Algebra & Number Theory

Volume 13
2019
No. 9

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2019 is US $385/year for the electronic version, and $590/year (+$60, 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. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.1983

Proof of a conjecture of Colliot-Thélène and a diophantine excision theorem

Jan Denef

We prove a conjecture of Colliot-Thélène that implies the Ax–Kochen Theorem on p-adic forms. We obtain it as an easy consequence of a diophantine excision theorem whose proof forms the body of the present paper.

1. Introduction

In this paper we prove the following conjecture from [Colliot-Thélène 2008].

1.1. ***Colliot-Thélène's conjecture.*** *Let $f : X \to Y$ be a dominant morphism of smooth proper geometrically integral varieties over a number field F, with geometrically integral generic fiber. Assume that for any nontrivial discrete valuation on the function field K of Y, with valuation ring $A \supset F$, there exists an integral regular A-scheme $\mathfrak{X}$, flat and proper over A, with generic fiber K-isomorphic to the generic fiber of f, and special fiber having an irreducible component of multiplicity* 1 *which is geometrically integral. Then the map $X(F_p) \to Y(F_p)$, induced by f, is surjective for almost all* (*nonarchimedean*) *places p of F.*

Here F_p denotes the p-adic completion of F, and with "almost all places of F" we mean "all but a finite number of places of F".

Actually we prove a stronger result, namely:

Main Theorem 1.2. *Let $f : X \to Y$ be a dominant morphism of smooth proper geometrically integral varieties over a number field F, with geometrically integral generic fiber. Assume for any modification $f' : X' \to Y'$ of f, with the same generic fiber as f, and X', Y' smooth over F, and for any prime divisor D' on Y', the following: the divisor $f'^*(D')$ on X' has an irreducible component C' with multiplicity* 1 *and geometrically integral generic fiber over D'* (*i.e., the morphism $C' \to D'$, induced by f', has geometrically integral generic fiber*). *Then the map $X(F_p) \to Y(F_p)$, induced by f, is surjective for almost all* (*nonarchimedean*) *places p of F.*

We say that f' is a *modification* of f if f' fits into a commutative square of morphisms of varieties, with vertical arrows f and f', and horizontal arrows birational proper morphisms $X' \to X$ and $Y' \to Y$; see Definition 2.1.

MSC2010: primary 14G20; secondary 11S05.
Keywords: Conjecture of Colliot-Thélène, Ax–Kochen Theorem, Diophantine equations, p-adic numbers, Forms in many variables, Diophantine geometry, Monomialization, Toroidalization of morphisms.

The conjecture of Colliot-Thélène is a direct consequence of Main Theorem 1.2, because any D' as in the theorem induces a discrete valuation on the function field of Y', which equals the function field of Y. Such a discrete valuation on the function field of Y is called a *divisorial valuation*. Moreover, if for this valuation there exists an A-scheme $\mathfrak{X}$ as in the conjecture, then the special fiber of any integral regular proper flat A-scheme $\mathfrak{X}'$, with the same generic fiber as $\mathfrak{X}$, has an irreducible component of multiplicity 1 which is geometrically integral. Indeed this is Proposition 3.9.(b) in Colliot-Thélène's lecture notes [2011]. Thus the hypotheses in the statement of the conjecture of Colliot-Thélène imply the hypotheses in the statement of Main Theorem 1.2, even if we restrict the assumption in the conjecture to divisorial valuations.

Note that Main Theorem 1.2 is substantially stronger than the conjecture of Colliot-Thélène, because it implies that the assumption in the conjecture is only required for divisorial discrete valuations on the function field of Y.

Colliot-Thélène [2008] proved the following: if $f : X \to Y$ is the universal family of all projective hypersurfaces of degree d in projective n-space over a number field F, with $n \geq d^2$, then f satisfies the hypotheses of the conjecture and hence also the hypotheses of Main Theorem 1.2. (A similar result also holds for complete intersections in projective space; see Theorem 2.2 of [Colliot-Thélène 2008]). Since our proof of Main Theorem 1.2 is purely algebraic geometric, this yields a new proof of the theorem of Ax and Kochen [1965] on p-adic forms, that does not rely on methods from mathematical logic. The theorem of Ax and Kochen states that for each $d \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for all primes $p > N$, each hypersurface of degree d in projective n-space over $\mathbb{Q}_p$, with $n \geq d^2$, has a $\mathbb{Q}_p$-rational point.

One of the motivations of Colliot-Thélène in formulating his conjecture was to obtain an algebraic geometric proof of the Ax–Kochen theorem that, unlike all previous ones, does not rely on methods from mathematical logic. At the same time, the author of the present paper also found another purely algebraic geometric proof of the Ax–Kochen theorem; see [Denef 2016]. Both proofs are based on the tameness theorem (see Section 4), which is proved in [Denef 2016] using the weak toroidalization theorem of Abramovich and Karu [2000] (extended to nonclosed fields [Abramovich et al. 2013]).

We prove Main Theorem 1.2 in Section 6, as an easy consequence of what we call a diophantine excision theorem. The proof of this Diophantine Excision Theorem 5.1 forms the body of the present paper and is contained in Section 5. It depends on the Tameness Theorem 4.1, which is treated in Section 4. Using mathematical logic one can give a simpler proof of Colliot-Thélène's conjecture (Section 1.1). However we don't see how to extend this to prove the stronger Main Theorem 1.2 or the Diophantine Excision Theorem 5.1. This alternative proof is given in Section 6.3. Preliminaries about modifications of morphisms and multiplicative residues are given in Sections 2 and 3.

A previous version of the present paper was posted on the ArXiv in 2011. In that version the diophantine excision theorem was called the diophantine purity theorem.

1.3. ***Terminology and notation.*** Throughout our paper, for ease of notation, we work with $\mathbb{Q}$ instead of an arbitrary number field. But all our results remain true replacing $\mathbb{Q}$ by any number field F, and the completions $\mathbb{Q}_p$ of $\mathbb{Q}$ by the nonarchimedean completions of F.

For any prime number p we denote the ring of p-adic integers by $\mathbb{Z}_p$, the field of p-adic numbers by $\mathbb{Q}_p$, and the field with p elements by $\mathbb{F}_p$. The p-adic valuation on $\mathbb{Q}_p$ is denoted by ord_p.

In the present paper, R will always denote a noetherian integral domain. A *variety over* R is an integral separated scheme of finite type over R. With a *morphism* of varieties over R we mean an R-morphism of schemes over R. A rational function x on a variety X over R is called *regular at* a point $P \in X$ if it belongs to the local ring $\mathcal{O}_{X,P}$ of X at P, and it is called *regular* if it is regular at each point of X.

Uniformizing parameters over R on a variety X over R, are regular rational functions on X that induce an étale morphism from X to an affine space over R.

A *reduced strict normal crossings divisor over* R on a smooth variety X over R is a closed subset D of X such that for any $P \in X$ there exist uniformizing parameters $x_1, \ldots, x_n$ over R on an open neighborhood of P, such that for any irreducible component C of D, containing P, there is an $i \in \{1, \ldots, n\}$ which generates the ideal of C in $\mathcal{O}_{X,P}$.

2. Modifications of morphisms

Definition 2.1. Let R be a noetherian integral domain, and X a variety over R. A *modification of* X is a proper birational morphism $X' \to X$ of varieties over R.

Let $f : X \to Y$ be a dominant morphism of varieties over R. A *modification of* f is a morphism $f' : X' \to Y'$ of varieties over R, which fits into a commutative diagram

$$\begin{array}{ccc} X' & \xrightarrow{\alpha} & X \\ {\scriptstyle f'}\downarrow & & \downarrow{\scriptstyle f} \\ Y' & \xrightarrow{\beta} & Y \end{array}$$

with α a modification of X, and β a modification of Y. This implies that f' is dominant. Clearly, if f is proper, then also f' is proper.

When $f : X \to Y$ is a dominant morphism of varieties over R, and $\beta : Y' \to Y$ is a modification of Y, then there exists a unique irreducible component X' of the fiber product $Y' \times_Y X$ that dominates X. Let f' and α be the restrictions to X' of the projections $Y' \times_Y X \to Y'$, and $Y' \times_Y X \to X$. Then α is a modification of X, and we call f' the *strict transform of* f *with respect to* β. Clearly f' is a modification of f. Such a modification is called a *strict modification of* f. Note that any strict modification of f' is also a strict modification of f.

2.2. ***Observations.*** (a) Let $f : X \to Y$ be a morphism of schemes of finite type over an excellent henselian discrete valuation ring R, with $Y \otimes_R \mathrm{Frac}(R)$ smooth, where $\mathrm{Frac}(R)$ denotes the fraction field of R. Let S be a closed subscheme of Y, containing no irreducible component of Y. If $Y(R) \setminus S(R) \subset f(X(R))$, then $Y(R) \subset f(X(R))$. Indeed this follows from Greenberg's theorem [1966], because $Y(R) \setminus S(R)$ is dense in $Y(R)$ with respect to the adic topology on $Y(R)$, since $Y \otimes_R \mathrm{Frac}(R)$ is smooth.

(b) Let $f : X \to Y$ be a dominant morphism of varieties over $\mathbb{Z}$, and let $f' : X' \to Y'$ be a strict modification of f. Assume that $Y \otimes \mathbb{Q}$ and $Y' \otimes \mathbb{Q}$ are smooth, and let p be a prime number. Then, the map $X(\mathbb{Z}_p) \to Y(\mathbb{Z}_p)$, induced by f, is surjective, if and only if the map $X'(\mathbb{Z}_p) \to Y'(\mathbb{Z}_p)$, induced by f', is surjective. This remains true when f' is a modification of f which is not strict, if we assume that also $X \otimes \mathbb{Q}$ is smooth. These claims follow directly from (a). Indeed, by (a) and the valuative criterion for properness, any modification of a variety V over $\mathbb{Z}$, with $V \otimes \mathbb{Q}$ smooth, induces a surjection on $\mathbb{Z}_p$-rational points.

Remark. We will often use (without mentioning) the following well known facts. Any morphism $f_0 : X_0 \to Y_0$ of varieties over $\mathbb{Q}$ has a model f over $\mathbb{Z}$. This means that f is a morphism $f : X \to Y$ of varieties over $\mathbb{Z}$ whose base change to $\mathbb{Q}$ is isomorphic to f_0. Combining this with Nagata's compactification theorem (see e.g., [Lütkebohmert 1993]), we see that we can choose f to be proper, when f_0 is proper. Two models of f_0 over $\mathbb{Z}$ become isomorphic after base change to $\mathbb{Z}[1/N]$, for some positive integer N. Hence, if f_0 is proper and f is a model of f_0 over $\mathbb{Z}$, then $f \otimes \mathbb{Z}[1/N]$ is proper for some $N \in \mathbb{N}$.

3. Multiplicative residues

Let R be a noetherian integral domain, and X a variety over R. Let A be any local R-algebra which is an integral domain. We denote by $\mathfrak{m}_A$ its maximal ideal, by $\mathrm{Frac}(A)$ its field of fractions, and by η_A the generic point of $\mathrm{Spec}(A)$. For any A-rational point $a \in X(A)$ on X we denote by $a \bmod \mathfrak{m}_A$ the $A/\mathfrak{m}_A$-rational point on X induced by a. For any $x \in \mathcal{O}_{X,a(\eta_A)}$ the pullback $a^*(x)$ of x to $\mathrm{Frac}(A)$ is denoted by $x(a) \in \mathrm{Frac}(A)$. Moreover, for $a, a' \in X(A)$ we write $a \equiv a' \bmod \mathfrak{m}_A$ to say that $a \bmod \mathfrak{m}_A = a' \bmod \mathfrak{m}_A$.

Definition 3.1. Let $z, z' \in \mathrm{Frac}(A)$. The elements z, z' have the *same multiplicative residue* if

$$z' \in z(1 + \mathfrak{m}_A).$$

Let $a, a' \in X(A)$ and let $x_1, \dots, x_r$ be rational functions on X. The points a, a' have the *same residues with respect to* $x_1, \dots, x_r$ if $a \equiv a' \bmod \mathfrak{m}_A$ and, for $i = 1, \dots, r$, the following two conditions hold:

(1) The rational function x_i is regular at $a(\eta_A)$ if and only if it is regular at $a'(\eta_A)$.

(2) $x_i(a), x_i(a') \in \mathrm{Frac}(A)$ have the same multiplicative residue if x_i is regular at both $a(\eta_A)$ and $a'(\eta_A)$.

The following lemma also appears in [Denef 2016].

Lemma 3.2. *Let X be an affine variety over R, and let $x_1, \dots, x_r$ be rational functions on X. Then there exist regular rational functions $x'_1, \dots, x'_{r'}$ on X such that for any local R-algebra A, which is an integral domain, and any $a, a' \in X(A)$ we have the following. The points a and a' have the same residues with respect to $x_1, \dots, x_r$ if they have the same residues with respect to $x'_1, \dots, x'_{r'}$.*

Proof. This is clear, by taking for $x'_1, \dots, x'_{r'}$ any finite list of regular rational functions on X which satisfies the following condition. For each $i \in \{1, \dots, r\}$ and each $P \in X$ with x_i regular at P, there are elements x'_j and x'_k in this list with $x_i = x'_j / x'_k$, and $x'_k(P) \neq 0$. Obviously, such a finite list exists if X is affine. $\square$

Lemma 3.3. *Let X be a variety over R, and let $z, x_1, \dots, x_r$ be regular rational functions on X. Assume that z can be written as a unit in $\Gamma(X, \mathcal{O}_X)$ times a monomial in the x_i. Then for any local R-algebra A, which is an integral domain, and any $a, a' \in X(A)$ we have the following. The points a and a' have the same residues with respect to z if they have the same residues with respect to $x_1, \dots, x_r$.*

Proof. Obvious, and left to the reader. □

Definition 3.4. Let X be a variety over $\mathbb{Z}$, and $x_1, \dots, x_r$ regular rational functions on X. Let p be a prime and let $z = (z_1, \dots, z_r) \in \mathbb{Z}_p^r$. We say that *the multiplicative residue of z is realizable with respect to $x_1, \dots, x_r$* if there exists $a \in X(\mathbb{Z}_p)$ such that $x_i(a)$ and z_i have the same multiplicative residue for each $i = 1, \dots, r$.

Definition 3.5. Let p be a prime. For any $w \in \mathbb{Q}_p$, the *angular component modulo p of w* is defined as

$$\overline{\mathrm{ac}}_p(w) := wp^{-\operatorname{ord}_p(w)} \mod p \in \mathbb{F}_p,$$

with the convention that $\overline{\mathrm{ac}}_p(0) := 0$.

Note that any $w, w' \in \mathbb{Q}_p$ have the same multiplicative residue if and only if they have the same p-adic valuation and the same angular component modulo p.

The following rather technical lemma will be used in the proof of the surjectivity criterion (Section 4.2). It is a direct consequence of the theorem of Pas [1989] on uniform p-adic quantifier elimination. The work of Pas is based on methods from mathematical logic. Below we give a purely algebraic geometric proof of this lemma which is based on embedded resolution of singularities.

Lemma 3.6. *Let X be a variety over $\mathbb{Z}$, and $x_1, \dots, x_r$ regular rational functions on X. There exists a finite partition of $\mathbb{N}^r$ such that for almost all primes p we have the following: "Let $z, z' \in \mathbb{Z}_p^r$. Assume that the p-adic valuations of z and z' are in a same stratum of the partition, and that $\overline{\mathrm{ac}}_p(z_i) = \overline{\mathrm{ac}}_p(z_i')$ for each i. Then the multiplicative residue of z is realizable with respect to $x_1, \dots, x_r$, if and only if the same holds for z'."*

Proof. Let D be the union of the zero loci of the regular rational functions $x_1, \dots, x_r$ on X, considered as a subset of X. Using embedded resolution of singularities of $D \otimes \mathbb{Q} \subset X \otimes \mathbb{Q}$, and induction on the dimension of $X \otimes \mathbb{Q}$, modifying X and inverting a finite number of primes, we may assume the following. The variety X is smooth over $\mathbb{Z}$, and D is a reduced strict normal crossings divisor over $\mathbb{Z}$ (in the sense of Section 1.3). This reduction is easily verified applying the valuative criterion of properness to the resolution morphism and using the induction hypothesis to take care of the exceptional locus of the resolution. More precisely, the induction hypothesis is applied to each component of the image in X of the exceptional locus, with each x_i replaced by its restriction to that component. Hence, by covering X with finitely many suitable open subschemes, we can further assume that X is affine, and that each x_i can be written as a unit u_i in $\Gamma(X, \mathcal{O}_X)$ times a monomial in uniformizing parameters $y_1, \dots, y_n$ over $\mathbb{Z}$ on X. As recalled in Section 1.3, this means that $y_1, \dots, y_n$ induce an étale morphism from X to affine n-space over $\mathbb{Z}$.

Let E be the matrix over $\mathbb{Z}$, with r rows and n columns, consisting of the exponents of these monomials, and let Δ be the linear map $\mathbb{Z}^n \to \mathbb{Z}^r$ determined by the matrix E. For each subset $S \subset \{1, \dots, n\}$, set

$$\Gamma_S := \{(\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n \mid \forall j : \alpha_j = 0 \Leftrightarrow j \in S\}.$$

Choose a finite partition of $\mathbb{N}^r$ such that each $\Delta(\Gamma_S)$ is a union of strata.

Let $z, z' \in \mathbb{Z}_p^r$ be as in the lemma and assume that the multiplicative residue of z is realizable with respect to $x_1, \dots, x_r$ by an element $a \in X(\mathbb{Z}_p)$. We have to show that the multiplicative residue of z' is also realizable. Note that each z_i is nonzero (because the p-adic valuation of z belongs to $\mathbb{N}^r$). Hence the multiplicative residue of z is realized by any element of $X(\mathbb{Z}_p)$ which is close enough to a. Thus we may suppose that $y_j(a) \neq 0$ for all $j = 1, \dots, n$. Let $S \subset \{1, \dots, n\}$ be such that $(\operatorname{ord}_p y_1(a), \dots, \operatorname{ord}_p y_n(a)) \in \Gamma_S$. Hence $\operatorname{ord}_p z = (\operatorname{ord}_p x_1(a), \dots, \operatorname{ord}_p x_r(a)) \in \Delta(\Gamma_S)$, since $u_i(a)$ is a unit in $\mathbb{Z}_p$ for each i. Because the p-adic valuations of z and z' are in a same stratum of the partition, also $\operatorname{ord}_p z'$ is an element of $\Delta(\Gamma_S)$. Hence there exists $\alpha' = (\alpha'_1, \dots, \alpha'_n) \in \Gamma_S$ with $\Delta(\alpha') = \operatorname{ord}_p(z')$. Note that $\alpha'_j = 0$ if and only if $\operatorname{ord}_p(y_j(a)) = 0$, because $\alpha' \in \Gamma_S$.

By Hensel's lemma, applied to the étale morphism induced by the $y_1, \dots, y_n$, there exists $a' \in X(\mathbb{Z}_p)$ with $a' \mod p = a \mod p$ and $\operatorname{ord}_p(y_j(a')) = \alpha'_j$ and $\overline{\operatorname{ac}}_p(y_j(a')) = \overline{\operatorname{ac}}_p(y_j(a))$, for $j = 1, \dots, n$. Indeed any element of $\mathbb{Z}_p^n$ which is congruent mod p to the image of a under this étale morphism, can be lifted to a point $a' \in X(\mathbb{Z}_p)$ congruent to a. Now we have that $\overline{\operatorname{ac}}_p(x_i(a')) = \overline{\operatorname{ac}}_p(x_i(a)) = \overline{\operatorname{ac}}_p(z_i) = \overline{\operatorname{ac}}_p(z'_i)$ and $\operatorname{ord}_p(x_i(a')) = (\Delta(\alpha'))_i = \operatorname{ord}_p(z'_i)$, because $u_i(a)$ and $u_i(a')$ are units in $\mathbb{Z}_p$ which are congruent mop p. Hence $x_i(a')$ and z'_i have the same multiplicative residue for $i = 1, \dots, r$. Thus the multiplicative residue of z' with respect to $x_1, \dots, x_r$ is realized by a'. This terminates the proof of the lemma. □

4. Tameness and the surjectivity criterion

The following result is a special case of the tameness theorem of [Denef 2016] (together with Remark 5.2 of that work).

Tameness Theorem 4.1. *Let $f : X \to Y$ be a morphism of varieties over $\mathbb{Z}$. Given rational functions $x_1, \dots, x_r$ on X, there exist rational functions $y_1, \dots, y_s$ on Y, such that for almost all primes p we have the following: "Any $b \in Y(\mathbb{Z}_p)$ having the same residues with respect to $y_1, \dots, y_s$ as an image $f(a')$, with $a' \in X(\mathbb{Z}_p)$, is itself an image of an $a \in X(\mathbb{Z}_p)$ with the same residues as a' with respect to $x_1, \dots, x_r$." Moreover, if Y is affine, then we can choose $y_1, \dots, y_s$ to be regular rational functions on Y.*

This special case, and the more general result in [Denef 2016], can be proved easily by using Basarab's theorem [1991] on elimination of quantifiers. The special case itself is also an easy consequence of the theorem of Pas [1989] on uniform p-adic quantifier elimination. The works of Pas and Basarab are based on methods from mathematical logic. However in [Denef 2016] we gave a purely algebraic geometric proof of the tameness theorem which is based on the weak toroidalization theorem of Abramovich and Karu [2000] (extended to nonclosed fields [Abramovich et al. 2013]).

We briefly sketch the geometric proof of the tameness theorem of [Denef 2016]. Using (weak) toroidalization of the morphism $f \otimes \mathbb{Q}$, and induction on the dimension of $X \otimes \mathbb{Q}$ (to take care of the exceptional loci of the modifications used to obtain a toroidalization), one easily reduces to the following case. The morphism $f \otimes \mathbb{Q}$ is toroidal, $X \otimes \mathbb{Q}$ and $Y \otimes \mathbb{Q}$ are smooth, and the zero loci and polar loci of $x_1, \ldots, x_r$, restricted to $X \otimes \mathbb{Q}$, are contained in the support of the toroidal divisor on $X \otimes \mathbb{Q}$. Then $f \otimes \mathbb{Q}$ is log-smooth with respect to the toroidal divisors. In that case the tameness theorem follows directly from a logarithmic version of Hensel's lemma. We refer to [Denef 2016] for the details. Note that the last sentence in the statement of the Tameness Theorem 4.1 is a direct formal consequence of Lemma 3.2. The relation with logarithmic geometry is investigated in [Cao 2015].

The following surjectivity criterion is based on the Tameness Theorem 4.1 and is essential for the proof of the Diophantine Excision Theorem 5.1.

4.2. ***Surjectivity criterion.*** Let $f : X \to Y$ be a morphism of varieties over $\mathbb{Z}$, with Y affine and $Y \otimes \mathbb{Q}$ smooth. Suppose that, given any regular rational functions $y_1, \ldots, y_s$ on Y and $M \in \mathbb{N}$, we have the following for almost all primes p. For each $b \in Y(\mathbb{Z}_p)$, with $\operatorname{ord}_p(y_i(b)) \leq M$ for $i = 1, \ldots, s$, there exists $a \in X(\mathbb{Z}_p)$ such that $f(a)$ and b have the same residues with respect to $y_1, \ldots, y_s$. If this condition is satisfied, then the map $X(\mathbb{Z}_p) \to Y(\mathbb{Z}_p)$, induced by f, is surjective for almost all primes p.

Proof. By the Tameness Theorem 4.1 applied to the morphism f and an empty list of rational functions on X, there exist a natural number N_1 and nonzero regular rational functions $y_1, \ldots, y_s$ on Y satisfying the conclusion of the tameness theorem for all primes $p > N_1$. By enlarging the list $y_1, \ldots, y_s$ we may assume that it contains a set of affine coordinates for Y.

Next we apply Lemma 3.6 firstly to the regular rational functions $y_1, \ldots, y_s$ on Y, and secondly also to the regular rational functions $y_1 \circ f, \ldots, y_s \circ f$ on X. This yields a natural number N_2 and a common partition $\mathcal{P}$ of $\mathbb{N}^s$ satisfying, for all primes $p > N_2$, the conclusion of Lemma 3.6 both for the y_i on Y and for the $y_i \circ f$ on X.

Choose a point in each stratum of this partition $\mathcal{P}$, and choose $M \in \mathbb{N}$ bigger than the p-adic valuations of the coordinates of these points. Choose a natural number N_3 such that the hypothesis of the surjectivity criterion holds for the above $y_1, \ldots, y_s$ and M, for all primes $p > N_3$.

From now on, let p be any prime bigger than N_1, N_2 and N_3, and let $b' \in Y(\mathbb{Z}_p)$. In order to prove the surjectivity criterion we will find an $a' \in X(\mathbb{Z}_p)$ with $f(a') = b'$. Because of observation 2.2.(a), we may assume that $y_i(b') \neq 0$ for $i = 1, \ldots, s$. Set $z' := (y_1(b'), \ldots, y_s(b'))$.

By our first application of Lemma 3.6 and the choice of M, there exists a point $b \in Y(\mathbb{Z}_p)$ such that the p-adic valuations of $z := (y_1(b), \ldots, y_s(b))$ and z' are in the same stratum of the partition $\mathcal{P}$, and $\overline{\mathrm{ac}}_p(y_i(b)) = \overline{\mathrm{ac}}_p(y_i(b'))$, and $\operatorname{ord}_p(y_i(b)) \leq M$ for $i = 1, \ldots, s$.

By the above mentioned instance of the hypothesis of the surjectivity criterion, there exists $a \in X(\mathbb{Z}_p)$ such that $f(a)$ and b have the same residues with respect to $y_1, \ldots, y_s$. Thus the p-adic valuations of $z'' := (y_1(f(a)), \ldots, y_s(f(a)))$ and z are equal and hence in the same stratum of $\mathcal{P}$ as these of z'. Moreover $\overline{\mathrm{ac}}_p(y_i(f(a))) = \overline{\mathrm{ac}}_p(y_i(b)) = \overline{\mathrm{ac}}_p(y_i(b'))$.

Thus, by our second application of Lemma 3.6, the multiplicative residue of z' is realizable with respect to $y_1 \circ f, \ldots, y_s \circ f$, because obviously the multiplicative residue of z'' is realizable with respect to these functions. This means that there exists an $a'' \in X(\mathbb{Z}_p)$ such that $y_i(f(a''))$ and $y_i(b')$ have the same multiplicative residue for $i = 1, \ldots, s$. Since the list $y_1, \ldots, y_s$ contains a set of affine coordinates for Y, this implies that $f(a'') \equiv b' \mod p$. Hence $f(a'')$ and b' have the same residues with respect to $y_1, \ldots, y_s$.

By our above mentioned application of the Tameness Theorem 4.1 we conclude that there exists an $a' \in X(\mathbb{Z}_p)$ with $f(a') = b'$. This terminates the proof of the surjectivity criterion. □

5. The Diophantine excision theorem

Diophantine Excision Theorem 5.1. *Let $f : X \to Y$ be a proper dominant morphism of varieties over $\mathbb{Z}$, with $Y \otimes \mathbb{Q}$ smooth. Assume that for each strict modification $f' : X' \to Y'$ of f, with $Y' \otimes \mathbb{Q}$ smooth, there exists a closed subscheme S' of Y', of codimension ≥ 2, such that for almost all primes p we have*

$$\{y \in Y'(\mathbb{Z}_p) \mid y \mod p \notin S'(\mathbb{F}_p)\} \subset f'(X'(\mathbb{Z}_p)).$$

Then the map $X(\mathbb{Z}_p) \to Y(\mathbb{Z}_p)$, induced by f, is surjective for almost all primes p.

We prove the diophantine excision theorem at the end of the present section, after two lemma's: Lemma 5.5 states that the hypothesis of the diophantine excision theorem, with Y affine, implies the hypothesis of the surjectivity criterion (Section 4.2). This is proved by using embedded resolution of singularities and extra blowups to reduce it to a special case implied by Lemma 5.3. This last mentioned lemma is proved by noetherian induction on closed subschemes of Y, and blowing up these subschemes. But first we mention some observations whose proofs are straightforward.

5.2. *Observations.* (a) Let $f : X \to Y$ be a proper dominant morphism of varieties over $\mathbb{Z}$, with $Y \otimes \mathbb{Q}$ smooth, satisfying the assumption of the diophantine excision theorem. If U is a nonempty open subscheme of Y, then also the morphism $f^{-1}(U) \to U$, induced by f, satisfies the assumption of the Diophantine Excision Theorem 5.1.

Proof. It suffices to show that any modification $\beta_0 : U' \to U$ of U, with $U' \otimes \mathbb{Q}$ smooth, factors as an open immersion $j : U' \to Y'$ composed with a modification $\beta : Y' \to Y$ of Y, with $Y' \otimes \mathbb{Q}$ smooth, and $j(U') = \beta^{-1}(U)$. To achieve this, let $\beta_1 : U' \to Y$ be the composition of β_0 with the inclusion $U \subset Y$. Apply Nagata's compactification theorem (see e.g., [Lütkebohmert 1993]) to factorize β_1 as an open immersion $j_2 : U' \hookrightarrow Y''$ composed with a proper morphism $\beta_2 : Y'' \to Y$ of $\mathbb{Z}$-varieties. This implies that β_2 is a modification of Y and that $U'' := j_2(U') = \beta_2^{-1}(U)$. Indeed the composition of the morphisms $U' \hookrightarrow \beta_2^{-1}(U) \to U$, induced by j_2 and β_2, is proper since it equals β_0; thus the open immersion $U' \hookrightarrow \beta_2^{-1}(U)$ is proper and hence surjective. There exists a resolution of $Y'' \otimes \mathbb{Q}$ which is a composition of blowups with smooth centers $C_1, \ldots, C_r$ that lie above $(Y'' \setminus U'') \otimes \mathbb{Q}$. Denote by $\bar{C}_1$ the closure of C_1 in Y''. Denote by $\bar{C}_2$ the closure of C_2 in the blowup of Y'' with center $\bar{C}_1$, and so on. Let $\pi : Y' \to Y''$ be the modification of Y'' obtained by composing the blowups with centers $\bar{C}_1, \ldots, \bar{C}_r$.

Then $Y' \otimes \mathbb{Q}$ is smooth. Moreover, π is an isomorphism above U''. This yields an open immersion $j_1 : U'' \to Y'$. Clearly $j := j_1 \circ j_2$ and $\beta := \beta_2 \circ \pi$ satisfy the required properties. □

(b) Let $f : X \to Y$ be a proper dominant morphism of varieties over $\mathbb{Z}$, with $Y \otimes \mathbb{Q}$ smooth, satisfying the assumption of the diophantine excision theorem. Let $f_1 : X_1 \to Y_1$ be a strict modification of f, with $Y_1 \otimes \mathbb{Q}$ smooth. Then f_1 satisfies the assumption of the diophantine excision theorem. This follows directly from the fact that any strict modification of f_1 is also a strict modification of f.

Lemma 5.3. *Let $f : X \to Y$ be a proper dominant morphism of varieties over $\mathbb{Z}$, with Y smooth over $\mathbb{Z}$, satisfying the assumption of the diophantine excision theorem. Let $h : Y \to \mathbb{A}^1_{\mathbb{Z}}$ be a smooth morphism. Then for almost all primes p we have the following: for each $b \in Y(\mathbb{Z}_p)$ there exists an $a \in X(\mathbb{Z}_p)$ such that $f(a) \equiv b \mod p$ and $h(f(a)) = h(b)$.*

Proof. By noetherian induction, it suffices to show that for any integral closed subscheme W of Y, there exists a nonempty open subscheme W_0 of W, such that, for almost all p, the assertion of the lemma holds for all $b \in Y(\mathbb{Z}_p)$ satisfying $b \mod p \in W_0$. If $W = Y$ then we can directly apply the assumption of the diophantine excision theorem, with $f' = f$, to find W_0. Hence we can assume that $W \subsetneq Y$. Moreover, by cutting away the nonsmooth locus of W and using Observation 5.2(a), we may also assume that W is smooth over $\mathbb{Z}$. Thus locally the ideal sheaf of W on Y can be generated by part of a set of local uniformizing parameters over $\mathbb{Z}$, hence its blowup is smooth over $\mathbb{Z}$ with exceptional locus a projective space bundle on W.

Let $\beta : Y' \to Y$ be the blowup of Y with center W, and let $f' : X' \to Y'$ be the strict transform of f with respect to β. Because the assumption of the diophantine excision theorem is assumed, there exists a closed subscheme S' of Y', of codimension ≥ 2, such that for almost all primes p we have

$$\{y \in Y'(\mathbb{Z}_p) \mid y \mod p \notin S'(\mathbb{F}_p)\} \subset f'(X'(\mathbb{Z}_p)). \tag{1}$$

To start, we take W_0 equal to W, but later on we will replace W_0 by a smaller nonempty open subscheme of W if necessary.

When the restriction of h to W is dominant, then making W_0 smaller if necessary, we may suppose that the restriction of h to W_0 is smooth. Whence the restriction of $h \circ \beta$ to $\beta^{-1}(W_0)$ is smooth, because the morphism from $\beta^{-1}(W_0)$ to W_0, induced by β, is smooth. This implies that $h \circ \beta$ is smooth at each point of $\beta^{-1}(W_0)$, by the smoothness criterion for morphisms of smooth schemes (Théorème 17.11.1 in [EGA IV$_4$ 1967]).

When the restriction of h to W is not dominant, then $h(W \otimes \mathbb{Q}) = \{P\}$, with P a closed point of $\mathbb{A}^1_{\mathbb{Q}}$. Let $[P]$ be the prime divisor on $\mathbb{A}^1_{\mathbb{Q}}$ consisting of this point P with multiplicity one. Because h is smooth, the multiplicity of $\beta^{-1}(W) \otimes \mathbb{Q}$, in the divisor $((h \circ \beta) \otimes \mathbb{Q})^*([P])$ on $Y' \otimes \mathbb{Q}$, equals 1. Let C' be the intersection of $\beta^{-1}(W) \otimes \mathbb{Q}$ with the union of the other irreducible components of this divisor. Clearly C' has codimension ≥ 2 in $Y' \otimes \mathbb{Q}$, and $(h \circ \beta) \otimes \mathbb{Q}$ is smooth at each point of $\beta^{-1}(W) \otimes \mathbb{Q} \setminus C'$. Enlarging S' if necessary, we may suppose that $C' \subset S'$. Whence, the singular locus of $h \circ \beta$ is disjoint from $(\beta^{-1}(W) \setminus S') \otimes \mathbb{Q}$. Hence shrinking W_0 if necessary, by inverting finitely many primes, we can assume that the singular locus of $h \circ \beta$ is disjoint from $\beta^{-1}(W_0) \setminus S'$.

Thus in either case, we can suppose that $h \circ \beta$ is smooth at each point of $\beta^{-1}(W_0) \setminus S'$.

Since $S' \subset Y'$ has codimension ≥ 2 and $\beta^{-1}(W) \subset Y'$ has codimension 1, the morphism from $\beta^{-1}(W) \setminus S'$ to W, induced by β, is dominant. Hence, replacing W_0 if necessary by a smaller open subscheme of W, we may suppose that $W_0 \subset \beta(Y' \setminus S')$.

Let p be a big enough prime, and consider any $b \in Y(\mathbb{Z}_p)$ satisfying $\bar{b} := b \mod p \in W_0$. Because the scheme-theoretic fiber of β over $\bar{b}$ is isomorphic to a projective space over $\mathbb{F}_p$, and its intersection with S' is contained in a hypersurface (of this fiber) with degree bounded independently of b, and because p is big enough, there exists a $\mathbb{F}_p$-rational point $\overline{b'}$ on $Y' \setminus S'$ with $\beta(\overline{b'}) = \bar{b}$. Since $h \circ \beta$ is smooth at $\overline{b'}$ and $(h \circ \beta)(\overline{b'}) = h(b) \mod p$, this point lifts to a point $b' \in Y'(\mathbb{Z}_p)$ with $b' \mod p = \overline{b'}$ and $(h \circ \beta)(b') = h(b)$. By (1) and because $b' \mod p \notin S'$, there exists $a' \in X'(\mathbb{Z}_p)$ with $f'(a') = b'$. Let $a \in X(\mathbb{Z}_p)$ be the image of a' under the natural morphism $X' \to X$. Then $f(a) = \beta(f'(a')) = \beta(b') \equiv b \mod p$, and $h(f(a)) = h(\beta(b')) = h(b)$. This terminates the proof of the lemma. □

Remark 5.4. The previous Lemma 5.3 also holds when there is no h involved, if we drop the requirement that $h(f(a)) = h(b)$. This follows formally from this lemma, using the observation in Section 5.2(a), by covering Y by finitely many small enough open subschemes on which there exists a smooth morphism to $\mathbb{A}^1_{\mathbb{Z}}$.

Lemma 5.5. *Let $f : X \to Y$ be a proper dominant morphism of varieties over $\mathbb{Z}$, with $Y \otimes \mathbb{Q}$ smooth, satisfying the assumption of the diophantine excision theorem. Let $y_1, \ldots, y_s$ be regular rational functions on Y, and $M \in \mathbb{N}$. Then for almost all primes p we have the following. For each $b \in Y(\mathbb{Z}_p)$, with $\mathrm{ord}_p(y_i(b)) \leq M$ for $i = 1, \ldots, s$, there exists $a \in X(\mathbb{Z}_p)$ such that $f(a)$ and b have the same residues with respect to $y_1, \ldots, y_s$.*

Proof. Let D be the union of the zero loci of the regular rational functions $y_1, \ldots, y_s$ on Y, considered as a subset of Y. Using embedded resolution of singularities of $D \otimes \mathbb{Q} \subset Y \otimes \mathbb{Q}$, modifying Y, without changing $Y \otimes \mathbb{Q} \setminus D \otimes \mathbb{Q}$, replacing f by its strict transform with respect to the modification of Y, and inverting a finite number of primes, we may assume the following. The variety Y is smooth over $\mathbb{Z}$, and D is a reduced strict normal crossings divisor over $\mathbb{Z}$ (in the sense of Section 1.3). This reduction is easily verified applying the valuative criterion of properness to the resolution morphism and using the observation in Section 5.2(b). Thus, by covering Y with finitely many suitable open subschemes, and using the observation in Section 5.2(a), we can further assume that Y is affine, and that each y_i can be written as a unit in $\Gamma(Y, \mathcal{O}_Y)$ times a monomial in uniformizing parameters over $\mathbb{Z}$ on X (i.e., regular functions on Y that induce an étale morphism to an affine space over $\mathbb{Z}$; see Section 1.3). Hence, by Lemma 3.3, we can moreover assume that $y_1, \ldots, y_s$ are part of a set of uniformizing parameters over $\mathbb{Z}$ on Y.

It remains now to prove the lemma in the special case that Y is smooth over $\mathbb{Z}$, and affine, say $Y = \mathrm{Spec}(A)$, and that $y_1, \ldots, y_s$ are part of a set of uniformizing parameters over $\mathbb{Z}$ on Y. We prove this special case by induction on M. Let p be a prime, big enough with respect to M and all data, and let $b \in Y(\mathbb{Z}_p)$ be any point with $\mathrm{ord}_p(y_i(b)) \le M$ for all $i = 1, \ldots, s$. If $M = 0$, then, in order to prove the lemma, it suffices to find $a \in X(\mathbb{Z}_p)$ with $f(a) \equiv b \mod p$. The existence of such an a follows from Remark 5.4. Thus we may suppose that $M > 0$ and that

$$I_0 := \{i \in \mathbb{N} \mid \mathrm{ord}_p(y_i(b)) > 0, 1 \le i \le s\} \ne \varnothing.$$

Choose $i_0 \in I_0$ such that $\mathrm{ord}_p(y_{i_0}(b)) = \mathrm{Min}_{i \in I_0} \mathrm{ord}_p(y_i(b))$.

Let $\pi : Y' \to Y$ be the blowup of the ideal sheaf on Y generated by all the y_i with $i \in I_0$. Consider the chart U on Y', defined as follows:

$$U := \mathrm{Spec}(A[(y_i/y_{i_0})_{i \in I_0}]) \xrightarrow{\pi} \mathrm{Spec}(A) = Y.$$

There exists a unique $b' \in U(\mathbb{Z}_p)$ with $\pi(b') = b$. Set $y'_i = y_i/y_{i_0}$ for $i \in I_0 \setminus \{i_0\}$, and $y'_i = y_i$ for the other $i \in \{1, \ldots, s\}$. One easily verifies that $y'_1, \ldots, y'_s$ are part of a set of uniformizing parameters over $\mathbb{Z}$ on U. Clearly, either $0 \le \mathrm{ord}_p(y'_i(b')) < M$, for all i, or $\mathrm{ord}_p(y'_i(b')) = 0$, for all $i \ne i_0$. We call these respectively the first case and the second case.

Let $f' : X' \to Y'$ be the strict transform of f with respect to the blowup $\pi : Y' \to Y$. In the first case, we apply the induction hypothesis to the morphism $f'^{-1}(U) \to U$ induced by f', and the regular rational functions $y'_1, \ldots, y'_s$ on U, to find $a' \in X'(\mathbb{Z}_p)$, with $f'(a') \in U(\mathbb{Z}_p)$, such that $f'(a')$ and b' have the same residues with respect to these functions on U. In the second case, we apply Lemma 5.3 to the morphism $f'^{-1}(U) \to U$ induced by f', and the morphism $U \to \mathbb{A}^1_{\mathbb{Z}}$ induced by y_{i_0}, to find $a' \in X'(\mathbb{Z}_p)$, with $f'(a') \in U(\mathbb{Z}_p)$, such that $f'(a') \equiv b' \mod p$ and $y_{i_0}(f'(a')) = y_{i_0}(b')$. Hence, also in the second case, $f'(a')$ and b' have the same residues with respect to $y'_1, \ldots, y'_s$, because $y'_i(b')$ is a unit in $\mathbb{Z}_p$ for all $i \ne i_0$.

Denote by a the image of a' under the natural map $X'(\mathbb{Z}_p) \to X(\mathbb{Z}_p)$. Then the points $f(a) = \pi(f'(a'))$ and $b = \pi(b')$ have the same residues with respect to $y_1, \ldots, y_s$. □

5.6. ***Proof of the Diophantine Excision Theorem 5.1.*** Using observation 5.2(a) we may assume that Y is affine. Then the Diophantine Excision Theorem 5.1 is a direct consequence of the above Lemma 5.5 and the surjectivity criterion (Section 4.2). □

6. Proof of the Main Theorem 1.2

In this section we show that the Main Theorem 1.2 is an easy consequence of the Diophantine Excision Theorem 5.1 and the following lemma, whose proof is rather straightforward.

Lemma 6.1. *Let $f : X \to Y$ be a proper dominant morphism of smooth varieties over $\mathbb{Z}$, with geometrically integral generic fiber. Assume for each $\mathbb{Z}$-flat prime divisor D on Y, that the divisor $f^*(D)$ on X has an irreducible component C with multiplicity 1 and geometrically integral generic fiber over D (i.e., the morphism $C \to D$, induced by f, has geometrically integral generic fiber). Then there exists a closed subscheme S of Y, of codimension ≥ 2, such that for almost all primes p we have*

$$\{y \in Y(\mathbb{Z}_p) \mid y \mod p \notin S(\mathbb{F}_p)\} \subset f(X(\mathbb{Z}_p)).$$

Proof. By Théorème 9.7.7 of [EGA IV$_3$ 1966] , there exists a reduced closed subscheme $E \subset Y$, of pure codimension 1, such that over the complement of E, the morphism f is smooth with geometrically integral fibers. Hence, for almost all primes p, any $y \in Y(\mathbb{Z}_p)$, with $y \mod p \notin E(\mathbb{F}_p)$, belongs to $f(X(\mathbb{Z}_p))$. Indeed this follows from Hensel's lemma and the Lang–Weil bound [Lang and Weil 1954].

For each irreducible component D of E we reason as follows. If D is not flat over $\mathrm{Spec}(\mathbb{Z})$, then $D(\mathbb{F}_p)$ is empty for almost all primes p. Suppose now that D is flat over $\mathrm{Spec}(\mathbb{Z})$. By assumption, the divisor $f^*(D)$ on X has an irreducible component C with multiplicity 1 and geometrically integral generic fiber over D. In particular, C dominates D. Hence there exists a reduced closed subscheme S of D, of codimension ≥ 1 in D, such that, over the complement of S, all fibers of $C \xrightarrow{f} D$ are geometrically integral and intersect the smooth locus of $f : X \to Y$. Indeed, f is smooth at the generic point of C, because C has multiplicity 1 in the divisor $f^*(D)$. Again by hensel's Lemma and the Lang–Weil bound [Lang and Weil 1954], we conclude for almost all primes p that any $y \in Y(\mathbb{Z}_p)$, with $y \mod p \in D(\mathbb{F}_p) \setminus S(\mathbb{F}_p)$, belongs to $f(X(\mathbb{Z}_p))$.

Taking the union of the subschemes S, obtained as above for each $\mathbb{Z}$-flat irreducible component D of E, we obtain a closed subscheme of Y, of codimension ≥ 2, that satisfies the conclusion of the lemma. □

6.2. ***Proof of Main Theorem 1.2.*** Let $f : X \to Y$ be a dominant morphism of smooth proper geometrically integral varieties over $\mathbb{Q}$, which satisfies the hypotheses of the main theorem. Choose a proper dominant morphism $\tilde{f} : \tilde{X} \to \tilde{Y}$ of smooth varieties over $\mathbb{Z}$, whose base change to $\mathbb{Q}$ is isomorphic to f. Because X and Y are proper, it suffices to prove that the map $X(\mathbb{Z}_p) \to Y(\mathbb{Z}_p)$, induced by $\tilde{f}$, is surjective for almost all primes p. By the Diophantine Excision Theorem 5.1, it suffices to prove that the morphism $\tilde{f}$ satisfies the assumption in the diophantine excision theorem, with f replaced by $\tilde{f}$.

Let $\tilde{f}' : \tilde{X}' \to \tilde{Y}'$ be any strict modification of $\tilde{f}$, with $\tilde{Y}' \otimes \mathbb{Q}$ smooth. Note that the generic fiber of $\tilde{f}'$ equals the one of $\tilde{f}$ and is contained in the smooth locus of $\tilde{X}'$, because the modification is strict and $\tilde{X}$ is smooth. We have to prove that there exists a closed subscheme S of $\tilde{Y}'$, of codimension ≥ 2, such that for almost all primes p we have

$$\{y \in \tilde{Y}'(\mathbb{Z}_p) \mid y \mod p \notin S(\mathbb{F}_p)\} \subset \tilde{f}'(\tilde{X}'(\mathbb{Z}_p)).$$

Composing $\tilde{f}'$ with a morphism whose base change to $\mathbb{Q}$ resolves the singularities of $\tilde{X}' \otimes \mathbb{Q}$ without changing the smooth locus of $\tilde{X}' \otimes \mathbb{Q}$, and inverting a finite number of primes, we see that to prove the above, we may assume that *the varieties $\tilde{X}'$ and $\tilde{Y}'$ are smooth over $\mathbb{Z}$, and $\tilde{f}'$ is a modification of $\tilde{f}$, with the same generic fiber as $\tilde{f}$*. But now $\tilde{f}'$ is not necessarily a strict modification of $\tilde{f}$ anymore.

Because, by assumption, f satisfies the hypotheses of the main theorem, it is easy to verify that $\tilde{f}'$ satisfies the hypotheses of Lemma 6.1 (with f replaced by $\tilde{f}'$). Hence this lemma implies the existence of a closed subscheme S of $\tilde{Y}'$ with the required properties. □

6.3. ***An alternative proof of Colliot-Thélène's Conjecture.*** Using model theory (mathematical logic) one can give a much simpler proof of Colliot-Thélène's conjecture in Section 1.1. However we don't see how to extend this to prove the stronger Main Theorem 1.2 or the Diophantine Excision Theorem 5.1. Moreover one of the motivations of Colliot-Thélène was to obtain a new proof of the Ax–Kochen

Theorem which does not rely on methods from mathematical logic. We briefly sketch this simpler proof of Colliot-Thélène's conjecture.

Assume the notation and hypotheses in the formulation of Colliot-Thélène's conjecture. Using a standard argument from model theory and the Ax–Kochen–Eršov transfer principle [Ax and Kochen 1965; Ershov 1965], we will first show that, in order to prove the conjecture, it suffices to show that the map $X(F[[t]]) \to Y(F[[t]])$, induced by f, is surjective for any pseudoalgebraically closed field F of characteristic zero. This goes as follows. If Colliot-Thélène's conjecture is false for f, then there exists an infinite set S of primes p for which the map from $X(\mathbb{Q}_p)$ to $Y(\mathbb{Q}_p)$, induced by f, is not surjective. Let K be the ultraproduct of all the fields $\mathbb{Q}_p$ with respect to an ultrafilter, on the set $\mathcal{P}$ of all primes, containing S and each subset of $\mathcal{P}$ with finite complement. Then the map from $X(K)$ to $Y(K)$, induced by f, is not surjective. Notice that K is a henselian valued field with residue field a pseudoalgebraically closed field F of characteristic zero (by [Lang and Weil 1954]), and value group elementary equivalent to $\mathbb{Z}$. Hence K is elementary equivalent to the field of fractions $F((t))$ of $F[[t]]$, by the Ax–Kochen-Eršov transfer principle which states that any two henselian valued fields are elementary equivalent if they have elementary equivalent value groups and elementary equivalent residue fields of characteristic zero. Thus the map from $X(F((t)))$ to $Y(F((t)))$, induced by f, is not surjective if Colliot-Thélène's conjecture is false for f. Since f is proper the same holds for $F((t))$ replaced by $F[[t]]$. We conclude that in order to prove the conjecture, it suffices to show that the map $X(F[[t]]) \to Y(F[[t]])$, induced by f, is surjective for any pseudoalgebraically closed field F of characteristic zero.

Let $y \in Y(F[[t]])$. We have to show that $y \in f(X(F[[t]]))$. Let s be the closed point of $\mathrm{Spec}(F[[t]])$. By slightly moving y and using Greenberg's theorem [1966], we may assume that the homomorphism $\mathcal{O}_{Y,y(s)} \to F[[t]]$ induced by y is injective. Composing this homomorphism with the standard valuation on $F[[t]]$, induces a discrete valuation v on the function field K of Y, with valuation ring say A.

If v is trivial, then $y(s)$ is the generic point of Y. Hence f is smooth at each point in the fiber of $y(s)$, and this fiber is geometrically integral. This implies that y lifts to a $F[[t]]$-rational point x on X, by Hensel's lemma and the assumption that F is pseudoalgebraically closed.

Thus we may assume that the discrete valuation v is not trivial. Hence there exists an integral regular A-scheme $\mathfrak{X}$ as in the formulation of Colliot-Thélène's conjecture. Note that y induces a $F[[t]]$-rational point $\tilde{y}$ on $\mathrm{Spec}(A)$, and a homomorphism $K \to F((t))$. Using the hypothesis about the special fiber of $\mathfrak{X}$, Hensel's Lemma, and the assumption that F is pseudoalgebraically closed, one easily verifies that $\tilde{y}$ lifts to a $F[[t]]$-rational point $\tilde{x}$ on $\mathfrak{X}$. Because the generic fiber of $\mathfrak{X}$ is K-isomorphic to the generic fiber of f, and because $\tilde{x}$ extends to a $F((t))$-rational point on $\mathfrak{X} \otimes K$, we find a $F((t))$-rational point on X, and hence, by the properness of X, also a $F[[t]]$-rational point x on X with $f(x) = y$. □

Acknowledgments

We thank J.-L. Colliot-Thélène, O. Gabber, and O. Wittenberg for a mutual conversation that resulted in the alternative proof (Section 6.3). We also thank D. Abramovich, R. Cluckers, J.-L. Colliot-Thélène, S.

D. Cutkosky, K. Karu, and O. Wittenberg for stimulating conversations and useful information. Moreover Cao's paper [2015] was helpful while producing the final version of the present paper.

Added in proof

Substantial refinements of the results in the present paper were recently obtained by D. Loughran, A. Skorobogatov and A. Smeets [Loughran et al. 2018].

References

[Abramovich and Karu 2000] D. Abramovich and K. Karu, "Weak semistable reduction in characteristic 0", *Invent. Math.* **139**:2 (2000), 241–273. MR Zbl

[Abramovich et al. 2013] D. Abramovich, J. Denef, and K. Karu, "Weak toroidalization over non-closed fields", *Manuscripta Math.* **142**:1-2 (2013), 257–271. MR Zbl

[Ax and Kochen 1965] J. Ax and S. Kochen, "Diophantine problems over local fields, I", *Amer. J. Math.* **87** (1965), 605–630. MR Zbl

[Basarab 1991] S. A. Basarab, "Relative elimination of quantifiers for Henselian valued fields", *Ann. Pure Appl. Logic* **53**:1 (1991), 51–74. MR Zbl

[Cao 2015] Y. Cao, "Denef's proof of Colliot-Thélène's conjecture", M2 mémoire, Univ. Paris-Sud, 2015, https://tinyurl.com/m2memcao.

[Colliot-Thélène 2008] J.-L. Colliot-Thélène, "Fibre spéciale des hypersurfaces de petit degré", *C. R. Math. Acad. Sci. Paris* **346**:1-2 (2008), 63–65. MR Zbl

[Colliot-Thélène 2011] J.-L. Colliot-Thélène, "Variétés presque rationnelles, leurs points rationnels et leurs dégénérescences", pp. 1–44 in *Arithmetic geometry* (Cetraro, Italy, 2007), edited by P. Corvaja and C. Gasbarri, Lecture Notes in Math. **2009**, Springer, 2011. MR Zbl

[Denef 2016] J. Denef, "Geometric proofs of theorems of Ax–Kochen and Eršov", *Amer. J. Math.* **138**:1 (2016), 181–199. MR Zbl

[EGA IV$_3$ 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[EGA IV$_4$ 1967] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl

[Ershov 1965] Y. L. Ershov, "On elementary theories of local fields", *Algebra i Logika Sem.* **4**:2 (1965), 5–30. In Russian. MR

[Greenberg 1966] M. J. Greenberg, "Rational points in Henselian discrete valuation rings", *Inst. Hautes Études Sci. Publ. Math.* **31** (1966), 59–64. MR Zbl

[Lang and Weil 1954] S. Lang and A. Weil, "Number of points of varieties in finite fields", *Amer. J. Math.* **76** (1954), 819–827. MR Zbl

[Loughran et al. 2018] D. Loughran, A. N. Skorobogatov, and A. Smeets, "Pseudo-split fibres and arithmetic surjectivity", preprint, 2018. arXiv

[Lütkebohmert 1993] W. Lütkebohmert, "On compactification of schemes", *Manuscripta Math.* **80**:1 (1993), 95–111. MR Zbl

[Pas 1989] J. Pas, "Uniform p-adic cell decomposition and local zeta functions", *J. Reine Angew. Math.* **399** (1989), 137–172. MR Zbl

Communicated by Bjorn Poonen
Received 2016-03-31 Revised 2019-01-10 Accepted 2019-05-14

jan.denef@kuleuven.be *Department of Mathematics, KULeuven, Leuven, Belgium*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.1997

Irreducible characters with bounded root Artin conductor

Amalia Pizarro-Madariaga

We prove that the best possible lower bound for the Artin conductor is exponential in the degree.

1. Introduction

Let K be an algebraic number field such that $K/\mathbb{Q}$ is Galois and let χ be the character of a linear representation of $\mathrm{Gal}(K/\mathbb{Q})$. We denote by f_χ the Artin conductor of χ. Odlyzko [1977] found lower bounds for f_χ by applying analytic methods to the Artin L-function. We have improved Odlyzko's lower bounds in [Pizarro-Madariaga 2011] by using explicit formulas for Artin L-functions. In particular, if χ is an irreducible character of $\mathrm{Gal}(K/\mathbb{Q})$ by assuming that $\chi\bar{\chi}$ satisfies the Artin conjecture, we obtained

$$f_\chi^{1/\chi(1)} \geq 4.73(1.648)^{(a_\chi - b_\chi)^2/\chi(1)^2} e^{-(13.34/\chi(1))^2},$$

where a_χ and b_χ are nonnegative integers giving the Γ-factors of the completed Artin L-function. Namely, $a_\chi + b_\chi = \chi(1)$ and $a_\chi - b_\chi = \chi(\sigma)$, with $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$ the complex conjugation. This bound is even better when we assume that $L(s, \chi\bar{\chi})$ satisfies the generalized Riemann hypothesis. We have to point out that, throughout this article, no additional hypothesis are needed.

A natural question now is how far from being optimal these bounds are. This problem has been studied for the discriminant of a number field. If $n_0 = r_1 + 2r_2$, let d_n be the minimal discriminant of the field K with degree n such that n is a multiple of n_0 and $r_1(K)$ and $r_2(K)$ are in the same ratio as r_1, r_2. Let $\alpha(r_1, r_2) = \liminf_{n\to\infty} d_n^{1/n}$. Martinet [1978] considered number fields with infinite 2-class field towers and proved that

$$\alpha(0, 1) < 93 \quad \text{and} \quad \alpha(1, 0) < 1059.$$

In this work, we follow this idea and consider a number field K with infinite p-class field tower for some prime p. Under some technical conditions on K, we find an upper bound (depending only on K) for the root Artin conductor of the irreducible characters of $\mathrm{Gal}(K_n/\mathbb{Q})$ (given by $f_\chi^{1/\chi(1)}$), where K_n is the Hilbert p-class field of K_{n-1} with $K_0 = K$.

This work is organized as follow. In Section 2, we propose a technique obtained from Clifford's theory which is useful to classify the irreducible characters of $\mathrm{Gal}(K_n/\mathbb{Q})$ in terms of a certain normal subgroup.

MSC2010: primary 11Y40; secondary 20G05.
Keywords: Artin character, Artin conductor, Hilbert class field tower.

This characterization is convenient in order to obtain upper bounds for root Artin conductors. In Section 3, we conclude that there exists an infinite sequence $\{\chi_n\}$ of irreducible Artin characters with $\chi_n(1) \to \infty$ and such that $f_{\chi_n}^{1/\chi_n(1)} \le C$, where $C > 0$ is an effective computable constant. In Section 4, we apply the results obtained in Section 2 and 3 to the number field $K = \mathbb{Q}(\zeta_{11} + \zeta_{11}^{-1}, \sqrt{2}, \sqrt{-23})$. This field was found by Martinet [1978] and has infinite 2-class field tower and lowest known discriminant. In particular, we prove that for each $n \ge 1$ it is possible to find an irreducible character of $\mathrm{Gal}(K_n/\mathbb{Q})$ with large degree and

$$f_\chi^{1/\chi(1)} \le C, \quad \text{where } C \le 11^4 \cdot 2^{15} \cdot 23.$$

2. Irreducible characters of large degree

In this section, we develop a technique to classify the irreducible characters of groups with a normal subgroup of prime index. Also, by using a result from [Isaacs 1976], we obtain conditions that ensure the existence of irreducible characters of large degree. We believe these results are of independent interest.

Let us consider a finite group G and a normal subgroup H of G. We denote the set of irreducible characters of G by $\mathrm{Irr}(G)$. If χ and θ are characters of G and H respectively, we denote the restriction of χ to H by $\mathrm{Res}_H^G \chi$ and the induced character of θ to G by $\mathrm{Ind}_H^G \theta$. If $\theta \in \mathrm{Irr}(H)$, we define the conjugate character to θ in G by $\theta^g : H \to \mathbb{C}$, where $\theta^g(h) = \theta(ghg^{-1})$. The inertia group of θ in G is given by

$$I_G(\theta) = \{g \in G : \theta^g = \theta\}.$$

G acts on $\mathrm{Irr}(H)$ by conjugation and $I_G(\theta)$ is the stabilizer of θ under this action. The next result of Clifford will be the main argument allowing us to give a classification of the irreducible characters of G.

Theorem 1 (Clifford, [Huppert 1998, page 253]). *Let H be a normal subgroup of G and $\theta \in \mathrm{Irr}(H)$, $\chi \in \mathrm{Irr}(G)$ such that θ is an irreducible constituent of $\mathrm{Res}_H^G \chi$, with $\langle \mathrm{Res}_H^G \chi, \theta \rangle = e > 0$. Suppose that $\theta = \theta^{g_1}, \theta^{g_2}, \dots, \theta^{g_t}$ are the distinct conjugates of θ in G. Assume also that*

$$G = \bigcup_{j=1}^{t} I_G(\theta) g_j, \quad \textit{with } t = [G : I_G(\theta)].$$

Then:

(a) $\mathrm{Res}_H^G \mathrm{Ind}_H^G \theta = |I_G(\theta)/H| \sum_{j=1}^t \theta^{g_j}$.

(b) $\langle \mathrm{Ind}_H^G \theta, \mathrm{Ind}_H^G \theta \rangle = |I_G(\theta)/H|$. *In particular*, $\mathrm{Ind}_H^G \theta \in \mathrm{Irr}(G)$ *if and only if* $I_G(\theta) = H$.

(c) $\mathrm{Res}_H^G \chi = e \sum_{j=1}^t \theta^{g_j}$. *In particular*,

$$\chi(1) = et\theta(1) \quad \textit{and} \quad \langle \mathrm{Res}_H^G \chi, \mathrm{Res}_H^G \chi \rangle = e^2 t.$$

Also, $e^2 \le |I_G(\theta)/H|$ *and* $e^2 t \le |G/H|$.

In order to ensure the existence of a sequence of irreducible characters of growing degrees, let us consider the following corollary which is given as an exercise in [Isaacs 1976, page 98]. The proof is a consequence of Clifford's theorem.

Corollary 2. *Let G be a group with a chain of normal subgroups*

$$1 = H_0 \trianglelefteq H_1 \trianglelefteq H_2 \cdots \trianglelefteq H_n = G$$

such that H_i/H_{i-1} is nonabelian for $i = 1, \ldots, n$. Then, there exists an irreducible character ϕ of G, such that $\phi(1) \geq 2^n$.

Now we state the following result which is crucial for the proof of Theorem 14.

Proposition 3. *Let H be a subgroup of a finite group G. Let $\theta \in \operatorname{Irr}(H)$. Then there exists $\rho \in \operatorname{Irr}(G)$ such that*:

(i) $\rho(1) \geq \theta(1)$.

(ii) $\langle \operatorname{Ind}_H^G \theta, \rho \rangle = a \geq 1$.

Proof. It is enough take ρ to be any irreducible constituent of $\operatorname{Ind}_H^G(\theta)$. □

We say that an irreducible character θ of H *is extendible to* G if there is an irreducible character χ of G such that $\operatorname{Res}_H^G \chi = \theta$. The following result gives us a criterion to decide when a character is extendible.

Theorem 4 [Gallagher 1962, page 225]. *Let G be a finite group with a normal subgroup H of prime index q in G. If $\theta \in \operatorname{Irr}(H)$ is invariant in G (i.e., $I_G(\theta) = G$), then θ is extendible to G.*

Lemma 5. *Suppose that G is a finite group with a normal subgroup H such that $[G : H] = q$, where q is a prime number. If $\theta \in \operatorname{Irr}(H)$, then the inertia group of θ is either*

(i) $I_G(\theta) = G$, *or*

(ii) $I_G(\theta) = H$.

Proof. See [Isaacs 1976, page 82]. □

Theorem 6. *Under the conditions of Lemma 5, let χ be an irreducible character of G. Then, either*

(i) $\operatorname{Res}_H^G \chi = \theta$, *for some $\theta \in \operatorname{Irr}(H)$ or*

(ii) $\chi = \operatorname{Ind}_H^G \theta$, *for some $\theta \in \operatorname{Irr}(H)$.*

Proof. Let $\chi \in \operatorname{Irr}(G)$ and take $\theta \in \operatorname{Irr}(H)$ an irreducible constituent of $\operatorname{Res}_H^G \chi$. The proof follows directly from Theorem 4, [Huppert 1998, Theorem 19.4] and Lemma 5. □

3. Estimation for the root Artin conductor of irreducible characters of G_n

Let L/M be a Galois extension and χ be the character of a linear representation of $\operatorname{Gal}(L/M)$. The Artin conductor attached to χ is given by the ideal

$$f_\chi = \prod_{\mathfrak{p} \nmid \infty} \mathfrak{p}^{f_\chi(\mathfrak{p})},$$

where

$$f_\chi(\mathfrak{p}) = \frac{1}{|G_0|} \sum_{j \geq 0} (|G_j|\chi(1) - \chi(G_j))$$

and G_i is the i-th ramification group of the local extension $L_{\mathfrak{b}}/M_{\mathfrak{p}}$ with $\mathfrak{b}$ a prime over $\mathfrak{p}$ and $\chi(G_j) = \sum_{g\in G_j}\chi(g)$.

It is well-known that if L is an unramified extension of M, then f_χ is the trivial ideal. Then, in order to find a family of irreducible representations with bounded root Artin conductor, let us consider a number field K with infinite p-class field tower for some prime p. Let K_n be the Hilbert p-class field of K_{n-1} with $K_0 = K$ and $G_n = \mathrm{Gal}(K_n/\mathbb{Q})$.

The main objective of this section is to prove that, under some conditions over K and applying the results of the previous section, there exists an upper bound for the root Artin conductor of the irreducible characters of G_n. This bound depends only on the base field K. In addition, we obtain that for each $n > 1$ it is possible to find an irreducible character of G_n with degree increasing with n.

Proposition 7. *Let K be a Galois extension of $\mathbb{Q}$ with infinite p-class field tower, for some prime p. Suppose that K has a subfield $\tilde{k}$ satisfying the following conditions*:

(a) *$\tilde{k}$ is Galois over $\mathbb{Q}$.*

(b) *$[\tilde{k}:\mathbb{Q}] = q$, with q a prime number.*

Let $\chi \in \mathrm{Irr}(G_n)$, where $G_n = \mathrm{Gal}(K_n/\mathbb{Q})$. If $\tilde{H}_n = \mathrm{Gal}(K_n/\tilde{k})$, then either

(i) *$\mathrm{Res}^{G_n}_{\tilde{H}_n}\chi = \theta$, for some $\theta \in \mathrm{Irr}(\tilde{H}_n)$, or*

(ii) *$\chi = \mathrm{Ind}^{G_n}_{\tilde{H}_n}\theta$, for some $\theta \in \mathrm{Irr}(\tilde{H}_n)$.*

Proof. The proof follows directly from Theorem 6 with $G = G_n$ and $H = \tilde{H}_n$. □

Proposition 8. *Let K be a number field with infinite p-class field tower for some prime p. If $T_n = \mathrm{Gal}(K_n/K)$, then for each $n \geq 1$ there exists $\phi \in \mathrm{Irr}(T_n)$ such that*

$$\phi(1) > 2^{(n-1)/2}.$$

Proof. Let us consider the following chain of subgroups. If n is even, we take for $1 \leq j \leq \frac{n}{2}$:

$$\begin{aligned}
H_0 &= \{1\}, & & \\
H_1 &= \mathrm{Gal}(K_n/K_{n-2}), & H_1/H_0 &\cong H_1 \\
H_2 &= \mathrm{Gal}(K_n/K_{n-4}), & H_2/H_1 &\cong \mathrm{Gal}(K_{n-2}/K_{n-4}) \\
&\vdots & & \\
H_j &= \mathrm{Gal}(K_n/K_{n-2j}), & H_j/H_{j-1} &\cong \mathrm{Gal}(K_{n-2(j-1)}/K_{n-2j}) \\
&\vdots & & \\
H_{n/2} &= T_n = \mathrm{Gal}(K_n/K), & H_{n/2}/H_{n/2-1} &\cong \mathrm{Gal}(K_2/K).
\end{aligned}$$

If $l < i-1$ then K_i/K_l is a nonabelian group, so by Corollary 2, there exists $\phi \in \mathrm{Irr}(T_n)$ with $\phi(1) \geq 2^{n/2} > 2^{(n-1)/2}$.

If n is odd, for $j < (n-1)/2$ we take H_j and H_j/H_{j-1} as in the even case. For $j = (n-1)/2$ we take $H_{(n-1)/2}{=}T_n$ and $H_{(n-1)/2}/H_{(n-1)/2-1} \cong \mathrm{Gal}(K_3/K)$. Hence, there exists $\phi \in \mathrm{Irr}(G)$ such that $\phi(1) > 2^{(n-1)/2}$. □

Corollary 9. *Let G_n be as in Proposition 7. Then for each $n > 1$, there exists $\chi \in \mathrm{Irr}(G_n)$ such that*

$$\chi(1) > 2^{(n-1)/2}.$$

Proof. Note that if $T_n = \mathrm{Gal}(K_n/K)$ has an irreducible character θ with $\theta(1) > 2^{(n-1)/2}$, then there exists $\chi \in \mathrm{Irr}(G)$ with $\chi(1) > 2^{(n-1)/2}$. In fact, let $\theta \in \mathrm{Irr}(T_n)$ with $\theta(1) > 2^{(n-1)/2}$ and choose $\chi \in \mathrm{Irr}(G_n)$ such that θ is an irreducible constituent of $\mathrm{Res}_{T_n}^{G_n} \chi$. By Theorem 1, $\chi(1) = et\theta(1)$, where $e = \langle \mathrm{Res}_{T_n}^{G_n} \chi, \theta\rangle$ and $t = [G_n : I_g(\theta)]$. As $e, t \geq 1$, then $\chi(1) \geq \theta(1) > 2^{(n-1)/2}$. □

Now, we obtain upper bounds for the root Artin conductor of irreducible characters of G_n.

Theorem 10. *Assume G_n as in Proposition 7 and $\chi \in \mathrm{Irr}(G_n)$:*

(i) *If $\mathrm{Res}_{\tilde{H}_n}^{G_n} \chi = \theta$, for some $\theta \in \mathrm{Irr}(\tilde{H}_n)$ then*

$$f_\chi^{1/\chi(1)} \leq |D_{\tilde{k}/\mathbb{Q}}| N_{\tilde{k}/\mathbb{Q}}(f_\theta)^{1/\theta(1)}.$$

(ii) *If $\chi = \mathrm{Ind}_{\tilde{H}_n}^{G_n} \theta$, for some $\theta \in \mathrm{Irr}(\tilde{H}_n)$ then*

$$f_\chi^{1/\chi(1)} = |D_{\tilde{k}/\mathbb{Q}}|^{1/q} N_{\tilde{k}/\mathbb{Q}}(f_\theta)^{1/q\theta(1)}.$$

Proof. In the first case, we have $\chi(1) = \theta(1)$ and

$$\mathrm{Ind}_{\tilde{H}_n}^{G_n} \theta = \sum_{i=1}^{q} \psi_i(1) \cdot \chi\psi_i,$$

where $\mathrm{Irr}(G_n/\tilde{H}_n) = \{\psi_1, \psi_2, \ldots, \psi_q\}$ (see [Huppert 1998, Theorem 19.5]). Since $G_n/\tilde{H}_n$ is isomorphic to the abelian group $\mathbb{Z}/q\mathbb{Z}$, it follows that $\mathrm{Ind}_{\tilde{H}_n}^{G_n} \theta = \sum_{i=1}^{q} \chi\psi_i$. The Artin conductor of this induced character is, on the one hand,

$$f_{\mathrm{Ind}_{\tilde{H}_n}^{G_n} \theta} = |D_{\tilde{k}/\mathbb{Q}}|^{\theta(1)} N_{\tilde{k}/\mathbb{Q}}(f_\theta),$$

where the ideal f_θ is the Artin conductor of θ. On the other hand, assuming that ψ_1 is the trivial character,

$$f_{\mathrm{Ind}_{\tilde{H}_n}^{G_n} \theta} = f_{\sum_{i=1}^{q} \chi\psi_i} = f_\chi \cdot \prod_{i=2}^{q} f_{\chi\psi_i}.$$

Now, combining these expressions we get

$$f_\chi = |D_{\tilde{k}/\mathbb{Q}}|^{\theta(1)} N_{\tilde{k}/\mathbb{Q}}(f_\theta) \cdot \left(\prod_{i=2}^{q} f_{\chi\psi_i}\right)^{-1},$$

so

$$f_\chi^{1/\chi(1)} \leq |D_{\tilde{k}/\mathbb{Q}}| N_{\tilde{k}/\mathbb{Q}}(f_\theta)^{1/\theta(1)}.$$

In the second case,

$$\chi(1) = [G_n : \tilde{H}_n]\theta(1) = q\theta(1)$$

and we can see that the root Artin conductor of χ is given by the expression

$$f_\chi^{1/\chi(1)} = |D_{\tilde{k}/\mathbb{Q}}|^{1/q} N_{\tilde{k}/\mathbb{Q}}(f_\theta)^{1/q\theta(1)}. \qquad \square$$

In order to obtain a bound for the root Artin conductors, we need the following result.

Lemma 11. *Assume K_n and K as in the Proposition 7. Let $\mathfrak{p}$ be a prime in $\tilde{k}$, with $\mathfrak{b}$ and $\mathfrak{q}$ primes over $\mathfrak{p}$ in K_n and K respectively. Let $G_i(K_{n,\mathfrak{b}}/\tilde{k}_\mathfrak{p})$ and $G_i(K_\mathfrak{q}/\tilde{k}_\mathfrak{p})$ be the i-th ramification groups of the local extensions $K_{n,\mathfrak{b}}/\tilde{k}_\mathfrak{p}$ and $K_\mathfrak{q}/\tilde{k}_\mathfrak{p}$. Then, for $i \geq 0$:*

(a) $G_i(K_{n,\mathfrak{b}}/K_\mathfrak{q}) = G_i(K_{n,\mathfrak{b}}/\tilde{k}_\mathfrak{p}) \cap G(K_{n,\mathfrak{b}}/K_\mathfrak{q}) = \{1\}$.

(b) $|G_i(K_{n,\mathfrak{b}}/\tilde{k}_\mathfrak{p})| = |G_i(K_\mathfrak{q}/\tilde{k}_\mathfrak{p})|$.

The proof of this lemma follows directly from properties of higher ramification groups (see for example [Neukirch 1999, pages 177–180]) and by the fact that K_n/K is an unramified extension.

Corollary 12. *There is an infinite sequence $\{\chi_n\}_{n\in\mathbb{N}}$ of irreducible Artin characters with $\chi_n(1) \to \infty$ and with*

$$f_{\chi_n}^{1/\chi_n(1)} \leq C,$$

where $C > 0$ is an effective computable constant.

Proof. By the Corollary 9 and Theorem 10, we know that for each n there is an irreducible character χ_n of G_n with $\chi_n(1) \to \infty$ and

$$f_{\chi_n}^{1/\chi_n(1)} \leq |D_{\tilde{k}/\mathbb{Q}}| N_{\tilde{k}/\mathbb{Q}}(f_\theta)^{1/\theta(1)},$$

for some $\theta \in \mathrm{Irr}(\tilde{H}_n)$. By the properties of the higher ramification groups stated in Lemma 11 and considering that the primes ramifying in K are the only ones that appears in $N_{\tilde{k}/\mathbb{Q}}(f_\theta)$, it is possible to find a constant $T > 0$ depending only on the base field K, such that $N_{\tilde{k}/\mathbb{Q}}(f_\theta) \leq T^{\theta(1)}$. Hence,

$$f_{\chi_n}^{1/\chi_n(1)} \leq |D_{\tilde{k}/\mathbb{Q}}| T := C. \qquad \square$$

Remark 13. As the referee pointed out, it is possible to avoid the hypothesis about the degree of $\tilde{k}/\mathbb{Q}$ and obtain the same type of bounds for the asymptotic behavior of $f_\chi^{1/\chi(1)}$. This is accomplished in Theorem 14 below.

Theorem 14. *Let K be a Galois extension of $\mathbb{Q}$ with infinite p-class field tower. Let $m = [K : \mathbb{Q}]$. Then there exists an infinite sequence $\{\chi_n\}_{n\in\mathbb{N}}$ of irreducible Artin characters such that $\chi_n(1) \to \infty$ and*

$$f_{\chi_n}^{1/\chi_n(1)} \leq |D_{K/\mathbb{Q}}|.$$

Proof. Let $G_n = \mathrm{Gal}(K_n/\mathbb{Q})$ and $T_n = \mathrm{Gal}(K_n/K)$. We can choose $\theta_n \in \mathrm{Irr}(T_n)$ as in Proposition 8. Then, by the Proposition 3, there exists $\chi_n \in \mathrm{Irr}(G_n)$ such that $\langle \mathrm{Ind}_{H_n}^{G_n} \theta_n, \chi_n \rangle = a \geq 1$ and with $\chi_n(1) \geq \theta_n(1)$, so

$$\chi_n(1) > 2^{(n-1)/2}.$$

Hence, by the properties of the Artin conductor we get

$$f_{\chi_n}^{a} \leq f_{\mathrm{Ind}_{H_n}^{G_n} \theta_n} = |D_{K/\mathbb{Q}}|^{\theta_n(1)},$$

and therefore,

$$f_{\chi_n}^{1/\chi_n(1)} \leq f_{\chi_n}^{a/\chi_n(1)} \leq |D_{K/\mathbb{Q}}|.$$

□

4. Number fields with infinite 2-class field tower

Golod and Shafarevich [1964] proved that a number field K has an infinite p-class field tower if the p-rank of the class group of K is large enough. In this case,

$$\alpha(r_1, r_2) \leq |D_K|^{1/[K:\mathbb{Q}]},$$

where D_K is the discriminant of K.

In addition, Martinet has constructed a number field with infinite Hilbert class field towers and lowest known root discriminant and proved that

$$\alpha(0, 1) < 93 \quad \text{and} \quad \alpha(1, 0) < 1059.$$

In particular, he found that $K = \mathbb{Q}(\zeta_{11} + \zeta_{11}^{-1}, \sqrt{2}, \sqrt{-23})$ has infinite 2-class field tower. Since $\tilde{k} = \mathbb{Q}(\zeta_{11} + \zeta_{11}^{-1})$ is a subfield of K of degree 5 over $\mathbb{Q}$, K satisfies the conditions of the Theorem 10. The discriminant of $\tilde{k}$ is

$$|D_{\tilde{k}/\mathbb{Q}}| = 14641 = 11^4$$

and the only rational primes that ramify in K are 2, 11 and 23. Using PARI/GP [PARI 2014], we can estimates the sizes of the higher ramification groups. Thus, we get the upper bound

$$N_{\tilde{k}/\mathbb{Q}}(f_\theta) \leq (2^{15}23)^{\theta(1)}.$$

With this estimation, we get the following explicit result:

Corollary 15. *For each $n \geq 1$, there exists a irreducible character χ_n such that $\chi_n(1) \to \infty$ and*

$$f_{\chi_n}^{1/\chi_n(1)} \leq C, \quad \text{where } C \leq 11^4 \cdot 2^{15} \cdot 23.$$

An open problem now is to improve the constant C.

Acknowledgements

This research was partially supported by the "Red Iberoamericana de Teoría de Números", ECOS-CONICYT grant 170022 "Explicit Arithmetic Geometry" and IdEx-Université de Bordeaux (credits to Benjamin Matschke). The author would like to thank Eduardo Friedman for suggesting this problem and his helpful advice. The author also thanks Bill Allombert, Yuri Bilu, Mariela Carvacho, Milton Espinoza and Andrea Vera for useful discussions. Finally, the author thanks the referees for the comments and in particular for suggesting Proposition 3 and Theorem 14. While working on this project, the author visited the ICMAT Madrid and IMB Bordeaux.

References

[Gallagher 1962] P. X. Gallagher, "Group characters and normal Hall subgroups", *Nagoya Math. J.* **21** (1962), 223–230. MR Zbl

[Golod and Shafarevich 1964] E. S. Golod and I. R. Shafarevich, "On the class field tower", *Izv. Akad. Nauk SSSR Ser. Mat.* **28** (1964), 261–272. In Russian. MR Zbl

[Huppert 1998] B. Huppert, *Character theory of finite groups*, de Gruyter Expos. Math. **25**, de Gruyter, Berlin, 1998. MR Zbl

[Isaacs 1976] I. M. Isaacs, *Character theory of finite groups*, Pure Appl. Math. **69**, Academic Press, New York, 1976. MR Zbl

[Martinet 1978] J. Martinet, "Tours de corps de classes et estimations de discriminants", *Invent. Math.* **44**:1 (1978), 65–73. MR Zbl

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundlehren der Math. Wissenschaften **322**, Springer, 1999. MR Zbl

[Odlyzko 1977] A. M. Odlyzko, "On conductors and discriminants", pp. 377–407 in *Algebraic number fields: L-functions and Galois properties* (Durham, UK, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR Zbl

[PARI 2014] PARI Group, PARI/GP version 2.7.1, 2014, Available at http://pari.math.u-bordeaux.fr/.

[Pizarro-Madariaga 2011] A. Pizarro-Madariaga, "Lower bounds for the Artin conductor", *Math. Comp.* **80**:273 (2011), 539–561. MR Zbl

Communicated by Brian Conrad
Received 2017-11-29 Revised 2019-05-22 Accepted 2019-06-25

amalia.pizarro@uv.cl *Instituto de Matemáticas, Universidad de Valparaíso, Chile*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2005

Frobenius–Perron theory of endofunctors

Jianmin Chen, Zhibin Gao, Elizabeth Wicks, James J. Zhang, Xiaohong Zhang and Hong Zhu

We introduce the Frobenius–Perron dimension of an endofunctor of a $\Bbbk$-linear category and provide some applications.

0. Introduction

The spectral radius (also called the Frobenius–Perron dimension) of a matrix is an elementary and extremely useful invariant in linear algebra, combinatorics, topology, probability and statistics. The Frobenius–Perron dimension has become a crucial concept in the study of fusion categories and representations of semisimple weak Hopf algebras since it was introduced by Etingof, Nikshych and Ostrik [Etingof et al. 2005] (also see [Etingof et al. 2004; 2015; Nikshych 2004]). In this paper several Frobenius–Perron type invariants are proposed to study derived categories, representations of finite dimensional algebras, and complexity of algebras and categories.

Throughout let $\Bbbk$ be an algebraically closed field, and let everything be over $\Bbbk$.

Definitions. The first goal is to introduce the Frobenius–Perron dimension of an endofunctor of a category. Here we only sketch the definition of $\operatorname{fpd}(\sigma)$ for an endofunctor σ of an abelian category $\mathcal{C}$ and the precise definition is given in Definition 2.3(2). Let $\phi := \{X_1, \dots, X_n\}$ be a finite subset of nonzero objects in $\mathcal{C}$ such that

$$\operatorname{Hom}_{\mathcal{C}}(X_i, X_j) = \begin{cases} \Bbbk & i = j, \\ 0 & i \neq j. \end{cases}$$

Let $\rho(A(\phi, \sigma))$ denote the spectral radius of the $n \times n$-matrix $[\dim \operatorname{Hom}_{\mathcal{C}}(X_i, \sigma(X_j))]_{n\times n}$. The *Frobenius–Perron dimension* of σ is defined to be

$$\operatorname{fpd}(\sigma) = \sup_{\phi}\{\rho(A(\phi, \sigma))\}$$

where ϕ ranges over all finite subsets of $\mathcal{C}$ satisfying the condition mentioned above. If an object V in a fusion category $\mathcal{C}$ is considered as the associated tensor endofunctor $V \otimes_{\mathcal{C}} -$, then our definition of the Frobenius–Perron dimension agrees with the definition given in [Etingof et al. 2005], see Example 2.11 for details. Our definition applies to the derived category of projective schemes and finite dimensional algebras, as well as other abelian and additive categories (Definitions 2.3 and 2.4). We also refer the reader to Section 2 for the following invariants of an endofunctor:

MSC2010: primary 16E10,18E30,16G60; secondary 16E35.
Keywords: Frobenius–Perron dimension, derived categories, embedding of categories, tame and wild dichotomy, complexity.

Frobenius–Perron growth (denoted by fpg).

Frobenius–Perron curvature (denoted by fpv).

Frobenius–Perron series (denoted by FP).

One can further define the above invariants for an abelian or a triangulated category. Note that the Frobenius–Perron dimension/growth/curvature of a category can be a noninteger, see Proposition 5.12(1), Example 8.7, and Remarks 5.13(5) for nonintegral values of fpd, fpg, and fpv respectively.

If $\mathfrak{A}$ is an abelian category, let $D^b(\mathfrak{A})$ denote the bounded derived category of $\mathfrak{A}$. On the one hand it is reasonable to call fpd a dimension function since

$$\operatorname{fpd}(D^b(\operatorname{Mod}-\Bbbk[x_1,\dots,x_n]))=n$$

(Proposition 4.3(1)), but on the other hand, one might argue that fpd should not be called a dimension function since

$$\operatorname{fpd}(D^b(\operatorname{coh}(\mathbb{P}^n)))=\begin{cases}1 & n=1,\\ \infty & n\geq 2,\end{cases}$$

(Propositions 6.5 and 6.7). In the latter case, fpd is an indicator of representation type of the category of $\operatorname{coh}(\mathbb{P}^n)$, namely, $\operatorname{coh}(\mathbb{P}^n)$ is tame if $n=1$, and is of wild representation type for all $n\geq 2$. A similar statement holds for projective curves in terms of genus (Proposition 6.5).

We can define the Frobenius–Perron ("fp") version of several other classical invariants:

fp global dimension (denoted by fpgldim, Definition 2.7(1)).

fp Kodaira dimension (denoted by fp κ) [Chen et al. 2019].

The first one is defined for all triangulated categories and the second one is defined for triangulated categories with Serre functor. In general, the fpgldim A does not agree with the classical global dimension of A (Theorem 7.8). The fp version of the Kodaira dimension agrees with the classical definition for smooth projective schemes [Chen et al. 2019].

Our second goal is to provide several applications.

Embeddings. In addition to the fact that the Frobenius–Perron dimension is an effective and sensible invariant of many categories, this invariant increases when the "size" of the endofunctors and categories increase.

Theorem 0.1. *Suppose $\mathcal{C}$ and $\mathcal{D}$ are $\Bbbk$-linear categories. Let $F:\mathcal{C}\to\mathcal{D}$ be a fully faithful functor. Let $\sigma_{\mathcal{C}}$ and $\sigma_{\mathcal{D}}$ be endofunctors of $\mathcal{C}$ and $\mathcal{D}$ respectively. Suppose that $F\circ\sigma_{\mathcal{C}}$ is naturally isomorphic to $\sigma_{\mathcal{D}}\circ F$. Then* $\operatorname{FP}(u,t,\sigma_{\mathcal{C}})\leq\operatorname{FP}(u,t,\sigma_{\mathcal{D}})$.

See Theorem 3.2 for the proof. By taking σ to be the suspension functor of a pretriangulated category [Neeman 2001, Definition 1.1.2], we have the following immediate consequence. (Note that the fp-dimension of a triangulated category $\mathcal{T}$ is defined to be $\operatorname{fpd}(\Sigma)$, where Σ is the suspension of $\mathcal{T}$.)

Corollary 0.2. *Let $\mathcal{T}_2$ be a pretriangulated category and $\mathcal{T}_1$ a full pretriangulated subcategory of $\mathcal{T}_2$. Then the following hold*:

(1) $\operatorname{fpd} \mathcal{T}_1 \le \operatorname{fpd} \mathcal{T}_2$.

(2) $\operatorname{fpg} \mathcal{T}_1 \le \operatorname{fpg} \mathcal{T}_2$.

(3) $\operatorname{fpv} \mathcal{T}_1 \le \operatorname{fpv} \mathcal{T}_2$.

(4) *If $\mathcal{T}_2$ has* fp-*subexponential growth*, *so does $\mathcal{T}_1$*.

Fully faithful embeddings of derived categories of projective schemes have been investigated in the study of Fourier–Mukai transforms, birational geometry, and noncommutative crepant resolutions (NCCRs) by Bondal and Orlov [2001; 2002], Van den Bergh [2004], Bridgeland [2002], Bridgeland, King and Reid [Bridgeland et al. 2001] and more.

Note that if $\operatorname{fpgldim}(\mathcal{T}) < \infty$, then $\operatorname{fpg}(\mathcal{T}) = 0$. If $\operatorname{fpg}(\mathcal{T}) < \infty$, then $\operatorname{fpv}(\mathcal{T}) \le 1$. Hence, fpd, fpgldim, fpg and fpv measure the "size", "representation type", or "complexity" of a triangulated category $\mathcal{T}$ at different levels. Corollary 0.2 has many consequences concerning nonexistence of fully faithful embeddings provided that we compute the invariants fpd, fpg and fpv of various categories efficiently.

Tame vs wild. Here we mention a couple of more applications. First we extend the classical trichotomy on the representation types of quivers to the fpd. A proof of the following theorem is given in Section 7.

Theorem 0.3. *Let Q be a finite quiver and let $\mathcal{Q}$ be the bounded derived category of finite dimensional left $\Bbbk Q$-modules*:

(1) *$\Bbbk Q$ is of finite representation type if and only if* $\operatorname{fpd} \mathcal{Q} = 0$.

(2) *$\Bbbk Q$ is of tame representation type if and only if* $\operatorname{fpd} \mathcal{Q} = 1$.

(3) *$\Bbbk Q$ is of wild representation type if and only if* $\operatorname{fpd} \mathcal{Q} = \infty$.

By the classical theorems of Gabriel [1972] and Nazarova [1973], the quivers of finite and tame representation types correspond to the ADE and $\tilde{A}\tilde{D}\tilde{E}$ diagrams respectively.

The above theorem fails for quiver algebras with relations (Proposition 5.12). As we have already seen, fpd is related to the "size" of a triangulated category, as well as, the representation types. We will see soon that fpg is also closely connected with the complexity of representations. When we focus on the representation types, we make some tentative definitions.

Let $\mathcal{T}$ be a triangulated category (such as $D^b(\operatorname{Mod}_{f.d.} - A)$):

(i) We call $\mathcal{T}$ fp-*trivial*, if $\operatorname{fpd} \mathcal{T} = 0$.

(ii) We call $\mathcal{T}$ fp-*tame*, if $\operatorname{fpd} \mathcal{T} = 1$.

(iii) We call $\mathcal{T}$ fp-*potentially wild*, if $\operatorname{fpd} \mathcal{T} > 1$. Further:

(a) $\mathcal{T}$ is fp-*finitely wild*, if $1 < \operatorname{fpd} \mathcal{T} < \infty$.

(b) $\mathcal{T}$ is fp-*locally-finitely wild*, if $\operatorname{fpd} \mathcal{T} = \infty$ and $\operatorname{fpd}^n(\mathcal{T}) < \infty$ for all n.

(c) $\mathcal{T}$ is fp-*wild*, if $\operatorname{fpd}^1 \mathcal{T} = \infty$.

There are other notions of tame/wildness in representation theory, see for example, [Geiss and Krause 2002; Drozd 2004]. Following the above definition, fpd provides a quantification of the tame-wild dichotomy. By Theorem 0.3, finite/tame/wild representation types of the path algebra $\Bbbk Q$ are equivalent to the fp-version of these properties of $\mathcal{Q}$. Let A be a quiver algebra with relations and let $\mathcal{A}$ be the derived category $D^b(\mathrm{Mod}_{f.d.} - A)$. Then, in general, finite/tame/wild representation types of A are NOT equivalent to the fp-version of these properties of $\mathcal{A}$ (Example 5.5). It is natural to ask

Question 0.4. For which classes of algebras A, is the fp-wildness of $\mathcal{A}$ equivalent to the classical and other wildness of A in representation theory literature?

Complexity. The complexity of a module or of an algebra is an important invariant in studying representations of finite dimensional algebras [Alperin and Evens 1981; Carlson 1996; Carlson et al. 1994; Guo et al. 2009]. Let A be the quiver algebra $\Bbbk Q/(R)$ with relations R. The *complexity* of A is defined to be the complexity of the A-module $T := A/\operatorname{Jac}(A)$, namely,

$$\operatorname{cx}(A) = \operatorname{cx}(T) := \limsup_{n\to\infty} \log_n(\dim \operatorname{Ext}^n_A(T, T)) + 1.$$

Let GKdim denote the Gelfand–Kirillov dimension of an algebra (see [Krause and Lenagan 1985] and [McConnell and Robson 1987, Chapter 8]). Under some reasonable hypotheses, one can show

$$\operatorname{cx}(A) = \operatorname{GKdim}\Bigl(\bigoplus_{n=0}^{\infty} \operatorname{Ext}^n_A(T, T)\Bigr).$$

It is easy to see that $\operatorname{cx}(A)$ is an derived invariant. We extend the definition of the complexity to any triangulated category (Definition 8.2(4)).

Theorem 0.5. *Let A be a finite dimensional quiver algebra $\Bbbk Q/(R)$ with relations R and let $\mathcal{A}$ be the bounded derived category of finite dimensional left A-modules. Then*

$$\operatorname{fpg}(\mathcal{A}) \le \operatorname{cx}(A) - 1.$$

This theorem is a consequence of Theorems 8.3 and 8.4(1). The equality $\operatorname{fpg}(\mathcal{A}) = \operatorname{cx}(A) - 1$ holds under some hypotheses (Theorem 8.4(2)).

Frobenius–Perron function. If $\mathcal{T}$ is a triangulated category with Serre functor S, we have an fp-function

$$\mathrm{fp} : \mathbb{Z}^2 \to \mathbb{R} \cup \{\pm\infty\}$$

which is defined by

$$\mathrm{fp}(a, b) := \operatorname{fpd}(\Sigma^a \circ S^b) \in \mathbb{R} \cup \{\pm\infty\}.$$

Then $\operatorname{fpd}(\mathcal{T})$ is the value of the fp-function at $(1, 0)$.

The fp-function for the projective line $\mathbb{P}^1$ and the quiver A_2 are given in the Examples 5.1 and 5.4 respectively.

The statements in Theorem 0.3, Questions 0.4 and 7.11 indicate that $\operatorname{fp}(1,0)$ predicts the representation type of $\mathcal{T}$ for certain triangulated categories. It is expected that values of the fp-function at other points in $\mathbb{Z}^2$ are sensitive to other properties of $\mathcal{T}$.

Properties. The paper contains some basic properties of fpd. Let us mention one of them, whose proof can be found in Proposition 3.6.

Proposition 0.6 (Serre duality). *Let $\mathcal{C}$ be a* Hom*-finite category with Serre functor S. Let σ be an endofunctor of $\mathcal{C}$:*

(1) *If σ has a right adjoint $\sigma^!$, then*

$$\operatorname{fpd}(\sigma) = \operatorname{fpd}(\sigma^! \circ S).$$

(2) *If σ is an equivalence with quasiinverse σ^{-1}, then*

$$\operatorname{fpd}(\sigma) = \operatorname{fpd}(\sigma^{-1} \circ S).$$

(3) *If $\mathcal{C}$ is n-Calabi–Yau, then we have a duality, for all i,*

$$\operatorname{fpd}(\Sigma^i) = \operatorname{fpd}(\Sigma^{n-i}).$$

Computations. Our third goal is to develop methods for computation. To use fp-invariants, we need to compute as many examples as possible. In general it is extremely difficult to calculate useful invariants for derived categories, as the definitions of these invariants are quite sophisticated. We develop some techniques for computing fp-invariants. In Sections 4–5, we compute the fp-dimension for some nontrivial examples.

Other significant applications. In addition to the results above, various Frobenius–Perron invariants of endofunctors have applications in study of other important objects/structures such as tensor triangulated categories in the sense of [Balmer 2005, Definition 1.1]. Let Q be a finite acyclic quiver and $\Bbbk Q$ be its path algebra. Let $\mathcal{T}_Q$ denote the bounded derived category $D^b(\mathrm{Mod}_{f.d.} - \Bbbk Q)$ of finite dimensional representations of Q. Note that every path algebra $\Bbbk Q$ of a finite quiver Q is naturally equipped with a weak bialgebra structure (where the coalgebra structure is similar to the one given in [Nikshych and Vainerman 2002, Example 2.5]), which implies that $\mathcal{T}_Q$ is a tensor triangulated category. One significant application of $\operatorname{fpv}(\sigma)$ Definition 2.3(4) (for various endofunctors σ) is to prove that two nonisomorphic acyclic finite quivers are not tensor triangulated equivalent. For example, let Q_1 and Q_2 be two nonisomorphic quivers of the same underlying ADE Dynkin graph. It is well-known that $\mathcal{T}_{Q_1}$ and $\mathcal{T}_{Q_2}$ are triangulated equivalent via Bernstein, Gelfand and Ponomarev reflection functors [Bernstein et al. 1973] (also see [Happel 1987]). Now using $\operatorname{fpv}(\sigma)$ it can be shown that $\mathcal{T}_{Q_1}$ and $\mathcal{T}_{Q_2}$ are *not* tensor triangulated equivalent. Details are given in [Zhang and Zhou $\geq$ 2019]. By using other known invariants such as the Balmer spectrum [2005], it is difficult for us to distinguish the tensor triangulated structures of $\mathcal{T}_{Q_1}$ and $\mathcal{T}_{Q_2}$ where these are triangulated equivalent.

Conventions.

(1) Usually Q means a quiver.

(2) $\mathcal{T}$ is a (pre-)triangulated category with suspension functor $\Sigma = [1]$.

(3) If A is an algebra over the base field $\Bbbk$, then $\mathrm{Mod}_{f.d.} - A$ denotes the category of finite dimensional left A-modules.

(4) If A is an algebra, then we use $\mathfrak{A}$ for the abelian category $\mathrm{Mod}_{f.d.} - A$.

(5) When $\mathfrak{A}$ is an abelian category, we use $\mathcal{A}$ for the bounded derived category $D^b(\mathfrak{A})$.

This paper is organized as follows. We provide background material in Section 1. The basic definitions are introduced in Section 2. Some basic properties are given in Section 3. We prove Theorem 0.1 and Proposition 0.6 in Section 3, see Theorem 3.2 and Proposition 3.6 respectively. Corollary 0.2 is an immediate consequence of Theorem 0.1. Section 4 deals with some derived categories of modules over commutative rings. In Section 5, we work out the fp-theories of the projective line and quiver A_2, as well as an example of nonintegral fpd. In Section 6, we develop some techniques to handle the fpd of projective curves and prove the tame-wild dichotomy of projective curves in terms of fpd. Theorem 0.3 is proved in Section 7 where representation types are discussed. Section 8 focuses on the complexity of algebras and categories and Theorem 0.5 is proved there. We continue to develop the fp-theory in our companion paper [Chen et al. 2019]. Some examples can be found in [Wicks 2019; Zhang and Zhou $\geq$ 2019].

1. Preliminaries

Classical definitions. Let A be an $n \times n$-matrix over complex numbers $\mathbb{C}$. The *spectral radius* of A is defined to be

$$\rho(A) := \max\{|r_1|, |r_2|, \ldots, |r_n|\} \in \mathbb{R}$$

where $\{r_1, r_2, \ldots, r_n\}$ is the complete set of eigenvalues of A. When each entry of A is a positive real number, $\rho(A)$ is also called the *Perron root* or the *Perron–Frobenius eigenvalue* of A. When applying ρ to the adjacency matrix of a graph (or a quiver), the spectral radius of the adjacency matrix is sometimes called the *Frobenius–Perron dimension* of the graph (or the quiver).

Let us mention a classical result concerning the spectral radius of simple graphs. A finite graph G is called *simple* if it has no loops and no multiple edges. Smith [1970] formulated the following result:

Theorem 1.1 [Dokuchaev et al. 2013, Theorem 1.3]. *Let G be a finite, simple, and connected graph with adjacency matrix A:*

(1) *$\rho(A) = 2$ if and only if G is one of the extended Dynkin diagrams of type $\tilde{A}\tilde{D}\tilde{E}$.*

(2) *$\rho(A) < 2$ if and only if G is one of the Dynkin diagrams of type ADE.*

To save space we refer to [Dokuchaev et al. 2013] and [Happel et al. 1980] for the diagrams of the ADE and $\tilde{A}\tilde{D}\tilde{E}$ quivers/graphs.

In order to include some infinite-dimensional cases, we extend the definition of the spectral radius in the following way.

Let $A := (a_{ij})_{n\times n}$ be an $n \times n$-matrix with entries a_{ij} in $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. Define $A' = (a'_{ij})_{n\times n}$ where

$$a'_{ij} = \begin{cases} a_{ij} & a_{ij} \neq \pm\infty, \\ x_{ij} & a_{ij} = \infty, \\ -x_{ij} & a_{ij} = -\infty. \end{cases}$$

In other words, we are replacing ∞ in the (i, j)-entry by a finite real number, called x_{ij}, in the (i, j)-entry. And every x_{ij} is considered as a variable or a function mapping $\mathbb{R} \to \mathbb{R}$.

Definition 1.2. Let A be an $n \times n$-matrix with entries in $\overline{\mathbb{R}}$. The *spectral radius* of A is defined to be

$$\rho(A) := \liminf_{\text{all } x_{ij}\to\infty} \rho(A') \in \overline{\mathbb{R}}. \tag{E1.2.1}$$

Remark 1.3. It also makes sense to use lim sup instead of lim inf in (E1.2.1). We choose to take lim inf in this paper.

Here is an easy example.

Example 1.4. Let $A = \left(\begin{smallmatrix} 1 & -\infty \\ 0 & 2 \end{smallmatrix}\right)$. Then $A' = \left(\begin{smallmatrix} 1 & -x_{12} \\ 0 & 2 \end{smallmatrix}\right)$. It is obvious that

$$\rho(A) = \lim_{x_{12}\to\infty} \rho(A') = \lim_{x_{12}\to\infty} 2 = 2.$$

$\Bbbk$-linear categories. If $\mathcal{C}$ is a $\Bbbk$-linear category, then $\mathrm{Hom}_{\mathcal{C}}(M, N)$ is a $\Bbbk$-module for all objects M, N in $\mathcal{C}$. If $\mathcal{C}$ is also abelian, then $\mathrm{Ext}^i_{\mathcal{C}}(M, N)$ are $\Bbbk$-modules for all $i \geq 0$. Let dim be the $\Bbbk$-vector space dimension.

Remark 1.5. One can generalize the notion of fpd to categories that are not $\Bbbk$-linear. Even when a category $\mathcal{C}$ is not $\Bbbk$-linear, it might still make sense to define a set map on the Hom-sets of the category $\mathcal{C}$, say

$$\partial : \{\mathrm{Hom}_{\mathcal{C}}(M, N) \mid M, N \in \mathcal{C}\} \to \mathbb{Z}_{\geq 0} \cup \{\infty\}.$$

We call such a map a *dimension function*. The definition of Frobenius–Perron dimension given in the next section can be modified using ∂ instead of dim to fit this very weak version of a dimension function.

