the right, and (7-4) is not satisfied for parity reasons, so we will have at most three new rows starting. Finally, if $\gamma_{i_0} = \gamma_{i_1} = 3$, neither (7-6) nor (7-5) can be satisfied for parity reasons, so we can move a 3 from the left to i_0, and have at most three starting rows.

The remaining cases are treated symmetrically, with rows starting replaced by rows ending. In each case, we see that the basic structure of the proof of Theorem 7.3 is preserved by our change of multidegree, so our linear independence is likewise preserved, yielding the desired statement. □

Corollary 7.7. *Assume that $\rho = 2$, $r = 6$ and we are in the "second 3-cycle" situation of Proposition 4.20. Suppose that in the default multidegree ω_{def}, we have the inequalities*

$$2a^{i_0}_{j_0-1} \leq c_{i_0} - 1, \quad \textit{and} \quad 2a^{i_1+1}_{j_0-1} \geq c_{i_1+1} + 1,$$

with exactly one of the two inequalities satisfied with equality. Then there exists an unimaginative multidegree ω' such that the (j_0-1, j_0-1) row does not have potentially present sections both left of i_0 and right of i_1 in multidegree ω', and such that all the potentially present sections are linearly independent.

Proof. The only row that has a semicolumn adding to $d-2$ is j_0-1 and it has two of them. From Proposition 5.10, the only row which could have disconnected potential support in some unimaginative multidegree is (j_0-1, j_0-1).

Suppose that in multidegree ω_{def}, we have

$$2a^{i_0}_{j_0-1} = c_{i_0} - 1, \quad \text{but} \quad 2a^{i_1+1}_{j_0-1} > c_{i_1+1} + 1.$$

We will show that we can always move a 3 from the left to a genus-1 column on or right of i_0, while preserving linear independence. This will eliminate potential support in the (j_0-1, j_0-1) row left of i_0, as desired. From the definition of $\lambda_{i,j}$ in Definition 5.1, this condition can be written as

$$2(j_0-1) + 1 - \gamma_{i_0-1} = 2\lambda_{i_0-1, j_0-1},$$

so in particular γ_{i_0-1} must be odd.

Case $\gamma_{i_0-1} = 1$. Then $j_0 - 1 = \lambda_{i_0-1, j_0-1}$. From the definition of default multidegree $\bar\lambda^1_{i_0-1} + \bar\lambda^2_{i_0-1} < 5$, which forces $j_0 - 1 = 1$, so $\bar\lambda_{i_0,1} = \lambda_{i_0-1,1} = 1$.

First, if i_1 is the genus-1 column immediately following i_0, we observe that if we move the first 3 to i_0, considering only the inequalities at i_0, there can be at most three rows with potential support starting at i_0: $(1, 2)$, $(2, 2)$ and $(0, j)$ for a unique $j > 2$: For the row (j, j') to be present, we need

$$\lambda_{i_0-1,j} + \lambda_{i_0-1,j'} < j + j' \leq 1 + \lambda_{i_0,j} + \lambda_{i_0,j'}.$$

From $j_0 - 1 = 1$, $\lambda_{i_0-1,1} = 1$ and the 3-cycle situation, $\lambda_{i_0,1} = 0$, $\lambda_{i_0-1,2} = 1$, $\lambda_{i_0,2} = 2$. So $(1, 2)$, $(2, 2)$ are potentially present and one pair $(0, j)$, $j > 2$. But in this case the actual potential support of $(1, 2)$ is connected and supported strictly to the right of i_1. Thus, there are in fact at most two rows with potential support starting at i_0, and neither of them involves the exceptional row (specifically, $j = 1$). So moving the first 3 to i_0 we will still be able to eliminate potentially present sections from left to right as before.

Next, suppose that i_1 is not the genus-1 column immediately following i_0, and denote this column by i. Suppose also that there is no degree-3 column between i_0 and i_1, so that in particular $\bar\lambda^1_i + \bar\lambda^2_i \leq 4$. We observe that we must also have $\bar\lambda_{i,0} = 1$, since we have $\bar\lambda_{i_0,1} = \bar\lambda_{i_0,2} = 1$, and we must have $\bar\lambda_{i_1-1,2} = \bar\lambda_{i_1-1,3} \geq 1$. So the only way we can avoid having a column of degree 3 before i_1 is if also $\bar\lambda_{i_1-1,0} = 1$. We can then apply Proposition 7.4 to move the first 3 to column i, and we will still obtain linear independence.

Finally, if we have a column of degree 3 between i_0 and i_1, say in column i, so that $\bar\lambda^1_i + \bar\lambda^2_i = 5$, then we claim that if we move the first 3 from the left to i_0, we will have at most two potentially present sections ending in column i, and at most two potentially present sections supported in the first column with $\bar\lambda^1_{i'} + \bar\lambda^2_{i'} = 6$. This will prove the desired statement, since we can then eliminate the potentially present sections starting from i' and moving both left and right from there. Checking the possible inequalities in column i', moving the 3 from the left to i_0 won't affect anything, so the argument for Theorem 7.3 implies *a priori* that there are at most three rows satisfying the inequalities at i' for potentially present sections to be supported there. We will check that there is always one such row which satisfies the inequalities at i', but does not in fact have potential support there. Because we have a 3 between i_0 and i_1, we must have $2a^{i_1+1}_{j_0-1} = c_{i_1+1} + 2$. If $i' < i_1$, the row in question is $(1, 1) = (j_0 - 1, j_0 - 1)$: indeed, in this situation we will have $a^{i''}_{(j_0-1,j_0-1)} = c_{i''}$ for all i'' with $i < i'' \le i_1$, so $(j_0 - 1, j_0 - 1)$ does satisfy the necessary inequalities at i', but its actual potential support (after moving the 3 to i_0) is strictly to the right of i_1. On the other hand, if $i' > i_1$, the row in question will be $(1, 2) = (j_0 - 1, j_0)$:

$$\text{we have}\quad a^{i_1+1}_{j_0-1} + a^{i_1+1}_{j_0} \le c_{i_1+1}, \quad \text{so}\quad a^{i_1}_{j_0-1} + a^{i_1}_{j_0} < c_{i_1},$$

and because the potential support is connected, it must be strictly left of i_1. However, we claim that we must have $a^{i_1+1}_{j_0-1} + a^{i_1+1}_{j_0} = c_{i_1+1}$, and that this must extend through the column i', so that the inequalities for potential support are satisfied at i'. Indeed, the only way this could fail is if $\delta_{i''} = j_0$ for some i'' with $i_0 < i'' < i'$. But we know that $\lambda_{i_0-1,j_0-1} = \lambda_{i_0-1,j_0} = 1$. So if $\delta_{i''} = j_0$ anywhere after i_0, it increases $\bar{\lambda}^1_{i''} + \bar{\lambda}^2_{i''}$ to at least 5. Thus, this could only happen for $i'' < i'$ if $i'' = i$, which then forces us to have $\bar{\lambda}^1_{i_0} = 3$ and $\bar{\lambda}^2_{i_0} = 1$. However, in this case, because we cannot have a gap between the $j_0 - 1$ and $j_0 + 1$ column at i_1, this would force us to also increase $\bar{\lambda}^1_{i''}$ to 4 before i_1, which violates our hypothesis that $i' > i_1$. Thus, in either situation we have shown that the column i' has at most two potentially present sections supported on it, and it remains to check that the column i has at most two potentially present sections ending on it. But we either have $\bar{\lambda}^1_{i-1} = 4$ and $\bar{\lambda}^2_{i-1} = 0$ or $\bar{\lambda}^1_{i-1} = 3$ and $\bar{\lambda}^2_{i-1} = 1$, and one can calculate directly that because we cannot have $\delta_i = 0$ or 4 in the second case, $\delta_i = 1$ in either case (recalling that by column i we have had a swap between rows 1 and 2), or $\delta_i = 3$ in the first case, the only rows with potential support ending in column i are $(1, 2)$ and $(0, j)$ for a unique value of j, yielding the desired statement.

Case $\gamma_{i_0-1} = 3$. Then either $j_0 - 1 = 2$ and $\lambda_{i_0-1,j_0-1} = 1$, or $j_0 - 1 = 3$ and $\lambda_{i_0-1,j_0-1} = 2$. First, suppose that $(j_0 - 1, j_0)$ has potential support strictly to the right of i_1, or equivalently, that there are no columns between i_0 and i_1 having degree 3, or with $\delta_i = j_0 - 1$ or j_0. In this case, if we move a 3 from the left to i_0, by Lemma 7.5 at most four rows satisfy the inequalities at i_0 to have potentially starting sections at i_0, and we see that these include $(j_0 - 1, j_0)$. But $(j_0 - 1, j_0)$ does not actually have potential support at i_0, so in this case we have at most three rows starting at i_0, and none of them involve the exceptional row (specifically, $j_0 - 1$), so we can eliminate this central block just as in Theorem 7.3, and we conclude we still have linear independence.

Now, the possibility that we have $\delta_i = j_0 - 1$ in between i_0 and i_1 is ruled out by the inequality $2a^{i_1+1}_{j_0-1} > c_{i_1+1} + 1$. If there is a column with $\delta_i = j_0$, but no column having degree 3 between i_0 and i_1, we will move the third degree-3 from the left to i_1, and the $(j_0 + 1, j_0 + 1) = (\delta_{i_1}, \delta_{i_1})$ row is supported strictly to the right of i_1. In addition (7-2) is ruled out by parity reasons, so by Lemma 7.5 we have at most three rows starting at i_1, and we also see that $(j_0 - 1, j_0 + 1)$ is not among them, as it will have potential support strictly to the right of i_1. Thus, no row involving j_0-1 (the exceptional row) has potential support starting at i_1, and in this case we can eliminate all potentially present sections just as in Theorem 7.3.

Next, suppose there is some column with degree 3 between i_0 and i_1, but no column with $\delta_i = j_0$. In this case, we will move the fourth 3 to the first column i with $\bar{\lambda}^2_i + \bar{\lambda}^3_i = 9$, and the third 3 to i_0. If $\bar{\lambda}^2_{i_1} + \bar{\lambda}^3_{i_1} < 9$, then according to Proposition 7.4, moving the fourth 3 doesn't disrupt linear independence, and then we are in exactly the same situation as the first case considered above, with (j_0-1, j_0) having potential support strictly to the right of i_1. On the other hand, if $\bar{\lambda}^2_{i_1} + \bar{\lambda}^3_{i_1} = 9$, we will still maintain linear independence, but for different reasons: we claim that will have at most three rows ending in the i-th column, no row ending in the first column i' with $\bar{\lambda}^2_{i'} + \bar{\lambda}^3_{i'} = 8$, and only two rows ending in the first column i'' with $\bar{\lambda}^2_{i''} + \bar{\lambda}^3_{i''} = 10$. Thus, we will be able to eliminate potentially present sections from the right, treating the columns from i' to i as a block to which to apply Lemma 6.3(3), and we will in this way eliminate all potentially present sections supported on either side of i_0. This leaves at most one potentially present section, which can then be eliminated. Thus, it suffices to prove the above claim. By the argument for Proposition 7.4, we have no potentially present section supported only in the i-th column, and at most three continuing from the previous column. So, there are at most three ending in the i-th column, as claimed. The fact that there are no rows ending in the i'-th column is immediate from Corollary 5.8 and Lemma 5.12. Finally, we know from the proof of Theorem 7.3 that there at most three rows satisfying the inequalities in column i'' to have potential support ending there. Moreover, we see that (j_0-1, j_0) is necessarily one of them. Indeed, since we have one column with degree 3 and none with $\delta_i = j_0 - 1$ or j_0 between i_0 and i_1, we see that $a^{i_1+1}_{(j_0-1,j_0)} = c_{i_1+1}$ even after changing the multidegree. But after i_1, any column with $\delta_i = j_0 - 1$ or j_0 will increase $\bar{\lambda}^2_i + \bar{\lambda}^3_i$, so this cannot occur strictly between i_1 and i'', and we conclude that $a^{i''}_{(j_0-1,j_0)} = c_{i''}$ as well. Since column i'' has degree 3, this means that (j_0-1, j_0) satisfies the inequalities to have potential support ending at i''. But again using that the fourth 3 is still left of i_1, the actual potential support of (j_0-1, j_0) is contained to the left of i_1, so we conclude that column i'' has at most two rows with potential support ending there, completing the proof of the claim.

It remains to analyze the possibility that we have a column of degree 3 and a column with $\delta_i = j_0$ in between i_0 and i_1. Recall that we have either

$$j_0 - 1 = 2 \quad \text{and} \quad \lambda_{i_0-1, j_0-1} = 1, \qquad \text{or} \qquad j_0 - 1 = 3 \quad \text{and} \quad \lambda_{i_0-1, j_0-1} = 2.$$

We first claim that in the latter case, we cannot have $\delta_i = j_0$ in between i_0 and i_1 without forcing there to be two columns of degree 3 in between, or equivalently, forcing $\bar{\lambda}^2_{i_1} + \bar{\lambda}^3_{i_1} \geq 10$. Indeed, since we cannot have a gap between $a^{i_0}_{j_0-1}$ and $a^{i_0}_{j_0}$ for the swap, we must have $\bar{\lambda}^2_{i_0} \geq 5$, and then for the same reason at i_1 we must have $\bar{\lambda}^2_{i_1} \geq 6$. But having some $\delta_i = j_0$ also requires $\bar{\lambda}^3_i \geq 4$, so we conclude that we would necessarily have $\bar{\lambda}^2_{i_1} + \bar{\lambda}^3_{i_1} \geq 10$, as claimed. Thus, it suffices to treat the situation that $\lambda_{i_0-1, j_0-1} = 1$. In this situation, we have $\bar{\lambda}^2_{i_1} + \bar{\lambda}^3_{i_1} \leq 6$, and we will move the third 3 to column i_0 and the fourth 3 to column i_1. We claim that we will have at most two rows with potentially present sections ending in i_1, and neither involves the exceptional row (specifically, $j_0 - 1$, which is 2). Thus, we will be able to eliminate all potentially present sections from the left and from the right of i_0, and finally eliminate the at most one potentially present section supported only at i_0. To verify the claim, we see that we necessarily have

$$5 \le \bar{\lambda}^1_{i_1} \le 7, \quad \bar{\lambda}^2_{i_1} = 3, \quad \text{and} \quad 1 \le \bar{\lambda}^3_{i_1} \le 3.$$

We compute that the only rows satisfying the inequalities to potentially end at i_1 are $(3, 4)$, $(0, 6)$, $(1, 5)$, $(1, 6)$ and $(3, 5)$, but by the uniqueness part of Corollary 5.9, we see that the only way we can have three of these occurring at once is if we have $(3, 4)$, $(0, 6)$ and $(1, 5)$. However, we also have that $(3, 4)$ can only end if $\bar{\lambda}^3_{i_1} \le 2$, $(0, 6)$ can only end if $\bar{\lambda}^1_{i_1} \le 6$, and $(1, 5)$ can only end if one of the preceding two inequalities is strict. But together these imply that $\bar{\lambda}^1_{i_1} + \bar{\lambda}^3_{i_1} \le 7$, meaning that we cannot have all the rows ending at i_1 under our hypothesis that the fourth 3 comes before i_1.

This concludes the case $\gamma_{i_0-1} = 3$.

Case $\gamma_{i_0-1} = 5$. In this case $j_0 - 1 = \lambda_{i_0-1, j_0-1} + 2$. From the definition of default multidegree, $\bar{\lambda}^2_{i_0-1} + \bar{\lambda}^3_{i_0-1} \ge 10$, so $j_0 - 1 \ge 4$. But, to allow for the double swap (there is a $j_0 + 1$ row), $j_0 - 1 \le 4$, so $j_0 - 1 = 4$. With an argument as in Lemma 7.5, if we move the fifth 3 to i_0, even if we obtain two rows involving $\delta_{i_0} = 5$ with potential support ending at i_0, there can be at most one more (necessarily of the form $(j, 6)$ for some j). Moreover, the $(4, 5)$ row is not one of these, as it will have potential support starting, not ending, at i_0. We can therefore still eliminate the block spanning from the first column with $\bar{\lambda}^2_i + \bar{\lambda}^3_i = 9$ to column i_0 just as before.

The case that $2a^{i_1+1}_{j_0-1} = c_{i_1+1} + 1$ but $2a^{i_0}_{j_0-1} < c_{i_0} - 1$ is handled completely symmetrically, completing the proof. □

8. An analysis of the degenerate case

To conclude the proof of the main theorem, we need multidegrees such that on the one hand, the potentially present sections are still linearly independent, and on the other hand, tensors coming from any exact linked linear series generate at least $\binom{r+2}{2} = 28$ linearly independent combinations of the potentially present sections. The key point is that even though there are cases where some row may not have any potentially present section occurring in the linked linear series in the chosen multidegree, in those cases there is more than one combination of sections from some other row. We first look at the behavior of mixed sections under tensor product

Lemma 8.1. *Suppose s, s' are mixed sections of multidegrees $\mathrm{md}_d(w)$ and $\mathrm{md}_d(w')$, and let $\mathrm{md}_{2d}(w'')$ be another multidegree. Then $f_{w+w',w''}(s \otimes s')$ lies in the potential ambient space in multidegree $\mathrm{md}(w'')$.*

Proof. By definition of mixed sections as sums, it suffices to treat the case that s is obtained purely from gluing together s^i_j for fixed j, and s' is obtained from gluing together $s^i_{j'}$ for fixed j'. But in this case the result is clear, since $f_{w+w',w''}(s \otimes s')$ must be a combination of potentially present sections from the (j, j') row. □

The following lemma is convenient for cutting down the number of possibilities to consider.

Lemma 8.2. *Let s, s' be mixed sections of multidegrees* $\mathrm{md}(w)$ *and* $\mathrm{md}(w')$ *and types* $(\vec{S}, \vec{j})$ *and* $(\vec{S}', \vec{j}')$ *respectively. Suppose that for some i with $1 < i < N$, we have*

$$\ell_1 \neq \ell_2 \quad \textit{and} \quad \ell'_1 \neq \ell'_2 \quad \textit{such that } i \in S_{\ell_1} \cap S_{\ell_2} \cap S'_{\ell'_1} \cap S'_{\ell'_2}.$$

Then for any unimaginative w'', the map $f_{w+w',w''}$ vanishes identically on Z_i.

If further either $\{j_{\ell_1}, j_{\ell_2}\} = \{j'_{\ell'_1}, j'_{\ell'_2}\}$ or $\{j_{\ell_1}, j_{\ell_2}\} \cap \{j'_{\ell'_1}, j'_{\ell'_2}\} = \varnothing$, then the same conclusion holds when $i = 1$ or $i = N$.

Proof. First consider the case $1 < i < N$, and write $w = (c_2, \dots, c_N)$ and $w' = (c'_2, \dots, c'_N)$. The hypotheses mean that w allows for support of both s^i_j and $s^i_{j'}$ for some distinct j, j', so $a^i_j, a^i_{j'} \geq c_i$ and $b^i_j, b^i_{j'} \geq d - c_{i+1}$. Without loss of generality, suppose $a^i_j < a^i_{j'}$. Then, either $a^i_j + b^i_j < d$ or $a^i_{j'} + b^i_{j'} < d$. If $b^i_j > b^i_{j'}$, then either $c_{i+1} \geq d - b^i_{j'} > d - b^i_j > a^i_j \geq c_i$ or $c_{i+1} \geq d - b^i_{j'} > a^i_{j'} > a^i_j \geq c_i$, so in either case we have $c_{i+1} \geq c_i + 2$. On the other hand, if $b^i_j < b^i_{j'}$, then $c_{i+1} \geq d - b^i_j > d - b^i_{j'} \geq a^i_{j'} > a^i_j \geq c_i$, so again $c_{i+1} \geq c_i + 2$. The same argument holds for w', so we conclude that $c_{i+1} + c'_{i+1} \geq c_i + c'_i + 4$, which implies that $f_{w+w',w''}$ vanishes on Z_i, since if we write $w'' = (c''_2, \dots, c''_N)$, the unimaginative hypothesis means that $c''_{i+1} \leq c''_i + 3$.

Next, if $i = 1$, the unimaginative hypothesis means that c_2 is equal to 2 or 3. It follows (see the proof of Theorem 7.3) that only the rows (0, 0), (0, 1), (1, 1) and (0, 2) can have potential support in the column, with not both (0, 0) and (1, 1) occurring. If $f_{w+w',w''}$ is nonzero on Z_1, then $f_{w+w',w''}(s \otimes s')$ must have $(j_{\ell_u}, j'_{\ell'_v})$ parts with potential support at $i = 1$ for $u = 1, 2$ and $v = 1, 2$, and this isn't possible if either $\{j_{\ell_1}, j_{\ell_2}\} = \{j'_{\ell'_1}, j'_{\ell'_2}\}$ or $\{j_{\ell_1}, j_{\ell_2}\} \cap \{j'_{\ell'_1}, j'_{\ell'_2}\} = \varnothing$. The case $i = N$ is symmetric. □

The cases of mixed sections appearing after each type of swap will require individual analysis. We treat the case of a single swap in the next proposition but we first give an example:

Example 8.3. Tables 2a and 2b together show the table T obtained from the tensor square of the limit linear series considered in Example 4.25, which has $r = 6$, $g = 22$, and $d = 25$.

We have highlighted the potentially present sections; note that the (2, 2) row contains two (the first in Table 2a, the second in Table 2b), while the rest all have a unique one. These two potentially present sections are thus treated separately in Theorem 7.3; the first appears as part of a block in the fifth and sixth columns which is eliminated using rule (iii) of Lemma 6.3, while the second occurs as the only new potentially present sections in the twelfth column, which is part of another block, extending from the seventh column to the sixteenth column, which is again eliminated using rule (iii), after all other potentially present sections have been eliminated on both the left and right. The only other block that requires rule (iii) contains the seventeenth and eighteenth columns, and is eliminated after the potentially present sections appearing to the right have all been dropped. Following the proof of Theorem 7.3, we see that we can eliminate all sections outside the aforementioned three blocks going inward from both the left and right ends, using only iterated applications of rule (i).

Observe that in the default multidegree, the (unique) potentially present section in row (2, 3) extends from the seventh column to the eleventh column. This means that if s'_3 and s''_3 have the smallest possible

col =	1		2		3		4		5		6		7		8		9		10		11	
(0, 0)	0	50	0	48	2	46	4	44	6	42	8	40	10	38	12	38	12	36	14	34	16	32
(0, 1)	1	48	2	47	3	45	5	43	7	41	9	39	11	37	13	36	14	34	16	33	17	31
(0, 2)	2	47	3	45	5	44	6	42	8	40	10	38	12	36	14	35	15	32	18	30	20	28
(1, 1)	2	46	4	46	4	44	6	42	8	40	10	38	12	36	14	34	16	32	18	32	18	30
(0, 3)	3	46	4	44	6	42	8	41	9	39	11	37	13	35	15	34	16	33	17	31	19	30
(1, 2)	3	45	5	44	6	43	7	41	9	39	11	37	13	35	15	33	17	30	20	29	21	27
(0, 4)	4	45	5	43	7	41	9	39	11	38	12	36	14	34	16	33	17	31	19	29	21	27
(1, 3)	4	44	6	43	7	41	9	40	10	38	12	36	14	34	16	32	18	31	19	30	20	29
(2, 2)	4	44	6	42	8	42	8	40	10	38	12	36	14	34	16	32	18	28	22	26	24	24
(0, 5)	5	44	6	42	8	40	10	38	12	36	14	35	15	33	17	32	18	30	20	28	22	26
(1, 4)	5	43	7	42	8	40	10	38	12	37	13	35	15	33	17	31	19	29	21	28	22	26
(2, 3)	5	43	7	41	9	40	10	39	11	37	13	35	15	33	17	31	19	29	21	27	23	26
(0, 6)	6	43	7	41	9	39	11	37	13	35	15	33	17	32	18	31	19	29	21	27	23	25
(1, 5)	6	42	8	41	9	39	11	37	13	35	15	34	16	32	18	30	20	28	22	27	23	25
(2, 4)	6	42	8	40	10	39	11	37	13	36	14	34	16	32	18	30	20	27	23	25	25	23
(3, 3)	6	42	8	40	10	38	12	38	12	36	14	34	16	32	18	30	20	30	20	28	22	28
(1, 6)	7	41	9	40	10	38	12	36	14	34	16	32	18	31	19	29	21	27	23	26	24	24
(2, 5)	7	41	9	39	11	38	12	36	14	34	16	33	17	31	19	29	21	26	24	24	26	22
(3, 4)	7	41	9	39	11	37	13	36	14	35	15	33	17	31	19	29	21	28	22	26	24	25
(2, 6)	8	40	10	38	12	37	13	35	15	33	17	31	19	30	20	28	22	25	25	23	27	21
(3, 5)	8	40	10	38	12	36	14	35	15	33	17	32	18	30	20	28	22	27	23	25	25	24
(4, 4)	8	40	10	38	12	36	14	34	16	34	16	32	18	30	20	28	22	26	24	24	26	22
(3, 6)	9	39	11	37	13	35	15	34	16	32	18	30	20	29	21	27	23	26	24	24	26	23
(4, 5)	9	39	11	37	13	35	15	33	17	32	18	31	19	29	21	27	23	25	25	23	27	21
(4, 6)	10	38	12	36	14	34	16	32	18	31	19	29	21	28	22	26	24	24	26	22	28	20
(5, 5)	10	38	12	36	14	34	16	32	18	30	20	30	20	28	22	26	24	24	26	22	28	20
(5, 6)	11	37	13	35	15	33	17	31	19	29	21	28	22	27	23	25	25	23	27	21	29	19
(6, 6)	12	36	14	34	16	32	18	30	20	28	22	26	24	26	24	24	26	22	28	20	30	18
(6, 6)	32	16	34	14	36	12	38	12	38	10	40	8	42	6	44	4	46	2	48	0	50	0
		47	3	45	5	43	7	41	9	38	12	36	14	33	17	31	19	29	21	27	23	25

Table 2a. This is the left side of the table T obtained from the tensor square of the limit linear series considered in Example 4.25, which has $r = 6$, $g = 22$, and $d = 25$. The right side is shown in Table 2b. We have also included the w corresponding to the default multidegree ω_{def} at the bottom of the table, and include not only the values c_i for $i = 2, \dots, 22$, but also $2d - c_i$ in the preceding subcolumns.

portions coming from the $j = 3$ row, so that s_3' only has nonzero s_3^i parts for $i \geq 8$ and s_3'' for $i \leq 10$, then the potentially present section for the (2, 3) row cannot come from either $s_2 \otimes s_3'$ or $s_2 \otimes s_3''$. This means that these sections (or more precisely, their images in multidegree ω_{def}) are forced to yield potentially present sections from the (2, 2) row, with $s_2 \otimes s_3'$ necessarily yielding the one supported from columns 5 through 7, and $s_2 \otimes s_3''$ necessarily yielding the one supported in column 12. Thus, we explicitly see the lack of a (2, 3) section being offset by the inclusion of two independent (2, 2) sections.

col =	12		13		14		15		16		17		18		19		20		21		22	
(0, 0)	18	30	20	28	22	26	24	24	26	24	26	22	28	20	30	18	32	16	34	14	36	12
(0, 1)	19	29	21	27	23	25	25	23	27	22	28	21	29	19	31	17	33	15	35	13	37	11
(0, 2)	22	27	23	25	25	23	27	21	29	20	30	18	32	16	34	15	35	13	37	11	39	9
(1, 1)	20	28	22	26	24	24	26	22	28	20	30	20	30	18	32	16	34	14	36	12	38	10
(0, 3)	20	28	22	26	24	24	26	22	28	21	29	19	31	18	32	16	34	14	36	12	38	10
(1, 2)	23	26	24	24	26	22	28	20	30	18	32	17	33	15	35	14	36	12	38	10	40	8
(0, 4)	23	25	25	24	26	22	28	20	30	19	31	17	33	15	35	13	37	12	38	10	40	8
(1, 3)	21	27	23	25	25	23	27	21	29	19	31	18	32	17	33	15	35	13	37	11	39	9
(2, 2)	26	24	26	22	28	20	30	18	32	16	34	14	36	12	38	12	38	10	40	8	42	6
(0, 5)	24	24	26	22	28	21	29	19	31	18	32	16	34	14	36	12	38	10	40	9	41	7
(1, 4)	24	24	26	23	27	21	29	19	31	17	33	16	34	14	36	12	38	11	39	9	41	7
(2, 3)	24	25	25	23	27	21	29	19	31	17	33	15	35	14	36	13	37	11	39	9	41	7
(0, 6)	25	23	27	21	29	19	31	18	32	17	33	15	35	13	37	11	39	9	41	7	43	6
(1, 5)	25	23	27	21	29	20	30	18	32	16	34	15	35	13	37	11	39	9	41	8	42	6
(2, 4)	27	22	28	21	29	19	31	17	33	15	35	13	37	11	39	10	40	9	41	7	43	5
(3, 3)	22	26	24	24	26	22	28	20	30	18	32	16	34	16	34	14	36	12	38	10	40	8
(1, 6)	26	22	28	20	30	18	32	17	33	15	35	14	36	12	38	10	40	8	42	6	44	5
(2, 5)	28	21	29	19	31	18	32	16	34	14	36	12	38	10	40	9	41	7	43	6	44	4
(3, 4)	25	23	27	22	28	20	30	18	32	16	34	14	36	13	37	11	39	10	40	8	42	6
(2, 6)	29	20	30	18	32	16	34	15	35	13	37	11	39	9	41	8	42	6	44	4	46	3
(3, 5)	26	22	28	20	30	19	31	17	33	15	35	13	37	12	38	10	40	8	42	7	43	5
(4, 4)	28	20	30	20	30	18	32	16	34	14	36	12	38	10	40	8	42	8	42	6	44	4
(3, 6)	27	21	29	19	31	17	33	16	34	14	36	12	38	11	39	9	41	7	43	5	45	4
(4, 5)	29	19	31	18	32	17	33	15	35	13	37	11	39	9	41	7	43	6	44	5	45	3
(4, 6)	30	18	32	17	33	15	35	14	36	12	38	10	40	8	42	6	44	5	45	3	47	2
(5, 5)	30	18	32	16	34	16	34	14	36	12	38	10	40	8	42	6	44	4	46	4	46	2
(5, 6)	31	17	33	15	35	14	36	13	37	11	39	9	41	7	43	5	45	3	47	2	48	1
(6, 6)	32	16	34	14	36	12	38	12	38	10	40	8	42	6	44	4	46	2	48	0	50	0
	25	23	27	21	29	19	31	17	33	14	36	12	38	9	41	7	43	5	45	3	47	

Table 2b. This is the right side of the table T obtained from the tensor square of the limit linear series considered in Example 4.25, which has $r = 6$, $g = 22$, and $d = 25$. The left side is shown in Table 2a. We have also included the w corresponding to the default multidegree ω_{def} at the bottom of the table, and include not only the values c_i for $i = 2, \ldots, 22$, but also $2d - c_i$ in the preceding subcolumns.

Proposition 8.4. *Suppose a limit linear series contains precisely one swap, occurring between the rows j_0, $j_0 - 1$ in column i_0. With notation as in Proposition 4.13, for any multidegree ω, the tensors pairs of the s_j for $j \neq j_0$, and s'_{j_0}, s''_{j_0} contain $\binom{r+2}{2}$ independent linear combinations of the potentially present sections.*

Proof. If $j \leq j'$ are both different from j_0, then s_j and $s_{j'}$ are in the linked linear series and contribute an $s_{(j,j'),i}$. This gives rise to $\binom{r}{2}$ potentially present sections, necessarily independent because they are supported in distinct rows. If $j \neq j_0$, $j_0 - 1$, there are three global sections $s_j \otimes s'_{j_0}$, $s_j \otimes s''_{j_0}$ and $s_j \otimes s_{j_0-1}$, each of which has nonzero image in multidegree ω. We claim that these three images must contain at least two distinct linear combinations of the $s_{(j,j_0),i}$ and $s_{(j,j_0-1),i}$. If $s_j \otimes s'_{j_0}$ has support in any columns greater than or equal to i_0, this necessarily includes a nonzero combination of the $s_{(j,j_0),i}$, which is distinct from the image of $s_j \otimes s_{j_0-1}$, and we are done. The same holds if $s_j \otimes s''_{j_0}$ has support in any columns less than or equal to i_0. The final case is that $s_j \otimes s'_{j_0}$ has support only in columns strictly less than i_0, and $s_j \otimes s''_{j_0}$ has support only in columns strictly greater than i_0. In this case, both may be linear combinations of the $s_{(j,j_0-1),i}$, but since their support is disjoint, they must be two distinct combinations, as desired.

Thus, we have produced $\binom{r}{2} + 2(r-1) = \binom{r+2}{2} - 3$ independent combinations of potentially present sections, supported among the rows (j, j') with $j \neq j_0 - 1, j_0$. Finally, we consider the tensors of $s_{j_0-1}, s'_{j_0}, s''_{j_0}$, and claim we obtain three distinct linear combinations, necessarily supported among the rows (j_0-1, j_0-1), (j_0-1, j_0), (j_0, j_0). Consider the images of $s'_{j_0} \otimes s'_{j_0}$, $s'_{j_0} \otimes s''_{j_0}$, and $s''_{j_0} \otimes s''_{j_0}$. If any of their images contain any portion of the (j_0, j_0) row, then considering $s_{j_0-1} \otimes s_{j_0-1}$, $s_{j_0-1} \otimes s'_{j_0}$, $s_{j_0-1} \otimes s'_{j_0}$. The same argument as above shows we obtain two distinct combinations of type (j_0-1, j_0-1) and/or (j_0-1, j_0), so we are done. But the only alternative is that the first three tensors come from the (j_0-1, j_0-1), (j_0-1, j_0) and (j_0-1, j_0-1) rows respectively, with the first and last having disjoint support. Thus, in this case these three are all linearly independent, and we again obtain the desired conclusion. □

When $\rho = 2$, there can be up to two swaps (Proposition 5.3) occurring on distinct columns $i_0 < i_1$ corresponding to genus 1 components. We will show that in each of the four possible configurations of the two swaps there are enough independent linear combinations of the potentially present sections (see Propositions 8.5, 8.9, 8.10, 8.12).

We find it convenient to introduce shorthand notation as follows: we will write for instance

$$s'_{j_0} \otimes s''_{j_0+1} = (j_0-1, j_0+1)_{\mathrm{L}} + (j_0-1, j_0)_{\mathrm{R}} + (j_0, j_0+1)$$

to indicate that the image of $s'_{j_0} \otimes s''_{j_0+1}$ in the relevant multidegree is a combination of potentially present sections from the (j_0-1, j_0+1), (j_0-1, j_0) and (j_0, j_0+1) rows, where the first is supported strictly left of i_0, the second strictly right of i_1 and the third has no restrictions on its support. We will also use subscripts C to denote support strictly between i_0 and i_1, LC to denote support strictly left of i_1 and CR to denote support strictly right of i_0.

Propositions 8.6, 8.7, 8.8 help us control the potential support of mixed sections using a left-weighted X_0.

Proposition 8.5. *Suppose that the limit linear series contains precisely two swaps, and both occur in the same pair of rows, say j_0, j_0-1 in columns i_0, i_1 ("repeated swap", see Proposition 4.16). Then for any unimaginative multidegree ω, the images in multidegree ω of the tensors of pairs of the s_j for $j \neq j_0, j_0-1$, and $s'_{j_0-1}, s''_{j_0-1}, s'_{j_0}, s''_{j_0}$ contain $\binom{r+2}{2}$ independent linear combinations of the potentially present sections.*

Proof. Just as in the proof of Proposition 8.4, for $j, j' \neq j_0, j_0-1$, the linked linear series contains s_j and $s_{j'}$, so the image of $s_j \otimes s_{j'}$ always gives a potentially present section from row (j, j').

Now consider $j \neq j_0, j_0-1$; we claim that $s_j \otimes s'_{j_0-1}, s_j \otimes s''_{j_0-1}, s_j \otimes s'_{j_0}, s_j \otimes s''_{j_0}$ cannot all coincide. Hence they have a two-dimensional span. Indeed, if $s_j \otimes s''_{j_0-1}$ coincides with $s_j \otimes s'_{j_0}$, they must be of the form $(j, j_0-1)_{\mathrm{L}} + (j, j_0)_{\mathrm{R}}$. But the former cannot occur in $s_j \otimes s''_{j_0}$, and the latter cannot occur in $s_j \otimes s'_{j_0-1}$, so one of the two will provide a second section.

It remains to show that we have at least three independent sections among all tensors of the s'_{j_0-1}, s''_{j_0-1}, s'_{j_0}, s''_{j_0}. We first consider the tensor squares of each of the four sections. According to Lemma 8.2, these can only contain types (j_0-1, j_0-1) and (j_0, j_0), with no type (j_0-1, j_0) appearing. Now, the possible (j_0, j_0) parts of $s'^{\otimes 2}_{j_0-1}$ and $s''^{\otimes 2}_{j_0-1}$ are disjoint, so we conclude that either these two are distinct, or they are of pure type (j_0-1, j_0-1). Similarly, the sections $s'^{\otimes 2}_{j_0}$ and $s''^{\otimes 2}_{j_0}$ are either distinct or of pure type (j_0, j_0). Thus, it suffices to show that we cannot have all of our tensors in the span of a single pair of sections, each of pure type (j_0-1, j_0-1) or (j_0, j_0). Now, $s'_{j_0} \otimes s''_{j_0}$ cannot have a (j_0-1, j_0-1) part, and $s'_{j_0-1} \otimes s''_{j_0-1}$ cannot have a (j_0, j_0) part, so the only possibility to consider is that one of our sections is purely of type (j_0-1, j_0-1), and the other is purely of type (j_0, j_0).

If the (j_0-1, j_0-1) part occurs in $s''_{j_0-1} \otimes s'_{j_0}$, it must be supported strictly to the left of i_0. Then $s''_{j_0-1} \otimes s''_{j_0}$ cannot have a (j_0-1, j_0-1) part, so must be of type (j_0, j_0), and the support must be strictly to the right of i_1. On the other hand, if the (j_0, j_0) part occurs in $s''_{j_0-1} \otimes s'_{j_0}$, it must again be supported strictly to the right of i_1, and then $s'_{j_0-1} \otimes s'_{j_0}$ cannot have a (j_0, j_0) part, so must be of type (j_0-1, j_0-1), again supported to the left of i_0. But in either case, $s'_{j_0-1} \otimes s''_{j_0}$ cannot be a linear combination of these two sections, as desired. □

Proposition 8.6. *Suppose that $\rho = 2$ and there are two swaps between rows j_0-1, j_0 and j_1-1, j_1 in columns $i_0 < i_1$ ("disjoint swap" of Proposition 4.14), Then, in an unimaginative multidegree, the potential support of every (j, j') is connected except possibly for (j_0-1, j_0-1), (j_1-1, j_1-1), and (j_0-1, j_1-1). Moreover, if (j_0-1, j_1-1) has disconnected potential support in multidegree ω, the potential support must be made up of two components, one contained strictly to the right of i_1, and one contained strictly to the left of i_0, and the potential support of (j_0-1, j_1) is contained strictly right of i_1-1, and the potential support of (j_0, j_1-1) is contained strictly left of i_0+1. Finally, if the potential support of (j_0-1, j_1) is contained strictly left of i_0, then (j_0-1, j_1-1) must also have a component of potential support contained strictly left of i_0, and if the potential support of (j_0, j_1-1) is contained strictly right of i_1, then (j_0-1, j_1-1) must also have a component of potential support contained strictly right of i_1.*

Proof. We write as usual $\omega = \mathrm{md}(w)$ with $w = (c_2, \ldots, c_g)$.

From Proposition 5.3, no rows are exceptional except row j_0-1 at i_0 and row j_1-1 at i_1. From Proposition 5.10, in multidegree ω, the potential support of every (j, j') is connected except possibly for (j_0-1, j_0-1), (j_1-1, j_1-1), and (j_0-1, j_1-1). and following the proof we see further that in order for (j_0-1, j_1-1) to have disconnected support, the support must be split between strictly right of i_1 and strictly left of i_0, as claimed. Next, if the potential support of (j_0-1, j_1-1) has a component lying strictly right of i_1, then

$$a^{i_1}_{(j_0-1, j_1)} = a^{i_1+1}_{(j_0-1, j_1)} - 1 = a^{i_1+1}_{(j_0-1, j_1-1)} - 2 > c_{i_1+1} - 2 \geq c_{i_1},$$

and the connectedness statement implies that the potential support of (j_0-1, j_1) is supported strictly to the right of i_1-1, as desired. The corresponding statement on support left of i_0 and i_0+1 follows similarly. Finally, if the potential support of (j_0-1, j_1) is contained strictly left of i_0, then

$$a^{i_0}_{(j_0-1,j_1-1)} < a^{i_0}_{(j_0-1,j_1)} < c_{i_0},$$

so (j_0-1, j_1-1) also has a component of potential support strictly left of i_0, as desired. The last statement on support strictly right of i_1 follows similarly. □

Proposition 8.7. *Suppose that $\rho = 2$ and there are two swaps in the "first 3-cycle" situation of Proposition 4.18. In an unimaginative multidegree, the potential support of every (j, j') is connected except possibly for (j_0-1, j_0-1), (j_0-1, j_0), and (j_0, j_0). Moreover, if for some j, the potential support of (j, j_0) has a component strictly to the left of i_0, then the potential support of (j, j_0-1) is entirely contained strictly to the left of i_0, and if the potential support of (j, j_0-1) has a component strictly to the right of i_1, then the potential support of (j, j_0) is entirely contained strictly to the right of i_1.*

Finally, if (j_0-1, j_0) has potential support contained entirely strictly to the left of i_1, then the potential support of (j_0-1, j_0+1) cannot be contained to the right of i_1; if it has potential support contained entirely strictly to the right of i_0, then the potential support of (j_0, j_0+1) cannot be contained to the left of i_0; and if it has potential support contained entirely strictly between i_0 and i_1, then (j_0-1, j_0-1) has potential support contained entirely strictly to the left of i_1, and (j_0, j_0) has potential support contained entirely strictly to the right of i_0.

Proof. Most of the argument is similar to Proposition 8.6. For the support of (j, j_0) to have a component strictly to the left of i_0 we must have $a^{i_0}_{(j,j_0)} \le c_{i_0}-1$, and then $a^{i_0}_{(j,j_0-1)} < c_{i_0}-1$. Arguing as in Proposition 8.6, we conclude that (even if $j = j_0-1$ or j_0) the support of (j, j_0-1) is connected and strictly to the left of i_0. The statement on support to the right of i_1 is proved in exactly the same way. For the last assertion, note that the (j_0-1, j_0-1) row has no support at i_1, and the (j_0, j_0) has no support at i_0, since both sum to $2d-4$ in the relevant columns. □

Proposition 8.8. *Suppose that X_0 is left-weighted, and that the rows j, j' have no exceptional behavior in any genus-0 columns. Then the image of $s_j \otimes s_{j'}$ in any unimaginative multidegree ω is equal to the leftmost potentially appearing section in the (j, j') row.*

Proof. The lack of exceptional behavior away from genus-1 components means that the $a^i_{(j,j')}$ are constant on the genus-0 components. The idea is then that the left-weighting means that the leftmost negative value of $a^i_{(j,j')} - c_i$ is repeated so many times that it must lead to a strict minimum of the partial sums. Compare the proof of Proposition 4.24, where in (4-1) we now replace d by $2d$ due to having passed to the tensor square. □

Proposition 8.9. *Suppose that $\rho = 2$, X_0 is left-weighted and we are in the "disjoint swap" case of Proposition 4.14, so that the limit linear series contains precisely two swaps in disjoint pairs of rows, say j_0, j_0-1 and j_1, j_1-1. Then for any unimaginative multidegree ω, choosing s'_{j_0} and s'_{j_1} as allowed* *by Proposition 4.24, the images in multidegree ω of the tensors of pairs of the s_j for $j \neq j_0, j_1$, and $s'_{j_0}, s''_{j_0}, s'_{j_1}, s''_{j_1}$ contain $\binom{r+2}{2}$ independent linear combinations of the potentially present sections.*

Proof. Without loss of generality, assume that $i_0 < i_1$. By Proposition 4.24, we may assume that s'_{j_1} is controlled, and that the j_1-part of s'_{j_1} does not contain any genus-1 component left of i_1. Every (j, j') has connected potential support unless $j, j' \in \{j_0 - 1, j_1 - 1\}$. Moreover, if $j, j' \neq j_0, j_0 - 1, j_1, j_1 - 1$, then we know that $f_{w_j + w'_j, w}(s_j \otimes s_{j'})$ is nonzero and composed of $s^i_{(j,j')}$. Now, suppose $j \neq j_0, j_0 - 1, j_1, j_1 - 1$. Then the same argument as in Proposition 8.4 shows that if we consider the images in multidegree ω of $s_j \otimes s_{j_0 - 1}$, $s_j \otimes s'_{j_0}$, and $s_j \otimes s''_{j_0}$, we either obtain one section of type $(j, j_0 - 1)$ and one with a contribution of type (j, j_0), or two sections of type $(j, j_0 - 1)$, but having disjoint support. The same holds with j_1 in place of j_0. Together, these produce $\binom{r-2}{2} + 4(r-3) = \binom{r+2}{2} - 10$ linearly independent combinations. It thus suffices to show that we have 10 linearly independent combinations coming from tensor products of pairs of the sections $s_{j_0-1}, s'_{j_0}, s''_{j_0}, s_{j_1-1}, s'_{j_1}, s''_{j_1}$. Just as in the proof of Proposition 8.4, tensor products of the first three sections yield three independent combinations, with contributions contained among the types $(j_0 - 1, j_0 - 1)$, $(j_0 - 1, j_0)$, and (j_0, j_0). Tensor products of the last three sections likewise yield three combinations, with j_1 replacing j_0 in the types.

It remains to consider the tensors with types contained among $(j_0 - 1, j_1 - 1)$, $(j_0 - 1, j_1)$, $(j_0, j_1 - 1)$ and (j_0, j_1). First suppose that $(j_0 - 1, j_1 - 1)$ has connected potential support in multidegree ω. Then just as in the single-swap case, at least one of $s_{j_0-1} \otimes s'_{j_1}$, $s_{j_0-1} \otimes s''_{j_1}$ must involve a $(j_0 - 1, j_1)$ part, and at least one of $s'_{j_0} \otimes s_{j_1-1}$, $s''_{j_0} \otimes s_{j_1-1}$ must involve a $(j_0, j_1 - 1)$ part. Since $s_{j_0-1} \otimes s_{j_1-1}$ is pure of type $(j_0 - 1, j_1 - 1)$, and all of these have unique potential support, we find that the span of these sections contains the (unique) pure types of each of $(j_0 - 1, j_1 - 1)$, $(j_0, j_1 - 1)$ and $(j_0 - 1, j_1)$. Thus, if we have anything with a nonzero part of type (j_0, j_1), this gives a fourth independent combination. On the other hand, if nothing has a (j_0, j_1) part, then we must have the following:

$$
\begin{aligned}
s'_{j_0} \otimes s''_{j_1} &= (j_0 - 1, j_1)_{\mathrm{L}} + (j_0, j_1 - 1)_{\mathrm{R}}, \\
s'_{j_0} \otimes s'_{j_1} &= (j_0 - 1, j_1 - 1)_{\mathrm{L}} + (j_0, j_1 - 1)_{\mathrm{LC}} \quad \text{and} \\
s''_{j_0} \otimes s''_{j_1} &= (j_0 - 1, j_1)_{\mathrm{CR}} + (j_0 - 1, j_1 - 1)_{\mathrm{R}}.
\end{aligned}
$$

First consider the possibility that the $(j_0 - 1, j_1)_{\mathrm{L}}$ part of $s'_{j_0} \otimes s''_{j_1}$ is nonzero. Then, by Proposition 8.6(1), we have that $(j_0 - 1, j_1 - 1)$ has support strictly left of i_0 too, which in turn means that $(j_0, j_1 - 1)$ can't have support strictly right of i_1. But this leaves no possibility for $s''_{j_0} \otimes s''_{j_1}$. On the other hand, if the $(j_0, j_1 - 1)_{\mathrm{R}}$ part of $s'_{j_0} \otimes s''_{j_1}$ is nonzero, we have that $(j_0 - 1, j_1 - 1)$ must have support strictly right of i_1, and hence that $(j_0 - 1, j_1)$ can't have support strictly left of i_0, leaving no possibility for $s'_{j_0} \otimes s'_{j_1}$. We conclude that it is not possible for these tensors not to have some (j_0, j_1) part, giving the desired four independent combinations when $(j_0 - 1, j_1 - 1)$ has connected potential support.

It remains to treat the case that $(j_0 - 1, j_1 - 1)$ has disconnected potential support in multidegree ω. Then Proposition 8.6 tells us that this potential support has two parts, contained strictly left of i_0 and right of i_1 respectively. Moreover, it says that the potential support of $(j_0 - 1, j_1)$ is contained strictly right of

$i_1 - 1$ and the potential support of $(j_0, j_1 - 1)$ is contained strictly left of $i_0 + 1$. This forces $s'_{j_0} \otimes s''_{j_1}$ to be of pure (j_0, j_1) type. Now, we observe that two of the sections $s_{j_0-1} \otimes s_{j_1-1}$, $s_{j_0-1} \otimes s'_{j_1}$, $s_{j_0-1} \otimes s''_{j_1}$ must be independent, either involving a $(j_0 - 1, j_1)$ part and a $(j_0 - 1, j_1 - 1)$ part, or two $(j_0 - 1, j_1 - 1)$ parts. Similarly, $s'_{j_0} \otimes s_{j_1-1}$ and $s''_{j_0} \otimes s_{j_1-1}$ must either involve a $(j_0, j_1 - 1)$ part or two $(j_0 - 1, j_1 - 1)$ parts. We see that the only way to avoid having four independent combinations would be if these five tensors are all of pure type $(j_0 - 1, j_1 - 1)$, necessarily achieving support independently both on the left and right. But we note that because the potential support of $(j_0, j_1 - 1)$ is contained strictly left of $i_0 + 1$, and because (in the disconnected support case) we must have $a^i_{(j_0-1, j_1-1)} = c_i$ for $i_0 < i \leq i_1$, the only way that $s''_{j_0} \otimes s_{j_0-1}$ can fail to have a $(j_0, j_1 - 1)$ part is if s''_{j_0} is not controlled, and more specifically if its j_0 portion does not extend more than halfway to the next genus-1 component after i_0. On the other hand, s'_{j_1} is controlled and has j_1 part not containing any genus-1 component smaller than i_1, so we conclude that in this situation its j_1 part is disjoint from the j_0 part of s''_{j_0}, and then $s''_{j_0} \otimes s'_{j_1} = (j_0, j_1 - 1) + (j_0 - 1, j_1)$, and gives a fourth independent combination. This completes the proof of the proposition. □

Proposition 8.10. *Suppose that $\rho = 2$ and we are in the "first 3-cycle" situation of Proposition 4.18. Choose an unimaginative multidegree ω such that the $(j_0 - 1, j_0)$ row has a unique potentially present section in multidegree ω, whose support does not contain i_0 or i_1 (use Corollary 7.6,). Then the images in multidegree ω of the tensors of pairs of the s_j for $j \neq j_0 + 1$, and $s'_{j_0+1}, s''_{j_0+1}, s'''_{j_0+1}$ contain $\binom{r+2}{2}$ independent linear combinations of the potentially present sections.*

Proof. We are assuming that there is one swap between the j_0-th and (j_0+1)-st on the i_0-th column, and a second swap between the (j_0-1)-st and (j_0+1)-st rows on the i_1-st column for some $i_1 > i_0$. We first show that for $j \neq j_0 - 1, j_0, j_0 + 1$, the sections

$$s_j \otimes s_{j_0-1}, \quad s_j \otimes s_{j_0}, \quad s_j \otimes s'_{j_0+1}, \quad s_j \otimes s''_{j_0+1}, \quad s_j \otimes s'''_{j_0+1}$$

must yield at least three independent combinations. But the first two tensors yield $(j, j_0 - 1)$ and (j, j_0) parts, so if any of the last three have any $(j, j_0 + 1)$ part, we obtain the desired independence. On the other hand, if not we find that

$$\begin{aligned} s_j \otimes s'_{j_0+1} &= (j, j_0 - 1)_{\mathrm{LC}} + (j, j_0)_{\mathrm{L}}; \\ s_j \otimes s''_{j_0+1} &= (j, j_0 - 1)_{\mathrm{R}} + (j, j_0)_{\mathrm{CR}}; \\ s_j \otimes s'''_{j_0+1} &= (j, j_0)_{\mathrm{L}} + (j, j_0 - 1)_{\mathrm{R}}. \end{aligned}$$

If the $(j, j_0)_{\mathrm{L}}$ part of the last tensor is nonzero, then by Proposition 8.7, the potential support of both the $(j, j_0 - 1)$ and (j, j_0) rows are connected and contained strictly to the left of i_0, leaving no possibility for the second tensor. But if the $(j, j_0 - 1)_{\mathrm{R}}$ part of the last tensor is nonzero, then similarly the potential support of both the $(j, j_0 - 1)$ and (j, j_0) rows are contained strictly to the right of i_1, leaving no possibility for the first tensor. Thus, we reach a contradiction, and conclude that we must obtain a $(j, j_0 + 1)$ part, giving the desired three independent combinations.

Next, we consider the 15 tensors arising from

$$s_{j_0-1}, \quad s_{j_0}, \quad s'_{j_0+1}, \quad s''_{j_0+1}, \quad s'''_{j_0+1};$$

we need to show that these yield 6 independent linear combinations.

By hypothesis, we have that the potential support of the (j_0-1, j_0) row is connected and does not contain i_0 or i_1, so we organize cases according to its support. First suppose that the support of the (j_0-1, j_0) row is entirely to the left of i_0; then according to Proposition 8.7, the same holds for the (j_0-1, j_0-1) row, and the (j_0-1, j_0+1) row cannot have its support to the right of i_1. We then see that $s_{j_0-1} \otimes s''_{j_0+1}$ cannot have any (j_0-1, j_0-1) or (j_0-1, j_0) parts, so must be of (j_0-1, j_0+1) type. Similarly, $s''_{j_0+1} \otimes s'''_{j_0+1}$ cannot have any (j_0-1, j_0-1), (j_0-1, j_0), or (j_0-1, j_0+1) parts, so it must contain (j_0, j_0+1) or (j_0+1, j_0+1) parts. In addition, the pair $s_{j_0} \otimes s''_{j_0+1}$ and $s_{j_0} \otimes s'''_{j_0+1}$ must contain either a (j_0, j_0+1) part, or two distinct (j_0, j_0) parts, supported left and right of i_0, respectively. Given that we always have (j_0-1, j_0-1), (j_0-1, j_0) and (j_0, j_0) parts, the only way we could fail to have produced six independent combinations is if $s''_{j_0+1} \otimes s'''_{j_0+1}$ has type (j_0, j_0+1), and we have only one (j_0, j_0) part. But then considering $s''^{\otimes 2}_{j_0+1}$ and $s'''^{\otimes 2}_{j_0+1}$ and using Lemma 8.2, we must produce a (j_0+1, j_0+1) part or two distinct (j_0, j_0) parts, so we necessarily obtain the sixth combination.

Similarly, if the potential support of the (j_0-1, j_0) row is entirely to the right of i_1, then Proposition 8.7 tells us that the same holds for (j_0, j_0), and that the potential support of the (j_0, j_0+1) row cannot be to the left of i_0. Then $s_{j_0} \otimes s'_{j_0+1}$ must be of (j_0, j_0+1) type, and $s'_{j_0+1} \otimes s'''_{j_0+1}$ must have (j_0-1, j_0+1) or (j_0+1, j_0+1) parts. The pair $s_{j_0-1} \otimes s'_{j_0+1}$ and $s_{j_0-1} \otimes s'''_{j_0+1}$ must contain either a (j_0-1, j_0+1) part, or two distinct (j_0-1, j_0-1) parts, and in either case the tensors $s'^{\otimes 2}_{j_0+1}$ and $s'''^{\otimes 2}_{j_0+1}$ (together with the usual tensors of s_{j_0-1} and s_{j_0}) must complete the six independent combinations.

Finally, if the potential support of the (j_0-1, j_0) row is between the i_0 and i_1 columns, then by Proposition 8.6, we know that the potential support of (j_0-1, j_0-1) is left of i_1 and the potential support of (j_0, j_0) is right of i_0. We then see that the tensors $s_{j_0-1} \otimes s'''_{j_0+1}$, $s_{j_0} \otimes s'''_{j_0+1}$, and $s'''^{\otimes 2}_{j_0+1}$ must be pure of types (j_0-1, j_0+1), (j_0, j_0+1), and (j_0+1, j_0+1) respectively, yielding the desired six combinations. □

Lemma 8.11. *Assume that $\rho = 2$ and $r = 6$ and we are in the "second 3-cycle" situation of Proposition 4.20. Then, there is an unimaginative multidegree ω_{def}, such that one of the following options is satisfied*:

(a) *the (j_0-1, j_0-1) row does not have potentially present sections both left of i_0 and right of i_1; or*

(b) $2a^{i_0}_{j_0-1} = c_{i_0} - 1$, *and* $2a^{i_1+1}_{j_0-1} = c_{i_1+1} + 1$; *or*

(c) $2a^{i_0}_{j_0-1} = c_{i_0} - 2$, *and* $2a^{i_1+1}_{j_0-1} = c_{i_1+1} + 2$, *and w has degree 2 in both i_0 and i_1.*

Proof. If (a) does not hold, the (j_0-1, j_0-1) row has support both left of i_0 and right of i_1. From Proposition 5.10, then

$$a^{i_0}_{(j_0-1, j_0-1)} < c_{i_0}, \quad a^{i_1+1}_{(j_0-1, j_0-1)} > c_{i_1}.$$

Write $a^{i_0}_{(j_0-1, j_0-1)} = c_{i_0} - k_0$, $a^{i_1+1}_{(j_0-1, j_0-1)} = c_{i_1} + k_1$. Denote by t the number of elliptic components from i_0, to i_1 (inclusive). As we are assuming that $\rho = 2$ and there are two swaps, there are no other exceptional columns. Therefore, between i_0 and i_1, $a^{i+1}_{j_0-1} = a^{i}_{j_0-1}$ if the component i is rational, $a^{i+1}_{j_0-1} \leq a^{i}_{j_0-1} + 1$ if it is elliptic with $i \neq i_0, i_1$ and $a^{i_k+1}_{j_0-1} = a^{i_k}_{j_0-1} + 2$, $k = 0, 1$. Therefore

$$a^{i_1+1}_{(j_0-1, j_0-1)} \leq a^{i_0}_{(j_0-1, j_0-1)} + 4 + 2t.$$

If ω_{def} is an unimaginative multidegree, then it has degree 0 on rational components, and 2 or 3 on elliptic components with γ_i the number of 3s in the first i components. Therefore

$$c_{i_1+1} = c_{i_0} + 2t + (\gamma_{i_1} - \gamma_{i_0-1}).$$

From these identities, it follows that

$$k_0 + k_1 \leq 4 - (\gamma_{i_1} - \gamma_{i_0-1}).$$

As by assumption, both k_0, k_1 are strictly positive, the options for the pair (k_0, k_1) are

$$(1, 1), \quad (1, 2), \quad (2, 1), \quad (2, 2).$$

The case $(1, 1)$ is option (b). In case $(2, 2)$, $\gamma_{i_1} - \gamma_{i_0-1} = 0$, therefore the degree on each elliptic component from i_0 to i_1 is 2. This is option (c). From Corollary 7.7, in cases $(1, 2)$, $(2, 1)$, with a suitable choice of multidegree, we are in case (a). Therefore, the result is proved □

Proposition 8.12. *Suppose that $\rho = 2$, X_0 is left-weighted and we are in the "second 3-cycle" situation of Proposition 4.20. Choose an unimaginative multidegree $w = (c_2, \dots, c_N)$ satisfying one of the conditions of Lemma 8.11. Then the images in multidegree $\operatorname{md}(w)$ of the tensors of pairs of the s_j for $j \neq j_0, j_0+1$, and $s'_{j_0}, s''_{j_0}, s'_{j_0+1}, s''_{j_0+1}, s'''$ contain $\binom{r+2}{2}$ independent linear combinations of the potentially present sections.*

Proof. By assumption, the limit linear series contains precisely two swaps, with one swap between the (j_0-1)-st and j_0-th rows occurring in the i_0-th column, and a second swap between the (j_0-1)-st and (j_0+1)-st rows in the i_1-st column for some $i_1 > i_0$. First suppose $j \neq j_0 - 1, j_0, j_0 + 1$; we show that we can always obtain three linearly independent combinations of potentially present sections from the rows $(j, j_0 - 1)$, (j, j_0) and $(j, j_0 + 1)$. $s_j \otimes s_{j_0-1}$ always yields a pure $(j, j_0 - 1)$ part. If $S'_2 = S''_4 = \{1, \dots, N\}$, then $s_j \otimes s'_{j_0}$ has a nonzero (j, j_0) part and no $(j, j_0 + 1)$ part, while $s_j \otimes s''_{j_0+1}$ has a nonzero $(j, j_0 + 1)$ part, so we get the desired three combinations. Otherwise, we have

$$\begin{aligned} s_j \otimes s'_{j_0} &= (j, j_0 - 1)_{\mathrm{L}} + (j, j_0), \\ s_j \otimes s''_{j_0} &= (j, j_0) + (j, j_0 + 1)_{\mathrm{R}'} + (j, j_0 - 1)_{\mathrm{CR}}, \\ s_j \otimes s'_{j_0+1} &= (j, j_0 - 1)_{\mathrm{LC}} + (j, j_0)_{\mathrm{L}'} + (j, j_0 + 1), \\ s_j \otimes s''_{j_0+1} &= (j, j_0 + 1) + (j, j_0 - 1)_{\mathrm{R}}, \\ s_j \otimes s''' &= (j, j_0) + (j, j_0 - 1)_{\mathrm{C}} + (j, j_0 + 1), \end{aligned}$$

where R' and L' denote possible support at and right of i_1 and at and left of i_0, respectively, and if $s_j \otimes s''_{j_0}$ has a nonzero $(j, j_0 + 1)$ part with support containing i_1, its (j, j_0) part must be nonzero, and similarly for the (j, j_0) and $(j, j_0 + 1)$ parts of $s_j \otimes s'_{j_0+1}$. Now, suppose that $(j, j_0 - 1)$ has connected potential support which is not contained strictly right of i_0. Then (j, j_0+1) cannot have any potential support strictly right of i_1 without also forcing (j, j_0-1) to have potential support strictly right of i_1, so the (j, j_0) part of $s_j \otimes s''_{j_0}$ must be nonzero. But then adding $s_j \otimes s''_{j_0+1} = (j, j_0+1)$ and $s_j \otimes s_{j_0-1}$ yields three independent sections. Similarly, if (j, j_0-1) has connected potential support not contained strictly left of i_1, then (j, j_0) cannot have potential support strictly left of i_0, so $s_j \otimes s'_{j_0+1}$ has nonzero (j, j_0+1) part, and adding $s_j \otimes s'_{j_0} = (j, j_0)$ and $s_j \otimes s_{j_0-1}$ yields the desired combinations. For connected potential support, the only remaining possibility is that (j, j_0-1) has potential support strictly between i_0 and i_1, in which case $s_j \otimes s'_{j_0} = (j, j_0)$ and $s_j \otimes s''_{j_0+1} = (j, j_0+1)$.

Finally, since $\rho = 2$, the only remaining possibility is that (j, j_0-1) has potential support both left of i_0 and right of i_1, and in this case we must have $a^{i_0}_{(j,j_0-1)} = c_{i_0} - 1$ and $a^{i_1+1}_{(j,j_0-1)} = c_{i_1+1} + 1$. Then (j, j_0+1) cannot have potential support strictly right of i_1, and (j, j_0) cannot have potential support strictly left of i_0, so as above we find that if the (j, j_0+1) part of $s_j \otimes s''_{j_0}$ is nonzero (necessarily with support at i_1), then the (j, j_0) part must also be nonzero, and if the (j, j_0) part of $s_j \otimes s'_{j_0+1}$ is nonzero, then the (j, j_0+1) part must also be nonzero. Now, we have $s_j \otimes s'_{j_0}$ and $s_j \otimes s''_{j_0+1}$ linearly independent always, and the only way they could fail to be independent from $s_j \otimes s'''$ is if either $s_j \otimes s'_{j_0} = (j, j_0)$ or $s_j \otimes s''_{j_0+1} = (j, j_0+1)$, while the only way they could fail to be independent from $s_j \otimes s_{j_0-1}$ if is either $s_j \otimes s'_{j_0} = (j, j_0-1)_{\mathrm{L}}$ or $s_j \otimes s''_{j_0+1} = (j, j_0-1)_{\mathrm{R}}$. If $s_j \otimes s'_{j_0} = (j, j_0)$ and $s_j \otimes s''_{j_0+1} = (j, j_0-1)_{\mathrm{R}}$, we see that $s_j \otimes s'_{j_0+1}$ necessarily gives a third independent combination, while if $s_j \otimes s'_{j_0} = (j, j_0-1)_{\mathrm{L}}$ and $s_j \otimes s''_{j_0+1} = (j, j_0+1)$, we see that $s_j \otimes s''_{j_0}$ necessarily gives a third independent combination.

It remains to show that we can get six independent combinations from the rows (j_0-1, j_0-1), (j_0-1, j_0), (j_0-1, j_0+1), (j_0, j_0), (j_0, j_0+1), and (j_0+1, j_0+1). If $S'_2 = S''_4 = \{1, \dots, N\}$, then we immediately get that the six tensors coming from $s_{j_0-1}, s'_{j_0}, s''_{j_0+1}$ are linearly independent, as desired. Otherwise, we will make use of the mixed section s''' to handle certain cases. For reference, we write out the form of all the relevant tensors of $s_{j_0-1}, s'_{j_0}, s''_{j_0}, s'_{j_0+1}, s''_{j_0+1}$ (we are making use of Lemma 8.2 in the case of self-tensors):

$$
\begin{aligned}
s_{j_0-1} \otimes s'_{j_0} &= (j_0-1, j_0-1)_{\mathrm{L}} + (j_0-1, j_0),\\
s_{j_0-1} \otimes s''_{j_0} &= (j_0-1, j_0) + (j_0-1, j_0+1)_{\mathrm{R}'} + (j_0-1, j_0-1)_{\mathrm{CR}},\\
s_{j_0-1} \otimes s'_{j_0+1} &= (j_0-1, j_0-1)_{\mathrm{LC}} + (j_0-1, j_0)_{\mathrm{L}'} + (j_0-1, j_0+1),\\
s_{j_0-1} \otimes s''_{j_0+1} &= (j_0-1, j_0+1) + (j_0-1, j_0-1)_{\mathrm{R}},\\
s'_{j_0} \otimes s'_{j_0} &= (j_0-1, j_0-1)_{\mathrm{L}} + (j_0, j_0),\\
s''_{j_0} \otimes s''_{j_0} &= (j_0, j_0) + (j_0+1, j_0+1)_{\mathrm{R}} + (j_0-1, j_0-1)_{\mathrm{CR}},\\
s'_{j_0} \otimes s''_{j_0} &= (j_0-1, j_0) + (j_0, j_0) + (j_0, j_0+1)_{\mathrm{R}'},\\
s'_{j_0+1} \otimes s'_{j_0+1} &= (j_0-1, j_0-1)_{\mathrm{LC}} + (j_0, j_0)_{\mathrm{L}} + (j_0+1, j_0+1),\\
s''_{j_0+1} \otimes s''_{j_0+1} &= (j_0+1, j_0+1) + (j_0-1, j_0-1)_{\mathrm{R}},\\
s'_{j_0+1} \otimes s''_{j_0+1} &= (j_0-1, j_0+1) + (j_0, j_0+1)_{\mathrm{L}'} + (j_0+1, j_0+1),
\end{aligned}
$$

$$s'_{j_0} \otimes s'_{j_0+1} = (j_0-1,\, j_0-1)_{\mathrm{L}} + (j_0-1,\, j_0)_{\mathrm{LC}} + (j_0-1,\, j_0+1)_{\mathrm{L}} + (j_0,\, j_0)_{\mathrm{L}'} + (j_0,\, j_0+1),$$
$$s'_{j_0} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)_{\mathrm{L}} + (j_0-1,\, j_0)_{\mathrm{R}} + (j_0,\, j_0+1),$$
$$s''_{j_0} \otimes s''_{j_0+1} = (j_0-1,\, j_0-1)_{\mathrm{R}} + (j_0-1,\, j_0)_{\mathrm{R}} + (j_0-1,\, j_0+1)_{\mathrm{CR}} + (j_0,\, j_0+1) + (j_0+1,\, j_0+1)_{\mathrm{R}'}.$$

As above, we separate out cases by the potential support of the $(j_0-1,\, j_0-1)$ row. Note that because the entries sum to $2d-4$ in both the i_0 and i_1 columns, the $(j_0-1,\, j_0-1)$ row cannot have any potential support in either of these columns in any unimaginative multidegree. First suppose the potential support is strictly to the left of i_0. In this case none of the relevant rows can have potential support extending right of i_1, so we get $s_{j_0-1} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)$, $s''_{j_0} \otimes s''_{j_0} = (j_0,\, j_0)$, and $s''_{j_0+1} \otimes s''_{j_0+1} = (j_0+1,\, j_0+1)$, and the $(j_0-1,\, j_0)$ part of $s_{j_0-1} \otimes s''_{j_0}$ must be nonzero. We also have $s'_{j_0} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)_{\mathrm{L}} + (j_0,\, j_0+1)$ and $s''_{j_0} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)_{\mathrm{CR}} + (j_0,\, j_0+1) + (j_0+1,\, j_0+1)_{\mathrm{R}'}$, where again the latter has to have nonzero $(j_0,\, j_0+1)$ part unless it is equal to $(j_0-1,\, j_0+1)_{\mathrm{CR}}$, so these must either yield a nonzero $(j_0,\, j_0+1)$ part, or two independent $(j_0-1,\, j_0+1)$ parts (which won't happen when $\rho = 2$), and in either case together with $s_{j_0-1} \otimes s_{j_0-1}$ we get the desired six independent combinations.

Similarly, if the potential support of the $(j_0-1,\, j_0-1)$ row is strictly to the right of i_1, we will have

$$s_{j_0-1} \otimes s'_{j_0} = (j_0-1,\, j_0), \quad s'_{j_0} \otimes s'_{j_0} = (j_0,\, j_0), \quad s'_{j_0+1} \otimes s'_{j_0+1} = (j_0+1,\, j_0+1),$$

with $s_{j_0-1} \otimes s'_{j_0+1}$ having a nonzero $(j_0-1,\, j_0+1)$ part, and $s'_{j_0} \otimes s''_{j_0+1} = (j_0-1,\, j_0)_{\mathrm{R}} + (j_0,\, j_0+1)$ and $s'_{j_0} \otimes s'_{j_0+1} = (j_0-1,\, j_0)_{\mathrm{LC}} + (j_0,\, j_0+1) + (j_0,\, j_0)_{\mathrm{L}'}$, and we again obtain six independent combinations in the same manner.

If the potential support of the $(j_0-1,\, j_0-1)$ row is strictly between i_0 and i_1, then none of the relevant rows can have support either left of i_0 or right of i_1, and we get $s_{j_0-1} \otimes s'_{j_0} = (j_0-1,\, j_0)$, $s_{j_0-1} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)$, $s'_{j_0} \otimes s'_{j_0} = (j_0,\, j_0)$, $s''_{j_0+1} \otimes s''_{j_0+1} = (j_0+1,\, j_0+1)$, and $s'_{j_0} \otimes s''_{j_0+1} = (j_0,\, j_0+1)$.

If the $(j_0-1,\, j_0-1)$ row has disconnected potential support to the left of i_0 and strictly between i_0 and i_1, then once again none of the relevant rows can have potential support extending right of i_1, and because $\rho = 2$ we must have $a^{i_0}_{(j_0-1,\, j_0-1)} = c_{i_0} - 1$, so none of the other relevant rows can have their potential support contained strictly left of i_0, either. Moreover, the $(j_0-1,\, j_0)$ row must have potential support containing i_0, so $s'_{j_0} \otimes s''_{j_0}$ cannot have any $(j_0-1,\, j_0)$ part, and its $(j_0,\, j_0)$ part must be nonzero. We then find that $s_{j_0-1} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1)$, $s''_{j_0+1} \otimes s''_{j_0+1} = (j_0+1,\, j_0+1)$, and $s'_{j_0} \otimes s''_{j_0+1} = (j_0,\, j_0+1)$. If the $(j_0-1,\, j_0)$ part of $s_{j_0-1} \otimes s''_{j_0}$ is nonzero, then these together with $s_{j_0-1} \otimes s_{j_0-1}$ give six independent combinations. Otherwise, we must have $s_{j_0-1} \otimes s''_{j_0} = (j_0-1,\, j_0-1)_{\mathrm{C}}$, and we see that $s_{j_0-1} \otimes s'_{j_0} = (j_0-1,\, j_0-1)_{\mathrm{L}} + (j_0-1,\, j_0)$ gives a sixth independent combination.

The situation is nearly the same if the $(j_0-1,\, j_0-1)$ row has disconnected potential support to the right of i_1 and strictly between i_0 and i_1. Here we instead obtain that $(j_0-1,\, j_0+1)$ must have potential support containing i_1, and thus that $s_{j_0-1} \otimes s'_{j_0} = (j_0-1,\, j_0)$, $s'_{j_0} \otimes s'_{j_0} = (j_0,\, j_0)$, and $s'_{j_0} \otimes s''_{j_0+1} = (j_0,\, j_0+1)$, with $s'_{j_0+1} \otimes s''_{j_0+1}$ having nonzero $(j_0+1,\, j_0+1)$ part. Then $s_{j_0-1} \otimes s'_{j_0+1}$ either has a nonzero $(j_0-1,\, j_0+1)$ part, or is equal to $(j_0-1,\, j_0-1)_{\mathrm{C}}$, and in either case we obtain a sixth combination, from $s_{j_0-1} \otimes s_{j_0-1}$ or $s_{j_0-1} \otimes s''_{j_0+1} = (j_0-1,\, j_0+1) + (j_0-1,\, j_0-1)_{\mathrm{R}}$, respectively.

If (j_0-1, j_0-1) has three components of potential support, necessarily left of i_0, strictly between i_0 and i_1, and right of i_1, then none of the relevant rows other than (j_0-1, j_0-1) can have potential support contained strictly left of i_0 or strictly right of i_1, and we also know that the potential support of the (j_0-1, j_0) (respectively, (j_0-1, j_0+1)) row contains i_0 (respectively, i_1). We then have that $s'_{j_0}\otimes s''_{j_0+1} = (j_0, j_0+1)$, and that $s'_{j_0}\otimes s''_{j_0}$ and $s'_{j_0+1}\otimes s''_{j_0+1}$ have nonzero (j_0, j_0) and (j_0+1, j_0+1) parts, respectively. We also have $s_{j_0-1}\otimes s'_{j_0} = (j_0-1, j_0-1)_{\mathrm{L}} + (j_0-1, j_0),$, $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0-1)_{\mathrm{R}} + (j_0-1, j_0+1)$, and $s_{j_0-1}\otimes s''' = (j_0-1, j_0) + (j_0-1, j_0-1)_{\mathrm{C}} + (j_0-1, j_0+1)$. To have a dependence between these, we need (at least one of) $s_{j_0-1}\otimes s'_{j_0} = (j_0-1, j_0)$ or $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0+1)$. On the other hand, to have a dependence between the first five and $s_{j_0-1}\otimes s_{j_0-1}$, we need $s_{j_0-1}\otimes s'_{j_0} = (j_0-1, j_0-1)_{\mathrm{L}}$ or $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0-1)_{\mathrm{R}}$. If $s_{j_0-1}\otimes s'_{j_0} = (j_0-1, j_0)$ and $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0-1)_{\mathrm{R}}$, we see that $s_{j_0-1}\otimes s'_{j_0+1}$ must have a nonzero $(j_0-1, j_0-1)_{\mathrm{LC}}$ or (j_0-1, j_0+1) part, and thus gives a sixth independent combination. On the other hand, if $s_{j_0-1}\otimes s'_{j_0} = (j_0-1, j_0-1)_{\mathrm{L}}$ and $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0+1)$, we see that $s_{j_0-1}\otimes s''_{j_0}$ must have a nonzero $(j_0-1, j_0-1)_{\mathrm{CR}}$ or (j_0-1, j_0) part, and again gives a sixth independent combination.