Frobenius–Perron dimension of a quiver. In this subsection we recall some known elementary definitions and facts.

Definition 1.6. Let Q be a quiver:

(1) If Q has finitely many vertices, then the *Frobenius–Perron dimension* of Q is defined to be

$$\operatorname{fpd} Q := \rho(A(Q))$$

where $A(Q)$ is the adjacency matrix of Q.

(2) Let Q be any quiver. The *Frobenius–Perron dimension* of Q is defined to be

$$\operatorname{fpd} Q := \sup\{\operatorname{fpd} Q'\}$$

where Q' runs over all finite subquivers of Q.

See [Erdmann and Solberg 2011, Propositions 2.1 and 3.2] for connections between fpd of a quiver and its representation types, as well as its complexity. We need the following well-known facts in linear algebra.

Lemma 1.7. (1) *Let B be a square matrix with nonnegative entries and let A be a principal minor of B. Then $\rho(A) \leq \rho(B)$.*

(2) *Let $A := (a_{ij})_{n\times n}$ and $B := (b_{ij})_{n\times n}$ be two square matrices such that $0 \leq a_{ij} \leq b_{ij}$ for all i, j. Then $\rho(A) \leq \rho(B)$.*

Let Q be a quiver with vertices $\{v_1, \dots, v_n\}$. An oriented cycle based at a vertex v_i is called *indecomposable* if it is not a product of two oriented cycles based at v_i. For each vertex v_i let θ_i be the number of indecomposable oriented cycles based at v_i. Define the *cycle number* of a quiver Q to be

$$\Theta(Q) := \max\{\theta_i \mid \forall i\}.$$

The following result should be well known.

Theorem 1.8. *Let Q be a quiver and let $\Theta(Q)$ be the cycle number of Q:*

(1) $\operatorname{fpd}(Q) = 0$ *if and only if* $\Theta(Q) = 0$, *namely, Q is acyclic.*

(2) $\operatorname{fpd}(Q) = 1$ *if and only if* $\Theta(Q) = 1$.

(3) $\operatorname{fpd}(Q) > 1$ *if and only if* $\Theta(Q) \geq 2$.

The proof is not hard, and to save space, it is omitted.

2. Definitions

Throughout the rest of the paper, let $\mathcal{C}$ denote a $\Bbbk$-linear category. A functor between two $\Bbbk$-linear categories is assumed to preserve the $\Bbbk$-linear structure. For simplicity, $\dim(A, B)$ stands for $\dim \operatorname{Hom}_{\mathcal{C}}(A, B)$ for any objects A and B in $\mathcal{C}$.

The set of finite subsets of nonzero objects in $\mathcal{C}$ is denoted by Φ and the set of subsets of n nonzero objects in $\mathcal{C}$ is denoted by Φ_n for each $n \geq 1$. It is clear that $\Phi = \bigcup_{n\geq 1} \Phi_n$. We do not consider the empty set as an element of Φ.

Definition 2.1. Let $\phi := \{X_1, X_2, \dots, X_n\}$ be a finite subset of nonzero objects in $\mathcal{C}$, namely, $\phi \in \Phi_n$. Let σ be an endofunctor of $\mathcal{C}$:

(1) The *adjacency matrix* of (ϕ, σ) is defined to be

$$A(\phi, \sigma) := (a_{ij})_{n\times n} \quad \text{where } a_{ij} := \dim(X_i, \sigma(X_j)) \ \forall i, j.$$

(2) An object M in $\mathcal{C}$ is called a *brick* [Assem et al. 2006, Definition 2.4, Chapter VII] if

$$\operatorname{Hom}_{\mathcal{C}}(M, M) = \Bbbk.$$

[Neeman 2001, Definition 1.1.2], an object M in $\mathcal{C}$ is called an *atomic* object if it is a brick and satisfies

$$\operatorname{Hom}_{\mathcal{C}}(M, \Sigma^{-i}(M)) = 0 \quad \forall i > 0. \tag{E2.1.1}$$

(3) $\phi \in \Phi$ is called a *brick set* (respectively, an *atomic set*) if each X_i is a brick (respectively, atomic) and

$$\dim(X_i, X_j) = \delta_{ij}$$

for all $1 \le i, j \le n$. The set of brick (respectively, atomic) n-object subsets is denoted by $\Phi_{n,b}$ (respectively, $\Phi_{n,a}$). We write $\Phi_b = \bigcup_{n\ge1} \Phi_{n,b}$ (respectively, $\Phi_a = \bigcup_{n\ge1} \Phi_{n,a}$). Define the *b-height* of $\mathcal{C}$ to be

$$h_b(\mathcal{C}) = \sup\{n \mid \Phi_{n,b} \text{ is nonempty}\}$$

and the *a-height* of $\mathcal{C}$ (when $\mathcal{C}$ is pretriangulated) to be

$$h_a(\mathcal{C}) = \sup\{n \mid \Phi_{n,a} \text{ is nonempty}\}.$$

Remarks 2.2. (1) A brick may not be atomic. Let A be the algebra

$$\Bbbk\langle x, y\rangle/(x^2, y^2 - 1, xy + yx).$$

This is a 4-dimensional Frobenius algebra (of injective dimension zero). There are two simple left A-modules

$$S_0 := A/(x, y-1), \quad \text{and} \quad S_1 := A/(x, y+1).$$

Let M_i be the injective hull of S_i for $i = 0, 1$. (Since A is Frobenius, M_i is projective.) There are two short exact sequences

$$0 \to S_0 \to M_0 \xrightarrow{f} S_1 \to 0 \quad \text{and} \quad 0 \to S_1 \xrightarrow{g} M_1 \to S_0 \to 0.$$

It is easy to check that $\operatorname{Hom}_A(M_i, M_j) = \Bbbk$ for all $0 \le i, j \le 1$. Let $\mathcal{A}$ be the derived category $D^b(\operatorname{Mod}_{f.d.} - A)$ and let X be the complex

$$\cdots \to 0 \to M_0 \xrightarrow{g\circ f} M_1 \to 0 \to \cdots$$

An easy computation shows that $\operatorname{Hom}_{\mathcal{A}}(X, X) = \Bbbk = \operatorname{Hom}_{\mathcal{A}}(X, X[-1])$. So X is a brick, but not atomic.

(2) A brick object is called a *Schur* object by several authors, see [Carroll and Chindris 2015; Chindris et al. 2015]. It is also called *endosimple* by others, see [van Roosmalen 2008; 2016].

(3) The definition of an atomic object in a triangulated category is similar to (and slightly weaker than) the definition of a point-object given by Bondal and Orlov [2001, Definition 2.1]. In particular, an atomic object only satisfies (ii) and (iii) of that definition with $k(P) = \Bbbk$. Note that a point-object is defined on a

triangulated category with Serre functor. In this paper we do not automatically assume the existence of a Serre functor in general.

Definition 2.3. Retain the notation as in Definition 2.1, and we use Φ_b as the testing objects. When $\mathcal{C}$ is a pretriangulated category, Φ_b is automatically replaced by Φ_a unless otherwise stated:

(1) The *n-th Frobenius–Perron dimension* of σ is defined to be

$$\operatorname{fpd}^n(\sigma) := \sup_{\phi \in \Phi_{n,b}} \{\rho(A(\phi, \sigma))\}.$$

If $\Phi_{n,b}$ is empty, then by convention, $\operatorname{fpd}^n(\sigma) = -\infty$.

(2) The *Frobenius–Perron dimension* of σ is defined to be

$$\operatorname{fpd}(\sigma) := \sup_n \{\operatorname{fpd}^n(\sigma)\} = \sup_{\phi \in \Phi_b} \{\rho(A(\phi, \sigma))\}.$$

(3) The *Frobenius–Perron growth* of σ is defined to be

$$\operatorname{fpg}(\sigma) := \sup_{\phi \in \Phi_b} \{\limsup_{n \to \infty} \log_n(\rho(A(\phi, \sigma^n)))\}.$$

By convention, $\log_n 0 = -\infty$.

(4) The *Frobenius–Perron curvature* of σ is defined to be

$$\operatorname{fpv}(\sigma) := \sup_{\phi \in \Phi_b} \{\limsup_{n \to \infty} (\rho(A(\phi, \sigma^n)))^{1/n}\}.$$

This is motivated by the concept of the *curvature* of a module over an algebra due to Avramov [1998].

(5) We say σ has fp-*exponential growth* (respectively, fp-*subexponential growth*) if $\operatorname{fpv}(\sigma) > 1$ (respectively, $\operatorname{fpv}(\sigma) \leq 1$).

In this above definition, we implicitly assume that

the isom-classes of brick objects (*respectively, atomic objects*) *form a set,*

otherwise, $\sup_{\phi \in \Phi_b}$ (respectively, $\sup_{\phi \in \Phi_a}$) is not defined. This assumption is automatic if the category $\mathcal{C}$ is essentially small. But, even when $\mathcal{C}$ is not essentially small, one can check the above assumption in many cases.

Sometimes we prefer to have all information from the Frobenius–Perron dimension. We make the following definition.

Definition 2.4. Let $\mathcal{C}$ be a category and σ be an endofunctor of $\mathcal{C}$:

(1) The *Frobenius–Perron theory* (or fp-theory) of σ is defined to be the set

$$\{\operatorname{fpd}^n(\sigma^m)\}_{n \geq 1, m \geq 0}.$$

(2) The *Frobenius–Perron series* (or fp-series) of σ is defined to be

$$\mathrm{FP}(u,t,\sigma) := \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} \mathrm{fpd}^n(\sigma^m)t^m u^n.$$

Remark 2.5. To define Frobenius–Perron dimension, one only needs have an assignment $\tau : \Phi_n \to M_{n\times n}(\mathrm{Mod}-\Bbbk)$, for every $n \geq 1$, satisfying the property

if ϕ_1 is a subset of ϕ_2, then $\tau(\phi_1)$ is a principal submatrix of $\tau(\phi_2)$.

Then we define the adjacency matrix of $\phi \in \Phi_n$ to be

$$A(\phi,\tau) = (a_{ij})_{n\times n} \quad \text{where } a_{ij} = \dim(\tau(\phi))_{ij}\ \forall i, j.$$

Then the Frobenius–Perron dimension of τ is defined in the same way as in Definition 2.3. If there is a sequence of τ_m, the Frobenius–Perron series of $\{\tau_m\}$ is defined in the same way as in Definition 2.4 by replacing σ^m by τ_m. See Example 2.6 next.

Example 2.6. (1) Let $\mathfrak{A}$ be a $\Bbbk$-linear abelian category. For each $m \geq 1$ and $\phi = \{X_1, \dots, X_n\}$, define

$$E^m : \phi \to (\mathrm{Ext}^m_{\mathfrak{A}}(X_i, X_j))_{n\times n}.$$

By convention, let $\mathrm{Ext}^0_{\mathfrak{A}}(X_i, X_j)$ denote $\mathrm{Hom}_{\mathfrak{A}}(X_i, X_j)$. Then, for each $m \geq 0$, one can define the Frobenius–Perron dimension of E^m as mentioned in Remark 2.5.

(2) Let $\mathfrak{A}$ be the $\Bbbk$-linear abelian category $\mathrm{Mod}_{f.d.}-A$ where A is a finite dimensional commutative algebra over a base field $\Bbbk$. For each $m \geq 1$ and $\phi = \{X_1, \dots, X_n\}$, define

$$T_m : \phi \to (\mathrm{Tor}^A_m(X_i, X_j))_{n\times n}.$$

By convention, let $\mathrm{Tor}^A_0(X_i, X_j)$ denote $X_i \otimes_A X_j$. Then, for each $m \geq 0$, one can define the Frobenius–Perron dimension of T_m as mentioned in Remark 2.5.

Definition 2.7. (1) Let $\mathfrak{A}$ be an abelian category. The *Frobenius–Perron dimension* of $\mathfrak{A}$ is defined to be

$$\mathrm{fpd}\,\mathfrak{A} := \mathrm{fpd}(E^1)$$

where $E^1 := \mathrm{Ext}^1_{\mathfrak{A}}(-,-)$ is defined as in Example 2.6(1). The *Frobenius–Perron theory* of $\mathfrak{A}$ is the collection

$$\{\mathrm{fpd}^m(E^n)\}_{m\geq 1, n\geq 0}$$

where $E^n := \mathrm{Ext}^n_{\mathcal{A}}(-,-)$ is defined as in Example 2.6(1).

(2) Let $\mathcal{T}$ be a pretriangulated category with suspension Σ. The *Frobenius–Perron dimension* of $\mathcal{T}$ is defined to be

$$\mathrm{fpd}\,\mathcal{T} := \mathrm{fpd}(\Sigma).$$

The *Frobenius–Perron theory* of $\mathcal{T}$ is the collection

$$\{\operatorname{fpd}^m(\Sigma^n)\}_{m\geq 1, n\in\mathbb{Z}}.$$

The fp-*global dimension* of $\mathcal{T}$ is defined to be

$$\operatorname{fpgldim}\mathcal{T} := \sup\{n \mid \operatorname{fpd}(\Sigma^n)\neq 0\}.$$

If $\mathcal{T}$ possesses a Serre functor S, the *Frobenius–Perron S-theory* of $\mathcal{T}$ is the collection

$$\{\operatorname{fpd}^m(\Sigma^n\circ S^w)\}_{m\geq 1, n, w\in\mathbb{Z}}.$$

Remarks 2.8. (1) The Frobenius–Perron dimension (respectively, Frobenius–Perron theory, fp-global dimension) can be defined for suspended categories [Keller and Vossieck 1987] and pre-n-angulated categories [Geiss et al. 2013] in the same way as Definition 2.7(2) since there is a suspension functor Σ.

(2) When $\mathfrak{A}$ is an abelian category, another way of defining the Frobenius–Perron dimension $\operatorname{fpd}\mathfrak{A}$ is as follows. We first embed $\mathfrak{A}$ into the derived category $D^b(\mathfrak{A})$. The suspension functor Σ of $D^b(\mathfrak{A})$ maps $\mathfrak{A}$ to $\mathfrak{A}[1]$ (so it is not a functor of $\mathfrak{A}$). The adjacency matrix $A(\phi,\Sigma)$ is still defined as in Definition 2.1(1) for brick sets ϕ in $\mathfrak{A}$. Then we can define

$$\operatorname{fpd}(\Sigma|_{\mathfrak{A}}) := \sup_{\phi\in\Phi_b,\phi\subset\mathfrak{A}}\{\rho(A(\phi,\Sigma))\}$$

as in Definition 2.3(2) by considering only the brick sets in $\mathfrak{A}$. Now $\operatorname{fpd}(\mathfrak{A})$ agrees with $\operatorname{fpd}(\Sigma|_{\mathfrak{A}})$.

The following lemma is clear.

Lemma 2.9. *Let $\mathfrak{A}$ be an abelian category and $n\geq 1$. Then $\operatorname{fpd}^n(D^b(\mathfrak{A}))\geq \operatorname{fpd}^n(\mathfrak{A})$. A similar statement holds for* fpd, fpg *and* fpv.

Proof. This follows from the fact that there is a fully faithful embedding $\mathfrak{A}\to D^b(\mathfrak{A})$ and that E^1 on $\mathfrak{A}$ agrees with Σ on $D^b(\mathfrak{A})$. $\square$

For any category $\mathcal{C}$ with an endofunctor σ, we define the *σ-quiver* of $\mathcal{C}$, denoted by $Q^{\sigma}_{\mathcal{C}}$, as follows:

(1) the vertex set of $Q^{\sigma}_{\mathcal{C}}$ consists of bricks in $\Phi_{1,b}$ in $\mathcal{C}$ (respectively, atomic objects in $\Phi_{1,a}$ when $\mathcal{C}$ is pretriangulated), and

(2) the arrow set of $Q^{\sigma}_{\mathcal{C}}$ consists of $n_{X,Y}$-arrows from X to Y, for all $X, Y\in\Phi_{1,b}$ (respectively, in $\Phi_{1,a}$), where $n_{X,Y}=\dim(X,\sigma(Y))$.

If σ is E^1, this quiver is denoted by $Q^{E^1}_{\mathcal{C}}$, which will be used in later sections.

The following lemma follows from the definition.

Lemma 2.10. *Retain the above notation. Then $\operatorname{fpd}\sigma\leq \operatorname{fpd} Q^{\sigma}_{\mathcal{C}}$.*

The fp-theory was motivated by the Frobenius–Perron dimension of objects in tensor or fusion categories [Etingof et al. 2015], see the following example.

Example 2.11. First we recall the definition of the Frobenius–Perron dimension given in [Etingof et al. 2015, Definitions 3.3.3 and 6.1.6]. Let $\mathcal{C}$ be a finite semisimple $\Bbbk$-linear tensor category. Suppose that $\{X_1, \dots, X_n\}$ is the complete list of nonisomorphic simple objects in $\mathcal{C}$. Since $\mathcal{C}$ is semisimple, every object X in $\mathcal{C}$ is a direct sum

$$X = \bigoplus_{i=1}^{n} X_i^{\oplus a_i}$$

for some integers $a_i \geq 0$. The tensor product on $\mathcal{C}$ makes its Grothendieck ring $\mathrm{Gr}(\mathcal{C})$ a $\mathbb{Z}_+$-ring [loc. cit., Definition 3.1.1]. For every object V in $\mathcal{C}$ and every j, write

$$V \otimes_{\mathcal{C}} X_j \cong \bigoplus_{i=1}^{n} X_i^{\oplus a_{ij}} \tag{E2.11.1}$$

for some integers $a_{ij} \geq 0$. In the Grothendieck ring $\mathrm{Gr}(\mathcal{C})$, the left multiplication by V sends X_j to $\sum_{i=1}^{n} a_{ij} X_i$. Then, by [loc. cit., Definition 3.3.3], the *Frobenius–Perron dimension* of V is defined to be

$$\mathrm{FPdim}(V) := \rho((a_{ij})_{n\times n}). \tag{E2.11.2}$$

In fact the Frobenius–Perron dimension is defined for any object in a $\mathbb{Z}_+$-ring.

Next we use Definition 2.3(2) to calculate the Frobenius–Perron dimension. Let σ be the tensor functor $V \otimes_{\mathcal{C}} -$ that is a $\Bbbk$-linear endofunctor of $\mathcal{C}$. If ϕ is a brick subset of $\mathcal{C}$, then ϕ is a subset of $\phi_n := \{X_1, \dots, X_n\}$. For simplicity, assume that ϕ is $\{X_1, \dots, X_s\}$ for some $s \leq n$. It follows from (E2.11.1) that

$$\mathrm{Hom}_{\mathcal{C}}(X_i, \sigma(X_j)) = \Bbbk^{\oplus a_{ij}} \quad \forall i, j.$$

Hence the adjacency matrix of (ϕ_n, σ) is

$$A(\phi_n, \sigma) = (a_{ij})_{n\times n}$$

and the adjacency matrix of (ϕ, σ) is a principal minor of $A(\phi_n, \sigma)$. By Lemma 1.7(1), $\rho(A(\phi, \sigma)) \leq \rho(A(\phi_n, \sigma))$. By Definition 2.3(2), the *Frobenius–Perron dimension* of the functor $\sigma = V \otimes_{\mathcal{C}} -$ is

$$\mathrm{fpd}(V \otimes_{\mathcal{C}} -) = \sup_{\phi\in\Phi_b} \{\rho(A(\phi, \sigma))\} = \rho(A(\phi_n, \sigma)) = \rho((a_{ij})_{n\times n}),$$

which agrees with (E2.11.2). This justifies calling $\mathrm{fpd}(V \otimes_{\mathcal{C}} -)$ the Frobenius–Perron dimension of V.

3. Basic properties

For simplicity, "Frobenius–Perron" is abbreviated to "fp".

Embeddings. It is clear that the fp-series and the fp-dimensions are invariant under equivalences of categories. We record this fact below. Recall that the Frobenius–Perron series $\mathrm{FP}(u, t, \sigma)$ of an endofunctor σ is defined in Definition 2.4(2).

Lemma 3.1. *Let $F : \mathcal{C} \to \mathcal{D}$ be an equivalence of categories. Let $\sigma_{\mathcal{C}}$ and $\sigma_{\mathcal{D}}$ be endofunctors of $\mathcal{C}$ and $\mathcal{D}$ respectively. Suppose that $F \circ \sigma_{\mathcal{C}}$ is naturally isomorphic to $\sigma_{\mathcal{D}} \circ F$. Then* $\mathrm{FP}(u, t, \sigma_{\mathcal{C}}) = \mathrm{FP}(u, t, \sigma_{\mathcal{D}})$.

Let $\overline{\mathbb{R}}_+$ denote the set of nonnegative real numbers union with $\{\pm\infty\}$. Let

$$f(u,t) := \sum_{m,n=0}^{\infty} f_{m,n} t^m u^n \quad \text{and} \quad g(u,t) := \sum_{m,n=0}^{\infty} g_{m,n} t^m u^n$$

be two elements in $\overline{\mathbb{R}}_+[[u,t]]$. We write $f \le g$ if $f_{m,n} \le g_{m,n}$ for all m, n.

Theorem 3.2. *Let $F : \mathcal{C} \to \mathcal{D}$ be a faithful functor that preserves brick subsets*:

(1) *Let $\sigma_{\mathcal{C}}$ and $\sigma_{\mathcal{D}}$ be endofunctors of $\mathcal{C}$ and $\mathcal{D}$ respectively. Suppose that $F \circ \sigma_{\mathcal{C}}$ is naturally isomorphic to $\sigma_{\mathcal{D}} \circ F$. Then* $\mathrm{FP}(u, t, \sigma_{\mathcal{C}}) \le \mathrm{FP}(u, t, \sigma_{\mathcal{D}})$.

(2) *Let $\tau_{\mathcal{C}}$ and $\tau_{\mathcal{D}}$ be assignments of $\mathcal{C}$ and $\mathcal{D}$ respectively satisfying the property in Remark 2.5. Suppose that $\rho(A(\phi, \tau_{\mathcal{C}})) \le \rho(A(F(\phi), \tau_{\mathcal{D}}))$ for all $\phi \in \Phi_{n,b}(\mathcal{C})$ and all n. Then* $\mathrm{FP}(u, t, \tau_{\mathcal{C}}) \le \mathrm{FP}(u, t, \tau_{\mathcal{D}})$.

Proof. (1) For every $\phi = \{X_1, \ldots, X_n\} \in \Phi_n(\mathcal{C})$, let $F(\phi)$ be $\{F(X_1), \ldots, F(X_n)\}$ in $\Phi_n(\mathcal{D})$. By hypothesis, if $\phi \in \Phi_{n,b}(\mathcal{C})$, then $F(\phi)$ is in $\Phi_{n,b}(\mathcal{D})$. Let $A = (a_{ij})$ (respectively, $B = (b_{ij})$) be the adjacency matrix of $(\phi, \sigma_{\mathcal{C}})$ (respectively, of $(F(\phi), \sigma_{\mathcal{D}})$). Then, by the faithfulness of F,

$$a_{ij} = \dim(X_i, \sigma_{\mathcal{C}}(X_j)) \le \dim(F(X_i), F(\sigma_{\mathcal{C}}(X_j))) = \dim(F(X_i), \sigma_{\mathcal{D}}(F(X_j))) = b_{ij}.$$

By Lemma 1.7(2),

$$\rho(A(\phi, \sigma_{\mathcal{C}})) =: \rho(A) \le \rho(B) := \rho(A(F(\phi), \sigma_{\mathcal{D}})). \tag{E3.2.1}$$

By definition,

$$\mathrm{fpd}^n(\sigma_{\mathcal{C}}) \le \mathrm{fpd}^n(\sigma_{\mathcal{D}}). \tag{E3.2.2}$$

Similarly, for all n, m, $\mathrm{fpd}^n(\sigma_{\mathcal{C}}^m) \le \mathrm{fpd}^n(\sigma_{\mathcal{D}}^m)$. The assertion follows.

(2) The proof of part (2) is similar. □

Theorem 0.1 follows from Theorem 3.2.

(a-)Hereditary algebras and categories. Recall that the global dimension of an abelian category $\mathfrak{A}$ is defined to be

$$\mathrm{gldim}\, \mathfrak{A} := \sup\{n \mid \mathrm{Ext}^n_{\mathfrak{A}}(X, Y) \ne 0, \text{ for some } X, Y \in \mathfrak{A}\}.$$

The global dimension of an algebra A is defined to be the global dimension of the category of left A-modules. An algebra (or an abelian category) is called *hereditary* if it has global dimension at most one.

There is a nice property concerning the indecomposable objects in the derived category of a hereditary abelian category (see [loc. cit., Section 2.5]).

Lemma 3.3. *Let $\mathfrak{A}$ be a hereditary abelian category. Then every indecomposable object in the derived category $D(\mathfrak{A})$ is isomorphic to a shift of an object in $\mathfrak{A}$.*

Note that every brick (or atomic) object in an additive category is indecomposable. Based on the property in the above lemma, we make a definition.

Definition 3.4. An abelian category $\mathfrak{A}$ is called *a-hereditary* (respectively, *b-hereditary*) if every atomic (respectively, brick) object X in the bounded derived category $D^b(\mathfrak{A})$ is of the form $M[i]$ for some object M in $\mathfrak{A}$ and $i \in \mathbb{Z}$. The object M is automatically a brick object in $\mathfrak{A}$.

By Lemma 3.11(2), if A is a finite dimensional local algebra, then the category $\mathrm{Mod}_{f.d.}-A$ of finite dimensional A-modules is a-hereditary. If A is not $\Bbbk$, then $\mathrm{Mod}_{f.d.}-A$ is not hereditary. Another such example is given in Lemma 4.1.

If α is an autoequivalence of an abelian category $\mathfrak{A}$, then it extends naturally to an autoequivalence, denoted by $\bar{\alpha}$, of the derived category $\mathcal{A} := D^b(\mathfrak{A})$. The main result in this subsection is the following. Recall that the b-height of $\mathfrak{A}$, denoted by $h_b(\mathfrak{A})$, is defined in Definition 2.1(3) and that the Frobenius–Perron global dimension of $\mathcal{A}$, denoted by $\operatorname{fpgldim}\mathcal{A}$, is defined in Definition 2.7(2).

Theorem 3.5. *Let $\mathfrak{A}$ be an a-hereditary abelian category with an auto-equivalence α. For each n, define $n' = \min\{n, h_b(\mathfrak{A})\}$. Let $\mathcal{A}$ be $D^b(\mathfrak{A})$:*

(1) *If $m < 0$ or $m > \operatorname{gldim}\mathfrak{A}$, then*

$$\operatorname{fpd}(\Sigma^m \circ \bar{\alpha}) = 0.$$

As a consequence, $\operatorname{fpgldim}\mathcal{A} \le \operatorname{gldim}\mathfrak{A}$.

(2) *For each n,*

$$\operatorname{fpd}^n(\alpha) \le \operatorname{fpd}^n(\bar{\alpha}) \le \max_{1 \le i \le n'} \{\operatorname{fpd}^i(\alpha)\}. \tag{E3.5.1}$$

If $\operatorname{gldim}\mathfrak{A} < \infty$, *then*

$$\operatorname{fpd}^n(\bar{\alpha}) = \max_{1 \le i \le n'} \{\operatorname{fpd}^i(\alpha)\}. \tag{E3.5.2}$$

(3) *Let $g := \operatorname{gldim}\mathfrak{A} < \infty$. Let β be the assignment $(X, Y) \to (\operatorname{Ext}^g_{\mathfrak{A}}(X, \alpha(Y)))$. Then*

$$\operatorname{fpd}^n(\Sigma^g \circ \bar{\alpha}) = \max_{1 \le i \le n'} \{\operatorname{fpd}^i(\beta)\}. \tag{E3.5.3}$$

(4) *For every hereditary abelian category $\mathfrak{A}$, we have* $\operatorname{fpd}(\mathcal{A}) = \operatorname{fpd}(\mathfrak{A})$.

Proof. (1) Since $\mathcal{A}$ is a-hereditary, every atomic object in $\mathcal{A}$ is of the form $M[i]$.

Case 1: $m < 0$. Write ϕ as $\{M_1[d_1], \ldots, M_n[d_n]\}$ where d_i is decreasing and M_i is in $\mathfrak{A}$. Then, for $i \le j$,

$$a_{ij} = \operatorname{Hom}_{\mathcal{A}}(M_i[d_i], (\Sigma^m \circ \bar{\alpha})M_j[d_j]) = \operatorname{Hom}_{\mathcal{A}}(M_i, \alpha(M_j)[d_j - d_i + m]) = 0$$

since $d_j - d_i + m < 0$. Thus the adjacency matrix $A := (a_{ij})_{n\times n}$ is strictly lower triangular. As a consequence, $\rho(A) = 0$. By definition, $\operatorname{fpd}(\Sigma^m \circ \bar{\alpha}) = 0$.

Case 2: $m > \operatorname{gldim}\mathfrak{A}$. Write ϕ as $\{M_1[d_1], \dots, M_n[d_n]\}$ where d_i is increasing and M_i is in $\mathfrak{A}$. Then, for $i \geq j$,

$$a_{ij} = \operatorname{Hom}_{\mathcal{A}}(M_i[d_i], (\Sigma^m \circ \bar{\alpha})M_j[d_j]) = \operatorname{Hom}_{\mathcal{A}}(M_i, \alpha(M_j)[d_j - d_i + m]) = 0$$

since $d_j - d_i + m > \operatorname{gldim}\mathfrak{A}$. Thus the adjacency matrix $A := (a_{ij})_{n\times n}$ is strictly upper triangular. As a consequence, $\rho(A) = 0$. By definition, $\operatorname{fpd}(\Sigma^m \circ \bar{\alpha}) = 0$.

(2) Let F be the canonical fully faithful embedding $\mathfrak{A} \to \mathcal{A}$. By Theorem 3.2 and (E3.2.2),

$$\operatorname{fpd}^n(\alpha) \leq \operatorname{fpd}^n(\bar{\alpha}).$$

For the other assertion, write ϕ as a disjoint union $\phi_{d_1} \cup \dots \cup \phi_{d_s}$ where d_i is strictly decreasing and the subset ϕ_{d_i} consists of objects of the form $M[d_i]$ for $M \in \mathfrak{A}$. For any objects $X \in \phi_{d_i}$ and $Y \in \phi_{d_j}$ for $i < j$, $\operatorname{Hom}_{\mathcal{A}}(X, Y) = 0$. Thus the adjacency matrix of $(\phi, \bar{\alpha})$ is of the form

$$A(\phi, \bar{\alpha}) = \begin{pmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ * & A_{22} & 0 & \cdots & 0 \\ * & * & A_{33} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ * & * & * & \cdots & A_{ss} \end{pmatrix} \tag{E3.5.4}$$

where each A_{ii} is the adjacency matrix $A(\phi_{d_i}, \bar{\alpha})$. For each ϕ_{d_i}, we have

$$A(\phi_{d_i}, \bar{\alpha}) = A(\phi_{d_i}[-d_i], \bar{\alpha}) = A(\phi_{d_i}[-d_i], \alpha)$$

which implies that

$$\rho(A_{ii}) \leq \operatorname{fpd}^{s_i}(\alpha) \leq \max_{1\leq j\leq n'} \operatorname{fpd}^j(\alpha)$$

where s_i is the size of A_{ii} and $n' = \min\{n, h_b(\mathfrak{A})\}$. By using the matrix (E3.5.4),

$$\rho(A(\phi, \bar{\alpha})) = \max_i\{\rho(A_{ii})\} \leq \max_{1\leq j\leq n'} \operatorname{fpd}^j(\alpha).$$

Then (E3.5.1) follows.

Suppose now that $g := \operatorname{gldim}\mathfrak{A} < \infty$. Let $\phi \in \Phi_{n,a}(\mathcal{A})$. Pick any $M \in \Phi_{1,b}(\mathfrak{A})$. Then, for $g' \gg g$, $\phi' := \phi \cup \{M[g']\} \in \Phi_{n+1,a}(\mathcal{A})$. By Lemma 1.7(1), $\rho(A(\phi', \bar{\alpha})) \geq \rho(A(\phi, \bar{\alpha}))$. Hence $\operatorname{fpd}^n(\bar{\alpha})$ is increasing as n increases. Therefore (E3.5.2) follows from (E3.5.1).

(3) The proof is similar to the proof of part (2). Let F be the canonical fully faithful embedding $\mathfrak{A} \to \mathcal{A}$. By Theorem 3.2(2) and (E3.2.2),

$$\operatorname{fpd}^n(\beta) \leq \operatorname{fpd}^n(\Sigma^g \circ \bar{\alpha}).$$

By the argument at the end of proof of part (2), $\operatorname{fpd}^n(\Sigma^g \circ \bar{\alpha})$ increases when n increases. Then

$$\max_{1\leq j\leq n'} \operatorname{fpd}^j(\beta) \leq \operatorname{fpd}^n(\Sigma^g \circ \bar{\alpha}).$$

For the other direction, write ϕ as a disjoint union $\phi_{d_1} \cup \cdots \cup \phi_{d_s}$ where d_i is strictly increasing and ϕ_{d_i} consists of objects of the form $M[d_i]$ for $M \in \mathfrak{A}$. For objects $X \in \phi_{d_i}$ and $Y \in \phi_{d_j}$ for $i < j$, $\mathrm{Hom}_{\mathcal{A}}(X, \Sigma^g(\alpha(Y))) = 0$. Let $\gamma = \Sigma^g \circ \bar{\alpha}$. Then the adjacency matrix of (ϕ, γ) is of the form (E3.5.4), namely,

$$A(\phi, \gamma) = \begin{pmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ * & A_{22} & 0 & \cdots & 0 \\ * & * & A_{33} & \cdots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ * & * & * & \cdots & A_{ss} \end{pmatrix}$$

where each A_{ii} is the adjacency matrix $A(\phi_{d_i}, \gamma)$. For each ϕ_{d_i}, we have

$$A(\phi_{d_i}, \gamma) = A(\phi_{d_i}[-d_i], \gamma) = A(\phi_{d_i}[-d_i], \beta)$$

which implies that

$$\rho(A_{ii}) \le \mathrm{fpd}^{s_i}(\beta) \le \max_{1 \le j \le n'} \mathrm{fpd}^j(\beta)$$

where s_i is the size of A_{ii}. By using matrix (E3.5.4),

$$\rho(A(\phi, \gamma)) = \max_i \{\rho(A_{ii})\} \le \max_{1 \le j \le n'} \mathrm{fpd}^j(\beta).$$

The assertion follows.

(4) Take α to be the identity functor of $\mathfrak{A}$ and $g = 1$ (since $\mathfrak{A}$ is hereditary). By (E3.5.3), we have

$$\mathrm{fpd}^n(\Sigma) = \max_{1 \le i \le n'} \{\mathrm{fpd}^i(E^1)\}.$$

By taking $\sup_n$, we obtain that $\mathrm{fpd}(E^1) = \mathrm{fpd}(\Sigma)$. The assertion follows. □

Categories with Serre functor. Recall from [Keller 2008, Section 2.6] that if a Hom-finite category $\mathcal{C}$ has a Serre functor S, then there is a natural isomorphism

$$\mathrm{Hom}_{\mathcal{C}}(X, Y)^* \cong \mathrm{Hom}_{\mathcal{C}}(Y, S(X))$$

for all $X, Y \in \mathcal{C}$. A (pre-)triangulated Hom-finite category $\mathcal{C}$ with Serre functor S is called *n-Calabi–Yau* if there is a natural isomorphism

$$S \cong \Sigma^n.$$

(In [Keller 2008, Section 2.6] it is called *weakly n-Calabi–Yau*.) We now prove Proposition 0.6.

Proposition 3.6 (Serre duality). *Let $\mathcal{C}$ be a* Hom*-finite category with Serre functor S. Let σ be an endofunctor of $\mathcal{C}$*:

(1) *If σ has a right adjoint $\sigma^!$, then*

$$\mathrm{fpd}(\sigma) = \mathrm{fpd}(\sigma^! \circ S).$$

(2) *If σ is an equivalence with quasiinverse σ^{-1}, then*

$$\operatorname{fpd}(\sigma) = \operatorname{fpd}(\sigma^{-1} \circ S).$$

(3) *If $\mathcal{C}$ is (pre-)triangulated and n-Calabi–Yau, then we have a duality*

$$\operatorname{fpd}(\Sigma^i) = \operatorname{fpd}(\Sigma^{n-i})$$

for all i.

Proof. (1) Let $\phi = \{X_1, \dots, X_n\} \in \Phi_{n,b}$ and let $A(\phi, \sigma)$ be the adjacency matrix with (i, j)-entry $a_{ij} = \dim(X_i, \sigma(X_j))$. By Serre duality,

$$a_{ij} = \dim(X_i, \sigma(X_j)) = \dim(\sigma(X_j), S(X_i)) = \dim(X_j, (\sigma^! \circ S)(X_i)),$$

which is the (j, i)-entry of the adjacency matrix $A(\phi, \sigma^! \circ S)$. Then $\rho(A(\phi, \sigma)) = \rho(A(\phi, \sigma^! \circ S))$. It follows from the definition that $\operatorname{fpd}^n(\sigma) = \operatorname{fpd}^n(\sigma^! \circ S)$ for all $n \geq 1$. The assertion follows from the definition.

(2) and (3) These are consequences of part (1). □

Opposite categories.

Lemma 3.7. *Let σ be an endofunctor of $\mathcal{C}$ and suppose that σ has a left adjoint σ^*. Consider σ^* as an endofunctor of the opposite category $\mathcal{C}^{\mathrm{op}}$ of $\mathcal{C}$. Then*

$$\operatorname{fpd}^n(\sigma|_{\mathcal{C}}) = \operatorname{fpd}^n(\sigma^*|_{\mathcal{C}^{\mathrm{op}}})$$

for all n.

Proof. Let $\phi := \{X_1, \dots, X_n\}$ be a brick subset of $\mathcal{C}$ (which is also a brick subset of $\mathcal{C}^{\mathrm{op}}$). Then

$$\dim_{\mathcal{C}}(X_i, \sigma(X_j)) = \dim_{\mathcal{C}}(\sigma^*(X_i), X_j) = \dim_{\mathcal{C}^{\mathrm{op}}}(X_j, \sigma^*(X_i)),$$

which implies that the adjacency matrix of σ^* as an endofunctor of $\mathcal{C}^{\mathrm{op}}$ is the transpose of the adjacency matrix of σ. The assertion follows. □

Definition 3.8. (1) Two pretriangulated categories $(\mathcal{T}_i, \Sigma_i)$, for $i = 1, 2$, are called fp-*equivalent* if

$$\operatorname{fpd}^n(\Sigma_1^m) = \operatorname{fpd}^n(\Sigma_2^m)$$

for all $n \geq 1$, $m \in \mathbb{Z}$:

(2) Two algebras are fp-*equivalent* if their bounded derived categories of finitely generated modules are fp-equivalent.

(3) Two pretriangulated categories with Serre functors $(\mathcal{T}_i, \Sigma_i, S_i)$, for $i = 1, 2$, are called fp-*S-equivalent* if

$$\operatorname{fpd}^n(\Sigma_1^m \circ S_1^k) = \operatorname{fpd}^n(\Sigma_2^m \circ S_2^k)$$

for all $n \geq 1$, $m, k \in \mathbb{Z}$.

Proposition 3.9. *Let $\mathcal{T}$ be a pretriangulated category*:

(1) *$\mathcal{T}$ and $\mathcal{T}^{\rm op}$ are* fp*-equivalent.*

(2) *Suppose S is a Serre functor of $\mathcal{T}$. Then $(\mathcal{T}, S)$ and $(\mathcal{T}^{\rm op}, S^{\rm op})$ are* fp*-S-equivalent.*

Proof. (1) Let Σ be the suspension of $\mathcal{T}$. Then $\mathcal{T}^{\rm op}$ is also pretriangulated with suspension functor being $\Sigma^{-1} = \Sigma^*$ (restricted to $\mathcal{T}^{\rm op}$). The assertion follows from Lemma 3.7.

(2) Note that the Serre functor of $\mathcal{T}^{\rm op}$ is equal to $S^{-1} = S^*$ (restricted to $\mathcal{T}^{\rm op}$). The assertion follows by Lemma 3.7. □

Corollary 3.10. *Let A be a finite dimensional algebra*:

(1) *A and $A^{\rm op}$ are* fp*-equivalent.*

(2) *Suppose A has finite global dimension. In this case, the bounded derived category of finite dimensional A-modules has a Serre functor. Then A and $A^{\rm op}$ are* fp*-S-equivalent.*

Proof. (1) Since A is finite dimensional, the $\Bbbk$-linear dual induces an equivalence of triangulated categories between $D^b(\mathrm{Mod}_{f.d.}-A)^{\rm op}$ and $D^b(\mathrm{Mod}_{f.d.}-A^{\rm op})$. The assertion follows from Proposition 3.9(1).

(2) The proof is similar, using Proposition 3.9(2) instead. □

There are examples where $\mathcal{T}$ and $\mathcal{T}^{\rm op}$ are not triangulated equivalent, see Example 3.12. In this paper, a $\Bbbk$-algebra A is called *local* if A has a unique maximal ideal $\mathfrak{m}$ and $A/\mathfrak{m} \cong \Bbbk$. The following lemma is easy and well known.

Lemma 3.11. *Let A be a finite dimensional local algebra over $\Bbbk$. Let $\mathfrak{A}$ be the category* $\mathrm{Mod}_{f.d.}-A$ *and $\mathcal{A}$ be $D^b(\mathfrak{A})$*:

(1) *Let X be an object in $\mathcal{A}$ such that* $\mathrm{Hom}_{\mathcal{A}}(X, X[-i]) = 0$ *for all $i > 0$. Then X is of the form $M[n]$ where M is an object in $\mathfrak{A}$ and $n \in \mathbb{Z}$.*

(2) *Every atomic object in $\mathcal{A}$ is of the form $M[n]$ where M is a brick object in $\mathfrak{A}$ and $n \in \mathbb{Z}$. Namely, $\mathfrak{A}$ is a-hereditary.*

Proof. (2) is an immediate consequence of part (1). We only prove part (1).

On the contrary we suppose that $H^m(X) \neq 0$ and $H^n(X) \neq 0$ for some $m < n$. Since X is a bounded complex, we can take m to be minimum of such integers and n to be the maximum of such integers. Since A is local, there is a nonzero map from $H^n(X) \to H^m(X)$, which induces a nonzero morphism in $\mathrm{Hom}_{\mathcal{A}}(X, X[m-n])$. This contradicts the hypothesis. □

Example 3.12. Let m, n be integers ≥ 2. Define $A_{m,n}$ to be the algebra

$$\Bbbk\langle x_1, x_2\rangle/(x_1^m, x_2^n, x_1x_2).$$

It is easy to see that $A_{m,n}$ is a finite dimensional local connected graded algebra generated in degree 1 (with $\deg x_1 = \deg x_2 = 1$). If $A_{m,n}$ is isomorphic to $A_{m',n'}$ as algebras, by [Bell and Zhang 2017, Theorem 1], these are isomorphic as graded algebras. Suppose $f : A_{m,n} \to A_{m',n'}$ is an isomorphism of graded algebras and write

$$f(x_1) = ax_1 + bx_2, \quad f(x_2) = cx_1 + dx_2.$$

Then the relation $f(x_1)f(x_2) = 0$ forces $b = c = 0$. As a consequence, $m = m'$ and $n = n'$. So we have proven that

(1) $A_{m,n}$ is isomorphic to $A_{m',n'}$ if and only if $m = m'$ and $n = n'$.

Next we claim that

(2) the derived category $D^b(\mathrm{Mod}_{f.d.} - A_{m,n})$ is not triangulated equivalent to $D^b(\mathrm{Mod}_{f.d.} - A^{\mathrm{op}}_{m,n})$, if $m \neq n$.

Let m, n, m', n' be integers ≥ 2. Suppose that $D^b(\mathrm{Mod}_{f.d.} - A_{m,n})$ is triangulated equivalent to $D^b(\mathrm{Mod}_{f.d.} - A_{m',n'})$. Since $A_{m,n}$ is local, by [Yekutieli 1999, Theorem 2.3], every tilting complex over $A_{m,n}$ is of the form $P[n]$ where P is a progenerator over $A_{m,n}$. As a consequence, $A_{m,n}$ is Morita equivalent to $A_{m',n'}$. Since both $A_{m,n}$ and $A_{m',n'}$ are local, Morita equivalence implies that $A_{m,n}$ is isomorphic to $A_{m',n'}$. By part (1), $m = m'$ and $n = n'$. In other words, if $(m, n) \neq (m', n')$, then $D^b(\mathrm{Mod}_{f.d.} - A_{m,n})$ is not triangulated equivalent to $D^b(\mathrm{Mod}_{f.d.} - A_{m',n'})$. As a consequence, if $m \neq n$, then $D^b(\mathrm{Mod}_{f.d.} - A_{m,n})$ is not triangulated equivalent to $D^b(\mathrm{Mod}_{f.d.} - A_{n,m})$. By definition, $A^{\mathrm{op}}_{m,n} \cong A_{n,m}$. Therefore claim (2) follows.

We can show that $D^b(\mathrm{Mod}_{f.d.} - A)$ is dual to $D^b(\mathrm{Mod}_{f.d.} - A^{\mathrm{op}})$ by using the $\Bbbk$-linear dual. In other words, $D^b(\mathrm{Mod}_{f.d.} - A)^{\mathrm{op}}$ is triangulated equivalent to $D^b(\mathrm{Mod}_{f.d.} - A^{\mathrm{op}})$. Therefore the following is a consequence of part (2).

(3) Suppose $m \neq n$ and let $\mathcal{A}$ be $D^b(\mathrm{Mod}_{f.d.} - A_{m,n})$. Then $\mathcal{A}$ is not triangulated equivalent to $\mathcal{A}^{\mathrm{op}}$. But by Proposition 3.9(1), $\mathcal{A}$ and $\mathcal{A}^{\mathrm{op}}$ are fp-equivalent.

4. Derived category over a commutative ring

Throughout this section A is a commutative algebra and $\mathcal{A} = D^b(\mathrm{Mod} - A)$. (In other sections $\mathcal{A}$ usually denotes $D^b(\mathrm{Mod}_{f.d.} - A)$.)

Lemma 4.1. *Let A be a commutative algebra. Let X be an atomic object in $\mathcal{A}$. Then X is of the form $M[i]$ for some simple A-module M and some $i \in \mathbb{Z}$. As a consequence*, Mod $-A$ *is a-hereditary.*

Proof. Consider X as a bounded above complex of projective A-modules. Since A is commutative, every $f \in A$ induces naturally a morphism of X by multiplication. For each i, $H^i(X)$ is an A-module. We have natural morphisms of A-algebras

$$A \to \mathrm{Hom}_{\mathcal{A}}(X, X) \to \mathrm{End}_A(H^i(X)).$$

By definition, $\operatorname{Hom}_{\mathcal{A}}(X, X) = \Bbbk$. Thus $\operatorname{Hom}_A(X, X) = A/\mathfrak{m}$ for some ideal $\mathfrak{m}$ of A that has codimension 1. Hence the A-action on $H^i(X)$ factors through the map $A \to A/\mathfrak{m}$. This means that $H^i(X)$ is a direct sum of $A/\mathfrak{m}$.

Let $n = \sup X$ and $m = \inf X$. Then $H^m(X) = (A/\mathfrak{m})^{\oplus s}$ and $H^n(X) = (A/\mathfrak{m})^{\oplus t}$ for some $s, t > 0$. If $m < n$, then

$$\operatorname{Hom}_{\mathcal{A}}(X, X[m-n]) \cong \operatorname{Hom}_{\mathcal{A}}(X[n], X[m]) \cong \operatorname{Hom}_A(H^n(X), H^m(X)) \neq 0$$

which contradicts (E2.1.1). Therefore $m = n$ and $X = M[n]$ for $M := H^n(X)$. Since X is atomic, M has only one copy of $A/\mathfrak{m}$. □

Lemma 4.2. *Let A be a noetherian commutative algebra. Let X and Y be two atomic objects in $\mathcal{A}$. Then* $\operatorname{Hom}_{\mathcal{A}}(X, Y) \neq 0$ *if and only if there is an ideal $\mathfrak{m}$ of A of codimension 1 such that $X \cong A/\mathfrak{m}[m]$ and $Y \cong A/\mathfrak{m}[n]$ for some* $0 \le n - m \le \operatorname{projdim} A/\mathfrak{m}$.

Proof. By Lemma 4.1, $X \cong A/\mathfrak{m}[m]$ for some ideal $\mathfrak{m}$ of codimension 1 and some integer m. Similarly, $Y \cong A/\mathfrak{n}[n]$ for ideal $\mathfrak{n}$ of codimension 1 and integer n.

Suppose $\operatorname{Hom}_{\mathcal{A}}(X, Y) \neq 0$. If $\mathfrak{m} \neq \mathfrak{n}$, then clearly $\operatorname{Hom}_{\mathcal{A}}(X, Y) = 0$. Hence $\mathfrak{m} = \mathfrak{n}$. Further, $\operatorname{Ext}^{n-m}_A(A/\mathfrak{m}, A/\mathfrak{m}) \cong \operatorname{Hom}_{\mathcal{A}}(X, Y) \neq 0$ implies that $0 \le n - m \le \operatorname{projdim} A/\mathfrak{m}$. The converse can be proved in a similar way by passing to a localization. □

If A is an affine commutative ring over $\Bbbk$, then every simple A-module is 1-dimensional. Hence $(A/\mathfrak{m})[i]$ is a brick (and atomic) object in $\mathcal{A}$ for every $i \in \mathbb{Z}$ and every maximal ideal $\mathfrak{m}$ of A. The fp-global dimension $\operatorname{fpgldim}(\mathcal{A})$ is defined in Definition 2.7(2).

Proposition 4.3. *Let A be an affine commutative domain of global dimension $g < \infty$:*

1. $\operatorname{fpd}(\mathcal{A}) = g$.
2. $\operatorname{fpd}(\Sigma^i) = \binom{g}{i}$ *for all* i.
3. $\operatorname{fpgldim}(\mathcal{A}) = g$.

Proof. (1) By Lemma 4.1, every atomic object is of the form $M[i]$ for some $M \cong A/\mathfrak{m}$ where $\mathfrak{m}$ is an ideal of codimension 1, and $i \in \mathbb{Z}$. It is well-known that

$$\dim \operatorname{Ext}^i_A(A/\mathfrak{m}, A/\mathfrak{m}) = \binom{g}{i} \quad \forall i. \tag{E4.3.1}$$

If $\mathfrak{m}_1$ and $\mathfrak{m}_2$ are two different maximal ideals, then

$$\operatorname{Ext}^i_A(A/\mathfrak{m}_1, A/\mathfrak{m}_2) = 0 \tag{E4.3.2}$$

for all i. Let ϕ be an atomic n-object subset. We can decompose ϕ into a disjoint union $\phi_{A/\mathfrak{m}_1} \cup \cdots \cup \phi_{A/\mathfrak{m}_s}$ where $\phi_{A/\mathfrak{m}}$ consists of objects of the form $A/\mathfrak{m}[i]$ for $i \in \mathbb{Z}$. It follows from (E4.3.2) that the adjacency matrix is a block-diagonal matrix. Hence, we only need to consider the case when $\phi = \phi_{A/\mathfrak{m}}$ after we use the reduction similar to the one used in the proof of Theorem 3.5. Let $\phi = \phi_{A/\mathfrak{m}} = \{A/\mathfrak{m}[d_1], \ldots, A/\mathfrak{m}[d_m]\}$

where d_i is increasing. By Lemma 4.2, we have $d_{i+1} - d_i > g$, or $d_i + g < d_{i+1}$, for all $i = 1, \dots, m-1$. Under these conditions, the adjacency matrix is lower triangular with each diagonal being g. Thus $\operatorname{fpd}(\Sigma) = g$.

The proof of (2) is similar and (3) is a consequence of (2). □

Suggested by Theorem 3.5, we could introduce some secondary invariants as follows. The *stabilization index* of a triangulated category $\mathcal{T}$ is defined to be

$$\operatorname{SI}(\mathcal{T}) = \min\{n \mid \operatorname{fpd}^{n'} \mathcal{T} = \operatorname{fpd} \mathcal{T}\ \forall n' \geq n\}.$$

The *global stabilization index* of $\mathcal{T}$ is defined to be

$$\operatorname{GSI}(\mathcal{T}) = \max\{\operatorname{SI}(\mathcal{T}') \mid \text{for all thick triangulated full subcategories } \mathcal{T}' \subseteq \mathcal{T}\}.$$

It is clear that both stabilization index and global stabilization index can be defined for an abelian category.

Similar to Proposition 4.3, one can show the following. Suppose that A is affine. For every i, let

$$d_i := \sup\{\dim \operatorname{Ext}^i_A(A/\mathfrak{m}, A/\mathfrak{m}) \mid \text{maximal ideals } \mathfrak{m} \subseteq A\}.$$

Proposition 4.4. *Let A be an affine commutative algebra. Then, for each i, $\operatorname{fpd}(\Sigma^i) = d_i < \infty$ and $\rho(A(\phi, \Sigma^i)) \leq d_i$ for all $\phi \in \Phi_{n,a}$. As a consequence, for each integer i, the following hold*:

(1) $\operatorname{fpd}(\Sigma^i) = \operatorname{fpd}^1(\Sigma^i)$. *Hence the stabilization index of $\mathcal{A}$ is* 1.

(2) $\operatorname{fpd}(\Sigma^i)$ *is a finite integer.*

Theorem 4.5. *Let A be an affine commutative algebra and $\mathcal{A}$ be $D^b(\operatorname{Mod} A)$. Let $\mathcal{T}$ be a triangulated full subcategory of $\mathcal{A}$ with suspension $\Sigma_{\mathcal{T}}$. Let i be an integer*:

(1) $\operatorname{fpd}(\Sigma^i_{\mathcal{T}}) = \operatorname{fpd}^1(\Sigma^i_{\mathcal{T}})$. *As a consequence, the global stabilization index of $\mathcal{A}$ is* 1.

(2) $\operatorname{fpd}(\Sigma^i_{\mathcal{T}})$ *is a finite integer.*

(3) *If $\mathcal{T}$ is isomorphic to $D^b(\operatorname{Mod}_{f.d.}-B)$ for some finite dimensional algebra B, then B is Morita equivalent to a commutative algebra.*

Proof. (1) and (2) are similar to Proposition 4.4.

(3) Since B is finite dimensional, it is Morita equivalent to a basic algebra. So we can assume B is basic and show that B is commutative. Write B as a $\Bbbk Q/(R)$ where Q is a finite quiver with admissible ideal $R \subseteq (\Bbbk Q)_{\geq 2}$. We will show that B is commutative.

First we claim that each connected component of Q consists of only one vertex. Suppose not. Then Q contains distinct vertices v_1 and v_2 with an arrow $\alpha : v_1 \to v_2$. Let S_1 and S_2 be the simple modules corresponding to v_1 and v_2 respectively. Then $\{S_1, S_2\}$ is an atomic set in $\mathcal{T}$. The arrow represents a nonzero element in $\operatorname{Ext}^1_B(S_1, S_2)$. Hence

$$\operatorname{Hom}_{\mathcal{T}}(S_1, S_2[1]) \cong \operatorname{Ext}^1_B(S_1, S_2) \neq 0.$$

By Lemma 4.2, S_1 is isomorphic to a complex shift of S_2. But this is impossible. Therefore, the claim holds.

It follows from the claim in the last paragraph that $B = B_1 \oplus \cdots \oplus B_n$ where each B_i is a finite dimensional local ring corresponding to a vertex, say v_i. Next we claim that each B_i is commutative. Without loss of generality, we can assume $B_i = B$.

Now let ι be the fully faithful embedding from

$$\iota : \mathcal{T} := D^b(\mathrm{Mod}_{f.d.} - B) \to \mathcal{A} := D^b(\mathrm{Mod} - A).$$

Let S be the unique simple left B-module. Then, by Lemma 4.1, there is a maximal ideal $\mathfrak{m}$ of A such that $\iota(S) = A/\mathfrak{m}[w]$ for some $w \in \mathbb{Z}$. After a shift, we might assume that $\iota(S) = A/\mathfrak{m}$. The left B-module B has a composition series such that each simple subquotient is isomorphic to S, which implies that, as a left A-module, $\iota(B)$ is generated by $A/\mathfrak{m}$ in $\mathcal{A}$. By induction on the length of B, one sees that, for every $n \in \mathbb{Z}$, $H^n(\iota(B))$ is a left $A/\mathfrak{m}^d$-module for some $d \gg 0$ (we can take $d = \mathrm{length}({}_B B)$). Since $\mathrm{Hom}_{\mathcal{A}}(\iota(B), \iota(B)[-i]) = \mathrm{Hom}_{\mathcal{T}}(B, B[-i]) = 0$ for all $i > 0$, the proof of Lemma 3.11(2) shows that $\iota(B) \cong M[m]$ for some left $A/\mathfrak{m}^d$-module M and $m \in \mathbb{Z}$. Since there are nonzero maps from S to B and from B to S, we have nonzero maps from $A/\mathfrak{m}$ to $\iota(B)$ and from $\iota(B)$ to $A/\mathfrak{m}$. This implies that $m = 0$. Since B is local (and then $B/\mathfrak{m}_B$ is 1-dimensional for the maximal ideal $\mathfrak{m}_B$), this forces that $M = A/I$ where I is an ideal of A containing $\mathfrak{m}^d$. Finally,

$$B^{\mathrm{op}} = \mathrm{End}_B(B) \cong \mathrm{End}_{\mathcal{A}}(A/I, A/I) = \mathrm{End}_A(A/I, A/I) \cong A/I$$

which is commutative. Hence B is commutative. □

5. Examples

In this section we give three examples.

Frobenius–Perron theory of projective line $\mathbb{P}^1 := \mathrm{Proj}\, \Bbbk[t_0, t_1]$.

Example 5.1. Let $\mathrm{coh}(\mathbb{P}^1) =: \mathfrak{A}$ denote the category of coherent sheaves on $\mathbb{P}^1$. We will calculate the fp dimension of various functors.

Proposition 5.1.1. *Every brick object X in $\mathfrak{A}$ (namely, satisfying $\mathrm{Hom}_{\mathbb{P}^1}(X, X) = \Bbbk$) is either $\mathcal{O}(m)$ for some $m \in \mathbb{Z}$ or $\mathcal{O}_p$ for some $p \in \mathbb{P}^1$.*

The above fact is well known and follows easily from Grothendieck theorem (see also [Brüning and Burban 2007, Example 3.18]).

Let ϕ be in $\Phi_{n,b}(\mathrm{coh}(\mathbb{P}^1))$. If $n = 1$ or ϕ is a singleton, then there are two cases: either $\phi = \{\mathcal{O}(m)\}$ or $\phi = \{\mathcal{O}_p\}$. Let E^1 be the functor $\mathrm{Ext}^1_{\mathbb{P}^1}(-, -)$. In the first case, $\rho(A(\phi, E^1)) = 0$ because $\mathrm{Ext}^1_{\mathbb{P}^1}(\mathcal{O}(m), \mathcal{O}(m)) = 0$, and in the second case, $\rho(A(\phi, E^1)) = 1$ because $\mathrm{Ext}^1_{\mathbb{P}^1}(\mathcal{O}_p, \mathcal{O}_p) = 1$.

If $|\phi| > 1$, then $\mathcal{O}(m)$ can not appear in ϕ as $\mathrm{Hom}_{\mathbb{P}^1}(\mathcal{O}(m), \mathcal{O}(m')) \neq 0$ and $\mathrm{Hom}_{\mathbb{P}^1}(\mathcal{O}(m), \mathcal{O}_p) \neq 0$ for all $m \leq m'$ and $p \in \mathbb{P}^1$. Hence, ϕ is a collection of $\mathcal{O}_p$ for finitely many distinct points p's. In this case, the adjacency matrix is the identity $n \times n$-matrix and $\rho(A(\phi, E^1)) = 1$. Therefore

$$\mathrm{fpd}^n(\mathrm{coh}(\mathbb{P}^1)) = \mathrm{fpd}(\mathrm{coh}(\mathbb{P}^1)) = 1 \tag{E5.1.1}$$

for all $n \geq 1$. Since $\operatorname{coh}(\mathbb{P}^1)$ is hereditary, by Theorem 3.5(3,4), we obtain that

$$\operatorname{fpd}^n(D^b(\operatorname{coh}(\mathbb{P}^1))) = \operatorname{fpd}(D^b(\operatorname{coh}(\mathbb{P}^1))) = 1 \tag{E5.1.2}$$

for all $n \geq 1$.

Let K_2 be the Kronecker quiver

$$\begin{tikzcd} \bullet \arrow[r, bend left] \arrow[r, bend right] & \bullet \end{tikzcd} \tag{E5.1.3}$$

By a result of Beilinson [1978], the derived category $D^b(\operatorname{Mod}_{f.d.} -\Bbbk K_2)$ is triangulated equivalent to $D^b(\operatorname{coh}(\mathbb{P}^1))$. As a consequence,

$$\operatorname{fpd}(D^b(\operatorname{Mod}_{f.d.} -\Bbbk K_2)) = \operatorname{fpd}(D^b(\operatorname{coh}(\mathbb{P}^1))) = 1. \tag{E5.1.4}$$

It is easy to see, or by Theorem 1.8(1),

$$\operatorname{fpd} K_2 = 0$$

where fpd of a quiver is defined in Definition 1.6.

This implies that

$$\operatorname{fpd}(D^b(\operatorname{Mod}_{f.d.} -\Bbbk K_2)) > \operatorname{fpd} K_2. \tag{E5.1.5}$$

Next we consider some general auto-equivalences of $D^b(\operatorname{coh}(\mathbb{P}^1))$. Let

$$(m) : \operatorname{coh}(\mathbb{P}^1) \to \operatorname{coh}(\mathbb{P}^1)$$

be the auto-equivalence induced by the shift of degree m of the graded modules over $\Bbbk[t_0, t_1]$ and let Σ be the suspension functor of $D^b(\operatorname{coh}(\mathbb{P}^1))$. Then the Serre functor S of $D^b(\operatorname{coh}(\mathbb{P}^1))$ is $\Sigma \circ (-2)$. Let σ be the functor $\Sigma^a \circ (b)$ for some $a, b \in \mathbb{Z}$. By Theorem 3.5(1),

$$\operatorname{fpd}^n(\Sigma^a \circ (b)) = 0 \quad \forall a \neq 0, 1.$$

For the rest we consider $a = 0$ or 1. By Theorem 3.5(2,3), we only need to consider fpd on $\operatorname{coh}(\mathbb{P}^1)$.

If ϕ is a singleton $\{\mathcal{O}(n)\}$, then the adjacency matrix is

$$A(\phi, \sigma) = \dim(\mathcal{O}, \Sigma^a \mathcal{O}(b)) = \begin{cases} 0 & a = 0, b < 0, \\ b+1 & a = 0, b \geq 0, \\ 0 & a = 1, b \geq -1, \\ -b-1 & a = 1, b < -1. \end{cases}$$

This follows from the well-known computation of $H^i_{\mathbb{P}^1}(\mathcal{O}(m))$ for $i = 0, 1$ and $m \in \mathbb{Z}$. (It also follows from a more general computation [Artin and Zhang 1994, Theorem 8.1].) If $\phi = \{\mathcal{O}_p\}$ for some $p \in \mathbb{P}^1$, then the adjacency matrix is

$$A(\phi, \sigma) = \dim(\mathcal{O}_p, \Sigma^a(\mathcal{O}_p)) = 1 \quad \text{for } a = 0, 1.$$

It is easy to see from the above computation that

$$\operatorname{fpd}^1(\Sigma^a \circ (b)) = \begin{cases} 1 & a=0, b<0, \\ b+1 & a=0, b\geq 0, \\ 1 & a=1, b\geq -1, \\ -b-1 & a=1, b<-1. \end{cases} \tag{E5.1.6}$$

Now we consider the case when $n>1$. If $\phi \in \Phi_{n,b}(\operatorname{coh}(\mathbb{P}^1))$, ϕ is a collection of $\mathcal{O}_p$ for finitely many distinct p's. In this case, the adjacency matrix $A(\phi, \Sigma^a \circ (b))$ is the identity $n\times n$-matrix for $a=0,1$, and $\rho(A(\phi,\sigma))=1$. Therefore

$$\operatorname{fpd}^n(\Sigma^a \circ (b)) = 1 \tag{E5.1.7}$$

for all $n>1$, when restricted to the category $\operatorname{coh}(\mathbb{P}^1)$.

It follows from Theorem 3.5(2,3) that:

Claim 5.1.2. *Consider $\Sigma^a \circ (b)$ as an endofunctor of $D^b(\operatorname{coh}(\mathbb{P}^1))$. For $a,b\in\mathbb{Z}$ and $n\geq 1$, we have*

$$\operatorname{fpd}^n(\Sigma^a \circ (b)) = \begin{cases} 0 & a\neq 0,1, \\ 1 & a=0, b<0, \\ b+1 & a=0, b\geq 0, \\ 1 & a=1, b\geq -1, \\ -b-1 & a=1, b<-1. \end{cases} \tag{E5.1.8}$$

Since $S=\Sigma\circ(-2)$, we have the following

$$\operatorname{fpd}^n(\Sigma^a \circ S^b) = \operatorname{fpd}^n(\Sigma^{a+b}\circ(-2b)) = \begin{cases} 0 & a+b\neq 0,1, \\ 1 & a+b=0, b>0, \\ -(2b-1) & a+b=0, b\leq 0, \\ 1 & a+b=1, b\leq 0, \\ 2b-1 & a+b=1, b>0. \end{cases} \tag{E5.1.9}$$

Claim 5.1.3. *Since $D^b(\operatorname{coh}(\mathbb{P}^1))$ and $D^b(\operatorname{Mod}_{f.d.}-\Bbbk K_2)$ are equivalent, (E5.1.9) agrees with the* fp-*theory of $D^b(\operatorname{Mod}_{f.d.}-\Bbbk K_2)$.*

Frobenius–Perron theory of the quiver A_2. We start with the following example.

Example 5.2. Let A be the $\mathbb{Z}$-graded algebra $\Bbbk[x]/(x^2)$ with $\deg x=1$. Let $\mathcal{C}:=\operatorname{gr}-A$ be the category of finitely generated graded left A-modules. Let $\sigma:=(-)$ be the degree shift functor of $\mathcal{C}$. It is clear that σ is an autoequivalence of $\mathcal{C}$. Let $\mathfrak{A}$ be the additive subcategory of $\mathcal{C}$ generated by $\sigma^n(A)=A(n)$ for all $n\in\mathbb{Z}$. Note that $\mathfrak{A}$ is not abelian and that every object in $\mathfrak{A}$ is of the form $\bigoplus_{n\in\mathbb{Z}} A(n)^{\oplus p_n}$ for some integers $p_n\geq 0$. Since the Hom-set in the graded module category consists of homomorphisms of degree zero, we have

$$\operatorname{Hom}_{\mathfrak{A}}(A, A(n)) = \begin{cases} \Bbbk & n=0,1, \\ 0 & \text{otherwise}. \end{cases}$$

In the following diagram each arrow represents a 1-dimensional Hom for all possible Hom-set for different objects $A(n)$

$$\cdots \to A(-2) \to A(-1) \to A(0) \to A(1) \to A(2) \to \cdots \tag{E5.2.1}$$

where the number of arrows from $A(m)$ to $A(n)$ agrees with $\dim \mathrm{Hom}(A(m), A(n))$. It is easy to see that the set of indecomposable objects is $\{A(n)\}_{n\in\mathbb{Z}}$, which is also the set of bricks in $\mathfrak{A}$.

Lemma 5.3. *Retain the notation as in Example 5.2. When restricting σ onto the category $\mathfrak{A}$, we have, for every $m \geq 1$,*

$$\mathrm{fpd}^m(\sigma^n) = \begin{cases} 1 & n = 0, 1, \\ 0 & \textit{otherwise.} \end{cases} \tag{E5.3.1}$$

Proof. When $n = 0$, (E5.3.1) is trivial. Let $n = 1$. For each set $\phi \in \Phi_{m,b}$, we can assume that $\phi = \{A(d_1), A(d_2,), \dots, A(d_m)\}$ for a strictly increasing sequence $\{d_i \mid i = 1, 2, \dots, m\}$. For any $i < j$, the (i, j)-entry of the adjacency matrix is

$$a_{ij} = \dim(A(d_i), A(d_j + 1)) = 0.$$

Thus $A(\phi, \sigma)$ is a lower triangular matrix with

$$a_{ii} = \dim(A(d_i), A(d_i + 1)) = 1.$$

Hence $\rho(A(\phi, \sigma)) = 1$. So $\mathrm{fpd}^m(\sigma) = 1$.

Similarly, $\mathrm{fpd}^m(\sigma^n) = 0$ when $n > 1$ as $\dim(A(d_i), A(d_i + 2)) = 0$ for all i.

Let $n < 0$. Let $\phi = \{A(d_1), A(d_2,), \dots, A(d_m)\} \in \Phi_{m,b}$ where d_i are strictly decreasing. Then $a_{ij} = \dim(A(d_i), A(d_j + n)) = 0$ for all $i \leq j$. Thus $\rho(A(\phi, \sigma^n)) = 0$ and (E5.3.1) follows in this case. □

Example 5.4. Consider the quiver A_2

$$\bullet_2 \to \bullet_1. \tag{E5.4.1}$$

Let P_i (respectively, I_i) be the projective (respectively, injective) left $\Bbbk A_2$-modules corresponding to vertices i, for $i = 1, 2$, It is well-known that there are only three indecomposable left modules over A_2, with Auslander–Reiten quiver (or AR-quiver, for short)

$$P_2 \to P_1 (= I_2) \to I_1 \tag{E5.4.2}$$

where each arrow represents a nonzero homomorphism (up to a scalar) [Schiffler 2014, Example 1.13, pages 24–25]. The AR-translation (or translation, for short) τ is determined by $\tau(I_1) = P_2$. Let $\mathcal{T}$ be $D^b(\mathrm{Mod}_{f.d.} - \Bbbk A_2)$. The Auslander–Reiten theory can be extended from the module category to the derived category. It is direct that, in $\mathcal{T}$, we have the AR-quiver of all indecomposable objects

$$\begin{tikzcd} & P_2[-1] \arrow[dr] & & I_1[-1] \arrow[dr] & & P_1 = I_2 \arrow[dr] & & P_2[1] \arrow[dr] & & I_1[1] & \\ \cdots & & P_1[-1] \arrow[ur] & & P_2 \arrow[ur] & & I_1 \arrow[ur] & & P_1[1] \arrow[ur] & & \cdots \end{tikzcd} \tag{E5.4.3}$$

The above represents all possible nonzero morphisms (up to a scalar) between nonisomorphic indecomposable objects in $\mathcal{T}$. Note that $\mathcal{T}$ has a Serre functor S and that the AR-translation τ can be extended to a functor of $\mathcal{T}$ such that $S = \Sigma \circ \tau$ [Reiten and Van den Bergh 2002, Proposition I.2.3] or [Crawley-Boevey 1992, Remarks(2), page 23]. After identifying

$$P_2[i] \leftrightarrow A(3i), \quad P_1[i] \leftrightarrow A(3i+1), \quad I_1[i] \leftrightarrow A(3i+2),$$

(E5.4.3) agrees with (E5.2.1). Using the above identification, at least when restricted to objects, we have

$$\Sigma(A(i)) \cong A(i+3), \tag{E5.4.4}$$

$$\tau(A(i)) \cong A(i-2), \tag{E5.4.5}$$

$$S(A(i)) \cong A(i+1). \tag{E5.4.6}$$

It follows from the definition of the AR-quiver [Auslander et al. 1995, VII] that the degree of τ is -2, see also [Assem et al. 2006, Picture on page 131]. Equation (E5.4.5) just means that the degree of τ is -2.

By (E5.4.6), the Serre functor S satisfies the property of σ defined in Example 5.2. By Lemma 5.3 or (E5.3.1), we have

$$\operatorname{fpd}^n(\Sigma^a \circ S^b) = \operatorname{fpd}^n(\sigma^{3a+b}) = \begin{cases} 1 & 3a+b = 0, 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore the fp-S-theory of $\mathcal{T}$ is given as above.

In particular, we have proven

$$\operatorname{fpgldim}(D^b(\mathrm{Mod}_{f.d.} - \Bbbk A_2)) = \operatorname{fpd}(D^b(\mathrm{Mod}_{f.d.} - \Bbbk A_2)) = \operatorname{fpd}(\Sigma) = 0,$$

which is less than $\operatorname{gldim} \Bbbk A_2 = 1$.

An example of nonintegral Frobenius–Perron dimension. In the next example, we "glue" K_2 in (E5.1.3) and A_2 in (E5.4.1) together.

Example 5.5. Let G_2 be the quiver

$$\begin{tikzcd} \underset{1}{\bullet} \arrow[r, bend left=50, "\beta"] \arrow[r, bend left=15, "\gamma"'] & \underset{2}{\bullet} \arrow[l, bend left=40, "\alpha"] \end{tikzcd} \tag{E5.5.1}$$

consisting of two vertices 1 and 2, with arrow $\alpha : 2 \to 1$ and $\beta, \gamma : 1 \to 2$ satisfying relations

$$R: \quad \beta\alpha = \gamma\alpha = 0, \quad \alpha\beta = \alpha\gamma = 0. \tag{E5.5.2}$$

Note that (G_2, R) is a quiver with relations. The corresponding quiver algebra with relations is a 5-dimensional algebra

$$A = \Bbbk e_1 + \Bbbk e_2 + \Bbbk\alpha + \Bbbk\beta + \Bbbk\gamma.$$

We can use the following matrix form to represent the algebra A

$$A = \begin{pmatrix} \Bbbk e_1 & \Bbbk\alpha \\ \Bbbk\beta + \Bbbk\gamma & \Bbbk e_2 \end{pmatrix}.$$

For each $i = 1, 2$, let S_i be the left simple A-module corresponding to the vertex i and P_i be the projective cover of S_i. Then $P_1 \cong Ae_1$ is isomorphic to the first column of A, namely $\left(\begin{smallmatrix} \Bbbk e_1 \\ \Bbbk\beta+\Bbbk\gamma \end{smallmatrix}\right)$, and $P_2 \cong Ae_2$ is isomorphic to the second column of A, namely $\left(\begin{smallmatrix} \Bbbk\alpha \\ \Bbbk e_2 \end{smallmatrix}\right)$.

We will show that the Frobenius–Perron dimension of the category of finite dimensional representations of (G_2, R) is $\sqrt{2}$, by using several lemmas below that contain some detailed computations.

Lemma 5.6. *Let $V = (V_1, V_2)$ be a representation of (G_2, R). Let $\overline{W} = \operatorname{im}\alpha$ and $K = \ker\alpha$. Take a $\Bbbk$-space decomposition $V_2 = W \oplus K$ where $W \cong \overline{W}$. Then there is a decomposition of (G_2, R)-representations $V \cong (\overline{W} \oplus T, W \oplus K) \cong (\overline{W}, W) \oplus (T, K)$ where α is the identity when restricted to W (and identifying W with $\overline{W}$) and is zero when restricted to K, where β and γ are zero when restricted to $\overline{W}$.*

Proof. Since $\overline{W} = \operatorname{im}\alpha$, $V_2 \cong W \oplus K$ where $K = \ker\alpha$ and $W \cong \overline{W}$. Write $V_1 = \overline{W} \oplus T$ for some $\Bbbk$-subspace $T \subseteq V_1$. The assertion follows by using the relations in (E5.5.2). □

Recall that A_2 is the quiver given in (E5.4.1) and K_2 is the Kronecker quiver given in (E5.1.3). By the above lemma, the subrepresentation (W, W) (where we identify $\overline{W}$ with W) is in fact a representation of $\left(\begin{smallmatrix} \Bbbk e_1 & \Bbbk\alpha \\ 0 & \Bbbk e_2 \end{smallmatrix}\right)(\cong \Bbbk A_2)$ and the subrepresentation (T, K) is a representation of $\left(\begin{smallmatrix} \Bbbk e_1 & 0 \\ \Bbbk\beta+\Bbbk\gamma & \Bbbk e_2 \end{smallmatrix}\right)(\cong \Bbbk K_2)$.

Let I_n be the $n \times n$-identity matrix. Let $Bl(\lambda)$ denote the block matrix

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

Lemma 5.7. *Suppose $\Bbbk$ is of characteristic zero. The following is a complete list of indecomposable representations of (G_2, R).*

(1) $P_2 \cong (\Bbbk, \Bbbk)$, *where* $\alpha = I_1$ *and* $\beta = \gamma = 0$.

(2) $X_n(\lambda) = (K, K)$ *with* $\dim K = n$, *where* $\alpha = 0$, $\beta = I_n$ *and* $\gamma = Bl(\lambda)$ *for some* $\lambda \in \Bbbk$.

(3) $Y_n = (K, K)$ *with* $\dim K = n$, *where* $\alpha = 0$, $\beta = Bl(0)$ *and* $\gamma = I_n$.

(4) $S_{2,n} = (T, K)$ *with* $\dim T = n$ *and* $\dim K = n + 1$, *where* $\alpha = 0$, $\beta = (I_n, 0)$ *and* $\gamma = (0, I_n)$.

(5) $S_{1,n} = (T, K)$ *with* $\dim T = n + 1$ *and* $\dim K = n$, *where* $\alpha = 0$, $\beta = (I_n, 0)^\tau$ *and* $\gamma = (0, I_n)^\tau$.

As a consequence, $\Bbbk G_2/(R)$ is of tame representation type (Definition 7.1).

Proof. (1) By Lemma 5.6, this is the only case that could happen when $\alpha \neq 0$. Now we assume $\alpha = 0$.

(2), (3), (4) and (5) If $\alpha = 0$, then we are working with representations of Kronecker quiver K_2 (E5.1.3). The classification follows from a classical result of Kronecker [Benson 1991, Theorem 4.3.2].

By (1)–(5), for each integer n, there are only finitely many 1-parameter families of indecomposable representations of dimension n. Therefore A is of tame representation type. □

The following is a consequence of Lemma 5.7 and a direct computation.

Lemma 5.8. *Retain the hypotheses of Lemma 5.7. The following is a complete list of brick representations of* (G_2, R):

(1) $P_2 \cong (\Bbbk, \Bbbk)$, *where* $\alpha = I_1$ *and* $\beta = \gamma = 0$.

(2) $X_1(\lambda) = (\Bbbk, \Bbbk)$, *where* $\alpha = 0$, $\beta = I_1$ *and* $\gamma = \lambda I_1$ *for some* $\lambda \in \Bbbk$.

(3) $Y_1 = (\Bbbk, \Bbbk)$, *where* $\alpha = 0$, $\beta = 0$ *and* $\gamma = I_1$.

(4) $S_{2,n}$ *for* $n \geq 0$.

(5) $S_{1,n}$ *for* $n \geq 0$.

The set $\Phi_{1,b}$ *consists of the above objects.*

Let $X_1(\infty)$ denote Y_1. We have the following short exact sequences of (G_2, R)-representations

$$0 \longrightarrow S_1 \longrightarrow P_2 \longrightarrow S_2 \longrightarrow 0,$$

$$0 \longrightarrow S_2 \longrightarrow X_1(\lambda) \longrightarrow S_1 \longrightarrow 0,$$

$$0 \longrightarrow S_2^{n+1} \longrightarrow S_{2,n} \longrightarrow S_1^n \longrightarrow 0,$$

$$0 \longrightarrow S_2^n \longrightarrow S_{1,n} \longrightarrow S_1^{n+1} \longrightarrow 0,$$

$$0 \longrightarrow S_2^2 \longrightarrow S_{2,n} \longrightarrow S_{1,n-1} \longrightarrow 0,$$

where $n \geq 1$ for the last exact sequence, and have the following nonzero Homs, where $A = \Bbbk G_2/(R)$:

$$\begin{aligned}
\operatorname{Hom}_A(X_1(\lambda), S_{1,n}) &\neq 0 & \forall n &\geq 1,\\
\operatorname{Hom}_A(S_{2,n}, X_1(\lambda)) &\neq 0 & \forall n &\geq 1,\\
\operatorname{Hom}_A(S_{2,m}, S_{2,n}) &\neq 0 & \forall m &\leq n,\\
\operatorname{Hom}_A(S_{1,n}, S_{1,m}) &\neq 0 & \forall m &\leq n,\\
\operatorname{Hom}_A(S_{2,n}, S_{1,m}) &\neq 0 & \forall m + n &\geq 1.
\end{aligned}$$

Lemma 5.9. *Retain the hypotheses of Lemma 5.7. The following is the complete list of zero hom-sets between brick representations of G_2 in both directions*:

(1) $\mathrm{Hom}_A(X_1(\lambda), X_1(\lambda')) = \mathrm{Hom}_A(X_1(\lambda'), X_1(\lambda)) = 0$ *if* $\lambda \neq \lambda'$ *in* $\Bbbk \cup \{\infty\}$.

(2) $\mathrm{Hom}_A(S_1, S_2) = \mathrm{Hom}_A(S_2, S_1) = 0$.

As a consequence, if $\phi \in \Phi_{n,b}$ *for some* $n \geq 2$, *then* $\phi = \{S_1, S_2\}$ *or* $\phi = \{X_1(\lambda_i)\}_{i=1}^n$ *for different parameters* $\{\lambda_1, \dots, \lambda_n\}$.

We also need to compute the Ext^1_A-groups.

Lemma 5.10. *Retain the hypotheses of Lemma 5.7. Let* $\lambda \neq \lambda'$ *be in* $\Bbbk \cup \{\infty\}$:

(1) $\mathrm{Ext}^1_A(X_1(\lambda), X_1(\lambda)) = \mathrm{Hom}_A(X_1(\lambda), X_1(\lambda)) = \Bbbk$.

(2) $\mathrm{Ext}^1_A(X_1(\lambda), X_1(\lambda')) = \mathrm{Hom}_A(X_1(\lambda), X_1(\lambda')) = 0$.

(3) $\begin{pmatrix} \mathrm{Ext}^1_A(S_1, S_1) & \mathrm{Ext}^1_A(S_1, S_2) \\ \mathrm{Ext}^1_A(S_2, S_1) & \mathrm{Ext}^1_A(S_2, S_2) \end{pmatrix} = \begin{pmatrix} 0 & \Bbbk^{\oplus 2} \\ \Bbbk & 0 \end{pmatrix}$.

(4) $\mathrm{Ext}^1_A(P_2, P_2) = 0$.

(5) $\dim \mathrm{Ext}^1_A(S_{2,n}, S_{2,n}) \leq 1$ *for all* n.

(6) $\dim \mathrm{Ext}^1_A(S_{1,n}, S_{1,n}) \leq 1$ *for all* n.

Remarks 5.11. In fact, one can show the following stronger version of Lemma 5.10(5) and (6):

(5′) $\mathrm{Ext}^1_A(S_{2,n}, S_{2,n}) = 0$ for all n.

(6′) $\mathrm{Ext}^1_A(S_{1,n}, S_{1,n}) = 0$ for all n.

Proof of Lemma 5.10. (1) and (2) Consider a minimal projective resolution of $X_1(\lambda)$

$$P_1 \to P_2 \xrightarrow{f_\lambda} P_1 \to X_1(\lambda) \to 0$$

where f_λ sends $e_2 \in P_2$ to $\gamma - \lambda\beta \in P_1$. More precisely, we have

$$\begin{pmatrix} \Bbbk e_1 \\ \Bbbk\beta + \Bbbk\gamma \end{pmatrix} \xrightarrow{e_1 \to \alpha} \begin{pmatrix} \Bbbk\alpha \\ \Bbbk e_2 \end{pmatrix} \xrightarrow{e_2 \to \gamma - \lambda\beta} \begin{pmatrix} \Bbbk e_1 \\ \Bbbk\beta + \Bbbk\gamma \end{pmatrix} \to P_1/(\Bbbk(\gamma - \lambda\beta)) \to 0.$$

Applying $\mathrm{Hom}_A(-, X_1(\lambda'))$ to the truncated projective resolution of the above, we obtain the following complex

$$\Bbbk \xleftarrow{0} \Bbbk \xleftarrow{g} \Bbbk \to 0.$$

If g is zero, this is case (1). If $g \neq 0$, this is case (2).

(3) The proof is similar to the above by considering minimal projective resolutions of S_1 and S_2.

(4) This is clear since P_2 is a projective module.

(5) and (6) Let S be either $S_{2,n}$ or $S_{1,n}$. By Example 5.1, $\mathrm{fpd}(\mathrm{Mod}_{f.d} - \Bbbk K_2) = 1$. This implies that

$$\dim \mathrm{Ext}^1_{\Bbbk K_2}(S, S) \leq 1$$

where S is considered as an indecomposable K_2-module.

Let us make a comment before we continue the proof. Following a more careful analysis, one can actually show that

$$\mathrm{Ext}^1_{\Bbbk K_2}(S, S) = 0.$$

Using this fact, the rest of the proof would show the assertions (5',6') in Remarks 5.11.

Now we continue the proof. There is a projective cover $P_1^b \xrightarrow{f} S$ so that $\ker f$ is a direct sum of finitely many copies of S_2. Since P_2 is the projective cover of S_2, we have a minimal projective resolution

$$\to P_2^a \to P_1^b \to S \to 0$$

for some a, b. In the category $\mathrm{Mod}_{f.d} - \Bbbk K_2$, we have a minimal projective resolution of S

$$0 \to S_2^a \to P_1^b \to S \to 0$$

where S_2 is a projective $\Bbbk K_2$-module. Hence we have a morphism of complexes

$$\begin{tikzcd} {} \arrow[r] & P_2^a \arrow[r] \arrow[d] & P_1^b \arrow[r] \arrow[d, "="] & S \arrow[r] \arrow[d, "="] & 0 \\ 0 \arrow[r] & S_2^a \arrow[r] & P_1^b \arrow[r] & S \arrow[r] & 0 \end{tikzcd}$$

Applying $\mathrm{Hom}_A(-, S)$ to above, we obtain that

$$\begin{tikzcd} \cdots & \mathrm{Hom}_A(P_2^a, S) \arrow[l] & \mathrm{Hom}_A(P_1^b, S) \arrow[l, "h"'] & \mathrm{Hom}_A(S, S) \arrow[l] & 0 \arrow[l] \\ \cdots & \mathrm{Hom}_A(S_2^a, S) \arrow[l] \arrow[u, "g"] & \mathrm{Hom}_A(P_1^b, S) \arrow[l, "f"'] \arrow[u, "="'] & \mathrm{Hom}_A(S, S) \arrow[l] \arrow[u, "="'] & 0 \arrow[l] \end{tikzcd}$$

Note that g is an isomorphism. Since $\dim \mathrm{Ext}^1_{\Bbbk K_2}(S, S) \le 1$, the cokernel of f has dimension at most 1. Since g is an isomorphism, the cokernel of h has dimension at most 1. This implies that $\mathrm{Ext}^1_A(S, S)$ has dimension at most 1. □

Proposition 5.12. *Let* $\mathfrak{A}$ *be the category* $\mathrm{Mod}_{f.d.} - A$ *where* A *is as in Example 5.5*:

(1)
$$\mathrm{fpd}^n\, \mathfrak{A} = \begin{cases} \sqrt{2} & n = 2, \\ 1 & n \neq 2. \end{cases}$$

As a consequence, $\mathrm{fpd}\, \mathfrak{A} = \sqrt{2}$.

(2) $\mathrm{SI}(\mathfrak{A}) = 2$.

(3) $\mathrm{fpd}\, \mathcal{A} \ge \sqrt{2}$.

Proof. (1) This is a consequence of Lemmas 5.9 and 5.10. Parts (2) and (3) follow from part (1). □

Remarks 5.13. Let A be the algebra given in Example 5.5. We list some facts, comments and questions:

(1) The algebra A is nonconnected $\mathbb{N}$-graded Koszul.

(2) The minimal projective resolutions of S_1 and S_2 are

$$\cdots \to P_1^{\oplus 4} \to P_2^{\oplus 4} \to P_1^{\oplus 2} \to P_2^{\oplus 2} \to P_1 \to S_1 \to 0,$$

and

$$\cdots \to P_2^{\oplus 4} \to P_1^{\oplus 2} \to P_2^{\oplus 2} \to P_1 \to P_2 \to S_2 \to 0.$$

(3) For $i \geq 0$, we have:

$$\operatorname{Ext}^i_A(S_1, S_1) = \begin{cases} \Bbbk^{\oplus 2^{i/2}} & i \text{ is even,} \\ 0 & i \text{ is odd.} \end{cases} \quad \operatorname{Ext}^i_A(S_1, S_2) = \begin{cases} 0 & i \text{ is even,} \\ \Bbbk^{\oplus 2^{(i+1)/2}} & i \text{ is odd.} \end{cases}$$
$$\operatorname{Ext}^i_A(S_2, S_2) = \begin{cases} \Bbbk^{\oplus 2^{i/2}} & i \text{ is even,} \\ 0 & i \text{ is odd.} \end{cases} \quad \operatorname{Ext}^i_A(S_2, S_1) = \begin{cases} 0 & i \text{ is even,} \\ \Bbbk^{\oplus 2^{(i-1)/2}} & i \text{ is odd.} \end{cases}$$

(4) One can check that every algebra of dimension 4 or less has either infinite or integral fpd. Hence, A is an algebra of smallest $\Bbbk$-dimension that has finite nonintegral (or irrational) fpd. It is unknown if there is a finite dimensional algebra A such that $\operatorname{fpd}(\operatorname{Mod}_{f.d.} - A)$ is transcendental.

(5) Several authors have studied the connection between tame-wildness and complexity [Bergh and Solberg 2010; Erdmann and Solberg 2011; Farnsteiner 2007; Feldvoss and Witherspoon 2011; Külshammer 2013; Rickard 1990]. The algebra A is probably the first explicit example of a tame algebra with infinite complexity.

(6) It follows from part (3) that the fp-curvature of $\mathcal{A} := D^b(\operatorname{Mod}_{f.d.} - A)$ is $\sqrt{2}$ (some details are omitted). As a consequence, $\operatorname{fpg}(\mathcal{A}) = \infty$. By Theorem 8.3, the complexity of A is ∞. We don't know what $\operatorname{fpd} \mathcal{A}$ is.

6. σ-decompositions

We fix a category $\mathcal{C}$ and an endofunctor σ. For a set of bricks B in $\mathcal{C}$ (or a set of atomic objects when $\mathcal{C}$ is triangulated), we define

$$\operatorname{fpd}^n|_B(\sigma) = \sup\{\rho(A(\phi, \sigma)) \mid \phi := \{X_1, \ldots, X_n\} \in \Phi_{n,b} \text{ and } X_i \in B \ \forall i\}.$$

Let $\Lambda := \{\lambda\}$ be a totally ordered set. We say a set of bricks B in $\mathcal{C}$ has a *σ-decomposition* $\{B^\lambda\}_{\lambda \in \Lambda}$ (based on Λ) if the following hold:

(1) B is a disjoint union $\bigcup_{\lambda \in \Lambda} B^\lambda$.

(2) If $X \in B^\lambda$ and $Y \in B^\delta$ with $\lambda < \delta$, $\operatorname{Hom}_{\mathcal{C}}(X, \sigma(Y)) = 0$.

The following lemma is easy.

Lemma 6.1. *Let n be a positive integer. Suppose that B has a σ-*decomposition *$\{B^\lambda\}_{\lambda \in \Lambda}$. Then*

$$\operatorname{fpd}^n|_B(\sigma) \leq \sup_{\lambda \in \Lambda, m \leq n} \{\operatorname{fpd}^m|_{B^\lambda}(\sigma)\}.$$

Proof. Let ϕ be a brick set that is used in the computation of $\operatorname{fpd}^n|_B(\sigma)$. Write

$$\phi = \phi_{\lambda_1} \cup \cdots \cup \phi_{\lambda_s} \tag{E6.1.1}$$

where λ_i is strictly increasing and $\phi_{\lambda_i} = \phi \cap B^{\lambda_i}$. For any objects $X \in \phi^{\lambda_i}$ and $Y \in \phi^{\lambda_j}$, where $\lambda_i < \lambda_j$, by definition, $\operatorname{Hom}_{\mathcal{C}}(X, \sigma(Y)) = 0$. Listing the objects in ϕ in the order that suggested by (E6.1.1), then the adjacency matrix of (ϕ, σ) is of the form

$$A(\phi,\sigma) = \begin{pmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ * & A_{22} & 0 & \cdots & 0 \\ * & * & A_{33} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ * & * & * & \cdots & A_{ss} \end{pmatrix}$$

where each A_{ii} is the adjacency matrix $A(\phi_{\lambda_i}, \sigma)$. By definition,

$$\rho(A_{ii}) \le \operatorname{fpd}^{s_i}|_{B^{\lambda_i}}(\sigma)$$

where s_i is the size of A_{ii}, which is no more than n. Therefore

$$\rho(A(\phi,\sigma)) = \max_i\{\rho(A_{ii})\} \le \sup_{\lambda\in\Lambda, m\le n}\{\operatorname{fpd}^m|_{B^\lambda}(\sigma)\}.$$

The assertion follows. □

We give some examples of σ-decompositions.

Example 6.2. Let $\mathfrak{A}$ be an abelian category and $\mathcal{A}$ be the derived category $D^b(\mathfrak{A})$. Let $[n]$ be the n-fold suspension Σ^n:

(1) Suppose that α is an endofunctor of $\mathfrak{A}$ and $\bar{\alpha}$ is the induced endofunctor of $\mathcal{A}$. For each $n \in \mathbb{Z}$, let $B^n := \{M[-n] \mid M \text{ is a brick in } \mathfrak{A}\}$ and $B := \bigcup_{n\in\mathbb{Z}} B^n$. If $M_i[-n_i] \in B^{n_i}$, for $i = 1, 2$, such that $n_1 < n_2$, then

$$\operatorname{Hom}_{\mathcal{A}}(M_1[-n_1], \bar{\alpha}(M_2[-n_2])) = \operatorname{Ext}^{n_1-n_2}_{\mathfrak{A}}(M_1, \alpha(M_2)) = 0.$$

Then B has a $\bar{\alpha}$-decomposition $\{B^n\}_{n\in\mathbb{Z}}$ based on $\mathbb{Z}$.

(2) Suppose $g := \operatorname{gldim}\mathfrak{A} < \infty$. Let σ be the functor $\Sigma^g \circ \bar{\alpha}$. For each $n \in \mathbb{Z}$, let $B^n := \{M[n] \mid M \text{ is a brick in } \mathfrak{A}\}$ and $B := \bigcup_{n\in\mathbb{Z}} B^n$. If $M_i[n_i] \in B^{n_i}$, for $i = 1, 2$, such that $n_1 < n_2$, then

$$\operatorname{Hom}_{\mathcal{A}}(M_1[n_1], \sigma(M_2[n_2])) = \operatorname{Ext}^{n_2-n_1+g}_{\mathfrak{A}}(M_1, \alpha(M_2)) = 0.$$

Then B has a σ-decomposition $\{B^n\}_{n\in\mathbb{Z}}$ based on $\mathbb{Z}$.

Example 6.3. Let C be a smooth projective curve and let $\mathfrak{A}$ be the category of coherent sheaves over C. Every coherent sheaf over C is a direct sum of a torsion subsheaf and a locally free subsheaf. Define

$$B^0 = \{T \text{ is a torsion brick object in } \mathfrak{A}\}, \quad B^{-1} = \{F \text{ is a locally free brick object in } \mathfrak{A}\}, \quad B = B^{-1} \cup B^0.$$

Let σ be the functor $E^1 := \mathrm{Ext}^1_{\mathfrak{A}}(-,-)$. If $F \in B^{-1}$ and $T \in B^0$, then

$$\mathrm{Ext}^1_{\mathfrak{A}}(F, T) = 0.$$

Hence, B has an E^1-decomposition based on the totally ordered set $\Lambda := \{-1, 0\}$.

The next example is given in [Brüning and Burban 2007].

Example 6.4. Let C be an elliptic curve. Let $\mathfrak{A}$ be the category of coherent sheaves over C and $\mathcal{A}$ be the derived category $D^b(\mathfrak{A})$.

First we consider coherent sheaves. Let Λ be the totally ordered set $\mathbb{Q} \cup \{+\infty\}$. The slope of a coherent sheaf $X \neq 0$ [loc. cit., Definition 4.6] is defined to be

$$\mu(X) := \frac{\chi(X)}{\mathrm{rk}(X)} \in \Lambda$$

where $\chi(X)$ is the Euler characteristic of X and $\mathrm{rk}(X)$ is the rank of X. If X and Y are bricks such that $\mu(X) < \mu(Y)$, by [loc. cit., Corollary 4.11], X and Y are semistable, and thus by [loc. cit., Proposition 4.9(1)], $\mathrm{Hom}_{\mathfrak{A}}(Y, X) = 0$. By Serre duality (namely, Calabi–Yau property),

$$\mathrm{Hom}_{\mathcal{A}}(X, Y[1]) = \mathrm{Ext}^1_{\mathfrak{A}}(X, Y) = \mathrm{Hom}_{\mathfrak{A}}(Y, X)^* = 0. \tag{E6.4.1}$$

Write $B = \Phi_{1,b}(\mathfrak{A})$ and B^λ be the set of (semistable) bricks with slope λ. Then $B = \bigcup_{\lambda\in\Lambda} B^\lambda$. By (E6.4.1), $\mathrm{Ext}^1_{\mathfrak{A}}(X, Y) = 0$ when $X \in B^\lambda$ and $Y \in B^\nu$ with $\lambda < \nu$. Hence B has an E^1-decomposition. By Lemma 6.1, for every $n \geq 1$,

$$\mathrm{fpd}^n(E^1) = \mathrm{fpd}^n\mid_B(E^1) \leq \sup_{\lambda\in\Lambda, m\leq n} \{\mathrm{fpd}^m\mid_{B^\lambda}(E^1)\}.$$

Next we compute $\mathrm{fpd}^n\mid_{B^\lambda}(E^1)$. Let SS^λ be the full subcategory of $\mathfrak{A}$ consisting of semistable coherent sheaves of slope λ. By [loc. cit., Summary], SS^λ is an abelian category that is equivalent to SS^∞. Therefore one only needs to compute $\mathrm{fpd}^n\mid_{B^\infty}(E^1)$ in the category SS^∞. Note that SS^∞ is the abelian category of torsion sheaves and every brick object in SS^∞ is of the form $\mathcal{O}_p$ for some $p \in C$. In this case, $A(\phi, E^1)$ is the identity matrix. Consequently, $\rho(A(\phi, E^1)) = 1$. This shows that $\mathrm{fpd}^n\mid_{B^\lambda}(E^1) = \mathrm{fpd}^n\mid_{B^\infty}(E^1) = 1$ for all $n \geq 1$. It is clear that $\mathrm{fpd}^n(E^1) \geq \mathrm{fpd}^n\mid_{B^\infty}(E^1) = 1$. Combining with Lemma 6.1, we obtain that $\mathrm{fpd}^n(E^1) = 1$ for all n. (The above approach works for functors other than E^1.)

Finally we consider the fp-dimension for the derived category $\mathcal{A}$. It follows from Theorem 3.5(3) that

$$\mathrm{fpd}^n(\Sigma) = \mathrm{fpd}^n(E^1) = 1$$

for all $n \geq 1$. By definition,

$$\mathrm{fpd}(\mathcal{A}) = \mathrm{fpd}(\mathfrak{A}) = 1.$$

As we explained before fpd is an indicator of the representation types of categories.

Drozd and Greuel [2001] studied a tame-wild dichotomy for vector bundles on projective curves and introduced the notion of VB-finite, VB-tame and VB-wild similar to the corresponding notion in the representation theory of finite dimensional algebras.

Let C be a connected smooth projective curve, Drozd and Greuel [2001] showed the following:

(a) C is VB-finite if and only if C is $\mathbb{P}^1$.

(b) C is VB-tame if and only if C is elliptic (that is, of genus 1).

(c) C is VB-wild if and only if C has genus $g \geq 2$.

We now prove an fp-version of [Drozd and Greuel 2001, Theorem 1.6]. We thank Max Lieblich for providing ideas in the proof of Proposition 6.5(3).

Proposition 6.5. *Suppose $\Bbbk = \mathbb{C}$. Let $\mathbb{X}$ be a connected smooth projective curve and let g be the genus of $\mathbb{X}$:*

(1) *If $g = 0$ or $\mathbb{X} = \mathbb{P}^1$, then* $\operatorname{fpd} D^b(\operatorname{coh}(\mathbb{X})) = 1$.

(2) *If $g = 1$ or $\mathbb{X}$ is an elliptic curve, then* $\operatorname{fpd} D^b(\operatorname{coh}(\mathbb{X})) = 1$.

(3) *If $g \geq 2$, then* $\operatorname{fpd} D^b(\operatorname{coh}(\mathbb{X})) = \infty$.

Proof. (1) The assertion follows from (E5.1.4).

(2) The assertion follows from Example 6.4.

(3) By Theorem 3.5(4), $\operatorname{fpd}(D^b(\operatorname{coh}(\mathbb{X}))) = \operatorname{fpd}(\operatorname{coh}(\mathbb{X}))$. Hence it suffices to show that $\operatorname{fpd}(\operatorname{coh}(\mathbb{X})) = \infty$.

For each n, let $\{x_i\}_{i=1}^n$ be a set of n distinct points on $\mathbb{X}$. By [Drozd and Greuel 2001, Lemma 1.7], we might further assume that $2x_i \not\sim x_j + x_k$ for all $i \neq j$, as divisors on $\mathbb{X}$. Write $\mathcal{E}_i := \mathcal{O}(x_i)$ for all i. By [loc. cit., page 11], $\operatorname{Hom}_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) = 0$ for all $i \neq j$, which is also a consequence of a more general result [Huybrechts and Lehn 1997, Proposition 1.2.7]. It is clear that $\operatorname{Hom}_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_i) = \Bbbk$ for all i. Let ϕ_n be the set $\{\mathcal{E}_1, \dots, \mathcal{E}_n\}$. Then it is a brick set of nonisomorphic vector bundles on $\mathbb{X}$ (which are stable with $\operatorname{rank}(\mathcal{E}_i) = \deg(\mathcal{E}_i) = 1$ for all i).

Define the sheaf $\mathcal{H}_{ij} = \mathcal{H}om(\mathcal{E}_i, \mathcal{E}_j)$ for all i, j. Then $\deg(\mathcal{H}_{ij}) = 0$. By the Riemann–Roch theorem, we have

$$\begin{aligned} 0 &= \deg(\mathcal{H}_{ij}) \\ &= \chi(\mathcal{H}_{ij}) - \operatorname{rank}(\mathcal{H}_{ij})\chi(\mathcal{O}_{\mathbb{X}}) \\ &= \dim \operatorname{Hom}_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) - \dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) - (1-g) \\ &= \delta_{ij} - \dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) + (g-1), \end{aligned}$$

which implies that $\dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) = g - 1 + \delta_{ij}$. This formula was also given in [Drozd and Greuel 2001, page 11 before Lemma 1.7] when $i \neq j$.

Define the matrix A_n with entries $a_{ij} := \dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{E}_i, \mathcal{E}_j) = g - 1 + \delta_{ij}$, which is the adjacency matrix of (ϕ_n, E^1). This matrix has entries g along the diagonal and entries $g-1$ everywhere else. Therefore the vector $(1, \ldots, 1)$ is an eigenvector for this matrix with eigenvalue $n(g-1)+1$. So $\rho(A_n) \geq n(g-1)+1 \geq n+1$. Since we can define ϕ_n for arbitrarily large n, we must have $\operatorname{fpd}(\operatorname{coh}(\mathbb{X})) = \infty$. $\square$

Question 6.6. Let $\mathbb{X}$ be a smooth irreducible projective curve of genus $g \geq 2$. Is $\operatorname{fpd}^n(\mathbb{X})$ finite for each n? If yes, do these invariants recover g?

Proposition 6.7. *Suppose $\Bbbk = \mathbb{C}$. Let $\mathbb{Y}$ be a smooth projective scheme of dimension at least 2. Then*

$$\operatorname{fpd}^1(\operatorname{coh}(\mathbb{Y})) = \operatorname{fpd}(\operatorname{coh}(\mathbb{Y})) = \operatorname{fpd}^1(D^b(\operatorname{coh}(\mathbb{Y}))) = \operatorname{fpd}(D^b(\operatorname{coh}(\mathbb{Y}))) = \infty.$$

Proof. It is clear that $\operatorname{fpd}^1(\operatorname{coh}(\mathbb{Y}))$ is smallest among these four invariants. It suffices to show that $\operatorname{fpd}^1(\operatorname{coh}(\mathbb{Y})) = \infty$.

It is well-known that $\mathbb{Y}$ contains an irreducible projective curve $\mathbb{X}$ of arbitrarily large (either geometric or arithmetic) genus, see, for example, [Ciliberto et al. 2016, Theorem 0.1] or [Chen 1997, Theorems 1 and 2]. Let $\mathcal{O}_{\mathbb{X}}$ be the coherent sheaf corresponding to the curve $\mathbb{X}$ and let g be the arithmetic genus of $\mathbb{X}$. In the abelian category $\operatorname{coh}(\mathbb{X})$, we have

$$\dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{O}_{\mathbb{X}}, \mathcal{O}_{\mathbb{X}}) = \dim H^1(\mathbb{X}, \mathcal{O}_{\mathbb{X}}) = g.$$

Since $\operatorname{coh}(\mathbb{X})$ is a full subcategory of $\operatorname{coh}(\mathbb{Y})$, we have

$$\dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{Y}}}(\mathcal{O}_{\mathbb{X}}, \mathcal{O}_{\mathbb{X}}) \geq \dim \operatorname{Ext}^1_{\mathcal{O}_{\mathbb{X}}}(\mathcal{O}_{\mathbb{X}}, \mathcal{O}_{\mathbb{X}}) = g.$$

By taking $\phi = \{\mathcal{O}_{\mathbb{X}}\}$, one sees that $\operatorname{fpd}^1(\operatorname{coh}(\mathbb{Y})) \geq \operatorname{fpd}^1(\operatorname{coh}(\mathbb{X})) \geq g$ for all such $\mathbb{X}$. Since g can be arbitrarily large, the assertion follows. $\square$

7. Representation types

Representation types. We first recall some known definitions and results.

Definition 7.1. Let A be a finite dimensional algebra:

(1) We say A is of *finite representation type* if there are only finitely many isomorphism classes of finite dimensional indecomposable left A-modules.

(2) We say A is *tame* or *of tame representation type* if it is not of finite representation type, and for every $n \in \mathbb{N}$, all but finitely many isomorphism classes of n-dimensional indecomposables occur in a finite number of one-parameter families.

(3) We say A is *wild* or *of wild representation type* if, for every finite dimensional $\Bbbk$-algebra B, the representation theory of B can be embedded into that of A.

The following is the famous trichotomy result due to Drozd [1980].

Theorem 7.2 (Drozd's trichotomy theorem). *Every finite dimensional algebra is either of finite, tame, or wild representation type.*

Remarks 7.3. (1) An equivalent and more precise definition of a wild algebra is the following. An algebra A is called *wild* if there is a faithful exact embedding of abelian categories

$$\mathrm{Emb}: \mathrm{Mod}_{f.d.}-\Bbbk\langle x, y\rangle \to \mathrm{Mod}_{f.d.}-A \tag{E7.3.1}$$

that preserves indecomposables and respects isomorphism classes (namely, $\mathrm{Emb}(X) \cong \mathrm{Emb}(Y)$ implies that $X \cong Y$).

(2) A stronger notion of wildness is the following. An algebra A is called *strictly wild*, also called *fully wild*, if Emb in part (1) is a fully faithful embedding.

(3) It is clear that strictly wild is wild. The converse is not true.

We collect some celebrated results in terms of representation types of path algebras.

Theorem 7.4. *Let Q be a finite connected quiver*:

(1) [Gabriel 1972] *The path algebra $\Bbbk Q$ is of finite representation type if and only if the underlying graph of Q is a Dynkin diagram of type ADE.*

(2) [Nazarova 1973; Donovan and Freislich 1973] *The path algebra $\Bbbk Q$ is of tame representation type if and only if the underlying graph of Q is an extended Dynkin diagram of type $\tilde{A}\tilde{D}\tilde{E}$.*

Our main goal in this section is to prove Theorem 0.3. We thank Klaus Bongartz for suggesting the following lemma (personal communication).

Lemma 7.5. *Let A be a finite dimensional algebra that is strictly wild. Then, for each integer $a > 0$, there is a finite dimensional brick left A-module N such that $\dim \mathrm{Ext}^1_A(N, N) \geq a$.*

Proof. Let V be the vector space $\bigoplus_{i=1}^{a} \Bbbk x_i$ and let B be the finite dimensional algebra $\Bbbk\langle V\rangle/(V^{\otimes 2})$. By [Bongartz 2016, Theorem 2(i)], there is a fully faithful exact embedding

$$\mathrm{Mod}_{f.d.}-B \to \mathrm{Mod}_{f.d.}-\Bbbk\langle x, y\rangle.$$

Since A is strictly wild, there is a fully faithful exact embedding

$$\mathrm{Mod}_{f.d.}-\Bbbk\langle x, y\rangle \to \mathrm{Mod}_{f.d.}-A.$$

Hence we have a fully faithful exact embedding

$$F: \mathrm{Mod}_{f.d.}-B \to \mathrm{Mod}_{f.d.}-A. \tag{E7.5.1}$$

Let S be the trivial B-module $B/B_{\geq 1}$. It follows from an easy calculation that $\dim \mathrm{Ext}^1_B(S, S) = \dim(V)^* = a$. Since F is fully faithful exact, F induces an injection

$$F: \mathrm{Ext}^1_B(S, S) \to \mathrm{Ext}^1_A(F(S), F(S)).$$

Thus $\dim \mathrm{Ext}^1_A(F(S), F(S)) \geq a$. Since S is simple, it is a brick. Hence, $F(S)$ is a brick. The assertion follows by taking $N = F(S)$. $\square$

Proposition 7.6. (1) *Let A be a finite dimensional algebra that is strictly wild, then*

$$\operatorname{fpd}^1(E^1) = \operatorname{fpd}(\mathfrak{A}) = \operatorname{fpd}(\mathcal{A}) = \infty.$$

(2) *If $A := \Bbbk Q$ is wild, then*

$$\operatorname{fpd}^1(E^1) = \operatorname{fpd}(\mathfrak{A}) = \operatorname{fpd}(\mathcal{A}) = \infty.$$

Proof. (1) For each integer a, by Lemma 7.5, there is a brick N in $\mathfrak{A}$ such that $\operatorname{Ext}^1_{\mathfrak{A}}(N, N) \geq a$. Hence $\operatorname{fpd}^1(E^1) \geq a$. Since a is arbitrary, $\operatorname{fpd}^1(E^1) = \infty$. Consequently, $\operatorname{fpd}(\mathfrak{A}) = \infty$. By Lemma 2.9, $\operatorname{fpd}(\mathcal{A}) = \infty$.

(2) It is well-known that a wild path algebra is strictly wild, see a comment of Gabriel [1975, page 149] or [Ariki 2005, Proposition 7]. The assertion follows from part (1). □

The following lemma is based on a well-understood AR-quiver theory for acyclic quivers of finite representation type and the hammock theory introduced by Brenner [1986]. We refer to [Ringel and Vossieck 1987] if the reader is interested in a more abstract version of the hammock theory.

For a class of quivers including all ADE quivers, there is a convenient (though not essential) way of positioning the vertices as in [Assem et al. 2006, Example IV.2.6]. A quiver Q is called *well-positioned* if the vertices of Q are located so that all arrows are strictly from the right to the left of the same horizontal distance. For example, the following quiver D_n is well positioned:

$$\begin{tikzcd}
 & & & & n-1 \arrow[dl] \\
1 & 2 \arrow[l] & \cdots \arrow[l] & n-2 \arrow[l] & \\
 & & & & n \arrow[ul]
\end{tikzcd}$$

Lemma 7.7. *Let Q be a quiver such that*

(i) *the underlying graph of Q is a Dynkin diagram of type A, or D, or E, and that*

(ii) *Q is well-positioned.*

Let $A = \Bbbk Q$ and let M, N be two indecomposable left A-modules in the AR-quiver of A. Then the following hold:

(1) *There is a standard way of defining the order or degree for indecomposable left A-modules M, denoted by* $\deg M$, *such that all arrows in the AR-quiver have degree* 1, *or equivalently, all arrows are from the left to the right of the same horizontal distance. As in* (E5.4.2), *when $Q = A_2$,* $\deg P_2 = 0$, $\deg P_1 = 1$ *and* $\deg I_1 = 2$.

(2) *If* $\operatorname{Hom}_A(M, N) \neq 0$, *then* $\deg M \leq \deg N$.

(3) *The degree of the AR-translation τ is -2.*

(4) *If* $\mathrm{Ext}^1_A(M, N) \neq 0$, *then* $\deg M \geq \deg N + 2$.

(5) *There is no oriented cycle in the* E^1*-quiver of* $\mathfrak{A} := \mathrm{Mod}_{f.d.} - \Bbbk Q$, *denoted by* $Q^{E^1}_{\mathfrak{A}}$, *defined before Lemma 2.10.*

(6) $\mathrm{fpd}(\mathfrak{A}) = 0$.

Proof. (1) This is a well-known fact in AR-quiver theory. For each given quiver Q as described in (i) and (ii), one can build the AR-quiver by using the Auslander–Reiten translation τ and *the knitting algorithm*, see [Schiffler 2014, Chapter 3]. Some explicit examples are given in [Gabriel 1980, Chapter 6] and [Schiffler 2014, Chapter 3].

(2) This follows from (1). Note that the precise dimension of $\mathrm{Hom}_A(M, N)$ can be computed by using hammock theory [Brenner 1986; Ringel and Vossieck 1987]. Some examples are given in [Schiffler 2014, Chapter 3].

(3) This follows from the definition of the translation τ in the AR-quiver theory [Auslander et al. 1995, VII]. See also, [Crawley-Boevey 1992, Remarks (2), page 23].

(4) By Serre duality, $\mathrm{Ext}^1_R(M, N) = \mathrm{Hom}_A(N, \tau M)^*$ [Reiten and Van den Bergh 2002, Proposition I.2.3] or [Crawley-Boevey 1992, Lemma 1, page 22]. If $\mathrm{Ext}^1_R(M, N) \neq 0$, then, by Serre duality and part (2), $\deg N \leq \deg \tau M = \deg M - 2$. Hence $\deg M \geq \deg N + 2$.

(5) In this case, every indecomposable module is a brick. Hence the E^1-quiver $Q^{E^1}_{\mathfrak{A}}$ has the same vertices as the AR-quiver. By part (4), if there is an arrow from M to N in the quiver $Q^{E^1}_{\mathfrak{A}}$, then $\deg M \geq \deg N+2$. This means that all arrows in $Q^{E^1}_{\mathfrak{A}}$ are from the right to the left. Therefore there is no oriented cycle in $Q^{E^1}_{\mathfrak{A}}$.

(6) This follows from part (5), Theorem 1.8(1) and Lemma 2.10. □

Theorem 7.8. *Let* Q *be a finite quiver whose underlying graph is a Dynkin diagram of type* ADE *and let* $A = \Bbbk Q$. *Then* $\mathrm{fpd}(\mathfrak{A}) = \mathrm{fpd}(\mathcal{A}) = \mathrm{fpgldim}(\mathcal{A}) = 0$.

Proof. Since the path algebra A is hereditary, $\mathfrak{A}$ is a-hereditary of global dimension 1. By Theorem 3.5(3), $\mathrm{fpd}(\mathfrak{A}) = \mathrm{fpd}(\mathcal{A})$. If Q_1 and Q_2 are two quivers whose underlying graphs are the same, then, by Bernstein–Gelfand–Ponomarev (BGP) reflection functors [Bernstein et al. 1973], $D^b(\mathrm{Mod}_{f.d.} - \Bbbk Q_1)$ and $D^b(\mathrm{Mod}_{f.d.} - \Bbbk Q_2)$ are triangulated equivalent. Hence we only need prove the statement for one representative. Now we can assume that Q satisfies the hypotheses (i) and (ii) of Lemma 7.7. By Lemma 7.7(6), $\mathrm{fpd}(\mathfrak{A}) = 0$. Therefore $\mathrm{fpd}(\mathcal{A}) = 0$, or equivalently, $\mathrm{fpd}(\Sigma) = 0$. By Theorem 3.5(1), $\mathrm{fpd}(\Sigma^i) = 0$ for all $i \neq 0, 1$. Therefore $\mathrm{fpgldim}(\mathcal{A}) = 0$. □

Weighted projective lines. To prove Theorem 0.3, it remains to show part (2) of the theorem. Our proof uses a result of [Chen et al. 2019] about weighted projective lines, which we now review. Details can be found in [Geigle and Lenzing 1987, Section 1].

For $t \geq 1$, let $\boldsymbol{p} := (p_0, p_1, \ldots, p_t)$ be a $(t+1)$-tuple of positive integers, called the *weight sequence*. Let $\boldsymbol{D} := (\lambda_0, \lambda_1, \ldots, \lambda_t)$ be a sequence of distinct points of the projective line $\mathbb{P}^1$ over $\Bbbk$. We normalize

$\boldsymbol{D}$ so that $\lambda_0 = \infty$, $\lambda_1 = 0$ and $\lambda_2 = 1$ (if $t \geq 2$). Let

$$S := \Bbbk[X_0, X_1, \dots, X_t]/(X_i^{p_i} - X_1^{p_1} + \lambda_i X_0^{p_0}, i = 2, \dots, t).$$

The image of X_i in S is denoted by x_i for all i. Let $\mathbb{L}$ be the abelian group of rank 1 generated by $\vec{x_i}$ for $i = 0, 1, \dots, t$ and subject to the relations

$$p_0 \vec{x_0} = \cdots = p_i \vec{x_i} = \cdots = p_t \vec{x_t} =: \vec{c}\,.$$

The algebra S is $\mathbb{L}$-graded by setting $\deg x_i = \vec{x_i}$. The corresponding *weighted projective line*, denoted by $\mathbb{X}(\boldsymbol{p}, \boldsymbol{D})$ or simply $\mathbb{X}$, is a noncommutative space whose category of coherent sheaves is given by the quotient category

$$\mathrm{coh}(\mathbb{X}) := \frac{\mathrm{gr}^{\mathbb{L}} - S}{\mathrm{gr}^{\mathbb{L}}_{f.d.} - S}$$

where $\mathrm{gr}^{\mathbb{L}} - S$ is the category of noetherian $\mathbb{L}$-graded left S-modules and $\mathrm{gr}^{\mathbb{L}}_{f.d.} - S$ is the full subcategory of $\mathrm{gr}^{\mathbb{L}} - S$ consisting of finite dimensional modules.

The weighted projective lines are classified into the following three classes:

$$\mathbb{X} \text{ is } \begin{cases} \textit{domestic} & \text{if } \boldsymbol{p} \text{ is } (p, q), (2, 2, n), (2, 3, 3), (2, 3, 4), (2, 3, 5); \\ \textit{tubular} & \text{if } \boldsymbol{p} \text{ is } (2, 3, 6), (3, 3, 3), (2, 4, 4), (2, 2, 2, 2); \\ \textit{wild} & \text{otherwise.} \end{cases}$$

Let $\mathbb{X}$ be a weighted projective curve. Let $\mathrm{Vect}(\mathbb{X})$ be the full subcategory of $\mathrm{coh}(\mathbb{X})$ consisting of all vector bundles. Similar to the elliptic curve case, Example 6.4, one can define the concepts of *degree*, *rank* and *slope* of a vector bundle on a weighted projective curve $\mathbb{X}$, see [Lenzing and Meltzer 1993, Section 2] for details. For each $\mu \in \mathbb{Q} \cup \{\infty\}$, let $\mathrm{Vect}_\mu(\mathbb{X})$ be the full subcategory of $\mathrm{Vect}(\mathbb{X})$ consisting of all vector bundles of slope μ.

Lemma 7.9. *Let $\mathbb{X} = \mathbb{X}(\boldsymbol{p}, \boldsymbol{D})$ be a weighted projective line*:

(1) $\mathrm{coh}(\mathbb{X})$ *is noetherian and hereditary.*

(2)
$$D^b(\mathrm{coh}(\mathbb{X})) \cong \begin{cases} D^b(\mathrm{Mod}_{f.d.} - \Bbbk\tilde{A}_{p,q}) & \textit{if } \boldsymbol{p} = (p, q), \\ D^b(\mathrm{Mod}_{f.d.} - \Bbbk\tilde{D}_n) & \textit{if } \boldsymbol{p} = (2, 2, n), \\ D^b(\mathrm{Mod}_{f.d.} - \Bbbk\tilde{E}_6) & \textit{if } \boldsymbol{p} = (2, 3, 3), \\ D^b(\mathrm{Mod}_{f.d.} - \Bbbk\tilde{E}_7) & \textit{if } \boldsymbol{p} = (2, 3, 4), \\ D^b(\mathrm{Mod}_{f.d.} - \Bbbk\tilde{E}_8) & \textit{if } \boldsymbol{p} = (2, 3, 5). \end{cases}$$

(3) *Let M be a generic simple object in* $\mathrm{coh}(\mathbb{X})$. *Then* $\mathrm{Ext}^1_{\mathbb{X}}(M, M) = 1$.

(4) $\mathrm{fpd}^1(\mathrm{coh}(\mathbb{X})) \geq 1$.

(5) *If $\mathbb{X}$ is tubular or domestic, then* $\mathrm{Ext}^1_{\mathbb{X}}(X, Y) = 0$ *for all* $X \in \mathrm{Vect}_{\mu'}(\mathbb{X})$ *and* $Y \in \mathrm{Vect}_\mu(\mathbb{X})$ *with* $\mu' < \mu$.

(6) *If $\mathbb{X}$ is domestic, then* $\mathrm{Ext}^1_{\mathbb{X}}(X, Y) = 0$ *for all* $X \in \mathrm{Vect}_{\mu'}(\mathbb{X})$ *and* $Y \in \mathrm{Vect}_\mu(\mathbb{X})$ *with* $\mu' \leq \mu$. *As a consequence*, $\mathrm{fpd}(\Sigma|_{\mathrm{Vect}_{\mu'}(\mathbb{X})}) = 0$ *for all* $\mu < \infty$.

(7) *Suppose $\mathbb{X}$ is tubular. Then every indecomposable vector bundle on $\mathbb{X}$ is semistable.*

(8) *Suppose $\mathbb{X}$ is tubular and let $\mu \in \mathbb{Q}$. Then each* $\mathrm{Vect}_{\mu}(\mathbb{X})$ *is a uniserial category. Accordingly indecomposables in* $\mathrm{Vect}_{\mu}(\mathbb{X})$ *decomposes into Auslander–Reiten components, which all are tubes of finite rank.*

Proof. (1) This is well known.

(2) [Geigle and Lenzing 1987, 5.4.1].

(3) Let M be a generic simple object. Then M is a brick and $\mathrm{Ext}^1(M, M) = 1$.

(4) Follows from (3) by taking $\phi := \{M\}$.

(5) This is [Schiffmann 2012, Corollary 4.34(i)] since tubular is also called elliptic in that work.

(6) This is [Schiffmann 2012, Comments after Corollary 4.34] since domestic is also called parabolic in that work. The consequence is clear.

(7) [Geigle and Lenzing 1987, Theorem 5.6(i)].

(8) [Geigle and Lenzing 1987, Theorem 5.6(iii)]. □

We will use the following result which is proved in [Chen et al. 2019].

Theorem 7.10. *Let $\mathbb{X}$ be a weighted projective line*:

(1) *If $\mathbb{X}$ is domestic*, *then* $\operatorname{fpd} D^b(\mathrm{coh}(\mathbb{X})) = 1$.

(2) *If $\mathbb{X}$ is tubular*, *then* $\operatorname{fpd} D^b(\mathrm{coh}(\mathbb{X})) = 1$.

(3) *If $\mathbb{X}$ is wild*, *then* $\operatorname{fpd} D^b(\mathrm{coh}(\mathbb{X})) \geq \dim \mathrm{Hom}_{\mathbb{X}}(\mathcal{O}_{\mathbb{X}}, \mathcal{O}_{\mathbb{X}}(\vec{\omega}))$ *where $\vec{\omega}$ is the dualizing element* [Geigle and Lenzing 1987, Section 1.2].

There is a similar statement for smooth complex projective curves (Proposition 6.5). The authors are interested in answering the following question.

Question 7.11. Let $\mathbb{X}$ be a wild weighted projective line. What is the exact value of $\operatorname{fpd}^n D^b(\mathrm{coh}(\mathbb{X}))$?

Tubes. The following example is studied in [Chen et al. 2019], which is dependent on direct linear algebra calculations.

Example 7.12. Let ξ be a primitive n-th root of unity. Let T_n be the algebra

$$T_n := \frac{\Bbbk\langle g, x\rangle}{(g^n - 1, xg - \xi gx)}.$$

This algebra can be expressed by using a group action. Let G be the group

$$\{g \mid g^n = 1\} \cong \mathbb{Z}/(n)$$

acting on the polynomial ring $\Bbbk[x]$ by $g \cdot x = \xi x$. Then T_n is naturally isomorphic to the skew group ring $\Bbbk[x] * G$. Let $\overrightarrow{A_{n-1}}$ denote the cycle quiver with n vertices, namely, the quiver with one oriented cycle connecting n vertices. It is also known that T_n is isomorphic to the path algebra of the quiver $\overrightarrow{A_{n-1}}$. Then $\mathrm{fpd}(\mathrm{Mod}_{f.d.} - T_n) = 1$ by [Chen et al. 2019].

Proof of Theorem 0.3. Part (1) follows from Theorems 7.4(1) and 7.8 and part (3) follows from Proposition 7.6(2). It remains to deal with part (2).

By Theorem 7.4(2), Q must be of type either $\overrightarrow{A_{n-1}}$, or $\tilde{A}_{p,q}$, or $\tilde{D}_n$, or $\tilde{E}_{6,7,8}$. If Q is of type $\overrightarrow{A_{n-1}}$, the assertion follows from Example 7.12. If Q is of type $\tilde{A}_{p,q}$, $\tilde{D}_n$, or $\tilde{E}_{6,7,8}$, the assertion follows from Lemma 7.9(2) and Theorem 7.10(1). □

8. Complexity

The concept of complexity was first introduced by Alperin and Evens [1981] in the study of group cohomology. Since then the study of complexity has been extended to finite dimensional algebras, Frobenius algebras, Hopf algebras and commutative algebras. First we recall the classical definition of the complexity for finite dimensional algebras and then give a definition of the complexity for triangulated categories. We give the following modified (but equivalent) version, which can be generalized.

Definition 8.1. Let A be a finite dimensional algebra and $T = A/J(A)$ where $J(A)$ is the Jacobson radical of A. Let M be a finite dimensional left A-module:

(1) The *complexity* of M is defined to be

$$\mathrm{cx}(M) := \limsup_{n\to\infty} \log_n(\dim \mathrm{Ext}^n_A(M, T)) + 1.$$

(2) The *complexity* of the algebra A is defined to be

$$\mathrm{cx}(A) := \mathrm{cx}(T).$$

In the original definition of *complexity* by Alperin and Evens [1981] and in most other papers, the dimension of n-syzygies is used instead of the dimension of the Ext^n-groups, but it is easy to see that the asymptotic behavior of these two series are the same, therefore these give rise to the same complexity. It is well-known that $\mathrm{cx}(M) \le \mathrm{cx}(A)$ for all finite dimensional left A-modules M. Next we introduce the notion of a complexity for a triangulated category which is partially motivated by the work in [Bao et al. 2019, Section 4].

Definition 8.2. Let $\mathcal{T}$ be a pretriangulated category. Let d be a real number:

(1) The left subcategory of complexity less than d is defined to be

$${}_d\mathcal{T} := \left\{X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d-1}} \dim \mathrm{Hom}_{\mathcal{T}}(X, \Sigma^n(Y)) = 0, \forall Y \in \mathcal{T}\right\}.$$

(2) The right subcategory of complexity less than d is defined to be

$$\mathcal{T}_d := \left\{X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d-1}} \dim \mathrm{Hom}_{\mathcal{T}}(Y, \Sigma^n(X)) = 0, \forall Y \in \mathcal{T}\right\}.$$

(3) The *complexity* of $\mathcal{T}$ is defined to be

$$\operatorname{cx}(\mathcal{T}) := \inf\{d \mid {}_d\mathcal{T} = \mathcal{T}\}.$$

(4) The *Frobenius–Perron complexity* of $\mathcal{T}$ is defined to be

$$\operatorname{fpcx}(\mathcal{T}) := \operatorname{fpg}(\Sigma) + 1.$$

Note that it is not hard to show that $\operatorname{cx}(\mathcal{T}) = \inf\{d \mid \mathcal{T}_d = \mathcal{T}\}$.

Theorem 8.3. *Let $\mathcal{T}$ be a pretriangulated category. Then* $\operatorname{fpcx}(\mathcal{T}) \leq \operatorname{cx}(\mathcal{T})$.

Proof. Let d be any number strictly larger than $\operatorname{cx}(\mathcal{T})$. We need to show that $\operatorname{fpcx}(\mathcal{T}) \leq d$.

Let $\phi \in \Phi_{m,a}$ be an atomic set and let $X := \bigoplus_{X_i \in \phi} X_i$. Then, by definition,

$$\lim_{n\to\infty} \frac{\dim \operatorname{Hom}_{\mathcal{T}}(X, \Sigma^n(X))}{n^{d-1}} = 0.$$

Then there is a constant C such that $\dim \operatorname{Hom}_{\mathcal{T}}(X, \Sigma^n(X)) < Cn^{d-1}$ for all $n > 0$. Since each X_i is a direct summand of X, we have

$$a_{ij}(n) := \dim \operatorname{Hom}_{\mathcal{T}}(X_i, \Sigma^n(X_j)) < Cn^{d-1}$$

for all i, j. This means that each entry $a_{ij}(n)$ in the adjacency matrix of $A(\phi, \Sigma^n)$ is less than Cn^{d-1}. Therefore $\rho(A(\phi, \Sigma^n)) < mCn^{d-1}$. By Definition 2.3(3), $\operatorname{fpg}(\Sigma) \leq d-1$. Thus $\operatorname{fpcx}(\mathcal{T}) \leq d$ as desired. $\square$

We will prove that the equality $\operatorname{fpcx}(\mathcal{T}) = \operatorname{cx}(\mathcal{T})$ holds under some extra hypotheses. Let A be a finite dimensional algebra with a complete list of simple left A-modules $\{S_1, \ldots, S_w\}$. We use n for any integer and i, j for integers between 1 and w. Define, for $i \leq j$,

$$p_{ij}(n) := \min\{\dim \operatorname{Ext}^n_A(S_i, S_j), \dim \operatorname{Ext}^n_A(S_j, S_i)\}$$

and

$$P_n := \max\{p_{ij}(n) \mid i \leq j\}.$$

We say A satisfies *averaging growth condition* if there are positive integers C and d, independent of the choices of n and (i, j), such that

$$\dim \operatorname{Ext}^n_A(S_i, S_j) \leq C \max\{P_{n-d}, P_{n-d+1}, \ldots, P_{n+d}\} \tag{E8.3.1}$$

for all n and all $1 \leq i, j \leq w$.

Theorem 8.4. *Let A be a finite dimensional algebra and $\mathcal{A} = D^b(\mathrm{Mod}_{f.d.}-A)$:*

(1) $\operatorname{cx}(A) = \operatorname{cx}(\mathcal{A})$. *As a consequence*, $\operatorname{cx}(A)$ *is a derived invariant.*

(2) *If A satisfies the averaging growth condition, then* $\operatorname{fpcx}(\mathcal{A}) = \operatorname{cx}(\mathcal{A}) = \operatorname{cx}(A)$. *As a consequence, if A is local or commutative, then* $\operatorname{fpcx}(\mathcal{A}) = \operatorname{cx}(\mathcal{A}) = \operatorname{cx}(A)$.

We will prove Theorem 8.4 after the next lemma.

Let $\mathcal{T}$ be a pretriangulated category with suspension Σ. We use X, Y, Z for objects in $\mathcal{T}$. Fix a family ϕ of objects in $\mathcal{T}$ and a positive number d. Define:

$$ {}_d(\phi) = \left\{ X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d-1}} \dim \operatorname{Hom}_{\mathcal{T}}(X, \Sigma^n(Y)) = 0, \forall Y \in \phi \right\}. \tag{E8.4.1}$$

$$ (\phi)_d = \left\{ X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d-1}} \dim \operatorname{Hom}_{\mathcal{T}}(Y, \Sigma^n(X)) = 0, \forall Y \in \phi \right\}. \tag{E8.4.2}$$

$$ {}^d(\phi) = \left\{ X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d}} \sum_{i\le n} \dim \operatorname{Hom}_{\mathcal{T}}(X, \Sigma^i(Y)) = 0, \forall Y \in \phi \right\}. \tag{E8.4.3}$$

$$ (\phi)^d = \left\{ X \in \mathcal{T} \mid \lim_{n\to\infty} \tfrac{1}{n^{d}} \sum_{i\le n} \dim \operatorname{Hom}_{\mathcal{T}}(Y, \Sigma^i(X)) = 0, \forall Y \in \phi \right\}. \tag{E8.4.4}$$

Lemma 8.5. *The following are full thick pretriangulated subcategories of $\mathcal{T}$ closed under direct summands*:

(1) ${}_d(\phi)$.

(2) $(\phi)_d$.

(3) ${}^d(\phi)$.

(4) $(\phi)^d$.

Proof. We only prove (1). The proofs of other parts are similar. Suppose $X \in {}_d(\phi)$. Using the fact $\lim_{n\to\infty} n^{d-1}/(n+1)^{d-1} = 1$, we see that $X[1] = \Sigma(X)$ is in ${}_d(\phi)$. Similarly, $X[-1]$ is in ${}_d(\phi)$. If $f : X_1 \to X_2$ be a morphism of objects in ${}_d(\phi)$, and let X_3 be the mapping cone of f, then, for each $Y \in \phi$, we have an exact sequence

$$\to \operatorname{Hom}_{\mathcal{T}}(X_1, \Sigma^{n-1}(Y)) \to \operatorname{Hom}_{\mathcal{T}}(X_3, \Sigma^{n}(Y)) \to \operatorname{Hom}_{\mathcal{T}}(X_2, \Sigma^{n}(Y)) \to$$

which implies that $X_3 \in {}_d(\phi)$. Therefore ${}_d(\phi)$ is a thick pretriangulated subcategory of $\mathcal{T}$. If $X \in {}_d(\phi)$ and $X = Y \oplus Z$, it is clear that $Y, Z \in {}_d(\phi)$. Therefore ${}_d(\phi)$ is closed under taking direct summands. □

Proof of Theorem 8.4. (1) Let $c = \operatorname{cx}(A)$. For every $d < c$, we have that

$$\limsup_{n\to\infty} \frac{\dim \operatorname{Ext}^n_A(T, T)}{n^{d-1}} = \infty$$

which implies that $T \notin {}_d\mathcal{A}$. Therefore $d \le \operatorname{cx}(\mathcal{A})$.

Conversely, let $d > c$. It follows from the definition that

$$\limsup_{n\to\infty} \frac{\dim \operatorname{Ext}^n_A(T, T)}{n^{d-1}} = 0.$$

This means that $T \in (\{T\})_d$. Since T generates $\mathcal{A}$, we have $\mathcal{A} = (\{T\})_d$. Again, since T generates $\mathcal{A}$, we have $\mathcal{A} = \mathcal{A}_d = {}_d\mathcal{A}$. By definition, $d \ge \operatorname{cx}(\mathcal{A})$ as desired.

(2) Assume that A satisfies the averaging growth condition. Let

$$\begin{aligned} c_1 &= \operatorname{fpcx}(\mathcal{A}), \\ c_2 &= \limsup_{n\to\infty} \log_n (C \max\{P_{n-d}, P_{n-d+1}, \dots, P_{n+d}\}) + 1, \\ c_3 &= \limsup_{n\to\infty} \log_n (P_n) + 1, \\ c_4 &= \operatorname{cx}(A) = \operatorname{cx}(\mathcal{A}). \end{aligned}$$

By calculus, we have $c_2 = c_3$. Let ϕ be the atomic set of simple objects $\{S_i\}_{i=1}^{w}$. Then $\rho(\phi, \Sigma^n) \geq p_{ij}(n)$, for all i, j, by Lemma 1.7(2). So $\rho(\phi, \Sigma^n) \geq P_n$. As a consequence, $c_1 \geq c_3$. Let $T = A/J = \bigoplus_{i=1}^{w} S_i^{d_i}$ for some finite numbers $\{d_i\}_{i=1}^{w}$. Let D be $\max_i \{d_i\}$. By the averaging growth condition, namely, (E8.3.1),

$$\begin{aligned} \dim \operatorname{Ext}_A^n(T, T) &= \sum_{i,j} d_i d_j \dim \operatorname{Ext}_A^n(S_i, S_j) \\ &\leq w^2 DC \max\{P_{n-d}, P_{n-d+1}, \dots, P_{n+d}\} \end{aligned}$$

which implies that $c_4 = \operatorname{cx}(A) = \operatorname{cx}(T) \leq c_2$. Combining with Theorem 8.3, we have $c_1 = c_2 = c_3 = c_4$ as desired.

If A is local, then there is only one simple module S_1. Then (E8.3.1) is automatic. If A is commutative, then $\operatorname{Ext}_A^i(S_i, S_j) = 0$ for all n and all $i \neq j$. Again, in this case, (E8.3.1) is obvious. The consequence follows from the main assertion. □

For all well-studied finite dimensional algebras A, (E8.3.1) holds. For example, the algebra A in Example 5.5 satisfies the averaging growth condition. This can be shown by using the computation given in Remarks 5.13(3). It is natural to ask if every finite dimensional algebra satisfies the averaging growth condition.

Theorem 0.5 follows easily from Theorems 8.3 and 8.4.

Proof of Theorem 0.5. By Definition 8.2(4), Theorems 8.3 and 8.4(1), we have

$$\operatorname{fpg}(\mathcal{A}) = \operatorname{fpcx}(\mathcal{A}) - 1 \leq \operatorname{cx}(\mathcal{A}) - 1 = \operatorname{cx}(A) - 1.$$

The assertion follows. □

Lemma 8.6. (1) *Let $\mathfrak{A}$ be an abelian category and $\mathcal{A} = D^b(\mathfrak{A})$. If* $\operatorname{gldim} \mathfrak{A} < \infty$, *then* $\operatorname{fpcx}(\mathcal{A}) = 0$.

(2) *Let $\mathcal{T}$ be a pretriangulated category. If* $\operatorname{fpgldim} \mathcal{T} < \infty$, *then* $\operatorname{fpcx}(\mathcal{T}) = 0$.

Proof. Both are easy and proofs are omitted. □

We conclude with examples of nonintegral fpg of a triangulated category.

Example 8.7. (1) Let α be any real number in $\{0\} \cup \{1\} \cup [2, \infty)$. By [Krause and Lenagan 1985, Theorem 1.8, or page 14], there is a finitely generated algebra R with $\operatorname{GKdim} R = \alpha$. More precisely, [Krause and Lenagan 1985, Theorem 1.8] implies that there is a 2-dimensional vector space $V \subset R$ that generates R such that, there are positive integers $a < b$, for every $n > 0$,

$$an^{\alpha} < \dim(\Bbbk 1 + V)^n < bn^{\alpha}.$$

Define a filtration $\mathcal{F}$ on R by

$$F_i R = (\Bbbk 1 + V)^i \quad \forall i.$$

Let A be the associated graded algebra $\operatorname{gr} R$ with respect to this grading. Then A is connected graded and generated by two elements in degree 1 and satisfying, for every $n > 0$,

$$an^{\alpha} < \sum_{i=0}^{n} \dim A_i < bn^{\alpha}. \tag{E8.7.1}$$

To match up with the definition of complexity, we further assume that there are $c < d$ such that, for every $n > 0$,

$$cn^{\alpha-1} < \dim A_n < dn^{\alpha-1}. \tag{E8.7.2}$$

This can be achieved, for example, by replacing A by its polynomial extension $A[t]$ (with $\deg t = 1$) and replacing α by $\alpha + 1$.

Next we make A a differential graded (dg) algebra by setting elements in A_i to have cohomological degree i and $d_A = 0$. For this dg algebra, we denote the derived category of left dg A-modules by $\mathcal{A}$. Let $\mathcal{O}$ be the object ${}_A A$ in $\mathcal{A}$. By the definition of the cohomological degree of A, we have

$$\operatorname{Hom}_{\mathcal{A}}(\mathcal{O}, \Sigma^i \mathcal{O}) = A_i \quad \forall i. \tag{E8.7.3}$$

Let $\mathcal{T}$ be the full triangulated subcategory of $\mathcal{A}$ generated by $\mathcal{O}$. (E8.7.3) implies that $\mathcal{O}$ is an atomic object. Now using (E8.7.3) together with (E8.7.2), we obtain that

$$\operatorname{fpcx}(\mathcal{T}) \geq \alpha. \tag{E8.7.4}$$

By (E8.7.2)-(E8.7.3), we have that, for every $d > \alpha$, $\mathcal{O} \in {}_d(\{\mathcal{O}\})$. Since $\mathcal{O}$ generates $\mathcal{T}$, we have ${}_d(\{\mathcal{O}\}) = \mathcal{T}$. The last equation means that $\mathcal{O} \in (\mathcal{T})_d$. Since $\mathcal{O}$ generates $\mathcal{T}$, we have $(\mathcal{T})_d = \mathcal{T}$. By definition, $d > \operatorname{cx}(\mathcal{T})$. Combining these with Theorem 8.3 and (E8.7.4), we have, for every $d > \alpha$,

$$\alpha \leq \operatorname{fpcx}(\mathcal{T}) \leq \operatorname{cx}(\mathcal{T}) < d$$

which implies that $\operatorname{fpcx}(\mathcal{T}) = \operatorname{cx}(\mathcal{T}) = \alpha$. This construction implies that

$$\operatorname{GKdim}\left(\bigoplus_{i=0}^{\infty} \operatorname{Hom}_{\mathcal{T}}(\mathcal{O}, \Sigma^i(\mathcal{O}))\right) = \operatorname{GKdim} A = \alpha. \tag{E8.7.5}$$

(2) We now consider an extreme case. Let $a := \{a_i\}_{i=0}^{\infty}$ be any sequence of nonnegative integers with $a_0 = 1$. Define B to be the dg algebra $\bigoplus B_i$ such that:

(i) $\dim B_i = a_i$ for all i. In particular, $B_0 = \Bbbk$. Elements in B_i have cohomological degree i.

(ii) $\left(\bigoplus_{i>0} B_i\right)^2 = 0$.

(iii) Differential $d_B = 0$.

In this case, $\operatorname{GKdim} B = 0$. Similar to part (1), the derived category of left dg B-modules is denoted by $\mathcal{B}$. Let $\mathcal{O}$ be the object ${}_B B$ in $\mathcal{B}$. Then

$$\operatorname{Hom}_{\mathcal{B}}(\mathcal{O}, \Sigma^i \mathcal{O}) = B_i \quad \forall i,$$

and $\mathcal{O}$ is an atomic object. Let $\mathcal{T}$ be the full triangulated subcategory of $\mathcal{B}$ generated by $\mathcal{O}$. The argument in part (1) shows that

$$\operatorname{fpcx}(\mathcal{T}) = \limsup_{n\to\infty} \log_n(a_n) + 1.$$

Now let r be any real number ≥ 1 and let

$$a_i = \begin{cases} 1 & i = 0, \\ \lfloor i^{r-1} \rfloor & i \geq 1. \end{cases}$$

Then we have $\operatorname{fpcx}(\mathcal{T}) = r$. Let r be any real number ≥ 1 and $a_i = \lfloor r^i \rfloor$ for all $i \geq 0$. Then

$$\operatorname{fpcx}(\mathcal{T}) = \begin{cases} 1 & r = 1, \\ \infty & r > 1. \end{cases}$$

Using a similar method (with details omitted), $\operatorname{fpv}(\mathcal{T}) = r$.

Acknowledgments

The authors would like to thank Klaus Bongartz, Christof Geiss, Ken Goodearl, Claus Michael Ringel, and Birge Huisgen-Zimmermann for many useful conversations on the subject, thank Max Lieblich for the proof of Proposition 6.5(3), and thank the referees for their very careful reading and useful comments. J. Chen was partially supported by the National Natural Science Foundation of China (Grant No. 11571286). Z. Gao was partially supported by the National Natural Science Foundation of China (Grant No. 61401381). E. Wicks and J.J. Zhang were partially supported by the US National Science Foundation (Grant Nos. DMS-1402863 and DMS-1700825). X.-H. Zhang was partially supported by the National Natural Science Foundation of China (Grant No. 11401328). H. Zhu was partially supported by a grant from Jiangsu overseas Research and Training Program for university prominent young and middle-aged Teachers and Presidents, China.

References

[Alperin and Evens 1981] J. L. Alperin and L. Evens, "Representations, resolutions and Quillen's dimension theorem", *J. Pure Appl. Algebra* **22**:1 (1981), 1–9. MR Zbl

[Ariki 2005] S. Ariki, "Hecke algebras of classical type and their representation type", *Proc. London Math. Soc.* (3) **91**:2 (2005), 355–413. MR Zbl

[Artin and Zhang 1994] M. Artin and J. J. Zhang, "Noncommutative projective schemes", *Adv. Math.* **109**:2 (1994), 228–287. MR Zbl

[Assem et al. 2006] I. Assem, D. Simson, and A. Skowroński, *Elements of the representation theory of associative algebras, I: Techniques of representation theory*, London Math. Soc. Student Texts **65**, Cambridge Univ. Press, 2006. MR Zbl

[Auslander et al. 1995] M. Auslander, I. Reiten, and S. O. Smalø, *Representation theory of Artin algebras*, Cambridge Stud. Adv. Math. **36**, Cambridge Univ. Press, 1995. MR Zbl

[Avramov 1998] L. L. Avramov, "Infinite free resolutions", pp. 1–118 in *Six lectures on commutative algebra* (Bellaterra, Spain, 1996), edited by J. M. Giral et al., Progr. Math. **166**, Birkhäuser, Basel, 1998. MR Zbl

[Balmer 2005] P. Balmer, "The spectrum of prime ideals in tensor triangulated categories", *J. Reine Angew. Math.* **588** (2005), 149–168. MR Zbl

[Bao et al. 2019] Y. Bao, J. He, and J. Zhang, "Pertinency of Hopf actions and quotient categories of Cohen–Macaulay algebras", *J. Noncommut. Geom.* **13**:2 (2019), 667–710. MR

[Beilinson 1978] A. A. Beilinson, "Coherent sheaves on $\mathbb{P}^n$ and problems in linear algebra", *Funktsional. Anal. i Prilozhen.* **12**:3 (1978), 68–69. In Russian; translated in *Funct. Anal. Appl.* **12**:3 (1978), 214–216. MR

[Bell and Zhang 2017] J. Bell and J. J. Zhang, "An isomorphism lemma for graded rings", *Proc. Amer. Math. Soc.* **145**:3 (2017), 989–994. MR Zbl

[Benson 1991] D. J. Benson, *Representations and cohomology, I: Basic representation theory of finite groups and associative algebras*, Cambridge Stud. Adv. Math. **30**, Cambridge Univ. Press, 1991. MR Zbl

[Van den Bergh 2004] M. Van den Bergh, "Three-dimensional flops and noncommutative rings", *Duke Math. J.* **122**:3 (2004), 423–455. MR Zbl

[Bergh and Solberg 2010] P. A. Bergh and Ø. Solberg, "Relative support varieties", *Q. J. Math.* **61**:2 (2010), 171–182. MR Zbl

[Bernstein et al. 1973] I. N. Bernstein, I. M. Gelfand, and V. A. Ponomarev, "Coxeter functors and Gabriel's theorem", *Uspehi Mat. Nauk* **28**:2(170) (1973), 19–33. In Russian; translated in *Russ. Math. Surv.* **28**:2 (1973), 17–32. MR Zbl

[Bondal and Orlov 2001] A. Bondal and D. Orlov, "Reconstruction of a variety from the derived category and groups of autoequivalences", *Compos. Math.* **125**:3 (2001), 327–344. MR Zbl

[Bondal and Orlov 2002] A. Bondal and D. Orlov, "Derived categories of coherent sheaves", pp. 47–56 in *Proc. Int. Congr. Math, II* (Beijing, 2002), edited by T. Li, Higher Ed. Press, Beijing, 2002. MR Zbl

[Bongartz 2016] K. Bongartz, "Representation embeddings and the second Brauer–Thrall conjecture", preprint, 2016. arXiv 1611.02017

[Brenner 1986] S. Brenner, "A combinatorial characterisation of finite Auslander–Reiten quivers", pp. 13–49 in *Representation theory, I* (Ottawa, 1984), edited by V. Dlab et al., Lecture Notes in Math. **1177**, Springer, 1986. MR Zbl

[Bridgeland 2002] T. Bridgeland, "Flops and derived categories", *Invent. Math.* **147**:3 (2002), 613–632. MR Zbl

[Bridgeland et al. 2001] T. Bridgeland, A. King, and M. Reid, "The McKay correspondence as an equivalence of derived categories", *J. Amer. Math. Soc.* **14**:3 (2001), 535–554. MR Zbl

[Brüning and Burban 2007] K. Brüning and I. Burban, "Coherent sheaves on an elliptic curve", pp. 297–315 in *Interactions between homotopy theory and algebra* (Chicago, 2004), edited by L. L. Avramov et al., Contemp. Math. **436**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Carlson 1996] J. F. Carlson, "The decomposition of the trivial module in the complexity quotient category", *J. Pure Appl. Algebra* **106**:1 (1996), 23–44. MR Zbl

[Carlson et al. 1994] J. F. Carlson, P. W. Donovan, and W. W. Wheeler, "Complexity and quotient categories for group algebras", *J. Pure Appl. Algebra* **93**:2 (1994), 147–167. MR Zbl

[Carroll and Chindris 2015] A. T. Carroll and C. Chindris, "Moduli spaces of modules of Schur-tame algebras", *Algebr. Represent. Theory* **18**:4 (2015), 961–976. MR Zbl

[Chen 1997] J. A. Chen, "On genera of smooth curves in higher-dimensional varieties", *Proc. Amer. Math. Soc.* **125**:8 (1997), 2221–2225. MR Zbl

[Chen et al. 2019] J. M. Chen, Z. B. Gao, E. Wicks, J. J. Zhang, X.-H. Zhang, and H. Zhu, "Frobenius–Perron theory for projective schemes", preprint, 2019. arXiv 1907.02221

[Chindris et al. 2015] C. Chindris, R. Kinser, and J. Weyman, "Module varieties and representation type of finite-dimensional algebras", *Int. Math. Res. Not.* **2015**:3 (2015), 631–650. MR Zbl

[Ciliberto et al. 2016] C. Ciliberto, F. Flamini, and M. Zaidenberg, "Gaps for geometric genera", *Arch. Math.* (*Basel*) **106**:6 (2016), 531–541. MR Zbl

[Crawley-Boevey 1992] W. Crawley-Boevey, "Lectures on representations of quivers", lecture notes, Oxford University, 1992, Available at https://www.math.uni-bielefeld.de/∼wcrawley/quivlecs.pdf.

[Dokuchaev et al. 2013] M. A. Dokuchaev, N. M. Gubareni, V. M. Futorny, M. A. Khibina, and V. V. Kirichenko, "Dynkin diagrams and spectra of graphs", *São Paulo J. Math. Sci.* **7**:1 (2013), 83–104. MR Zbl

[Donovan and Freislich 1973] P. Donovan and M. R. Freislich, *The representation theory of finite graphs and associated algebras*, Carleton Math. Lecture Notes **5**, Carleton Univ., Ottawa, 1973. MR Zbl

[Drozd 1980] J. A. Drozd, "Tame and wild matrix problems", pp. 242–258 in *Representation theory, II* (Ottawa, 1979), edited by V. Dlab and P. Gabriel, Lecture Notes in Math. **832**, Springer, 1980. MR Zbl

[Drozd 2004] Y. A. Drozd, "Derived tame and derived wild algebras", *Algebra Discrete Math.* **1** (2004), 57–74. MR Zbl

[Drozd and Greuel 2001] Y. A. Drozd and G.-M. Greuel, "Tame and wild projective curves and classification of vector bundles", *J. Algebra* **246**:1 (2001), 1–54. MR Zbl

[Erdmann and Solberg 2011] K. Erdmann and Ø. Solberg, "Radical cube zero selfinjective algebras of finite complexity", *J. Pure Appl. Algebra* **215**:7 (2011), 1747–1768. MR Zbl

[Etingof et al. 2004] P. Etingof, S. Gelaki, and V. Ostrik, "Classification of fusion categories of dimension pq", *Int. Math. Res. Not.* **2004**:57 (2004), 3041–3056. MR Zbl

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR Zbl

[Etingof et al. 2015] P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik, *Tensor categories*, Math. Surv. Monogr. **205**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Farnsteiner 2007] R. Farnsteiner, "Tameness and complexity of finite group schemes", *Bull. London Math. Soc.* **39**:1 (2007), 63–70. MR Zbl

[Feldvoss and Witherspoon 2011] J. Feldvoss and S. Witherspoon, "Support varieties and representation type of self-injective algebras", *Homology Homotopy Appl.* **13**:2 (2011), 197–215. MR Zbl

[Gabriel 1972] P. Gabriel, "Unzerlegbare Darstellungen, I", *Manuscripta Math.* **6** (1972), 71–103. Correction in **6**:3 (1972), 309. MR Zbl

[Gabriel 1975] P. Gabriel, "Représentations indécomposables", exposé 444, pp. 143–169 in *Séminaire Bourbaki,* 1973/1974, Lecture Notes in Math. **431**, 1975. MR Zbl

[Gabriel 1980] P. Gabriel, "Auslander–Reiten sequences and representation-finite algebras", pp. 1–71 in *Representation theory, I* (Ottawa, 1979), edited by V. Dlab and P. Gabriel, Lecture Notes in Math. **831**, Springer, 1980. MR Zbl

[Geigle and Lenzing 1987] W. Geigle and H. Lenzing, "A class of weighted projective curves arising in representation theory of finite-dimensional algebras", pp. 265–297 in *Singularities, representation of algebras, and vector bundles* (Lambrecht, Germany, 1985), edited by G.-M. Greuel and G. Trautmann, Lecture Notes in Math. **1273**, Springer, 1987. MR Zbl

[Geiss and Krause 2002] C. Geiss and H. Krause, "On the notion of derived tameness", *J. Algebra Appl.* **1**:2 (2002), 133–157. MR Zbl

[Geiss et al. 2013] C. Geiss, B. Keller, and S. Oppermann, "n-angulated categories", *J. Reine Angew. Math.* **675** (2013), 101–120. MR Zbl

[Guo et al. 2009] J. Y. Guo, A. H. Li, and Q. X. Wu, "Selfinjective Koszul algebras of finite complexity", *Acta Math. Sin.* (*Engl. Ser.*) **25**:12 (2009), 2179–2198. MR Zbl

[Happel 1987] D. Happel, "On the derived category of a finite-dimensional algebra", *Comment. Math. Helv.* **62**:3 (1987), 339–389. MR Zbl

[Happel et al. 1980] D. Happel, U. Preiser, and C. M. Ringel, "Binary polyhedral groups and Euclidean diagrams", *Manuscripta Math.* **31**:1-3 (1980), 317–329. MR Zbl

[Huybrechts and Lehn 1997] D. Huybrechts and M. Lehn, *The geometry of moduli spaces of sheaves*, Aspects Math. **E31**, Vieweg & Sohn, Braunschweig, Germany, 1997. MR Zbl

[Keller 2008] B. Keller, "Calabi–Yau triangulated categories", pp. 467–489 in *Trends in representation theory of algebras and related topics*, edited by A. Skowroński, Eur. Math. Soc., Zürich, 2008. MR Zbl

[Keller and Vossieck 1987] B. Keller and D. Vossieck, "Sous les catégories dérivées", *C. R. Acad. Sci. Paris Sér. I Math.* **305**:6 (1987), 225–228. MR Zbl

[Krause and Lenagan 1985] G. R. Krause and T. H. Lenagan, *Growth of algebras and Gelfand–Kirillov dimension*, Res. Notes Math. **116**, Pitman, Boston, 1985. MR Zbl

[Külshammer 2013] J. Külshammer, "Representation type of Frobenius–Lusztig kernels", *Q. J. Math.* **64**:2 (2013), 471–488. Correction in **66**:4 (2015), 1139. MR Zbl

[Lenzing and Meltzer 1993] H. Lenzing and H. Meltzer, "Sheaves on a weighted projective line of genus one, and representations of a tubular algebra", pp. 313–337 in *Representations of algebras* (Ottawa, 1992), edited by V. Dlab and H. Lenzing, CMS Conf. Proc. **14**, Amer. Math. Soc., Providence, RI, 1993. MR Zbl

[McConnell and Robson 1987] J. C. McConnell and J. C. Robson, *Noncommutative Noetherian rings*, Grad. Stud. in Math. **30**, Wiley, Chichester, England, 1987. MR Zbl

[Nazarova 1973] L. A. Nazarova, "Representations of quivers of infinite type", *Izv. Akad. Nauk SSSR Ser. Mat.* **37** (1973), 752–791. In Russian; translated in *Math. USSR-Izv.* **7**:4 (1973), 749–792. MR Zbl

[Neeman 2001] A. Neeman, *Triangulated categories*, Ann. Math. Stud. **148**, Princeton Univ. Press, 2001. MR Zbl

[Nikshych 2004] D. Nikshych, "Semisimple weak Hopf algebras", *J. Algebra* **275**:2 (2004), 639–667. MR Zbl

[Nikshych and Vainerman 2002] D. Nikshych and L. Vainerman, "Finite quantum groupoids and their applications", pp. 211–262 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Math. Sci. Res. Inst. Publ. **43**, Cambridge Univ. Press, 2002. MR Zbl

[Reiten and Van den Bergh 2002] I. Reiten and M. Van den Bergh, "Noetherian hereditary abelian categories satisfying Serre duality", *J. Amer. Math. Soc.* **15**:2 (2002), 295–366. MR Zbl

[Rickard 1990] J. Rickard, "The representation type of self-injective algebras", *Bull. London Math. Soc.* **22**:6 (1990), 540–546. MR Zbl

[Ringel and Vossieck 1987] C. M. Ringel and D. Vossieck, "Hammocks", *Proc. London Math. Soc.* (3) **54**:2 (1987), 216–246. MR Zbl

[van Roosmalen 2008] A.-C. van Roosmalen, "Abelian 1-Calabi–Yau categories", *Int. Math. Res. Not.* **2008**:6 (2008), art. id. rnn003. MR Zbl

[van Roosmalen 2016] A.-C. van Roosmalen, "Numerically finite hereditary categories with Serre duality", *Trans. Amer. Math. Soc.* **368**:10 (2016), 7189–7238. MR Zbl

[Schiffler 2014] R. Schiffler, *Quiver representations*, Springer, 2014. MR Zbl

[Schiffmann 2012] O. Schiffmann, "Lectures on Hall algebras", pp. 1–141 in *Geometric methods in representation theory, II* (Grenoble, France, 2008), edited by M. Brion, Sémin. Congr. **24**, Soc. Math. France, Paris, 2012. MR Zbl

[Smith 1970] J. H. Smith, "Some properties of the spectrum of a graph", pp. 403–406 in *Combinatorial structures and their spplications* (Calgary, 1969), edited by R. Guy et al., Gordon and Breach, New York, 1970. MR Zbl

[Wicks 2019] E. Wicks, "Frobenius–Perron theory of modified ADE bound quiver algebras", *J. Pure Appl. Algebra* **223**:6 (2019), 2673–2708. MR Zbl

[Yekutieli 1999] A. Yekutieli, "Dualizing complexes, Morita equivalence and the derived Picard group of a ring", *J. London Math. Soc.* (2) **60**:3 (1999), 723–746. MR Zbl

[Zhang and Zhou $\geq$ 2019] J. J. Zhang and J.-H. Zhou, "Frobenius–Perron theory of representations of quivers", in preparation.

Communicated by J. Toby Stafford
Received 2018-04-04 Revised 2019-04-04 Accepted 2019-06-25

chenjianmin@xmu.edu.cn *School of Mathematical Sciences, Xiamen University, China*

gaozhibin@xmu.edu.cn *Department of Communication Engineering, Xiamen University, China*

lizwicks@uw.edu *Department of Mathematics, University of Washington, Seattle, WA, United States*

zhang@math.washington.edu *Department of Mathematics, University of Washington, Seattle, WA, United States*

zhang-xiaohong@t.shu.edu.cn *College of Sciences, Ningbo University of Technology, China*

zhuhongazhu@aliyun.com *Department of Information Sciences, the School of Mathematics and Physics, Changzhou University, China*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2057

Positivity of anticanonical divisors and F-purity of fibers

Sho Ejiri

We prove that given a flat generically smooth morphism between smooth projective varieties with F-pure closed fibers, if the source space is Fano, weak Fano or a variety with nef anticanonical divisor, respectively, then so is the target space. We also show that, in arbitrary characteristic, a generically smooth surjective morphism between smooth projective varieties cannot have nef and big relative anticanonical divisor, if the target space has positive dimension.

1. Introduction

Let X be a smooth projective variety over an algebraically closed field. The positivity of the anticanonical divisor $-K_X$ on X is an important notion that helps us know certain geometric properties of X. Let $f: X \to Y$ be a surjective morphism from X to another smooth projective variety Y. Kollár, Miyaoka and Mori [Kollár et al. 1992, Corollary 2.9] proved that, under the assumption that f is smooth, if X is a Fano variety, that is $-K_X$ is ample, then so is Y. It follows from an analogous argument that, under the same assumption, if $-K_X$ is nef, then so is $-K_Y$ [Miyaoka 1993; Fujino and Gongyo 2014, Theorem 1.1; Debarre 2001, Corollary 3.15(a)]. Based on these results, Yasutake asked "what positivity condition is passed from $-K_X$ to $-K_Y$?" Some answers to this question are known in characteristic 0. Fujino and Gongyo [2012, Theorem 1.1] proved that, under the assumption that f is smooth, if X is a weak Fano variety, that is $-K_X$ is nef and big, then so is Y. Birkar and Chen [2016, Theorem 1.1] showed that, under the same assumption, if $-K_X$ is semiample, then so is $-K_Y$. Furthermore, similar but weaker results hold even if f is not smooth (but the characteristic of k is still 0). For example, a result of Prokhorov and Shokurov [2009, Lemma 2.8] (see also [Fujino and Gongyo 2012, Corollary 3.3]) implies that if $-K_X$ is nef and big, then $-K_Y$ is big. Chen and Zhang [2013, Main Theorem] also proved that if $-K_X$ is nef, then $-K_Y$ is pseudoeffective.

In contrast, little was known about the positive characteristic case. In this paper, assuming that the geometric generic fiber has only F-pure or strongly F-regular singularities, we prove that (generalizations of) the statements above hold in positive characteristic, except the one concerning semiampleness. F-purity and strong F-regularity are mild singularities defined in terms of Frobenius splitting properties

MSC2010: primary 14D06; secondary 14J45.
Keywords: Fano variety, weak Fano variety, anticanonical divisor, restricted base locus, augmented base locus.

(Definition 2.4), which have a close connection to log canonical and Kawamata log terminal singularities, respectively.

Suppose that the base field k is an algebraically closed field of characteristic $p > 0$. Let $f : X \to Y$ be a surjective morphism between smooth projective varieties, Δ an effective $\mathbb{Q}$-divisor on X with index $\operatorname{ind}\Delta$, and D a $\mathbb{Q}$-divisor on Y. Let $X_{\bar{\eta}}$ denote the geometric generic fiber of f. Then our main theorem is stated as follows:

Theorem 1.1 (Theorem 4.1). *Let y be a scheme-theoretic point in Y such that the following conditions hold*:

(i) $\dim f^{-1}(y) = \dim X - \dim Y$.

(ii) *The support of Δ does not contain any irreducible component of $f^{-1}(y)$.*

(iii) *$(X_{\bar{y}}, \Delta|_{X_{\bar{y}}})$ is F-pure, where $X_{\bar{y}}$ is the geometric fiber over y.*

Suppose that $p \nmid \operatorname{ind}(\Delta)$ and $-(K_X+\Delta+f^*D)$ is nef. Then y is not in the Zariski closure of $\boldsymbol{B}_-(-(K_Y+D))$.

Here, $\boldsymbol{B}_-$ denotes the restricted base locus (Definition 2.8). This locus is empty (resp. has nonempty complement) if and only if the divisor is nef (resp. pseudoeffective). Theorem 1.1 implies, in the case of $\Delta = 0$, that every closed fiber over $\boldsymbol{B}_-(-(K_Y + D))$ is "bad", where "bad" means the fiber is not F-pure or has dimension larger than that of the general fiber.

The following two theorems are corollaries of Theorem 1.1.

Theorem 1.2 (Corollary 4.5). *Suppose that conditions* (i)–(iii) *in Theorem 1.1 hold for every closed point in Y*:

(1) *Assume $p \nmid \operatorname{ind}(\Delta)$. If $-(K_X + \Delta + f^*D)$ is nef, then so is $-(K_Y + D)$.*

(2) *If $-(K_X + \Delta + f^*D)$ is ample, then so is $-(K_Y + D)$.*

Theorem 1.3 (Corollary 4.6). *Suppose that $(X_{\bar{\eta}}, \Delta|_{X_{\bar{\eta}}})$ is F-pure*:

(1) *If $p \nmid \operatorname{ind}(\Delta)$ and $-(K_X + \Delta + f^*D)$ is nef, then $-(K_Y + D)$ is pseudoeffective.*

(2) *If $-(K_X + \Delta + f^*D)$ is ample, then $-(K_Y + D)$ is big.*

(3) *If $(X_{\bar{\eta}}, \Delta|_{X_{\bar{\eta}}})$ is strongly F-regular and $-(K_X + \Delta + f^*D)$ is nef and big, then $-(K_Y + D)$ is big.*

Theorem 1.2 is a generalization of [Kollár et al. 1992, Corollary 2.9] and [Debarre 2001, Corollary 3.15] in positive characteristic. One can also recover [Kollár et al. 1992, Corollary 2.9] in characteristic zero from Theorem 1.2, using reduction to characteristic p. Our proof relies on a study of the positivity of direct image sheaves in terms of the Grothendieck trace of the relative Frobenius morphism. This is completely different from the proof [Kollár et al. 1992, Corollary 2.9] that is an application of their great study regarding rational curves on varieties. Theorem 1.3 should be compared with [Prokhorov and Shokurov 2009, Lemma 2.8] and [Chen and Zhang 2013, Main Theorem].

The following two theorems are direct consequences of Theorems 1.2 and 1.3.

Theorem 1.4 (Corollary 4.7). *Suppose that $(X_{\bar{\eta}}, \Delta|_{X_{\bar{\eta}}})$ is F-pure. If $p \nmid \mathrm{ind}(\Delta)$ and $K_X + \Delta$ is numerically equivalent to $f^*(K_Y + L)$ for some $\mathbb{Q}$-divisor L on Y, then L is pseudoeffective.*

Theorem 1.5 (Corollary 4.9). *Suppose that f is a flat morphism such that every closed fiber is F-pure and the geometric generic fiber is strongly F-regular. If X is a weak Fano variety, that is, $-K_X$ is nef and big, then so is Y.*

Theorem 1.5 is a positive characteristic counterpart of [Fujino and Gongyo 2012, Theorem 1.1].

For another application of Theorem 1.1, we return to the situation where k is of arbitrary characteristic. Suppose that $f : X \to Y$ is a generically smooth surjective morphism between smooth projective varieties of positive dimension.

Theorem 1.6 (Corollary 4.10 and Theorem 5.4). *The relative anticanonical divisor $-K_{X/Y}$ cannot be both nef and big.*

Theorem 1.7 (Corollary 4.11 and Theorem 5.5). *Suppose that $\omega_{X_{\bar{\eta}}}^{-m}$ is globally generated for an integer $m > 0$. Then $f_*\omega_{X/Y}^{-m}$ is not big in the sense of Definition 2.6.*

In both the theorems, the characteristic zero case is proved by reduction to positive characteristic. Theorem 1.6 includes [Kollár et al. 1992, Corollary 2.8] which states that $-K_{X/Y}$ is not ample. Theorem 1.7 generalizes [Miyaoka 1993, Corollary 2'] which says that if $\omega_{X/Y}^{-1}$ is f-ample and $\omega_{X/Y}^{-m}$ is f-free for some $m \in \mathbb{Z}_{>0}$, then $f_*\omega_{X/Y}^{-m}$ is not an ample vector bundle.

Notation. Let k be a field. A *k-scheme* is a separated scheme of finite type over k. A *variety* is an integral k-scheme. Let $\varphi : S \to T$ be a morphism of k-schemes and T' a T-scheme. Then, $S_{T'}$ and $\varphi_{T'} : S_{T'} \to T'$ denote the fiber product $S \times_T T'$ and its second projection, respectively. For a Cartier or $\mathbb{Q}$-Cartier divisor D on S (resp. an $\mathcal{O}_S$-module $\mathcal{G}$), the pullback of D (resp. $\mathcal{G}$) to $S_{T'}$ is written as $D_{T'}$ (resp. $\mathcal{G}_{T'}$) if it is well defined. Similarly, for a homomorphism $\alpha : \mathcal{F} \to \mathcal{G}$ of $\mathcal{O}_S$-modules, $\alpha_{T'} : \mathcal{F}_{T'} \to \mathcal{G}_{T'}$ is the pullback of α to $S_{T'}$. Assume that k is of characteristic $p > 0$. We say that k is *F-finite* if the field extension k/k^p is finite. Let X be a k-scheme. Then, $F_X : X \to X$ denotes the absolute Frobenius morphism of X. We often write the source of F_X^e as X^e. Let $f : X \to Y$ be a morphism between k-schemes. The same morphism is denoted by $f^{(e)} : X^e \to Y^e$ when we regard X (resp. Y) as X^e (resp. Y^e). We define the e-th relative Frobenius morphism of f to be the morphism $F_{X/Y}^{(e)} := (F_X^e, f^{(e)}) : X^e \to X \times_Y Y^e =: X_{Y^e}$. We write the localization of $\mathbb{Z}$ at $(p) = p\mathbb{Z}$ as $\mathbb{Z}_{(p)}$.

2. Preliminaries

2A. ***Relative Frobenius morphisms and trace maps.*** In this subsection, given a morphism between varieties, we consider the relative Frobenius morphism and its trace map. Let k be an F-finite field and $f : X \to Y$ a morphism from a pure dimensional Gorenstein k-scheme X to a regular variety Y. For each

$d, e \in \mathbb{Z}_{>0}$ we use the notation defined by the following commutative diagram:

$$
\begin{tikzcd}
X^{de} \arrow[d] \arrow[dd, bend left, "F^{(d(e-i))}_{X^{di}/Y^{di}}"] \arrow[dddd, bend right, "F^{(de)}_{X/Y}"'] \arrow[ddddd, bend right=60, "f^{(de)}"'] \\
\vdots \arrow[d] \\
X^{di}_{Y^{de}} \arrow[d] \arrow[dd, bend left, "F^{(di)}_{X_{Y^{d(e-i)}}/Y^{d(e-i)}}"] \arrow[ddd, bend left=60, "f^{(di)}_{Y^{de}}"] \\
\vdots \arrow[d] \\
X_{Y^{de}} \arrow[d] \\
Y^{de}
\end{tikzcd}
\qquad
\begin{tikzcd}
\ddots \arrow[dr, "F_X^d"] & & & \\
\cdots \arrow[r] & X^{2d} \arrow[d, "F^{(d)}_{X^d/Y^d}"'] \arrow[dr, "F_X^d"] & & \\
\cdots \arrow[r] & X^d_{Y^{2d}} \arrow[r] \arrow[d, "F^{(d)}_{X_{Y^d}/Y^d}"'] & X^d \arrow[d, "F^{(d)}_{X/Y}"'] \arrow[dr, "F_X^d"] & \\
\cdots \arrow[r] & X_{Y^{2d}} \arrow[r] \arrow[d, "f_{Y^{2d}}"'] & X_{Y^d} \arrow[r, "(F_Y^d)_X"'] \arrow[d, "f_{Y^d}"'] & X \arrow[d, "f"'] \\
\cdots \arrow[r, "F_Y^d"'] & Y^{2d} \arrow[r, "F_Y^d"'] & Y^d \arrow[r, "F_Y^d"'] & Y
\end{tikzcd}
$$

Since F_Y is flat, every horizontal morphism in the diagram is a Gorenstein morphism, so every object is a pure dimensional Gorenstein k-scheme. Let ω_X be the dualizing sheaf on X. Let $\mathrm{Tr}_{F^{(1)}_{X/Y}} : F^{(1)}_{X/Y\,*}\omega_{X^1} \to \omega_{X_{Y^1}}$ denote the morphism obtained by applying the functor $\mathscr{H}om_{\mathcal{O}_{X_{Y^1}}}((\,), \omega_{X_{Y^1}})$ to the natural morphism $F^{(1)\,\#}_{X/Y} : \mathcal{O}_{X_{Y^1}} \to F^{(1)}_{X/Y\,*}\mathcal{O}_{X^1}$. Take a Cartier divisor K_X satisfying $\mathcal{O}_X(K_X) \cong \omega_X$. Set $K_{X/Y} := K_X - f^*K_Y$. For each $e \in \mathbb{Z}_{>0}$ we define

$$
\begin{aligned}
\phi^{(1)}_{X/Y} &:= \mathrm{Tr}_{F^{(1)}_{X/Y}} \otimes \mathcal{O}_{X_{Y^1}}(-K_{X_{Y^1}}) : F^{(1)}_{X/Y\,*}\mathcal{O}_{X^1}((1-p)K_{X^1/Y^1}) \to \mathcal{O}_{X_{Y^1}}, \quad \text{and} \\
\phi^{(e+1)}_{X/Y} &:= (\phi^{(e)}_{X/Y})_{Y^{e+1}} \circ F^{(e)}_{X_{Y^1}/Y^1\,*}(\phi^{(1)}_{X^e/Y^e} \otimes \mathcal{O}_{X^e_{Y^{e+1}}}((1-p^e)K_{X^e_{Y^{e+1}}/Y^{e+1}})) \\
&\qquad\qquad : F^{(e+1)}_{X/Y\,*}\mathcal{O}_X((1-p^{e+1})K_{X^{e+1}/Y^{e+1}}) \to \mathcal{O}_{X_{Y^{e+1}}}.
\end{aligned}
$$

Let E be an effective Cartier divisor on X, let $a > 0$ be an integer not divisible by p, and let $d > 0$ be the minimum integer satisfying $a|(p^d - 1)$. Note that an integer $e > 0$ satisfies $a|(p^e - 1)$ if and only if $d|e$. Set $\Delta := E \otimes \frac{1}{a} \in \mathrm{Car}(X) \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)}$. For each $e \in d\mathbb{Z}_{>0}$ we define

$$
\begin{aligned}
&\mathcal{L}^{(e)}_{(X/Y,\Delta)} := \mathcal{O}_{X^e}((1-p^e)(K_{X^e/Y^e} + \Delta^e)) \subseteq \mathcal{O}_{X^e}((1-p^e)K_{X^e/Y^e}), \\
&\phi^{(d)}_{(X/Y,\Delta)} : F^{(d)}_{X/Y\,*}\mathcal{L}^{(d)}_{(X/Y,\Delta)} \to F^{(d)}_{X/Y\,*}\mathcal{O}_{X^d}((1-p^d)K_{X^d/Y^d}) \xrightarrow{\phi^{(d)}_{X/Y}} \mathcal{O}_{X_{Y^d}}, \quad \text{and} \\
&\phi^{(e+d)}_{(X/Y,\Delta)} := (\phi^{(e)}_{(X/Y,\Delta)})_{Y^e} \circ F^{(e)}_{X_{Y^d}/Y^d\,*}(\phi^{(d)}_{(X^e/Y^e,\Delta^e)} \otimes (\mathcal{L}^{(e)}_{(X/Y,\Delta)})_{Y^{e+d}}) : F^{(e+d)}_{X/Y\,*}\mathcal{L}^{(e+d)}_{(X/Y,\Delta)} \to \mathcal{O}_{X_{Y^{e+d}}}.
\end{aligned}
$$

In order to generalize the definitions above, we recall the notion of generalized divisors on a k-scheme. Let X be a k-scheme of pure dimension satisfying S_2 and G_1. An *AC divisor* (or *almost Cartier divisor*) on

X is a reflexive coherent subsheaf of the sheaf of total quotient rings on X that is invertible in codimension one (see [Hartshorne 1994; Miller and Schwede 2012]). For an AC divisor D, the coherent sheaf defining D is denoted by $\mathcal{O}_X(D)$. The set of AC divisors $\mathrm{WSh}(X)$ has a structure of additive group [Hartshorne 1994, Corollary 2.6]. A $\mathbb{Z}_{(p)}$-*AC divisor* (resp. $\mathbb{Q}$-*AC divisor*) is an element of $\mathrm{WSh}(X) \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)}$ (resp. $\mathrm{WSh}(X) \otimes_{\mathbb{Z}} \mathbb{Q}$). An AC divisor D is said to be *effective* if $\mathcal{O}_X \subseteq \mathcal{O}_X(D)$, and a $\mathbb{Z}_{(p)}$-AC (resp. $\mathbb{Q}$-AC) divisor Δ is said to be *effective* if $\Delta = D \otimes r$ for an effective AC divisor D and an $r \geq 0$. For two AC divisors D and E, the notation $D \leq E$ means that $E - D$ is effective. We use the same notation for $\mathbb{Z}_{(p)}$-AC (resp. $\mathbb{Q}$-AC) divisors.

Remark 2.1. Each of the natural morphisms

$$\mathrm{WSh}(X) \xrightarrow{(\ )\otimes 1} \mathrm{WSh}(X) \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)} \to \mathrm{WSh}(X) \otimes_{\mathbb{Z}} \mathbb{Q}$$

is not necessarily injective (see Example 2.2). Let D and E be AC divisors. Then, D and E are equal as $\mathbb{Z}_{(p)}$-AC (resp. $\mathbb{Q}$-AC) divisors if and only if $mD = mE$ for some nonzero $m \in \mathbb{Z} \setminus p\mathbb{Z}$ (resp. $m \in \mathbb{Z}$). Furthermore, D and E can be equal as $\mathbb{Z}_{(p)}$-AC (resp. $\mathbb{Q}$-AC) divisors even when D is effective but E is not (see Example 2.2).

Example 2.2 [Corti 1992, (16.1.2)]. Set $X := \operatorname{Spec} k[x, y, z, z^{-1}]/(x^n - zy^n)$ for an $n \geq 2$. Note that X is integral and Gorenstein but not normal. Let D and E be AC divisors on X defined by $x^{-1}\mathcal{O}_X$ and $y^{-1}\mathcal{O}_X$, respectively. For an $m \geq 1$, one has

$$mD = mE \iff x^{-m}\mathcal{O}_X = y^{-m}\mathcal{O}_X \iff n \mid m.$$

Hence, we see that

- $D \neq E$ as AC divisors,
- $D \otimes 1 = E \otimes 1$ as $\mathbb{Z}_{(p)}$-AC divisors if and only if $p \nmid n$, and
- $D \otimes 1 = E \otimes 1$ as $\mathbb{Q}$-AC divisors.

Furthermore, $D - E$ is not effective but $n(D - E) = 0$ is effective.

Remark 2.3. Let E and K be two AC divisors, take an $\varepsilon \in \mathbb{Z}_{(p)}$ (resp. $\varepsilon \in \mathbb{Q}$) and set $\Delta := E \otimes \varepsilon$. When we consider the $\mathbb{Z}_{(p)}$-AC (resp. $\mathbb{Q}$-AC) divisor $K + \Delta$, for each $m \in \mathbb{Z}$ with $\varepsilon m \in \mathbb{Z}$, we let $m(K + \Delta)$ denote the AC divisor $mK + (\varepsilon m)E$.

Let $f : X \to Y$ be a morphism from a pure dimensional k-scheme X to a regular variety Y and assume that X satisfies S_2 and G_1. Let E be an effective AC divisor on X, and fix a Gorenstein open subset $U \subseteq X$ such that $\operatorname{codim}_X(X \setminus U) \geq 2$ and $E|_U$ is Cartier. Let $U \xhookrightarrow{\iota} X$ denote the open immersion. Then, for each $e \in \mathbb{Z}_{>0}$, we have the following commutative diagram:

$$\begin{tikzcd}
U^e \arrow[r, "F^{(e)}_{U/Y}"] \arrow[d, hook, "\iota^{(e)}"'] & U_{Y^e} \arrow[r] \arrow[d, hook, "\iota_{Y^e}"] & U \arrow[d, hook, "\iota"] \\
X^e \arrow[r, "F^{(e)}_{X/Y}"] \arrow[dr, "f^{(e)}"'] & X_{Y^e} \arrow[r] \arrow[d, "f_{Y^e}"] & X \arrow[d, "f"] \\
& Y^e \arrow[r, "F^e_Y"'] & Y
\end{tikzcd}$$

Take an integer $a > 0$ not divisible by p, set $\Delta := E \otimes \frac{1}{a}$, and let d be the minimum positive integer satisfying $a \mid (p^d - 1)$. For each $e \in d\mathbb{Z}_{>0}$, we define

$$\mathcal{L}^{(e)}_{(X/Y,\Delta)} := \iota^{(e)}{}_*\mathcal{L}^{(e)}_{(U/Y,\Delta|_U)} \quad \text{and} \quad \phi^{(e)}_{(X/Y,\Delta)} := \iota_{Y^e *}(\phi^{(e)}_{(U/Y,\Delta|_U)}) : F^{(e)}_{X/Y}{}_*\mathcal{L}^{(e)}_{(X/Y,\Delta)} \to \mathcal{O}_{X_{Y^e}}.$$

Fix $e \in d\mathbb{Z}_{>0}$. Set $\mathcal{L}^{(e)}_{(X,\Delta)} := \mathcal{O}_{X^e}((1-p^e)(K_{X^e} + \Delta^e))$. We can define the morphism $\phi^{(e)}_{(X,\Delta)} : F^e_{X*}\mathcal{L}^{(e)}_{(X,\Delta)} \to \mathcal{O}_X$ by a procedure similar to the one above, replacing $F^{(e)}_{X/Y}$ by F^e_X. In the case where k is perfect and $Y = \operatorname{Spec} k$, one may identify, respectively, F^e_X, $\mathcal{L}^{(e)}_{(X,\Delta)}$ and $\phi^{(e)}_{(X,\Delta)}$ with $F^{(e)}_{X/Y}$, $\mathcal{L}^{(e)}_{(X/Y,\Delta)}$ and $\phi^{(e)}_{(X/Y,\Delta)}$.

We next introduce singularities of pairs defined by the Grothendieck trace map of the Frobenius morphism.

Definition 2.4. Let X be a k-scheme of pure dimension satisfying S_2 and G_1, and let Δ be an effective $\mathbb{Q}$-AC divisor on X:

(i) We say that (X, Δ) is *F-pure* if for each $e \in \mathbb{Z}_{>0}$ and every effective AC divisor E' with $\Delta' := E' \otimes 1/(p^e - 1) \leq \Delta$, the morphism

$$\phi^{(e)}_{(X,\Delta')} : F^e_{X*}\mathcal{O}_X((1-p^e)(K_X + \Delta')) \to \mathcal{O}_X$$

is surjective.

(ii) [Schwede 2008, Definition 3.1] Assume that X is a normal variety. We say that (X, Δ) is *strongly F-regular* if every effective Cartier divisor D on X, the morphism

$$F^e_{X*}\mathcal{O}_X(\lfloor (1-p^e)(K_X+\Delta) - D \rfloor) = F^e_{X*}\mathcal{O}_X((1-p^e)(K_X+\Delta')) \xrightarrow{\phi^{(e)}_{(X,\Delta')}} \mathcal{O}_X$$

is surjective for some $e \in \mathbb{Z}_{>0}$, where $\Delta' := \frac{1}{p^e-1}\lfloor (p^e-1)\Delta + D \rfloor$. Here, $\lfloor\ \rfloor$ denotes the round down.

We simply say that X is *F-pure* (resp. *strongly F-regular*) if $(X, 0)$ is F-pure (resp. strongly F-regular).

Remarks 2.5. (1) With the notation as above, assume that X is normal and affine. Then the above definition of F-purity is equivalent to that in [Hara and Watanabe 2002]. This follows from the fact that $\lfloor (p^e-1)\Delta \rfloor$ is the greatest element of the set S of all divisors E' such that $E' \leq (p^e-1)\Delta$.

(2) When X is not normal, S in (1) does not necessarily have a greatest element. Let X, D and E be as in Example 2.2 with $p \nmid n$ and $n \nmid l$, where $l := p^e - 1$. Then, $\Delta := D \otimes 1 = E \otimes 1$ as $\mathbb{Z}_{(p)}$-divisors and $lD \neq lE$. If S has a greatest element G, then $G \geq lD$ and $G \otimes 1 = l\Delta$, from which one can get $G = lD$. In the same way, we get $G = lE$, so $lD = lE$, a contradiction.

(3) Let (X, Δ) be a strongly F-regular pair, and Δ' an effective $\mathbb{Q}$-divisor on X. Then there is $\varepsilon \in \mathbb{Q}_{>0}$ such that $(X, \Delta + \varepsilon\Delta')$ is again strongly F-regular.

2B. *Weak positivity.* In this subsection, we recall the notion of weak positivity introduced by Viehweg [1983]. The definition employed in this paper is slightly different from Viehweg's original one. We work over a field k in this subsection.

Definition 2.6. Let Y be a quasiprojective normal variety, let $\mathcal{G}$ and $\mathcal{G}'$ be coherent sheaves on Y, and let $\mathcal{H}$ be an ample line bundle on Y. Take a subset S of the underlying topological space of Y such that the stalk of $\mathcal{G}$ at any point in S is free, i.e., there is an open subset $Y_0 \subset Y$ such that $S \subseteq Y_0$ and $\mathcal{G}|_{Y_0}$ is locally free:

(i) We say that a morphism $\mathcal{G} \to \mathcal{G}'$ is *surjective over* S if S and the support of the cokernel do not intersect.

(ii) We say that $\mathcal{G}$ is *globally generated over* S if the natural morphism $H^0(Y, \mathcal{G}) \otimes_k \mathcal{O}_Y \to \mathcal{G}$ is surjective over S.

(iii) We say that $\mathcal{G}$ is *weakly positive over* S if for every $\alpha \in \mathbb{Z}_{>0}$, there is $\beta \in \mathbb{Z}_{>0}$ such that $(S^{\alpha\beta}\mathcal{G})^{**} \otimes \mathcal{H}^{\beta}$ is globally generated over S. Here, $S^{\alpha\beta}(\ )$ and $(\ )^{**}$ denote the $\alpha\beta$-th symmetric product and the double dual, respectively.

(iv) We say that $\mathcal{G}$ is *big over* S if there is $\gamma \in \mathbb{Z}_{>0}$ such that $(S^{\gamma}\mathcal{G}) \otimes \mathcal{H}^{-1}$ is weakly positive over S.

We simply say that $\mathcal{G}$ is *generically globally generated* if $\mathcal{G}$ is globally generated over $\{\eta\}$, where η is the generic point of Y. Furthermore, we simply say that $\mathcal{G}$ is *weakly positive* (resp. *big*) if it is weakly positive (resp. big) over $\{\eta\}$.

Remark 2.7. Let Y, $\mathcal{G}$, S and $\mathcal{H}$ be as above:

(1) The above definitions are independent of the choice of $\mathcal{H}$ [Viehweg 1995, Lemma 2.14].

(2) Suppose that $\mathcal{G}$ is a vector bundle over a smooth projective curve Y. Then $\mathcal{G}$ is weakly positive (resp. big) over Y if and only if $\mathcal{G}$ is nef (resp. ample).

2C. *Augmented and restricted base loci.* In this subsection, we recall the definition of the augmented and restricted base locus of a $\mathbb{Q}$-Cartier divisor. In this subsection, we work over a field k.

Definition 2.8 [Ein et al. 2006; Mustaţă 2013]. Let Y be a quasiprojective variety and D a $\mathbb{Q}$-Cartier divisor on Y:

(i) The *stable base locus* $\boldsymbol{B}(D)$ of D is defined as the reduced base locus of mD for sufficiently large and divisible integer m.

(ii) The *augmented base locus* is given by

$$\boldsymbol{B}_{+}(D) := \bigcap_{A} \boldsymbol{B}(D-A),$$

where A runs over all the ample $\mathbb{Q}$-Cartier divisors on Y.

(iii) The *restricted base locus* (also called the *nonnef locus* or the *diminished base locus*) is defined by

$$\boldsymbol{B}_{-}(D) := \bigcup_{A} \boldsymbol{B}(D+A),$$

where A runs over all the ample $\mathbb{Q}$-Cartier divisors on Y.

Remarks 2.9. (1) In [Ein et al. 2006], the variety Y is assumed to be projective.

(2) Assume that Y is projective. Then the following hold:

- $\boldsymbol{B}_{+}(D) = \varnothing$ if and only if D is ample [loc. cit., Example 1.7].
- $\boldsymbol{B}_{+}(D) \neq Y$ if and only if D is big [loc. cit., Example 1.7].
- $\boldsymbol{B}_{-}(D) = \varnothing$ if and only if D is nef [loc. cit., Example 1.18].
- When k is uncountable, $\boldsymbol{B}_{-}(D) \neq Y$ if and only if D is pseudoeffective [Mustaţă 2013, Section 2].

(3) Assume that Y is a normal projective variety and D is Cartier. Let S be a subset of the underlying topological space of Y. Then, the weak positivity (resp. bigness) of $\mathcal{O}_Y(D)$ is equivalent to saying that $S \cap \boldsymbol{B}_{-}(D) = \varnothing$ (resp. $S \cap \boldsymbol{B}_{+}(D) = \varnothing$).

The next lemma can be proved in the same way as in the proof of [Ein et al. 2006, Proposition 1.19].

Lemma 2.10 [Ein et al. 2006, Propositions 1.19]. *Let the notation be as in Definition 2.8. Let H be an ample $\mathbb{Q}$-Cartier divisor on Y and $\{a_m\}$ a sequence of positive rational numbers that converges to zero. Then $\boldsymbol{B}_{-}(D) = \bigcup_m \boldsymbol{B}(D + a_m H)$.*

3. Auxiliary lemmas

In this section, we prove several lemmas used in the proofs of the main theorems. Throughout this section, the base field k is assumed to be an F-finite field of characteristic $p > 0$.

Lemma 3.1. *Let W be a normal quasiprojective variety and $W_0 \subseteq W$ an open subset. Let $\mathcal{H}$ be an ample line bundle on W and $\mathcal{G}$ a coherent sheaf on W such that $\mathcal{G}|_{W_0}$ is locally free. Then there exists $m_0 \in \mathbb{Z}_{>0}$ such that, for every $m \geq m_0$, there exists a homomorphism $\theta : \bigoplus^n \mathcal{G} \to \mathcal{H}^m$ of $\mathcal{O}_W$-modules which is surjective over W_0.*

Proof. Take $m \in \mathbb{Z}_{>0}$ such that $\mathcal{G}^* \otimes \mathcal{H}^m$ is generated by its global sections. Since W_0 is Noetherian, we get a surjective morphism $\theta' : \bigoplus^n \mathcal{O}_W \twoheadrightarrow \mathcal{G}^* \otimes \mathcal{H}^m$ for some $n \in \mathbb{Z}_{>0}$. We then obtain

$$\bigoplus^n \mathcal{G} \cong \left(\bigoplus^n \mathcal{O}_W \right) \otimes \mathcal{G} \xrightarrow{\theta' \otimes \mathcal{G}} (\mathcal{G}^* \otimes \mathcal{H}^m) \otimes \mathcal{G} \cong \mathcal{H}om(\mathcal{G}, \mathcal{H}^m) \otimes \mathcal{G} \to \mathcal{H}^m.$$

Here, the last morphism is the evaluation morphism, which is surjective over W_0, since $\mathcal{G}|_{W_0}$ is locally free. Hence, the composite of the above morphisms is the desired morphism. $\square$

Lemma 3.2. *Let W be a normal quasiprojective variety and D a Cartier divisor on W. Let $W_0 \subseteq W$ be an open subset, $\mathcal{E}$ a coherent sheaf that is globally generated over W_0, and $\mathcal{G}$ a coherent sheaf on W such that $\mathcal{G}|_{W_0}$ is free. Suppose that there exists a morphism*

$$\varphi : \mathcal{E} \otimes \left(\bigotimes^{p^e} \mathcal{G} \right) \to \mathcal{O}_W(D) \otimes (F_W^e)^* \mathcal{G}$$

that is surjective over W_0. Then $W_0 \cap \boldsymbol{B}_-(D) = \varnothing$.

Proof. Obviously, we may assume that $\mathcal{E} = \bigoplus^d \mathcal{O}_W$ for some $d \in \mathbb{Z}_{>0}$. Then $\mathcal{E} \otimes \big(\bigotimes^{p^e} \mathcal{G}\big) \cong \bigoplus^d \big(\bigotimes^{p^e} \mathcal{G}\big)$. Take an ample line bundle $\mathcal{H}$ on W such that $\mathcal{G} \otimes \mathcal{H}$ is globally generated. Since we have $\big(\bigotimes^{p^e} \mathcal{G}\big) \otimes \mathcal{H}^{p^e} \cong \bigotimes^{p^e} (\mathcal{G} \otimes \mathcal{H})$ and $((F_Y^e)^* \mathcal{G}) \otimes \mathcal{H}^{p^e} \cong (F_Y^e)^* (\mathcal{G} \otimes \mathcal{H})$, replacing $\mathcal{G}$ (resp. φ) by $\mathcal{G} \otimes \mathcal{H}$ (resp. $\varphi \otimes \mathcal{H}^{p^e}$), we may assume that $\mathcal{G}$ is globally generated. Let $S(\mathcal{G}) \subseteq \mathbb{Q}$ be the set of rational numbers r satisfying the following condition: there is $h \in \mathbb{Z}_{>0}$ such that $p^h r \in \mathbb{Z}$ and that the sheaf

$$\mathcal{O}_W(p^h r D) \otimes (F_W^h)^* \mathcal{G}$$

is globally generated over W_0. We then have $0 \in S(\mathcal{G})$. We prove that $S(\mathcal{G})$ is *not* bounded from above. Choose $r \in S$ and $h \in \mathbb{Z}_{>0}$ so that the above conditions hold. We then have the following sequence of morphisms:

$$\begin{aligned} \bigoplus^d \left(\bigotimes^{p^e} (\mathcal{O}_W(p^h r D) \otimes (F_W^h)^* \mathcal{G}) \right) &\cong \mathcal{O}_W(p^{e+h} r D) \otimes (F_W^h)^* \left(\bigoplus^d \left(\bigotimes^{p^e} \mathcal{G} \right) \right) \\ &\xrightarrow{\psi} \mathcal{O}_W(p^{e+h} r D) \otimes (F_W^h)^* (\mathcal{O}_W(D) \otimes (F_W^e)^* \mathcal{G}) \\ &\cong \mathcal{O}_W((p^h + p^{e+h} r) D) \otimes (F_W^{e+h})^* \mathcal{G} \end{aligned}$$

Here, $\psi := ((F_W^h)^* \varphi) \otimes \mathcal{O}_W(p^h r D)$, so it is surjective over W_0, which implies that the last sheaf is globally generated over W_0. We then see that $1/p^e + r = (p^h + p^{e+h} r)/p^{e+h} \in S(\mathcal{G})$, and hence $S(\mathcal{G})$ can not be bounded from above. Next, we show the assertion. Lemma 3.1 shows that we have an ample Cartier divisor H on W and a morphism $\theta : \bigoplus^n \mathcal{G} \to \mathcal{H} := \mathcal{O}_W(H)$ that is surjective over W_0. One can easily check that $S(\mathcal{G}) \subseteq S(\mathcal{H})$, so $S(\mathcal{H})$ is also not bounded from above. Take $0 < r \in S(\mathcal{H})$. Then for some $h \gg 0$, the sheaf $\mathcal{O}_W(p^h r D) \otimes (F_W^h)^* \mathcal{H} \cong \mathcal{O}_W\big(p^h r \big(D + \frac{1}{r} H\big)\big)$ is globally generated over W_0, and so $\boldsymbol{B}\big(D + \frac{1}{r} H\big) \subseteq W \setminus W_0$. Hence, we conclude from Lemma 2.10 that $\boldsymbol{B}_-(D) \subseteq W \setminus W_0$. $\square$

Before stating the next lemma, we recall Keeler's vanishing theorem, which is a relative version of Fujita's vanishing theorem.

Theorem 3.3 [Keeler 2003, Theorem 1.5]. *Let $f : X \to Y$ be a projective morphism between Noetherian schemes, $\mathcal{F}$ a coherent sheaf on X, and $\mathcal{L}$ an f-ample line bundle on X. Then there exists $m_0 \in \mathbb{Z}_{>0}$ such that*

$$R^i f_* (\mathcal{F} \otimes \mathcal{L}^m \otimes \mathcal{N}) = 0$$

for each $m \geq m_0$ and every f-nef line bundle $\mathcal{N}$ on X.

Note that, in the situation of the theorem, a line bundle on X is said to be f-nef if the restriction to each fiber is nef.

Lemma 3.4. *Let $f : X \to Y$ be a surjective morphism between projective varieties, and $\mathcal{A}$ an ample line bundle on X. Then there exists $m_0 \in \mathbb{Z}_{>0}$ such that $f_*(\mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N})$ is generated by its global sections for each $m \geq m_0$ and every nef line bundle $\mathcal{N}$ on X.*

Proof. Let $\mathcal{H}$ be an ample line bundle on Y that is generated by its global sections. Set $n := \dim Y$. Take $m_1 > 0$ so that $\mathcal{A}^{m_1} \otimes f^*\mathcal{H}^{-n}$ is nef. Applying Theorem 3.3, we get $m_2 > 0$ such that $H^i(X, \mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N}) = 0$ and $R^i f_*(\mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N}) = 0$ for each $m \geq m_2$, $i > 0$ and every nef line bundle $\mathcal{N}$ on X. The Leray spectral sequence then implies that $H^i(Y, f_*(\mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N})) = 0$ for each $i > 0$. Fix $m \geq m_0 := m_1 + m_2$ and a nef line bundle $\mathcal{N}$ on X, and set $\mathcal{M} := \mathcal{A}^m \otimes \mathcal{N}$. We then have $\mathcal{F} \otimes \mathcal{M} \otimes f^*\mathcal{H}^{-i} \cong \mathcal{F} \otimes \mathcal{A}^{m-m_1} \otimes (\mathcal{A}^{m_1} \otimes f^*\mathcal{H}^{-n}) \otimes f^*\mathcal{H}^{n-i}$, so the above argument tells us that the i-th cohomology of $f_*(\mathcal{F} \otimes \mathcal{M} \otimes f^*\mathcal{H}^{-i}) \cong (f_*(\mathcal{F} \otimes \mathcal{M})) \otimes \mathcal{H}^{-i}$ vanishes for $0 < i \leq n$. This means that $f_*(\mathcal{F} \otimes \mathcal{M})$ is 0-regular with respect to $\mathcal{H}$, so this sheaf is generated by its global sections thanks to the Castelnuovo–Mumford regularity (see [Lazarsfeld 2004, Theorem 1.8.5]). ☐

Lemma 3.5. *Let $g : V \to W$ be a surjective projective morphism from a k-scheme V to a variety W, let $\mathcal{A}$ be a g-ample line bundle on V, and let $\mathcal{F}$ be a coherent sheaf on V that is flat over W. Then, there exists $m_0 \in \mathbb{Z}_{>0}$ such that $g_*(\mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N})$ is locally free for each $m \geq m_0$ and every g-nef line bundle $\mathcal{N}$ on V.*

Proof. By Theorem 3.3, there is $m_0 \in \mathbb{Z}_{>0}$ such that $R^i g_*(\mathcal{F} \otimes \mathcal{A}^m \otimes \mathcal{N}) = 0$ for each $m \geq m_0$, $i > 0$ and every g-nef line bundle $\mathcal{N}$ on V. Fix $m \geq m_0$ and a g-nef line bundle $\mathcal{N}$ on V, and set $\mathcal{M} := \mathcal{A}^m \otimes \mathcal{N}$. For each $i \geq 0$, define the function $h^i : W \to \mathbb{Z}$ by $h^i(w) := \dim_{k(w)} H^i(V_w, (\mathcal{F} \otimes \mathcal{M})|_{V_w})$. By the choice of m_0 and cohomology and base change (see [Hartshorne 1977, Theorem III 12.11]), we obtain that $h^i = 0$ for each $i > 0$, so $\chi((\mathcal{F} \otimes \mathcal{M})|_{V_w}) = h^0(w)$ for every $w \in W$. Then [Hartshorne 1977, Theorem III 9.9 and its proof] implies that h^0 is constant. Hence, our claim follows from Grauert's theorem (see [Hartshorne 1977, Corollary III 12.9]). ☐

Lemma 3.6. *Let the notation be as in Lemma 3.5. Let $\mathcal{L}$ be a line bundle on V:*

(1) *If $\mathcal{L}$ is g-free, then there exists $n_0 \in \mathbb{Z}_{>0}$ such that the natural morphism*

$$g_*\mathcal{L}^m \otimes g_*(\mathcal{F} \otimes \mathcal{L}^n \otimes g^*\mathcal{P}) \to g_*(\mathcal{F} \otimes \mathcal{L}^{m+n} \otimes g^*\mathcal{P})$$

is surjective for each $n \geq n_0$, $m > 0$ and every line bundle $\mathcal{P}$ on W.

(2) *If $\mathcal{L}$ is g-ample and g-free, then there exists $n_0 \in \mathbb{Z}_{>0}$ such that the natural morphism*

$$g_*\mathcal{L}^m \otimes g_*(\mathcal{F} \otimes \mathcal{L}^n \otimes \mathcal{N}) \to g_*(\mathcal{F} \otimes \mathcal{L}^{m+n} \otimes \mathcal{N})$$

is surjective for each $n \geq n_0$, $m > 0$ and every g-nef line bundle $\mathcal{N}$ on V.