It remains to analyze the case that (j_0-1, j_0-1) has two components of potential support, one left of i_0, and the other right of i_1. By hypothesis, we only have to address the case that $a^{i_0}_{(j_0-1, j_0-1)} = c_{i_0}-2$ and $a^{i_1+1}_{(j_0-1, j_0-1)} = c_{i_1+1}+2$, and that we have degree 2 in both i_0 and i_1. In this situation, the (j_0-1, j_0) row has potential support strictly left of i_0, but none of the other relevant rows do, and the (j_0, j_0) row must have support containing i_0 and extending left to at least the previous genus-1 component. Similarly, the (j_0-1, j_0+1) row has potential support strictly right of i_1, but none of the other relevant rows do, and the (j_0+1, j_0+1) row has support containing i_1 and extending to the right to at least the next genus-1 component. We also see that the potential support of (j_0, j_0+1) must be contained between i_0 and i_1 inclusive, and cannot be equal solely to i_0 or to i_1. In particular, $s'_{j_0}\otimes s''_{j_0+1}$ cannot have a (j_0-1, j_0) or (j_0-1, j_0+1) part, so must be equal to (j_0, j_0+1).

Now, $s_{j_0-1}\otimes s_{j_0-1} = (j_0-1, j_0-1)_{\mathrm{L}}$ because X_0 is left-weighted, and we begin by considering the case that no tensor has a $(j_0-1, j_0-1)_{\mathrm{R}}$ part. Then we must have $s_{j_0-1}\otimes s''_{j_0+1} = (j_0-1, j_0+1)$, $s''_{j_0+1}\otimes s''_{j_0+1} = (j_0+1, j_0+1)$, $s''_{j_0}\otimes s''_{j_0} = (j_0, j_0)$, and we also see that $s_{j_0-1}\otimes s''_{j_0}$ must be (j_0-1, j_0), because it could only have a $(j_0-1, j_0+1)_{\mathrm{R}'}$ part if the j_0 part of s''_{j_0} extends through i_1, and in this case the fact that X_0 is left-weighted gives us that $s_{j_0-1}\otimes s''_{j_0} = (j_0-1, j_0)$ regardless. Thus, we obtain the desired six independent combinations in this case.

On the other hand, if any tensor has a $(j_0-1, j_0-1)_{\mathrm{R}}$ part, we need to produce only three more independent combinations, and we consider the four tensors,

$$s'_{j_0}\otimes s''_{j_0} = (j_0-1, j_0) + (j_0, j_0), \qquad s'_{j_0+1}\otimes s''_{j_0+1} = (j_0-1, j_0+1) + (j_0+1, j_0+1),$$
$$s_{j_0-1}\otimes s''' = (j_0-1, j_0) + (j_0-1, j_0+1), \qquad s'''\otimes s''' = (j_0, j_0) + (j_0+1, j_0+1).$$

These must have at least a three-dimensional span unless they collapse into equal pairs, and there are two possibilities for this: either $s'_{j_0}\otimes s''_{j_0} = s_{j_0-1}\otimes s''' = (j_0-1, j_0)$ and $s'_{j_0+1}\otimes s''_{j_0+1} = s'''\otimes s''' = (j_0+1, j_0+1)$, or $s'_{j_0}\otimes s''_{j_0} = s'''\otimes s''' = (j_0, j_0)$ and $s'_{j_0+1}\otimes s''_{j_0+1} = s_{j_0-1}\otimes s''' = (j_0-1, j_0+1)$. Moreover, Proposition 4.24

implies that the j_0-part of s'_{j_0} doesn't contain any genus-1 components left of i_0. Then we necessarily have $s'_{j_0} \otimes s''_{j_0} = (j_0 - 1, j_0)$, so only the first possibility above can occur. Now, in general we have $s''_{j_0} \otimes s''' = (j_0, j_0) + (j_0 - 1, j_0 + 1)_{\mathrm{CR}} + (j_0 + 1, j_0 + 1)_{\mathrm{R}'} + (j_0 - 1, j_0 - 1)_{\mathrm{C}} + (j_0 - 1, j_0)_{\mathrm{CR}} + (j_0, j_0 + 1)$, which in our case simplifies to $s''_{j_0} \otimes s''' = (j_0, j_0) + (j_0 - 1, j_0 + 1)_{\mathrm{CR}} + (j_0, j_0 + 1) + (j_0 + 1, j_0 + 1)_{\mathrm{R}'}$.

If this has nonzero (j_0, j_0) or $(j_0 - 1, j_0 + 1)$ term, we have our sixth independent combination. On the other hand, if the $(j_0 + 1, j_0 + 1)$ term is nonzero, the $(j_0, j_0 + 1)$ term must also be. Because the potential support of $(j_0, j_0 + 1)$ must end no later than i_1 and cannot be supported solely at i_1, if the $(j_0, j_0 + 1)$ term of $s''_{j_0} \otimes s'''$ is nonzero, this means that the j_0 part of s''_{j_0} must extend to cover all of $(j_0, j_0 + 1)$ (note that the proof of Lemma 8.1 indicates that a $(j_0, j_0 + 1)$ part has to come from either a j_0 part of s''_{j_0} and a $(j_0 + 1)$ part of s''' or vice versa, but not some mixture of the two). But we know that this contains at least one genus-1 component strictly right of i_0, so since the support of (j_0, j_0) ends at i_0, and X_0 is left-weighted, we conclude that we would have to have $s''_{j_0} \otimes s''_{j_0} = (j_0, j_0)$ in this case. Thus, in all cases we obtain the desired six independent combinations. □

9. Proof of the main theorem

We are now ready to prove Theorem 1.1. The main point is that if we have a smoothing family $\pi : X \to B$ as in Remark 3.2, and a generic linear series $(\mathscr{L}_\eta, V_\eta)$, which after base change and blowup we may assume is rational on the generic fiber, we can apply the linked linear series construction both to $(\mathscr{L}_\eta, V_\eta)$ and to $(W_\eta, \mathscr{L}_\eta^{\otimes 2})$, where W_η is the image of the multiplication map (1-1) of sections $s' \in V_{\omega'}$ and $s'' \in V_{\omega''}$. As in the discussion following Proposition 3.12 of [Liu et al. 2021], for any multidegree ω of total degree $2d$, and any multidegrees ω', ω'' of total degree d, $f_{\omega'+\omega'',\omega}(s' \otimes s'')$ lies in W_ω. Thus, in order to give a lower bound on the rank of (1-1), we can choose many different ω', ω'' and s', s'', and show that they span a certain-dimensional subspace of $(\mathscr{L}^{\otimes 2})_\omega$.

Theorem 9.1. *We assume characteristic* 0. *Fix g, r, d with $r \geq 3$ and $\rho = 1$ or $\rho = 2$.*

If $\rho = 1$, suppose that for every generic chain of rational and elliptic curves X_0 and every refined limit $\mathfrak{g}^r_d$ on X_0, there is a multidegree ω such that the potentially present sections in multidegree ω are linearly independent.

If $\rho = 2$, suppose that for every left-weighted generic chain of elliptic and rational curves X_0 of total genus g and every refined limit $\mathfrak{g}^r_d$ on X_0, there is an unimaginative $w = (c_2, \ldots, c_N)$ such that the potentially present sections in multidegree $\mathrm{md}(w)$ *are linearly independent*,

Then the strong maximal rank conjecture holds for (g, r, d), and more specifically, if we define $\mathcal{D}_{(g,r,d)} \subseteq \mathcal{M}_g$ to be the set of curves which have a $\mathfrak{g}^r_d$ for which (1-1) *is not injective, then the closure in $\overline{\mathcal{M}}_g$ of $\mathcal{D}_{g,r,d}$ does not contain a general chain of genus*-1 *curves.*

Proof. According to the above discussion together with Theorem 3.4 and Proposition 3.10, we need to show that an arbitrary exact linked linear series on X_0 lying over a refined limit linear series admits some multidegree ω such that the combined images $f_{\omega'+\omega'',\omega}(s' \otimes s'')$ span an $\binom{r+2}{2}$-dimensional space. For the ω in the statement, it then suffices to show that these sections give $\binom{r+2}{2}$ independent combinations of the potentially present sections.

When $\rho = 1$, from Proposition 5.3 , we can have at most one swap. If we have no swaps, we obtain the desired independence directly from the independence of the potentially present sections, using Proposition 4.9. On the other hand, if we have a single swap, Proposition 8.4 states that there are $\binom{r+2}{2}$-linearly independent combinations of the potentially present sections in some unimaginative multidegree. We have proved the statement for all X_0 at once, so the stronger assertion on the closure of $\mathcal{D}_{(g,r,d)}$ follows (compare with the proof of Proposition 3.13 of [Liu et al. 2021]).

If $\rho = 2$, from Proposition 5.3, there are at most two swaps. If there are no swaps or one swap, the proof above still works. If there are two swaps, we can use Propositions 8.5, 8.9, 8.10 together with Corollary 7.6 and Proposition 8.12 together with Lemma 8.11 to guarantee the existence of $\binom{r+2}{2}$- linearly independent combinations of the potentially present sections in some unimaginative multidegree.

When $\rho = 2$, we assume X_0 is left-weighted. This forces us to consider only special directions of approach to X_0 in $\overline{\mathcal{M}}_g$. Recalling that being left-weighted is preserved under the insertions of genus-0 chains which occur from base change and then blow up to resolve the resulting singularities, we conclude that for suitable smoothing families, the generic fiber cannot carry a $\mathfrak{g}^r_d$ for which (1-1) is not injective, as desired. □

Theorem 1.1 follows now from Theorem 9.1 together with Theorem 7.3 using that $\rho = 1$ when $g = 22$ and $\rho = 2$ when $g = 23$ and that in both cases, $r = 6$.

Remark 9.2. In our arguments for the $g = 23$ case, we used the $\rho = 2$ hypothesis in two distinct ways: first, to limit the number of swaps occurring to two, but then also to control the behavior of the rest of the limit linear series when two swaps did occur, for instance limiting the number of possibilities for rows having disconnected potential support. This may appear discouraging from the point of view of generalizing to cases with higher ρ, but as ρ increases, one also obtains more flexibility in choosing multidegrees while still maintaining linear independence of the potentially present sections. Indeed, we are taking advantage of this phenomenon already in the $\rho = 2$ case with Corollary 7.6.

Acknowledgements

We would like to thank Gavril Farkas for helpful contextual conversations and anonymous referees for suggesting improvements to the presentation of the paper.

References

[Aprodu and Farkas 2011] M. Aprodu and G. Farkas, "Koszul cohomology and applications to moduli", pp. 25–50 in *Grassmannians, moduli spaces and vector bundles* (Cambridge, MA, 2006), edited by D. A. Ellwood and E. Previato, Clay Math. Proc. **14**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Bakker and Farkas 2018] B. Bakker and G. Farkas, "The Mercat conjecture for stable rank 2 vector bundles on generic curves", *Amer. J. Math.* **140**:5 (2018), 1277–1295. MR Zbl

[Ballico 2012] E. Ballico, "Embeddings of general curves in projective spaces: the range of the quadrics", *Lith. Math. J.* **52**:2 (2012), 134–137. MR Zbl

[Eisenbud and Harris 1986] D. Eisenbud and J. Harris, "Limit linear series: basic theory", *Invent. Math.* **85**:2 (1986), 337–371. MR Zbl

[Eisenbud and Harris 1987] D. Eisenbud and J. Harris, "The Kodaira dimension of the moduli space of curves of genus ≥ 23", *Invent. Math.* **90**:2 (1987), 359–387. MR Zbl

[Esteves and Osserman 2013] E. Esteves and B. Osserman, "Abel maps and limit linear series", *Rend. Circ. Mat. Palermo* (2) **62**:1 (2013), 79–95. MR Zbl

[Farkas 2009a] G. Farkas, "Birational aspects of the geometry of $\bar{\mathcal{M}}_g$", pp. 57–110 in *Geometry of Riemann surfaces and their moduli spaces*, edited by L. Ji et al., Surv. Differ. Geom. **14**, Int. Press, Somerville, MA, 2009. MR Zbl

[Farkas 2009b] G. Farkas, "The global geometry of the moduli space of curves", pp. 125–147 in *Algebraic geometry*, *I* (Seattle, WA, 2005), edited by D. Abramovich et al., Proc. Sympos. Pure Math. **80.1**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Farkas 2009c] G. Farkas, "Koszul divisors on moduli spaces of curves", *Amer. J. Math.* **131**:3 (2009), 819–867. MR Zbl

[Farkas 2018] G. Farkas, "Effective divisors on Hurwitz spaces and moduli of curves", preprint, 2018, available at https://www.mathematik.hu-berlin.de/~farkas/divisor_m23.pdf.

[Farkas and Ortega 2011] G. Farkas and A. Ortega, "The maximal rank conjecture and rank two Brill–Noether theory", *Pure Appl. Math. Q.* **7**:4 (2011), 1265–1295. MR Zbl

[Farkas et al. 2020] G. Farkas, D. Jensen, and S. Payne, "The Kodaira dimensions of $\bar{\mathcal{M}}_{22}$ and $\bar{\mathcal{M}}_{23}$", preprint, 2020. arXiv 2005.00622

[Farkas et al. 2024] G. Farkas, D. Jensen, and S. Payne, "The nonabelian Brill–Noether divisor on $\bar{\mathcal{M}}_{13}$ and the Kodaira dimension of $\overline{\mathcal{R}}_{13}$", *Geom. Topol.* **28**:2 (2024), 803–866. MR Zbl

[Harris 1984] J. Harris, "On the Kodaira dimension of the moduli space of curves, II: The even-genus case", *Invent. Math.* **75**:3 (1984), 437–466. MR Zbl

[Harris and Mumford 1982] J. Harris and D. Mumford, "On the Kodaira dimension of the moduli space of curves", *Invent. Math.* **67**:1 (1982), 23–88. MR Zbl

[Jensen and Payne 2016] D. Jensen and S. Payne, "Tropical independence, II: The maximal rank conjecture for quadrics", *Algebra Number Theory* **10**:8 (2016), 1601–1640. MR Zbl

[Jensen and Payne 2018] D. Jensen and S. Payne, "On the strong maximal rank conjecture in genus 22 and 23", preprint, 2018. arXiv 1808.01285

[Larson 2017] E. Larson, "The maximal rank conjecture", preprint, 2017. arXiv 1711.04906

[Larson 2020] E. Larson, "The maximal rank conjecture for sections of curves", *J. Algebra* **555** (2020), 223–245. MR Zbl

[Liu et al. 2021] F. Liu, B. Osserman, M. Teixidor i Bigas, and N. Zhang, "Limit linear series and ranks of multiplication maps", *Trans. Amer. Math. Soc.* **374**:1 (2021), 367–405. MR Zbl

[Osserman 2006] B. Osserman, "A limit linear series moduli scheme", *Ann. Inst. Fourier* (*Grenoble*) **56**:4 (2006), 1165–1205. MR Zbl

[Osserman 2014] B. Osserman, "Limit linear series moduli stacks in higher rank", preprint, 2014. arXiv 1405.2937

[Teixidor i Bigas 2003] M. Teixidor i Bigas, "Injectivity of the symmetric map for line bundles", *Manuscripta Math.* **112**:4 (2003), 511–517. MR Zbl

[Teixidor i Bigas 2023] M. Teixidor i Bigas, "Ranks of maps of vector bundles", *Rev. Mat. Complut.* (online publication July 2023).

Communicated by Bjorn Poonen
Received 2019-10-08 Revised 2022-08-19 Accepted 2023-09-11

fuliu@math.ucdavis.edu *Department of Mathematics, University of California, Davis, Davis, CA, United States*

brian.osserman@gmail.com *Department of Mathematics, University of California, Davis, Davis, CA, United States*

mteixido@tufts.edu *Department of Mathematics, Tufts University, Medford, MA, United States*

zhangnaizhen@gmail.com *Fairleigh Dickinson University, Vancouver Campus, Vancouver, BC, Canada*

ALGEBRA AND NUMBER THEORY 18:8 (2024)
https://doi.org/10.2140/ant.2024.18.1465

Unramifiedness of weight 1 Hilbert Hecke algebras

Shaunak V. Deo, Mladen Dimitrov and Gabor Wiese

We prove that the Galois pseudo-representation valued in the mod p^n cuspidal Hecke algebra for GL(2) over a totally real number field F, of parallel weight 1 and level prime to p, is unramified at any place above p. The same is true for the noncuspidal Hecke algebra at places above p whose ramification index is not divisible by $p-1$. A novel geometric ingredient, which is also of independent interest, is the construction and study, in the case when p ramifies in F, of generalised Θ-operators using Reduzzi and Xiao's generalised Hasse invariants, including especially an injectivity criterion in terms of minimal weights.

Introduction

Starting with Wiles [1995] and Taylor and Wiles [1995], $R = \mathbb{T}$ theorems have been developed and taken a role as cornerstones in number theory. They provide both the existence of Galois representations with values in Hecke algebras satisfying prescribed local properties and modularity lifting theorems. The state of $R = \mathbb{T}$ theorems for 2-dimensional representations in residual characteristic p of the absolute Galois group $\mathrm{G}_{\mathbb{Q}}$ of $\mathbb{Q}$ and Hecke algebras acting on elliptic modular forms is quite satisfactory. In particular, the notoriously difficult case of Galois representations that are unramified at an odd prime p has been settled by ground-breaking work of Calegari and Geraghty [2018], in which they show that those correspond to modular forms of weight 1. More precisely, given an odd irreducible representation $\bar{\rho} : \mathrm{G}_{\mathbb{Q}} \to \mathrm{GL}_2(\bar{\mathbb{F}}_p)$ unramified outside a finite set of places S not containing p, they show that

$$R^S_{\mathbb{Q},\bar{\rho}} \xrightarrow{\sim} \mathbb{T}^{(1)}_{\bar{\rho}},$$

where $R^S_{\mathbb{Q},\bar{\rho}}$ is the universal deformation ring parametrising deformations of $\bar{\rho}$ which are unramified outside S and $\mathbb{T}^{(1)}_{\bar{\rho}}$ is the local component at $\bar{\rho}$ of a weight 1 Hecke algebra of a certain level prime to p.

MSC2020: primary 11F80; secondary 11F25, 11F33, 11F41, 11G18, 14G35.
Keywords: Hilbert modular forms, Galois representations, Hecke algebras, Theta-operators, weight one.

In this article, we address the corresponding question for parallel weight 1 Hilbert modular forms over a totally real field F of degree $d=[F:\mathbb{Q}]\geqslant 2$ and ring of integers $\mathfrak{o}$. We focus on the construction of the Galois (pseudo-)representation with values in the parallel weight 1 Hecke algebra with p-power torsion coefficients and proving its local ramification properties. In particular, given a finite set S of places in F relatively prime to p and a totally odd irreducible representation $\bar{\rho}:\mathrm{G}_F\to \mathrm{GL}_2(\bar{\mathbb{F}}_p)$ unramified outside S we show that there exists a surjective homomorphism

$$R^S_{F,\bar{\rho}}\twoheadrightarrow \mathbb{T}^{(1)}_{\bar{\rho}},$$

where $R^S_{F,\bar{\rho}}$ is the universal deformation ring parametrising deformations of $\bar{\rho}$ which are unramified outside S and $\mathbb{T}^{(1)}_{\bar{\rho}}$ is the local component at $\bar{\rho}$ of a weight 1 Hecke algebra of a certain level prime to p (see Corollary 3.10 for a precise statement).

Let $M_\kappa(\mathfrak{n},R)$ be the R-module of Hilbert modular forms of parallel weight $\kappa\geqslant 1$ and prime to p level $\mathfrak{n}$ over a $\mathbb{Z}_p$-algebra R, as in Definition 2.1. This R-module is equipped with a commuting family of Hecke operators $T_\mathfrak{q}$ as well as with diamond operators $\langle\mathfrak{q}\rangle$ for all primes $\mathfrak{q}$ of F not dividing $\mathfrak{n}$. Let $K/\mathbb{Q}_p$ be a finite extension containing the images of all embeddings of F in $\overline{\mathbb{Q}}_p$, and let $\mathcal{O}$ be its valuation ring, ϖ a uniformiser and $\mathbb{F}=\mathcal{O}/\varpi$ its residue field. We put $M_\kappa(\mathfrak{n},K/\mathcal{O})=\varinjlim_n M_\kappa(\mathfrak{n},\mathcal{O}/\varpi^n)$ and define the parallel weight 1 Hecke algebra

$$\mathbb{T}^{(1)}=\mathrm{im}(\mathcal{O}[T_\mathfrak{q},\langle\mathfrak{q}\rangle]_{\mathfrak{q}\nmid\mathfrak{n}p}\to \mathrm{End}_\mathcal{O}(M_1(\mathfrak{n},K/\mathcal{O}))),$$

as well as its cuspidal quotient $\mathbb{T}^{(1)}_{\mathrm{cusp}}$ acting faithfully on the submodule of parallel weight 1 cuspforms. We can now state the main results of this article. Let $p\mathfrak{o}=\prod_{\mathfrak{p}\mid p}\mathfrak{p}^{e_\mathfrak{p}}$ with $e_\mathfrak{p}\geqslant 1$. We emphasise that there is no restriction on the ramification of p in F.

Theorem 0.1. *There exists a $\mathbb{T}^{(1)}$-valued pseudo-representation $P^{(1)}$ of G_F of degree 2 which is unramified at all primes $\mathfrak{q}$ not dividing $\mathfrak{n}p$ and satisfies*

$$P^{(1)}(\mathrm{Frob}_\mathfrak{q})=(T_\mathfrak{q},\langle\mathfrak{q}\rangle).$$

Moreover, if $p-1$ does not divide $e_\mathfrak{p}$ for some $\mathfrak{p}\mid p$, then $P^{(1)}$ is also unramified at $\mathfrak{p}$ and satisfies

$$P^{(1)}(\mathrm{Frob}_\mathfrak{p})=(T_\mathfrak{p},\langle\mathfrak{p}\rangle),$$

in particular $T_\mathfrak{p}\in\mathbb{T}^{(1)}$.

Finally, the pseudo-representation $P^{(1)}_{\mathrm{cusp}}$ obtained after composing $P^{(1)}$ with the natural surjection $\mathbb{T}^{(1)}\to\mathbb{T}^{(1)}_{\mathrm{cusp}}$ is unramified at all $\mathfrak{p}\mid p$ and satisfies

$$P^{(1)}_{\mathrm{cusp}}(\mathrm{Frob}_\mathfrak{p})=(T_\mathfrak{p},\langle\mathfrak{p}\rangle).$$

The strategy of the proof is based on the *doubling* method developed in [Wiese 2014], further simplified and conceptualised in [Dimitrov and Wiese 2020] and [Calegari and Geraghty 2018]. The parallel weight 1 Hilbert modular forms over $\mathcal{O}/\varpi^n$ can be mapped into some higher weight in two ways, per prime $\mathfrak{p}$

dividing p, either by multiplication by a suitable power of the total Hasse invariant, or by applying a V-operator. That *doubling* map is used by Calegari and Specter [2019] to prove an analogue of Theorem 0.1 when $F = \mathbb{Q}$, for which they successfully develop the notion of a *p-ordinary* pseudo-representation. In that case, one knows by a result of Katz that the doubling map is injective. Furthermore, the existence of the Hecke operator T_p acting on weight 1 modular forms and the knowledge of its precise effect on the q-expansion (both due to Gross) allow one to show that the image of the doubling map is contained in the p-ordinary part of the higher weight space.

The existence of an optimally integral Hecke operator $T_\mathfrak{p}$ acting on parallel weight 1 Hilbert modular forms with arbitrary coefficients having the desired effect on their q-expansions (see [Diamond 2021] improving on and correcting previous works such as [Emerton et al. 2017] and [Dimitrov and Wiese 2020]) allows us to adapt the overall Calegari–Specter strategy to the Hilbert modular setting, while slightly generalising and clarifying some aspects of their arguments (see Section 3), the main challenge being to prove the injectivity of the doubling map. Note that the simple calculation in [Dimitrov and Wiese 2020] showing injectivity after restriction to an eigenspace is insufficient as the Hecke algebra modulo p need not be semisimple. Instead, we observe that the injectivity of the doubling map would follow from the injectivity of a certain generalised Θ-operator, introduced in the foundational work of Andreatta and Goren [2005] for Hilbert modular forms in characteristic p defined over the Deligne–Pappas moduli space. When p is unramified in F, the theory of partial Θ-operators was also developed by Diamond and Sasaki [2023] in a more general setting using a slightly different approach from that of [Andreatta and Goren 2005]. However, when p is ramified in F, the results of [Diamond and Sasaki 2023] do not apply, while those of [Andreatta and Goren 2005] are not sufficiently precise for our purposes, as the Hilbert modular forms defined over the Deligne–Pappas model "miss" some weights, and as a consequence the injectivity result of the latter paper is not optimal. In order to tackle this problem, we go back to the root of the problem and work with the Pappas–Rapoport moduli space, which does not miss any weight.

Capitalising on the theory of generalised Hasse invariants developed by Reduzzi and Xiao [2017] in this context, we carefully revisit [Andreatta and Goren 2005] and develop in Section 1E the needed theory of generalised Θ-operators over the Pappas–Rapoport moduli space and prove a refined injectivity criterion in terms of the minimal weights. In particular, we show that the generalised Θ-operators are indeed injective on parallel weight 1 Hilbert modular forms provided their weight is minimal at $\mathfrak{p}$. By the recent works of Diamond and Kassaei [2017; 2023] (see Section 1C) weight 1 Hilbert modular forms having "nonminimal" weight at $\mathfrak{p}$ could only possibly exist when $p-1$ divides $e_\mathfrak{p}$, and are products of forms of partial weight 0 at $\mathfrak{p}$ with generalised Hasse invariants.

In order to show the vanishing of the space of Katz cuspforms of partial weight 0 at $\mathfrak{p}$, and thus complete the proof of the last parts of the Theorem, in Section 1D we construct a partial Frobenius endomorphism $\Phi_{\mathfrak{p}^e}$ of this space and show that it is simultaneously injective and nilpotent. Our construction is inspired by the one in [Diamond and Sasaki 2023, Section 9.8] in the case when p is unramified in F. We also compute its effect on q-expansions, which is crucially used in our proof and, in order to avoid having to switch between different cusps, we only study the partial Frobenius operator of an appropriate power of $\mathfrak{p}$, rather than of $\mathfrak{p}$ itself.

In the language of linear representations, we prove the following result, which can be seen as a first step towards an $R = \mathbb{T}$ theorem.

Corollary (Corollary 3.10). *For every non-Eisenstein maximal ideal* $\mathfrak{m}$ *of* $\mathbb{T}^{(1)}$ (*see Definition 3.8*) *there exists a representation*

$$\rho_{\mathfrak{m}} : \mathrm{G}_F \to \mathrm{GL}_2(\mathbb{T}^{(1)}_{\mathfrak{m}}),$$

unramified at all primes $\mathfrak{q}$ *not dividing* $\mathfrak{n}$ *such that* $\mathrm{tr}(\rho_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = T_{\mathfrak{q}}$ *and* $\det(\rho_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = \langle \mathfrak{q} \rangle$.

We believe that our modest contribution to the theory of generalised Θ-operators in the setting of the Pappas–Rapoport splitting model is worthwhile on its own, beyond the application to our main theorem. On our way to the injectivity criterion, we also explore some related themes, such as the relation between Hilbert modular forms defined over the Pappas–Rapoport model with those defined over the Deligne–Pappas model, and the q-expansion and vanishing loci of the generalised Hasse invariants defined by Reduzzi and Xiao. We hope that it bridges the gap between many existing references in the literature and also clarifies some important aspects of the theory of mod p Hilbert modular forms. In the meantime, motivated by geometric Serre weight conjectures, Diamond [2023] extended the techniques of [Diamond and Sasaki 2023] to also construct partial Θ-operators which have an optimal effect on weights in the case where p ramifies in F. Moreover, Diamond [2023] generalised the construction of the partial Frobenius operators (our partial Frobenius operator $\Phi_{\mathfrak{p}^e}$ is essentially Diamond's $V^e_{\mathfrak{p}}$). Note that Diamond also describes kernels of partial Θ-operators in terms of images of his partial Frobenius maps $V_{\mathfrak{p}}$; see [loc. cit., Theorem 9.1.1]. However, our construction is less technical because we restrict to the Rapoport locus and we only consider the case of weights 0 at $\mathfrak{p}$.

Notation. Throughout the paper, we will use the following notation. We let F be a totally real number field of degree $d \geqslant 2$ and ring of integers $\mathfrak{o}$. We denote by $\overline{\mathbb{Q}} \subset \mathbb{C}$ the subfield of algebraic numbers and denote by $\mathrm{G}_F = \mathrm{Gal}(\overline{\mathbb{Q}}/F)$ the absolute Galois group of F. For every prime $\mathfrak{q}$ of F we denote by $\mathrm{Frob}_{\mathfrak{q}} \in \mathrm{G}_F$ a fixed choice of an arithmetic Frobenius at $\mathfrak{q}$. Let $\mathfrak{p}$ be a prime of F dividing p. Fixing an embedding ι_p of $\overline{\mathbb{Q}}$ into a fixed algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ allows one to see the absolute Galois group $\mathrm{G}_{F_{\mathfrak{p}}} = \mathrm{Gal}(\overline{\mathbb{Q}}_p/F_{\mathfrak{p}})$ of $F_{\mathfrak{p}}$ as a decomposition subgroup of G_F at $\mathfrak{p}$, and we let $\mathrm{I}_{\mathfrak{p}}$ denote its inertia subgroup. Furthermore, we fix a finite extension $K/\mathbb{Q}_p$ containing the images of all embeddings of F in $\overline{\mathbb{Q}}_p$, and let $\mathcal{O}$ be its valuation ring, ϖ a uniformiser and $\mathbb{F} = \mathcal{O}/(\varpi)$ its residue field.

For a prime $\mathfrak{p}$ of F dividing p, denote the residue field of $F_{\mathfrak{p}}$ by $\mathbb{F}_{\mathfrak{p}}$ and the ring of Witt vectors of $\mathbb{F}_{\mathfrak{p}}$ by $W(\mathbb{F}_{\mathfrak{p}})$. We also let $f_{\mathfrak{p}}$ and $e_{\mathfrak{p}}$ denote the residue and the ramification index of $\mathfrak{p}$, respectively. Let Σ be the set of infinite places of F, which we view as embeddings $F \hookrightarrow \overline{\mathbb{Q}}_p$ via ι_p. We have a natural partitioning $\Sigma = \coprod_{\mathfrak{p}|p} \Sigma_{\mathfrak{p}}$ where $\Sigma_{\mathfrak{p}}$ contains exactly those embeddings inducing the place $\mathfrak{p}$. For $\sigma \in \Sigma_{\mathfrak{p}}$, we denote by $\bar{\sigma}$ its restriction to the maximal unramified subfield of $F_{\mathfrak{p}}$ or, equivalently, the induced embedding of $\mathbb{F}_{\mathfrak{p}(\sigma)}$ into $\bar{\mathbb{F}}_p$. Furthermore, we let $\overline{\Sigma}_{\mathfrak{p}} = \{\bar{\sigma} \mid \sigma \in \Sigma_{\mathfrak{p}}\}$ and $\overline{\Sigma} = \{\bar{\sigma} \mid \sigma \in \Sigma\} = \coprod_{\mathfrak{p}|p} \overline{\Sigma}_{\mathfrak{p}}$. As a general rule, elements of Σ will be called σ whereas τ usually designates an element of $\overline{\Sigma}$. In both cases, $\mathfrak{p}(\sigma)$ and $\mathfrak{p}(\tau)$ denotes the underlying prime ideal. When either σ or τ is clear from the context, we will just denote this prime ideal by $\mathfrak{p}$. In particular, an element $\tau \in \overline{\Sigma}_{\mathfrak{p}}$ denotes both an embedding $\mathbb{F}_{\mathfrak{p}(\tau)} \hookrightarrow \mathbb{F}$ and the corresponding p-adic one $W(\mathbb{F}_{\mathfrak{p}(\tau)}) \hookrightarrow \mathcal{O}$. Denoting the absolute arithmetic Frobenius on $\mathbb{F}$ by ϕ, we have $\overline{\Sigma}_{\mathfrak{p}} = \{\phi^j \circ \tau \mid j \in \mathbb{Z}\} \simeq \mathbb{Z}/f_{\mathfrak{p}}\mathbb{Z} \simeq \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$ for any choice $\tau \in \overline{\Sigma}_{\mathfrak{p}}$. For any $\tau \in \overline{\Sigma}_{\mathfrak{p}}$, we let $\Sigma_\tau = \{\sigma \in \Sigma_{\mathfrak{p}} \mid \bar{\sigma} = \tau\} = \{\sigma_{\tau,i} \mid 1 \leqslant i \leqslant e_{\mathfrak{p}}\}$, where the numbering is chosen in an arbitrary, but fixed way. As an abbreviation, we write $\widetilde{\tau} = \sigma_{\tau,e_{\mathfrak{p}}}$.

Let $\mathcal{C}$ be a fixed set of representatives, all relatively prime to p, for the narrow class group of F.

1. Hilbert modular forms in finite characteristic

This section refines the theory of Θ-operators developed by Andreatta and Goren [2005], when p ramifies in F, in the setting of Hilbert modular forms defined over the Pappas–Rapoport splitting models for Hilbert modular varieties with the aim of proving the injectivity of the doubling map in Section 2. Along the way, we will need the generalised Hasse invariants of Reduzzi and Xiao [2017], results of Diamond and Kassaei [2017; 2023] about minimal weights as well as a partial Frobenius operator generalised from [Diamond and Sasaki 2023].

Throughout this section we fix an ideal $\mathfrak{n}$ of $\mathfrak{o}$ relatively prime to p and having a prime factor which does not divide $6\mathfrak{d}$, where $\mathfrak{d}$ denotes the different of F.

1A. ***Pappas–Rapoport splitting models for Hilbert modular varieties.*** Since we allow our base field F to ramify at p, we have to be careful with the model we choose for our Hilbert modular variety.

Fix $\mathfrak{c} \in \mathcal{C}$. We first consider the functor from the category of locally Noetherian $\mathbb{Z}_p$-schemes to the category of sets which assigns to a scheme S the set of isomorphism classes of tuples (A, λ, μ) where:

(i) A is a *Hilbert–Blumenthal abelian variety* (*HBAV*) over S, i.e., an abelian S-scheme of relative dimension d, together with a ring embedding $\mathfrak{o} \hookrightarrow \mathrm{End}_S(A)$.

(ii) λ is a $\mathfrak{c}$-polarisation of A/S, i.e., an isomorphism $\lambda : A^\vee \to A \otimes_{\mathfrak{o}} \mathfrak{c}$ of HBAV's over S such that the induced isomorphism $\mathrm{Hom}_{\mathfrak{o}}(A, A \otimes_{\mathfrak{o}} \mathfrak{c}) \simeq \mathrm{Hom}_{\mathfrak{o}}(A, A^\vee)$ sends elements of $\mathfrak{c}$ (resp. of the cone $\mathfrak{c}_+$ of its totally positive elements) to symmetric elements (resp. to polarisations),

(iii) μ is a $\mu_{\mathfrak{n}}$-level structure on A, i.e., an $\mathfrak{o}$-linear closed embedding of S-schemes $\mu : \mu_{\mathfrak{n}} \to A$, where $\mu_{\mathfrak{n}}$ denotes the Cartier dual of the constant group scheme $\mathfrak{o}/\mathfrak{n}$ over S.

Under our assumption on $\mathfrak{n}$ above, this functor is representable by a $\mathbb{Z}_p$-scheme $\mathcal{X}^{\mathrm{DP}}$ of finite type, called the *Deligne–Pappas moduli space*; see [Andreatta and Goren 2005, Remark 3.3] and [Dimitrov and Tilouine 2004, Lemma 1.4].

Suppose now that A is an HBAV over a locally Noetherian $\mathcal{O}$-scheme S with structure map $s : A \to S$ and let $\Omega^1_{A/S}$ be the sheaf of relative differentials of A over S. Define

$$\omega_S = s_* \Omega^1_{A/S},$$

i.e., ω_S is the sheaf of invariant differentials of A over S. Consider the decomposition

$$\mathfrak{o}\otimes_{\mathbb{Z}}\mathcal{O}_S=(\mathfrak{o}\otimes_{\mathbb{Z}}\mathbb{Z}_p)\otimes_{\mathbb{Z}_p}\mathcal{O}_S=\prod_{\mathfrak{p}\mid p}\mathfrak{o}_{\mathfrak{p}}\otimes_{\mathbb{Z}_p}\mathcal{O}_S=\prod_{\tau\in\overline{\Sigma}}\mathfrak{o}_{\mathfrak{p}(\tau)}\otimes_{W(\mathbb{F}_{\mathfrak{p}(\tau)}),\tau}\mathcal{O}_S. \tag{1}$$

It implies that we have a corresponding decomposition

$$\omega_S=\bigoplus_{\tau\in\overline{\Sigma}}\omega_{S,\tau}. \tag{2}$$

The sheaf $\omega_{S,\tau}$ is locally free over S of rank $e_{\mathfrak{p}(\tau)}$; see [Reduzzi and Xiao 2017, Section 2.2]. Note that on $\omega_{S,\tau}$, the action of $W(\mathbb{F}_{\mathfrak{p}(\tau)})\subset\mathfrak{o}_{\mathfrak{p}(\tau)}$ is via τ. Fix a uniformiser $\varpi_{\mathfrak{p}(\tau)}$ of $\mathfrak{o}_{\mathfrak{p}(\tau)}$. From the product decomposition above, we get an action of $\mathfrak{o}_{\mathfrak{p}(\tau)}$ on $\omega_{S,\tau}$. Denote the action of $\varpi_{\mathfrak{p}(\tau)}$ on $\omega_{S,\tau}$ by $[\varpi_{\mathfrak{p}(\tau)}]$.

We are now ready to present the Pappas–Rapoport model. Consider the functor from the category of locally Noetherian $\mathcal{O}$-schemes to the category of sets which assigns to a scheme S the set of isomorphism classes of tuples $(A,\lambda,\mu,(\mathcal{F}_{\mathfrak{p}})_{\mathfrak{p}\mid p})$ where (A,λ,μ) is as above and for all $\mathfrak{p}\mid p$, $\mathcal{F}_{\mathfrak{p}}$ is a collection $(\mathcal{F}_\tau^i)_{\tau\in\overline{\Sigma}_{\mathfrak{p}},0\leqslant i\leqslant e_{\mathfrak{p}}}$ of $\mathfrak{o}\otimes\mathcal{O}_S$-modules, which are locally free as $\mathcal{O}_S$-modules, such that:

- $0=\mathcal{F}_\tau^0\subset\cdots\subset\mathcal{F}_\tau^{e_{\mathfrak{p}}}=\omega_{S,\tau}$.
- For any $\sigma=\sigma_{\tau,i}\in\Sigma_\tau$, the $\mathcal{O}_S$-module $\omega_{S,\tau,i}=\omega_{S,\sigma}=\mathcal{F}_\tau^i/\mathcal{F}_\tau^{i-1}$ is locally free of rank 1 and annihilated by $[\varpi_{\mathfrak{p}}]-\sigma(\varpi_{\mathfrak{p}})$. Note that the numbering here depends on the one for Σ_τ.

This functor is representable by a smooth $\mathcal{O}$-scheme $\mathcal{X}$ of finite type called the *Pappas–Rapoport moduli space*; see [Reduzzi and Xiao 2017, Proposition 2.4] and [Dimitrov and Tilouine 2004, Lemma 1.4].

In order to better understand the relation between the Deligne–Pappas and the Pappas–Rapoport moduli spaces, we recall that the *Rapoport locus* $\mathcal{X}^{\mathrm{Ra}}$ is the open subscheme of $\mathcal{X}^{\mathrm{DP}}$ classifying HBAV's $s:A\to S$ satisfying the following condition introduced by Rapoport: $s_*\Omega^1_{A/S}$ is a locally free $\mathfrak{o}\otimes_{\mathbb{Z}}\mathcal{O}_S$-module of rank 1. Then $\mathcal{X}^{\mathrm{Ra}}$ is the smooth locus of $\mathcal{X}^{\mathrm{DP}}$ and its complement is supported in the special fibre and has codimension at least 2 in it. The forgetful map $\mathcal{X}\to\mathcal{X}^{\mathrm{DP}}_{\mathcal{O}}$ induces an isomorphism on the open subscheme $\mathcal{X}^{\mathrm{Ra}}_{\mathcal{O}}$; see [Reduzzi and Xiao 2017, Proposition 2.4]. If p is unramified in F, the different schemes agree: $\mathcal{X}=\mathcal{X}^{\mathrm{Ra}}_{\mathcal{O}}=\mathcal{X}^{\mathrm{DP}}_{\mathcal{O}}$; see [Reduzzi and Xiao 2017, Section 1].

Let $\mathcal{A}$ be the universal abelian scheme over $\mathcal{X}$ with structure morphism $s:\mathcal{A}\to\mathcal{X}$. Let $\omega_{\mathcal{X}}=s_*\Omega^1_{\mathcal{A}/\mathcal{X}}$. Note that the restriction of $\omega_{\mathcal{X}}$ to $\mathcal{X}^{\mathrm{Ra}}_{\mathcal{O}}$ is a locally free sheaf of rank 1 over $\mathfrak{o}\otimes_{\mathbb{Z}}\mathcal{O}_{\mathcal{X}^{\mathrm{Ra}}_{\mathcal{O}}}$. As abbreviation we write $\omega,\omega_\tau,\omega_{\tau,i},\omega_\sigma$ for $\omega_{\mathcal{X}},\omega_{\mathcal{X},\tau},\omega_{\mathcal{X},\tau,i},\omega_{\mathcal{X},\sigma}$. In particular, for each $\tau\in\overline{\Sigma}$, the sheaf ω_τ is equipped with a filtration the graded pieces of which are the invertible sheaves ω_σ for $\sigma\in\Sigma_\tau$. In [Reduzzi and Xiao 2017] this is referred to as the *universal filtration*. We point out explicitly that the last graded piece $\omega_{\tilde{\tau}}$ is a quotient of ω_τ.

Next we give, following Katz, a geometric definition of the space of Hilbert modular forms.

Definition 1.1. A *Katz Hilbert modular form* of weight $k=\sum_{\sigma\in\Sigma}k_\sigma\sigma\in\mathbb{Z}[\Sigma]$, level $\mathfrak{n}$ and coefficients in an $\mathcal{O}$-algebra R is a global section of the line bundle $\omega^{\otimes k}=\bigotimes_{\sigma\in\Sigma}\omega_\sigma^{\otimes k_\sigma}$ over $\mathcal{X}\times_{\mathcal{O}}R$. We will denote by $M_k^{\mathrm{Katz}}(\mathfrak{c},\mathfrak{n};R)$ the corresponding R-module.

Its R-submodule of cuspforms $S_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; R)$ consists of those Katz Hilbert modular forms that vanish along the cuspidal divisor of any toroidal compactification of $\mathcal{X} \times_{\mathcal{O}} R$; see [Reduzzi and Xiao 2017, Section 2.11].

As $\mathcal{X}$ admits toroidal compactifications (see [Reduzzi and Xiao 2017, Section 2.11]) which are smooth and proper over $\mathcal{O}$ and to which ω_σ extends for all $\sigma \in \Sigma$, the Koecher principle implies, in view of [Stacks 2005–, Tag 02O5], that $M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; R)$ is a finitely generated R-module.

Remark 1.2. When the weight $k \in \mathbb{Z}[\Sigma]$ is *parallel*, i.e., $k_\sigma = \kappa \in \mathbb{Z}$ for all $\sigma \in \Sigma$, one also could define a *Katz Hilbert modular form* of parallel weight $\kappa \in \mathbb{Z}$, level $\mathfrak{n}$ and coefficients in a $\mathbb{Z}_p$-algebra R as a global section of the line bundle $\left(\bigwedge^d s_*\Omega^1_{\mathcal{A}/\mathcal{X}^{\mathrm{DP}}}\right)^{\otimes\kappa}$ over $\mathcal{X}^{\mathrm{DP}} \times_{\mathbb{Z}_p} R$. By Zariski's main theorem applied to the proper birational map $\mathcal{X} \to \mathcal{X}^{\mathrm{DP}}_{\mathcal{O}}$ between normal varieties, this would lead to the same space as in Definition 1.1.

1B. ***Generalised Hasse invariants.*** From this point onwards we will work over $\mathbb{F}$. Let X be the Pappas–Rapoport moduli space over $\mathbb{F}$, i.e., the special fibre $\mathcal{X} \times_{\mathcal{O}} \mathbb{F}$ of $\mathcal{X}$. There is a natural morphism $X \to \mathcal{X}^{\mathrm{DP}} \times_{\mathbb{Z}_p} \mathbb{F}$ obtained by forgetting the filtrations. Let $X^{\mathrm{Ra}} = \mathcal{X}^{\mathrm{Ra}} \times_{\mathbb{Z}_p} \mathbb{F}$. We have the equality

$$\mathfrak{o} \otimes_{\mathbb{Z}} \mathbb{F} = \prod_{\mathfrak{p}|p} \mathfrak{o}_{\mathfrak{p}} \otimes_{\mathbb{Z}_p} \mathbb{F} \simeq \prod_{\tau \in \overline{\Sigma}} \mathfrak{o}_{\mathfrak{p}(\tau)} \otimes_{W(\mathbb{F}_{\mathfrak{p}(\tau)}),\tau} \mathbb{F} = \prod_{\tau \in \overline{\Sigma}} \mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}}), \tag{3}$$

coming from (1). Note that the last equality of (3) depends on the choice of the uniformiser $\varpi_{\mathfrak{p}(\tau)}$ of $\mathfrak{o}_{\mathfrak{p}(\tau)}$, made in the previous subsection for every $\tau \in \overline{\Sigma}$, and allows us to view ω_τ as an $\mathcal{O}_X[x]/(x^{e_{\mathfrak{p}}})$-module. If S is a locally Noetherian $\mathbb{F}$-scheme and A is an HBAV over S satisfying the Rapoport condition, then $\omega_{S,\tau}$ is a locally free $\mathcal{O}_S[x]/(x^{e_{\mathfrak{p}(\tau)}})$-module of rank 1. Hence, there is a unique filtration on $\omega_{S,\tau}$ satisfying the Pappas–Rapoport conditions given by $x^{e_{\mathfrak{p}(\tau)}-i}\omega_{S,\tau}$ for $0 \leqslant i \leqslant e_{\mathfrak{p}(\tau)}$. We point out again that the definition of $\mathcal{X}$ depends on the numbering of the embeddings in Σ_τ fixed above, but that X is independent of any such choice; see also [Reduzzi and Xiao 2017, Remark 2.3].

If $\mathfrak{p} \mid p$ and $\tau \in \overline{\Sigma}_{\mathfrak{p}}$, then suppose the universal filtration on ω_τ is given by $(\mathcal{F}_\tau^i)_{0 \leqslant i \leqslant e_{\mathfrak{p}}}$. We now recall Reduzzi and Xiao's constructions of generalised Hasse invariants h_σ given in [Reduzzi and Xiao 2017]. Let $\mathfrak{p} \mid p$ and $\tau \in \overline{\Sigma}_{\mathfrak{p}}$ and assume first that $2 \leqslant i \leqslant e_{\mathfrak{p}}$. There is a map $\mathcal{F}_\tau^i \to \mathcal{F}_\tau^{i-1}$ which sends a local section z of $\mathcal{F}_\tau^i$ to the section $x \cdot z$ of $\mathcal{F}_\tau^{i-1}$, where the action of x is given by $[\varpi_{\mathfrak{p}(\tau)}]$. Hence, we get a map $\mathcal{F}_\tau^i/\mathcal{F}_\tau^{i-1} \to \mathcal{F}_\tau^{i-1}/\mathcal{F}_\tau^{i-2}$ inducing a section $h_{\tau,i} = h_{\sigma_{\tau,i}}$ of $\omega_{\tau,i-1} \otimes \omega_{\tau,i}^{-1}$ over X. This h_σ is the *generalised Hasse invariant at* $\sigma = \sigma_{\tau,i}$; see [Reduzzi and Xiao 2017, Construction 3.3] and [Emerton et al. 2017, Section 2.11] for more details. As $(\omega_\tau)_{|X^{\mathrm{Ra}}}$ is a locally free sheaf over $\mathcal{O}_{X^{\mathrm{Ra}}}[x]/(x^{e_{\mathfrak{p}}})$ of rank 1, we have $(\mathcal{F}_\tau^i)_{|X^{\mathrm{Ra}}} = (x^{e_{\mathfrak{p}}-i}\omega_\tau)_{|X^{\mathrm{Ra}}}$. It follows that $h_{\tau,i}$ is a nowhere vanishing section over X^{Ra} and multiplication by $h_{\tau,i}$ induces an isomorphism between $(\omega_{\tau,i})_{|X^{\mathrm{Ra}}}$ and $(\omega_{\tau,i-1})_{|X^{\mathrm{Ra}}}$.

For the case $i = 1$, the *generalised Hasse invariant* $h_{\tau,1}$ is defined as a global section over X of $\omega_{\phi^{-1}\circ\tau,e_{\mathfrak{p}}}^{\otimes p} \otimes \omega_{\tau,1}^{\otimes -1}$; see [Reduzzi and Xiao 2017, Construction 3.6] for more details. We let $h_\tau = \prod_{\sigma\in\Sigma_\tau} h_\sigma = \prod_{i=1}^{e_{\mathfrak{p}}} h_{\tau,i}$. It is a modular form of weight $p \cdot \widetilde{\phi^{-1}\circ\tau} - \widetilde{\tau}$.

Remark 1.3. Let A be the universal abelian scheme over X and $\mathrm{Ver} : A^{(p)} \to A$ be the Verschiebung morphism, where $A^{(p)} = A \times_{\mathbb{F},\phi} \mathbb{F}$. It induces maps $\omega_\tau \to \omega^{(p)}_{\phi^{-1}\circ\tau}$ and further $\mathcal{F}_\tau^{e_\mathfrak{p}}/\mathcal{F}_\tau^{e_\mathfrak{p}-1} \to (\mathcal{F}^{e_\mathfrak{p}}_{\phi^{-1}\circ\tau}/\mathcal{F}^{e_\mathfrak{p}-1}_{\phi^{-1}\circ\tau})^{(p)}$. Note that $\mathcal{F}_\tau^{e_\mathfrak{p}}/\mathcal{F}_\tau^{e_\mathfrak{p}-1} = \omega_{\tau,e_\mathfrak{p}}$, $(\mathcal{F}^{e_\mathfrak{p}}_{\phi^{-1}\circ\tau}/\mathcal{F}^{e_\mathfrak{p}-1}_{\phi^{-1}\circ\tau})^{(p)} = \omega^{\otimes p}_{\phi^{-1}\circ\tau,e_\mathfrak{p}}$ and the resulting section of $\omega^{\otimes p}_{\phi^{-1}\circ\tau,e_\mathfrak{p}} \otimes \omega^{\otimes -1}_{\tau,e_\mathfrak{p}}$ over X is precisely given by h_τ; see [Reduzzi and Xiao 2017, Lemma 3.8]. Moreover, its restriction to X^{Ra} coincides with Andreatta and Goren's *partial Hasse invariant* constructed in [Andreatta and Goren 2005, Definition 7.12]. In particular, when p is unramified in F, the generalised Hasse invariants constructed by Reduzzi and Xiao are the same as the partial Hasse invariants constructed by Andreatta and Goren.

We will now determine the geometric q-expansions of these generalised Hasse invariants. We will mostly follow conventions of [Dimitrov 2004, Section 8]. Let $\infty_\mathfrak{c}$ be the standard infinity cusp whose Tate object is given by $(\mathbb{G}_m \otimes_\mathbb{Z} \mathfrak{c}^*)/q^\mathfrak{o}$; see [Dimitrov and Wiese 2020, Section 2.3]. Here $\mathfrak{c}^* = \mathfrak{c}^{-1}\mathfrak{d}^{-1}$. Let $X^\wedge$ be the formal completion of a toroidal compactification of X along the divisor at the cusp $\infty_\mathfrak{c}$; see [Dimitrov 2004, Theorem 8.6]. By [loc. cit.], the pull back of ω to $X^\wedge$ is canonically isomorphic to $\mathcal{O}_{X^\wedge} \otimes \mathfrak{c}$. Choosing an identification

$$\mathbb{F} \otimes \mathfrak{c} \xrightarrow{\sim} \mathbb{F} \otimes \mathfrak{o} \tag{4}$$

one can canonically identify $\omega_{\tau\,|\,X^\wedge}$ with $\tau(\mathcal{O}_{X^\wedge} \otimes \mathfrak{o}) = \mathcal{O}_{X^\wedge}[x]/(x^{e_{\mathfrak{p}(\tau)}})$ (see (3)). A global section of ω_τ over $X^\wedge$ is an element of

$$\Bigl\{ \sum_{\xi \in \mathfrak{c}_+ \cup \{0\}} a_\xi q^\xi \;\Big|\; a_\xi \in \mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}}) \text{ and } a_{u^2\xi} = \tau(u) a_\xi,\, \forall u \in \mathfrak{o}^\times,\, u-1 \in \mathfrak{n} \Bigr\},$$

whereas a section z of $\omega_{\tau,i}$ over $X^\wedge$ is an element of

$$\Bigl\{ x^{e_{\mathfrak{p}(\tau)}-i} \cdot \sum_{\xi \in \mathfrak{c}_+ \cup \{0\}} b_\xi q^\xi \;\Big|\; b_\xi \in \mathbb{F} \text{ and } b_{u^2\xi} = \tau(u) b_\xi,\, \forall u \in \mathfrak{o}^\times,\, u-1 \in \mathfrak{n} \Bigr\}$$

whose q-expansion is given by $\sum_{\xi \in \mathfrak{c}_+ \cup \{0\}} b_\xi q^\xi$ with respect to the choice of basis of $\omega_{\tau,i\,|\,X^\wedge}$ corresponding to $x^{e_{\mathfrak{p}(\tau)}-i}$.

Lemma 1.4. *Let $\mathfrak{p} \mid p$, $\tau \in \overline{\Sigma}_\mathfrak{p}$. Then for every $1 \leqslant i \leqslant e_\mathfrak{p}$, the geometric q-expansion of the generalised Hasse invariant $h_{\tau,i}$ at $\infty_\mathfrak{c}$ is* 1. *In particular, it does not vanish at any cusp.*

Proof. When $i > 1$, as $x \cdot z$ is a section of $\omega_{\tau,i-1}$ having by definition the same q-expansion, one concludes that $h_{\tau,i}$ has q-expansion 1, thus proving the claim in that case. In the remaining case of $i = 1$, we observe that the q-expansion of $h_\tau = \prod_{i=1}^{e_\mathfrak{p}(\tau)} h_{\tau,i}$ at $\infty_\mathfrak{c}$ is 1 by Remark 1.3 and [Andreatta and Goren 2005, Proposition 7.14]. Hence the q-expansion of $h_{\tau,1}$ at $\infty_\mathfrak{c}$ is 1. Finally, since the $h_{\tau,i}$ can be defined in any level, we deduce their nonvanishing at all cusps from the nonvanishing at $\infty_\mathfrak{c}$. □

We now collect some properties of the generalised Hasse invariants that will be used in the sequel. Let $Z_\sigma \subset X$ be the divisor of h_σ and, in order to shorten the notation, we let $Z_{\tau,i} = Z_{\sigma_{\tau,i}}$.

Lemma 1.5. *The complement of X^{Ra} in X coincides with $\bigcup_{\tau\in\overline{\Sigma}}\bigcup_{i=2}^{e_{\mathfrak{p}(\tau)}} Z_{\tau,i}$. Moreover, for any $I\subseteq\Sigma$, the intersection $\bigcap_{\sigma\in I} Z_\sigma$ is, either empty, or equidimensional of dimension $d-|I|$. In particular, the zero loci of two different generalised Hasse invariants do not have a common divisor.*

Proof. The first claim has been established in [Emerton et al. 2017, Proposition 2.13 (2)]. For the second, if $\cap_{\sigma\in I} Z_\sigma$ is nonempty, then the tangent space computation in [Reduzzi and Xiao 2017, Theorem 3.10] ensures the correct dimension. □

Remark 1.6. Diamond and Kassaei also prove Lemma 1.5 and obtain in addition the nonemptiness of the intersection; see [Diamond and Kassaei 2023, Proposition 5.8]. Here we sketch a constructive proof, following ideas of Andreatta and Goren [2003], if $e_{\mathfrak{p}(\tau)}$ is odd for all $\tau\in\overline{\Sigma}$.

Let $A=E\otimes_{\mathbb{Z}}\mathfrak{o}^*$, where E is a supersingular elliptic curve over $\mathbb{F}$. We see, as in [Andreatta and Goren 2003, Proof of Theorem 10.1], that $\omega_{A,\tau}\simeq\mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}})$ for all $\tau\in\overline{\Sigma}$. Let $\mathrm{Frob}_A : A\to A^{(p)}$ be the Frobenius map and $H=\ker(\mathrm{Frob}_A)[\prod_{\mathfrak{p}|p}\mathfrak{p}^{[e_\mathfrak{p}/2]}]$. By imitating the calculations of [Andreatta and Goren 2003, Section 8] (more specifically [Andreatta and Goren 2003, Proposition 6.5, Lemmas 8.6, 8.9, Proposition 8.10]), one sees that if $A^{(1)}=A/H$, then

$$\omega_{A^{(1)},\tau}\simeq x^{[e_{\mathfrak{p}(\tau)}/2]}\cdot\mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}})\bigoplus x^{e_{\mathfrak{p}(\tau)}-[e_{\mathfrak{p}(\tau)}/2]}\cdot\mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}})\quad\text{for all }\tau\in\overline{\Sigma}. \tag{5}$$

Note that $A^{(1)}$ is a $\mathfrak{c}'$-polarised HBAV over $\mathbb{F}$ for some $\mathfrak{c}'\in\mathcal{C}$. Let $\mathfrak{a}\subset\mathfrak{o}$ be an ideal relatively prime to p such that $\mathfrak{a}\mathfrak{c}'$ and $\mathfrak{c}$ represent the same element in the narrow class group of F. Let $H^{(1)}$ be an $\mathfrak{o}$-invariant subgroup scheme of $A^{(1)}[\mathfrak{a}]$ isomorphic to $\mathfrak{o}/\mathfrak{a}$ and let $A^{(2)}=A^{(1)}/H^{(1)}$. By [Kisin and Lai 2005, Section 1.9], $A^{(2)}$ is a $\mathfrak{c}$-polarised HBAV over $\mathbb{F}$ and since $\mathfrak{a}$ is relatively prime to p, we have $\omega_{A^{(2)}}=\omega_{A^{(1)}}$. Endowing each $\omega_{A^{(2)},\tau}$ with the "alternating" filtration between the two summands in (5) yields a point in $\bigcap_{\tau\in\overline{\Sigma}}\bigcap_{i=2}^{e_{\mathfrak{p}(\tau)}} Z_{\tau,i}$, showing that the latter is nonempty.

If $e_{\mathfrak{p}(\tau)}$ is odd, then the filtration on $\omega_{A^{(2)},\tau}$ described above is unique. Moreover, as $A^{(2)}$ is supersingular (i.e., its p-torsion subgroup has no étale component), the map $\omega_{A^{(2)},\tau}\to\omega_{A^{(2)},\phi^{-1}\circ\tau}$ induced by the Verschiebung morphism is the zero map. Hence, we conclude, using the structure of $\omega_{A^{(2)},\tau}$ and the definition of the Hasse invariant $h_{\tau,1}$, that any such point also lies in $Z_{\tau,1}$. Thus, if $e_{\mathfrak{p}(\tau)}$ is odd for all $\tau\in\overline{\Sigma}$, then we get a point in $\bigcap_{\tau\in\overline{\Sigma}}\bigcap_{i=1}^{e_{\mathfrak{p}(\tau)}} Z_{\tau,i}$.

We illustrate the weights of the generalised and partial Hasse invariants in Table 1, where we let $\tau\in\overline{\Sigma}$ and write $e=e_{\mathfrak{p}(\tau)}$ as abbreviation.

One of the advantages of Definition 1.1 is that it allows us to define mod p Hilbert modular forms in *any* weight $k=\sum_{\sigma\in\Sigma}k_\sigma\sigma\in\mathbb{Z}[\Sigma]$, while the definition in [Andreatta and Goren 2005] was *missing* some weights when p ramifies in F, namely theirs are indexed by $\overline{\Sigma}$, instead of Σ. Indeed, the space of modular forms introduced by Andreatta and Goren [2005, Propisition 5.5] is

$$M_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c},\mathfrak{n};\mathbb{F})=\mathrm{H}^0\Bigl(X^{\mathrm{Ra}},\bigotimes_{\tau\in\overline{\Sigma}}\omega_{\tilde{\tau}}^{k_\tau}\Bigr),\quad\text{where }\bar{k}=\sum_{\tau\in\overline{\Sigma}}k_\tau\tau\in\mathbb{Z}[\overline{\Sigma}]. \tag{6}$$

	weights										
	$\phi^{-1}\circ\tau$			τ					$\phi\circ\tau$		
	$\cdots$	$e-1$	e	1	2	$\cdots$	$e-1$	e	1	2	$\cdots$
$h_{\phi^{-1}\circ\tau,e}$		1	-1								
$h_{\tau,1}$			p	-1							
$h_{\tau,2}$				1	-1						
$\vdots$					$\ddots$	$\ddots$					
$h_{\tau,e-1}$						1	-1				
$h_{\tau,e}$							1	-1			
$h_{\phi\circ\tau,1}$								p	-1		
$h_{\phi\circ\tau,2}$									1	-1	
h_τ			p					-1			

Table 1. Weights of Hasse invariants.

We will denote by $S^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ the subspace of $M^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ consisting of cuspforms, which are defined as modular forms such that the constant coefficient of the q-expansion at every cusp vanishes. If $k=\sum_{\sigma\in\Sigma}k_\sigma\sigma\in\mathbb{Z}[\Sigma]$, then for every $\tau\in\overline{\Sigma}$, let $k_\tau=\sum_{\sigma\in\Sigma_\tau}k_\sigma$ and define $\bar{k}:=\sum_{\tau\in\overline{\Sigma}}k_\tau\tau\in\mathbb{Z}[\overline{\Sigma}]$. We let

$$H^{\mathrm{RX}}_k=\prod_{\tau\in\overline{\Sigma}}\prod_{i=2}^{e_{\mathfrak{p}(\tau)}}h_{\tau,i}^{\sum_{j=1}^{i-1}k_{\tau,j}},\tag{7}$$

where $k_{\tau,j}=k_{\sigma_{\tau,j}}$. In view of the table of weights of the generalised Hasse invariants, for every $\tau\in\overline{\Sigma}$, the (τ,i)-component of the weight of f/H^{RX}_k is 0 if $1\leqslant i\leqslant e_{\mathfrak{p}(\tau)}-1$ and the $\widetilde{\tau}=(\tau,e_{\mathfrak{p}(\tau)})$-component is k_τ. Since H^{RX}_k is invertible on X^{Ra}, we obtain the following result.

Lemma 1.7. *The restriction from X to X^{Ra} yields an injection of $M^{\mathrm{Katz}}_k(\mathfrak{c},\mathfrak{n};\mathbb{F})$ into $M^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ sending f to f/H^{RX}_k.*

A converse is described in Lemma 1.12 below.

1C. ***Minimal weights.*** We recall the notion of *minimal weight* of a mod p Hilbert modular form.

Definition 1.8. We define the *minimal weight* of $0\neq f\in M^{\mathrm{Katz}}_k(\mathfrak{c},\mathfrak{n};\mathbb{F})$ to be the unique weight k' such that $f=g\cdot\prod_{\sigma\in\Sigma}h_\sigma^{n_\sigma}$, where $g\in M^{\mathrm{Katz}}_{k'}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ and the integers $(n_\sigma)_{\sigma\in\Sigma}$ are as large as possible.

Lemma 1.9. *The notion of minimal weight is well defined.*

Proof. First note that Z_σ is nonempty for every $\sigma\in\Sigma$. Indeed, this follows from [Diamond and Kassaei 2023, Corollary 5.7]. Alternatively, we have shown in Remark 1.6 that $Z_{\tau,i}$ is nonempty for every $\tau\in\overline{\Sigma}$ and $2\leq i\leq e_{\mathfrak{p}(\tau)}$. Moreover, it is well known that the zero locus of $h_\tau=\prod_{i=1}^{e_{\mathfrak{p}(\tau)}}h_{\tau,i}$ in X^{Ra} is nonempty for every $\tau\in\overline{\Sigma}$; see [Andreatta and Goren 2005, Corollary 8.18]. By Lemma 1.5, the Hasse invariants

$h_{\tau,i}$ with $\tau \in \overline{\Sigma}$ and $2 \leq i \leq e_{\mathfrak{p}(\tau)}$ are invertible on X^{Ra}. Therefore, it follows that the divisor $Z_{\tau,1}$ is nonempty for every $\tau \in \overline{\Sigma}$.

Recall from Lemma 1.5 that the zero loci of two different generalised Hasse invariants do not have a common divisor. Let j_σ be the order of vanishing of a Hilbert modular form $f \neq 0$ on Z_σ. So, if we divide f by $\prod_{\sigma \in \Sigma} h_\sigma^{j_\sigma}$, we get the modular form g needed in Definition 1.8. Hence, it follows that the notion of minimal weight is indeed well defined; see also the proof of [Andreatta and Goren 2005, Theorem 8.19] and [Diamond and Kassaei 2023, Section 8]. □

Remark 1.10. When p is unramified in F (i.e., $\Sigma = \overline{\Sigma}$), the notion of minimal weights from Definition 1.8 is the same as the one introduced by Andreatta and Goren [2005, Section 8.20]. On the other hand, when $\mathfrak{p}$ is ramified, multiplying $0 \neq f \in M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ having minimal weight k with arbitrary powers of generalised Hasse invariants ($h_{\tau,i}$ with $2 \leqslant i \leqslant e_{\mathfrak{p}}$) yields forms sharing the same $\bar{k}$ but whose weights are not minimal anymore.

Diamond and Kassaei [2017; 2023] define the *minimal cone* by

$$C^{\min} = \left\{ \sum_{\tau \in \overline{\Sigma}} \sum_{i=1}^{e_{\mathfrak{p}(\tau)}} k_{\tau,i}\sigma_{\tau,i} \in \mathbb{Q}[\Sigma] \;\middle|\; \forall \tau \in \overline{\Sigma}, \forall 1 \leqslant i < e_{\mathfrak{p}(\tau)}, k_{\tau,i+1} \geqslant k_{\tau,i},\, pk_{\tau,1} \geqslant k_{\phi^{-1}\circ\tau, e_{\mathfrak{p}(\tau)}} \right\}.$$

Regarding the minimal weights for Hilbert modular forms, Diamond and Kassaei prove the following result in [Diamond and Kassaei 2017, Corollary 5.3], when p is unramified in F, and in [Diamond and Kassaei 2023, Corollary 8.2], when p is ramified in F.

Proposition 1.11 (Diamond and Kassaei). *The minimal weight of* $0 \neq f \in M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ *belongs to* $C^{\min}$.

The minimal weights allow us to further elaborate on the relation between the modular forms defined by Andreatta and Goren [2005] and those of Definition 1.1.

Lemma 1.12. *Let* $\bar{k} \in \mathbb{Z}[\overline{\Sigma}]$. *There is a finite subset* $K \subset C^{\min}$ *such that for every* $f \in M_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$, *there is* $k' \in K$, *a modular form* $g \in M_{k'}^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ *and a product of generalised Hasse invariants* $H = \prod_{\tau \in \overline{\Sigma}} \prod_{i=1}^{e_{\mathfrak{p}(\tau)}} h_{\tau,i}^{j_{\tau,i}}$ *with* $j_{\tau,i} \in \mathbb{Z}$ *and* $j_{\tau,1} \geqslant 0$, *such that the restriction to* X^{Ra} *of* $g \cdot H$ *equals* f. *In particular,* f *and* g *have the same geometric* q*-expansion at the cusp* $\infty_{\mathfrak{c}}$.

Proof. The result is trivial for $f = 0$. Seeing $0 \neq f \in M_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ as a meromorphic section of the line bundle $\bigotimes_{\tau \in \overline{\Sigma}} \omega_{\tilde{\tau}}^{\otimes k_\tau}$ over X, we let $j_{\tau,i} \in \mathbb{Z}$ be the order of vanishing of f along the divisor $Z_{\tau,i}$ defined by the Hasse invariant $h_{\tau,i}$ for $\tau \in \overline{\Sigma}$ and $1 \leqslant i \leqslant e_{\mathfrak{p}(\tau)}$. As f is holomorphic on X^{Ra}, which intersects every irreducible component of $Z_{\tau,1}$ nontrivially by Lemma 1.5, we deduce that $j_{\tau,1} \geqslant 0$. Dividing f by $H = \prod_{\tau \in \overline{\Sigma}} \prod_{i=1}^{e_{\mathfrak{p}(\tau)}} h_{\tau,i}^{j_{\tau,i}}$ yields a holomorphic section on all of X, i.e., a Katz modular form g in a weight k' which is by construction minimal, hence belongs to $C^{\min}$ by Proposition 1.11. As the q-expansions of all generalised Hasse invariants at the cusp $\infty_{\mathfrak{c}}$ equal 1 by Lemma 1.4, both f and g have the same q-expansion.

We next prove that given $\bar{k}$, there are only finitely many $k' \in C^{\min}$ that can appear for nonzero modular forms in $M_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ via the method in the previous paragraph. Since dividing by $h_{\tau,1}$ (for any $\tau \in \overline{\Sigma}$)

subtracts $(p-1)$ from the sum of the weights, whereas multiplying or dividing by $h_{\tau,i}$ for $\tau \in \overline{\Sigma}$ and $2 \leqslant i \leqslant e_{\mathfrak{p}(\tau)}$ leaves that sum unchanged, we deduce that $\sum_{\sigma\in\Sigma} k'_\sigma \leqslant \sum_{\tau\in\overline{\Sigma}} k_\tau$. As in the language of [Diamond and Kassaei 2023], the minimal cone is contained in the standard cone, we have $k'_\sigma \geqslant 0$ for all $\sigma \in \Sigma$ and the claimed finiteness follows. □

The finiteness of K in Lemma 1.12 yields the following result.

Corollary 1.13. *The $\mathbb{F}$-vector space $M_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ is finite dimensional.*

We now further use the work of Diamond and Kassaei to study the minimality of the weight for modular forms of parallel weight one.

Corollary 1.14. *Suppose $f \in M_1^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ is a nonzero Hilbert modular form and k is its minimal weight. Then, for any prime $\mathfrak{p} \mid p$, either $k_\sigma = 1$ for all $\sigma \in \Sigma_{\mathfrak{p}}$ (in that case, we say that the weight is* ***minimal at*** $\mathfrak{p}$*), or $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$, the latter case being possible only if $(p-1)$ divides $e_{\mathfrak{p}}$.*

Proof. By Proposition 1.11, we know that $k \in C^{\min}$. By definition of $C^{\min}$ one has $k_\sigma \geqslant 0$ for all $\sigma \in \Sigma$ and, moreover, if $k_\sigma = 0$ with $\sigma \in \Sigma_{\mathfrak{p}}$ for some $\mathfrak{p} \mid p$, then $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$.

We assume for the rest of this proof that $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$. Denote the weight of the Hasse invariant $h_{\tau,i}$ by $w_{\tau,i}$. By the definition of the minimal weight, there exist integers $n_{\tau,i} \geqslant 0$ such that

$$\sum_{\sigma\in\Sigma_{\mathfrak{p}}} \sigma = \sum_{\tau\in\overline{\Sigma}_{\mathfrak{p}}} \sum_{i=1}^{e_{\mathfrak{p}}} n_{\tau,i} w_{\tau,i}. \tag{8}$$

From the description of $w_{\tau,i}$ (see Table 1), it follows that for all $i \geqslant 2$ one has $n_{\tau,i} = n_{\tau,i-1} + 1$ and furthermore $p n_{\tau,1} = n_{\phi^{-1}\circ\tau,1} + e_{\mathfrak{p}}$. It is then easy to find that $n_{\tau,1} = e_{\mathfrak{p}}/(p-1)$ for all $\tau \in \overline{\Sigma}_{\mathfrak{p}}$, showing that $p-1$ divides $e_{\mathfrak{p}}$. □

The following result, the proof of which will be completed in the next subsection, shows that one can be more precise when restricting to cuspforms.

Proposition 1.15. *Let $\mathfrak{p}$ be a prime of F dividing p. Let $k = \sum_{\sigma\in\Sigma} k_\sigma \sigma \in \mathbb{Z}[\Sigma]$ be a weight such that $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$. Then $S_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F}) = 0$.*

Proof. By Lemma 1.7, $S_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ injects into $S_{\bar{k}}^{\mathrm{AG}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$, which is zero by Proposition 1.22. Alternatively, if there is a unique prime $\mathfrak{p}$ of F dividing p, then $k = 0$ and Koecher's principle applied to an embedding of the connected scheme X in a toroidal compactification implies that $\mathrm{H}^0(X, \mathcal{O}_X)$ consists only of forms which are constant, thus it does not contain any nonzero cuspforms. □

Corollary 1.16. *The weight of any nonzero parallel weight 1 cuspform is minimal.*

Proof. Let f be a nonzero cuspform of parallel weight 1 and let k be its minimal weight. Suppose the minimal weight k is not $\sum_{\sigma\in\Sigma} \sigma$. Then by Corollary 1.14 we already know that there exists $\mathfrak{p} \mid p$ such that $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$. Moreover, as the generalised Hasse invariants do not vanish at any cusp (see Lemma 1.4), we have constructed a *nonzero cuspform* of weight k, contradicting Proposition 1.15. This proves the corollary. □

Remark 1.17. Confusion may arise from the fact that parallel weight 1 forms in our sense have weight exponents $e_{\mathfrak{p}(\tau)}$ when seen as a modular form in $M^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ as $\bigwedge^{e_{\mathfrak{p}(\tau)}}\omega_\tau \simeq \omega_{\tilde{\tau}}^{\otimes e_{\mathfrak{p}(\tau)}}$ for $\tau\in\overline{\Sigma}$ over the Rapoport locus (see Lemma 1.7).

1D. ***Partial Frobenius operator.*** Fix $\mathfrak{c}\in\mathcal{C}$ and let $e\in\mathbb{N}$ be such that $\mathfrak{p}^{ee_{\mathfrak{p}}}=(\alpha)$ with $\alpha\in\mathfrak{o}_+$ and $\alpha\equiv p^e\equiv 1 \pmod{\mathfrak{n}}$. Also let $\beta\in\mathfrak{o}_+$ such that $p^e=\alpha\cdot\beta$. In order to lighten notation, we let $Y=X^{\mathrm{Ra}}$ denote the Rapoport locus and let $s:\mathcal{A}\to Y$ be the universal $\mathfrak{c}$-polarised HBAV endowed with $\mu_{\mathfrak{n}}$-level structure. Let $\mathcal{A}^{(p^e)}=\mathcal{A}\times_{Y,\mathrm{Fr}^e}Y$ be the base change by the e-th power of absolute Frobenius $\mathrm{Fr}:Y\to Y$. The e-th power of Verschiebung then defines an isogeny over Y

$$\mathcal{A}^{(p^e)}\xrightarrow{\mathrm{Ver}^e}\mathcal{A},$$

the kernel of which we denote by H. It is a finite group scheme with an $\mathfrak{o}/(p^e)$-action. Hence we can apply the Chinese remainder theorem to obtain the direct product decomposition $H=H_{\mathfrak{p}}\times H'_{\mathfrak{p}}$, where $H_{\mathfrak{p}}=H[\alpha]$ is the $\mathfrak{p}$-component of H and $H'_{\mathfrak{p}}=H[\beta]$ is the product of $\mathfrak{p}'$-components of H for all $\mathfrak{p}'\neq\mathfrak{p}$ dividing p. We now define the abelian variety

$$\mathcal{B}=\mathcal{A}^{(p^e)}/H'_{\mathfrak{p}},$$

through which Ver^e factors, leading to an isogeny over Y

$$\begin{array}{ccc}\mathcal{B} & \xrightarrow{V_{\mathcal{A}}} & \mathcal{A}. \\ & {\scriptstyle t}\searrow \quad \swarrow{\scriptstyle s} & \\ & Y & \end{array} \tag{9}$$

Lemma 1.18. *The abelian variety $\mathcal{B}$ inherits a $\mu_{\mathfrak{n}}$-level structure and an $\alpha^{-1}\mathfrak{c}$-polarisation λ_α.*

Proof. As $\mathcal{A}^{(p^e)}\to\mathcal{B}$ is a p-primary isogeny, the $\mu_{\mathfrak{n}}$-level structure on $\mathcal{A}^{(p^e)}$ yields one on $\mathcal{B}$.

Regarding the polarisation, following a suggestion of the referee (see also [Kisin and Lai 2005, Section 1.9]), we claim that the kernel of the composed isogeny

$$\delta:\mathcal{B}\otimes_{\mathfrak{o}}\mathfrak{c}\xrightarrow{V_{\mathcal{A}}\otimes 1}\mathcal{A}\otimes_{\mathfrak{o}}\mathfrak{c}\xrightarrow{\lambda^{-1}}\mathcal{A}^\vee\xrightarrow{V_{\mathcal{A}}^\vee}\mathcal{B}^\vee$$

equals the α-torsion of $\mathcal{B}\otimes_{\mathfrak{o}}\mathfrak{c}$, i.e., $\ker(\delta)$ is α-torsion and has the same order as $(\mathcal{B}\otimes_{\mathfrak{o}}\mathfrak{c})[\alpha]$. As the order of finite flat group schemes is locally constant, it suffices to check this pointwise on the ordinary locus of Y which is dense.

As Ver^e is étale at an ordinary closed point $y\in Y$, its kernel is isomorphic to the constant group scheme given by $\mathfrak{o}/p^e\mathfrak{o}$, whence $\ker(V_{\mathcal{A}_y})\simeq\mathfrak{o}/\alpha\mathfrak{o}$. Consequently, the kernel of the dual isogeny $V_{\mathcal{A}_y}^\vee$ is isomorphic to the Cartier dual $\mu_{\alpha\mathfrak{o}}$ of $\mathfrak{o}/\alpha\mathfrak{o}$. This gives us a short exact sequence of finite flat commutative group schemes

$$0\to\mathfrak{c}/\alpha\mathfrak{c}\simeq\ker(V_{\mathcal{A}_y})\otimes_{\mathfrak{o}}\mathfrak{c}\hookrightarrow\ker(\delta_y)\xrightarrow{\lambda^{-1}\circ(V_{\mathcal{A}_y}\otimes 1)}\ker(V_{\mathcal{A}_y}^\vee)\simeq\mu_{\alpha\mathfrak{o}}\to 0.$$

As the connected-étale sequence of any finite flat group scheme over a perfect field splits (see [Tate 1997, Section 3.7]), we deduce that $\ker(\delta_y)$ is isomorphic to the group scheme $(\mathfrak{c}/\alpha\mathfrak{c}) \times \mu_{\alpha\mathfrak{o}}$. In particular $\ker(\delta_y)$ is α-torsion and has the same order as $(\mathcal{B}_y \otimes_{\mathfrak{o}} \mathfrak{c})[\alpha]$. Therefore, it follows that $\ker(\delta) = (\mathcal{B} \otimes_{\mathfrak{o}} \mathfrak{c})[\alpha]$.

Since $(\mathcal{B} \otimes_{\mathfrak{o}} \mathfrak{c})/(\mathcal{B} \otimes_{\mathfrak{o}} \mathfrak{c})[\alpha]$ is canonically isomorphic to $\mathcal{B} \otimes_{\mathfrak{o}} (\alpha^{-1}\mathfrak{c})$, we deduce an isomorphism $\mathcal{B} \otimes_{\mathfrak{o}} (\alpha^{-1}\mathfrak{c}) \xrightarrow{\sim} \mathcal{B}^{\vee}$ the inverse of which is the desired $\alpha^{-1}\mathfrak{c}$-polarisation λ_α. □

We now verify that the HBAV $\mathcal{B}/Y$ satisfies the Rapoport condition. Recall that $\omega_{\mathcal{A}/Y} = s_* \Omega^1_{\mathcal{A}/Y}$ and $\omega_{\mathcal{B}/Y} := t_* \Omega^1_{\mathcal{B}/Y}$.

Lemma 1.19. *For any $\tau \in \overline{\Sigma} \setminus \overline{\Sigma}_{\mathfrak{p}}$, the map $V^*_{\mathcal{A},\tau} : \omega_{\mathcal{A}/Y,\tau} \to \omega_{\mathcal{B}/Y,\tau}$ is an isomorphism.*

On the other hand, if $\tau \in \overline{\Sigma}_{\mathfrak{p}}$, then the isogeny $\mathcal{A}^{(p^e)} \to \mathcal{B}$ induces an isomorphism

$$\omega_{\mathcal{B}/Y,\tau} \simeq \omega_{\mathcal{A}^{(p^e)}/Y,\tau} \simeq (\mathrm{Fr}^e)^* \omega_{\mathcal{A}/Y,\phi^{-e}\circ\tau}.$$

Proof. Let $r_{\mathfrak{p}}$ be the projection of $\mathfrak{o}/(p^e)$ on its $\mathfrak{p}$-primary component and let $\gamma \in \mathfrak{o}$ be such that its image in $\mathfrak{o}/(p^e)$ represents $r_{\mathfrak{p}}$. The image of $\gamma' = 1 - \gamma$ in $\mathfrak{o}/(p^e)$ represents the complementary idempotent $r'_{\mathfrak{p}} = 1 - r_{\mathfrak{p}}$. As γ' kills $\ker(V_{\mathcal{A}})$, the isogeny $\mathcal{B} \xrightarrow{\gamma'\cdot} \mathcal{B}$ factors through $V_{\mathcal{A}}$, yielding a factorisation

$$\begin{tikzcd} \omega_{\mathcal{B}/Y} \arrow[r] \arrow[rr, bend left, "\gamma'\cdot"] & \omega_{\mathcal{A}/Y} \arrow[r, "V_{\mathcal{A}}^*"] & \omega_{\mathcal{B}/Y} \end{tikzcd}$$

If $\mathfrak{p}' \mid p$ and $\mathfrak{p}' \neq \mathfrak{p}$, the projection of γ' on the $\mathfrak{p}'$-component of $\mathfrak{o}/(p^e)$ is 1. Hence, it induces the identity on the $\mathfrak{p}'$-component of $\omega_{\mathcal{B}/Y}$. So the map $V^*_{\mathcal{A}}$ is split on the $\mathfrak{p}'$-component and hence $\omega_{\mathcal{B}/Y,\tau}$ is isomorphic to a direct summand of $\omega_{\mathcal{A}/Y,\tau}$ for all $\tau \in \overline{\Sigma}_{\mathfrak{p}'}$. Recall that both $\omega_{\mathcal{B}/Y,\tau}$ and $\omega_{\mathcal{A}/Y,\tau}$ are locally free sheaves over Y of the same rank. Therefore, after passing to their stalks, we conclude that $V^*_{\mathcal{A},\tau}$ is an isomorphism for all $\tau \in \overline{\Sigma}_{\mathfrak{p}'}$.

Similarly, as γ annihilates the kernel of the isogeny $\mathcal{A}^{(p^e)} \to \mathcal{B}$, we obtain an isomorphism between $\omega_{\mathcal{A}^{(p^e)}/Y,\tau}$ and $\omega_{\mathcal{B}/Y,\tau}$ for all $\tau \in \overline{\Sigma}_{\mathfrak{p}}$. This proves the lemma. □

We get a $\mathfrak{c}$-polarisation on $\mathcal{B}$ from the $\alpha^{-1}\mathfrak{c}$-polarisation λ_α (which is obtained in Lemma 1.18) after identifying $\alpha^{-1}\mathfrak{c}$ with $\mathfrak{c}$ by multiplication by α. Thus, using Lemma 1.19, the universal property of $\mathcal{A} \to Y$ yields a Cartesian diagram

$$\begin{tikzcd} \mathcal{B} \arrow[r] \arrow[d, "t"] \arrow[dr, phantom, "\square"] & \mathcal{A}, \arrow[d, "s"] \\ Y \arrow[r, "\phi_\alpha"] & Y, \end{tikzcd} \tag{10}$$

from which we deduce a natural isomorphism $\phi_\alpha^* \omega_{\mathcal{A}/Y} \xrightarrow{\sim} \omega_{\mathcal{B}/Y}$ of $\mathfrak{o} \otimes \mathcal{O}_Y$-modules.

Let $k = \sum_{\sigma\in\Sigma} k_\sigma \sigma \in \mathbb{Z}[\Sigma]$ be a weight such that $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$. By Lemma 1.19

$$V^*_{\mathcal{A}} : \omega^{\otimes k}_{\mathcal{A}/Y} \xrightarrow{\sim} \omega^{\otimes k}_{\mathcal{B}/Y}. \tag{11}$$

Definition 1.20. Let $k = \sum_{\sigma\in\Sigma} k_\sigma \sigma \in \mathbb{Z}[\Sigma]$ be a weight such that $k_\sigma = 0$ for all $\sigma \in \Sigma_{\mathfrak{p}}$. The partial Frobenius operator $\Phi_{\mathfrak{p}^e}$ is defined as the composition of the adjunction morphism coming from (10) with (11)

$$\Phi_{\mathfrak{p}^e} : \mathrm{H}^0(Y, \omega_{\mathcal{A}/Y}^{\otimes k}) \xrightarrow{\phi_\alpha^*} \mathrm{H}^0(Y, \omega_{\mathcal{B}/Y}^{\otimes k}) \xrightarrow[\sim]{(V_{\mathcal{A}}^*)^{-1}} \mathrm{H}^0(Y, \omega_{\mathcal{A}/Y}^{\otimes k}).$$

We next study the effect of $\Phi_{\mathfrak{p}^e}$ on q-expansions. To this end, we recall the definition and properties of Tate objects. For fractional ideals $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ of $\mathfrak{o}$ such that $\mathfrak{a}\mathfrak{b} \subset \mathfrak{c}$ and a cone C in $\mathfrak{c}_+^*$, we let

$$T_{\mathfrak{a},\mathfrak{b}} = (\mathbb{G}_m \otimes \mathfrak{a}^*)/q^{\mathfrak{b}} \to \bar{S}_C^\circ = \mathrm{Spec}(R_C^\circ)$$

be the Tate HBAV over the Noetherian algebra $R_C^\circ \supset \mathbb{F}[\![q^\xi, \xi \in \mathfrak{c}_+]\!]$ (for more details we refer to [Dimitrov 2004, Section 2], where R_C° is denoted by $R_C^\wedge \otimes_{R_C} R$). It is equipped with a $\mu_{\mathfrak{n}}$-level structure which depends on the choice of an isomorphism between $\mathfrak{a}/\mathfrak{n}\mathfrak{a}$ and $\mathfrak{o}/\mathfrak{n}$. Moreover, the natural isomorphism

$$\lambda_{\mathfrak{a},\mathfrak{b}} : T_{\mathfrak{a},\mathfrak{b}}^\vee = T_{\mathfrak{b},\mathfrak{a}} \to T_{\mathfrak{a},\mathfrak{b}} \otimes_{\mathfrak{o}} (\mathfrak{a}\mathfrak{b}^{-1})$$

endows $T_{\mathfrak{a},\mathfrak{b}}$ with a canonical $\mathfrak{a}\mathfrak{b}^{-1}$-polarisation. Note that $T_{\mathfrak{c},\mathfrak{o}} = (\mathbb{G}_m \otimes \mathfrak{c}^*)/q^{\mathfrak{o}}$ is a Tate HBAV at the standard cusp $\infty_{\mathfrak{c}}$ of Y fitting, by universality of $\mathcal{A}/Y$, into a Cartesian diagram

$$\begin{array}{ccc} T_{\mathfrak{c},\mathfrak{o}} & \longrightarrow & \mathcal{A} \\ \big\downarrow & \square & \big\downarrow s \\ \bar{S}_C^\circ & \xrightarrow{\alpha_Y} & Y. \end{array} \tag{12}$$

This gives a natural isomorphism $\alpha_Y^* \omega_{\mathcal{A}/Y} \simeq \omega_{T_{\mathfrak{c},\mathfrak{o}}/\bar{S}_C^\circ}$ and further, by adjunction and choice of canonical trivialisations using (4), we obtain a q-expansion map at the cusp $\infty_{\mathfrak{c}}$:

$$\mathrm{H}^0(Y, \omega_{\mathcal{A}/Y}^{\otimes k}) \xrightarrow{\alpha_Y^*} \mathrm{H}^0(\bar{S}_C^\circ, \omega_{T_{\mathfrak{c},\mathfrak{o}}/\bar{S}_C^\circ}^{\otimes k}) \simeq R_C^\circ.$$

Next we describe Ver^e on Tate objects. Define $T_{\mathfrak{c},\mathfrak{o}}^{(p^e)} = T_{\mathfrak{c},\mathfrak{o}} \times_{\bar{S}_C^\circ, \mathrm{Fr}^e} \bar{S}_C^\circ$ as the base change by the e-th power of absolute Frobenius. Note that $T_{\mathfrak{c},\mathfrak{o}}^{(p^e)} = T_{p^e\mathfrak{c},\mathfrak{o}}$ and the relative Frobenius map $\mathrm{Frob}_{T_{\mathfrak{c},\mathfrak{o}}}^e : T_{\mathfrak{c},\mathfrak{o}} \to T_{p^e\mathfrak{c},\mathfrak{o}}$ is the map induced by the inclusion $p^e\mathfrak{c} \hookrightarrow \mathfrak{c}$. The p^e-th Verschiebung is the dual of the p^e-th relative Frobenius on $T_{\mathfrak{c},\mathfrak{o}}^\vee$, i.e., $\mathrm{Ver}^e = (\mathrm{Frob}_{T_{\mathfrak{c},\mathfrak{o}}^\vee}^e)^\vee$. We do not identify $((T_{\mathfrak{c},\mathfrak{o}}^\vee)^{(p^e)})^\vee$ with $T_{\mathfrak{c},\mathfrak{o}}^{(p^e)}$ (as is usually done while defining Verschiebung) in order to get the desired maps on Tate objects. In particular, $\mathrm{Frob}_{T_{\mathfrak{c},\mathfrak{o}}^\vee}^e : T_{\mathfrak{o},\mathfrak{c}} \to T_{p^e\mathfrak{o},\mathfrak{c}}$ is the map induced by the inclusion $p^e\mathfrak{o} \hookrightarrow \mathfrak{o}$. Therefore, its dual map $\mathrm{Ver}^e : T_{\mathfrak{c},p^e\mathfrak{o}} \to T_{\mathfrak{c},\mathfrak{o}}$ is the natural projection obtained by going modulo $q^{\mathfrak{o}}$.

Our next aim is to specialise ϕ_α to the Tate objects. It follows from the previous paragraph that the base change of (9) to $\bar{S}_C^\circ$ is given by the following commutative diagram:

$$\begin{array}{ccc} T_{\mathfrak{c},\alpha\mathfrak{o}} & \xrightarrow{V_T} & T_{\mathfrak{c},\mathfrak{o}} \\ & \searrow \quad \swarrow & \\ & \bar{S}_C^\circ & \end{array}$$

where the map V_T is the natural projection obtained by going modulo $q^{\mathfrak{o}}$. Combining with (10), we get the following Cartesian diagram:

$$\begin{array}{ccccc} T_{\mathfrak{c},\alpha\mathfrak{o}} & \longrightarrow & \mathcal{B} & \longrightarrow & \mathcal{A} \\ \big\downarrow & \square & \big\downarrow t & \square & \big\downarrow s \\ \bar{S}^{\circ}_{C} & \xrightarrow{\alpha_Y} & Y & \xrightarrow{\phi_\alpha} & Y \end{array} \tag{13}$$

On the other hand, considering the $\mathfrak{c}$-polarised HBAV $T_{\mathfrak{c},\mathfrak{o}}$ over $\bar{S}^{\circ}_{\alpha C}$ gives a Cartesian diagram:

$$\begin{array}{ccccc} T_{\mathfrak{c},\alpha\mathfrak{o}} & \longrightarrow & T_{\mathfrak{c},\mathfrak{o}} & \longrightarrow & \mathcal{A} \\ \big\downarrow & \square & \big\downarrow & \square & \big\downarrow s \\ \bar{S}^{\circ}_{C} & \xrightarrow{f_\alpha} & \bar{S}^{\circ}_{\alpha C} & \xrightarrow{\alpha'_Y} & Y, \end{array} \tag{14}$$

where f_α is induced by the morphism $R^{\circ}_{\alpha C} \to R^{\circ}_{C}$ sending q^{ξ} to $q^{\alpha\xi}$. We would like to emphasise that αC is considered as a cone in $\mathfrak{c}^{*}_{+}$, hence the dual cone (used in the construction of $R^{\circ}_{\alpha C}$, see [Dimitrov 2004, Section 2]) is considered as a cone in $\mathfrak{c}$ (and not in $\alpha^{-1}\mathfrak{c}$). In particular, the morphism $R^{\circ}_{\alpha C} \to R^{\circ}_{C}$, $q^{\xi} \mapsto q^{\alpha\xi}$ is not étale.

Lemma 1.21. *Under the notation developed above*, $\phi_\alpha \circ \alpha_Y = \alpha'_Y \circ f_\alpha$.

Proof. The proof proceeds by showing that the $\mathfrak{c}$-polarisation and $\mu_{\mathfrak{n}}$-level structure on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ obtained from the base change in (13) coincide with the ones obtained from the base change in (14). This, along with the universality of $\mathcal{A}/Y$, implies $\phi_\alpha \circ \alpha_Y = \alpha'_Y \circ f_\alpha$.

As Mumford's construction of Tate objects presented in [Dimitrov 2004, Section 2] is functorial in $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c})$ and C, we deduce that the $\mathfrak{c}$-polarisation on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ arising from (14) is obtained from $\lambda_{\mathfrak{c},\alpha\mathfrak{o}}$ after identifying $\alpha^{-1}\mathfrak{c}$ with $\mathfrak{c}$ by multiplication by α.

We now derive the $\mathfrak{c}$-polarisation on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ via the base change in (13). To do this, we proceed as in the proof of Lemma 1.18 to first obtain an $\alpha^{-1}\mathfrak{c}$-polarisation on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ from $\lambda_{\mathfrak{c},\mathfrak{o}}$. From the proof of Lemma 1.18, it follows that the kernel of the isogeny

$$T_{\mathfrak{c},\alpha\mathfrak{o}} \otimes_{\mathfrak{o}} \mathfrak{c} \to T_{\mathfrak{c},\mathfrak{o}} \otimes_{\mathfrak{o}} \mathfrak{c} \to T_{\mathfrak{o},\mathfrak{c}} \to T_{\alpha\mathfrak{o},\mathfrak{c}}.$$

is just the α-torsion of $T_{\mathfrak{c},\alpha\mathfrak{o}} \otimes_{\mathfrak{o}} \mathfrak{c}$. Here the first map is induced by V_T (the natural projection given by going modulo $q^{\mathfrak{o}}$), the second map is $\lambda^{-1}_{\mathfrak{c},\mathfrak{o}}$, and the final map is induced by the inclusion $\alpha\mathfrak{o} \subset \mathfrak{o}$. Therefore, this composition of maps induces an isomorphism

$$\lambda : T^{\vee}_{\mathfrak{c},\alpha\mathfrak{o}} = T_{\alpha\mathfrak{o},\mathfrak{c}} \xrightarrow{\sim} (T_{\mathfrak{c},\alpha\mathfrak{o}} \otimes_{\mathfrak{o}} \mathfrak{c}) \otimes_{\mathfrak{o}} \alpha^{-1}\mathfrak{o} = T_{\mathfrak{c},\alpha\mathfrak{o}} \otimes_{\mathfrak{o}} \alpha^{-1}\mathfrak{c},$$

which is the $\alpha^{-1}\mathfrak{c}$-polarisation on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ induced from $\lambda_{\mathfrak{c},\mathfrak{o}}$. From the description of the maps above, it follows that $\lambda = \lambda_{\mathfrak{c},\alpha\mathfrak{o}}$. Hence, the $\mathfrak{c}$-polarisation on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ via the base change in (13) is obtained from $\lambda_{\mathfrak{c},\alpha\mathfrak{o}}$ by identifying $\alpha^{-1}\mathfrak{c}$ with $\mathfrak{c}$ by multiplication by α. Therefore, it follows that these two $\mathfrak{c}$-polarisations on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ coincide.

As $\alpha \equiv 1 \pmod{\mathfrak{n}}$, the multiplication by α map preserves the $\mu_{\mathfrak{n}}$-level structure of $T_{\mathfrak{c},\mathfrak{o}}$. Hence, the $\mu_{\mathfrak{n}}$-level structure on $T_{\mathfrak{c},\alpha\mathfrak{o}}$ induced by f_α is same as the one coming from the quotient map $T_{\mathfrak{c},p^e\mathfrak{o}} \to T_{\mathfrak{c},\alpha\mathfrak{o}}$. This concludes the proof of the lemma. □

We are now ready to compute the effect of $\Phi_{\mathfrak{p}^e}$ on q-expansions at $\infty_{\mathfrak{c}}$.

Proposition 1.22. *Let $\bar{k} = \sum_{\tau\in\overline{\Sigma}} k_\tau \tau \in \mathbb{Z}[\overline{\Sigma}]$ such that $k_\tau = 0$ for all $\tau \in \overline{\Sigma}_{\mathfrak{p}}$. Then the map $\Phi_{\mathfrak{p}^e}$ defines an endomorphism of $M^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$, sending $f = \sum_{\xi\in\mathfrak{c}_+} a_\xi q^\xi$ to $\Phi_{\mathfrak{p}^e}(f) = \sum_{\xi\in\mathfrak{c}_+} a_\xi q^{\alpha\xi}$. In particular, the restriction of $\Phi_{\mathfrak{p}^e}$ to $S^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$ is injective and nilpotent, hence $S^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F}) = \{0\}$.*

Proof. By definition, the q-expansion of $\Phi_{\mathfrak{p}^e}(f)$ at $\infty_{\mathfrak{c}}$ is the image of f under the map $\mathrm{H}^0(Y, \omega^{\otimes k}_{\mathcal{A}/Y}) \to \mathrm{H}^0(\bar{S}^\circ_C, \omega^{\otimes k}_{T_{\mathfrak{c},\alpha\mathfrak{o}}/\bar{S}^\circ_C})$ coming from the Cartesian diagram (13), followed by $(V^*_T)^{-1}$. By Lemma 1.21, one can use instead the Cartesian diagram (14). Hence, the q-expansion of $\Phi_{\mathfrak{p}^e}(f)$ at $\infty_{\mathfrak{c}}$ can be obtained as the image of f under the adjunction morphism

$$\mathrm{H}^0(Y, \omega^{\otimes k}_{\mathcal{A}/Y}) \xrightarrow{\alpha'_Y{}^*} \mathrm{H}^0(\bar{S}^\circ_{\alpha C}, \omega^{\otimes k}_{T_{\mathfrak{c},\mathfrak{o}}/\bar{S}^\circ_{\alpha C}}) \xrightarrow{f^*_\alpha} \mathrm{H}^0(\bar{S}^\circ_C, \omega^{\otimes k}_{T_{\mathfrak{c},\alpha\mathfrak{o}}/\bar{S}^\circ_C})$$

followed by the map $(V^*_T)^{-1} : \mathrm{H}^0(\bar{S}^\circ_C, \omega^{\otimes k}_{T_{\mathfrak{c},\alpha\mathfrak{o}}/\bar{S}^\circ_C}) \xrightarrow{\sim} \mathrm{H}^0(\bar{S}^\circ_C, \omega^{\otimes k}_{T_{\mathfrak{c},\mathfrak{o}}/\bar{S}^\circ_C})$. Since the q-expansion $\sum_{\xi\in\mathfrak{c}_+} a_\xi q^\xi$ of f at the cusp $\infty_{\mathfrak{c}}$ is independent of a particular choice of a cone, it is given by the image of f under the map $\alpha'_Y{}^* : \mathrm{H}^0(Y, \omega^{\otimes k}_{\mathcal{A}/Y}) \to \mathrm{H}^0(\bar{S}^\circ_{\alpha C}, \omega^{\otimes k}_{T_{\mathfrak{c},\mathfrak{o}}/\bar{S}^\circ_{\alpha C}})$. As f_α is induced by the map sending q^ξ to $q^{\alpha\xi}$ we deduce that $f^*_\alpha\big(\sum_{\xi\in\mathfrak{c}_+} a_\xi q^\xi\big) = \sum_{\xi\in\mathfrak{c}_+} a_\xi q^{\alpha\xi}$. Finally, as V_T is induced from the identity map on the torus $\mathbb{G}_m \otimes \mathfrak{c}^*$, the morphism V^*_T is the identity on the q-expansions, i.e., $(V^*_T)^{-1}\big(\sum_{\xi\in\mathfrak{c}_+} a_\xi q^{\alpha\xi}\big) = \sum_{\xi\in\mathfrak{c}_+} a_\xi q^{\alpha\xi}$, yielding the desired formula.

The rest follows from the q-expansion principle and the finite dimensionality of $S^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$. □

1E. *Refined injectivity criterion for Θ-operators.* The purpose of this section is to extend the definition of the Andreatta–Goren operators $\Theta^{\mathrm{AG}}_\tau$ for $\tau \in \overline{\Sigma}$ and prove an injectivity criterion refining [Andreatta and Goren 2005, Proposition 15.10] when p ramifies in F. Given $f \in M^{\mathrm{Katz}}_k(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$, by Lemma 1.7 and the discussion after it, $\Theta^{\mathrm{AG}}_\tau(f/H^{\mathrm{RX}}_k)$ defines a meromorphic section over X, whose poles lie outside X^{Ra}. A careful study of the order at these poles will first show that multiplication by H^{RX}_k leads to a holomorphic section and further allow us to establish Proposition 1.28 (injectivity criterion). If p is unramified in F, our Θ-operators coincide exactly with those of Andreatta and Goren, and in that case everything that we prove here has already been proved in [Andreatta and Goren 2005]; see also [Diamond and Sasaki 2023].

The construction of $\Theta^{\mathrm{AG}}_\tau$ goes via the Kummer cover. By definition, the ordinary locus X^{ord} of X^{Ra} is endowed with a Galois cover $X(\mu_{(p)})^{\mathrm{ord}} \to X^{\mathrm{ord}}$ with group $(\mathfrak{o}/(p))^\times$, where $X(\mu_{(p)})$ is the Deligne–Pappas moduli space of level $p\mathfrak{n}$. Taking the quotient by the p-Sylow subgroup yields a cover $\pi : X^{\mathrm{Kum}} \to X^{\mathrm{ord}}$ with group $\prod_{\mathfrak{p}|p}(\mathfrak{o}/\mathfrak{p})^\times$, called the *Kummer cover*. Let $\widetilde{\pi} : \widetilde{X} \to X$ be the normal closure of X in X^{Kum}. It can be described explicitly using the generalised Hasse invariants as follows. For $\tau \in \overline{\Sigma}$, we write $\mathfrak{p} = \mathfrak{p}(\tau)$, $f = f_{\mathfrak{p}}$ and we let

$$H_\tau = \prod_{j=0}^{f-1} (h_{\phi^{-j}\circ\tau})^{p^j}. \tag{15}$$

It is a modular form of weight $(p^f-1)\widetilde{\tau}$ and X^{Kum} is obtained by adjoining a (p^f-1)-th root s_τ of it for all $\tau\in\overline{\Sigma}$. The nowhere vanishing section s_τ provides a trivialisation of the line bundle $\pi^*\omega_{\widetilde{\tau}}$ over X^{Kum} (see [Andreatta and Goren 2005, Definition 7.4]), and by definition of the normalisation, it also defines a section over $\widetilde{X}$; see [Andreatta and Goren 2005, Proposition 7.9]. As

$$H^p_{\phi^{-1}\circ\tau}=H_\tau\cdot(h_\tau)^{p^f-1},\tag{16}$$

$H^p_{\phi^{-1}\circ\tau}/H_\tau$ is a (p^f-1)-th power, this construction does not depend on the choice of $\tau\in\overline{\Sigma}_{\mathfrak{p}}$.

Next we describe the Kodaira–Spencer maps. By [Reduzzi and Xiao 2017, Theorem 2.9] there is a decomposition

$$\Omega^1_{X/\mathbb{F}}=\bigoplus_{\tau\in\overline{\Sigma}}\Omega^1_{X/\mathbb{F},\tau}\tag{17}$$

where each $\Omega^1_{X/\mathbb{F},\tau}$ is endowed with a filtration whose successive subquotients are naturally isomorphic to $\omega^{\otimes 2}_{\tau,i}$ with $1\leqslant i\leqslant e_{\mathfrak{p}}$ for $\mathfrak{p}=\mathfrak{p}(\tau)$ in descending order, i.e., $\omega^{\otimes 2}_{\widetilde{\tau}}=\omega^{\otimes 2}_{\tau,e_{\mathfrak{p}}}$ is naturally a quotient. As the map $\pi:X^{\mathrm{Kum}}\to X^{\mathrm{ord}}$ is étale, we have $\Omega^1_{X^{\mathrm{Kum}}/\mathbb{F}}=\pi^*\Omega^1_{X^{\mathrm{ord}}/\mathbb{F}}$, the elements of which we view as meromorphic sections of the sheaf $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F}}$ over $\widetilde{X}$. Given a section of $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F}}$, we denote by a subscript $\tau\in\overline{\Sigma}$ its projection onto the τ-component via (17). Consider the surjective map

$$\mathrm{KS}_\tau:\widetilde{\pi}^*\Omega^1_{X/\mathbb{F},\tau}\to\widetilde{\pi}^*\omega^{\otimes 2}_{\widetilde{\tau}}.$$

Definition 1.23. Let $k=\sum_{\sigma\in\Sigma}k_\sigma\sigma\in\mathbb{Z}[\Sigma]$ and $k_\tau=\sum_{\sigma\in\Sigma_\tau}k_\sigma$ for $\tau\in\overline{\Sigma}$. Let $f\in M^{\mathrm{Katz}}_k(\mathfrak{c},\mathfrak{n};\mathbb{F})$. Recall that $f/H^{\mathrm{RX}}_k\in M^{\mathrm{AG}}_{\bar{k}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ (see Lemma 1.7). We put

$$H^{\mathrm{AG}}_k=\prod_{\tau\in\overline{\Sigma}}s_\tau^{k_\tau}\quad\text{and}\quad H_k=H^{\mathrm{AG}}_k\cdot\pi^*(H^{\mathrm{RX}}_k).$$

Similarly to [Andreatta and Goren 2005, Definition 7.19], we further put

$$r(f)=\pi^*(f/H^{\mathrm{RX}}_k)/H^{\mathrm{AG}}_k=\pi^*(f)/H_k\in\mathrm{H}^0(X^{\mathrm{Kum}},\mathcal{O}_{X^{\mathrm{Kum}}})$$

where we restricted f and H^{RX}_k to X^{ord}.

Definition 1.24. For $\tau\in\overline{\Sigma}$, we define the *generalised* Θ*-operator* acting on $f\in M^{\mathrm{Katz}}_k(\mathfrak{c},\mathfrak{n};\mathbb{F})$ as

$$\Theta_\tau(f)=\mathrm{KS}_\tau(d(r(f))_\tau)\cdot H_k\cdot\pi^*(h_\tau)=H^{\mathrm{RX}}_k\cdot\Theta^{\mathrm{AG}}_\tau\left(\frac{f}{H^{\mathrm{RX}}_k}\right),$$

viewed as an element of $\mathrm{H}^0(X^{\mathrm{Kum}},\pi^*\omega^{\otimes k'})$ where:

(1) If $\tau\neq\phi^{-1}\circ\tau$, then

$$k'_\sigma=\begin{cases}k_\sigma+1 & \text{if }\sigma=\sigma_{\tau,e_{\mathfrak{p}(\tau)}}=\widetilde{\tau},\\ k_\sigma+p & \text{if }\sigma=\sigma_{\phi^{-1}\circ\tau,e_{\mathfrak{p}(\tau)}}=\widetilde{\phi^{-1}\circ\tau},\\ k_\sigma & \text{otherwise.}\end{cases}$$

(2) If $\tau = \phi^{-1} \circ \tau$, then

$$k'_\sigma = \begin{cases} k_\sigma + p + 1 & \text{if } \sigma = \sigma_{\tau, e_{\mathfrak{p}(\tau)}} = \widetilde{\tau}, \\ k_\sigma & \text{otherwise.} \end{cases}$$

We will now prove that $\Theta_\tau(f)$ yields an element of $M^{\mathrm{Katz}}_{k'}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$. In order to prove this, we proceed as in [Andreatta and Goren 2005] to calculate the poles of $d(r(f))_\tau$ along the divisors of the generalised Hasse invariants.

Using the trivialisation of the line bundles $\pi^*\omega_{\widetilde{\tau}}$ given by the sections s_τ and the generalised Hasse invariants $h_{\tau,i}$'s for $i > 1$, we get trivialisations of $\pi^*\omega_{\tau,i}$ for all $\tau \in \overline{\Sigma}$ and $1 \leqslant i \leqslant e_{\mathfrak{p}(\tau)}$. Using these trivialisations we can view the pullbacks $\widetilde{\pi}^* h_{\tau,i}$ and $\widetilde{\pi}^* h_\tau$ of Hasse invariants as functions over $\widetilde{X}$ (see [Andreatta and Goren 2005, Section 12.32] for more details), whose differentials are denoted by $d(h_{\tau,i})$ and $d(h_\tau)$, respectively (viewed as meromorphic sections of $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F}}$).

For $\tau' \in \overline{\Sigma}$, we let $\widetilde{Z}$ be an irreducible component of the effective Weil divisor of $\widetilde{X}$ associated to $\widetilde{\pi}^*(h_{\tau'})$ (see Section 1B). From the construction of $\widetilde{X}$ and [Andreatta and Goren 2005, Section 9.3, Proposition 9.4] (see also [Andreatta and Goren 2005, Section 12.32]), we can choose a uniformiser δ at the generic point of $\widetilde{Z}$ such that $\delta^{p^{f_{\mathfrak{p}(\tau')}}-1} = H_{\tau'}$ (see (15) for the definition of $H_{\tau'}$). We fix this choice from now on and let v_δ be the corresponding normalised discrete valuation. For the sake of readability, we will often drop $\widetilde{\pi}^*$ from the notation when pulling back Hilbert modular forms, especially generalised Hasse invariants; for instance, we usually write $v_\delta(h_\tau)$ for $v_\delta(\widetilde{\pi}^*(h_\tau))$.

We will first calculate $v_\delta((d\delta)_\tau)$, where $(d\delta)_\tau$ is viewed as a meromorphic section of $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F},\tau}$ over $\widetilde{X}$. We also prove some complementary results which will be used in the proof of the injectivity criterion.

Lemma 1.25. (i) *Let $\tau \in \overline{\Sigma}$ different from τ'. Then $(d\delta)_\tau = 0$.*

(ii) *There is a unique $1 \leqslant i_0 \leqslant e_{\mathfrak{p}(\tau')}$ (depending on $\widetilde{Z}$) such that $v_\delta(h_{\tau',i_0}) = p^{f_{\mathfrak{p}(\tau')}} - 1$ and $v_\delta(h_{\tau',i}) = 0$ for all $i \neq i_0$. Moreover, $v_\delta(h_{\tau'}) = p^{f_{\mathfrak{p}(\tau')}} - 1$ and $v_\delta(h_\tau) = 0$ if $\tau \neq \tau'$.*

(iii) *$v_\delta(d(h_{\tau',i})) \geqslant 0$ for all i and for i_0 found in (ii), $v_\delta(d(h_{\tau',i_0})) = 0$.*

(iv) *$v_\delta(s_{\tau'}) = 1$, $v_\delta(s_\tau) = p^j$ if $\tau = \phi^j \circ \tau'$ and $v_\delta(s_\tau) = 0$ if $\tau \neq \phi^j \circ \tau'$ for any integer j.*

(v) *$v_\delta((d\delta)_{\tau'}) = 2 - p^{f_{\mathfrak{p}(\tau')}}$.*

(vi) *$(d\delta)_{\tau'} = D + g \cdot (d(h_{\tau',i_0}))_{\tau'}$ where $g = -\delta^{2-p^{f_{\mathfrak{p}(\tau')}}} \cdot \bigl(\prod_{j=1}^{f_{\mathfrak{p}(\tau')}-1} (h_{\phi^{-j}\circ\tau'})^{p^j}\bigr) \cdot \bigl(\prod_{j\neq i_0} h_{\tau',j}\bigr)$ and D is a meromorphic section of $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F},\tau'}$ such that $v_\delta(D) \geqslant 0$.*

(vii) *$\mathrm{KS}_{\tau'}(d(h_{\tau',e_{\mathfrak{p}(\tau')}})_{\tau'}) \neq 0$ and if $\widetilde{Z}$ is an irreducible component of the effective Weil divisor of $\widetilde{X}$ associated with $\widetilde{\pi}^*(h_{\tau',e_{\mathfrak{p}(\tau')}})$, then $v_\delta(\mathrm{KS}_{\tau'}(d(h_{\tau',e_{\mathfrak{p}(\tau')}})_{\tau'}) s_{\tau'}^{-2}) = -2$.*

Proof. Recall that we have chosen δ such that $\delta^{p^{f_{\mathfrak{p}(\tau')}}-1} = H_{\tau'}$. Hence

$$-\delta^{p^{f_{\mathfrak{p}(\tau')}}-2} d\delta = \left(\prod_{j=1}^{f_{\mathfrak{p}(\tau')}-1} (h_{\phi^{-j}\circ\tau'})^{p^j} \right) \cdot d(h_{\tau'}).$$

Since X^{Ra} is Zariski dense in X, it follows from [Andreatta and Goren 2005, Lemma 12.34] that $(d(h_{\tau'}))_\tau = 0$ if $\tau \neq \tau'$. Hence, we get $d\delta = (d\delta)_{\tau'}$, which implies that $(d\delta)_\tau = 0$ if $\tau \neq \tau'$. Now

$$d(h_{\tau'}) = d\left(\prod_{i=1}^{e_{\mathfrak{p}(\tau')}} h_{\tau',i}\right) = \sum_{i=1}^{e_{\mathfrak{p}(\tau')}} \left(\prod_{j\neq i} h_{\tau',j}\right) \cdot d(h_{\tau',i}).$$

Since δ is a uniformiser at the generic point of $\widetilde{Z}$, there is a unique i_0 such that $v_\delta(h_{\tau',i_0}) > 0$. Note that $v_\delta(h_{\tau',i}) = 0$ for $i \neq i_0$ and $v_\delta(d(h_{\tau',i})) \geqslant 0$ for all i. Moreover, it follows from [Reduzzi and Xiao 2017, Theorem 3.10] that $v_\delta(d(h_{\tau',i_0})) = 0$. So (ii) and (iii) follow from the discussion above. Combining this with (16) gives (iv). Hence, $v_\delta(d(h_{\tau'})) = v_\delta(\prod_{i\neq i_0} h_{\tau',i}(d(h_{\tau',i_0}))) = 0$ from which (v) follows and combining this with (ii) and (iii) gives us (vi).

We will now prove statement (vii). Recall that, by [Reduzzi and Xiao 2017, Theorem 2.9], $\Omega^1_{X/\mathbb{F},\tau'}$ admits a canonical filtration whose successive subquotients are (isomorphic to) $\omega^{\otimes 2}_{\tau',i}$ with $1 \leqslant i \leqslant e_{\mathfrak{p}(\tau')}$. Recall that $\mathrm{KS}_{\tau'}$ is the surjective map from $\widetilde{\pi}^*\Omega^1_{X/\mathbb{F},\tau'}$ onto its first subquotient $\widetilde{\pi}^*\omega^{\otimes 2}_{\tau',e_{\mathfrak{p}(\tau')}}$. On the other hand, by [Reduzzi and Xiao 2017, Theorem 3.10], $\Omega^1_{Z_{\tau',e_{\mathfrak{p}(\tau')}}/\mathbb{F},\tau'}$ admits a canonical filtration whose successive subquotients are (isomorphic to) $\omega^{\otimes 2}_{\tau',i}$ with $1 \leqslant i < e_{\mathfrak{p}(\tau')}$. Here $Z_{\tau',e_{\mathfrak{p}(\tau')}} \subset X$ is the divisor of $h_{\tau',e_{\mathfrak{p}(\tau')}}$. Therefore, we conclude that $\mathrm{KS}_{\tau'}(d(h_{\tau',e_{\mathfrak{p}(\tau')}})_{\tau'}) \neq 0$. Since $s_{\tau'}$ gives a trivialisation of the line bundle $\pi^*\omega_{\widetilde{\tau}'}$, $v_\delta(\mathrm{KS}_{\tau'}(d(h_{\tau',e_{\mathfrak{p}(\tau')}})_{\tau'})s_{\tau'}^{-2})$ is well defined. We conclude $v_\delta(\mathrm{KS}_{\tau'}(d(h_{\tau',e_{\mathfrak{p}(\tau')}})_{\tau'})s_{\tau'}^{-2}) = -2$ by combining [Reduzzi and Xiao 2017, Theorem 3.10] with (iii) and (iv); see also [Andreatta and Goren 2005, Proposition 12.37]. This concludes the proof of the lemma. □

In order to compute $v_\delta(d(r(f))_\tau)$, it is sufficient to work in the discrete valuation ring obtained by localising at the generic point of $\widetilde{Z}$. Letting $r(f) = \frac{u}{\delta^n}$, with $v_\delta(u) = 0$, we have

$$d(r(f))_\tau = \frac{(du)_\tau}{\delta^n} - \frac{nu(d\delta)_\tau}{\delta^{n+1}}. \tag{18}$$

Lemma 1.26. *Let* $\tau \in \overline{\Sigma}$ *and let* $f \in M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{F})$. *Then*

(i) $v_\delta((du)_\tau) \geqslant \inf\{0, v_\delta((d\delta)_\tau)\}$,

(ii) $v_\delta(d(r(f))_\tau) \geqslant v_\delta(r(f))$, *if* $\tau \neq \tau'$,

(iii) $v_\delta(d(r(f))_{\tau'}) \geqslant v_\delta(r(f)) - (p^{f_{\mathfrak{p}(\tau')}} - 2)$, *if* $p \mid v_\delta(r(f))$,

(iv) $v_\delta(d(r(f))_{\tau'}) = v_\delta(r(f)) - (p^{f_{\mathfrak{p}(\tau')}} - 1)$, *if* $p \nmid v_\delta(r(f))$.

Proof. Let $\tau \in \overline{\Sigma}$. The proof of (i) is similar to the proof of [Andreatta and Goren 2005, Proposition 12.35] and we reproduce parts of it here. Let B (resp. $\widetilde{B}$) be the local ring of the generic point of $\widetilde{\pi}(\widetilde{Z})$ (resp. of $\widetilde{Z}$). As in [loc. cit., Corollary 9.6], we have $B \subset B^{\mathrm{et}} \subset \widetilde{B}$ where B^{et} is étale over B and $\widetilde{B} = B^{\mathrm{et}}[\delta]$. Writing $u = \sum_{h=0}^{p^{f_{\mathfrak{p}(\tau')}}-2} u_h\delta^h$ with $u_h \in B^{\mathrm{et}}$, we have

$$(du)_\tau = \sum_{h=0}^{p^{f_{\mathfrak{p}(\tau')}}-2} (\delta^h(du_h)_\tau + hu_h\delta^{h-1}(d\delta)_\tau).$$

Now $\delta^h(du_h)_\tau$ lies in $\widetilde{B}\otimes_{B^{\mathrm{et}}}\Omega^1_{B^{\mathrm{et}}/\mathbb{F}}$, which is the same as $\widetilde{\pi}^*\Omega^1_{B/\mathbb{F}}$ since B^{et} is étale over B. Hence $\delta^h(du_h)_\tau$ has no poles and (i) follows. Combining this inequality with (18) and Lemma 1.25(i), (v) gives us the other parts of the lemma. □

Finally we are ready to prove that $\Theta_\tau(f)$ is also a mod p Hilbert modular form.

Proposition 1.27. *Let $\tau\in\overline{\Sigma}$, $f\in M_k^{\mathrm{Katz}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ and k' be as in Definition 1.24. Then $\Theta_\tau(f)$ descends to a global section of the line bundle $\omega^{\otimes k'}$ over X^{ord}, and further extends to a section over X, yielding an element $\Theta_\tau(f)\in M_{k'}^{\mathrm{Katz}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$.*

Proof. The descent follows by applying [Andreatta and Goren 2005, Theorem 12.39] to f/H_k^{RX}.

As KS_τ is a surjective map of locally free sheaves with a locally free kernel over the normal scheme $\widetilde{X}$, the orders of the poles of $\mathrm{KS}_\tau(d(r(f))_\tau)$ are less than or equal to the orders of the poles of $d(r(f))_\tau$, i.e., $v_\delta(\mathrm{KS}_\tau(d(r(f))_\tau))\geqslant v_\delta(d(r(f))_\tau)$ (see the proof of [loc. cit., Proposition 12.37] for more details). Note that $r(f)\cdot H_k=\pi^*(f)$ has no poles on $\widetilde{X}$, i.e., $v_\delta(\pi^*f)\geqslant 0$. Combining Lemma 1.25(ii) and Lemma 1.26, we get that $\Theta_\tau(f)$ has no poles over $\widetilde{X}$. Hence, the section obtained by descending from X^{Kum} to X^{ord} extends to all of X and is thus a Hilbert modular form. □

The effect of Θ_τ on the geometric q-expansions of Hilbert modular forms will be used in Section 2B and can be described as follows. The identification (4), used in defining the geometric q-expansion $\sum_{\xi\in\mathfrak{c}_+\cup\{0\}}a_\xi(f)q^\xi$ of f at the cusp $\infty_\mathfrak{c}$, allows one to consider the map

$$\bar{\tau}_\mathfrak{c}:\mathbb{F}\otimes\mathfrak{c}\xrightarrow{\sim}\mathbb{F}\otimes\mathfrak{o}\twoheadrightarrow\mathbb{F}[x]/(x^{e_{\mathfrak{p}(\tau)}})\twoheadrightarrow\mathbb{F},$$

where the middle map is given by the idempotent at τ. By [Andreatta and Goren 2005, Corollary 12.40] we obtain the following q-expansion at the cusp $\infty_\mathfrak{c}$:

$$\Theta_\tau(f)=\sum_{\xi\in\mathfrak{c}_+}\bar{\tau}_\mathfrak{c}(1\otimes\xi)a_\xi q^\xi. \tag{19}$$

The proof of our main theorem uses the injectivity of Θ_τ on certain mod p Hilbert modular forms.

Proposition 1.28. *Let $f\in M_k^{\mathrm{Katz}}(\mathfrak{c},\mathfrak{n};\mathbb{F})$ and let $\tau\in\overline{\Sigma}_\mathfrak{p}$. Suppose $p\nmid k_{\tau,e_\mathfrak{p}}$ and $h_{\tau,e_\mathfrak{p}}$ does not divide f. Then $\Theta_\tau(f)\neq 0$.*

In particular, if the weight of $f\neq 0$ is minimal at $\mathfrak{p}$, and $p\nmid k_{\tau,e_\mathfrak{p}}$, then $\Theta_\tau(f)\neq 0$.

Proof. We follow the proof of [Andreatta and Goren 2005, Proposition 15.10]. Let δ be the uniformiser at the generic point of an irreducible component of the Weil divisor of $\widetilde{X}$ attached to $\widetilde{\pi}^*h_{\tau,e_\mathfrak{p}}$ chosen just before Lemma 1.25. As $v_\delta(\widetilde{\pi}^*(f))=0$, using Lemma 1.25(ii),(iv) (see also [loc. cit., Proposition 15.9]) we deduce

$$n:=v_\delta(r(f))=-\sum_{j=0}^{f-1}p^jk_{\phi^j\circ\tau}-(p^{f_\mathfrak{p}}-1)\sum_{j=1}^{e_\mathfrak{p}-1}k_{\tau,j}\equiv -k_{\tau,e_\mathfrak{p}}\pmod p.$$

Hence $p \nmid n$ and Lemma 1.26 (i) shows that the right most term in (18) has a strictly lower valuation than the other term on the right hand side. Thus, Lemma 1.25 (vi) shows that

$$d(r(f))_\tau = D' - \frac{nu\delta^{2-p^{f_{\mathfrak{p}(\tau)}}} \cdot \left(\prod_{j=1}^{f_{\mathfrak{p}(\tau)}-1} (h_{\phi^{-j}\circ\tau})^{p^j}\right) \cdot \left(\prod_{j \neq e_{\mathfrak{p}}} h_{\tau,j}\right)}{\delta^{n+1}} (dh_{\tau,e_{\mathfrak{p}}})_\tau,$$

where D' is a meromorphic section of $(\widetilde{\pi}^*\Omega^1_{X/\mathbb{F}})_\tau$ and the right most term has a strictly smaller valuation than D'. Combining this with Lemma 1.25 (vii), we get that $\mathrm{KS}_\tau(d(r(f))_\tau) \neq 0$. This implies that $\Theta_\tau(f) \neq 0$. □

Remark 1.29. When p is unramified in F, Proposition 1.28 can also be deduced from [Diamond and Sasaki 2023, Theorem 8.2.2] whose proof is different. Furthermore, Diamond and Sasaki [2023, Theorem 9.8.2] also determine the kernel of Θ_τ in terms of the partial Frobenius operator at τ that they define. Meanwhile, the case when p is ramified in F has been treated in [Diamond 2023]. Proposition 1.28 follows from [Diamond 2023, Theorem 5.2.1] and the kernel of Θ_τ is described in terms of partial Frobenius operators in [loc. cit., Corollary 9.1.2].

2. Doubling and Hecke algebras

2A. ***Hilbert modular forms of parallel weight 1.*** It is important to distinguish between Katz Hilbert modular forms defined on the fine moduli space and those on the coarse quotient by the action of the totally positive units of $\mathfrak{o}$. The latter enjoy the good Hecke theory for GL(2) and are the natural objects to study in relation with two dimensional Galois representations; see [Dimitrov and Wiese 2020]. In this section, we will define Hilbert modular forms of parallel weight building on Definition 1.1. Even though we give a definition valid in all levels $\mathfrak{n}$ that are prime to p, we nevertheless need to consider the following condition (which is stronger than the one we imposed in Section 1) expressing that $\mathfrak{n}$ is sufficiently divisible:

$\mathfrak{n}$ is divisible by a prime above a prime number q splitting completely in $F(\sqrt{\epsilon} \mid \epsilon \in \mathfrak{o}_+^\times)$ and such that $q \equiv -1 \pmod{4\ell}$ for all prime numbers ℓ such that $[F(\mu_\ell) : F] = 2$. (20)

This condition ensures that $\mathcal{X}^{\mathrm{DP}}$ is a scheme on which $[\epsilon] \in E = \mathfrak{o}_+^\times/\{\epsilon \in \mathfrak{o}^\times \mid \epsilon - 1 \in \mathfrak{n}\}^2$ acts properly and discontinuously by sending (A, λ, μ) to $(A, \epsilon\lambda, \mu)$; see [Dimitrov 2009, Lemma 2.1(iii)]. For any $\mathfrak{c} \in \mathcal{C}$, any $\mathbb{Z}_p$-algebra R, and any parallel weight k, this induces an action of E on $M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; R)$, whose invariants are denoted by $M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; R)^E$. The following definition is equivalent to the one used in [Dimitrov and Wiese 2020, Section 2.2].

Definition 2.1. If $\mathfrak{n}$ satisfies (20), then the space of Hilbert modular forms over a $\mathbb{Z}_p$-algebra R of (parallel) weight $\kappa \in \mathbb{Z}$ and level $\mathfrak{n}$ is given by

$$M_\kappa(\mathfrak{n}, R) = \bigoplus_{\mathfrak{c} \in \mathcal{C}} M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; R)^E,$$

where $k = \sum_{\sigma\in\Sigma} \kappa\sigma$. For a general level $\mathfrak{n}$, let $\mathfrak{q}_1 \neq \mathfrak{q}_2$ be primes such that both $\mathfrak{n}\mathfrak{q}_1$ and $\mathfrak{n}\mathfrak{q}_2$ satisfy (20) and define

$$M_\kappa(\mathfrak{n}, R) = M_\kappa(\mathfrak{n}\mathfrak{q}_1, R) \cap M_\kappa(\mathfrak{n}\mathfrak{q}_2, R),$$

where the intersection can be taken in $M_\kappa(\mathfrak{n}\mathfrak{q}_1\mathfrak{q}_2, R)$. Note that the primes $\mathfrak{q}_1$ and $\mathfrak{q}_2$ can be chosen from a set of primes of positive density and that the definition does not depend on this choice.

For $f \in M_\kappa(\mathfrak{n}, R)$, we let $\sum_{\mathfrak{b}\in\mathcal{I}\cup\{(0)\}} a(\mathfrak{b}, f)q^{\mathfrak{b}}$ be the adelic q-expansion of f, where $\mathcal{I}$ denotes the group of fractional ideals of F; see [Dimitrov and Wiese 2020, Section 2.6].

We denote by $S_\kappa(\mathfrak{n}, R)$ the R-submodule of $M_\kappa(\mathfrak{n}, R)$ consisting of Hilbert modular cuspforms.

For $f \in M_\kappa(\mathfrak{n}, R)$ and $\mathfrak{c} \in \mathcal{C}$, we will let $f_{\mathfrak{c}}$ denote the corresponding E-invariant element of $M_k^{\mathrm{Katz}}(\mathfrak{c}, \mathfrak{n}; \mathbb{R})$ or, equivalently, its geometric q-expansion at the cusp $\infty_{\mathfrak{c}}$; see [Dimitrov and Wiese 2020, Section 2.5]. Recall that when $\mathfrak{n}$ satisfies (20), $M_\kappa(\mathfrak{n}, R)$ is endowed with Hecke and diamond operators; see [Dimitrov and Wiese 2020, Sections 3.1–3.3]. When $\mathfrak{n}$ is not sufficiently divisible, Hecke and diamond operators exist nonetheless because they stabilise the intersection $M_\kappa(\mathfrak{n}\mathfrak{q}_1, R) \cap M_\kappa(\mathfrak{n}\mathfrak{q}_2, R)$, where the auxiliary primes $\mathfrak{q}_1, \mathfrak{q}_2$ may be chosen appropriately. When it is not clear from the context, a superscript between parentheses indicates the weight of the space of Hilbert modular forms on which an operator acts, e.g., $T_{\mathfrak{p}}^{(1)}$. Since we are interested in torsion coefficients, we let $M_\kappa(\mathfrak{n}, K/\mathcal{O}) = \varinjlim_n M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)$, where the inductive limit is taken by identifying $M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)$ with the subspace $M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n) \otimes_{\mathcal{O}} (\varpi\mathcal{O})$ of $M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^{n+1})$.

2B. *Doubling.* We shall rely on the following lifting result.

Lemma 2.2. *Suppose that* $\mathfrak{n}$ *satisfies* (20). *There exists a* $\kappa_0 \in \mathbb{Z}$ *such that for all* $\kappa \geqslant \kappa_0$ *and all* $n \in \mathbb{N}$, *the natural map*

$$M_\kappa(\mathfrak{n}, \mathcal{O}) \otimes_{\mathcal{O}} \mathcal{O}/\varpi^n \to M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)$$

is a Hecke equivariant isomorphism.

Proof. The proof of [Dimitrov and Wiese 2020, Lemma 2.2] works unchanged after replacing $\mathbb{Z}_p$ by $\mathcal{O}$ and p by ϖ^n. □

We also need a generalisation of the total Hasse invariant modulo ϖ^n.

Lemma 2.3. *Suppose that* $\mathfrak{n}$ *satisfies* (20). *For every* $n \in \mathbb{N}$, *there is a* $\kappa_n \in \mathbb{N}$ *such that* $(\kappa_n - 1)$ *is a multiple of* $(p-1)p^{n-1}$, *and a modular form* $h_n \in M_{\kappa_n - 1}(\mathfrak{n}, \mathcal{O}/\varpi^n)$ *having* q*-expansion equal to* 1 *at* $\infty_{\mathfrak{c}}$ *for all* $\mathfrak{c} \in \mathcal{C}$. *In particular, it does not vanish at any cusp.*

Proof. Let $h \in M_{p-1}(\mathfrak{n}, \mathbb{F})$ be the usual Hasse invariant see [Dimitrov and Wiese 2020, Section 3.4]. For r such that $r(p-1)$ is big enough to apply Lemma 2.2, the modular form $h^r \in M_{r(p-1)}(\mathfrak{n}, \mathcal{O})$ has q-expansion congruent to 1 modulo ϖ at each cusp $\infty_{\mathfrak{c}}$, $\mathfrak{c} \in \mathcal{C}$. A big enough power of it satisfies the required congruence relation and condition on the weight. □

The theory of generalised Θ-operators presented in Section 1E allows us to prove the following result.

Lemma 2.4. *Assume such an f exists. Assume that $\mathfrak{n}$ satisfies* (20). *Then there does not exist any $0 \neq f \in M_1(\mathfrak{n}, \mathbb{F})$ such that $f_\mathfrak{c}$ has minimal weight at a fixed prime $\mathfrak{p}$ dividing p* (*see Corollary 1.14*) *for all $\mathfrak{c} \in \mathcal{C}$ and such that $a(\mathfrak{b}, f) = 0$ for all ideals $\mathfrak{b} \subset \mathfrak{o}$ not divisible by $\mathfrak{p}$.*

Proof. The minimality of the weight at $\mathfrak{p}$ implies that there exists a $\tau \in \overline{\Sigma}_\mathfrak{p}$ such that $h_{\bar{\tau}}$ does not divide $f_\mathfrak{c}$ (the proof of Corollary 1.14 implies that this is true for all $\tau \in \overline{\Sigma}_\mathfrak{p}$). Let $\mathfrak{b} = (\xi)\mathfrak{c}^{-1}$. Then, by definition, $a_\xi = a(\mathfrak{b}, f)$ and this is zero unless $\mathfrak{p} \mid \mathfrak{b}$, in which case $\mathfrak{p} \mid (\xi)$. Thus, it follows that $\bar{\tau}_\mathfrak{c}(1 \otimes \xi) = 0$. By (19), this shows that the geometric q-expansion of $\Theta_\tau(f_\mathfrak{c})$ vanishes at $\infty_\mathfrak{c}$ for all $\mathfrak{c} \in \mathcal{C}$, i.e., $\Theta_\tau(f_\mathfrak{c}) = 0$, contradicting the injectivity criterion from Proposition 1.28. □

For $\mathfrak{p} \mid p$ and $n \in \mathbb{N}$, we define the $V_\mathfrak{p}$-operator by (see [Diamond 2021], improving on and correcting previous works such as [Emerton et al. 2017] and [Dimitrov and Wiese 2020], for the definition of $T_\mathfrak{p}^{(1)}$)

$$V_{\mathfrak{p},n} = \langle \mathfrak{p} \rangle^{-1}(h_n T_\mathfrak{p}^{(1)} - T_\mathfrak{p}^{(\kappa_n)} h_n)$$

with h_n and κ_n from Lemma 2.3. A simple computation on q-expansions (see [Dimitrov and Wiese 2020, Proposition 3.6]) shows that $V_{\mathfrak{p},n}$ has the following effect on adelic q-expansions:

$$a((0), V_{\mathfrak{p},n} f) = a((0), f)[\mathfrak{p}^{-1}],$$
$$a(\mathfrak{r}, V_{\mathfrak{p},n} f) = a(\mathfrak{p}^{-1}\mathfrak{r}, f)$$

for nonzero ideals $\mathfrak{r} \subseteq \mathfrak{o}$.

Proposition 2.5. *Let $\mathfrak{p} \mid p$ be a prime and assume that $\mathfrak{n}$ satisfies* (20).

(i) *If $f \in S_1(\mathfrak{n}, K/\mathcal{O})$ and $a(\mathfrak{b}, f) = 0$ for all ideals $\mathfrak{b} \subset \mathfrak{o}$ not divisible by $\mathfrak{p}$, then $f = 0$.*

(ii) *For all $n \in \mathbb{N}$, the "doubling map"*

$$(h_n, V_{\mathfrak{p},n}) : S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)^{\oplus 2} \xrightarrow{(f,g) \mapsto h_n f + V_{\mathfrak{p},n} g} M_{\kappa_n}(\mathfrak{n}, \mathcal{O}/\varpi^n)$$

is injective and compatible with the Hecke operators $T_\mathfrak{q}$ for $\mathfrak{q} \nmid \mathfrak{n}p$. The Hecke operator $T_\mathfrak{p}^{(\kappa_n)}$ acts on the image by the formula $T_\mathfrak{p}^{(\kappa_n)} \circ (h_n, V_{\mathfrak{p},n}) = (h_n, V_{\mathfrak{p},n}) \circ \left(\begin{smallmatrix} T_\mathfrak{p}^{(1)} & 1 \\ -\langle \mathfrak{p} \rangle & 0 \end{smallmatrix}\right)$. In particular, the image $W_{\mathfrak{p},n}$ of $(h_n, V_{\mathfrak{p},n})$ lies in the $\mathfrak{p}$-ordinary part of $M_{\kappa_n}(\mathfrak{n}, \mathcal{O}/\varpi^n)$ and is stable under all Hecke operators $T_\mathfrak{q}$ for $\mathfrak{q} \nmid \mathfrak{n}p$.

If $(p-1)$ does not divide $e_\mathfrak{p}$, then the same statements hold after replacing the spaces $S_1(\mathfrak{n}, K/\mathcal{O})$ and $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ by $M_1(\mathfrak{n}, K/\mathcal{O})$ and $M_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$, respectively.

Proof. (i) For $f \in S_1(\mathfrak{n}, \mathcal{O}/\varpi)$, the claim is precisely the content of Lemma 2.4, in view of Corollaries 1.14 and 1.16. The induction step from $n-1$ to n follows from the q-expansion principle and the exact sequence

$$0 \to S_1(\mathfrak{n}, \mathcal{O}/\varpi) \otimes_\mathcal{O} \varpi^{n-1}\mathcal{O} \to S_1(\mathfrak{n}, \mathcal{O}/\varpi^n) \to S_1(\mathfrak{n}, \mathcal{O}/\varpi^{n-1}).$$

(ii) The injectivity follows from (i) applied to the first component of an element in the kernel. The matrix is obtained from a calculation as in [Dimitrov and Wiese 2020, Lemma 3.7]. □

2C. ***Hecke algebras.*** For $\kappa \geqslant 1$ and $n \in \mathbb{N}$, we consider the following complete Artinian (resp. Noetherian) semi-local $\mathcal{O}$-algebras

$$\begin{aligned} \mathbb{T}_n^{(\kappa)} &= \mathrm{im}\big(\mathcal{O}[T_{\mathfrak{q}}, \langle \mathfrak{q}\rangle]_{\mathfrak{q}\nmid \mathfrak{n}p} \to \mathrm{End}_{\mathcal{O}}(M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n))\big), \\ \mathbb{T}_{\mathrm{cusp},n}^{(\kappa)} &= \mathrm{im}\big(\mathcal{O}[T_{\mathfrak{q}}, \langle \mathfrak{q}\rangle]_{\mathfrak{q}\nmid \mathfrak{n}p} \to \mathrm{End}_{\mathcal{O}}(S_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n))\big), \text{ resp.,} \\ \mathbb{T}^{(\kappa)} &= \mathrm{im}\big(\mathcal{O}[T_{\mathfrak{q}}, \langle \mathfrak{q}\rangle]_{\mathfrak{q}\nmid \mathfrak{n}p} \to \mathrm{End}_{\mathcal{O}}(M_\kappa(\mathfrak{n}, K/\mathcal{O}))\big) = \varprojlim_n \mathbb{T}_n^{(\kappa)}, \\ \mathbb{T}_{\mathrm{cusp}}^{(\kappa)} &= \mathrm{im}\big(\mathcal{O}[T_{\mathfrak{q}}, \langle \mathfrak{q}\rangle]_{\mathfrak{q}\nmid \mathfrak{n}p} \to \mathrm{End}_{\mathcal{O}}(S_\kappa(\mathfrak{n}, K/\mathcal{O}))\big) = \varprojlim_n \mathbb{T}_{\mathrm{cusp},n}^{(\kappa)}. \end{aligned} \tag{21}$$

Note that they all contain $\langle \mathfrak{p}\rangle$ for $\mathfrak{p} \mid p$ since p is relatively prime to $\mathfrak{n}$. Moreover, the restriction to the cusp space gives surjective morphisms $\mathbb{T}_n^{(\kappa)} \twoheadrightarrow \mathbb{T}_{\mathrm{cusp},n}^{(\kappa)}$ and $\mathbb{T}^{(\kappa)} \twoheadrightarrow \mathbb{T}_{\mathrm{cusp}}^{(\kappa)}$. We also consider the torsion free Hecke $\mathcal{O}$-algebra:

$$\mathbb{T}_{\mathcal{O}}^{(\kappa)} = \mathrm{im}\big(\mathcal{O}[T_{\mathfrak{q}}, \langle \mathfrak{q}\rangle]_{\mathfrak{q}\nmid \mathfrak{n}p} \to \mathrm{End}_{\mathcal{O}}(M_\kappa(\mathfrak{n}, \mathcal{O}))\big).$$

Let I_n be the annihilator of $\mathbb{T}_{\mathcal{O}}^{(\kappa)}$ acting on $M_\kappa(\mathfrak{n}, \mathcal{O}) \otimes_{\mathcal{O}} (\mathcal{O}/\varpi^n)$. Then we have natural surjective ring homomorphisms

$$\mathbb{T}_n^{(\kappa)} \twoheadrightarrow \mathbb{T}_{\mathcal{O}}^{(\kappa)}/I_n \quad \text{and} \quad \mathbb{T}^{(\kappa)} \twoheadrightarrow \mathbb{T}_{\mathcal{O}}^{(\kappa)},$$

the latter coming from the fact that the intersection $\bigcap_n I_n$ is zero. For sufficiently large κ, both homomorphisms are isomorphisms due to Lemma 2.2. However, this need no longer be true in our principal case of interest $\kappa = 1$ since the inclusions

$$M_1(\mathfrak{n}, \mathcal{O}) \otimes_{\mathcal{O}} (\mathcal{O}/\varpi^n) \hookrightarrow M_1(\mathfrak{n}, \mathcal{O}/\varpi^n) \quad \text{and} \quad M_1(\mathfrak{n}, \mathcal{O}) \otimes_{\mathcal{O}} (K/\mathcal{O}) \hookrightarrow M_1(\mathfrak{n}, K/\mathcal{O})$$

need not be isomorphisms, in general. The kernel of $\mathbb{T}^{(1)} \twoheadrightarrow \mathbb{T}_{\mathcal{O}}^{(1)}$ is a finitely generated torsion $\mathcal{O}$-module, which is isomorphic to the kernel of $\mathbb{T}_n^{(1)} \to \mathbb{T}_{\mathcal{O}}^{(1)}/I_n$ for $n \in \mathbb{N}$ sufficiently large. Recall that multiplication by the Hasse invariant h_n allows us to see $M_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ inside $M_{\kappa_n}(\mathfrak{n}, \mathcal{O}/\varpi^n)$ equivariantly for all Hecke operators $T_{\mathfrak{q}}$ and $\langle \mathfrak{q}\rangle$ for $\mathfrak{q}\nmid \mathfrak{n}p$, yielding a surjection $\mathbb{T}_n^{(\kappa_n)} \twoheadrightarrow \mathbb{T}_n^{(1)}$ (see Proposition 2.5). For a prime $\mathfrak{p} \mid p$, consider also the Hecke algebra:

$$\widetilde{\mathbb{T}}_n^{(\kappa_n)} = \mathbb{T}_n^{(\kappa_n)}[T_{\mathfrak{p}}^{(\kappa_n)}] \subset \mathrm{End}_{\mathcal{O}}(M_{\kappa_n}(\mathfrak{n}, \mathcal{O}/\varpi^n)). \tag{22}$$

Corollary 2.6. *Suppose that* $\mathfrak{n}$ *satisfies* (20). *Let* $\mathfrak{p} \mid p$. *Then for any* $n \in \mathbb{N}$, *there is a surjection sending* $T_{\mathfrak{p}}^{(\kappa_n)}$ *to* U (*considered as a polynomial variable*):

$$\widetilde{\mathbb{T}}_n^{(\kappa_n)} \twoheadrightarrow \mathbb{T}_{\mathrm{cusp},n}^{(1)}[T_{\mathfrak{p}}^{(1)}, U]/(U^2 - T_{\mathfrak{p}}^{(1)}U + \langle \mathfrak{p}\rangle).$$

The same statement holds after replacing $\mathbb{T}_{\mathrm{cusp},n}^{(1)}$ *by* $\mathbb{T}_n^{(1)}$, *provided* $(p-1)\nmid e_{\mathfrak{p}}$.

Proof. The injection from Proposition 2.5 gives a morphism $\widetilde{\mathbb{T}}_n^{(\kappa_n)} \to \mathrm{End}_{\mathcal{O}}(S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)^{\oplus 2})$ of $\mathcal{O}$-algebras compatible with $T_{\mathfrak{q}}$ and $\langle \mathfrak{q}\rangle$ for all $\mathfrak{q}\nmid \mathfrak{n}p$ and, hence, we get a surjection $\mathbb{T}_n^{(\kappa_n)} \twoheadrightarrow \mathbb{T}_{\mathrm{cusp},n}^{(1)}$. Moreover, $T_{\mathfrak{p}}^{(\kappa_n)}$ acts on the image of $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)^{\oplus 2}$ via the matrix $\left(\begin{smallmatrix} T_{\mathfrak{p}}^{(1)} & 1 \\ -\langle \mathfrak{p}\rangle & 0 \end{smallmatrix}\right)$, whence it is annihilated by its characteristic polynomial $U^2 - T_{\mathfrak{p}}^{(1)}U + \langle \mathfrak{p}\rangle$ and does not satisfy any nontrivial linear relation over $\mathbb{T}_n^{(1)}[T_{\mathfrak{p}}^{(1)}]$, thus proving the existence of the desired homomorphism. For the surjectivity, let us observe that the image of $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)^{\oplus 2}$ is contained in the $T_\mathfrak{p}^{(\kappa_n)}$-ordinary subspace of $M_{\kappa_n}(\mathfrak{n}, \mathcal{O}/\varpi^n)$, and that the endomorphism $T_\mathfrak{p}^{(\kappa_n)} + \langle\mathfrak{p}\rangle(T_\mathfrak{p}^{(\kappa_n)})^{-1}$ of the latter space acts on the former as $\left(\begin{smallmatrix} T_\mathfrak{p}^{(1)} & 0 \\ 0 & T_\mathfrak{p}^{(1)} \end{smallmatrix}\right)$. Finally, assuming $(p-1) \nmid e_\mathfrak{p}$ allows us to apply Proposition 2.5 with $M_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ instead of $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$, leading to the validity of the result with $\mathbb{T}_{\mathrm{cusp},n}^{(1)}$ replaced by $\mathbb{T}_n^{(1)}$. □

3. Pseudo-representations for weight 1 Hecke algebras

3A. ***Pseudo-representations of degree 2.*** In this section, we recall some definitions due to Chenevier [2014] and Calegari and Specter [2019].

Definition 3.1. Let R be a complete Noetherian local $\mathcal{O}$-algebra with maximal ideal $\mathfrak{m}$ and residue field $R/\mathfrak{m} = \mathbb{F}$ considered with its natural $\mathfrak{m}$-adic topology. An *R-valued pseudo-representation of degree* 2 of G_F is a tuple $P = (T, D)$ consisting of continuous maps $T, D : \mathrm{G}_F \to R$ such that

(i) D is a group homomorphism $\mathrm{G}_F \to R^\times$,

(ii) $T(1) = 2$ and $T(gh) = T(hg) = T(g)T(h) - D(g)T(g^{-1}h)$ for all $g, h \in \mathrm{G}_F$.

We extend $T : \mathrm{G}_F \to R$ to an R-linear map $R[\mathrm{G}_F] \to R$ and we denote this map by T as well.

Given $g \in \mathrm{G}_F$, we define $D(g-1) := D(g) - T(g) + 1$.

The pseudo-representation $P = (T, D)$ is said to be *unramified at* $\mathfrak{p}$ if $D(h-1) = T(g(h-1)) = 0$ for all $g \in \mathrm{G}_F$ and all $h \in \mathrm{I}_\mathfrak{p}$.

Any continuous representation $\rho : \mathrm{G}_F \to \mathrm{GL}_2(R)$ yields a degree 2 pseudo-representation $P_\rho = (\mathrm{tr} \circ \rho, \det \circ \rho)$. The converse is true when the semisimple representation $\bar{\rho} : \mathrm{G}_F \to \mathrm{GL}_2(\mathbb{F})$ corresponding to the residual pseudo-representation is absolutely irreducible; see [Chenevier 2014, Theorem 2.22]. Further, if ρ is unramified outside a finite set of places S, then so is P_ρ. Again, the converse is true in the residually absolutely irreducible case. This can be seen by applying [loc. cit.] to the Galois group of the maximal extension of F unramified outside S over F.

We introduce a notion of ordinarity inspired from Calegari and Specter [2019].

Definition 3.2. Let $\widetilde{P} = (P, \alpha_\mathfrak{p})$ with $P = (T, D)$ a degree 2 pseudo-representation of G_F over R and $\alpha_\mathfrak{p} \in R$ a root of $X^2 - T(\mathrm{Frob}_\mathfrak{p})X + D(\mathrm{Frob}_\mathfrak{p}) \in R[X]$.

We say that $\widetilde{P}$ is *ordinary at* $\mathfrak{p}$ of weight $\kappa \geqslant 1$, if for all $h, h' \in \mathrm{I}_\mathfrak{p}$ and all $g \in \mathrm{G}_F$ we have

(i) $D(h-1) = 0$ and $T(h-1) = \chi_p^{\kappa-1}(h) - 1$, where χ_p denotes the p-adic cyclotomic character,

(ii) $T\big(g(h - \chi_p^{\kappa-1}(h))(h' \mathrm{Frob}_\mathfrak{p} - \alpha_\mathfrak{p})\big) = 0$.

Remark 3.3. Note that our notion of $\mathfrak{p}$-ordinary pseudo-representations implies the one of Calegari and Specter [2019, Definition 2.5]. Let $P = (T, D) : \mathrm{G}_F \to R^2$ be a degree 2 pseudo-representation and let $(\overline{T}, \overline{D}) : \mathrm{G}_F \to \mathbb{F}^2$ be its residual pseudo-representation. Suppose there exists a lift $\mathrm{Frob}_\mathfrak{p} \in \mathrm{G}_{F_\mathfrak{p}}$ of the arithmetic Frobenius at $\mathfrak{p}$ such that the polynomial $X^2 - \overline{T}(\mathrm{Frob}_\mathfrak{p})X + \overline{D}(\mathrm{Frob}_\mathfrak{p})$ has *distinct* roots in $\mathbb{F}$. Then $(P, \alpha_\mathfrak{p})$ is a $\mathfrak{p}$-ordinary pseudo-representation in the sense of Definition 3.2 if and only if it is

$\mathfrak{p}$-ordinary in the sense of Calegari–Specter. However, if this hypothesis does not hold, then we expect that the two notions are not equivalent.