Proof. We first show that (2) implies (1). Since $\mathcal{L}$ is g-free, g can be decomposed as

$$g : V \xrightarrow{\sigma} Z \xrightarrow{\tau} W$$

and $\mathcal{L} \cong \sigma^*\mathcal{M}$ for a τ-ample and τ-free line bundle $\mathcal{M}$ on Z. Then we have

$$\begin{tikzcd}
(\tau_*\mathcal{M}^m) \otimes \tau_*((\sigma_*\mathcal{F}) \otimes \mathcal{M}^n \otimes \tau^*\mathcal{P}) \arrow[r] \arrow[d] & \tau_*((\sigma_*\mathcal{F}) \otimes \mathcal{M}^{m+n} \otimes \tau^*\mathcal{P}) \arrow[dd, "\cong"] \\
(\tau_*\sigma_*\sigma^*\mathcal{M}^m) \otimes \tau_*((\sigma_*\mathcal{F}) \otimes \mathcal{M}^n \otimes \tau^*\mathcal{P}) \arrow[d, "\cong"] & \\
\tau_*\sigma_*\mathcal{L}^m \otimes \tau_*\sigma_*(\mathcal{F} \otimes \mathcal{L}^n \otimes \sigma^*\tau^*\mathcal{P}) \arrow[r] & \tau_*\sigma_*(\mathcal{F} \otimes \mathcal{L}^{m+n} \otimes \sigma^*\tau^*\mathcal{P})
\end{tikzcd}$$

Here, the isomorphisms follow from the projection formula. If (2) holds, then the top horizontal arrow is surjective, and hence so is the bottom horizontal arrow, so (1) holds. We show (2). Theorem 3.3 tells us that we have $n_0 \in \mathbb{Z}_{>0}$ such that for each $n \geq n_0$ and every g-nef line bundle $\mathcal{N}$ on V, the sheaf $\mathcal{F} \otimes \mathcal{L}^n \otimes \mathcal{N}$ is 0-regular with respect to $\mathcal{L}$ and g. Hence, in the case when $m = 1$, our claim follows from the relative Castelnuovo–Mumford regularity (see [Lazarsfeld 2004, Example 1.8.24]). Using this repeatedly, we see that the natural morphism

$$\left(\bigotimes^m g_*\mathcal{L}\right) \otimes g_*(\mathcal{F} \otimes \mathcal{L}^n \otimes \mathcal{N}) \to g_*(\mathcal{F} \otimes \mathcal{L}^{m+n} \otimes \mathcal{N})$$

is surjective for each $m \in \mathbb{Z}_{>0}$. This morphism factors through

$$(g_*\mathcal{L}^m) \otimes g_*(\mathcal{F} \otimes \mathcal{L}^n \otimes \mathcal{N}) \to g_*(\mathcal{F} \otimes \mathcal{L}^{m+n} \otimes \mathcal{N}),$$

which completes the proof. □

Lemma 3.7. *Let V be a k-scheme of pure dimension satisfying S_2 and G_1, let W be a regular variety, and let $g : V \to W$ be a flat projective morphism. Let $E \geq 0$ be an AC divisor on V such that $aK_V + E$ is Cartier for some $a \in \mathbb{Z}_{>0}$ with $p \nmid a$. Set $\Delta := E \otimes \frac{1}{a}$. Let $U \subseteq V$ be a Gorenstein open subset. Suppose that the codimension of $(V \setminus U)|_{V_w}$ (resp. $E|_{V_w}$) in V_w is at least 2 (resp. 1) for every $w \in W$:*

(1) [Patakfalvi et al. 2018, Corollary 3.31] *The set $W_0 := \{w \in W \mid (V_{\overline{w}}, \overline{\Delta|_{U_{\overline{w}}}})$ is F-pure$\}$ is an open subset of W. Here, $V_{\overline{w}}$ is the geometric fiber over w and $\overline{\Delta|_{U_{\overline{w}}}}$ is the $\mathbb{Z}_{(p)}$-AC divisor on $V_{\overline{w}}$ obtained as the unique extension of $\Delta|_{U_{\overline{w}}}$.*

(2) *Assume that W_0 is nonempty. Set $V_0 := g^{-1}(W_0)$. Let $\mathcal{A}$ be a line bundle on V such that $\mathcal{A}|_{V_0}$ is $g|_{V_0}$-ample. Then there exists $m_0 \in \mathbb{Z}_{>0}$ such that*

$$g_{W^e*}(\phi^{(e)}_{(V/W,\Delta)} \otimes \mathcal{A}^m_{W^e} \otimes \mathcal{N}_{W^e}) : g^{(e)}{}_*(\mathcal{L}^{(e)}_{(V/W,\Delta)} \otimes \mathcal{A}^{p^e m} \otimes \mathcal{N}^{p^e}) \to g_{W^e*}(\mathcal{A}^m_{W^e} \otimes \mathcal{N}_{W^e})$$

is surjective over W_0 for each $e \in \mathbb{Z}_{>0}$ with $a \mid (p^e - 1)$, $m \geq m_0$ and every line bundle $\mathcal{N}$ on V whose restriction $\mathcal{N}|_{V_0}$ to V_0 is $g|_{V_0}$-nef

Proof. One can prove (1) by the same argument as that in [Patakfalvi et al. 2018]. We prove (2). Let $d > 0$ be the minimum integer such that $a \mid (p^d - 1)$. For simplicity, let $\phi^{(e)}$ (resp. $\mathcal{L}^{(e)}$) denote $\phi^{(e)}_{(V/W,\Delta)}$ (resp. $\mathcal{L}^{(e)}_{(V/W,\Delta)}$) for each $e \in d\mathbb{Z}_{>0}$. Replacing W by W_0, we may assume that $W_0 = W$. The morphism $\phi^{(e)}|_{V_{\overline{w}}} \cong \phi^{(e)}_{(V_{\overline{w}}/\overline{w},\Delta_{\overline{w}})}$ is then surjective for every $w \in W$ and $e \in d\mathbb{Z}_{>0}$, so $\phi^{(e)}$ is surjective. Applying

Theorem 3.3 to the kernel of $\phi^{(d)}$, we obtain $m'_0 \in \mathbb{Z}_{>0}$ such that $g_{W^d *}(\phi^{(d)} \otimes \mathcal{A}^m_{W_d} \otimes \mathcal{N}_{W_d})$ is surjective for every $m \geq m'_0$ and g-nef line bundle $\mathcal{N}$ on V. Take $m_0 \geq m'_0$ so that $m_0 A - (K_{V/W} + \Delta)$ is g-nef, where A is a Cartier divisor on V satisfying $\mathcal{O}_V(A) \cong \mathcal{A}$. We fix $m \geq m_0$ and a g-nef line bundle $\mathcal{N}$ on V. Set $\mathcal{M} := \mathcal{A}^m \otimes \mathcal{N}$. We show that $\psi^{(e)} := g_{W^e *}(\phi^{(e)} \otimes \mathcal{M}_{W^e})$ is surjective for each $e \in d\mathbb{Z}_{>0}$. We have already seen that $\psi^{(d)}$ is surjective. Take $e \in d\mathbb{Z}_{>0}$. Assuming the surjectivity of $\psi^{(e)}$, we show that $\psi^{(e+d)}$ is surjective. By the definition of $\phi^{(e+d)}$, we have

$$\begin{aligned}\psi^{(e+d)} &= g_{W^{e+d} *}(\phi^{(e+d)} \otimes \mathcal{M}_{W^{e+d}}) \\ &\cong (F^d_W)^*(g_{W^e *}(\phi^{(e)} \otimes \mathcal{M}_{W^e})) \circ g^{(e)}_{W^{e+d} *}(\phi^{(d)}_{(V^e/W^e, \Delta^e)} \otimes (\mathcal{L}^{(e)} \otimes \mathcal{M}^{p^e})_{W^{e+d}}) \\ &= (F^d_W)^*(\psi^{(e)}) \circ g^{(e)}_{W^{e+d} *}(\phi^{(d)}_{(V^e/W^e, \Delta^e)} \otimes (\mathcal{L}^{(e)} \otimes \mathcal{M}^{p^e})_{W^{e+d}}).\end{aligned}$$

Note that the surjectivity of $\psi^{(e)}$ induces that of $(F^d_W)^*(\psi^{(e)})$, so we only need to show that

$$g^{(e)}_{W^{e+d} *}(\phi^{(d)}_{(V^e/W^e, \Delta^e)} \otimes (\mathcal{L}^{(e)} \otimes \mathcal{M}^{p^e})_{W^{e+d}})$$

is surjective. We can rewrite this morphism as

$$g_{W^d *}(\phi^{(d)} \otimes (\mathcal{O}_V((1-p^e)(K_{V/W} + \Delta)) \otimes \mathcal{M}^{p^e})_{W^d})$$

identifying $g^{(e)} : V^e \to W^e$ with $g : V \to W$. This morphism is surjective if

$$\mathcal{P} := \mathcal{O}_V((1-p^e)(K_{V/W} + \Delta)) \otimes \mathcal{M}^{p^e} \otimes \mathcal{A}^{-m_0}$$

is g-nef, by the choice of m'_0. This g-nefness follows from the isomorphisms

$$\mathcal{P} \cong \mathcal{O}_V((1-p^e)(K_{V/W}+\Delta)) \otimes \mathcal{A}^{mp^e - m_0} \otimes \mathcal{N}^{p^e} \cong \mathcal{O}_V((p^e-1)(m_0 A - (K_{V/W}+\Delta))) \otimes \mathcal{A}^{(m-m_0)p^e} \otimes \mathcal{N}^{p^e}$$

and the choice of m_0. □

4. Main theorems and corollaries

In this section, we prove the main theorems. After this, we give several corollaries.

4A. ***Main theorems.*** In this subsection, we prove Theorems 4.1 and 4.2. *Throughout this subsection, we use the following notation:*

Fix an F-finite field k. Let $f : X \to Y$ be a surjective projective morphism from a pure dimensional quasiprojective k-scheme X satisfying S_2 and G_1 to a normal quasiprojective variety Y. Let E be an effective AC-divisor on X and $a > 0$ an integer such that $aK_X + E$ is Cartier. Set $\Delta := E \otimes a^{-1}$. Let $U \subset X$ be the Gorenstein locus and $W \subseteq Y$ the maximal regular open subset such that $g := f|_V : V \to W$ is flat, where $V := f^{-1}(W)$. Suppose that there exists a scheme-theoretic point $w \in W$ with the following properties:

(i) $\operatorname{codim}_{X_w}(X_w \setminus U_w) \geq 2$.

(ii) Supp E does not contain any irreducible component of X_w.

(iii) $(X_{\overline{w}}, \overline{\Delta|_{U_{\overline{w}}}})$ is F-pure.

Here, $X_{\overline{w}}$ (resp. $U_{\overline{w}}$) is the geometric fiber of f (resp. $f|_U$) over w, and $\overline{\Delta|_{U_{\overline{w}}}}$ is the $\mathbb{Z}_{(p)}$-AC divisor on $X_{\overline{w}}$ obtained as the unique extension of $\Delta|_{U_{\overline{w}}}$. Let D be a $\mathbb{Q}$-Cartier divisor on Y.

We now have the following commutative diagram whose squares are cartesian:

$$\begin{tikzcd}
U_{\overline{w}} \arrow[rr] \arrow[d, hook, "(\mathrm{codim}\geq 2)"] & & \underset{}{\overset{(\text{Gorenstein})}{U}} \arrow[d, hook, "(\mathrm{codim}\geq 2)"] \\
X_{\overline{w}} \arrow[r] \arrow[d, "f_{\overline{w}}"] & V \arrow[r, hook] \arrow[d, "g", "(\text{flat})"'] & X \arrow[d, "f"] \\
\overline{w} := \operatorname{Spec} \overline{k(w)} \arrow[r] & \underset{(\text{regular})}{W} \arrow[r, hook] & Y
\end{tikzcd}$$

Here, "(codim ≥ 2)" means the morphism is an open immersion whose complement is of codimension at least 2.

In this situation, we prove the following two theorems:

Theorem 4.1. *Let the notation be as above. Assume that X is projective and $K_X + \Delta$ is $\mathbb{Q}$-Cartier*:

(1) *If $p\nmid a$ and $-(K_X+\Delta+f^*D)$ is nef, then the closure of $\boldsymbol{B}_-(-(K_W+D|_W))$ in Y does not contain w.*

(2) *If K_X is $\mathbb{Q}$-Cartier and $-(K_X+\Delta+f^*D)$ is ample, then $\boldsymbol{B}_+(-(K_W+D|_W))$ does not contain w.*

Theorem 4.2. *Let the notation be as above. Take $b \in \mathbb{Z}_{>0}$ so that bD is Cartier and set $\mathcal{M} := \mathcal{O}_X(-ab(K_X+\Delta+f^*D))$. Assume that $\mathcal{M}|_U$ is $f|_U$-free for some open subset $U \subseteq X$ containing $f^{-1}(w)$. If $p\nmid a$ and if $f_*\mathcal{M}$ is weakly positive over $\{w\}$, then $w \notin \boldsymbol{B}_-(-(K_W+D|_W))$.*

Remarks 4.3. (1) When X is normal, we may choose a as the Cartier index of $K_X+\Delta$.

(2) When $w \in W$ is the generic point, assumptions (i) and (ii) above hold. However, assumption (iii) does not necessarily hold even if X is smooth and $\Delta = 0$.

(3) If X is a variety, then $\operatorname{codim}_Y(Y\setminus W) \geq 2$. Furthermore, if $\operatorname{codim}_Y(Y\setminus W)\geq 2$ and K_Y is $\mathbb{Q}$-Cartier, then $\boldsymbol{B}_-(-(K_W+D|_W)) = \boldsymbol{B}_-(-(K_Y+D))|_W$.

Remark 4.4. Take $m \in \mathbb{Z}$. In the following proof, $m\Delta$ (resp. mf^*D) denotes $\frac{m}{a}E$ (resp. $f^*(mD)$) if $a \mid m$ (resp. mD is Cartier). Note that there may be two distinct AC divisors on X that are equal as $\mathbb{Z}_{(p)}$-AC divisors.

Proof of Theorem 4.1. We first show that (1) implies (2). By the assumption in (2), Δ is $\mathbb{Q}$-Cartier. Write $a = mp^c$, where $m, c \in \mathbb{Z}_{\geq 0}$ with $p\nmid m$. Take $e \gg 0$. Put $a' := m(p^e+1)$. Then $p\nmid a'$. Set $\Delta' := (p^{e-c}E)\otimes\frac{1}{a'}$. We then have $\Delta' = p^{e-c}a/a'\Delta = p^e/(p^e+1)\Delta \leq \Delta$, so (X, Δ') satisfies assumption (iii). Since $e \gg 0$, we may assume that $-(K_X+\Delta'+f^*D) = -(K_X+\Delta+f^*D) - 1/(p^e+1)\Delta$ is ample. Let H be an ample $\mathbb{Q}$-Cartier divisor on Y such that $-(K_X+\Delta'+f^*(D+H))$ is nef. Then, (1) implies that

$w \notin \boldsymbol{B}_-(-(K_W + D|_W + H|_W))$. Putting $\Gamma := -(K_W + D|_W)$, we obtain that $\boldsymbol{B}_+(\Gamma) \subseteq \boldsymbol{B}\big(\Gamma - \frac{1}{2}H|_W\big) \subseteq \boldsymbol{B}_-(\Gamma - H|_W) \subseteq W \setminus \{w\}$.

We begin the proof of (1). Define W_0 to be the subset of points in W satisfying conditions (i)–(iii). We first claim that $W_0 \subseteq W$ is open. Lemma 3.7 (1) tells us that (iii) is an open condition on W. Set $r := \dim X - \dim Y$ and $Z := (X \setminus U)_{\mathrm{red}}$. Then, condition (i) (resp. (ii)) is equivalent to saying that $\dim Z_w \le r - 2$ (resp. $\dim E_w \le r - 1$). Hence, our claim follows from Chevalley's theorem [EGA IV$_3$ 1966, Corollaire 13.1.5], which says that the function $\delta(w) := \dim Z_w$ (resp. $\delta(w) := \dim E_w$) on W is upper semicontinuous. Next, let us introduce some notation:

(n1) Take $d \in \mathbb{Z}_{>0}$ with $a \mid d$ such that dD and $d(K_X + \Delta)$ are Cartier.

(n2) Let $\mathcal{A}'$ be an ample line bundle on X and put $\mathcal{A}'|_V := \mathcal{A}$.

(n3) Denote $g_*\mathcal{A}$ by $\mathcal{G}$.

Lemmas 3.4–3.7 tell us that, by replacing $\mathcal{A}'$, we may assume that the following conditions hold:

(a1) For every nef line bundle $\mathcal{N}'$ on X and each $0 \le i < d$ with $a \mid i$, the sheaf

$$g_*(\mathcal{O}_V(-i(K_V + \Delta|_V)) \otimes \mathcal{A} \otimes \mathcal{N}) \cong (f_*(\mathcal{O}_X(-i(K_X + \Delta)) \otimes \mathcal{A}' \otimes \mathcal{N}'))\big|_W$$

is a locally free sheaf generated by its global sections, where $\mathcal{N} := \mathcal{N}'|_V$. In particular, $\mathcal{G} = g_*\mathcal{A}$ is locally free (Lemmas 3.4 and 3.5).

(a2) For a g-nef line bundle $\mathcal{N}$ on V, the natural morphism

$$\mathcal{G} \otimes g_*(\mathcal{A}^n \otimes \mathcal{N}) = (g_*\mathcal{A}) \otimes g_*(\mathcal{A}^n \otimes \mathcal{N}) \to g_*(\mathcal{A}^{n+1} \otimes \mathcal{N})$$

is surjective for each $n \in \mathbb{Z}_{>0}$ (Lemma 3.6).

(a3) For each $e \in \mathbb{Z}_{>0}$ with $a \mid (p^e - 1)$, there exists a morphism

$$g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)} \otimes \mathcal{A}^{p^e}) \to g_{W^e *}\mathcal{A}_{W^e} \cong (F_W^e)^* g_*\mathcal{A} = (F_W^e)^*\mathcal{G}$$

that is surjective over W_0 (Lemma 3.7).

We continue to introduce some notation:

(n4) Take an ample Cartier divisor H on W such that for each $0 \le i < d$, there is a surjective morphism $\bigoplus^t \omega_Y^{-i} \otimes \mathcal{G} \to \mathcal{H} := \mathcal{O}_W(H)$. Such an H exists as shown in Lemma 3.1.

(n5) Fix $e \in \mathbb{Z}_{>0}$ with $a \mid (p^e - 1)$ and write $p^e - 1 = dq + r$ for $q, r \in \mathbb{Z}$ with $0 \le r < d$. Note that $a \mid r$.

(n6) Set $\mathcal{N}'$ to be the nef line bundle $\mathcal{O}_X(-dq(K_X + \Delta + f^*D))$ and

$$\mathcal{N} := \mathcal{N}'|_V \cong \mathcal{O}_V(-dq(K_V + \Delta|_V + g^*(D|_W))).$$

Also, set $\mathcal{P} := \mathcal{O}_W(-dq(K_W + D|_W))$.

(n7) Recall that $\mathcal{L}^{(e)}_{(V/W,\Delta|_V)} := \mathcal{O}_V((1 - p^e)(K_{V/W} + \Delta|_V))$. Here, we identify V^e (resp. W^e) with V (resp. W). Note that $\mathcal{L}_{(V/W,\Delta|_V)}$ is g-nef in this situation.

We prove the assertion. We have

$$\mathcal{N}\otimes g^*\mathcal{P}^{-1} \underset{(n6)}{\cong} \mathcal{O}_V(-dq(K_{V/W}+\Delta|_V)) \underset{(n7)}{\cong} \mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{O}_V(r(K_{V/W}+\Delta|_V)).$$

We then obtain

$$\mathcal{O}_V(-r(K_V+\Delta|_V))\otimes\mathcal{A}\otimes\mathcal{N}\cong\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A}\otimes g^*(\mathcal{P}\otimes\omega_W^{-r}),$$

so the projection formula implies that

$$g_*(\mathcal{O}_V(-r(K_V+\Delta|_V))\otimes\mathcal{A}\otimes\mathcal{N})\cong g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A})\otimes\mathcal{P}\otimes\omega_W^{-r}. \tag{$*$}$$

It then follows from (a1) that the right-hand side is globally generated. Hence, we may apply Lemma 3.2 to the composition of the following morphisms which are surjective over W_0:

$$\begin{aligned}\left(\bigotimes^{p^e}\mathcal{G}\right)\otimes\bigoplus^{t}((g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A}))\otimes\mathcal{P}\otimes\omega_Y^{-r}) &\cong \left(\bigotimes^{p^e-1}\mathcal{G}\right)\otimes(g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A}))\otimes\mathcal{P}\otimes\left(\bigoplus^{t}\mathcal{G}\otimes\omega_Y^{-r}\right)\\ &\xrightarrow{(n4)} \left(\bigotimes^{p^e-1}\mathcal{G}\right)\otimes(g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A}))\otimes\mathcal{P}\otimes\mathcal{H}\\ &\xrightarrow{(a2)} (g_*(\mathcal{L}^{(e)}_{(V/W,\Delta|_V)}\otimes\mathcal{A}^{p^e}))\otimes\mathcal{P}\otimes\mathcal{H}\\ &\xrightarrow{(a3)} ((F_Y^e)^*\mathcal{G})\otimes\mathcal{P}\otimes\mathcal{H}.\end{aligned}$$

Note that $\mathcal{P}\otimes\mathcal{H}\cong\mathcal{O}_W\big(dq\big(-(K_W+D|_W)+\frac{1}{dq}H\big)\big)$. Thus, we obtain

$$\boldsymbol{B}\big(-(K_Y+D)+\tfrac{2}{dq}H\big) \underset{\text{by def}}{\subseteq} \boldsymbol{B}_-\big(-(K_Y+D)+\tfrac{1}{dq}H\big) \underset{\text{Lemma 3.2}}{\subseteq} W\setminus W_0.$$

Since $\dfrac{2}{dq}$ goes to zero as $e\to\infty$, we conclude from Lemma 2.10 that $\boldsymbol{B}_-(-(K_Y+D))\subseteq W\setminus W_0$. □

Proof of Theorem 4.2. Replacing $f:X\to Y$ by $g:V\to W$, we may assume that $X=V$ and $Y=W$. Then f is flat. Set $Y_0:=Y\setminus(f(X\setminus U))$. Note that $Y_0\subseteq Y$ is an open subset containing w. We may assume $U=f^{-1}(Y_0)$ by shrinking U. Put $f_0:=f_{Y_0}:U\to Y_0$. Since $\mathcal{M}|_U$ is f_0-free by assumption, f_0 can be decomposed as

$$f_0:U\xrightarrow{\sigma}T\xrightarrow{\tau}Y_0$$

and $\mathcal{M}|_U\cong\sigma^*\mathcal{R}$ for a τ-ample line bundle $\mathcal{R}$ on T. For each $c\in\mathbb{Z}_{>0}$, the projection formula says that $\sigma_*(\mathcal{M}^c|_U)\cong(\sigma_*\mathcal{O}_U)\otimes\mathcal{R}^c$, so we get

$$(f_*\mathcal{M}^c)|_{Y_0}\cong(f_0)_*(\mathcal{M}^c|_U)\cong\tau_*\sigma_*(\mathcal{M}^c|_U)\cong\tau_*((\sigma_*\mathcal{O}_U)\otimes\mathcal{R}^c).$$

The last sheaf is locally free if $c\gg0$, as shown in [Hartshorne 1977, Theorem III 9.9 and its proof]. Fix such a c. Replacing b by bc, we may assume that $(f_*\mathcal{M})|_{Y_0}$ is locally free. We then have a closed subset $Z\subset Y$ of codimension at least 2 such that $Y_0\subseteq Y\setminus Z$ and $(f_*\mathcal{M})|_{Y\setminus Z}$ is locally free. Shrinking Y to $Y\setminus Z$, we may assume that $f_*\mathcal{M}$ is locally free. Take $\alpha\in\mathbb{Z}_{>0}$ and an ample Cartier divisor H on Y. Set

$\mathcal{H} := \mathcal{O}_Y(H)$. Then, there is $\beta \in \mathbb{Z}_{>0}$ such that $(S^{\alpha\beta} f_*\mathcal{M}) \otimes \mathcal{H}^\beta$ is globally generated over $\{w\}$. We may assume that this sheaf is globally generated over Y_0, shrinking Y_0 to a neighborhood of w. We set

(n8) $d := ab\alpha\beta$

and use notation (n2) and (n3) in the proof of Theorem 4.1. Assume that $\mathcal{A}$ satisfies conditions (a2) and (a3). Furthermore, we add the following notation:

(n9) Take $n_0 \in \mathbb{Z}_{>0}$ with $a \mid n_0$ such that for each $n \geq n_0$, $0 \leq i < d$ and every line bundle $\mathcal{Q}$ on Y, the natural morphism

$$(f_*\mathcal{M}) \otimes f_*(\mathcal{M}^n \otimes \mathcal{O}_X(-i(K_X+\Delta)) \otimes \mathcal{A} \otimes f^*\mathcal{Q}) \to f_*(\mathcal{M}^{n+1} \otimes \mathcal{O}_X(-i(K_X+\Delta)) \otimes \mathcal{A} \otimes f^*\mathcal{Q})$$

is surjective over Y_0. We can find such an n_0 by Lemma 3.6.

(n10) Choose $\nu \in \mathbb{Z}_{>0}$ so that

- $\mathcal{H}^\nu \otimes f_*\mathcal{O}_X(-i(K_X+\Delta))$ is generated by its global sections for each $i \in a\mathbb{Z}$ with $abn_0 \leq i < abn_0 + d$, and
- there is a morphism $\bigoplus^t \mathcal{G} \to \mathcal{H}^\nu$ that is surjective over Y_0.

The existence of such a ν is ensured by Lemma 3.1.

(n11) Fix $e \in \mathbb{Z}_{>0}$ with $a \mid (p^e - 1)$ and write $p^e - 1 = dq + r$ for $q, r \in \mathbb{Z}$ with $abn_0 \leq r < abn_0 + d$. Note that $a \mid r$.

We also use notation (n6) and (n7) in the proof of Theorem 4.1. Then,

$$\mathcal{N} = \mathcal{O}_X(-dq(K_X+\Delta+f^*D)) \cong \mathcal{M}^{\alpha\beta q} \quad \text{and}$$
$$\mathcal{O}_X(-r(K_X+\Delta)) \cong \mathcal{M}^{n_0} \otimes \mathcal{O}_X(-(abn_0 - r)(K_X+\Delta)) \otimes g^*\mathcal{O}_Y(abn_0 D).$$

Note that $0 \leq abn_0 - r < d$. Therefore, the morphisms

$$\begin{aligned}
&(S^q(\mathcal{H}^\beta \otimes S^{\alpha\beta} f_*\mathcal{M})) \otimes \mathcal{H}^\nu \otimes f_*(\mathcal{O}_X(-r(K_X+\Delta)) \otimes \mathcal{A}) \\
&\qquad \cong \mathcal{H}^{\beta q+\nu} \otimes (S^q(S^{\alpha\beta} f_*\mathcal{M})) \otimes f_*(\mathcal{O}_X(-r(K_X+\Delta)) \otimes \mathcal{A}) \\
&\qquad \xrightarrow{(n9)} \mathcal{H}^{\beta q+\nu} \otimes f_*(\mathcal{O}_X(-r(K_X+\Delta)) \otimes \mathcal{A} \otimes \mathcal{M}^{\alpha\beta q}) \\
&\qquad \cong \mathcal{H}^{\beta q+\nu} \otimes f_*(\mathcal{O}_X(-r(K_X+\Delta)) \otimes \mathcal{A} \otimes \mathcal{N}) \\
&\qquad \underset{(*)}{\cong} \mathcal{H}^{\beta q+\nu} \otimes f_*(\mathcal{L}^{(e)}_{(X/Y,\Delta)} \otimes \mathcal{A}) \otimes \mathcal{P} \otimes \omega_Y^{-r}
\end{aligned}$$

are surjective over Y_0. Here, the last isomorphism is $(*)$ in the proof of Theorem 4.1. By the choice of β and ν, we see that the first sheaf is globally generated over Y_0, and hence so is the last sheaf. Now, we have the following sequence of morphisms that are surjective over Y_0:

$$\begin{aligned}
&\Big(\bigotimes^{p^e}\mathcal{G}\Big)\otimes\bigoplus^{t}(\mathcal{H}^{\beta q+\nu}\otimes f_*(\mathcal{L}^{(e)}_{(X/Y,\Delta)}\otimes\mathcal{A})\otimes\mathcal{P}\otimes\omega_Y^{-r})\\
&\qquad\cong\Big(\bigotimes^{p^e-1}\mathcal{G}\Big)\otimes(f_*(\mathcal{L}^{(e)}_{(X/Y,\Delta)}\otimes\mathcal{A}))\otimes\mathcal{P}\otimes\Big(\bigoplus^{t}\mathcal{G}\otimes\omega_Y^{-r}\Big)\otimes\mathcal{H}^{\beta q+\nu}\\
&\qquad\xrightarrow{(n10)}\Big(\bigotimes^{p^e-1}\mathcal{G}\Big)\otimes(f_*(\mathcal{L}^{(e)}_{(X/Y,\Delta)}\otimes\mathcal{A}))\otimes\mathcal{P}\otimes\mathcal{H}^{\beta q+2\nu}\\
&\qquad\xrightarrow{(a2)}(f_*(\mathcal{L}^{(e)}_{(X/Y,\Delta)}\otimes\mathcal{A}^{p^e}))\otimes\mathcal{P}\otimes\mathcal{H}^{\beta q+2\nu}\\
&\qquad\xrightarrow{(a3)}((F_Y^e)^*\mathcal{G})\otimes\mathcal{P}\otimes\mathcal{H}^{\beta q+2\nu}.
\end{aligned}$$

Note that $\mathcal{P}\otimes\mathcal{H}^{\beta q+2\nu}\cong\mathcal{O}_X(dq(-(K_Y+D)+(\beta q+2\nu)/(dq)H))$. Replacing e by some larger one if necessary, we may assume that $\beta q>2\nu$. Then,

$$\boldsymbol{B}\Big(-(K_Y+D)+\frac{2\beta q}{dq}H\Big)\underset{\text{by def}}{\subseteq}\boldsymbol{B}_-\Big(-(K_Y+D)+\frac{\beta q+2\nu}{dq}H\Big)\underset{\text{Lemma 3.2}}{\subseteq}Y\setminus Y_0.$$

Since $\dfrac{2\beta q}{dq}=\dfrac{2}{ab\alpha}$ goes to zero as $\alpha\to\infty$, we conclude from Lemma 2.10 that $\boldsymbol{B}_-(-(K_Y+D))\subseteq Y\setminus Y_0$. $\square$

4B. *Corollaries.* In this subsection, we give several corollaries of the main theorems. *Throughout this subsection, we use the following notation:*

Let $f:X\to Y$ be a surjective morphism between regular projective varieties over an F-finite field, Δ an effective $\mathbb{Q}$-divisor on X, and a the Cartier index of Δ. Let D be a $\mathbb{Q}$-divisor on Y. Let $\bar{\eta}$ denote the geometric generic point of Y.

Corollary 4.5. *Assume that f is flat. Suppose that* Supp Δ *does not contain any component of any fiber, and $(X_{\bar{y}},\Delta_{\bar{y}})$ is F-pure for every point $y\in Y$:*

(1) *If $p\nmid a$ and if $-(K_X+\Delta+f^*D)$ is nef, then so is $-(K_Y+D)$.*

(2) *If $-(K_X+\Delta+f^*D)$ is ample, then so is $-(K_Y+D)$.*

Proof. This follows from Theorem 4.1 and Remarks 2.9 immediately. $\square$

The author learned the proof of Corollary 4.6(3) below from professor Yoshinori Gongyo.

Corollary 4.6. *Assume that $(X_{\bar{\eta}},\Delta_{\bar{\eta}})$ is F-pure:*

(1) *If $p\nmid a$ and if $-(K_X+\Delta+f^*D)$ is nef, then $-(K_Y+D)$ is pseudoeffective.*

(2) *If $-(K_X+\Delta+f^*D)$ is ample, then $-(K_Y+D)$ is big.*

(3) *If $(X_{\bar{\eta}},\Delta_{\bar{\eta}})$ is strongly F-regular and if $-(K_X+\Delta+f^*D)$ is nef and big, then $-(K_Y+D)$ is big.*

Proof. By Remarks 2.9, we see that (1) and (2) follow from (1) and (2) of Theorem 4.1, respectively. We prove (3). By Kodaira's lemma, there is a $\mathbb{Q}$-divisor $\Delta'\geq\Delta$ on X such that $-(K_X+\Delta'+f^*D)$ is ample and $(X_{\bar{\eta}},\Delta'_{\bar{\eta}})$ is again strongly F-regular. Hence (3) follows from (2). $\square$

Corollary 4.7. *Assume that $(X_{\bar{\eta}},\Delta_{\bar{\eta}})$ is F-pure. If $p\nmid a$ and if $K_X+\Delta$ is numerically equivalent to $f^*(K_Y+L)$ for some $\mathbb{Q}$-divisor L on Y, then L is pseudoeffective.*

Proof. Set $D := -(K_Y + L)$. Then, $K_X + \Delta + f^*D$ is numerically trivial, so it is nef. Hence, by Corollary 4.6(1), we obtain the assertion. $\square$

Remarks 4.8. (1) In the situation of Corollary 4.7, it is known that if $(X_{\bar\eta}, \Delta_{\bar\eta})$ is globally F-split, then $\kappa(L) \geq 0$ (see [Das and Schwede 2017, Theorem B] or [Ejiri 2017, Theorem 3.18]). Of course, $(X_{\bar\eta}, \Delta_{\bar\eta})$ is not necessary globally F-split even if $X_{\bar\eta}$ is a smooth curve and $\Delta = 0$. At the same time, Chen and Zhang proved that the relative canonical divisor of an elliptic fibration has nonnegative Kodaira dimension [Chen and Zhang 2015, 3.2].

(2) In the case when $\dim Y = 1$, Corollary 4.7 follows from a result of Patakfalvi [2014, Theorem 1.6].

Corollary 4.9. *Assume that f is flat and every geometric fiber is F-pure*:

(1) *If X is a Fano variety, that is, $-K_X$ is ample, then so is Y.*

(2) *Suppose that the geometric generic fiber of f is strongly F-regular. If X is a weak Fano variety, that is, $-K_X$ is nef and big, then so is Y.*

Proof. Putting $\Delta = 0$ and $D = 0$, we see that the assertions follow from Corollaries 4.5(2) and 4.6(3). $\square$

Corollary 4.10. *Assume that Y has positive dimension*:

(1) *If $(X_{\bar\eta}, \Delta_{\bar\eta})$ is F-pure, then $-(K_{X/Y} + \Delta)$ is not ample.*

(2) *If $(X_{\bar\eta}, \Delta_{\bar\eta})$ is strongly F-regular, then $-(K_{X/Y} + \Delta)$ cannot be both nef and big.*

Proof. Set $D := -K_Y$. Then $-(K_X + \Delta + f^*D) = -(K_{X/Y} + \Delta)$. Since $-(K_Y + D) = 0$ is not big, the assertions follows from Corollary 4.6(2) and (3). $\square$

Corollary 4.11. *Assume that Y has positive dimension. Suppose that $\mathcal{O}_X(-ab(K_X + \Delta))|_{X_\eta}$ is globally generated for some $b \in \mathbb{Z}_{>0}$. If $p \nmid a$ and if $(X_{\bar\eta}, \Delta_{\bar\eta})$ is F-pure, then $f_*\mathcal{O}_X(-ab(K_{X/Y} + \Delta))$ is not big.*

Proof. Set $\mathcal{G}_{(-l)} := f_*\mathcal{O}_X(al(K_X + \Delta))$ for each $l \in \mathbb{Z}$. Suppose that $\mathcal{G}_{(b)}$ is big. Then, $\mathcal{H}^{-1} \otimes S^\gamma \mathcal{G}_{(b)}$ is weakly positive for some $\gamma \in \mathbb{Z}_{>0}$ and an ample line bundle $\mathcal{H}$ on Y. Take $n_0 \in \mathbb{Z}_{>0}$ so that the natural morphism $\mathcal{G}_{(b)} \otimes \mathcal{G}_{(n)} \to \mathcal{G}_{(b+n)}$ is generically surjective for each $n \geq n_0$. We can find such an n_0 by Lemma 3.6. Choose $\nu \in \mathbb{Z}_{>0}$ so that $\mathcal{H}^\nu \otimes \mathcal{G}_{(n_0)}$ is globally generated. Fix $l \in \mathbb{Z}$ with $l > \nu$. Using the natural morphism $S^l(S^\gamma \mathcal{G}_{(b)}) \otimes \mathcal{G}_{(n_0)} \to \mathcal{G}_{(bl\gamma+n_0)}$ and [Viehweg 1995, Lemma 2.16], we see that $\mathcal{H}^{\nu-l} \otimes \mathcal{G}_{(bl\gamma+n_0)}$ is weakly positive. Let H be a Cartier divisor on Y satisfying $\mathcal{O}_Y(H) \cong \mathcal{H}$ and set $H' := (l-\nu)/(a(bl\gamma+n_0))H$. The projection formula then shows that

$$\mathcal{H}^{\nu-l} \otimes \mathcal{G}_{(bl\gamma+n_0)} \cong f_*\mathcal{O}_X(-a(bl\gamma+n_0)(K_{X/Y} + \Delta + f^*H')).$$

It then follows from Theorem 4.2 that $\boldsymbol{B}_-(-H') \neq \varnothing$, i.e., $-H'$ is pseudoeffective, a contradiction. $\square$

5. Results in arbitrary characteristic

In this section we generalize several results in Section 4 to arbitrary characteristic. In particular, we prove the characteristic zero counterparts of Corollaries 4.10 and 4.11 (Theorems 5.4 and 5.5). We also deal with a morphism that is special but not necessarily smooth, and show that the image of a Fano variety is again a Fano variety.

To begin with, let us recall the following definition:

Definition 5.1. Let X be a normal variety over a field k of characteristic zero, and Δ an effective $\mathbb{Q}$-Weil divisor on X. Let (X_R, Δ_R) be a model of (X, Δ) over a finitely generated $\mathbb{Z}$-subalgebra R of k. We say that (X, Δ) is of *dense F-pure type* (resp. *strongly F-regular type*) if there exists a dense (resp. dense open) subset $S \subseteq \operatorname{Spec} R$ such that (X_μ, Δ_μ) is *F-pure* (resp. strongly F-regular) for all closed points $\mu \in S$.

Remark 5.2. The above definition can be generalized in an obvious way to the case where X is a finite disjoint union of varieties over k.

Theorem 5.3 [Takagi 2004, Corollary 3.4]. *Let X be a normal variety over a field of characteristic zero, and Δ an effective $\mathbb{Q}$-Weil divisor on X such that $K_X + \Delta$ is $\mathbb{Q}$-Cartier. Then (X, Δ) is klt if and only if it is of strongly F-regular type.*

Theorem 5.4. *Let k be an algebraically closed field of characteristic zero. Let $f : X \to Y$ be a surjective morphism between smooth projective varieties of positive dimension, and Δ an effective $\mathbb{Q}$-divisor on X. If (X_y, Δ_y) is of dense F-pure type (resp. klt) for every general closed point $y \in Y$, then $-(K_{X/Y} + \Delta)$ cannot be ample (resp. both nef and big).*

Proof. Let X_R, Δ_R, Y_R, y_R and f_R be models of X, Δ, Y, y and f over a finitely generated $\mathbb{Z}$-algebra R, respectively. We may assume that $(X_R)_{y_R}$ is a model of X_y over R. We first suppose that (X_y, Δ_y) is of dense F-pure type for a general closed point $y \in Y$. Then, there is a dense subset $S \subseteq \operatorname{Spec} R$ such that $((X_y)_\mu, \Delta_\mu)$ is F-pure for every $\mu \in S$. Note that $(X_y)_\mu \cong (X_\mu)_{y_\mu}$ and $(\Delta_y)_\mu = (\Delta_\mu)_{y_\mu}$. Corollary 4.10 then implies that $-(K_{X_\mu/Y_\mu} + \Delta_\mu)$ is not ample, which means that $-(K_{X/Y} + \Delta)$ is not ample. We next suppose that (X_y, Δ_y) is klt for every general closed point $y \in Y$. If $-(K_{X/Y} + \Delta)$ is nef and big, then by Kodaira's lemma, there is $\Delta' \geq \Delta$ such that $-(K_{X/Y} + \Delta')$ is ample and (X_y, Δ'_y) is klt for a general closed point $y \in Y$. However, Theorem 5.3 tells us that (X_y, Δ'_y) is of dense F-pure type, which contradicts the above arguments. $\square$

Theorem 5.5. *Let k be an algebraically closed field of characteristic zero. Let $f : X \to Y$ be a surjective morphism between smooth projective varieties of positive dimension, and Δ an effective $\mathbb{Q}$-divisor on X. Assume that (X_y, Δ_y) is of dense F-pure type for a general closed point $y \in Y$. Let $\bar{\eta}$ be a geometric generic point of Y. If $\mathcal{O}_X(-m(K_{X/Y} + \Delta))|_{X_{\bar{\eta}}}$ is globally generated for some $m > 0$ such that $m\Delta$ is integral, then $f_*\mathcal{O}_X(-m(K_{X/Y} + \Delta))$ is not big.*

Proof. Set $\mathcal{G} := f_*\mathcal{O}_X(-m(K_{X/Y} + \Delta))$ and $r := \operatorname{rank} \mathcal{G}$. Since $y \in Y$ is general, f is flat at every point in $f^{-1}(y)$ and $\dim H^0(X_y, -m(K_{X_y} + \Delta_y)) = r$. Let X_R, Δ_R, Y_R, y_R and f_R be models of X, Δ, Y, y and f, respectively. By replacing R if necessary, we may assume that $f_{R*}\mathcal{O}_{X_R}(-m(K_{X_R/Y_R}+\Delta_R))$ and $(X_R)_{y_R}$ are respectively models of $\mathcal{G}$ and X_y. We may further assume that $\dim H^0((X_\mu)_{y_\mu}, -m(K_{X_\mu}+\Delta_\mu)_{y_\mu}) = r$ for every $\mu \in \operatorname{Spec} R$. Then, [Hartshorne 1977, Corollary 12.9] implies that the natural morphism

$$\mathcal{G}_\mu = f_{R*}\mathcal{O}_{X_R}(-m(K_{X_R/Y_R}+\Delta_R))|_{Y_\mu} \to f_{\mu*}\mathcal{O}_{X_\mu}(-m(K_{X_\mu/Y_\mu}+\Delta_\mu))$$

is surjective over y_μ. Since $f_{\mu*}\mathcal{O}_{X_\mu}(-m(K_{X_\mu/Y_\mu}+\Delta_\mu))$ is not big as shown in Corollary 4.11, $\mathcal{G}_\mu$ is also not big. Hence, the lemma below completes the proof. □

Lemma 5.6. *Let $\mathcal{G}$ be a torsion-free coherent sheaf on a smooth quasiprojective variety Y over an algebraically closed field of characteristic zero. Let Y_R and $\mathcal{G}_R$ be models of Y and $\mathcal{G}$ respectively over a finitely generated $\mathbb{Z}$-algebra R. If $\mathcal{G}$ is big, then there exists a dense open subset $S \subseteq \operatorname{Spec} R$ such that $\mathcal{G}_\mu$ is big for every $\mu \in S$.*

Proof. Let $Z \subset Y$ be a closed subset of codimension at least 2 such that $\mathcal{G}|_{Y\setminus Z}$ is locally free. Replacing Y by $Y \setminus Z$, we may assume that $\mathcal{G}$ is locally free. By the definition, we have $\gamma \in \mathbb{Z}_{>0}$ such that $\mathcal{H}^{-1} \otimes S^\gamma \mathcal{G}$ is weakly positive for some ample line bundle $\mathcal{H}$ on Y. Then, there is $\beta \in \mathbb{Z}_{>0}$ such that $\mathcal{H}^\beta \otimes S^{2\beta}(\mathcal{H}^{-1} \otimes S^\gamma \mathcal{G}) \cong \mathcal{H}^{-\beta} \otimes S^{2\beta}(S^\gamma \mathcal{G})$ is generically globally generated. Using the natural morphism $S^{2\beta}(S^\gamma \mathcal{G}) \to S^{2\beta+\gamma}\mathcal{G}$, we see that $\mathcal{F} := \mathcal{H}^{-\beta} \otimes S^{2\beta+\gamma}\mathcal{G}$ is generically globally generated, i.e., there is a morphism $\theta : \bigoplus^t \mathcal{O}_Y \to \mathcal{F}$ that is surjective over a dense open subset $V \subseteq Y$, where $t \in \mathbb{Z}_{>0}$. Let θ_R, H_R and V_R be models of θ, H and V over R, respectively. Replacing R if necessary, we may assume that θ_R is surjective over V_R. Thus for every closed point $\mu \in \operatorname{Spec} R$, the morphism $\theta_\mu : \bigoplus^t \mathcal{O}_{X_\mu} \to \mathcal{H}_\mu^{-\beta} \otimes S^{2\beta+\gamma}\mathcal{G}_\mu$ is surjective over V_μ, which means that $\mathcal{G}_\mu$ is big. □

Kollár, Miyaoka and Mori [1992, Corollary 2.9] (compare [Miyaoka 1993, Theorem 3]) proved that images of Fano varieties under smooth morphisms are again Fano varieties. The rest of this paper is devoted to extending this result to toroidal morphisms.

Definition 5.7 [Abramovich and Karu 2000; Kawamata 2002]. Let k be an algebraically closed field of arbitrary characteristic:

(i) Let X be a normal variety and U an open subset of X. We say that the embedding $U \subseteq X$ is *toroidal* if for every closed point $x \in X$, there exists
   - a toric variety V with torus T,
   - a closed point $v \in V$ and
   - an isomorphism $\hat{\mathcal{O}}_{X,x} \cong \hat{\mathcal{O}}_{V,v}$ of complete local k-algebras such that the ideal of $B := X \setminus U$ maps isomorphically to that of $V \setminus T$.

   Such a pair (V, v) is called a *local model* at $x \in X$. The pair (X, B) is often called a *toroidal variety*.

(ii) Let (X, B) and (Y, C) be toroidal varieties. A *toroidal morphism* $f : (X, B) \to (Y, C)$ is a dominant morphism $f : X \to Y$ with $f(X \setminus B) \subseteq Y \setminus C$ such that for every closed point $x \in X$, there exist
   - local models (V, v) and (W, w) at x and $y := f(x)$, respectively, and

- a toric morphism $g : V \to W$ such that the following diagram commutes:

$$\begin{tikzcd} \hat{\mathcal{O}}_{X,x} \arrow[r, "\cong"] & \hat{\mathcal{O}}_{V,v} \\ \hat{\mathcal{O}}_{Y,y} \arrow[u, "f^{\#}"] \arrow[r, "\cong"] & \hat{\mathcal{O}}_{W,w} \arrow[u, "\hat{g}^{\#}"'] \end{tikzcd}$$

The next theorem is a generalization of [Kollár et al. 1992, Corollary 2.9].

Theorem 5.8. *Let k be an algebraically closed field of any characteristic $p \ge 0$. Let $f : X \to Y$ be a surjective morphism between smooth projective varieties and B a reduced divisor on X. Let Δ be a $\mathbb{Q}$-divisor on X such that $0 \le \Delta \le B$ and that $a\Delta$ is integral for some $0 < a \in \mathbb{Z} \setminus p\mathbb{Z}$. Assume that the following conditions hold*:

(i) *f induces a toroidal morphism $f : (X, B) \to (Y, \varnothing)$.*

(ii) *f is equidimensional.*

(iii) *Every closed fiber of f is reduced.*

(iv) Supp Δ *does not contain any irreducible component of any fiber.*

In this situation, if $-(K_X + \Delta + f^*D)$ is ample for some $\mathbb{Q}$-divisor D on Y, then so is $-(K_Y + D)$.

Proof. Let $x_\lambda \in X$ be a closed point and set $y_\lambda := f(x_\lambda)$. By assumption (i), there is a local model (V_λ, v_λ) (resp. (W_λ, w_λ)) at x_λ (resp. y_λ) and a toric morphism $g_\lambda : V_\lambda \to W_\lambda$. Using Artin's approximation theorem [1969, Corollary 2.6], we obtain a commutative diagram

$$\begin{tikzcd} X \arrow[d, "f"'] & T_\lambda \arrow[l, "\rho_\lambda"'] \arrow[r, "\mu_\lambda"] \arrow[d, "h_\lambda"'] & V_\lambda \arrow[d, "g_\lambda"] \\ Y & U_\lambda \arrow[l, "\sigma_\lambda"'] \arrow[r, "\nu_\lambda"] & W_\lambda \end{tikzcd} \tag{$*$}$$

such that

- T_λ and U_λ are varieties,
- all the horizontal morphisms are étale, and
- there is a closed point $t_\lambda \in T_\lambda$ such that $\rho_\lambda(t_\lambda) = x_\lambda$ and $\mu_\lambda(t_\lambda) = v_\lambda$.

Let $k[v_1, v_1^{-1}, \dots, v_m, v_m^{-1}]$ (resp. $k[w_1, w_1^{-1}, \dots, w_n, w_n^{-1}]$) be the coordinate ring of the torus of V_λ (resp. W_λ). Set $t_i := \mu_\lambda^* v_i$ and $u_i := \nu_\lambda^* w_i$ for each i. We then see from assumptions (ii) and (iii) that

$$h_\lambda^* u_j = h_\lambda^* \nu_\lambda^* w_j = \mu_\lambda^* g_\lambda^* w_j = \mu_\lambda^* \prod_{l_{j-1} < i \le l_j} v_i = \prod_{l_{j-1} < i \le l_j} t_i$$

for $j = 1, \dots, n$, where $0 = l_0 < l_1 < \cdots < l_n \le m$.

Shrinking T_λ if necessary, we may assume that for any closed point $t \in T_\lambda$, there are $a_1, \dots, a_m \in k$ such that $\mathfrak{m}_t = (t_1 - a_1, \dots, t_m - a_m)$. We may also assume a similar condition for U_λ and $u_1, \dots, u_n$.

Let Λ be a *finite* set of λ such that $X = \bigcup_{\lambda \in \Lambda} \rho_\lambda(T_\lambda)$. When $p = 0$, one can check that diagram $(*)$ can be reduced to characteristic $p \gg 0$ for all $\lambda \in \Lambda$ simultaneously. For this reason, we consider the case of $p > 0$.

We see from [Matsumura 1986, Corollary of Theorem 23.1] that f is flat. Therefore, in order to apply Corollary 4.5, we only need to show that $(Z, \Delta|_Z)$ is F-pure for every closed fiber Z of f. This holds if $(S, (\rho_\lambda^*\Delta)|_S)$ is F-pure for every closed fiber S of h_λ, since ρ_λ is étale. Fix a closed fiber S over $u \in U_\lambda$ and a closed point $t \in T_\lambda$ contained in S. Then, there are $a_1, \dots, a_m, b_1, \dots, b_n \in k$ such that $\mathfrak{m}_t = (t_1 - a_1, \dots, t_m - a_m)$ and $\mathfrak{m}_u = (u_1 - b_1, \dots, u_n - b_n)$. Put $t_i' := t_i - a_i$ and $u_i' := u_i - b_i$ for each i. We then have

$$h_\lambda^* u_j' = \prod_{l_{j-1} < i \le l_j} (t_i' + a_i) - \prod_{l_{j-1} < i \le l_j} a_i$$

for $j = 1, \dots, n$. Set $\delta := \prod_{l_n < i \le m} t_i$. Now, one can easily check that

- the sequence $(h_\lambda^* u_1', \dots, h_\lambda^* u_n', \delta)$ is $\mathcal{O}_{T_\lambda, t}$-regular, and
- $(h_\lambda^* u_1' \cdots h_\lambda^* u_n')^{q-1} \cdot \delta^{q-1} \notin \mathfrak{m}_t^{[q]}$ for every $q = p^e$.

Then, [Hara and Watanabe 2002, Corollary 2.7] tells us that $(S, \operatorname{div}(\delta)|_S)$ is F-pure around t. Since $\rho_\lambda^* \Delta \le \operatorname{div}(\delta)$, we conclude that $(S, (\rho_\lambda^*\Delta)|_S)$ is F-pure around t. $\square$

Example 5.9. Let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the canonical basis of $\mathbb{R}^3$. For integers $m, n \ge 0$, we define $\boldsymbol{v}_{m,n} := (1, m, n) \in \mathbb{R}^3$. Let $\Sigma_{m,n}$ be the fan consisting of all the faces of the following cones:

$$\langle \boldsymbol{v}_{m,n}, \boldsymbol{e}_2, \boldsymbol{e}_2 + \boldsymbol{e}_3 \rangle, \langle \boldsymbol{v}_{m,n}, \boldsymbol{e}_2 + \boldsymbol{e}_3, \boldsymbol{e}_3 \rangle, \langle \boldsymbol{v}_{m,n}, -\boldsymbol{e}_2, \boldsymbol{e}_3 \rangle, \langle \boldsymbol{v}_{m,n}, \boldsymbol{e}_2, -\boldsymbol{e}_3 \rangle, \langle \boldsymbol{v}_{m,n}, -\boldsymbol{e}_2, -\boldsymbol{e}_3 \rangle,$$
$$\langle -\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_2 + \boldsymbol{e}_3 \rangle, \langle -\boldsymbol{e}_1, \boldsymbol{e}_2 + \boldsymbol{e}_3, \boldsymbol{e}_3 \rangle, \langle -\boldsymbol{e}_1, -\boldsymbol{e}_2, \boldsymbol{e}_3 \rangle, \langle -\boldsymbol{e}_1, \boldsymbol{e}_2, -\boldsymbol{e}_3 \rangle, \langle -\boldsymbol{e}_1, -\boldsymbol{e}_2, -\boldsymbol{e}_3 \rangle.$$

Let $X_{m,n}$ be the smooth toric 3-fold corresponding to the fan $\Sigma_{m,n}$ with respect to the lattice $\mathbb{Z}^3 \subset \mathbb{R}^3$. Then $X_{m,n}$ is a Fano variety if and only if $m, n \in \{0, 1\}$. The projection $\mathbb{R}^3 \to \mathbb{R}^2 : (x, y, z) \mapsto (x, y)$ induces a toric morphism $f : X_{m,n} \to Y_m$ from $X_{m,n}$ to the Hirzebruch surface $Y_m := \mathbb{P}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(m))$. Set $\Delta = 0$. Then one can check that f satisfies the assumptions of Theorem 5.8, but it is not smooth. Hence by Theorem 5.8, we see that Y_m is a Del Pezzo surface if $m = 0, 1$. In fact, it is well known that Y_m is a Del Pezzo surface if and only if $m = 0, 1$.

Acknowledgements

The author is greatly indebted to his supervisor Professor Shunsuke Takagi for suggesting several problems studied in this paper and for much valuable advice. He wishes to express his gratitude to Professor Yoshinori Gongyo for showing him the proof of Corollary 4.9(2) and for fruitful conversations. He is grateful to Professors Zsolt Patakfalvi and Hiromu Tanaka for suggesting improvements of several results in a preprint version of this paper. He would like to thank Professors Caucher Birkar, Yifei Chen and Osamu Fujino for stimulating discussions and helpful comments. He also would like to thank Professor Akiyoshi Sannai, and Doctors Takeru Fukuoka, Akihiro Kanemitsu, Eleonora Anna Romano, Kenta Sato, Antoine Song and Fumiaki Suzuki for useful comments. Finally, he is grateful to the referees for a careful reading and for suggesting an improvement of Theorem 5.8. He was supported by JSPS KAKENHI Grant Number 15J09117 and the Program for Leading Graduate Schools, MEXT, Japan.

References

[Abramovich and Karu 2000] D. Abramovich and K. Karu, "Weak semistable reduction in characteristic 0", *Invent. Math.* **139**:2 (2000), 241–273. MR Zbl

[Artin 1969] M. Artin, "Algebraic approximation of structures over complete local rings", *Inst. Hautes Études Sci. Publ. Math.* **36** (1969), 23–58. MR Zbl

[Birkar and Chen 2016] C. Birkar and Y. Chen, "Images of manifolds with semi-ample anti-canonical divisor", *J. Algebraic Geom.* **25**:2 (2016), 273–287. MR Zbl

[Chen and Zhang 2013] M. Chen and Q. Zhang, "On a question of Demailly–Peternell–Schneider", *J. Eur. Math. Soc.* **15**:5 (2013), 1853–1858. MR Zbl

[Chen and Zhang 2015] Y. Chen and L. Zhang, "The subadditivity of the Kodaira dimension for fibrations of relative dimension one in positive characteristics", *Math. Res. Lett.* **22**:3 (2015), 675–696. MR Zbl

[Corti 1992] A. Corti, "Adjunction of log divisors", pp. 171–182 in *Flips and abundance for algebraic threefolds* (Salt Lake City, UT, 1991), edited by J. Kollár, Astérisque **211**, Soc. Math. France, Paris, 1992. Zbl

[Das and Schwede 2017] O. Das and K. Schwede, "The F-different and a canonical bundle formula", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **17**:3 (2017), 1173–1205. MR Zbl

[Debarre 2001] O. Debarre, *Higher-dimensional algebraic geometry*, Springer, 2001. MR Zbl

[EGA IV_3 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[Ein et al. 2006] L. Ein, R. Lazarsfeld, M. Mustaţă, M. Nakamaye, and M. Popa, "Asymptotic invariants of base loci", *Ann. Inst. Fourier* (*Grenoble*) **56**:6 (2006), 1701–1734. MR Zbl

[Ejiri 2017] S. Ejiri, "Weak positivity theorem and Frobenius stable canonical rings of geometric generic fibers", *J. Algebraic Geom.* **26**:4 (2017), 691–734. MR Zbl

[Fujino and Gongyo 2012] O. Fujino and Y. Gongyo, "On images of weak Fano manifolds", *Math. Z.* **270**:1-2 (2012), 531–544. MR Zbl

[Fujino and Gongyo 2014] O. Fujino and Y. Gongyo, "On images of weak Fano manifolds, II", pp. 201–207 in *Algebraic and complex geometry* (Hannover, Germany, 2012), edited by A. Frühbis-Krüger et al., Springer Proc. Math. Stat. **71**, Springer, 2014. MR Zbl

[Hara and Watanabe 2002] N. Hara and K.-I. Watanabe, "F-regular and F-pure rings vs. log terminal and log canonical singularities", *J. Algebraic Geom.* **11**:2 (2002), 363–392. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Math. **52**, Springer, 1977. MR Zbl

[Hartshorne 1994] R. Hartshorne, "Generalized divisors on Gorenstein schemes", *K-Theory* **8**:3 (1994), 287–339. MR Zbl

[Kawamata 2002] Y. Kawamata, "On algebraic fiber spaces", pp. 135–154 in *Contemporary trends in algebraic geometry and algebraic topology* (Tianjin, 2000), edited by S.-S. Chern et al., Nankai Tracts Math. **5**, World Sci., River Edge, NJ, 2002. MR Zbl

[Keeler 2003] D. S. Keeler, "Ample filters of invertible sheaves", *J. Algebra* **259**:1 (2003), 243–283. MR Zbl

[Kollár et al. 1992] J. Kollár, Y. Miyaoka, and S. Mori, "Rational connectedness and boundedness of Fano manifolds", *J. Differential Geom.* **36**:3 (1992), 765–779. MR Zbl

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, II: Positivity for vector bundles, and multiplier ideals*, Ergebnisse der Mathematik (3) **49**, Springer, 2004. MR Zbl

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Adv. Math. **8**, Cambridge Univ. Press, 1986. MR Zbl

[Miller and Schwede 2012] L. E. Miller and K. Schwede, "Semi-log canonical vs F-pure singularities", *J. Algebra* **349**:1 (2012), 150–164. MR Zbl

[Miyaoka 1993] Y. Miyaoka, "Relative deformations of morphisms and applications to fibre spaces", *Comment. Math. Univ. St. Paul.* **42**:1 (1993), 1–7. MR Zbl

[Mustaţă 2013] M. Mustaţă, "The non-nef locus in positive characteristic", pp. 535–551 in *A celebration of algebraic geometry* (Cambridge, MA, 2011), edited by B. Hassett et al., Clay Math. Proc. **18**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Patakfalvi 2014] Z. Patakfalvi, "Semi-positivity in positive characteristics", *Ann. Sci. École Norm. Sup.* (4) **47**:5 (2014), 991–1025. MR Zbl

[Patakfalvi et al. 2018] Z. Patakfalvi, K. Schwede, and W. Zhang, "F-singularities in families", *Algebr. Geom.* **5**:3 (2018), 264–327. MR Zbl

[Prokhorov and Shokurov 2009] Y. G. Prokhorov and V. V. Shokurov, "Towards the second main theorem on complements", *J. Algebraic Geom.* **18**:1 (2009), 151–199. MR Zbl

[Schwede 2008] K. Schwede, "Generalized test ideals, sharp F-purity, and sharp test elements", *Math. Res. Lett.* **15**:6 (2008), 1251–1261. MR Zbl

[Takagi 2004] S. Takagi, "An interpretation of multiplier ideals via tight closure", *J. Algebraic Geom.* **13**:2 (2004), 393–415. MR Zbl

[Viehweg 1983] E. Viehweg, "Weak positivity and the additivity of the Kodaira dimension for certain fibre spaces", pp. 329–353 in *Algebraic varieties and analytic varieties* (Tokyo, 1981), edited by S. Iitaka, Adv. Stud. Pure Math. **1**, North-Holland, Amsterdam, 1983. MR Zbl

[Viehweg 1995] E. Viehweg, *Quasi-projective moduli for polarized manifolds*, Ergebnisse der Mathematik (3) **30**, Springer, 1995. MR Zbl

Communicated by Shigefumi Mori
Received 2018-05-22 Revised 2019-05-09 Accepted 2019-06-13

s-ejiri@cr.math.sci.osaka-u.ac.jp *Department of Mathematics, Graduate School of Science, Osaka University, Japan*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2081

A probabilistic approach to systems of parameters and Noether normalization

Juliette Bruce and Daniel Erman

We study systems of parameters over finite fields from a probabilistic perspective and use this to give the first effective Noether normalization result over a finite field. Our central technique is an adaptation of Poonen's closed point sieve, where we sieve over higher dimensional subvarieties, and we express the desired probabilities via a zeta function-like power series that enumerates higher dimensional varieties instead of closed points. This also yields a new proof of a recent result of Gabber, Liu and Lorenzini (2015) and Chinburg, Moret-Bailly, Pappas and Taylor (2017) on Noether normalizations of projective families over the integers.

Given an n-dimensional projective scheme $X \subseteq \mathbb{P}^r$ over a field, Noether normalization says that we can find homogeneous polynomials that induce a finite morphism $X \to \mathbb{P}^n$. Such a morphism is determined by a system of parameters, namely by choosing homogeneous polynomials $f_0, f_1, \ldots, f_n$ of degree d where $X \cap V(f_0, f_1, \ldots, f_n) = \varnothing$. Such a system of polynomials $f_0, f_1, \ldots, f_n$ is a system of parameters on the homogeneous coordinate ring of X. More generally, for $k \leq n$ we say that $f_0, f_1, \ldots, f_k$ *are parameters on* X if

$$\dim \mathbb{V}(f_0, f_1, \ldots, f_k) \cap X = \dim X - (k+1).$$

By convention, the empty set has dimension -1.

Over an infinite field any generic choice of $\leq n+1$ linear polynomials will automatically be parameters on X. Over a finite field we can ask:

Questions 1.1. Let $\mathbb{F}_q$ be a finite field and $X \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ be an n-dimensional closed subscheme:

(1) What is the probability that a random choice $f_0, f_1, \ldots, f_k$ of polynomials of degree d will be parameters on X?

(2) Can one effectively bound the degrees d for which such a finite morphism exists?

We will provide new insight into these questions by studying the distribution of systems of parameters from both a geometric and probabilistic viewpoint.

The first author was partially supported by the NSF GRFP under Grant No. DGE-1256259. The second author was partially supported by NSF grants DMS-1302057 and DMS-1601619.
MSC2010: primary 13B02; secondary 11G25, 14D10, 14G10, 14G15.
Keywords: Noether normalization, system of parameters, closed point sieve.

For the geometric side, we fix a field $\boldsymbol{k}$ and let $S=\boldsymbol{k}[x_0,x_1,\dots,x_r]$ be the coordinate ring of $\mathbb{P}^r_{\boldsymbol{k}}$. We write S_d for the vector space of degree d polynomials in S. In Section 4, we define a scheme $\mathscr{D}_{k,d}(X)$ parametrizing collections that do not form parameters. The $\boldsymbol{k}$-points of $\mathscr{D}_{k,d}(X)$ are

$$\mathscr{D}_{k,d}(X)(\boldsymbol{k})=\{(f_0,f_1,\dots,f_k)\text{ that are } \textit{not} \text{ parameters on } X\}\subset \underbrace{S_d\times\cdots\times S_d}_{k+1\text{ copies}}.$$

We prove an elementary bound on the codimension of these closed subschemes of the affine space $S_d^{\oplus k+1}$.

Theorem 1.2. *Let $X\subseteq\mathbb{P}^r_{\boldsymbol{k}}$ be an n-dimensional closed subscheme. We have*:

$$\operatorname{codim}\mathscr{D}_{k,d}(X)=\begin{cases}\geq\binom{n-k+d}{n-k} & \textit{if } k<n,\\ =1 & \textit{if } k=n.\end{cases}$$

This generalizes several results from the literature: the case $k=n$ is a classical result about Chow forms [Gelfand et al. 1994, 3.2.B]. For $d=1$ and $k<n$, the bound is sharp, by a classical result about determinantal varieties.[1] The bound for the case $k=0$ appears in [Benoist 2011, Lemme 3.3]. If $k<n$, then the codimension grows as $d\to\infty$ and this factors into our asymptotic analysis over finite fields. It also leads to a uniform convergence result that allows us to go from a finite field to $\mathbb{Z}$.

For the probabilistic side, we work over a finite field $\mathbb{F}_q$ and compute the asymptotic probability that random polynomials $f_0,f_1,\dots,f_k$ of degree d are parameters on X. The following result, which follows from known results in the literature, shows that there is a bifurcation between the $k=n$ and $k<n$ cases, reflecting Theorem 1.2.

Theorem 1.3 [Bucur and Kedlaya 2012; Poonen 2013]. *Let $X\subseteq\mathbb{P}^r_{\mathbb{F}_q}$ be an n-dimensional closed subscheme. The asymptotic probability that random polynomials $f_0,f_1,\dots,f_k$ of degree d are parameters on X is*

$$\lim_{d\to\infty}\operatorname{Prob}(f_0,f_1,\dots,f_k \textit{ of degree } d \textit{ are parameters on } X)=\begin{cases}1 & \textit{if } k<n,\\ \zeta_X(n+1)^{-1} & \textit{if } k=n,\end{cases}$$

where $\zeta_X(s)$ is the arithmetic zeta function of X.

The maximal case $k=n$ follows from the $k=m+1$ case of Bucur and Kedlaya [2012, Theorem 1.2] (though they assume that X is smooth, their proof does not need that assumption when $k=m+1$) and is proven using Poonen's closed point sieve. Moreover, the result in both cases could be derived from a slight modification of [Poonen 2013, Proof of Theorem 2.1]. See also [Charles and Poonen 2016, Corollary 1.4] for a similar result.

The main results in our paper stem from a deeper investigation of the cases where $k<n$, as the limiting value of 1 is only the beginning of the story. In the following theorem, we use $|Z|$ to denote the number of irreducible components of a scheme Z, and we write $\dim Z\equiv k$ if Z is equidimensional of dimension k.

[1] See [Bruns and Vetter 1988, Theorem 2.5] for a modern statement and proof. That result has a complicated history, discussed in [Bruns and Vetter 1988, Section 2.E], with some cases dating as far back as [Macaulay 1916, Section 53].

Theorem 1.4. *Let $X \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ be a projective scheme of dimension n. Fix e and let $k < n$. The probability that random polynomials $f_0, f_1, \dots, f_k$ of degree d are parameters on X is*

$$\operatorname{Prob}\left(\begin{array}{c} f_0, f_1, \dots, f_k \text{ of degree } d \\ \text{are parameters on } X \end{array}\right) = 1 - \sum_{\substack{Z \subseteq X \text{ reduced} \\ \dim Z \equiv n-k \\ \deg Z \leq e}} (-1)^{|Z|-1} q^{-(k+1)h^0(Z, \mathcal{O}_Z(d))} + o(q^{-e(k+1)\binom{n-k+d}{n-k}}).$$

Theorem 1.4 illustrates that the probability of finding a sequence $f_0, f_1, \dots, f_k$ of parameters on X is intimately tied to the codimension k geometry of X. Note that, by basic properties of the Hilbert polynomial, as $d \to \infty$ we have

$$h^0(Z, \mathcal{O}_Z(d)) = \frac{\deg(Z)}{(n-k)!} d^{n-k} + o(d^{n-k}) = \deg(Z) \binom{n-k+d}{n-k} + o(d^{n-k}).$$

It follows that the term $q^{-(k+1)h^0(Z, \mathcal{O}_Z(d))}$ lies in $o(q^{-e(k+1)\binom{n-k+d}{n-k}})$ if and only if $\deg(Z) > e$.

For instance, setting $e = 1$, the sum simplifies to $1 - N \cdot q^{-(k+1)\binom{n-k+d}{n-k}} + o(q^{-(k+1)\binom{n-k+d}{n-k}})$, where N is the number of $(n-k)$-dimensional linear subspaces lying in X. It would thus be more difficult to find parameters on a variety X containing lots of linear spaces, as illustrated in Example 8.1. More generally, the probability of finding parameters for $k < n$ depends on a power series that counts the number of $(n-k)$-dimensional subvarieties of varying degrees, in analogue with the appearance of the zeta function in the $k = n$ case.

Our approach to Theorem 1.4 is motivated by a simple observation: $f_0, f_1, \dots, f_k$ fail to be parameters if and only if they all vanish along some $(n-k)$-dimensional subvariety of X. We thus develop an analogue of Poonen's sieve where closed points are replaced by $(n-k)$-dimensional varieties. Sieving over higher dimensional varieties presents new challenges, especially bounding the error. This error depends on the Hilbert function of these varieties, and one key innovation is a uniform lower bound for Hilbert functions given in Lemma 3.1.

This perspective also leads to our second main result: an answer to Questions 1.1.(2) where the bound is in terms of the sum of the degrees of the irreducible components. If $X \subseteq \mathbb{P}^r$ has minimal irreducible components $V_1, V_2, \dots, V_s$ (considered with the reduced scheme structure), then we define $\widehat{\deg}(X) := \sum_{i=1}^s \deg(V_i)$ (see Definition 2.2). We set $\log_q 0 = -\infty$.

Theorem 1.5. *Let $X \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ where $\dim X = n$. If $\max\{d, \frac{q}{d^n}\} \geq \widehat{\deg}(X)$ and*

$$d > \log_q \widehat{\deg}(X) + \log_q n + n \log_q d$$

then there exist $f_0, f_1, \dots, f_n$ of degree d^{n+1} inducing a finite morphism $\pi : X \to \mathbb{P}^n_{\mathbb{F}_q}$.

The bound is asymptotically optimal in q. Namely, if we fix $\widehat{\deg}(X)$, then as $q \to \infty$, the bound becomes $d = 1$. Thus, a linear Noether normalization exists if $q \gg \widehat{\deg}(X)$. For a fixed q, we expect the bound could be significantly improved. (Even the case $\dim X = 0$ would be interesting, as it is related to Kakeya type problems over finite fields [Ellenberg and Erman 2016; Ellenberg et al. 2010].)

Theorem 1.5 provides the first explicit bound for Noether normalization over a finite field. (One could potentially derive an explicit bound from Nagata's argument [1962, Chapter I.14], though the inductive nature of that construction would at best yield a bound that is multiply exponential in the largest degree of a defining equation of X.)

After computing the probabilities over finite fields, we combine these analyses and characterize the distribution of parameters on projective B-schemes where $B = \mathbb{Z}$ or $\mathbb{F}_q[t]$. We use standard notions of density for a subset of a free B-module; see Definition 7.1.

Corollary 1.6. *Let $B = \mathbb{Z}$ or $\mathbb{F}_q[t]$. If $X \subseteq \mathbb{P}^r_B$ is a closed subscheme whose general fiber over B has dimension n, then*

$$\lim_{d\to\infty} \operatorname{Density}\left\{ \begin{array}{l} f_0, f_1, \dots, f_k \textit{ of degree } d \textit{ that} \\ \textit{restrict to parameters on } X_p \textit{ for all } p \end{array} \right\} = \begin{cases} 1 & \textit{if } k < n, \\ 0 & \textit{if } k = n \textit{ and all } d. \end{cases}$$

The density over B thus equals the product over all the fibers of the asymptotic probabilities over $\mathbb{F}_q$. In the case $B = \mathbb{Z}$, our proof relies on Ekedahl's infinite Chinese remainder theorem [Ekedahl 1991, Theorem 1.2] combined with Proposition 5.1, which illustrates uniform convergence in p for the asymptotic probabilities in Theorem 1.3. In the case $B = \mathbb{F}_q[t]$, we use Poonen's analogue of Ekedahl's result [Poonen 2003, Theorem 3.1].

When $k = n$, an analogue of Corollary 1.6 for smoothness is given by Poonen [2004, Theorem 5.13]. Moreover, while it is unknown if there are any smooth hypersurfaces of degree > 2 over $\mathbb{Z}$ (see for example the discussion in [Poonen 2009]), the density zero subset from Corollary 1.6 turns out to be nonempty for large d. This leads to a new proof of a recent result about uniform Noether normalizations.

Corollary 1.7. *Let $B = \mathbb{Z}$ or $\mathbb{F}_q[t]$. Let $X \subseteq \mathbb{P}^r_B$ be a closed subscheme. If each fiber of X over B has dimension n, then for some d, there exist homogeneous polynomials $f_0, f_1, \dots, f_n \in B[x_0, x_1, \dots, x_r]$ of degree d inducing a finite morphism $\pi : X \to \mathbb{P}^n_B$.*

Corollary 1.7 is a special case of a recent result of Chinburg, Moret-Bailly, Pappas and Taylor [2017, Theorem 1.2] and of Gabber, Liu and Lorenzini [2015, Theorem 8.1]. This corollary can fail when B is any of $\mathbb{Q}[t]$ or $\mathbb{Z}[t]$ or $\mathbb{F}_q[s, t]$, as in those cases, the Picard group of a finite cover of Spec B can fail to be torsion. See Section 8 for explicit examples and counterexamples and see [Chinburg et al. 2017; Gabber et al. 2015] for generalizations and applications.

There are a few earlier results related to Noether normalization over the integers. For instance [Moh 1979] shows that Noether normalizations of semigroup rings always exist over $\mathbb{Z}$; and [Nagata 1962, Theorem 14.4] implies that given a family over any base, one can find a Noether normalization over an open subset of the base. Relative Noether normalizations play a key role in [Achinger 2015, Section 5]. There is also the incorrect claim in [Zariski and Samuel 1960, page 124] that Noether normalizations exist over any infinite domain (see [Abhyankar and Kravitz 2007]). Brennan and Epstein [2011] analyze the distribution of systems of parameters from a different perspective, introducing the notion of a generic matroid to relate various different systems of parameters. In addition, after our paper was posted, work of

Charles [2017] on arithmetic Bertini theorems appeared which, under the additional hypothesis that X is integral and flat, implies a stronger version of Corollary 1.6 where one also obtains bounds on the norms of the functions.

This paper is organized as follows. Section 2 gathers background results and Section 3 involves a key lower bound on Hilbert functions. Section 4 contains our geometric analysis of parameters including a proof of Theorem 1.2. Sections 5 and 6 contain the probabilistic analysis of parameters over finite fields: Section 5 proves Theorem 1.3 and Theorem 1.5 while Section 6 gives the more detailed description via an analogue of the zeta function enumerating $(n-k)$-dimensional subvarieties, including the proof of Theorem 1.4. Section 7 contains our analysis over $\mathbb{Z}$ including proofs of Corollaries 1.6 and 1.7 and related corollaries. Section 8 contains examples.

2. Background

In this section, we gather some algebraic and geometric facts that we will cite throughout.

Lemma 2.1. *Let $\boldsymbol{k}$ be a field and let R be a $(k+1)$-dimensional graded $\boldsymbol{k}$-algebra where $R_0 = \boldsymbol{k}$. If $f_0, f_1, \dots, f_k$ are homogeneous elements of degree d and $R/\langle f_0, f_1, \dots, f_k\rangle$ has finite length, then the extension $\boldsymbol{k}[z_0, z_1, \dots, z_k] \to R$ given by $z_i \mapsto f_i$ is a finite extension.*

Proof. See [Bruns and Herzog 1993, Theorem 1.5.17]. □

This lemma implies that if $X \subseteq \mathbb{P}^r_{\boldsymbol{k}}$ has dimension n, and if $f_0, f_1, \dots, f_n$ are parameters on X, then the map $\phi\colon X \to \mathbb{P}^n_{\boldsymbol{k}}$ given by sending $x \mapsto [f_0(x) : f_1(x) : \cdots : f_n(x)]$ is a finite morphism. In particular, if R is the homogeneous coordinate of X, then the ideal $\langle f_0, f_1, \dots, f_n\rangle \subseteq R$ has finite colength, and thus the base locus of ϕ is the empty set. In other words, ϕ defines a genuine morphism. Moreover, the lemma shows that the corresponding map of coordinate rings $\phi^\sharp : R \to \boldsymbol{k}[z_0, z_1, \dots, z_n]$ is finite, and this implies that ϕ is finite.

Definition 2.2. Let $X \subseteq \mathbb{P}^r$ be a projective scheme with minimal irreducible components $V_1, \dots, V_s$ (considered with the reduced scheme structure). We define $\widehat{\deg}(X) := \sum_{i=1}^s \deg(V_i)$. For a subscheme $X' \subseteq \mathbb{A}^r$ with projective closure $\overline{X}' \subseteq \mathbb{P}^r$ we define $\widehat{\deg}(X') := \widehat{\deg}(\overline{X}')$.

This provides a notion of degree which ignores nonreduced structure but takes into account components of lower dimension. Similar definitions have appeared in the literature: for instance, in the language of [Bayer and Mumford 1993, Section 3], we would have $\widehat{\deg}(X) = \sum_{j=0}^{\dim X} \text{geom-deg}_j(X)$.

Lemma 2.3. *Let $\boldsymbol{k}$ be any field and let $X \subseteq \mathbb{A}^r_{\boldsymbol{k}}$. Let $f_0, f_1, \dots, f_t$ be polynomials in $\boldsymbol{k}[x_1, \dots, x_r]$. If $X' = X \cap \mathbb{V}(f_0, f_1, \dots, f_t)$, then $\widehat{\deg}(X') \leq \widehat{\deg}(X) \cdot \prod_{i=0}^t \deg(f_i)$.*

Proof. This follows from the refined version of Bezout's theorem [Fulton 1984, Example 12.3.1]. □

3. A uniform lower bound on Hilbert functions

For a subscheme of $\mathbb{P}^r$, the Hilbert function in degree d is controlled by the Hilbert polynomial, at least if d is very large related to some invariants of the subscheme. We analyze the Hilbert function at the other extreme, where the degree of the subscheme may be much larger than d. The following lemma, which applies to subschemes of arbitrarily high degree, provides uniform lower bounds that are crucial to bounding the error in our sieves.

Lemma 3.1. *Let $\boldsymbol{k}$ be an arbitrary field and fix some $e \geq 0$. Let $V \subseteq \mathbb{P}^r_{\boldsymbol{k}}$ be any closed, m-dimensional subscheme of degree $> e$ with homogeneous coordinate ring R:*

(1) *We have* $\dim R_d \geq h^0(\mathbb{P}^m, \mathcal{O}_{\mathbb{P}^m}(d))$ *for all d.*

(2) *For any $0 < \epsilon < 1$, there exists a constant C depending only on m and ϵ (but not on d or $\boldsymbol{k}$ or R) such that*

$$\dim R_d > (e+\epsilon)\cdot h^0(\mathbb{P}^m, \mathcal{O}_{\mathbb{P}^m}(d))$$

for all $d \geq Ce^{m+1}$.

Proof. If $\boldsymbol{k}'$ is a field extension of $\boldsymbol{k}$, then the Hilbert series of R is the same as the Hilbert series of $R \otimes_{\boldsymbol{k}} \boldsymbol{k}'$. We can thus assume that $\boldsymbol{k}$ is an infinite field. For part (1), we simply take a linear Noether normalization $\boldsymbol{k}[t_0, t_1, \ldots, t_m] \subseteq R$ of the ring R [Eisenbud 1995, Theorem 13.3]. This yields $\boldsymbol{k}[t_0, t_1, \ldots, t_m]_d \subseteq R_d$, giving the statement about Hilbert functions.

We prove part (2) of the lemma by induction on m. Let $S = \boldsymbol{k}[x_0, x_1, \ldots, x_r]$ and let $I_V \subseteq S$ be the saturated, homogeneous ideal defining V. Thus $R = S/I_V$. If $m = 0$, then we have $\dim R_d \geq \min\{d+1, \deg V\} \geq \min\{d+1, e+1\}$ which is at least $e+\epsilon$ for all $d \geq e$. This proves the case $m = 0$, where the constant C can be chosen to be 1.

Now assume the claim holds for all closed subschemes of dimension less than m. Let $V \subset \mathbb{P}^r_{\boldsymbol{k}}$ be a closed subscheme with $\dim V = m \geq 1$. Fix $0 < \epsilon < 1$. Since we are working over an infinite field, [Eisenbud 1995, Lemma 13.2(c)] allows us to choose a linear form ℓ that is a nonzero divisor in R. This yields a short exact sequence $0 \to R(-1) \xrightarrow{\cdot\ell} R \to R/\ell \to 0$. Since $R/\ell = S/(I_V + \langle \ell \rangle)$, this yields the equality

$$\dim R_i = \dim R_{i-1} + \dim(S/(I_V + \langle \ell \rangle))_i. \tag{1}$$

Letting $W = V \cap V(\ell)$ we know that $\dim W = m-1$ and $\deg W = \deg V$. Moreover, if I_V is the saturated ideal defining V and if I_W is the saturated ideal defining W, then since I_W contains $I_V + \langle \ell \rangle$, we have $\dim(S/(I_V + \langle \ell \rangle))_i \geq \dim(S/I_W)_i$. Combining with (1) yields

$$\dim R_i \geq \dim R_{i-1} + \dim(S/I_W)_i. \tag{2}$$

Now, by induction, in the case $m-1$ and $\epsilon' := \frac{1+\epsilon}{2}$, there exists C' depending on ϵ' and $m-1$ (or equivalently depending on ϵ and m) where

$$\dim(S/I_W)_i \geq (e+\epsilon')\binom{m-1+i}{m-1} \tag{3}$$

for all $i \geq C'e^m$. Now let $d \geq C'e^m$. Iteratively applying (2) for $i = d, d-1, d-2, \ldots, \lceil C'e^m \rceil$, we obtain:

$$\dim R_d \geq \dim R_{\lceil C'e^m \rceil - 1} + \sum_{i=\lceil C'e^m \rceil}^{d} \dim(S/I_W)_i.$$

By dropping the $\dim R_{\lceil C'e^m \rceil - 1}$ term and applying (3), we conclude that

$$\dim R_d \geq \sum_{i=\lceil C'e^m \rceil}^{d} (e+\epsilon')\binom{m-1+i}{m-1}.$$

The identity $\sum_{i=a}^{b}\binom{i+k}{k} = \binom{b+k+1}{k+1} - \binom{a+k}{k+1}$ implies that $\sum_{i=\lceil C'e^m \rceil}^{d}(e+\epsilon')\binom{m-1+i}{m-1}$ can be rewritten as $(e+\epsilon')\left(\binom{m+d}{m} - \binom{m-1+\lceil C'e^m \rceil}{m}\right)$. There exists a constant C depending on ϵ and m so that $(\epsilon'-\epsilon)\binom{m+d}{m} = \left(\frac{1}{2}-\frac{\epsilon}{2}\right)\binom{m+d}{m} \geq (e+\epsilon')\binom{m-1+\lceil C'e^m \rceil}{m}$ for all $d \geq \lceil Ce^{m+1} \rceil$. Thus, for all $d \geq \lceil Ce^{m+1} \rceil$ we have

$$\dim R_d \geq (e+\epsilon')\binom{m+d}{m} - (\epsilon'-\epsilon)\binom{m+d}{d} = (e+\epsilon)\binom{m+d}{m}. \qquad \square$$

Remark 3.2. Asymptotically in e, the bound of Ce^2 is the best possible for curves. For instance, let $C \subseteq \mathbb{P}^r$ be a curve of degree $(e+1)$ lying inside some plane $\mathbb{P}^2 \subseteq \mathbb{P}^r$. Let R be the homogeneous coordinate ring of C. If $d \geq e$ then the Hilbert function is given by

$$\dim R_d = (e+1)d - \tfrac{e^2-e}{2}.$$

Thus, if we want $\dim R_d \geq (e+\epsilon)(d+1)$, we will need to let $d \geq (e^2+e+2\epsilon)/(2(1-\epsilon)) \approx \frac{1}{2}e^2$. It would be interesting to know if the bound Ce^{m+1} is the best possible for higher dimensional varieties.

4. Geometric analysis

In this section we analyze the geometric picture for the distribution of parameters on X. The basic idea behind the proof of Theorem 1.2 is that $f_0, f_1, \ldots, f_k$ fail to be parameters on X if and only if they all vanish along some $(n-k)$-dimensional subvariety of X. Since the Hilbert polynomial of a $(n-k)$-dimensional variety grows like d^{n-k}, when we restrict a degree d polynomial f_j to such a subvariety, it can be written in terms of $\approx d^{n-k}$ distinct monomials. The polynomial f_j will all vanish along the subvariety if and only if all of the $\approx d^{n-k}$ coefficients vanish. This rough estimate explains the growth of the codimension of $\mathscr{D}_{k,d}(X)$ as $d \to \infty$.

We begin by constructing the schemes $\mathscr{D}_{k,d}(X)$. Fix $X \subseteq \mathbb{P}^r_{\boldsymbol{k}}$ a closed subscheme of dimension n over a field $\boldsymbol{k}$. Given $k < n$ and $d > 0$, let $\mathscr{A}_{k,d}$ be the affine space $H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(d))^{\oplus k+1}$ and $\boldsymbol{k}[c_{0,1}, \ldots, c_{k,\binom{r+d}{d}}]$ be the corresponding polynomial ring. We enumerate the monomials in $H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(d))$ as $m_1, \ldots, m_{\binom{r+d}{d}}$, and then define the universal polynomial

$$F_i := \sum_{j=1}^{\binom{r+d}{d}} c_{i,j} m_j \in \boldsymbol{k}[c_{0,1}, \ldots, c_{k,\binom{r+d}{d}}] \otimes_{\boldsymbol{k}} \boldsymbol{k}[x_0, x_1, \ldots, x_r].$$

Given a closed point $c \in \mathscr{A}_{k,d}$ we can specialize $F_0, F_1, \ldots, F_k$ and obtain polynomials $f_0, f_1, \ldots, f_k \in \kappa(c)[x_0, x_1, \ldots, x_r]$, where $\kappa(c)$ is the residue field of c. We will thus identify each element of $\mathscr{A}_{k,d}(\boldsymbol{k})$ with a collection of polynomials $\boldsymbol{f} = (f_0, f_1, \ldots, f_k) \in \boldsymbol{k}[x_0, x_1, \ldots, x_r]$.

Now define $\Sigma_{k,d}(X) \subseteq X \times \mathscr{A}_{k,d}$ via the equations $F_0, F_1, \ldots, F_k$. Consider the second projection $p_2 \colon \Sigma_{k,d}(X) \to \mathscr{A}_{k,d}$. Given a point $\boldsymbol{f} = (f_0, f_1, \ldots, f_k) \in \mathscr{A}_{k,d}$, the fiber $p_2^{-1}(\boldsymbol{f}) \subseteq X$ can be identified with the points lying in $X \cap \mathbb{V}(f_0, f_1, \ldots, f_k)$. For generic choices of $\boldsymbol{f}$ (after passing to an infinite field if necessary) the polynomials $f_0, f_1, \ldots, f_k$ will define an ideal of codimension $k+1$, and thus the fiber $p_2^{-1}(\boldsymbol{f})$ will have dimension $n-k-1$.

There is a closed sublocus $\mathscr{D}_{k,d}(X) \subsetneq \mathscr{A}_{k,d}$ where the dimension of the fiber is at least $n-k$, and we give $\mathscr{D}_{k,d}(X)$ the reduced scheme structure. It follows that $\mathscr{D}_{k,d}(X)$ parametrizes collections $\boldsymbol{f} = (f_0, f_1, \ldots, f_k)$ of degree d polynomials which fail to be parameters on X.

Remark 4.1. If we fix $X_{\mathbb{Z}} \subseteq \mathbb{P}^r_{\mathbb{Z}}$, then we can follow the same construction to obtain a scheme $\mathscr{D}_{k,d}(X_{\mathbb{Z}}) \subseteq \mathscr{A}_{k,d}$. Writing $X_{\boldsymbol{k}}$ as the pullback $X \times_{\operatorname{Spec} \mathbb{Z}} \operatorname{Spec} \boldsymbol{k}$, we observe that the equations defining $\Sigma_{k,d}(X_{\boldsymbol{k}})$ are obtained by pulling back the equations defining $\Sigma_{k,d}(X_{\mathbb{Z}})$. It follows that $\mathscr{D}_{k,d}(X_{\mathbb{Z}}) \times_{\operatorname{Spec} \mathbb{Z}} \operatorname{Spec}(\boldsymbol{k})$ has the same set-theoretic support as $\mathscr{D}_{k,d}(X_{\boldsymbol{k}})$.

Definition 4.2. We let $\mathscr{D}^{\mathrm{bad}}_{k,d}(X)$ be the locus of points in $\mathscr{D}_{k,d}(X)$ where $f_0, f_1, \ldots, f_{k-1}$ already fail to be parameters on X and let $\mathscr{D}^{\mathrm{good}}_{k,d}(X) := \mathscr{D}_{k,d}(X) \setminus \mathscr{D}^{\mathrm{bad}}_{k,d}(X)$. We set $\mathscr{D}^{\mathrm{bad}}_{0,d}(X) = \varnothing$.

Remark 4.3. We have a factorization:

$$\mathscr{A}_{k,d} \to \mathscr{A}_{k-1,d} \times \mathscr{A}_{0,d}$$
$$(f_0, f_1, \ldots, f_k) \mapsto ((f_0, f_1, \ldots, f_{k-1}), f_k).$$

We let $\pi : \mathscr{D}_{k,d}(X) \to \mathscr{A}_{k-1,d}$ be the induced projection, which will we use to work inductively.

Proof of Theorem 1.2. First consider the case $k = n$. There is a natural rational map from $\mathscr{A}_{n,d}$ to the Grassmanian $\operatorname{Gr}(n+1, S_d)$ given by sending the point $(f_0, f_1, \ldots, f_n) \in \mathscr{A}_{n,d}$ to the linear space that those polynomials span. Inside of the Grassmanian, the locus of choices of $(f_0, f_1, \ldots, f_n)$ that all vanish on a point of X is a divisor in the Grassmanian defined by the Chow form; see [Gelfand et al. 1994, 3.2.B]. The preimage of this hypersurface in $\mathscr{A}_{n,d}$ is a hypersurface contained in $\mathscr{D}_{n,d}(X)$, and thus $\mathscr{D}_{n,d}(X)$ has codimension 1.

For $k < n$, we will induct on k. Let $k = 0$. A polynomial f_0 will fail to be a parameter on X if and only if $\dim X = \dim(X \cap \mathbb{V}(f_0))$. This happens if and only if f_0 is a zero divisor on a top-dimensional component of X. Let V be the reduced subscheme of some top-dimensional irreducible component of X and let $\mathcal{I}_V$ be the defining ideal sheaf of V. Then the set of zero divisors of degree d on V will form a linear subspace in $\mathscr{A}_{0,d}$ corresponding to the elements of the vector subspace $H^0(\mathcal{I}_V(d))$. The codimension of $H^0(\mathcal{I}_V(d)) \subseteq S_d$ is precisely given by the Hilbert function of the homogeneous coordinate ring of V in degree d. By applying Lemma 3.1(1), we conclude that for all d this linear space has codimension at least $\binom{n+d}{d}$. Since $\mathscr{D}_{0,d}(X)$ is the union of these linear spaces over all top-dimensional components of X, this proves that $\operatorname{codim} \mathscr{D}_{0,d}(X) \geq \binom{n+d}{d}$.

Take the induction hypothesis that we have proven the statement for $\mathscr{D}_{j,d}(X')$ for all $X' \subseteq \mathbb{P}^r$ and all $j \leq k-1$. We separate $\mathscr{D}_{k,d}(X) = \mathscr{D}_{k,d}^{\text{bad}}(X) \sqcup \mathscr{D}_{k,d}^{\text{good}}(X)$ and will show that each locus has sufficiently large codimension. We begin with $\mathscr{D}_{k,d}^{\text{bad}}(X)$. By using the factorization from Remark 4.3, we can realize $\mathscr{D}_{k,d}^{\text{bad}}(X) \subseteq \mathscr{A}_{k,d} \cong \mathscr{A}_{k-1,d} \times \mathscr{A}_{0,d}$. By definition of $\mathscr{D}_{k,d}^{\text{bad}}(X)$, the image of $\mathscr{D}_{k,d}^{\text{bad}}(X)$ in $\mathscr{A}_{k-1,d} \times \mathscr{A}_{0,d}$ is $\mathscr{D}_{k-1,d}(X) \times \mathscr{A}_{0,d}$. It follows that

$$\operatorname{codim}(\mathscr{D}_{k,d}^{\text{bad}}(X), \mathscr{A}_{k,d}) = \operatorname{codim}(\mathscr{D}_{k-1,d}(X), \mathscr{A}_{k-1,d}) \geq \binom{n-k+1+d}{n-k+1} \geq \binom{n-k+d}{n-k},$$

where the middle inequality follows by induction.

Now consider an arbitrary point $\boldsymbol{f} = (f_0, f_1, \ldots, f_k)$ in $\mathscr{D}_{k,d}^{\text{good}}(X)$. By definition, $f_0, f_1, \ldots, f_{k-1}$ are parameters on X, and thus $\pi(\boldsymbol{f}) \in \mathscr{A}_{k-1,d} \setminus \mathscr{D}_{k-1,d}(X)$. Using the splitting of Remark 4.3, the fiber of $\mathscr{D}_{k,d}^{\text{good}}(X)$ over $\boldsymbol{f}$ can be identified with $\mathscr{D}_{0,d}(X')$ where $X' := X \cap \mathbb{V}(f_0, f_1, \ldots, f_{k-1})$. Since $(f_0, f_1, \ldots, f_{k-1}) \notin \mathscr{D}_{k-1,d}(X)$, we have that $\dim X' = n-k$. The inductive hypothesis thus guarantees that $\operatorname{codim} \mathscr{D}_{0,d}(X') \geq \binom{\dim X'+d}{d} = \binom{n-k+d}{d}$. □

5. Probabilistic analysis, I: Proof of Theorem 1.3

The main result of this section is Proposition 5.1 which provides an effective bound for finding parameters, and which we will use to prove Theorem 1.5. We also use this to give a new proof of Theorem 1.3 for $k < n$. Throughout this section, we let $X \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ be a projective scheme of dimension n over a finite field $\mathbb{F}_q$. Recall that $S_d = H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(d))$. We define

$$\operatorname{Par}_{d,k} = \{f_0, f_1, \ldots, f_k \text{ that are parameters on } X\} \subset S_d^{k+1}.$$

In Theorem 1.3, we compute the following limit (which a priori might not exist):

$$\lim_{d\to\infty} \operatorname{Prob}(f_0, f_1, \ldots, f_k \text{ of degree } d \text{ are parameters on } X) := \lim_{d\to\infty} \frac{\#\operatorname{Par}_{d,k}}{\#S_d^{k+1}}.$$

Proposition 5.1. *If $k < n$ then*

$$\operatorname{Prob}(f_0, f_1, \ldots, f_k \text{ of degree } d \text{ are parameters on } X) \geq 1 - \widehat{\deg}(X)(1+d+d^2+\cdots+d^k)q^{-\binom{n-k+d}{n-k}}.$$

Proof. We induct on k and largely follow the structure of the proof of Theorem 1.2. First, let $k=0$. A polynomial f_0 will fail to be a parameter on X if and only if it is a zero divisor on a top-dimensional component V of X. There are at most $\widehat{\deg}(X)$ many such components. As argued in the proof of Theorem 1.2, the set of zero divisors on V corresponds to the elements of $H^0(\mathbb{P}^r, \mathcal{I}_V(d))$ which has codimension at least $\binom{n+d}{d}$ in S_d. It follows that

$$\operatorname{Prob}(f_0 \text{ of degree } d \text{ is not a parameter on } X) \leq \widehat{\deg}(X)q^{-\binom{n+d}{d}}.$$

Now consider the induction step. We will separately compute the probability that $\boldsymbol{f} = (f_0, f_1, \ldots, f_k)$ lies in $\mathscr{D}_{k,d}^{\text{bad}}(X)$ and the probability that $\boldsymbol{f}$ lies in $\mathscr{D}_{k,d}^{\text{good}}(X)$. By definition, the projection π maps $\mathscr{D}_{k,d}^{\text{bad}}(X)$

onto $\mathscr{D}_{k-1,d}(X)$, and by induction

$$\begin{aligned}\operatorname{Prob}(\pi(\boldsymbol{f}) \in \mathscr{P}_{k-1,d}(X)(\mathbb{F}_q)) &\le \widehat{\deg}(X)(1+d+d^2+\cdots+d^{k-1})q^{-\binom{n-k+1+d}{n-k+1}} \\ &\le \widehat{\deg}(X)(1+d+d^2+\cdots+d^{k-1})q^{-\binom{n-k+d}{n-k}}.\end{aligned}$$

We now assume $\boldsymbol{f} \notin \mathscr{D}^{\text{bad}}_{k,d}(X)$. We thus have that $f_0, f_1, \ldots, f_{k-1}$ are parameters on X. As in the proof of Theorem 1.2, the fiber $\pi^{-1}(\boldsymbol{f})$ can be identified with $\mathscr{D}_{0,d}(X')$ where $X' := X \cap \mathbb{V}(f_0, f_1, \ldots, f_{k-1})$. By construction $\dim X' = n-k$ and by Lemma 2.3, $\widehat{\deg}(X') \le \widehat{\deg}(X) \cdot d^k$. Our inductive hypothesis thus implies that

$$\operatorname{Prob}\begin{pmatrix} (f_0, f_1, \ldots, f_k) \in \mathscr{D}_{k,d}(X)(\mathbb{F}_q) \text{ given that} \\ (f_0, f_1, \ldots, f_{k-1}) \notin \mathscr{D}_{k-1,d}(X)(\mathbb{F}_q) \end{pmatrix} \le \widehat{\deg}(X')q^{-\binom{n-k+d}{n-k}} \le \widehat{\deg}(X)\cdot d^k q^{-\binom{n-k+d}{n-k}}.$$

Combining the estimates for $\mathscr{D}^{\text{bad}}_{k,d}(X)$ and $\mathscr{D}^{\text{good}}_{k,d}(X)$ yields the proposition. □

Proof of Theorem 1.3. If $k < n$, then we apply Proposition 5.1 to obtain

$$\lim_{d\to\infty} \operatorname{Prob}\begin{pmatrix} f_0, f_1, \ldots, f_k \text{ of degree } d \\ \text{are parameters on } X \end{pmatrix} \ge \lim_{d\to\infty} 1 - \widehat{\deg}(X)(d^0+d^1+\cdots+d^k)q^{-\binom{n-k+d}{n-k}} = 1.$$

Now let $k = n$. For completeness, we summarize the proof of [Bucur and Kedlaya 2012, Theorem 1.2]. We fix e, which will go to ∞, and separate the argument into low, medium, and high degree cases.

Low degree argument. For a zero dimensional subscheme Y, we have that S_d surjects on $H^0(Y, \mathcal{O}_Y(d))$ when $d \ge \deg Y - 1$ [Poonen 2004, Lemma 2.1]. So if $d > \deg P - 1$, the probability that $f_0, f_1, \ldots, f_n$ all vanish at a closed point $P \in X$ is $1 - q^{-(n+1)\deg P}$. If $Y \subseteq X$ is the union of all points of degree $\le e$, and if $d \ge \deg Y - 1$, then the surjection onto $H^0(Y, \mathcal{O}_Y(d))$ implies that the probabilities at the points $P \in Y$ behave independently. This yields:

$$\operatorname{Prob}\begin{pmatrix} f_0, f_1, \ldots, f_n \text{ of degree } d \text{ are parameters on } X \\ \text{at all points } P \in X \text{ where } \deg(P) \le e \end{pmatrix} = \prod_{\substack{P\in X \\ \deg(P)\le e}} 1 - q^{-(n+1)\deg P}.$$

Medium degree argument. Our argument is nearly identical to [Poonen 2004, Lemma 2.4], and covers all points whose degree lies in the range $\left[e+1, \frac{d}{n+1}\right]$. For any such point $P \in X$, S_d surjects onto $H^0(P, \mathcal{O}_P(d))$ and thus the probability that $f_0, f_1, \ldots, f_n$ all vanish at P is $q^{-\ell(n+1)}$. By [Lang and Weil 1954], $\#X(\mathbb{F}_{q^\ell}) \le Kq^{\ell n}$ for some constant K independent of ℓ. We have

$$\begin{aligned}\operatorname{Prob}\begin{pmatrix} f_0, f_1, \ldots, f_n \text{ of degree } d \text{ all vanish} \\ \text{at some } P \in X \text{ where } e < \deg(P) \le \left\lfloor \frac{d}{n+1} \right\rfloor \end{pmatrix} &\le \sum_{\ell=e+1}^{\left\lfloor \frac{d}{n+1}\right\rfloor} \#X(\mathbb{F}_{q^\ell})q^{-\ell(n+1)} \\ &\le \sum_{\ell=e+1}^{\infty} Kq^{\ell n}q^{-(n+1)\ell} \\ &= \frac{Kq^{-e-1}}{1-q^{-1}}.\end{aligned}$$

This tends to 0 as $e \to \infty$, and therefore does not contribute to the asymptotic limit.

High degree argument. By the case when $k = n-1$, we may assume that $f_0, f_1, \dots, f_{n-1}$ form a system of parameters with probability $1 - o(1)$. So we let V be one of the irreducible components of this intersection (over $\mathbb{F}_q$) and we let R be its homogeneous coordinate ring. If $\deg V \leq \frac{d}{n+1}$, then it can be ignored as we considered such points in the low and medium degree cases. Hence, we can assume $\deg V > \frac{d}{n+1}$. Since $\dim R_\ell \geq \min\{\ell+1, \deg R\}$ for all ℓ, the probability that f_n vanishes along V is at most $q^{-\lfloor d/(n+1)\rfloor - 1}$. Hence the probability of vanishing on some high degree point is bounded by $O(d^n q^{-\lfloor d/(n+1)\rfloor -1})$ which is $o(1)$ as $d \to \infty$.

Combining the various parts as $e \to \infty$, we see that the low degree argument converges to $\zeta_X(n+1)^{-1}$ and the contributions from the medium and high degree points go to 0. $\square$

Remark 5.2. It might be interesting to consider variants of Theorem 1.3 that allow imposing conditions along closed subschemes, similar to Poonen's Bertini with Taylor coefficients [Poonen 2004, Theorem 1.2]. For instance, [Kedlaya 2005, Theorem 1] might be provable by such an approach, though this would be more complicated than the original proof.

Proposition 5.1 yields an effective bound on the degree of a full system of parameters over a finite field. Sharper bounds can be obtained if one allows the f_i to have different degrees.