Let $\overline{P} = (\overline{T}, \overline{D}) : \mathrm{G}_F \to \mathbb{F}^2$ be a fixed degree 2 pseudo-representation unramified outside $\mathfrak{n}p\infty$. Denote by $P^{\mathrm{ps}} = (T^{\mathrm{ps}}, D^{\mathrm{ps}}) : \mathrm{G}_F \to (R^{\mathrm{ps}})^2$ the universal deformation of $\overline{P}$ unramified outside $\mathfrak{n}p\infty$ in the category of complete Noetherian local $\mathcal{O}$-algebras with residue field $\mathbb{F}$ and consider the quotient

$$R^{\mathrm{ps}}[X]/(X^2 - T^{\mathrm{ps}}(\mathrm{Frob}_{\mathfrak{p}})X + D^{\mathrm{ps}}(\mathrm{Frob}_{\mathfrak{p}})) \twoheadrightarrow R^{\mathrm{ord}} \tag{23}$$

which classifies pairs $(P, \alpha_{\mathfrak{p}})$ such that P is a deformation of $\overline{P}$ unramified outside $\mathfrak{n}p\infty$ and $(P, \alpha_{\mathfrak{p}})$ is ordinary at $\mathfrak{p}$ of weight κ. The universal ring R^{ord}, classifying deformations of $\overline{P}$ which are unramified outside $\mathfrak{n}p\infty$ and are ordinary at $\mathfrak{p}$ of weight κ, is the quotient of the ring $R^{\mathrm{ps}}[X]/(X^2 - T^{\mathrm{ps}}(\mathrm{Frob}_{\mathfrak{p}})X + D^{\mathrm{ps}}(\mathrm{Frob}_{\mathfrak{p}}))$ by the ideal generated by the set

$$\big\{D^{\mathrm{ps}}(h-1),\, T^{\mathrm{ps}}(h-1) - \chi_p^{\kappa-1}(h) + 1,\, T^{\mathrm{ps}}\big(g(h - \chi_p^{\kappa-1}(h))(h' \,\mathrm{Frob}_{\mathfrak{p}} - X)\big) \mid h, h' \in \mathrm{I}_{\mathfrak{p}},\, g \in \mathrm{G}_F\big\}$$

and a direct computation shows that R^{ord} is independent of the choice of $\mathrm{Frob}_{\mathfrak{p}}$.

Note that R^{ord} is a finite R^{ps}-algebra. As R^{ps} is a local ring, it follows that R^{ord} is a semi-local ring and all of its maximal ideals contain the unique maximal ideal $\mathfrak{m}^{\mathrm{ps}}$ of R^{ps}. After going modulo $\mathfrak{m}^{\mathrm{ps}}$ in R^{ord}, it is easy to see, using the description of the ideal from the previous paragraph, that the number of maximal ideals of R^{ord} is the number of distinct $\alpha \in \mathbb{F}$ such that $(\overline{P}, \alpha)$ is a $\mathfrak{p}$-ordinary pseudo-representation of weight κ.

Now suppose $\overline{P}$ is unramified at $\mathfrak{p}$ and $\kappa \equiv 1 \pmod{p-1}$. Then we have

$$\overline{T}\big(g(h - \chi_p^{\kappa-1}(h))(h' \,\mathrm{Frob}_{\mathfrak{p}} - X)\big) = \overline{T}(g(h-1)h' \,\mathrm{Frob}_{\mathfrak{p}}) - X\overline{T}(g(h-1)) = \overline{T}(h' \,\mathrm{Frob}_{\mathfrak{p}}\, g(h-1)) = 0.$$

Here we are repeatedly using the fact that $\overline{T}(g(h-1)) = 0$ for all $g \in G_F$ and $h \in \mathrm{I}_{\mathfrak{p}}$, which is a consequence of the assumption that $\overline{P}$ is unramified at $\mathfrak{p}$. Thus, in this case, we see that $(\overline{P}, \alpha)$ is a $\mathfrak{p}$-ordinary pseudo-representation of weight κ if and only if α is a root of the polynomial $X^2 - \overline{T}(\mathrm{Frob}_{\mathfrak{p}})X + \overline{D}(\mathrm{Frob}_{\mathfrak{p}})$. Hence, in this case, R^{ord} is a semi-local Noetherian ring with two maximal ideals if the polynomial $X^2 - \overline{T}(\mathrm{Frob}_{\mathfrak{p}})X + \overline{D}(\mathrm{Frob}_{\mathfrak{p}})$ has two distinct roots and it is a local Noetherian ring otherwise.

3B. ***Existence of an ordinary Hecke algebra-valued pseudo-representation.*** We continue to use the notation from Section 2. In particular, suppose that $\mathfrak{n}$ satisfies (20). Let $\mathfrak{m}$ be any maximal ideal of $\mathbb{T}^{(1)}$ (or equivalently of $\mathbb{T}_n^{(1)}$ for some n) and denote also by $\mathfrak{m}$ the maximal ideals of $\mathbb{T}^{(\kappa_n)}$ and $\mathbb{T}_n^{(\kappa_n)}$ defined as the pull-back of $\mathfrak{m} \subset \mathbb{T}_n^{(1)}$.

Lemma 3.4. *There exists a $\mathbb{T}_{n,\mathfrak{m}}^{(\kappa_n)}$-valued pseudo-representation $P_{n,\mathfrak{m}}^{(\kappa_n)}$ of G_F of degree 2 which is unramified at all primes $\mathfrak{q} \nmid \mathfrak{n}p$ and $P_{n,\mathfrak{m}}^{(\kappa_n)}(\mathrm{Frob}_{\mathfrak{q}}) = (T_{\mathfrak{q}}, \langle \mathfrak{q} \rangle)$. In particular, after replacing $\mathcal{O}$ by a finite unramified extension, there exists a unique semisimple Galois representation*

$$\bar{\rho}_{\mathfrak{m}} : \mathrm{G}_F \to \mathrm{GL}_2(\mathbb{T}^{(1)}/\mathfrak{m})$$

unramified outside $\mathfrak{n}p\infty$ satisfying

$$\mathrm{tr}(\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = T_{\mathfrak{q}} \pmod{\mathfrak{m}} \quad \textit{and} \quad \det(\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = \langle \mathfrak{q} \rangle \pmod{\mathfrak{m}}$$

for all primes $\mathfrak{q} \nmid \mathfrak{n}p$.

Proof. After enlarging K, we may assume that it contains all the eigenvalues of $\mathbb{T}^{(\kappa_n)}$ acting on $M_{\kappa_n}(\mathfrak{n}, \mathcal{O})$. The $\mathcal{O}$-algebra $\mathbb{T}^{(\kappa_n)}$ generated by the Hecke operators outside the level and p is torsion-free and reduced, hence $\mathbb{T}^{(\kappa_n)}_{\mathfrak{m}} \otimes_{\mathcal{O}} K = \prod_{g \in \mathcal{N}} K$ where $\mathcal{N}$ denotes the set of newforms occurring in $M_{\kappa_n}(\mathfrak{n}, \mathcal{O})_{\mathfrak{m}}$. As is well known, one can attach to each such eigenform g a G_F-pseudo-representation P_g of degree 2 unramified outside $\mathfrak{n}p\infty$ such that $P_g(\mathrm{Frob}_{\mathfrak{q}}) = (a(\mathfrak{q}, g), \psi_g(\mathfrak{q})\,\mathrm{N}(\mathfrak{q})^{\kappa_n - 1})$ for all $\mathfrak{q} \nmid \mathfrak{n}p$, where $\langle \mathfrak{q} \rangle g = \psi_g(\mathfrak{q}) g$. Since the natural homomorphism $\mathbb{T}^{(\kappa_n)}_{\mathfrak{m}} \to \mathbb{T}^{(\kappa_n)}_{\mathfrak{m}} \otimes_{\mathcal{O}} K$ is injective, in view of the Chebotarev Density Theorem, we obtain a $\mathbb{T}^{(\kappa_n)}_{\mathfrak{m}}$-valued G_F-pseudo-representation $P^{(\kappa_n)}_{\mathfrak{m}}$ unramified outside $\mathfrak{n}p\infty$ such that $P^{(\kappa_n)}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}}) = (T_{\mathfrak{q}}, \langle \mathfrak{q} \rangle\,\mathrm{N}(\mathfrak{q})^{\kappa_n - 1})$ for all $\mathfrak{q} \nmid \mathfrak{n}p$; see [Chenevier 2014, Corollary 1.14].

Note that $\mathrm{N}(\mathfrak{q})^{\kappa_n - 1} \equiv 1 \pmod{\varpi^n}$. Composing $P^{(\kappa_n)}_{\mathfrak{m}}$ with the surjection $\mathbb{T}^{(\kappa_n)}_{\mathfrak{m}} \twoheadrightarrow \mathbb{T}^{(\kappa_n)}_{n,\mathfrak{m}}$, we get the desired pseudo-representation. Finally $\mathbb{T}^{(\kappa_n)}_n/\mathfrak{m} = \mathbb{T}^{(1)}_n/\mathfrak{m}$ along with [Chenevier 2014, Theorem A] finishes the proof of the lemma. □

Let $R^{\mathrm{ps}}_{\mathfrak{m}}$ be the universal deformation ring of the corresponding degree 2 pseudo-representation $\bar{P}_{\mathfrak{m}} = (\mathrm{tr} \circ \bar{\rho}_{\mathfrak{m}}, \det \circ \bar{\rho}_{\mathfrak{m}})$ unramified outside $\mathfrak{n}p\infty$ in the category of complete Noetherian local $\mathcal{O}$-algebras with residue field $\mathbb{F}$ (chosen large enough in order to contain the residue field of $\mathbb{T}^{(1)}_{\mathfrak{m}}$). Using the surjection $\mathbb{T}^{(\kappa_n)}_{\mathfrak{m}} \twoheadrightarrow \mathbb{T}^{(\kappa_n)}_{n,\mathfrak{m}} \twoheadrightarrow \mathbb{T}^{(1)}_{n,\mathfrak{m}}$ and then passing to the projective limit $\mathbb{T}^{(1)}_{\mathfrak{m}} = \varprojlim_n \mathbb{T}^{(1)}_{n,\mathfrak{m}}$, we obtain the following result.

Corollary 3.5. *For any maximal ideal* $\mathfrak{m}$ *of* $\mathbb{T}^{(1)}$, *there exists a* $\mathbb{T}^{(1)}_{\mathfrak{m}}$*-valued pseudo-representation* $P^{(1)}_{\mathfrak{m}}$ *of* G_F *of degree* 2 *which is unramified for all primes* $\mathfrak{q} \nmid \mathfrak{n}p$ *and* $P^{(1)}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}}) = (T_{\mathfrak{q}}, \langle \mathfrak{q} \rangle)$. *It yields a surjection* $R^{\mathrm{ps}}_{\mathfrak{m}} \twoheadrightarrow \mathbb{T}^{(1)}_{\mathfrak{m}}$.

Note that for a maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{(\kappa_n)}_n$, the algebra $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\mathfrak{m}}$ is in general only semi-local (see (22)). By the main result of [Dimitrov and Wiese 2020], $\bar{\rho}_{\mathfrak{m}}$ is unramified at $\mathfrak{p}$, allowing us to consider the ideal

$$\tilde{\mathfrak{m}} = \big(\mathfrak{m}, (T^{(\kappa_n)}_{\mathfrak{p}})^2 - \mathrm{tr}(\widehat{\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}})})T^{(\kappa_n)}_{\mathfrak{p}} + \det(\widehat{\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}})})\big) \subset \widetilde{\mathbb{T}}^{(\kappa_n)}_n, \tag{24}$$

where $\mathrm{tr}(\widehat{\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}})})$ and $\det(\widehat{\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}})})$ are some lifts of $\mathrm{tr}(\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}}))$ and $\det(\bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{p}}))$, respectively in $\mathbb{T}^{(\kappa_n)}_n$. Note that the ideal $\tilde{\mathfrak{m}}$ does not depend on the choices of these lifts.

Let $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\tilde{\mathfrak{m}}}$ be the completion of $\widetilde{\mathbb{T}}^{(\kappa_n)}_n$ with respect to $\tilde{\mathfrak{m}}$. The algebra $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\tilde{\mathfrak{m}}}$ then has at most two local components. Let $R^{\mathrm{ord}}_{\mathfrak{m}}$ be the universal $\mathcal{O}$-algebra classifying deformations of $\bar{P}_{\mathfrak{m}}$ unramified outside primes dividing $\mathfrak{n}p\infty$ and ordinary at $\mathfrak{p}$ of weight 1 (see (23)).

Lemma 3.6. *There exists a* $\mathfrak{p}$*-ordinary* $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\tilde{\mathfrak{m}}}$*-valued pseudo-representation* $\widetilde{P}^{(\kappa_n)}_{n,\mathfrak{m}} = (P^{(\kappa_n)}_{n,\mathfrak{m}}, T^{(\kappa_n)}_{\mathfrak{p}})$ *of degree* 2 *and weight* 1 *of* G_F *such that* $P^{(\kappa_n)}_{n,\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}}) = (T_{\mathfrak{q}}, \langle \mathfrak{q} \rangle)$ *for all* $\mathfrak{q} \nmid \mathfrak{n}p$. *It yields a surjection* $R^{\mathrm{ord}}_{\mathfrak{m}} \twoheadrightarrow \widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\tilde{\mathfrak{m}}}$.

Proof. Let $\widetilde{\mathbb{T}}^{(\kappa_n)} = \mathbb{T}^{(\kappa_n)}[T^{(\kappa_n)}_{\mathfrak{p}}]$ and denote also by $\tilde{\mathfrak{m}}$ the ideal of $\widetilde{\mathbb{T}}^{(\kappa_n)}$ defined as the pull-back of $\tilde{\mathfrak{m}} \subset \widetilde{\mathbb{T}}^{(\kappa_n)}_n$. Let $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\tilde{\mathfrak{m}}}$ be the completion of $\widetilde{\mathbb{T}}^{(\kappa_n)}$ with respect to $\tilde{\mathfrak{m}}$.

We have $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\tilde{\mathfrak{m}}} \otimes_{\mathcal{O}} K = \prod_{g \in \widetilde{\mathcal{N}}} K$, where $\widetilde{\mathcal{N}}$ denotes the subset of $\mathcal{N}$ (see the proof of Lemma 3.4) consisting of newforms occurring in $M_{\kappa_n}(\mathfrak{n}, \mathcal{O})_{\tilde{\mathfrak{m}}}$. As $\mathfrak{p}$ does not divide $\mathfrak{n}$, any $g \in \widetilde{\mathcal{N}}$ is an eigenvector for $T^{(\kappa_n)}_{\mathfrak{p}}$ (resp. $\langle \mathfrak{p} \rangle$) whose eigenvalue $a(\mathfrak{p}, g)$ (resp. $\psi_g(\mathfrak{p})$) is necessarily a p-adic unit by (24), i.e., g is $\mathfrak{p}$-ordinary. By a result due to Hida and Wiles, when g is ordinary at all places dividing p, and to Saito

[2009] and Skinner [2009] in general, p-adic Galois representation ρ_g attached to g is ordinary at $\mathfrak{p}$, i.e., its restriction to $\mathrm{G}_{F_\mathfrak{p}}$ has a one-dimensional unramified quotient on which $\mathrm{Frob}_\mathfrak{p}$ acts by the (unique) p-adic unit root $\alpha_{\mathfrak{p},g}$ of the Hecke polynomial $X^2 - a(\mathfrak{p}, g)X + \psi_g(\mathfrak{p})\,\mathrm{N}(\mathfrak{p})^{\kappa_n - 1}$. This implies that $\alpha_{\mathfrak{p},g}$ is also a root of $X^2 - \mathrm{tr}(\rho_g)(\mathrm{Frob}_\mathfrak{p})X + \det(\rho_g)(\mathrm{Frob}_\mathfrak{p})$, for *any* choice of a Frobenius element $\mathrm{Frob}_\mathfrak{p} \in \mathrm{G}_{F_\mathfrak{p}}$. Thus, the pseudo-representation $P_g = (\mathrm{tr}(\rho_g), \det(\rho_g))$ is $\mathfrak{p}$-ordinary of weight κ_n with respect to $\alpha_{\mathfrak{p},g}$ in the sense of Definition 3.2.

Since $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\widetilde{\mathfrak{m}}}$ is a semi-local finite $\mathcal{O}$-algebra, applying Hensel's lemma to each local component shows that the polynomial $X^2 - T_\mathfrak{p}^{(\kappa_n)}X + \langle \mathfrak{p} \rangle\,\mathrm{N}(\mathfrak{p})^{\kappa_n - 1}$ admits a unique unit root U in $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\widetilde{\mathfrak{m}}}$. By the Chebotarev density theorem, gluing the $\mathfrak{p}$-ordinary pseudo-representations $\widetilde{P}_g = (P_g, \alpha_{\mathfrak{p},g})$ for all $g \in \widetilde{\mathcal{N}}$ gives us a $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\widetilde{\mathfrak{m}}}$-valued $\mathfrak{p}$-ordinary pseudo-representation $(P_\mathfrak{m}^{(\kappa_n)}, U)$ of weight κ_n such that $P_\mathfrak{m}^{(\kappa_n)}(\mathrm{Frob}_\mathfrak{q}) = (T_\mathfrak{q}, \langle \mathfrak{q} \rangle\,\mathrm{N}(\mathfrak{q})^{\kappa_n - 1})$ for all $\mathfrak{q} \nmid \mathfrak{n}p$. We have $\chi_p^{\kappa_n - 1}(g) \equiv 1 \pmod{\varpi^n}$ for all $g \in G_F$. Hence, the reduction of $(P_\mathfrak{m}^{(\kappa_n)}, U)$ to $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\widetilde{\mathfrak{m}}}$ is a $\mathfrak{p}$-ordinary pseudo-representation of weight 1. Note that by Hensel's lemma, U reduces to $T_\mathfrak{p}^{(\kappa_n)}$ in $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\widetilde{\mathfrak{m}}}$, since the former (resp. the latter) is the unique unit root of $X^2 - T_\mathfrak{p}^{(\kappa_n)}X + \langle \mathfrak{p} \rangle\,\mathrm{N}(\mathfrak{p})^{\kappa_n - 1}$ in $\widetilde{\mathbb{T}}^{(\kappa_n)}_{\widetilde{\mathfrak{m}}}$ (resp. in $\widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\widetilde{\mathfrak{m}}}$). As $\mathrm{N}(\mathfrak{q})^{\kappa_n - 1} \equiv 1 \pmod{\varpi^n}$ for all $\mathfrak{q} \nmid \mathfrak{n}p$, this completes the proof of the lemma. □

3C. *Proof of the main theorem.* In the proof of Theorem 0.1 we can assume without loss of generality that $\mathfrak{n}$ satisfies (20), because given any prime q, the Hecke algebra in level $\mathfrak{n}$ is a quotient of the one in level $\mathfrak{n}q$. Moreover, since the algebra $\mathbb{T}^{(1)}$ is semi-local, equal to the product of $\mathbb{T}^{(1)}_\mathfrak{m}$ where $\mathfrak{m}$ runs over its maximal ideals, it is enough to prove the theorem after localisation at $\mathfrak{m}$.

Recall that in Corollary 3.5 we constructed a $\mathbb{T}^{(1)}_\mathfrak{m}$-valued pseudo-representation $P^{(1)}_\mathfrak{m} = (T, D)$ of G_F, whose image under the surjective homomorphism $\mathbb{T}^{(1)}_\mathfrak{m} \twoheadrightarrow \mathbb{T}^{(1)}_{\mathrm{cusp},\mathfrak{m}} \twoheadrightarrow \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}$ will be denoted by $P^{(1)}_{n,\mathfrak{m}} = (T_n, D_n)$, for $n \in \mathbb{N}$. This gives the first row of the following commutative diagram:

$$\begin{array}{ccccc} R^{\mathrm{ps}}_\mathfrak{m} & \xrightarrow{\text{Corollary 3.5}} & \mathbb{T}^{(1)}_{\mathrm{cusp},\mathfrak{m}} & \longrightarrow & \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}} \\ \downarrow & & & & \downarrow \\ R^{\mathrm{ord}}_\mathfrak{m} & \xrightarrow{\text{Lemma 3.6}} & \widetilde{\mathbb{T}}^{(\kappa_n)}_{n,\widetilde{\mathfrak{m}}} & \xrightarrow{\text{Corollary 2.6}} & \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}, U]/(U^2 - T^{(1)}_\mathfrak{p}U + \langle \mathfrak{p} \rangle). \end{array} \tag{25}$$

The morphisms in the second row come from Lemma 3.6 and Corollary 2.6. Combining them, we see that $\widetilde{P}^{(1)}_{n,\mathfrak{m}} = (P^{(1)}_{n,\mathfrak{m}}, U)$ is a $\mathfrak{p}$-ordinary pseudo-representation of weight 1.

We now perform the key "doubling" step, as presented in [Calegari and Specter 2019, Proposition 2.10], and slightly improved upon, since the surjectivity of the composed map $R^{\mathrm{ord}}_\mathfrak{m} \to \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}, U]/(U^2 - T^{(1)}_\mathfrak{p}U + \langle \mathfrak{p} \rangle)$ will not be used in the sequel. One has

$$\mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}, U]/(U^2 - T^{(1)}_\mathfrak{p}U + \langle \mathfrak{p} \rangle) = \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}] \oplus U \cdot \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}].$$

Since $\widetilde{P}^{(1)}_{n,\mathfrak{m}}$ is ordinary at $\mathfrak{p}$ of weight 1, for all $g \in \mathrm{G}_F$ and $h \in \mathrm{I}_\mathfrak{p}$ the following equality holds:

$$T_n(gh\,\mathrm{Frob}_\mathfrak{p}) - T_n(g\,\mathrm{Frob}_\mathfrak{p}) = U(T_n(gh) - T_n(g)) \in \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}] \cap U\mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T^{(1)}_\mathfrak{p}] = \{0\},$$

hence $T_n(gh) = T_n(g)$, i.e., $P^{(1)}_{n,\mathfrak{m}}$ is unramified at $\mathfrak{p}$.

Note that U satisfies the following relations

$$U^2 - T_{\mathfrak{p}}^{(1)} U + \langle \mathfrak{p} \rangle = 0 \quad \text{and} \quad U^2 - T_n(\mathrm{Frob}_{\mathfrak{p}}) U + D_n(\mathrm{Frob}_{\mathfrak{p}}) = 0$$

in the ring $\mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}[T_{\mathfrak{p}}^{(1)}, U]/(U^2 - T_{\mathfrak{p}}^{(1)} U + \langle \mathfrak{p} \rangle)$. Indeed, the second relation follows from the fact that $(P^{(1)}_{n,\mathfrak{m}}, U)$ is a $\mathfrak{p}$-ordinary pseudo-representation of weight 1. As the former polynomial is minimal, one obtains the desired equality $(T_n(\mathrm{Frob}_{\mathfrak{p}}), D_n(\mathrm{Frob}_{\mathfrak{p}})) = (T_{\mathfrak{p}}^{(1)}, \langle \mathfrak{p} \rangle)$, in particular $T_{\mathfrak{p}}^{(1)} \in \mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}$. Letting n vary finishes the proof of Theorem 0.1 for $\mathbb{T}^{(1)}_{\mathrm{cusp},\mathfrak{m}}$.

In order to obtain the theorem for $\mathbb{T}^{(1)}_{\mathfrak{m}}$, we replace $\mathbb{T}^{(1)}_{\mathrm{cusp},\mathfrak{m}}$ by $\mathbb{T}^{(1)}_{\mathfrak{m}}$, $\mathbb{T}^{(1)}_{\mathrm{cusp},n,\mathfrak{m}}$ by $\mathbb{T}^{(1)}_{n,\mathfrak{m}}$, and $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ by $M_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ throughout. The arguments continue to work if we assume that $p-1$ does not divide $e_{\mathfrak{p}}$, which is used in Corollary 2.6.

Corollary 3.7. *Let $\mathfrak{p} \mid p$. Then $T_{\mathfrak{p}}^{(1)} \in \mathbb{T}^{(1)}_{\mathrm{cusp}}$, i.e., for all $n \in \mathbb{N}$, the Hecke operator $T_{\mathfrak{p}}^{(1)}$ acts on $S_1(\mathfrak{n}, \mathcal{O}/\varpi^n)$ by an element of $\mathbb{T}^{(1)}_{\mathrm{cusp},n}$. Moreover, if $(p-1) \nmid e_{\mathfrak{p}}$, then one also has $T_{\mathfrak{p}}^{(1)} \in \mathbb{T}^{(1)}$.*

3D. *Non-Eisenstein ideals.*

Definition 3.8. A maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{(\kappa)}$ (or of $\mathbb{T}^{(\kappa)}_n$) is called Eisenstein if the corresponding $(\overline{\mathbb{T}^{(\kappa)}_{\mathfrak{m}}/\mathfrak{m}})$-valued pseudo-representation of G_F is the sum of two $(\overline{\mathbb{T}^{(\kappa)}_{\mathfrak{m}}/\mathfrak{m}})$-valued characters, where $\overline{\mathbb{T}^{(\kappa)}_{\mathfrak{m}}/\mathfrak{m}}$ is an algebraic closure of $\mathbb{T}^{(\kappa)}_{\mathfrak{m}}/\mathfrak{m}$.

We now prove that in the non-Eisenstein case it suffices to consider the cuspidal Hecke algebra.

Proposition 3.9. *The localisation of the natural surjection $\mathbb{T}^{(\kappa)}_n \twoheadrightarrow \mathbb{T}^{(\kappa)}_{n,\mathrm{cusp}}$ at any non-Eisenstein maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{(\kappa)}_n$ is an isomorphism.*

Proof. It suffices to prove that the localisation of $M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)/S_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)$ at a non-Eisenstein ideal vanishes. By multiplication by a suitable power of h_n which does not vanish at any cusp (see Lemma 2.3), we can assume that κ is sufficiently large so that Lemma 2.2 applies yielding $M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n) = M_\kappa(\mathfrak{n}, \mathcal{O}) \otimes_{\mathcal{O}} (\mathcal{O}/\varpi^n)$. Hence the natural Hecke equivariant morphism

$$M_\kappa(\mathfrak{n}, \mathcal{O})/S_\kappa(\mathfrak{n}, \mathcal{O}) \to M_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)/S_\kappa(\mathfrak{n}, \mathcal{O}/\varpi^n)$$

is surjective. The former, however, can be Hecke equivariantly embedded into $M_\kappa(\mathfrak{n}, \mathbb{C})/S_\kappa(\mathfrak{n}, \mathbb{C})$ which is well known to be generated by Eisenstein series whose Galois representations are reducible. This proves the proposition. □

Henceforth we assume $\mathfrak{m}$ to be a non-Eisenstein ideal of $\mathbb{T}^{(1)}$, so that the corresponding residual Galois representation $\bar{\rho}_{\mathfrak{m}}$ is absolutely irreducible. Therefore, by combining Theorem 0.1 with a result of Chenevier [2014, Theorem 2.22], we get a representation

$$\rho_{\mathfrak{m}} : \mathrm{G}_F \to \mathrm{GL}_2(\mathbb{T}^{(1)}_{\mathfrak{m}}),$$

unramified outside $\mathfrak{n}p\infty$, and uniquely characterised by the property that for all primes $\mathfrak{q} \nmid \mathfrak{n}p$ one has $\mathrm{tr}(\rho_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = T_{\mathfrak{q}}$ and $\det(\rho_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}})) = \langle \mathfrak{q} \rangle$. By combining Theorem 0.1 with Proposition 3.9, we deduce that the pseudo-representation $P^{(1)}_{\mathfrak{m}}$ is unramified at all primes $\mathfrak{p} \mid p$ and, by the discussion after

Definition 3.1, we conclude that $\rho_{\mathfrak{m}}$ is unramified at these primes as well. Let S be the set of places of F dividing $\mathfrak{n}\infty$ and let $R^S_{F,\bar{\rho}_{\mathfrak{m}}}$ be the universal deformation ring of $\bar{\rho}_{\mathfrak{m}}$ unramified outside S in the category of complete Noetherian local $\mathcal{O}$-algebras with residue field $\mathbb{F}$. Hence $\rho_{\mathfrak{m}}$ induces an $\mathcal{O}$-algebra homomorphism $R^S_{F,\bar{\rho}_{\mathfrak{m}}} \to \mathbb{T}^{(1)}_{\mathfrak{m}}$.

As $\mathbb{T}^{(1)}$ is generated by $T_{\mathfrak{q}}$ and $\langle \mathfrak{q} \rangle$ for $\mathfrak{q} \nmid \mathfrak{n}p$ as an $\mathcal{O}$-algebra, we obtain the following result.

Corollary 3.10. *There exists a surjective homomorphism $R^S_{F,\bar{\rho}_{\mathfrak{m}}} \twoheadrightarrow \mathbb{T}^{(1)}_{\mathfrak{m}}$ of $\mathcal{O}$-algebras.*

Acknowledgements

The authors are indebted to Fabrizio Andreatta, Adel Betina, Frank Calegari, Fred Diamond, Payman Kassaei, Sheng-Chi Shih and Liang Xiao for many clarifying explanations and discussions. The debt to the works of Andreatta and Goren [2005] and Reduzzi and Xiao [2017] is evident.

The authors would particularly like to thank the anonymous referee for the careful reading, pointing out several inaccuracies and for the insightful comments which helped tremendously improve the manuscript.

The research leading to this article is jointly funded by the Agence National de Recherche ANR-18-CE40-0029 and the Fonds National de Recherche Luxembourg INTER/ANR/18/12589973 in the project *Galois representations*, *automorphic forms and their L-functions* (GALF).

Deo and Dimitrov would like to thank IISER Pune and ICTS Bangalore, where part of the work was done, for their hospitality, namely during the program *Perfectoid spaces* (Code: ICTS/perfectoid2019/09). Deo was partially supported by a Young Investigator Award from the Infosys Foundation, Bangalore and also by the DST FIST program - 2021 [TPN - 700661].

References

[Andreatta and Goren 2003] F. Andreatta and E. Z. Goren, "Geometry of Hilbert modular varieties over totally ramified primes", *Int. Math. Res. Not.* **2003**:33 (2003), 1786–1835. MR Zbl

[Andreatta and Goren 2005] F. Andreatta and E. Z. Goren, *Hilbert modular forms: mod p and p-adic aspects*, Mem. Amer. Math. Soc. **819**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Calegari and Geraghty 2018] F. Calegari and D. Geraghty, "Modularity lifting beyond the Taylor–Wiles method", *Invent. Math.* **211**:1 (2018), 297–433. MR Zbl

[Calegari and Specter 2019] F. Calegari and J. Specter, "Pseudorepresentations of weight one are unramified", *Algebra Number Theory* **13**:7 (2019), 1583–1596. MR Zbl

[Chenevier 2014] G. Chenevier, "The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings", pp. 221–285 in *Automorphic forms and Galois representations*, *I* (Durham, 2011), edited by F. Diamond et al., Lond. Math. Soc. Lect. Note Ser. **414**, Cambridge Univ. Press, 2014. MR Zbl

[Diamond 2021] F. Diamond, "Kodaira–Spencer isomorphisms and degeneracy maps on Iwahori-level Hilbert modular varieties: the saving trace", preprint, 2021. arXiv 2111.10160

[Diamond 2023] F. Diamond, "Geometric weight-shifting operators on Hilbert modular forms in characteristic p", *J. Inst. Math. Jussieu* **22**:4 (2023), 1871–1930. MR Zbl

[Diamond and Kassaei 2017] F. Diamond and P. L. Kassaei, "Minimal weights of Hilbert modular forms in characteristic p", *Compos. Math.* **153**:9 (2017), 1769–1778. MR Zbl

[Diamond and Kassaei 2023] F. Diamond and P. L. Kassaei, "The cone of minimal weights for mod p Hilbert modular forms", *Int. Math. Res. Not.* **2023**:14 (2023), 12148–12171. MR Zbl

[Diamond and Sasaki 2023] F. Diamond and S. Sasaki, "A Serre weight conjecture for geometric Hilbert modular forms in characteristic p", *J. Eur. Math. Soc.* **25**:9 (2023), 3453–3536. MR Zbl

[Dimitrov 2004] M. Dimitrov, "Compactifications arithmétiques des variétés de Hilbert et formes modulaires de Hilbert pour $\Gamma_1(\mathfrak{c}, \mathfrak{n})$", pp. 527–554 in *Geometric aspects of Dwork theory*, *I* (Padova, Italy, 2001), edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR Zbl

[Dimitrov 2009] M. Dimitrov, "On Ihara's lemma for Hilbert modular varieties", *Compos. Math.* **145**:5 (2009), 1114–1146. MR Zbl

[Dimitrov and Tilouine 2004] M. Dimitrov and J. Tilouine, "Variétés et formes modulaires de Hilbert arithmétiques pour $\Gamma_1(\mathfrak{c}, \mathfrak{n})$", pp. 555–614 in *Geometric aspects of Dwork theory*, *II* (Padova, Italy, 2001), edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR Zbl

[Dimitrov and Wiese 2020] M. Dimitrov and G. Wiese, "Unramifiedness of Galois representations attached to Hilbert modular forms mod p of weight 1", *J. Inst. Math. Jussieu* **19**:2 (2020), 281–306. MR Zbl

[Emerton et al. 2017] M. Emerton, D. Reduzzi, and L. Xiao, "Unramifiedness of Galois representations arising from Hilbert modular surfaces", *Forum Math. Sigma* **5** (2017), art. id. e29. MR Zbl

[Kisin and Lai 2005] M. Kisin and K. F. Lai, "Overconvergent Hilbert modular forms", *Amer. J. Math.* **127**:4 (2005), 735–783. MR Zbl

[Reduzzi and Xiao 2017] D. A. Reduzzi and L. Xiao, "Partial Hasse invariants on splitting models of Hilbert modular varieties", *Ann. Sci. École Norm. Sup.* (4) **50**:3 (2017), 579–607. MR Zbl

[Saito 2009] T. Saito, "Hilbert modular forms and p-adic Hodge theory", *Compos. Math.* **145**:5 (2009), 1081–1113. MR Zbl

[Skinner 2009] C. Skinner, "A note on the p-adic Galois representations attached to Hilbert modular forms", *Doc. Math.* **14** (2009), 241–258. MR Zbl

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Tate 1997] J. Tate, "Finite flat group schemes", pp. 121–154 in *Modular forms and Fermat's last theorem* (Boston, MA, 1995), edited by G. Cornell et al., Springer, 1997. MR Zbl

[Taylor and Wiles 1995] R. Taylor and A. Wiles, "Ring-theoretic properties of certain Hecke algebras", *Ann. of Math.* (2) **141**:3 (1995), 553–572. MR Zbl

[Wiese 2014] G. Wiese, "On Galois representations of weight one", *Doc. Math.* **19** (2014), 689–707. MR Zbl

[Wiles 1995] A. Wiles, "Modular elliptic curves and Fermat's last theorem", *Ann. of Math.* (2) **141**:3 (1995), 443–551. MR Zbl

Communicated by Frank Calegari
Received 2021-11-11 Revised 2023-07-11 Accepted 2023-09-18

shaunakdeo@iisc.ac.in *Department of Mathematics, Indian Institute of Science, Bangalore, India*

mladen.dimitrov@univ-lille.fr *University of Lille, CNRS, UMR 8524 – Laboratoire Paul Painleve, Lille, France*

gabor.wiese@uni.lu *Department of Mathematics, University of Luxembourg, Grand Duchy of Luxembourg*

ALGEBRA AND NUMBER THEORY 18:8 (2024)
https://doi.org/10.2140/ant.2024.18.1497

Failure of the local-global principle for isotropy of quadratic forms over function fields

Asher Auel and V. Suresh

We prove the failure of the local-global principle, with respect to discrete valuations, for isotropy of quadratic forms in 2^n variables over function fields of transcendence degree $n \geq 2$ over an algebraically closed field of characteristic $\neq 2$. Our construction involves the generalized Kummer varieties considered by Borcea and by Cynk and Hulek as well as new results on the nontriviality of unramified cohomology of products of elliptic curves over discretely valued fields.

Introduction

The Hasse–Minkowski theorem states that if a quadratic form q over a number field is isotropic over every completion, then q is isotropic. This is the first, and most famous, instance of the *local-global principle* for isotropy of quadratic forms. Already for a function field of transcendence degree one over a number field, Witt [1935] found examples of the failure of the local-global principle for isotropy of quadratic forms in 3 variables (and also 4). Lind [1940] and Reichardt [1942], and later Cassels [1963], found examples of the failure of the local-global principle for isotropy of pairs of quadratic forms in 4 variables over $\mathbb{Q}$ (see [Aitken and Lemmermeyer 2011] for a detailed account), giving examples of quadratic forms over the function field $\mathbb{Q}(t)$ by an application of the Amer–Brumer theorem [Leep 2007; Elman et al. 2008, Theorem 17.14]. Cassels, Ellison, and Pfister [Cassels et al. 1971] found examples in 4 variables over the function field $\mathbb{R}(x, y)$.

Here, we are interested in the failure of the local-global principle for isotropy of quadratic forms over function fields of higher transcendence degree over algebraically closed fields. All our fields will be assumed to be of characteristic $\neq 2$ and all our quadratic forms nondegenerate. A quadratic form is called isotropic if it admits a nontrivial zero. If K is a field and v is a discrete valuation on K, we denote by K_v the fraction field of the completion (with respect to the v-adic topology) of the valuation ring of v. When we speak of the *local-global principle for isotropy of quadratic forms*, sometimes referred to as the strong Hasse principle, in a given dimension d over a given field K, we mean the following statement:

> If q is a quadratic form in d variables over K and q is isotropic over K_v for every discrete valuation v on K, then q is isotropic over K.

MSC2020: primary 11E04, 11R58, 12G05, 14G12; secondary 14H52, 14J32, 14J40, 19D45.
Keywords: quadratic forms, local-global principle, Hasse principle, function fields, unramified cohomology, elliptic curves.

Our main result is the following.

Theorem 1. *The local-global principle for isotropy of quadratic forms fails to hold in dimension 2^n over any function field K of transcendence degree $n \geq 2$ over an algebraically closed field k of characteristic $\neq 2$ other than possibly the algebraic closure of a finite field.*

Previously, only the case of $n = 2$ was known, with the first explicit examples over $K = \mathbb{C}(x, y)$ appearing in [Kim and Roush 1991], and later in [Bevelacqua 2004] and [Jaworski 2001]. For a construction, using algebraic geometry, over any transcendence degree 2 function field over an algebraically closed field of characteristic 0, see [Auel 2013; Auel et al. 2015, Section 6]. In a previous version of this work, Theorem 1 was proved in the case of complex rational function fields, and left as a conjecture. Though we no longer need to make use of it, in Section 6, we also prove a "geometric presentation lemma" of general interest about the existence of double covers of varieties admitting nontrivial unramified cohomology in maximal degree, which was conjectured in an earlier version of this work and was shown to imply Theorem 1.

We recall that by Tsen–Lang theory [Lang 1952, Theorem 6], such function fields are C_n-fields, hence have u-invariant 2^n, and thus all quadratic forms of dimension $> 2^n$ are already isotropic, thus we provide counterexamples to the local-global principle in the maximal dimension in which they could occur.

We mention that in the case of transcendence degree $n = 1$, where $K = k(X)$ for a smooth projective curve X over an algebraically closed field k, the local-global principle for isotropy of binary quadratic forms (the "global square theorem") holds when the genus of X is zero and fails when X has positive genus, see Remark 5.3.

Finally, when k is the algebraic closure of a finite field, our methods no longer work. Though one can use other techniques to handle the case of transcendence degree $n = 2$ (see Remark 5.4), proving the failure of the local-global principle for quadratic forms over function fields K of transcendence degree $n \geq 3$ over $\bar{\mathbb{F}}_p$ remains an open problem. Our method relies on proving the nontriviality of certain unramified cohomology classes in top degree, see Section 6. Already for $n = 3$, the existence of threefolds over $\mathbb{F}_p$ or $\bar{\mathbb{F}}_p$ admitting nontrivial unramified cohomology in degree 3 is an open problem related to the integral Tate conjecture, see [Colliot-Thélène and Kahn 2013, Question 5.4].

Our result relies on two new ingredients and one very useful trick. The trick, due to Bogomolov [1995] and outlined in Section 1, is a kind of refinement of the existence of transcendence bases, and allows us to reduce the construction of counterexamples to the local-global principle over general function fields to the case of rational function fields. Next, our construction over rational function fields makes use of so-called generalized Kummer varieties, first considered by Borcea [1992] and developed by Cynk and Hulek [2007], which are constructed as quotients of products of elliptic curves and are birationally double covers of rational varieties. Finally, we prove a new result (Theorem 3.3) on the nontriviality of unramified cohomology on products of elliptic curves, which provides an arithmetic generalization of a result of Gabber [Colliot-Thélène 2002, Appendice], see also Colliot-Thélène [2019].

1. Bogomolov's trick

Let K/k be a finitely generated field extension. Recall that K/k admits a finite *transcendence basis*, i.e., a set of elements $x_1, \dots, x_n \in K$ that are algebraically independent over k and such that $K/k(x_1, \dots, x_n)$ is a finite extension. The cardinality of any transcendence basis is equal to the transcendence degree of K/k.

A *projective model* of K/k is an integral projective k-variety X whose field of rational functions is k-isomorphic to K. By the classical Chow's lemma, every finitely generated field extension admits a projective model, where the dimension of the model coincides with the transcendence degree of the extension.

The following statement, a refinement of the existence of transcendence bases, can be traced back to Bogomolov, in the course of the proof of [Bogomolov 1995, Theorem 1.1], cf. [Bogomolov and Tschinkel 2012, Proposition 20].

Lemma 1.1 (Bogomolov's trick). *Let K/k be a finitely generated extension of transcendence degree n. Assume that k is infinite and that a projective model of K/k admits a smooth k-point. Then for any prime number p, there exists a transcendence basis $x_1, \dots, x_n \in K$ such that $K/k(x_1, \dots, x_n)$ is finite of degree prime to p.*

We remark that by the Lang–Nishimura theorem, see [Lang 1954; Nishimura 1955] and also [Reichstein and Youssin 2000, Proposition A.6], the existence of a smooth k-point on a projective model of K/k implies that any other projective model admits a k-point. The condition that a projective model admits a smooth k-point also implies that any model is geometrically integral and generically smooth, see [Stacks 2005–, Lemma 0CDW and Lemma 056V]. In particular, if k is algebraically closed, then any projective model of K/k admits a smooth k-point.

Proof. As above, since a projective model of K/k is geometrically integral, it is geometrically reduced, and hence K/k is separably generated by a result of MacLane, see [Eisenbud 1995, Theorem A1.3]. Hence, as in [Hartshorne 1977, Proposition I.4.9], there exists a projective hypersurface model $X \subset \mathbb{P}^{n+1}$ of K/k. Let d be the degree of X. If $d = 1$, then $X = \mathbb{P}^n$ and there is nothing to prove, so we can assume that $d > 1$.

Projection from a k-point in the complement of X (using that k is infinite) yields a dominant rational map $X \dashrightarrow \mathbb{P}^n$ of degree d. Indeed, it is dominant since the fibers of the projection are the intersections of X with the lines through the point, and such intersections are always nonempty, cf. [Hartshorne 1977, Theorem I.7.2]. Moreover, it is generically finite of degree d since any line through the point cannot be contained in X, hence must intersect X in a zero-dimensional scheme, which has length d. Similarly, projection from a smooth k-point P of X yields a dominant rational map $X \dashrightarrow \mathbb{P}^n$ of degree $d - 1$. Indeed, since P is a smooth point, the tangent space to X at P has codimension 1 in $\mathbb{P}^{n+1}$, hence (again using that k is infinite) the general line in $\mathbb{P}^{n+1}$ through P meets X transversally at P and thus intersects X in a nonempty zero-dimensional scheme of degree d containing P as an irreducible component. Then the general fiber of this projection, which is the complement of P in the intersection of X with a general line through P, is nonempty (using $d > 1$) and has length $d-1$, cf. [Harris 1992, Example 18.16]. Since d and $d-1$ are relatively prime, no prime number p can divide both, hence the associated extension of function fields $K = k(X)/k(\mathbb{P}^n) = k(x_1, \ldots, x_n)$ can be chosen of degree prime to p. □

We remark that the hypothesis on a projective model admitting a k-point is essential. For example, if K/k is the function field of a smooth plane conic X with no k-point, then there is no presentation of K as an odd degree extension of a rational function field $k(x)$. Indeed, X cannot acquire rational points over rational function fields (see [Elman et al. 2008, Lemma 7.15]) or extensions of odd degree (by Springer's theorem), but does acquire a rational point over its own function field.

We have the following immediate corollary of Bogomolov's trick.

Corollary 1.2. *Let K be a finitely generated field of transcendence degree n over an algebraically closed field k. Then there exists a transcendence basis $x_1, \ldots, x_n \in K$ such that $K/k(x_1, \ldots, x_n)$ is of odd degree.*

With this in mind, we now explain how Springer's theorem allows us to reduce the construction of counterexamples to the local-global principle for isotropy of quadratic forms over general function fields to the case of rational function fields.

Proposition 1.3. *Let q be a nondegenerate quadratic form over a field K' and let K/K' be a finite extension of odd degree. If q is a counterexample to the local-global principle for isotropy over K', then q_K is such a counterexample over K.*

Proof. By Springer's theorem, since q is anisotropic over K' and K/K' has odd degree, then q_K is anisotropic over K. To show that q_K is locally isotropic over K, let v be a discrete valuation on K, which then lies over a discrete valuation v' on K'. Since the completion K_v is a finite extension of the completion $K'_{v'}$ and since q is isotropic over $K'_{v'}$, we see that q_K is isotropic over K_v. □

2. Unramified cohomology of function fields

We now recall the notion of unramified cohomology, introduced in [Colliot-Thélène and Ojanguren 1989], restricting ourselves to mod 2 coefficients. Readers should consult the excellent survey [Colliot-Thélène 1995] for further details. Let k be a field of characteristic $\neq 2$ and K/k be a finitely generated extension. By a *discrete valuation v on K/k* we mean a rank 1 discrete valuation v on K that is trivial on k.

For each discrete valuation v on K/k with residue field $\kappa(v)$, recall the residue map in Galois cohomology

$$\partial_v : H^n(K, \mu_2^{\otimes n}) \to H^{n-1}(\kappa(v), \mu_2^{\otimes n-1})$$

which arises from the Gysin sequence associated to the closed point in the spectrum of the valuation ring R_v of v, see [Colliot-Thélène 1995, Section 3.3]. The residue map is uniquely determined by the property that $\partial_v((u_1)\cdots(fu_{n-1})\cdot(\pi_v)) = (\bar{u}_1)\cdots(\bar{u}_{n-1})$, where π_v is a uniformizer and $u_1, \ldots, u_{n-1}$ are units of R_v, and $\bar{u}$ means the image of a unit in $\kappa(v)$. The degree n unramified cohomology of K/k is defined by

$$H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n}) = \bigcap_v \ker(\partial_v : H^n(K, \mu_2^{\otimes n}) \to H^{n-1}(\kappa(v), \mu_2^{\otimes n-1}))$$

where the intersection ranges over all discrete valuations v on K/k. We say that an element $\alpha \in H^n(K, \mu_2^{\otimes n})$ is *unramified* if it belongs to $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n})$.

We recall two results about discrete valuations on rational function fields that will be useful later.

Proposition 2.1. (a) *Let k be a field and $K = k(x_1, \dots, x_n)$ a rational function field over k with $n \geq 1$. For each $1 \leq m \leq n$, there exists a discrete valuation v on K/k satisfying $v(x_i) = 1$ for all $1 \leq i \leq m$ and $v(x_i) = 0$ for all $m+1 \leq i \leq n$.*

(b) *Let k_0 be a field with a discrete valuation v_0 and residue field κ_0. Then there exists a discrete valuation v on the rational function field $K_0 = k_0(x_1, \dots, x_n)$, extending v_0 on k_0, and with residue field $\kappa_0(x_1, \dots, x_n)$.*

Proof. For (a), let A be the localization of $k[x_1, \cdots, x_n]$ at the prime ideal $(x_1, \cdots, x_m)$. Then $R = A[y_1, \cdots, y_{m-1}]/(x_m - x_1 y_1, \cdots, x_m - x_{m-1} y_{m-1})$ is an integral domain with field of fractions isomorphic to K. Furthermore, the ideal $\mathfrak{p}$ of R generated by the images of $x_1, \dots, x_m$ is a prime ideal and $R_{\mathfrak{p}}$ is a discrete valuation ring. The valuation on K/k given by this discrete valuation ring has the required properties. Geometrically, this corresponds to blowing up the model $\mathbb{P}^n$ of K/k along the linear subspace defined by $x_1 = \cdots = x_m = 0$.

For (b), letting $R_0 \subset k_0$ be the valuation ring of v_0 and π_0 a uniformizer, we take the discrete valuation v on K_0 associated to the prime ideal in $R_0[x_1, \dots, x_n] \subset K_0$ generated by π_0. By construction, the residue field of v on K_0 is $\kappa_0(x_1, \dots, x_n)$. Geometrically, this corresponds to the special fiber of the model $\mathbb{P}^n_{R_0}$. □

3. Generalized Kummer varieties

In this section, we review a construction, considered in the context of modular Calabi–Yau varieties [Cynk and Hulek 2007, Section 2; Cynk and Schütt 2009], of a generalized Kummer variety attached to a product of elliptic curves. This recovers, in dimension 2, the Kummer K3 surface associated to a decomposable abelian surface, and in dimension 3, a class of Calabi–Yau threefolds of CM type considered by Borcea [1992, Section 3]. We also prove some results about the unramified cohomology groups in top degree of products of elliptic curves and their associated generalized Kummer varieties.

Let $E_1, \dots, E_n$ be elliptic curves over an algebraically closed field k of characteristic $\neq 2$ and let $Y = E_1 \times \cdots \times E_n$. Let σ_i denote the negation automorphism on E_i and $E_i \to \mathbb{P}^1$ the associated quotient branched double cover. We extend each σ_i to an automorphism of Y by acting trivially on each E_j for $j \neq i$; the subgroup $G \subset \mathrm{Aut}(Y)$ they generate is an elementary abelian 2-group. Consider the exact sequence of abelian groups

$$1 \to H \to G \xrightarrow{\Pi} \mathbb{Z}/2 \to 0,$$

where Π is defined by sending each σ_i to 1. Then the product of the double covers $Y \to \mathbb{P}^1 \times \cdots \times \mathbb{P}^1$ is the quotient by G and we denote by $Y \to X$ the quotient by the subgroup H. The intermediate quotient $X \to \mathbb{P}^1 \times \cdots \times \mathbb{P}^1$ is a double cover, branched over a reducible divisor of type $(4, \dots, 4)$. For $n = 2$, this divisor is the union of 4 vertical fibers and 4 horizontal fibers of $\mathbb{P}^1 \times \mathbb{P}^1$ meeting in 16 points.

We point out that X is a singular degeneration of smooth Calabi–Yau varieties that (geometrically) admits a smooth Calabi–Yau model, see [Cynk and Hulek 2007, Corollary 2.3; Cynk and Schütt 2009, Section 4]. For $n = 2$, the minimal resolution of X is indeed isomorphic to the Kummer K3 surface $\mathrm{Kum}(E_1 \times E_2)$.

Given nontrivial classes $\gamma_i \in H^1_{\text{ét}}(E_i, \mu_2)$, we consider the cup product

$$\gamma = \gamma_1 \cdots \gamma_n \in H^n_{\text{ét}}(Y, \mu_2^{\otimes n}) \tag{1}$$

and its image in $H^n_{\text{ur}}(k(Y)/k, \mu_2^{\otimes n})$ under restriction to the generic point. These classes have been studied in [Colliot-Thélène 2002]. We remark that γ is in the image of the restriction map

$$H^n(k(\mathbb{P}^1 \times \cdots \times \mathbb{P}^1), \mu_2^{\otimes n}) \to H^n(k(Y), \mu_2^{\otimes n})$$

in Galois cohomology since each γ_i is in the image of the restriction map $H^1(k(\mathbb{P}^1), \mu_2) \to H^1(k(E_i), \mu_2)$.

We make this more explicit as follows. Corresponding to each double cover $E_i \to \mathbb{P}^1$, choose a Weierstrass equation in Legendre form

$$y_i^2 = x_i(x_i - 1)(x_i - \lambda_i) \tag{2}$$

where x_i is a coordinate on $\mathbb{P}^1$ and $\lambda_i \in k \setminus \{0, 1\}$, see [Silverman 1986, III.1.7]. Then the branched double cover $X \to \mathbb{P}^1 \times \cdots \times \mathbb{P}^1$ is birationally defined by the equation

$$y^2 = \prod_{i=1}^{n} x_i(x_i - 1)(x_i - \lambda_i) = f(x_1, \dots, x_n) \tag{3}$$

where $y = y_1 \cdots y_n$ in $k(Y)$, see [Cynk and Schütt 2009, Section 3]. Up to an automorphism, we can, and henceforth will, choose the Legendre forms so that the image of γ_i under the map $H^1_{\text{ét}}(E_i, \mu_2) \to H^1(k(E_i), \mu_2)$ coincides with the square class $(x_i) \in k(E_i)/k(E_i)^{\times 2} = H^1(k(E_i), \mu_2)$ of the rational function x_i, which is then visibly in the image of the restriction map $H^1(k(\mathbb{P}^1), \mu_2) \to H^1(k(E_i), \mu_2)$. Hence we see that the (ramified) cup product class $\xi = (x_1) \cdots (x_n) \in H^n(k(x_1, \cdots, x_n), \mu_2^{\otimes n})$, restricts to the unramified class $\gamma \in H^n_{\text{ur}}(k(Y)/k, \mu_2^{\otimes n})$.

The first main result of this section is that the class ξ already restricts to an unramified class over the quadratic extension $k(X)$. We prove a more general result that can be viewed as a higher dimensional generalization of [Colliot-Thélène 1995, Section 1].

Proposition 3.1. *Let k be an algebraically closed field of characteristic $\neq 2$ and $K = k(x_1, \dots, x_n)$ a rational function field over k. For $1 \le i \le n$, let $f_i(x_i) \in k[x_i]$ be polynomials of even degree satisfying $f_i(0) \neq 0$, and let $f = \prod_{i=1}^n x_i f_i(x_i)$. Then the restriction of the class $\xi = (x_1) \cdots (x_n) \in H^n(K, \mu_2^{\otimes n})$ to $H^n(K(\sqrt{f}), \mu_2^{\otimes n})$ is unramified with respect to all discrete valuations.*

Proof. Let $L = K(\sqrt{f})$ and v a discrete valuation on L with valuation ring R, maximal ideal $\mathfrak{m}$, and residue field κ. Write ξ_L for the restriction of ξ to $H^n(L, \mu_2^{\otimes n})$.

Suppose $v(x_i) < 0$ for some i. Let d_i be the degree of f_i and consider the reciprocal polynomial $f_i^*(x_i) = x_i^{d_i} f_i\big(\frac{1}{x_i}\big)$, so that $x_i f_i(x_i) = x_i^{d_i+1} \cdot \frac{1}{x_i} f_i^*\big(\frac{1}{x_i}\big)$. Since d_i is even, we have that the polynomials $x_i f_i(x_i)$ and $\frac{1}{x_i} f_i^*\big(\frac{1}{x_i}\big)$ have the same class in $K^\times/K^{\times 2}$.

Thus, up to replacing, for all i with $v(x_i) < 0$, the polynomial f_i by f_i^* in the definition of f and replacing x_i by $\frac{1}{x_i}$, we can assume that $v(x_i) \geq 0$ for all i without changing the extension L/K. Hence $k[x_1, \dots, x_n] \subset R_v$.

Consider $\mathfrak{p} = k[x_1, \dots, x_n] \cap \mathfrak{m}$. Then $\mathfrak{p}$ is a prime ideal of $k[x_1, \dots, x_n]$ whose residue field $\kappa(\mathfrak{p})$ is a subfield of κ. Let $K_\mathfrak{p}$ be the completion of K at $\mathfrak{p}$ and L_v the completion of L at v. Then $K_\mathfrak{p}$ is a subfield of L_v.

If $v(x_i) = 0$ for all i, then ξ_L is unramified at v. So suppose that $v(x_i) \neq 0$ for some i. By reindexing $x_1, \dots, x_n$, we assume that there exists $m \geq 1$ such that $v(x_i) > 0$ for $1 \leq i \leq m$ and $v(x_i) = 0$ for $m+1 \leq i \leq n$, i.e., we have $x_1, \dots, x_m \in \mathfrak{p}$ and $x_{m+1}, \dots, x_n \notin \mathfrak{p}$. In particular, the transcendence degree of $\kappa(\mathfrak{p})$ over k is $\leq n - m$.

First, suppose $f_i(x_i) \in \mathfrak{p}$ for some $m+1 \leq i \leq n$. Since $f_i(x_i)$ is a product of linear factors in $k[x_i]$, we have that $x_i - a_i \in \mathfrak{p}$ for some $a_i \in k$, with $a_i \neq 0$ since $f_i(0) \neq 0$. Thus the image of x_i in $\kappa(\mathfrak{p})$ is equal to a_i and hence is a square in $K_\mathfrak{p}$. In particular, x_i is a square in L_v, thus ξ_L is trivial (hence unramified) at v, cf. [Colliot-Thélène and Ojanguren 1989, Proposition 1.4].

Now, suppose that $f_i(x_i) \notin \mathfrak{p}$ for all $m+1 \leq i \leq n$. Then for each $1 \leq i \leq m$, we see that since $x_i \in \mathfrak{p}$ and $f_i(0) \neq 0$, we have $f_i(x_i) \notin \mathfrak{p}$. Consequently, we can assume that $f = x_1 \cdots x_m u$ for some $u \in k[x_1, \dots, x_n] \setminus \mathfrak{p}$. We remark that $f = x_1 \cdots x_m u$ is a square in L, so that $(x_1 \cdots x_m) = (u)$ in $H^1(L, \mu_2)$.

For $m = 1$, we see that $\xi_L = (u) \cdot (x_2) \cdots (x_n)$ is unramified at v since u and $x_2, \dots, x_n$ are units at v.

For $m > 1$, a computation with symbols

$$(x_1) \cdots (x_m) = (x_1) \cdots (x_{m-1}) \cdot (x_1 \cdots x_m) = (x_1) \cdots (x_{m-1}) \cdot (u) \in H^m(L, \mu_2^{\otimes m})$$

shows that $\xi_L = (x_1) \cdots (x_{m-1}) \cdot (u) \cdot (x_{m+1}) \cdots (x_n)$. Since u and $x_{m+1}, \dots, x_n$ are units at v, computing with the Galois cohomology residue homomorphism $\partial_v : H^n(L, \mu_2^{\otimes n}) \to H^{n-1}(\kappa(v), \mu_2^{\otimes n-1})$ from Section 2 shows that

$$\partial_v((x_1) \cdots (x_{m-1}) \cdot (u) \cdot (x_{m+1}) \cdots (x_n)) = \alpha \cdot (\bar{u}) \cdot (\bar{x}_{m+1}) \cdots (\bar{x}_n)$$

for some $\alpha \in H^{m-2}(\kappa(v), \mu_2^{\otimes m-2})$, where for any $h \in k[x_1, \dots, x_n]$, we write $\bar{h}$ for the image of h in $\kappa(\mathfrak{p}) \subset \kappa$. Since the transcendence degree of $\kappa(\mathfrak{p})$ over k is $\leq n - m$ and k is algebraically closed, we have that $\kappa(\mathfrak{p})$ has 2-cohomological dimension $\leq n - m$ by [Serre 2002, II.4.2 Proposition 11], so that $H^{n-m+1}(\kappa(\mathfrak{p}), \mu_2^{\otimes n-m+1}) = 0$. Since $\bar{u}, \bar{x}_i \in \kappa(\mathfrak{p})$, we then have that $(\bar{u}) \cdot (\bar{x}_{m+1}) \cdots (\bar{x}_n)$ is trivial. In particular, $\partial_v(\xi_L)$ is trivial, and hence ξ_L is unramified at v. Finally, we have shown that the restriction ξ_L is unramified at all discrete valuations on L. ☐

As an immediate consequence, we deduce the fact that the class ξ restricts to an unramified class over $k(X) = k(x_1, \dots, x_n)(\sqrt{f})$, where f is as in (3).

Proposition 3.2. *Let $E_1, \dots, E_n$ be elliptic curves over an algebraically closed field k of characteristic $\neq 2$, given in the Legendre form* (2), *with $K = k(x_1, \dots, x_n)$. Then the restriction of the class $\xi = (x_1) \cdots (x_n)$ in $H^n(K, \mu_2^{\otimes n})$ to $H^n(k(X), \mu_2^{\otimes n})$ is unramified at all discrete valuations.*

This unramified class on $k(X)$ restricts to the class γ on $k(Y)$ in (1), so without loss of generality, we will also call it γ. Finally, we will need conditions ensuring that our class γ is nontrivial over $k(X)$. For this, we must choose the elliptic curves $E_1, \dots, E_n$ more carefully, and we will then show that γ is nontrivial over $k(Y)$, hence is nontrivial over $k(X)$. We proceed as follows.

First, we choose a subfield $k_0 \subset k$ admitting a discrete valuation v_0. This is possible unless k is the algebraic closure of a finite field; this is why we must henceforth assume that k is not the algebraic closure of a finite field. Then we choose E_i defined over k_0 with Weierstrass equation (2) satisfying $v_0(\lambda_i) > 0$. Finally, we appeal to the following arithmetic version, which was inspired by Bogomolov [1992, Section 7], of a result of Gabber [Colliot-Thélène 2002, Appendice].

Theorem 3.3. *Let k_0 be a field with a discrete valuation v_0 whose residue field has characteristic $\neq 2$. Let $E_1, \dots, E_n$ be elliptic curves over k_0 given in the Legendre form* (2), *with $v_0(\lambda_i) > 0$ for all $1 \le i \le n$. Let $Y = E_1 \times \cdots \times E_n$ and k/k_0 be an algebraically closed extension. Then the class $\gamma \in H^n(k(Y), \mu_2^{\otimes n})$ in* (1) *is nontrivial.*

Proof. Let $K_0 = k_0(x_1, \dots, x_n)$ and let γ_0 be the restriction of the class $\xi_0 = (x_1) \cdots (x_n) \in H^n(K_0, \mu_2^{\otimes n})$ to $H^n(k_0(Y), \mu_2^{\otimes n})$. Letting κ_0 be the residue field of v_0, by Proposition 2.1(b) we can extend v_0 to a discrete valuation on K_0 with residue field $\kappa_0(x_1, \dots, x_n)$. We remark that each $x_i \in K_0$ is a unit with respect to this valuation. Since $k_0(Y)/K_0$ is a finite separable extension, we can further extend this valuation to a discrete valuation $\tilde{v}$ on $k_0(Y)$. Writing

$$k_0(Y) = k_0(x_1, \cdots, x_n)(\sqrt{x_1(x_1-1)(x_1-\lambda_1)}, \cdots, \sqrt{x_n(x_n-1)(x_n-\lambda_n)})$$

then since $\tilde{v}(\lambda_i) > 0$ and $\tilde{v}(x_i) = 0$ for all i, we have that the residue field of $\tilde{v}$ is

$$\tilde{\kappa} = \kappa_0(x_1, \cdots, x_n)(\sqrt{x_1 - 1}, \cdots, \sqrt{x_n - 1}).$$

Since each x_i is a unit at $\tilde{v}$, the class γ_0 is unramified at $\tilde{v}$, and has specialization $\tilde{\xi}_0 = (x_1) \cdots (x_n) \in H^n(\tilde{\kappa}, \mu_2^{\otimes n})$.

We now argue that $\tilde{\xi}_0$ is nontrivial, hence that γ_0 is nontrivial. To this end, by Proposition 2.1(a) there is a valuation v_n on $\kappa_0(x_1, \cdots, x_n)$ such that $v_n(x_i) = 0$ for $1 \le i \le n-1$ and $v_n(x_n) = 1$, and we denote by $\tilde{v}_n$ an extension to $\tilde{\kappa}$, which is separable over $\kappa_0(x_1, \dots, x_n)$ and unramified at $\tilde{v}_n$. Thus $\tilde{v}_n$ is trivial on the subfield

$$\tilde{\kappa}_n = \kappa_0(x_1, \cdots, x_{n-1})(\sqrt{x_1 - 1}, \cdots, \sqrt{x_{n-1} - 1})$$

and satisfies $\tilde{v}_n(x_n) = 1$. Then the residue field of $\tilde{v}_n$ is $\tilde{\kappa}_n(\sqrt{-1})$ and the residue of the class $\tilde{\xi}_0$ at $\tilde{v}_n$ is simply $(x_1)\cdots(x_{n-1})$. Repeatedly taking residues using this process, we arrive at the class $(x_1) \in H^1(\kappa_0(x_1)(\sqrt{-1}, \sqrt{x_1 - 1}), \mu_2)$, which is nontrivial, hence $\tilde{\xi}_0$ is nontrivial. Thus $\gamma_0 \in H^n(k_0(Y), \mu_2^{\otimes n})$ is nontrivial.

Now let k/k_0 be any algebraically closed field extension and let $\bar{k}_0$ be the algebraic closure of k_0 in k. First, we show that the restriction of γ_0 to $H^n(\bar{k}_0(Y), \mu_2^{\otimes n})$ is nontrivial. This is equivalent to the restriction of γ_0 to $H^n(l_0(Y), \mu_2^{\otimes n})$ being nontrivial for every finite algebraic extension l_0/k_0. Letting w_0 be an extension of v_0 to l_0, we still have that $w_0(\lambda_i) > 0$ for all i, so we can apply what we have already proved. Second, since γ_0 is unramified, its restriction to $H^n(\bar{k}_0(Y), \mu_2^{\otimes n})$ and further to $H^n(k(Y), \mu_2^{\otimes n})$, remains unramified and coincides with the class γ. Then we can appeal to the rigidity property for unramified cohomology, which implies that the restriction map $H^n_{\mathrm{ur}}(\bar{k}_0(Y)/\bar{k}, \mu_2^{\otimes n}) \to H^n_{\mathrm{ur}}(k(Y)/k, \mu_2^{\otimes n})$ is an isomorphism, see [Colliot-Thélène 1995, Section 4.4], showing that γ is nontrivial. □

Additional aspects and applications of the argument in the proof of Theorem 3.3 will be the subject of forthcoming work [Auel and Suresh 2023]. In particular, $\mu_2^{\otimes n}$ coefficients can be replaced by $\mu_\ell^{\otimes n}$ coefficients for any positive integer ℓ prime to the residue characteristic of k_0. We content ourselves with giving one application here, which is a new proof of (a generalization of) Gabber's result [Colliot-Thélène 2002, Appendice].

Corollary 3.4. *Let k be a field of characteristic $\neq 2$ and K/k an algebraically closed extension. Let $E_1, \dots, E_n$ be elliptic curves over K whose j-invariants are algebraically independent over k. Let $Y = E_1 \times \cdots \times E_n$. Then the class $\gamma \in H^n(K(Y), \mu_2^{\otimes n})$ in (1) is nontrivial.*

Proof. Since K is algebraically closed, each elliptic curve E_i can be put into Legendre form (2). Hence Y is defined over the field $k_0 = k(\lambda_1, \dots, \lambda_n)$. Since the j-invariant of E_i is a rational function in λ_i, the algebraic independence of $j(E_1), \dots, j(E_n)$ over k implies the algebraic independence of $\lambda_1, \dots, \lambda_n$ over k. By Proposition 2.1(a), there exists a discrete valuation v_0 on k_0 such that $v_0(\lambda_i) > 0$ for all i, and then we can apply Theorem 3.3. □

4. Hyperbolicity over a quadratic extension

Let K be a field of characteristic $\neq 2$. We will need the following result about isotropy of quadratic forms, generalizing a well-known result in the dimension four case, see [Scharlau 1985, Chapter 2, Lemma 14.2].

Proposition 4.1. *Let q be a quadratic form over K of dimension divisible by 4 and discriminant d, and let $L = K(\sqrt{d})$. If q is hyperbolic over L then q is isotropic over K.*

Proof. If $d \in K^{\times 2}$, then $K = L$ and there is nothing to prove, so suppose $d \notin K^{\times 2}$. To get a contradiction, we will assume q is anisotropic. Since q_L is hyperbolic, we then have $q \simeq \langle 1, -d \rangle \otimes q_1$ for some quadratic form q_1 over K, see [Scharlau 1985, Chapter 2, Theorem 5.2]. Since the dimension of q is divisible by four, the dimension of q_1 is divisible by two, and a computation of the discriminant shows that $d \in K^{\times 2}$, which is a contradiction. □

For $n \geq 1$ and $a_1, \dots, a_n \in K^\times$, recall the n-fold Pfister form

$$\langle\langle a_1, \dots, a_n \rangle\rangle = \langle 1, -a_1 \rangle \otimes \cdots \otimes \langle 1, -a_n \rangle$$

and the associated symbol $(a_1) \cdots (a_n)$ in the Galois cohomology group $H^n(K, \mu_2^{\otimes n})$. Then $\langle\langle a_1, \dots, a_n \rangle\rangle$ is hyperbolic if and only if $\langle\langle a_1, \dots, a_n \rangle\rangle$ is isotropic if and only if $(a_1) \cdots (a_n)$ is trivial. For the fact that isotropic Pfister forms are hyperbolic, see [Scharlau 1985, Chapter 4, Corollary 1.5]. The fact that the triviality of $(a_1) \cdots (a_n)$ implies the hyperbolicity of $\langle\langle a_1, \dots, a_n \rangle\rangle$ is a consequence of the Milnor conjectures for the Witt group, as proved by Voevodsky [2003] and Orlov, Vishik, Voevodsky [Orlov et al. 2007].

For $d \in K^\times$ and $n \geq 2$, we will consider quadratic forms of discriminant d related to n-fold Pfister forms, as follows. Write $\langle\langle a_1, \dots, a_n \rangle\rangle$ as $q_0 \perp \langle (-1)^n a_1 \cdots a_n \rangle$, then define

$$\langle\langle a_1, \dots, a_n; d \rangle\rangle = q_0 \perp \langle (-1)^n a_1 \cdots a_n d \rangle.$$

For example:

$$\langle\langle a; d \rangle\rangle = \langle 1, -ad \rangle,$$

$$\langle\langle a, b; d \rangle\rangle = \langle 1, -a, -b, abd \rangle,$$

$$\langle\langle a, b, c; d \rangle\rangle = \langle 1, -a, -b, -c, ab, ac, bc, -abcd \rangle,$$

for $n = 1, 2, 3$, respectively. We remark that every quadratic form of dimension 4 is similar to one of this type. We also remark that $\langle\langle a_1, \dots, a_n; d \rangle\rangle$ becomes isomorphic to $\langle\langle a_1, \dots, a_n \rangle\rangle$ over $K(\sqrt{d})$. In general, these quadratic forms are examples of *twisted Pfister forms* in the sense of Hoffmann [1996].

Proposition 4.2. *Assume $n \geq 2$. If $q = \langle\langle a_1, \dots, a_n; d \rangle\rangle$ and $L = K(\sqrt{d})$ then q is isotropic if and only if q_L is isotropic if and only if $(a_1) \cdots (a_n) \in H^n(L, \mu_2^{\otimes n})$ is trivial.*

Proof. If q is isotropic then q_L is isotropic. If q_L is isotropic, then as mentioned above, it is hyperbolic as it is a Pfister form, hence by Proposition 4.1 (since q has dimension 2^n and $n \geq 2$), q is isotropic over K. As previously mentioned above (and consequence of the Milnor conjectures), $(a_1) \cdots (a_n) \in H^n(L, \mu_2^{\otimes n})$ is trivial if and only if the Pfister form q_L is isotropic. □

This generalizes a well-known result about quadratic forms of dimension 4, see [Scharlau 1985, Chapter 2, Lemma 14.2].

5. Failure of the local global principle

In this section, we prove our main Theorem 1 by providing a construction of quadratic forms over function fields that are locally isotropic yet globally anisotropic. First we prove a general result about the generalized Pfister forms in Section 4.

Proposition 5.1. *Let k be an algebraically closed field of characteristic $\neq 2$ and K/k a finitely generated extension of transcendence degree $n \geq 2$. Let $a_1, \dots, a_n, d \in K^\times$ be such that the symbol $(a_1) \cdots (a_n)$ in $H^n(K, \mu_2^{\otimes n})$ becomes unramified over $L = K(\sqrt{d})$. Then the quadratic form $q = \langle\langle a_1, \dots, a_n; d \rangle\rangle$ is locally isotropic over K.*

Proof. Let v be a discrete valuation on K and w an extension to L, with completions K_v and L_w and residue fields $\kappa(v)$ and $\kappa(w)$, respectively. By assumption, the restriction of the symbol $(a_1)\cdots(a_n)$ to $H^n(L, \mu_2^{\otimes n})$ is unramified at w. By cohomological purity for discrete valuation rings (see [Colliot-Thélène 1995, Section 3.3]) we have a surjective map $H^n_{\text{ét}}(R_w, \mu_2^{\otimes n}) \to H^n_{\text{ur}}(L_w/k, \mu_2^{\otimes n})$ where $R_w \subset L_w$ is the valuation ring. By proper base change (see [SGA 4_3 1973, XII.5.5], see also a general result of Gabber [Stacks 2005–, Tag 09ZI]), we have an isomorphism $H^n_{\text{ét}}(R_w, \mu_2^{\otimes n}) \cong H^n(\kappa(w), \mu_2^{\otimes n})$. Since $\kappa(w)/k$ has transcendence degree $< n$ by Abhyankar's inequality [1956, Corollary 1(1)] and k is algebraically closed, we have that $\kappa(w)$ has 2-cohomological dimension $< n$ by [Serre 2002, II.4.2 Proposition 11]. From all this, we deduce that $H^n_{\text{ur}}(L_w/k, \mu_2^{\otimes n}) = 0$. In particular, the symbol $(a_1)\cdots(a_n)$ has trivial restriction to $H^n(L_w, \mu_2^{\otimes n})$. Thus by Proposition 4.2, we have that q_{K_v} is isotropic. Finally, as this holds for every discrete valuation v on K, the quadratic form q is locally isotropic over K. □

Now, we will utilize our constructions in Section 3. Let k be an algebraically closed field of characteristic $\neq 2$ that is not the algebraic closure of a finite field. Let $k_0 \subset k$ be a subfield with a discrete valuation v_0 whose residue field has characteristic $\neq 2$. Let $E_1, \dots, E_n$ be elliptic curves over k_0 given in the Legendre form (2), with $v_0(\lambda_i) > 0$ for all $1 \le i \le n$. Let $X \to \mathbb{P}^1 \times \cdots \times \mathbb{P}^1$ be the double cover defined by $y^2 = \prod_{i=1}^n x_i(x_i - 1)(x_i - \lambda_i) = f(x_1, \dots, x_n)$ in (3), and consider the quadratic form

$$q = \langle\langle x_1, \dots, x_n; f \rangle\rangle \tag{4}$$

over the rational function field $k(\mathbb{P}^1 \times \cdots \times \mathbb{P}^1) = k(x_1, \dots, x_n)$, as in Section 4.

Our main result is that for $n \ge 2$, the quadratic form q shows the failure of the local-global principle for isotropy, with respect to all discrete valuations, for quadratic forms of dimension 2^n over $k(x_1, \dots, x_n)$.

Theorem 5.2. *Let k be an algebraically closed field of characteristic $\neq 2$ that is not the algebraic closure of a finite field and assume $n \ge 2$. The quadratic form $q = \langle\langle x_1, \dots, x_n; f \rangle\rangle$ as in* (4) *is anisotropic over $k(x_1, \dots, x_n)$ yet is isotropic over the completion at every discrete valuation.*

Proof. Write $K = k(x_1, \dots, x_n)$ and $L = K(\sqrt{f})$. By Proposition 3.2, the restriction of the symbol $(x_1)\cdots(x_n) \in H^n(K, \mu_2^{\otimes n})$ to L is unramified. Hence Proposition 5.1 implies that q is locally isotropic at every discrete valuation on K.

The restriction of the symbol $(x_1)\cdots(x_n) \in H^n(K, \mu_2^{\otimes n})$ to L is nontrivial since its further restriction to $k(E_1 \times \cdots \times E_n)$ is nontrivial by Theorem 3.3. Hence Proposition 4.2 implies that q is anisotropic over K. □

Proof of Theorem 1. Using Bogomolov's trick (see Corollary 1.2), we find $x_1, \dots, x_n \in K$ such that $K/k(x_1, \dots, x_n)$ has odd degree. If q is as in (4), then by Theorem 5.2, q is anisotropic yet locally isotropic over $k(x_1, \dots, x_n)$. Finally, by Proposition 1.3, q is a counterexample to the local-global principle for isotropy over K. □

To give an explicit example, let $a, b, c \in \overline{\mathbb{Q}} \setminus \{0, 1\}$ be any algebraic integers all divisible by a common odd prime ideal in a number field containing them. For example, take $a = b = c = 3$. Then over the function field $K = \mathbb{C}(x, y, z)$, the quadratic form

$$q = \langle 1, x, y, z, xy, xz, yz, (x-1)(y-1)(z-1)(x-a)(y-b)(z-c)\rangle$$

is isotropic over every completion K_v associated to a discrete valuation v on K, and yet q is anisotropic over K.

Remark 5.3. Let k be any algebraically closed field of characteristic $\neq 2$. When K/k is a finitely generated field of transcendence degree 1, then $K = k(X)$ for a smooth projective curve X over k. Any binary quadratic form q over K is similar to $\langle\langle a\rangle\rangle = \langle 1, -a\rangle$ for some $a \in K^\times$, and q is isotropic if and only if a is a square. For any discrete valuation v on K, we have that q is isotropic over K_v if and only if a is a square in K_v, equivalently (since k is algebraically closed and characteristic $\neq 2$), $v(a)$ is even. Thus if q is locally isotropic at all discrete valuations on K then the divisor of the rational function a on X can be written as $2D$ for a divisor D on X. The divisor class of D is 2-torsion in $\mathrm{Pic}(X)$ and it is trivial if and only if a is a square in K. Conversely, if X admits a nontrivial 2-torsion element of $\mathrm{Pic}(X)$, then twice this element is the divisor of a rational function $a \in K$ and the local-global principle fails for $\langle\langle a\rangle\rangle$. Thus the local-global principle for isotropy fails for K if and only if the Picard group of X admits a nontrivial 2-torsion element, equivalently (again, since k is algebraically closed and characteristic $\neq 2$), the genus of X is positive. Equivalently, the local-global principle for isotropy holds for quadratic forms over K if and only if K/k is purely transcendental.

In fact, we see that Proposition 4.2 (and hence Proposition 5.1) is false for $n = 1$ by considering the trivial class in $H^1(K, \mu_2)$ and $d \in K^\times$ any nonsquare.

Remark 5.4. When k is the algebraic closure of a finite field of characteristic $\neq 2$, Theorem 5.2 still holds for $n = 2$ assuming that the elliptic curves E_1 and E_2 are not isogenous. Indeed, by Proposition 3.2, the restriction of the symbol $(x_1)\cdot(x_2) \in H^2(K, \mu_2^{\otimes 2})$ to L is still unramified, and the only thing left to verify is that it is nontrivial. We can check this by further restriction to $k(E_1 \times E_2)$, where the symbol is the restriction to the generic point of a class in $H^1_{\text{ét}}(E_1, \mu_2) \otimes H^1_{\text{ét}}(E_2, \mu_2)$ by Section 3. However, standard computations of the Brauer group of $E_1 \times E_2$, see [Skorobogatov and Zarhin 2012, Section 3], show that if E_1 is not isogenous to E_2, then in fact $\mathrm{Br}(E_1 \times E_2) \cong H^1_{\text{ét}}(E_1, \mu_2) \otimes H^1_{\text{ét}}(E_2, \mu_2)$, so that each such cup product class is indeed nontrivial in the Brauer group. Then, as before, Proposition 5.1 implies that the local-global principle for isotropy fails for q as in (4) over K, hence Theorem 1 also holds in this case.

6. A geometric presentation lemma

The method for producing locally isotropic but globally anisotropic quadratic forms of dimension 2^n over function fields of transcendence degree n presented in this work is different from the one employed in [Auel et al. 2015, Section 6] for $n = 2$. There, we first proved a kind of geometric presentation lemma about the existence of nontrivial unramified cohomology (in degree 2) over quadratic extensions. Specifically, using Hodge theory, we proved [loc. cit., Proposition 6.4] that given any smooth projective surface S over an algebraically closed field of characteristic zero, there exists a double cover $T \to S$ with T smooth and $H^2_{\mathrm{ur}}(k(T)/k, \mu_2^{\otimes 2}) = \mathrm{Br}(T)[2] \neq 0$. It has been an open question ever since whether such a geometric presentation lemma holds for unramified cohomology in higher degree.

Conjecture 6.1. *Let K be a finitely generated field of transcendence degree n over an algebraically closed field k of characteristic $\neq 2$. Then either $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n}) \neq 0$ or there exists a separable quadratic extension L/K such that $H^n_{\mathrm{ur}}(L/k, \mu_2^{\otimes n}) \neq 0$.*

Assuming this conjecture, we can give a more direct proof of the existence of quadratic forms representing a failure of the local-global principle for isotropy without using the construction involving generalized Kummer varieties in Section 3.

Proposition 6.2. *Let K be a finitely generated field of transcendence degree n over an algebraically closed field k of characteristic $\neq 2$. If Conjecture 6.1 holds for K, then the local-global principle for isotropy of quadratic forms fails to hold in dimension 2^n over K.*

Before proceeding with the proof of Proposition 6.2, we recall a standard application of the Milnor conjectures for the Witt group. Since we could not find a suitable reference, we also provide a proof.

Lemma 6.3. *Let K be a field of characteristic $\neq 2$. If K is a C_n-field then every element in $H^n(K, \mu_2^{\otimes n})$ is a symbol.*

Proof. By the Milnor conjectures for the Witt group, as proved by Voevodsky [2003] and Orlov, Vishik, Voevodsky [2007], there exists a surjective homomorphism $e_n : I^n(K) \to H^n(K, \mu_2^{\otimes n})$ taking n-fold Pfister forms to symbols, where $I^n(K)$ is the n-th power of the fundamental ideal of the Witt group of K. Thus it suffices to prove that every element in $I^n(K)$ is represented by a Pfister form. Let q be an anisotropic quadratic form representing a class in $I^n(K)$. By the Arason–Pfister Hauptsatz (see [Scharlau 1985, Chapter 4, Theorem 5.6]), q has dimension $\geq 2^n$, but since we are assuming that K is a C_n-field, every quadratic form of dimension $> 2^n$ is isotropic, hence q has dimension 2^n.

Now we recall that every anisotropic form q of dimension 2^n in $I^n(K)$ is similar to a Pfister form over (any field) K, see [Kahn 2008, Corollaire 4.3.7]. Indeed, let $K(q)$ be the function field of the projective quadric defined by q. Then $q_{K(q)} \in I^n(K(q))$. Since q is isotropic over $K(q)$, the anisotropic part of $q_{K(q)}$ over $K(q)$ has dimension smaller than 2^n, hence by the Arason–Pfister Hauptsatz must be zero, thus q is hyperbolic over $K(q)$. Being anisotropic over K and hyperbolic over $K(q)$, the quadratic form q is thus similar to a Pfister form over K, see [Scharlau 1985, Chapter 4, Theorem 5.4(i)].

Since K is assumed to be a C_n-field, and $I^{n+1}(K)$ is additively generated by $(n+1)$-fold Pfister forms by [loc. cit., Chapter 4, Lemma 5.5], which are hyperbolic as soon as they are isotropic by [loc. cit., Chapter 4, Corollary 1.5], we conclude that $I^{n+1}(K) = 0$. We now argue that any quadratic form in $I^n(K)$ that is similar to a Pfister form is actually a Pfister form. Indeed, if ψ is any Pfister form in $I^n(K)$ and $a \in K^\times$, then $\langle\langle a \rangle\rangle \otimes \psi = \psi \perp -a\psi$ is in $I^{n+1}(K) = 0$, hence $\psi \cong a\psi$. Thus our anisotropic quadratic form q in $I^n(K)$ is a Pfister form, proving the desired statement. □

We do not know, in the spirit of [Serre 2002, II.4.5 Remark 3] and [Krashen and Matzri 2015], whether the statement of Lemma 6.3 holds for Galois cohomology modulo ℓ for primes $\ell \neq 2$.

Proof of Proposition 6.2. First, by Lemma 6.3, every element in $H^n(K, \mu_2^{\otimes n})$ is a symbol since K is a C_n-field by Tsen–Lang theory [Lang 1952]. Proposition 5.1 (applied with $d = 1$) implies that for any symbol $(a_1) \cdots (a_n)$ in $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n})$, the n-fold Pfister form $\langle\langle a_1, \ldots, a_n \rangle\rangle$ is locally isotropic. If we assume that $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n}) \neq 0$, then taking a nontrivial unramified symbol $(a_1) \cdots (a_n)$, the n-fold Pfister form $\langle\langle a_1, \ldots, a_n \rangle\rangle$ is locally isotropic but is anisotropic by Proposition 4.2, giving a counterexample to the local-global principle for isotropy over K.

Now assume that $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n}) = 0$ and that $H^n_{\mathrm{ur}}(L/k, \mu_2^{\otimes n}) \neq 0$ for some separable quadratic extension $L = K(\sqrt{d})$ of K. By Tsen–Lang theory (e.g., [Serre 2002, II.4.5]), L is also a C_n-field, hence by Lemma 6.3 every element in $H^n(L, \mu_2^{\otimes n})$ is a symbol. Thus we can choose a nontrivial unramified symbol $(a_1) \cdots (a_n) \in H^n_{\mathrm{ur}}(L/k, \mu_2^{\otimes n})$. Since the corestriction map $H^n(L, \mu_2^{\otimes n}) \to H^n(K, \mu_2^{\otimes n})$ preserves unramified cohomology, and we have assumed that $H^n_{\mathrm{ur}}(K/k, \mu_2^{\otimes n}) = 0$, we see that the corestriction of $(a_1) \cdots (a_n)$ is trivial. By the restriction-corestriction exact sequence for Galois cohomology, see [Arason 1975, Satz 4.5] or [Serre 2002, I Section 2 Exercise 2], we have that $(a_1) \cdots (a_n)$ is in the image of the restriction map $H^n(K, \mu_2^{\otimes n}) \to H^n(L, \mu_2^{\otimes n})$, and thus we can take $a_1, \ldots, a_n \in K^\times$. Then by Proposition 5.1, the twisted Pfister form $\langle\langle a_1, \ldots, a_n; d \rangle\rangle$ is locally isotropic over K but globally anisotropic. □

However, under the hypothesis in which we prove Theorem 1, namely, that k is not the algebraic closure of a finite field, our method allows us to prove Conjecture 6.1.

Theorem 6.4. *Let k be an algebraically closed field of characteristic $\neq 2$. If k is not the algebraic closure of a finite field then Conjecture 6.1 holds for any finitely generated field K of transcendence degree n over k.*

Proof. By Bogomolov's trick (Corollary 1.2) we consider K as an extension K/K_0 of odd degree over a rational function field $K_0 = k(x_1, \ldots, x_n)$. By Theorem 3.3, the symbol $(x_1) \cdots (x_n) \in H^n(K_0, \mu_2^{\otimes n})$ is nontrivial over the (separable) quadratic extension $L_0 = K_0(\sqrt{f})$ for $f \in K$ defined by (3). Since K/K_0 and L/K_0 have relatively prime degree, $L = K \otimes_{K_0} L_0$ is a quadratic extension of K and L/L_0 has odd degree. Thus by a standard restriction-corestriction argument, the symbol $(x_1) \cdots (x_n)$ remains nontrivial when restricted from L_0 to L. By Proposition 3.2, it is unramified over L_0, hence it remains unramified over L. □

Remark 6.5. When k is the algebraic closure of a finite field of characteristic $\neq 2$, then Conjecture 6.1 holds for $n = 2$. Indeed, following the proof of Theorem 6.4, we only need to show that $(x_1) \cdot (x_2) \in H^2(K_0, \mu_2^{\otimes n})$ is nontrivial over the quadratic extension $L_0 = K_0(\sqrt{f})$, which follows from Remark 5.4.

Thus we have reduced Conjecture 6.1 to k the algebraic closure of a finite field. However, the construction of nontrivial higher degree unramified cohomology on varieties over a finite field (or the algebraic closure of a finite field) is an open problem. In degree 3, this is related to the integral Tate conjecture.

Currently, there are no known smooth projective threefolds over a finite field with nontrivial unramified cohomology in degree 3; investigating this is a favorite problem of Colliot-Thélène, see [Colliot-Thélène and Kahn 2013, Question 5.4]. The smallest known dimensions in which such varieties exist is 5 (see [Pirutka 2011]), and recently, 4 (see [Scavia and Suzuki 2022]). Of course, one wonders whether the cup product class on a product of three elliptic curves, as in Section 3, is nontrivial over a finite field. One might also investigate the same class on the associated generalized Kummer variety over a finite field.

Acknowledgments

We would like to thank the organizers of the summer school *ALGAR*: *Quadratic forms and local-global principles*, at the University of Antwerp, Belgium, July 3–7, 2017, where the authors obtained a preliminary version of the result. We also acknowledge the Shapiro Visitor Program in the Department of Mathematics at Dartmouth College, which enabled a prolonged visit, where the authors obtained the final version of the result. We would also like to thank Fedor Bogomolov, Jean-Louis Colliot-Thélène, David Leep, and Parimala for very helpful discussions. Auel received partial support from NSA Young Investigator grant H98230-16-1-0321, Simons Foundation grant 712097, and National Science Foundation grant DMS-2200845; Suresh from National Science Foundation grant DMS-1463882.

References

[Abhyankar 1956] S. Abhyankar, "On the valuations centered in a local domain", *Amer. J. Math.* **78**:2 (1956), 321–348. MR Zbl

[Aitken and Lemmermeyer 2011] W. Aitken and F. Lemmermeyer, "Counterexamples to the Hasse principle", *Amer. Math. Monthly* **118**:7 (2011), 610–628. MR Zbl

[Arason 1975] J. K. Arason, "Cohomologische invarianten quadratischer Formen", *J. Algebra* **36**:3 (1975), 448–491. MR Zbl

[Auel 2013] A. Auel, "Failure of the local-global principle for isotropy of quadratic forms over surfaces", *Oberwolfach Rep.* **10**:2 (2013), 1823–1825. Part of the conference report Quadratic forms and linear algebraic groups.

[Auel and Suresh 2023] A. Auel and V. Suresh, "Unramified cohomology of products with an elliptic curve", in preparation, 2023.

[Auel et al. 2015] A. Auel, R. Parimala, and V. Suresh, "Quadric surface bundles over surfaces", *Doc. Math.* Merkurjev 60th birtday issue (2015), 31–70. MR Zbl

[Bevelacqua 2004] A. J. Bevelacqua, "Four dimensional quadratic forms over $F(X)$ where $I_t^3 F(X) = 0$ and a failure of the strong Hasse principle", *Comm. Algebra* **32**:3 (2004), 855–877. MR Zbl

[Bogomolov 1992] F. A. Bogomolov, "Stable cohomology of groups and algebraic varieties", *Mat. Sb.* **183**:5 (1992), 3–28. In Russian; translated in *Russian Acad. Sci. Sb. Math.* **76**:1 (1993), 1–21. MR Zbl

[Bogomolov 1995] F. A. Bogomolov, "On the structure of Galois groups of the fields of rational functions", pp. 83–88 in *K-theory and algebraic geometry: connections with quadratic forms and division algebras* (Santa Barbara, CA, 1992), edited by B. Jacob and A. Rosenberg, Proc. Sympos. Pure Math. **58.2**, Amer. Math. Soc., Providence, RI, 1995. MR Zbl

[Bogomolov and Tschinkel 2012] F. Bogomolov and Y. Tschinkel, "Introduction to birational anabelian geometry", pp. 17–63 in *Current developments in algebraic geometry*, edited by L. Caporaso et al., Math. Sci. Res. Inst. Publ. **59**, Cambridge Univ. Press, 2012. MR Zbl

[Borcea 1992] C. Borcea, "Calabi–Yau threefolds and complex multiplication", pp. 489–502 in *Essays on mirror manifolds*, edited by S.-T. Yau, Int. Press, Hong Kong, 1992. MR Zbl

[Cassels 1963] J. W. S. Cassels, "Arithmetic on curves of genus 1, V: Two counterexamples", *J. Lond. Math. Soc.* **38** (1963), 244–248. MR Zbl

[Cassels et al. 1971] J. W. S. Cassels, W. J. Ellison, and A. Pfister, "On sums of squares and on elliptic curves over function fields", *J. Number Theory* **3** (1971), 125–149. MR Zbl

[Colliot-Thélène 1995] J.-L. Colliot-Thélène, "Birational invariants, purity and the Gersten conjecture", pp. 1–64 in *K-theory and algebraic geometry: connections with quadratic forms and division algebras* (Santa Barbara, CA, 1992), edited by B. Jacob and A. Rosenberg, Proc. Sympos. Pure Math. **58.1**, Amer. Math. Soc., Providence, RI, 1995. MR Zbl

[Colliot-Thélène 2002] J.-L. Colliot-Thélène, "Exposant et indice d'algèbres simples centrales non ramifiées", *Enseign. Math.* (2) **48**:1-2 (2002), 127–146. With an appendix by Ofer Gabber. MR Zbl

[Colliot-Thélène 2019] J.-L. Colliot-Thélène, "Cohomologie non ramifiée dans le produit avec une courbe elliptique", *Manuscripta Math.* **160**:3-4 (2019), 561–565. MR Zbl

[Colliot-Thélène and Kahn 2013] J.-L. Colliot-Thélène and B. Kahn, "Cycles de codimension 2 et H^3 non ramifié pour les variétés sur les corps finis", *J. K-Theory* **11**:1 (2013), 1–53. MR Zbl

[Colliot-Thélène and Ojanguren 1989] J.-L. Colliot-Thélène and M. Ojanguren, "Variétés unirationnelles non rationnelles: au-delà de l'exemple d'Artin et Mumford", *Invent. Math.* **97**:1 (1989), 141–158. MR Zbl

[Cynk and Hulek 2007] S. Cynk and K. Hulek, "Higher-dimensional modular Calabi–Yau manifolds", *Canad. Math. Bull.* **50**:4 (2007), 486–503. MR Zbl

[Cynk and Schütt 2009] S. Cynk and M. Schütt, "Generalised Kummer constructions and Weil restrictions", *J. Number Theory* **129**:8 (2009), 1965–1975. MR Zbl

[Eisenbud 1995] D. Eisenbud, *Commutative algebra: with a view toward algebraic geometry*, Grad. Texts in Math. **150**, Springer, 1995. MR Zbl

[Elman et al. 2008] R. Elman, N. Karpenko, and A. Merkurjev, *The algebraic and geometric theory of quadratic forms*, Amer. Math. Soc. Colloq. Publ. **56**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Harris 1992] J. Harris, *Algebraic geometry: a first course*, Grad. Texts in Math. **133**, Springer, 1992. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. **52**, Springer, 1977. MR Zbl

[Hoffmann 1996] D. W. Hoffmann, "Twisted Pfister forms", *Doc. Math.* **1** (1996), 67–102. MR Zbl

[Jaworski 2001] P. Jaworski, "On the strong Hasse principle for fields of quotients of power series rings in two variables", *Math. Z.* **236**:3 (2001), 531–566. MR Zbl

[Kahn 2008] B. Kahn, *Formes quadratiques sur un corps*, Cours Spécialisés **15**, Soc. Math. France, Paris, 2008. MR Zbl

[Kim and Roush 1991] K. H. Kim and W. Roush, "Quadratic forms over $\mathbb{C}[t_1, t_2]$", *J. Algebra* **140**:1 (1991), 65–82. MR Zbl

[Krashen and Matzri 2015] D. Krashen and E. Matzri, "Diophantine and cohomological dimensions", *Proc. Amer. Math. Soc.* **143**:7 (2015), 2779–2788. MR Zbl

[Lang 1952] S. Lang, "On quasi algebraic closure", *Ann. of Math.* (2) **55** (1952), 373–390. MR Zbl

[Lang 1954] S. Lang, "Some applications of the local uniformization theorem", *Amer. J. Math.* **76** (1954), 362–374. MR Zbl

[Leep 2007] D. B. Leep, "The Amer–Brumer theorem over arbitrary fields", preprint, 2007, available at https://www.ms.uky.edu/~leep/Amer-Brumer_theorem.pdf.

[Lind 1940] C.-E. Lind, *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven vom Geschlecht Eins*, dissertation, University of Uppsala, 1940.

[Nishimura 1955] H. Nishimura, "Some remarks on rational points", *Mem. Coll. Sci. Univ. Kyoto Ser. A. Math.* **29** (1955), 189–192. MR Zbl

[Orlov et al. 2007] D. Orlov, A. Vishik, and V. Voevodsky, "An exact sequence for $K_*^M/2$ with applications to quadratic forms", *Ann. of Math.* (2) **165**:1 (2007), 1–13. MR Zbl

[Pirutka 2011] A. Pirutka, "Sur le groupe de Chow de codimension deux des variétés sur les corps finis", *Algebra Number Theory* **5**:6 (2011), 803–817. MR Zbl

[Reichardt 1942] H. Reichardt, "Einige im Kleinen überall lösbare, im Grossen unlösbare diophantische Gleichungen", *J. Reine Angew. Math.* **184** (1942), 12–18. MR Zbl

[Reichstein and Youssin 2000] Z. Reichstein and B. Youssin, "Essential dimensions of algebraic groups and a resolution theorem for G-varieties", *Canad. J. Math.* **52**:5 (2000), 1018–1056. MR Zbl

[Scavia and Suzuki 2022] F. Scavia and F. Suzuki, "Non-algebraic geometrically trivial cohomology classes over finite fields", preprint, 2022. arXiv 2206.12732

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundl. Math. Wissen. **270**, Springer, 1985. MR Zbl

[Serre 2002] J.-P. Serre, *Galois cohomology*, Springer, 2002. MR Zbl

[SGA 4_3 1973] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 3: Exposés IX–XIX* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **305**, Springer, 1973. MR Zbl

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Grad. Texts in Math. **106**, Springer, 1986. MR Zbl

[Skorobogatov and Zarhin 2012] A. N. Skorobogatov and Y. G. Zarhin, "The Brauer group of Kummer surfaces and torsion of elliptic curves", *J. Reine Angew. Math.* **666** (2012), 115–140. MR Zbl

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Voevodsky 2003] V. Voevodsky, "Motivic cohomology with $\mathbb{Z}/2$-coefficients", *Publ. Math. Inst. Hautes Études Sci.* **98** (2003), 59–104. MR Zbl

[Witt 1935] E. Witt, "Über ein Gegenbeispiel zum Normensatz", *Math. Z.* **39**:1 (1935), 462–467. MR Zbl

Communicated by Jean-Louis Colliot-Thélène
Received 2022-08-09 Revised 2023-08-16 Accepted 2023-09-18

asher.auel@dartmouth.edu *Department of Mathematics, Dartmouth College, Hanover, NH, United States*

suresh.venapally@emory.edu *Department of Mathematics, Emory University, Atlanta, GA, United States*

ALGEBRA AND NUMBER THEORY 18:8 (2024)
https://doi.org/10.2140/ant.2024.18.1515

Application of a polynomial sieve: beyond separation of variables

Dante Bonolis and Lillian B. Pierce

Let a polynomial $f \in \mathbb{Z}[X_1, \dots, X_n]$ be given. The square sieve can provide an upper bound for the number of integral $\boldsymbol{x} \in [-B, B]^n$ such that $f(\boldsymbol{x})$ is a perfect square. Recently this has been generalized substantially: first to a power sieve, counting $\boldsymbol{x} \in [-B, B]^n$ for which $f(\boldsymbol{x}) = y^r$ is solvable for $y \in \mathbb{Z}$; then to a polynomial sieve, counting $\boldsymbol{x} \in [-B, B]^n$ for which $f(\boldsymbol{x}) = g(y)$ is solvable, for a given polynomial g. Formally, a polynomial sieve lemma can encompass the more general problem of counting $\boldsymbol{x} \in [-B, B]^n$ for which $F(y, \boldsymbol{x}) = 0$ is solvable, for a given polynomial F. Previous applications, however, have only succeeded in the case that $F(y, \boldsymbol{x})$ exhibits separation of variables, that is, $F(y, \boldsymbol{x})$ takes the form $f(\boldsymbol{x}) - g(y)$. In the present work, we present the first application of a polynomial sieve to count $\boldsymbol{x} \in [-B, B]^n$ such that $F(y, \boldsymbol{x}) = 0$ is solvable, in a case for which F does not exhibit separation of variables. Consequently, we obtain a new result toward a question of Serre, pertaining to counting points in thin sets.

1. Introduction

Fix an integer $m \geq 2$ and integers $d, e \geq 1$. Consider the polynomial

$$F(Y, \boldsymbol{X}) = Y^{md} + Y^{m(d-1)} f_1(\boldsymbol{X}) + \cdots + Y^m f_{d-1}(\boldsymbol{X}) + f_d(\boldsymbol{X}), \tag{1-1}$$

in which for each $1 \leq i \leq d$, $f_i \in \mathbb{Z}[X_1, \dots, X_n]$ is a form with $\deg f_i = m \cdot e \cdot i$. We are interested in counting

$$N(F, B) := \left|\{\boldsymbol{x} \in [-B, B]^n \cap \mathbb{Z}^n : \exists y \in \mathbb{Z} \text{ such that } F(y, \boldsymbol{x}) = 0\}\right|.$$

Trivially, $N(F, B) \ll B^n$; our main result proves a nontrivial upper bound. We assume in what follows that $f_d \not\equiv 0$, since otherwise $(0, \boldsymbol{X})$ is a solution to $F(Y, \boldsymbol{X}) = 0$ for all $\boldsymbol{X}$, and then $B^n \ll N(F, B) \ll B^n$. (Throughout, we use the convention that $A \ll_\kappa B$ if there exists a constant C, possibly depending on κ, such that $|A| \leq CB$.)

Theorem 1.1. *Fix $n \geq 3$. Fix integers $m \geq 2$ and $e, d \geq 1$. Let F be defined as in* (1-1), *with $f_d \not\equiv 0$. Suppose the weighted hypersurface $V(F(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ defined by $F(Y, \boldsymbol{X}) = 0$ is nonsingular over $\mathbb{C}$. Then*

$$N(F, B) \ll B^{n-1+\frac{1}{n+1}} (\log B)^{\frac{n}{n+1}}.$$

The implicit constant may depend on n, m, d, e, but is otherwise independent of F.

MSC2020: 11D45, 11D85, 11N36.
Keywords: thin sets, polynomial sieve.

The main progress achieved in Theorem 1.1 is for $n \geq 4$, $e \geq 2$, $d \geq 2$. The requirement that $n \geq 3$ occurs since a key step, Proposition 5.2, is not true for $n = 2$ (see Remark 5.4). In any case, for $n = 2, 3$ the result of Theorem 1.1 is superseded by results of Broberg in [5], as described below in (1-14) and (1-15). When $e = 1$, the variety $V(F(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \ldots, 1)$ is unweighted, so that in the setting of Theorem 1.1, to bound $N(F, B)$ it is equivalent to count points $[Y : X_1 : \cdots : X_n]$ with $|Y|, |X_i| \ll B$ on a nonsingular projective hypersurface of degree at least 2 in $\mathbb{P}^n$. Then the result of Theorem 1.1 (in the stronger form $N(F, B) \ll_{m,d,n,\varepsilon} B^{n-1+\varepsilon}$) has already been obtained by work of Heath-Brown and Browning, appearing in [6; 9; 10; 26; 27], as summarized by Salberger in [42]. Finally, when $d = 1$, the result of Theorem 1.1 (aside from uniformity in the coefficients of F) follows from recent work of the first author in [2] (see Remark 3.2).

The condition $m \geq 2$ is applied in two ways: first, in the construction of certain sieve weights (see Section 1.2 and the proof of Lemma 1.2), and second, in Section 3.3 when we pass from the weighted variety to an unweighted variety. For illustration, we also describe how an alternative approach to the sieve lemma, conditional on GRH, can be devised when $m = 1$ (see Section 3.2 and Remark 1.3).

Bounding $N(F, B)$ relates to a question of Serre on counting integral points in thin sets. Let $\mathcal{V}$ denote the affine variety

$$\mathcal{V} = \{(Y, \boldsymbol{X}) \in \mathbb{A}^{n+1} : F(Y, \boldsymbol{X}) = 0\}, \tag{1-2}$$

and consider the projection

$$\pi : \mathcal{V} \to \mathbb{A}^n, \quad (y, \boldsymbol{x}) \mapsto \boldsymbol{x}. \tag{1-3}$$

Under the hypotheses of Theorem 1.1, the set $Z = \pi(\mathcal{V}(\mathbb{Q}))$ is a *thin set of type II* in $\mathbb{A}^n_{\mathbb{Q}}$, in the nomenclature of Serre. Serre has posed a general question that can be interpreted in our present setting as asking whether it is possible to prove that

$$N(F, B) \ll B^{n-1} (\log B)^c \tag{1-4}$$

for some c. Previous work by Broberg [5] nearly settled Serre's conjecture for thin sets of type II in $\mathbb{P}^{n-1}$ for $n = 2, 3$; see (1-14) and (1-15) below. For $n \geq 4$, Theorem 1.1 represents new progress toward resolving Serre's question for certain thin sets of type II. Note that as $n \to \infty$, the bound in Theorem 1.1 approaches a bound of the strength (1-4). We provide general background on Serre's question, and state precisely how Theorem 1.1 relates to previous literature on this question, in Section 1.1 and Section 1.2.

To prove Theorem 1.1, we develop an appropriate polynomial sieve lemma, and then bound each contribution to the sieve using analytic, algebraic, and geometric ideas. A novel feature of this work is that we do not assume that $F(Y, \boldsymbol{X})$ exhibits separation of variables: that is, when $d \geq 2$, $F(Y, \boldsymbol{X})$ of the form (1-1) cannot in general be written as $F(Y, \boldsymbol{X}) = g(Y) - G(\boldsymbol{X})$ for polynomials g, G. A formal polynomial sieve lemma has been formulated previously in a level of generality that does not require separation of variables; see [8; 13]. However, in those works it has so far only been applied to count points on a variety that does exhibit separation of variables. To our knowledge, Theorem 1.1 is the first application of a polynomial sieve to produce an upper bound for $N(F, B)$ in a case without separation of variables. We state precisely how Theorem 1.1 relates to previous literature on so-called square, power, and polynomial sieves in Section 1.2.

A second strength of Theorem 1.1 is that the exponent in the upper bound for $N(F, B)$ is independent of e, where we recall that as a function of $\boldsymbol{X}$, F has highest degree $m \cdot e \cdot d$. For any given $\boldsymbol{x} \in [-B, B]^n$ such that $F(Y, \boldsymbol{x}) = 0$ is solvable, one observes that any solution y to $F(y, \boldsymbol{x}) = 0$ must satisfy $y \ll B^e$, and there can be at most md solutions y for the given $\boldsymbol{x}$ (or, equivalently, preimages under the projection π in (1-3)), since the coefficient of Y^{md} in $F(Y, \boldsymbol{X})$ is nonzero. Thus an alternative method to bound $N(F, B)$ (up to an implicit constant depending on md) would be to count all $(n+1)$-tuples $\{(y, \boldsymbol{x}) : y \ll B^e, x_i \ll B : F(y, \boldsymbol{x}) = 0\}$. Other potential methods might be sensitive to the role of e or size of d, m (see for example Remark 1.4), while in contrast both the method and the result of Theorem 1.1 do not depend on e (aside from a possible implicit constant).

Third, we note that the result of Theorem 1.1 is independent of the coefficients of F; the implicit constant depends only on F in terms of its degree. To accomplish this, we adapt a strategy of [27], also recently applied in a similar setting in [3], to show that either $N(F, B)$ is already acceptably small, or $\|F\| \ll B^{(mde)^{n+2}}$. In the latter case, we then show that any dependence on $\|F\|$ in the sieve method is at most logarithmic, which we show is allowable for the result in Theorem 1.1.

1.1. ***Context of Theorem 1.1 within the study of Serre's question on thin sets.*** Here we recall the notion of thin sets defined by Serre in [46, §9.1 p. 121] and [45, p. 19]. Let k be a field of characteristic zero and let V be an irreducible algebraic variety in $\mathbb{P}^n_k$ (respectively $\mathbb{A}^n_k$). A subset M of $V(k)$ is said to be a projective (respectively, affine) thin set of type I if there is a closed subset $W \subset V$, $W \neq V$, with $M \subset W(k)$ (i.e., M is not Zariski dense in V). A subset M of $V(k)$ is said to be a projective (respectively, affine) thin set of type II if there is an irreducible projective (respectively, affine) algebraic variety X with $\dim X = \dim V$, and a generically surjective morphism $\pi : X \to V$ of degree $d \geq 2$ with $M \subset \pi(X(k))$. Any thin set is a finite union of thin sets of type I and thin sets of type II. From now on we consider only $k = \mathbb{Q}$, although Serre's treatment considers any number field.

Given a thin set $M \subset \mathbb{A}^n_{\mathbb{Q}}$, define the counting function

$$M(B) := \bigl|\{\boldsymbol{x} \in M \cap \mathbb{Z}^n : \max_{1 \leq i \leq n} |x_i| \leq B\}\bigr|,$$

so that trivially $M(B) \ll B^n$ for all $B \geq 1$. A theorem of Cohen [16] (see also [46, Chapter 13, Theorem 1, p. 177]) shows that

$$M(B) \ll_M B^{n-1/2} (\log B)^{\gamma} \qquad \text{for some } \gamma < 1, \tag{1-5}$$

where $\ll_M$ denotes that the implicit constant can depend on the coefficients of the equations defining M. As Serre remarks, this bound is essentially optimal, since the thin set

$$M = \{\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{Z}^n : x_1 \text{ is a square}\} \tag{1-6}$$

has $M(B) \gg B^{n-1/2}$. However, this M arises from a morphism that is singular; it is reasonable to expect that the result can be improved under an appropriate nonsingularity assumption (such as in the setting of Theorem 1.1).

Now let $M \subset \mathbb{P}^{n-1}_{\mathbb{Q}}$ be a thin set in projective space. Define the height function $H(x)$ for $x = [x_1 : \cdots : x_n] \in \mathbb{P}^{n-1}_{\mathbb{Q}}$ such that $(x_1, \dots, x_n) \in \mathbb{Z}^n$ and $\gcd(x_1, \dots, x_n) = 1$ by $H(x) = \max_{1 \le i \le n} |x_i|$. Define the associated counting function

$$M_H(B) = \{x \in M(\mathbb{Q}) : H(x) \le B\}$$

so that trivially $M_H(B) \ll B^n$. Serre deduces in [46, Chapter 13, Theorem 3] from an application of (1-5) that

$$M_H(B) \ll_M B^{n-1/2}(\log B)^{\gamma} \qquad \text{for some } \gamma < 1. \tag{1-7}$$

Serre raises a general question in [46, p. 178]: is it possible to prove that

$$M_H(B) \ll B^{n-1}(\log B)^{c} \tag{1-8}$$

for some c? (The set (1-6) is not an example of a thin set here because if $M = \{[x_1^2 : x_2 : \cdots : x_n]\} \subset \mathbb{P}^{n-1}_{\mathbb{Q}}$ then for any $x_1 \ne 0$,

$$[x_1 : x_2 : \cdots : x_n] = x_1[x_1 : x_2 : \cdots : x_n] = [x_1^2 : x_1x_2 : \cdots : x_1x_n] \in M,$$

so that $M \supset \mathbb{P}^{n-1}_{\mathbb{Q}}$.)

1.1.1. *Results for thin sets of type I.* If Z is an irreducible projective variety in $\mathbb{P}^{n-1}_{\mathbb{Q}}$ of degree $d \ge 2$, Serre deduces from (1-7) that $Z_H(B) \ll_Z B^{\dim Z + 1/2}(\log B)^{\gamma}$ for some $\gamma < 1$. Serre asks if it is possible to prove that $Z_H(B) \ll_Z B^{\dim Z}(\log B)^{c}$ for some c. (This question is raised in both [46, p. 178] and [45, p. 27]. Serre provides an example of a quadric for which a logarithmic factor necessarily arises. See also the question in the case of a hypersurface in Heath-Brown [24, p. 227], formally stated in both nonuniform and uniform versions as [27, Conjectures 1 and 2].) This is now called the *dimension growth conjecture* (in the terminology of [7]), and is often described as the statement that

$$Z_H(B) \ll_{Z,\varepsilon} B^{\dim Z + \varepsilon} \qquad \text{for every } \varepsilon > 0. \tag{1-9}$$

A refined version, credited to Heath-Brown and known as the *uniform dimension growth conjecture*, is the statement that

$$Z_H(B) \ll_{n, \deg Z, \varepsilon} B^{\dim Z + \varepsilon} \qquad \text{for every } \varepsilon > 0. \tag{1-10}$$

In the case that $Z \subset \mathbb{P}^{n-1}_{\mathbb{Q}}$ is a nonsingular projective hypersurface of degree $d \ge 2$, as mentioned before, combined works of Browning and Heath-Brown have proved (1-10) for all $n \ge 3$. More generally, Browning, Heath-Brown and Salberger proved (1-10) for all geometrically integral varieties of degree $d = 2$ and $d \ge 6$ (see [27] and [12], respectively). Recent work of Salberger has proved (1-9) in all remaining cases, and has even proved the uniform version (1-10) for $d \ge 4$ [43]. See [14] for a helpful survey, statements of open questions, and new progress such as an explicit bound $Z_H(B) \le C d^E B^{\dim Z}$

when $\deg Z = d \geq 5$, for a certain $C = C(n)$ and $E = E(n)$. The resolution of the dimension growth conjecture means that attention now turns to thin sets of type II, the subject of the present article.

1.1.2. *Results for thin sets of type II.* We turn to the case of thin sets of type II, our present focus. Given a finite cover $\phi : X \to \mathbb{P}^{n-1}$ over $\mathbb{Q}$ with $n \geq 2$, X irreducible and ϕ of degree at least 2, set

$$N_B(\phi) = |\{P \in X(\mathbb{Q}) : H(\phi(P)) \leq B\}| \tag{1-11}$$

for the standard height function above. Serre's question asks whether

$$N_B(\phi) \ll_{\phi,n} B^{n-1}(\log B)^c \qquad \text{for some } c, \tag{1-12}$$

or in a uniform version,

$$N_B(\phi) \ll_{\deg \phi,n} B^{n-1}(\log B)^c \qquad \text{for some } c. \tag{1-13}$$

For $n = 2, 3$ work of Broberg via the determinant method proves cases of Serre's conjecture up to the logarithmic factor [5]. Precisely, for $\phi : X \to \mathbb{P}^1$ of degree $r \geq 2$, Broberg proves

$$N_B(\phi) \ll_{\phi,\varepsilon} B^{2/r+\varepsilon} \qquad \text{for any } \varepsilon > 0. \tag{1-14}$$

For $\phi : X \to \mathbb{P}^2$ of degree r, Broberg proves

$$N_B(\phi) \ll_{\phi,\varepsilon} B^{2+\varepsilon} \text{ for } r \geq 3, \qquad N_B(\phi) \ll_{\phi,\varepsilon} B^{9/4+\varepsilon} \text{ for } r = 2, \text{ for any } \varepsilon > 0. \tag{1-15}$$

For $n \geq 4$, the question remains open whether one can achieve $N_B(\phi) \ll B^{n-1+\varepsilon}$ for all $\varepsilon > 0$, although we record some progress on this for specific types of ϕ in Section 1.2.

Now recall the setting of Theorem 1.1 in this paper, and the affine variety $\mathcal{V} \subset \mathbb{A}^{n+1}$ defined in (1-2) according to the polynomial $F(Y, \boldsymbol{X})$. Under the hypotheses of Theorem 1.1, we have:

(i) The variety $\mathcal{V}$ is irreducible (see Remark 3.3).

(ii) The projection π has degree $dm > 1$ since $m \geq 2$.

Thus $Z = \pi(\mathcal{V}(\mathbb{Q}))$ is a thin set of type II in $\mathbb{A}^n_{\mathbb{Q}}$, and in particular Cohen's result (1-5) implies that

$$Z(B) = N(F, B) \ll_F B^{n-1/2}(\log B)^\gamma, \tag{1-16}$$

following the same reasoning as [46, Chapter 13, Theorem 2, p. 178]. Or, interpreting the setting of Theorem 1.1 as counting points on a finite cover ϕ of $\mathbb{P}^{n-1}$ as in (1-11), this shows

$$N_B(\phi) \ll N(F, B) \ll_\phi B^{n-1/2}(\log B)^\gamma.$$

Our new work, Theorem 1.1, improves on (1-16) for each $n \geq 3$, for F of the form (1-1) with $V(F(Y, \boldsymbol{X}))$ nonsingular, and approaches a uniform bound of the strength (1-13) as $n \to \infty$.

1.2. ***Context of Theorem 1.1 within sieve methods.*** We now recall a few recent developments of sieve methods in the context of counting solutions to Diophantine equations, with a particular focus on progress toward Serre's conjecture for type II sets, as described above.

1.2.1. *Square sieve.* Let $f(\boldsymbol{X}) \in \mathbb{Z}[X_1, \ldots, X_n]$ be a fixed polynomial. Let $\mathcal{B}$ be a "box," such as $[-B, B]^n$ or more generally $\prod_i [-B_i, B_i]$. In [25], Heath-Brown codified the square sieve to count the number of integral values $\boldsymbol{x} \in \mathcal{B}$ such that $f(\boldsymbol{x}) = y^2$ is solvable over $\mathbb{Z}$, building on a method of Hooley [31]. At its heart was a formal sieve lemma involving a character sum with Legendre symbols. Heath-Brown applied this in particular to improve the error term in an asymptotic for the number of consecutive square-free numbers in a range. In [40], Pierce developed a stronger version of the square sieve, with a sieving set comprised of products of two primes rather than primes; this effectively allows the underlying modulus to be larger relative to the box $\mathcal{B}$, by factoring the modulus and using the q-analogue of van der Corput differencing. Pierce applied this to prove a nontrivial upper bound for 3-torsion in class groups of quadratic fields [40]; Heath-Brown subsequently used this sieve method to prove there are finitely many imaginary quadratic fields having class group of exponent 5 [28]; Bonolis and Browning applied it to prove a uniform bound for counting rational points on hyperelliptic fibrations [3].

1.2.2. *Power sieve.* The square sieve has been generalized to a power sieve, in order to count integral values $\boldsymbol{x} \in \mathcal{B}$ with $f(\boldsymbol{x}) = y^r$ solvable, for a fixed $r \geq 2$. Recall the question of bounding $N_{\boldsymbol{B}}(\phi)$ as in (1-12). For any $n \geq 2$, in the special case that ϕ is a nonsingular cyclic cover of degree $r \geq 2$, Munshi observed this can be reduced to counting the number of integral values $\boldsymbol{x} \in [-B, B]^n$ with $F(x_1, \ldots, x_n) = y^r$ solvable, for a nonsingular form F of degree mr for some $m \geq 1$. To bound this, Munshi developed a formal sieve lemma involving a character sum in terms of multiplicative Dirichlet characters [39]. Munshi applied it to prove that

$$\left|\{\boldsymbol{x} \in [-B, B]^n : F(\boldsymbol{x}) = y^r \text{ is solvable over } \mathbb{Z}\}\right| \ll B^{n-1+\frac{1}{n}} (\log B)^{\frac{n-1}{n}} \tag{1-17}$$

Consequently, this proved $N_{\boldsymbol{B}}(\phi) \ll B^{n-1+\frac{1}{n}} (\log B)^{\frac{n-1}{n}}$ for nonsingular cyclic covers. (See [2, Remark 1] for a note on the history of this result; the exponents stated here are slightly different from those presented in [39].)

In [29] Heath-Brown and Pierce have strengthened the power sieve, by using a sieving set comprised of products of primes, generalizing the approach of [40]. They used this method to prove that for any polynomial $f(\boldsymbol{X}) \in \mathbb{Z}[X_1, \ldots, X_n]$ of degree $d \geq 3$ with nonsingular leading form, and for any $r \geq 2$,

$$|\{\boldsymbol{x} \in [-B, B]^n : f(\boldsymbol{x}) = y^r \text{ is solvable over } \mathbb{Z}\}| \ll \begin{cases} B^{n-1+\frac{n(8-n)+4}{6n+4}} (\log B)^2, & 2 \leq n \leq 8, \\ B^{n-1+\frac{1}{2n+10}} (\log B)^2, & n = 9, \\ B^{n-1-\frac{n-10}{2n+10}} (\log B)^2, & n \geq 10. \end{cases} \tag{1-18}$$

This proves Serre's conjecture (1-12) for $N_{\boldsymbol{B}}(\phi)$, for all nonsingular cyclic covers, for $n \geq 10$. Indeed, the bound achieved is even smaller than the general conjecture, which is reasonable due to the imposed nonsingularity assumption.

Independently, Brandes also developed a power sieve in [4], applied to counting sums and differences of power-free numbers.

1.2.3. *Polynomial sieve: with separation of variables.* The next significant generalization addressed counting $\boldsymbol{x} \in \mathcal{B}$ for which $g(y) = f(\boldsymbol{x})$ is solvable, for appropriate polynomials g, f. Here, a quite general framework for a polynomial sieve lemma was developed by Browning in [8]. Specifically, in that work, Browning applied the polynomial sieve lemma to count x_1, x_2 such that $g(y) = f(x_1, x_2)$ is solvable, for particular functions f, g, that enabled an application showing the sparsity of like sums of a quartic polynomial of one variable.

Bonolis [2] further developed a polynomial sieve lemma with a character sum involving trace functions. Applying this, he proved that for any polynomial $g \in \mathbb{Z}[Y]$ of degree $r \geq 2$, and any irreducible form $F \in \mathbb{Z}[X_1, \dots, X_n]$ of degree $e \geq 2$ such that the projective hypersurface $V(F)$ defined by $F = 0$ is nonsingular over $\mathbb{C}$, then

$$\left|\{\boldsymbol{x} \in [-B, B]^n : F(\boldsymbol{x}) = g(y) \text{ is solvable over } \mathbb{Z}\}\right| \ll B^{n-1+\frac{1}{n+1}} (\log B)^{\frac{n}{n+1}}. \tag{1-19}$$

(This improves (1-17) and recovers the result initially stated in [39]; see [2, Remark 1].) This can also be seen as an improvement on Cohen's theorem (1-16) for a special type of thin set (defined as the image of $\mathcal{V} = \{(y, \boldsymbol{x}) \in \mathbb{A}^{n+1} : F(\boldsymbol{x}) - g(y) = 0\}$ under $(y, \boldsymbol{x}) \mapsto \boldsymbol{x}$, under the assumption that $V(F)$ defines a nonsingular projective hypersurface). The special case of our Theorem 1.1 when $d = 1$ follows from [2, Theorem 1.1]; see Remark 3.2.

Notably, the method employed in [2] to prove (1-19) was the first to demonstrate nontrivial averaging over pairs of primes in the sieving set, and exploiting such a strategy is central to the strength of our main theorem. We explain explicitly the advantage of such averaging in equations (1-25) and (1-26), below. For now, we simply state abstractly that any polynomial sieve method tests the solvability of the desired equation modulo p for primes in a chosen sieving set $\mathcal{P}$. The outcome of applying a sieve lemma (such as Lemma 1.2 below) is that one must bound from above an expression roughly of the form $|\mathcal{P}|^{-2} \sum_{p \neq q \in \mathcal{P}} T(p, q)$, where $T(p, q)$ studies the solvability of the desired equation modulo pairs $p \neq q \in \mathcal{P}$. Previous to [2], papers applying any type of polynomial sieve produced an upper bound for $|T(p, q)|$ that was uniform over p, q and then summed trivially over $p \neq q \in \mathcal{P}$. Instead, averaging nontrivially over p, q exploits the fact that $T(p, q)$ is typically smaller than its worst (largest) upper bound.

Most recently, a geometric generalization of Browning's polynomial sieve lemma has been developed over function fields by Bucur, Cojocaru, Lalín and the second author in [13]. They pose an analogue of Serre's question (1-8) in that setting (also raised by Browning and Vishe [11]), and apply a polynomial sieve to prove a bound of analogous strength to (1-19), in the special case of nonsingular cyclic covers in a function field setting. It remains an interesting open question to achieve a stronger bound such as (1-18), or to prove results for finite covers that are noncyclic, in such a function field setting.

1.2.4. *Polynomial sieve: without separation of variables.* So far we have mentioned applications of a sieve lemma to count solutions to $G(Y, \boldsymbol{X}) = 0$ when G separates variables as $G(Y, \boldsymbol{X}) = g(Y) - f(\boldsymbol{X})$ for some polynomials g, f. More generally, it is reasonable to ask — and this is a motivation for the present paper — whether an appropriate polynomial sieve can be employed to count solutions to equations of the form $G(Y, \boldsymbol{X}) = 0$ where $G(Y, \boldsymbol{X}) \in \mathbb{Z}[Y, X_1, \dots, X_n]$ is a polynomial of degree D of the form

$$G(Y, \boldsymbol{X}) = Y^D + Y^{D-1} f_1(\boldsymbol{X}) + \cdots + Y f_{D-1}(\boldsymbol{X}) + f_D(\boldsymbol{X}), \tag{1-20}$$

where each f_i is a form of degree $i \cdot e$, and we assume that the weighted hypersurface $V(G(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ defined by $G(Y, \boldsymbol{X}) = 0$ is nonsingular. Define

$$N(G, B) := \left|\{\boldsymbol{x} \in [-B, B]^n : \exists y \in \mathbb{Z} \text{ such that } G(y, \boldsymbol{x}) = 0\}\right|.$$

Under the assumption $f_D \not\equiv 0$, the aim is to improve on the trivial bound $N(G, B) \ll B^n$. To be clear, the formal sieve lemmas appearing in [8; 13] include this level of generality, but have only been applied to prove a bound for $N(G, B)$ when separation of variables occurs. In this paper we accomplish the first application of the polynomial sieve without assuming separation of variables, but under the additional assumption that the degree D of $G(Y, \boldsymbol{X})$ defined in (1-20) factors as $D = md$ for some $m \geq 2$, and all powers of Y that appear are divisible by m. (To see why this restriction is useful, see the proof of Lemma 1.2; for an alternative approach when $m = 1$, conditional on GRH, see Remark 1.3 and Section 3.2.)

The strength of our approach hinges on a particular formulation of the polynomial sieve, given in Lemma 1.2. It is worthwhile to compare our formulation with the polynomial sieve presented in [8, Theorem 1.1]. In [8, Theorem 1.1], the sieve weight system, adapted to counting solutions to (1-20), is defined as follows:

$$w_{p,\mathrm{Bro}}(\boldsymbol{k}) = \alpha + (\nu_p(\boldsymbol{k}) - 1)(D - \nu_p(\boldsymbol{k})),$$

in which $\nu_p(\boldsymbol{k}) = |\{y \in \mathbb{F}_p : G(y, \boldsymbol{k}) = 0 \in \mathbb{F}_p\}|$. (These weights are then applied in an inequality analogous to (3-1) below, to derive a sieve lemma.) Consequently, if $G(Y, \boldsymbol{k}) = 0$ is solvable over $\mathbb{Z}$, the conditions $1 \leq \nu_p(\boldsymbol{k}) \leq D$ and $\alpha > 0$ guarantee that $w_{p,\mathrm{Bro}}(\boldsymbol{k}) > 0$ *for any* p. In our approach, we consider simpler weights:

$$w_p(\boldsymbol{k}) = \nu_p(\boldsymbol{k}) - 1.$$

Thus, in our situation, if $G(Y, \boldsymbol{k}) = 0$ is solvable over $\mathbb{Z}$, we can only conclude that $w_p(\boldsymbol{k}) \geq 0$. However, it is still possible to establish that $w_p(\boldsymbol{k}) > 0$ for *a positive proportion of primes*, which suffices for our application. (Precisely, we obtain $\omega_p(\boldsymbol{k}) > 0$ for those $p \equiv 1 \pmod{m}$ where $m \geq 2$; see (3-2) in the proof of Lemma 1.2.)

The simplicity of our weight system turns out to be crucial for bounding the terms that appear in the polynomial sieve lemma. In the setting of the polynomial $F(Y, \boldsymbol{X})$ as in (1-1), our main task will be to prove square root cancellation for the sum

$$\sum_{\substack{(z,\boldsymbol{a}) \in \mathbb{F}_p^{n+1} \\ F(z^e, \boldsymbol{a}) = 0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u} \rangle),$$

for generic $\boldsymbol{a} \in \mathbb{F}_p^n$, which can be accomplished by exploiting the smoothness of the variety $V(F(Z^e, \boldsymbol{X}))$. On the other hand, if we were to adopt [8, Theorem 1.1], the presence of the factor $(\nu_p(\boldsymbol{k}))^2$ would lead to the exponential sum

$$\sum_{\substack{(z_1,z_2,\boldsymbol{a})\in\mathbb{F}_p^{n+2}\\ F(z_1^e,\boldsymbol{a})=0\\ F(z_2^e,\boldsymbol{a})=0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle),$$

which is more challenging to handle, due to the highly singular nature of the variety $V(F(Z_1^e, \boldsymbol{X})) \cap V(F(Z_2^e, \boldsymbol{X}))$.

1.3. ***Overview of the method.*** We now provide an overview of our method, highlighting four key aspects of our strategy. To prove a nontrivial upper bound for $N(F, B)$ via a sieve, we introduce a smooth nonnegative function $W : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined by $W(\boldsymbol{x}) = w(\boldsymbol{x}/B)$, where w is an infinitely differentiable, compactly supported function that is $\equiv 1$ on $[-1, 1]^n$, and supported in $[-2, 2]^n$. Define the smoothed counting function

$$\mathcal{S}(F, B) := \sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^n\\ F(y,\boldsymbol{k})=0 \text{ solvable}}} W(\boldsymbol{k}), \tag{1-21}$$

which sums over $\boldsymbol{k} \in \mathbb{Z}^n$ such that there exists $y \in \mathbb{Z}$ with $F(y, \boldsymbol{k}) = 0$. By construction

$$N(F, B) \leq \mathcal{S}(F, B),$$

and we may focus on proving a nontrivial upper bound for $\mathcal{S}(F, B)$. We employ the following sieve lemma, which we prove in Section 3.1. Here and throughout, given a polynomial f, we let $\|f\|$ denote the maximum absolute value of any coefficient of f.

Lemma 1.2 (polynomial sieve lemma). *Let $e, d \geq 1$ and $m \geq 2$ be integers. Consider the polynomial*

$$F(Y, \boldsymbol{X}) = Y^{md} + Y^{m(d-1)} f_1(\boldsymbol{X}) + \cdots + Y^m f_{d-1}(\boldsymbol{X}) + f_d(\boldsymbol{X}),$$

under the assumption that $f_d \not\equiv 0$, and that $\deg f_i = m \cdot e \cdot i$ *for each $1 \leq i \leq d$.*

Let $B \geq 1$ and define a smooth weight W supported in $[-2B, 2B]^n$ and $\equiv 1$ on $[-B, B]^n$, as above. Let $\mathcal{P} \subset \{p \equiv 1 \bmod m\}$ be a finite set of primes $p \in [Q, 2Q]$, with cardinality P. Suppose that $Q = B^\kappa$ for some fixed $0 < \kappa \leq 1$ and that $P \gg Q/\log Q$. Suppose also that

$$P \gg_{m,e,d} \max\{\log \|f_d\|, \log B\}. \tag{1-22}$$

For each $\boldsymbol{k} \in \mathbb{Z}^n$ and $p \in \mathcal{P}$ define

$$\nu_p(\boldsymbol{k}) = |\{y \in \mathbb{F}_p : F(y, \boldsymbol{k}) = 0 \ (mod\ p)\}|.$$

Then

$$\mathcal{S}(F, B) \ll_{m,e,d} \sum_{\boldsymbol{k}: f_d(\boldsymbol{k})=0} W(\boldsymbol{k}) + \frac{1}{P}\sum_{\boldsymbol{k}} W(\boldsymbol{k}) + \frac{1}{P^2}\sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \left|\sum_{\boldsymbol{k}} W(\boldsymbol{k})(\nu_p(\boldsymbol{k})-1)(\nu_q(\boldsymbol{k})-1)\right|.$$

Remark 1.3. We observe that the same lemma holds for $m = 1$, conditional on GRH, with (1-22) replaced by $Q \gg_{m,e,d} \max\{(\log \|F\|)^{\alpha_0}, (\log B)^{\alpha_0}\}$ for some $\alpha_0 > 2$. For the sake of illustration, we demonstrate this in Section 3.2, although we do not apply such a conditional result in this paper.

We now point out four key aspects of our method for applying this sieve lemma to prove Theorem 1.1. First, for all $\boldsymbol{k}$ and for all primes p, $\nu_p(\boldsymbol{k}) \le md$; this is because Y^{md} has coefficient 1 in $F(Y, \boldsymbol{X})$, so that for all values of $\boldsymbol{k}$, $F(Y, \boldsymbol{k})$ is of degree md as a polynomial in Y. On the other hand, in the proof of the lemma, we use the assumption that each prime in the sieving set has $p \equiv 1 \pmod{m}$ in order to provide a lower bound $\nu_p(\boldsymbol{k}) - 1 \ge m - 1 > 0$ for many $\boldsymbol{k}$, motivating our requirement that $m \ge 2$. This is the first novelty of our method for dealing with a case in which the variables $Y, \boldsymbol{X}$ are not "separated."

For each pair of primes $p \ne q \in \mathcal{P}$, the sieve lemma leads us to study

$$T(p,q) := \sum_{\boldsymbol{k} \in \mathbb{Z}^n} W(\boldsymbol{k})(\nu_p(\boldsymbol{k}) - 1)(\nu_q(\boldsymbol{k}) - 1). \tag{1-23}$$

After an application of the Poisson summation formula, we see that

$$T(p,q) = \left(\frac{1}{pq}\right)^n \sum_{\boldsymbol{u} \in \mathbb{Z}^n} \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq),$$

where

$$g(\boldsymbol{u}, pq) := \sum_{\boldsymbol{a} \,(\mathrm{mod}\, pq)} (\nu_p(\boldsymbol{a}) - 1)(\nu_q(\boldsymbol{a}) - 1) e_{pq}(\langle \boldsymbol{a}, \boldsymbol{u}\rangle). \tag{1-24}$$

Here we write each coordinate of $\boldsymbol{a}$ in terms of its residue class modulo pq, and $e_{pq}(t) = e^{2\pi i t/pq}$. After showing that $g(\boldsymbol{u}, pq)$ satisfies a multiplicativity relation, we can focus on the case of prime modulus, and study

$$g(\boldsymbol{u}, p) := \sum_{\boldsymbol{a} \in \mathbb{F}_p^n} (\nu_p(\boldsymbol{a}) - 1) e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle).$$

We show that the main task to bound $g(\boldsymbol{u}, p)$ is to bound the exponential sum

$$\sum_{\substack{(y,\boldsymbol{a}) \in \mathbb{F}_p^{n+1} \\ F(y,\boldsymbol{a}) = 0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle).$$

Here we highlight a second aspect: the fact that the polynomial $F(Y, \boldsymbol{X})$ is not homogeneous motivates a more sophisticated approach to bounding this sum (see Remark 4.6). Given a polynomial H, let $V(H)$ denote the corresponding variety $\{H = 0\}$, and let $\langle \boldsymbol{X}, \boldsymbol{U}\rangle = \sum_i X_i U_i$. Roughly speaking, for each prime p we divide $\boldsymbol{u} \in \mathbb{Z}^n$ into three cases: a *type zero* case when $\boldsymbol{u} \equiv 0 \pmod{p}$, a *good* case when $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is not tangent to $V(F(Y, \boldsymbol{X}))$ over $\overline{\mathbb{F}}_p$, and finally a *bad* case in which $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is tangent to $V(F(Y, \boldsymbol{X}))$ over $\overline{\mathbb{F}}_p$. (More precisely, we reformulate this in terms of varieties in unweighted projective space.) In the type zero case, we can only show that $g(\boldsymbol{0}, p) \ll p^{n-1/2}$, but such cases are sparse. In the remaining two cases, we apply a version of the Weil bound to $g(\boldsymbol{u}, p)$, obtaining $g(\boldsymbol{u}, p) \ll p^{n/2}$ if $\boldsymbol{u}$ is good and $g(\boldsymbol{u}, p) \ll p^{n/2+1/2}$ if $\boldsymbol{u}$ is bad (Proposition 4.2).

A third crucial aspect arises when we assemble this information efficiently inside the third term on the right-hand side of the sieve lemma, namely

$$\frac{1}{P^2} \sum_{p \neq q \in \mathcal{P}} |T(p,q)| \ll \frac{1}{P^2 Q^{2n}} \sum_{p \neq q \in \mathcal{P}} \sum_{\boldsymbol{u} \in \mathbb{Z}^n} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq) \right|. \tag{1-25}$$

In many earlier applications of the power sieve or polynomial sieve to count solutions to Diophantine equations, the strategy has been to bound $|T(p,q)|$ uniformly over $p \neq q$ and simply sum trivially over $p \neq q$. However, recent work of the first author demonstrated how to take advantage of nontrivial averaging over the sum of $p \neq q \in \mathcal{P}$; see [2]. In this paper, we also average nontrivially over $p \neq q$ and this contributes to the strength of our main theorem.

In order to average nontrivially over $p \neq q \in \mathcal{P}$, we quantify the fact that there cannot be many triples $\boldsymbol{u}, p, q$ for which $\boldsymbol{u}$ is simultaneously bad for both p and q. Roughly speaking, we characterize the dual variety of the original hypersurface $V(F(Y, \boldsymbol{X}))$ according to an irreducible polynomial $G(U_Y, U_1, \dots, U_n)$, and observe that $G(0, \boldsymbol{u}) \neq 0$ precisely when the hyperplane $V(\langle \boldsymbol{u}, \boldsymbol{X} \rangle)$ is not tangent to $V(F(Y, \boldsymbol{X}))$ over $\mathbb{C}$. Then we reverse the order of summation in the right-hand side of (1-25), writing it as

$$\frac{1}{P^2 Q^{2n}} \sum_{\boldsymbol{u} \in \mathbb{Z}^n} \sum_{p \neq q \in \mathcal{P}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq) \right|. \tag{1-26}$$

The sum over $\boldsymbol{u}$ can be split into case (a) where $G(0, \boldsymbol{u}) \neq 0$ and case (b) where $G(0, \boldsymbol{u}) = 0$. In case (a), we show $\boldsymbol{u}$ is bad modulo p and q only if p and q divide the (nonzero) value of a certain resultant polynomial; thus there can only be very few such p, q.

A fourth key aspect arises in case (b), for which $\boldsymbol{u}$ is bad for all primes (since the value of the resultant is zero). To compensate, we show that there are not too many $\boldsymbol{u}$ for which $G(0, \boldsymbol{u}) = 0$. This step is one of the significant novelties of the paper. It requires understanding not the variety $V(G(U_Y, \boldsymbol{U}))$ but $V(G(U_Y, \boldsymbol{U})) \cap V(U_Y)$, the intersection with the hyperplane $U_Y = 0$. To tackle this, we show that any polynomial divisor of $G(0, \boldsymbol{U})$ has degree at least 2 (Proposition 5.2), so that we can apply strong bounds of Heath-Brown [27] and Pila [41] to count solutions to $G(0, \boldsymbol{u}) = 0$ (see (5-18)). To prove the key result in Proposition 5.2, we employ a geometric argument to show that given a nonsingular projective hypersurface X and a projective line ℓ not contained in X, the generic hyperplane containing ℓ is not tangent to X. This statement, proved in Section 6 via a strategy suggested by Per Salberger, is critical to the method and the ultimate strength of Theorem 1.1.

Remark 1.4. It would be interesting to consider bounding $N(F, B)$, in the setting of Theorem 1.1, by other methods. As mentioned earlier, one approach is to count all $(n+1)$-tuples $\{(y, \boldsymbol{x}) \in \mathbb{Z}^{n+1} : y \ll B^e, x_i \ll B : F(y, \boldsymbol{x}) = 0\}$, for example, by applying the determinant method. Since the range of y depends on e, such a direct approach is likely to produce a bound for $N(F, B)$ with an exponent depending on e. Alternatively, one could fix $x_2, \dots, x_n$ (with $\approx B^{n-1}$ such choices) and consider the resulting equation as a projective curve in variables y, x_1. Supposing that the resulting curve is generically of degree dme, an application of Bombieri-Pila [1] could count (y, x_1) in the square $[-B^e, B^e]^2$. This could ultimately lead to a total bound of the form $N(F, B) \ll B^{n-1} \cdot B^{e/dme+\varepsilon} = B^{n-1+1/dm+\varepsilon}$. This putative outcome appears independent of e, but the method has overcounted x_1 in the range B^e; nevertheless, such an approach could be advantageous for large d, m.

1.4. ***Notation.*** We use $e_q(t) = e^{2\pi i t/q}$. We denote $\boldsymbol{X} = (X_1, \dots, X_n)$, $\boldsymbol{U} = (U_1, \dots, U_n)$. Moreover, for two vectors $\boldsymbol{s} = (s_1, \dots s_n)$, $\boldsymbol{t} = (t_1, \dots, t_n)$, we define $\langle \boldsymbol{s}, \boldsymbol{t} \rangle = \sum_{i=1}^n s_i t_i$. We let $\|F\|$ denote the absolute value of the maximum coefficient in a polynomial $F \in \mathbb{Z}[X_1, \dots, X_n]$; similarly $\|\boldsymbol{X}\| = \max_{1 \le i \le n} |X_i|$ for $\boldsymbol{X} \in \mathbb{Z}^n$.

2. Reduction to remove dependence on $\|F\|$

Recall that Theorem 1.1 states that the upper bound for $N(F, B)$ is only dependent on the degree of F, and not on the coefficients of F. In fact, the sieve methods we apply prove an upper bound for $N(F, B)$ that *can* depend on $\|F\|$. In this section we show by alternative methods that we may assume that $\|F\| \ll B^{(mde)^{n+2}}$. The method does not rely on assuming $m \ge 2$ in (1-1), and so without any additional trouble we may work more generally in the setting of (1-20).

Lemma 2.1. *Let $V(G(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ be defined by*

$$G(Y, \boldsymbol{X}) = Y^D + Y^{D-1} f_1(\boldsymbol{X}) + \cdots + Y f_{D-1}(\boldsymbol{X}) + f_D(\boldsymbol{X})$$

with each f_i a form of $\deg f_i = i \cdot e$, for fixed $D, e \ge 1$ and $n \ge 1$. Assume that $f_D \not\equiv 0$ and the weighted hypersurface $V(G(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ is absolutely irreducible. Then either

$$\|G\| \ll B^{(De)^{n+2}},$$

or $N(G, B) \ll_{n,D,e} B^{n-1}$.

Remark 2.2. Under the hypotheses of Theorem 1.1, for F as in (1-1), $V(F(Y, \boldsymbol{X}))$ is absolutely irreducible (following similar reasoning to Remark 3.3). As a result of this lemma, we can obtain the bound claimed in Theorem 1.1 as long as all later dependence on $\|F\|$ is at most logarithmic in $\|F\|$, which we track as the argument proceeds.

Proof. The method of proof follows [27, Theorem 4], or the recent similar result [3, Lemma 2.1]. Fix $n, D, e \ge 1$. We start by considering the set of monomials

$$\mathcal{E} := \left\{ Y^{d_Y} X_1^{d_1} \cdots X_n^{d_n} : d_Y e + \sum_{i=1}^n d_i = De \right\},$$

in which the degrees $d_Y, d_1, \dots, d_n$ vary over all nonnegative integers satisfying $d_Y e + \sum d_i = De$. It is easy to see that $|\mathcal{E}| \le (De)^{n+1}$.

Let $B \geq 1$ be fixed. Let $\boldsymbol{v}$ denote coordinates $(y, x_1, \ldots, x_n)$ and let $\{\boldsymbol{v}_1, \ldots \boldsymbol{v}_N\}$ enumerate the set of points that are solutions to $G(Y, \boldsymbol{X}) = 0$, with each of the last n coordinates of $\boldsymbol{v}_j$ lying in $[-B, B]$. Note that these count each $\boldsymbol{X} \in [-B, B]^n$ for which $G(Y, \boldsymbol{X})$ is solvable at least once, so that $N(G, B) \leq N \leq D \cdot N(G, B)$. (For the upper bound, we recall that the coefficient of Y^D in $G(Y, \boldsymbol{X})$ is nonzero, so that any given $\boldsymbol{X}$ can correspond to at most D such Y.) Then, we construct the $N \times |\mathcal{E}|$ matrix

$$\boldsymbol{C} = (\boldsymbol{v}_i^{\boldsymbol{e}})_{\substack{1 \leq i \leq N \\ \boldsymbol{e} \in \mathcal{E}}}.$$

Notice that $\operatorname{rank} \boldsymbol{C} \leq |\mathcal{E}| - 1$, since the vector $\boldsymbol{a} \in \mathbb{Z}^{|\mathcal{E}|} \setminus \{0\}$ whose entries correspond to the coefficients of $G(Y, \boldsymbol{X})$ is such that $\boldsymbol{C}\boldsymbol{a} = \boldsymbol{0}$. Moreover, $\boldsymbol{a}$ is primitive since the coefficient associated to Y^D is 1. Now the strategy is to find another nonzero vector $\boldsymbol{b}$ in the nullspace of $\boldsymbol{C}$ and show that if $\boldsymbol{b}$ is in the span of $\boldsymbol{a}$ then $\|G\|$ is small, and if $\boldsymbol{b}$ is not in the span of $\boldsymbol{a}$ then we have an improved count for $N(G, B)$. We may assume henceforward that $|\mathcal{E}| \leq N$, since otherwise we already have the upper bound $N(G, B) \leq N \leq |\mathcal{E}| \leq (De)^{n+1}$, which suffices for the lemma.

If $\operatorname{rank} \boldsymbol{C} \leq |\mathcal{E}| - 2$, then the nullspace has dimension at least 2, and we can take $\boldsymbol{b} \in \mathbb{Z}^{|\mathcal{E}|}$ to be any element in the nullspace that is not in the span of $\boldsymbol{a}$. Let $H(Y, \boldsymbol{X})$ be the polynomial defined by the coefficients corresponding to the vector $\boldsymbol{b}$ and consider the polynomial $R(\boldsymbol{X}) = \operatorname{Res}(G(Y, \boldsymbol{X}), H(Y, \boldsymbol{X}))$, which is a polynomial in $\boldsymbol{X}$ of degree $\ll_{D,e,n} 1$. (See, e.g., [21, Ch 12], which we apply to take the resultant of two polynomials in the variable Y, whose coefficients are determined by $\boldsymbol{X}$.) We claim that $R(\boldsymbol{X}) \not\equiv 0$: indeed, if $R(\boldsymbol{X}) \equiv 0$, then G and H would share an irreducible component. Since $G(Y, \boldsymbol{X}) = 0$ is irreducible, and $\deg H \leq De = \deg G$, it would follow that G is a constant multiple of H, but this is not possible since we are assuming that $\boldsymbol{a}$ and $\boldsymbol{b}$ are not proportional. Thus $R(\boldsymbol{X}) \not\equiv 0$. Moreover, observe that for any $\boldsymbol{x} \in \mathbb{Z}^n$

$$R(\boldsymbol{x}) = 0 \iff G(Y, \boldsymbol{x}) \text{ and } H(Y, \boldsymbol{x}) \text{ have a common root.}$$

Note that any $\boldsymbol{x}$ such that $G(y, \boldsymbol{x}) = 0$ is solvable contributes at least one row to the matrix $\boldsymbol{C}$; each such row also corresponds to a solution to $H(y, \boldsymbol{x}) = 0$. Thus it follows that

$$\begin{aligned} N(G, B) &= |\{\boldsymbol{x} \in [-B, B]^n : \exists y \in \mathbb{Z} \text{ such that } G(y, \boldsymbol{x}) = H(y, \boldsymbol{x}) = 0\}| \\ &\leq |\{\boldsymbol{x} \in [-B, B]^n : \ R(\boldsymbol{x}) = 0\}| \\ &\ll_{n,D,e,} B^{n-1}, \end{aligned}$$

with an implicit constant independent of the coefficients of R, via an application of a trivial counting bound for the nonzero polynomial R. (This bound is sometimes called the Schwartz-Zippel bound, and a proof can be found in [27, Theorem 1]; we remark that although in that context the polynomial under consideration is absolutely irreducible, the method of proof only requires that it is not identically zero.)

The remaining case is when $\operatorname{rank} \boldsymbol{C} = |\mathcal{E}| - 1$, so that all $|\mathcal{E}| \times |\mathcal{E}|$ minors vanish, but at least one $(|\mathcal{E}| - 1) \times (|\mathcal{E}| - 1)$ minor does not; we claim there is a nonzero $\boldsymbol{b} \in \mathbb{Z}^{|\mathcal{E}|}$ in the nullspace of $\boldsymbol{C}$ such that

$|\boldsymbol{b}| = O(B^{De|\mathcal{E}|}) = O(B^{(De)^{n+2}})$. If so, then since $\boldsymbol{a}$ is primitive (and $\boldsymbol{b}$ must be proportional to $\boldsymbol{a}$) it follows that $|\boldsymbol{a}| \le |\boldsymbol{b}| \ll B^{(De)^{n+2}}$. This shows that $\|G\| \ll B^{(De)^{n+2}}$ as claimed.

An appropriate $\boldsymbol{b}$ can be constructed with entries that are $(|\mathcal{E}|-1)\times(|\mathcal{E}|-1)$ minors, so that the size estimate $|\boldsymbol{b}| = O(B^{De|\mathcal{E}|})$ follows from the fact that each entry of $\boldsymbol{C}$ is $O(B^{De})$. For completeness, we sketch this construction. Without loss of generality, we can let $\boldsymbol{C}'$ denote the top $|\mathcal{E}|\times|\mathcal{E}|$ submatrix in $\boldsymbol{C}$, and assume that the minor $\boldsymbol{C}'_{1,1}$ (obtained by omitting the first row and first column of $\boldsymbol{C}'$) is nonzero. Define a vector $\boldsymbol{b}$ as follows: for each $1 \le j \le |\mathcal{E}|$, define the entry b_j to be the $(1, j)$-th cofactor of $\mathbf{C}'$; in particular $b_1 \ne 0$ so $\boldsymbol{b}$ is nonzero, and $|\boldsymbol{b}| = O(B^{De(|\mathcal{E}|-1)}) = O(B^{De|\mathcal{E}|})$. We now show that $\boldsymbol{b}$ is in the nullspace of $\boldsymbol{C}$. Let $\boldsymbol{r}_i$ denote the i-th row of $\boldsymbol{C}$; then for each $1 \le i \le N$,

$$\boldsymbol{r}_i \cdot \boldsymbol{b} = \det \begin{pmatrix} \boldsymbol{r}_i \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_{|\mathcal{E}|} \end{pmatrix} = 0. \tag{2-1}$$

Indeed, for $i = 1$ or $i > |\mathcal{E}|$, up to sign, $\boldsymbol{r}_i \cdot \boldsymbol{b}$ is an $|\mathcal{E}|\times|\mathcal{E}|$ minor of $\boldsymbol{C}$, and all such minors vanish since $\operatorname{rank}\boldsymbol{C} < |\mathcal{E}|$. For $2 \le i \le |\mathcal{E}|$, the matrix (2-1) has two identical rows. Thus $\boldsymbol{C}\boldsymbol{b} = \mathbf{0}$. □

3. Preliminaries on the sieve lemma

In this section we gather together two preliminary steps: first, we prove the sieve inequality in Lemma 1.2; for $m = 1$ we provide an alternative proof, conditional on GRH. Second, we formulate an equivalent nonsingularity condition in unweighted projective space. We also make preliminary remarks on the sieving set.