Corollary 5.3. (1) *If d_1 satisfies $d_1^{n-1} q^{-d_1-1} < (n \cdot \widehat{\deg}(X))^{-1}$, then there exist $g_0, g_1, \dots, g_{n-1}$ of degree d_1 that are parameters on X.*

(2) *Let X' be 0-dimensional. If $\max\{d_2+1, q\} \geq \widehat{\deg}(X')$ then there exists a degree d_2 parameter on X'.*

Proof. Applying Proposition 5.1 in the case $k = n-1$ yields (1). For (2), let f be a random degree d polynomial and let $P \in X'$ be a closed point. Since the dimension of the image of S_d in $H^0(P, \mathcal{O}_P(d))$ is at least $\min\{d+1, \deg P\}$, the probability that f vanishes at P is at worst $q^{-\min\{d+1, \deg P\}}$ which is at least q^{-1}. It follows that the probability that a degree d function vanishes on some point of X' is at worst $\sum_{P \in X'} q^{-1} \leq \widehat{\deg}(X') q^{-1}$. Thus if $q > \widehat{\deg}(X')$, this happens with probability strictly less than 1. On the other hand, if $d+1 \geq \widehat{\deg}(X')$ then polynomials of degree d surject onto $H^0(X', \mathcal{O}_{X'}(d))$ and hence we can find a parameter on X' by choosing a polynomial that restricts to a unit on X'. $\square$

Proof of Theorem 1.5. If $\dim X = 0$, then we can directly apply Corollary 5.3(2) to find a parameter of degree d. So we assume $n := \dim X > 0$. Since $d > \log_q \widehat{\deg}(X) + \log_q n + n \log_q d$ it follows that $(n \cdot \widehat{\deg}(X))^{-1} > q^{-d} d^n > q^{-d-1} d^{n-1}$. Applying Corollary 5.3(1), we find $g_0, g_1, \dots, g_{n-1}$ in degree d that are parameters on X. Let $X' = X \cap V(g_0, g_1, \dots, g_{n-1})$. Since $\max\{d, \frac{q}{d^n}\} \geq \widehat{\deg}(X)$ it follows that $\max\{d^{n+1}, q\} \geq d^n \widehat{\deg}(X) \geq \widehat{\deg}(X')$, and Corollary 5.3(2) yields a parameter g_n of degree d^{n+1} on X'. Thus $g_0^{d^n}, g_1^{d^n}, \dots, g_{n-1}^{d^n}, g_n$ are parameters of degree d^{n+1} on X. $\square$

6. Probabilistic analysis, II: The error term and proof of Theorem 1.4

In this section, we let $k < n$ and we analyze the error terms in Theorem 1.3 more precisely. In particular, we prove Theorem 1.4, which shows that the probabilities are controlled by the probability of vanishing along an $(n-k)$-dimensional subvariety, with varieties of lowest degree contributing the most.

Our proof of Theorem 1.4 adapts Poonen's sieve in a couple of key ways. The first big difference is that instead of sieving over closed points, we will sieve over $(n-k)$-dimensional subvarieties of X; this is because polynomials $f_0, f_1, \dots, f_k$ will fail to be parameters on X only if they all vanish along some $(n-k)$-dimensional subvariety.

The second difference is that the resulting probability formula will not be a product of local factors. This is because the values of a function can never be totally independent along two higher dimensional varieties with a nontrivial intersection. For instance, Lemma 6.1 shows that the probability that a degree d polynomial vanishes along a line is $q^{-(d+1)}$, but the probability of vanishing along two lines that intersect in a point is $q^{-(2d+1)} > (q^{-(d+1)})^2$.

The following result characterizes the individual probabilities arising in our sieve.

Lemma 6.1. *If $Z \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ is a reduced, projective scheme over a finite field $\mathbb{F}_q$ with homogeneous coordinate ring R then*

$$\operatorname{Prob}(f_0, f_1, \dots, f_k \text{ of degree } d \text{ all vanish along } Z) = \left(\frac{1}{\# R_d}\right)^{k+1}.$$

If d is at least the Castelnuovo–Mumford regularity of the ideal sheaf of Z, then

$$\operatorname{Prob}(f_0, f_1, \dots, f_k \text{ of degree } d \text{ all vanish along } Z) = q^{-(k+1)h^0(Z, \mathcal{O}_Z(d))}.$$

Proof. Let $I \subseteq S$ be the homogeneous ideal defining Z, so that $R = S/I$. An element $h \in S_d$ vanishes along Z if and only if it restricts to 0 in R_d i.e., if and only if it lies in I_d. Since we have an exact sequence of $\mathbb{F}_q$-vector spaces:

$$0 \to I_d \to S_d \to R_d \to 0$$

we obtain

$$\operatorname{Prob}(h \text{ vanishes on } Z) = \frac{\# I_d}{\# S_d} = \frac{1}{\# R_d}.$$

For $k+1$ elements of S_d, the probabilities of vanishing along Z are independent and this yields the first statement of the lemma.

We write $\tilde{I}$ for the ideal sheaf of Z. If d is at least the regularity of $\tilde{I}$, then $H^1(\mathbb{P}^r_{\mathbb{F}_q}, \tilde{I}(d)) = 0$. Hence there is a natural isomorphism between R_d and $H^0(Z, \mathcal{O}_Z(d))$. Thus, we have

$$\frac{1}{\# R_d} = q^{-h^0(Z, \mathcal{O}_Z(d))},$$

yielding the second statement. $\square$

Proof of Theorem 1.4. Throughout the proof, we set $\epsilon_{e,k}$ to be the error term for a given e and k, namely $\epsilon_{e,k} := q^{-e(k+1)\binom{n-k+d}{n-k}}$. We also set:

$$\mathrm{Par}_{d,k} := \{f_0, f_1, \ldots, f_k \text{ are parameters on } X\}$$
$$\mathrm{Low}_{d,k,e} := \left\{ \begin{array}{l} f_0, f_1, \ldots, f_k \text{ all vanish along a variety } Z \\ \text{where } \dim Z = (n-k) \text{ and } \deg(Z) \leq e \end{array} \right\}$$
$$\mathrm{Med}_{d,k,e} := \left\{ \begin{array}{l} (f_0, f_1, \ldots, f_k) \notin \mathrm{Low}_{d,k,e} \text{ which all vanish along a variety } Z \\ \text{where } \dim Z = (n-k) \text{ and } e < \deg(Z) \leq e(k+1) \end{array} \right\}$$
$$\mathrm{High}_{d,k,e} := \left\{ \begin{array}{l} (f_0, f_1, \ldots, f_k) \notin \mathrm{Low}_{d,ke} \cup \mathrm{Med}_{d,k,e} \text{ which all vanish along} \\ \text{a variety } Z \text{ where } \dim Z = (n-k) \text{ and } e(k+1) < \deg(Z) \end{array} \right\}.$$

Note that if $f_0, f_1, \ldots, f_k$ all vanish along a variety of dimension $> n-k$ then they will also all vanish along a high degree variety, and hence we do not need to count this case separately. For $\boldsymbol{f} = f_0, f_1, \ldots, f_k \in S_d^{k+1}$, we thus have

$$\begin{aligned} \mathrm{Prob}(\boldsymbol{f} \in \mathrm{Par}_{d,k}) &= 1 - \mathrm{Prob}(\boldsymbol{f} \in \mathrm{Low}_{d,k,e} \cup \mathrm{Med}_{d,k,e} \cup \mathrm{High}_{d,k,e}) \\ &= 1 - \mathrm{Prob}(\boldsymbol{f} \in \mathrm{Low}_{d,k,e}) - \mathrm{Prob}(\boldsymbol{f} \in \mathrm{Med}_{d,k,e}) - \mathrm{Prob}(\boldsymbol{f} \in \mathrm{High}_{d,k,e}). \end{aligned}$$

It thus suffices to show that

$$\mathrm{Prob}(\boldsymbol{f} \in \mathrm{Low}_{d,k,e}) = \sum_{\substack{Z \subseteq X \text{reduced} \\ \dim Z \equiv n-k \\ \deg Z \leq e}} (-1)^{|Z|-1} q^{-(k+1)h^0(Z, \mathcal{O}_Z(d))} + o(\epsilon_{e,k})$$

and that $\mathrm{Prob}(\boldsymbol{f} \in \mathrm{Med}_{d,k,e})$ and $\mathrm{Prob}(\boldsymbol{f} \in \mathrm{High}_{d,k,e})$ are each in $o(\epsilon_{e,k})$.

We proceed by induction on k. When $k = 0$ the condition that f_0 is a parameter on X is equivalent to f_0 not vanishing along a top-dimensional component of X. Thus, combining Lemma 6.1 with an inclusion/exclusion argument implies the exact result:

$$\mathrm{Prob}(f_0 \in \mathrm{Par}_{d,0}) = 1 - \sum_{\substack{Z \subseteq X \text{reduced} \\ \dim Z \equiv n-k}} (-1)^{|Z|-1} q^{-h^0(Z, \mathcal{O}_Z(d))}.$$

By basic properties of the Hilbert polynomial, as $d \to \infty$ we have

$$h^0(Z, \mathcal{O}_Z(d)) = \frac{\deg(Z)}{n!} d^n + o(d^n) = \deg(Z) \binom{n+d}{d} + o(d^n).$$

Hence for the fixed degree bound e, we obtain

$$\begin{aligned}\operatorname{Prob}(f \in \operatorname{Par}_{d,0}) &= 1 - \sum_{\substack{Z\subseteq X\text{reduced}\\ \dim Z\equiv n-k\\ \deg Z\le e}} (-1)^{|Z|-1} q^{-h^0(Z,\mathcal{O}_Z(d))} - \sum_{\substack{Z\subseteq X\text{reduced}\\ \dim Z\equiv n-k\\ \deg Z> e}} (-1)^{|Z|-1} q^{-h^0(Z,\mathcal{O}_Z(d))}\\ &= 1 - \sum_{\substack{Z\subseteq X\text{reduced}\\ \dim Z\equiv n-k\\ \deg Z\le e}} (-1)^{|Z|-1} q^{-h^0(Z,\mathcal{O}_Z(d))} + o(\epsilon_{e,0}).\end{aligned}$$

We now consider the induction step. Let $\boldsymbol{f} = (f_0, f_1, \dots, f_k)$ drawn randomly from S_d^{k+1}. Here we separate into low, medium, and high degree cases.

Low degree argument. Let $\boldsymbol{V}_{k,e}$ denote the set of integral projective varieties $V \subseteq X$ of dimension $n-k$ and degree $\le e$. We have $\boldsymbol{f} \in \operatorname{Low}_{d,k,e}$ if and only if $\boldsymbol{f}$ vanishes on some $V \in \boldsymbol{V}_{k,e}$. Since $\boldsymbol{V}_{k,e}$ is a finite set, we may use an inclusion-exclusion argument to get

$$\operatorname{Prob}(\boldsymbol{f} \in \operatorname{Low}_{d,k,e}) = \sum_{\substack{Z\subseteq X \text{ a union of}\\ V\in \boldsymbol{V}_{k,e}}} (-1)^{|Z|-1} \operatorname{Prob}(f_0, f_1, \dots, f_k \text{ of degree } d \text{ all vanish along } Z).$$

If $\deg Z > e$ then Lemma 6.1 implies that those terms can be absorbed into the error term $o(\epsilon_{e,k})$. Moreover, assuming that Z is a union of $V \in \boldsymbol{V}_{k,e}$ satisfying $\deg(Z) \le e$ is equivalent to assuming Z is reduced and equidimensional of dimensional $n-k$. We thus have

$$= \sum_{\substack{Z\subseteq X \text{ reduced}\\ \dim Z\equiv n-k\\ \deg Z\le e}} (-1)^{|Z|-1} \operatorname{Prob}(f_0, f_1, \dots, f_k \text{ of degree } d \text{ all vanish along } Z) + o(\epsilon_{e,k}).$$

Medium degree argument. We know that $\operatorname{Prob}(\boldsymbol{f} \in \operatorname{Med}_{d,k,e})$ is bounded by the sum of the probabilities that f vanishes along some irreducible variety V in $\boldsymbol{V}_{k,e(k+1)} \setminus \boldsymbol{V}_{k,e}$.

$$\operatorname{Prob}(\boldsymbol{f} \in \operatorname{Med}_{d,k,e}) \le \sum_{Z\in \boldsymbol{V}_{k,e(k+1)}\setminus \boldsymbol{V}_{k,e}} \operatorname{Prob}(f_0, f_1, \dots, f_k \text{ of degree } d \text{ all vanish along } Z).$$

Lemma 6.1 implies that each summand on the right-hand side lies in $o(\epsilon_{e,k})$. This sum is finite and thus $\operatorname{Prob}(\boldsymbol{f} \in \operatorname{Med}_{d,k,e})$ is in $o(\epsilon_{e,k})$.

High degree argument. Proposition 5.1 implies that $f_0, f_1, \dots, f_{k-1}$ are parameters on X with probability $1 - o(q^{-\binom{n-k+1+d}{d}}) \ge 1 - o(\epsilon_{e,k})$ for any e. Hence we may restrict our attention to the case where $f_0, f_1, \dots, f_{k-1}$ are parameters on X.

Let $V_1, V_2, \dots, V_s$ be the irreducible components of $X' := X \cap \mathbb{V}(f_0, f_1, \dots, f_{k-1})$ that have dimension $n-k$. We have that $f_0, f_1 \dots, f_k$ fail to be parameters on X if and only if f_k vanishes on some V_i. We can assume that f_k does not vanish on any V_i where $\deg V_i \le e(k+1)$ as we have already accounted for this possibility in the low and medium degree cases. After possibly relabeling the components, we let $V_1, V_2, \dots, V_t$ be the components of degree $> e(k+1)$ and $X'' = V_1 \cup V_2 \cup \dots \cup V_t$. Using Lemma 2.3, we compute $\widehat{\deg}(X'') \le \widehat{\deg}(X') = \widehat{\deg}(X) \cdot d^k$. It follows that X'' has at most $\widehat{\deg}(X) d^k/(e(k+1))$ irreducible components.

Now for the key point: since the value of d is not necessarily larger than the Castelnuovo–Mumford regularity of V_i, we cannot use a Hilbert polynomial computation to bound the probability that f_k vanishes along V_i. Instead, we use the lower bound for Hilbert functions obtained in Lemma 3.1. Let $\epsilon = \frac{1}{2}$, though any choice of ϵ would work. We write $R(V_i)$ for the homogeneous coordinate ring of V_i. For any $1 \le i \le t$, Lemmas 3.1 and 6.1 yield

$$\mathrm{Prob}(f_k \text{ of degree } d \text{ vanishes along } V_i) = q^{-\dim R(V_i)_d} \le q^{-(e(k+1)+\epsilon)\binom{n-k+d}{n-k}}$$

whenever $d \ge Ce^{k+1}$. Combining this with our bound on the number of irreducible components of X'' gives $\mathrm{Prob}(\boldsymbol{f} \in \mathrm{High}_{d,k,e}) \le \frac{1}{e(k+1)} \widehat{\deg} X d^k q^{-(e(k+1)+\epsilon)\binom{n-k+d}{n-k}}$ which is in $o(\epsilon_{e,k})$. □

Corollary 6.2. *Let $X \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ be an n-dimensional closed subscheme and let $k < n$. Then*

$$\lim_{d\to\infty} q^{(k+1)\binom{n-k+d}{n-k}} \mathrm{Prob}\begin{pmatrix} f_0, f_1, \ldots, f_k \textit{ of degree } d \\ \textit{are } \underline{\textit{not}} \textit{ parameters on } X \end{pmatrix} = \#\{(n-k)\textit{-planes } L \subseteq \mathbb{P}^r_{\mathbb{F}_q} \textit{ such that } L \subseteq X\}.$$

Proof. Let N denote the number of $(n-k)$-planes $L \subseteq \mathbb{P}^r_{\mathbb{F}_q}$ such that $L \subseteq X$. Choosing $e = 1$ in Theorem 1.4, we compute that

$$\mathrm{Prob}(f_0, f_1, \ldots, f_k \text{ of degree } d \text{ are parameters on } X) = 1 - Nq^{-(k+1)\binom{n-k+d}{n-k}} + o(q^{-(k+1)\binom{n-k+d}{n-k}}).$$

It follows that

$$\mathrm{Prob}(f_0, f_1, \ldots, f_k \text{ of degree } d \text{ are } \underline{\text{not}} \text{ parameters on } X) = Nq^{-(k+1)\binom{n-k+d}{n-k}} + o(q^{-(k+1)\binom{n-k+d}{n-k}}).$$

Dividing both sides by $q^{-(k+1)\binom{n-k+d}{n-k}}$ and taking the limit as $d \to \infty$ yields the corollary. □

7. Passing to $\mathbb{Z}$ and $\mathbb{F}_q[t]$

In this section we prove Corollaries 1.6 and 1.7.

Definition 7.1. Let $B = \mathbb{Z}$ or $\mathbb{F}_q[t]$ and fix a finitely generated, free B-module B^s and a subset $\mathcal{S} \subseteq B^s$. Given $a \in B^s$ we write $a = (a_1, a_2, \ldots, a_s)$. The *density* of $\mathcal{S} \subseteq B^s$ is

$$\mathrm{Density}(\mathcal{S}) := \begin{cases} \lim_{N\to\infty} \frac{\#\{a \in \mathcal{S} \mid \max\{|a_i|\} \le N\}}{\#\{a \in \mathbb{Z}^s \mid \max\{|a_i|\} \le N\}} & \text{if } B = \mathbb{Z}, \\ \lim_{N\to\infty} \frac{\#\{a \in \mathcal{S} \mid \max\{\deg a_i\} \le N\}}{\#\{a \in \mathbb{F}_q[t]^s \mid \max\{\deg a_i\} \le N\}} & \text{if } B = \mathbb{F}_q[t]. \end{cases}$$

Proof of Corollary 1.6. For clarity, we will prove the result over $\mathbb{Z}$ in detail and at the end, mention the necessary adaptations for $\mathbb{F}_q[t]$.

We first let $k < n$. Given degree d polynomials $f_0, f_1, \ldots, f_k$ with integer coefficients and a prime p, let $\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_k$ be the reduction of these polynomials mod p. Then $\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_k$ will be parameters on X_p if and only if the point $\bar{\boldsymbol{f}} = (\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_k)$ lies $\mathscr{D}_{d,k}(X_{\mathbb{F}_p})$. As noted in Remark 4.1, this is equivalent to asking that $\overline{f}$ is an $\mathbb{F}_p$-point of $\mathscr{D}_{k,d}(X_{\mathbb{Z}})$. Thus, we may apply [Ekedahl 1991, Theorem 1.2] to $\mathscr{D}_{d,k}(X_{\mathbb{Z}}) \subseteq \mathscr{A}_{k,d}$ (using $M = 1$) to conclude that

$$\text{Density}\left\{\begin{array}{l} f_0, f_1, \dots, f_k \text{ of degree } d \\ \text{that restrict to parameters on } X_p \text{ for all } p \end{array}\right\} = \prod_p \text{Prob}\left(\begin{array}{l} f_0, f_1, \dots, f_k \text{ of degree } d \\ \text{restrict to parameters on } X_p \end{array}\right).$$

Applying Proposition 5.1 to estimate the individual factors; we have:

$$\begin{aligned}\text{Density}\left\{\begin{array}{l} f_0, f_1, \dots, f_k \text{ of degree } d \text{ that} \\ \text{restrict to parameters on } X_p \text{ for all } p \end{array}\right\} &= \lim_{d\to\infty} \prod_p \text{Prob}\left(\begin{array}{l} f_0, f_1, \dots, f_k \text{ of degree } d \\ \text{restrict to parameters on } X_p \end{array}\right) \\ &\geq \lim_{d\to\infty} \prod_p (1 - \widehat{\deg}(X_p)(1+d+\cdots+d^k)p^{-\binom{n-k+d}{n-k}}).\end{aligned}$$

Lemma 7.2 shows that there is an integer D where $D \geq \widehat{\deg}(X_p)$ for all p. Moreover, $1+d+\cdots+d^k \leq kd^k$ for all d, and hence:

$$\geq \lim_{d\to\infty} \prod_p (1 - Dkd^k p^{-\binom{n-k+d}{n-k}}).$$

For $d \gg 0$ we can make $Dkd^k p^{-\binom{n-k+d}{n-k}} \leq p^{-d/2}$ for all p simultaneously. Using $\zeta(n)$ for the Riemann zeta function, we get:

$$\geq \lim_{d\to\infty} \prod_p (1 - p^{-d/2}) \geq \lim_{d\to\infty} \zeta(d/2)^{-1} = 1.$$

We now consider the case $k = n$. This follows by a "low degree argument" exactly analogous to [Poonen 2004, Theorem 5.13]. Fix a large integer N and let Y be the union of all closed points $P \in X$ whose residue field $\kappa(P)$ has cardinality at most N. Since Y is a finite union of closed points, we see that for $d \gg 0$, there is a surjection

$$H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(d)) \to H^0(Y, \mathcal{O}_Y(d)) \cong \bigoplus_{\substack{P\in X \\ \#\kappa(P)\leq N}} H^0(P, \mathcal{O}_P(d)) \to 0.$$

It follows that we have a product formula

$$\text{Density}\left\{\begin{array}{l} f_0, f_1, \dots, f_n \text{ of degree } d \text{ do not all} \\ \text{vanish on a point } P \text{ with } \#\kappa(P) \leq N \end{array}\right\} = \prod_{P\in X, \#\kappa(P)\leq N} \left(1 - \frac{1}{\#\kappa(P)^{n+1}}\right)$$

This is certainly an upper bound on the density of $f_0, f_1, \dots, f_n$ that are parameters on X_p for all p. As $N \to \infty$ the right-hand side approaches $\zeta_X(n+1)^{-1}$. However, since the dimension of X is $n+1$, this zeta function has a pole at $s = n+1$ [Serre 1965, Theorems 1 and 3(a)]. Hence this asymptotic density equals 0. This completes the proof over $\mathbb{Z}$.

Over $\mathbb{F}_q[t]$, the key adaptation is to use [Poonen 2003, Theorem 3.1] in place of Ekedahl's result. Poonen's result is stated for a pair of polynomials, but it applies equally well to n-tuples of polynomials such as the n-tuples defining $\mathscr{D}_{k,d}(X)$. In particular, one immediately reduces to proving an analogue of

[Poonen 2003, Lemma 5.1], for n-tuples of polynomials which are irreducible over $\mathbb{F}_q(t)$ and which have gcd equal to 1; but the $n = 2$ version of the lemma then implies the $n \geq 2$ versions of the lemma.[2] The rest of our argument over $\mathbb{Z}$ works over $\mathbb{F}_q[t]$. $\square$

Lemma 7.2. *Let $X \subseteq \mathbb{P}^r_B$ be any closed subscheme. There is an integer D where $D \geq \widehat{\deg}(X_s)$ for all $s \in \operatorname{Spec} B$.*

Proof. First we take a flattening stratification for X over B [EGA IV$_4$ 1967, Corollaire 6.9.3]. Within each stratum, the maximal degree of a minimal generator is semicontinuous, and we can thus find a degree e where X_s is generated in degree e for all $s \in \operatorname{Spec} B$. By [Bayer and Mumford 1993, Proposition 3.5], we then obtain that $\widehat{\deg}(X) \leq \sum_{j=0}^{n} e^{r-j}$. In particular defining $D := re^r$ will suffice. $\square$

To prove Corollary 1.7, we use Corollary 1.6 to find a submaximal collection $f_0, f_1, \ldots, f_{n-1}$ which restrict to parameters on X_s for all $s \in \operatorname{Spec} B$. This cuts X down to a scheme $X' = X \cap \mathbb{V}(f_0, f_1, \ldots, f_{n-1})$ with 0-dimensional fibers over each point s. When $B = \mathbb{Z}$, such a scheme is essentially a union of orders in number fields, and we find the last element f_n by applying classical arithmetic results about the Picard groups of rings of integers of number fields. When $B = \mathbb{F}_q[t]$, we use similar facts about Picard groups of affine curves over $\mathbb{F}_q$.

An example illustrates this approach. Let $X = \mathbb{P}^1_{\mathbb{Z}} = \operatorname{Proj}(\mathbb{Z}[x, y])$. A polynomial of degree d will be a parameter on X as long as the $d + 1$ coefficients are relatively prime. Thus as $d \to \infty$, the density of these choices will go to 1. However, once we have fixed one such parameter, say $5x - 3y$, it is much harder to find an element that will restrict to a parameter on $\mathbb{Z}[x, y]/(5x - 3y)$ modulo p for all p. In fact, the only possible choices are the elements which restrict to units on $\operatorname{Proj}(\mathbb{Z}[x, y]/(5x - 3y))$. Among the linear forms, these are

$$\pm(7x - 4y) + c(5x - 3y) \text{ for any } c \in \mathbb{Z}.$$

Hence, these elements arise with density zero, and yet they form a nonempty subset.

Lemmas 7.3 and 7.4 below are well-known to experts, but we sketch the proofs for clarity.

Lemma 7.3. *If $X' \subseteq \mathbb{P}^r_{\mathbb{Z}}$ is closed and finite over* $\operatorname{Spec}(\mathbb{Z})$, *then* $\operatorname{Pic}(X')$ *is finite.*

Proof. We first reduce to the case where X' is reduced. Let $\mathcal{N} \subseteq \mathcal{O}_{X'}$ be the nilradical ideal. If X' is nonreduced then there is some integer $m > 1$ for which $\mathcal{N}^m = 0$. Let X'' be the closed subscheme defined by $\mathcal{N}^{m-1}$. We have a short exact sequence $0 \to \mathcal{N}^{m-1} \to \mathcal{O}^*_{X'} \to \mathcal{O}^*_{X''} \to 1$ where the first map sends $f \mapsto 1 + f$. Since X' is affine and noetherian and $\mathcal{N}^{m-1}$ is a coherent ideal sheaf, we have that $H^1(X', \mathcal{N}^{m-1}) = H^2(X', \mathcal{N}^{m-1}) = 0$ [Hartshorne 1977, Theorem III.3.7]. Taking cohomology of the above sequence thus yields an isomorphism $\operatorname{Pic}(X') \cong \operatorname{Pic}(X'')$. Iterating this argument, we may assume X' is reduced.

We now have $X' = \operatorname{Spec}(B)$ where B is a finite, reduced $\mathbb{Z}$-algebra. If Q is a minimal prime of B, then B/Q is either zero dimensional or an order in a number field, and hence has a finite Picard group [Neukirch 1999, Theorem I.12.12]. If B has more than one minimal prime, then we let Q' be the

[2] We thank Bjorn Poonen for pointing out this reduction.

intersection of all of the minimal primes of B except for Q, and we again have an exact sequence in cohomology

$$\cdots \to (B/(Q+Q'))^* \to \mathrm{Pic}(X') \to \mathrm{Pic}(B/Q) \oplus \mathrm{Pic}(B/Q') \to \cdots$$

Since $(B/(Q+Q'))^*$ is a finite set, and since B/Q and B/Q' have fewer minimal primes than B, we may use induction to conclude that $\mathrm{Pic}(X')$ is finite. □

Lemma 7.4. *If C is an affine curve over $\mathbb{F}_q$, then $\mathrm{Pic}(C)$ is finite.*

Proof. If C fails to be integral, then an argument entirely analogous to the proof of Lemma 7.3 reduces us to the case C is integral. We next assume that C is nonsingular and integral, and that $\bar{C}$ is the corresponding nonsingular projective curve. Since C is affine we have $\mathrm{Pic}(C) = \mathrm{Pic}^0(C) \subseteq \mathrm{Pic}^0(\bar{C}) \cong \mathrm{Jac}(\bar{C})(\mathbb{F}_q)$, the last of which is a finite group. If C is singular, then the finiteness of $\mathrm{Pic}(C)$ follows from the nonsingular case by a minor adaptation of the proof of [Neukirch 1999, Proposition I.12.9]. □

Proof of Corollary 1.7. By Corollary 1.6, for $d \gg 0$ we can find polynomials $f_0, f_1, \dots, f_{n-1}$ of degree d that restrict to parameters on X_s for all $s \in \mathrm{Spec}\, B$. Let $X' := \mathbb{V}(f_0, f_1, \dots, f_{n-1}) \cap X$, which is finite over B by construction. Let A be the finite B-algebra where $\mathrm{Spec}\, A = X'$. Lemma 7.3 or 7.4 implies that $H^0(X', \mathcal{O}_{X'}(e)) = A$ for some e. We can thus find a polynomial f_n of degree e mapping onto a unit in the B-algebra A. It follows that $\mathbb{V}(f_n) \cap X' = \varnothing$. Replace f_i by f_i^e for $i = 0, \dots, n-1$ and replace f_n by f_n^d. Then we have $f_0, f_1, \dots, f_n$ of degree $d' := de$ and restricting to parameters on X_s for all $s \in \mathrm{Spec}(B)$ simultaneously.

We thus obtain a proper morphism $\pi : X \to \mathbb{P}^n_B$ where $X_s \to \mathbb{P}^n_{\kappa(s)}$ is finite for all s. Since π is quasifinite and proper, it is finite by [EGA IV$_3$ 1966, Théorème 8.11.1]. □

The following generalizes Corollary 1.7 to other graded rings.

Corollary 7.5. *Let $B = \mathbb{Z}$ or $\mathbb{F}_q[t]$ and let R be a graded, finite type B-algebra where $\dim R \otimes_{\mathbb{Z}} \mathbb{F}_p = n+1$ for all p. Then there exist $f_0, f_1, \dots, f_n$ of degree d for some d such that $B[f_0, f_1, \dots, f_n] \subseteq R$ is a finite extension.*

Proof. After replacing R by a high degree Veronese subring R', we may assume that R' is generated in degree one and contains no R'_+-torsion submodule, where $R'_+ \subseteq R'$ is the homogeneous ideal of strictly positive degree elements. Let $r+1$ be the number of generators of R'_1. Then there is a surjection $\phi \colon B[x_0, x_1, \dots, x_r] \to R'$ inducing an embedding of $X := \mathrm{Proj}(R') \subseteq \mathbb{P}^r_B$. Since R' contains no R'_+-torsion submodule, the kernel of ϕ will be saturated with respect to $(x_0, x_1, \dots, x_r)$ and hence R' will equal the homogeneous coordinate ring of X. Choosing $f_0, f_1, \dots, f_n$ as in Corollary 1.7, it follows that $B[f_0, f_1, \dots, f_n] \subseteq R'$ is a finite extension, and thus so is $B[f_0, f_1, \dots, f_n] \subseteq R$. □

8. Examples

Example 8.1. By Corollary 6.2, it is more difficult to randomly find parameters on surfaces that contain lots of lines. Consider $\mathbb{V}(xyz) \subset \mathbb{P}^3$ which contains substantially more lines than $\mathbb{V}(x^2+y^2+z^2) \subset \mathbb{P}^3$.

Using [Macaulay2] to select 1,000,000 random pairs (f_0, f_1) of polynomials of degree two, the proportion that failed to be systems of parameters were

	$\mathbb{V}(xyz)$	$\mathbb{V}(x^2+y^2+z^2)$
$\mathbb{F}_2$	.2638	.1179
$\mathbb{F}_3$	.0552	.0059
$\mathbb{F}_5$	.0063	.0004

Example 8.2. Let $X \subseteq \mathbb{P}^3_{\mathbb{F}_q}$ be a smooth cubic surface. Over the algebraic closure X has 27 lines, but it has between 0 and 27 lines defined over $\mathbb{F}_q$. For example, working over $\mathbb{F}_4$, the Fermat cubic surface X' defined by $x^3+y^3+z^3+w^3$ has 27 lines, while the cubic surface X defined by $x^3+y^3+z^3+aw^3$ where $a \in \mathbb{F}_4 \setminus \mathbb{F}_2$ has no lines defined over $\mathbb{F}_4$ [Debarre et al. 2017, Section 3]. It will thus be more difficult to find parameters on X than on X'. Using [Macaulay2] to select 100,000 random pairs (f_0, f_1) of polynomials of degree two, 0.62% failed to be parameters on X whereas no choices whatsoever failed to be parameters on X'. This is in line with the predictions from Corollary 6.2; for instance, in the case of X, we have $27 \cdot 4^{-2\cdot 3} \approx 0.66\%$.

Example 8.3. Let $X = [1:4] \cup [3:5] \cup [4:5] = \mathbb{V}((4x-y)(5x-3y)(5x-4y)) \subseteq \mathbb{P}^1_{\mathbb{Z}}$ and let R be the homogeneous coordinate ring of X. The fibers are 0-dimensional so finding a Noether normalization $X \to \mathbb{P}^0_{\mathbb{Z}}$ is equivalent to finding a single polynomial f_0 that restricts to a unit on each of the points simultaneously. We can find such an f_0 of degree d if and only if the induced map of free $\mathbb{Z}$-modules $\mathbb{Z}[x, y]_d \to R_d$ is surjective. A computation in [Macaulay2] shows that this happens if and only if d is divisible by 60.

Example 8.4. Let $R = \mathbb{Z}[x]/(3x^2 - 5x) \cong \mathbb{Z} \oplus \mathbb{Z}\left[\frac{1}{3}\right]$. This is a flat, finite type $\mathbb{Z}$-algebra where every fiber has dimension 0, yet it is not a finite extension of $\mathbb{Z}$. However, if we take the projective closure of $\mathrm{Spec}(R)$ in $\mathbb{P}^1_{\mathbb{Z}}$, then we get $\mathrm{Proj}(\bar{R})$ where $\bar{R} = \mathbb{Z}[x, y]/(3x^2 - 5xy)$. If we then choose $f_0 := 4x - 7y$, we see that $\mathbb{Z}[f_0] \subseteq \bar{R}$ is a finite extension of graded rings.

Example 8.5. Let $\boldsymbol{k}$ be a field and let $X = [1:1+t] \cup [1-t:1] = \mathbb{V}((y-(1+t)x)(x-(1-t)y)) \subseteq \mathbb{P}^1_{\boldsymbol{k}[t]}$. Let R be the homogeneous coordinate ring of X. In degree d, we have the map $\phi_d : \boldsymbol{k}[t][x, y]_d \cong \boldsymbol{k}[t]^{d+1} \to R_d \cong \boldsymbol{k}[t]^2$. Choosing the standard basis $x^d, x^{d-1}y, \ldots, y^d$ for the source of ϕ_d, and the two points of X for the target, we can represent ϕ_d by the matrix

$$\begin{pmatrix} 1 & 1+t & (1+t)^2 & \cdots & (1+t)^d \\ (1-t)^d & (1-t)^{d-1} & (1-t)^{d-2} & \cdots & 1 \end{pmatrix}.$$

It follows that $\operatorname{im}\phi_d = \operatorname{im}\left(\begin{smallmatrix} t^2 & (1+t)^d \\ 0 & 1 \end{smallmatrix}\right) = \operatorname{im}\left(\begin{smallmatrix} t^2 & 1+dt \\ 0 & 1 \end{smallmatrix}\right)$. The image of ϕ_d thus contains a unit if and only if the characteristic of $\boldsymbol{k}$ is p and $p \mid d$. In particular, if $\boldsymbol{k} = \mathbb{Q}$, then we cannot find a polynomial f_0 inducing a finite map $X \to \mathbb{P}^0_{\mathbb{Q}[t]}$.

Example 8.6. Let $\boldsymbol{k}$ be any field, let $B = \boldsymbol{k}[s, t]$, and let $X = [s:1] \cup [1:t] = \mathbb{V}((x-sy)(y-tx)) \subseteq \mathbb{P}^1_B$. We claim that for any $d > 0$, there does not exist a polynomial that restricts to a parameter on X_b for each point $b \in B$. Assume for contradiction that we had such an $f = \sum_{i=0}^{d} c_i s^i t^{d-i}$ with $c_i \in B$. After scaling, we obtain

$$f([s:1]) = c_0 s^d + c_1 s^{d-1} + \cdots + c_d = 1 \quad \text{and} \quad f([1:t]) = c_0 + c_1 t + \cdots + c_d t^d = \lambda$$

where $\lambda \in B^* = \boldsymbol{k}^*$. Substituting for c_d we obtain

$$f([1:t]) = c_0 + c_1 t + \cdots + c_{d-1} t^{d-1} + (1 - (c_0 s^d + c_1 s^{d-1} + \cdots + c_{d-1} s)) t^d = \lambda,$$

which implies that

$$\begin{aligned} \lambda - t^d &= c_0 + c_1 t + \cdots + c_{d-1} t^{d-1} - (c_0 s^d + c_1 s^{d-1} + \cdots + c_{d-1} s) t^d \\ &= (c_0 - c_0 s^d t^d) + (c_1 t - c_1 s^{d-1} t^d) + \cdots + (c_{d-1} t^{d-1} - c_{d-1} s t^d) \\ &= (1 - st) h(s,t) \end{aligned}$$

where $h(s,t) \in \boldsymbol{k}[s,t]$. This implies that $\lambda - t^d$ is divisible by $(1 - st)$, which is a contradiction.

Acknowledgements

We thank Joe Buhler, Nathan Clement, David Eisenbud, Jordan S. Ellenberg, Benedict Gross, Moisés Herradón Cueto, Craig Huneke, Kiran Kedlaya, Brian Lehmann, Dino Lorenzini, Bjorn Poonen, Anurag Singh, Melanie Matchett Wood, and the anonymous referees for their helpful conversations and comments. The computer algebra system [Macaulay2] provided valuable assistance throughout our work.

References

[Abhyankar and Kravitz 2007] S. S. Abhyankar and B. Kravitz, "Two counterexamples in normalization", *Proc. Amer. Math. Soc.* **135**:11 (2007), 3521–3523. MR Zbl

[Achinger 2015] P. Achinger, "$K(\pi, 1)$-neighborhoods and comparison theorems", *Compos. Math.* **151**:10 (2015), 1945–1964. MR Zbl

[Bayer and Mumford 1993] D. Bayer and D. Mumford, "What can be computed in algebraic geometry?", pp. 1–48 in *Computational algebraic geometry and commutative algebra* (Cortona, Italy, 1991), edited by D. Eisenbud and L. Robbiano, Sympos. Math. **34**, Cambridge Univ. Press, 1993. MR Zbl

[Benoist 2011] O. Benoist, "Le théorème de Bertini en famille", *Bull. Soc. Math. France* **139**:4 (2011), 555–569. MR Zbl

[Brennan and Epstein 2011] J. P. Brennan and N. Epstein, "Noether normalizations, reductions of ideals, and matroids", *Proc. Amer. Math. Soc.* **139**:8 (2011), 2671–2680. MR Zbl

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Stud. Adv. Math. **39**, Cambridge Univ. Press, 1993. MR Zbl

[Bruns and Vetter 1988] W. Bruns and U. Vetter, *Determinantal rings*, Monograf. de Mat. **45**, Inst. Matemática Pura e Aplicada, Rio de Janeiro, 1988. MR Zbl

[Bucur and Kedlaya 2012] A. Bucur and K. S. Kedlaya, "The probability that a complete intersection is smooth", *J. Théor. Nombres Bordeaux* **24**:3 (2012), 541–556. MR Zbl

[Charles 2017] F. Charles, "Arithmetic ampleness and an arithmetic Bertini theorem", preprint, 2017. arXiv

[Charles and Poonen 2016] F. Charles and B. Poonen, "Bertini irreducibility theorems over finite fields", *J. Amer. Math. Soc.* **29**:1 (2016), 81–94. MR Zbl

[Chinburg et al. 2017] T. Chinburg, L. Moret-Bailly, G. Pappas, and M. J. Taylor, "Finite morphisms to projective space and capacity theory", *J. Reine Angew. Math.* **727** (2017), 69–84. MR Zbl

[Debarre et al. 2017] O. Debarre, A. Laface, and X. Roulleau, "Lines on cubic hypersurfaces over finite fields", pp. 19–51 in *Geometry over nonclosed fields*, edited by F. Bogomolov et al., Springer, 2017. MR Zbl

[EGA IV$_3$ 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[EGA IV$_4$ 1967] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl

[Eisenbud 1995] D. Eisenbud, *Commutative algebra: with a view toward algebraic geometry*, Grad. Texts in Math. **150**, Springer, 1995. MR Zbl

[Ekedahl 1991] T. Ekedahl, "An infinite version of the Chinese remainder theorem", *Comment. Math. Univ. St. Paul.* **40**:1 (1991), 53–59. MR Zbl

[Ellenberg and Erman 2016] J. S. Ellenberg and D. Erman, "Furstenberg sets and Furstenberg schemes over finite fields", *Algebra Number Theory* **10**:7 (2016), 1415–1436. MR Zbl

[Ellenberg et al. 2010] J. S. Ellenberg, R. Oberlin, and T. Tao, "The Kakeya set and maximal conjectures for algebraic varieties over finite fields", *Mathematika* **56**:1 (2010), 1–25. MR Zbl

[Fulton 1984] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik (3) **2**, Springer, 1984. MR Zbl

[Gabber et al. 2015] O. Gabber, Q. Liu, and D. Lorenzini, "Hypersurfaces in projective schemes and a moving lemma", *Duke Math. J.* **164**:7 (2015), 1187–1270. MR Zbl

[Gelfand et al. 1994] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, Boston, 1994. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. **52**, Springer, 1977. MR Zbl

[Kedlaya 2005] K. S. Kedlaya, "More étale covers of affine spaces in positive characteristic", *J. Algebraic Geom.* **14**:1 (2005), 187–192. MR Zbl

[Lang and Weil 1954] S. Lang and A. Weil, "Number of points of varieties in finite fields", *Amer. J. Math.* **76** (1954), 819–827. MR Zbl

[Macaulay 1916] F. S. Macaulay, *The algebraic theory of modular systems*, Cambridge Tracts Math. and Math. Phys. **19**, Cambridge Univ. Press, 1916. Zbl

[Macaulay2] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", Available at http://www.math.uiuc.edu/Macaulay2.

[Moh 1979] T. T. Moh, "On a normalization lemma for integers and an application of four colors theorem", *Houston J. Math.* **5**:1 (1979), 119–123. MR Zbl

[Nagata 1962] M. Nagata, *Local rings*, Intersci. Tracts Pure Appl. Math. **13**, Intersci., New York, 1962. MR Zbl

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundlehren der Math. Wissenschaften **322**, Springer, 1999. MR Zbl

[Poonen 2003] B. Poonen, "Squarefree values of multivariable polynomials", *Duke Math. J.* **118**:2 (2003), 353–373. MR Zbl

[Poonen 2004] B. Poonen, "Bertini theorems over finite fields", *Ann. of Math.* (2) **160**:3 (2004), 1099–1127. MR Zbl

[Poonen 2009] B. Poonen et al., "Smooth proper scheme over $\mathbb{Z}$", 2009, Available at https://mathoverflow.net/questions/9576/smooth-proper-scheme-over-z/9605. Discussion on MathOverflow website.

[Poonen 2013] B. Poonen, "Extending self-maps to projective space over finite fields", *Doc. Math.* **18** (2013), 1039–1044. MR Zbl

[Serre 1965] J.-P. Serre, "Zeta and L functions", pp. 82–92 in *Arithmetical algebraic geometry* (West Lafayette, IN, 1963), edited by O. F. G. Schilling, Harper & Row, New York, 1965. MR Zbl

[Zariski and Samuel 1960] O. Zariski and P. Samuel, *Commutative algebra, II*, Van Nostrand, Princeton, 1960. MR Zbl

Communicated by Kiran S. Kedlaya
Received 2018-05-23 Revised 2018-12-18 Accepted 2019-06-27

juliette.bruce@math.wisc.edu Department of Mathematics, University of Wisconsin, Madison, WI, United States

derman@math.wisc.edu Department of Mathematics, University of Wisconsin, Madison, WI, United States

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2103

The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures

Terence Tao and Joni Teräväinen

We study the asymptotic behaviour of higher order correlations

$$\mathbb{E}_{n\leq X/d}\, g_1(n+ah_1)\cdots g_k(n+ah_k)$$

as a function of the parameters a and d, where $g_1,\dots,g_k$ are bounded multiplicative functions, $h_1,\dots,h_k$ are integer shifts, and X is large. Our main structural result asserts, roughly speaking, that such correlations asymptotically vanish for almost all X if $g_1\cdots g_k$ does not (weakly) pretend to be a twisted Dirichlet character $n\mapsto\chi(n)n^{it}$, and behave asymptotically like a multiple of $d^{-it}\chi(a)$ otherwise. This extends our earlier work on the structure of logarithmically averaged correlations, in which the d parameter is averaged out and one can set $t=0$. Among other things, the result enables us to establish special cases of the Chowla and Elliott conjectures for (unweighted) averages at almost all scales; for instance, we establish the k-point Chowla conjecture $\mathbb{E}_{n\leq X}\lambda(n+h_1)\cdots\lambda(n+h_k)=o(1)$ for k odd or equal to 2 for all scales X outside of a set of zero logarithmic density.

1. Introduction

The Chowla and Elliott conjectures. Define a 1-*bounded multiplicative function* to be a function $g:\mathbb{N}\to\mathbb{D}$ from the natural numbers $\mathbb{N}:=\{1,2,\dots\}$ to the unit disk $\mathbb{D}:=\{z\in\mathbb{C}:|z|\leq 1\}$ satisfying $g(nm)=g(n)g(m)$ whenever n,m are coprime. If in addition $g(nm)=g(n)g(m)$ for all $n,m\in\mathbb{N}$, we say that g is *completely multiplicative*. In addition, we adopt the convention that $g(n)=0$ when n is zero or a negative integer.

This paper is concerned with the structure of higher order correlations of such functions. To describe our results, we need some notation for a number of averages.

Definition 1.1 (averaging notation). Let $f:A\to\mathbb{C}$ be a function defined on a nonempty finite set A:

(i) (Unweighted averages) We define

$$\mathbb{E}_{n\in A} f(n):=\frac{\sum_{n\in A} f(n)}{\sum_{n\in A} 1}.$$

MSC2010: primary 11N37; secondary 37A45.
Keywords: multiplicative functions, Chowla's conjecture, Elliott's conjecture.

(ii) (Logarithmic averages) If A is a subset of the natural numbers $\mathbb{N}$, we define

$$\mathbb{E}^{\log}_{n\in A} f(n) := \frac{\sum_{n\in A} f(n)/n}{\sum_{n\in A} 1/n}.$$

(iii) (Doubly logarithmic averages) If A is a subset of the natural numbers $\mathbb{N}$, we define

$$\mathbb{E}^{\log\log}_{n\in A} f(n) := \frac{\sum_{n\in A} f(n)/(n\log(1+n))}{\sum_{n\in A} 1/(n\log(1+n))}.$$

Of course, the symbol n can be replaced here by any other free variable. For any real number $X\ge 1$, we use $\mathbb{E}_{n\le X} f(n)$ as a synonym for $\mathbb{E}_{n\in\mathbb{N}\cap[1,X]} f(n)$, and similarly for $\mathbb{E}^{\log}_{n\le X} f(n)$ and $\mathbb{E}^{\log\log}_{n\le X} f(n)$. If we use the symbol p (or p_1, p_2, etc.) instead of n, we implicitly restrict p to the set of primes $\boldsymbol{P} := \{2, 3, 5, 7, \dots\}$, thus for instance for $X\ge 2$, $\mathbb{E}_{p\le X} f(p)$ is a synonym for $\mathbb{E}_{p\in\boldsymbol{P}\cap[2,X]} f(p)$, and similarly for $\mathbb{E}^{\log}_{p\le X} f(p)$.

Remark 1.2. The use of $\log(1+n)$ in the $\mathbb{E}^{\log\log}$ notation instead of $\log n$ is only in order to avoid irrelevant divergences at $n=1$, and the shift by 1 may otherwise be ignored. Because of the prime number theorem, prime averages such as $\mathbb{E}_{p\le X} f(p)$ are often of "comparable strength" to logarithmic averages $\mathbb{E}^{\log}_{n\le X} f(n)$, and similarly logarithmic prime averages such as $\mathbb{E}^{\log}_{p\le X} f(p)$ are of comparable strength to $\mathbb{E}^{\log\log}_{n\le X} f(n)$. See Lemma 2.6 for a more precise statement.

Following Granville and Soundararajan [2008], given two 1-bounded multiplicative functions $f, g : \mathbb{N}\to\mathbb{D}$, and $X\ge 1$, we define the *pretentious distance* $\mathbb{D}(f, g; X)$ between f and g up to scale X by the formula

$$\mathbb{D}(f, g; X) := \left(\sum_{p\le X} \frac{1-\operatorname{Re}(f(p)\overline{g(p)})}{p}\right)^{1/2}.$$

It is conjectured that multiple correlations of 1-bounded multiplicative functions should asymptotically vanish unless all of the functions involved "pretend" to be twisted Dirichlet characters in the sense of the pretentious distance. More precisely, the following conjecture is essentially due to Elliott.

Conjecture 1.3 (Elliott conjecture). *Let $g_1, \dots, g_k : \mathbb{N}\to\mathbb{D}$ be 1-bounded multiplicative functions for some $k\ge 1$. Assume that there exists $j\in\{1, \dots, k\}$ such that for every Dirichlet character χ one has*

$$\inf_{|t|\le X} \mathbb{D}(g_j, n\mapsto \chi(n)n^{it}; X) \to \infty \tag{1}$$

as $X\to\infty$.

(i) (*Unweighted Elliott conjecture*) *If $h_1, \dots, h_k\in\mathbb{Z}$ are distinct integers, then*

$$\lim_{X\to\infty} \mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k) = 0.$$

(ii) (*Logarithmically averaged Elliott conjecture*) *If $h_1, \dots, h_k\in\mathbb{Z}$ are distinct integers, then*

$$\lim_{X\to\infty} \mathbb{E}^{\log}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k) = 0.$$

Conjecture 1.3(i) was first stated by Elliott [1992; 1994], with condition (1) weakened to the assertion that $\mathbb{D}(g_j, n \mapsto \chi(n)n^{\mathrm{it}}; X) \to \infty$ for each fixed t, with no uniformity in t assumed. However, it was shown in [Matomäki et al. 2015] that this version of the conjecture fails for a technical reason. By summation by parts, Conjecture 1.3(i) implies Conjecture 1.3(ii). At present, both forms of the Elliott conjecture are known for $k = 1$ (thanks to Halász's theorem [1971]), while the $k = 2$ case of the logarithmic Elliott conjecture was established in [Tao 2016]. Specialising the above conjecture to the case of the Liouville function λ,[1] we recover the following conjecture of Chowla [1965], together with its logarithmically averaged form.

Conjecture 1.4 (Chowla conjecture). *Let $k \geq 1$ be a natural number*:

(i) (*Unweighted Chowla conjecture*) *If $h_1, \dots, h_k \in \mathbb{Z}$ are distinct integers*, *then*

$$\lim_{X\to\infty} \mathbb{E}_{n\leq X}\lambda(n+h_1)\cdots\lambda(n+h_k) = 0.$$

(ii) (*Logarithmically averaged Chowla conjecture*) *If $h_1, \dots, h_k \in \mathbb{Z}$ are distinct integers*, *then*

$$\lim_{X\to\infty} \mathbb{E}^{\log}_{n\leq X}\lambda(n+h_1)\cdots\lambda(n+h_k) = 0.$$

Note that for $k = 1$, the unweighted Chowla conjecture is equivalent to the prime number theorem, while the logarithmically averaged 1-point Chowla conjecture has a short elementary proof. No further cases of the unweighted Chowla conjecture are currently known, but the logarithmically averaged Chowla conjecture has been established for $k = 2$ in [Tao 2016] and for all odd values of k in [Tao and Teräväinen 2019] (with a second proof given in [Tao and Teräväinen 2018]). The logarithmically averaged Chowla conjecture is also known to be equivalent to the logarithmically averaged form of a conjecture of Sarnak [2010]; see [Tao 2017a]. See also [Matomäki et al. 2015] for a version of Elliott's conjecture where one averages over the shifts h_i. One can also formulate an analogous version of Chowla's conjecture for the Möbius function, for which very similar results are known.[2]

In [Tao and Teräväinen 2019], we obtained the following special case of the logarithmically averaged Elliott conjecture (Conjecture 1.4(ii)). We say that a 1-bounded multiplicative function $f : \mathbb{N} \to \mathbb{D}$ *weakly pretends* to be another 1-bounded multiplicative function $g : \mathbb{N} \to \mathbb{D}$ if

$$\lim_{X\to\infty} \frac{1}{\log\log X}\mathbb{D}(f, g; X)^2 = 0$$

or equivalently

$$\sum_{p\leq X} \frac{1 - \mathrm{Re}(f(p)\overline{g(p)})}{p} = o(\log\log X).$$

[1] For the definitions of the standard multiplicative functions used in this paper, see page 1.

[2] If one generalises the Chowla conjecture by using affine forms $a_i n + h_i$ instead of shifts $n + h_i$, then a simple sieving argument can be used to show the equivalence of such generalised Chowla conjectures for the Liouville function and their counterparts for the Möbius function; we leave the details to the interested reader.

Theorem 1.5 (special case of logarithmically averaged Elliott [Tao and Teräväinen 2019, Corollary 1.6]). *Let $k \geq 1$, and let $g_1, \ldots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions such that the product $g_1 \cdots g_k$ does not weakly pretend to be any Dirichlet character $n \mapsto \chi(n)$. Then for any integers $h_1, \ldots, h_k$, one has*

$$\lim_{X\to\infty} \mathbb{E}^{\log}_{n\leq X} g_1(n+h_1)\cdots g_k(n+h_k) = 0.$$

In particular this establishes the logarithmically averaged Chowla conjecture for odd values of k. This result was also recently used by Frantzikinakis and Host [2019] to control the Furstenberg measure-preserving systems associated to 1-bounded multiplicative functions, and to establish a version of the logarithmic Sarnak conjecture where the Möbius function $\mu(n)$ is replaced by a 1-bounded multiplicative function $g(n)$ and the topological dynamical system involved is assumed to be uniquely ergodic.

Theorem 1.5 was deduced from a more general structural statement about the correlation sequence $a \mapsto \lim_{X\to\infty} \mathbb{E}^{\log}_{n\leq X} g_1(n+ah_1)\cdots g_k(n+ah_k)$ for 1-bounded multiplicative functions $g_1, \ldots, g_k$, where one now permits the product $g_1 \cdots g_k$ to weakly pretend to be a Dirichlet character. Here one runs into the technical difficulty that the asymptotic limits $\lim_{X\to\infty} \mathbb{E}^{\log}_{n\leq X}$ are not known *a priori* to exist. To get around this difficulty, the device of *generalised limit functionals* was employed.[3] By a generalised limit functional we mean a bounded linear functional $\lim^*_{X\to\infty} : \ell^\infty(\mathbb{N}) \to \mathbb{C}$ which agrees with the ordinary limit functional $\lim_{X\to\infty}$ on convergent sequences, maps nonnegative sequences to nonnegative numbers, and which obeys the bound

$$\Big|\lim^*_{X\to\infty} f(X)\Big| \leq \limsup_{X\to\infty} |f(n)|$$

for all bounded sequences f. As is well known, the existence of such generalised limits follows from the Hahn–Banach theorem. With these notations, we proved in [Tao and Teräväinen 2019, Theorem 1.1] the following:

Theorem 1.6 (structure of logarithmically averaged correlation sequences). *Let $k \geq 1$, and let $h_1, \ldots, h_k$ be integers and $g_1, \ldots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions. Let $\lim^*_{X\to\infty}$ be a generalised limit functional. Let $f : \mathbb{Z} \to \mathbb{D}$ denote the function*

$$f(a) := \lim^*_{X\to\infty} \mathbb{E}^{\log}_{n\leq X} g_1(n+ah_1)\cdots g_k(n+ah_k). \tag{2}$$

(i) *If the product $g_1 \cdots g_k$ does not weakly pretend to be a Dirichlet character, then f is identically zero.*

(ii) *If instead the product $g_1 \cdots g_k$ weakly pretends to be a Dirichlet character χ, then f is the uniform limit of periodic functions F_i, each of which is* χ-isotypic *in the sense that $F_i(ab) = F_i(a)\chi(b)$ whenever a is an integer and b is an integer coprime to the periods of F_i and χ.*

Among other things, Theorem 1.6 yields Theorem 1.5 as a direct corollary. Theorem 1.5 in turn can be used to establish various results about the distribution of consecutive values of 1-bounded multiplicative

[3] Alternatively, one could employ ultrafilter limits, or pass to subsequences in which all limits of interest exist. The latter approach is for instance the one adopted in [Frantzikinakis 2017; Frantzikinakis and Host 2018; 2019].

functions; to give just one example, in [Tao and Teräväinen 2019, Corollary 7.2] it was used to show that every sign pattern in $\{-1,+1\}^3$ occurred with logarithmic density $\frac{1}{8}$ amongst the Liouville sign patterns $(\lambda(n), \lambda(n+1), \lambda(n+2))$.

From logarithmic averages to almost all ordinary averages. It would be desirable if many of the above results for logarithmically averaged correlations such as $\mathbb{E}^{\log}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k)$ could be extended to their unweighted counterparts such as $\mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k)$. However, such extensions cannot be automatic, since for instance the logarithmic averages $\mathbb{E}^{\log}_{n\le X} n^{it}$ converge to 0 for $t\neq 0$, but the unweighted averages $\mathbb{E}_{n\le X} n^{it}$ diverge. Similarly, the statement $\mathbb{E}^{\log}_{n\le X}\lambda(n)=o(1)$ has a short and simple elementary proof,[4] whereas the unweighted analogue $\mathbb{E}_{n\le X}\lambda(n)=o(1)$ is equivalent to the prime number theorem and its proofs are more involved. Moreover, one can show[5] that if, for example, the correlation limit $\lim_{X\to\infty}\mathbb{E}_{n\le X}\lambda(n)\lambda(n+1)$ exists, then it has to be equal to 0, which means that proving the mere existence of the limit captures the difficulty in the two-point unweighted Chowla conjecture.

Nevertheless, there are some partial results of this type in which control on logarithmic averages can be converted to control on unweighted averages for a subsequence of scales X. For instance, in [Gomilko et al. 2018] it is shown using ergodic theory techniques that if the logarithmically averaged Chowla conjecture holds for all k, then there exists an increasing sequence of scales X_i such that the Chowla conjecture for all k holds for X restricted to these scales. This was refined in a blog post of Tao [2017b], where it was shown by an application of the second moment method that if the logarithmically averaged Chowla conjecture held for some even order $2k$, then the Chowla conjecture for order k would hold for all scales X outside of an exceptional set $\mathcal{X}\subset\mathbb{N}$ of logarithmic density zero, by which we mean that

$$\lim_{X\to\infty}\mathbb{E}^{\log}_{n\le X} 1_{\mathcal{X}}(n)=0.$$

Unfortunately, as the only even number for which the logarithmically averaged Chowla conjecture is currently known to hold is $k=2$, this only recovers (for almost all scales) the $k=1$ case of the unweighted Chowla conjecture, which was already known from the prime number theorem.

At present, the restriction to logarithmic averaging in many of the above results is needed largely because it supplies (via the "entropy decrement argument") a certain approximate dilation invariance, which roughly speaking asserts the approximate identity

$$g_1(p)\cdots g_k(p)\mathbb{E}^{\log}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k)\approx \mathbb{E}^{\log}_{n\le X} g_1(n+ph_1)\cdots g_k(n+ph_k)$$

for "most" primes p, and for extremely large values of X; see for instance [Frantzikinakis and Host 2019, Theorem 3.2] for a precise form of this statement, with a proof essentially provided in [Tao and

[4]One can for example prove this by writing $\mathbb{E}^{\log}_{n\le X}\lambda(n)=-\mathbb{E}_{p\le y}\mathbb{E}^{\log}_{n\le X}\lambda(n)p1_{p|n}+o_{y\to\infty}(1)$, and then using the Turán–Kubilius inequality to get rid of the $p1_{p|n}$ factor; we leave the details to the interested reader.

[5]More generally, one can use partial summation to show that, for any bounded real-valued sequence $a:\mathbb{N}\to\mathbb{R}$, if $\lim_{X\to\infty}\mathbb{E}^{\log}_{n\le X}a(n)=\alpha$, then there exists an increasing sequence X_i such that $\lim_{i\to\infty}\mathbb{E}_{n\le X_i}a(n)=\alpha$. In particular, if the logarithmic Elliott conjecture holds, then the ordinary Elliott conjecture also holds in the case of real-valued functions along some subsequence of scales (which may depend on the functions involved).

Teräväinen 2019, Section 3]. However, an inspection of the entropy decrement argument reveals that it also provides an analogous identity for unweighted averages, namely that

$$g_1(p)\cdots g_k(p)\mathbb{E}_{n\le X}g_1(n+h_1)\cdots g_k(n+h_k)\approx \mathbb{E}_{n\le X/p}g_1(n+ph_1)\cdots g_k(n+ph_k) \tag{3}$$

for "most" primes p, and "most" extremely large values of X; see Proposition 2.3 for a precise statement. By using this form of the entropy decrement argument, we are able to obtain the following analogue of Theorem 1.6 for unweighted averages, which is the main technical result of our paper and is proven in Section 2.

Theorem 1.7 (structure of unweighted correlation sequences). *Let $k\ge 1$, and let $h_1,\dots,h_k$ be integers and $g_1,\dots,g_k:\mathbb{N}\to\mathbb{D}$ be 1-bounded multiplicative functions. Let $\lim^*_{X\to\infty}$ be a generalised limit functional. For each real number $d>0$, let $f_d:\mathbb{Z}\to\mathbb{D}$ denote the function*

$$f_d(a):=\lim^*_{X\to\infty}\mathbb{E}_{n\le X/d}g_1(n+ah_1)\cdots g_k(n+ah_k). \tag{4}$$

(i) *If the product $g_1\cdots g_k$ does not weakly pretend to be any twisted Dirichlet character $n\mapsto\chi(n)n^{\mathrm{it}}$, then*

$$\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\le X}|f_d(a)|=0$$

for all integers a.

(ii) *If instead the product $g_1\cdots g_k$ weakly pretends to be a twisted Dirichlet character $n\mapsto\chi(n)n^{\mathrm{it}}$, then there exists a function $f:\mathbb{Z}\to\mathbb{D}$ such that*

$$\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\le X}|f_d(a)-f(a)d^{-\mathrm{it}}|=0 \tag{5}$$

for all integers a. Furthermore, f is the uniform limit of χ-isotypic periodic functions F_i.[6]

We have defined f_d for all real numbers $d>0$ for technical reasons, but we will primarily be interested in the behaviour of f_d for natural numbers d; for instance, the averages $\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\le X}$ appearing in the above theorem are restricted to this case.

Roughly speaking, the logarithmic correlation sequence $f(a)$ appearing in Theorem 1.6 is analogous to the average $\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\le X}f_d(a)$ of the sequences appearing here (ignoring for this discussion the question of whether the limits exist). These averages vanish when $t\ne 0$ in Theorem 1.7, and one basically recovers a form of Theorem 1.6; but, as the simple example of averaging the single 1-bounded multiplicative function $n\mapsto n^{\mathrm{it}}$ already shows, in the $t\ne 0$ case it is possible for the $f_d(a)$ to be nonzero while the logarithmically averaged counterpart $f(a)$ vanishes.

By combining Theorem 1.7 with a simple application of the Hardy–Littlewood maximal inequality, we can obtain several new cases of the unweighted Elliott and Chowla conjectures at almost all scales, as follows.

[6] That is, we have $F_i(ab)=F_i(a)\chi(b)$ for any integers a and b with b coprime to the periods of F_i and χ.

Corollary 1.8 (some cases of the unweighted Elliott conjecture at almost all scales). *Let $k \geq 1$, and let $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions. Suppose that the product $g_1 \cdots g_k$ does not weakly pretend to be any twisted Dirichlet character $n \mapsto \chi(n)n^{it}$.*

(i) *For any $h_1, \dots, h_k \in \mathbb{Z}$ and $\varepsilon > 0$, one has*

$$|\mathbb{E}_{n \leq X} g_1(n+h_1) \cdots g_k(n+h_k)| \leq \varepsilon$$

for all natural numbers X outside of a set $\mathcal{X}_\varepsilon$ of logarithmic Banach density zero, in the sense that

$$\lim_{\omega \to \infty} \sup_{X \geq \omega} \mathbb{E}^{\log}_{X/\omega \leq n \leq X} 1_{\mathcal{X}_\varepsilon}(n) = 0. \tag{6}$$

(ii) *There is a set $\mathcal{X}_0$ of logarithmic density zero, such that*

$$\lim_{X \to \infty; X \notin \mathcal{X}_0} \mathbb{E}_{n \leq X} g_1(n+h_1) \cdots g_k(n+h_k) = 0$$

for all $h_1, \dots, h_k \in \mathbb{Z}$.

Remark 1.9. We note that Corollary 1.8 can be generalised to the case of dilated correlations

$$\mathbb{E}_{n \leq X} g_1(q_1 n + h_1) \cdots g_k(q_k n + h_k),$$

where $q_1, \dots, q_k \in \mathbb{N}$. To see this, one applies exactly the same trick related to Dirichlet character expansions as in [Tao and Teräväinen 2019, Appendix A]. Similarly, Corollary 1.13 below generalises to the dilated case. We leave the details to the interested reader.

Remark 1.10. We see by partial summation that if $f : \mathbb{N} \to \mathbb{C}$ is any bounded function such that for every $\varepsilon > 0$ we have $|\lim_{X \to \infty; X \notin \mathcal{X}_\varepsilon} \mathbb{E}_{n \leq X} f(n)| \leq \varepsilon$ for some set $\mathcal{X}_\varepsilon \subset \mathbb{N}$ of logarithmic Banach density 0, then we also have the logarithmic correlation result $\limsup_{X \to \infty} |\mathbb{E}^{\log}_{X/\omega(X) \leq n \leq X} f(n)| \ll \varepsilon$ for any function $1 \leq \omega(X) \leq X$ tending to infinity. Thus Corollary 1.8 is a strengthening of our earlier result [Tao and Teräväinen 2019, Corollary 1.6] on logarithmic correlation sequences. Similarly, Corollary 1.13 below is a strengthening of [Tao 2016, Corollary 1.5].

Remark 1.11. The logarithmic density (or logarithmic Banach density) appearing in Corollaries 1.8 and 1.13 is the right density to consider in this problem. Namely, if one could show that the set $\mathcal{X}_0$ has *asymptotic* density 0, then $[1, \infty) \setminus \mathcal{X}_0$ would intersect every interval $[x, (1+\varepsilon)x]$ for all large x, which would easily imply (together with (56) below) that the unweighted correlation converges to zero without any exceptional scales.

Remark 1.12. The twisted Dirichlet characters $\chi(n)n^{it}$ appear both in Conjecture 1.3 and in Theorems 1.6 and 1.7. However, there is an interesting distinction as to how they appear; in Conjecture 1.3, t is allowed to be quite large (as large as X) and $\chi(n)n^{it}$ is associated to just a single multiplicative function g_j, while in Theorems 1.6 and 1.7, the quantity t is independent of X and is now associated to the product $g_1 \cdots g_k$.

The dependence of t on X in Conjecture 1.3(i) is necessary,[7] as is shown in [Matomäki et al. 2015]; roughly speaking, the individual g_j can oscillate like n^{it_j} for various large t_j in such a fashion that these oscillations largely cancel and produce nontrivial correlations in the product $g_1(n+h_1)\cdots g_k(n+h_k)$. Meanwhile, Theorem 1.7 asserts in some sense that the shifted product $g_1(n+h_1)\cdots g_k(n+h_k)$ oscillates "similarly to" the unshifted product $g_1(n)\cdots g_k(n)$, so in particular if the latter began oscillating like n^{it} for increasingly large values of t then the former product should exhibit substantial cancellation.

The proof of Corollary 1.8 is found in Section 3. So far, all of our results have concerned correlations where the product of the multiplicative functions involved is nonpretentious. In the case of two-point correlations, however, we can prove Corollary 1.8 under the mere assumption that one of the multiplicative functions involved is nonpretentious, thus upgrading the logarithmic two-point Elliott conjecture in [Tao 2016] to an unweighted version at almost all scales.

Corollary 1.13 (the binary unweighted Elliott conjecture at almost all scales). *Let $g_1, g_2 : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions, such that there exists $j \in \{1, 2\}$ for which (1) holds as $X \to \infty$ for every Dirichlet character χ.*

(i) *For any distinct $h_1, h_2 \in \mathbb{Z}$ and $\varepsilon > 0$, one has*

$$|\mathbb{E}_{n\le X} g_1(n+h_1)g_2(n+h_2)| \le \varepsilon$$

for all natural numbers X outside of a set $\mathcal{X}_\varepsilon$ of logarithmic Banach density zero (in the sense of (6)).

(ii) *There is a set $\mathcal{X}_0$ of logarithmic density zero such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} \mathbb{E}_{n\le X} g_1(n+h_1)g_2(n+h_2) = 0$$

for all distinct $h_1, h_2 \in \mathbb{Z}$.

When specialised to the case of the Liouville function, the previous corollaries produce the following almost-all result.

Corollary 1.14 (some cases of the unweighted Chowla conjecture at almost all scales). *There is an exceptional set $\mathcal{X}_0$ of logarithmic density zero, such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} \mathbb{E}_{n\le X} \lambda(n+h_1)\cdots\lambda(n+h_k) = 0$$

for all natural numbers k that are either odd or equal to 2, and for any distinct integers $h_1, \dots, h_k$. The same result holds if one replaces one or more of the copies of the Liouville function λ with the Möbius function μ.

[7] In the case of the *logarithmically averaged* Conjecture 1.3(ii), in contrast, (1) might not be a necessary assumption, since the sequence of bad scales constructed in [Matomäki et al. 2015, Theorem B.1] is sparse and thus does not influence logarithmic averages.

We establish these results in Section 3. One can use these corollaries to extend some previous results involving the logarithmic density of sign patterns to now cover unweighted densities of sign patterns at almost all scales. For instance, by inserting Corollary 1.14 into the proof of [Tao and Teräväinen 2019, Corollary 1.10(i)], one obtains the following.

Corollary 1.15 (Liouville sign patterns of length three). *There is an exceptional set $\mathcal{X}_0$ of logarithmic density zero, such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} \mathbb{E}_{n\le X} 1_{(\lambda(n),\lambda(n+1),\lambda(n+2))=(\epsilon_0,\epsilon_1,\epsilon_2)} = \tfrac{1}{8}$$

for all sign patterns $(\epsilon_0, \epsilon_1, \epsilon_2) \in \{-1, 1\}^3$.

Similarly several other results in [Tao and Teräväinen 2019] and in [Teräväinen 2018] can be generalised. For example, the result [Teräväinen 2018, Theorem 1.16] on the largest prime factors of consecutive integers can be upgraded to the following form.

Corollary 1.16 (the largest prime factors of consecutive integers at almost all scales). *Let $P^+(n)$ be the largest prime factor of n with $P^+(1) := 1$. Then there is an exceptional set $\mathcal{X}_0$ of logarithmic density* 0, *such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} \mathbb{E}_{n\le X} 1_{P^+(n)<P^+(n+1)} = \tfrac{1}{2}. \tag{7}$$

The same equality with ordinary limit in place of the almost-all limit is an old conjecture formulated in the correspondence of Erdős and Turán [Sós 2002, pages 100–101; Erdős 1979]. We remark on the proof of Corollary 1.16 in Remark 3.3. In [Teräväinen 2018, Theorem 1.6] it was proved that (7) holds for the logarithmic average $\mathbb{E}^{\log}_{n\le X}$ (without any exceptional scales).

It would of course be desirable if we could upgrade "almost all scales" to "all scales" in the above results. We do not know how to do so in general, however there is one exceptional (though conjecturally nonexistent) case in which this is possible, namely if there are unusually few sign patterns in the multiplicative functions of interest. We illustrate this principle with the following example.

Theorem 1.17 (few sign patterns implies binary Chowla conjecture). *Suppose that for every $\varepsilon > 0$, there exist arbitrarily large natural numbers K such that the set $\{(\lambda(n+1), \dots, \lambda(n+K)) : n \in \mathbb{N}\} \subset \{-1, +1\}^K$ of sign patterns of length K has cardinality less than* $\exp(\varepsilon K/\log K)$. *Then, for any natural number h, one has*

$$\lim_{X\to\infty} \mathbb{E}_{n\le X}\lambda(n)\lambda(n+h) = 0.$$

Remark 1.18. The best known lower bounds for the number $s(K)$ of sign patterns of length K for the Liouville function are very far from $\exp(\varepsilon K/\log K)$. It was shown by Matomäki, Radziwiłł and Tao [Matomäki et al. 2016] that $s(K) \ge K+5$, and Frantzikinakis and Host [2018] showed that $s(K)/K \to \infty$ as $K \to \infty$, but the rate of growth is inexplicit in that result. This was very recently improved to $s(K) \gg K^2$ by McNamara [2019]. If one assumes the Chowla conjecture (in either the unweighted or logarithmically averaged forms), it is not difficult to conclude that in fact $s(K) = 2^K$ for all K.

We prove this result in Section 5. Roughly speaking, the reason for this improvement is that the entropy decrement argument that is crucially used in the previous arguments becomes significantly stronger under the hypothesis of few sign patterns. A similar result holds for the odd order cases of the Chowla conjecture if one assumes the sign pattern control for *all* large K (rather than for a sequence of arbitrarily large K) by adapting the arguments in [Tao and Teräväinen 2018], but we do not do so here. It is also possible to strengthen this theorem in a number of further ways (for instance, restricting attention to sign patterns that occur with positive upper density, or to extend to other 1-bounded multiplicative functions than the Liouville function), but we again do not do so here.

One should view Theorem 1.17 as stating that if there is ”too much structure” in the Liouville sequence (in the sense that it has a small number of sign patterns), then the binary Chowla conjecture holds. This is somewhat reminiscent of various statements in analytic number theory that rely on the assumption of a Siegel zero; for example, Heath-Brown [1983] proved that if there are Siegel zeros, then the twin prime conjecture (which is connected to the two-point Chowla conjecture) holds. Nevertheless, the proof of Theorem 1.17 does not resemble that in [Heath-Brown 1983].

Isotopy formulae. The conclusion of Theorem 1.7(ii) asserts, roughly speaking, that $f_d(a)$ “behaves like” a multiple of $\chi(a)d^{-\mathrm{it}}$ in a certain asymptotic sense. The following corollary of that theorem makes this intuition a bit more precise.

Theorem 1.19 (isotopy formulae). *Let $k \geq 1$, let $h_1, \dots, h_k$ be integers and $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions. Suppose that the product $g_1 \cdots g_k$ weakly pretends to be a twisted Dirichlet character $n \mapsto \chi(n)n^{\mathrm{it}}$.*

(i) (*Archimedean isotopy*) *There exists an exceptional set $\mathcal{X}_0$ of logarithmic density zero, such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} (\mathbb{E}_{n\leq X} g_1(n+h_1)\cdots g_k(n+h_k) - q^{\mathrm{it}}\mathbb{E}_{n\leq X/q} g_1(n+h_1)\cdots g_k(n+h_k)) = 0$$

for all rational numbers $q > 0$.

(ii) (*Nonarchimedean isotopy*) *There exists an exceptional set $\mathcal{X}_0$ of logarithmic density zero, such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} (\mathbb{E}_{n\leq X} g_1(n-ah_1)\cdots g_k(n-ah_k) - \chi(-1)\mathbb{E}_{n\leq X} g_1(n+ah_1)\cdots g_k(n+ah_k)) = 0$$

for all integers a.

Remark 1.20. This generalises [Tao and Teräväinen 2019, Theorem 1.2(iii)], which implies $f(-a) = \chi(-1)f(a)$ where $f(a)$ is a generalised limit of a logarithmic correlation defined in (2) (indeed, Theorem 1.19(ii) implies by partial summation that $f(-a) = \chi(-1)f(a)$ in the notation of (2)). In [Tao and Teräväinen 2019], we only considered logarithmically averaged correlations, and for such averages Theorem 1.19(i) does not make sense, as logarithmic averages are automatically slowly varying. However, for unweighted averages Theorem 1.19(i) gives nontrivial information about the behaviour of the correlation at nearby scales.

We give the proof of Theorem 1.19 in Section 4. We show in that section that, perhaps surprisingly, the nonarchimedean isotopy formula (Theorem 1.19(ii)) allows us to evaluate the correlations of some multiplicative functions whose product *does pretend* to be a Dirichlet character. Among other things, we use the isotopy formula to prove a version of the even order logarithmic Chowla conjectures where we twist one of the copies of the Liouville function by a carefully chosen Dirichlet character and the shifts of λ are consecutive.

Corollary 1.21 (even order correlations of a twisted Liouville function). *Let $k \geq 4$ be an even integer, and let χ be an odd Dirichlet character of period $k-1$ (there are $\varphi(k-1)/2$ such characters). Then there exists an exceptional set $\mathcal{X}_0$ of logarithmic density* 0, *such that*

$$\lim_{X\to\infty;\, X\notin\mathcal{X}_0} \mathbb{E}_{n\leq X}\,\chi(n)\lambda(n)\lambda(n+a)\cdots\lambda(n+(k-1)a) = 0 \tag{8}$$

for all integers a.

By partial summation, we see from (8) that we have the logarithmic correlation result

$$\lim_{X\to\infty} \mathbb{E}^{\log}_{n\leq X}\,\chi(n)\lambda(n)\lambda(n+1)\cdots\lambda(n+k-1) = 0,$$

which is already new. We stated Corollary 1.21 only for even k, but of course the result also holds for odd k by Corollary 1.8.

The assumption that χ is an odd character is crucial above, as will be seen in Section 4; the isotopy formulae are not able to say anything about the untwisted even order correlations of the Liouville function.

We likewise show in Section 4 that the archimedean isotopy formula (Theorem 1.19(i)) gives a rather satisfactory description of the *limit points* of the correlations

$$\mathbb{E}_{n\leq X}\, g_1(n+h_1)\cdots g_k(n+h_k), \tag{9}$$

where the product $g_1\cdots g_k$ weakly pretends to be a twisted Dirichlet character $n\mapsto \chi(n)n^{\mathrm{it}}$ with $t\neq 0$. Indeed, our Theorem 4.2 shows that once one continuously excludes the scales at which the correlation (9) is close to zero, the argument of the quantity in (9) is in a sense uniformly distributed on the unit circle. This uniform distribution is indeed expected when g_j are pretentious; for example, one has $\mathbb{E}_{n\leq X} n^{\mathrm{it}} = X^{\mathrm{it}}/(1+it)+o(1)$, which uniformly distributes on the circle of radius $1/|1+it|$ with respect to logarithmic density.

Proof ideas. We now briefly describe (in informal terms) the proof strategy for Theorem 1.7, which follows the ideas in [Tao and Teräväinen 2019], but now contains some "archimedean" arguments (relating to the archimedean characters $n\mapsto n^{\mathrm{it}}$) in addition to the "nonarchimedean" arguments in [loc. cit.] (that related to the Dirichlet characters $n\mapsto\chi(n)$). The new features compared to [loc. cit.] include extensive use of the fact that the correlations $f_d(a)$ are "slowly varying" in terms of d (this is made precise in formula (16)), and the use of this to derive "approximate quasimorphism properties" for certain quantities related to these correlations (these are detailed below). We then prove that the approximate quasimorphisms are very close to actual quasimorphisms (which in our case are Dirichlet characters or archimedean characters), which eventually leads to the desired conclusions.

As already noted, one key ingredient is (a rigorous form of) the approximate identity (3) that arises from the entropy decrement argument. In terms of the correlation functions $f_d(a)$, this identity takes the (heuristic) form

$$f_{dp}(a)G(p) \approx f_d(ap)$$

for any integers a, d and "most" p, where $G := g_1 \cdots g_k$; see Proposition 2.3 for a precise statement. Compared to [loc. cit.], the main new difficulty is the dependence of f_d on the d parameter.

Assuming for simplicity that G has modulus 1 (which is the most difficult case), we thus have

$$f_{dp}(a) \approx f_d(ap)\overline{G(p)}$$

for any integers a, d and "most" p. Iterating this leads to

$$f_{p_1 p_2}(a) \approx f_1(ap_1 p_2)\overline{G(p_1)}\,\overline{G(p_2)} \tag{10}$$

for "most" primes p_1, p_2 (more precisely, the difference between the two sides of the equation is $o(1)$ when suitably averaged over p_1, p_2; see Corollary 2.4). On the other hand, results from ergodic theory (such as [Leibman 2015; Le 2018]) give control on the function $f_1(a)$, describing it (up to negligible errors) as a nilsequence, which can then be decomposed further into a periodic piece $f_{1,0}$ and an "irrational" component. The irrational component was already shown in [Tao and Teräväinen 2019] to give a negligible contribution to the (10) after performing some averaging in p_1, p_2, thanks to certain bilinear estimates for nilsequences. As such, one can effectively replace f_1 here by the periodic component $f_{1,0}$ (see (19) for a precise statement).

We thus reach the relation

$$f_{p_1 p_2}(a) \approx f_{1,0}(ap_1 p_2)\overline{G(p_1)}\,\overline{G(p_2)}$$

for "most" p_1, p_2. Let q be the period of $f_{1,0}$. If we pick two large primes $p_1 \equiv c \pmod q$ and $p_1' \equiv bc \pmod q$ for arbitrary $b, c \in (\mathbb{Z}/q\mathbb{Z})^\times$ with $p_1 \approx p_1'$ (using the prime number theorem), we get

$$f_{1,0}(acp_2)\overline{G(p_1)} \approx f_{1,0}(abcp_2)\overline{G(p_1')},$$

for "most" p_1, p_1', p_2, since the averages $f_d(a)$ are slowly varying as a function of d (see (16) for the precise meaning of this). Choosing $p_2 \equiv 1 \pmod q$, we see that the quotient $f_{1,0}(ac)/f_{1,0}(abc)$ is independent of a (since p_1, p_1' were independent of a). Substituting then $a = a_1$ and $a = a_2$ to the quotient, we get the approximate identity

$$f_{1,0}(a_1 c)f_{1,0}(a_2 bc) \approx f_{1,0}(a_1 bc)f_{1,0}(a_2 c); \tag{11}$$

see Proposition 2.7 for a precise version of this, where we need to average over c to make the argument rigorous. We may assume that $f(a_0) \neq 0$ for some a_0, as otherwise there is nothing to prove, and this leads to $f_{1,0}(a_0) \neq 0$. Taking $a_1 \equiv a_0 c^{-1} \pmod q$, $a_2 \equiv a_0 \pmod q$ in (11), we are led to

$$f_{1,0}(a_0) f_{1,0}(a_0 bc) \approx f_{1,0}(a_0 b) f_{1,0}(a_0 c)$$

Thus, the function $\psi(x) = f_{1,0}(a_0 x)/f_{1,0}(a_0)$ satisfies the approximate quasimorphism equation

$$\psi(b_1 b_2) \approx \psi(b_1)\psi(b_2)$$

for $b_1, b_2 \in (\mathbb{Z}/q\mathbb{Z})^\times$ ranging in the invertible residue classes in $\mathbb{Z}/q\mathbb{Z}$ and some unknown function $\psi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ (to make the above deductions rigorous, we need to take as $\psi(x)$ an averaged version of $x \mapsto f_{1,0}(a_0 x)/f_{1,0}(a_0)$). Moreover, the function $\psi(x)$ takes values comparable to 1. Of course, Dirichlet characters obey the quasimorphism equation exactly; and we can use standard "cocycle straightening" arguments to show conversely that any solution to the quasimorphism equation must be very close to a Dirichlet character χ (see Lemma 2.8 for a precise statement). This will be used to show that $f_{1,0}$ and f_d are essentially χ-isotypic.

Once this isotopy property is established, one can then return to (10) and analyse the dependence of various components of (10) on the archimedean magnitudes of p_1, p_2 rather than their residues mod q. One can eventually transform this equation again to the quasimorphism equation, but this time on the multiplicative group $\mathbb{R}^+$ rather than $(\mathbb{Z}/q\mathbb{Z})^\times$ (also, the functions ψ will be "log-Lipschitz" in a certain sense). Now it is the archimedean characters $n \mapsto n^{\mathrm{it}}$ that are the model solutions of this equation, and we will again be able to show that all other solutions to this equation are close to an archimedean character (see Lemma 2.10 for a precise statement). Once one has extracted both the Dirichlet character χ and the archimedean character $n \mapsto n^{\mathrm{it}}$ in this fashion, the rest of Theorem 1.7 can be established by some routine calculations.

Notation. We use the usual asymptotic notation $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the bound $|X| \le CY$ for some constant C. If C needs to depend on parameters, we will denote this by subscripts, thus for instance $X \ll_k Y$ denotes the estimate $|X| \le C_k Y$ for some C_k depending on k. We also write $o_{n\to\infty}(Y)$ for a quantity bounded in magnitude by $c(n)Y$ for some $c(n)$ that goes to zero as $n \to \infty$ (holding all other parameters fixed). For any set $\mathcal{X} \subset \mathbb{N}$ with infinite complement, we define the limit operator $\lim_{X\to\infty; X\notin\mathcal{X}} f(X)$ as $\lim_{n\to\infty} f(x_n)$, where $x_1, x_2, \ldots$ are the elements of the complement $\mathbb{N} \setminus \mathcal{X}$ in strictly increasing order.

We use $a\ (q)$ to denote the residue class of a modulo q. If E is a set, we write 1_E for its indicator function, thus $1_E(n) = 1$ when $n \in E$ and $1_E(n) = 0$ otherwise.

We use the following standard multiplicative functions throughout the paper:

- The *Liouville function* λ, which is the 1-bounded completely multiplicative function with $\lambda(p) = -1$ for all primes p.
- The *Möbius function* μ, which is equal to λ at square-free numbers and 0 elsewhere.
- *Dirichlet characters* χ, which are 1-bounded completely multiplicative functions of some period q, with $\chi(n)$ nonzero precisely when n is coprime to q.

- *Archimedean characters* $n \mapsto n^{\mathrm{it}}$, where t is a real number.
- *Twisted Dirichlet characters* $n \mapsto \chi(n) n^{\mathrm{it}}$, which are the product of a Dirichlet character and an archimedean character.

In the arguments that follow, asymptotic averages of various types feature frequently, so we introduce some abbreviations for them.

Definition 1.22 (asymptotic averaging notation). If $f : \mathbb{N} \to \mathbb{C}$ is a function, we define the asymptotic average

$$\mathbb{E}_{n \in \mathbb{N}} f(n) := \lim_{X \to \infty} \mathbb{E}_{n \le X} f(n)$$

provided that the limit exists. We adopt the convention that assertions such as $\mathbb{E}_{n \in \mathbb{N}} f(n) = \alpha$ are automatically false if the limit involved does not exist. Similarly define $\mathbb{E}^{\log}_{n \in \mathbb{N}} f(n)$ and $\mathbb{E}^{\log\log}_{n \in \mathbb{N}} f(n)$. If $f : \boldsymbol{P} \to \mathbb{C}$ is a function, we similarly define

$$\mathbb{E}_{p \in \boldsymbol{P}} f(p) := \lim_{X \to \infty} \mathbb{E}_{p \le X} f(p) \quad \text{and} \quad \mathbb{E}^{\log}_{p \in \boldsymbol{P}} f(p) := \lim_{X \to \infty} \mathbb{E}^{\log}_{p \le X} f(p).$$

Moreover, given a generalised limit functional $\lim^*_{X \to \infty}$, we define the corresponding asymptotic limits $\mathbb{E}^*_{n \in \mathbb{N}}$, $\mathbb{E}^{\log,*}_{n \in \mathbb{N}}$, $\mathbb{E}^{\log\log,*}_{n \in \mathbb{N}}$, $\mathbb{E}^*_{p \in \boldsymbol{P}}$, $\mathbb{E}^{\log,*}_{p \in \boldsymbol{P}}$ by replacing the ordinary limit functional by the generalised limit, thus for instance

$$\mathbb{E}^{\log,*}_{n \in \mathbb{N}} f(n) := \lim^*_{X \to \infty} \mathbb{E}^{\log}_{n \le X} f(n).$$

If an ordinary asymptotic limit such as $\mathbb{E}^{\log}_{n \in \mathbb{N}} f(n)$ exists, then $\mathbb{E}^{\log,*}_{n \in \mathbb{N}} f(n)$ will attain the same value; but the latter limit exists for all bounded sequences f, whereas the ordinary limit need not exist. In later parts of the paper we will also need an additional generalised limit $\lim^{**}_{X \to \infty}$, and one can then define generalised asymptotic averages such as $\mathbb{E}^{\log,**}_{n \in \mathbb{N}} f(n)$ accordingly.

Remark 1.23. If f is a bounded sequence and α is a complex number, a standard summation by parts exercise shows that the statement $\mathbb{E}_{n \in \mathbb{N}} f(n) = \alpha$ implies $\mathbb{E}^{\log}_{n \in \mathbb{N}} f(n) = \alpha$, which in turn implies $\mathbb{E}^{\log\log}_{n \in \mathbb{N}} f(n) = \alpha$, and similarly $\mathbb{E}_{p \in \boldsymbol{P}} f(p) = \alpha$ implies $\mathbb{E}^{\log}_{p \in \boldsymbol{P}} f(p) = \alpha$; however, the converse implications can be highly nontrivial or even false. For instance, as mentioned earlier, it is not difficult to show that $\mathbb{E}^{\log}_{n \in \mathbb{N}} n^{\mathrm{it}} = 0$ for any $t \neq 0$, but the limit $\mathbb{E}_{n \in \mathbb{N}} n^{\mathrm{it}}$ does not exist. (On the other hand, from the prime number theorem and partial summation one has $\mathbb{E}_{p \in \boldsymbol{P}} p^{\mathrm{it}} = 0$.) In the same spirit, if A is the set of integers whose decimal expansion has leading digit 1, then one easily computes "Benford's law" $\mathbb{E}^{\log}_{n \in \mathbb{N}} 1_A(n) = (\log 2)/(\log 10)$, whereas $\mathbb{E}_{n \in \mathbb{N}} 1_A(n)$ fails to exist.

2. Proof of main theorem

In this section we establish Theorem 1.7. We first establish a version of the Furstenberg correspondence principle.

Proposition 2.1 (Furstenberg correspondence principle). *Let the notation and hypotheses be as in Theorem 1.7. Then for any real number $d > 0$, there exist random functions $\mathbf{g}_1^{(d)}, \ldots, \mathbf{g}_k^{(d)} : \mathbb{Z} \to \mathbb{D}$ and a random profinite integer $\mathbf{n}^{(d)} \in \hat{\mathbb{Z}}$,*[8] *all defined on a common probability space $\Omega^{(d)}$, such that*

$$\mathbb{E}^{(d)} F(((\mathbf{g}_i^{(d)}(h))_{1\le i\le k, -N\le h\le N}, \mathbf{n}^{(d)}\ (q))) = \lim^*_{X\to\infty} \mathbb{E}_{n\le X/d} F((g_i(n+h))_{1\le i\le k,-N\le h\le N}, n\ (q))$$

for any natural numbers N, q and any continuous function $F : \mathbb{D}^{k(2N+1)} \times \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$, where $\mathbb{E}^{(d)}$ denotes the expectation on the probability space $\Omega^{(d)}$. Furthermore, the random variables $\mathbf{g}_1^{(d)}, \ldots, \mathbf{g}_k^{(d)} : \mathbb{Z} \to \mathbb{D}$ and $\mathbf{n}^{(d)} \in \hat{\mathbb{Z}}$ are a stationary process, by which we mean that for any natural number N, the joint distribution of $(\mathbf{g}_i^{(d)}(n+h))_{1\le i\le k,-N\le h\le N}$ and $\mathbf{n}^{(d)} + n$ does not depend on n as n ranges over the integers.

Proof. Up to some minor notational changes, this is essentially [Tao and Teräväinen 2019, Proposition 3.1], applied once for each value of d. The only difference is that the logarithmic averaging $\mathbb{E}^{\log}_{x_m/w_m\le n\le x_m}$ there has been replaced by the nonlogarithmic averaging $\mathbb{E}_{n\le X/d}$. However, an inspection of the arguments reveal that the proof of the proposition is essentially unaffected by this change. □

Let $G : \mathbb{N} \to \mathbb{D}$ denote the multiplicative function $G := g_1 \cdots g_k$. We now adapt the entropy decrement arguments from [Tao and Teräväinen 2019, Section 3] to establish the approximate relation

$$f_d(ap) \approx f_{dp}(a)G(p) \tag{12}$$

for integers a, real numbers $d > 0$, and "most" primes p.

Fix a, d, and let p be a prime. From (4) we have

$$f_{dp}(a)G(p) = \lim^*_{x\to\infty} \mathbb{E}_{n\le x/dp}\, g_1(p)g_1(n+ah_1)\cdots g_k(p)g_k(n+ah_k).$$

From multiplicativity, we can write $g_j(p)g_j(n+ah_j)$ as $g_j(pn+aph_j)$ unless $n = -ah_j\ (p)$. The latter case contributes $O\big(\frac{1}{p}\big)$ to the above limit (where we allow implied constants to depend on k), thus

$$f_{dp}(a)G(p) = \lim^*_{x\to\infty} \mathbb{E}_{n\le x/dp}\, g_1(pn+aph_1)\cdots g_k(pn+aph_k) + O\big(\tfrac{1}{p}\big).$$

If we now make pn rather than n the variable of summation, we conclude that

$$f_{dp}(a)G(p) = \lim^*_{x\to\infty} \mathbb{E}_{n\le x/d}\, g_1(n+aph_1)\cdots g_k(n+aph_k)p1_{p\,|\,n} + O\big(\tfrac{1}{p}\big).$$

Comparing this with (4), we conclude that

$$f_{dp}(a)G(p) - f_d(ap) = \lim^*_{x\to\infty} \mathbb{E}_{n\le x/d}\, g_1(n+aph_1)\cdots g_k(n+aph_k)(p1_{p\,|\,n}-1) + O\big(\tfrac{1}{p}\big)$$

and hence by Proposition 2.1

$$f_{dp}(a)G(p) - f_d(ap) = \mathbb{E}^{(d)} \mathbf{g}_1^{(d)}(aph_1)\cdots \mathbf{g}_k^{(d)}(aph_k)(p1_{p\,|\,\mathbf{n}^{(d)}}-1) + O\big(\tfrac{1}{p}\big). \tag{13}$$

[8] The *profinite integers* $\hat{\mathbb{Z}}$ are the inverse limit of the cyclic groups $\mathbb{Z}/q\mathbb{Z}$, with the weakest topology that makes the reduction maps $n \mapsto n\ (q)$ continuous. This is a compact abelian group and therefore it has a well-defined probability Haar measure.

On the other hand, by repeating the proof of [Tao and Teräväinen 2019, Theorem 3.6] verbatim (see also [loc. cit., Remark 3.7]), we have the following general estimate:

Proposition 2.2 (entropy decrement argument). *Let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_k : \mathbb{Z} \to \mathbb{D}$ be random functions and $\boldsymbol{n} \in \hat{\mathbb{Z}}$ be a stationary process, let $a, h_1, \dots, h_k$ be integers, and let $0 < \varepsilon < \frac{1}{2}$ be real. Then one has*

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |\mathbb{E}\boldsymbol{g}_1(aph_1) \cdots \boldsymbol{g}_k(aph_k)(p1_{p \mid \boldsymbol{n}} - 1)| \le \varepsilon$$

for all natural numbers m outside of an exceptional set $\mathcal{M}$ obeying the bound

$$\sum_{m \in \mathcal{M}} \tfrac{1}{m} \ll_{a,h_1,\dots,h_k} \varepsilon^{-4} \log \tfrac{1}{\varepsilon}. \tag{14}$$

Note that the bound (14) is uniform in the random functions $\boldsymbol{g}_1, \dots, \boldsymbol{g}_k$ (although the set $\mathcal{M}$ may depend on these functions). Summing the result over different dyadic scales gives us the following version of (12).

Proposition 2.3 (approximate isotopy). *Let the notation and hypotheses be as in Theorem 1.7. Let a be an integer, and let $\varepsilon > 0$ be real. Then for sufficiently large P, we have*

$$\sup_{d>0} \mathbb{E}^{\log}_{p \le P} |f_{dp}(a)G(p) - f_d(ap)| \le \varepsilon$$

where the supremum is over positive reals.

A key technical point for our application is that while P may depend on a, ε, it can be taken to be uniform in d.

Proof. Let a, ε, P be as in the proposition, and let $d > 0$. We may assume that $\varepsilon > 0$ is small. By the prime number theorem, we have

$$\mathbb{E}^{\log}_{p \le P} |f_{dp}(a)G(p) - f_d(ap)| \ll \mathbb{E}^{\log}_{m \le (\log P)/(\log 2)} \mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a)G(p) - f_d(ap)|.$$

By (13) and Proposition 2.2, we have

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a)G(p) - f_d(ap)| \le \varepsilon^2$$

for all m outside of an exceptional set $\mathcal{M}_{a,\varepsilon,d}$ obeying the bound

$$\sum_{m \in \mathcal{M}_{a,\varepsilon,d}} \tfrac{1}{m} \ll_{a,h_1,\dots,h_k} \varepsilon^{-8} \log \tfrac{1}{\varepsilon}.$$

In the exceptional set $\mathcal{M}_{a,\varepsilon,d}$, we use the trivial bound

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a)G(p) - f_d(ap)| \ll 1$$

to conclude that

$$\mathbb{E}^{\log}_{p \le P} |f_{dp}(a)G(p) - f_d(ap)| \ll \varepsilon^2 + O_{a,h_1,\dots,h_k}\left(\frac{\varepsilon^{-8} \log 1/\varepsilon}{\log \log P}\right),$$

and the claim follows by choosing P large in terms of $a, \varepsilon, h_1, \dots, h_k$. $\square$

As in [Tao and Teräväinen 2019], we iterate this approximate formula to obtain:

Corollary 2.4. *For any integer a one has*

$$\limsup_{P_1\to\infty}\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)G(p_1)G(p_2)-f_1(ap_1p_2)|=0.$$

Proof. Let a be an integer, let $\varepsilon>0$ be real, let P_1 be sufficiently large depending on a,ε, and let P_2 be sufficiently large depending on a,ε,P_1. From Proposition 2.3 one has

$$\mathbb{E}^{\log}_{p_1\le P_1}|f_{p_1p_2}(a)G(p_1)-f_{p_2}(ap_1)|\ll\varepsilon$$

for all primes p_2, and hence

$$\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)G(p_1)G(p_2)-f_{p_2}(ap_1)G(p_2)|\ll\varepsilon.$$

On the other hand, from a second application of Proposition 2.3 one has

$$\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_2}(ap_1)G(p_2)-f_1(ap_1p_2)|\ll\varepsilon$$

for all $p_1\le P_1$, and hence

$$\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_2}(ap_1)G(p_2)-f_1(ap_1p_2)|\ll\varepsilon.$$

From the triangle inequality we thus have

$$\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)G(p_1)G(p_2)-f_1(ap_1p_2)|\ll\varepsilon$$

under the stated hypotheses on ε, P_1, P_2. Taking limit superior in P_2 and then in P_1, we conclude that

$$\limsup_{P_1\to\infty}\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)G(p_1)G(p_2)-f_1(ap_1p_2)|\ll\varepsilon$$

for any $\varepsilon>0$, and the claim follows. □

Next, we have the following structural description of f_1.

Proposition 2.5. *Let f_1 be as in Theorem 1.7. For any $\varepsilon>0$, one can write*

$$f_1=f_{1,0}+g$$

where $f_{1,0}=f_{1,0}^{(\varepsilon)}$ is periodic, and the error $g=g^{(\varepsilon)}$ obeys the bilinear estimate

$$\mathbb{E}_{p_1\le x}\mathbb{E}_{p_2\le y}\alpha_{p_1}\beta_{p_2}g(ap_1p_2)\ll\varepsilon \tag{15}$$

as well as the logarithmic counterpart

$$\mathbb{E}^{\log}_{p_1\le x}\mathbb{E}^{\log}_{p_2\le y}\alpha_{p_1}\beta_{p_2}g(ap_1p_2)\ll\varepsilon$$

whenever a is a nonzero integer, x is sufficiently large depending on a,ε; y is sufficiently large depending on x,a,ε; and $\alpha_{p_1},\beta_{p_2}=O(1)$ are bounded sequences.

Proof. We freely use the notation from [Tao and Teräväinen 2019, Sections 4–5]. By summation by parts it suffices to obtain a decomposition obeying (15). By repeating the proof of [loc. cit., Corollary 4.6] verbatim,[9] we can write

$$f_1 = f_{1,1} + f_{1,2}$$

where $f_{1,1}$ is a nilsequence of some finite degree D, and $f_{1,2}$ obeys the asymptotic

$$\lim_{x\to\infty} \mathbb{E}_{p\le x} |f_{1,2}(ap)| = 0$$

for any nonzero integer a. We can now neglect the $f_{1,2}$ term as it can be absorbed into the g error. Next, applying [loc. cit., Proposition 5.6], we can decompose

$$f_{1,1} = f_{1,0} + \sum_{i=1}^{D} \sum_{j=1}^{J_i} c_{i,j}\chi_{i,j}$$

for some periodic function $f_{1,0}$, some nonnegative integers $J_1, \dots, J_D$, some irrational nilcharacters $\chi_{i,j}$ of degree i, and some linear functionals $c_{i,j}$. Using [loc. cit., Lemma 5.8] (noting that if χ is an irrational nilcharacter, then so is $\chi(a\cdot)$) we see that each of the terms $c_{i,j}\chi_{i,j}$ can be absorbed into the error term g. The claim then follows from the triangle inequality. □

Finally, we record a simple log-Lipschitz estimate

$$|f_{d_1}(a) - f_{d_2}(a)| \le 2|\log d_1 - \log d_2| \tag{16}$$

for any integer a and any real $d_1, d_2 > 0$; this follows by using (4) and the triangle inequality to estimate $|f_{d_1}(a) - f_{d_2}(a)| \le 2|d_1 - d_2|/\max\{d_1, d_2\}$ and then the mean value theorem to $x \mapsto \log x$.

We return to the proof of Theorem 1.7. If we have

$$\limsup_{X\to\infty} \mathbb{E}^{\log\log}_{d\le X} |f_d(a)| = 0$$

for all a, then the claim follows by setting $f = 0$, so we may assume without loss of generality that there exists an integer a_0 such that

$$\limsup_{X\to\infty} \mathbb{E}^{\log\log}_{d\le X} |f_d(a_0)| > 0.$$

Thus, by the Hahn–Banach theorem, we may find a generalised limit $\lim^{**}_{X\to\infty}$ (which may or may not be equal to the previous generalised limit $\lim^{*}_{X\to\infty}$) such that

$$\lim^{**}_{X\to\infty} \mathbb{E}^{\log\log}_{d\le X} |f_d(a_0)| > 0,$$

and thus using the generalised limit asymptotic notation associated to $\lim^{**}_{X\to\infty}$ (see page 1), we have

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_d(a_0)| \gg 1. \tag{17}$$

[9] In [loc. cit., Corollary 4.6], a was required to be a natural number rather than a nonzero integer, however one can easily adapt the arguments to the case of negative a with only minor modifications (in particular, one has to modify the definition of $\boldsymbol{X}_m$ slightly to allow l to be negative).