3.1. ***Proof of the polynomial sieve lemma.*** To prove Lemma 1.2, observe that

$$\mathcal{S}(F, B) = \sum_{\boldsymbol{k}:f_d(\boldsymbol{k})=0} W(\boldsymbol{k}) + \sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^n: \\ f_d(\boldsymbol{k})\ne 0 \\ F(y,\boldsymbol{k})=0 \text{ solvable}}} W(\boldsymbol{k}),$$

since within the first term, $y = 0$ is always a solution to $F(y, \boldsymbol{k}) = 0$. We consider the weighted sum

$$\sum_{\boldsymbol{k}:f_d(\boldsymbol{k})\ne 0} W(\boldsymbol{k}) \Big(\sum_{p\in\mathcal{P}} (\nu_p(\boldsymbol{k}) - 1) \Big)^2. \tag{3-1}$$

Fix $\boldsymbol{k}$ such that $f_d(\boldsymbol{k}) \ne 0$ and the polynomial $F(Y, \boldsymbol{k})$ is solvable over $\mathbb{Z}$, so that there exists $y_0 \in \mathbb{Z}$ such that $F(y_0, \boldsymbol{k}) = 0$. For any $p \in \mathcal{P}$ such that $p \nmid f_d(\boldsymbol{k})$, then $y_0 \not\equiv 0 \bmod p$. Then since $p \equiv 1 \bmod m$, and due to the structure of F in (1-1), we have that $\{y_0, \gamma_p y_0, \dots, \gamma_p^{m-1} y_0\}$ are distinct solutions of $F(Y, \boldsymbol{k}) \equiv 0 \pmod p$, where $\gamma_p^m \equiv 1 \bmod p$ and γ_p is a primitive m-th root of unity in $\mathbb{F}_p$. In particular, for such p, $\nu_p(\boldsymbol{k}) \ge m$. Consequently, for each $\boldsymbol{k}$ such that $f_d(\boldsymbol{k}) \ne 0$ and $F(Y, \boldsymbol{k})$ is solvable, we have that

$$\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1)\ge(m-1)\sum_{p\in\mathcal{P},\,p\nmid f_d(\boldsymbol{k})}1\gg_m P-\sum_{p\in\mathcal{P},\,p\mid f_d(\boldsymbol{k})}1\ge(1/2)P, \tag{3-2}$$

as long as $P\gg_{m,e,d}\max\{\log\|f_d\|,\log B\}$. The last step follows since the number $\omega(f_d(\boldsymbol{k}))$ of distinct prime divisors of $f_d(\boldsymbol{k})\ne 0$ is at most

$$\begin{aligned}\omega(f_d(\boldsymbol{k}))&\ll\log(f_d(\boldsymbol{k}))/\log\log(f_d(\boldsymbol{k}))\\&\ll\log(\|f_d\|B^{dem})\\&\ll_{m,e,d}\log\|f_d\|+\log B.\end{aligned}$$

Thus the last inequality in (3-2) holds as long as

$$P\gg_{m,e,d}\max\{\log\|f_d\|,\log B\}, \tag{3-3}$$

leading to the corresponding hypothesis in the lemma.

From (3-2) and the nonnegativity of the weight W, we see that

$$P^2\sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^n:\\ f_d(\boldsymbol{k})\ne0\\ F(y,\boldsymbol{k})=0\text{ solvable}}}W(\boldsymbol{k})\ll\sum_{\boldsymbol{k}:f_d(\boldsymbol{k})\ne0}W(\boldsymbol{k})\Big(\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1)\Big)^2\le\sum_{\boldsymbol{k}}W(\boldsymbol{k})\Big(\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1)\Big)^2.$$

Opening the square on the right-hand side, the contribution from $p=q\in\mathcal{P}$ is

$$\sum_{p\in\mathcal{P}}\sum_{\boldsymbol{k}}W(\boldsymbol{k})(\nu_p(\boldsymbol{k})-1)^2\ll_{m,d}P\sum_{\boldsymbol{k}}W(\boldsymbol{k}),$$

since $\nu_p(\boldsymbol{k})\le md$ for all $\boldsymbol{k}$, as previously mentioned. The contribution from $p\ne q\in\mathcal{P}$ is bounded in absolute value by

$$\sum_{p\ne q\in\mathcal{P}}\Big|\sum_{\boldsymbol{k}}W(\boldsymbol{k})(\nu_p(\boldsymbol{k})-1)(\nu_q(\boldsymbol{k})-1)\Big|.$$

Assembling all these terms, we see that Lemma 1.2 is proved.

Remark 3.1. When we apply Lemma 1.2 to prove Theorem 1.1, we can assume that $\|f_d\|\le\|F\|\ll B^{(mde)^{n+2}}$, by Lemma 2.1. This will allow us to verify that (3-3) holds for our choice of sieving set, as we will verify in Section 7 when we choose Q in (7-4).

3.2. ***Alternative proof when $m=1$, conditional on GRH.*** Recall from Section 1.2.4 the general problem of counting $\boldsymbol{x}\in[-B,B]^n$ such that $G(y,\boldsymbol{x})=0$ is solvable in $\mathbb{Z}$, with $G(Y,\boldsymbol{X})$ of degree D as in (1-20). In our main work in this paper, we assume that $D=md$ with $m\ge2$ and that G is a polynomial in Y^m. This additional structure allowed us to choose a sieving set $\mathcal{P}\subset[Q,2Q]$ of primes $p\equiv1\pmod m$, so that all the m-th roots of unity are present in $\mathbb{F}_p$, for each $p\in\mathcal{P}$. With this property, we could define sieve weights that exhibit an appropriate lower bound in the form (3-2) for most $\boldsymbol{k}$ in the support of $W(\boldsymbol{k})$ and a positive proportion of primes.

Nevertheless, we can proceed by a different argument to develop a sieve lemma to bound the number of $\boldsymbol{x} \in [-B, B]^n$ such that $G(y, \boldsymbol{x}) = 0$ is solvable over $\mathbb{Z}$, with no condition on the degree D; that is, to prove a version of Lemma 1.2 in the case $m = 1$. As a first step, we naturally try to introduce a system of weights, according to a fixed set of primes. Let us take $\mathcal{P} = \{Q \leq p \leq 2Q : p \text{ prime}\}$ for some parameter Q to be chosen optimally with respect to B. In particular, by the prime number theorem, $|\mathcal{P}| \gg Q(\log Q)^{-1}$ for all $Q \gg 1$. Fix $\boldsymbol{k} \in \mathbb{Z}^n$. For each prime $p \in \mathcal{P}$, set

$$\nu_p(\boldsymbol{k}) = |\{y \in \mathbb{F}_p : G(y, \boldsymbol{k}) = 0 \ (\mathrm{mod}\ p)\}|.$$

Since $G(y, \boldsymbol{k})$ contains the term y^D, it is not the zero polynomial in y, and $\nu_p(\boldsymbol{k}) \leq D$. Consider, as in the proof of Lemma 1.2 above, the weighted sum

$$\sum_{\boldsymbol{k}: f_D(\boldsymbol{k}) \neq 0} W(\boldsymbol{k}) \Big(\sum_{p \in \mathcal{P}} (\nu_p(\boldsymbol{k}) - 1) \Big)^2. \tag{3-4}$$

In order to deduce a sieve lemma, we need a lower bound for the arithmetic weight (the squared term), for those $\boldsymbol{k}$ for which $f_D(\boldsymbol{k}) \neq 0$ and $G(Y, \boldsymbol{k}) = 0$ is solvable over $\mathbb{Z}$.

Here is one approach. Let $\boldsymbol{k}$ be fixed, with $f_D(\boldsymbol{k}) \neq 0$ and $G(Y, \boldsymbol{k}) = 0$ solvable over $\mathbb{Z}$, and $\boldsymbol{k}$ in the support of W. Then $G(Y, \boldsymbol{k}) = (Y - y_0)\tilde{g}_{\boldsymbol{k}}(Y)$ for some $y_0 \in \mathbb{Z} \setminus \{0\}$ and some (monic) $\tilde{g}_{\boldsymbol{k}}(Y) \in \mathbb{Z}[Y]$ of degree $D - 1$. For each such $\boldsymbol{k}$, we can obtain a suitable lower bound for the arithmetic weight in (3-4) as long as for a positive proportion of $p \in \mathcal{P}$, $\tilde{g}_{\boldsymbol{k}}$ has a root over $\mathbb{F}_p$. Let $g_{\boldsymbol{k}}$ be an irreducible factor of $\tilde{g}_{\boldsymbol{k}}$. Let $F_{\boldsymbol{k}}$ denote the splitting field of $g_{\boldsymbol{k}}$ over $\mathbb{Q}$, say $F_{\boldsymbol{k}} = \mathbb{Q}(\alpha_{\boldsymbol{k}})$. Since $g_{\boldsymbol{k}}$ is irreducible, then it is the minimal polynomial of $\alpha_{\boldsymbol{k}}$ in $\mathbb{Z}[Y]$, and it is separable (since we are working over characteristic zero), and the splitting field is Galois over $\mathbb{Q}$. By Dedekind's theorem, for all $p \nmid [\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]$, $g_{\boldsymbol{k}}$ splits completely over $\mathbb{F}_p$ precisely when $(p) = p\mathcal{O}_{F_{\boldsymbol{k}}}$ splits completely in $F_{\boldsymbol{k}}$; see, e.g., [37, Theorem 27, p. 79]. Then

$$\sum_{p \in \mathcal{P}} (\nu_p(\boldsymbol{k}) - 1) = \sum_{p \in \mathcal{P}} \big|\{y \in \mathbb{F}_p : \tilde{g}_{\boldsymbol{k}}(y) = 0\}\big| \geq \sum_{p \in \mathcal{P}} \big|\{y \in \mathbb{F}_p : g_{\boldsymbol{k}}(y) = 0\}\big|.$$

If $g_{\boldsymbol{k}}$ is linear in $\mathbb{Z}[Y]$, this sum is of size $|\mathcal{P}|$, which suffices. If $\deg g_{\boldsymbol{k}} \geq 2$, we continue to argue that

$$\begin{aligned} \sum_{p \in \mathcal{P}} (\nu_p(\boldsymbol{k}) - 1) &\geq \deg(g_{\boldsymbol{k}}) \big|\{p \in \mathcal{P} : g_{\boldsymbol{k}}(Y) \text{ completely split over } \mathbb{F}_p\}\big| \\ &\geq \big|\{p \in \mathcal{P} : p\mathcal{O}_{F_{\boldsymbol{k}}} \text{ splits completely in } F_{\boldsymbol{k}}\}\big| - \big|\{p \in \mathcal{P} : p \mid [\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]\}\big|. \end{aligned} \tag{3-5}$$

Let

$$\pi_{\boldsymbol{k}}(Q) = \big|\{p \leq Q : p\mathcal{O}_{F_{\boldsymbol{k}}} \text{ splits completely in } F_{\boldsymbol{k}}\}\big|$$

and $N(\boldsymbol{k}) = |\{p \mid [\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]\}|$. The Chebotarev density theorem, in the unconditional form of [34, Theorem 1.3], shows that

$$\left| \pi_{\boldsymbol{k}}(Q) - \frac{1}{|G_{\boldsymbol{k}}|} \frac{Q}{\log Q} \right| = \frac{1}{|G_{\boldsymbol{k}}|} \frac{Q^{\beta_0}}{\log Q^{\beta_0}} + O_{D,A}(Q(\log Q)^{-A}) \tag{3-6}$$

for every $A \geq 2$, as long as $Q \geq \exp(10 \deg F_{\boldsymbol{k}} (\log |D(F_{\boldsymbol{k}})|)^2)$. Here $G_{\boldsymbol{k}}$ is the Galois group $\mathrm{Gal}(F_{\boldsymbol{k}}/\mathbb{Q})$, $D(F_{\boldsymbol{k}})$ is the discriminant of the splitting field $F_{\boldsymbol{k}}/\mathbb{Q}$, and $\deg F_{\boldsymbol{k}} = \deg |F_{\boldsymbol{k}}/\mathbb{Q}|$ is the degree of the extension. The implicit constant in the error term depends only on A and $\deg F_{\boldsymbol{k}} = |G_{\boldsymbol{k}}| \leq (D-1)!$. The real number $1/2 < \beta_0 < 1$, if it exists, is the (real, simple) exceptional zero of the associated Dedekind zeta function $\zeta_{F_{\boldsymbol{k}}}$; if no exceptional zero exists, that term does not appear in the result.

In particular, under the assumption of GRH for $\zeta_{F_{\boldsymbol{k}}}$, Lagarias and Odlyzko's Theorem 1.1 in [34] (in the refined form of Serre [44, Theorem 4]) shows that for any $Q > 2$, the entire right-hand side of (3-6) may be replaced by

$$O(|G_{\boldsymbol{k}}|^{-1} Q^{1/2} \log(|D(F_{\boldsymbol{k}})| Q^{\deg F_{\boldsymbol{k}}})) = O_D(Q^{1/2} \log Q) + O_D(Q^{1/2} \log |D(F_{\boldsymbol{k}})|),$$

in which the implied constant is absolute and effectively computable. There exists a constant $Q_0(D)$ depending only on D such that the first term is $\leq \frac{1}{4} \frac{1}{(D-1)!} Q (\log Q)^{-1}$ for all $Q \geq Q_0(D)$. The second term is also $\leq \frac{1}{4} \frac{1}{(D-1)!} Q (\log Q)^{-1}$ if for example $Q \geq Q_1(D)(\log D(F_{\boldsymbol{k}}))^{\alpha_0}$ for a constant $Q_1(D)$ and some fixed $\alpha_0 > 2$. This shows that under GRH, for all $Q \gg_D (\log D(F_{\boldsymbol{k}}))^{\alpha_0}$ some fixed $\alpha_0 > 2$,

$$\pi_{\boldsymbol{k}}(Q) - \pi_{\boldsymbol{k}}(Q/2) \gg_D Q/\log Q \gg_D |\mathcal{P}|. \tag{3-7}$$

Two tasks remain in order to complete a lower bound for (3-5): (i) to bound $D(F_{\boldsymbol{k}})$ from above, so that the lower bound $Q \gg_D (\log D(F_{\boldsymbol{k}}))^{\alpha_0}$ can be made uniform over $\boldsymbol{k}$, and (ii) to count

$$N(\boldsymbol{k}) = |\{p | [\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]\}| = \omega([\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]) \ll \log[\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]] / \log\log[\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]].$$

We note the relation

$$D(F_{\boldsymbol{k}})[\mathcal{O}_{F_{\boldsymbol{k}}} : \mathbb{Z}[\alpha_{\boldsymbol{k}}]]^2 = \mathrm{Disc}(g_{\boldsymbol{k}}), \tag{3-8}$$

which holds by [38, Remark 2.25 and equation (8) on p. 38]. (Since $g_{\boldsymbol{k}}$ was assumed to be irreducible and we are in characteristic zero, then $g_{\boldsymbol{k}}$ is separable and $\mathrm{Disc}(g_{\boldsymbol{k}}) \neq 0$.) Thus for both remaining tasks, it suffices to bound $\mathrm{Disc}(g_{\boldsymbol{k}})$ from above, since by (3-8) both

$$N(\boldsymbol{k}) \ll \log \mathrm{Disc}\,(g_{\boldsymbol{k}}), \qquad \log D(F_{\boldsymbol{k}}) \leq \log \mathrm{Disc}\,(g_{\boldsymbol{k}}).$$

Now Disc $(g_{\boldsymbol{k}})$ (the resultant of $g_{\boldsymbol{k}}(Y)$ and $g'_{\boldsymbol{k}}(Y)$, as defined in [21, Chapter 13, Proposition 1.1]) is a polynomial in the coefficients of $g_{\boldsymbol{k}}$ with degree bounded in terms of D. The coefficients of $g_{\boldsymbol{k}}$ are polynomials in $\boldsymbol{k}$ and the coefficients of $G(Y, \boldsymbol{X})$ with degree at most D. Since we only consider $\boldsymbol{k}$ in the support of W, $|\boldsymbol{k}| \ll B$, and the coefficients of $g_{\boldsymbol{k}}$ are $\ll \|G\| B^D$. Thus

$$\log \mathrm{Disc}\,(g_{\boldsymbol{k}}) \ll_D \log \|G\| + \log B.$$

In combination with (3-7), we can conclude in (3-5) that for some constant C_D,

$$\sum_{p \in \mathcal{P}} (\nu_p(\boldsymbol{k}) - 1) \gg_D Q/\log Q - C_D(\log \|G\| + \log B),$$

for all $Q \geq C'_D \max\{(\log \|G\|)^{\alpha_0}, (\log B)^{\alpha_0}\}$ for some $\alpha_0 > 2$. By taking C'_D sufficiently large, we achieve $\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1) \gg |\mathcal{P}| = P$. This shows that, conditional on GRH,

$$P^2 \sum_{\substack{\boldsymbol{k}\in\mathbb{Z}^n:\\ f_D(\boldsymbol{k})\neq 0\\ G(y,\boldsymbol{k})=0 \text{ solvable}}} W(\boldsymbol{k}) \ll \sum_{\boldsymbol{k}:f_D(\boldsymbol{k})\neq 0} W(\boldsymbol{k})\Big(\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1)\Big)^2 \leq \sum_{\boldsymbol{k}} W(\boldsymbol{k})\Big(\sum_{p\in\mathcal{P}}(\nu_p(\boldsymbol{k})-1)\Big)^2.$$

From here, the remainder of the proof used above for Lemma 1.2 can be repeated, and this completes the proof of the claim in Remark 1.3.

3.3. ***Associated variety in unweighted projective space.*** It is a hypothesis of Theorem 1.1 that the weighted hypersurface $V(F(Y,\boldsymbol{X})) \subset \mathbb{P}(e,1,\dots,1)$, defined by $F(Y,\boldsymbol{X})=0$, is nonsingular over $\mathbb{C}$. It is convenient to relate $V(F(Y,\boldsymbol{X}))$ to a variety in unweighted projective space. We claim that for

$$F(Y,\boldsymbol{X}) = Y^{dm} + Y^{(d-1)m} f_1(\boldsymbol{X}) + \cdots + f_d(\boldsymbol{X}),$$

then $V(F(Y,\boldsymbol{X})) \subset \mathbb{P}(e,1,\dots,1)$ is nonsingular if and only if $V(F(Z^e,\boldsymbol{X})) \subset \mathbb{P}^n$ is nonsingular. Here, we again apply the assumption $m \geq 2$. Indeed the weighted projective variety is nonsingular if and only if the only solution of

$$\begin{cases} F(Y,\boldsymbol{X}) = 0, \\ \dfrac{\partial F}{\partial Y}(Y,\boldsymbol{X}) = \sum_{i=0}^{d-1} f_i(\boldsymbol{X})\cdot m(d-i)Y^{m(d-i)-1} = 0, \\ \dfrac{\partial F}{\partial X_j}(Y,\boldsymbol{X}) = 0, \quad j=1,\dots,n, \end{cases} \tag{3-9}$$

on $\mathbb{A}^{n+1}$ is the point $P = \boldsymbol{0}$. (By convention we set $f_0(\boldsymbol{X}) = 1$.) Similarly, the projective variety $V(F(Z^e,\boldsymbol{X}))$ is nonsingular if and only if the only solution of

$$\begin{cases} F(Z^e,\boldsymbol{X}) = 0, \\ \dfrac{\partial F}{\partial Z}(Z^e,\boldsymbol{X}) = \sum_{i=0}^{d-1} f_i(\boldsymbol{X})\cdot me(d-i)Z^{em(d-i)-1} = 0, \\ \dfrac{\partial F}{\partial X_j}(Z^e,\boldsymbol{X}) = 0, \quad j=1,\dots,n, \end{cases} \tag{3-10}$$

on $\mathbb{A}^{n+1}$ is the point $P = \boldsymbol{0}$. Moreover, note that

$$\frac{\partial F}{\partial Y}(Y,\boldsymbol{X}) = mY^{m-1}\sum_{i=0}^{d-1} f_i(\boldsymbol{X})(d-i)Y^{m(d-i-1)}, \tag{3-11}$$

$$\frac{\partial F}{\partial Z}(Z^e,\boldsymbol{X}) = emZ^{em-1}\sum_{i=0}^{d-1} f_i(\boldsymbol{X})(d-i)Z^{em(d-i-1)}.$$

We will momentarily use this to confirm that if $m \geq 2$, a nonzero solution (say $P = (y, \boldsymbol{x}) \in \mathbb{A}^{n+1}$) to (3-9) exists if and only if a solution (namely $Q = (y^{1/e}, \boldsymbol{x}) \in \mathbb{A}^{n+1}$) to (3-10) exists.

To clarify the role of the assumption $m \geq 2$, let us briefly make a general observation. In general, let a polynomial $G(Y, \boldsymbol{X})$ be given as in (1-20) and assume $V(G(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ is nonsingular; we may assume $e \geq 2$ (since otherwise the variety is already unweighted). Then we claim $V(G(Z^e, \boldsymbol{X}))$ is nonsingular (as a projective variety) if and only if $V(G(Y, \boldsymbol{X})) \cap V(Y)$ is nonsingular (as a weighted projective variety). By the chain rule,

$$\frac{\partial G}{\partial Z}(Z^e, \boldsymbol{X}) = eZ^{e-1}\left(\frac{\partial G}{\partial Y}\right)(Z^e, \boldsymbol{X}).$$

Observe that

$$\begin{aligned}\operatorname{Sing}(V(G(Z^e, \boldsymbol{X}))) &= \{(z, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{Z,\boldsymbol{X}} G(z^e, \boldsymbol{x}) = \boldsymbol{0}\} \\ &= \{(0, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{\boldsymbol{X}} G(0, \boldsymbol{x}) = \boldsymbol{0}\} \cup \{(z, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{Y,\boldsymbol{X}} G(z^e, \boldsymbol{x}) = \boldsymbol{0}\} \\ &= \{(0, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{\boldsymbol{X}} G(0, \boldsymbol{x}) = \boldsymbol{0}\} \cup \varnothing \qquad (3\text{-}12)\end{aligned}$$

under the assumption that $V(G(Y, \boldsymbol{X}))$ is nonsingular. On the other hand, by the Jacobian criterion,

$$\operatorname{Sing}(V(G(Y, \boldsymbol{X})) \cap V(Y)) = \{(0, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{\boldsymbol{X}} G(0, \boldsymbol{x}) = \boldsymbol{0}\}.$$

(Here we have used that $G(0, \boldsymbol{X})$ is itself homogeneous in $\boldsymbol{X}$, so that $\nabla_{\boldsymbol{X}} G(0, \boldsymbol{X}) = 0$ implies $G(0, \boldsymbol{X}) = 0$ by Euler's identity.) Since the singular sets are identical, this proves the claim.

Let us apply this in our case with G taken to be the polynomial $F(Y, \boldsymbol{X})$, with $V(F(Y, \boldsymbol{X}))$ assumed to be nonsingular. We consider whether there are any $(0, \boldsymbol{x}) \in \mathbb{P}^n$ such that $\nabla_{\boldsymbol{X}} F(0, \boldsymbol{x}) = 0$. Supposing such $(0, \boldsymbol{x})$ exists, it must be the case that $(\partial F/\partial Y)(0, \boldsymbol{x}) \neq 0$, since otherwise $(0, \boldsymbol{x})$ would be a singular point on $V(F(Y, \boldsymbol{X}))$. If $m \geq 2$, then due to the leading factor Y^{m-1} in (3-11), any point $(0, \boldsymbol{x}) \in \mathbb{P}^n$ must lead to $(\partial F/\partial Y)(0, \boldsymbol{x}) = 0$. Consequently there can be no such $(0, \boldsymbol{x})$, and $\operatorname{Sing}(V(F(Y, \boldsymbol{X})) \cap V(Y))$ must be empty. Hence by the general argument above, so is $\operatorname{Sing}(V(F(Z^e, \boldsymbol{X})))$. In conclusion, if $m \geq 2$, $V(F(Y, \boldsymbol{X}))$ being nonsingular implies $V(F(Z^e, \boldsymbol{X}))$ is nonsingular.

However if $m = 1$, there is no leading factor of Y in (3-11), and indeed at $(0, \boldsymbol{x})$, (3-11) evaluates to $f_{d-1}(\boldsymbol{x})$. Thus points $(0, \boldsymbol{x})$ for which $f_{d-1}(\boldsymbol{x}) \neq 0$ and $\nabla_{\boldsymbol{X}} F(0, \boldsymbol{x}) = 0$ can lead to singular points on $V(F(Y, \boldsymbol{X})) \cap V(Y)$ and hence to singular points on $F(F(Z^e, \boldsymbol{X}))$. (Nevertheless, there cannot be too many singular points, as we will observe in (4-1) below that the singular locus has at most dimension 0.)

In the other direction, suppose that $V(F(Z^e, \boldsymbol{X}))$ is nonsingular, so that as computed in (3-12),

$$\operatorname{Sing}(V(F(Z^e, \boldsymbol{X}))) = \{(0, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{\boldsymbol{X}} F(0, \boldsymbol{x}) = \boldsymbol{0}\} \cup \{(z, \boldsymbol{x}) \in \mathbb{P}^n : \nabla_{Y,\boldsymbol{X}} F(z^e, \boldsymbol{x}) = \boldsymbol{0}\}$$

is empty. If there were a point $(y, \boldsymbol{x})$ in $\operatorname{Sing}(V(Y, \boldsymbol{X}))$ then if $y = 0$ this would produce an element in the first set on the right-hand side, while if $y \neq 0$ then taking $z = y^{1/e}$ (working over $\mathbb{C}$) would produce a point in the second set on the right-hand side. Thus $V(F(Y, \boldsymbol{X}))$ must be nonsingular (and here we did not need to apply $m \geq 2$).

Remark 3.2. In the special case that $d = 1$, then $F(Y, \boldsymbol{X}) = Y^m + f_1(\boldsymbol{X})$. Thus $V(F(Y, \boldsymbol{X})) \subset \mathbb{P}(e, 1, \dots, 1)$ is nonsingular if and only if $V(Z^{em} + f_1(\boldsymbol{X})) \subset \mathbb{P}^n$ is nonsingular, with $f_1 \not\equiv 0$ homogeneous of degree em. This occurs if and only if $V(f_1(\boldsymbol{X})) \subset \mathbb{P}^{n-1}$ is nonsingular; in this special case, the problem we consider falls in the scope of the work in [2, Theorem 1.1], which proves this case of Theorem 1.1. Our method of proof works regardless, so we allow $d = 1$ as we continue.

Remark 3.3. Recall the affine hypersurface $\mathcal{V} \subset \mathbb{A}^{n+1}_{\mathbb{C}}$ defined in (1-2) according to the polynomial $F(Y, \boldsymbol{X})$. We note that $\mathcal{V}$ is irreducible under the conditions of Theorem 1.1. Suppose it is reducible, so that $F(Y, \boldsymbol{X}) = G(Y, \boldsymbol{X})H(Y, \boldsymbol{X})$ for some nonconstant polynomials. Then $F(Z^e, \boldsymbol{X}) = G(Z^e, \boldsymbol{X})H(Z^e, \boldsymbol{X})$ so that the projective variety $V(F(Z^e, \boldsymbol{X}))$ is reducible. Consequently, by [13, Lemma 11.1], $V(F(Z^e, \boldsymbol{X}))$ is singular, which is a contradiction because by the discussion above, $V(F(Y, \boldsymbol{X}))$ is nonsingular if and only if $V(F(Z^e, \boldsymbol{X}))$ is nonsingular.

3.4. ***Initial considerations of the sieving set.*** We suppose that $Q = B^\kappa$ for some $0 < \kappa \le 1$ to be chosen later (see (7-4)). We will choose a sieving set

$$\mathcal{P} \subset [Q, 2Q]$$

comprised of primes with certain properties. In the special case that $(e, m) = 1$, it is sensible to restrict our attention to a set $\mathcal{P}_0$ of primes in $[Q, 2Q]$ such that

(i) $p \equiv 1 \pmod{m}$ (recalling $m \ge 2$) and

(ii) $p \equiv 2 \bmod e$, and

(iii) the reduction of $V(F(Y, \boldsymbol{X}))$ as a weighted variety over $\overline{\mathbb{F}}_p$ is nonsingular.

The first criterion (i) we have used in the proof of the sieve lemma (Lemma 1.2). The second criterion (ii) ensures that $(e, p-1) = 1$ so that every $y \in \mathbb{F}_p$ satisfies $y = z^e$ for some $z \in \mathbb{F}_p$. Then for each $p \in \mathcal{P}$, we can simply consider the reduction $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\mathbb{F}_p}$ in place of the weighted variety, so that (iii) is equivalent to

(iii′) the reduction of $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is nonsingular.

By the Chinese remainder theorem and the Siegel–Walfisz theorem on primes in arithmetic progressions, under the assumption that $(e, m) = 1$, there are $\gg_{m,e} Q/\log Q$ primes that satisfy (i) and (ii) in any dyadic region $[Q, 2Q]$, for all Q sufficiently large. We could then choose the sieving set $\mathcal{P}_0$ to be the subset of such primes for which (iii′) holds; the remaining task is to show there are sufficiently few primes that violate (iii′).

Recall from Section 3.3 that $V(F(Y, \boldsymbol{X}))$ is nonsingular over $\mathbb{C}$ (as a weighted projective variety) if and only if $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n$ is nonsingular over $\mathbb{C}$. Thus under the hypothesis of Theorem 1.1, the latter is nonsingular, and consequently there are no nontrivial simultaneous solutions of the system (3-10), and thus the resultant

$$r := \operatorname{Res}\left(F, \frac{\partial F}{\partial Z}, \frac{\partial F}{\partial X_1}, \dots, \frac{\partial F}{\partial X_n}\right)$$

of those $n+2$ polynomials in $n+1$ variables is a nonzero integer. Moreover, by [21, Chapter 13, Proposition 1.1], r is a polynomial in the coefficients of F with degree bounded in terms of m, e, d. By [15, Section IV], the reduction $V_p(F(Z^e, \boldsymbol{X}))$ of $V(F(Z^e, \boldsymbol{X}))$ modulo p is singular precisely when $p|r$, which can only occur for at most $\omega(r)$ primes, where

$$\omega(r) \ll \log r / \log\log r \ll_{m,e,d} \log \|F\|. \tag{3-13}$$

(Notice that the argument in this paragraph made no assumption on the relative primality of e and m.)

In particular, if $(e, m) = 1$, then as long as Q is sufficiently large, say $Q \gg_{m,e,d} (\log \|F\|)^{1+\delta_0}$ for any fixed $\delta_0 > 0$ or even $Q \gg_{m,e,d} (\log \|F\|)(\log\log \|F\|)$, we can conclude that $|\mathcal{P}_0| \gg_{m,e,d} Q/\log Q$. After we choose Q to be a certain power of B (see (7-4)), this will only require a lower bound on B that is on the order of a power of $\log \|F\|$, which we will see can be accommodated by the bound on the right-hand side of our claim in Theorem 1.1.

These remarks all apply in the case that $(e, m) = 1$. However, we can also argue more generally without this assumption, as we demonstrate in the next section, by working not with $V(F(Z^e, \boldsymbol{X}))$ as above, but with a finite collection of varieties W_i, defined according to $F(\gamma^i z^e, \boldsymbol{X}) = 0$ in $\mathbb{F}_p$, for a certain primitive root $\gamma \in \mathbb{F}_p^\times$ (see Lemma 4.3). Thus we postpone our definition of the sieving set, in general, until the end of the next section.

4. Estimates for exponential sums

In this section we apply the Weil bound to prove an upper bound for the exponential sum $g(\boldsymbol{u}, p)$ (see (1-24)) in the case that $\boldsymbol{u}$ is each of three types: type zero, good, or bad modulo p (Definition 4.1). At the end, in Section 4.2 we then define the sieving set $\mathcal{P}$.

We note the multiplicativity condition

$$g(\boldsymbol{u}, pq) := \sum_{\boldsymbol{a} \bmod pq} (\nu_p(\boldsymbol{a}) - 1)(\nu_q(\boldsymbol{a}) - 1) e_{pq}(\langle \boldsymbol{a}, \boldsymbol{u}\rangle) = g(\bar{q}\boldsymbol{u}, p) g(\bar{p}\boldsymbol{u}, q),$$

where $q\bar{q} \equiv 1 \bmod p$, and $p\bar{p} \equiv 1 \bmod q$. This leads us to study the key exponential sums with prime modulus:

$$g(\boldsymbol{u}, p) := \sum_{\boldsymbol{a} \in \mathbb{F}_p^n} (\nu_p(\boldsymbol{a}) - 1) e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle).$$

Let p be a fixed prime of good reduction for $F(Z^e, \boldsymbol{X})$, so that $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is a nonsingular projective hypersurface. For any point $P \in V(F(Z^e, \boldsymbol{X}))$, let $T_{\boldsymbol{P}} \subseteq \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ denote the projective tangent space to $V(F(Z^e, \boldsymbol{X}))$ at P. A linear space L is tangent to $V(F(Z^e, \boldsymbol{X}))$ at P if $T_{\boldsymbol{P}} \subseteq L$; if L is a hyperplane, this is equivalent to P being a singular point of $V(F(Z^e, \boldsymbol{X})) \cap L$ (see [20, p. 57]).

Given $\boldsymbol{u} \in \mathbb{Z}^n$ with $\boldsymbol{u} \not\equiv \boldsymbol{0} \pmod p$, if $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is not tangent to $V(F(Z^e, \boldsymbol{X}))$ at any point (i.e., they intersect transversely), we simply say $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is not tangent to $V(F(Z^e, \boldsymbol{X}))$; otherwise, we will say they are tangent (and as we will discuss below in (4-1), there are at most finitely many points at which they are tangent).

Using this terminology, we will classify $\boldsymbol{u} \in \mathbb{Z}^n$ in terms of three cases:

Definition 4.1. For $\boldsymbol{u} \in \mathbb{Z}^n$ and $p \in \mathcal{P}$ we say that:

(i) $\boldsymbol{u}$ is of type zero mod p if $\boldsymbol{u} \equiv \mathbf{0} \pmod{p}$,

(ii) $\boldsymbol{u}$ is good mod p if $\boldsymbol{u} \not\equiv \mathbf{0} \pmod{p}$ and $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is not tangent to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$,

(iii) $\boldsymbol{u}$ is bad mod p if $\boldsymbol{u} \not\equiv \mathbf{0} \pmod{p}$, and $V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is tangent to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$.

(The fact that we define these types in relation to $V(F(Z^e, \boldsymbol{X}))$, is justified by Lemma 4.4, below.) The main result of this section is the following:

Proposition 4.2. *Assume that $p > 2$ is a prime of good reduction for $F(Z^e, \boldsymbol{X})$, that is $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is nonsingular.*

(i) *If $\boldsymbol{u}$ is type zero modulo p then $g(\boldsymbol{u}, p) \ll p^{n-1/2}$;*

(ii) *If $\boldsymbol{u}$ is good modulo p then $g(\boldsymbol{u}, p) \ll p^{n/2}$;*

(iii) *If $\boldsymbol{u}$ is bad modulo p then $g(\boldsymbol{u}, p) \ll p^{(n+1)/2}$.*

The implied constants can depend on n, m, e, d, but are independent of $\|F\|, \boldsymbol{u}, p$.

In a final step of the proof, we will apply the property that if $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n$ is nonsingular, any hyperplane L has

$$\dim\{P \in V(F(Z^e, \boldsymbol{X})) : T_P \subseteq L\} = \dim(\mathrm{Sing}(V(F(Z^e, \boldsymbol{X})) \cap L)) \leq 0. \tag{4-1}$$

Here, by $\dim(\mathrm{Sing}(V))$ we mean the dimension of the singular locus of a variety $V \subset \mathbb{P}^n$. We will apply this in (4-3) over $\overline{\mathbb{F}}_p$ for p a prime of good reduction for $F(Z^e, \boldsymbol{X})$. The result (4-1) is a special case of Zak's theorem on tangencies as in [20, Theorem 7.1, Remark 7.5], valid over any algebraically closed field, or [33, Lemma 3], valid over any perfect field. More simply, in our setting (4-1) can be shown directly, and we do so in Remark 4.5.

As preparation for proving Proposition 4.2, we transform $g(\boldsymbol{u}, p)$ into an exponential sum over solutions to $F(y, \boldsymbol{a}) = 0$ by writing

$$\begin{aligned} g(\boldsymbol{u}, p) &= \sum_{\boldsymbol{a} \in \mathbb{F}_p^n} \nu_p(\boldsymbol{a}) e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle) - \sum_{\boldsymbol{a} \in \mathbb{F}_p^n} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle) \\ &= -\delta_{\boldsymbol{u}=\mathbf{0}} \cdot p^n + \sum_{\boldsymbol{a} \in \mathbb{F}_p^n} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle) \sum_{\substack{y \in \mathbb{F}_p \\ F(y, \boldsymbol{a}) = 0}} 1 \\ &= -\delta_{\boldsymbol{u}=\mathbf{0}} \cdot p^n + \sum_{\substack{(y, \boldsymbol{a}) \in \mathbb{F}_p^{n+1} \\ F(y, \boldsymbol{a}) = 0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle), \end{aligned}$$

where $\delta_{\boldsymbol{u}=\boldsymbol{0}} = 1$ if $\boldsymbol{u} \equiv \boldsymbol{0} \pmod{p}$ and is 0 otherwise. The task now is to estimate the sum

$$g(\boldsymbol{u}, p) + \delta_{\boldsymbol{u}=\boldsymbol{0}} \cdot p^n = \sum_{\substack{(y,\boldsymbol{a}) \in \mathbb{F}_p^{n+1} \\ F(y,\boldsymbol{a})=0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u} \rangle).$$

A barrier to doing this efficiently is that the polynomial $F(Y, \boldsymbol{X})$ is not homogeneous (see Remark 4.6). Recall the definition of $F(Y, \boldsymbol{X})$ in (1-1), and recall the integer $e \geq 1$ fixed in that definition. As a first step, we prove:

Lemma 4.3. *Fix a prime $p > 2$. Let $f = (e, p-1)$, and let $\gamma \in \mathbb{F}_p^\times$ be a primitive f-th root of unity. Then*

$$\sum_{(y,\boldsymbol{a}) \in W} e_p(\langle \boldsymbol{a}, \boldsymbol{u} \rangle) = \frac{1}{f} \sum_{i=0}^{f-1} \sum_{(z,\boldsymbol{a}) \in W_i} e_p(\langle \boldsymbol{a}, \boldsymbol{u} \rangle),$$

where

$$W = \{(y, \boldsymbol{a}) \in \mathbb{F}_p^{n+1} : F(y, \boldsymbol{a}) = 0\},$$
$$W_i = \{(z, \boldsymbol{a}) \in \mathbb{F}_p^{n+1} : F(\gamma^i z^e, \boldsymbol{a}) = 0\}, \quad \textit{for } i = 0, \ldots, f-1.$$

(This lemma replaces the remarks in Section 3.4 that applied in the special case $(e, p-1) = 1$.)

Proof. We start by claiming that for any $y \in \mathbb{F}_p^\times$ there exists an unique $i \in \{0, \ldots, f-1\}$ and some $z \in \mathbb{F}_p^\times$ such that $y = \gamma^i z^e$: we write $e = \ell k$ where

$$(\ell, q) = 1 \text{ for any } q \,|\, (p-1), \qquad k = \frac{e}{\ell}.$$

Note that then $f | k$ and also there exists some integer N such that $k | (f^N)$. Since γ is a generator for the group $\mathbb{F}_p^\times / \mathbb{F}_p^{\times f}$, then for any $y \in \mathbb{F}_p^\times$ there exists an unique $i \in \{0, \ldots, f-1\}$ and $z_1 \in \mathbb{F}_p^\times$ such that $y = \gamma^i z_1^f$. On the other hand, we can apply the same principle to z_1, finding an unique $j \in \{0, \ldots, f-1\}$ and $z_2 \in \mathbb{F}_p^\times$ such that $z_1 = \gamma^j z_2^f$. Thus, $y = \gamma^i z_1^f = \gamma^i (\gamma^j z_2^f)^f = \gamma^i z_2^{f^2}$. Iterating this process N times, we can find $z_N \in \mathbb{F}_p^\times$ such that $y = \gamma^i z_N^{f^N}$ with $k | f^N$. Then, $y = \gamma^i (z_N^{f^N/k})^k$. On the other hand, since $(\ell, p-1) = 1$, we have that $z_N^{f^N/k} = z^\ell$ for some $z \in \mathbb{F}_p^\times$, so that $y = \gamma^i z^{\ell k} = \gamma^i z^e$ and this proves the claim. Moreover, note that once we have obtained z such that $y = \gamma^i z^e$ then we can multiply z by any f-th root of unity, so that there are f such values z.

Next, for any $i \in \{0, \ldots, f-1\}$ we can consider the map

$$\varphi_i : W_i \longrightarrow W \quad (z, \boldsymbol{a}) \mapsto (\gamma^i z^e, \boldsymbol{a}).$$

From this, we deduce that if $(y, \boldsymbol{a})$ is in the image of φ_i then

$$|\varphi_i^{-1}(y, \boldsymbol{a})| = \begin{cases} f & \text{if } y \neq 0, \\ 1 & \text{if } y = 0. \end{cases}$$

On the other hand, if $(0, \boldsymbol{a}) \in W$, then $(0, \boldsymbol{a}) \in W_i$ for each of $i = 0, \ldots, f-1$. The result follows. □

When we apply Lemma 4.3 it will be convenient to treat all cases analogously as i varies; to do so we will employ the following lemma.

Lemma 4.4. *Fix $e \geq 1$ and recall $F(Y, \boldsymbol{X})$ from* (1-1). *Let p be a prime, and let $\boldsymbol{u} \in \overline{\mathbb{F}}_p^n$. Then for any $\alpha \in \overline{\mathbb{F}}_p^\times$ the variety $V(F(\alpha Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is isomorphic to $V(F(Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$. In particular, for $\boldsymbol{u} = \boldsymbol{0}$, we conclude $V(F(\alpha Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is isomorphic to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$.*

Proof. Let $\beta \in \overline{\mathbb{F}}_p^\times$ be such that $\beta^e = \alpha$. Then the change of variables $(Z, \boldsymbol{X}) \mapsto (\beta Z, \boldsymbol{X})$ induces an isomorphism between $V(F(Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ and $V(F(\alpha Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$. □

4.1. *Proof of Proposition 4.2.* We are now ready to prove our main result of this section, Proposition 4.2. In the following, we denote $f = (e, p-1)$. An application of Lemma 4.3 leads to

$$g(\boldsymbol{u}, p) = -\delta_{\boldsymbol{u}=\boldsymbol{0}} p^n + \frac{1}{f} \sum_{i=0}^{f-1} \sum_{(z,\boldsymbol{a}) \in W_i} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle). \tag{4-2}$$

4.1.1. *Type zero case.* Assume $\boldsymbol{u} \equiv \boldsymbol{0} \pmod{p}$. The right-hand side of (4-2) becomes

$$g(\boldsymbol{0}, p) = -p^n + \frac{1}{f} \sum_{i=0}^{f-1} \sum_{(z,\boldsymbol{a}) \in W_i} 1 = -p^n + \frac{1}{f} \sum_{i=0}^{f-1} |W_i|.$$

By definition, for any $i = 0, \ldots, f-1$ the set W_i is the set of the $\mathbb{F}_p$-points on the affine variety $V(F(\gamma^i Z^e, \boldsymbol{X})) \subset \mathbb{A}^{n+1}_{\mathbb{F}_p}$. By hypothesis, p is of good reduction for $V(F(Z^e, \boldsymbol{X}))$, so $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is nonsingular. Then by Lemma 4.4, we have that $V(F(\gamma^i Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is a nonsingular variety for each $i = 0, \ldots, f-1$ (and in particular is absolutely irreducible over $\overline{\mathbb{F}}_p$), and certainly $V(F(\gamma^i Z^e, \boldsymbol{X}))$ is defined over $\mathbb{F}_p$. Thus the Lang-Weil bound [35] implies that (counting projectively)

$$|V(F(\gamma^i Z^e, \boldsymbol{X}))(\mathbb{F}_p)| = p^{n-1} + O_{m,e,d}(p^{n-1-1/2}) \qquad \text{for each } i = 0, \ldots, f-1,$$

so that $|W_i| = p^n + O_{m,e,d,}(p^{n-1/2})$ for each $i = 0, \ldots, f-1$. Thus we may conclude that $g(\boldsymbol{0}, p) \ll p^{n-1/2}$.

4.1.2. *Good/bad case.* Assume $\boldsymbol{u} \not\equiv \boldsymbol{0} \pmod{p}$; we may initially argue the good and the bad cases together. The right hand side of (4-2) becomes

$$g(\boldsymbol{u}, p) = \frac{1}{f} \sum_{i=0}^{f-1} \sum_{(z,\boldsymbol{a}) \in W_i} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle).$$

In either the good or the bad case, it suffices to estimate each sum

$$g_i(\boldsymbol{u}, p) = \sum_{(z,\boldsymbol{a}) \in W_i} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle), \qquad \text{for } i = 0, .., f-1.$$

First we prove that for any $\alpha \in \mathbb{F}_p^\times$, $g_i(\boldsymbol{u}, p) = g_i(\alpha\boldsymbol{u}, p)$. Indeed

$$\begin{aligned} g_i(\alpha\boldsymbol{u}, p) &= \sum_{(z,\boldsymbol{a})\in W_i} e_p(\langle \boldsymbol{a}, \alpha\boldsymbol{u}\rangle) = \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} e_p(\langle \boldsymbol{a}, \alpha\boldsymbol{u}\rangle) \\ &= \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} e_p(\langle \alpha\boldsymbol{a}, \boldsymbol{u}\rangle) = \sum_{\substack{(t,\boldsymbol{b})\in\mathbb{F}_p^{n+1}\\ \overline{\alpha}^{med} F(\gamma^i t^e,\boldsymbol{b})=0}} e_p(\langle \boldsymbol{b}, \boldsymbol{u}\rangle) \\ &= \sum_{\substack{(t,\boldsymbol{b})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i t^e,\boldsymbol{b})=0}} e_p(\langle \boldsymbol{b}, \boldsymbol{u}\rangle) = g_i(\boldsymbol{u}, p), \end{aligned}$$

where in the fourth step we use the change of variables $(z, \boldsymbol{a}) = (\overline{\alpha}t, \overline{\alpha}\boldsymbol{b})$, for $\alpha\overline{\alpha} \equiv 1 \pmod p$. Hence

$$\begin{aligned} (p-1)g_i(\boldsymbol{u}, p) &= \sum_{\alpha\in\mathbb{F}_p^\times} g_i(\alpha\boldsymbol{u}, p) \\ &= \sum_{\alpha\in\mathbb{F}_p^\times} \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} e_p(\langle \boldsymbol{a}, \alpha\boldsymbol{u}\rangle) \\ &= \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} \sum_{\alpha\in\mathbb{F}_p^\times} e_p(\alpha\langle \boldsymbol{a}, \boldsymbol{u}\rangle) = \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} \sum_{\alpha\in\mathbb{F}_p} e_p(\alpha\langle \boldsymbol{a}, \boldsymbol{u}\rangle) - \sum_{\substack{(z,\boldsymbol{a})\in\mathbb{F}_p^{n+1}\\ F(\gamma^i z^e,\boldsymbol{a})=0}} 1 \\ &= p(p-1)\big|(V(F(\gamma^i Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{u}, \boldsymbol{X}\rangle))(\mathbb{F}_p)\big| - (p-1)\big|V(F(\gamma^i Z^e, \boldsymbol{X})(\mathbb{F}_p)\big| + (p-1), \end{aligned}$$

where in the last step we have passed to counting points over $\mathbb{F}_p$ in the projective sense. Applying [32, Appendix by N. Katz, Theorem 1], we have that

$$|V(F(\gamma^i Z^e, \boldsymbol{X}))(\mathbb{F}_p)| = \sum_{j=0}^{n-1} p^j + O_{n,m,e,d}(p^{\frac{n+\delta_i}{2}}),$$

$$|(V(F(\gamma^i Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{u}, \boldsymbol{X}\rangle))(\mathbb{F}_p)| = \sum_{j=0}^{n-2} p^j + O_{n,m,e,d}(p^{\frac{n-1+\delta_{i,\boldsymbol{u}}}{2}}),$$

where $\delta_i = \dim(\mathrm{Sing}(V(F(\gamma^i Z^e, \boldsymbol{X}))))$ and $\delta_{i,\boldsymbol{u}} = \dim(\mathrm{Sing}(V(F(\gamma^i Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{u}, \boldsymbol{X}\rangle)))$.

On the other hand, Lemma 4.4 implies that $\delta_i = \delta_0$ and $\delta_{i,\boldsymbol{u}} = \delta_{0,\boldsymbol{u}}$ for each i. Moreover, $\delta_0 = -1$ since we are assuming that p is of good reduction for $V(F(Z^e, \boldsymbol{X}))$. Thus, we obtain

$$g_i(\boldsymbol{u}, p) = O(p^{\frac{n+1+\delta_{0,\boldsymbol{u}}}{2}}), \tag{4-3}$$

with an implicit constant depending only on n, m, e, d. Finally, by (4-1),

$$\delta_{0,\boldsymbol{u}} = \begin{cases} 0 & \text{if } V(\langle \boldsymbol{u}, \boldsymbol{X}\rangle) \text{ is tangent to } V(F(Z^e, \boldsymbol{X})), \\ -1 & \text{otherwise,} \end{cases}$$

and this completes the proof of the good and bad cases in Proposition 4.2.

Remark 4.5. This remark justifies (4-1). Let $V = V(H(\boldsymbol{X})) \subset \mathbb{P}^n$ be a nonsingular hypersurface and $L = V(\langle \boldsymbol{a}, \boldsymbol{X}\rangle)$ be a hyperplane. We may suppose without loss of generality that $a_1 \neq 0$. By the Jacobian criterion, $\mathrm{Sing}(V \cap L)$ is the set of points on the intersection $V \cap L$ for which the $(n+1)\times 2$ matrix with columns ∇H and $\boldsymbol{a}$ has rank 1. Consequently, $\mathrm{Sing}(V \cap L) \subset W$ where

$$W = V \cap V(g_2) \cap \cdots \cap V(g_n),$$

in which for each $i = 2, \dots, n$,

$$g_i(\boldsymbol{X}) = a_1 \frac{\partial H}{\partial X_i}(\boldsymbol{X}) - a_i \frac{\partial H}{\partial X_1}(\boldsymbol{X}).$$

On the other hand, $W \cap V(\partial H/\partial X_1) = \mathrm{Sing}(V) = \varnothing$ under the hypothesis that V is nonsingular. Consequently, $\dim W \le 0$, implying $\dim(\mathrm{Sing}(V \cap L)) \le 0$, as desired.

Remark 4.6. It is worth remarking what we have gained from the arguments in this section. Briefly, suppose $\boldsymbol{u} \not\equiv 0 \pmod p$ and consider

$$g(\boldsymbol{u}, p) = \sum_{\substack{(y,\boldsymbol{a}) \in \mathbb{F}_p^{n+1} \\ F(y,\boldsymbol{a})=0}} e_p(\langle \boldsymbol{a}, \boldsymbol{u}\rangle).$$

To work directly with this sum rather than passing through the dissection into the components W_i as we did above, we would first need to homogenize the polynomial $F(Y, \boldsymbol{x})$, say defining a homogeneous polynomial

$$\tilde{F}(T, Y, \boldsymbol{X}) = T^{md(e-1)} Y^{md} + \cdots + T^{m(e-1)} Y^m f_{d-1}(\boldsymbol{X}) + f_d(\boldsymbol{X}).$$

(Here we suppose that $e \ge 2$ for this example.) Then observe that $[1 : 0 : \cdots : 0]$ is a singular point on $V(\tilde{F}(T, Y, \boldsymbol{X})) \subset \mathbb{P}^{n+1}$. Consequently, if one proceeded to estimate $g(\boldsymbol{u}, p)$, roughly analogous to the approach in (4-3), by counting points on the complete intersection described by

$$V(\tilde{F}(T, Y, \boldsymbol{X})) \cap V(\langle \boldsymbol{u}, \boldsymbol{X}\rangle) \cap V(T = 1),$$

the role of $\delta_{0,\boldsymbol{u}}$ in the exponent is now played by a dimension that is always at least 0, ultimately leading to a result that is larger by a factor of $p^{1/2}$ than the results we obtain in Proposition 4.2.

4.2. ***Choice of the sieving set.*** We can now continue the discussion initiated in Section 3.4, and choose the sieving set. We suppose that $Q = B^\kappa$ for some $1/2 \le \kappa \le 1$ to be chosen later (see (7-4)). We choose the sieving set

$$\mathcal{P} \subset [Q, 2Q]$$

comprised of all primes in this range such that (i) $p \equiv 1 \pmod m$ (recalling $m \ge 2$), and (iii′) the reduction $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\overline{\mathbb{F}}_p}$ is nonsingular.

By the Siegel–Walfisz theorem on primes in arithmetic progressions, there are $\gg_m Q/\log Q$ primes such that $p \equiv 1 \pmod m$ in any dyadic region $[Q, 2Q]$, for all $Q \gg_m 1$ sufficiently large, which we assume is a condition met henceforward. We recall from (3-13) that at most $O_{m,e,d}(\log \|F\|)$ primes fail (iii′). We henceforward assume that

$$Q \gg_{m,e,d} (\log \|F\|)(\log\log \|F\|) \tag{4-4}$$

for an appropriately large implied constant, so that consequently

$$P = |\mathcal{P}| \gg_m Q/\log Q - C_{m,e,d}(\log \|F\|) \gg_{m,e,d} Q/\log Q. \tag{4-5}$$

When we finally choose Q as a power of B, (4-4) will impose a lower bound on B; we defer this to (7-4).

5. Estimating the main sieve term: the bad-bad case

This section is the technical heart of the paper. We show how to bound the most difficult contribution to the sieve, which occurs when $\boldsymbol{u}$ is bad with respect to two primes $p \neq q \in \mathcal{P}$. (We reserve the treatment of all other cases, when $\boldsymbol{u}$ is either type zero, or good with respect to at least one of these primes, to Section 7; these remaining cases are significantly easier.)

We recall from the sieve lemma, Lemma 1.2, that $\mathcal{S}(F, B)$ is bounded above by a sum of three terms. The first two terms can be bounded simply:

$$\sum_{\boldsymbol{k}: f_d(\boldsymbol{k})=0} W(\boldsymbol{k}) + \frac{1}{P}\sum_{\boldsymbol{k}} W(\boldsymbol{k}) \ll B^{n-1} + B^n P^{-1}. \tag{5-1}$$

Here the first term follows from the Schwartz-Zippel trivial bound $\ll_{n,e,d} B^{n-1}$ for the number of zeroes of f_d with $\boldsymbol{k} \in \operatorname{supp}(W)$, since $f_d \not\equiv 0$ (see, e.g., [27, Theorem 1], which as mentioned before has a method of proof that applies even if f_d is not absolutely irreducible). We will call the remaining, third, term on the right-hand side of the sieve lemma the main sieve term.

Now we are ready to estimate the main sieve term, which after an application of Poisson summation inside the definition (1-23) of $T(p, q)$ is

$$\begin{aligned}\frac{1}{P^2}\sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} |T(p,q)| &= \frac{1}{P^2}\sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \left(\frac{1}{pq}\right)^n \left|\sum_{\boldsymbol{u}} \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \\ &\ll \frac{1}{P^2 Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\boldsymbol{u}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right|. \end{aligned} \tag{5-2}$$

We will apply Proposition 4.2 to bound $g(\boldsymbol{u}, pq)$, according to the "type" of $\boldsymbol{u}$ modulo p and q, respectively; this leads to cases we can abbreviate as zero-zero, zero-good, zero-bad, good-good, good-bad, and bad-bad. Unsurprisingly, the greatest difficulty is to bound the contribution of the bad-bad case, and we focus on this first, returning to the other cases in Section 7.

Recall that W is a nonnegative function with $W(\boldsymbol{u}) = w(\boldsymbol{u}/B)$ for an infinitely differentiable, nonnegative function w that is $\equiv 1$ on $[-1,1]$ and vanishes outside of $[-2,2]$. Thus $\hat{W}(\boldsymbol{u}) = B^n \hat{w}(B\boldsymbol{u})$ and $\hat{w}(\boldsymbol{u})$ has rapid decay in $\boldsymbol{u}$, so that

$$|\hat{W}(\boldsymbol{u})| \ll B^n \prod_{i=1}^{n} (1+|u_i|B)^{-M} \tag{5-3}$$

for any $M \geq 1$; we will for example specify a lower bound on M at (5-22) and can certainly always assume $M \geq 2n$. In particular, we will later apply the fact that for any $B, L \geq 1$,

$$\sum_{\boldsymbol{u} \in \mathbb{Z}^n} |\hat{W}(\boldsymbol{u}/L)| \ll \max\{B^n, L^n\}. \tag{5-4}$$

5.1. *The dual variety.* To consider any bad case, it is useful to consider certain facts about the dual variety. Recall that $m \geq 2$ and $d, e \geq 1$, and

$$F(Y, \boldsymbol{X}) = Y^{md} + Y^{m(d-1)} f_1(\boldsymbol{X}) + \cdots + f_d(\boldsymbol{X}), \tag{5-5}$$

in which for each $1 \leq i \leq d$, f_i is a polynomial in $\mathbb{Z}[X_1, \dots, X_n]$ with $\deg f_i = m \cdot e \cdot i$. By hypothesis, the variety defined by $F(Y, \boldsymbol{X}) = 0$ in weighted projective space, denoted $V(F(Y, \boldsymbol{X})) \subset \mathbb{P}_{\mathbb{C}}(e, 1, \dots, 1)$, is nonsingular. Recall from Section 3.3 that $V(F(Y, \boldsymbol{X})) \subset \mathbb{P}_{\mathbb{C}}(e, 1, \dots, 1)$ is nonsingular if and only if $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\mathbb{C}}$ is nonsingular. The dual variety $V^* = V(F(Z^e, \boldsymbol{X}))^* \subset \mathbb{P}^n_{\mathbb{C}}$ of a hypersurface is a hypersurface. We denote by

$$G(U_Y, U_1, \dots, U_n) \tag{5-6}$$

the irreducible homogeneous polynomial such that $V(G) = V^*$ (see, e.g., [13, Proposition 11.2, Appendix]). Recall that $\deg F(Z^e, \boldsymbol{X}) = mde$; by [19, Proposition 2.9],

$$\deg G = mde(mde-1)^{n-1} \geq 2.$$

In our analysis of the bad-bad case in Section 5.2, our strategy is to divide our analysis depending on whether $\boldsymbol{u}$ has the property $G(0, \boldsymbol{u}) \neq 0$ or $G(0, \boldsymbol{u}) = 0$. In the first case, we now show via an explicit constructive argument that

$$|\{p : \boldsymbol{u} \text{ is bad modulo } p\}| \ll_{n,m,e,d} \log(\|F\| \|\boldsymbol{u}\|). \tag{5-7}$$

Let us prove this. A given $\boldsymbol{u}$ has the property $G(0, \boldsymbol{u}) \neq 0$ if and only if the hyperplane $V(\langle \boldsymbol{u}, \boldsymbol{X} \rangle) \subset \mathbb{P}^n_{\mathbb{C}}$ is not tangent to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\mathbb{C}}$; that is, if and only if for any $[z : \boldsymbol{x}] \in V(F(Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)$, the matrix

$$\begin{pmatrix} \frac{\partial F}{\partial Z}(z^e, \boldsymbol{x}) & 0 \\ \frac{\partial F}{\partial X_1}(z^e, \boldsymbol{x}) & u_1 \\ \vdots & \\ \frac{\partial F}{\partial X_n}(z^e, \boldsymbol{x}) & u_n \end{pmatrix} \tag{5-8}$$

has maximal rank (i.e., at least one 2×2 minor is nonvanishing). Now define $n+2$ polynomials in $Z, X_1, \dots, X_n$, with integral coefficients (depending on $\boldsymbol{u}$) as follows: set

$$H_{0,\boldsymbol{u}}(Z, \boldsymbol{X}) = F(Z^e, \boldsymbol{X}), \qquad H_{n+1,\boldsymbol{u}}(Z, \boldsymbol{X}) = \langle \boldsymbol{X}, \boldsymbol{u} \rangle,$$

and for $1 \le i \le n$ set

$$H_{i,\boldsymbol{u}}(Z, \boldsymbol{X}) = \begin{cases} \det \begin{pmatrix} \frac{\partial F}{\partial Z}(z^e, \boldsymbol{x}) & 0 \\ \frac{\partial F}{\partial X_1}(z^e, \boldsymbol{x}) & u_1 \end{pmatrix} & \text{for } i = 1, \\ \det \begin{pmatrix} \frac{\partial F}{\partial X_{i-1}}(z^e, \boldsymbol{x}) & u_{i-1} \\ \frac{\partial F}{\partial X_i}(z^e, \boldsymbol{x}) & u_i \end{pmatrix} & \text{for } 2 \le i \le n. \end{cases}$$

Then define the resultant (see [21, Chapter 13])

$$R(\boldsymbol{u}) = \operatorname{Res}(H_{0,\boldsymbol{u}}, H_{1,\boldsymbol{u}}, \dots, H_{n+1,\boldsymbol{u}}). \tag{5-9}$$

The following are all equivalent:

(1) $\boldsymbol{u}$ has the property that $V(\langle \boldsymbol{u}, \boldsymbol{X} \rangle)$ is tangent to $V(F(Z^e, \boldsymbol{X}))$.

(2) For some $[z : \boldsymbol{x}] \in V(F(Z^e, \boldsymbol{X})) \cap V(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)$, (5-8) has rank < 2.

(3) The polynomials $H_{i,\boldsymbol{u}}(Z, \boldsymbol{X})$ (for $0 \le i \le n+1$) share a common (nonzero) root.

(4) $R(\boldsymbol{u}) = 0$.

Now we consider the analogues of these statements for each p. Fix a prime p. For a polynomial $L \in \mathbb{Z}[\boldsymbol{U}]$, let $\bar{L}$ denote its reduction modulo p. By definition, $\boldsymbol{u}$ is bad modulo p precisely when $\overline{H}_{i,\boldsymbol{u}}$ (for $0 \le i \le n+1$) have a common nontrivial root modulo p, that is if and only if $p \mid \operatorname{Res}(\overline{H}_{0,\boldsymbol{u}}, \dots, \overline{H}_{n+1,\boldsymbol{u}})$. By [15, Section IV], as a polynomial in $\boldsymbol{U}$,

$$\operatorname{Res}(\overline{H}_{0,\boldsymbol{U}}, \dots, \overline{H}_{n+1,\boldsymbol{U}}) = \overline{R}(\boldsymbol{U}),$$

where R is defined as in (5-9). (That is, the resultant of the reductions modulo p is the reduction modulo p of the resultant.) Thus for each $\boldsymbol{u}$ such that $G(0, \boldsymbol{u}) \neq 0$ so that $R(\boldsymbol{u}) \neq 0$, we can conclude that

$$\left|\{p : \boldsymbol{u} \text{ is bad modulo } p\}\right| = \omega(\operatorname{Res}(H_{0,\boldsymbol{u}}, \dots, H_{n+1,\boldsymbol{u}})),$$

where $\omega(r)$ indicates the number of distinct prime divisors of an integer r; we recall in particular that $\omega(r) \ll (\log r)/(\log \log r)$. By [21, Chapter 13, Proposition 1.1], the resultant is a homogeneous polynomial in the coefficients of the forms $H_{0,\boldsymbol{u}}, \dots, H_{n+1,\boldsymbol{u}}$ (with degree bounded in terms of n, m, e, d). Thus, for every value of $\boldsymbol{u}$ such that $G(0, \boldsymbol{u}) \neq 0$ so that $\operatorname{Res}(H_{0,\boldsymbol{u}}, \dots, H_{n+1,\boldsymbol{u}})$ is a nonzero integer,

$$\omega(\operatorname{Res}(H_{0,\boldsymbol{u}}, \dots, H_{n+1,\boldsymbol{u}})) \ll_{n,m,e,d} \log(\|F\| \|\boldsymbol{u}\|). \tag{5-10}$$

Finally, if $G(0, \boldsymbol{u}) = 0$, then the hyperplane $V(\langle \boldsymbol{u}, \boldsymbol{X} \rangle) \subset \mathbb{P}^n_{\mathbb{C}}$ is tangent to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n_{\mathbb{C}}$ so that (5-8) has rank 1 over $\mathbb{C}$; consequently $\boldsymbol{u}$ is bad for all primes p. Thus in this latter case, we will instead focus on showing there are sufficiently few solutions to $G(0, \boldsymbol{u}) = 0$.

Remark 5.1. It is a common occurrence that one requires the fact that there are "quite few" primes of bad reduction for a variety of the form $\mathcal{V} \cap \{u_0 X_0 + \cdots u_n X_n = 0\}$ for some variety $\mathcal{V}$ and parameter $(u_0, u_1, \ldots, u_n)$ of interest, in this case $V(G)$ with G describing the dual of F, and $u_0 = 0$. The fact that our result (5-7) depends only logarithmically on $\|F\|$ is important for our ultimate deduction that the implicit constant in Theorem 1.1 is independent of $\|F\|$; see the application in Section 5.2.1. This motivated the explicit argument we gave above. Alternatively, we thank Per Salberger for pointing out that the useful references [17, pp. 95–98] and [18] also provide similar constructions leading to explicit results of the form (5-10) and hence (5-7). We remark that if we did not require logarithmic dependence on $\|F\|$, one could apply a result such as [13, Proposition 11.5(3), Appendix] to conclude immediately that for all sufficiently large primes (in an inexplicit sense), $\boldsymbol{u}$ is bad modulo p precisely when $p | G(0, \boldsymbol{u})$ (so that $|\{p : \boldsymbol{u} \text{ is bad modulo p}\}| \ll_G \log \|\boldsymbol{u}\|$ when $G(0, \boldsymbol{u}) \neq 0$), but with dependence on G and hence on F that has not been made explicit, and so does not immediately suffice for our application.

5.2. ***Bad-bad case.*** We use the above facts to control the contribution of the bad-bad case to the sieve, which by Proposition 4.2 is bounded by

$$\frac{1}{P^2 Q^{2n}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q}} \sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq) \right| \ll \frac{Q^{n+1}}{P^2 Q^{2n}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q}} \sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) \right|. \tag{5-11}$$

We start by exchanging the order of summation between $\boldsymbol{u}$ and the primes p, q, and then splitting the sum as

$$\sum_{\boldsymbol{u} \in \mathbb{Z}^n} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) \right| = \sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ G(0,\boldsymbol{u}) = 0}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} + \sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ G(0,\boldsymbol{u}) \neq 0}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} .$$

In this section, we will prove that the contribution from $G(0, \boldsymbol{u}) \neq 0$ is

$$\sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ G(0,\boldsymbol{u}) \neq 0}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) \right| \ll_{n,m,e,d} Q^{2n} (\log B)^2. \tag{5-12}$$

On the other hand, we will prove that the contribution from $G(0, \boldsymbol{u}) = 0$ is

$$\sum_{\substack{\boldsymbol{u} \in \mathbb{Z}^n \\ G(0,\boldsymbol{u}) = 0}} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left| \hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) \right| \ll_\varepsilon P^2 \left(Q^{2n} B^{-\alpha(M-1)} + B^n \left(\frac{Q^2}{B^{1-\alpha}} \right)^{n-2+\frac{1}{3}+\varepsilon} \right), \tag{5-13}$$

for a small $0 < \alpha < 1$ of our choice, and any $\varepsilon > 0$. Once we have proved these two inequalities, we will wrap up the contribution of the bad-bad case in Section 5.2.3.

5.2.1. *The case* $G(0,\boldsymbol{u})\neq 0$. Proving (5-12) is quite simple; by the decay (5-3) for $\hat{W}$ and the bound (5-10) for counting p,q,

$$\sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})\neq 0}}\ \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q\\ \boldsymbol{u}\text{ bad mod } p\\ \boldsymbol{u}\text{ bad mod } q}}\left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right)\right| \ll B^n \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})\neq 0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}\omega(R(\boldsymbol{u}))^2$$

$$\ll B^n\sum_{\boldsymbol{u}\in\mathbb{Z}^n}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}(\log(\|F\|\|\boldsymbol{u}\|))^2$$

$$\ll_{n,m,e,d} Q^{2n}(\log B)^2.$$

Here we have used the fact that $Q=B^\kappa$ with $1/2\leq\kappa\leq 1$ (so that $Q^{2n}\gg B^n$), and the fact from Lemma 2.1 that in the only case we need to consider, $\log\|F\|\ll_{m,e,d}\log B$. This proves (5-12) with an implied constant independent of $\|F\|$.

5.2.2. *The case* $G(0,\boldsymbol{u})=0$. Proving (5-13) is a key novel aspect of our proof. Note that if $G(0,\boldsymbol{u})=0$, then $\boldsymbol{u}$ is bad mod p for all $p\in\mathcal{P}$. Then

$$\sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\ \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q\\ \boldsymbol{u}\text{ bad mod } p\\ \boldsymbol{u}\text{ bad mod } q}}\left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right)\right| \ll B^nP^2\sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}. \tag{5-14}$$

Let $0<\alpha<1$ be a parameter to be chosen later and consider the cube

$$\mathcal{C}_\alpha=[-Q^2/B^{1-\alpha},Q^2/B^{1-\alpha}]^n\subset\mathbb{R}^n.$$

This is slightly larger than the "essential support" of the sum over $\boldsymbol{u}$, so that outside this box we can exploit decay more efficiently. We will ultimately prove that

$$\sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}\ll_\varepsilon Q^{2n}B^{-n}B^{-\alpha(M-1)}+\left(\frac{Q^2}{B^{1-\alpha}}\right)^{n-2+1/3+\varepsilon}, \tag{5-15}$$

for any $\varepsilon>0$. We split the sum as

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}+\sum_{\substack{\boldsymbol{u}\notin\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}. \tag{5-16}$$

In the second sum in (5-16), we can exploit decay:

$$\sum_{\substack{\boldsymbol{u}\notin\mathcal{C}_\alpha\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}\ll\sum_{j=1}^n\sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0\\ |u_j|>Q^2/B^{1-\alpha}}}\prod_{i=1}^n\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}\ll\left(\frac{Q^2}{B}\right)^n\frac{1}{B^{\alpha(M-1)}}.$$

The contribution of these $\boldsymbol{u}$ to (5-14) is thus $\ll Q^{2n}P^2B^{-\alpha(M-1)}$ for $0<\alpha<1$ and any $M\geq 2n$; this contributes the first term in (5-13).

It remains to deal with the first sum appearing on the right-hand side of (5-16), summing over $\boldsymbol{u}\in\mathcal{C}_\alpha$ such that $G(0,\boldsymbol{u})=0$. Here we show that there are few solutions to $G(0,\boldsymbol{u})=0$. Recall the definition of the form G from Section 5.1. Consider $V(G(0,\boldsymbol{U}))\subset\mathbb{P}^{n-1}$ defined by $G(0,\boldsymbol{U})=0$ as a function of $\boldsymbol{U}$. (First notice that $G(0,\boldsymbol{U})$ is not identically zero; indeed, if it were then we would conclude that $\{U_Y=0\}\subset\{G(U_Y,U_1,\dots,U_n)=0\}$. Recalling that $G(U_Y,\boldsymbol{U})$ is irreducible, both these projective varieties have dimension $n-1$ so that in fact we must have $\{G=0\}=\{U_Y=0\}$. But this is impossible, since G has degree >1.) Thus $V(G(0,\boldsymbol{U}))\subset\mathbb{P}^{n-1}_{\mathbb{C}}$ is a projective variety of dimension $n-2$ and $\deg G(0,\boldsymbol{U})=\deg G(U_Y,\boldsymbol{U})\geq 2$. Moreover, let us decompose $G(0,\boldsymbol{U})$ into irreducible components, i.e., by writing

$$G(0,\boldsymbol{U})=\prod_{\ell=1}^{L}G_\ell(\boldsymbol{U}),\tag{5-17}$$

where $G_\ell(\boldsymbol{U})$ is an irreducible polynomial for each $\ell\leq L$ (and $L\ll_{n,m,e,d}1$). Set $d_\ell:=\deg G_\ell$. We have

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}\prod_{i=1}^{n}\left(1+\frac{B|u_i|}{Q^2}\right)^{-M}\leq\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G(0,\boldsymbol{u})=0}}1\leq\sum_{\ell=1}^{L}\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G_\ell(\boldsymbol{u})=0}}1.$$

In the next section, we shall prove:

Proposition 5.2. *Let $n\geq 3$. For the homogeneous polynomial $G(U_Y,U_1,\dots,U_n)\in\mathbb{C}[U_Y,U_1,\dots,U_n]$ defined in* (5-6), *$G(0,U_1,\dots,U_n)$ contains no linear factor, that is, we cannot write $G(0,\boldsymbol{U})=L(\boldsymbol{U})\tilde{H}(\boldsymbol{U})$ for any linear form $L(\boldsymbol{U})\in\mathbb{C}[U_1,\dots,U_n]$.*

Remark 5.3. As a consequence of Proposition 5.2, $G(0,U_1,\dots,U_n)$ contains no factor in one or two variables. For suppose that in the notation of (5-17) some factor $G_\ell(\boldsymbol{U})$ (after an appropriate $GL_n(\mathbb{C})$ change of variables) can be written as a polynomial $g_1(U_1)$ or $g_2(U_1,U_2)$. Then $g_1(U_1)$ is a monomial, hence a product of linear factors, contradicting the proposition. Alternatively, any form $g_2(U_1,U_2)$ factors over $\mathbb{C}$ into homogeneous linear factors in U_1,U_2, as a consequence of the fundamental theorem of algebra applied to $g_2(1,t)\in\mathbb{C}[t]$, followed by noting $g_2(U_1,U_2)=U_1^{\deg g_2}g_2(1,U_2/U_1)$. This again would contradict the proposition. (Since the statement of Proposition 5.2 is false if $n=2$, see Remark 5.4 for an alternative approach for $n=2$.)

The crucial point is that Proposition 5.2 implies that for each $\ell=1,\dots,L$ the degree $d_\ell\geq 2$ (and G_ℓ depends on at least 3 variables). By [27, Theorem 2] and [41, Theorem A], we have, for any $\varepsilon>0$,

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n\\ G_\ell(\boldsymbol{u})=0}}1\ll_\varepsilon\begin{cases}(Q^2/B^{1-\alpha})^{n-2+\varepsilon} & \text{if } d_\ell=2,\\ (Q^2/B^{1-\alpha})^{n-2+\frac{1}{d_\ell}+\varepsilon} & \text{if } d_\ell>2.\end{cases}\tag{5-18}$$

Within these results, the implied constant is independent of $\|F\|$ in each case. In particular, we may conclude that for each $\ell = 1, \dots, L$,

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{C}_\alpha\cap\mathbb{Z}^n \\ G_\ell(0,\boldsymbol{u})=0}} 1 \ll_\varepsilon \left(\frac{Q^2}{B^{1-\alpha}}\right)^{n-2+\frac{1}{3}+\varepsilon}.$$

Thus the total contribution of these terms to (5-14) is

$$\ll_\varepsilon B^n P^2 \left(\frac{Q^2}{B^{1-\alpha}}\right)^{n-2+\frac{1}{3}+\varepsilon}.$$

This contributes the second term in (5-13), and hence (5-13) is proved.

5.2.3. *Conclusion of the bad-bad sieve term.* From (5-12) and (5-13) we conclude that the total contribution of the bad-bad case (5-11) to the sieve is

$$\frac{Q^{n+1}}{P^2 Q^{2n}}\left(Q^{2n}(\log B)^2 + Q^{2n}P^2B^{-\alpha(M-1)} + B^nP^2\left(\frac{Q^2}{B^{1-\alpha}}\right)^{n-2+\frac{1}{3}+\varepsilon}\right)$$
$$\ll_{\varepsilon'} Q^n\left(QP^{-2}(\log B)^2 + QB^{-\alpha(M-1)} + \left(\frac{B^{\frac{5}{3}+g(\alpha)+\varepsilon'}}{Q^{\frac{7}{3}+\varepsilon'}}\right)\right), \quad (5\text{-}19)$$

where $g(\alpha) = \alpha\big(n - \frac{5}{3} + \varepsilon'\big)$, for any $\varepsilon' > 0$. To simplify the third term above, henceforward we assume $Q = B^\kappa$ with

$$\tfrac{3}{4} \le \kappa \le 1. \quad (5\text{-}20)$$

Then the above is

$$\ll_{\varepsilon'} Q^n(QP^{-2}(\log B)^2 + QB^{-\alpha(M-1)} + B^{-\frac{1}{12}+g(\alpha)+\varepsilon'}), \quad (5\text{-}21)$$

for any $\varepsilon' > 0$. In the first term on the right-hand side, we observe by (4-5) that $P \gg Q/\log Q$ so that

$$QP^{-2}(\log B)^2 \ll Q^{-1}(\log B)^4 \ll B^{-3/4}(\log B)^4.$$

In the second term, we can choose $\alpha = \frac{1}{24}(n - \frac{5}{3} + \varepsilon')^{-1}$ so $g(\alpha) = \frac{1}{24}$, and set $M \ge \max\{2n, \alpha^{-1} + 1\}$. Regarding the third term, so far this is true for any $\varepsilon' > 0$; let us take $\varepsilon' = 1/100$, say. We conclude that

$$\frac{Q^{n+1}}{P^2Q^{2n}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n \\ G(0,\boldsymbol{u})=0}} \sum_{\substack{p,q\in\mathcal{P} \\ p\neq q \\ \boldsymbol{u} \text{ bad mod } p \\ \boldsymbol{u} \text{ bad mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right)\right| \ll Q^n\big(B^{-3/4}(\log B)^4 + QB^{-1} + B^{-\frac{1}{24}+\frac{1}{100}}\big) \ll Q^n, \quad (5\text{-}22)$$

since $B \ge Q$. The implied constant is independent of $\|F\|$. (Here we could even obtain a term that is $o(Q^n)$, but this will not change our main theorem, since the good-good contribution to the sieve is $O(Q^n)$.) This completes the treatment of the bad-bad contribution to the sieve, except for the proof of Proposition 5.2, which we provide in the next section. Then in Section 7 we show that the contributions of all the other types to the sieve are also dominated by $\ll Q^n$, and then conclude the proof of our main theorem.