For future reference we record the following convenient lemma relating the averaging operator $\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}$ with $\mathbb{E}^{\log,**}_{p\in\boldsymbol{P}}$:

Lemma 2.6 (comparing averages over integers and primes). *Let $f : \mathbb{N} \to \mathbb{C}$ be a function which is bounded log-Lipschitz in the sense that there is a constant C such that $|f(d)| \leq C$ and $|f(d) - f(d')| \leq C|\log d - \log d'|$ for all $d, d' \in \mathbb{N}$. Then for any natural number a, one has*

$$\limsup_{X\to\infty} |\mathbb{E}^{\log\log}_{d\leq X} f(d) - \mathbb{E}^{\log}_{p\leq X} f(ap)| = 0,$$

so in particular

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} f(d) = \mathbb{E}^{\log,**}_{p\in\boldsymbol{P}} f(ap).$$

Proof. We allow implied constants to depend on C, a. Let $\varepsilon > 0$, and assume X is sufficiently large depending on C, ε. Then from the prime number theorem and the bounded log-Lipschitz property we have

$$\begin{aligned}
\mathbb{E}^{\log}_{p\leq X} f(ap) &= \frac{1}{\log\log X} \sum_{p\leq X} \frac{f(ap)}{p} + O(\varepsilon) \\
&= \frac{1}{\log\log X} \sum_{d\leq X} \frac{1}{\varepsilon d} \sum_{d\leq p\leq(1+\varepsilon)d} \frac{f(ap)}{p} + O(\varepsilon) \\
&= \frac{1}{\log\log X} \sum_{d\leq X} \frac{1}{\varepsilon d} \sum_{d\leq p\leq(1+\varepsilon)d} \frac{f(ad)}{d} + O(\varepsilon) \\
&= \frac{1}{\log\log X} \sum_{d\leq X} \frac{f(ad)}{d\log(2+d)} + O(\varepsilon).
\end{aligned}$$

Again by the bounded log-Lipschitz property, we have

$$f(ad) = \frac{1}{a} \sum_{ad\leq d'<a(d+1)} f(d') + O(1/d),$$

and inserting this into the preceding computation, we get

$$\begin{aligned}
\mathbb{E}^{\log}_{p\leq X} f(ap) &= \frac{1}{\log\log X} \sum_{d'\leq aX} f(d') \cdot \frac{1}{a} \sum_{d'/a-1<d\leq d'/a} \frac{1}{d\log(2+d)} + O(\varepsilon) \\
&= \frac{1}{\log\log X} \sum_{d'\leq X} \frac{f(d')}{d'\log(2+d')} + O(\varepsilon).
\end{aligned}$$

Taking the absolute value of the difference of the two sides of this equation, applying $\limsup_{X\to\infty}$ and then sending $\varepsilon \to 0$, we obtain the claim. ☐

Now, let $\varepsilon > 0$ be a sufficiently small parameter. If one had

$$\sum_p \frac{1 - |g_j(p)|}{p} = \infty$$

for some $1 \leq j \leq k$, then by Wirsing's theorem [1967] as in [Tao and Teräväinen 2019, Section 6] one would have $f_d(a) = 0$ for all a, d. Thus we may assume that

$$\sum_p \frac{1 - |g_j(p)|}{p} < \infty$$

for all j, which implies in particular that one has

$$1 - \varepsilon \leq |G(p)| \leq 1 \tag{18}$$

for all but finitely many p. For any integer a, we see from Corollary 2.4 that

$$\limsup_{P_1 \to \infty} \limsup_{P_2 \to \infty} \mathbb{E}^{\log}_{p_1 \leq P_1} \mathbb{E}^{\log}_{p_2 \leq P_2} |f_{p_1 p_2}(a) G(p_1) G(p_2) - f_1(a p_1 p_2)| \ll \varepsilon.$$

By (18) we then have

$$\limsup_{P_1 \to \infty} \limsup_{P_2 \to \infty} \mathbb{E}^{\log}_{p_1 \leq P_1} \mathbb{E}^{\log}_{p_2 \leq P_2} |f_{p_1 p_2}(a) - \overline{G(p_1)}\,\overline{G(p_2)} f(a p_1 p_2)| \ll \varepsilon$$

Applying Proposition 2.5, we conclude that

$$\limsup_{P_1 \to \infty} \limsup_{P_2 \to \infty} \mathbb{E}^{\log}_{p_1 \leq P_1} \mathbb{E}^{\log}_{p_2 \leq P_2} |f_{p_1 p_2}(a) - \overline{G(p_1)}\,\overline{G(p_2)} f_{1,0}(a p_1 p_2)| \ll \varepsilon. \tag{19}$$

In particular we have

$$\mathbb{E}^{\log,**}_{p_1 \in \boldsymbol{P}} \mathbb{E}^{\log,**}_{p_2 \in \boldsymbol{P}} |f_{p_1 p_2}(a) - \overline{G(p_1)}\,\overline{G(p_2)} f_{1,0}(a p_1 p_2)| \ll \varepsilon. \tag{20}$$

Heuristically, (20) asserts the approximation

$$f_{p_1 p_2}(a) \approx \overline{G(p_1)}\,\overline{G(p_2)} f_{1,0}(a p_1 p_2) \tag{21}$$

for "most" a, p_1, p_2. This turns out to be a remarkably powerful approximate equation, giving a lot of control on the functions G, f_d, and $f_{1,0}$. Roughly speaking, we will be able to show that the only way to solve (21) (in a manner compatible with (17) and (16)) is if $G(p) \approx \chi(p) p^{\mathrm{i}t}$, $f_d(a) \approx f(a) d^{-\mathrm{i}t}$, and $f_{1,0} \approx f$ for some χ-isotypic q-periodic function f. Conversely, it is easy to see that if G, f_d, $f_{1,0}$ are of the above form, then they obey (21).

We first use (20) to control $f_{1,0}$. Let q denote the period of $f_{1,0}$ (which depends on ε); by abuse of notation, we view $f_{1,0}$ as a function on $\mathbb{Z}/q\mathbb{Z}$ as well as on $\mathbb{Z}$. We then have:

Proposition 2.7 (initial control on $f_{1,0}$). *Let a_0 be as in (17). We have*

$$\mathbb{E}_{c \in (\mathbb{Z}/q\mathbb{Z})^\times} |f_{1,0}(a_0 c)| \gg 1. \tag{22}$$

Furthermore, for any integers a_1, a_2 and any natural number b coprime to q, we have

$$\mathbb{E}_{c \in (\mathbb{Z}/q\mathbb{Z})^\times} |f_{1,0}(a_1 c) f_{1,0}(a_2 b c) - f_{1,0}(a_1 b c) f_{1,0}(a_2 c)| \ll \varepsilon. \tag{23}$$

Proof. By Lemma 2.6, (17) and (16), we see that

$$\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{p_1p_2}(a_0)| = \mathbb{E}^{\log\log,**}_{d}|f_d(a_0)| \gg 1$$

for any p_1, and hence

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{p_1p_2}(a_0)| \gg 1.$$

On the other hand, from (20) we have

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{p_1p_2}(a_0) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_0p_1p_2)| \ll \varepsilon. \tag{24}$$

From the triangle inequality, we have

$$|f_{p_1p_2}(a_0)| \ll |f_{1,0}(a_0p_1p_2)| + |f_{p_1p_2}(a_0) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_0p_1p_2)|,$$

and hence (since ε is assumed small)

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{1,0}(a_0p_1p_2)| \gg 1.$$

By the periodicity of $f_{1,0}$ and the prime number theorem in arithmetic progressions, we conclude (22).

Next, let a_1, a_2, b be as in the proposition. Applying (20) twice, we see that

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{p_1p_2}(a_1) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_1p_1p_2)| \ll \varepsilon \tag{25}$$

and

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in\boldsymbol{P}}|f_{p_1p_2}(a_2) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_2p_1p_2)| \ll \varepsilon. \tag{26}$$

We now eliminate the functions $f_{p_1p_2}$ and G from these estimates. As in the proof of Lemma 2.6, we can use the prime number theorem in arithmetic progressions to rearrange the left-hand side of (25) as

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2<(1+\varepsilon)d;\,p_2=c\ (q)}|f_{p_1p_2}(a_1) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_1p_1p_2)| + O(\varepsilon)$$

and hence after a change of variables $c\mapsto bc$ (and renaming p_2 as p_2')

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2'<(1+\varepsilon)d;\,p_2'=bc\ (q)}|f_{p_1p_2'}(a_1) - \overline{G(p_1)}\,\overline{G(p_2')}f_{1,0}(a_1p_1p_2')| \ll \varepsilon.$$

From (16), we have $f_{p_1p_2}(a_1), f_{p_1p_2'}(a_1) = f_{p_1d}(a_1) + O(\varepsilon)$; from the periodicity of $f_{1,0}$ we also have $f_{1,0}(a_1p_1p_2) = f_{1,0}(a_1cp_1)$ and $f_{1,0}(a_1p_1p_2') = f_{1,0}(a_1bcp_1)$. We conclude that

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2<(1+\varepsilon)d;\,p_2=c\ (q)}|f_{p_1d}(a_1) - \overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a_1cp_1)| \ll \varepsilon$$

and

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2'<(1+\varepsilon)d;\,p_2'=bc\ (q)}|f_{p_1d}(a_1) - \overline{G(p_1)}\,\overline{G(p_2')}f_{1,0}(a_1bcp_1)| \ll \varepsilon$$

and hence by the triangle inequality and (18) we have

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2<(1+\varepsilon)d;\,p_2=c\ (q)} \\ \times\mathbb{E}_{d\le p_2'<(1+\varepsilon)d;\,p_2'=bc\ (q)}|\overline{G(p_2)}f_{1,0}(a_1cp_1) - \overline{G(p_2')}f_{1,0}(a_1bcp_1)| \ll \varepsilon.$$

We have thus eliminated $f_{p_1p_2}$ and one factor of G; we still seek to eliminate the other factor of G. To do this, we replace a_1 by a_2 in the above analysis to obtain

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2<(1+\varepsilon)d;\,p_2=c\ (q)} \times \mathbb{E}_{d\le p_2'<(1+\varepsilon)d;\,p_2'=bc\ (q)}|\overline{G(p_2)}f_{1,0}(a_2cp_1)-\overline{G(p_2')}f_{1,0}(a_2bcp_1)|\ll\varepsilon.$$

At this point, let us note that $|f_{1,0}(a)|\ll 1$ for $a\in\mathbb{Z}$. To see this, we use Corollary 2.4 to conclude that

$$\limsup_{P_1\to\infty}\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1;\,p_1\equiv 1\ (q)}\mathbb{E}^{\log}_{p_2\le P_2;\,p_2\equiv 1\ (q)}|f_{p_1p_2}(a)-\overline{G(p_1)}\,\overline{G(p_2)}f_{1,0}(a)|=0.$$

Then from the triangle inequality, (18), and the trivial bound $|f_{p_1p_2}(a)|\ll 1$ we reach the conclusion $|f_{1,0}(a)|\ll 1$.

Next observe the identity

$$\begin{aligned}\overline{G(p_2)}&(f_{1,0}(a_1cp_1)f_{1,0}(a_2bcp_1)-f_{1,0}(a_1bcp_1)f_{1,0}(a_2cp_1))\\ &= f_{1,0}(a_2bcp_1)(\overline{G(p_2)}f_{1,0}(a_1cp_1)-\overline{G(p_2')}f_{1,0}(a_1bcp_1))\\ &\quad - f_{1,0}(a_1bcp_1)(\overline{G(p_2)}f_{1,0}(a_2cp_1)-\overline{G(p_2')}f_{1,0}(a_2bcp_1));\end{aligned}$$

we thus have from the triangle inequality, the boundedness of $|f_{1,0}(a)|$, and (18) that

$$\begin{aligned}|f_{1,0}&(a_1cp_1)f_{1,0}(a_2bcp_1)-f_{1,0}(a_1bcp_1)f_{1,0}(a_2cp_1)|\\ &\ll|\overline{G(p_2)}f_{1,0}(a_1cp_1)-\overline{G(p_2')}f_{1,0}(a_1bcp_1)|+|\overline{G(p_2)}f_{1,0}(a_2cp_1)-\overline{G(p_2')}f_{1,0}(a_2bcp_1)|\end{aligned}$$

for all but finitely many p_1, p_2, and thus by further application of the triangle inequality

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}\mathbb{E}_{d\le p_2<(1+\varepsilon)d;\,p_2=c\ (q)} \times \mathbb{E}_{d\le p_2'<(1+\varepsilon)d;\,p_2'=bc\ (q)}|f_{1,0}(a_1cp_1)f_{1,0}(a_2bcp_1)-f_{1,0}(a_1bcp_1)f_{1,0}(a_2cp_1)|\ll\varepsilon.$$

As the expression being averaged does not depend on d, p_2, p_2', this bound simplifies to

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_1cp_1)f_{1,0}(a_2bcp_1)-f_{1,0}(a_1bcp_1)f_{1,0}(a_2cp_1)|\ll\varepsilon$$

and by the prime number theorem in arithmetic progressions and the periodicity of $f_{1,0}$, this simplifies further (see Lemma 2.6) to give the desired bound (23). □

Let a be an integer, and let b be coprime to q. Applying (23) with $a_1=a$ and $a_2=a_0c'$ for c' coprime to q, and averaging, we conclude that

$$\mathbb{E}_{c'\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(ac)f_{1,0}(a_0bcc')-f_{1,0}(abc)f_{1,0}(a_0cc')|\ll\varepsilon$$

and hence

$$\mathbb{E}_{c'\in(\mathbb{Z}/q\mathbb{Z})^\times}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\left|f_{1,0}(ac)f_{1,0}(a_0bcc')\overline{f_{1,0}(a_0cc')}-f_{1,0}(abc)|f_{1,0}(a_0cc')|^2\right|\ll\varepsilon.$$

By the triangle inequality, this implies that

$$\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\Big|f_{1,0}(ac)\mathbb{E}_{c'\in(\mathbb{Z}/q\mathbb{Z})^\times}f_{1,0}(a_0bcc')\overline{f_{1,0}(a_0cc')}-f_{1,0}(abc)\mathbb{E}_{c'\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0cc')|^2\Big|\ll\varepsilon.$$

Making the change of variables $c''=cc'$, this is

$$\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}\Big|f_{1,0}(ac)\mathbb{E}_{c''\in(\mathbb{Z}/q\mathbb{Z})^\times}f_{1,0}(a_0bc'')\overline{f_{1,0}(a_0c'')}-f_{1,0}(abc)\mathbb{E}_{c''\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0c'')|^2\Big|\ll\varepsilon.$$

If we define the function $\psi:(\mathbb{Z}/q\mathbb{Z})^\times\to\mathbb{C}$ by

$$\psi(b):=\frac{\mathbb{E}_{c''\in(\mathbb{Z}/q\mathbb{Z})^\times}f_{1,0}(a_0bc'')\overline{f_{1,0}(a_0c'')}}{\mathbb{E}_{c''\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0c'')|^2}$$

then by (22) and Cauchy–Schwarz, we have $\psi(b)=O(1)$ for all $b\in(\mathbb{Z}/q\mathbb{Z})^\times$, and

$$\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(ac)\psi(b)-f_{1,0}(abc)|\ll\varepsilon \tag{27}$$

for all $a\in\mathbb{Z}/q\mathbb{Z}$ and $b\in(\mathbb{Z}/q\mathbb{Z})^\times$.

By definition, $\psi(1)=1$. Next, we establish an approximate multiplicativity property of ψ, known as the *quasimorphism* property [Kotschick 2004] in the literature. If $b_1,b_2\in(\mathbb{Z}/q\mathbb{Z})^\times$, then from three applications of (27) one has

$$\begin{aligned}\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0c)\psi(b_1)-f_{1,0}(a_0b_1c)|&\ll\varepsilon\\ \mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0b_1c)\psi(b_2)-f_{1,0}(a_0b_1b_2c)|&\ll\varepsilon\\ \mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0c)\psi(b_1b_2)-f_{1,0}(a_0b_1b_2c)|&\ll\varepsilon.\end{aligned}$$

Applying the triangle inequality (after multiplying the first inequality by $|\psi(b_2)|$, we conclude that

$$\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(a_0c)(\psi(b_1b_2)-\psi(b_1)\psi(b_2))|\ll\varepsilon$$

and hence by (22) we have the *quasimorphism equation*

$$\psi(b_1b_2)=\psi(b_1)\psi(b_2)+O(\varepsilon).$$

We now apply a stability theorem to replace this quasimorphism on $(\mathbb{Z}/q\mathbb{Z})^\times$ by a homomorphism (i.e., a Dirichlet character).

Lemma 2.8 (stability of Dirichlet characters). *Let $\varepsilon>0$, and let $\psi:(\mathbb{Z}/q\mathbb{Z})^\times\to\mathbb{C}$ be a function obeying the bound $\psi(b)=O(1)$ for all $b\in\mathbb{Z}/q\mathbb{Z}$, the identity $\psi(1)=1$, and the quasimorphism equation $\psi(b_1b_2)=\psi(b_1)\psi(b_2)+O(\varepsilon)$ for all $b_1,b_2\in(\mathbb{Z}/q\mathbb{Z})^\times$. Then there exists a Dirichlet character $\chi:(\mathbb{Z}/q\mathbb{Z})^\times\to\mathbb{S}^1$ of period q such that $\psi(b)=\chi(b)+O(\varepsilon)$ for all $b\in(\mathbb{Z}/q\mathbb{Z})^\times$.*

This lemma is a special case of Kazhdan [1982],[10] and also follows from [Balog et al. 2013, Proposition 5.3] (which cites [Babai et al. 2003] for a more general result), but for the convenience of the reader we give a self-contained proof here.

[10] We thank Assaf Naor for this reference. Ben Green also pointed out to us the closely related fact that the bounded cohomology of amenable groups is trivial; see for instance [Frigerio 2017, Theorem 3.7].

Proof. We can assume that ε is smaller than any given positive absolute constant, as the claim is trivial otherwise. Since $1 = \psi(1) = \psi(b)\psi(b^{-1}) + O(\varepsilon)$ and $\psi(b^{-1}) = O(1)$, we see that $1 \ll |\psi(b)| \ll 1$ for all $b \in (\mathbb{Z}/q\mathbb{Z})^\times$. We introduce the cocycle $\rho : (\mathbb{Z}/q\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ by defining $\rho(b_1, b_2)$ for $b_1, b_2 \in (\mathbb{Z}/q\mathbb{Z})^\times$ to be the unique complex number of size $O(\varepsilon)$ such that

$$\psi(b_1 b_2) = \psi(b_1)\psi(b_2)\exp(\rho(b_1, b_2)); \tag{28}$$

this is well-defined for ε small enough. For $b_1, b_2, b_3 \in (\mathbb{Z}/q\mathbb{Z})^\times$, we have

$$\psi(b_1b_2b_3) = \psi(b_1b_2)\psi(b_3)\exp(\rho(b_1b_2, b_3)) = \psi(b_1)\psi(b_2)\psi(b_3)\exp(\rho(b_1, b_2) + \rho(b_1b_2, b_3))$$

and

$$\psi(b_1b_2b_3) = \psi(b_1)\psi(b_2b_3)\exp(\rho(b_2, b_3)) = \psi(b_1)\psi(b_2)\psi(b_3)\exp(\rho(b_1, b_2b_3) + \rho(b_2, b_3))$$

which on taking logarithms yields (for ε small enough) the *cocycle equation*

$$\rho(b_1, b_2) + \rho(b_1b_2, b_3) = \rho(b_1, b_2b_3) + \rho(b_2, b_3).$$

Averaging in b_3, we conclude the *coboundary equation*

$$\rho(b_1, b_2) + \phi(b_1b_2) = \phi(b_1) + \phi(b_2)$$

where $\phi(b) := \mathbb{E}_{b_3 \in (\mathbb{Z}/q\mathbb{Z})^\times}\rho(b, b_3)$. If we then define the function $\chi : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ by

$$\chi(b) := \psi(b)\exp(\phi(b)),$$

then $\psi(b) = \chi(b) + O(\varepsilon)$ for all $b \in (\mathbb{Z}/q\mathbb{Z})^\times$, and from (28) we have

$$\chi(b_1b_2) = \chi(b_1)\chi(b_2)$$

for all $b_1, b_2 \in (\mathbb{Z}/q\mathbb{Z})^\times$, thus $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ is a homomorphism and therefore a Dirichlet character of period q. The claim follows. □

Let χ be the Dirichlet character of period q provided by the above lemma, then from (27) and the triangle inequality we have the *approximate isotopy equation*

$$\mathbb{E}_{c \in (\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(ac)\chi(b) - f_{1,0}(abc)| \ll \varepsilon$$

for all $a \in \mathbb{Z}/q\mathbb{Z}$ and $b \in (\mathbb{Z}/q\mathbb{Z})^\times$. We can rearrange this as

$$\mathbb{E}_{c \in (\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(ac) - \overline{\chi(b)}f_{1,0}(abc)| \ll \varepsilon$$

and average in b to conclude that

$$\mathbb{E}_{c \in (\mathbb{Z}/q\mathbb{Z})^\times}|f_{1,0}(ac) - \tilde{f}(ac)| \ll \varepsilon \tag{29}$$

for all a, where $\tilde{f} : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ is the function

$$\tilde{f}(a) := \mathbb{E}_{b \in (\mathbb{Z}/q\mathbb{Z})^\times}\overline{\chi(b)}f_{1,0}(ab).$$

Observe that $\tilde f$ is χ-isotypic in the sense that

$$\tilde f(ab) = \chi(b)\tilde f(a)$$

whenever $a \in \mathbb{Z}/q\mathbb{Z}$ and $b \in (\mathbb{Z}/q\mathbb{Z})^\times$.

From (29) and (22), one has

$$\mathbb{E}_{c\in(\mathbb{Z}/q\mathbb{Z})^\times} |\tilde f(a_0 c)| \gg 1$$

and hence by the χ-isotypy of $\tilde f$

$$|\tilde f(a_0)| \gg 1. \tag{30}$$

Now we work to control f_d. Let a be an integer. From (29) and the prime number theorem in arithmetic progressions, we have

$$\limsup_{P_1\to\infty} \limsup_{P_2\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log}_{p_2\le P_2} |f_{1,0}(ap_1p_2) - \tilde f(ap_1p_2)| \ll \varepsilon$$

From this, (18), (19), and the triangle inequality, we conclude that

$$\limsup_{P_1\to\infty} \limsup_{P_2\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log}_{p_2\le P_2} |f_{p_1p_2}(a) - \overline{G(p_1)}\,\overline{G(p_2)}\tilde f(ap_1p_2)| \ll \varepsilon.$$

Using the χ-isotopy of $\tilde f$, we can write this as

$$\limsup_{P_1\to\infty} \limsup_{P_2\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log}_{p_2\le P_2} |f_{p_1p_2}(a) - \bar G\chi(p_1)\bar G\chi(p_2)\tilde f(a)| \ll \varepsilon. \tag{31}$$

This has the following useful consequence.

Lemma 2.9 (isotopy). *Let the notation be as above. Let a be an integer and let b be an integer coprime to q. Then we have*

$$\limsup_{X\to\infty} \mathbb{E}^{\log\log}_{d\le X} |f_d(ab) - \chi(b) f_d(a)| \ll \varepsilon.$$

Proof. It suffices to prove the claim with an arbitrary generalised limit $\lim^*_{X\to\infty}$ in place of $\limsup_{X\to\infty}$. From (31) we have

$$\mathbb{E}^{\log,*}_{p_1\in\boldsymbol{P}} \mathbb{E}^{\log,*}_{p_2\in\boldsymbol{P}} |f_{p_1p_2}(a) - \bar G\chi(p_1)\bar G\chi(p_2)\tilde f(a)| \ll \varepsilon$$

and

$$\mathbb{E}^{\log,*}_{p_1\in\boldsymbol{P}} \mathbb{E}^{\log,*}_{p_2\in\boldsymbol{P}} |f_{p_1p_2}(ab) - \bar G\chi(p_1)\bar G\chi(p_2)\tilde f(ab)| \ll \varepsilon.$$

As $\tilde f$ is isotypic, $\tilde f(ab) = \chi(b)\tilde f(a)$. From the triangle inequality and (18), we conclude that

$$\mathbb{E}^{\log,*}_{p_1\in\boldsymbol{P}} \mathbb{E}^{\log,*}_{p_2\in\boldsymbol{P}} |f_{p_1p_2}(ab) - \chi(b) f_{p_1p_2}(a)| \ll \varepsilon.$$

On the other hand, since $d \mapsto |f_d(ab) - \chi(b) f_d(a)|$ is bounded log-Lipschitz by (16), by Lemma 2.6 for any p_1 we have

$$\mathbb{E}^{\log,*}_{p_2\in\boldsymbol{P}} |f_{p_1p_2}(ab) - \chi(b) f_{p_1p_2}(a)| = \mathbb{E}^{\log\log,*}_{d\in\mathbb{N}} |f_d(ab) - \chi(b) f_d(a)|,$$

and now the claim now follows by taking the average $\mathbb{E}^{\log,*}_{p_1\in \boldsymbol{P}}$ on both sides. $\square$

Now we derive another consequence of (31). Let $x>0$ be a positive real, and let a be an integer. From (31) we have

$$\mathbb{E}^{\log,**}_{p_1'\in \boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}}|f_{p_1'p_2}(a)-\bar{G}\chi(p_1')\bar{G}\chi(p_2)\tilde{f}(a)|\ll\varepsilon.$$

By the prime number theorem, this can also be written as

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\mathbb{E}_{xp_1\le p_1'\le(1+\varepsilon)xp_1}\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}}|f_{p_1'p_2}(a)-\bar{G}\chi(p_1')\bar{G}\chi(p_2)\tilde{f}(a)|\ll\varepsilon.$$

From (18) we have

$$1-\varepsilon\le|\bar{G}\chi(p_1)|,|\bar{G}\chi(p_1')|,|\bar{G}\chi(p_2)|\le 1$$

for all but finitely many p_1, p_1', p_2, so that

$$\begin{aligned}&|f_{p_1'p_2}(a)-\bar{G}\chi(p_1')G\bar{\chi}(p_1)f_{p_1p_2}(a)|\\ &\qquad\ll|f_{p_1p_2}(a)-\bar{G}\chi(p_1)\bar{G}\chi(p_2)\tilde{f}(a)|+|f_{p_1'p_2}(a)-\bar{G}\chi(p_1')\bar{G}\chi(p_2)\tilde{f}(a)|+O(\varepsilon).\end{aligned}$$

Thus by the triangle inequality we have

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\mathbb{E}_{xp_1\le p_1'\le(1+\varepsilon)xp_1}\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}}|f_{p_1'p_2}(a)-\bar{G}\chi(p_1')G\bar{\chi}(p_1)f_{p_1p_2}(a)|\ll\varepsilon.$$

From (16) we have $f_{p_1'p_2}(a)=f_{xp_1p_2}(a)+O(\varepsilon)$ (recall that f_d is defined for any real $d>0$), thus

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\mathbb{E}_{xp_1\le p_1'\le(1+\varepsilon)xp_1}\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}}|f_{xp_1p_2}(a)-\bar{G}\chi(p_1')G\bar{\chi}(p_1)f_{p_1p_2}(a)|\ll\varepsilon$$

and thus by the triangle inequality

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}}|f_{xp_1p_2}(a)-\alpha_{p_1}(x)f_{p_1p_2}(a)|\ll\varepsilon,$$

where

$$\alpha_{p_1}(x):=\mathbb{E}_{xp_1\le p_1'<(1+\varepsilon)xp_1}\bar{G}\chi-p_1')G\bar{\chi}(p_1).$$

By Lemma 2.6, this implies that

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}|f_{xd}(a)-\alpha_{p_1}(x)f_d(a)|\ll\varepsilon \tag{32}$$

which by the triangle inequality implies that

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}|f_{xd}(a)-\alpha(x)f_d(a)|\ll\varepsilon \tag{33}$$

where

$$\alpha(x):=\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}}\alpha_{p_1}(x).$$

By construction, we have $\alpha(x)=O(1)$ for all x. Setting $a=a_0$ in (33) and using (16) to write $f_{xd}(a)=f_d(a)+O(\varepsilon)$ for $|x-1|\le\varepsilon$, we deduce from (17) that $\alpha(x)=1+O(\varepsilon)$ for $|x-1|\le\varepsilon$.

Next, for $x, y > 0$, we have the estimates

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{xd}(a_0) - \alpha(x) f_d(a_0)| \ll \varepsilon$$
$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{xyd}(a_0) - \alpha(y) f_{xd}(a_0)| \ll \varepsilon$$
$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{xyd}(a_0) - \alpha(xy) f_d(a_0)| \ll \varepsilon,$$

which by the triangle inequality and (17) implies the quasimorphism equation

$$\alpha(xy) = \alpha(x)\alpha(y) + O(\varepsilon).$$

We now require the archimedean analogue of Lemma 2.8 (which is also a special case of the results of [Kazhdan 1982]).

Lemma 2.10 (stability of archimedean characters). *Let $\alpha : (0, +\infty) \to \mathbb{C}$ be any function obeying the bound $\alpha(x) = O(1)$ for all $x > 0$, such that $\alpha(x) = 1 + O(\varepsilon)$ when $|x-1| \le \varepsilon$, and $\alpha(xy) = \alpha(x)\alpha(y) + O(\varepsilon)$ for all $x, y > 0$. Then there exists a real number t such that $\alpha(x) = x^{-\mathrm{it}} + O(\varepsilon)$ for all $x > 0$.*

Proof. As before, we can assume ε is smaller than any given positive constant, as the claim is trivial otherwise. Since $\alpha(1) = 1 + O(\varepsilon)$ and $\alpha(1) = \alpha(x)\alpha(1/x) + O(\varepsilon)$, we have the bounds $1 \ll |\alpha(x)| \ll 1$ for all x. By construction, we also have $\alpha(xy) = \alpha(x) + O(\varepsilon)$ whenever $1 \le y \le 1 + \varepsilon$. By replacing α with the discretised version

$$\alpha_1(x) := \begin{cases} \alpha\left(\varepsilon^2 \left\lfloor \frac{x}{\varepsilon^2} \right\rfloor\right) & x \ge \varepsilon, \\ \alpha\left(\frac{1}{n}\right) & x \in \left(\frac{1}{n+1}, \frac{1}{n}\right],\ 0 < x < \varepsilon, \end{cases}$$

we may assume that α is Lebesgue measurable. The function α_1 continues to enjoy the same properties as α, since $\alpha_1(x) = \alpha(x) + O(\varepsilon)$ for all $x > 0$. To simplify notation, we denote α_1 by α in what follows.

We introduce the cocycle $\rho : (0, +\infty) \times (0, +\infty) \to \mathbb{C}$ by defining $\rho(x_1, x_2)$ for $x_1, x_2 > 0$ to be the unique complex number of size $O(\varepsilon)$ such that

$$\alpha(x_1 x_2) = \alpha(x_1)\alpha(x_2)\exp(\rho(x_1, x_2)); \tag{34}$$

this is well-defined and measurable for ε small enough. Arguing exactly as in the proof of Lemma 2.8, we obtain the cocycle equation

$$\rho(x_1, x_2) + \rho(x_1 x_2, x_3) = \rho(x_1, x_2 x_3) + \rho(x_2, x_3).$$

Taking an asymptotic logarithmic average in x_3, we conclude the coboundary equation

$$\rho(x_1, x_2) + \phi(x_1 x_2) = \phi(x_1) + \phi(x_2) \tag{35}$$

where

$$\phi(x) := \widetilde{\lim_{M\to\infty}} \frac{1}{\log M} \int_1^M \rho(x, x_3) \frac{dx_3}{x_3}.$$

If we then define the function $\tilde{\alpha} : (0, +\infty) \to \mathbb{C}$ by

$$\tilde{\alpha}(x) := \alpha(x)\exp(\phi(x))$$

then $\tilde{\alpha}(x) = \alpha(x) + O(\varepsilon)$ for all $x > 0$, and from (34) and (35) we have

$$\tilde{\alpha}(xy) = \tilde{\alpha}(x)\tilde{\alpha}(y)$$

for all $x, y > 0$, thus $\tilde{\alpha} : (0, +\infty) \to \mathbb{C}$ is a homomorphism. Also, by construction one has $\tilde{\alpha}(x) = O(1)$ for all x, so $\tilde{\alpha}$ in fact takes values in the unit circle $\mathbb{S}^1$. We have $\tilde{\alpha}(x) = 1 + O(\varepsilon)$ when $|x - 1| \le \varepsilon$, and we will use this additional information to show that $\tilde{\alpha}(x) = x^{\mathrm{it}}$ for some real t and all $x > 0$.

If $|x - 1| \le \varepsilon/n$ for some natural number n, then $\tilde{\alpha}(x)^n, \tilde{\alpha}(x) = 1 + O(\varepsilon)$, which implies that $\tilde{\alpha}(x) = 1 + O(\varepsilon/n)$. This implies that $\tilde{\alpha}(x) = 1 + O(|x - 1|)$, and so $\tilde{\alpha}$ is continuous at 1 and hence continuous on all of $(0, +\infty)$. Next, if $x_0 := 1 + \varepsilon$ then we have $\tilde{\alpha}(x_0) = x_0^{\mathrm{it}}$ for some $t = O(1)$; taking roots we conclude that $\tilde{\alpha}(x_0^{1/n}) = (x_0^{1/n})^{\mathrm{it}}$ for all natural numbers n, and hence $\tilde{\alpha}(x_0^{m/n}) = (x_0^{m/n})^{\mathrm{it}}$ for all natural numbers n and integers m. By continuity we conclude that $\tilde{\alpha}(x) = x^{\mathrm{it}}$ for all $x \in (0, +\infty)$, as required. □

From the above lemma, we conclude that there is a real number t with the property that for every integer a and real $x > 0$, one has

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{xd}(a) - x^{-\mathrm{it}} f_d(a)| \ll \varepsilon. \tag{36}$$

In particular, for every prime p_1, one has

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{p_1 d}(a_0) - p_1^{-\mathrm{it}} f_d(a_0)| \ll \varepsilon,$$

and thus

$$\mathbb{E}^{\log,**}_{p_1\in\boldsymbol{P}} \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{p_1 d}(a_0) - p_1^{-\mathrm{it}} f_d(a_0)| \ll \varepsilon \tag{37}$$

On the other hand, from Proposition 2.3 one has that if P_1 is sufficiently large depending on a_0, ε, then

$$\sup_{d>0} \mathbb{E}^{\log}_{p_1\le P_1} |f_{p_1 d}(a_0)G(p_1) - f_d(a_0 p_1)| \ll \varepsilon.$$

Hence on averaging in d and taking limits in the d average and then in the p_1 average, we conclude that

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_{p_1 d}(a_0)G(p_1) - f_d(a_0 p_1)| \ll \varepsilon. \tag{38}$$

Meanwhile, from Lemma 2.9 we have

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_d(a_0 p_1) - \chi(p_1) f_d(a_0)| \ll \varepsilon$$

for all sufficiently large p_1, and thus

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |f_d(a_0 p_1) - \chi(p_1) f_d(a_0)| \ll \varepsilon. \tag{39}$$

Applying the triangle inequality to (37), (38), (39), we obtain

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |G(p_1) - \chi(p_1) p_1^{\mathrm{it}}||f_d(a_0)| \ll \varepsilon$$

and hence by (17) we have

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} |G(p_1)-\chi(p_1)p_1^{\mathrm{it}}| \ll \varepsilon.$$

To summarise the above analysis, we have shown that for every $\varepsilon>0$ there exists a Dirichlet character $\chi=\chi_\varepsilon$ and a real number $t=t_\varepsilon$ such that

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} |G(p_1)-\chi_\varepsilon(p_1)p_1^{it_\varepsilon}| \ll \varepsilon.$$

A priori, the character χ_ε and the real number t_ε depend on ε. But if $\varepsilon,\varepsilon'>0$ are sufficiently small, we have from the triangle inequality that

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} |\chi_{\varepsilon'}(p_1)p_1^{it_{\varepsilon'}}-\chi_\varepsilon(p_1)p_1^{it_\varepsilon}| \ll \varepsilon+\varepsilon'.$$

But from the prime number theorem in arithmetic progressions and partial summation, we see that the left-hand side is $\gg 1$ unless $t_\varepsilon=t_{\varepsilon'}$ and the Dirichlet characters are *cotrained* in the sense that they are both induced from the same primitive character χ. We conclude that there exists a primitive character χ independent of ε, and a real number t_0 independent of ε, such that $t_\varepsilon=t_0$ and χ_ε is induced from χ for ε sufficiently small. In particular, as $\chi_\varepsilon(p_1)$ and $\chi(p_1)$ agree for all but $O_\varepsilon(1)$ primes p_1, we have for each $\varepsilon>0$ that

$$\limsup_{P_1\to\infty} \mathbb{E}^{\log}_{p_1\le P_1} |G(p_1)-\chi(p_1)p_1^{it_0}| \ll \varepsilon$$

and thus

$$\mathbb{E}^{\log}_{p_1\in \boldsymbol{P}} |G(p_1)-\chi(p_1)p_1^{it_0}| = 0. \tag{40}$$

Thus G weakly pretends to be the twisted Dirichlet character $n\mapsto n^{it_0}\chi(n)$. This (vacuously) establishes part (i) of Theorem 1.7.

Now let $\varepsilon>0$ be small, and let a be an integer. From (31) (and the fact that χ_ε is induced from χ), and making the dependence of $\tilde f_\varepsilon$ on ε explicit, we have

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}} \mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}} |f_{p_1p_2}(a)-\bar G\chi(p_1)\bar G\chi(p_2)\tilde f_\varepsilon(a)| \ll \varepsilon$$

and hence by (40) and the triangle inequality

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}} \mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}} |f_{p_1p_2}(a)-p_1^{-it_0}p_2^{-it_0}\tilde f_\varepsilon(a)| \ll \varepsilon$$

or equivalently

$$\mathbb{E}^{\log,**}_{p_1\in \boldsymbol{P}} \mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}} |(p_1p_2)^{it_0}f_{p_1p_2}(a)-\tilde f_\varepsilon(a)| \ll \varepsilon.$$

Applying (16), Lemma 2.6 and (36) (where we can in fact take $\varepsilon\to 0$, since the deduction succeeding this formula shows that $t=t_0$ is independent of ε), we have

$$\mathbb{E}^{\log,**}_{p_2\in \boldsymbol{P}} |(p_1p_2)^{it_0}f_{p_1p_2}(a)-\tilde f_\varepsilon(a)| = \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |(p_1d)^{it_0}f_{p_1d}(a)-\tilde f_\varepsilon(a)| = \mathbb{E}^{\log\log,**}_{d\in\mathbb{N}} |d^{it_0}f_d(a)-\tilde f_\varepsilon(a)|$$

for any p_1, and hence

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}|d^{it_0} f_d(a)-\tilde{f}_\varepsilon(a)| \ll \varepsilon.$$

We thus see from the triangle inequality that

$$|\tilde{f}_\varepsilon(a)-\tilde{f}_{\varepsilon'}(a)| \ll \varepsilon+\varepsilon'$$

and so $\tilde{f}_\varepsilon$ converges uniformly to a limit f with

$$|\tilde{f}_\varepsilon(a)-f(a)| \ll \varepsilon \tag{41}$$

and thus by the triangle inequality, we have

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}|d^{it_0} f_d(a)-f(a)| \ll \varepsilon$$

whenever $\varepsilon>0$, which gives

$$\mathbb{E}^{\log\log,**}_{d\in\mathbb{N}}|d^{it_0} f_d(a)-f(a)| = 0. \tag{42}$$

From (17) we see in particular that $f(a_0)\neq 0$. By construction, each $\tilde{f}_\varepsilon$ is χ-isotypic in the sense that $\tilde{f}_\varepsilon(ab)=\chi(b)\tilde{f}_\varepsilon(a)$ whenever a,b are integers with b coprime to the periods of both χ and $\tilde{f}_\varepsilon$. Hence, what remains to be shown is that (42) holds also when taking the average with respect to the ordinary limit.

Now let $\varepsilon>0$ be arbitrary. Inserting (40) into (31), we see that

$$\limsup_{P_1\to\infty}\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)-(p_1p_2)^{-it_0}\tilde{f}_\varepsilon(a)| \ll \varepsilon$$

and hence by (41) and sending $\varepsilon\to 0$ we get

$$\limsup_{P_1\to\infty}\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)-(p_1p_2)^{-it_0}f(a)| = 0.$$

For any $\varepsilon>0$ and any P_1 large enough in terms of ε, we apply Lemma 2.6, Proposition 2.3, formula (40) and Lemma 2.9 to write

$$\begin{aligned}
&\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log}_{p_2\le P_2}|f_{p_1p_2}(a)-(p_1p_2)^{-it_0}f(a)|\\
&\qquad=\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log\log}_{d\le P_2}|f_{p_1d}(a)-(p_1d)^{-it_0}f(a)|\\
&\qquad=\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log\log}_{d\le P_2}|\overline{G(p_1)}f_d(ap_1)-(p_1d)^{-it_0}f(a)|+O(\varepsilon)\\
&\qquad=\limsup_{P_2\to\infty}\mathbb{E}^{\log}_{p_1\le P_1}\mathbb{E}^{\log\log}_{d\le P_2}|p_1^{-it_0}\overline{\chi}(p_1)f_d(ap_1)-(p_1d)^{-it_0}f(a)|+O(\varepsilon)\\
&\qquad=\limsup_{P_2\to\infty}\mathbb{E}^{\log\log}_{d\le P_2}|f_d(a)-d^{-it_0}f(a)|+O(\varepsilon),
\end{aligned}$$

and hence, sending $\varepsilon\to 0$, we obtain

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}}|f_d(a)-d^{-it_0}f(a)| = 0.$$

This establishes part (ii) of Theorem 1.7 (recalling as before that as G weakly pretends to be a twisted Dirichlet character $n \mapsto \chi(n)n^{\mathrm{it}}$, it can only weakly pretend to be another twisted Dirichlet character $n \mapsto \chi'(n)n^{it'}$ if $t = t'$ and χ, χ' are cotrained).

3. Proofs of corollaries

In this section we use Theorem 1.7 to prove Corollaries 1.8, 1.13, 1.14. We begin with Corollary 1.8.

Proof of Corollary 1.8. Suppose the claim failed, then we can find $k, g_1, \dots, g_k$ as in that corollary, as well as $h_1, \dots, h_k \in \mathbb{Z}$ and $\varepsilon > 0$, such that the set

$$\mathcal{X} := \{X \in \mathbb{N} : |\mathbb{E}_{n \le X} g_1(n+h_1) \cdots g_k(n+h_k)| > \varepsilon\}$$

does not have logarithmic Banach density zero. In particular, one can find sequences $X_i \ge \omega_i \to \infty$ and $0 < \delta < \frac{1}{2}$ such that

$$\mathbb{E}^{\log}_{X_i/\omega_i \le X \le X_i} 1_{\mathcal{X}}(x) \ge \delta \tag{43}$$

for all i.

Intuitively, if the exceptional set $\mathcal{X}$ was big in the sense of (43), there would have to be a lot of "points of density" of $\mathcal{X}$ (in a sense to be specified later). To make this rigorous, we introduce for each i the function $a_i : \mathbb{R} \to [0, 1]$ given by

$$a_i(s) := \sum_{X_i/\omega_i \le X \le X_i : X \notin \mathcal{X}} 1_{\log(X-1) < s \le \log X}.$$

Note that $a_i(s)$ is the indicator function of the event that there exists an integer $X \notin \mathcal{X}$ with $X \in [e^s, e^s + 1)$ and $X_i/\omega_i \le X \le X_i$.

The function a_i is a piecewise constant function supported on an interval of length $(1 + o_{i\to\infty}(1)) \log \omega_i$ and has integral

$$\begin{aligned}\int_{\mathbb{R}} a_i(s)\, ds &= \sum_{X_i/\omega_i \le X \le X_i : X \notin \mathcal{X}} \log \frac{X}{X-1} \\ &= \Bigg(\sum_{X_i/\omega_i \le X \le X_i : X \notin \mathcal{X}} \frac{1}{X}\Bigg) + O(1) \\ &= \log \omega_i + O(1) - \sum_{X_i/\omega_i \le X \le X_i : X \in \mathcal{X}} \frac{1}{X} \\ &\le (1 - \delta + o_{i\to\infty}(1)) \log \omega_i.\end{aligned}$$

We introduce the one-sided Hardy–Littlewood maximal function

$$Ma_i(s) := \sup_{r>0} \frac{1}{r} \int_{s-r}^{s} a_i(s')\, ds'.$$

It is a well-known consequence of the rising sun lemma [Riesz 1932] that one has the Hardy–Littlewood maximal inequality

$$m(\{s \in \mathbb{R} : Ma_i(s) \geq \lambda\}) \leq \frac{1}{\lambda} \int_{\mathbb{R}} a_i(s)\, ds$$

for any $\lambda > 0$, where m denotes Lebesgue measure. Applying this with $\lambda := (1-\delta)^{1/2}$, we conclude that

$$m(\{s \in \mathbb{R} : Ma_i(s) \geq (1-\delta)^{1/2}\}) \leq ((1-\delta)^{1/2} + o_{i\to\infty}(1)) \log \omega_i.$$

In particular, one can find a real number s_i with

$$\log X_i - ((1-\delta)^{1/2} + o_{i\to\infty}(1)) \log \omega_i \leq s_i \leq \log X_i \tag{44}$$

such that

$$Ma_i(s_i) < (1-\delta)^{1/2}$$

which implies that

$$\int_{s_i - r}^{s_i} a_i(t)\, dt \leq (1-\delta)^{1/2} r \tag{45}$$

for all $r > 0$. Informally, the estimate (45) asserts that the natural number $\lfloor \exp(s_i) \rfloor$ is a "multiplicative point of density" for the exceptional set $\mathcal{X}$.

By passing to subsequences, and using a diagonalisation argument, we may assume that the limits

$$f_d(a) := \lim_{i\to\infty} \mathbb{E}_{n \leq \lfloor \exp(s_i) \rfloor / d}\, g_1(n + ah_1) \cdots g_k(n + ah_k), \tag{46}$$

exist for every natural number d and integer a. In particular, the limit of the right-hand side of (46) is the same along any generalised limit $\lim^{**}$. If we now apply Theorem 1.7(i) to a generalised limit of the form

$$\lim^{*}_{X\to\infty} f(X) := \lim^{**}_{i\to\infty} f(\lfloor \exp(s_i) \rfloor),$$

where $\lim^{**}$ is any generalised limit, we conclude that

$$\mathbb{E}^{\log\log}_{d \in \mathbb{N}} |f_d(1)| = 0.$$

Thus, if we let $\mu > 0$ denote a small constant (depending on δ, ε) to be chosen later, and D is sufficiently large depending on μ, we have

$$\mathbb{E}^{\log\log}_{d \leq D} |f_d(1)| \leq \mu.$$

Thus by the triangle inequality

$$\limsup_{i\to\infty} \mathbb{E}^{\log\log}_{d \leq D} |\mathbb{E}_{n \leq \lfloor \exp(s_i) \rfloor / d}\, g_1(n + h_1) \cdots g_k(n + h_k)| \leq \mu,$$

and hence for all sufficiently large i (depending on $\delta, \varepsilon, \mu, D$) we find

$$\mathbb{E}^{\log\log}_{d \leq D} |\mathbb{E}_{n \leq \lfloor \exp(s_i) \rfloor / d}\, g_1(n + h_1) \cdots g_k(n + h_k)| \leq 2\mu.$$

This implies

$$\sum_{\log D \le d \le D} \frac{1}{d \log d} |\mathbb{E}_{n \le \lfloor \exp(s_i) \rfloor / d} g_1(n+h_1) \cdots g_k(n+h_k)| \ll \mu \log\log D,$$

say. In particular, by Markov's inequality one has

$$|\mathbb{E}_{n \le \lfloor \exp(s_i) \rfloor / d} g_1(n+h_1) \cdots g_k(n+h_k)| \le \frac{\varepsilon}{2} \tag{47}$$

for all $\log D \le d \le D$ outside of an exceptional set $\mathcal{D}_i$ with

$$\sum_{d \in \mathcal{D}_i} \frac{1}{d \log d} \ll \frac{\mu}{\varepsilon} \log\log D. \tag{48}$$

If $\log D \le d \le D$ lies outside of $\mathcal{D}_i$, then one has

$$|\mathbb{E}_{n \le X} g_1(n+h_1) \cdots g_k(n+h_k)| < \varepsilon$$

for all X between $\exp(s_i)/(d+1) - 1$ and $\exp(s_i)/d + 1$. In particular, all such X lie outside of $\mathcal{X}$. Using (44) (which places $\exp(s_i)/d$ below X_i and well above X_i/ω_i), we conclude that

$$a_i(t) = 1$$

on the interval $[s_i - \log(d+1), s_i - \log(d)]$. In particular,

$$\int_d^{d+1} a_i(s_i - \log u)\, du = 1.$$

For $d \in \mathcal{D}_i$ we use the trivial bound

$$\int_d^{d+1} a_i(s_i - \log u)\, du \ge 0.$$

From (48) we conclude that

$$\sum_{\log D \le d \le D} \frac{1}{d \log d} \int_d^{d+1} a_i(s_i - \log u)\, du \ge \left(1 - O\left(\frac{\mu}{\varepsilon}\right)\right) \log\log D. \tag{49}$$

The left-hand side, up to errors that can be absorbed into the $O\left(\frac{\mu}{\varepsilon}\right) \log\log D$ term, can be rewritten as

$$\int_{\log D}^{D} a_i(s_i - \log u) \frac{du}{u \log u}$$

which by the change of variables $s = s_i - \log u$ becomes

$$\int_{s_i - \log D}^{s_i - \log\log D} a_i(s) \frac{ds}{s_i - s}.$$

However, from Fubini's theorem and (45) we have

$$\begin{aligned}
\int_{s_i-\log D}^{s_i-\log\log D} a_i(s)\frac{ds}{s_i-s} &= \int_{s_i-\log D}^{s_i-\log\log D} a_i(s)\left(\int_{s_i-\log D}^{s}\frac{dt}{(s_i-t)^2}+\frac{1}{\log D}\right)ds\\
&= \int_{s_i-\log D}^{s_i-\log\log D}\left(\int_{t}^{s_i-\log\log D} a_i(s)\,ds\right)\frac{dt}{(s_i-t)^2}+\frac{1}{\log D}\int_{s_i-\log D}^{s_i-\log\log D} a_i(s)\,ds\\
&\le \int_{s_i-\log D}^{s_i-\log\log D}\left(\int_{t}^{s_i} a_i(s)\,ds\right)\frac{dt}{(s_i-t)^2}+\frac{1}{\log D}\int_{s_i-\log D}^{s_i} a_i(s)\,ds\\
&\le \int_{s_i-\log D}^{s_i-\log\log D}(1-\delta)^{1/2}(s_i-t)\frac{dt}{(s_i-t)^2}+\frac{1}{\log D}(1-\delta)^{1/2}\log D\\
&= (1-\delta)^{1/2}(\log\log D-\log\log\log D+1)
\end{aligned}$$

and the right-hand side is equal to $(1-\delta)^{1/2}\log\log D$ up to errors that can be absorbed into the $O\left(\frac{\mu}{\varepsilon}\right)\log\log D$ term. For μ small enough, this gives a contradiction when compared with (49), proving Corollary 1.8(i).

We are left with proving part (ii) of Corollary 1.8. Since sets of logarithmic Banach density zero automatically have logarithmic density zero, we already know from Corollary 1.8(i) that for each tuple $(h_1,\dots,h_k)$ of integers and every $m \ge 1$, there is a set $\mathcal{X}_{h_1,\dots,h_k,m}$ of logarithmic density zero such that

$$|\mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k)| \le \frac{1}{m}$$

for all X outside of $\mathcal{X}_{h_1,\dots,h_k,m}$. Since the number of tuples $(h_1,\dots,h_k,m)$ is countable, a standard diagonalisation construction then gives a further set $\mathcal{X}_0$, still of logarithmic density zero, such that for each $h_1,\dots,h_k,m$, all but finitely many of the elements of $\mathcal{X}_{h_1,\dots,h_k,m}$ are contained in $\mathcal{X}_0$. For instance, one could remove finitely many elements from $\mathcal{X}_{h_1,\dots,h_k,m}$ to create a subset $\mathcal{X}'_{h_1,\dots,h_k,m}$ with the property that

$$\mathbb{E}^{\log}_{X\le Y} 1_{\mathcal{X}'_{h_1,\dots,h_k,m}}(X) \le 2^{-h_1-\dots-h_k-m}$$

for all $Y \ge 1$, and then take $\mathcal{X}_0$ to be the union of all the $\mathcal{X}'_{h_1,\dots,h_k,m}$, which thus differs from a finite union of these sets by a set of arbitrarily small logarithmic density (and finite unions of the sets $\mathcal{X}'_{h_1,\dots,h_k,m}$ have logarithmic density 0). By construction one then has

$$\limsup_{X\to\infty;\, X\notin\mathcal{X}_0} |\mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k)| \le \frac{1}{m}$$

for all $h_1,\dots,h_k,m$, and the claim follows. □

Remark 3.1. An inspection of the above argument shows that one could have replaced the sequence $n \mapsto g_1(n+h_1)\cdots g_1(n+h_k)$ by any other bounded sequence $n \mapsto F(n)$ for which the analogue of Theorem 1.7(i) holds, or more precisely that

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}}|\lim^*_{X\to\infty}\mathbb{E}_{n\le X/d}F(n)| = 0$$

for any generalised limit $\lim^*_{X\to\infty}$.

Next we prove Corollary 1.13.

Proof of Corollary 1.13. By Corollary 1.8, we are done unless $g_1 g_2$ weakly pretends to be a twisted Dirichlet character $n \mapsto \chi(n) n^{\mathrm{it}}$, so suppose that this is indeed the case for some χ and t. Then for any generalised limit $\lim^*_{X\to\infty}$, the corresponding correlations $f_d(a)$ defined by (4) obey the property (5) for some function $f : \mathbb{Z} \to \mathbb{D}$. If this function f was vanishing at $a = 1$ for every choice of the generalised limit, then one could repeat the proof of Corollary 1.8 to obtain the claim (see Remark 3.1). Thus suppose instead that we can find a generalised limit $\lim^*_{X\to\infty}$ such that $f(1) \neq 0$ for the function f provided by Theorem 1.7(ii). By (5) and the triangle inequality, this implies that

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} f_d(1) d^{\mathrm{it}} = f(1) \neq 0.$$

In particular, for D sufficiently large, one has

$$|\mathbb{E}^{\log\log}_{d\le D} f_d(1) d^{\mathrm{it}}| \gg 1$$

and hence by summation by parts we have

$$|\mathbb{E}^{\log}_{d\le D} f_d(1) d^{\mathrm{it}}| \gg 1$$

for a sequence of arbitrarily large D. If D obeys the above estimate, then by (4) we have

$$|\lim^*_{X\to\infty} \mathbb{E}^{\log}_{d\le D} d^{\mathrm{it}} \mathbb{E}_{n\le X/d} g_1(n+h_1) g_2(n+h_2)| \gg 1$$

and thus there exist arbitrarily large X such that

$$|\mathbb{E}^{\log}_{d\le D} d^{\mathrm{it}} \mathbb{E}_{n\le X/d} g_1(n+h_1) g_2(n+h_2)| \gg 1.$$

This implies that

$$|\mathbb{E}^{\log}_{d\le D} d^{\mathrm{it}} \mathbb{E}_{cX/d\le n\le X/d} g_1(n+h_1) g_2(n+h_2)| \gg 1$$

for some small constant $c > 0$ (not depending on D and X). This yields

$$\left| \sum_{\log D\le d\le D} d^{\mathrm{it}} \sum_{cX/d\le n\le X/d} g_1(n+h_1) g_2(n+h_2) \right| \gg X \log D$$

The left-hand side can be rearranged (discarding negligible errors, assuming D is large enough) as

$$\left| \sum_{cX/D\le n\le X/\log D} \left(\sum_{cX/n\le d\le X/n} d^{\mathrm{it}} \right) g_1(n+h_1) g_2(n+h_2) \right| \gg X \log D.$$

By summation by parts, for $cX/D \le n \le X/\log D$ we have

$$\sum_{cX/n\le d\le X/n} d^{\mathrm{it}} = \alpha n^{-\mathrm{it}} \frac{X}{n} + o_{D\to\infty}(1), \quad \alpha = \frac{X^{\mathrm{it}} - (cX)^{\mathrm{it}} \cdot c}{1+it},$$

where in particular the quantity α is bounded and is independent of n. For D large enough, we conclude that

$$\left|\sum_{cX/D\le n\le X/\log D}\frac{n^{-\mathrm{it}}}{n}g_1(n+h_1)g_2(n+h_2)\right|\gg\log D.$$

and hence

$$|\mathbb{E}^{\log}_{X/D\le n\le X}n^{-\mathrm{it}}g_1(n+h_1)g_2(n+h_2)|\gg 1.$$

Approximating $n^{-\mathrm{it}}$ by $(n+h_1)^{-\mathrm{it}}$, we conclude that there exist arbitrarily large D such that

$$|\mathbb{E}^{\log}_{X/D\le n\le X}(n+h_1)^{-\mathrm{it}}g_1(n+h_1)g_2(n+h_2)|\gg 1$$

for arbitrarily large X. But this contradicts the $k=2$ case of the logarithmically averaged Elliott conjecture [Tao 2016, Corollary 1.5] applied to the functions $n\mapsto n^{-\mathrm{it}}g_1(n)$ and $n\mapsto g_2(n)$ (note that the hypothesis (1) for g_1 implies the same hypothesis for $n\mapsto n^{-\mathrm{it}}g_1(n)$). This completes the proof of part (i) of Corollary 1.13.

Part (ii) of Corollary 1.13 is then deduced from Corollary 1.13(i) using precisely the same diagonalisation argument that was used to deduce Corollary 1.8(ii) from Corollary 1.8(i). □

Remark 3.2. The above argument shows more generally that if the logarithmically averaged Elliott conjecture[11] (resp. Chowla conjecture) is proven for a given value of k, then the unweighted form of the Elliott conjecture (resp. Chowla conjecture) for that value of k holds at almost all scales. (Note in the case of the Chowla conjecture that the parameter t will vanish, since $\lambda^k=1$ for even k and $\lambda^k=\lambda$ does not pretend to be any twisted Dirichlet character for odd k.)

Remark 3.3. With small modifications, we can adapt the above proofs to prove Corollary 1.16. Firstly, by approximating the indicator function $1_{P^+(n)<P^+(n+1)}$ as in [Teräväinen 2018, Section 4] by a linear combination of indicator functions of the form $1_{P^+(n)<n^\alpha,P^+(n+1)<n^\beta}$, we can reduce the proof to showing

$$\lim_{X\to\infty;X\notin\mathcal{X}_0}\mathbb{E}_{n\le X}1_{P^+(n)<n^\alpha}1_{P^+(n+1)<n^\beta}=\rho\left(\frac{1}{\alpha}\right)\rho\left(\frac{1}{\beta}\right),\tag{50}$$

where $\rho(\cdot)$ is the Dickmann function and $\alpha,\beta\in(0,1)$ are any rational numbers. Since the set of rationals is countable, by a diagonal argument (as in the proof of Corollary 1.13(ii)) it suffices to prove (50) with α,β fixed. One starts by proving a version of the structural theorem (Theorem 1.7) in the case of the functions $g_1(n)=1_{P^+(n)<n^\alpha}$, $g_2(n)=1_{P^+(n)<n^\beta}$; these are not quite multiplicative functions, but they can be approximated as $1_{P^+(n)<n^\alpha}=1_{P^+(n)<X^\alpha}+O(1_{P^+(n)\in[(X/\log X)^\alpha,X^\alpha]})$ for $n\in[X/\log X,X]$. The $O(\cdot)$ term has negligible contribution in the entropy decrement argument by standard estimates on smooth numbers, so the proof of Proposition 2.3 goes through for the generalised limits associated to the correlations of g_1 and g_2 with $G\equiv 1$ (so certainly (18) holds). We did not use the specific properties of

[11] One needs the variant where we sum over $X/\omega(X)\le n\le X$ rather than $n\le X$.

g_1, g_2 anywhere else in the proof of Theorem 1.7, so that proof goes through, giving

$$\mathbb{E}_{d\in\mathbb{N}}|\lim_{X\to\infty}^{*}\mathbb{E}_{n\leq X/d}\,g_1(n)g_2(n+1)-c^*|=0 \tag{51}$$

for all generalised limits $\lim^*$ and some constant c^* depending on $\lim^*$. From [Teräväinen 2018, proof of Corollary 1.19], we have a logarithmic version of (50), so following the proof of Corollary 1.13 verbatim, we see that $c^*=\rho(1/\alpha)\rho(1/\beta)$. Then from Remark 3.1 we deduce (50). We leave the details to the interested reader.

Proof of Corollary 1.14. We observe from Corollary 1.8(i) (for odd k) or Corollary 1.13(ii) (for $k=2$) that for any distinct integers $h_1,\dots,h_k$ and $\varepsilon>0$, one has

$$|\mathbb{E}_{n\leq X}\lambda(n+h_1)\cdots\lambda(n+h_k)|\leq\varepsilon$$

for all X outside of a set $\mathcal{X}_{k,\varepsilon}$ of logarithmic Banach density zero, and hence also of logarithmic density zero. The claim then follows by the same diagonalisation argument used to prove Corollary 1.8(ii) and Corollary 1.13(ii). □

4. Consequences of the isotopy formulae

Before proving the isotopy formula in the form of Theorem 1.19, let us state a variant of it that involves the quantities $f_d(a)$ present in Theorem 1.7. In what follows, a sequence b_n of integers is said to be *asymptotically rough* if for any given prime p, one has $p\nmid b_n$ for all sufficiently large n. For instance, any increasing sequence of primes is asymptotically rough, as is the sequence $-1,-1,-1,\dots$.

Lemma 4.1. *Let the notation and hypotheses be as in Theorem 1.7. Let $n\mapsto\chi(n)n^{\mathrm{it}}$ be a twisted Dirichlet character that weakly pretends to be $g_1\cdots g_k$, if one exists; otherwise, choose χ and t arbitrarily. Let a be an arbitrary integer:*

(i) (*Archimedean isotopy*) *For any natural number h, one has*

$$\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\leq X}|f_{hd}(a)-h^{-\mathrm{it}}f_d(a)|=0.$$

(ii) (*Nonarchimedean isotopy*) *For any asymptotically rough sequence b_n of natural numbers, one has*

$$\lim_{n\to\infty}\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\leq X}|f_d(ab_n)-\chi(b_n)f_d(a)|=0.$$

In particular, since the sequence $b_n=-1$ is asymptotically rough, one has

$$\lim_{X\to\infty}\mathbb{E}^{\log\log}_{d\leq X}|f_d(-a)-\chi(-1)f_d(a)|=0. \tag{52}$$

A variant of Lemma 4.1(ii) (for logarithmic averaging, and with b_n specialised to the primes in an arithmetic progression 1 (q) for q a period of χ) was obtained in [Frantzikinakis and Host 2019, Corollary 3.7].

Proof. We may assume without loss of generality that $g_1 \cdots g_k$ weakly pretends to be $n \mapsto \chi(n)n^{\mathrm{it}}$, as the claims follow from Theorem 1.7(i) otherwise. Extracting out the contribution to (5) from multiples of h, we see that

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} |f_{hd}(a) - f(a)(hd)^{-\mathrm{it}}| = 0,$$

and also by (5) we have

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} |f_{d}(a) - f(a)d^{-\mathrm{it}}| = 0.$$

Now the claim follows from the triangle inequality.

To prove Claim (ii),[12] we observe from (5) that

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} |f_{d}(ab_n) - f(ab_n)d^{-\mathrm{it}}| = 0$$

for all n, and

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} |f_{d}(a) - f(a)d^{-\mathrm{it}}| = 0.$$

Putting together the above two equalities we have

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} |f_{d}(ab_n) - \chi(b_n)f_d(a)| = |f(ab_n) - \chi(b_n)f(a)| \tag{53}$$

By Theorem 1.7, f is the uniform limit of χ-isotypic periodic functions F_i. For each such F_i, we have $F_i(ab_n) = \chi(b_n)F_i(a)$ for all sufficiently large n, since the sequence b_n is asymptotically rough. Thus also $f(ab_n) = \chi(b_n)f(a) + o_{n\to\infty}(1)$. Combining this with (53), the claim follows. □

We then use Lemma 4.1 to deduce the isotopy formulae (Theorem 1.19).

Proof of Theorem 1.19. We start with the proof of (i). By a diagonalisation argument, similarly as in the proof of Corollary 1.8(ii), it suffices to show that for any fixed rational $q > 0$ there exists a set $\mathcal{X}_{0,q}$ of logarithmic density 0 such that the claim holds with $\mathcal{X}_{0,q}$ in place of $\mathcal{X}_0$. Next, we argue that it suffices to consider the case $q \in \mathbb{N}$. Suppose that the case $q \in \mathbb{N}$ has been established, and let $q = a/b$ with $a, b \in \mathbb{N}$. Then if $\mathcal{X}_{0,q} := (1/b)\mathcal{X}_{0,a} \cup (1/b)\mathcal{X}_{0,b}$ (which is still a set of logarithmic density zero), we have

$$\begin{aligned}\lim_{X\to\infty;\, X\notin\mathcal{X}_{0,q}} &\Big(\mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k) - \big(\tfrac{a}{b}\big)^{\mathrm{it}}\mathbb{E}_{n\le bX/a} g_1(n+h_1)\cdots g_k(n+h_k)\Big)\\ &= \lim_{X\to\infty;\, X\notin\mathcal{X}_{0,q}} \Big(b^{-\mathrm{it}}\mathbb{E}_{n\le bX} g_1(n+h_1)\cdots g_k(n+h_k) - \big(\tfrac{a}{b}\big)^{\mathrm{it}}\mathbb{E}_{n\le bX/a} g_1(n+h_1)\cdots g_k(n+h_k)\Big) = 0.\end{aligned}$$

Hence we may assume from now on that $q \in \mathbb{N}$. Observe that the statement of Lemma 4.1(i) with $a = 1$ can be written in the form

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}} \Big|\lim^{*}_{X\to\infty} \mathbb{E}_{n\le X/d}(g_1(n+h_1)\cdots g_k(n+h_k) - q^{-\mathrm{it}}\mathbb{E}_{b\in\mathbb{Z}/q\mathbb{Z}} g_1(qn+b+h_1)\cdots g_k(qn+b+h_k))\Big| = 0$$

[12] Note that Lemma 2.9 does not directly imply Claim (ii), since the Dirichlet character present in that lemma depends on the error ε.

for every generalised limit $\lim^*_{X\to\infty}$. By following the proof of Corollary 1.8(i) verbatim (see also Remark 3.1), this leads to

$$\lim_{X\to\infty; X\notin\mathcal{X}_{0,q}} \mathbb{E}_{n\le X}(g_1(n+h_1)\cdots g_k(n+h_k) - q^{-\mathrm{it}}\mathbb{E}_{b\in\mathbb{Z}/q\mathbb{Z}}g_1(qn+b+h_1)\cdots g_k(qn+b+h_k)) = 0 \quad (54)$$

for some set $\mathcal{X}_{0,q}$ of logarithmic density zero. But rewriting (54), it becomes the identity asserted in Theorem 1.19(i).

We turn to the proof of part (ii), which is similar. Again by a diagonalisation argument, it suffices to prove the statement for fixed a rather than all a. From Lemma 4.1(ii) we have

$$\mathbb{E}^{\log\log}_{d\in\mathbb{N}}|\lim^*_{X\to\infty}\mathbb{E}_{n\le X/d}(g_1(n-ah_1)\cdots g_k(n-ah_k) - \chi(-1)g_1(n+ah_1)\cdots g_k(n+ah_k))| = 0$$

for every generalised limit $\lim^*_{X\to\infty}$. Just as in the proof of part (i) of the Theorem, by the proof of Corollary 1.8(i) (see Remark 3.1) we get

$$\lim_{X\to\infty; X\notin\mathcal{X}_{0,a}}|\mathbb{E}_{n\le X}(g_1(n-ah_1)\cdots g_k(n-ah_k) - \chi(-1)g_1(n+ah_1)\cdots g_k(n+ah_k))| = 0$$

for some set $\mathcal{X}_{0,a}$ of logarithmic density zero, and this is what we wished to prove. □

Morally speaking, the archimedean isotopy formula implies that the argument of the correlation (9) becomes equidistributed at large scales whenever $t \neq 0$. Unfortunately we cannot quite establish this claim as stated, because of the discontinuous nature of the complex argument function. However, if we insert a continuous mollifier to remove this discontinuity, we can obtain equidistribution. More precisely, we have the following result.

Theorem 4.2 (equidistribution of argument away from zero). *Let $k \ge 1$, let $h_1, \dots, h_k$ be integers and $g_1, \dots, g_k : \mathbb{N} \to \mathbb{D}$ be 1-bounded multiplicative functions. Suppose that the product $g_1 \cdots g_k$ weakly pretends to be a twisted Dirichlet character $n \mapsto \chi(n)n^{\mathrm{it}}$, where $t \neq 0$. Let us denote*

$$S(X) := \mathbb{E}_{n\le X} g_1(n+h_1)\cdots g_k(n+h_k).$$

Let $\psi : \mathbb{C} \to \mathbb{C}$ be a continuous function that vanishes in a neighbourhood of the origin, and let

$$\bar{\psi}(z) := \frac{1}{2\pi}\int_0^{2\pi} \psi(e^{i\theta}z)\,d\theta$$

be ψ averaged over rotations around the origin. Then we have

$$\mathbb{E}^{\log}_{X\in\mathbb{N}}\psi(S(X)) - \bar{\psi}(S(X)) = 0.$$

Proof. Since S is bounded, we may assume that ψ is compactly supported. By replacing ψ by $\psi - \bar{\psi}$ we may assume that $\bar{\psi} = 0$. Approximating ψ uniformly by partial Fourier series (e.g., using Fejér summation) in the angular variable, and using linearity, we may assume that ψ takes the form $\psi(re^{i\theta}) = \Psi(r)e^{ik\theta}$ for some nonzero integer k and some continuous compactly supported function Ψ that vanishes in a neighbourhood of the origin (cf. the standard proof of the Weyl equidistribution criterion [1916]). In particular we have the isotopy formula

$$\psi(\omega z) = \omega^k \psi(z) \tag{55}$$

for all $z \in \mathbb{C}$ and $\omega \in S^1$.

Let $q > 1$ be an integer to be chosen later. From Theorem 1.19(i), outside of an exceptional set $\mathcal{X}_0$ of logarithmic density zero, we have

$$\lim_{X\to\infty; X\notin\mathcal{X}_0} S(X) - q^{\mathrm{it}} S(X/q) = 0.$$

From (55) and the uniform continuity of ψ, this implies that

$$\lim_{X\to\infty; X\notin\mathcal{X}_0} \psi(S(X)) - q^{ikt} \psi(S(X/q)) = 0.$$

Taking logarithmic averages, we conclude that

$$\mathbb{E}^{\log}_{X\in\mathbb{N}} \psi(S(X)) - q^{ikt}\psi(S(X/q)) = 0.$$

On the other hand, in analogy to (16), we have the log-Lipschitz bound

$$|S(x) - S(y)| \ll |\log x - \log y|. \tag{56}$$

We can use this and the uniform continuity of ψ to estimate, for X_0 large enough,

$$\begin{aligned}
\mathbb{E}^{\log}_{X\le X_0} \psi(S(X/q)) &= \mathbb{E}^{\log}_{X\le X_0/q} \mathbb{E}_{0\le b<q} \psi(S(X+b/q)) + o(1) \\
&= \mathbb{E}^{\log}_{X\le X_0/q} \psi(S(X)) + o(1) \\
&= \mathbb{E}^{\log}_{X\le X_0} \psi(S(X)) + o(1).
\end{aligned}$$

Hence

$$\mathbb{E}^{\log}_{X\in\mathbb{N}} \psi(S(X)) - \psi(S(X/q)) = 0.$$

By the triangle inequality, we conclude that

$$(1 - q^{ikt})\mathbb{E}^{\log}_{X\in\mathbb{N}} \psi(S(X)) = 0.$$

Since $t \neq 0$, we can select q so that $q^{ikt} \neq 1$ for all $k \in \mathbb{N}$. The claim follows. $\square$

Suppose that $\Psi : [0, +\infty) \to [0, +\infty)$ is a nonnegative continuous function vanishing near the origin, and let $I \subset \mathbb{R}/2\pi\mathbb{Z}$ be an arc in the unit circle $\mathbb{R}/2\pi\mathbb{Z}$. Applying Theorem 4.2 to upper and lower approximants to the discontinuous function $z \mapsto \Psi(|z|)1_I(\arg(z))$, and taking limits, we conclude that

$$\mathbb{E}^{\log}_{X\in\mathbb{N}} \left(1_I(\arg(S(X))) - \frac{|I|}{2\pi}\right)\Psi(|S(X)|) = 0$$

where $|I|$ denotes the length of I. Informally, this asserts that the argument $\arg(S(X))$ is uniformly distributed in the unit circle, so long as one inserts a continuous weight of the form $\Psi(|S(X)|)$. It would be more aesthetically pleasing if we could replace this weight with a discontinuous cutoff such as $1_{|S(X)|\geq\varepsilon}$, but we were unable to exclude the possibility that $|S(X)|$ lingers very close to ε for very many scales X, with the event that $|S(X)|\geq\varepsilon$ being coupled in some arbitrary fashion to the argument of $S(X)$, leading to essentially no control on the argument of $S(X)$ restricted to the event $|S(X)|\geq\varepsilon$. On the other hand, if one was able to show that $S(X)$ did not concentrate at the origin in the sense that

$$\limsup_{X_0\to\infty}\mathbb{E}^{\log}_{X\leq X_0}1_{|S(X)|\leq\varepsilon}\to 0$$

as $\varepsilon\to 0$, then the above arguments do show that

$$\mathbb{E}^{\log}_{X\in\mathbb{N}}1_I(\arg(S(X)))=\frac{|I|}{2\pi}$$

for all intervals I, so that $\arg(S(X))$ is indeed asymptotically equidistributed on the unit circle. Alternatively, by selecting the cutoff ε using the pigeonhole principle to ensure that $|S(X)|$ does not linger too often in a neighbourhood of ε, one can prove statements such as the following: If $\delta>0$, then for all sufficiently large X_0 outside of a set of logarithmic density zero, one can find $0<\varepsilon\leq\delta$ with the approximate equidistribution property

$$\mathbb{E}^{\log}_{X\leq X_0}\left(1_I(\arg(S(X)))-\frac{|I|}{2\pi}\right)1_{|S(X)|\geq\varepsilon}\leq\delta$$

for all intervals $I\subset\mathbb{R}/2\pi\mathbb{Z}$. We leave the proof of this assertion to the interested reader.

Now we investigate the consequences of the nonarchimedean isotopy formula (Theorem 1.19(ii)). Many of these consequences tell us that the correlation (9) tends to 0 along almost all scales also in some cases that are not covered by Corollary 1.8(i).

Definition 4.3. We say that a tuple $(g_1,\dots,g_k)$ of functions is *reflection symmetric* if $g_i=g_{k+1-i}$ for all $1\leq i\leq(k+1)/2$. Similarly, we say that a tuple $(h_1,\dots,h_k)$ of integers is *progression-like* if $h_1+h_k=h_i+h_{k+1-i}$ for all $1\leq i\leq(k+1)/2$. In particular, all arithmetic progressions are progression-like.

Theorem 4.4. *Let $k\geq 1$ and let $h_1,\dots,h_k$ be integers. Suppose that χ is an odd Dirichlet character (i.e., $\chi(-1)=-1$) with $\chi(n+h_1+h_k)=\chi(n)$ for all n. Let $g_1,\dots,g_k:\mathbb{N}\to\mathbb{D}$ be multiplicative functions such that the product $g_1\cdots g_k$ weakly pretends to be a Dirichlet character ψ with ψ even. Suppose additionally that the tuple $(g_1,\dots,g_k)$ is reflection symmetric and that the tuple $(h_1,\dots,h_k)$ is progression-like. Then there exists an exceptional set $\mathcal{X}_0$ of logarithmic density 0, such that*

$$\lim_{X\to\infty;\,X\notin\mathcal{X}_0}\mathbb{E}_{n\leq X}\chi(n)g_1(n+h_1)g_2(n+h_2)\cdots g_k(n+h_k)=0.$$

Proof. Note that the function $g_1\cdots g_k\chi$ weakly pretends to be $\psi\chi$, which is an odd character. Hence by Theorem 1.19(ii) there exists some set $\mathcal{X}_0$ of logarithmic density 0, such that for $X\notin\mathcal{X}_0$ we have

$$\mathbb{E}_{n\leq X}\chi(n)g(n)g_1(n+h_1)\cdots g_k(n+h_k)=-\mathbb{E}_{n\leq X}\chi(n)g_1(n-h_1)g(n-h_2)\cdots g(n-h_k)+o(1).$$

By translation invariance, the periodicity assumption $\chi(n+h_1+h_k)=\chi(n)$, and the progression-likeness of $(h_1,\dots,h_k)$, the latter expression equals

$$-\mathbb{E}_{n\le X}\chi(n+h_1+h_k)g_1(n+h_k)g_2(n+h_1+h_k-h_2)\cdots g_k(n+h_1)+o(1)$$
$$=-\mathbb{E}_{n\le X}\chi(n)g_1(n+h_k)g_2(n+h_{k-1})\cdots g_k(n+h_1)+o(1).$$

Since the tuple $(g_1,\dots,g_k)$ is reflection symmetric, this equals the original correlation with a minus sign, proving the statement. □

Corollary 1.21 is an immediate consequence of Theorem 4.4.

Proof of Corollary 1.21. Taking $g_1=\cdots=g_k=\lambda$ and $(h_1,\dots,h_k)=(0,a,\dots,(k-1)a)$ in Theorem 4.4, we readily obtain the claim. □

In other words, the shifted products of the Liouville function can be shown to be orthogonal to some suitable Dirichlet characters also when there is an even number of shifts. As already mentioned, also the weaker, logarithmic version of Corollary 1.21 is new.

The next theorem is in the same spirit as Theorem 4.4, but with somewhat different conditions.

Theorem 4.5. *Let $k\ge 1$ be odd, and let $g_1,\dots,g_k:\mathbb{N}\to\mathbb{D}$ be multiplicative functions such that the product $g_1\cdots g_k$ weakly pretends to be a Dirichlet character χ with χ odd. Suppose also that the tuple $(g_1,\dots,g_k)$ is reflection symmetric and that $(h_1,\dots,h_k)$ is a progression-like tuple of integers. Then there exists an exceptional set $\mathcal{X}_0$ of logarithmic density 0, such that*

$$\lim_{X\to\infty;\,X\notin\mathcal{X}_0}\mathbb{E}_{n\le X}g_1(n+h_1)g_2(n+h_2)\cdots g_k(n+h_k)=0.$$

Proof. As with Theorem 4.4, this follows directly from the isotopy formula (Theorem 1.19) and translation invariance. □

This theorem can for example be applied to the variants

$$\lambda_q(n):=e\left(\frac{2\pi i\Omega(n)}{q}\right)$$

of the Liouville function that take values in the q-th roots of unity. Here $\Omega(n)$ is the number of prime factors of n with multiplicities. We obtain the following.

Corollary 4.6. *Let $k\ge 1$ be odd, $q\in\mathbb{N}$, and let χ be an odd Dirichlet character. Then there exists an exceptional set $\mathcal{X}_0$ of logarithmic density 0, such that*

$$\lim_{X\to\infty;\,X\notin\mathcal{X}_0}\mathbb{E}_{n\le X}\lambda_q(n)\chi(n)\lambda_q(n+a)\chi(n+a)\cdots\lambda_q(n+(k-1)a)\chi(n+(k-1)a)=0.$$

Proof. We apply Theorem 4.5 with $g_j(n)=g(n):=\chi(n)\lambda_q(n)$ and $(h_1,\dots,h_k)=(0,\dots,(k-1)a)$. Then if $q\nmid k$, the function $g^k\chi^k$ does not weakly pretend to be any twisted Dirichlet character, since g^k does not do so. In this case, we may appeal to Corollary 1.8(i) to obtain the claim. Suppose then that $q\mid k$. Then g^k weakly pretends to be χ^k, which is an odd character, so Theorem 4.5 is applicable. □

Example 4.7. Let χ_3 be the odd Dirichlet character of modulus 3 and χ_8 any odd Dirichlet character of modulus 8. Then from Corollary 4.6 and partial summation, for any sequences $1 \le \omega_m \le x_m$ of reals tending to infinity we have

$$\lim_{m\to\infty} \mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} \chi_3(n)\lambda_3(n)\lambda_3(n+3)\lambda_3(n+6) = 0$$

and

$$\lim_{m\to\infty} \mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} \lambda_3(n)\lambda_3(n+2)\lambda_3(n+4)\chi_8(n+6) = 0.$$

This is seen by applying the corollary to the functions $g_j(n) = \lambda_3(n)\chi_3(n)$ with $a = 3$ and $g_j(n) = \lambda_3(n)\chi_8(n)$ with $a = 2$ and using $n(n+2)(n+4) \equiv n+6 \pmod 8$ for n odd.

We then turn to bounding more general correlations of multiplicative functions where the shifts involved no longer form a progression-like tuple. In the case of triple correlations, we obtain savings that are explicit but nevertheless far from the desired $o(1)$ bound.

Theorem 4.8 (savings in logarithmic three-point Elliott conjecture). *Let $g : \mathbb{N} \to \mathbb{D}$ be a multiplicative function, and let h_1, h_2, h_3 be distinct integers. Suppose that g is nonpretentious in the sense that*

$$\liminf_{X\to\infty} \inf_{|t|\le X} \mathbb{D}(g, n \mapsto \chi(n)n^{it}, x) = \infty$$

for every Dirichlet character χ. Then for any sequences $1 \le \omega_m \le x_m$ tending to infinity we have

$$\limsup_{m\to\infty} |\mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} g(n+h_1)g(n+h_2)g(n+h_3)| \le \frac{1}{\sqrt{2}}. \tag{57}$$

Remark 4.9. This looks superficially similar to [Klurman and Mangerel 2018, Lemma 5.3] (and also to [Tao and Teräväinen 2019, Proposition 7.1], which achieves the better upper bound of $\frac{1}{2}$ for real-valued multiplicative functions). However, importantly, the shifts h_i are allowed to be arbitrary here, while in the aforementioned results they had to form an arithmetic progression for the method to work.

Proof. If h_1, h_2, h_3 is an arithmetic progression, we may apply [Klurman and Mangerel 2018, Lemma 5.3], so we may henceforth suppose that h_1, h_2, h_3 is not an arithmetic progression.

If the function g^3 does not weakly pretend to be any Dirichlet character, we get the bound 0 for the lim sup by [Tao and Teräväinen 2019, Theorem 1.2(ii)]. Suppose then that g^3 weakly pretends to be some character χ. By the isotopy formula (Theorem 1.19), partial summation and translation invariance, we have

$$\begin{aligned}\mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} &g(n+h_1)g(n+h_2)g(n+h_3)\\ &= \chi(-1)\mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} g(n+h_1)g(n+h_1+h_3-h_2)g(n+h_3) + o_{m\to\infty}(1).\end{aligned} \tag{58}$$

In particular, the first part of (58) is the average of both parts of the equation. Hence, the average on the left-hand side of (57) is up to $o_{m\to\infty}(1)$ bounded by

$$\begin{aligned}\tfrac{1}{2}|\mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} &g(n+h_1)g(n+h_3)(g(n+h_2)+\chi(-1)g(n+h_1+h_3-h_2))|\\ &\le \tfrac{1}{2}\mathbb{E}^{\log}_{x_m/\omega_m \le n \le x_m} |g(n+h_2)+\chi(-1)g(n+h_1+h_3-h_2))|.\end{aligned}$$

By the Cauchy–Schwarz inequality, this is bounded by

$$\tfrac{1}{2}(\mathbb{E}^{\log}_{x_m/\omega_m\le n\le x_m}|g(n+h_2)+\chi(-1)g(n+h_1+h_3-h_2))|^2)^{1/2}$$
$$\le \tfrac{1}{2}(\mathbb{E}^{\log}_{x_m/\omega_m\le n\le x_m}(2+2\chi(-1)\operatorname{Re} i(g(n+h_2)\bar{g}(n+h_1+h_3-h_2))))^{1/2}.$$

Since $h_2 \neq h_1+h_3-h_2$ by assumption, we can apply [Tao 2016, Corollary 1.5] to see that the term involving real parts contributes $o_{m\to\infty}(1)$. Then we indeed get a bound of $1/\sqrt{2}+o_{m\to\infty}(1)$ for the correlation. □

Remark 4.10. For specific multiplicative functions one can do slightly better by not applying Cauchy–Schwarz. For example, in the case $g(n)=\lambda_3(n)$ one gets a bound of $\frac{2}{3}$ for the correlation by using the fact (following from [Tao 2016, Corollary 1.5]) that $(\lambda_3(n), \lambda_3(n+h))$ takes for fixed $h\neq 0$ each of the possible 9 values with logarithmic density $\frac{1}{9}$.

5. The case of few sign patterns

In this section we prove Theorem 1.17. Assume the hypotheses of that theorem. Let h be a natural number. By the Hahn–Banach theorem, it suffices to show that

$$\mathbb{E}^*_{n\in\mathbb{N}}\lambda(n)\lambda(n+h)=0$$

for every generalised limit $\lim^*_{X\to\infty}$. Accordingly, let us fix such a limit. As usual, we introduce the correlation sequences

$$f_d(a) := \lim^*_{X\to\infty} \mathbb{E}_{n\le X/d}\lambda(n)\lambda(n+ah) \tag{59}$$

for every real $d>0$. Our task is now to show that

$$f_1(1)=0.$$

Proposition 2.3 (noting that $G(p)=1$ in our case) establishes the approximate isotopy formula

$$\sup_{d>0}\mathbb{E}^{\log}_{p\le P}|f_{dp}(a)-f_d(ap)|\le\varepsilon$$

whenever $\varepsilon>0$ and P is sufficiently large depending on ε. But because of our hypothesis of few sign patterns, we can obtain a stronger result in which the logarithmic weighting on the averages is removed.

Proposition 5.1 (improved approximate isotopy formula). *Let $f_d(a)$ be as in (59), let $\varepsilon>0$, and let a be a natural number. Assume the hypotheses of Theorem 1.17. Then there exist arbitrarily large m such that*

$$\sup_{d>0}\mathbb{E}_{2^m\le p<2^{m+1}}|f_{dp}(a)-f_d(ap)|\le\varepsilon.$$

This formula also applies for negative a, but in this argument we only require the case of positive a (in fact, for the binary correlations considered here, we only need the case $a=1$).

Proof. This will be a modification of the arguments in [Tao and Teräväinen 2019, Section 3], and we freely use the notation from that paper.

Let $d > 0$ be real, let a be an integer, and let m be a large integer to be chosen later. We allow implied constants to depend on h, a, but they will remain uniform in d, m, ε. From (13) we have the formula

$$f_{dp}(a) - f_d(ap) = \mathbb{E}^{(d)} \boldsymbol{g}^{(d)}(0) \boldsymbol{g}^{(d)}(aph)(p 1_{p \mid \boldsymbol{n}^{(d)}} - 1) + O(\varepsilon)$$

for all $2^m \le p < 2^{m+1}$, if m is sufficiently large depending on ε, and where $\boldsymbol{g}^{(d)} = \boldsymbol{g}_1^{(d)} = \boldsymbol{g}_2^{(d)}$ and $\boldsymbol{n}^{(d)}$ are the random variables provided by Proposition 2.1 (with $g_1 = g_2 = \lambda$). We can thus write the expression

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a) - f_d(ap)|$$

as

$$\mathbb{E}^{(d)} \mathbb{E}_{2^m \le p < 2^{m+1}} c_p \boldsymbol{g}^{(d)}(0) \boldsymbol{g}^{(d)}(aph)(p 1_{p | \boldsymbol{n}^{(d)}} - 1) + O(\varepsilon)$$

for some sequence of complex numbers c_p with $|c_p| \le 1$. By stationarity we can also write this expression as

$$\mathbb{E}^{(d)} \mathbb{E}_{1 \le l \le 2^m} \mathbb{E}_{2^m \le p < 2^{m+1}} c_p \boldsymbol{g}^{(d)}(l) \boldsymbol{g}^{(d)}(l + aph)(p 1_{\boldsymbol{n}^{(d)} = -l \ (p)} - 1) + O(\varepsilon)$$

and thus

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a) - f_d(ap)| = \mathbb{E}^{(d)} F(\boldsymbol{X}^{(d)}, \boldsymbol{Y}^{(d)})$$

where $\boldsymbol{X}^{(d)} = \boldsymbol{X}_m^{(d)} \in \{-1, +1\}^{(2ah+1)2^m}$, $\boldsymbol{Y}^{(d)} = \boldsymbol{Y}_m^{(d)} \in \prod_{2^m \le p < 2^{m+1}} \mathbb{Z}/p\mathbb{Z}$ are the random variables

$$\boldsymbol{X}^{(d)} := (\boldsymbol{g}^{(d)}(l))_{1 \le l \le (2ah+1)2^m} \quad \text{and} \quad \boldsymbol{Y}^{(d)} := (\boldsymbol{n}^{(d)} \ (p))_{2^m \le p < 2^{m+1}}$$

and $F : \{-1, +1\}^{(2ah+1)2^m} \times \prod_{2^m \le p < 2^{m+1}} \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is the function

$$F((g_l)_{1 \le l \le (2ah+1)2^m}, (n_p)_{2^m \le p < 2^{m+1}}) := \mathbb{E}_{1 \le l \le 2^m} \mathbb{E}_{2^m \le p < 2^{m+1}} c_p g_l g_{l+aph}(p 1_{n_p = -l \ (p)} - 1).$$

Repeating the arguments in [Tao and Teräväinen 2019, Section 3] verbatim (but without the additional conditioning on the $\boldsymbol{Y}_{<m}$ random variable), we conclude that

$$\mathbb{E}^{(d)} |F(\boldsymbol{X}^{(d)}, \boldsymbol{Y}^{(d)})| \le \varepsilon$$

unless we have the mutual information bound

$$\boldsymbol{I}(\boldsymbol{X}^{(d)} : \boldsymbol{Y}^{(d)}) > \varepsilon^5 \frac{2^m}{m}.$$

At this point we deviate from the arguments in [Tao and Teräväinen 2019, Section 3] by using the trivial bound

$$\boldsymbol{I}(\boldsymbol{X}^{(d)} : \boldsymbol{Y}^{(d)}) \le \boldsymbol{H}(\boldsymbol{X}^{(d)})$$

to conclude that we will have the desired bound

$$\mathbb{E}_{2^m \le p < 2^{m+1}} |f_{dp}(a) - f_d(ap)| \le \varepsilon$$

whenever $\boldsymbol{X}^{(d)}$ obeys the entropy bound

$$\boldsymbol{H}(\boldsymbol{X}^{(d)}) \leq \varepsilon^5 \frac{2^m}{m}.$$

By Jensen's inequality, this bound will hold if $\boldsymbol{X}^d$ attains at most $\exp(\varepsilon^5 2^m/m)$ values with positive probability. Using the correspondence principle (Proposition 2.1), this claim in turn is equivalent to the number of possible sign patterns $(\lambda(n+l))_{1\leq l\leq (2ah+1)2^m}$ not exceeding $\exp(\varepsilon^5 2^m/m)$; note that this assertion does not depend on d, so we in fact obtain the uniform bound

$$\sup_{d>0} \mathbb{E}_{2^m \leq p < 2^{m+1}} |f_{dp}(a) - f_d(ap)| \leq \varepsilon$$

in this case. But by the hypothesis of Theorem 1.17, this assertion holds for arbitrarily large values of m. □

Now we establish Theorem 1.17. By the above proposition, for any $\varepsilon > 0$, there exist arbitrarily large m such that

$$f_1(1) = \mathbb{E}_{2^m \leq p < 2^{m+1}} f_P(p) + O(\varepsilon),$$

where $P := 2^m$. By (59), it suffices to show that

$$\limsup_{X\to\infty} |\mathbb{E}_{P\leq p<2P} \mathbb{E}_{n\leq X/P} \lambda(n)\lambda(n+ph)| \ll \varepsilon$$

for sufficiently large P. But this follows from the results in [Tao 2016, Section 3], specifically Lemmas 3.6 and 3.7 and equation (2.9) of that paper[13] (see also Remark 3.8 for a simplification in the case of the Liouville function). We remark that the equation [loc. cit., (2.8)] relies crucially on the Matomäki–Radziwiłł theorem [Matomäki and Radziwiłł 2016] (as applied in [Matomäki et al. 2015]).

Remark 5.2. A similar argument also gives the odd order cases of the Chowla conjecture if one strengthens the hypothesis of Theorem 1.17 to hold for *all* sufficiently large K, rather than for arbitrarily large K, by using the arguments in [Tao and Teräväinen 2018, Section 3] (but with the exceptional sets $\mathcal{M}_1$ in those arguments now being empty, and using unweighted averaging in n rather than logarithmic averaging). We leave the details to the interested reader.

Acknowledgements

We thank the anonymous referees for careful reading of the paper and for useful comments.

T. Tao was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

J. Teräväinen was supported by UTUGS Graduate School and project number 293876 of the Academy of Finland. He thanks UCLA for excellent working conditions during a visit there in April 2018, during which a large proportion of this work was completed.

[13] When applying these results, note that (in the notation of [Tao 2016, Lemma 3.6]) the length of the sum over $j \in [1, H]$ and the range $\mathcal{P}_H$ of primes p do not need to be controlled by the same parameter H.

References

[Babai et al. 2003] L. Babai, K. Friedl, and A. Luckács, "Near representations of finite groups", preprint, 2003, Available at https://tinyurl.com/babainear.

[Balog et al. 2013] A. Balog, A. Granville, and K. Soundararajan, "Multiplicative functions in arithmetic progressions", *Ann. Math. Qué.* **37**:1 (2013), 3–30. MR Zbl

[Chowla 1965] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Math. Appl. **4**, Gordon and Breach, New York, 1965. MR Zbl

[Elliott 1992] P. D. T. A. Elliott, "On the correlation of multiplicative functions", *Notas Soc. Mat. Chile* **11**:1 (1992), 1–11. MR Zbl

[Elliott 1994] P. D. T. A. Elliott, *On the correlation of multiplicative and the sum of additive arithmetic functions*, Mem. Amer. Math. Soc. **538**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Erdős 1979] P. Erdős, "Some unconventional problems in number theory", pp. 73–82 in *Journées arithmétiques de Luminy* (Luminy, France, 1978), Astérisque **61**, Soc. Math. France, Paris, 1979. MR Zbl

[Frantzikinakis 2017] N. Frantzikinakis, "Ergodicity of the Liouville system implies the Chowla conjecture", *Discrete Anal.* **2017** (2017), art. id. 19. MR Zbl

[Frantzikinakis and Host 2018] N. Frantzikinakis and B. Host, "The logarithmic Sarnak conjecture for ergodic weights", *Ann. of Math.* (2) **187**:3 (2018), 869–931. MR Zbl

[Frantzikinakis and Host 2019] N. Frantzikinakis and B. Host, "Furstenberg systems of bounded multiplicative functions and applications", *Int. Math. Res. Not.* (online publication February 2019).

[Frigerio 2017] R. Frigerio, *Bounded cohomology of discrete groups*, Math. Surv. Monogr. **227**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Gomilko et al. 2018] A. Gomilko, D. Kwietniak, and M. Lemańczyk, "Sarnak's conjecture implies the Chowla conjecture along a subsequence", pp. 237–247 in *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, edited by S. Ferenczi et al., Lecture Notes in Math. **2213**, Springer, 2018. MR Zbl

[Granville and Soundararajan 2008] A. Granville and K. Soundararajan, "Pretentious multiplicative functions and an inequality for the zeta-function", pp. 191–197 in *Anatomy of integers* (Montreal, 2006), edited by J.-M. De Koninck et al., CRM Proc. Lecture Notes **46**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Halász 1971] G. Halász, "On the distribution of additive and the mean values of multiplicative arithmetic functions", *Studia Sci. Math. Hungar.* **6** (1971), 211–233. MR Zbl

[Heath-Brown 1983] D. R. Heath-Brown, "Prime twins and Siegel zeros", *Proc. Lond. Math. Soc.* (3) **47**:2 (1983), 193–224. MR Zbl

[Kazhdan 1982] D. Kazhdan, "On ε-representations", *Israel J. Math.* **43**:4 (1982), 315–323. MR Zbl

[Klurman and Mangerel 2018] O. Klurman and A. P. Mangerel, "Rigidity theorems for multiplicative functions", *Math. Ann.* **372**:1-2 (2018), 651–697. MR Zbl

[Kotschick 2004] D. Kotschick, "What is... a quasi-morphism?", *Notices Amer. Math. Soc.* **51**:2 (2004), 208–209. MR Zbl

[Le 2018] A. Le, "Nilsequences and multiple correlations along subsequences", *Ergodic Theory Dynam. Systems* (online publication October 2018).

[Leibman 2015] A. Leibman, "Nilsequences, null-sequences, and multiple correlation sequences", *Ergodic Theory Dynam. Systems* **35**:1 (2015), 176–191. MR Zbl

[Matomäki and Radziwiłł 2016] K. Matomäki and M. Radziwiłł, "Multiplicative functions in short intervals", *Ann. of Math.* (2) **183**:3 (2016), 1015–1056. MR Zbl

[Matomäki et al. 2015] K. Matomäki, M. Radziwiłł, and T. Tao, "An averaged form of Chowla's conjecture", *Algebra Number Theory* **9**:9 (2015), 2167–2196. MR Zbl

[Matomäki et al. 2016] K. Matomäki, M. Radziwiłł, and T. Tao, "Sign patterns of the Liouville and Möbius functions", *Forum Math. Sigma* **4** (2016), art. id. e14. MR Zbl

[McNamara 2019] R. McNamara, "Sarnak's conjecture for sequences of almost quadratic word growth", preprint, 2019. arXiv

[Riesz 1932] F. Riesz, "Sur un théorème de maximum de mm. Hardy et Littlewood", *J. Lond. Math. Soc.* **7**:1 (1932), 10–13. MR Zbl

[Sarnak 2010] P. Sarnak, "Three lectures on the Möbius function randomness and dynamics", preprint, 2010, Available at https://tinyurl.com/sarthree.

[Sós 2002] V. T. Sós, "Turbulent years: Erdős in his correspondence with Turán from 1934 to 1940", pp. 85–146 in *Paul Erdős and his mathematics, I* (Budapest, 1999), edited by G. Halász et al., Bolyai Soc. Math. Stud. **11**, János Bolyai Math. Soc., Budapest, 2002. MR Zbl

[Tao 2016] T. Tao, "The logarithmically averaged Chowla and Elliott conjectures for two-point correlations", *Forum Math. Pi* **4** (2016), art. id. e8. MR Zbl

[Tao 2017a] T. Tao, "Equivalence of the logarithmically averaged Chowla and Sarnak conjectures", pp. 391–421 in *Number theory: Diophantine problems, uniform distribution and applications*, edited by C. Elsholtz and P. Grabner, Springer, 2017. MR Zbl

[Tao 2017b] T. Tao, "The logarithmically averaged and non-logarithmically averaged Chowla conjectures", blog post, 2017, Available at terrytao.wordpress.com/2017/10/20.

[Tao and Teräväinen 2018] T. Tao and J. Teräväinen, "Odd order cases of the logarithmically averaged Chowla conjecture", *J. Théor. Nombres Bordeaux* **30**:3 (2018), 997–1015. MR Zbl

[Tao and Teräväinen 2019] T. Tao and J. Teräväinen, "The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures", *Duke Math. J.* **168**:11 (2019), 1977–2027. MR

[Teräväinen 2018] J. Teräväinen, "On binary correlations of multiplicative functions", *Forum Math. Sigma* **6** (2018), art. id. e10. MR Zbl

[Weyl 1916] H. Weyl, "Über die Gleichverteilung von Zahlen mod. Eins", *Math. Ann.* **77**:3 (1916), 313–352. MR Zbl

[Wirsing 1967] E. Wirsing, "Das asymptotische Verhalten von Summen über multiplikative Funktionen, II", *Acta Math. Acad. Sci. Hungar.* **18** (1967), 411–467. MR Zbl

Communicated by Andrew Granville
Received 2018-09-08 Revised 2019-06-24 Accepted 2019-07-24

tao@math.ucla.edu *Department of Mathematics, University of California Los Angeles, Los Angeles, CA, United States*

joni.teravainen@maths.ox.ac.uk *Mathematical Institute, University of Oxford, United Kingdom*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2151

VI-modules in nondescribing characteristic, part I

Rohit Nagpal

Let VI be the category of finite dimensional $\mathbb{F}_q$-vector spaces whose morphisms are injective linear maps and let $\boldsymbol{k}$ be a noetherian ring. We study the category of functors from VI to $\boldsymbol{k}$-modules in the case when q is invertible in $\boldsymbol{k}$. Our results include a structure theorem, finiteness of regularity, and a description of the Hilbert series. These results are crucial in the classification of smooth irreducible $\mathrm{GL}_\infty(\mathbb{F}_q)$-representations in nondescribing characteristic which is contained in Part II of this paper (*VI-modules in nondescribing characteristic*, part II, arxiv:1810.04592).

1. Introduction

Fix a commutative noetherian ring $\boldsymbol{k}$. Set $\mathbb{F} = \mathbb{F}_q$, and let GL_n be the n-th general linear group over $\mathbb{F}$. Roughly speaking, the aim of this paper is to study the behavior of sequences, whose n-th member is a $\boldsymbol{k}[\mathrm{GL}_n]$-module, as n approaches infinity (the "generic case"). As n varies, every prime appears as a divisor of the size of GL_n. But surprisingly, it is possible to avoid most of the complications of modular representation theory in the generic case after inverting just one prime, namely the characteristic of $\mathbb{F}$. We assume throughout that q is invertible in $\boldsymbol{k}$, and we call this the "nondescribing characteristic" assumption.

We obtain these sequences in the form of VI-modules. A VI-module M is a functor

$$M\colon\ \mathrm{VI} \to \mathrm{Mod}_{\boldsymbol{k}},$$

where VI is the category of finite dimensional $\mathbb{F}$-vector spaces with injective linear maps. Clearly, $\mathrm{GL}_n = \mathrm{Aut}_{\mathrm{VI}}(\mathbb{F}^n)$ acts on $M(\mathbb{F}^n)$. Thus M can be thought of as a sequence whose n-th member is a $\boldsymbol{k}[\mathrm{GL}_n]$-module. This sequence could be arbitrary if we do not impose any finiteness conditions on M. But there is a natural notion of "finite generation" in the category of VI-modules. This paper analyzes finitely generated VI-modules. Here is a sample theorem that we prove (it extends [Gan and Watterlond

MSC2010: primary 20C33; secondary 13D45, 20J05.
Keywords: VI-modules, FI-modules, finite general linear groups, representation stability.