Remark 5.4 (the case $n = 2$). The method of this paper applies for $n = 2$ up until Proposition 5.2; arguing as in Remark 5.3 shows that $G(0, U_1, U_2)$ factors over $\mathbb{C}$ into homogeneous linear factors in U_1, U_2, so that proposition is false for $n = 2$. Thus in the nomenclature of (5-17), each degree $d_\ell = 1$, and the estimate (5-18) is replaced by $(Q^2/B^{1-\alpha})^{n-1}$. Thus (5-19) is replaced by

$$Q^n\big(QP^{-2}(\log B)^2 + QB^{-\alpha(M-1)} + B^{(n-1)\alpha+1}Q^{-1}\big) \ll Q^{n+1},$$

upon taking $\alpha = 0$ and using $Q \gg B^{1/2}$. Ultimately, arguing in this way for $n = 2$ leads to the choice $Q = B^{1/2}(\log B)^{1/2}$ and the outcome $S(F, B) \ll B^{n-1+1/2}(\log B)^{1/2}$, which is essentially no better than (1-16), aside from the fact that we can remove the dependence on $\|F\|$ in the implicit constant. In any case, Broberg's results (1-14) and (1-15) supersede the outcome of the methods of this paper for $n = 2, 3$.

6. Proof of Proposition 5.2

In this section we prove the critical Proposition 5.2 that allows us to deduce all factors in $G(0, \boldsymbol{U})$ have at least degree 2, so that we can apply the nontrivial bounds of Heath-Brown and Pila in (5-18). We thank Per Salberger for suggesting the following strategy to prove the proposition.

Let $n \geq 3$. Suppose to the contrary that $G(0, \boldsymbol{U})$ contains a linear factor, that is,

$$G(0, \boldsymbol{U}) = L(\boldsymbol{U})\tilde{H}(\boldsymbol{U}) \tag{6-1}$$

for some linear form L. Then by a linear change of variables we can reduce to the case in which we may assume that $L(\boldsymbol{U}) = U_1$, and conclude that

$$G(0, \boldsymbol{U}) = U_1 H(\boldsymbol{U})$$

for some homogeneous polynomial H. Then *any* point $(0, 0, u_2, \dots, u_n) \in \{U_Y = U_1 = 0\} \subset \mathbb{P}^n$ satisfies $G(0, \boldsymbol{U}) = 0$ and thus defines a tangent hyperplane to $V(F(Z^e, \boldsymbol{X})) \subset \mathbb{P}^n$, given by

$$u_2X_2 + \dots + u_nX_n = 0.$$

In particular, for all $[u_2 : \dots : u_n] \in \mathbb{P}^{n-2}$, this hyperplane contains the line ℓ given by $X_2 = \dots = X_n = 0$ in $\mathbb{P}^n$. We note that this line ℓ is not contained in $V(F(Z^e, \boldsymbol{X}))$, since for example in the coordinates $[U_Y : U_1 : U_2 : \dots : U_n]$ we see that the point $[1 : 0 : 0 : \dots : 0] \in \ell$ but $[1 : 0 : 0 : \dots : 0] \notin V$, since in the definition of F the coefficient of Z^{mde} is 1. Thus under the assumption (6-1) we have shown that the generic hyperplane through ℓ is tangent to $V(F(Z^e, \boldsymbol{X}))$. We will see this is impossible, and our assumption (6-1) is false (so that Proposition 5.2 is verified), by the following proposition.

Proposition 6.1. *Let $n \geq 3$. Let $X \subset \mathbb{P}^n$ be a nonsingular hypersurface and let ℓ be a line not contained in X. Then the generic hyperplane in $\mathbb{P}^n$ containing ℓ is not tangent to X.*

Let X be given as in the proposition. Without loss of generality we can make a change of coordinates so that

$$\ell = \{X_2 = \cdots = X_n = 0\}.$$

Let $F \in \mathbb{C}[X_0, X_1, \ldots, X_n]$ be such that $X = \{F = 0\}$, and let D denote the degree of F. Our strategy is to construct the blow-up of X along the zero-dimensional subvariety $Z \subset X$, where we define

$$Z = \ell \cap X \subset \mathbb{P}^n.$$

Under the hypothesis that ℓ is not contained in X, then $\deg Z \leq D$. We also define the open set

$$U := X \setminus Z.$$

To prove the proposition, we first notice that we can parametrize the hyperplanes containing ℓ in $\mathbb{P}^n$ by points in $\mathbb{P}^{n-2}$ using the map

$$\mathbb{P}^{n-2} \to \{H \subset \mathbb{P}^n : \deg H = 1,\ \ell \subset H\}, \quad [v_2 : \cdots : v_n] \mapsto \{v_2 X_2 + \cdots + v_n X_n = 0\}.$$

Thus, it will suffice to show that there exists an open set $V \subset \mathbb{P}^{n-2}$ such that for all $\boldsymbol{v} = [v_2 : \cdots : v_n] \in V$,

$$X \cap \{v_2 X_2 + \cdots + v_n X_n = 0\}$$

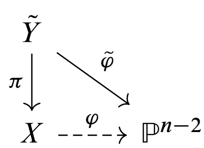
is smooth, so that in particular the hyperplane $\{v_2 X_2 + \cdots + v_n X_n = 0\} \subset \mathbb{P}^n$ is not tangent to X. We will prove this in two steps, first focusing on the intersection of the hyperplane with the open set $U = X \setminus Z$, and then focusing on the intersection of the hyperplane with the finite set of points in Z. In agreement with the citations we apply in what follows, from now on we will use the terminology "regular" for a scheme instead of "smooth." For a nonsingular hypersurface such as X, these notions are identical by the Jacobian criterion [36, Chapter 4, Theorem 2.19 and Example 2.10]; more generally, the notions are equivalent for any algebraic variety over a perfect field, and in particular over $\mathbb{C}$ [36, Chapter 4, Corollary 3.33].

Define a rational map $\varphi : X \dashrightarrow \mathbb{P}^{n-2}$ given by

$$\varphi : [X_0 : X_1 : X_2 : \cdots : X_n] \mapsto [X_2 : \cdots : X_n].$$

This is a regular map on U. We claim that there exists a projective variety $\tilde{Y}$ and two morphisms $\pi : \tilde{Y} \to X$, and $\tilde{\varphi} : \tilde{Y} \to \mathbb{P}^{n-2}$ such that:

(i) The diagram



is commutative.

(ii) The morphism π restricts to an isomorphism $\pi : \pi^{-1}(U) \to U$.

(iii) The projective variety $\tilde{Y}$ is regular.

Let us assume this claim for now and see how to conclude the proof of the proposition. Since $\tilde{Y}$ is regular, we can apply Kleiman's Bertini theorem [23, Chapter III, Corollary 10.9] to the morphism $\tilde{\varphi} : \tilde{Y} \to \mathbb{P}^{n-2}$, and deduce that given a generic hyperplane $H \subset \mathbb{P}^{n-2}$, $\tilde{\varphi}^{-1}(H) \subseteq \tilde{Y}$ is regular. Let us fix one of these generic hyperplanes, and call it

$$H = \{u_2 X_2 + \cdots + u_n X_n = 0\} \subset \mathbb{P}^{n-2}.$$

By the choice of H, $\tilde{\varphi}^{-1}(H) \cap \pi^{-1}(U)$ is nonsingular. Recall that π is an isomorphism when restricted to the open set $\pi^{-1}(U)$. Thus we also learn that

$$\begin{aligned}\pi(\tilde{\varphi}^{-1}(H) \cap \pi^{-1}(U)) = \pi(\tilde{\varphi}^{-1}(H)) \cap U = \varphi^{-1}(H) \cap U \\ = \{[x_0 : x_1 : x_2 : \cdots : x_n] \in U : u_2 x_2 + \cdots + u_n x_n = 0\}\end{aligned}$$

is regular. Since such H are generic in $\mathbb{P}^{n-2}$, we conclude that there is an open set $V_1 \subset \mathbb{P}^{n-2}$ such that for all $\boldsymbol{v} = [v_2 : \cdots : v_n] \in V_1$, the intersection

$$U \cap \{v_2 X_2 + \cdots + v_n X_n = 0\}$$

is regular.

Let us next focus on the intersection of the hyperplane with the set Z. For any $P \in Z$, a hyperplane $\{v_2 X_2 + \cdots + v_n X_n = 0\}$ with $[v_2 : \cdots : v_n] \in \mathbb{P}^{n-2}$ is tangent to X at P if the Jacobian matrix at P,

$$J_{\boldsymbol{v}}(P) = \begin{pmatrix} \frac{\partial F}{\partial X_0}(P) & 0 \\ \frac{\partial F}{\partial X_1}(P) & 0 \\ \frac{\partial F}{\partial X_2}(P) & v_2 \\ \vdots & \vdots \\ \frac{\partial F}{\partial X_n}(P) & v_n \end{pmatrix},$$

has rank ≤ 1. From this it is clear that if either $\frac{\partial F}{\partial X_0}(P) \neq 0$ or $\frac{\partial F}{\partial X_1}(P) \neq 0$ then rank $J_{\boldsymbol{v}}(P) = 2$ for any $\boldsymbol{v} \in \mathbb{P}^{n-2}$. On the other hand, if $\frac{\partial F}{\partial X_0}(P) = \frac{\partial F}{\partial X_1}(P) = 0$ then $\mathrm{rank}_{\boldsymbol{v}}(P) \leq 1$ if and only if $\boldsymbol{v} = \big[\frac{\partial F}{\partial X_2}(P) : \cdots : \frac{\partial F}{\partial X_n}(P)\big]$ since we are assuming that X is a nonsingular hypersurface. For each $P \in Z$ we define

$$C_P = \begin{cases} \{\big[\frac{\partial F}{\partial X_2}(P) : \cdots : \frac{\partial F}{\partial X_n}(P)\big]\} & \text{if } \frac{\partial F}{\partial X_0}(P) = \frac{\partial F}{\partial X_1}(P) = 0, \\ \varnothing & \text{otherwise.} \end{cases}$$

If we define $V_P = \mathbb{P}^{n-2} \setminus C_P$, it follows that for any $\boldsymbol{v} \in V_P$ the intersection

$$X \cap \{v_2 X_2 + \cdots + v_n X_n = 0\}$$

is regular at P.

Finally consider the set

$$V = V_1 \cap \bigcap_{P \in Z} V_P.$$

Since $\deg Z \le D$, then V is a nonempty open subset of $\mathbb{P}^{n-2}$. For each $\boldsymbol{v} \in V$, the hyperplane $v_2x_2 + \cdots + v_nx_n = 0$ contains ℓ, and

$$\{v_2X_2 + \cdots + v_nX_n = 0\} \cap (U \cup Z) = \{v_2X_2 + \cdots + v_nX_n = 0\} \cap X$$

is regular, or equivalently, nonsingular; thus $\{v_2X_2 + \cdots + v_nX_n = 0\}$ is not tangent to X. This completes the proof of Proposition 6.1, except for the proof of properties (i), (ii), and (iii) in the claim.

We now prove the claim of properties (i), (ii) and (iii). From the rational map $\varphi : X \dashrightarrow \mathbb{P}^{n-2}$ given by

$$\varphi : [X_0 : X_1 : X_2 : \cdots : X_n] \mapsto [X_2 : \cdots : X_n],$$

we consider the graph $\Gamma = \Gamma_\varphi$ of the map φ,

$$\Gamma = \{(\boldsymbol{x}, \varphi(\boldsymbol{x})) : \boldsymbol{x} \in U\} \subset X \times \mathbb{P}^{n-2}.$$

Define the Zariski closure $\tilde{X} = \overline{\Gamma} \subset X \times \mathbb{P}^{n-2}$. Define the projection map $\pi' : \tilde{X} \to X$ acting by $(\boldsymbol{x}, \varphi(\boldsymbol{x})) \to (\boldsymbol{x})$. Then the blow-up is $\tilde{X}$ along with a morphism φ' such that

$$\begin{array}{ccc} \tilde{X} & & \\ {\scriptstyle \pi'}\downarrow & \searrow^{\varphi'} & \\ X & \overset{\varphi}{\dashrightarrow} & \mathbb{P}^{n-2} \end{array}$$

is a commutative diagram (see, e.g., [22, Chapter 7, p. 82]). Moreover, from the definition of the blow-up it follows that π' restricts to an isomorphism $\pi' : (\pi')^{-1}(U) \to U$, i.e., $\tilde{X}$ satisfies properties (i) and (ii), but it might be singular. To resolve this, we apply Hironaka's resolution of singularities: as a consequence of [30, Theorem 1] (see also [30, p. 112]), there is a projective variety $\tilde{Y}$ and a morphism $f : \tilde{Y} \to \tilde{X}$ such that f is an isomorphism when restricted to the inverse image $f^{-1}(V)$ of the open set V of the regular points of $\tilde{X}$, and such that $\tilde{Y}$ is regular. Then the claim follows by taking $\pi = \pi' \circ f$, $\tilde{\varphi} = \varphi' \circ f$ and observing that $(\pi')^{-1}(U) \subset V$.

7. Concluding arguments

In Section 5 we proved that the contribution of the bad-bad terms to the sieve is $\ll Q^n$. We now turn to analyzing the contributions of the other types, as defined in Definition 4.1. We will treat these in three sections; in each case we apply the relevant bound for $|g(\boldsymbol{u}, pq)|$ from Proposition 4.2 and the bound (5-4) for $\hat{W}$. Once we have treated these cases, we proceed in Section 7.4 to choose the parameter Q, and conclude the proof of Theorem 1.1.

7.1. ***Zero-type cases.*** We first consider any case in which $\boldsymbol{u}$ is zero-type modulo p, divided into cases according to whether $\boldsymbol{u}$ is zero-type, good, or bad modulo q. The contribution of the first case (upon setting $\boldsymbol{u} = pq\boldsymbol{v}$ and applying (5-4)) is

$$\frac{1}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ zero mod } p\\ \boldsymbol{u}\text{ zero mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \ll \frac{Q^{2n-1}}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\boldsymbol{v}\in\mathbb{Z}^n} \left|\hat{W}(\boldsymbol{v})\right| \ll B^n Q^{-1}.$$

The contribution of the second case (upon setting $\boldsymbol{u} = p\boldsymbol{v}$, applying (5-4) with $L = Q < B$) is

$$\frac{1}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ zero mod } p\\ \boldsymbol{u}\text{ good mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \ll \frac{Q^{n-1/2}Q^{n/2}P^2}{P^2Q^{2n}} \sum_{\boldsymbol{v}\in\mathbb{Z}^n} \left|\hat{W}\left(\frac{\boldsymbol{v}}{Q}\right)\right| \ll B^n Q^{-n/2-1/2}.$$

The contribution of the third case (upon setting $\boldsymbol{u} = p\boldsymbol{v}$, applying (5-4) with $L = Q < B$) is

$$\frac{1}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ zero mod } p\\ \boldsymbol{u}\text{ bad mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \ll \frac{Q^{n-1/2}Q^{n/2+1/2}P^2}{P^2Q^{2n}} \sum_{\boldsymbol{v}\in\mathbb{Z}^n} \left|\hat{W}\left(\frac{\boldsymbol{v}}{Q}\right)\right| \ll B^n Q^{-n/2}.$$

As long as $n \geq 2$, all these cases contribute at most $\ll B^n Q^{-1}$ to the sieve, which is acceptable.

7.2. ***Good-good case.*** The contribution to the sieve from the good-good case is:

$$\frac{1}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ good mod } p\\ \boldsymbol{u}\text{ good mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \ll \frac{Q^nP^2}{P^2Q^{2n}} \sum_{\boldsymbol{u}\in\mathbb{Z}^n} \left|\hat{W}\left(\frac{\boldsymbol{u}}{Q^2}\right)\right| \ll Q^n,$$

after applying (5-4) with $L = Q^2 > B$, since under the assumption (5-20), $\kappa \geq 1/2$.

7.3. ***Good-bad case.*** The contribution to the sieve from the good-bad case is

$$\frac{1}{P^2Q^{2n}} \sum_{\substack{p,q\in\mathcal{P}\\ p\neq q}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ good mod } p\\ \boldsymbol{u}\text{ bad mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right) g(\boldsymbol{u}, pq)\right| \ll \frac{Q^{n+1/2}}{P^2Q^{2n}} \sum_{p\in\mathcal{P}} \sum_{q\neq p\in\mathcal{P}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ \boldsymbol{u}\text{ bad mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right)\right|. \tag{7-1}$$

Here we proceed by imitating the key step from Section 5 for the bad-bad case, and sum over q before summing over $\boldsymbol{u}$. We again define $G(U_Y, \boldsymbol{U})$ as in (5-6), and let $R(\boldsymbol{u})$ denote the resultant (5-9), so that

$$\sum_{p\in\mathcal{P}} \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})\neq 0}} \sum_{\substack{q\neq p\in\mathcal{P}\\ \boldsymbol{u}\text{ bad mod } q}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{pq}\right)\right| \ll P \sum_{\substack{\boldsymbol{u}\in\mathbb{Z}^n\\ G(0,\boldsymbol{u})\neq 0}} \left|\hat{W}\left(\frac{\boldsymbol{u}}{Q^2}\right)\right| \omega(R(\boldsymbol{u})) \ll_{n,m,e,d} PQ^{2n}\log B,$$

with an implied constant independent of $\|F\|$ (in the first case of Lemma 2.1), by arguing as in the proof of (5-12).

Notice that in the good-bad case, we do not need to consider a possible contribution from those $\boldsymbol{u}$ for which $G(0, \boldsymbol{u}) = 0$: when $G(0, \boldsymbol{u}) = 0$, then all q have the property that $\boldsymbol{u}$ is bad for q, whereas by definition in the good-bad case, $\boldsymbol{u}$ is good for at least one prime. In total, the contribution to the sieve from the good-bad case is thus

$$\frac{Q^{n+1/2}}{P^2 Q^{2n}} \cdot P Q^{2n} (\log B) \ll Q^{n+1/2} P^{-1} (\log B) \ll Q^n,$$

since $Q = B^\kappa$ for some $1/2 \le \kappa \le 1$ and under our acting assumption (4-4), by (4-5), $P \gg Q/\log Q$. Thus we can conclude that the total contribution of the good-bad case (7-1) of the sieve is $\ll Q^n$, with an implied constant independent of $\|F\|$ (in the first case of Lemma 2.1).

7.4. ***Final conclusion of the sieve and choice of parameters.*** We now assemble all the terms of the main sieve term in (5-2): we can conclude that

$$\frac{1}{P^2} \sum_{\substack{p,q \in \mathcal{P} \\ p \neq q}} |T(p,q)| \ll B^n Q^{-1} + Q^n. \tag{7-2}$$

The first term is from all zero-type cases, and the last term includes the good-good, good-bad, and bad-bad cases. We apply this in the sieve lemma, along with the bound (5-1) for the two simple terms in the sieve, to conclude that (in the first case of Lemma 2.1) our counting function admits the bound

$$\mathcal{S}(F, B) \ll_{n,m,e,d} (B^{n-1} + B^n P^{-1} + B^n Q^{-1} + Q^n) \ll (B^n P^{-1} + Q^n). \tag{7-3}$$

Choose

$$Q = B^{n/(n+1)} (\log B)^{1/(n+1)}. \tag{7-4}$$

The requirement (5-20) is met for all $n \ge 3$. (If $n = 2$, then this argument leads to the choice $Q \approx B^{2/3}$, which does not suffice to prove sufficient decay in the bad-bad case; see Remark 5.4.) Recall from (4-4) and (4-5) that

$$P = |\mathcal{P}| \gg_{m,e,d} Q (\log Q)^{-1} \gg_{n,m,e,d} B^{\frac{n}{n+1}} (\log B)^{-\frac{n}{n+1}}$$

as long as

$$Q \gg_{m,e,d} (\log \|F\|)(\log \log \|F\|). \tag{7-5}$$

Recall also that we require $P \gg_{m,e,d} \max\{\log \|f_d\|, \log B\}$ in Lemma 1.2. Certainly the first condition is satisfied under the assumption (7-5). The second condition is satisfied for Q as in (7-4) for all $B \gg_n 1$.

To meet the requirement (7-5) for Q as chosen in (7-4), it suffices to require that

$$B \gg_{m,e,d} (\log \|F\| \log \log \|F\|)^{\frac{n+1}{n}}.$$

For such B, the conclusion of the sieve process in (7-3) shows that

$$\mathcal{S}(F, B) \ll_{n,m,e,d} B^{n-1+\frac{1}{n+1}} (\log B)^{\frac{n}{n+1}},$$

where the implicit constant is independent of $\|F\|$. This suffices for Theorem 1.1. Finally, for all $B \ll_{m,e,d} (\log\|F\| \log\log\|F\|)^{\frac{n+1}{n}}$, we apply the trivial bound

$$\begin{aligned}\mathcal{S}(F,B) \ll_n B^n \ll_{n,m,e,d} (\log\|F\| \log\log\|F\|)^{n+1} \ll (\log\|F\|)^{n+2} \\ \ll_{n,m,d,e} (\log B)^{n+2} \ll_n B^{n-1+\frac{1}{n+1}} (\log B)^{\frac{n}{n+1}}.\end{aligned}$$

Here we applied the fact from Lemma 2.1 that in the case it remains to prove Theorem 1.1, $\|F\| \ll B^{(mde)^{n+2}}$ so that $\log\|F\| \ll_{n,m,d,e} \log B$. This completes the proof of Theorem 1.1.

Acknowledgements

The authors thank T. Browning for suggesting the application of the polynomial sieve to smooth coverings and for useful discussions, and J. Lyczak for many helpful remarks. In addition, the authors thank P. Salberger for suggesting a strategy to prove Proposition 5.2, and both Salberger and an anonymous referee for helpful remarks on an earlier version of this manuscript. The authors credit ChatGPT with expository edits to the last two paragraphs of Section 1.2.4.

Bonolis has been supported by FWF grant P 32428-N35. Pierce has been partially supported by NSF DMS-2200470 and NSF CAREER grant DMS-1652173, a Sloan Research Fellowship, and a Joan and Joseph Birman Fellowship. The authors thank the Hausdorff Center for Mathematics for hosting a productive collaboration visit and the RTG DMS-2231514. Pierce thanks HCM for hosting a visit as a Bonn Research Chair.

References

[1] E. Bombieri and J. Pila, "The number of integral points on arcs and ovals", *Duke Math. J.* **59**:2 (1989), 337–357. MR Zbl

[2] D. Bonolis, "A polynomial sieve and sums of Deligne type", *Int. Math. Res. Not.* **2021**:2 (2021), 1096–1137. MR Zbl

[3] D. Bonolis and T. Browning, "Uniform bounds for rational points on hyperelliptic fibrations", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **24**:1 (2023), 173–204. Correction in **24**:4 (2023), 2501–2504. MR Zbl

[4] J. Brandes, "Sums and differences of power-free numbers", *Acta Arith.* **169**:2 (2015), 169–180. MR Zbl

[5] N. Broberg, "Rational points on finite covers of $\mathbb{P}^1$ and $\mathbb{P}^2$", *J. Number Theory* **101**:1 (2003), 195–207. MR Zbl

[6] T. D. Browning, "A note on the distribution of rational points in threefolds", *Quart. J. Math.* **54** (2003), 33–39.

[7] T. D. Browning, *Quantitative arithmetic of projective varieties*, Progr. Math. **277**, Birkhäuser, Basel, 2009. MR Zbl

[8] T. D. Browning, "The polynomial sieve and equal sums of like polynomials", *Int. Math. Res. Not.* **2015**:7 (2015), 1987–2019. MR Zbl

[9] T. D. Browning and D. R. Heath-Brown, "The density of rational points on non-singular hypersurfaces, I", *Bull. Lond. Math. Soc.* **38**:3 (2006), 401–410. MR Zbl

[10] T. D. Browning and D. R. Heath-Brown, "The density of rational points on non-singular hypersurfaces, II", *Proc. Lond. Math. Soc.* (3) **93**:2 (2006), 273–303. MR Zbl

[11] T. D. Browning and P. Vishe, "Rational points on cubic hypersurfaces over $\mathbb{F}_q(t)$", *Geom. Funct. Anal.* **25**:3 (2015), 671–732. MR Zbl

[12] T. D. Browning, D. R. Heath-Brown, and P. Salberger, "Counting rational points on algebraic varieties", *Duke Math. J.* **132**:3 (2006), 545–578. MR Zbl

[13] A. Bucur, A. C. Cojocaru, M. N. Lalín, and L. B. Pierce, "Geometric generalizations of the square sieve, with an application to cyclic covers", *Mathematika* **69**:1 (2023), 106–154. MR Zbl

[14] W. Castryck, R. Cluckers, P. Dittmann, and K. H. Nguyen, "The dimension growth conjecture, polynomial in the degree and without logarithmic factors", *Algebra Number Theory* **14**:8 (2020), 2261–2294. MR Zbl

[15] M. Chardin, "The resultant via a Koszul complex", pp. 29–39 in *Computational algebraic geometry* (Nice, Italy, 1992), edited by F. Eyssette and A. Galligo, Progr. Math. **109**, Birkhäuser, Boston, MA, 1993. MR Zbl

[16] S. D. Cohen, "The distribution of Galois groups and Hilbert's irreducibility theorem", *Proc. Lond. Math. Soc.* (3) **43**:2 (1981), 227–250. MR Zbl

[17] D. A. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, 2nd ed., Grad. Texts in Math. **185**, Springer, 2005. MR Zbl

[18] M. Demazure, "Résultant, discriminant", *Enseign. Math.* (2) **58**:3-4 (2012), 333–373. MR Zbl

[19] D. Eisenbud and J. Harris, 3264 *and all that: a second course in algebraic geometry*, Cambridge Univ. Press, 2016. MR Zbl

[20] W. Fulton and R. Lazarsfeld, "Connectivity and its applications in algebraic geometry", pp. 26–92 in *Algebraic geometry* (Chicago, IL, 1980), edited by A. Libgober and P. Wagreich, Lecture Notes in Math. **862**, Springer, 1981. MR Zbl

[21] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, Boston, MA, 1994. MR Zbl

[22] J. Harris, *Algebraic geometry: a first course*, Grad. Texts in Math. **133**, Springer, 1992. MR Zbl

[23] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. **52**, Springer, 1977. MR Zbl

[24] D. R. Heath-Brown, "Cubic forms in ten variables", *Proc. Lond. Math. Soc.* (3) **47**:2 (1983), 225–257. MR Zbl

[25] D. R. Heath-Brown, "The square sieve and consecutive square-free numbers", *Math. Ann.* **266**:3 (1984), 251–259. MR Zbl

[26] D. R. Heath-Brown, "The density of rational points on nonsingular hypersurfaces", *Proc. Indian Acad. Sci. Math. Sci.* **104**:1 (1994), 13–29. MR Zbl

[27] D. R. Heath-Brown, "The density of rational points on curves and surfaces", *Ann. of Math.* (2) **155**:2 (2002), 553–595. MR Zbl

[28] D. R. Heath-Brown, "Imaginary quadratic fields with class group exponent 5", *Forum Math.* **20**:2 (2008), 275–283. MR Zbl

[29] D. R. Heath-Brown and L. B. Pierce, "Counting rational points on smooth cyclic covers", *J. Number Theory* **132**:8 (2012), 1741–1757. MR Zbl

[30] H. Hironaka, "Resolution of singularities of an algebraic variety over a field of characteristic zero, I", *Ann. of Math.* (2) **79**:1 (1964), 109–203. MR Zbl

[31] C. Hooley, "On the representations of a number as the sum of four cubes, I", *Proc. Lond. Math. Soc.* (3) **36**:1 (1978), 117–140. MR Zbl

[32] C. Hooley, "On the number of points on a complete intersection over a finite field", *J. Number Theory* **38**:3 (1991), 338–358. MR Zbl

[33] N. M. Katz, "Estimates for 'singular' exponential sums", *Int. Math. Res. Not.* **1999**:16 (1999), 875–899. MR Zbl

[34] J. C. Lagarias and A. M. Odlyzko, "Effective versions of the Chebotarev density theorem", pp. 409–464 in *Algebraic number fields*: L*-functions and Galois properties* (Durham, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR Zbl

[35] S. Lang and A. Weil, "Number of points of varieties in finite fields", *Amer. J. Math.* **76** (1954), 819–827. MR Zbl

[36] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Grad. Texts in Math. **6**, Oxford Univ. Press, 2002. MR Zbl

[37] D. A. Marcus, *Number fields*, Springer, 1977. MR Zbl

[38] J. S. Milne, "Algebraic number theory", preprint, version 3.08, 2020, available at https://www.jmilne.org/math/CourseNotes/ant.html.

[39] R. Munshi, "Density of rational points on cyclic covers of $\mathbb{P}^n$", *J. Théor. Nombres Bordeaux* **21**:2 (2009), 335–341. MR Zbl

[40] L. B. Pierce, "A bound for the 3-part of class numbers of quadratic fields by means of the square sieve", *Forum Math.* **18**:4 (2006), 677–698. MR Zbl

[41] J. Pila, "Density of integral and rational points on varieties", pp. 183–187 in *Columbia University Number Theory Seminar* (New York, 1992), Astérisque **228**, Soc. Math. France, Paris, 1995. MR Zbl

[42] P. Salberger, "On the density of rational and integral points on algebraic varieties", *J. Reine Angew. Math.* **606** (2007), 123–147. MR Zbl

[43] P. Salberger, "Counting rational points on projective varieties", *Proc. Lond. Math. Soc.* (3) **126**:4 (2023), 1092–1133. MR Zbl

[44] J.-P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Études Sci. Publ. Math.* **54** (1981), 323–401. MR Zbl

[45] J.-P. Serre, *Topics in Galois theory*, Res. Notes in Math. **1**, Jones & Bartlett, Boston, MA, 1992. MR Zbl

[46] J.-P. Serre, *Lectures on the Mordell–Weil theorem*, 3rd ed., Vieweg & Sohn, Braunschweig, Germany, 1997. MR Zbl

Communicated by Philippe Michel
Received 2022-09-06 Revised 2023-06-19 Accepted 2023-10-31

dante.bonolis@duke.edu *Mathematics Department, Duke University, Durham, NC 27708, United States*

pierce@math.duke.edu *Mathematics Department, Duke University, Durham, NC 27708, United States*

ALGEBRA AND NUMBER THEORY 18:8 (2024)
https://doi.org/10.2140/ant.2024.18.1557

Functorial embedded resolution via weighted blowings up

Dan Abramovich, Michael Temkin and Jarosław Włodarczyk

We provide a simple procedure for resolving, in characteristic 0, singularities of a variety X embedded in a smooth variety Y by repeatedly blowing up the worst singularities, in the sense of stack-theoretic weighted blowings up. No history, no exceptional divisors, and no logarithmic structures are necessary to carry this out; the steps are explicit geometric operations requiring no choices; and the resulting algorithm is efficient.

A similar result was discovered independently by McQuillan (2020).

1. Introduction

1.1. ***Classical embedded resolution.*** All known methods to canonically (or functorially) resolve singularities of a variety X are embedded: first, one locally embeds X into a smooth *ambient variety* Y, and then gradually improves the transforms (either proper or weak) X_i of X by a series of *basic modifications* $\cdots Y_2 \to Y_1 \to Y_0 = Y$ such that each Y_i is smooth. In fact, the embedded framework was already used by Hironaka [1964a; 1964b], then, based on Hironaka's and Giraud's works, canonical methods were introduced by Bierstone-Milman [1997] and Villamayor [1989], and the full functoriality with respect to smooth morphisms was achieved by Schwartz [1992] and Włodarczyk [2005]. Note that it suffices to construct a functorial resolution étale-locally as globalization follows from the reembedding principle

This research is supported by BSF grants 2014365 and 2018193, ERC Consolidator Grant 770922 - BirNonArchGeom, and NSF grant DMS-1759514.
MSC2020: 14A20, 14E15, 14M25.
Keywords: resolution of singularities, weighted blowings up, algebraic stacks.

(recalled in Section 8.1) and functoriality. We refer to the methods developed in these papers, as well as [Encinas and Hauser 2002; Encinas and Villamayor 2003; Kollár 2007], etc., the *classical methods*.

Basic modifications in the classical methods are blowings up with smooth centers $V_i \subset Y_i$, in particular, this is the way to guarantee that Y_{i+1} is also smooth. Most naturally, one would like to choose V_i to be the worst singularity locus for a natural singularity invariant $\mathrm{inv}_{(Y,X)} : X \to I$ with values in a well-ordered set so that each blowing up improves the invariant. This would lead to a simplest resolution method controlled by a geometrically meaningful singularity measure. But it was common knowledge for decades that this dream is unrealizable; see, for instance, [Kollár 2007, Example 3.6.1] and Section 1.8.

Starting with Hironaka's work, the classical methods use history and the choice of V_i, as well as the invariant $\mathrm{inv}_{(Y_i,X_i)}$, depends on the whole earlier resolution process rather than just on (Y_i, X_i). In particular, only the center $V_0 \subset Y_0$ of the first blowup of the process, sometimes known as the "year-zero center", and invariant $\mathrm{inv}_{(X_0,Y_0)}$, possess a clear geometric meaning.

Remark 1.1.1. (i) The composed sequence $Y_n \to Y$ can be canonically realized as a single blowing up along a highly nonreduced center V, but this is a rather useless presentation, no clear connection between the geometry of V and the singularities of X is known, and it is even unclear how to show that $\mathrm{Bl}_V(Y)$ is smooth if not by a direct computation.

(ii) In classical methods, a basic embedded resolution operates with weak (or principal) transforms, so the intermediate X_i may have new components contained in the exceptional divisor, the center V_i does not have to lie in the i-th proper (or strict) transform X_i^{st} and though $X_{i+1}^{\mathrm{st}} \to X_i^{\mathrm{st}}$ is a blowing up, its center $V_i \cap X_i^{\mathrm{st}}$ may be singular. Using basic resolution and Hilbert–Samuel function one can develop a much more technical method, usually called strong resolution, which operates with proper transforms and hence satisfies $V_i \subset X_i$ and $X_{i+1} = \mathrm{Bl}_{V_i}(X_i)$. It outputs a desingularization blowing up sequence $X_n \to \cdots X_0$ whose centers are smooth.

As noted in Section 1.3, our desingularization Theorem 1.2.2 works directly with proper transforms, and thus achieves strong resolution without the need for further reductions.

1.2. ***Statement of main results.*** In this paper we show that the unrealizable dream becomes possible (probably, even the most natural solution) once one enlarges the pool of basic modifications to the class of weighted blowings up along smooth centers. In fact, just the classical year-zero blowings up with correct weights, which are encoded in the classical year zero invariant, does the job! A similar result was obtained independently by McQuillan [2020].

The stumbling block all these years was the fact that weighted blowings up were not a legitimate tool in embedded resolution because the output ambient variety may be singular. Recently, it was discovered that such blowings up possess a smooth stack-theoretic refinement, and this makes them an absolutely kosher embedded resolution tool at the price of working with Deligne–Mumford stacks instead of varieties. Since the construction is étale-local and the coarse moduli space can be easily resolved (using simple combinatorial tools going under the name "destackification"), this is not a real burden; see Section 1.6 and Theorem 8.1.3.

Definition 1.2.1. By a *DM pair* (X, Y) we mean a quasicompact Deligne–Mumford stack Y smooth *over a field of characteristic zero* and a closed substack $X \subset Y$.

To extend the pool of blowings up we introduce in Section 2 *valuative* $\mathbb{Q}$*-ideals*, providing a convenient formalism to work with Hironaka's idealistic exponents. The basic examples are called *centers*, locally they are of the form $J = (x_1^{a_1}, \dots, x_k^{a_k})$, where $a_i \in \mathbb{Q}_{>0}$ and $x_1, \dots, x_n$ is a regular system of parameters. A center is called *reduced* if $w_i = 1/a_i$ are natural numbers with $\gcd(w_1, \dots, w_k) = 1$. Note that a usual ideal and its normalization give rise to the same valuative ideal, and $J^l = (x_1^{la_1}, \dots, x_k^{la_k})$ as valuative ideals for $l \in \mathbb{N}$. In particular, there is a unique reduced center $\bar{J}$ such that $J = \bar{J}^l$ with $l \in \mathbb{Q}$. In Section 3 we associate to any center J a blowing up $\mathrm{Bl}\,\bar{J}(Y)$, which is a smooth stack-theoretic enhancement of the classical weighted blowing up along $x_1, \dots, x_k$ with weights $w_1, \dots, w_k$. Such blowings up are compatible with smooth morphisms $f : Y' \to Y$, that is, $\mathrm{Bl}_{f^{-1}\bar{J}}(Y') = \mathrm{Bl}\,\bar{J}(Y) \times_Y Y'$. In particular, $\mathrm{Bl}\,\bar{J}(Y) \to (Y)$ is an isomorphism outside of $V(J) := V(x_1, \dots, x_k)$, so the proper transform of closed subschemes is defined as usual.

Theorem 1.2.2 (a step towards resolution). *There is a construction which associates to each DM pair* (X, Y), *with* X *nonempty, a semicontinuous function* $\mathrm{inv}_{(X,Y)} : X \to \Xi_m$ *with values in a well-ordered set* Ξ_m *and a reduced center* $\bar{J} = \bar{J}(X, Y)$ *with the associated blowing up* $F_1(X, Y) : Y_1 \to Y$ *and proper transform* $X_1 \subset Y_1$ *such that the following conditions hold*:

(1) *The vanishing locus*: $V(\bar{J})$ *is precisely the locus where* $\mathrm{inv}_{(X,Y)}$ *attains its maximal value* $\mathrm{maxinv}(X, Y)$.

(2) *The invariant drops*: $\mathrm{maxinv}(X_1, Y_1) < \mathrm{maxinv}(X, Y)$.

(3) *Functoriality*: *for any smooth morphism* $f : Y' \to Y$ *with* $X' = X \times_Y Y'$, *one has that* $\mathrm{inv}_{(X',Y')} = \mathrm{inv}_{(X,Y)} \circ f$. *Furthermore, either* $f^{-1}\bar{J}(X, Y) = (1)$, *or* $\bar{J}(X', Y') = f^{-1}\bar{J}(X, Y)$ *and hence* $(X'_1, Y'_1) = (X_1, Y_1) \times_Y Y_1$.

The set Ξ_m does not depend on X or Y, only on $m := \dim Y$. It is a well-ordered subset $\Xi_m \subset \mathbb{Q}^{\le m}$ of the set of sequences of length at most m, described in the context of Theorem 1.2.5 and in Section 5.1. Moreover, as m varies these sets are nested: $\Xi_m \subset \Xi_{m+1}$ allowing for the necessary comparison in (3).

The index 1 of $F_1(X, Y)$ indicates that it is a one-step operation on the way to a final product; the final product is achieved when X is empty so F_1 does not exist.

Since Ξ_m is well-ordered, composing the one-step partial resolution blowings up $F_1(X_i, Y_i) : Y_{i+1} \to Y_i$ one obtains a sequence $(X_l, Y_l) \to \cdots \to (X_0, Y_0) = (X, Y)$ with $X_l = \varnothing$.

The full weighted embedded resolution is obtained by stopping this process once a center containing an irreducible component of X is chosen, and here an equicodimensionality condition has to be imposed. From the description of the invariant below one sees that the minimal invariant locus is precisely the largest codimension component of the smooth locus of X, hence the theorem immediately implies

Corollary 1.2.3 (weighted resolution). *For a DM pair* (X, Y) *let* $F(X, Y) : (X_n, Y_n) \to \cdots \to (X_0, Y_0) = (X, Y)$ *denote the maximal sequence of blowings up* $F_1(X_i, Y_i)$ *whose centers are nowhere dense in* X_i. *In particular*, $X_n \to X$ *is proper and birational*:

(1) *If X is generically reduced and of constant codimension in Y, then X_n is smooth.*

(2) *If, in addition, $Y' \to Y$ is a smooth morphism and $X' = X \times_Y Y'$, then the sequence $F(X', Y')$ is obtained from $F(X, Y) \times_Y Y'$ by removing all blowings up with empty centers. In particular, $(X'_{n'}, Y'_{n'}) = (X_n, Y_n) \times_Y Y'$.*

Remark 1.2.4 (functorial formulation). One can spell out the results in terms of functors on categories. This is not used in the paper, so we only outline the formulation: F_1 can be viewed as a partial resolution endofunctor on the category of DM pairs with smooth surjective morphisms. Its birational stabilization $F = F_1^{\circ n}$ gives rise to a resolution endofunctor on the category of generically reduced DM pairs of constant codimension with arbitrary smooth morphisms:

- (nonembedded resolution) Using standard arguments, one deduces nonembedded resolution — see Theorem 8.1.1.
- (principalization) Theorem 1.2.2 relies on principalization of ideals on Deligne–Mumford stacks. See Theorem 6.3.1, where strict transforms in Theorem 1.2.2 and Corollary 1.2.3 are replaced by weak transforms.
- (coarse resolution) The reader may wonder about the coarse moduli spaces when Y is a variety. As we note in Section 8.2, the stacks Y_i and X_n have finite abelian stabilizers, hence their coarse moduli spaces $\underline{Y}_i$ and $\underline{X}_n$ have finite abelian quotient singularities. These are eminently resolvable, see Section 1.6 and Theorem 8.1.3. The transformations $\underline{Y}_{i+1} \to \underline{Y}_i$ are best described as the coarse transformations of the weighted blowings up $Y_{i+1} \to Y_i$.

Finally, we provide a very simple and geometric characterization of the invariant $\mathrm{inv}_{(X,Y)}$ and center $\bar{J}(X, Y)$, and we view this as a part of our main results. We will always order local parameters at a point p giving a center $J = (x_1^{a_1}, \ldots, x_k^{a_k})$ so that $a_1 \le a_2 \le \cdots$ and set $\mathrm{inv}_J(p) = (a_1, \ldots, a_k) \in \mathbb{Q}^{\le m} = \bigsqcup_{k=0}^m \mathbb{Q}^k$. We provide the set of invariants with the natural lexicographic order, where shorter sequences are declared to be of larger order.

Theorem 1.2.5. *Let (X, Y) be a variety pair, $I \subset \mathcal{O}_Y$ the ideal of X and $p \in X$ a point:*

(1) *There exists a neighborhood $p \in U$ and a center J on U such that $p \in V(J)$, $I|_U \subseteq J$ and $\mathrm{inv}_J(p)$ is maximal possible among such pairs (U, J). Moreover for all $p' \in V(J)$ we have $\mathrm{inv}_J(p') = \mathrm{inv}_J(p)$ and is locally maximal at p'. In particular, the invariant $\mathrm{inv}_{(X,Y)}(p) = \mathrm{inv}_J(p)$ is well defined and upper semicontinuous.*

(2) *The localization J_p is unique and does not depend on the choice of (U, J).*

(3) *If $\mathrm{inv}_J(p) = (a_1, \ldots, a_k)$, then the numbers $b_1 = a_1$ and $b_l = a_l \prod_{i=1}^{l-1} b_i!$ for $2 \le l \le k$ are integers.*

The theorem is stated for varieties as it is local in nature. The theorem immediately implies that the set Ξ_m of actual invariants is well ordered, and there exists a unique center $J = J(X, Y)$ whose invariant is $\mathrm{maxinv}(X, Y)$, whose vanishing locus is the maximality locus of $\mathrm{inv}_{(X,Y)}$ and such that $V(I_X) \supseteq V(J)$. The center $\bar{J}(X, Y)$ is simply the reduction of J. So, what the algorithm really does — it blows up the unique center J such that $V(I_X) \supseteq V(J)$ and inv_J is maximal possible. Loosely speaking, this is just the center of maximal invariant contained in X.

1.3. *Our quest for the present algorithm.* Several times during our study, we were positively surprised by the properties of the algorithm presented here.

In [Abramovich et al. 2020a] we extended the pool of smooth blowings up in the logarithmic setting, and this allowed to produce algorithms with better efficiency and functoriality properties. That work required to consider stack-theoretic blowings up with nontrivial weights for monomial parameters, so our next project was to study what is the natural resolution algorithm that uses weighted blowings up with arbitrary weights. Our expectation was that the algorithm will be more efficient than the classical ones, but we did not expect at all that it would not require the history of prior operations, a "memoryless" algorithm visibly improving singularities by each weighted blowing up, as it turned out to be. The paradigm that such things do not exist was too strong. We do not know if there exists a simpler algorithm, or a faster one, or a more geometrically informative one, but to the best of our knowledge currently there is not even a conjecture in that direction.

Another surprise is that the algorithm shares common features with the strong resolution methods and uses proper transforms. In particular, the centers (with an appropriate formalism) are contained in X itself, so it can even be interpreted as a nonembedded algorithm and described without using an ambient manifold. In fact, while proving the principalization we show that already the weak transform reduces the invariant on each blowing up, hence the same is true for the proper transform. Since the algorithm is "memoryless", independent of the history of prior operations, this allows to work with proper transforms as well. No need to use Hilbert–Samuel function and much of the usual classical machinery.

Remark 1.3.1. In fact, our method produces a sequence of stack-theoretic modifications $X_n \to \cdots \to X_0 = X$ with a smooth source $\mathcal{F}_{\mathrm{ner}}(X) = X_n$ such that each $f_i : X_{i+1} \to X_i$ satisfies the following two properties:

(i) The method is "memoryless": f_i depends only on X_i.

(ii) The resolution is strong in the extended (stack-theoretic) meaning: each f_i is a stack-theoretic blowing up of a weighted smooth center.

Since we only introduce a narrow class of weighted blowings up — blowings up of *smooth* varieties along *smooth* weighted centers, our interpretation of the second property is the following naive one: (locally) X_i embeds into a smooth stack Y and f_i is the proper transform of a weighted blowing up $g : Y' \to Y$ along a weighted smooth center $\mathcal{J}$ which contains the ideal $\mathcal{I}_{X_i}$ of $\mathcal{O}_Y$. However, Quek and Rydh [2021] define weighted blowings up of arbitrary schemes along arbitrary Rees algebras, not necessarily smooth, and establish their basic properties. In particular, in the formalism of [Quek and Rydh 2021], f_i is indeed the strict transform of g and it is the weighted blowing up of the restriction of $\mathcal{J}$ onto X_i.

Finally, the choice of the center fits and clarifies very well the classical constructions, see Section 1.5. Loosely speaking, we just take the year-zero center with correct weights predicted by the year-zero invariant.

1.4. ***Weighted blowings up, stacks, and resolutions.*** Weighted blowings up in a scheme theoretic sense have been used in birational geometry (as well as many other subjects in mathematics) for a long time. Varchenko used them to characterize the log canonical threshold of a surface; see [Varčenko 1976; Kollár et al. 2004, Theorem 6.40]. Reid [1980; 2002] employs them in the foundation of canonical singularities and in the geometry of surfaces. Kawamata [1992] used them to relate discrepancies to indices. Martín-Morales [2013; 2014] uses them to efficiently study monodromy zeta functions as well as explicit $\mathbb{Q}$-desingularizations of certain singularities. Artal Bartolo, Martín-Morales, and Ortigas-Galindo [Artal Bartolo et al. 2012; 2014] further study the geometry of surfaces. All this on top of the enormous literature on weighted projective spaces.

All these authors show that weighted blowings up are remarkably efficient in computing invariants of singularities. In [Martín-Morales 2013; 2014], they are shown, in a wide class of examples, to be remarkably efficient in finding $\mathbb{Q}$-resolutions, namely modifications with at most quotient singularities.

Most relevant to the present paper, Panazzolo [2006] used scheme theoretic weighted blowings up to simplify foliations in dimension three, and McQuillan and Panazzolo [2013] revisited the problem using stack theoretic blowings up. In particular it is shown there that weighted blowings up are unavoidable for their goals. The paper [McQuillan and Panazzolo 2013] led to the paper [McQuillan 2020] concurrent to ours.

In our work, stack theoretic modification appeared in [Abramovich et al. 2020a] and shown to be unavoidable for functoriality of logarithmic resolution, leading us to investigate weighted blowings up in general.

1.5. ***Invariants and parameters.*** The notation for the present invariant $\mathrm{inv}_{\mathcal{I}}(p)$ in [Abramovich et al. 2020a] was $a_1 \cdot \mathrm{inv}_{\mathcal{I}_X, a_1}(p)$, and extends to arbitrary ideal sheaves on logarithmic orbifolds. Here it is applied solely when Y is smooth with trivial logarithmic structure.

Both this invariant and our center of blowing up are present in earlier work:

This invariant $(a_1, \dots, a_k)$ is closely related to invariants developed in earlier papers on resolution of singularities, in particular [Bierstone and Milman 1997] and [Włodarczyk 2005]. In fact $(a_1, \dots, a_k)$ is determined by a sequence $(b_1, \dots, b_k)$ of integers, which is "interspersed" in Bierstone and Milman's richer invariant $(H_1, s_1, b_2, \dots, b_k, s_k)$. Here b_1 is determined by the Hilbert–Samuel function H_1 and the $s_i = 0$ since no divisors are present — our invariant is in essence the classical "year zero invariant". Invariants of similar nature are already introduced in [Hironaka 1964b].

The center $J = (x_1^{a_1}, \dots, x_k^{a_k})$ can be interpreted in terms of Newton polyhedra, and as such it appears in [Youssin 1990, Section 1], with a closely related precedent in [Hironaka 1967]. The local parameters $x_1, \dots, x_k$ in the definition of J were already introduced in [Bierstone and Milman 1997; Encinas and Villamayor 2003; Włodarczyk 2005; Abramovich et al. 2020a] as a sequence of iterated hypersurfaces of maximal contact for appropriate coefficient ideals, see Section 5.1. In this paper we prove the necessary properties of the invariant $\mathrm{inv}_{\mathcal{I}}(p)$ and the center J, but many of these properties are directly implied by these cited works.

In earlier work the ideal $(x_1, \dots, x_k)$ was used to locally define the unique center of blowing up satisfying appropriate admissibility and functoriality properties for resolution using smooth blowings up. A central observation here is that the center $(x_1^{a_1}, \dots, x_k^{a_k})$ is uniquely defined as a valuative $\mathbb{Q}$-ideal, see Theorem 5.3.1(3).

As recalled below, in general, after blowing up the reduced ideal $(x_1, \dots, x_k)$, the invariant does not drop, and may increase; Earlier work enhanced this invariant by including data of exceptional divisors and their history, or more recently, logarithmic structures. Another central observation here is that, with the use of weighted blowings up, no history, no exceptional divisors, and no logarithmic structures are necessary.

1.6. ***Tools and methods.*** The present treatment requires the theory of Deligne–Mumford stacks. The reader is assumed to be comfortable with their basic notions, such as coherent sheaves and coarse moduli spaces, though there is little harm in viewing a stack as "locally the quotient of a variety by the action of a finite group", in which case coherent sheaves are represented by equivariant sheaves on the variety, and the coarse moduli space is the schematic (or algebraic space) quotient.

An application of Bergh's destackification theorem [Bergh 2017, Theorem 1.2] or its generalization [Bergh and Rydh 2019, Theorem B] allows one to replace $X_n \subset Y_n$ by a smooth embedded scheme $X'_n \subset Y'_n$ projective over $X \subset Y$, giving a resolution in the schematic sense, see Theorem 8.1.3. Alternatively the coarse moduli space admits only abelian quotient singularities (see Section 8.2) and can be resolved directly by combinatorial methods; see [Bogomolov 1992; Abramovich and de Jong 1997; Abramovich et al. 2002; 2020c; Włodarczyk 2003; Illusie and Temkin 2014; Włodarczyk 2022]. Both destackification and this resolution process apply in arbitrary characteristics, as the stabilizer group-schemes involved are tame.[1]

Our center J can be identified as an *idealistic exponent*, see [Hironaka 1977], which we present here through the slightly more flexible formalism of *valuative $\mathbb{Q}$-ideals*, see Section 2.2, or equivalently equivariant ideals in the h topology, see Section 2.5. This formalism allows us to show with little effort that centers are unique and functorial. We believe the formalism, which is inspired by existing work on $\mathbb{Q}$-ideals, graded families of ideals, and B-divisors, is the correct formalism to consider ideals with rational multiplicities up to blowings up, a topic permeating birational geometry.

We provide a proof of the theorem based on existing theory of resolution of singularities, using concepts and methods from [Hironaka 1964a; 1964b; Villamayor 1989; Bierstone and Milman 1997;

[1]We remind the reader that, by a theorem of de Jong [1997, Corollary 5.15], as stated in [Bergh and Rydh 2019, Theorem 1.4], any variety X over a field of any characteristic admits a purely inseparable alteration $X' \to X$ with X' the coarse moduli space of a smooth Deligne–Mumford stack $\mathcal{X}'$. Thus, if the field is perfect, resolution of X is reduced to the combination of destackification of a possibly wild Deligne–Mumford stack $\mathcal{X}'$ and the resolution of a purely inseparable cover of a smooth scheme X' — using Frobenius we can realize a modification of X as a purely inseparable alteration of X'.

2008; Encinas and Villamayor 2003; 2007; Włodarczyk 2005; Kollár 2007], among others. The reader is assumed to be familiar with the introductory material in [Kollár 2007, 3.1–3.2]. We explicitly use [loc. cit., Theorem 3.67] (or [Bierstone and Milman 2008, Lemma 3.3]), [Kollár 2007, Theorem 3.92], and [loc. cit., Proposition 3.99], and our terminology (maximal contact, coefficient ideals) is consistent with Kollár's (and others') treatment.

1.7. ***Concurrent and future work.*** As indicated before, Theorem 1.2.2 was discovered independently by McQuillan [2020].

The present paper is a beginning for several other works, all requiring additional techniques.

The present treatment does not address logarithmic resolutions, a critical requirement of birational geometry. As Section 8.3 shows this does not follow by accident. The necessary modifications were worked out by Quek [2022]. This requires, in addition to the present methods, bringing in the theory of logarithmic structures as in [Abramovich et al. 2020a]. A variant of Quek's work using smooth Artin stacks is provided in [Abramovich and Quek 2021]. A variant using only smooth Deligne–Mumford stacks is provided in [Włodarczyk 2023]. The work [Włodarczyk 2023] provides an alternative view on the current work, representing the stacks as global quotients of varieties with torus actions.

The present results were discovered along the way of our work [Abramovich et al. 2020b], addressing resolution of singularities in families and semistable reduction, again using the logarithmic theory of [Abramovich et al. 2020a]. The chapter [Temkin 2023] indicates how the present methods should be introduced into that project, and carried out in the appropriate generality of *quasiexcellent schemes*, to deduce results in other geometric categories of interest, as is done in [Temkin 2012; Abramovich and Temkin 2019]. McQuillan's method [2020] is developed in the generality of quasiexcellent schemes.

Further discussion of these and other aspects is included in the volume [Abramovich et al. 2023].

1.8. ***Examples: comparing smooth and weighted blowings up.***

1.8.1. *Blowing up without weights.* It is well-known that there exists no classical "memoryless algorithm" which blows up smooth centers and is compatible with smooth morphisms in the sense of Theorem 1.2.2(3); for example, see [Kollár 2007, Claim 3.6.3]. We give here slightly different examples.

Consider first the 3-dimensional singularity

$$x^2 = y_1 y_2 y_3.$$

The singular locus consists of the three lines $x = y_i = y_j = 0$, for $i \neq j$, meeting at the origin. Due to the group of permutations acting on the singularity the only possible *invariant* smooth center is the origin: $\{x = y_1 = y_2 = y_3 = 0\}$, but its blowing up leads to the three points with singularities identical to the original one, occurring on the three y_i-charts. Writing

$$x = x' y_3', \quad y_1 = y_1' y_3', \quad y_2 = y_2' y_3', \quad \text{and} \quad y_3 = y_3'$$

we get, after clearing out y_3^2, the equation

$$x'^2 = y_1' y_2' y_3'$$

in the new coordinates.

Thus functorial embedded desingularization by smooth blowings up, using no additional structure — called "history" by some authors — is simply impossible, as it may lead to an infinite cycle.[2]

This paucity of functorial centers leads to choices which are far from optimal, and resulting in *worse* singularities.

Consider the equation

$$x^2 = y_1^a y_2^a y_3^a,$$

with $a \geq 2$ instead. The origin is again the unique possible functorial center, and leads to a singularity of the form $x^2 = y_1^a y_2^a y_3^{3a-2}$ in the y_3-chart. This visibly is a worse singularity.

1.8.2. *Weighted blowing up.* The main reason for working with smooth centers in Hironaka's approach is that we want to keep the ambient space Y smooth.

A birational geometer knows that the singularity $x^2 = y_1 y_2 y_3$ asks for the blowing up of $J = (x^2, y_1^3, y_2^3, y_3^3)$. This is the observation used by the authors mentioned in Section 1.4 above. But a weighted blowing up in the *schematic* sense gives rise to a singular ambient space Y, with abelian quotient singularities. For the classical algorithm this is a nonstarter.

As explained in Section 3, we use instead the *stack theoretic* weighted blowing up of the associated *reduced* center — in the example $J^{1/6} = (x^{1/3}, y_1^{1/2}, y_2^{1/2}, y_3^{1/2})$. The chart corresponding to y_3 is of the form

$$[\operatorname{Spec} \mathbb{C}[x', y_1', y_2', u]/\boldsymbol{\mu}_2],$$

evidently smooth, where

$$y_3 = u^2, \quad x = x'u^3, \quad y_1 = y_1'u^2, \quad y_2 = y_2'u^2,$$

and $\boldsymbol{\mu}_2 = \pm 1$ acts by $(x', y_1', y_2', u) \mapsto (-x', y_1', y_2', -u)$. The general equations, and their derivation, are given in Section 3.

Plugging this into the original equation $x^2 = y_1 y_2 y_3$ we get $u^6 x'^2 = u^6 y_1' y_2'$, where the factor u^6 is exceptional, with proper transform

$$x'^2 = y_1' y_2'.$$

[2]To resolve this, in Hironaka's classical algorithm one must encode $y_3' = 0$ as an exceptional divisor — this is quite natural and useful. One must then note that upon restriction to the first maximal contact $x = 0$ the ideal $y_1' y_1' y_3'$ factors an exceptional "monomial" part y_3'. Unfortunately in general the monomial part makes it impossible to proceed with transverse maximal contact. One must then separate it from the order-2 locus with a resolution subroutine sometimes called "the monomial stage". Only then one can find further maximal contact elements and proceed.

In other words, the vector of degrees (2, 3, 3, 3) is reduced to (2, 2, 2), an immediate and visible improvement. One more blowing up resolves the singularities (in the category of Deligne–Mumford stacks).[3]

Similarly, our general algorithm, which requires no knowledge of prior steps taken, assigns to a singularity a canonical weighted blowing up which improves actual singularities, rather than an intricate additional auxiliary structure. Consequently the natural centers and resulting valuations are much better suited for computations of various birational invariants, such as log canonical thresholds, as recalled in Section 1.4.

1.9. ***Efficiency.*** As most algorithms in algebraic geometry, our algorithm is woefully expensive computationally, and can only be carried out in low dimension and degree. One source for computational costs here is the use of the iterated factorial in the construction of invariant and centers. Still, empirically the improvements are significant. Already in [Abramovich et al. 2020a] we showed how more limited use of stack-theoretic blowings up leads to a vast improvement in efficiency. In examples the present algorithm is remarkably efficient, with great improvements even on [loc. cit.]. For instance, in the example above, two weighted blowings up suffice. Cases of interest which were out of reach for computer calculation are now computed. This adds to the evidence recalled in Section 1.4. Our process is explicitly computable, and an implementation in SINGULAR [Decker et al. 2019] is available in [Lee et al. 2020].

2. Valuative ideals, idealistic exponents, and centers

To simplify the exposition we will mainly work with schemes. All intermediate constructions can be extended to Deligne–Mumford stacks, étale topology and geometric points via étale descent, but we will use this only in the main statements and constructions, including centers, weighted blowings up and resolution invariants. For completeness, we provide in remarks and complementary sections some additional material with only outlined arguments; it will not be used and can be safely ignored if the reader prefers.

2.1. ***Zariski–Riemann spaces.*** Given an integral noetherian scheme Y we are interested in understanding ideals, and more generally $\mathbb{Q}$-ideals, as they behave after arbitrary blowing up. For instance the ideals (x^2, y^2) and (x^2, xy, y^2) coincide after blowing up the origin, and a formalism in which they are the same object is desirable. We propose to work with the Zariski–Riemann space $\boldsymbol{ZR}(Y)$ of Y, the projective limit of all projective birational transformations of Y, whose points consist of all valuation rings R of $K(Y)$ extending to a morphism $\operatorname{Spec} R \to Y$.

The space $\boldsymbol{ZR}(Y)$ carries a constant sheaf $K = K(Y)$, a subsheaf of rings $\mathcal{O}$ with stalk at v consisting of the valuation ring R_v, and a sheaf of ordered groups $\Gamma = K^*/\mathcal{O}^*$ such that $v : K^* \to \Gamma$ is the valuation. The image $v(\mathcal{O} \setminus \{0\}) =: \Gamma_+ \subset \Gamma$ is the valuation monoid consisting of nonnegative sections of Γ.

The space $\boldsymbol{ZR}(Y)$ is quasicompact; see [Temkin 2010, Proposition 3.2.1]. If $Y = \bigcup Y_i$ is reduced but possibly reducible with irreducible components Y_i, we define $\boldsymbol{ZR}(Y) := \bigsqcup \boldsymbol{ZR}(Y_i)$.

[3] Hironaka's classical algorithm requires many more blowings up, and, as indicated in the previous note, is quite technically involved.

Remark 2.1.1. While Theorem 1.2.2 is applied to Deligne–Mumford stacks $X \subset Y$, functoriality means that we can always work on an étale cover by a scheme $\tilde{X} \subset \tilde{Y}$: the resolution step $F_1(X \subset Y)$ is obtained by étale descent from $F_1(\tilde{X} \subset \tilde{Y})$. In particular we need not introduce $\boldsymbol{ZR}(Y)$ for a stack. Nevertheless we note that such $\boldsymbol{ZR}(Y)$ can be constructed as well, be it by étale descent, or directly as a limit, or as a suitably normalized fibered product of Y with the Zariski–Riemann space of the coarse moduli space.

2.2. *Valuative $\mathbb{Q}$-ideals.*

Definition 2.2.1. (1) By a *valuative ideal* on Y we mean a section $\gamma \in H^0(\boldsymbol{ZR}(Y), \Gamma_+)$. Every ideal $\mathcal{I}$ on every birational model $Y' \to Y$, proper over Y, defines a valuative ideal that we denote $v(\mathcal{I})$ by taking the minimal element of the image of $\mathcal{I}$ in Γ_+.

(2) The group $\Gamma_{\mathbb{Q}} = \Gamma \otimes \mathbb{Q}$ is also ordered. We denote the monoid of nonnegative elements by $\Gamma_{\mathbb{Q}+}$. By a *valuative $\mathbb{Q}$-ideal* we mean a section $\gamma \in H^0(\boldsymbol{ZR}(Y), \Gamma_{\mathbb{Q}+})$.

(3) Any dominant morphism $f : Z \to Y$ induces a map $\boldsymbol{ZR}(Z) \to \boldsymbol{ZR}(Y)$. For a valuative $\mathbb{Q}$-ideal γ on Y its image under the induced map $\Gamma_Y \otimes \mathbb{Q} \to \Gamma_Z \otimes \mathbb{Q}$ will be denoted $f^{-1}(\gamma)$ and called the *preimage of γ on Z*.

Ideals with the same integral closure have the same valuative ideal. Every valuative ideal γ defines an ideal sheaf $\mathcal{I}'_\gamma$ on every modification Y' of Y by taking $\mathcal{I}'_\gamma := \{f \in \mathcal{O}_{Y'} \mid v(f) \geq \gamma_v \forall v\}$, which is automatically integrally closed. We will use only the ideal $\mathcal{I}_\gamma$ thus defined on Y itself.

The definition of $\mathcal{I}_\gamma$ extends to valuative $\mathbb{Q}$-ideals. Conversely, there is a convenient way to consider $\mathbb{Q}$-ideals, extending the definition of $v(\mathcal{I})$: given a finite collection $f_i \in \mathcal{O}_Y$ and $a_i \in \mathbb{Q}_{>0}$ we write

$$(f_1^{a_1}, \dots, f_k^{a_k}) := (\min\{a_i \cdot v(f_i)\})_v \in H^0(\boldsymbol{ZR}(Y), \Gamma_{\mathbb{Q}+}) \tag{1}$$

for the naturally associated valuative $\mathbb{Q}$-ideal. When a_i are integers this coincides with $v(f_1^{a_1}, \dots, f_k^{a_k})$.

Remark 2.2.2. As was pointed out by D. Rydh, valuative $\mathbb{Q}$-ideals are equivalent to effective $\mathbb{Q}$-Cartier divisors on $\boldsymbol{ZR}(X)$. Indeed, any section γ of Γ_+ is locally the image of an element of $\mathcal{O}$, and since $\boldsymbol{ZR}(X)$ is quasicompact, finitely many such representatives suffice. Moreover, taking a common birational model $Y' \to Y$ over which all the representative sections are regular, we find that γ is an invertible ideal on Y'. Allowing denominators, any valuative $\mathbb{Q}$-ideal γ is written, using the notation of (1), locally on the model Y' as $\gamma = (f^a)$.

2.3. *Complements: idealistic exponents.* A valuative $\mathbb{Q}$-ideal which is represented locally on Y itself as $(f_1^{a_1}, \dots, f_k^{a_k})$ is an *idealistic exponent*. This notion coincides with Hironaka's [1977, Definition 3] by [loc. cit., Remark (2.2)]. Hironaka's notation $(\mathcal{J}, b)$, with $\mathcal{J} \subset \mathcal{O}_Y$, $b \in \mathbb{N}$ translates to the valuative $\mathbb{Q}$-ideal $\mathcal{J}^{1/b}$. Hironaka's definition of pullback of an idealistic exponent under a dominant morphism $Y' \to Y$ extends to an arbitrary valuative $\mathbb{Q}$-ideal.

As indicated in the next section, these are related to Rees algebras [Encinas and Villamayor 2007] or graded families of ideals [Lazarsfeld 2004, Section 2.4.B]. This relationship was pursued in greater depth by Quek [2022].

2.4. ***Centers and admissibility.***

Definition 2.4.1. (1) By a *center* J on a regular scheme Y we mean a valuative $\mathbb{Q}$-ideal for which there is an affine covering $Y = \cup U_i$ and regular systems of parameters $(x_1^{(i)}, \ldots, x_k^{(i)}) = (x_1, \ldots, x_k)$ on U_i such that $J_{U_i} = (x_1^{a_1}, \ldots, x_k^{a_k})$ for some $a_j \in \mathbb{Q}_{>0}$ independent of i.

(2) A center J is *admissible* for a valuative $\mathbb{Q}$-ideal β if $J_v \leq \beta_v$ for all v. A center is *admissible* for an ideal $\mathcal{I}$ if it is admissible for the associated valuative $\mathbb{Q}$-ideal $v(\mathcal{I})$, in which case we use the suggestive notation $\mathcal{I} \subseteq J$.

(3) The center J is *reduced* if $w_i = 1/a_i$ are positive integers with $\gcd(w_1, \ldots, w_k) = 1$. For any center J we write $\bar{J} = (x_1^{1/w_1}, \ldots, x_k^{1/w_k})$ for the unique reduced center such that $\bar{J}^\ell = J$ for some $\ell \in \mathbb{Q}_{>0}$.

In Section 3 below we define the blowing up of $(x_1^{1/w_1}, \ldots, x_k^{1/w_k})$. In Section 5.2 we show how admissibility is manifested in terms of this blowing up, and becomes very much analogous to the notion used in earlier resolution algorithms.

Remark 2.4.2. Using the coordinates as in (1), the center J corresponds to a unique monomial valuation associated to the cocharacter

$$(a_1^{-1}, \ldots, a_k^{-1}, 0, \ldots, 0),$$

where $v\left(\prod x_i^{c_i}\right) = \sum_{i=1}^k c_i/a_i$.

The definition of centers extends to stacks similarly to usual ideals.

Definition 2.4.3. Let Y be a Deligne–Mumford stack:

(1) By $\mathrm{Cov}(Y)$ we denote the category of étale covers $Y' \to Y$ with Y' a scheme and Y-morphisms between the covers.

(2) A *center* J on Y is a compatible family of centers J' on the elements Y' of $\mathrm{Cov}(Y)$: for any morphism $f : Y'' \to Y'$ in $\mathrm{Cov}(Y)$ one has $f^{-1}J' = J''$.

Partial regular families of parameters are preserved by preimages under smooth and, more generally, regular morphisms (see [Stacks 2005–, 07R6]), hence we have:

Lemma 2.4.4. *If $f : Y' \to Y$ is a regular morphism of regular schemes and J is a center on Y, then $f^{-1}J$ is a center on Y'.*

Remark 2.4.5. In fact, the inverse is also true: is f is surjective, γ is a valuative $\mathbb{Q}$-ideal and $f^{-1}\gamma$ is a center, then γ is a center. As a corollary one obtains an extension of these claims to stacks and the claim that a center on a stack can be defined using a single presentation rather than the whole category $\mathrm{Cov}(Y)$. However, we will not need these natural but not completely trivial results.

2.5. ***Complements: relation with the h topology.*** The following observation is not used in the paper, so we just outline it without proof. Valuative $\mathbb{Q}$-ideals are closely related to what we call equivariant ideals in the h topology, where Zariski open coverings and alterations generate a cofinal collection of coverings; see [Voevodsky 1996, Definition 3.1.5 and Theorem 3.1.9]. The structure sheaf $\mathcal{O}_{X_h}$ is the sheafification of the presheaf $U \mapsto \Gamma(\mathcal{O}_U)$. In fact, $\mathcal{O}_{X_h}(U) = \Gamma(\mathcal{O}_{U^{sn}})$, where U^{sn} is the seminormalized reduction of U; see [Huber and Jörder 2014, Proposition 4.5]. Any finitely generated ideal $\mathcal{J} \subseteq \mathcal{O}_{X_h}$ is generated by ideals $J_i \subseteq \mathcal{O}_{Y_i}$ on a Zariski cover $Y' = \cup Y_i$ of an alteration $Y' \to Y$. Refining the alteration we can achieve that pullbacks of J_i agree on the intersections, so $\mathcal{J}$ comes from an ideal J' on Y' and hence yields a valuative ideal γ' on Y'. Refining Y' further we can achieve that $Y' \to Y$ is a Galois alteration, namely it splits into a composition of a Galois cover $Y' \to Y''$, with Galois group G, and a generically radicial alteration $Y'' \to Y$. On the level of sets $\boldsymbol{ZR}(Y')/G = \boldsymbol{ZR}(Y'') = \boldsymbol{ZR}(Y)$, hence γ' comes from a valuative $\mathbb{Q}$-ideal γ if and only if γ' is G-equivariant. In fact, the latter happens if and only if one can choose Y' and J' so that already J' is G-equivariant.

3. Weighted blowings up

Stack theoretic projective spectra were considered informally by Miles Reid, introduced officially in [Abramovich and Hassett 2011] to study moduli spaces of varieties, and treated in Olsson's book [2016, Section 10.2.7].

The manuscript by Quek and Rydh [2021] provides foundations for stack-theoretic blowings up. The presentation here is rather terse as complete details already appear there. The local equations we present here can be found in [Kollár et al. 2004, page 167], where they are developed for the study of log canonical thresholds. The graded algebras we present below are special cases of the graded families of ideals discussed in [Lazarsfeld 2004, Section 2.4.B], especially Example 2.4.8.

From now on Y is a smooth Deligne–Mumford stack over a field k of characteristic zero. In Sections 3–5, if not said to the contrary, Y is also assumed to be a variety.

3.1. ***Graded algebras and their $\mathcal{P}$roj.*** Given a quasicoherent graded algebra $\mathcal{A} = \bigoplus_{m \geq 0} \mathcal{A}_m$ on Y with associated $\mathbb{G}_m$-action defined by $(t, s) \mapsto t^m s$ for $s \in \mathcal{A}_m$ we define its stack-theoretic projective spectrum to be

$$\mathcal{P}\mathrm{roj}_Y \mathcal{A} := [(\mathrm{Spec}_{\mathcal{O}_Y} \mathcal{A} \setminus S_0)/\mathbb{G}_m],$$

where the vertex S_0 is the zero scheme of the ideal $\bigoplus_{m>0} \mathcal{A}_m$; see [Quek and Rydh 2021, Section 1.2]. When $\mathcal{A}_1$ is coherent and generates $\mathcal{A}$ over $\mathcal{A}_0$ this agrees with the construction in [Hartshorne 1977, II.7, page 160]; see [Quek and Rydh 2021, Corollary 1.6.2]. As usual $\mathcal{P}\mathrm{roj}_Y \mathcal{A}$ carries an invertible sheaf $\mathcal{O}_{\mathcal{P}\mathrm{roj}_Y \mathcal{A}}(1)$ corresponding to the graded module $\mathcal{A}(1)$. When $\mathcal{A}$ is finitely generated over $\mathcal{O}_Y$ with coherent graded components the resulting morphism $\mathcal{P}\mathrm{roj}_Y \mathcal{A} \to Y$ is proper; see [Quek and Rydh 2021, Proposition 1.6.1(ii)].

3.2. ***Rees algebras of ideals.*** If $\mathcal{I}$ is an ideal on Y, its Rees algebra is $\mathcal{A}_{\mathcal{I}} := \bigoplus_{m\geq 0} \mathcal{I}^m$, and the blowing up of $\mathcal{I}$ is $Y' = \mathrm{Bl}_Y(\mathcal{I}) := \mathcal{P}\mathrm{roj}_Y(\mathcal{A}_{\mathcal{I}})$. It is the universal birational map making $\mathcal{I}\mathcal{O}_{Y'}$ invertible, in this case $Y' \to Y$ projective; see definition [Hartshorne 1977, II.7, page 163].

3.3. ***Rees algebras of valuative $\mathbb{Q}$-ideals.***

Definition 3.3.1. (1) Given a valuative $\mathbb{Q}$-ideal γ we define its *Rees algebra* to be

$$\mathcal{A}_\gamma := \bigoplus_{m\in\mathbb{N}} \mathcal{I}_{m\gamma}.$$

(2) The *blowing up* of γ is defined to be $Y' = \mathrm{Bl}_Y(\gamma) := \mathcal{P}\mathrm{roj}_Y\, \mathcal{A}_\gamma$.

At least when $\gamma = (f_1^{a_1}, \dots, f_k^{a_k})$ is an idealistic exponent, $Y' \to Y$ satisfies a corresponding universal property; see [Quek and Rydh 2021, Proposition 3.5.3]. Since we will not use this property in this paper, we just mention that the valuative $\mathbb{Q}$-ideal $E = \gamma\mathcal{O}_{Y'}$, in a suitable sense of Zariski–Riemann spaces of stacks, or as an h-ideal, becomes an *invertible ideal sheaf* on Y'. We only show this below for the blowing up of a center.

Note that if $Y_1 \to Y$ is flat and $Y_1' = \mathrm{Bl}_Y(\gamma\mathcal{O}_{Y_1})$ then $Y_1' = Y' \times_Y Y_1$.

3.4. ***Weighted blowings up: local equations.*** Now consider the situation where γ is a center of the special form $J = (x_1^{1/w_1}, \dots, x_k^{1/w_k})$, with $w_i \in \mathbb{N}$. In this case the algebra $\mathcal{A}_\gamma = \bigoplus_{m\in\mathbb{N}} \mathcal{I}_{m\gamma}$, with $\mathcal{I}_{m\gamma} = \big(x_i^{b_1} \cdots x_n^{b_n} \mid \sum w_i b_i \geq m\big)$ is finitely generated. It is the integral closure inside $\mathcal{O}_Y[T, T^{-1}]$ of the simpler algebra with generators $(x_i)T^{w_i}$. We can therefore describe $\mathrm{Bl}_Y(J) = \mathrm{Bl}_Y(\gamma)$, which deserves to be called a stack-theoretic weighted blowing up, explicitly in local coordinates, as follows [Quek and Rydh 2021, Corollary 4.4.4]:

The chart associated to x_1 has local variables $u, x_2', \dots, x_n'$, where

- $x_1 = u^{w_1}$,
- $x_i' = x_i/u^{w_i}$ for $2 \leq i \leq k$, and
- $x_j' = x_j$ for $j > k$.