2018, Theorem 1.7] away from characteristic zero, and also improves some cases of [Sam and Snowden 2017a, Corollary 8.3.4]):

Theorem 1.1 (q-polynomiality of dimension). *Assume that $\boldsymbol{k}$ is a field in which q is invertible. Let M be a finitely generated* VI*-module. Then there exists a polynomial P such that $\dim_{\boldsymbol{k}} M(\mathbb{F}^n) = P(q^n)$ for large enough n.*

The result above is a consequence of our main structural result that we prove about finitely generated VI-modules. Given a VI-module M and a vector space X, we can define a new VI-module $\Sigma^X M$ by

$$\Sigma^X M(Z) = M(X+Z).$$

We call this new VI-module the *shift* of M by X. Our main result roughly says that the shift of a finitely generated module by a vector space of large enough dimension has a very simple description. To make it precise, note that there is a natural restriction functor

$$\mathrm{Mod}_{\mathrm{VI}} \to \prod_{n\geq 0} \mathrm{Mod}_{\boldsymbol{k}[\mathrm{GL}_n]}.$$

This functor admits a left adjoint $\mathcal{I}$. We call a VI-module *induced* if it is of the form $\mathcal{I}(W)$ for some W. A VI-module that admits a finite filtration whose graded pieces are induced is called *semiinduced*. We now state our main theorem.

Theorem 1.2 (the shift theorem). *Assume that q is invertible in $\boldsymbol{k}$. Let M be a finitely generated* VI*-module. Then $\Sigma^X M$ is semiinduced if the dimension of X is large enough.*

Idea behind the shift theorem. The shift theorem is proven by induction on the degree of generation. To make the induction hypothesis work, we construct a "categorical derivation" in the monoidal category of Joyal and Street [1995]. To make it precise, let VB be the category of finite dimensional $\mathbb{F}$-vector spaces with bijective linear maps. Joyal and Street considered a monoidal structure[1] $\otimes_{\mathrm{VB}}$ on $\mathrm{Mod}_{\mathrm{VB}}$ given by

$$(M \otimes_{\mathrm{VB}} N)(Y) = \bigoplus_{X\leq Y} M(Y/X) \otimes_{\boldsymbol{k}} N(X).$$

We construct a categorical derivation $\overline{\Sigma}$ on $(\mathrm{Mod}_{\mathrm{VB}}, \otimes_{\mathrm{VB}})$. In other words, $\overline{\Sigma}$ satisfies

$$\overline{\Sigma}(M\otimes N) = (\overline{\Sigma}M \otimes N)\bigoplus(M\otimes \overline{\Sigma}N).$$

As pointed out to us by Steven Sam, there is an algebra object $\boldsymbol{A}$ in $(\mathrm{Mod}_{\mathrm{VB}}, \otimes_{\mathrm{VB}})$ such that the category of VI-module is equivalent to the category of $\boldsymbol{A}$-modules. Under this equivalence, induced modules are $\boldsymbol{A}$-modules of the form $\boldsymbol{A} \otimes_{\mathrm{VB}} W$. Our categorical derivation shows that if we apply the cokernel of $\mathrm{id} \to \overline{\Sigma}$ to an induced module then we obtain another induced module of strictly smaller degree of generation. This is what makes our inductive proof work. But there is a caveat. Everything said and done

[1] It is shown in [Joyal and Street 1995] that this category is actually a braided monoidal category if $\boldsymbol{k}$ is a field of characteristic zero. But we don't need the braiding, and so we don't need the characteristic zero assumption

in this paragraph so far is true without any restrictions on the characteristic. On the other hand, the shift theorem is false if we drop the nondescribing characteristic assumption.

The category $\mathrm{Mod}_{\mathrm{VI}}$ naturally contains a localizing subcategory $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ whose members are called torsion VI-modules. Given a VI-module M, we denote the maximal torsion submodule of M by $\Gamma(M)$. The functor Γ is left exact, and its right derived functor is denoted $\mathrm{R}\Gamma$. A crucial technical ingredient in our proof of the shift theorem is the following criterion for semiinduced modules.

Theorem 1.3. *Assume that q is invertible in $\boldsymbol{k}$. Let M be a finitely generated* VI*-module. Then M is semiinduced if and only if* $\mathrm{R}\Gamma(M)=0$.

That a semiinduced M satisfies $\mathrm{R}\Gamma(M)=0$ is easy to prove and doesn't require any assumptions on the characteristic. But the converse requires the nondescribing characteristic assumption in two crucial and separate places: (1) $\bar{\Sigma}$ is exact and (2) $\bar{\Sigma}$ commutes with Γ. (1) is immediate from our construction of $\bar{\Sigma}$ but (2) requires an interesting combinatorial identity (which appears in the proof of Lemma 4.26).

The last ingredient of our proof is a recent theorem proved independently by Putman and Sam [2017] and Sam and Snowden [2017a] which resolved a long-standing conjecture of Lannes and Schwartz.

Theorem 1.4 [Putman and Sam 2017; Sam and Snowden 2017a]. *Suppose $\boldsymbol{k}$ is an arbitrary noetherian ring* (*the nondescribing characteristic assumption is not needed*). *Then the category of* VI*-modules is locally noetherian.*

We also need the following immediate corollary of this theorem, which provides us control over the torsion part of a module.

Corollary 1.5 [Putman and Sam 2017; Sam and Snowden 2017a]. *Suppose $\boldsymbol{k}$ is an arbitrary noetherian ring* (*the nondescribing characteristic assumption is not needed*). *Let M be a finitely generated* VI*-module. Then $\Gamma(M)(X)=0$ if the dimension of X is large enough.*

All these ingredients above allow us to show by induction on the degree of generation that $\bar{\Sigma}^n M$ is semiinduced if n is large enough. The shift theorem then follows from it.

Some consequences of the shift theorem. To start with, Theorem 1.1 is a consequence of the shift theorem simply because induced modules can be easily seen to satisfy q-polynomiality of dimension. If we drop the nondescribing characteristic assumption, and assume that $\boldsymbol{k}=\mathbb{F}$, then $M(X)=X$ defines a finitely generated VI-module. This implies that q-polynomiality fails in equal characteristic, and so the shift theorem must also fail. Below we list some more consequences.

Theorem 1.6 (finiteness of local cohomology). *Assume that q is invertible in $\boldsymbol{k}$. Let M be a finitely generated* VI*-module. Then we have the following*:

(a) *For each i, the module $\mathrm{R}^i\Gamma(M)$ is finitely generated. In particular, $\mathrm{R}^i\Gamma(M)(X)=0$ if the dimension of X is large enough.*

(b) $\mathrm{R}^i\Gamma(M)=0$ *for i large enough.*

The theorem above extends Corollary 1.5 to the higher derived functors of Γ. We use this theorem, and an argument similar to the one for FI-modules as in [Nagpal et al. 2018], to bound the regularity. In particular, we provide a bound on the regularity in terms of the degrees of the local cohomology.

Theorem 1.7 (finiteness of regularity). *Assume that q is invertible in $\boldsymbol{k}$. Let M be a finitely generated* VI*-module. Then M has finite Castelnuovo–Mumford regularity.*

Gan and Watterlond [2018] have shown that, when $\boldsymbol{k}$ is an algebraically closed field of characteristic zero, then any finitely generated VI-module exhibits "representation stability", a phenomenon described by Church and Farb [2013]. Representation stability for VI-modules also follows from a recent result of Gadish [2017, Corollary 1.13]. We prove representation stability in a more systematic way. We believe that our method can be used to write down a virtual specht stability statement away from characteristic zero as done for FI-modules by Harman [2017]. In contrast to this, the methods in [Gan and Watterlond 2018] or [Gadish 2017] use characteristic zero assumption in an essential way. Below, we only state a part of the result to avoid giving a full definition of representation stability here (for full definition, see page 2182).

Theorem 1.8 [Gan and Watterlond 2018, Theorem 1.6]. *Assume that $\boldsymbol{k}$ is an algebraically closed field of characteristic zero. Let M be a finitely generated* VI*-module. Then the length of the $\boldsymbol{k}[\mathrm{GL}_n]$-module $M(\mathbb{F}^n)$ stabilizes in n.*

We also obtain the following new theorem in characteristic zero.

Theorem 1.9 (finiteness of injective dimension). *Assume that $\boldsymbol{k}$ is a field of characteristic zero. Then the following holds in* $\mathrm{Mod}_{\mathrm{VI}}$:

(a) *Every projective is injective.*

(b) *Every torsion-free injective is projective.*

(c) *Every finitely generated module has finite injective dimension.*

Along the way, we classify all indecomposable injectives in characteristic zero, and we also classify indecomposable torsion injectives when $\boldsymbol{k}$ is an arbitrary noetherian ring.

Relations to other works. Recently, Kuhn [2015] has analyzed a similar but simpler (of lower Krull dimension) category of VA-modules, where VA is the category of finite dimensional $\mathbb{F}$ vector spaces.

Theorem 1.10 [Kuhn 2015, Theorem 1.1]. *In the nondescribing characteristic*, $\mathrm{Mod}_{\mathrm{VA}}$ *is equivalent to the product category* $\prod_{n\geq 0}\mathrm{Mod}_{\boldsymbol{k}[\mathrm{GL}_n]}$. *In particular, if $\boldsymbol{k}$ is a field then* $\mathrm{Mod}_{\mathrm{VA}}$ *is of Krull dimension zero.*

A folklore result says that one recovers the representation theory of the symmetric groups from the representation theory of the finite general linear group over $\mathbb{F}_q$ by setting $q = 1$. We observe a similar phenomenon between FI-modules and VI-modules: all the results we have for VI-modules in the nondescribing characteristic are true for FI-modules in all characteristic (FI-modules encode sequences of representations of the symmetric groups; see [Church et al. 2015]). In other words, the proofs for the results on FI-modules are degenerate cases of the proofs for the corresponding results on VI-modules in the nondescribing characteristic. But we point out that (1) many of our ideas are copied from the corresponding ideas on FI-modules and (2) we know a lot more about FI-modules, for example, all the questions that we pose below have been solved for FI-modules. We have tried to summarize throughout the text where each crucial idea has been borrowed from, but here is a list of references that contain analogs of our results — [Church 2016; Church and Ellenberg 2017; Church et al. 2014; 2015; Djament 2016; Djament and Vespa 2019; Li 2016; Li and Ramos 2018; Nagpal 2015; Nagpal et al. 2018; Ramos 2018; Sam and Snowden 2016].

A higher dimension category of similar representation theoretic nature whose structure is well understood is the category of FI_d-modules; see [Sam and Snowden 2017b; 2019].

Further comments and questions. Theorem 1.8 implies that every finitely generated object in the category

$$\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{gen}} := \mathrm{Mod}_{\mathrm{VI}} \,/\, \mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$$

of generic VI-modules is of finite length, that is, the Krull dimension of $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{gen}}$ is zero. In a subsequent paper [Nagpal 2018], we shall prove that the same holds in the nondescribing characteristic (where $\boldsymbol{k}$ is still assumed to be a field) by providing a complete set of irreducibles of the generic category. Determining Krull dimension in equal characteristic ($\boldsymbol{k} = \mathbb{F}$) is related to an old open problem called the strong artinian conjecture [Powell 1998; 2000].

Sam and Snowden have proven that the categories of torsion and the generic FI-modules are equivalent in characteristic zero [Sam and Snowden 2016, Theorem 3.2.1], and such a phenomenon seem to appear in some other categories as well (for example, see [Sam and Snowden 2015] and [Nagpal et al. 2016] for the category of $\mathrm{Sym}(\mathrm{Sym}^2)$-modules). We have the following question along the same lines:

Question 1.11. Assume that $\boldsymbol{k}$ is of characteristic zero. Is there an equivalence of categories $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}} \cong \mathrm{Mod}_{\mathrm{VI}}^{\mathrm{gen}}$?

Remark 1.12. After the release of the first draft of this paper, Gan, Li and Xi have positively answered the question above; see [Gan et al. 2017]. We note that they used the shift theorem (Theorem 1.2) nontrivially; see [Gan et al. 2017, Lemma 4.1].

Our result provides bounds on the Castelnuovo–Mumford regularity in terms of the local cohomology. But we have not been able to bound local cohomology in terms of the degrees of generation and relation. An analogous question for FI-modules has already been answered [Church and Ellenberg 2017, Theorem A]; also see [Church 2016; Li 2016; Li and Ramos 2018, Theorem E] for more results on this. We also note that, in characteristic zero, Miller and Wilson have provided bounds on the higher syzygies for a similar category called VIC-modules; see [Miller and Wilson 2018, Theorem 2.26].

Question 1.13. Let M be a VI-module generated in degrees $\leq t_0$ and whose syzygies are generated in degrees $\leq t_1$. Is there a number n depending only on t_0 and t_1 such that $\Gamma(M)(X) = 0$ for every vector space X of dimension larger than n.

Remark 1.14. After the release of the first draft of this paper, Gan and Li have positively answered the question above; see [Gan and Li 2017]. We note that they used the shift theorem (Theorem 1.2) nontrivially. Along the way, they also made all the bounds in the current paper explicit in terms of t_0 and t_1; see [Gan and Li 2017, Theorem 1.1]. Bounds in the current paper are in terms of degrees of the local cohomology groups.

The question below is a VI-module analog of [Li and Ramos 2018, Conjecture 1.3] which has been resolved for FI-modules in [Nagpal et al. 2018].

Question 1.15. Is the Castelnuovo–Mumford regularity of a VI-module exactly $\max_i(\deg \mathrm{R}^i\Gamma(M)+i)$ where i varies over the finitely many values for which $\mathrm{R}^i\Gamma(M)$ is nonzero?

Outline of the paper. In Section 2, we provide an overview of VI-modules. In particular, we sketch an equivalence between $\mathrm{Mod}_{\mathrm{VI}}$ and the module category of an algebra object $\boldsymbol{A}$ in the monoidal category of Joyal and Street, and we recall some formalism of local cohomology and saturation from [Sam and Snowden 2019]. In Section 3, we prove some formal properties of induced and semiinduced modules that we need. These properties are formal in the sense that they have nothing much to do with VI-modules and are true (with appropriate definitions) in several other categories (for example, $\mathrm{Mod}_{\mathrm{FI}}$, $\mathrm{Mod}_{\mathrm{FI}_d}$ or $\mathrm{Mod}_{\mathrm{VIC}}$). We decided to include a short section and collect these formal results at one place. The meat of the paper is contained in Section 4 where we prove the shift theorem. The last section (Section 5) contains all the consequences of the shift theorem.

2. Overview of VI-modules

Notation. We work over a unital commutative ring $\boldsymbol{k}$. For a nonnegatively graded $\boldsymbol{k}$-module M, we define $\deg M$ to be the least integer $n \geq -1$ such that $M_k = 0$ for $k > n$, and $\deg M = \infty$ if no such n exists.

We fix a finite field $\mathbb{F}$ of cardinality q, and assume that all vector spaces are over $\mathbb{F}$. For a vector space X, we denote the group of automorphisms of X by $\mathrm{Aut}(X)$ or $\mathrm{GL}(X)$. When the dimension of X is n, we also denote these groups by GL_n. We denote the trivial vector space by 0, and we shall simply write $X \preceq Y$ whenever $\dim_{\mathbb{F}} X \leq \dim_{\mathbb{F}} Y$.

The monoidal category of Joyal and Street. We denote, by VB, the category of finite dimensional vector spaces with isomorphisms. A VB-module is a functor from VB to $\mathrm{Mod}_{\boldsymbol{k}}$. VB-modules form a category $\mathrm{Mod}_{\mathrm{VB}}$ which is naturally equivalent to the product category $\prod_{n\geq 0}\mathrm{Mod}_{\boldsymbol{k}[\mathrm{GL}_n]}$. In particular, a VB-module is naturally a nonnegatively graded $\boldsymbol{k}$-module. We denote, by V_d, the VB-module satisfying

$$V_d(X) = \begin{cases} V(X) & \text{if } \dim_{\mathbb{F}} X = d, \\ 0 & \text{if } \dim_{\mathbb{F}} X \neq d. \end{cases}$$

If $V = V_d$, we say that V is supported in degree d. Given VB-modules M, N we define an external product $\otimes_{\mathrm{VB}}$ by

$$(M \otimes_{\mathrm{VB}} N)(Y) = \bigoplus_{X \leq Y} M(Y/X) \otimes_{\boldsymbol{k}} N(X).$$

Then $\otimes_{\mathrm{VB}}$ turns $\mathrm{Mod}_{\mathrm{VB}}$ into a monoidal category; see [Joyal and Street 1995, Section 2].

The algebra A. Let $\boldsymbol{A}$ be the VB-module such that $\boldsymbol{A}_n = \boldsymbol{k}$ is the trivial representation of GL_n for each n. We have a map $\boldsymbol{A} \otimes_{\mathrm{VB}} \boldsymbol{A} \to \boldsymbol{A}$ given by

$$a \otimes b \in \boldsymbol{A}(Y/X) \otimes_{\boldsymbol{k}} \boldsymbol{A}(X) \mapsto ab \in \boldsymbol{A}(Y).$$

This turns $\boldsymbol{A}$ into an algebra object in the monoidal category $(\mathrm{Mod}_{\mathrm{VB}}, \otimes_{\mathrm{VB}})$. We denote the category of $\boldsymbol{A}$-modules by $\mathrm{Mod}_{\boldsymbol{A}}$. The VB-module $\boldsymbol{k} = \boldsymbol{A}/\boldsymbol{A}_+$ is naturally an $\boldsymbol{A}$-module. As usual, the *degree of generation* of an $\boldsymbol{A}$-module M is defined to be $\deg \boldsymbol{k} \otimes_{\boldsymbol{A}} M$. We shall denote $\deg \mathrm{Tor}_i^{\boldsymbol{A}}(\boldsymbol{k}, M)$ by $t_i(M)$, and so the degree of generation of M is $t_0(M)$. We say that an $\boldsymbol{A}$-module is *presented in finite degrees* if $t_0(M)$ and $t_1(M)$ are finite.

Definition of a VI-module. We denote, by VI, the category of finite dimensional vector spaces with injective linear maps. A VI-module is a functor from VI to $\mathrm{Mod}_{\boldsymbol{k}}$. We denote the category of VI-modules by $\mathrm{Mod}_{\mathrm{VI}}$. Let M be a VI-module. A VI morphism $f\colon X \to Y$ induces a map $M(X) \to M(Y)$ which we denote by $f_\star$. The VI-module M restricts to a VB-module and admits a natural map $\boldsymbol{A} \otimes_{\mathrm{VB}} M \to M$ given by

$$a \otimes b \in \boldsymbol{A}(Y/X) \otimes_{\boldsymbol{k}} M(X) \mapsto a\iota_\star(b) \in M(Y)$$

where $\iota\colon X \to Y$ is the inclusion. Conversely, if M is an $\boldsymbol{A}$-module and $f\colon X \to Y$ is a VI-morphism, then we have a map $f_\star\colon M(X) \to M(Y)$ given by the composite

$$M(X) \to M(f(X)) \xrightarrow{1\otimes -} A(Y/f(X)) \otimes_{\boldsymbol{k}} M(f(X)) \to M(Y)$$

where the first map comes from VB-module structure on M and the last map comes from $\boldsymbol{A}$-module structure on M. It is easy to see that the above discussion describes an equivalence of categories.

Proposition 2.1. $\mathrm{Mod}_{\mathrm{VI}}$ *is equivalent to* $\mathrm{Mod}_{\boldsymbol{A}}$.

We shall not distinguish between VI-modules and $\boldsymbol{A}$-modules. In particular, notions like degree of generation makes sense for VI-modules. We explain degree of generation from the VI perspective now. Given a VB-module V, we can upgrade it to a VI-module by declaring that all VI-morphisms that are not isomorphisms acts on V by 0. This defines a functor $\Psi^{\uparrow}\colon \mathrm{Mod}_{\mathrm{VB}} \to \mathrm{Mod}_{\mathrm{VI}}$. We define $\mathrm{H}_0^{\mathrm{VI}}$ to be the left adjoint to $\Psi^{\uparrow}$. Let M be a VI-module. Denote the smallest VI-submodule containing $M(Y)$ for $Y \prec X$ by $M_{\prec X}$. Then $\mathrm{H}_0^{\mathrm{VI}}(M)$ is given explicitly by

$$\mathrm{H}_0^{\mathrm{VI}}(M)(X) = (M/M_{\prec X})(X).$$

The functor $\mathrm{H}_0^{\mathrm{VI}}$ (called VI-*homology*) is same as the functor $\mathrm{Tor}_0^{\boldsymbol{A}}(\boldsymbol{k}, -) = \boldsymbol{k} \otimes_{\boldsymbol{A}} -$ under the equivalence above. We shall use the notation $\mathrm{H}_i^{\mathrm{VI}}(-)$ instead of $\mathrm{Tor}_i^{\boldsymbol{A}}(\boldsymbol{k}, -)$. Here are some basic results on VI-homology.

Proposition 2.2. *We have* $\mathrm{H}_0^{\mathrm{VI}}(M_{\prec d}) = \mathrm{H}_0^{\mathrm{VI}}(M)_{<d}$. *In particular, if* $n < m$ *then the natural map* $\mathrm{H}_0^{\mathrm{VI}}(M_{\prec n}) \to \mathrm{H}_0^{\mathrm{VI}}(M_{\prec m})$ *is just the inclusion map* $\mathrm{H}_0^{\mathrm{VI}}(M)_{<n} \to \mathrm{H}_0^{\mathrm{VI}}(M)_{<m}$.

Proposition 2.3. *Let M be a* VI*-module, and $f\colon M \to N$ be a morphism of* VI*-modules. Then we have the following*:

(a) $\mathrm{H}_0^{\mathrm{VI}}(M) = 0$ *if and only if* $M = 0$.

(b) $\mathrm{H}_0^{\mathrm{VI}}(f)$ *is an epimorphism if and only if f is an epimorphism.*

(c) *Suppose $t_0(M) \le d$ and $N(X) = 0$ for $X \prec \mathbb{F}^d$. Then $\mathrm{H}_0^{\mathrm{VI}}(f) = 0$ if and only if $f = 0$.*

Proof. Part (a) is just the Nakayama lemma and (b) follows from (a) and the right exactness of $\mathrm{H}_0^{\mathrm{VI}}$. For part (c), suppose $\mathrm{H}_0^{\mathrm{VI}}(f) = 0$. By part (a), it suffices to show that $\mathrm{H}_0^{\mathrm{VI}}(\operatorname{im} f) = 0$. First suppose X is a vector space of dimension at most d. Since $N(Y) = 0$ for all $Y \prec X$, the map $f(X)\colon M(X) \to N(X)$ factors through the projection $M(X) \to \mathrm{H}_0^{\mathrm{VI}}(M)(X)$ and $N(X)$ is naturally isomorphic to $\mathrm{H}_0^{\mathrm{VI}}(N)(X)$. This shows that

$$\mathrm{H}_0^{\mathrm{VI}}(\operatorname{im} f)(X) = (\operatorname{im} \mathrm{H}_0^{\mathrm{VI}}(f))(X) = 0.$$

Next suppose X is a vector space of dimension bigger than d. Since $M \to \operatorname{im} f$ is a surjection and $\mathrm{H}_0^{\mathrm{VI}}$ is right exact we see that $t_0(M) \le d \implies \mathrm{H}_0^{\mathrm{VI}}(\operatorname{im} f)(X) = 0$. Thus $\mathrm{H}_0^{\mathrm{VI}}(\operatorname{im} f) = 0$, completing the proof. □

Local cohomology and saturation. Let M be a VI-module. We say that an element $x \in M(X)$ is *torsion* if there exists an injective linear map $f\colon X \to Y$ such that $f_\star(x) = 0$. A VI-module is torsion if it consists entirely of torsion elements. We denote the maximal torsion submodule of M by $\Gamma(M)$, the i-th right derived functor of Γ by $\mathrm{R}^i\Gamma$, and the degree of $\mathrm{R}^i\Gamma(M)$ by $h_i(M)$. Let $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ be the category of torsion VI-modules. It is easy to see that $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}} \subset \mathrm{Mod}_{\mathrm{VI}}$ is a localizing subcategory. Let $\mathrm{T}\colon \mathrm{Mod}_{\mathrm{VI}} \to \mathrm{Mod}_{\mathrm{VI}} / \mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ be the corresponding localization functor and S be its right adjoint (the section functor). We define *saturation* of M to be the composition $\boldsymbol{S}(M) = \mathrm{ST}(M)$. We denote the i-th right derived functor of $\boldsymbol{S}$ by $\mathrm{R}^i\boldsymbol{S}$.

We refer the readers to [Sam and Snowden 2019, Section 4] where the formalism of local cohomology and saturation is discussed in quite generality. This formalism needed an assumption which in our case is the following:

$$\text{Injective objects of } \mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}} \text{ remain injective in } \mathrm{Mod}_{\mathrm{VI}}. \tag{*}$$

We note here that both $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ and $\mathrm{Mod}_{\mathrm{VI}}$ are Grothendieck abelian categories, and so both contain enough injectives.

Lemma 2.4. *The assumption (*), as above, holds. In particular, the injective hull (as* VI*-modules) of a torsion module is torsion.*

Proof. The first assertion follows immediately from Theorem 1.4 and [Sam and Snowden 2019, Proposition 4.18]. Now suppose that M is a torsion VI-module. Then we can embed M into an injective object I in $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$. By (*), I is injective in $\mathrm{Mod}_{\mathrm{VI}}$, and so I contains the injective hull of M. The second assertion is immediate from this. □

Lemma 2.5. *If I is injective in* $\mathrm{Mod}_{\mathrm{VI}}$, *then $\Gamma(I)$ is also injective in* $\mathrm{Mod}_{\mathrm{VI}}$. *In particular, if M is a torsion* VI*-module, then $\mathrm{R}^i\Gamma(M) = 0$ for $i > 0$.*

Proof. Since I is injective and contains $\Gamma(I)$, it follows that I contains the injective hull of $\Gamma(I)$. By the previous lemma and the maximality of $\Gamma(I)$, we conclude that $\Gamma(I)$ is its own injective hull. This proves the first assertion.

The first assertion implies that if M is a torsion module then it admits an injective resolution $M \to I^\bullet$ such that each I^i is torsion. Since Γ is identity on torsion modules, we see that $\Gamma(I^\bullet) = I^\bullet$. The second assertion follows. □

Corollary 2.6. *Let T be an object of the right derived category* $\mathrm{D}^+(\mathrm{Mod}_{\mathrm{VI}})$ *which can be represented by a complex of torsion* VI*-modules. Then* $\mathrm{R}\Gamma(T) \cong T$, *and* $\mathrm{R}\boldsymbol{S}(T) = 0$.

We now state a result from [Sam and Snowden 2019] that we need.

Proposition 2.7 [Sam and Snowden 2019, Proposition 4.6]. *Let $M \in \mathrm{D}^+(\mathrm{Mod}_{\mathrm{VI}})$. Then we have an exact triangle*

$$\mathrm{R}\Gamma(M) \to M \to \mathrm{R}\boldsymbol{S}(M) \to$$

where the first two maps are the canonical ones.

We call a VI-module M *derived saturated* if $M \to \mathrm{R}\boldsymbol{S}(M)$ is an isomorphism in $\mathrm{D}^+(\mathrm{Mod}_{\mathrm{VI}})$, or equivalently $\mathrm{R}\Gamma(M) = 0$ (see the proposition above).

3. Induced and semiinduced VI-modules

The aim of this section is to prove some formal properties of induced and semiinduced modules. The restriction map $\Psi_\downarrow\colon \mathrm{Mod}_{\mathrm{VI}} \to \mathrm{Mod}_{\mathrm{VB}}$ admits a left adjoint $\mathrm{Mod}_{\mathrm{VB}} \to \mathrm{Mod}_{\mathrm{VI}}$ denoted $\mathcal{I}$, which is exact. By definition of $\mathcal{I}$, we have the adjunction

$$\mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(\mathcal{I}(V), M) = \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VB}}}(V, M). \qquad (*)$$

We call VI-modules of the form $\mathcal{I}(V)$ *induced*. If V is supported in degree d we say that $\mathcal{I}(V)$ is induced from degree d. Moreover, when V_d is a VB-module isomorphic to $\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VB}}(\mathbb{F}^d, -)]$ then we denote $\mathcal{I}(V)$ by simply $\mathcal{I}(d)$. By Yoneda lemma, we have $\mathcal{I}(d) = \boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(\mathbb{F}^d, -)]$. We have the following alternative descriptions for $\mathcal{I}(V)$:

$$\mathcal{I}(V) = \boldsymbol{A} \otimes_{\mathrm{VB}} V, \quad \text{and} \quad \mathcal{I}(V) = \bigoplus_{d \geq 0} \mathcal{I}(d) \otimes_{\boldsymbol{k}[\mathrm{Aut}(\mathbb{F}^d)]} V(\mathbb{F}^d).$$

Proposition 3.1. *The composite functor* $\mathrm{H}_0^{\mathrm{VI}}\mathcal{I}$ *is naturally isomorphic to the identity functor on* VB*-modules. The counit* $\mathcal{I}\Psi_\downarrow \to \mathrm{id}$ *is an epimorphism on any* VI*-module.*

Proof. The first assertion is clear because composing $\boldsymbol{k} \otimes_{\boldsymbol{A}} -$ with $\boldsymbol{A} \otimes_{\mathrm{VB}} -$ yields $\boldsymbol{k} \otimes_{\mathrm{VB}} -$, which is naturally isomorphic to the identity functor. Alternatively, by adjointness of $\mathcal{I}$ and $\mathrm{H}_0^{\mathrm{VI}}$, we have

$$\mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VB}}}(\mathrm{H}_0^{\mathrm{VI}}\mathcal{I}(M), N) = \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VB}}}(M, \Psi_\downarrow\Psi^\uparrow N) = \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VB}}}(M, N),$$

and so the result follows by the uniqueness of left adjoints. For the second assertion, it suffices to check that $\Psi_\downarrow$ is faithful, which is trivial. ☐

A useful thing to note is that if M is a VI-module and $f: V \to M$ is a map of VB-modules then the image of the corresponding map $g: \mathcal{I}(V) \to M$ is the smallest VI-submodule of M containing the image of f. In particular, if $V(X) \to M(X)$ is surjective then $g(X)$ is surjective.

Proposition 3.2. *$\mathcal{I}(V)$ is a projective* VI*-module if and only if V is a projective* VB*-module. All projective* VI*-modules are of the form $\mathcal{I}(V)$.*

Proof. Each of $\mathcal{I}$ and $\mathrm{H}_0^{\mathrm{VI}}$ is left adjoint to an exact functor ($\Psi_\downarrow$ and $\Psi^\uparrow$ respectively), so both of them preserve projectives [Weibel 1994, Proposition 2.3.10]. Since $\mathrm{H}_0^{\mathrm{VI}}\mathcal{I} = \mathrm{id}$ (Proposition 3.1), we conclude that $\mathcal{I}(V)$ is projective if and only if V is projective.

For the second assertion, let P be a projective VI-module. By Proposition 3.1, there is a natural surjection $\phi: \mathcal{I}\Psi_\downarrow(P) \to P$, and since P is projective it admits a section s. Let $\psi: \mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(P) \to P$ be the map given by $\psi = \phi \circ \mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(s)$. It suffices to show that ψ is an isomorphism. By Proposition 3.1, we have

$$\mathrm{H}_0^{\mathrm{VI}}(\psi) \cong \mathrm{H}_0^{\mathrm{VI}}(\phi \circ s) = \mathrm{H}_0^{\mathrm{VI}}(\mathrm{id}) = \mathrm{id}\,.$$

Thus, by Proposition 2.3, ψ is surjective. Since P is projective we have a short exact sequence

$$0 \to \mathrm{H}_0^{\mathrm{VI}}(\ker\psi) \to \mathrm{H}_0^{\mathrm{VI}}(\mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(P)) \xrightarrow{\mathrm{H}_0^{\mathrm{VI}}(\psi)\cong\mathrm{id}} \mathrm{H}_0^{\mathrm{VI}}(P) \to 0.$$

In particular, $\mathrm{H}_0^{\mathrm{VI}}(\ker\psi) = 0$. Thus, by Proposition 2.3, we conclude that ψ is an isomorphism. This completes the proof. ☐

Corollary 3.3. $\mathrm{Mod}_{\mathrm{VI}}$ *has enough projectives.*

Proof. Clearly, $\mathrm{Mod}_{\mathrm{VB}} \cong \prod_{n\ge 0} \mathrm{Mod}_{\boldsymbol{k}[\mathrm{GL}_n]}$ has enough projectives. Now let M be a VI-module and let $P \to \Psi_\downarrow(M)$ be a surjection from a projective VB-module P. Then, the composite $\mathcal{I}(P) \to \mathcal{I}\Psi_\downarrow(M) \to M$ is a surjection (Proposition 3.1) and $\mathcal{I}(P)$ is projective (Proposition 3.2), completing the proof. ☐

Proposition 3.4. *$\mathrm{H}_i^{\mathrm{VI}}(\mathcal{I}(V)) = 0$ for $i > 0$ and is isomorphic to V for $i = 0$. In particular, $t_0(\mathcal{I}(V)) = \deg V$, and $\mathcal{I}(V)$ is presented in finite degrees if and only if $\deg(V) < \infty$.*

Proof. Let $P_\bullet \to V$ be a projective resolution of V as a VB-module. Then $\mathcal{I}(P_\bullet)$ is a projective resolution of $\mathcal{I}(V)$ (Proposition 3.2). The assertion now follows by applying $\mathrm{H}_0^{\mathrm{VI}}(-)$ and noting that $\mathrm{H}_0^{\mathrm{VI}}\mathcal{I} = \mathrm{id}$ (Proposition 3.1). ☐

Proposition 3.5. *Let $\mathcal{I}(U)$, $\mathcal{I}(V)$ be* VI*-modules induced from d. Then $\mathrm{H}_0^{\mathrm{VI}}$ induces an isomorphism*

$$\mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(\mathcal{I}(U), \mathcal{I}(V)) \to \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VB}}}(U, V),$$

whose inverse is given by $\mathcal{I}$.

Proof. By Proposition 3.1, $\mathrm{H}_0^{\mathrm{VI}}\mathcal{I} = \mathrm{id}$. Conversely, suppose $f \in \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(\mathcal{I}(U), \mathcal{I}(V))$. Then, again by Proposition 3.1, $\mathrm{H}_0^{\mathrm{VI}}(f - \mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(f)) = 0$. Thus, by Proposition 2.3(3), we conclude that $f - \mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(f) = 0$, completing the proof. ☐

Proposition 3.6. *Kernel and cokernel of a map of* VI*-modules induced from d are induced from d. An extension of* VI*-modules induced from d is induced from d.*

Proof. Let $f\colon \mathcal{I}(U) \to \mathcal{I}(V)$ be a map of VI-modules. Then by the previous proposition, there is a $g\colon U \to V$ such that $f = \mathcal{I}(g)$. Since $\mathcal{I}$ is exact, we have $\ker f = \mathcal{I}(\ker g)$ and $\operatorname{coker} f = \mathcal{I}(\operatorname{coker} g)$, proving the first assertion. For the second assertion, let M be an extension of $\mathcal{I}(U)$ and $\mathcal{I}(V)$. Let $P_\bullet \to U$ and $Q_\bullet \to V$ be projective resolutions of U and V such that P_i and Q_i are all supported in degree d. By the horseshoe lemma and Proposition 3.2, $\mathcal{I}(P_\bullet + Q_\bullet)$ is a projective resolution of M. By the first assertion, M is induced from d. □

Proposition 3.7. *Let $\mathcal{I}(W)$ be a module induced from d. And let M be a submodule of $\mathcal{I}(W)$ generated in degrees $\leq d$. Then M is isomorphic to $\mathcal{I}(M_d)$. In particular, M is induced from d.*

Proof. Since M is generated in degree d and $M_k \subset \mathcal{I}(W)_k = 0$ for $k < d$, we have $\mathrm{H}_0^{\mathrm{VI}}(M) = M_d$. It follows that the natural map $f\colon \mathcal{I}(M_d) \to M$ is a surjection (Proposition 2.3). Composing it with the inclusion $M \to \mathcal{I}(W)$, we obtain a map $g\colon \mathcal{I}(M_d) \to \mathcal{I}(W)$. By construction, $\mathrm{H}_0^{\mathrm{VI}}(g)$ is the natural inclusion $M_d \to W$. Thus by the Proposition 3.5, we have

$$\ker(g) = \ker(\mathcal{I}\mathrm{H}_0^{\mathrm{VI}}(g)) = \mathcal{I}(\ker(\mathrm{H}_0^{\mathrm{VI}}(g))) = \mathcal{I}(0) = 0.$$

This implies that f is injective, completing the proof. □

Proposition 3.8. *Let M be a* VI*-module. Then*:

(a) *M is generated in degrees $\leq d$ if and only if it admits a surjection $\mathcal{I}(V) \to M$ with* $\deg V \leq d$.

(b) *M is presented in finite degrees if and only if there is an exact sequence*

$$\mathcal{I}(W) \to \mathcal{I}(V) \to M \to 0$$

such that $\deg V, \deg W < \infty$.

Proof. (a) Suppose there is a surjection $\mathcal{I}(V) \to M$. Since $\mathrm{H}_0^{\mathrm{VI}}$ is right exact, we have a surjection $V \to \mathrm{H}_0^{\mathrm{VI}}(M)$. This shows that $\deg V \leq d \implies t_0(M) \leq d$. Conversely, suppose $t_0 \leq d$. Let V be the VB-module with $\deg V \leq d$ satisfying $V(X) = M(X)$ for $\dim X \leq d$. By construction, we have a surjection $V \to \mathrm{H}_0^{\mathrm{VI}}(M)$. By Nakayama lemma, the natural map $\mathcal{I}(V) \to M$ is a surjection, completing the proof.

(b) First suppose M is presented in finite degrees. Then by part (a), there is a surjection $f\colon \mathcal{I}(V) \to M$ with $\deg V < \infty$. It suffices to show that the kernel of f is generated in finite degrees. But this follows from the long exact sequence corresponding to $\mathrm{H}_0^{\mathrm{VI}}$. Conversely, if there is an exact sequence

$$\mathcal{I}(W) \to \mathcal{I}(V) \to M \to 0$$

such that $\deg V, \deg W < \infty$. Then by part (a), M and the kernel of $\mathcal{I}(V) \to M$ are generated in finite degrees. Again, the long exact sequence corresponding to $\mathrm{H}_0^{\mathrm{VI}}$ finishes the proof (see Proposition 3.4). □

Semiinduced modules. We call a module *semiinduced* if it admits a finite filtration whose graded pieces (successive quotients) are induced modules that are generated in finite degrees.

Lemma 3.9. *Suppose* $\mathrm{H}_1^{\mathrm{VI}}(Q) = 0$ *and assume that* $\mathrm{H}_0^{\mathrm{VI}}(Q)$ *is concentrated in degree* d. *Then* Q *is induced from* d. *In particular*, Q *is homology acyclic.*

Proof. By the assumption, $Q_d = \mathrm{H}_0^{\mathrm{VI}}(Q)$. This implies that there is a natural surjection $\phi\colon M := \mathcal{I}(\mathrm{H}_0^{\mathrm{VI}}(Q)) \to Q$ which induces an isomorphism $\mathrm{H}_0^{\mathrm{VI}}(M) \to \mathrm{H}_0^{\mathrm{VI}}(Q)$. By the assumption that $\mathrm{H}_1^{\mathrm{VI}}(Q) = 0$ and Nakayama's lemma, we see that the kernel of ϕ is trivial. This shows that Q is induced from d. The statement that Q is homology acyclic follows from Proposition 3.4. $\square$

The proof of the following proposition is motivated by a very similar theorem of Ramos for FI-modules [2018, Theorem B].

Proposition 3.10. *Let* M *be a module generated in finite degrees. Then* M *is homology acyclic if and only if* M *is semiinduced. More generally, if* $\mathrm{H}_1^{\mathrm{VI}}(M) = 0$ *then the graded pieces* (*successive quotients* $Q^i := M_{\preceq i}/M_{\prec i}$) *of the natural filtration*

$$0 \subset M_{\preceq 0} \subset \cdots \subset M_{\preceq d} = M$$

are induced (*more precisely*, Q^i *is induced from* i).

Proof. By Proposition 3.4, if M is semiinduced then it satisfies $\mathrm{H}_i^{\mathrm{VI}}(M) = 0$ for $i > 0$, and is thus acyclic. The reverse inclusion follows from the second assertion which we now prove by induction on $d := t_0(M)$. Note that $\mathrm{H}_0^{\mathrm{VI}}(Q^i)$ is concentrated in degree i, and $\mathrm{H}_0^{\mathrm{VI}}(M_{\prec d})$ injects into $\mathrm{H}_0^{\mathrm{VI}}(M_{\preceq d})$ (Proposition 2.2). Thus applying $\mathrm{H}_0^{\mathrm{VI}}(-)$ to the exact sequence

$$0 \to M_{\prec d} \to M \to Q^d \to 0$$

shows that $\mathrm{H}_1^{\mathrm{VI}}(Q^d) = 0$. By Lemma 3.9, Q^d is induced from d, and hence acyclic. Thus $\mathrm{H}_1^{\mathrm{VI}}(M_{\prec d}) = 0$. The rest follows by induction. $\square$

Corollary 3.11. *Suppose* M *is semiinduced module generated in degree* $\leq d$. *Then the graded pieces* (*successive quotients* $Q^i := M_{\preceq i}/M_{\prec i}$) *of the natural filtration*

$$0 \subset M_{\preceq 0} \subset \cdots \subset M_{\preceq d} = M$$

are induced (*more precisely*, Q^i *is induced from* i).

4. The shift theorem

The aim of this section is to prove our main result — the shift theorem.

The shift and the difference functors, I. The category of $\mathbb{F}$-vector spaces (and in particular, VI) has a symmetric monoidal structure $+$ given by the direct sum of vector spaces. It allows us to define a shift functor τ^X on $\mathbb{F}$-vector spaces (or on VI) by

$$\tau^X(Z) = X + Z.$$

Moreover, for any $\mathbb{F}$-linear map $\ell\colon X \to Y$, we have a natural transformation $\tau^\ell\colon \tau^X \to \tau^Y$ given by $\tau^\ell(Z) = \ell + \mathrm{id}_Z$.

We say that a morphism $f\colon \mathbb{F}^d \to X + Z$ is of *X-rank k* if the dimension of $(X + \operatorname{im} f)/X$ is k (clearly, X-rank of f is at most d). In other words, k is the least integer such that there are VI-morphisms $g\colon \mathbb{F}^d \to X + \mathbb{F}^k$ and $h\colon \mathbb{F}^k \to Z$ satisfying $f = \tau^X(h)g$. We call any decomposition of the form $f = \tau^X(h)g$ as above, an *(X, k)-decomposition* of f. The following lemma is immediate from basic linear algebra.

Lemma 4.1. *Let $\tau^X(h_1)g_1 = \tau^X(h_2)g_2$ are two (X, k)-decompositions of $f\colon \mathbb{F}^d \to X + Z$. Then there is a unique $\sigma \in \mathrm{GL}_k$ such that $g_2 = \tau^X(\sigma)g_1$ and $h_2 = h_1\sigma^{-1}$.*

Let $D^d_k(X, Z)$ be the free $\boldsymbol{k}$-module on morphisms $f\colon \mathbb{F}^d \to X + Z$ of X-rank k. Then $D^d_k(X, Z)$ is a VI-module in both of the arguments X and Z, and has a natural action of GL_d on the right.

Lemma 4.2. *We have the following*:

(a) $D^d_k(X, \mathbb{F}^k)$ *is a free $\boldsymbol{k}[\mathrm{GL}_k]$-module.*

(b) $D^d_k(X, Z) = \boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(\mathbb{F}^k, Z)] \otimes_{\boldsymbol{k}[\mathrm{GL}_k]} D^d_k(X, \mathbb{F}^k)$.

(c) *Given a* VI*-morphism $\ell\colon X \to Y$, the natural map*

$$\ell_\star\colon D^d_k(X, Z) \to D^d_k(Y, Z)$$

given by $f \mapsto \tau^Z(\ell)f$ is a split injection of VI*-modules in the variable Z.*

Proof. The first two parts are immediate from the previous lemma. Since $\ell\colon X \to Y$ is an injection, it admits an $\mathbb{F}$-linear section $s\colon Y \to X$ (which may not be an injection). This defines a map $\psi\colon D^d_k(Y, Z) \to D^d_k(X, Z)$ given by

$$f \mapsto \begin{cases} \tau^Z(s)f & \text{if } \tau^Z(s)f \text{ is injective,} \\ 0 & \text{if } \tau^Z(s)f \text{ is not injective.} \end{cases}$$

This map is clearly functorial in Z and is a section to $\ell_\star$, finishing the proof. □

The functor τ^X induces an exact functor Σ^X, which we again call the *shift functor*, on $\mathrm{Mod}_{\mathrm{VI}}$ given by $(\Sigma^X M)(Y) = M(\tau^X(Y)) = M(X + Y)$. An element $\phi \in \mathrm{Aut}(Y)$ acts on $(\Sigma^X M)(Y) = M(X + Y)$ where the action is induced by $\tau^X(\phi)$. Similarly, there is an action of $\mathrm{Aut}(X)$ on $\Sigma^X M(Y)$.

Proposition 4.3. *We have the following*:

(a) $\Sigma^X \mathcal{I}(d) = \bigoplus_{0 \le k \le d} \mathcal{I}(k) \otimes_{\boldsymbol{k}[\mathrm{GL}_k]} D^d_k(X, \mathbb{F}^k)$.

(b) $\Sigma^X \mathcal{I}(W) = \Sigma^X \mathcal{I}(d) \otimes_{\boldsymbol{k}[\mathrm{GL}_d]} W = \bigoplus_{0 \le k \le d} \mathcal{I}(k) \otimes_{\boldsymbol{k}[\mathrm{GL}_k]} D^d_k(X, \mathbb{F}^k) \otimes_{\boldsymbol{k}[\mathrm{GL}_d]} W$.

Where W is any $\boldsymbol{k}[\mathrm{GL}_d]$-module. In particular, shift of an induced module is induced, and shift of a projective VI*-module is projective.*

Proof. Since every VI-morphism $f\colon \mathbb{F}^d \to X+Z$ is of X-rank k at most d, we have an isomorphism $\Sigma^X \mathcal{I}(d)(Z) = \bigoplus_{0\le k\le d} D_k^d(X, Z)$. This isomorphism is clearly functorial in Z. The rest follows from the previous lemma. $\square$

Corollary 4.4. *The shift of an induced (semiinduced) $\mathcal{C}$-module is induced (respectively semiinduced). The category of modules generated (presented) in finite degrees is stable under shift. In particular, $t_0(\Sigma^X M) \le t_0(M)$.*

Proof. Exactness of the shift functor and the previous proposition yields the first assertion. The second assertion follows from Proposition 3.8 and the previous proposition. $\square$

Suppose $\ell \in \mathrm{Hom}_{\mathrm{VI}}(X, Y)$, and $\tau^\ell\colon \tau^X \to \tau^Y$ be the corresponding natural transformation. If M is a VI-module, then τ^ℓ naturally induces a map $\Sigma^\ell\colon \Sigma^X M \to \Sigma^Y M$ which is functorial in M. We denote the cokernel of this map by $\Delta^\ell M$. When $X=0$, we simply denote this cokernel by Δ^Y, or simply Δ if we also have $\dim_{\mathbb{F}} Y = 1$.

Proposition 4.5. *Let W be a* VB*-module. Then $\Sigma^\ell\colon \Sigma^X \mathcal{I}(W) \to \Sigma^Y \mathcal{I}(W)$ is split injective and $\Delta^\ell \mathcal{I}(W)$ is an induced module.*

Proof. If $f\colon \mathbb{F}^d \to X+Z$ is of X-rank k then $\tau^Z(\ell) f$ is clearly of Y-rank k. Thus $\ell_\star$ takes the k-th direct summand of $\Sigma^X \mathcal{I}(d)(Z) = \bigoplus_{0\le k\le d} D_k^d(X, Z)$ to the k-th direct summand of $\Sigma^Y \mathcal{I}(d)(Z) = \bigoplus_{0\le k\le d} D_k^d(Y, Z)$, and is functorial in Z. Thus it suffices to show that the map $\ell_\star\colon D_k^d(X, Z) \to D_k^d(Y, Z)$ is split and the cokernel is induced. That it is split is proven in Lemma 4.2(c), and that the cokernel is induced follows from Lemma 4.2(b) and Proposition 3.6. This proves the result when $W = \boldsymbol{k}[\mathrm{Hom}_{\mathrm{VB}}(\mathbb{F}^d, -)]$. The general result follows by observing that tensoring preserves split injections. $\square$

The following basic result is easy to establish.

Proposition 4.6. *Let $\ell \in \mathrm{Hom}_{\mathrm{VI}}(X, Y)$ and M be a* VI*-module. Then:*

(a) *The shift commutes with Γ. In particular, $h_0(\Sigma^X M) = \max(h_0(M) - \dim X, -1)$.*

(b) *The kernel of $\Sigma^\ell\colon \Sigma^X M \to \Sigma^Y M$ is a torsion module of degree $h_0(\Sigma^X M)$. In particular, $\Sigma^\ell\colon \Sigma^X M \to \Sigma^Y M$ is injective if $\dim X > h_0(M)$.*

The shift and the difference functors, II. We define another shift-like functor $\overline{\Sigma}$ which has better formal properties than Σ. We first set some notation. Let $\mathcal{F}$ be a flag on a vector space Z given by

$$0 = Z_0 \subset Z_1 \subset \cdots \subset Z_n = Z.$$

We call the stabilizer of $\mathcal{F}$ in $\mathrm{GL}(Z)$ the parabolic subgroup corresponding to $\mathcal{F}$ and denote it by $\boldsymbol{P}(\mathcal{F})$. The *unipotent radical* of $\boldsymbol{P}(\mathcal{F})$ is the kernel of the natural map

$$\boldsymbol{P}(\mathcal{F}) \to \prod_{i=1}^{n} \mathrm{GL}(Z_i/Z_{i-1})$$

and is denoted by $\boldsymbol{U}(\mathcal{F})$. Fix a maximal flag

$$0 = X_0 \subset X_1 \subset \cdots \subset X_n = X.$$

In particular, n is equal to the dimension of X. Set $Z_0 = 0$ and $Z_{i+1} = X_i + Z$ for $i \geq 0$. Denote the unipotent radical corresponding to the flag

$$0 = Z_0 \subset Z_1 \subset \cdots \subset Z_{n+1} = X + Z$$

by $\boldsymbol{U}_X(Z)$. Then $\boldsymbol{U}_X$ given by $Z \mapsto \boldsymbol{U}_X(Z) \cong Z^{\dim X} \rtimes \boldsymbol{U}_X(0)$ is clearly a VI-group, that is, $\boldsymbol{U}_X$ is a functor from VI to groups. This is in contrast with $Z \mapsto \mathrm{GL}(Z)$, which does not define a VI-group. We define $\overline{\Sigma}^X$ on $\mathrm{Mod}_{\mathrm{VI}}$ (or $\mathrm{Mod}_{\mathrm{VB}}$) by $\overline{\Sigma}^X M = (\Sigma^X M)_{\boldsymbol{U}_X}$, that is,

$$\overline{\Sigma}^X M(Z) = M(X+Z)_{\boldsymbol{U}_X(Z)}.$$

It is not hard to see that if M is a VI-module then $\overline{\Sigma}^X M$ is a VI-module. In fact, all we need to check is that for every VI-morphism $f\colon Z \to Z'$, $a \in M(X+Z)$ and $\sigma \in \boldsymbol{U}_X(Z)$ there exists a $\sigma' \in \boldsymbol{U}_X(Z')$ such that $\tau^X(f)_\star(\sigma_\star a - a) = \sigma'_\star \tau^X(f)_\star a - \tau^X(f)_\star a$. But one can simply take σ' to be $f_\star \sigma$ (the last expression makes sense because $\boldsymbol{U}_X$ is a VI-group) and check that the equation holds. Thus $\overline{\Sigma}^X\colon \mathrm{Mod}_{\mathrm{VI}} \to \mathrm{Mod}_{\mathrm{VI}}$ is a functor. Here we have suppressed the choice of flag on X. We drop the superscript X from Σ^X (or $\overline{\Sigma}^X$) when X is of dimension 1.

Suppose we are given an $\ell \in \mathrm{Hom}_{\mathrm{VI}}(X, Y)$ and maximal flags of X and Y such that ℓ takes the flag on X to an initial segment of the flag on Y. Any $\sigma \in \boldsymbol{U}_Y(Z)$ stabilizes $\ell(X) + Z$ and hence can be identified with an element of $\boldsymbol{U}_X(Z)$. This induces a surjection $\ell^\star\colon \boldsymbol{U}_Y \to \boldsymbol{U}_X$ of VI-groups. If M is a VI-module then we can make $\boldsymbol{U}_Y$ act on $\Sigma^X M$ via $\ell^\star$. Moreover, the map $\Sigma^\ell\colon \Sigma^X M \to \Sigma^Y M$ is $\boldsymbol{U}_Y$-equivariant. We define $\overline{\Sigma}^\ell = \Sigma^\ell_{\boldsymbol{U}_Y}$ and $\overline{\Delta}^\ell = \Delta^\ell_{\boldsymbol{U}_Y}$. Clearly, we have $(\Sigma^X M)_{\boldsymbol{U}_Y} = \overline{\Sigma}^X M$. So $\overline{\Sigma}^\ell$ is a map from $\overline{\Sigma}^X M$ to $\overline{\Sigma}^Y M$. It is not hard to see that $\overline{\Sigma}^\ell$ is a map of VI-modules. When $X = 0$, there is a unique map $\ell \in \mathrm{Hom}_{\mathrm{VI}}(X, Y)$, so in this case we drop the notation Σ^ℓ and simply call the map $M \to \overline{\Sigma}^Y M$ the natural map. We now note down some basic properties of $\overline{\Sigma}$ that we will use.

Lemma 4.7. *In the nondescribing characteristic, if Σ^ℓ is injective then so is $\overline{\Sigma}^\ell$. In particular, $\overline{\Sigma}^\ell$ is injective if* $\dim X > h_0(M)$.

Proof. This is clear because the size of the group $\boldsymbol{U}_Y(Z)$ is invertible in $\boldsymbol{k}$ for each Z, and Σ^ℓ is $\boldsymbol{U}_Y$-equivariant. □

The lemma immediately implies the following proposition.

Proposition 4.8. *Let ℓ be the unique map from 0 to X. In the nondescribing characteristic, the kernel of the map $\overline{\Sigma}^{\ell}\colon M \to \overline{\Sigma}^X M$ is torsion. In particular, if M is torsion-free then $\overline{\Sigma}^{\ell}$ is injective.*

Proposition 4.9. $t_0(\overline{\Sigma}^X M) \leq t_0(\Sigma^X M) \leq t_0(M)$.

Proof. The first inequality follows from the surjection $\Sigma^X M \to \overline{\Sigma}^X M$. The second is proven in Corollary 4.4. □

Remark 4.10. It is not true that $\Sigma^X \overline{\Sigma}^Y = \overline{\Sigma}^Y \Sigma^X$. In general, we only have a surjection $\overline{\Sigma}^Y \Sigma^X M \to \Sigma^X \overline{\Sigma}^Y M$. Since we have suppressed the data of the flag on $X+Y$ from $\overline{\Sigma}^{X+Y}$, we will be careful to never interchange X and Y. We adopt the convention that an initial segment of the maximal flag on $X+Y$ forms an initial segment of a maximal flag on Y (and not X).

Proposition 4.11. *We have the following natural isomorphisms:*

(a) $\Sigma^{X+Y} = \Sigma^Y \Sigma^X$.

(b) $\overline{\Sigma}^{X+Y} = \overline{\Sigma}^Y \overline{\Sigma}^X$.

In particular, Σ^X is isomorphic to $(\dim X)$-fold iterate of Σ. The same holds for $\overline{\Sigma}^X$.

Proof. Part (a) is trivial. Note that we have a short exact sequence of VI-groups

$$1 \to \Sigma^Y \boldsymbol{U}_X \to \boldsymbol{U}_{X+Y} \to i_X(\boldsymbol{U}_Y) \to 1$$

where $i_X(Z)\colon \mathrm{GL}(Y+Z) \to \mathrm{GL}(X+Y+Z)$ is the natural map. Part (b) now follows from

$$\overline{\Sigma}^{X+Y} M = (\Sigma^{X+Y} M)_{\boldsymbol{U}_{X+Y}} = ((\Sigma^{X+Y} M)_{\Sigma^Y \boldsymbol{U}_X})_{i_X(\boldsymbol{U}_Y)} = (\Sigma^Y \overline{\Sigma}^X M)_{i_X(\boldsymbol{U}_Y)} = \overline{\Sigma}^Y \overline{\Sigma}^X M. \qquad \square$$

The following proposition is the most crucial for our purpose.

Proposition 4.12. *Let X be a vector space of dimension one. Then $\overline{\Sigma}^X$ is a categorical derivation, that is, we have*

$$\overline{\Sigma}^X (M \otimes N) = (\overline{\Sigma}^X M \otimes N) \bigoplus (M \otimes \overline{\Sigma}^X N).$$

In particular, $\overline{\Sigma}\mathcal{I}(V) = \mathcal{I}(V) \oplus \mathcal{I}(\overline{\Sigma} V)$ and $\overline{\Delta}\mathcal{I}(V) = \mathcal{I}(\overline{\Sigma} V)$.

Proof. Let $V \leq W + X$. Then either V is contained in W and $\boldsymbol{U}_X(W)$ acts trivially on V, or there is an element $\sigma \in \boldsymbol{U}_X(W)$ such that σV is of the form $V' + X$ for some subspace V' of W. Moreover, if $\tau V = V'' + X$ for some $V'' \leq W$ and $\tau \in \boldsymbol{U}_X(W)$ then we must have $V' + X = V'' + X$. This shows that

$\sigma^{-1}\tau \in \boldsymbol{U}_X(V')$. Thus we have

$$
\begin{aligned}
&\overline{\Sigma}^X(M\otimes N)(W)\\
&\quad=(M\otimes N)(W+X)_{\boldsymbol{U}_X(W)}\\
&\quad=\Bigl(\bigoplus_{V\le W+X} M((W+X)/V)\otimes_{\boldsymbol{k}} N(V)\Bigr)_{\boldsymbol{U}_X(W)}\\
&\quad=\Bigl(\bigoplus_{V\le W} M((W+X)/V)\otimes_{\boldsymbol{k}} N(V)\Bigr)_{\boldsymbol{U}_X(W)}\bigoplus\Bigl(\bigoplus_{V'\le W} M((W+X)/(V'+X))\otimes_{\boldsymbol{k}} N(V'+X)\Bigr)_{\boldsymbol{U}_X(V')}\\
&\quad=\Bigl(\bigoplus_{V\le W} M(W/V+X)_{\boldsymbol{U}_X(W/V)}\otimes_{\boldsymbol{k}} N(V)\Bigr)\bigoplus\Bigl(\bigoplus_{V'\le W} M(W/V')\otimes_{\boldsymbol{k}} N(V'+X)_{\boldsymbol{U}_X(V')}\Bigr)\\
&\quad=(\overline{\Sigma}^X M\otimes N)(W)\bigoplus(M\otimes\overline{\Sigma}^X N)(W).
\end{aligned}
$$

This completes the proof of the first assertion. For the second assertion, just note that $\mathcal{I}(V)=\boldsymbol{A}\otimes V$ and apply the previous part. □

We have the following basic observations.

Lemma 4.13. *Let $A, B\colon \mathcal{C}_1\to\mathcal{C}_2$ be two functors between Grothendieck categories. Suppose there is a natural transformation $\Psi\colon A\to B$ such that $\Psi(P)$ is an isomorphism for each projective object $P\in\mathcal{C}_1$. If A, B are right exact then $\Psi(M)$ is an isomorphism for each $M\in\mathcal{C}_1$.*

Lemma 4.14. *Let A, B, C be right exact functors between two Grothendieck categories $\mathcal{C}_1, \mathcal{C}_2$. Suppose there are natural transformations*

$$A\xrightarrow{\Psi}B\xrightarrow{\Phi}C$$

such that for each projective $P\in\mathcal{C}_1$, the composite $A(P)\to B(P)\to C(P)$ vanishes. Then Φ factors through $\operatorname{coker}(\Psi)$.

Part (b) of the proposition below is motivated by the footnote in [Church 2016].

Proposition 4.15. *Let X and Y be vector spaces of dimension one. We have the following equality of functors*:

(a) $\Sigma^X\Delta^Y=\Delta^Y\Sigma^X$.

(b) $\mathrm{H}_0^{\mathrm{VI}}\overline{\Delta}=\overline{\Sigma}\mathrm{H}_0^{\mathrm{VI}}$.

Proof. (a) is identical to [Djament and Vespa 2019, Proposition 1.4(5)]. We provide a proof sketch here. In the following natural commutative diagram the vertical arrows are isomorphisms and so the cokernel of the horizontal maps are also isomorphic:

$$
\begin{array}{ccc}
\Sigma^Y & \longrightarrow & \Sigma^Y\Sigma^X\\
\big\downarrow & & \big\downarrow\\
\Sigma^Y & \longrightarrow & \Sigma^X\Sigma^Y
\end{array}
$$

This shows that

$$\Sigma^Y \Delta^X = \operatorname{coker}(\Sigma^Y \to \Sigma^Y \Sigma^X) = \operatorname{coker}(\Sigma^Y \to \Sigma^X \Sigma^Y) = \Delta^X \Sigma^Y,$$

completing the proof of (a).

(b) Composing the natural transformation $\mathrm{id} \to \mathrm{H}_0^{\mathrm{VI}}$ with $\bar{\Sigma}$ we obtain $\bar{\Sigma} \to \bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}}$. Since $\mathrm{H}_0^{\mathrm{VI}}\bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}} = \bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}}$, we obtain a transformation $\mathrm{H}_0^{\mathrm{VI}}\bar{\Sigma} \to \bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}}$. We shall now apply Lemma 4.14 to the composite

$$\mathrm{H}_0^{\mathrm{VI}} \to \mathrm{H}_0^{\mathrm{VI}}\bar{\Sigma} \to \bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}}.$$

To check the hypothesis of the lemma, it is enough to assume that $P = \mathcal{I}(V)$ where V is concentrated in degree d (Proposition 3.2). Evaluating the composite above at P yields

$$V \to V \oplus \bar{\Sigma} V \to \bar{\Sigma} V.$$

From degree considerations, hypothesis of Lemma 4.14 is satisfied. Thus we conclude that there is a natural transformation $\mathrm{H}_0^{\mathrm{VI}}\bar{\Delta} \to \bar{\Sigma}\mathrm{H}_0^{\mathrm{VI}}$. By Lemma 4.13 and Proposition 3.2, this transformation is an isomorphism. This completes the proof. □

Remark 4.16. There does not seem to be an equivalence between $\bar{\Sigma}^X\bar{\Delta}^Y$ and $\bar{\Delta}^Y\bar{\Sigma}^X$. This is in contrast with the case of FI-modules.

We denote the kernel of the natural transformation $\mathrm{id} \to \bar{\Sigma}^X$ by κ^X.

Proposition 4.17. *In the nondescribing characteristic, we have* $\mathrm{L}_1\bar{\Delta}^X = \kappa^X$, *and* $\mathrm{L}_i\bar{\Delta}^X = 0$ *for* $i > 1$.

Proof. The proof is the same as that of [Church and Ellenberg 2017, Lemma 4.7], where $\bar{\Sigma}^X$ plays the role of S. We provide a proof sketch here. Given a VI-module M, we can find a presentation

$$0 \to K \to F \to M \to 0,$$

where F is a projective VI-module, and K is torsion-free. The corresponding long exact sequence for the right exact functor $\bar{\Delta}^X$ implies that $\mathrm{L}_1\bar{\Delta}^X(M) = \ker(\bar{\Delta}^X K \to \bar{\Delta}^X F)$. Note that $F \to \Sigma^X F$ is injective, as F is torsion-free. By Lemma 4.7, we conclude that $F \to \bar{\Sigma}^X F$ is injective. Thus we have the following commutative diagram:

$$\begin{array}{ccccccccc}
 & & K & \longrightarrow & \bar{\Sigma}^X K & \longrightarrow & \bar{\Delta}^X K & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & F & \longrightarrow & \bar{\Sigma}^X F & \longrightarrow & \bar{\Delta}^X F & \longrightarrow & 0
\end{array}$$

Applying the snake lemma, we see that

$$\ker(\bar{\Sigma}^X K \to \bar{\Sigma}^X F) = 0 \to \mathrm{L}_1\bar{\Delta}^X(M) \to M \to \bar{\Sigma}^X M \to \bar{\Delta}^X M \to 0.$$

This shows that $\mathrm{L}_1\bar{\Delta}^X(M) = \kappa^X(M)$, finishing the proof of the first assertion. By dimension shifting, we have $\mathrm{L}_2\bar{\Delta}^X(M) = \mathrm{L}_1\bar{\Delta}^X(K) = \kappa^X(K)$. Since K is torsion-free, we see that $\mathrm{L}_2\bar{\Delta}^X(M) = 0$. Since M is arbitrary it follows that $\mathrm{L}_i\bar{\Delta}^X = 0$ for $i > 1$. □

The following lemma is proven in a similar way as [Djament and Vespa 2019, Proposition 1.4(7)].

Lemma 4.18. *Let M be a* VI*-module, and X, Y be vector spaces. We have an exact sequence of the form*

$$\overline{\Delta}^Y M \to \overline{\Delta}^{X+Y} M \to \overline{\Sigma}^Y \overline{\Delta}^X M \to 0.$$

Moreover, in the nondescribing characteristic, this can be extended to

$$0 \to \kappa^Y M \to \kappa^{X+Y} M \to \overline{\Sigma}^Y \kappa^X M \to \overline{\Delta}^Y M \to \overline{\Delta}^{X+Y} M \to \overline{\Sigma}^Y \overline{\Delta}^X M \to 0.$$

Proof. Let $\ell\colon 0 \to Y$, $\ell'\colon 0 \to X$ and $\ell''\colon 0 \to X+Y$ be natural maps. Then we have composable maps $\overline{\Sigma}^\ell\colon M \to \overline{\Sigma}^Y M$ and $\overline{\Sigma}^Y \overline{\Sigma}^{\ell'}\colon \overline{\Sigma}^Y M \to \overline{\Sigma}^Y \overline{\Sigma}^X M$, where the composite is $(\overline{\Sigma}^Y \overline{\Sigma}^{\ell'}) \circ \overline{\Sigma}^\ell = \overline{\Sigma}^{\ell''}$. Two composable morphisms u, v in an abelian category induce an exact sequence [Mac Lane 1963, Exercise 6, Section II.5]

$$0 \to \ker(u) \to \ker(v \circ u) \to \ker(v) \to \operatorname{coker}(u) \to \operatorname{coker}(v \circ u) \to \operatorname{coker}(v) \to 0.$$

Set $u = \overline{\Sigma}^\ell$ and $v = \overline{\Sigma}^Y \overline{\Sigma}^{\ell'}$. Since $\overline{\Sigma}^Y$ is right exact we see that $\operatorname{coker} v = \overline{\Sigma}^Y \overline{\Delta}^X M$ and the first assertion follows. In nondescribing characteristic, $\overline{\Sigma}^Y$ is exact. Thus we have $\ker(v) = \overline{\Sigma}^Y \kappa^X M$. This finishes the proof. □

Corollary 4.19. *Let X and Y be vector spaces, and fix maximal flags on X and Y. Let $\ell \in \operatorname{Hom}_{\mathrm{VI}}(X, Y)$ be a map that takes the maximal flag on X to an initial segment of the flag on Y. Then $t_0(\overline{\Delta}^\ell M) < t_0(M)$.*

Proof. Choose a complement Z of $\ell(X)$ in Y. Then the maximal flag on Y will induce a maximal flag on Z. We can identify ℓ with $\tau^X(\ell')$ where $\ell'\colon 0 \to Z$. This shows that $\overline{\Sigma}^\ell = \overline{\Sigma}^X \overline{\Delta}^{\ell'} = \overline{\Sigma}^X \overline{\Delta}^Z$. Thus by Corollary 4.4, it is enough to show that $t_0(\overline{\Delta}^Z M) < t_0(M)$. By the previous lemma, it suffices to prove it in the case when $\dim Z = 1$. But in this case, we have $t_0(\overline{\Delta}^Z M) = \deg(\overline{\Sigma} \mathrm{H}_0^{\mathrm{VI}}(M)) < t_0(M)$ (see Proposition 4.15). This completes the proof. □

Derived saturated objects. Our aim here is to show that the semiinduced modules are always derived saturated, and that the converse holds in the nondescribing characteristic. We recall that a module M is derived saturated if and only if $\mathrm{R}\Gamma(M) = 0$ (Proposition 2.7).

Lemma 4.20. *The natural map $\Sigma(\mathrm{R}^i\Gamma)(M) \to (\mathrm{R}^i\Gamma)\Sigma M$ is an isomorphism. Equivalently, Σ preserves Γ-acyclic objects.*

Proof. We follow the argument in [Djament 2016, Proposition A.3] to prove our assertion. The proof is by induction on i. The base case $i = 0$ is immediate as Σ commutes with Γ. Suppose that $i > 0$, and that the result has been proven for $j < i$.

We first apply a dimension shifting argument to see that the natural map $\Sigma(\mathrm{R}^i\Gamma)(M) \to (\mathrm{R}^i\Gamma)\Sigma M$ is injective. To see this, consider any exact sequence

$$0 \to M \to I \to N \to 0$$

where I is an injective. This yields a commutative diagram

$$\begin{tikzcd}
\Sigma(\mathrm{R}^{i-1}\Gamma)I \arrow[r] \arrow[d] & \Sigma(\mathrm{R}^{i-1}\Gamma)(N) \arrow[r] \arrow[d] & \Sigma(\mathrm{R}^{i}\Gamma)(M) \arrow[r] \arrow[d] & 0 \arrow[d] \\
(\mathrm{R}^{i-1}\Gamma)\Sigma I \arrow[r] & (\mathrm{R}^{i-1}\Gamma)\Sigma N \arrow[r] & (\mathrm{R}^{i}\Gamma)\Sigma M \arrow[r] & (\mathrm{R}^{i}\Gamma)\Sigma I
\end{tikzcd}$$

whose rows are exact. By induction, the first two vertical arrows are isomorphisms. Thus by the four lemma, we see that the third vertical arrow is injective.

By Lemma 2.5, we see that $\mathrm{R}^k\Gamma N = 0$ whenever $k > 0$ and N is a torsion module. Thus, for any $i > 0$, $\mathrm{R}^i\Gamma(M/\Gamma(M)) = \mathrm{R}^i\Gamma M$. Given a vector space X, we have the following natural exact sequence

$$0 \to M/\Gamma(M) \to \Sigma^X M \to \Delta^X M \to 0.$$

By the corresponding long exact sequence for Γ, we obtain the following exact sequence

$$\mathrm{R}^{i-1}\Gamma(\Sigma^X M) \to \mathrm{R}^{i-1}\Gamma(\Delta^X M) \to \mathrm{R}^i\Gamma(M) \to \Sigma^X \mathrm{R}^i\Gamma(M),$$

where the exactness comes from the injectivity of the map $\Sigma^X(\mathrm{R}^i\Gamma)(M) \to (\mathrm{R}^i\Gamma)\Sigma^X M$ proved in the previous paragraph. We conclude that

$$\ker(\mathrm{R}^i\Gamma(M) \to (\mathrm{R}^i\Gamma)\Sigma^X M) = \operatorname{coker}(\mathrm{R}^{i-1}\Gamma(\Sigma^X M) \to \mathrm{R}^{i-1}\Gamma(\Delta^X M)).$$

Since Σ is exact, and commutes with Σ^X and Δ^X (Proposition 4.15), we see that

$$\begin{aligned}
\Sigma \ker(\mathrm{R}^i\Gamma(M) \to (\mathrm{R}^i\Gamma)\Sigma^X M) &= \Sigma \operatorname{coker}(\mathrm{R}^{i-1}\Gamma(\Sigma^X M) \to \mathrm{R}^{i-1}\Gamma(\Delta^X M)) \\
&= \operatorname{coker}(\mathrm{R}^{i-1}\Gamma(\Sigma^X \Sigma M) \to \mathrm{R}^{i-1}\Gamma(\Delta^X \Sigma M)) \qquad \text{(by induction)} \\
&= \ker(\mathrm{R}^i\Gamma(\Sigma M) \to (\mathrm{R}^i\Gamma)\Sigma^X \Sigma M)
\end{aligned}$$

Thus Σ commutes with $\ker(\mathrm{id} \to \Sigma^X) \circ (\mathrm{R}^i\Gamma)$ for any X. Since X is arbitrary and Σ is cocontinuous, we see that Σ commutes with $\mathrm{R}^i\Gamma$. This finishes the proof. □

The following result is motivated by [Djament 2016, Proposition 1.1].

Proposition 4.21. *If F is an induced* VI*-module, then* $\mathrm{R}^i\Gamma(F) = 0$ *for* $i \geq 0$.

Proof. We have the following natural commutative diagram:

$$\begin{tikzcd}
\mathrm{R}^i\Gamma(F) \arrow[r, "\Sigma^\ell"] \arrow[dr, "\mathrm{R}^i\Gamma(\Sigma^\ell)"'] & \Sigma^X \mathrm{R}^i\Gamma(F) \arrow[d] \\
& \mathrm{R}^i\Gamma(\Sigma^X F)
\end{tikzcd}$$

where ℓ is the map from 0 to X. Since Σ^ℓ, applied to F, is split-injective (Proposition 4.5), we see that $\mathrm{R}^i\Gamma(\Sigma^\ell)$ is injective. By the previous lemma, the vertical map is an isomorphism. Thus the map $\Sigma^\ell\colon \mathrm{R}^i\Gamma(F) \to \Sigma^X\mathrm{R}^i\Gamma(F)$ is injective as well. Since X is arbitrary, we see that $\mathrm{R}^i\Gamma(F)$ is torsion-free. By definition, $\mathrm{R}^i\Gamma(F)$ is also a torsion VI-module. Hence $\mathrm{R}^i\Gamma(F) = 0$. □

Corollary 4.22. *Semiinduced modules are derived saturated.*

Corollary 4.23. *In a short exact sequence, if two of the objects are semiinduced then so is the third.*

Proof. Let $0 \to L \to M \to N \to 0$ be an exact sequence of modules presented in finite degrees. Then there exists a d such that L, M, N are generated in degree $\leq d$. We proceed by induction on d. First suppose that N is semiinduced. In this case, $\mathrm{H}_1^{\mathrm{VI}}(L) = 0$ if and only if $\mathrm{H}_1^{\mathrm{VI}}(M) = 0$. So the result follows from Proposition 3.10. Now suppose that L and M are semiinduced. By the previous corollary, N is derived-saturated. In particular, N is torsion-free. We claim that

$$0 \to L_{\prec d} \to M_{\prec d} \to N_{\prec d} \to 0$$

is an exact sequence. To see this, first note that we have a natural exact sequence

$$0 \to L \cap M_{\prec d} \to M_{\prec d} \to N_{\prec d} \to 0,$$

and that $L_{\prec d} \subset L \cap M_{\prec d}$. Now suppose, if possible, x is in $L \cap M_{\prec d}$ but not in $L_{\prec d}$. Then there exists a $y \in M(\mathbb{F}^{d-1})$ and a VI-morphism f such that $f_\star(y) = x \in L$. Since x is not in $L_{\prec d}$, we see that $y \notin L(\mathbb{F}^{d-1})$. Let $\bar{x}$, $\bar{y}$ be the images of x and y in N. Then $\bar{y} \neq 0$, but $f_\star(\bar{y}) = \bar{x} = 0$. This contradicts the fact that N is torsion-free, proving the claim.

By induction, $N_{\prec d}$ is semiinduced. Thus it suffices to show that $N/N_{\prec d}$ is induced from d. By applying the snake lemma to the diagram,

$$\begin{tikzcd}
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0 \\
0 \arrow[r] & L_{\prec d} \arrow[r] \arrow[u] & M_{\prec d} \arrow[r] \arrow[u] & N_{\prec d} \arrow[r] \arrow[u] & 0
\end{tikzcd}$$

we obtain an exact sequence

$$0 \to L_{\preceq d}/L_{\prec d} \to M_{\preceq d}/M_{\prec d} \to N_{\preceq d}/N_{\prec d} \to 0.$$

Since the first two objects in this exact sequence are induced from d, so is the third (Proposition 3.6). This completes the proof. □

Question 4.24. Let A, B, N be semiinduced modules and assume that $A, B \subset N$. Then is it true that $A \cap B$ is semiinduced?

The case of nondescribing characteristic. We now assume that we are in the nondescribing characteristic and prove the converse of Corollary 4.22. Along the way, we show that $\bar{\Sigma}$ commutes with Γ which, indeed, is a crucial step of our proof.

Lemma 4.25. *Let V be a $\boldsymbol{k}[G]$-module, and assume that the size of G is invertible in $\boldsymbol{k}$. Let x be an element of V_G, and let $\tilde{x}$ be a lift of x in V. Then*

(a) $1/|G| \sum_{\sigma \in G} \sigma \tilde{x}$ *in another lift of x.*

(b) $x = 0$ *if and only if* $\sum_{\sigma \in G} \sigma \tilde{x} = 0$.

Proof. This is a standard result. $\square$

Lemma 4.26. *Let M be a torsion-free* VI*-module, and let X be a vector space. Then $\bar{\Sigma}^X M$ is torsion-free.*

Proof. We may assume that X is of dimension one (Proposition 4.11). Let Y be another vector space of dimension one. It suffices to show that the map $f_\star\colon \bar{\Sigma}^X M(Z) \to \bar{\Sigma}^X M(Z+Y)$ induced by the inclusion $f\colon Z \to Z+Y$ is injective for every Z. Suppose $f_\star(x)=0$ for some x. By the previous lemma, there is a lift $\tilde{x} \in \Sigma^X M(Z) = M(X+Z)$ of x which is invariant with respect to $\boldsymbol{U}_X(Z)$. Since $f_\star(x)=0$ and $f_\star(\tilde{x}) \in \Sigma^X M(Z+Y) = M(X+Z+Y)$ is a lift of $f_\star(x)$, the previous lemma tells us that

$$\sum_{\sigma \in \boldsymbol{U}_X(Y+Z)} \sigma f_\star(\tilde{x}) = 0.$$

But $\boldsymbol{U}_X(Y+Z) = \boldsymbol{U}_X(Y) \times \boldsymbol{U}_X(Z)$ and $\tilde{x}$ is invariant with respect to $\boldsymbol{U}_X(Z)$, and so we conclude that

$$\sum_{\sigma \in \boldsymbol{U}_X(Y)} \sigma f_\star(\tilde{x}) = 0.$$

Let W be the VB module given by $\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VB}}(X' \oplus Z, -)]$ where X' is a one-dimensional space. Fix an isomorphism $\alpha\colon X'+Z \to X+Z$. Then $[\alpha]$ is a generator of the VI-module $\mathcal{I}(W)$. There is a unique map $\psi\colon \mathcal{I}(W) \to M$ which takes $[\alpha]$ to $\tilde{x}$. Let N be the VI-submodule of $\mathcal{I}(W)$ generated by $\sum_{\sigma \in \boldsymbol{U}_X(Y)} \sigma f_\star([\alpha])$. Then the equation at the end of the last paragraph shows that ψ factors through the projection $\mathcal{I}(W) \to \mathcal{I}(W)/N$. We claim that $\psi = 0$. Since M is torsion-free and ψ factors through $\mathcal{I}(W)/N$, it suffices to show that $\mathcal{I}(W)/N$ is a torsion module. Fix an isomorphism $h\colon Y \to X$. Let S be the collection consisting of $q-1$ automorphisms of $X+Y+Z$ that fix Z, send Y to X via h, and send X to Y via a nonzero multiple of h^{-1}. Then the following equation can be easily verified:

$$\Bigg(\sum_{\tau \in \boldsymbol{U}_Y(X)} \tau - \sum_{\tau \in S} \tau\Bigg)\Bigg(\sum_{\sigma \in \boldsymbol{U}_X(Y)} \sigma f_\star([\alpha])\Bigg) = q f_\star([\alpha]).$$

Since q is invertible, the above equation shows that $f_\star([\alpha]) \in N$. This shows that $\mathcal{I}(W)/N$ is torsion, and so $\psi = 0$. This implies that $x = 0$, completing the proof. $\square$

Proposition 4.27. *$\bar{\Sigma}$ commutes with Γ.*

Proof. Let M be a VI-module, and X be a vector space of dimension one so that $\bar{\Sigma} = \bar{\Sigma}^X$. Since $\bar{\Sigma}$ is exact and $\Gamma M \subset M$, we see that $\bar{\Sigma}\Gamma M \subset \Gamma\bar{\Sigma} M$. For the reverse inclusion, first note that $M/\Gamma M$ is torsion-free. Thus by the previous lemma and the exactness of $\bar{\Sigma}$, we see that

$$\bar{\Sigma}(M/\Gamma M) = (\bar{\Sigma} M)/(\bar{\Sigma}\Gamma M)$$

is torsion-free, and so the torsion part $\Gamma\bar{\Sigma} M$ of $\bar{\Sigma} M$ is contained in $\bar{\Sigma}\Gamma(M)$, completing the proof. $\square$

We now focus on showing that $\bar{\Sigma}$ preserves Γ-acyclic objects. We need a couple of lemma.

Lemma 4.28 [Djament 2016, Corollaire A.4]. *Let M be a* VI*-module, and let n be a nonnegative integer. Then the following are equivalent*:

(a) $\mathrm{R}^k\Gamma(M) = 0$ *for* $0 \le k \le n$.

(b) *For each* $0 \le k \le n$ *and vector spaces* $X_1, \dots, X_k$, *the* VI-*module* $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}M$ *is torsion-free.*

Proof. We prove the assertion by induction on n. The base case $n = 0$ is trivial. Assume now that $n > 0$, and that the assertion holds for smaller values of n.

Suppose first that (b) holds. Then, by induction, $\mathrm{R}^k\Gamma(M) = 0$ for $0 \le k < n$. In particular, M is torsion free. So for any vector space X, we have a short exact sequence:

$$0 \to M \to \Sigma^X M \to \Delta^X M \to 0.$$

By induction, $\mathrm{R}^k\Gamma(\Delta^X M) = 0$ for $0 \le k < n$. Thus the long exact sequence corresponding to the exact sequence above yields that $\mathrm{R}^n\Gamma(M) \to \mathrm{R}^n\Gamma(\Sigma^X M)$ is injective. We have the following natural commutative diagram:

$$\begin{tikzcd}
\mathrm{R}^n\Gamma(M) \arrow[r, "\Sigma^\ell"] \arrow[dr, "\mathrm{R}^n\Gamma(\Sigma^\ell)"'] & \Sigma^X\mathrm{R}^n\Gamma(M) \arrow[d] \\
& \mathrm{R}^n\Gamma(\Sigma^X M)
\end{tikzcd}$$

where ℓ is the map from 0 to X. Since the vertical map is an isomorphism (Lemma 4.20), we conclude that the horizontal map is injective as well. Since this holds for each X and $\mathrm{R}^n\Gamma(M)$ is a torsion module, we have $\mathrm{R}^n\Gamma(M) = 0$. Thus (a) holds.

Conversely, suppose that (a) holds. Since $n > 0$, the module M is torsion-free. So for any vector space X, we have a short exact sequence

$$0 \to M \to \Sigma^X M \to \Delta^X M \to 0.$$

The corresponding long exact sequence yields $\mathrm{R}^k\Gamma(\Sigma^X M) \cong \mathrm{R}^k\Gamma(\Delta^X M)$ for $0 \le k < n$. By Lemma 4.20, we conclude that $\mathrm{R}^k\Gamma(\Delta^X M) = 0$ for $0 \le k < n$. Now (b) follows immediately from the induction hypothesis. This completes the proof. $\square$

Lemma 4.29. *Let* M *be a* VI-*module and let* $X, X_1, \dots, X_k$ *be vector spaces. Suppose that the* VI-*module* $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}M$ *is torsion-free. Then* $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}\overline{\Sigma}^X M$ *is torsion-free.*

Proof. By Proposition 4.15, we see that $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}\Sigma^X M = \Sigma^X\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}M$. Set $N = \Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}M$, and note that

$$\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}\overline{\Sigma}^X M(Z) = \Sigma^X\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}M(Z)_{U_X\left(\sum_{i=1}^k X_i+Z\right)} = N(X+Z)_{U_X\left(\sum_{i=1}^k X_i+Z\right)}.$$

Set $\boldsymbol{V}_X(-) = \boldsymbol{U}_X\left(\sum_{i=1}^k X_i + -\right)$. We now follow the proof of Lemma 4.26 closely.

We may assume without loss of generality that X is of dimension one. Let Y be another vector space of dimension one. It suffices to show that the map $f_\star\colon \Sigma^X N(Z)_{\boldsymbol{V}_X(Z)} \to \Sigma^X N(Z+Y)_{\boldsymbol{V}_X(Z+Y)}$ induced by the inclusion $f\colon Z \to Z+Y$ is injective for every Z. Suppose $f_\star(x) = 0$ for some x. By Lemma 4.25, there is a lift $\tilde{x} \in \Sigma^X N(Z) = N(X+Z)$ of x which is invariant with respect to $\boldsymbol{V}_X(Z)$. Since $f_\star(x) = 0$ and $f_\star(\tilde{x}) \in \Sigma^X N(Z+Y) = N(X+Z+Y)$ is a lift of $f_\star(x)$, Lemma 4.25 tells us that

$$\sum_{\sigma \in \boldsymbol{V}_X(Y+Z)} \sigma f_\star(\tilde{x}) = 0.$$

But $\boldsymbol{V}_X(Y+Z) = \boldsymbol{U}_X(Y) \times \boldsymbol{V}_X(Z)$ and $\tilde{x}$ is invariant with respect to $\boldsymbol{V}_X(Z)$, and so we conclude that

$$\sum_{\sigma \in \boldsymbol{U}_X(Y)} \sigma f_\star(\tilde{x}) = 0.$$

Let W be the VB module given by $\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VB}}(X' \oplus Z, -)]$ where X' is a one-dimensional space. Fix an isomorphism $\alpha\colon X' + Z \to X + Z$. Then $[\alpha]$ is a generator of the VI-module $\mathcal{I}(W)$. There is a unique map $\psi\colon \mathcal{I}(W) \to N$ which takes $[\alpha]$ to $\tilde{x}$. Let N' be the VI-submodule of $\mathcal{I}(W)$ generated by $\sum_{\sigma \in \boldsymbol{U}_X(Y)} \sigma f_\star([\alpha])$. Then the equation at the end of the last paragraph shows that ψ factors through the projection $\mathcal{I}(W) \to \mathcal{I}(W)/N'$. We claim that $\psi = 0$. Since M is torsion-free and ψ factors through $\mathcal{I}(W)/N'$, it suffices to show that $\mathcal{I}(W)/N'$ is a torsion module. This has already been established in the proof of Lemma 4.26. So $\psi = 0$. This implies that $x = 0$, completing the proof. □

Proposition 4.30. *The functor $\overline{\Sigma}$ preserves Γ-acyclic objects.*

Proof. Let M be a Γ-acyclic object. By Proposition 2.7 and Lemma 2.5, the VI-module $M/\Gamma(M)$ is derived saturated. By Lemma 4.28, for each $k \geq 0$ and vector spaces $X_1, \dots, X_k$, the VI-module $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}(M/\Gamma(M))$ is torsion-free. By the previous lemma, for each $k \geq 0$ and vector spaces $X_1, \dots, X_k$, the VI-module $\Delta^{X_1}\Delta^{X_2}\cdots\Delta^{X_k}\overline{\Sigma}(M/\Gamma(M))$ is torsion-free. By Lemma 4.28 again, $\overline{\Sigma}(M/\Gamma(M))$ is derived saturated. Since Γ commutes with $\overline{\Sigma}$ (Proposition 4.27), we see that $\overline{\Sigma}M/\Gamma(\overline{\Sigma}M)$ is derived saturated. By Lemma 2.5, $\overline{\Sigma}M$ is Γ-acyclic, completing the proof. □

The following question is quite natural:

Question 4.31. Do either Σ or $\overline{\Sigma}$ preserve injective objects? Note, a positive answer is known in the $q = 1$ (FI-modules) case; see [Gan 2017].

Lemma 4.32. *If M is derived saturated, then so are $\overline{\Sigma}M$ and $\overline{\Delta}M$.*

Proof. Since $\overline{\Sigma}$ commutes with Γ (Proposition 4.27) and preserves Γ-acyclic objects (Proposition 4.30), we have $\mathrm{R}\Gamma\overline{\Sigma}M = \overline{\Sigma}\mathrm{R}\Gamma M = 0$. Thus by Proposition 2.7, we see that $\overline{\Sigma}M$ is derived saturated. The result about $\overline{\Delta}M$ follows from the exact sequence (see Proposition 4.8)

$$0 \to M \to \overline{\Sigma}M \to \overline{\Delta}M \to 0. \qquad \square$$

Lemma 4.33 (nonvanishing coinvariants). *Suppose $K \leq H \leq G$ are finite groups. Let W be a $\boldsymbol{k}[H]$-module such that K acts trivially on W. Then for any $\boldsymbol{k}[G]$-submodule V of $\mathrm{Ind}_H^G W$, we have $V_K = 0 \iff V = 0$.*

Proof. Let $\mathcal{T} = \{\tau_1, \dots, \tau_n\}$ be a full set of representatives in G of the left coset space G/H. We assume that $\tau_1 = 1_G$. Any element $x \in \operatorname{Ind}_H^G W$ can be thought of as a function $x \colon \mathcal{T} \to W$, and the action of $\sigma \in G$ on x is given by $(\sigma x)(\tau_{n_i}) = h_i x(\tau_i)$ where $h_i \in H$ and n_i are uniquely determined by the equation $\sigma\tau_i = \tau_{n_i} h_i$. As a special case, we note that if $\sigma \in K$, then we have $\sigma = \sigma\tau_1 = \tau_1 h_1 = h_1 \in K$. Since K acts trivially on W, we conclude that $(\sigma x)(\tau_1) = \sigma(x(\tau_1)) = x(\tau_1)$.

Assume now that V is nontrivial. Let $x \in V$ be a nonzero element. As in the previous paragraph, we think of x as a function from $\mathcal{T}$ to W. Since G acts transitively on G/H, there exists a $\sigma \in G$ such that σx is nonzero on τ_1. Now suppose, if possible, the image of σx in V_K is 0. Then σx can be written as

$$\sigma x = \sum_j (x_j - \sigma_j x_j)$$

where x_j are in $\operatorname{Ind}_H^G W$, and σ_j are in K. By the previous paragraph, $(x_j - \sigma_j x_j)(\tau_1) = 0$ for each j. It follows that $(\sigma x)(\tau_1) = 0$, which is a contradiction. This completes the proof. □

Lemma 4.34. *Let M be a derived saturated submodule of a semiinduced module P. Then $t_0(M) \le t_0(P)$.*

Proof. We proceed by induction on $d = t_0(P)$. Denote the induced module $P/P_{\prec d}$ by I and its submodule $(M + P_{\prec d})/P_{\prec d}$ by N. Suppose first that N is an induced submodule of I. In this case, we have $t_0(N) \le t_0(I) = d$. Using the exact sequence

$$0 \to M \cap P_{\prec d} \to M \to N \to 0$$

we see that $M \cap P_{\prec d}$ is a derived saturated submodule of $P_{\prec d}$. By induction, we have $t_0(M \cap P_{\prec d}) \le d - 1$, and it follows that $t_0(M) \le d = t_0(P)$. Thus we can assume that N is not an induced module. In this case, there exists an $r > d$ such that $\mathrm{H}_0^{\mathrm{VI}}(N)_r$ is nonzero. Pick the least such r. We claim that $\mathrm{H}_0^{\mathrm{VI}}(N)(\mathbb{F}^r)$ is a $\boldsymbol{k}[\mathrm{GL}(\mathbb{F}^r)]$-submodule of $\mathcal{I}(\mathrm{H}_0^{\mathrm{VI}}(I/N)_d)(\mathbb{F}^r)$. To see this, let N' be the submodule of $\mathcal{I}(W)$ generated by N_d. By Proposition 3.7, we have $N' = \mathcal{I}(N_d)$. By minimality of r, we have $N' = N_{\prec r} \subset N$. The claim now follows from the following:

$$\begin{aligned} \mathrm{H}_0^{\mathrm{VI}}(N)(\mathbb{F}^r) &= (N/N_{\prec r})(\mathbb{F}^r) \\ &= (N/\mathcal{I}(N_d))(\mathbb{F}^r) \\ &\subset (I/\mathcal{I}(N_d))(\mathbb{F}^r) \\ &= \mathcal{I}((I/N)_d)(\mathbb{F}^r) \qquad \text{(By Proposition 3.6)} \\ &= \mathcal{I}(\mathrm{H}_0^{\mathrm{VI}}(I/N)_d)(\mathbb{F}^r). \end{aligned}$$

Let $A + B + X$ be a decomposition of $\mathbb{F}^r$ such that $\dim A = d$ and $\dim X = 1$. Set $W = \mathrm{H}_0^{\mathrm{VI}}(I/N)(A)$. Clearly, W is a $\boldsymbol{k}[\mathrm{GL}(A)]$-module. Let H be the subgroup of $\mathrm{GL}(\mathbb{F}^r)$ that stabilizes A. There is a natural surjection $\phi \colon H \to \mathrm{GL}(A)$. We let H act on W via this surjection. Since $\boldsymbol{U}_X(A+B)$ lies in the kernel of ϕ, we see that $\boldsymbol{U}_X(A+B)$ acts trivially on W. We also have

$$\mathcal{I}(\mathrm{H}_0^{\mathrm{VI}}(I/N)_d)(\mathbb{F}^r) = \operatorname{Ind}_H^{\mathrm{GL}(\mathbb{F}^r)} W.$$

By the previous lemma, we conclude that $(\overline{\Sigma}\mathrm{H}_0^{\mathrm{VI}}(N))_{r-1}$ is nonzero. Since $\mathrm{H}_0^{\mathrm{VI}}$ is right exact, it follows that $(\overline{\Sigma}\mathrm{H}_0^{\mathrm{VI}}(M))_{r-1}$ is nonzero. By Proposition 4.15, we see that $t_0(\overline{\Delta}M) \geq r-1 > d-1$. But by Lemma 4.32, $\overline{\Delta}M$ is a derived saturated submodule of $\overline{\Delta}P$, which contradicts the inductive hypothesis. This contradiction completes the proof. □

The following argument is motivated by [Nagpal and Snowden 2018, Proposition 2.9].

Proposition 4.35. *Let M be a module generated in finite degrees. If M is derived saturated then it admits a resolution $F_\bullet \to M$ of length at most $t_0(M)+1$ where each F_i is an induced module generated in finite degrees.*

Proof. Let $d = t_0(M)$, and let r be the least number such that $\mathrm{H}_0^{\mathrm{VI}}(M)$ is nontrivial in degree r. We prove by induction on $d-r$ that there is a resolution $F_\bullet \to M$ of length at most $d-r+1$. Let $F_0 = \bigoplus_{0\leq k\leq d} \mathcal{I}(V_k)$ where $V_k = M_k$. We note that $\mathrm{H}_0^{\mathrm{VI}}(M)_r = V_r = \mathrm{H}_0^{\mathrm{VI}}(F_0)_r$ and $\mathrm{H}_0^{\mathrm{VI}}(M)_k = 0 = \mathrm{H}_0^{\mathrm{VI}}(F_0)_k$ for $k<r$. By construction, $t_0(F_0) \leq d$ and there is a surjection $\psi\colon F_0 \to M$. Clearly, we have $\mathrm{H}_0^{\mathrm{VI}}(\ker(\psi))_k = 0$ for $k \leq r$. Since both M and F_0 are derived saturated, we see that $\ker(\psi)$ is derived saturated as well. By the previous lemma, $t_0(\ker(\psi)) \leq d$. Thus by induction on $d-r$, $\ker(\psi)$ admits a resolution of the desired format. We can append F_0 to this resolution to get a resolution of M, completing the proof. □

Theorem 4.36. *Assume that we are in the nondescribing characteristic. Let M be a module generated in finite degrees. Then M is derived saturated if and only if it is semiinduced.*

Proof. Corollary 4.22 shows that semiinduced modules are derived saturated. The other implication follows from the previous proposition and Corollary 4.23. □

An FI-module analog of the result above has been proven in [Djament 2016, Theorem A.9].

The shift theorem. Here we assume that $\boldsymbol{k}$ is a noetherian ring.

Theorem 4.37 [Putman and Sam 2017; Sam and Snowden 2017a]. *The category of* VI*-modules over a noetherian ring is locally noetherian. In particular, if M is a finitely generated* VI*-module over $\boldsymbol{k}$ then $\Gamma(M)$ is supported in finitely many degrees.*

We now state and prove our main theorem (an FI-module analog has been proven by the author in [Nagpal 2015, Theorem A]).

Theorem 4.38 (the shift theorem). *Assume that we are in the nondescribing characteristic, and let M be a finitely generated* VI*-module. Then the following hold:*

(a) *$\overline{\Sigma}^n M$ and $\Sigma^n M$ are semiinduced for large enough n.*

(b) *There exists a finite length complex $I^\bullet$ supported in nonnegative degrees with the following properties:*

- *$I^0 = M$.*
- *I^i is semiinduced for $i>0$.*
- *$I^n = 0$ for $n > t_0(M)+1$.*
- *$\mathrm{H}^i(I^\bullet)$ is supported in finitely many degrees for each i.*

We need a lemma.

Lemma 4.39. *Let Y be fixed vector space, and N be a torsion* VI*-module. For a vector space X, let ℓ_X denote the map from* 0 *to X. If $\overline{\Sigma}^Y\overline{\Sigma}^{\ell_X}\colon \overline{\Sigma}^Y N \to \overline{\Sigma}^Y\overline{\Sigma}^X N$ is an injection for all X then $\overline{\Sigma}^Y N = 0$.*

Proof. Suppose, if possible, $\overline{\Sigma}^Y N(Z)$ is nontrivial for some vector space Z, and pick a nonzero element $x \in \overline{\Sigma}^Y N(Z)$. Let $\tilde{x}$ be a lift of x in $N(Y+Z)$. Since N is torsion, there is a vector space X such that for every linear injection $f\colon Y+Z \to X+Y+Z$ the induced map $f_\star\colon N(Y+Z) \to N(Y+X+Z)$ takes $\tilde{x}$ to zero. But this shows that $\overline{\Sigma}^Y\overline{\Sigma}^{\ell_X}$ takes x to zero, contradicting the injectivity hypothesis. This completes the proof. □

Proof of Theorem 4.38. We first prove that $\overline{\Sigma}^n M$ is semiinduced for large enough n. We do this by induction on $t_0(M)$. By Theorem 4.37, $h_0(M) < \infty$. Let X be a nontrivial vector space. Then the cokernel $\overline{\Delta}^X M$ of $M \to \overline{\Sigma}^X M$ satisfies $t_0(\overline{\Delta}^X M) < t_0(M)$ (Corollary 4.19). Moreover, the kernel $K = \kappa^X(M)$ of $M \to \overline{\Sigma}^X M$ is a torsion-module supported in degrees $\le h_0(M)$ (Lemma 4.7).

We claim that $\overline{\Sigma}^Y\overline{\Delta}^X M$ is semiinduced for large enough Y which is independent of $\dim X$. To see this, suppose that X is of dimension g. Since $t_0(\overline{\Delta}^X M) < t_0(M)$, the induction hypothesis implies that there exists a number k_g such that $\overline{\Sigma}^Y\overline{\Delta}^X M$ is semiinduced whenever the dimension of Y is larger than k_g. Pick a t larger than $h_0(M)$ and k_1, and assume that the dimension of Y is at least t. Then $\overline{\Sigma}^Y K = \overline{\Sigma}^Y\kappa^X(M) = 0$, and so Lemma 4.18 yields the following exact sequence

$$0 \to \overline{\Delta}^Y M \to \overline{\Delta}^{X+Y} M \to \overline{\Sigma}^Y\overline{\Delta}^X M \to 0.$$

Now suppose X is of dimension 1. Then the last term in this exact sequence is semiinduced as $t > k_1$. We conclude that $\overline{\Sigma}^{Y'}\overline{\Delta}^Y M$ is semiinduced if and only if $\overline{\Sigma}^{Y'}\overline{\Delta}^{Y+X} M$ is semiinduced (Corollary 4.23). In other words, we may assume $k_{t+1} = k_t$ for any $t > \max(h_0(M), k_1)$. Thus if Y is of dimension larger than $h_0(M)$ and k_i for $1 \le i \le \max(h_0(M), k_1) + 1$, then $\overline{\Sigma}^Y\overline{\Delta}^X M$ is semiinduced for all X. This proves the claim.

Let Y be large enough such that $\overline{\Sigma}^Y\overline{\Delta}^X M$ is semiinduced for all X, and assume that the dimension of Y is larger than $h_0(M)$. Then $\overline{\Sigma}^Y K = \overline{\Sigma}^Y\kappa^X(M) = 0$, and so we have an exact sequence

$$0 \to \overline{\Sigma}^Y M \to \overline{\Sigma}^Y\overline{\Sigma}^X M \to \overline{\Sigma}^Y\overline{\Delta}^X M \to 0.$$

By Corollary 4.22 and Proposition 2.7, we see that $\mathrm{R}\Gamma(\overline{\Sigma}^Y\overline{\Delta}^X M) = 0$. Thus by the exact sequence above, we conclude that $\mathrm{R}^i\Gamma(\overline{\Sigma}^Y M) \cong \mathrm{R}^i\Gamma(\overline{\Sigma}^Y\overline{\Sigma}^X M)$ where the isomorphism is given by $\mathrm{R}^i\Gamma(\overline{\Sigma}^Y\overline{\Sigma}^\ell)$ where $\ell\colon 0 \to X$ is the unique map. By Proposition 4.27, we see that

$$\mathrm{R}^i\Gamma(\overline{\Sigma}^Y\overline{\Sigma}^\ell -) = \overline{\Sigma}^Y\overline{\Sigma}^\ell\mathrm{R}^i\Gamma(-).$$

This shows that the map

$$\overline{\Sigma}^Y\overline{\Sigma}^\ell\colon \overline{\Sigma}^Y\mathrm{R}^i\Gamma(M) \to \overline{\Sigma}^Y\overline{\Sigma}^X\mathrm{R}^i\Gamma(M)$$

is an isomorphism for each X. The previous lemma implies that $\overline{\Sigma}^Y \mathrm{R}^i\Gamma(M) = 0$. Thus $\mathrm{R}^i\Gamma(\overline{\Sigma}^Y M) = 0$ for all i (Proposition 4.27). By Proposition 2.7 and Theorem 4.36, $\overline{\Sigma}^Y M$ is semiinduced. Thus $\overline{\Sigma}^n M$ is semiinduced for large n (see Proposition 4.11).