The group $\boldsymbol{\mu}_{w_1}$ acts through

$$(u, x_2', \dots, x_k') \mapsto (\zeta_{w_1} u, \zeta_{w_1}^{-w_2} x_2', \dots, \zeta_{w_1}^{-w_k} x_k')$$

and trivially on x_j', $j > k$, giving an étale local isomorphism of the chart with

$$[\mathrm{Spec}\, k[u, x_2', \dots, x_n']/\boldsymbol{\mu}_{w_1}].$$

It is easy to see that these charts glue to a stack-theoretic modification $Y' \to Y$ with a smooth Y' and its coarse space is the classical (singular) weighted blowing up.

Write $E = (u)$ for the exceptional ideal. Then $v(E) = (x_1^{1/w_1}, \dots, x_k^{1/w_k})$, and this persists on all charts, in other words the center $(x_1^{1/w_1}, \dots, x_k^{1/w_k})$ becomes an invertible ideal sheaf on Y'.

We sometimes, but not always, insist on $\gcd(w_1, \dots, w_k) = 1$, in which case the center is *reduced*. We will however need to consider the proper transform of the locus $H = \{x_1 = 0\}$, where it may happen that $\gcd(w_2, \dots, w_k) \neq 1$. The relationships are summarized by the following lemma, which uses the construction of the root stack $Y(E^{1/c})$ along a divisor E (for a treatment on a stack see [Abramovich and Fantechi 2016, Section 1.1]) and follows from [Quek and Rydh 2021, Corollary 3.2.1] or by considering the charts:

Lemma 3.4.1. *If $J' = (x_1^{1/w_1}, \dots, x_k^{1/w_k})$ and $J'' = (x_1^{1/cw_1}, \dots, x_k^{1/cw_k})$ with w_i, c positive integers, and if $Y', Y'' \to Y$ are the corresponding blowings up, with E', E'' the exceptional divisors, then $Y'' = Y'(\sqrt[c]{E'})$ is the root stack of Y' along E'.*

Write $H = \{x_1 = 0\}$, and $H' \to H$ the blowing up of the **reduced** *center $\bar{J}'_H$ associated to $J'_H := (x_2^{1/w_2}, \dots, x_k^{1/w_k})$, with exceptional E_H. Then the proper transform $\tilde{H}' \to H$ of H via the blowing up of J'' is the root stack $H'(\sqrt[(cc')]{E_H})$ of H' along $E_H \subset H'$, where $c' = \gcd(w_2, \dots, w_k)$. Therefore $\tilde{H}'$ is the blowing up of ${\bar{J}'_H}^{1/(cc')}$.*

3.5. *Derivation of equations.* Let us derive the description in Section 3.4 above, in a manner similar to [Quek and Rydh 2021, Lemma 1.3.1]. Write $y_i = x_i T^{w_i}$. The x_1-chart is the stack $[\operatorname{Spec} \mathcal{A}[y_1^{-1}]/\mathbb{G}_m]$. The slice $W_1 := \operatorname{Spec} \mathcal{A}[y_1^{-1}]/(y_1 - 1)$ is stabilized by $\boldsymbol{\mu}_{w_1}$, so the embedding $W_1 \subset \operatorname{Spec} \mathcal{A}[y_1^{-1}]$ gives rise to a morphism $\phi : [W_1/\boldsymbol{\mu}_{w_1}] \to [\operatorname{Spec} \mathcal{A}[y_1^{-1}]/\mathbb{G}_m]$. This is an isomorphism: the equation $u^{w_1} = x_1$ describes a $\boldsymbol{\mu}_{w_1}$-torsor on $\operatorname{Spec} \mathcal{A}[y_1^{-1}]$ mapping to W_1 equivariantly via $T \mapsto u^{-1}$. The resulting morphism $\operatorname{Spec} \mathcal{A}[y_1^{-1}] \to [W_1/\boldsymbol{\mu}_{w_1}]$ descends to $[\operatorname{Spec} \mathcal{A}[y_1^{-1}]/\mathbb{G}_m] \to [W_1/\boldsymbol{\mu}_{w_1}]$ which is an inverse to ϕ.

It thus remains to show that $[W_1/\boldsymbol{\mu}_{w_1}]$ has the local description above. Since $T^{-w_1} = y_1^{-1} x_1 \in \mathcal{A}[y_1^{-1}]$ and $\mathcal{A}$ is integrally closed in $\mathcal{O}_Y[T, T^{-1}]$ we have $u := T^{-1} \in \mathcal{A}[y_1^{-1}]$, and its restriction to W_1 satisfies $u^{w_1} = x_1$. For $i = 2, \dots, k$ we write x'_i for the restriction of y_i, obtaining $x'_i = x_i/u^{w_i}$. Now W_1 is normal and finite birational over $\operatorname{Spec} k[u, x'_2, \dots, x'_n]$, hence they are isomorphic.

3.6. *Complements: local toric description of weighted blowings up* [Quek and Rydh 2021, Section 4.5.5]**.** Again working locally, assume that $Y = \operatorname{Spec} k[x_1, \dots, x_n]$. It is the affine toric variety associated to the monoid $\mathbb{N}^n \subset \sigma = \mathbb{R}^n_{\geq 0}$. Here the generator e_i of $\mathbb{N}^n$ corresponds to the monomial valuation v_i associated to the divisor $x_i = 0$, namely $v_i(x_j) = \delta_{ij}$.

The monomial x_i^{1/w_i} defines the linear function on σ whose value on $(b_1, \dots, b_n)$ is its valuation b_i/w_i. The ideal $(x_1^{1/w_1}, \dots, x_k^{1/w_k})$ thus defines the piecewise linear function $\min_i\{b_i/w_i\}$, which becomes linear precisely on the star subdivision $\Sigma = v_{\bar{J}} \star \sigma$ with

$$v_{\bar{J}} = (w_1, \dots, w_k, 0, \dots, 0).$$

This defines the scheme theoretic weighted blowing up $\bar{Y}'$; see [Reid 1980, Section 4]. Note that this cocharacter $v_{\bar{J}}$ is a multiple of the valuation associated to the exceptional divisor of the center.

Since $v_{\bar{J}}$ is assumed integral, we can apply the theory of toric stacks [Borisov et al. 2005; Fantechi et al. 2010; Geraschenko and Satriano 2015a; 2015b; Gillam and Molcho 2015]. We have a smooth toric stack $Y' \to \overline{Y}'$ associated to the same fan Σ with the cone $\sigma_i = \langle v_{\bar{J}}, e_1, \dots, \hat{e}_i, \dots, e_n \rangle$ endowed with the sublattice $N_i \subset N$ generated by the elements $v_{\bar{J}}, e_1, \dots, \hat{e}_i, \dots, e_n$, for all $i = 1, \dots, k$. This toric stack is precisely the stack theoretic weighted blowing up $Y' \to Y$. One can derive the equations in Section 3.4 from this toric picture.

4. Coefficient ideals

In this section we recall some notions from the classical embedded resolution. By Y we denote a smooth k-variety.

4.1. ***Graded algebra and coefficient ideals.*** Fix an ideal $\mathcal{I} \subset \mathcal{O}_Y$ and an integer $a > 0$. We use the notation of [Abramovich et al. 2020a], except that we use the saturated coefficient ideal as in [Kollár 2007; Abramovich et al. 2020b], which is consistent with the Rees algebra approach of [Encinas and Villamayor 2007]:

Definition 4.1.1. (1) Consider the graded subalgebra $\mathcal{G} = \mathcal{G}(\mathcal{I}, a) \subseteq \mathcal{O}_Y[T]$ generated by placing $\mathcal{D}^{\leq a-i}\mathcal{I}$ in degree i. Its graded pieces are

$$\mathcal{G}_j = \sum_{\sum_{i=0}^{a-1}(a-i)\cdot b_i \geq j} \mathcal{I}^{b_0} \cdot (\mathcal{D}^{\leq 1}\mathcal{I})^{b_1} \cdots (\mathcal{D}^{\leq a-1}\mathcal{I})^{b_{a-1}},$$

where the sum runs over all monomials in the ideals $\mathcal{I}, \dots, \mathcal{D}^{\leq a-1}\mathcal{I}$ of weighted degree

$$\sum_{i=0}^{a-1}(a-i)\cdot b_i \geq j.$$

(2) Let $\mathcal{I} \subset \mathcal{O}_Y$ and $a \geq 1$ an integer. Define the *coefficient ideal*

$$C(\mathcal{I}, a) := \mathcal{G}_{a!}.$$

The product rule, and the trivial inclusion $\mathcal{D}^{\leq 1}\mathcal{D}^{\leq a-1}\mathcal{I} \subset (1)$, imply that $\mathcal{D}\mathcal{G}_{k+1} \subset \mathcal{G}_k$ for $k \geq 0$. The formation of $\mathcal{G}$ and $C(\mathcal{I}, a)$ is functorial for smooth morphisms: if $Y_1 \to Y$ is smooth then $C(\mathcal{I}, a)\mathcal{O}_{Y_1} = C(\mathcal{I}\mathcal{O}_{Y_1}, a)$. This follows since the formation of $\mathcal{D}^{\leq 1}\mathcal{I}$, ideal product, and ideal sum are all functorial.

4.2. ***Maximal contact.*** For the rest of the section we assume that $\mathcal{I} \subset \mathcal{O}_Y$ has maximal order $\leq a$. Recall that an element $x \in \mathcal{D}^{\leq a-1}\mathcal{I}$ which is a regular parameter at $p \in Y$ is called *a maximal contact element* at p, and its vanishing locus *a maximal contact hypersurface* at p. In general, maximal contact only exists locally. For completeness, any parameter is a maximal contact element for the unit ideal.

The coefficient ideal combines sufficient information from derivatives of $\mathcal{I}$ so that when one restricts $C(\mathcal{I}, a)$ to a hypersurface of maximal contact H no information necessary for resolution is lost. For example, this is manifested in the equivalence (in the sense of [Bierstone and Milman 1997]) of $(\mathcal{I}, a)$ and $C(\mathcal{I}, a)|_H$.

4.3. ***Invariance.*** Now consider $\mathcal{I} \subset \mathcal{O}_Y$ and assume that $x_1 \in \mathcal{D}^{\leq a-1}\mathcal{I}$ is a maximal contact element at $p \in Y$. The ideals $\mathcal{G}_i$ enjoy a strong invariance property summarized in the following theorem:[4]

Theorem 4.3.1. *Let x_1 and x_1' be maximal contact elements at p, and $x_2, \dots, x_n \in \mathcal{O}_{Y,p}$ such that $(x_1, x_2, \dots, x_n)$ and $(x_1', x_2, \dots, x_n)$ are both regular sequences of parameters. There is a scheme $\tilde{Y}$ with point $\tilde{p} \in \tilde{Y}$ and two morphisms $\phi, \phi' : \tilde{Y} \to Y$ with $\phi(\tilde{p}) = \phi'(\tilde{p}) = p$, both étale at p, satisfying*

(1) $\phi^* x_1 = \phi'^* x_1'$,

(2) $\phi^* x_i = \phi'^* x_i$ *for* $i = 2, \dots, n$, *and*

(3) $\phi^* \mathcal{G}_i = \phi'^* \mathcal{G}_i$.

This is [Kollár 2007, Theorem 3.92], generalizing [Włodarczyk 2005, Lemma 3.5.5].[5]

4.4. ***Formal decomposition.*** We now pass to formal completions. Fixing a field of coefficients $k_p = k(p) \hookrightarrow \hat{\mathcal{O}}_{Y,p}$ and extending to a regular sequence of parameters we have $\hat{\mathcal{O}}_{Y,p} = k_p[\![x_1, \dots, x_n]\!]$. We use the reduction homomorphism $k_p[\![x_1, \dots, x_n]\!] \to k_p[\![x_2, \dots, x_n]\!]$ and the inclusion $k_p[\![x_2, \dots, x_n]\!] \to k_p[\![x_1, \dots, x_n]\!]$.

We have $\mathcal{G}_j = (x_1^j) + (x_1^{j-1})\mathcal{G}_1 + \dots + (x_1)\mathcal{G}_{j-1} + \mathcal{G}_j$ since the ideal on the left contains every term on the right. Write $\bar{\mathcal{C}}_j = \mathcal{G}_j k_p[\![x_2, \dots, x_n]\!] \subset k_p[\![x_2, \dots, x_n]\!]$ via the reduction homomorphism sending x_1 to 0, and $\widetilde{\mathcal{C}}_j = \bar{\mathcal{C}}_j k_p[\![x_1, \dots, x_n]\!] \subset k_p[\![x_1, \dots, x_n]\!]$ its image via inclusion. We hope the reader can distinguish the notation $\bar{\mathcal{C}}_j$ from $\widetilde{\mathcal{C}}_j$.

Proposition 4.4.1. *Denoting the completions $\hat{\mathcal{G}}_j = \mathcal{G}_j \hat{\mathcal{O}}_{Y,p}$ and $\hat{C}(\mathcal{I}, a) = C(\mathcal{I}, a)\hat{\mathcal{O}}_{Y,p}$, we have*

$$\hat{\mathcal{G}}_j = (x_1^j) + (x_1^{j-1})\widetilde{\mathcal{C}}_1 + \dots + (x_1)\widetilde{\mathcal{C}}_{j-1} + \widetilde{\mathcal{C}}_j,$$

in particular

$$\hat{C}(\mathcal{I}, a) = (x_1^{a!}) + (x_1^{a!-1}\widetilde{\mathcal{C}}_1) + \dots + (x_1\widetilde{\mathcal{C}}_{a!-1}) + \widetilde{\mathcal{C}}_{a!}.$$

Proof. We write $x = x_1$. Apply induction on j, noting that $\hat{\mathcal{G}}_0 = (1)$ so that we may start with $(1) = \widetilde{\mathcal{C}}_0$ and inductively assume the equality holds up to $j - 1$.

For an integer $M > j$ the ideals $\hat{\mathcal{G}}_j \supset (x^M)$ are stable under the linear operator $x\partial/\partial x$. Hence the quotient $\hat{\mathcal{G}}_j/(x^M)$ inherits a linear action, with m-eigenspaces we denote $x^m \cdot \hat{\mathcal{G}}_j^{(m)} \subset x^m k_p[\![x_2, \dots, x_n]\!]$, giving

$$\hat{\mathcal{G}}_j/(x^{M+1}) = \hat{\mathcal{G}}_j^{(0)} \oplus x \cdot \hat{\mathcal{G}}_j^{(1)} \oplus \dots \oplus x^m \cdot \hat{\mathcal{G}}_j^{(m)} \oplus \dots \oplus x^M \cdot \hat{\mathcal{G}}_j^{(M)},$$

with $\hat{\mathcal{G}}_j^{(m)} \subset k_p[\![x_2, \dots, x_n]\!]$ and equality holding for $m \geq j$. Note that $\hat{\mathcal{G}}_j^{(0)} = \bar{\mathcal{C}}_j$.

[4] The reader familiar with [Kollár 2007, Section 3.53] will recognize that $\mathcal{G}_i$ are all *MC-invariant*: $\mathcal{G}_1 \cdot \mathcal{D}^{\leq 1}\mathcal{G}_i \subset \mathcal{G}_i$, hence they are *homogeneous* in the sense of [Włodarczyk 2005].

[5] These are the easier properties of coefficient ideals. We emphasize that we do not require the harder part (4) of [Włodarczyk 2005, Lemma 3.5.5] or [Kollár 2007, Theorem 3.97] describing the behavior after a sequence of blowings up.

The subspaces $\hat{\mathcal{G}}_j^{(m)} \subset k_p[\![x_2, \dots, x_n]\!]$ are independent of the choice of $M \geq m$. Moreover $x^j \cdot \hat{\mathcal{G}}_j^{(m)} \subset \hat{\mathcal{G}}_j \cap x^j \cdot k_p[\![x_2, \dots, x_n]\!]$, so that

$$\hat{\mathcal{G}}_j^{(m)} = \frac{\partial^j}{\partial x^j}(x^j \cdot \hat{\mathcal{G}}_j^{(m)}) \subset \hat{\mathcal{G}}_{j-m} \cap k_p[\![x_2, \dots, x_n]\!] \subset \bar{\mathcal{C}}_{j-m}.$$

Taking ideals we obtain

$$\hat{\mathcal{G}}_j \subset \hat{\mathcal{G}}_j^{(0)} + (x)\widetilde{\mathcal{C}}_{j-1} + \cdots + (x^{j-1})\widetilde{\mathcal{C}}_1 + (x^j).$$

Induction gives

$$(x)\widetilde{\mathcal{C}}_{j-1} + \cdots + (x^{j-1})\widetilde{\mathcal{C}}_1 + (x^j) = (x)\hat{\mathcal{G}}_{j-1} \subset \hat{\mathcal{G}}_j.$$

Together with $\bar{\mathcal{C}}_j = \hat{\mathcal{G}}_j^{(0)} \subset \hat{\mathcal{G}}_j$ the equality follows. □

By [Kollár 2007, Proposition 3.99] we have $(\mathcal{D}^{\leq j} C(\mathcal{I}, a))^{a!} \subset C(\mathcal{I}, a)^{a!-j}$. This implies:

Corollary 4.4.2. $$(\widetilde{\mathcal{C}}_{a!-j})^{a!} \subset \widetilde{\mathcal{C}}_{a!}^{a!-j}.$$

5. Invariants, local centers, and admissibility

In this section we continue to work on a smooth variety Y and fix an ideal $\mathcal{I} \subseteq \mathcal{O}_Y$. All definitions and results will be local at a point p, and to simplify notation we will use the same letter Y after passing to a neighborhood, where a maximal contact at p is defined (a pedantic reader can simply work with the localization $Y_p = \operatorname{Spec}(\mathcal{O}_{Y,p})$ instead).

5.1. *Existence of invariants and centers.*

Definition 5.1.1. (1) For an ideal $\mathcal{I} \subset \mathcal{O}_Y$ and sequence of parameters $x_1, \dots, x_k$ at p one defines $\mathcal{I}[1] = \mathcal{I}$ and recursively ideals $\mathcal{I}[i]$ and integers b_i by setting $b_i = \operatorname{ord}_p(\mathcal{I}[i])$ and $\mathcal{I}[i+1] = C(\mathcal{I}[i], b_i)|_{V(x_1, \dots, x_i)}$, ending with either $k = 1, \mathcal{I} = (1)$ or $\mathcal{I}[k+1] = 0$. The sequence of parameters $x_1, \dots, x_k$ at p is called a *maximal contact sequence* if each x_i is a maximal contact for $(\mathcal{I}[i], b_i)$ at p.

(2) To a maximal contact sequence we associate the invariant $\operatorname{inv}_{\mathcal{I}}(p) = (a_1, \dots, a_k)$, where $a_i = b_i / \prod_{j=1}^{i-1} b_j!$ and the center $J = J_p(\mathcal{I}) = (x_1^{a_1}, \dots, x_k^{a_k})$.

Obviously, a maximal contact sequence exists, and it is empty if and only if $\mathcal{I} = 0$ at p, in which case we also have that $J = 0$ and $\operatorname{inv} = ()$ is empty. The other extreme occurs when $\mathcal{I} = (1)$, in which case $J = (1)$ and $\operatorname{inv} = (0)$. Note also that $\operatorname{inv}_{\mathcal{I}[1]}(p) = (a_1, \operatorname{inv}_{\mathcal{I}[2]}(p)/(a_1 - 1)!)$ the concatenation, and $x_2, \dots, x_k$ are lifts of the parameters for $\mathcal{I}[2]$. In the notation of Section 4.4, $\mathcal{I}[2] = \bar{\mathcal{C}}_{a_1!}$.

The invariant and center in Definition 5.1.1(2) require the choice of a maximal contact sequence. The goal of Section 5 is to prove that the invariant and the center (as a valuative $\mathbb{Q}$-ideal) are independent of the choice of maximal contacts. This is at once a consequence and a generalization of Theorem 4.3.1.

A posteriori, this will also imply that $\operatorname{inv}_{\mathcal{I}}(p)$ is the maximal invariant of a center admissible for $\mathcal{I}$ at p and J is the unique center of maximal invariant admissible for $\mathcal{I}$ at p — a characterization which can be used as a choice-free definition.

Remark 5.1.2. The string $(b_1, \dots, b_k)$ was used as a singularity invariant in [Abramovich et al. 2020a], but it is its rescaling $(a_1, \dots, a_k)$ which gives a natural definition of the canonical center J independent of choices.

We order the set of invariants lexicographically, with truncated sequences considered larger, for instance

$$(1, 1, 1) < (1, 1, 2) < (1, 2, 1) < (1, 2) < (2, 2, 1).$$

The invariant takes values in a well-ordered subset Ξ_n, $n = \dim Y$, since it is order-equivalent to $(b_1, \dots, b_k)$. Explicitly write $\Xi_1 = \mathbb{N}^{\geq 1}$ and

$$\Xi_n = \Xi_1 \sqcup \bigsqcup_{a \geq 1} \{a\} \times \frac{\Xi_{n-1}}{(a-1)!}.$$

In particular, the denominators are bounded in terms of the previous entries of the invariant.

Theorem 5.1.3 [Abramovich et al. 2020a]. *Keep the above notation, then*:

(1) *The invariant* $\mathrm{inv}_{\mathcal{I}}(p)$ *is independent of the choices.*

(2) *The invariant function* $\mathrm{inv}_{\mathcal{I}} : Y \to \Xi_n$ *is constructible and upper-semicontinuous.*

(3) *The invariant is functorial for smooth morphisms*: *if* $f : Y' \to Y$ *is smooth and* $\mathcal{I}' = f^{-1}\mathcal{I}$, *then* $\mathrm{inv}_{\mathcal{I}'} = \mathrm{inv}_{\mathcal{I}} \circ f$.

Proof. (3) Since both $\mathrm{ord}_p(\mathcal{I})$ and the formation of coefficient ideals are functorial for smooth morphisms, the invariant is functorial for smooth morphisms, once parameters are chosen.

(1) We now show that the choices of maximal contacts do not change the invariant. The integer $a_1 = \mathrm{ord}_p(\mathcal{I}) = \max\{a : \mathcal{I}_p \subseteq \mathfrak{m}_p^a\}$ requires no choices. Given a regular sequence of parameters $(x_1, \dots, x_n)$ extending $(x_1, \dots, x_k)$, and given another maximal contact element x_1', we may choose constants t_i, and replace $x_2, \dots, x_n$ by $x_2 + t_2 x_1, \dots, x_n + t_n x_1$ so that also $(x_1', x_2, \dots, x_n)$ is a regular sequence of parameters.

Taking étale $\phi, \phi' : \tilde{Y} \to Y$ as in Theorem 4.3.1, we have $\phi^*\mathcal{I}[2] = \phi'^*\mathcal{I}[2]'$, where $\mathcal{I}[2]'$ is defined using x_1'. By induction $a_2, \dots, a_k$ are independent of choices. Hence $(a_1, \dots, a_k)$ is independent of choices.

(2) Since the closed subscheme $V(\mathcal{D}^{\leq a-1}\mathcal{I})$ is the locus where $\mathrm{ord}_p(\mathcal{I}) \geq a$, the order is constructible and upper-semicontinuous. The subscheme $V(\mathcal{D}^{\leq a_1 - 1}\mathcal{I})$ is contained in $V(x_1)$ on which $\mathrm{inv}_p(\mathcal{I}[2])$ is constructible and upper-semicontinuous by induction, hence $\mathrm{inv}_p(\mathcal{I})$ is constructible and upper-semicontinuous. □

Remark 5.1.4. Theorem 5.1.3(3) allows to extend the definition of inv to the case of smooth stacks Y. Indeed, if $\mathcal{I}$ is an ideal, choose a smooth presentation $p_{1,2} : Y_1 \rightrightarrows Y_0$ of Y and let $\mathcal{I}_i \subseteq \mathcal{O}_{Y_i}$ be the pullbacks of $\mathcal{I}$. Then $\mathrm{inv}_{\mathcal{I}_1} = \mathrm{inv}_{\mathcal{I}_0} \circ p_i$ for $i = 1, 2$, hence $\mathrm{inv}_{\mathcal{I}_0}$ factors through $Y_0 \to |Y|$ uniquely. A similar argument shows that the induced map $\mathrm{inv}_{\mathcal{I}} : |Y| \to \Xi_n$ is independent of the presentation.

Concerning the independence of J, we note the following consequence of Theorem 4.3.1:

Lemma 5.1.5. *If x_1' is another maximal contact element such that $(x_1', x_2, \ldots, x_n)$ is a regular sequence of parameters at p, then $J' = (x_1'^{a_1}, x_2^{a_2}, \ldots, x_k^{a_k})$ is also a center associated to $\mathcal{I}$ at p.*

Proof. As above, this follows since $\phi^*\mathcal{I}[2] = \phi'^*\mathcal{I}[2]'$, where $\mathcal{I}[2]'$ is defined using x_1'. □

5.2. *Admissibility of centers.* As in earlier work on resolution of singularities, *admissibility* allows flexibility in studying the behavior of ideals under blowings up of centers. This becomes important when an ideal is related to the sum of ideals with different invariants of their own, but all admitting a common admissible center.

In this section we assume that a_1 is a positive integer and $a_i \le a_{i+1}$. We deliberately do not assume $(a_1, \ldots, a_k)$ is $\operatorname{inv}_p(\mathcal{I})$ — see Remark 5.3.2.

5.2.1. *Admissibility and blowing up.* Recall that by Definition 2.4.1(2), a center $J = (x_1^{a_1}, \ldots, x_k^{a_k})$ is $\mathcal{I}$-admissible at p if the inequality $(x_1^{a_1}, \ldots, x_k^{a_k}) \le v(\mathcal{I})$ of valuative $\mathbb{Q}$-ideals is satisfied on a neighborhood of p.

Very much in analogy to the notion used in earlier resolution algorithms, this can be described in terms of the associated weighted blowing up $Y' = \mathrm{Bl}_{\bar{J}}(Y) \to Y$ along $\bar{J} := (x_1^{1/w_1}, \ldots, x_k^{1/w_k})$ as follows: let $E = \bar{J}\mathcal{O}_{Y'}$, which is an invertible ideal sheaf. Note that since $a_1 w_1$ is an integer also $J\mathcal{O}_{Y'} = E^{a_1 w_1}$ is an invertible ideal sheaf. Therefore $J = (x_1^{a_1}, \ldots, x_k^{a_k})$ is $\mathcal{I}$-admissible if and only if $E^{a_1 w_1}$ is $\mathcal{I}\mathcal{O}_{Y'}$ admissible, if and only if $\mathcal{I}\mathcal{O}_{Y'} = E^{a_1 w_1}\mathcal{I}'$, with $\mathcal{I}'$ an ideal.

Definition 5.2.2. In the situation as above, $\mathcal{I}'$ is called the *weak transform* of $\mathcal{I}$ under the weighted blowing up.

We will only use this operation when J is the center associated to $\mathcal{I}$, which is shown to be $\mathcal{I}$-admissible below.

Remark 5.2.3. In terms of its monomial valuation, J is admissible for $\mathcal{I}$ if and only if $v_J(f) \ge 1$ for all $f \in \mathcal{I}$. This means that if $f = \sum c_{\bar{\alpha}} x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ then $\sum_{i=1}^k \alpha_i / a_i \ge 1$ whenever $c_{\bar{\alpha}} \ne 0$. This is convenient for testing admissibility, as long as one remembers that $v_{J^m} = v_J / m$.

If $Y_1 \to Y$ is smooth and J is $\mathcal{I}$-admissible then $J\mathcal{O}_{Y_1}$ is $\mathcal{I}\mathcal{O}_{Y_1}$-admissible, with the converse holding when $Y_1 \to Y$ is surjective.

5.2.4. *Working with rescaled centers.* For induction to work in the arguments below, it is worthwhile to consider blowings up of centers of the form

$$\bar{J}^{1/c} := (x_1^{1/(w_1 c)}, \ldots, x_k^{1/(w_k c)})$$

for a positive integer c. We also use the notation $J^\alpha := (x_1^{a_1 \alpha}, \ldots, x_k^{a_k \alpha})$ throughout — this being an equality of valuative $\mathbb{Q}$-ideals.

5.2.5. *Basic properties.* The description in Section 5.2.1 of the monomial valuation of J immediately provides the following lemmas:

Lemma 5.2.6. *If J is both $\mathcal{I}_1$-admissible and $\mathcal{I}_2$-admissible then J is $\mathcal{I}_1+\mathcal{I}_2$-admissible. If J is $\mathcal{I}$-admissible then J^k is $\mathcal{I}^k$-admissible. More generally if J^{c_j} is $\mathcal{I}_j$-admissible then $J^{\sum c_j}$ is $\prod \mathcal{I}_j$-admissible.*

Indeed if $v_J(f) \geq 1$ and $v_J(g) \geq 1$ then $v_J(f+g) \geq 1$ and $v_J(f^{c_1} \cdot g^{c_2}) \geq c_1 + c_2$, etc.

Lemma 5.2.7. *If J is $\mathcal{I}$-admissible then $J' = J^{(a_1-1)/a_1}$ is $\mathcal{D}(\mathcal{I})$-admissible. If $a_1 > 1$ and $J^{(a_1-1)/a_1}$ is $\mathcal{I}$-admissible then J is $x_1\mathcal{I}$-admissible.*

Proof. For the first statement note that if $\sum_{i=1}^k \alpha_i/a_i \geq 1$ and $\alpha_j \geq 1$ then

$$v_J\left(\frac{\partial(x_1^{\alpha_1}\cdots x_n^{\alpha_n})}{\partial x_j}\right) = \sum_{i=1}^k \alpha_i/a_i - 1/a_j \geq 1 - 1/a_1,$$

so

$$v_{J'}\left(\frac{\partial(x_1^{\alpha_1}\cdots x_n^{\alpha_n})}{\partial x_j}\right) \geq 1,$$

as needed. The other statement is similar. □

As in Section 4.4 by $k_p = k(p)$ we denote a fixed field of coefficients.

Lemma 5.2.8. *For $\mathcal{I}_0 \subset k_p[\![x_2,\dots,x_n]\!]$ write $\tilde{\mathcal{I}}_0 = \mathcal{I}_0 k_p[\![x_1,\dots,x_n]\!]$. Assume $(x_2^{a_2},\dots,x_k^{a_k})$ is $\mathcal{I}_0$-admissible. Then $(x_1^{a_1},\dots,x_k^{a_k})$ is $\tilde{\mathcal{I}}_0$-admissible.*

Proof. Here for generators of $\tilde{\mathcal{I}}_0$ we have $\sum_{i=1}^k \alpha_i/a_i = \sum_{i=2}^k \alpha_i/a_i$. □

Lemma 5.2.9. *J is $\mathcal{I}$-admissible if and only if $J^{(a_1-1)!}$ is $C(\mathcal{I}, a_1)$-admissible.*

Proof. When $\mathcal{I}$ has order $< a_1$ then J is not admissible for $\mathcal{I}$ and $J^{(a_1-1)!}$ is not admissible for $\mathcal{C}(\mathcal{I}, a_1) = (1)$. When $\mathcal{I}$ has order $\geq a_1$ this combines Lemmas 5.2.6 and 5.2.7 for the terms defining $C(\mathcal{I}, a_1)$. □

This statement is only relevant, and will only be used, when $\mathcal{I}$ has order a_1. If $a_1 < a := \mathrm{ord}(\mathcal{I})$ then $J^{(a_1-1)!}$ is in general not $C(\mathcal{I}, a)$-admissible. For instance $J = (x_1)$ is admissible for $\mathcal{I} = (x_1x_2)$ but not for $C(\mathcal{I}, 2) = (x_1^2, x_1x_2, x_2^2)$.

Lemma 5.2.10. *Assume $(x_1, x_2, \dots, x_n)$ and $(x_1', x_2, \dots, x_n)$ are both regular sequences of parameters, and suppose $(x_1^{a_1}, x_2^{a_2}, \dots, x_k^{a_k}) \leq v(x_1'^{a_1})$. Then $(x_1^{a_1}, x_2^{a_2}, \dots, x_k^{a_k}) = (x_1'^{a_1}, x_2^{a_2}, \dots, x_k^{a_k})$ as centers.*

Proof. We may rescale a_i and assume they are all integers. The inequality $(x_1^{a_1}, x_2^{a_2}, \dots, x_k^{a_k}) \leq v(x_1'^{a_1})$ implies that $x_1'^{a_1}$ lies in the integral closure $(x_1^{a_1}, x_2^{a_2}, \dots, x_k^{a_k})^{\mathrm{int}}$, hence

$$(x_1'^{a_1}, x_2^{a_2}, \dots, x_k^{a_k})^{\mathrm{int}} \subset (x_1^{a_1}, x_2^{a_2}, \dots, x_k^{a_k})^{\mathrm{int}}.$$

Since these two ideals have the same Hilbert–Samuel functions they coincide. □

5.3. ***Unique admissibility of $J_p(\mathcal{I})$.*** Finally, we prove the second main result of Section 5 in addition to Theorem 5.1.3.

Theorem 5.3.1. *Let Y be a smooth variety, $p \in Y$ a point, and $\mathcal{I} \subseteq \mathcal{O}_Y$ an ideal with* $\mathrm{inv}_{\mathcal{I}}(p) = (a_1, \dots, a_k)$:

(1) *If $x_1, \dots, x_k$ is a maximal contact sequence at p and $J = (x_1^{a_1}, \dots, x_k^{a_k})$ the corresponding center, then J is $\mathcal{I}$-admissible at p.*

(2) $$\mathrm{inv}_{\mathcal{I}}(p) = \max_{(x_1'^{b_1}, \dots, x_k'^{b_k}) \leq v(\mathcal{I})} (b_1, \dots, b_k),$$
in other words, $\mathrm{inv}_{\mathcal{I}}(p)$ *is the maximal invariant of a center admissible for $\mathcal{I}$.*

(3) *Locally at p, J is the unique admissible center with invariant* $\mathrm{inv}_{\mathcal{I}}(p)$*. In particular, it is in fact independent of the maximal contact sequence $(x_1, \dots, x_k)$.*

(4) *Locally at p, any point p' with* $\mathrm{inv}_{\mathcal{I}}(p') = \mathrm{inv}_{\mathcal{I}}(p)$ *lies in $V(J)$.*

Proof. We first prove (1). We can work on formal completions as the usual admissibility is equivalent to the formal one: J is dominated by $\mathcal{I}$ at p if and only if the completion $\hat{J} = J\hat{\mathcal{O}}_{Y,p}$ is dominated by $\hat{\mathcal{I}} = \mathcal{I}\hat{\mathcal{O}}_{Y,p}$. Applying Lemma 5.2.9, we replace $\mathcal{I}$ by $\mathcal{C} = C(\mathcal{I}, a_1)$ and rescale the invariant up to $a_1!$. Recall that by Proposition 4.4.1

$$\hat{\mathcal{C}} = (x_1^{a_1!}) + (x_1^{a_1!-1}\widetilde{\mathcal{C}}_1) + \cdots + (x_1\widetilde{\mathcal{C}}_{a_1!-1}) + \widetilde{\mathcal{C}}_{a_1!}.$$

The inductive hypothesis implies that $\hat{J}^{(a_1-1)!}$ is $\overline{\mathcal{C}}_{a_1!}$-admissible. By Lemma 5.2.8 $\hat{J}^{(a_1-1)!}$ is $\widetilde{\mathcal{C}}_{a_1!}$-admissible. By Corollary 4.4.2 and Lemma 5.2.7 $\hat{J}^{(a_1-1)!}$ is $(x_1^{a_1!-j}\widetilde{\mathcal{C}}_j)$-admissible, so by Lemma 5.2.6 $\hat{J}^{(a_1-1)!}$ is $\hat{\mathcal{C}}$-admissible, as needed.

We prove (2) and (3) simultaneously. Assume $(b_1, \dots, b_m) \geq (a_1, \dots, a_k)$. If $J' = (x_1'^{b_1}, \dots, x_k'^{b_k})$ is admissible for $\mathcal{I}$ then $b_1 \leq a_1$. Since our chosen center J has $b_1 = a_1$ this maximum is achieved. Let $\ell = \max\{i : b_i = a_1\} \geq 1$. Evaluating $J' < v(\mathcal{I}) \leq v(x^{a_1})$ at the divisorial valuation of $x_1 = 0$ we have that $x_1 \in (x_1', \dots, x_\ell') + \mathfrak{m}_p^2$, and after reordering we get that $(x_1, x_2', \dots, x_n')$ is a regular system of parameters. By Lemma 5.2.10 we may write $J' = (x_1^{a_1}, x_2'^{b_2}, \dots, x_k'^{b_k})$. Working on formal completions we may replace x_i' by a suitable $x_i' + \alpha x_1$ so we may assume $x_i' \in k_p[\![x_2, \dots, x_n]\!]$.

As in the proof of (1) above, we may replace $\mathcal{I}$ and $\mathrm{inv}_{\mathcal{I}}(p)$ by the coefficient ideal $\mathcal{C} = C(\mathcal{I}, a_1)$ and the rescaled invariant $(a_1 - 1)!(a_1, \dots, a_k)$, and for the formal completions one has

$$\hat{\mathcal{C}} = (x_1^{a_1!}) + (x_1^{a_1!-1}\widetilde{\mathcal{C}}_1) + \cdots + (x_1\widetilde{\mathcal{C}}_{a_1!-1}) + \widetilde{\mathcal{C}}_{a_1!}.$$

By induction $(a_1 - 1)!(a_2, \dots, a_k)$ is the maximal invariant for $\overline{\mathcal{C}}_{a_1!}$, with unique center $(x_2^{a_2}, \dots, x_k^{a_k})$. By functoriality, the invariant is maximal for $\widetilde{\mathcal{C}}_{a_1!}$. But $J' = (x_1^{a_1}, x_2'^{b_2}, \dots, x_k'^{b_k}) < v(\widetilde{\mathcal{C}}_{a_1!})$ is equivalent to $(x_2'^{b_2}, \dots, x_k'^{b_k}) < v(\widetilde{\mathcal{C}}_{a_1!})$. It follows that $(a_1 - 1)!(a_1, \dots, a_k)$ is the maximal invariant of a center admissible for $C(\mathcal{I}, a_1)$, with unique center J.

Finally, (4) follows from the same induction using the classical fact that a maximal contact to $(\mathcal{I}, a_1)$ contains all neighboring points p' with $\mathrm{ord}_{p'}(\mathcal{I}) = a_1$ (for example, see the proof of Theorem 5.1.3(2)). □

Remark 5.3.2. (1) Stated in terms of the monomial valuation v_J associated to J, the theorem says it is the unique monomial valuation with lexicographically *minimal* weights $(w_1, \dots, w_n)$ satisfying $v(\mathcal{I}) = 1$.

(2) As an example for the added flexibility provided by admissibility, the center (x_1^6, x_2^6) is $(x_1^3x_2^3)$-admissible because this is the corresponding invariant, but also $(x_1^5, x_2^{15/2})$ is admissible. This second center becomes important when one considers instead the ideal $(x_1^5+x_1^3x_2^3)$, or even $(x_1^5+x_1^3x_2^3+x_2^8)$, whose invariant is $\left(5, \frac{15}{2}\right)$, as described in Section 7 below.

Corollary 5.3.3. *We have* $\operatorname{inv}_{\mathcal{I}^k}(p) = k \cdot \operatorname{inv}_{\mathcal{I}}(p)$ *and* $\operatorname{inv}_{C(\mathcal{I},a_1)}(p) = (a_1-1)!\operatorname{inv}_{\mathcal{I}}(p)$ *when* $a_1 = \operatorname{ord}_p(\mathcal{I})$.

Proof. Indeed J^k is admissible for $\mathcal{I}^k$ if and only if J is admissible for $\mathcal{I}$, and Lemma 5.2.9 provides the analogous statement for the coefficient ideal. □

6. Principalization and resolution

6.1. ***The maximal center.*** Our local construction of centers $J_p(\mathcal{I})$ can be globalized as follows along the maximality locus of the invariant.

Theorem 6.1.1. (1) *For any smooth variety Y and an ideal $\mathcal{I} \subseteq \mathcal{O}_Y$ there exists a unique $\mathcal{I}$-admissible center $J = J(\mathcal{I})$ such that* $\operatorname{inv}_J = \max \operatorname{inv}_{\mathcal{I}}$ *and $p \in V(J)$ if and only if* $\operatorname{inv}_{\mathcal{I}}(p) = \max \operatorname{inv}_{\mathcal{I}}$.

(2) *Compatibility with smooth morphisms $f : Y' \to Y$: either $f^{-1}J(\mathcal{I}) = (1)$, or $f^{-1}J(\mathcal{I}) = J(\mathcal{I}')$, where $\mathcal{I}' = f^{-1}\mathcal{I}$.*

(3) *If Y is a smooth stack of finite type over a field of characteristic zero and $\mathcal{I}$ is an ideal on Y, then associating to each presentation $f : Y' \to Y$ the center $J(f^{-1}\mathcal{I})$ one obtains a center on Y, which will be denoted $J(\mathcal{I})$.*

Proof. Uniqueness in (1) follows from the local uniqueness in Theorem 5.3.1(3). Moreover, it implies that it suffices to establish the existence locally at p. If $\operatorname{inv}_{\mathcal{I}}(p) = \max \operatorname{inv}_{\mathcal{I}}$ then locally at p such a center is provided by Theorem 5.3.1, and otherwise the center is empty in a neighborhood of p.

Recall that the invariant is compatible with arbitrary smooth morphisms by Theorem 5.1.3(3). If $\operatorname{maxinv}(\mathcal{I}') < \operatorname{maxinv}(\mathcal{I})$, then the invariant at any $p' \in f(Y')$ is smaller than $\operatorname{maxinv}(\mathcal{I})$, and hence $V(J) \cap f(Y') = \varnothing$ and $f^{-1}(J) = (1)$. If $\operatorname{maxinv}(\mathcal{I}') = \operatorname{maxinv}(\mathcal{I})$, then the center $f^{-1}(J)$ satisfies the condition defining $J(\mathcal{I}')$. Since such a center is unique by (1), we obtain (2). Finally, (3) is a straightforward consequence of (2). □

Definition 6.1.2. The center $J = J(\mathcal{I})$ defined by Theorem 6.1.1 will be called *the maximal $\mathcal{I}$-admissible center*.

6.2. ***The invariant drops.*** The main miracle about the maximal $\mathcal{I}$-admissible center is that blowing it up one automatically reduces the invariant of the weak transform of $\mathcal{I}$ (see Definition 5.2.2). For inductive reasons we prefer to prove a slightly stronger claim:

Theorem 6.2.1. *Assume that Y is a smooth k-stack and $\mathcal{I} \neq (1)$ is a coherent ideal on Y, and $c > 0$ a natural number. Consider the blowing up $f_c : Y'_c = \mathrm{Bl}_{\bar{J}^{1/c}}(Y)$ of the rescaled reduction $\bar{J}^{1/c}$ of the maximal $\mathcal{I}$-admissible center $J = J(\mathcal{I})$, and let $\mathcal{I}' = E^{-a_1 w_1 c} f_c^{-1}\mathcal{I}$ be the weak transform of $\mathcal{I}$, where* $\mathrm{maxinv}(\mathcal{I}) = (a_1, \ldots, a_k)$ *and* $(w_1, \ldots, w_k)$ *are the corresponding weights. Then* $\mathrm{maxinv}(\mathcal{I}') < \mathrm{maxinv}(\mathcal{I})$.

Proof. All players in the assertion are compatible with surjective smooth morphisms by Theorems 5.1.3(3) and 6.1.1(2), hence we can replace Y and $\mathcal{I}$ by an étale cover and the pullback of $\mathcal{I}$. Thus, we can assume that Y is a scheme and it suffices to prove that if $p \in Y$ satisfies $\mathrm{inv}_{\mathcal{I}}(p) = (a_1, \ldots, a_k)$, then any $p' \in f_c^{-1}(p)$ satisfies $\mathrm{inv}_{\mathcal{I}'}(p') < (a_1, \ldots, a_k)$. In particular, working locally at p we can assume that $J = (x_1^{a_1}, \ldots, x_k^{a_k})$ for a maximal contact sequence $(x_1, \ldots, x_k)$, and hence $\bar{J}^{1/c} := (x_1^{1/(w_1 c)}, \ldots, x_k^{1/(w_k c)})$.

If $k = 0$ the ideal is (0) and there is nothing to prove. When $k = 1$ the ideal is $(x_1^{a_1})$, which becomes exceptional with weak transform $\mathcal{I}' = (1)$. We now assume $k > 1$.

Again using Proposition 4.4.1, we choose formal coordinates, work with $\widetilde{\mathcal{C}} := \hat{C}(\mathcal{I}, a_1)$, and write

$$\widetilde{\mathcal{C}} = (x_1^{a_1!}) + (x_1^{a_1!-1}\widetilde{\mathcal{C}}_1) + \cdots + (x_1\widetilde{\mathcal{C}}_1) + \widetilde{\mathcal{C}}_{a_1!}.$$

Writing $\widetilde{\mathcal{C}}\mathcal{O}_{Y'_c} = E^{a_1! w_1 c}\widetilde{\mathcal{C}}'$, we will first show that $\mathrm{inv}_{p'}(\widetilde{\mathcal{C}}') < (a_1 - 1)! \cdot (a_1, a_2, \ldots, a_k)$ for all points p' over p.

Write $H = \{x_1 = 0\}$, and $H' \to H$ the blowing up of the reduced center $\bar{J}_H$ associated to $J_H := (x_2^{a_2}, \ldots, x_k^{a_k})$. By Lemma 3.4.1 the proper transform $\tilde{H}' \to H$ of H via the blowing up of $\bar{J}$ is the blowing up of $\bar{J}_H^{1/(cc')}$, allowing for induction.

We now inspect the behavior on different charts. On the x_1-chart we have $x_1 = u^{w_1 c}$ so the first term becomes $(x_1^{a_1!}) = E^{a_1! w_1 c} \cdot (1)$ and $\mathrm{inv}_{p'}\widetilde{\mathcal{C}}' = \mathrm{inv}(1) = 0$.[6] This implies that on all other charts it suffices to consider $p' \in \tilde{H}' \cap E$, as all other points belong to the x_1-chart. By the inductive assumption, for such points we have

$$\mathrm{inv}_{p'}((\bar{\mathcal{C}}_{a_1!})') < (a_1 - 1)! \cdot (a_2, \ldots, a_k).$$

Note that the term $(x_1^{a_1!})$ in $\widetilde{\mathcal{C}}$ is transformed, via $x_1 = u^{w_1 c} x'_1$ to the form $E^{a_1! w_1 c}(x_1'^{a_1!})$. It follows that $\mathrm{ord}_{p'}(\widetilde{\mathcal{C}}') \leq a_1!$, and if $\mathrm{ord}_{p'}(\widetilde{\mathcal{C}}') < a_1!$ then a fortiori $\mathrm{inv}_{p'}(\widetilde{\mathcal{C}}') < \mathrm{inv}_p(\widetilde{\mathcal{C}})$.

If on the other hand $\mathrm{ord}_{p'}(\widetilde{\mathcal{C}}') = a_1!$ then the variable x'_1 is a maximal contact element. Using the inductive assumption we compute

$$\mathrm{inv}_{p'}((x_1'^{a_1!}) + (\widetilde{\mathcal{C}}_{a_1!})') = (a_1!, \mathrm{inv}_{p'}((\bar{\mathcal{C}}_{a_1!})')) < (a_1!, \mathrm{inv}_{p'}(\bar{\mathcal{C}}_{a_1!})) = (a_1 - 1)!(a_1, \ldots, a_k).$$

Since $\widetilde{\mathcal{C}}'$ includes this ideal, we obtain again $\mathrm{inv}_{p'}(\widetilde{\mathcal{C}}') < \mathrm{inv}_p(\widetilde{\mathcal{C}})$, as claimed.

We deduce that $\mathrm{inv}_{p'}(\mathcal{I}') < \mathrm{inv}_p(\mathcal{I})$ as well. As in [Kollár 2007, Theorem 3.67; Bierstone and Milman 2008, Lemma 3.3; Abramovich et al. 2020a; 2020b], we have the inclusions $\mathcal{I}'^{(a_1-1)!} \subset \widetilde{\mathcal{C}}' \subset \hat{C}(\mathcal{I}', a_1)$,[7]

[6] This reflects the fact that before passing to the coefficient ideal $\mathrm{ord}(\mathcal{I}') < a_1$ on this chart — it need not become a unit ideal in general!

[7] These are the "easy" inclusions — which hold even in the logarithmic situation.

hence $\mathrm{ord}_{p'}(\mathcal{I}') \leq a_1$. We may again assume x_1' is a maximal contact element and $\mathrm{ord}_{p'}(\mathcal{I}') = a_1$. By Theorem 5.3.1(2)

$$\mathrm{inv}_{p'}(\mathcal{I}'^{(a_1-1)!}) \geq \mathrm{inv}_{p'}(\widetilde{\mathcal{C}}') \geq \mathrm{inv}_{p'}(\hat{C}(\mathcal{I}', a_1)).$$

By Corollary 5.3.3 we have $\mathrm{inv}_{p'}(\mathcal{I}'^{(a_1-1)!}) = \mathrm{inv}_{p'}(\hat{C}(\mathcal{I}', a_1))$ giving equalities throughout, hence

$$\mathrm{inv}_{p'}(\mathcal{I}') = \frac{1}{(a_1-1)!}\mathrm{inv}_{p'}(\widetilde{\mathcal{C}}') < \frac{1}{(a_1-1)!}\mathrm{inv}_{p}(\widetilde{\mathcal{C}}) = \mathrm{inv}_p(\mathcal{I}),$$

as needed. □

6.3. *The principalization theorem.* It remains to summarize our results. First, we obtain principalization. Given a pair $(Y, \mathcal{I})$ consisting of a smooth Deligne–Mumford k-stack Y and an ideal $\mathcal{I} \subset \mathcal{O}_{Y_{\text{ét}}}$, let $J = J(\mathcal{I})$ be the maximal $\mathcal{I}$-admissible center with reduction $\bar{J}$, let $Y_1 = \mathrm{Bl}\,\bar{J}(Y)$ and let $\mathcal{I}_1$ be the weak transform of $\mathcal{I}$. We set $\mathcal{P}_1(Y, \mathcal{I}) = (Y_1, \mathcal{I}_1)$.

Theorem 6.3.1 (principalization). (1) *Partial principalization*:

- (a) $\mathcal{P}_1$ *reduces the invariant*: $\mathrm{maxinv}(\mathcal{I}_1) < \mathrm{maxinv}(\mathcal{I})$.
- (b) *If $f : Y' \to Y$ is smooth and $\mathcal{I}' = f^{-1}\mathcal{I}$, then either $\mathcal{P}_1(Y, \mathcal{I})$ pullbacks to the empty blowing up of Y', or $\mathcal{P}_1(Y', f^{-1}\mathcal{I}) = \mathcal{P}_1(Y, \mathcal{I}) \times_Y Y'$. In particular, $\mathcal{P}_1$ is compatible with surjective smooth morphisms.*

(2) *Full principalization*:

- (a) *The sequence $(Y_{i+1}, \mathcal{I}_{i+1}) = \mathcal{P}_1(Y_i, \mathcal{I}_i)$ starting with $(Y_0, \mathcal{I}_0) = (Y, \mathcal{I})$ stabilizes and this happens at the smallest n with $\mathcal{I}_n = (1)$.*
- (b) *The principalization blowings up sequence $\mathcal{P}(Y, \mathcal{I}) : Y_n \to \cdots \to Y$ is compatible with arbitrary smooth morphisms $f : Y' \to Y$: the sequence $\mathcal{P}(Y', f^{-1}\mathcal{I}')$ is obtained from $\mathcal{P}(Y, \mathcal{I}) \times_Y Y'$ by removing all empty blowings up.*

Proof. Claim (1) is covered by Theorems 6.2.1 and 6.1.1(2). Since the set of invariants Ξ_n is well-ordered and attains its minimum (0) on the trivial ideal (1), we obtain (2a). The functoriality of $\mathcal{P}$ follows from the functoriality of $\mathcal{P}_1$. □

As a corollary we deduce embedded resolution. This is a standard reduction, except the fact that here we are also able to replace the weak transform by the proper transform.

Proof of Theorems 1.2.2 and 1.2.5. Given a DM pair $X \subset Y$ consider the ideal $\mathcal{I} = \mathcal{I}_X$ defining X in Y, and set $\mathrm{inv}_{(X,Y)} = \mathrm{inv}_{\mathcal{I}}$ and $J(X, Y) = J(\mathcal{I})$. Thus, we take $Y_1 \to Y$ to be the same weighted blowing up as in $\mathcal{P}_1(Y, \mathcal{I})$ and take X_1 to be the proper transform of X. Since $\mathcal{I}_1 \subseteq \mathcal{I}_{X_1}$ we obtain that $\mathrm{maxinv}(X_1, Y_1) \leq \mathrm{maxinv}(\mathcal{I}_1)$. Therefore all assertions of the two theorems follow from the properties of the center $\mathcal{J}(\mathcal{I})$ and the invariant $\mathrm{inv}_{\mathcal{I}}$ proven in Theorems 6.1.1 and 6.3.1. □

Note that in the deduction of Corollary 1.2.3 we also used that if X is of codimension c at $p \in Y$, then $\mathrm{inv}_{\mathcal{I}}(p) \geq (1, \ldots, 1)$ of length c, and the equality holds if and only if X is smooth at p.

7. An example

Consider the plane curve

$$X = V(x^5 + x^3y^3 + y^k)$$

with $k \geq 5$. Its resolution depends on whether or not $k \geq 8$.

7.1. ***The case k ≥ 8.*** This curve is singular at the origin p. We have $a_1 = \mathrm{ord}_p(\mathcal{I}_X) = 5$. Since $\mathcal{D}^{\leq 4}\mathcal{I} = (x, y^2)$ we may take $x_1 = x$ and $H = V(x)$. A direct computation provides the coefficient ideal

$$C(\mathcal{I}_X, 5)|_H = (\mathcal{D}^{\leq 3}(\mathcal{I}_X)|_H)^{120/2} = (y^{180}),$$

with $b_2 = 180$ and $a_2 = \frac{180}{4!} = \frac{15}{2}$. Rescaling, we need to take the weighted blowup of $\bar{J} = (x^{1/3}, y^{1/2})$:

- In the x-chart we have $x = u^3$, $y = u^2y'$, giving

  $$Y'_x = [\mathrm{Spec}\, k[u, y']/\boldsymbol{\mu}_3],$$

  the action given by $(u, y') \mapsto (\zeta_3 u, \zeta_3 y')$. The equation of X becomes

  $$u^{15}(1 + y'^3 + u^{2k-15}y'^k),$$

  with proper transform $X'_x = V(1 + y'^3 + u^{2k-15}y'^k)$ smooth.
- In the y-chart we have $y = v^2$, $x = v^3x'$, giving

  $$Y'_y = [\mathrm{Spec}\, k[x', v]/\boldsymbol{\mu}_2],$$

  the action given by $(x', v) \mapsto (-x', -v)$. The equation of X becomes $v^{15}(x'^5 + x'^3 + v^{2k-15})$, with proper transform $X'_y = V(x'^5 + x'^3 + v^{2k-15})$.

  Note that X'_y is smooth when $k = 8$. Otherwise it is singular at the origin with invariant $(3, 2k-15)$, which is lexicographically strictly smaller than $\left(5, \frac{15}{2}\right)$; A single weighted blowing up resolves the singularity.

7.2. ***The case k ≤ 7.*** Consider now the same equation with $k = 7$ (the cases $k = 5, 6$ being similar). We still take $a_1 = 5$, $x_1 = x$ and $H = V(x)$. This time

$$C(\mathcal{I}_X)|_H = ((\mathcal{I}_X)|_H)^{120/5} = (y^{168}),$$

with $b_2 = 7 \cdot (4!)$ and $a_2 = 7$. We take the weighted blowup of $J = (x^{1/7}, y^{1/5})$:

- In the x-chart we have $x = u^7$, $y = u^5y'$, giving

  $$Y'_x = [\mathrm{Spec}\, k[u, y']/\boldsymbol{\mu}_7],$$

  the action given by $(u, y') \mapsto (\zeta_7 u, \zeta_7^{-5} y')$. The equation of X becomes

  $$u^{35}(1 + uy'^3 + y'^7),$$

  with proper transform $X'_x = V(1 + uy'^3 + y'^7)$ smooth.

- In the y-chart we have $y = v^5$, $x = v^7 x'$, giving

$$Y'_y = [\operatorname{Spec} k[x', v]/\boldsymbol{\mu}_5],$$

the action given by $(x', v) \mapsto (\zeta_5^{-7} x', \zeta_5 v)$. The equation of X becomes $v^{35}(x'^5 + v x'^3 + 1)$, with smooth proper transform $X'_y = V(x'^5 + v x'^3 + 1)$.

8. Further comments

8.1. ***Nonembedded resolution.*** Given two embeddings $X \subset Y_1$ and $X \subset Y_2$ such that $\dim_p(Y_1) = \dim_p(Y_2)$ for all $p \in X$, the two embeddings are étale locally equivalent. By functoriality the embedded resolutions of $X \subset Y_1$ and $X \subset Y_2$ are étale locally isomorphic, hence the resolutions $X'_1 \to X$ and $X'_2 \to X$ coincide.

Our resolutions also satisfy the reembedding principle [Abramovich et al. 2020a, proposition 2.12.3]: given an embedding $Y \subset Y_1 := Y \times \operatorname{Spec} k[x_0]$ and $\operatorname{inv}_p(\mathcal{I}_{X \subset Y}) = (a_1, \ldots, a_k)$ with parameters $(x_1, \ldots, x_k)$ we have $\operatorname{inv}_p(\mathcal{I}_{X \subset Y_1}) = (1, a_1, \ldots, a_k)$ with parameters $(x_0, x_1, \ldots, x_k)$. The proper transform X'_1 of X in Y'_1 is disjoint from the x_0-chart, and on every other chart we have $Y'_1 = Y' \times \operatorname{Spec} k[x_0]$ so that $X'_1 = X'$ and induction applies.

Since every pure-dimensional stack can be étale locally embedded in pure codimension, we deduce:

Theorem 8.1.1 (nonembedded resolution). *There is a functor F_{ner} associating to a pure-dimensional reduced stack X of finite type over a characteristic-0 field k a proper, generically representable and birational morphism $F_{\mathrm{ner}}(X) \to X$ with $F_{\mathrm{ner}}(X)$ regular. This is functorial for smooth morphisms: if $X_1 \to X$ is smooth then $F_{\mathrm{ner}}(X_1) = F_{\mathrm{ner}}(X) \times_X X_1$.*

Remark 8.1.2. Of course one can deduce functorial resolution of X which is not pure dimensional just by applying F_{ner} to the normalization of X. One can also use other operations to separate components, for example, the disjoint union of the schematic closures of the generic points of X does the job.

Carefully using Bergh's destackification theorem we also obtain:

Theorem 8.1.3 (coarse resolution). *There is a construction F_{crs} associating to a pure-dimensional reduced stack X of finite type over a characteristic-0 field k a **projective** birational morphism $F_{\mathrm{crs}}(X) \to X$ with $F_{\mathrm{crs}}(X)$ smooth. This is functorial for smooth representable morphisms $X_1 \to X$, namely, $F_{\mathrm{crs}}(X_1) = F_{\mathrm{crs}}(X) \times_X X_1$.*

Proof. We apply [Bergh and Rydh 2019, Theorem 7.1], using $F_{\mathrm{ner}}(X) \to X \to \operatorname{Spec} k$ for $X \to T \to S$ in that theorem. This provides a projective morphism $F_{\mathrm{ner}}(X)' \to F_{\mathrm{ner}}(X)$, functorial for smooth morphisms $X_1 \to X$, such that the relative coarse moduli space $F_{\mathrm{ner}}(X)' \to \underline{F_{\mathrm{ner}}(X)'} \to X$ is projective over X, and such that $F_{\mathrm{ner}}(X)'$ and $\underline{F_{\mathrm{ner}}(X)'}$ are regular. We may take $F_{\mathrm{crs}}(X) = \underline{F_{\mathrm{ner}}(X)'}$. □

Remark 8.1.4. In general, $F_{\mathrm{crs}}(X)$ is only representable (even projective) over X, but not over k. This implies that when X is an algebraic space (or projective) so is $F_{\mathrm{crs}}(X)$. Of course one can replace in the construction relative destackification by absolute destackification. In such a case, the resulting resolution of the coarse moduli space $\underline{X}$ would be an algebraic space, but the construction would not be compatible with smooth morphisms.

8.2. ***Note on stabilizers.*** Even though Bergh's destackification is known for tame stacks, one might wonder about the stabilizers occurring in our resolution. We note, however, that the stabilizers of a weighted blowing up locally embed in $I_Y \times \mathbb{G}_m$, where I_Y denotes the inertia stack of Y. We therefore have that the stabilizers of Y_n locally embed in $I_Y \times \mathbb{G}_m^n$. In particular, if Y is a scheme then Y_n has abelian inertia, and its coarse moduli space has abelian quotient singularities.

8.3. ***Note on exceptional loci.*** We show by way of an example that the exceptional loci produced in our algorithm do not necessarily have normal crossings with centers.

Consider $\mathcal{I} = (x^2yz + yz^4) \subset \mathbb{C}[x, y, z]$. Then $\operatorname{maxinv}(\mathcal{I}) = (4, 4, 4)$ is attained at the origin with center (x^4, y^4, z^4) and reduced center (x, y, z). In the z-chart one obtains the ideal $(y_3(x_3^2 + z))$. The new invariant is $(2, 2)$ with reduced center $(y_3, x_3^2 + z)$, which is tangent to the exceptional $z = 0$.

The methods of [Abramovich et al. 2020a] suggest using the logarithmic derivative in z, resulting in the invariant $(3, 3, \infty)$ with center $(y_3^3, x_3^3, z^{3/2})$ and reduced Kummer center $(y_3, x_3, z^{1/2})$. This reduces logarithmic invariants respecting logarithmic, hence exceptional, divisors. A general algorithm is worked out in [Quek 2022].

Acknowledgements

We again mention that Theorem 1.2.2 was discovered independently by McQuillan [2020]. We thank Johannes Nicaise for bringing that to our attention. We thank Michael McQuillan for discussions of past, present and future projects and comparison of approaches (we concluded that the present key results are the same but methods quite different). We thank the referees for a very detailed reports that helped us to improve the paper. We also thank Jochen Heinloth, Milena Hering, János Kollár, Marc Levine, Ronald van Luijk, Ming Hao Quek, David Rydh, and Eugenii Shustin for questions, discussions, and illuminating both theory and existing literature.

References

[Abramovich and de Jong 1997] D. Abramovich and A. J. de Jong, "Smoothness, semistability, and toroidal geometry", *J. Algebraic Geom.* **6**:4 (1997), 789–801. MR Zbl

[Abramovich and Fantechi 2016] D. Abramovich and B. Fantechi, "Orbifold techniques in degeneration formulas", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **16**:2 (2016), 519–579. MR Zbl

[Abramovich and Hassett 2011] D. Abramovich and B. Hassett, "Stable varieties with a twist", pp. 1–38 in *Classification of algebraic varieties*, edited by C. Faber et al., Eur. Math. Soc., Zürich, 2011. MR Zbl

[Abramovich and Quek 2021] D. Abramovich and M. H. Quek, "Logarithmic resolution via multi-weighted blow-ups", preprint, 2021. arXiv 2112.06361

[Abramovich and Temkin 2019] D. Abramovich and M. Temkin, "Functorial factorization of birational maps for qe schemes in characteristic 0", *Algebra Number Theory* **13**:2 (2019), 379–424. MR Zbl

[Abramovich et al. 2002] D. Abramovich, K. Karu, K. Matsuki, and J. Włodarczyk, "Torification and factorization of birational maps", *J. Amer. Math. Soc.* **15**:3 (2002), 531–572. MR Zbl

[Abramovich et al. 2020a] D. Abramovich, M. Temkin, and J. Włodarczyk, "Principalization of ideals on toroidal orbifolds", *J. Eur. Math. Soc.* **22**:12 (2020), 3805–3866. MR Zbl

[Abramovich et al. 2020b] D. Abramovich, M. Temkin, and J. Włodarczyk, "Relative desingularization and principalization of ideals", preprint, 2020. arXiv 2003.03659

[Abramovich et al. 2020c] D. Abramovich, M. Temkin, and J. Włodarczyk, "Toroidal orbifolds, destackification, and Kummer blowings up", *Algebra Number Theory* **14**:8 (2020), 2001–2035. MR Zbl

[Abramovich et al. 2023] D. Abramovich, A. Frühbis-Krüger, M. Temkin, and J. Włodarczyk, *New techniques in resolution of singularities*, Oberwolfach Seminars **50**, Birkhäuser, Cham, 2023. MR Zbl

[Artal Bartolo et al. 2012] E. Artal Bartolo, J. Martín-Morales, and J. Ortigas-Galindo, "$\mathbb{Q}$-resolutions and intersection numbers", pp. 13–23 in *Eleventh International Conference Zaragoza-Pau on Applied Mathematics and Statistics* (Jaca, Spain, 2010), edited by J. Giacomoni et al., Monogr. Mat. García Galdeano **37**, Prensas Univ. Zaragoza, 2012. MR Zbl

[Artal Bartolo et al. 2014] E. Artal Bartolo, J. Martín-Morales, and J. Ortigas-Galindo, "Intersection theory on abelian-quotient V-surfaces and $\mathbb{Q}$-resolutions", *J. Singul.* **8** (2014), 11–30. MR Zbl

[Bergh 2017] D. Bergh, "Functorial destackification of tame stacks with abelian stabilisers", *Compos. Math.* **153**:6 (2017), 1257–1315. MR Zbl

[Bergh and Rydh 2019] D. Bergh and D. Rydh, "Functorial destackification and weak factorization of orbifolds", preprint, 2019. arXiv 1905.00872

[Bierstone and Milman 1997] E. Bierstone and P. D. Milman, "Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant", *Invent. Math.* **128**:2 (1997), 207–302. MR Zbl

[Bierstone and Milman 2008] E. Bierstone and P. D. Milman, "Functoriality in resolution of singularities", *Publ. Res. Inst. Math. Sci.* **44**:2 (2008), 609–639. MR Zbl

[Bogomolov 1992] F. A. Bogomolov, "Stable cohomology of groups and algebraic varieties", *Mat. Sb.* **183**:5 (1992), 3–28. In Russian; translated in *Russian Acad. Sci. Sb. Math.* **76**:1 (1993), 1–21. MR Zbl

[Borisov et al. 2005] L. A. Borisov, L. Chen, and G. G. Smith, "The orbifold Chow ring of toric Deligne–Mumford stacks", *J. Amer. Math. Soc.* **18**:1 (2005), 193–215. MR Zbl

[Decker et al. 2019] W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann, "Singular, a computer algebra system for polynomial computations", 2019, available at http://www.singular.uni-kl.de.

[Encinas and Hauser 2002] S. Encinas and H. Hauser, "Strong resolution of singularities in characteristic zero", *Comment. Math. Helv.* **77**:4 (2002), 821–845. MR Zbl

[Encinas and Villamayor 2003] S. Encinas and O. Villamayor, "A new proof of desingularization over fields of characteristic zero", *Rev. Mat. Iberoam.* **19**:2 (2003), 339–353. MR Zbl

[Encinas and Villamayor 2007] S. Encinas and O. Villamayor, "Rees algebras and resolution of singularities", pp. 63–85 in *Proceedings of the XVIth Latin American Algebra Colloquium* (Colonia del Sacramento, Uruguay, 2005), edited by W. Ferrer Santos et al., Rev. Mat. Iberoam., Madrid, 2007. MR Zbl

[Fantechi et al. 2010] B. Fantechi, E. Mann, and F. Nironi, "Smooth toric Deligne–Mumford stacks", *J. Reine Angew. Math.* **648** (2010), 201–244. MR Zbl

[Geraschenko and Satriano 2015a] A. Geraschenko and M. Satriano, "Toric stacks, I: The theory of stacky fans", *Trans. Amer. Math. Soc.* **367**:2 (2015), 1033–1071. MR Zbl

[Geraschenko and Satriano 2015b] A. Geraschenko and M. Satriano, "Toric stacks, II: Intrinsic characterization of toric stacks", *Trans. Amer. Math. Soc.* **367**:2 (2015), 1073–1094. MR Zbl

[Gillam and Molcho 2015] W. D. Gillam and S. Molcho, "A theory of stacky fans", preprint, 2015. arXiv 1512.07586

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. **52**, Springer, 1977. MR Zbl

[Hironaka 1964a] H. Hironaka, "Resolution of singularities of an algebraic variety over a field of characteristic zero, I", *Ann. of Math.* (2) **79**:1 (1964), 109–203. MR Zbl

[Hironaka 1964b] H. Hironaka, "Resolution of singularities of an algebraic variety over a field of characteristic zero, II", *Ann. of Math.* (2) **79**:2 (1964), 205–326. MR Zbl

[Hironaka 1967] H. Hironaka, "Characteristic polyhedra of singularities", *J. Math. Kyoto Univ.* **7** (1967), 251–293. MR Zbl

[Hironaka 1977] H. Hironaka, "Idealistic exponents of singularity", pp. 52–125 in *Algebraic geometry* (Baltimore, MD, 1976), edited by J.-I. Igusa, Johns Hopkins Univ. Press, Baltimore, MD, 1977. MR Zbl

[Huber and Jörder 2014] A. Huber and C. Jörder, "Differential forms in the h-topology", *Algebr. Geom.* **1**:4 (2014), 449–478. MR Zbl

[Illusie and Temkin 2014] L. Illusie and M. Temkin, "Gabber's modification theorem (absolute case)", pp. 103–160 in *Travaux de Gabber sur l'uniformisation locale et la cohomologie étale des schémas quasi-excellents*, edited by L. Illusie et al., Astérisque **363-364**, Soc. Math. France, Paris, 2014. MR Zbl

[de Jong 1997] A. J. de Jong, "Families of curves and alterations", *Ann. Inst. Fourier* (*Grenoble*) **47**:2 (1997), 599–621. MR Zbl

[Kawamata 1992] Y. Kawamata, "The minimal discrepancy of a 3-fold terminal singularity", (1992). Appendix to V. V. Shokurov, "Three-dimensional log perestroikas", *Izv. Ross. Akad. Nauk Ser. Mat.* **56**:1 (1992), 105–203. MR Zbl

[Kollár 2007] J. Kollár, *Lectures on resolution of singularities*, Ann. of Math. Stud. **166**, Princeton Univ. Press, 2007. MR Zbl

[Kollár et al. 2004] J. Kollár, K. E. Smith, and A. Corti, *Rational and nearly rational varieties*, Cambridge Stud. Adv. Math. **92**, Cambridge Univ. Press, 2004. MR Zbl

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, I: Classical setting: line bundles and linear series*, Ergebnisse der Math. (3) **48**, Springer, 2004. MR Zbl

[Lee et al. 2020] J. Lee, A. Frühbis-Krüger, and D. Abramovich, "resweighted.lib, a Singular library for weighted resolution", 2020, available at https://www.math.brown.edu/dabramov/PAPERS/01code/resweighted.lib.

[Martín-Morales 2013] J. Martín-Morales, "Monodromy zeta function formula for embedded $\mathbb{Q}$-resolutions", *Rev. Mat. Iberoam.* **29**:3 (2013), 939–967. MR Zbl

[Martín-Morales 2014] J. Martín-Morales, "Embedded $\mathbb{Q}$-resolutions for Yomdin–Lê surface singularities", *Israel J. Math.* **204**:1 (2014), 97–143. MR Zbl

[McQuillan 2020] M. McQuillan, "Very functorial, very fast, and very easy resolution of singularities", *Geom. Funct. Anal.* **30**:3 (2020), 858–909. MR Zbl

[McQuillan and Panazzolo 2013] M. McQuillan and D. Panazzolo, "Almost étale resolution of foliations", *J. Differential Geom.* **95**:2 (2013), 279–319. MR Zbl

[Olsson 2016] M. Olsson, *Algebraic spaces and stacks*, Amer. Math. Soc. Colloq. Publ. **62**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Panazzolo 2006] D. Panazzolo, "Resolution of singularities of real-analytic vector fields in dimension three", *Acta Math.* **197**:2 (2006), 167–289. MR Zbl

[Quek 2022] M. H. Quek, "Logarithmic resolution via weighted toroidal blow-ups", *Algebr. Geom.* **9**:3 (2022), 311–363. MR Zbl

[Quek and Rydh 2021] M. H. Quek and D. Rydh, "Weighted blow-ups", preprint, 2021, available at https://people.kth.se/~dary/weighted-blowups20220329.pdf.

[Reid 1980] M. Reid, "Canonical 3-folds", pp. 273–310 in *Journées de Géometrie Algébrique d'Angers* (Angers, France, 1979), edited by A. Beauville, Sijthoff & Noordhoff, Germantown, MD, 1980. MR Zbl

[Reid 2002] M. Reid, "Graded rings and varieties in weighted projective space", preprint, 2002, available at https://homepages.warwick.ac.uk/~masda/surf/more/grad.pdf.

[Schwartz 1992] A. J. Schwartz, *Functorial smoothing of morphisms in equal characteristic* 0, Ph.D. thesis, Harvard University, 1992, available at https://www.proquest.com/docview/303970976.

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Temkin 2010] M. Temkin, "Stable modification of relative curves", *J. Algebraic Geom.* **19**:4 (2010), 603–677. MR Zbl

[Temkin 2012] M. Temkin, "Functorial desingularization of quasi-excellent schemes in characteristic zero: the nonembedded case", *Duke Math. J.* **161**:11 (2012), 2207–2254. MR Zbl

[Temkin 2023] M. Temkin, "Relative and logarithmic resolution of singularities", pp. 139–218 in *New techniques in resolution of singularities*, Oberwolfach Seminars **50**, Birkhäuser, Cham, 2023. MR Zbl

[Varčenko 1976] A. N. Varčenko, "Newton polyhedra and estimates of oscillatory integrals", *Funkcional. Anal. i Priložen.* **10**:3 (1976), 13–38. In Russian; translated in *Functional Anal. Appl.* **18**:3 (1976), 175–196. MR Zbl

[Villamayor 1989] O. Villamayor, "Constructiveness of Hironaka's resolution", *Ann. Sci. École Norm. Sup.* (4) **22**:1 (1989), 1–32. MR Zbl

[Voevodsky 1996] V. Voevodsky, "Homology of schemes", *Selecta Math.* (*N.S.*) **2**:1 (1996), 111–153. MR Zbl

[Włodarczyk 2003] J. Włodarczyk, "Toroidal varieties and the weak factorization theorem", *Invent. Math.* **154**:2 (2003), 223–331. MR Zbl

[Włodarczyk 2005] J. Włodarczyk, "Simple Hironaka resolution in characteristic zero", *J. Amer. Math. Soc.* **18**:4 (2005), 779–822. MR Zbl

[Włodarczyk 2022] J. Włodarczyk, "Functorial resolution except for toroidal locus: toroidal compactification", *Adv. Math.* **407** (2022), art. id. 108551. MR Zbl

[Włodarczyk 2023] J. Włodarczyk, "Weighted resolution of singularities: a Rees algebra approach", pp. 219–317 in *New techniques in resolution of singularities*, Oberwolfach Seminars **50**, Birkhäuser, Cham, 2023. MR

[Youssin 1990] B. Youssin, *Newton polyhedra without coordinates*, Mem. Amer. Math. Soc. **433**, Amer. Math. Soc., Providence, RI, 1990. MR Zbl

Communicated by János Kollár
Received 2023-03-15 Revised 2023-07-06 Accepted 2023-09-03

dan_abramovich@brown.edu *Department of Mathematics, Brown University, Providence, RI, United States*

temkin@math.huji.ac.il *Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel*

wlodar@purdue.edu *Department of Mathematics, Purdue University, West Lafayette, IN, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 18 No. 8 2024