To prove that $\Sigma^n M$ is semiinduced for large enough n we need part (b), which we now prove by induction on $t_0(M)$. Let Y be a vector space such that $\overline{\Sigma}^Y M$ is semiinduced, and $\ell\colon 0 \to Y$ be the unique map. Set $I^0 = M$, $I^1 = \overline{\Sigma}^Y M$ where the map $I^0 \to I^1$ is $\overline{\Sigma}^\ell$. The cokernel of this map is $\overline{\Delta}^\ell M$. We have $t_0(\overline{\Delta}^\ell M) < t_0(M)$ (Corollary 4.19). By induction, there is a complex $J^\bullet$ of length at most $t_0(M)$ with $J^0 = \overline{\Delta}^\ell M$, J^i semiinduces for $i > 0$, and $\mathrm{H}^i(J^\bullet)$ finitely supported for each i. Now set $I^i = J^{i-1}$ for $i \geq 2$, and note that we can naturally append these to $I^0 \to I^1$ to get a complex $I^\bullet$. Clearly, this $I^\bullet$ has all the required properties. This proves part (b).

Finally, we show that $\Sigma^n M$ is semiinduced for large enough n. For this let $I^\bullet$ be the complex as in part (b). Let n be large enough such that $\deg \mathrm{H}^i(I^\bullet) < n$ for all i. By construction, $\Sigma^n I^\bullet$ is exact and $\Sigma^n I^i$ are semiinduced for $i > 0$ (shift of a semiinduced module is semiinduced; Corollary 4.4). By Corollary 4.23, $\Sigma^n I^0 = \Sigma^n M$ is semiinduced. This completes the proof. □

Remark 4.40. The proof of part (b) above shows that if $M \to N$ is a map of finitely generated VI-modules then we can find complexes $I^\bullet$ and $J^\bullet$ for M and N respectively (with all the properties as mentioned in part (b)) and a natural map $I^\bullet \to J^\bullet$ extending the map $M \to N$.

Remark 4.41. It is easy to see that the shift theorem together with Corollary 4.23 imply that $\mathrm{Mod}_{\mathrm{VI}}$ is locally noetherian. Since we have only used Corollary 1.5 in our proof, it follows that Theorem 1.4 is equivalent to its corollary.

5. Some consequences of the shift theorem

Unless otherwise mentioned, we assume that we are in the nondescribing characteristic, and that $\boldsymbol{k}$ is noetherian.

Stable degree and the q-polynomiality of dimension. We define the *stable degree* of a VI-module M, denoted $\delta(M)$, by

$$\delta(M) := \inf_{n\geq 0} t_0(\overline{\Sigma}^n M).$$

This is an invariant associated to VI-module with several useful properties that we prove below. An invariant with the same name, but for FI-modules, is discussed in [Church et al. 2018, Section 2].

Proposition 5.1. *Let M be a finitely generated module. We have the following*:

(a) *If M is semiinduced, then $\delta(M) = t_0(M)$.*

(b) *$\delta(M)$ is the common value of $t_0(\overline{\Sigma}^n M)$ for $n \gg 0$.*

(c) *$\delta(M)$ is the common value of $t_0(\Sigma^n M)$ for $n \gg 0$.*

(d) *$\delta(M) = \delta(\Sigma^n M) = \delta(\overline{\Sigma}^n M)$ for any $n \geq 0$.*

(e) $\delta(M) \le t_0(M) < \infty$.

(f) *If* $0 \to L \to M \to N \to 0$ *is a short exact sequence*, $\delta(M) = \max(\delta(L), \delta(N))$.

(g) *If* K *is a subquotient of* M, $\delta(K) \le \delta(M)$.

(h) *If* T *is a torsion submodule of* M, *then* $\delta(M/T) = \delta(M)$.

(i) *The cokernel* $\bar{\Delta}^X M$ *of the natural map* $M \to \bar{\Sigma}^X M$ *satisfies* $\delta(\bar{\Delta}^X M) < \max(\delta(M), 0)$.

Proof. (a) First suppose that $M = \mathcal{I}(V)$ is induced. From the equalities $\bar{\Sigma}\mathcal{I}(V) = \mathcal{I}(V) \oplus \mathcal{I}(\bar{\Sigma} V)$ (Proposition 4.12) and $t_0(\mathcal{I}(V)) = \deg V$, we see that $\delta(M) = t_0(\bar{\Sigma}^n M) = t_0(M)$. Since induced modules are acyclic with respect to $\mathrm{H}_0^{\mathrm{VI}}$ (Proposition 3.10) and $\bar{\Sigma}$ is exact, we conclude that the result holds for semiinduced modules as well.

(b)–(e) Since $t_0(\bar{\Sigma}^n M)$ is a decreasing function of n (Proposition 4.9), we see that $\delta(M) = \delta(\bar{\Sigma}^n M)$ for any n. By the shift theorem (Theorem 4.38) and part (a), we conclude that $\delta(M)$ is the common value of $t_0(\bar{\Sigma}^n M)$ for $n \gg 0$. Let a be large such that $\bar{\Sigma}^a M$ is semiinduced and n be large such that $\Sigma^n M$ is semiinduced (use the shift theorem again). Then we have an injection $\Sigma^n M \to \Sigma^n \bar{\Sigma}^a M$. By Corollary 4.23, Proposition 3.10 and part (a), we see that $t_0(\Sigma^n M) \le t_0(\Sigma^n \bar{\Sigma}^a M) = \delta(\bar{\Sigma}^a M) = \delta(M)$. Conversely, since we also have $t_0(\Sigma^n M) \ge t_0(\bar{\Sigma}^n M)$, we see that part (c) holds. Part (d) follows from (b) and (c) once we note that $t_0(\Sigma^n M)$ and $t_0(\bar{\Sigma}^n M)$ are decreasing functions of n (Proposition 4.9). Part (e) is trivial from this discussion.

(f)–(h) Choose n large enough that $\Sigma^n L$, $\Sigma^n M$, and $\Sigma^n N$ are semiinduced. Since semiinduced modules are homology-acyclic, we have a short exact sequence

$$0 \to \mathrm{H}_0^{\mathrm{VI}}(\Sigma^n L) \to \mathrm{H}_0^{\mathrm{VI}}(\Sigma^n M) \to \mathrm{H}_0^{\mathrm{VI}}(\Sigma^n N) \to 0.$$

Thus, $t_0(\Sigma^n M) = \max(t_0(\Sigma^n L), t_0(\Sigma^n L))$, which implies the claim in light of part (c). Part (g) is a consequence of part (f). For part (h), note that T is supported in finitely many degrees (Theorem 4.37). By part (d), $\delta(T) = 0$. Part (h) now follows from Part (f).

(i) First suppose that M is semiinduced. Then by Corollary 4.23, $\bar{\Delta}^X M$ is semiinduced. By Corollary 4.19, we see that $t_0(\bar{\Delta}^X M) < t_0(M)$. By part (a), we conclude that $\delta(\bar{\Delta}^X M) < \delta(M)$. Thus the result holds for semiinduced modules. Now suppose that M is not semiinduced. Let Y be large so that $\bar{\Sigma}^Y M$ is semiinduced. We have an exact sequence

$$0 \to M/\kappa^Y(M) \to \bar{\Sigma}^Y M \to \bar{\Delta}^Y M \to 0.$$

Applying $\bar{\Delta}^X$, we obtain the following exact sequence:

$$(\mathrm{L}_1 \bar{\Delta}^X)(\bar{\Delta}^Y M) \to \bar{\Delta}^X(M/\kappa^Y(M)) \to \bar{\Delta}^X \bar{\Sigma}^Y M \to \bar{\Delta}^X \bar{\Delta}^Y M \to 0.$$

The first term of this sequence is torsion (Proposition 4.17). Thus by parts (g) and (h), we see that

$$\delta(\bar{\Delta}^X(M/\kappa^Y(M))) \le \delta(\bar{\Delta}^X \bar{\Sigma}^Y M) < \delta(\bar{\Sigma}^Y M) = \delta(M).$$

Now consider the exact sequence

$$\bar{\Delta}^X \kappa^Y(M) \to \bar{\Delta}^X M \to \bar{\Delta}^X(M/\kappa^Y(M)) \to 0.$$

Since the first term is torsion, we conclude that $\delta(\overline{\Delta}^X M) = \delta(\overline{\Delta}^X(M/\kappa^Y(M))) < \delta(M)$. This completes the proof. $\square$

Corollary 5.2. *Let $I^\bullet$ be the complex as in Theorem 4.38. Then we may assume that $t_0(I^1) = \delta(M)$, and $t_0(I^i) \le \delta(M) - i \le t_0(M) - i$ for $i > 1$.*

Proof. This follows from the construction of $I^\bullet$ and the properties of the stable degree. $\square$

Lemma 5.3. *Assume that $\boldsymbol{k}$ is a field. Let $\mathcal{I}(V)$ be a module induced from d. Then*

$$\dim_{\boldsymbol{k}} \mathcal{I}(V)(\mathbb{F}^n) = \frac{(q^n-1)(q^n-q)\cdots(q^n-q^{d-1})}{(q^d-1)(q^d-q)\cdots(q^d-q^{d-1})} \dim_{\boldsymbol{k}} V(\mathbb{F}^d)$$

for every $n \ge 0$. In particular, there is a polynomial $P \in \boldsymbol{Q}[X]$ such that $\dim_{\boldsymbol{k}} \mathcal{I}(V)(\mathbb{F}^n) = P(q^n)$ for every $n \ge 0$.

Proof. This easily follows from the equality $\mathcal{I}(V)(\mathbb{F}^n) = \boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(\mathbb{F}^d, \mathbb{F}^n)] \otimes_{\boldsymbol{k}[\mathrm{GL}_d]} V(\mathbb{F}^d)$. $\square$

Theorem 5.4 (q-polynomiality of dimension). *Assume that $\boldsymbol{k}$ is a field. Let M be a finitely generated* VI*-module. Then there exists a polynomial P of degree $\delta(M)$ such that $\dim_{\boldsymbol{k}} M(\mathbb{F}^n) = P(q^n)$ for $n \gg 0$.*

Proof. Let a be large enough such that $N := \Sigma^a M$ is semiinduced. By Proposition 5.1, we have $t_0(N) = \delta(M)$. Set $d = \delta(M)$. By Corollary 3.11, $N_{\preceq i}/N_{\prec i}$ is induced from i, and $N_{\preceq d}/N_{\prec d}$ is nonzero. By the previous lemma, there exists a polynomial P such that $\dim_{\boldsymbol{k}} N(\mathbb{F}^n) = P(q^n)$ for every $n \ge 0$. This shows that $\dim_{\boldsymbol{k}} M(\mathbb{F}^n) = P(q^{n-a})$ for $n \ge a$, completing the proof. $\square$

Remark 5.5. The least a such that $\Sigma^a M$ is semiinduced is exactly equal to $h^{\max}(M) + 1$ where

$$h^{\max}(M) = \max_{i \ge 0} h^i(M)$$

is the maximum of all local cohomology degrees. This follows easily from Theorem 4.36, and the fact that Γ commutes with Σ. We shall prove in the next section that $h^i(M) = 0$ for $i > \delta(M) + 1$. Thus in the proof above, we have $\dim_{\boldsymbol{k}} M(\mathbb{F}^n) = P(q^n)$ for $n > \max_{0 \le i \le \delta(M)+1} h^i(M) = h^{\max}(M)$.

Finiteness of local cohomology and regularity. Let D be the full triangulated subcategory of the bounded derived category $\mathrm{D}^{\mathrm{b}}(\mathrm{Mod}_{\mathrm{VI}})$ consisting of those objects that are represented by finite complexes with finitely generated cohomologies.

Proposition 5.6. *Let M be an object of* D*. Then*:

(a) $\mathrm{R}\Gamma(M)$ *is in* D *and can be represented by a finite complex of finitely generated torsion modules.*

(b) $\mathrm{R}\boldsymbol{S}(M)$ *is in* D *and can be represented by a finite complex of finitely generated induced modules.*

(c) $\mathrm{R}^i\Gamma(M)$ *is finitely generated for each i and vanishes if $i \gg 0$.*

(d) *There is an exact triangle*

$$\mathrm{R}\Gamma(M) \to M \to \mathrm{R}\boldsymbol{S}(M) \to .$$

Proof. By the shift theorem (Theorem 4.38) and Remark 4.40, we have an exact triangle of the form

$$T \to M \to F \to$$

in D such that T is represented by a finite complex of finitely generated torsion modules and F is represented by a finite complex of finitely generated semiinduced modules (see [Nagpal et al. 2018, Lemma 2.3] for more details). By Proposition 4.35, F is quasiisomorphic to a finite complex of finitely generated induced modules. By Corollary 2.6 and Proposition 4.21, we have $\mathrm{R}\Gamma(T) \cong T$ and $\mathrm{R}\Gamma(F) = 0$. Thus by applying $\mathrm{R}\Gamma$ to the triangle above yields $T \cong \mathrm{R}\Gamma(M)$. By Corollary 2.6, Theorem 4.36, we see that $\mathrm{R}\boldsymbol{S}(T) = 0$ and $\mathrm{R}\boldsymbol{S}(F) \cong F$. Thus by applying $\mathrm{R}\boldsymbol{S}$ to the triangle above yields $\mathrm{R}\boldsymbol{S}(M) \cong F$. The proof is now complete by Proposition 2.7. □

The FI-module analog of the theorem below has been studied in [Sam and Snowden 2016].

Theorem 5.7 (finiteness of local cohomology). *Let M be a finitely generated* VI*-module. Then* $\mathrm{R}\Gamma(M)$ *and* $\mathrm{R}\boldsymbol{S}(M)$ *are objects of* D *and are supported in nonnegative degrees. Moreover, we have the following*

(a) $\mathrm{R}^i\Gamma(M) = 0$ *if* $i > \delta(M) + 1$.

(b) $\mathrm{R}^i\boldsymbol{S}(M) = 0$ *if* $i > \delta(M)$.

(c) *We have an exact sequence* $0 \to \Gamma(M) \to M \to \boldsymbol{S}(M) \to \mathrm{R}^1\Gamma(M) \to 0$.

(d) $\mathrm{R}^{i+1}\Gamma(M) \cong \mathrm{R}^i\boldsymbol{S}(M)$ *for* $i \geq 1$.

Proof. Let $I = I^\bullet$ be the complex as in the shift theorem (Theorem 4.38). Then I is supported in nonnegative degrees and $I^i = 0$ if $i > \delta(M) + 1$ (see Proposition 5.1 part (i) and the construction of $I^\bullet$). We may take T, as in the proof of Proposition 5.6, to be equal (or quasiisomorphic; see [Nagpal et al. 2018, Lemma 2.3]) to I. This shows that part (a) holds. The rest is immediate from Proposition 2.7. □

Corollary 5.8. *Let $I^\bullet$ be the complex as in Theorem 4.38. Then* $\mathrm{R}^i\Gamma(M) = \mathrm{H}^i(I^\bullet)$.

Lemma 5.9. *There is a resolution of the* VI*-module* $\boldsymbol{k} = \boldsymbol{A}/\boldsymbol{A}_+$ *of the form* $\mathcal{I}(\mathrm{St}_\bullet) \to \boldsymbol{k} \to 0$, *where* St_d *denote the Steinberg representation of* GL_d.

Proof. We refer the reader to [Charney 1984, page 7] where an argument for split Steinberg representation is given. The argument for the Steinberg representation is similar. □

Lemma 5.10. *Let M be a finitely generated torsion module, and suppose* $\deg M = d$. *Then* $t_i(M) - i \leq d$ *for all* $i \geq 0$.

Proof. Since induced modules are homology-acyclic (Proposition 3.10), the previous lemma implies that $\mathrm{H}_i^{\mathrm{VI}}(M) = \mathrm{Tor}_i^{\boldsymbol{A}}(\boldsymbol{k}, M) = \mathrm{H}_i(\mathcal{I}(\mathrm{St}_\bullet) \otimes_{\boldsymbol{A}} M)$. Clearly, $\mathcal{I}(\mathrm{St}_i) \otimes_{\boldsymbol{A}} M = \mathrm{St}_i \otimes_{\mathrm{VB}} M$ is supported in degrees $\leq d + i$. The result follows. □

For a finitely generated VI-module M, let $r(M) = \max_{0 \leq i \leq \delta(M)+1}(h^i(M) + i)$. The following argument is based on [Nagpal et al. 2018, Corollary 2.5].

Theorem 5.11 (finiteness of regularity). *Let M be a finitely generated* VI*-module. Then $t_i(M)-i \le r(M)$ for all $i>0$. In particular, M has finite Castelnuovo–Mumford regularity.*

Proof. By Theorem 5.7 and the previous lemma, we see that $t_i(\mathrm{R}\Gamma(M))-i \le r(M)$. Since $\mathrm{R}\boldsymbol{S}(M)$ is supported in nonnegative cohomological degrees (which we think of as nonpositive homological degrees), we conclude that $t_i(\mathrm{R}\boldsymbol{S}(M))=0$ for $i>0$ (Proposition 3.10). The exact triangle $\mathrm{R}\Gamma(M)\to M\to \mathrm{R}\boldsymbol{S}(M)\to$ of Proposition 5.6 implies that $t_i(M)\le \max(t_i(\mathrm{R}\Gamma(M)), t_i(\mathrm{R}\boldsymbol{S}(M)))$. Thus for $i>0$, we obtain $t_i(M)-i\le r(M)$. This completes the proof. □

Representation stability in characteristic zero. In this section, we assume that $\boldsymbol{k}$ is an algebraically closed field of characteristic 0. We first recall a parametrization of irreducible representations of GL_n over $\boldsymbol{k}$, we follow [Zelevinsky 1981, Section 9]. Let $\mathcal{C}_n$ be the isomorphism classes of cuspidal representations (irreducible representations which cannot be obtained via a parabolic induction) of GL_n and set $\mathcal{C}=\sqcup_{n\ge 1}\mathcal{C}_n$. If $\rho\in\mathcal{C}_n$, we set $|\rho|=n$. Let $\mathcal{P}$ be the set of partitions. Given a partition λ, we set $|\lambda|=n$ if λ is a partition of n. Given a function $\boldsymbol{\mu}\colon \mathcal{C}\to\mathcal{P}$, we set $|\boldsymbol{\mu}|=\sum_{x\in\mathcal{C}}|x||\boldsymbol{\mu}(x)|$. The isomorphism classes of irreducible representations of GL_n are in bijection with the set of functions $\boldsymbol{\mu}$ satisfying $|\boldsymbol{\mu}|=n$. We fix an irreducible representation $M_{\boldsymbol{\mu}}$ corresponding to each partition function $\boldsymbol{\mu}$.

Let $\iota\in\mathcal{C}_1$ be the trivial representation of GL_1. For a partition function $\boldsymbol{\mu}$ with $\boldsymbol{\mu}(\iota)=\lambda$, we define another partition function $\boldsymbol{\mu}[n]$ by

$$\boldsymbol{\mu}[n](\rho)=\begin{cases}(n-|\boldsymbol{\mu}|,\lambda_1,\lambda_2,\dots) & \text{if } \rho=\iota,\\ \boldsymbol{\mu}(\rho) & \text{if } \rho\ne\iota.\end{cases}$$

This definition makes sense only if $n\ge|\boldsymbol{\mu}|+\lambda_1$.

Let

$$M_0\xrightarrow{\phi_0}M_1\xrightarrow{\phi_1}M_2\xrightarrow{\phi_2}\cdots$$

be a sequence such that each M_n is a $\boldsymbol{k}[\mathrm{GL}_n]$-module and each ϕ_n is GL_n-equivariant. Following [Gan and Watterlond 2018] which, in turn, is based on [Church and Farb 2013], we call such a sequence *representation stable of degree d starting at N* if the following three conditions are satisfied for every $n\ge N$:

(RS1) *Injectivity*: The map $\phi_n\colon M_n\to M_{n+1}$ is injective.

(RS2) *Surjectivity*: The GL_{n+1} orbits of $\phi_n(M_n)$ span all of M_{n+1}.

(RS3) *Multiplicities*: There is a decomposition

$$M_n=\bigoplus_{\boldsymbol{\mu}}M_{\boldsymbol{\mu}[n]}^{\oplus c(\boldsymbol{\mu})}$$

where the multiplicities $0\le c(\boldsymbol{\mu})<\infty$ do not depend on n, and $c(\boldsymbol{\mu})=0$ if $|\boldsymbol{\mu}|>d$.

We now prove and improve [Gan and Watterlond 2018, Theorem 1.6].

Theorem 5.12 (representation stability). *Let M be a finitely generated* VI*-module. Denote $M(\mathbb{F}^n)$ by M_n, and let $\phi_n\colon M_n \to M_{n+1}$ be the map induced by the natural inclusion $\mathbb{F}^n \hookrightarrow \mathbb{F}^{n+1}$. Then the sequence*

$$M_0 \xrightarrow{\phi_0} M_1 \xrightarrow{\phi_1} M_2 \xrightarrow{\phi_2} \cdots$$

is representation stable of degree $\delta(M)$ starting at $N := \max(h^{\max}(M)+1, 2t_0(M))$.

Proof. Since $h^0(M) < N$, we see that (RS1) holds. Similarly, $t_0(M) \le N$ implies that (RS2) holds. Now we prove (RS3). Let $I^\bullet$ be the complex as in Theorem 4.38. Then $I^\bullet(\mathbb{F}^n)$ is exact if $n > h^{\max}(M)$ (Corollary 5.8). Since $I^0 = M$, it suffices to prove (RS3) for I^i for each $i > 0$. We may also assume that $t_0(I^1) = \delta(M)$, and $t_0(I^i) \le \delta(M) - i \le t_0(M) - i$ for $i > 1$ (Corollary 5.2). Thus it suffices to show (RS3) for a semiinduced module generated in degrees $\le \delta(M)$. By Proposition 3.2, every semiinduced module is induced in characteristic zero. Thus we are reduced to showing (RS3) for a finitely generated induced module generated in degrees $\le \delta(M)$. This follows from Pieri's formula (see [Gan and Watterlond 2018, Lemma 2.8]), completing the proof. □

Classification of indecomposable injectives in characteristic zero. We first classify torsion-free injectives in the proposition below. We repeatedly use the fact that in characteristic zero, every induced module is projective (Proposition 3.2), and so every semiinduced module is, in fact, induced.

Proposition 5.13. *Every induced* (*and hence semiinduced*) VI*-module is injective in* $\mathrm{Mod}_{\mathrm{VI}}$. *A torsion-free injective* VI*-module is induced.*

Proof. Let $\mathcal{I}(W)$ be a finitely generated induced module. Note that VI-modules form a locally noetherian category (Theorem 4.37), and so any direct sum of injective modules is injective. Since any induced module is a direct sum of finitely generated induced modules, it suffices to show that $\mathcal{I}(W)$ is injective.

We start by showing that $\mathrm{Ext}^1(Q, \mathcal{I}(W)) = 0$ for any finitely generated module Q. This is equivalent to showing that any short exact sequence of the form

$$0 \to \mathcal{I}(W) \to M \to Q \to 0$$

splits. Thus it suffices to show that any injection $f\colon \mathcal{I}(W) \to M$ admits a section whenever M is finitely generated. Let X be a vector space of large enough dimension so that $\Sigma^X M$ is semiinduced (Theorem 4.38). Let $\ell\colon 0 \to X$ be the unique map. Exactness of Σ^X and the commutativity of the diagram

$$\begin{tikzcd} \Sigma^X \mathcal{I}(W) \arrow[r, "\Sigma^X(f)"] & \Sigma^X M \\ \mathcal{I}(W) \arrow[u, "\Sigma^\ell"] \arrow[r, "f"] & M \arrow[u, "\Sigma^\ell"'] \end{tikzcd}$$

shows that $\Sigma^\ell f\colon \mathcal{I}(W) \to \Sigma^X M$ is injective. By Corollary 4.23, the cokernel of $\Sigma^\ell f$ is semiinduced. By Proposition 3.2 and the characteristic 0 hypothesis, every semiinduced module is projective. Hence $\Sigma^\ell f$ admits a section s. Then $s\Sigma^\ell$ is a section of f, as required.

Next, let $M \subset N$ be arbitrary VI-modules, and $\Phi\colon M \to \mathcal{I}(W)$ be an arbitrary map. We will show that Φ extends to N which finishes the proof of the first assertion. We follow the proof of Baer's criterion as in [Stacks 2005–, Tag 0AVF]. By Zorn's lemma, it suffices to show that if $M \subsetneq N$ then Φ extends to a submodule $M' \subset N$ which properly contains M. For this, pick an $x \in N \setminus M$, and let M' be the submodule of N generated by M and x. Then $x \in N(\mathbb{F}^d)$ for some d. Note that

$$Q := \{f \in \mathcal{I}(d)\colon fx \in M\}$$

is a VI-submodule of $\mathcal{I}(d)$. By the previous paragraph, we have $\mathrm{Ext}^1(\mathcal{I}(d)/Q, \mathcal{I}(W)) = 0$. Thus the map $\psi\colon Q \to \mathcal{I}(W)$ given by $f \mapsto \Phi(fx)$ extends to a map $\tilde{\psi}\colon \mathcal{I}(d) \to \mathcal{I}(W)$. Now consider the map $\tilde{\Psi}\colon M \oplus \mathcal{I}(d) \to \mathcal{I}(W)$ given by

$$(y, f) \mapsto \Phi(y) - \tilde{\psi}(f).$$

The kernel of this map contains the kernel of the natural map $M \oplus \mathcal{I}(d) \to N$ given by $(y, f) \mapsto y + fx$. Thus $\tilde{\Psi}$ factors through a map $\Psi\colon M' \to \mathcal{I}(W)$. It is easy to check that this map extends Φ. This concludes the proof of the first assertion.

Let I be an arbitrary torsion-free injective module. Then by the shift theorem, I embeds into a direct sum J of induced modules. Since I is injective, the embedding $I \to J$ splits. This shows that the injection $I_{\preceq d} \to J_{\preceq d}$ is split as well, and so $I_{\preceq d}$ is injective and torsion-free. It follows that $\mathrm{R}\Gamma(I_{\preceq d}) = 0$, and so $I_{\preceq d}$ must be derived saturated. Thus $I_{\preceq d}$ is induced (Theorem 4.36). since colimits are exact and $I = \varinjlim_d I_{\preceq d}$, we see that I is a direct sum of induced modules, concluding the proof of the second assertion. □

We now classify torsion injectives. For this we do not need any assumption on $\boldsymbol{k}$ (noetherianity is still needed but the nondescribing characteristic assumption is not needed). So assume that $\boldsymbol{k}$ is an arbitrary noetherian ring. Let $\mathcal{V}$ be a monoidal category. Given two functors $F_1\colon \mathcal{C} \to \mathcal{V}$ and $F_2\colon \mathcal{C}^{\mathrm{op}} \to \mathcal{V}$ there is a natural notion of a tensor product $F_1 \otimes_{\mathcal{C}} F_2$ (we refer the readers to [Palmquist and Newell 1971] for details). More explicitly, if $\mathcal{C} = \mathrm{VI}$ and $\mathcal{V} = (\mathrm{Mod}_{\boldsymbol{k}}, \otimes_{\boldsymbol{k}})$, then $F_1 \otimes_{\mathrm{VI}} F_2$ is given by the following $\boldsymbol{k}$-module

$$\Bigl(\bigoplus_{X \in \mathrm{Obj(VI)}} F_1(X) \otimes_{\boldsymbol{k}} F_2(X)\Bigr)/\langle f_\star(v) \otimes w - v \otimes f^\star(w)\colon f \in \mathrm{Hom}_{\mathrm{VI}}(X, Y),\, v \in F_1(X),\, w \in F_2(Y)\rangle.$$

The following lemma is elementary.

Lemma 5.14. *$\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(-, \mathbb{F}^d)]$ is a projective $\mathrm{VI}^{\mathrm{op}}$-module. Moreover, for any VI-module N, we have*

$$\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(-, \mathbb{F}^d)] \otimes_{\mathrm{VI}} N = N(\mathbb{F}^d).$$

Let E be a $\boldsymbol{k}[\mathrm{GL}_d]$-module. We denote by $\check{\mathcal{I}}(E)$ the VI-module given by

$$\mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(-, \mathbb{F}^d)], E).$$

$\check{\mathcal{I}}(E)$ is clearly a torsion VI-module (note that $\check{\mathcal{I}}(E)(Y) = 0$ for $Y \succ \mathbb{F}^d$).

Proposition 5.15. *For any $\boldsymbol{k}[\mathrm{GL}_d]$-module, we have*

$$\mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(M, \check{\mathfrak{I}}(E)) = \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(M(\mathbb{F}^d), E).$$

In particular, if E is an injective $\boldsymbol{k}[\mathrm{GL}_d]$-module then $\check{\mathfrak{I}}(E)$ is an injective VI*-module.*

Proof. By the tensor-hom adjunction, we have

$$\begin{aligned}\mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(M, \check{\mathfrak{I}}(E)) &= \mathrm{Hom}_{\mathrm{Mod}_{\mathrm{VI}}}(M, \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(-, \mathbb{F}^d)], E)) \\ &= \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(-, \mathbb{F}^d)] \otimes_{\mathrm{VI}} M, E) \\ &= \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(M(\mathbb{F}^d), E)\end{aligned}$$

where the last equality follows from the previous lemma. If E is injective, the functor given by

$$M \mapsto \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(M(\mathbb{F}^d), E)$$

is exact, and hence $\check{\mathfrak{I}}(E)$ is injective. □

For a VI-module M, we denote the maximal submodule supported in degrees $\leq d$ by $M^{\preceq d}$.

Proposition 5.16. *Suppose $I = \check{\mathfrak{I}}(E)$. Then $I(\mathbb{F}^d) \cong E$. Moreover, $I^{\prec d} = 0$ and $I^{\preceq d} = I$.*

Proof. Clearly, $I(\mathbb{F}^d) = \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(\boldsymbol{k}[\mathrm{GL}_d], E) \cong E$. For the second statement, it suffices to show that if Ψ is a nonzero element of $I(X)$, then $g_\star(\Psi)$ is nonzero for any $g \in \mathrm{Hom}_{\mathrm{VI}}(X, Y)$ with $Y \preceq \mathbb{F}^d$. So suppose $\Psi \in I(X) = \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(\boldsymbol{k}[\mathrm{Hom}_{\mathrm{VI}}(X, \mathbb{F}^d)], E)$. If Ψ is nonzero then there exists an $h \in \mathrm{Hom}_{\mathrm{VI}}(X, \mathbb{F}^d)$ such that $\Psi(h) \neq 0$. Let $f \in \mathrm{Hom}_{\mathrm{VI}}(X, \mathbb{F}^d)$ be such that $fg = h$. Now $(g_\star(\Psi))(f) = \Psi(fg) = \Psi(h) \neq 0$. Thus $g_\star(\Psi)$ is nonzero completing the proof. □

A *principal injective* of type d is a VI-module of the form $\check{\mathfrak{I}}(E)$ where E is an injective $\boldsymbol{k}[\mathrm{GL}_d]$-module. By Proposition 5.16, the degree d part of a principal injective of type d is an injective $\boldsymbol{k}[\mathrm{GL}_d]$-module.

Lemma 5.17. *Let M be a* VI*-module. Then $M^{\preceq d}/M^{\prec d}$ injects into a principal injective I of type d. In fact, if E is the injective hull of $M^{\preceq d}(\mathbb{F}^d)$ as a $\boldsymbol{k}[\mathrm{GL}_d]$-module, then we may take $I = \check{\mathfrak{I}}(E)$.*

Proof. Let $N = M^{\preceq d}/M^{\prec d}$. Then N is supported in degree $\leq d$, and by definition of N, $\ell_\star \colon N(X) \to N(\mathbb{F}^d)$ is injective for any X and any $\ell \in \mathrm{Hom}_{\mathrm{VI}}(X, \mathbb{F}^d)$. Thus if $f \colon N \to I$ is a map, then f is injective if and only if $f(\mathbb{F}^d) \colon N(\mathbb{F}^d) \to I(\mathbb{F}^d)$ is injective. Now let $\iota \colon N(\mathbb{F}^d) \to E$ be the injective-hull of $N(\mathbb{F}^d) = M^{\preceq d}(\mathbb{F}^d)$ as a $\boldsymbol{k}[\mathrm{GL}_d]$-module. Then by Proposition 5.15, ι induces a map $\iota_\star \colon N \to \check{\mathfrak{I}}(E)$. By our previous argument, it suffices to show that it is injective in degree d. But in degree d, this map is given by the image of ι under the natural adjunction isomorphism $\mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(N(\mathbb{F}^d), E) \to \mathrm{Hom}_{\boldsymbol{k}[\mathrm{GL}_d]}(N(\mathbb{F}^d), \check{\mathfrak{I}}(E)(\mathbb{F}^d))$ (see Proposition 5.16) and hence is injective. □

Proposition 5.18. *Suppose M is supported in degrees $\leq d$. Let E_k be the injective-hull of $M^{\preceq k}(\mathbb{F}^k)$ as a $\boldsymbol{k}[\mathrm{GL}_k]$-module. Then M embeds into the injective module $\bigoplus_{k \leq d} \check{\mathfrak{I}}(E_k)$.*

Proof. If a module is supported in degree $\leq d$, then it admits a filtration with modules of the form $M^{\preceq k}/M^{\prec k}$ with $k \leq d$. The proposition now follows from Lemma 5.17 and the horseshoe lemma. □

Proposition 5.19. *A direct sum of injectives is injective. If M is any torsion module and E_k is the injective-hull of $M^{\preceq k}(\mathbb{F}^k)$ as a $\boldsymbol{k}[\mathrm{GL}_d]$-module, then M embeds into the injective module $\bigoplus_{k\geq 0} \check{\mathfrak{I}}(E_k)$.*

Proof. It is a standard fact that in a locally noetherian category a direct sum of injectives is an injective. Thus the first statement follows (Theorem 4.37). Now let M be a torsion module. Then $M = \varinjlim_d M^{\preceq d}$ is a filtered colimit of modules supported in finitely many degrees. Since $\mathrm{Mod}_{\mathrm{VI}}$ is a Grothendieck category, filtered colimits are exact. Hence the result follows from Proposition 5.18. □

Proposition 5.20. *A torsion module is injective in $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ if and only if it is isomorphic to a direct sum of principal injectives. In particular, a torsion module is injective in $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$ if and only if it is injective in $\mathrm{Mod}_{\mathrm{VI}}$.*

Proof. By the previous proposition, a direct sum of principal injectives is injective. Let I be a torsion injective. Then by the previous proposition again, I admits an embedding $f\colon I \to J := \bigoplus_{k\geq 0} \check{\mathfrak{I}}(E_k)$ where E_k is the injective-hull of $I^{\preceq k}(\mathbb{F}^k)$ as a $\boldsymbol{k}[\mathrm{GL}_k]$-module. Since I is injective in $\mathrm{Mod}_{\mathrm{VI}}^{\mathrm{tors}}$, f admits a section s. This implies that $I^{\preceq k}/I^{\prec k}$ is a direct summand of $J^{\preceq k}/J^{\prec k} = \check{\mathfrak{I}}(E_k)$. Thus $(I^{\preceq k}/I^{\prec k})(\mathbb{F}^k) = I^{\preceq k}(\mathbb{F}^k)$ is a direct summand of $\check{\mathfrak{I}}(E_k)(\mathbb{F}^k) = E_k$. Since a direct summand of injective module is injective, we see that $I^{\preceq k}(\mathbb{F}^k)$ is injective, and hence is equal to its injective hull E_k. Thus if $K = \mathrm{coker}(f)$, then $(K^{\preceq k}/K^{\prec k})(\mathbb{F}^k) = 0$ for each k. By Nakayama's lemma, $K = 0$. This shows that f is an isomorphism, completing the proof. □

We are now ready to prove our main theorem on classification of indecomposable injectives. Note that the FI-module analog of this result is proved in [Sam and Snowden 2016, Theorem 4.3.4].

Theorem 5.21 (classification of indecomposable injectives). *Assume that $\boldsymbol{k}$ is a field of characteristic zero. Every injective is a direct sum of a torsion-free injective and a torsion injective. Moreover, we have the following:*

(a) *The set of torsion-free indecomposable injectives consists of modules of the form $\mathfrak{I}(E)$ where E (or, more precisely, $E(\mathbb{F}^d)$) is an irreducible $\boldsymbol{k}[\mathrm{GL}_d]$-module for some d.*

(b) *The set of torsion indecomposable injectives consists of modules of the form $\check{\mathfrak{I}}(E)$ where E is an irreducible $\boldsymbol{k}[\mathrm{GL}_d]$-module for some d.*

Proof. In light of Lemma 2.5, every injective is a direct sum of a torsion injective and a torsion-free injective. Part (a) follows from Proposition 5.13, and part (b) follows from Proposition 5.20. □

Finiteness of injective dimension in characteristic zero.

Lemma 5.22. *Let M be a finitely generated torsion module. Then M has finite injective dimension.*

Proof. We prove the assertion by induction on $d = h^0(M)$. We have an exact sequence

$$0 \to M^{\prec d} \to M \to M^{\preceq d}/M^{\prec d} \to 0.$$

Since $h^0(M^{\prec d}) < d$, the induction hypothesis implies that $M^{\prec d}$ has finite injective dimension. By the horseshoe lemma, it suffices to prove that $M^{\preceq d}/M^{\prec d}$ has finite injective dimension. For that, let

$E = M^{\preceq d}/M^{\prec d}(\mathbb{F}^d)$. Since we are in characteristic zero, E is an injective $\boldsymbol{k}[\mathrm{GL}_d]$-module. As in Lemma 5.17, there is an embedding $\iota\colon M^{\preceq d}/M^{\prec d} \to \check{\mathfrak{I}}(E)$ which induces an isomorphism in degree d. This shows that $h^0(\mathrm{coker}(\iota)) < d$. By induction, $\mathrm{coker}(\iota)$ has finite injective dimension. Since $\check{\mathfrak{I}}(E)$ is injective, we conclude that $M^{\preceq d}/M^{\prec d}$ has finite injective dimension, concluding the proof. □

The FI-module analog of the following result is proved in [Sam and Snowden 2016, Theorem 4.3.1].

Theorem 5.23 (finiteness of injective dimension). *Every finitely generated* VI-*module has finite injective dimension.*

Proof. Let M be a finitely generated VI-module. By Proposition 5.6, there exists an exact triangle

$$X \to M \to F \to$$

where X is a finite length complex of finitely generated torsion modules and F is a finite length complex of finitely generated semiinduced modules. In characteristic zero, every semiinduced module is injective. Thus is suffices to show that every finitely generated torsion module has finite injective dimension. But this is the content of the previous lemma. This finishes the proof. □

Acknowledgments

We would like to thank Steven Sam for pointing out connections between the VI-modules and the category of Joyal and Street [1995]. This helped us obtain a clean proof of Proposition 4.12. We thank Nate Harman for useful conversations on virtual specht stability. We thank Nir Gadish, Liping Li, Peter Patzt, Eric Ramos, Steven Sam, Andrew Snowden and an anonymous referee for pointing out several errors and suggesting numerous improvements on the first draft.

References

[Charney 1984] R. Charney, "On the problem of homology stability for congruence subgroups", *Comm. Algebra* **12**:17-18 (1984), 2081–2123. MR Zbl

[Church 2016] T. Church, "Bounding the homology of FI-modules", preprint, 2016. arXiv

[Church and Ellenberg 2017] T. Church and J. S. Ellenberg, "Homology of FI-modules", *Geom. Topol.* **21**:4 (2017), 2373–2418. MR Zbl

[Church and Farb 2013] T. Church and B. Farb, "Representation theory and homological stability", *Adv. Math.* **245** (2013), 250–314. MR Zbl

[Church et al. 2014] T. Church, J. S. Ellenberg, B. Farb, and R. Nagpal, "FI-modules over Noetherian rings", *Geom. Topol.* **18**:5 (2014), 2951–2984. MR Zbl

[Church et al. 2015] T. Church, J. S. Ellenberg, and B. Farb, "FI-modules and stability for representations of symmetric groups", *Duke Math. J.* **164**:9 (2015), 1833–1910. MR Zbl

[Church et al. 2018] T. Church, J. Miller, R. Nagpal, and J. Reinhold, "Linear and quadratic ranges in representation stability", *Adv. Math.* **333** (2018), 1–40. MR Zbl

[Djament 2016] A. Djament, "Des propriétés de finitude des foncteurs polynomiaux", *Fund. Math.* **233**:3 (2016), 197–256. MR Zbl

[Djament and Vespa 2019] A. Djament and C. Vespa, "Foncteurs faiblement polynomiaux", *Int. Math. Res. Not.* **2019**:2 (2019), 321–391. MR Zbl

[Gadish 2017] N. Gadish, "Categories of FI type: a unified approach to generalizing representation stability and character polynomials", *J. Algebra* **480** (2017), 450–486. MR Zbl

[Gan 2017] W. L. Gan, "On the negative-one shift functor for FI-modules", *J. Pure Appl. Algebra* **221**:5 (2017), 1242–1248. MR Zbl

[Gan and Li 2017] W. L. Gan and L. Li, "Bounds on homological invariants of VI-modules", preprint, 2017. arXiv

[Gan and Watterlond 2018] W. L. Gan and J. Watterlond, "A representation stability theorem for VI-modules", *Algebr. Represent. Theory* **21**:1 (2018), 47–60. MR Zbl

[Gan et al. 2017] W. L. Gan, L. Li, and C. Xi, "An application of Nakayama functor in representation stability theory", preprint, 2017. arXiv

[Harman 2017] N. Harman, "Virtual Specht stability for *FI*-modules in positive characteristic", *J. Algebra* **488** (2017), 29–41. MR Zbl

[Joyal and Street 1995] A. Joyal and R. Street, "The category of representations of the general linear groups over a finite field", *J. Algebra* **176**:3 (1995), 908–946. MR Zbl

[Kuhn 2015] N. J. Kuhn, "Generic representation theory of finite fields in nondescribing characteristic", *Adv. Math.* **272** (2015), 598–610. MR Zbl

[Li 2016] L. Li, "Upper bounds of homological invariants of FI_G-modules", *Arch. Math.* (*Basel*) **107**:3 (2016), 201–211. MR Zbl

[Li and Ramos 2018] L. Li and E. Ramos, "Depth and the local cohomology of FI_G-modules", *Adv. Math.* **329** (2018), 704–741. MR Zbl

[Mac Lane 1963] S. Mac Lane, *Homology*, Grundlehren der Math. Wissenschaften **114**, Academic Press, New York, 1963. MR Zbl

[Miller and Wilson 2018] J. Miller and J. C. H. Wilson, "Quantitative representation stability over linear groups", *Int. Math. Res. Not.* (online publication October 2018).

[Nagpal 2015] R. Nagpal, *FI-modules and the cohomology of modular representations of symmetric groups*, Ph.D. thesis, University of Wisconsin-Madison, 2015, Available at https://search.proquest.com/docview/1681518656.

[Nagpal 2018] R. Nagpal, "VI-modules in non-describing characteristic, II", preprint, 2018. arXiv

[Nagpal and Snowden 2018] R. Nagpal and A. Snowden, "Periodicity in the cohomology of symmetric groups via divided powers", *Proc. Lond. Math. Soc.* (3) **116**:5 (2018), 1244–1268. MR Zbl

[Nagpal et al. 2016] R. Nagpal, S. V Sam, and A. Snowden, "Noetherianity of some degree two twisted commutative algebras", *Selecta Math.* (*N.S.*) **22**:2 (2016), 913–937. MR Zbl

[Nagpal et al. 2018] R. Nagpal, S. V Sam, and A. Snowden, "Regularity of FI-modules and local cohomology", *Proc. Amer. Math. Soc.* **146**:10 (2018), 4117–4126. MR Zbl

[Palmquist and Newell 1971] J. F. Palmquist and D. C. Newell, "Bifunctors and adjoint pairs", *Trans. Amer. Math. Soc.* **155** (1971), 293–303. MR Zbl

[Powell 1998] G. M. L. Powell, "The structure of indecomposable injectives in generic representation theory", *Trans. Amer. Math. Soc.* **350**:10 (1998), 4167–4193. MR Zbl

[Powell 2000] G. M. L. Powell, "On Artinian objects in the category of functors between $\mathbb{F}_2$-vector spaces", pp. 213–228 in *Infinite length modules* (Bielefeld, Germany, 1998), edited by H. Krause and C. M. Ringel, Birkhäuser, Basel, 2000. MR Zbl

[Putman and Sam 2017] A. Putman and S. V Sam, "Representation stability and finite linear groups", *Duke Math. J.* **166**:13 (2017), 2521–2598. MR Zbl

[Ramos 2018] E. Ramos, "Homological invariants of FI-modules and FI_G-modules", *J. Algebra* **502** (2018), 163–195. MR Zbl

[Sam and Snowden 2015] S. V Sam and A. Snowden, "Stability patterns in representation theory", *Forum Math. Sigma* **3** (2015), art. id. e11. MR Zbl

[Sam and Snowden 2016] S. V Sam and A. Snowden, "GL-equivariant modules over polynomial rings in infinitely many variables", *Trans. Amer. Math. Soc.* **368**:2 (2016), 1097–1158. MR Zbl

[Sam and Snowden 2017a] S. V Sam and A. Snowden, "Gröbner methods for representations of combinatorial categories", *J. Amer. Math. Soc.* **30**:1 (2017), 159–203. MR Zbl

[Sam and Snowden 2017b] S. V Sam and A. Snowden, "Regularity bounds for twisted commutative algebras", preprint, 2017. arXiv

[Sam and Snowden 2019] S. V Sam and A. Snowden, "GL-equivariant modules over polynomial rings in infinitely many variables, II", *Forum Math. Sigma* **7** (2019), art. id. e5. MR Zbl

[Stacks 2005–] P. Belmans, A. J. de Jong, et al., "The Stacks project", electronic reference, 2005–, Available at http://stacks.math.columbia.edu.

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Stud. Adv. Math. **38**, Cambridge Univ. Press, 1994. MR Zbl

[Zelevinsky 1981] A. V. Zelevinsky, *Representations of finite classical groups: a Hopf algebra approach*, Lecture Notes in Math. **869**, Springer, 1981. MR Zbl

Communicated by Victor Reiner
Received 2018-11-20 Revised 2019-05-30 Accepted 2019-07-22

rohitna@gmail.com *Department of Mathematics, University of Chicago, IL*

Current address: *Department of Mathematics, University of Michigan, Ann Arbor, MI, United States*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2191

Degree of irrationality of very general abelian surfaces

Nathan Chen

The degree of irrationality of a projective variety X is defined to be the smallest degree of a rational dominant map to a projective space of the same dimension. For abelian surfaces, Yoshihara computed this invariant in specific cases, while Stapleton gave a sublinear upper bound for very general polarized abelian surfaces (A, L) of degree d. Somewhat surprisingly, we show that the degree of irrationality of a very general polarized abelian surface is uniformly bounded above by 4, independently of the degree of the polarization. This result disproves part of a conjecture of Bastianelli, De Poi, Ein, Lazarsfeld, and Ullery.

1. Introduction

Given a projective variety X of dimension n which is not rational, one can try to quantify how far it is from being rational. When $n = 1$, the natural invariant is the *gonality* of a curve C, defined to be the smallest degree of a branched covering $C' \to \mathbb{P}^1$ (where C' is the normalization of C). One generalization of gonality to higher dimensions is the *degree of irrationality*, defined as:

$$\mathrm{irr}(X) := \min\{\delta > 0 \mid \exists \text{ degree } \delta \text{ rational dominant map } X \dashrightarrow \mathbb{P}^n\}.$$

Recently, there has been significant progress in understanding the case of hypersurfaces of large degree [Bastianelli 2017; Bastianelli et al. 2014; 2017]. The history behind the development of these ideas is described in [Bastianelli et al. 2017]. The results of those works depend on the positivity of the canonical bundles of the varieties in question, so it is interesting to consider what happens in the K_X-trivial case. Our purpose here is to prove the somewhat surprising fact that the degree of irrationality of a very general polarized abelian surface is uniformly bounded above, independently of the degree of the polarization.

To be precise, let $A = A_d$ be an abelian surface carrying a polarization $L = L_d$ of type $(1, d)$ and assume that $\mathrm{NS}(A) \cong \mathbb{Z}[L]$. An argument of Stapleton [2017] showed that there is a positive constant C such that

$$\mathrm{irr}(A) \le C \cdot \sqrt{d}$$

for $d \gg 0$, and it was conjectured in [Bastianelli et al. 2017] that equality holds asymptotically. Our main result shows that this is maximally false:

MSC2010: primary 14K99; secondary 14E05.
Keywords: Irrationality, abelian surface.
Research partially supported by the National Science Foundation under the Stony Brook/SCGP RTG grant DMS-1547145.

Theorem 1.1. *For an abelian surface $A = A_d$ with Picard number $\rho = 1$, one has*

$$\operatorname{irr}(A) \leq 4.$$

As far as we can see, the conjecture of [Bastianelli et al. 2017] for very general polarized K3 surfaces (S_d, B_d) of genus d — namely, that there exist positive constants C_1, C_2 satisfying $C_1 \cdot \sqrt{d} \leq \operatorname{irr}(S_d) \leq C_2 \cdot \sqrt{d}$ for $d \gg 0$ — remains plausible. Here, B_d is an ample line bundle generating $\operatorname{Pic}(S_d)$ with $B_d^2 = 2d - 2$.

For an abelian variety A of dimension n, it has been shown in [Alzati and Pirola 1992] that $\operatorname{irr}(A) \geq n+1$. When A is an abelian surface, we give a geometric proof of the fact that $\operatorname{irr}(A) \geq 3$ in Lemma 3.1. Yoshihara proved that $\operatorname{irr}(A) = 3$ for abelian surfaces A containing a smooth curve of genus 3 [Yoshihara 1996]. On a related note, Voisin [2018] showed that the covering gonality of a very general abelian variety A of dimension n is bounded from below by $f(n)$, where $f(n)$ grows like $\log(n)$, and this lower bound was subsequently improved to $\lceil \frac{1}{2}n + 1 \rceil$ by Martin [2019]. The covering gonality is defined as the minimum integer $c > 0$ such that given a general point $x \in A$, there exists a curve C passing through x with gonality c.

In the proof of our theorem, assuming as we may that L is symmetric, we consider the space $H^0(A, \mathcal{O}_A(2L))^+$ of even sections of $\mathcal{O}_A(2L)$. By imposing suitable multiplicities at the two-torsion points of A, we construct a subspace $V \subset H^0(A, \mathcal{O}_A(2L))^+$ which numerically should define a rational map from A to a surface $S \subset \mathbb{P}^N$. Using bounds on the degree of the map and the degree of S, we construct a rational covering $A \dashrightarrow \mathbb{P}^2$ of degree 4. The main difficulty is to deal with the possibility that $\mathbb{P}_{\text{sub}}(V)$ has a fixed component. Our approach was inspired in part by the work of Bauer [1994; 1998; 1999].

2. Set-up

Let $A = A_d$ be an abelian surface with $\rho(A) = 1$. Assume $\operatorname{NS}(A) \cong \mathbb{Z}[L]$ where L is a polarization of type $(1, d)$ for some fixed $d \geq 1$, so that $L^2 = 2d$ and $h^0(L) = d$. Let

$$\iota : A \to A, \quad x \mapsto -x$$

be the inverse morphism and let $Z = \{p_1, \ldots, p_{16}\}$ be the set of two-torsion points of A (fixed points of ι). We may assume that L is symmetric — that is, $\iota^* \mathcal{O}_A(L) \cong \mathcal{O}_A(L)$ — by replacing L with a suitable translate. In particular, the cyclic group of order two acts on $H^0(A, \mathcal{O}_A(2L))$. The space of *even* sections $H^0(A, \mathcal{O}_A(2L))^+$ of the line bundle $\mathcal{O}_A(2L)$ (sections s with the property that $\iota^* s = s$) has dimension

$$h^0(A, 2L)^+ = 2d + 2$$

(see [Lange and Birkenhake 1992, Corollary 4.6.6]). Since an even section of $\mathcal{O}_A(2L)$ vanishes to even order at any two-torsion point, it is at most

$$1 + 3 + \cdots + (2m - 1) = m^2$$

conditions for an even section to vanish to order $2m$ at a fixed point $p \in Z$ (see [Bauer 1994] and the Appendix to [Bauer 1998] for more details).

Fix any integer solutions $a_1, \dots, a_{16} \geq 0$ to the equation

$$\sum_{i=1}^{16} a_i^2 = 2d - 2,$$

with $a_{15} = 0 = a_{16}$ (this last assumption will be useful in Corollary 3.4). Such a solution always exists by Lagrange's four-square theorem. Let $V \subset H^0(A, \mathcal{O}_A(2L))^+$ be the space of even sections vanishing to order at least $2a_i$ at each point p_i, so that

$$\dim V \geq 2d + 2 - \sum_{i=1}^{16} a_i^2 \geq 4.$$

Projectivizing via subspaces, let $\mathfrak{d} = \mathbb{P}_{\text{sub}}(V) \subseteq |2L|^+$ be the corresponding linear system of divisors, whose dimension is $N := \dim \mathfrak{d} \geq 3$. Write

$$d_i := \operatorname{mult}_{p_i} D$$

for a general divisor $D \in \mathfrak{d}$, so that $d_i \geq 2a_i$.

Remark 2.1. From [Lange and Birkenhake 1992, Section 4.8], it follows that sections of V are pulled back from the singular Kummer surface A/ι, so any divisor $D \in \mathfrak{d}$ is symmetric, i.e., $\iota(D) = D$.

Let $\varphi : A \dashrightarrow \mathbb{P}^N$ be the rational map given by the linear system $\mathfrak{d}$ above (if $\mathfrak{d}$ has a fixed component F, take $\mathfrak{d} - F$), and write $S := \overline{\operatorname{Im}(\varphi)}$ for the image of φ. Regardless of whether or not $\mathfrak{d}$ has a fixed component, we find that:

Proposition 2.2. *$S \subset \mathbb{P}^N$ is an irreducible and nondegenerate surface.*

Proof. Suppose for the sake of contradiction that $\overline{\operatorname{Im}(\varphi)}$ is a nondegenerate curve C. Then $\deg C \geq 3$ since $N \geq 3$, and a hyperplane section of $C \subset \mathbb{P}^N$ pulls back to a divisor with at least three irreducible components. This contradicts the fact that any divisor $D(\sim_{\text{lin}} 2L) \in \mathfrak{d}$ has at most two irreducible components since $\operatorname{NS}(A) \cong \mathbb{Z}[L]$. So the image of φ is a surface. □

The following lemma will also be useful:

Lemma 2.3. *Let $\varphi : X \dashrightarrow \mathbb{P}^n$ be a rational map from a surface X to a projective space of dimension $n \geq 2$, and suppose that its image $S := \overline{\operatorname{Im}(\varphi)} \subset \mathbb{P}^n$ has dimension 2. Let $\mathfrak{d}$ be the linear system corresponding to φ (assuming $\mathfrak{d}$ has no base components). Then for any $D \in \mathfrak{d}$,*

$$\deg \varphi \cdot \deg S \leq D^2.$$

Proof. The indeterminacy locus of φ is a finite set. □

3. Degree bounds

We first begin with an observation, which holds for an arbitrary abelian surface:

Lemma 3.1. *There are no rational dominant maps $A \dashrightarrow \mathbb{P}^2$ of degree* 2.

Proof. Suppose there exists such a map f. We have the following diagram

$$\begin{tikzcd} & & A^{[2]} \arrow[r, "s"] & A \\ A \arrow[r, dashed, "f"] & \mathbb{P}^2 \arrow[ur, dashed, "g"] \arrow[r, dashed, "h"'] & K^{[2]}(A) \arrow[u, hook] & \!\!\!\!\!\!\!\!\!\!\!\!\!=: s^{-1}(0) \end{tikzcd}$$

where g is the pullback map on 0-cycles, $A^{[2]}$ is the Hilbert scheme of 2 points on A, and s is given by summation composed with the Hilbert–Chow morphism. Since the rational map $s \circ g$ can be extended to a morphism (see [Lange and Birkenhake 1992, Theorem 4.9.4]), it must be constant. So $\overline{\mathrm{Im}(g)}$ is contained in a fiber $s^{-1}(0)$, which is a smooth Kummer K3 surface $K^{[2]}(A)$. Since g is injective, it descends to an injective (and hence birational) map $h : \mathbb{P}^2 \dashrightarrow K^{[2]}(A)$, yielding a contradiction. □

We will now study the numerical properties of the linear series $\mathfrak{d}$ constructed in the previous section. There are two possibilities for $\mathfrak{d}$; either (i) $\mathfrak{d}$ has no fixed component or (ii) $\mathfrak{d}$ has a fixed component, denoted by $F \neq 0$. In fact, with a little more work one can show that the second case does not actually occur; see Remark 3.5.

In the second case, let $\mathfrak{b}$ be the movable component of $\mathfrak{d}$, so that we may write every divisor $D \in \mathfrak{d}$ as

$$D = F + M \quad \text{where } M \in \mathfrak{b}.$$

By definition, $\dim \mathfrak{d} = \dim \mathfrak{b}$. Since $\mathrm{NS}(A) \cong \mathbb{Z}[L]$, $D \sim_{\mathrm{lin}} 2L$ implies $F, M \sim_{\mathrm{alg}} L$ and are irreducible effective divisors for all $M \in \mathfrak{b}$. Choose a general divisor $M \in \mathfrak{b}$ and write

$$m_i := \mathrm{mult}_{p_i} M \quad \text{and} \quad f_i := \mathrm{mult}_{p_i} F,$$

so that $d_i = m_i + f_i \geq 2a_i$ for all i. We claim that F must be symmetric as a divisor. If not, then

$$\iota(M) + \iota(F) = \iota(D) = D = M + F \quad \text{for all } D \in \mathfrak{d}.$$

This implies that $M = \iota(F)$ and $F = \iota(M)$ for all $M \in \mathfrak{b}$, which would mean that M must also be fixed, leading to a contradiction. Hence, F must be symmetric, and likewise for all $M \in \mathfrak{b}$.

We first need an intermediate estimate:

Proposition 3.2. *Assume $\mathfrak{d}$ has a fixed component $F \neq 0$. Keeping the notation as above,*

$$\sum_{i=1}^{16} m_i^2 \geq 2d - 8.$$

Proof. The idea here is to use the Kummer construction to push our fixed curve F onto a K3 surface and apply Riemann–Roch. This is analogous to a proof of Bauer's [1999, Theorem 6.1]. Consider the smooth Kummer K3 surface K associated to A:

$$\begin{tikzcd} E \subset \hat{A} \arrow[r, "\gamma"] \arrow[d, "\pi"'] & \hat{A}/\{1,\sigma\} =: K \\ Z \subset A & \end{tikzcd}$$

where π is the blow-up of A along the collection of two-torsion points Z. Since the points in Z are ι-invariant, ι lifts to an involution σ on $\hat{A}$ and the quotient K is a smooth K3 surface. Let E_i denote the exceptional curve over $p_i \in Z$, so that $E = \sum_{i=1}^{16} E_i$ is the exceptional divisor of π. Since F is symmetric, its strict transform

$$\hat{F} = \pi^* F - \sum_{i=1}^{16} f_i E_i,$$

descends to an irreducible curve $\bar{F} \subset K$. We claim that

$$h^0(K, \mathcal{O}_K(\bar{F})) = 1.$$

In fact, if the linear system $|\mathcal{O}_K(\bar{F})|$ were to contain a pencil, then this would give us a pencil of symmetric curves in $|\mathcal{O}_A(F)|$ with the same multiplicities at the two-torsion points, which contradicts F being a fixed component of $\mathfrak{d}$.

On the other hand, it is well-known that an irreducible curve $\bar{F}$ on a K3 surface with $h^0(K, \bar{F}) = 1$ satisfies $(\bar{F})^2 = -2$, so

$$-4 = 2(\bar{F})^2 = (\gamma^* \bar{F})^2 = (\hat{F})^2 = F^2 - \sum_{i=1}^{16} f_i^2 = 2d - \sum_{i=1}^{16} f_i^2 \tag{1}$$

combined with $\sum_{i=1}^{16} f_i m_i \leq \sum_{i=1}^{16} \left(\frac{d_i}{2}\right)^2$ yields

$$\sum_{i=1}^{16} d_i^2 = \sum_{i=1}^{16} (f_i^2 + m_i^2 + 2 f_i m_i) \leq 2d + 4 + \sum_{i=1}^{16} m_i^2 + \tfrac{1}{2} \sum_{i=1}^{16} d_i^2.$$

After rearranging the terms, we find that

$$\sum_{i=1}^{16} m_i^2 \geq -2d - 4 + \tfrac{1}{2} \sum_{i=1}^{16} d_i^2 \geq -2d - 4 + 2 \sum_{i=1}^{16} a_i^2 = 2d - 8 \tag{2}$$

for a general divisor $D = F + M \in \mathfrak{d}$, which is the desired inequality. □

As an immediate consequence:

Theorem 3.3. *Keeping the notation as before, let $\varphi : A \dashrightarrow \mathbb{P}^N$ be the rational map corresponding to $\mathfrak{d}$ (or $\mathfrak{b}$ if $F \neq 0$), with image S. Then*

$$\deg \varphi \cdot \deg S \leq 8. \tag{3}$$

Proof. By applying Proposition 2.2 and blowing-up A along the collection of two-torsion points Z to resolve some of the base points of $\mathfrak{d}$, we arrive at the diagram

$$\begin{tikzcd} \hat{A} := \mathrm{Bl}_Z A \arrow[d, "\pi"'] \arrow[dr, dashed, "\psi"] & \\ A \arrow[r, dashed, "\varphi"'] & S \subset \mathbb{P}^N. \end{tikzcd}$$

(i) If the linear system $\mathfrak{d}$ has no fixed components, the divisors corresponding to ψ are of the form

$$\hat{D} \sim_{\mathrm{lin}} \pi^* D - \sum_{i=1}^{16} d_i E_i,$$

where $\hat{D}$ denotes the strict transform of D. By Lemma 2.3 applied to ψ, we have

$$\deg \varphi \cdot \deg S = \deg \psi \cdot \deg S \le \hat{D}^2 = 4L^2 - \sum_{i=1}^{16} d_i^2 \le 4\left(2d - \sum_{i=1}^{16} a_i^2\right) = 8.$$

(ii) If the linear system $\mathfrak{d}$ has a fixed component $F \neq 0$, replace $\hat{D}$ and d_i in the equation above with $\hat{M}$ and m_i, respectively. Proposition 3.2 then gives an analogous bound. □

Corollary 3.4. *There exists a rational dominant map $\varphi : A \dashrightarrow \mathbb{P}^2$ of degree* 4.

Proof. From Remark 2.1, it follows that $\varphi : A \dashrightarrow S \subset \mathbb{P}^N$ factors through the quotient $A \to A/\iota$, so $\deg \varphi$ must be even. In addition, $\deg S \ge 2$ since S is nondegenerate. By Lemma 3.1, it is impossible for S to be rational together with $\deg \varphi = 2$, so $\{\deg \varphi = 2, \deg S = 2, 3\}$ is ruled out by the classification of quadric and cubic surfaces (using the fact that $\rho(A) = 1$).

Together with the upper bound $\deg \varphi \cdot \deg S \le 8$ given by Theorem 3.3, there are two possibilities:

$$\{\deg \varphi = 4,\ \deg S = 2\} \quad \text{and} \quad \{\deg \varphi = 2,\ \deg S = 4\}.$$

Either of these imply the equality in (3), so that we have a morphism $\mathrm{Bl}_Z A \to S$ which fits into the diagram:

$$\begin{tikzcd} E_i \subset \mathrm{Bl}_Z A \arrow[r, "\gamma"] \arrow[d, "\pi"'] \arrow[dr] & K \supset G_i \arrow[d, "\alpha"] \\ A \arrow[r, dashed, "\varphi"'] & S \subset \mathbb{P}^N \end{tikzcd}$$

where K is the smooth Kummer K3 surface, γ is a branched cover of degree 2, and $G_i := \gamma(E_i)$ are (-2)-curves.

In the first case where $\deg \varphi = 4$ and $\deg S = 2$, note that S is rational. In the second case where $\deg \varphi = 2$ and $\deg S = 4$, recall that we chose the multiplicities a_i so that $a_{15} = 0 = a_{16}$. Thus, equality in (3) forces either $d_{15} = 0 = d_{16}$ or $m_{15} = 0 = m_{16}$. This implies that the curves G_{15}, G_{16} are contracted and their images q_{15}, q_{16} under α are double points on S since α is a birational morphism. Projection from a general $(N-3)$-plane containing one but not both of the q_i defines a rational map $A \dashrightarrow \mathbb{P}^2$ of degree 2 (if q_{15} is a cone point of S, pick a general plane passing through q_{16}, and vice versa), which contradicts Lemma 3.1. $\square$

This immediately implies Theorem 1.1.

Remark 3.5. The case when $\mathfrak{d}$ has a fixed component $F \neq 0$ cannot occur. To see this, suppose $F \neq 0$ and note that the two cases given in Corollary 3.4 imply that equality must hold throughout the proof of Proposition 3.2. In particular, $d_i = m_i + f_i$ and $\sum_{i=1}^{16} f_i m_i = \sum_{i=1}^{16} \left(\frac{d_i}{2}\right)^2$ implies $f_i = m_i$ for all i. Combining this with (1) and (2) gives

$$2d + 4 = \sum_{i=1}^{16} f_i^2 = \sum_{i=1}^{16} m_i^2 = 2d - 8,$$

which is a contradiction.

Acknowledgements

I would like to thank my advisor Robert Lazarsfeld for suggesting the conjecture and for his encouragement and guidance throughout the formulation of the results in this paper. I would also like to thank Frederik Benirschke, Matthew Dannenberg, Mohamed El Alami, François Greer, Samuel Grushevsky, Ljudmila Kamenova, Yoon-Joo Kim, Radu Laza, John Sheridan, David Stapleton, and Ruijie Yang for engaging in valuable discussions. Finally, I am grateful to the referee for reviewing the paper and offering helpful comments.

References

[Alzati and Pirola 1992] A. Alzati and G. P. Pirola, "On the holomorphic length of a complex projective variety", *Arch. Math. (Basel)* **59**:4 (1992), 398–402. MR Zbl

[Bastianelli 2017] F. Bastianelli, "On irrationality of surfaces in $\mathbb{P}^3$", *J. Algebra* **488** (2017), 349–361. MR Zbl

[Bastianelli et al. 2014] F. Bastianelli, R. Cortini, and P. De Poi, "The gonality theorem of Noether for hypersurfaces", *J. Algebraic Geom.* **23**:2 (2014), 313–339. MR Zbl

[Bastianelli et al. 2017] F. Bastianelli, P. De Poi, L. Ein, R. Lazarsfeld, and B. Ullery, "Measures of irrationality for hypersurfaces of large degree", *Compos. Math.* **153**:11 (2017), 2368–2393. MR Zbl

[Bauer 1994] T. Bauer, "Projective images of Kummer surfaces", *Math. Ann.* **299**:1 (1994), 155–170. MR Zbl

[Bauer 1998] T. Bauer, "Seshadri constants and periods of polarized abelian varieties", *Math. Ann.* **312**:4 (1998), 607–623. MR Zbl

[Bauer 1999] T. Bauer, "Seshadri constants on algebraic surfaces", *Math. Ann.* **313**:3 (1999), 547–583. MR Zbl

[Lange and Birkenhake 1992] H. Lange and C. Birkenhake, *Complex abelian varieties*, Grundlehren der Math. Wissenschaften **302**, Springer, 1992. MR Zbl

[Martin 2019] O. Martin, "On a conjecture of Voisin on the gonality of very general abelian varieties", preprint, 2019. arXiv

[Stapleton 2017] D. Stapleton, *The degree of irrationality of very general hypersurfaces in some homogeneous spaces*, Ph.D. thesis, Stony Brook University, 2017, Available at https://search.proquest.com/docview/1972010253.

[Voisin 2018] C. Voisin, "Chow rings and gonality of general abelian varieties", preprint, 2018. arXiv

[Yoshihara 1996] H. Yoshihara, "Degree of irrationality of a product of two elliptic curves", *Proc. Amer. Math. Soc.* **124**:5 (1996), 1371–1375. MR Zbl

Communicated by Gavril Farkas
Received 2019-02-17 Revised 2019-05-17 Accepted 2019-06-25

nathan.chen@stonybrook.edu *Department of Mathematics, Stony Brook University, Stony Brook, NY, United States*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2199

Lower bounds for the least prime in Chebotarev

Andrew Fiori

In this paper we show there exists an infinite family of number fields L, Galois over $\mathbb{Q}$, for which the smallest prime p of $\mathbb{Q}$ which splits completely in L has size at least $(\log(|D_L|))^{2+o(1)}$. This gives a converse to various upper bounds, which shows that they are best possible.

1. Introduction

The purpose of this note is to prove the following result.

Theorem 1. *There exists an infinite family of number fields L, Galois over $\mathbb{Q}$, for which the smallest prime p of $\mathbb{Q}$ which splits completely in L has size at least*

$$(1+o(1))\left(\frac{3e^{\gamma}}{2\pi}\right)^2\left(\frac{\log(|D_L|)\log(2\log\log(|D_L|))}{\log\log(|D_L|)}\right)^2$$

as the absolute discriminant D_L of L over $\mathbb{Q}$, tends to infinity.

The result is independent of the generalized Riemann hypothesis. The result complements the existing literature on what is essentially a converse problem, stated generally as:

Problem. Let K be a number field, and L be a Galois extension of K, for any conjugacy class $\mathcal{C}$ in $\Gamma(L/K)$, the Galois group of L/K, show that the smallest (in norm) unramified degree one prime $\mathfrak{p}$ of K for which the conjugacy class $\mathrm{Frob}_{\mathfrak{p}}$ is $\mathcal{C}$ is *small* relative to $|D_L|$, the absolute discriminant of L/K.

Solutions to this problem have important applications in the explicit computation of class groups (see [Belabas et al. 2008]) where smaller is better. Some of the history of just how small we can get is summarized below:

- Lagarias and Odlyzko [1977] showed $N_{K/\mathbb{Q}}(\mathfrak{p}) < (\log(|D_L|))^{2+o(1)}$ conditionally on GRH.
- Bach and Sorenson [1996] gave an explicit constant C so that $N_{K/\mathbb{Q}}(\mathfrak{p}) < C(\log(|D_L|))^2$ conditionally on GRH.
- Lagarias, Montgomery, and Odlyzko [Lagarias et al. 1979] showed there is a constant A such that $N_{K/\mathbb{Q}}(\mathfrak{p}) < |D_L|^A$.

The author thanks the University of Lethbridge for providing a stimulating environment to conduct this work and in particular Nathan Ng and Habiba Kadiri for directing him towards this project. The author would also like to thank the referee whose suggestions simplified the proof of our main result.
MSC2010: primary 11R44; secondary 11R29.
Keywords: Chebotarev, class groups.

- Zaman [2017] showed $N_{K/\mathbb{Q}}(\mathfrak{p}) < |D_L|^{40}$ for D_L sufficiently large.
- Kadiri, Ng and Wong [Kadiri et al. 2019] improved this to $N_{K/\mathbb{Q}}(\mathfrak{p}) < |D_L|^{16}$ for D_L sufficiently large.
- Ahn and Kwon [2019] showed $N_{K/\mathbb{Q}}(\mathfrak{p}) < |D_L|^{12577}$ for all L.

By the above, Theorem 1 and the GRH bound above are best possible up to the exact $o(1)$ term.

Remark. The family under consideration will be a subfamily of the Hilbert class fields of quadratic imaginary extensions of $\mathbb{Q}$. All of the Galois groups will be generalized dihedral groups, and in the family the degree of the extensions goes to infinity.

We also would like to point out the work of Sandari [2018, Section 1.3] where some similar features of this family are remarked on in a different context.

2. Proofs

We first recall a few basic facts from algebraic number theory and class field theory.

Lemma 2. *Let $K = \mathbb{Q}(\sqrt{-d})$ where $d = |\mathrm{disc}(K)|$, let $\mathfrak{p}$ be a principal prime ideal of K. If we have $N_{K/\mathbb{Q}}(\mathfrak{p}) = (p)$ then p is a norm of $\mathcal{O}_K$ and hence $p \geq \frac{d}{4}$.*

Proof. Assuming $\mathfrak{p}$ is principally generated by x, then $N_{K/\mathbb{Q}}(\mathfrak{p})$ is principally generated by $N_{K/\mathbb{Q}}(x)$. As norms from K are positive, this gives that p must be a norm.

We next note that for $x + y\sqrt{-d} \in \mathcal{O}_K$ the expression $N_{K/\mathbb{Q}}(x + y\sqrt{-d}) = x^2 + dy^2$ cannot be prime if $y = 0$. Now, because $\mathcal{O}_k \subset \frac{1}{2}\mathbb{Z} + \frac{\sqrt{-d}}{2}\mathbb{Z}$ we conclude that if the norm is a prime, then $y \geq \frac{1}{2}$, from which it follows that if p is a norm then $p \geq \frac{d}{4}$. □

Lemma 3. *Let $K = \mathbb{Q}(\sqrt{-d})$ where $d = |\mathrm{disc}(K)|$, suppose that H is the Hilbert class field of K. If p is a prime of $\mathbb{Z}$ which splits completely in H, then p splits in K as $(p) = \mathfrak{p}_1\mathfrak{p}_1$ where both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are principal and $N_{K/\mathbb{Q}}(\mathfrak{p}_i) = (p)$. In particular, by the previous lemma $p \geq \frac{d}{4}$.*

Proof. The first claim is clear because ramification degrees, inertia degrees and hence splitting degrees are multiplicative in towers. That $\mathfrak{p}_i$ must be principal is a consequence of class field theory. Principal ideals for $\mathcal{O}_K$ map to the trivial Galois element for the Galois group of the Hilbert class field. However, for unramified prime ideals this map gives Frobenius. As the Frobenius element is trivial precisely when the inertial degree is 1, equivalently for Galois fields when the prime splits completely, we conclude the result. □

Remark 4. Denote by χ_d the quadratic Dirichlet character with fundamental discriminant $-d$. The main idea of the proof is to use the class number formula with lower bounds for $L(1, \chi_d)$. Using Siegel's ineffective bound gives

$$d = h_K^{2+o(1)} = \log|D_H|^{2+o(1)}.$$

To obtain our precise result we refine the $o(1)$ term using extreme values of $L(1, \chi_d)$.

Lemma 5. *Let $K = \mathbb{Q}(\sqrt{-d})$ where $d = |\mathrm{disc}(K)| > 16$, suppose that H is the Hilbert class field of K. Then*

$$\log|D_H| = h_K \log(d) = \tfrac{1}{\pi} L(1, \chi_d)\sqrt{d}\log(d)$$

where h_K is the class number of K, D_H is the discriminant of H and χ_d is the quadratic Dirichlet character with fundamental discriminant $-d$.

Proof. The first equality is immediate from the multiplicativity of the discriminant in towers, the second follows from the analytic class number formula

$$h_K = \frac{\sqrt{d}}{\pi} L(1, \chi_d). \qquad \square$$

The estimates on the extreme values of $L(1, \chi_d)$ which we need are the following.

Theorem 6. *There exists a family of quadratic imaginary fields $K = \mathbb{Q}(\sqrt{-d})$ where $d = |\mathrm{disc}(K)|$ such that for χ_d, the quadratic Dirichlet character with fundamental discriminant $-d$, we have*

$$L(1, \chi_d) < (1 + o(1)) \frac{\pi^2}{6e^\gamma \log\log(d)}.$$

A result of this sort was originally proven by Littlewood [1928] conditional on the generalized Riemann hypothesis, his result was proven unconditionally by Paley [1932] the version stated here follows from the work of Chowla [1949]. It is possible that the work of Granville and Soundararajan [2003] can further refine the constants in the above, and consequently those in Theorem 1.

The following proof includes several significant simplifications suggested by the referee. We would like to thank them for these valuable suggestions.

Proof of Theorem 1. We consider the family of fields $L = H_K$ where K is a field from the infinite family of Theorem 6 for which $d > 16$. To complete the proof we introduce some notation, define

$$x_d = L(1, \chi_d)\log\log(d) \quad \text{and} \quad f_d(x) = \frac{x\sqrt{d}\log(d)}{\pi \log\log(d)}.$$

Then by our choice of d we have

$$x_d < \frac{\pi^2}{6e^\gamma} + o(1)$$

and by Lemma 5 we have

$$\log|D_L| = f_d(x_d).$$

Now because the function $y \mapsto y\log(2\log(y))/\log(y)$ is increasing for $y > e$ and as

$$f_d(x_d) = \log|D_L| = h_K \log(d) \geq \log(16) > e$$

it follows that

$$\begin{aligned}\frac{\log|D_L|\log(2\log\log|D_L|)}{\log\log|D_L|} &= \frac{f_d(x_d)\log(2\log(f_d(x_d)))}{\log(f_d(x_d))} \\ &\le \frac{f_d\big(\frac{\pi^2}{6e^\gamma}+o(1)\big)\log\big(2\log\big(f_d\big(\frac{\pi^2}{6e^\gamma}+o(1)\big)\big)\big)}{\log\big(f_d\big(\frac{\pi^2}{6e^\gamma}+o(1)\big)\big)} \\ &\le (1+o(1))\frac{\pi}{3e^\gamma}\sqrt{d}.\end{aligned}$$

Combining the above with the bounds $p \ge \frac{d}{4}$ from Lemma 3 we obtain the result. $\square$

3. Numerics

Table 1 illustrates the phenomenon by giving the ratio

$$\text{Ratio} = p\Big/\left(\frac{3e^\gamma}{2\pi}\right)^2\left(\frac{\log(|D_L|)\log(2\log\log(|D_L|))}{\log\log(|D_L|)}\right)^2$$

for an example of a the Hilbert class field of a quadratic imaginary field of each class number less than 100 with large discriminant.

Note that in Table 1 we have $K = \mathbb{Q}(\sqrt{-d})$ and $|D_L| = d^{h_K}$.

References

[Ahn and Kwon 2019] J.-H. Ahn and S.-H. Kwon, "An explicit upper bound for the least prime ideal in the Chebotarev density theorem", *Ann. Inst. Fourier (Grenoble)* **69**:3 (2019), 1411–1458. Zbl

[Bach and Sorenson 1996] E. Bach and J. Sorenson, "Explicit bounds for primes in residue classes", *Math. Comp.* **65**:216 (1996), 1717–1735. MR Zbl

[Belabas et al. 2008] K. Belabas, F. Diaz y Diaz, and E. Friedman, "Small generators of the ideal class group", *Math. Comp.* **77**:262 (2008), 1185–1197. MR Zbl

[Chowla 1949] S. Chowla, "Improvement of a theorem of Linnik and Walfisz", *Proc. London Math. Soc.* (2) **50** (1949), 423–429. MR Zbl

[Granville and Soundararajan 2003] A. Granville and K. Soundararajan, "The distribution of values of $L(1, \chi_d)$", *Geom. Funct. Anal.* **13**:5 (2003), 992–1028. MR Zbl

[Kadiri et al. 2019] H. Kadiri, N. Ng, and P.-J. Wong, "The least prime ideal in the Chebotarev density theorem", *Proc. Amer. Math. Soc.* **147**:6 (2019), 2289–2303. MR Zbl

[Lagarias and Odlyzko 1977] 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, UK, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR Zbl

[Lagarias et al. 1979] J. C. Lagarias, H. L. Montgomery, and A. M. Odlyzko, "A bound for the least prime ideal in the Chebotarev density theorem", *Invent. Math.* **54**:3 (1979), 271–296. MR Zbl

[Littlewood 1928] J. E. Littlewood, "On the class-number of the corpus $P(\sqrt{-k})$", *Proc. London Math. Soc.* (2) **27**:1 (1928), 358–372. MR Zbl

[Paley 1932] R. E. A. C. Paley, "A theorem on characters", *J. London Math. Soc.* **7**:1 (1932), 28–32. MR Zbl

[Sardari 2018] N. T. Sardari, "The least prime number represented by a binary quadratic form", preprint, 2018. arXiv

[Zaman 2017] A. Zaman, "Bounding the least prime ideal in the Chebotarev density theorem", *Funct. Approx. Comment. Math.* **57**:1 (2017), 115–142. MR Zbl

h_K	d	p	Ratio	h_K	d	p	Ratio	h_K	d	p	Ratio
1	163	41	4.1557	34	189883	47491	2.2528	67	652723	163181	1.9030
2	427	107	2.4287	35	210907	52727	2.3373	68	819163	204791	2.2546
3	907	227	2.1188	36	217627	54409	2.2819	69	888427	222107	2.3556
4	1555	389	1.9476	37	158923	39733	1.6620	70	811507	202877	2.1215
5	2683	673	2.0276	38	289963	72493	2.6454	71	909547	227387	2.2823
6	3763	941	1.9222	39	253507	63377	2.2500	72	947923	236981	2.3061
7	5923	1481	2.1071	40	260947	65239	2.2034	73	886867	221717	2.1227
8	6307	1579	1.7569	41	296587	74149	2.3513	74	951043	237763	2.2001
9	10627	2657	2.1729	42	280267	70067	2.1445	75	916507	229127	2.0792
10	13843	3461	2.2386	43	300787	75209	2.1838	76	1086187	271549	2.3521
11	15667	3917	2.0939	44	319867	79967	2.2079	77	1242763	310693	2.5821
12	17803	4451	1.9938	45	308323	77081	2.0542	78	1004347	251087	2.0958
13	20563	5147	1.9503	46	462883	115727	2.7990	79	1333963	333491	2.6208
14	30067	7517	2.3373	47	375523	93887	2.2489	80	1165483	291371	2.2775
15	34483	8623	2.3173	48	335203	83813	1.9638	81	1030723	257687	2.0011
16	31243	7817	1.9050	49	393187	98297	2.1693	82	1446547	361637	2.6277
17	37123	9281	1.9719	50	389467	97367	2.0743	83	1074907	268729	1.9851
18	48427	12107	2.2225	51	546067	136519	2.6772	84	1225387	306347	2.1765
19	38707	9677	1.6747	52	439147	109789	2.1422	85	1285747	321443	2.2210
20	58507	14627	2.1572	53	425107	106277	2.0124	86	1534723	383681	2.5366
21	61483	15373	2.0614	54	532123	133033	2.3604	87	1261747	315437	2.0941
22	85507	21377	2.5024	55	452083	113021	1.9839	88	1265587	316403	2.0564
23	90787	22697	2.4308	56	494323	123581	2.0737	89	1429387	357347	2.2395
24	111763	27941	2.6847	57	615883	153991	2.4279	90	1548523	387137	2.3529
25	93307	23327	2.1425	58	586987	146749	2.2565	91	1391083	347771	2.1002
26	103027	25759	2.1714	59	474307	118583	1.8204	92	1452067	363017	2.1371
27	103387	25847	2.0351	60	662803	165701	2.3566	93	1475203	368801	2.1244
28	126043	31511	2.2543	61	606643	151667	2.1185	94	1587763	396943	2.2212
29	166147	41539	2.6760	62	647707	161947	2.1768	95	1659067	414767	2.2638
30	134467	33617	2.1037	63	991027	247759	3.0559	96	1684027	421009	2.2501
31	133387	33347	1.9698	64	693067	173267	2.1783	97	1842523	460633	2.3882
32	164803	41201	2.2263	65	703123	175781	2.1443	98	2383747	595939	2.9359
33	222643	55661	2.7216	66	958483	239623	2.7278	99	1480627	370159	1.9012

Table 1. Examples of smallest split primes in Hilbert class fields of $\mathbb{Q}(\sqrt{-d})$.

Communicated by Andrew Granville
Received 2019-03-01 Revised 2019-05-08 Accepted 2019-06-27

andrew.fiori@uleth.ca *Department of Mathematics and Computer Science, University of Lethbridge, AB, Canada*

ALGEBRA AND NUMBER THEORY 13:9 (2019)
dx.doi.org/10.2140/ant.2019.13.2205

Brody hyperbolicity of base spaces of certain families of varieties

Mihnea Popa, Behrouz Taji and Lei Wu

We prove that quasi-projective base spaces of smooth families of minimal varieties of general type with maximal variation do not admit Zariski dense entire curves. We deduce the fact that moduli stacks of polarized varieties of this sort are Brody hyperbolic, answering a special case of a question of Viehweg and Zuo. For two-dimensional bases, we show analogous results in the more general case of families of varieties admitting a good minimal model.

1. Introduction

The purpose of this paper is to establish a few results related to the hyperbolicity of base spaces of families of smooth complex varieties having maximal variation. Our study is motivated by the conjecturally degenerate behavior of entire curves inside the moduli P_h of polarized manifolds, corresponding to the moduli functor $\mathscr{P}_h$ which associates to a variety V the set $\mathscr{P}_h(V)$ of pairs $(f : U \to V, \mathscr{H})$, where f is a smooth projective morphism whose fibers have semiample canonical bundle and $\mathscr{H}$ is an f-ample line bundle with Hilbert polynomial h, up to isomorphisms and fiberwise numerical equivalence. The coarse moduli spaces P_h were shown to be quasi-projective schemes by Viehweg [1995].

1A. ***Families of minimal varieties of general type.*** The first result partially answers a question of Viehweg and Zuo [2003, Question 0.2], who established in their fundamental paper the analogous result in the case of moduli of canonically polarized manifolds (i.e., those whose canonical bundle is ample) [Viehweg and Zuo 2003, Theorem 0.1].

Theorem 1.1. *Let $f_U : U \to V$ be a smooth family of polarized manifolds in $\mathscr{P}_h(V)$ (in particular with semiample canonical bundle), with fibers of* general type *and with V quasi-projective, such that the induced morphism $\sigma : V \to P_h$ is quasi-finite. Then V is Brody hyperbolic, that is any holomorphic map $\gamma : \mathbb{C} \to V$ is constant.*

The question in [Viehweg and Zuo 2003] asks whether the same holds for moduli of arbitrary polarized varieties, i.e., not necessarily of general type. While this was our original goal, in the general case we have not been able to overcome difficulties related to vanishing theorems. We do however give a positive

MP was partially supported by the NSF grant DMS-1700819.
MSC2010: primary 14C30; secondary 14D07, 14E30, 14J10, 14J15, 14J29.
Keywords: Brody hyperbolicity, minimal models, moduli of polarized varieties, varieties of general type, Green–Griffiths–Lang's conjecture, Hodge modules.

answer to an even more general version of this question when V is a surface; see Corollary 1.5. Note that the more restrictive property of algebraic hyperbolicity, involving algebraic maps from curves and abelian varieties, has been known in great generality. It was established by Kovács [2000] for moduli of canonically polarized manifolds, and then by a combination of Viehweg and Zuo [2001] and Popa and Schnell [2017] for families admitting good minimal models. See also [Migliorini 1995] for families of surfaces.

Theorem 1.1 is a direct consequence of the following result regarding the base spaces of smooth families of minimal manifolds of general type that have maximal variation. Recall first that the exceptional locus of V is defined as

$$\mathrm{Exc}(V) := \overline{\left(\bigcup_{\gamma} \gamma(\mathbb{C})\right)},$$

where the union is taken over all nonconstant holomorphic maps $\gamma : \mathbb{C} \to V$, and the closure is in the Zariski topology.

Theorem 1.2. *Let $f_U : U \to V$ be a smooth projective morphism of smooth, quasi-projective varieties. Assume that f_U has maximal variation, and that its fibers are minimal manifolds of general type. Then the exceptional locus* $\mathrm{Exc}(V)$ *is a proper subset of V. In particular, every holomorphic map $\gamma : \mathbb{C} \to V$ is algebraically degenerate, that is the image of γ is not Zariski dense.*

For the general definition of the variation $\mathrm{Var}(f)$ of a family, we refer to [Viehweg 1983]. We are only concerned with maximal variation, $\mathrm{Var}(f) = \dim V$, which means that the very general fiber can only be birational to countably many other fibers; cf. also Lemma 3.11. For families coming from maps to moduli schemes, maximal variation simply means that the moduli map $V \to M$ is generically finite.

The theorem above is of course especially relevant for families of surfaces, where the minimality assumption becomes unnecessary, as one can pass to smooth minimal models in families. Recall that Gieseker [1977] has constructed a coarse moduli space M parametrizing birational isomorphism classes of surfaces of general type.

Corollary 1.3. *Let $f_U : U \to V$ be a smooth projective family of surfaces of general type with maximal variation. Then* $\mathrm{Exc}(V)$ *is a proper subset of V. If moreover the family comes from a quasi-finite map $V \to M$ to the moduli space of surfaces of general type, then V is Brody hyperbolic.*

1B. ***A tour of related problems and literature.*** Statements as in Theorems 1.1 and 1.2 are conjecturally expected to be consequences of a different property of a more algebraic flavor, which is the subject of Viehweg's hyperbolicity conjecture; itself a generalization of a conjecture of Shafaravich. Roughly speaking, Viehweg predicted that for families with maximal variation, a log smooth compactification (Y, D) of V is of log general type. The proof of the original statement of the conjecture, in the canonically polarized case, was established in important special cases in [Viehweg and Zuo 2002; Kebekus and Kovács 2008a; 2008b; 2010; Patakfalvi 2012], and was recently completed by Campana and Păun [2019, Theorem 8.1]; for a more detailed overview of this body of work and for further references, please see

[Popa and Schnell 2017, Section 1.2]. The statement was subsequently extended to families whose geometric generic fiber admits a good minimal model, so in particular to families of varieties of general type, by the first author and Schnell [Popa and Schnell 2017, Theorem A]. On the other hand, the conjecture of Green and Griffiths [1980] and Lang [1986], predicts that for a pair (Y, D) of log general type, the image of any entire curve

$$\gamma : \mathbb{C} \to V$$

is algebraically degenerate, where $V = Y \setminus D$.

It is worth noting that the hypotheses of the conjectures and results discussed above cannot be removed, at least not without imposing further restrictions. On one hand, the smoothness of the family is necessary, as it is well known that there exist non-smooth varying families of stable varieties (e.g., Lefschetz pencils) parametrized by $\mathbb{P}^1$. On the other hand, hyperbolicity may fail when the fibers have Kodaira dimension $-\infty$; for example, in [Javanpeykar and Loughran 2018] the authors exhibit examples of smooth maximally varying families of Fano threefolds parametrized by abelian surfaces.

We also note briefly that, besides the subject treated here, there are further important geometric and arithmetic aspects of the Shafarevich and Lang type conjectures. For instance, the geometric version of Shafarevich's conjecture has a boundedness component as well; a higher dimensional version for families of canonically polarized varieties was proved by Kovács and Lieblich [2010]. On the other hand, just as with entire curves, on the arithmetic side Lang's conjecture predicts that, over a number field, the set of rational points is not Zariski dense in a variety of general type. Not much is known in terms of general statements, besides of course Faltings' proof [1983] of the Mordell conjecture stating that there are only finitely many rational points on curves of genus at least two.

Going back to our main topic, in the canonically polarized case the problem of hyperbolicity of moduli stacks has a rich history from the purely analytic point of view. For the moduli stack $\mathcal{M}_g$ of compact Riemann surfaces of genus g, results of Ahlfors [1961], Royden [1975] and Wolpert [1986] show that the holomorphic sectional curvature of the Weil–Petersson metric on the base of a family admitting a quasi-finite map to $\mathcal{M}_g$, with $g \geq 2$, is negative and bounded away from zero. In particular, such base spaces are Brody hyperbolic. In higher dimensions, thanks to Aubin–Yau's solution to Calabi's conjecture, one studies equivalently families of compact complex manifolds admitting a smooth Kähler–Einstein metric with negative Ricci curvature. The first breakthrough in this direction was achieved by Siu's computation [1986] of the curvature of the Weil–Petersson metric on the moduli via the Kähler–Einstein metric of the fibers of the family (see also [Schumacher 2012]). To and Yeung [2015] built upon Siu's work to prove the Kobayashi hyperbolicity of moduli stacks of canonically polarized manifolds and thus gave a new proof of the Brody hyperbolicity of such moduli stacks (see also [To and Yeung 2018] for the Ricci-flat case). We also refer the reader to [Schumacher 2018, Theorem 9]. A different proof of this result has been established by Berndtsson, Păun and Wang [Berndtsson et al. 2017]. Recently, based on results we prove here and methods from the works above, Deng has extended Kobayashi hyperbolicity to effectively parametrized families of minimal manifolds of general type, and pseudo-Kobayashi hyperbolicity to families of polarized manifolds with maximal variation, in [Deng 2018a; 2018b].

To go beyond the canonically polarized case, in this paper we take a different path based on the approach of Viehweg and Zuo, where the key first step is to refine the Hodge theoretic constructions of [Viehweg and Zuo 2003] (and subsequently [Popa and Schnell 2017]), with the ultimate goal of "generically" endowing any complex line $\mathbb{C}$ in V with a metric with sufficiently negative curvature; this is the content of Section 2. The next step, presented in Section 3, is to extend this metric to a singular metric on $\mathbb{C}$ whose curvature current violates the singular Ahlfors–Schwarz inequality. A review of the line of work that has inspired this approach to hyperbolicity can be found at the end of [Viehweg and Zuo 2003, Section 1].

1C. *Two-dimensional parameter spaces in the general case.* As mentioned at the outset, the results in Theorem 1.1 and Theorem 1.2 are expected to hold for families of manifolds of lower Kodaira dimension as well, assuming that they have semiample canonical bundle or, more generally, admit a good minimal model (this last condition also includes the case of arbitrary fibers of general type).

On a related note, in [Popa and Schnell 2017, Theorem A] it is shown that the base V of any smooth family whose geometric generic fiber admits a good minimal model, and which has maximal variation, is of log general type. Thus the Green–Griffiths–Lang conjecture again predicts hyperbolicity properties for V. Note that when $\dim V = 1$, the two properties are equivalent, and had already been established in [Viehweg and Zuo 2001]. We finish the paper by establishing such results in the case when V is two-dimensional.

Theorem 1.4. *Let $f_U : U \to V$ be a smooth family of projective manifolds, with maximal variation. Assume that V is a quasi-projective surface*:

(1.4.1) *If the geometric generic fiber of f has a good minimal model, then every entire curve $\gamma : \mathbb{C} \to V$ is algebraically degenerate.*

(1.4.2) *Moreover, if the fibers are of general type, then the exceptional locus* $\mathrm{Exc}(V)$ *is a proper subset of V.*

As a consequence of Theorem 1.4, we can extend Theorem 1.1 to the case of moduli of polarized manifolds, not necessarily of general type, as long as V is two-dimensional.

Corollary 1.5. *Let V be a quasi-projective surface admitting a morphism $\sigma : V \to P_h$ induced by a smooth family $f_U : U \to V$ in $\mathscr{P}_h(V)$. If σ is quasi-finite, then V is Brody hyperbolic.*

1D. *Outline of the argument.* The paper follows the beautiful strategy towards proving hyperbolicity for parameter spaces that was developed in the series of works of Viehweg and Zuo [2001; 2002; 2003]. It relies also on the extension to Hodge modules provided in [Popa and Schnell 2017] of some Hodge-theoretic constructions in these papers, which in turns enables the level of generality we consider. Here are the key steps; in each of them we describe what is the new input needed in order to go beyond the canonically polarized case in [Viehweg and Zuo 2003].

(1) First, one constructs a special Hodge theoretic object on a compactification Y of (a birational model of) the base V, namely a graded subsheaf $(\mathscr{F}_\bullet, \theta_\bullet)$ of a Higgs bundle $(\mathscr{E}_\bullet, \theta_\bullet)$ associated to a Deligne canonical extension of a variation of Hodge structure (VHS) supported outside of a simple normal crossing divisor $D+S$, where $D = Y \setminus V$. The system $\mathscr{F}_\bullet$ encodes the data of maximal variation and has positivity properties due to general Hodge theory.

A large part of the construction follows ideas from [Popa and Schnell 2017]; a key ingredient is the use of Hodge module extensions of VHS, necessary especially when the fibers are no longer assumed to have semiample canonical bundle. A detailed discussion of the construction can be found in [Popa and Schnell 2017, Introduction and Section 2] (see also [Popa 2018] for an overview).[1] However, we make some modifications that lead to an a priori slightly different Higgs sheaf $(\mathscr{F}_\bullet, \theta_\bullet)$; the reason is that we crucially need the induced map $\mathscr{T}_Y \to \mathscr{F}_0^\vee \otimes \mathscr{F}_1$ to coincide generically with the Kodaira–Spencer map of the original family. This can be accomplished when the fibers are minimal of general type by appealing to a vanishing theorem due to Bogomolov and Sommese. (The construction for canonically polarized fibers in [Viehweg and Zuo 2003] appeals to Kodaira–Nakano vanishing, which may fail to hold in this context.) We note that this is the only point in the paper where it is necessary to work with minimal varieties of general type, and which needs to be overcome in order to answer the Viehweg–Zuo question in the arbitrary polarized case.

Given a holomorphic map $\gamma\colon \mathbb{C} \to V$, this construction eventually allows us to produce, for each $m \geq 0$, morphisms

$$\tau_m\colon \mathscr{T}_{\mathbb{C}}^{\otimes m} \to \gamma^*(\mathscr{L}^{-1} \otimes \mathscr{E}_m),$$

where $\mathscr{L}$ is a big and nef line bundle on Y, positive on V, and $\mathscr{E}_\bullet$ is the Higgs bundle mentioned above. This is all done in Section 2.

(2) For the next step, in the case of Viehweg's hyperbolicity conjecture the point was to apply a powerful criterion detecting the log general type property, due to Campana and Păun [2019]. In the present case of Brody hyperbolicity, this step is by contrast of an analytic, and in some sense more elementary flavor. Using the relationship with the Kodaira–Spencer map mentioned above, one shows that for some $m \geq 1$ the map τ_m factors through

$$\tau_m\colon \mathscr{T}_{\mathbb{C}}^{\otimes m} \to \gamma^*\mathscr{L}^{-1} \otimes \mathscr{N}_{(\gamma,m)},$$

where $\mathscr{N}_{(\gamma,m)}$ is defined as the kernel of the generalized Kodaira–Spencer map

$$\gamma^*\mathscr{E}_m \to \gamma^*\mathscr{E}_{m+1} \otimes \Omega^1_{\mathbb{C}}(P),$$

with $P = \gamma^{-1}(S)$. As in [Viehweg and Zuo 2003], we use this, together with results about the curvature of Hodge metrics, in order to construct a sufficiently negative singular metric on $\mathbb{C}$ which violates the Ahlfors–Schwarz inequality.

[1] We also take the opportunity in the Appendix to write down a reduction step to the simple normal crossings case; this was stated in [Viehweg and Zuo 2003] and [Popa and Schnell 2017] in the respective settings, but the concrete details were not included.

The relaxation of the assumption on the fibers of the family again creates technical difficulties compared to the situation in [Viehweg and Zuo 2003], where one could work with Hodge theoretic objects with finite monodromy around the components of S. We consider instead a further perturbation along S, which allows us to construct the singular metric we need using only the well-known growth estimates for Hodge metrics at the boundary given in [Schmid 1973] and [Cattani et al. 1986]. This does not require any further knowledge about the monodromy, and so has the advantage of giving a slightly simplified argument, in a more general situation. All of this is discussed in Section 3B–Section 3D.

(3) When the base V of the family is a surface, one does not need to appeal to the connection with the Kodaira–Spencer map mentioned in (1). Consequently the requirement that the fibers be minimal of general type, or even have semiample canonical bundle, can be dropped (meaning that we may assume only that the geometric generic fiber has a good minimal model), noting however that for the Hodge theoretic constructions we now necessarily have to use the more abstract Hodge module version. Instead, we follow a different approach by using the map τ_1 in order to produce a foliation on V such that $\gamma(\mathbb{C})$ is contained in one of its leaves. Given that by [Popa and Schnell 2017] we know that V is of log general type, we can then appeal to a result of McQuillan [1998] on the degeneracy of such entire curves, and to an extension to the logarithmic case in [El Goul 2003], in order to obtain a contradiction. This is the subject of Section 3E.

2. Hodge-theoretic constructions

2A. ***Relative (graded) Higgs sheaves.*** We start with a brief discussion of Higgs sheaves with logarithmic poles. We consider the relative setting, which will be necessary for technical reasons later on, though most of the time the constructions are needed in the absolute setting. Suppose X and Y are smooth quasi-projective varieties, and $f : X \to Y$ is a smooth morphism of relative dimension d, with D a reduced relative normal crossing divisor over Y.

Recall that an f-relative graded Higgs sheaf with log poles along D is a pair $(\mathscr{E}_\bullet, \theta_\bullet)$ such that:

(2.0.1) $\mathscr{E}_\bullet$ is a $\mathbb{Z}$-graded $\mathscr{O}_X$-module, with grading bounded from below.

(2.0.2) $\theta_\bullet$ is a grading-preserving $\mathscr{O}_X$-linear morphism

$$\theta_\bullet : \mathscr{E}_\bullet \to \Omega^1_{X/Y}(\log D) \otimes \mathscr{E}_{\bullet+1}$$

satisfying $\theta_\bullet \wedge \theta_\bullet = 0$, where $\Omega^1_{X/Y}(\log D)$ is the sheaf of relative 1-forms with logarithmic poles along D; it is called the Higgs field of the sheaf.

A (relative) Higgs sheaf is called a (relative) Higgs bundle if it consists of $\mathscr{O}_X$-modules that are locally free of finite rank. If f is trivial, then we get the usual notions of a Higgs sheaf or Higgs bundle. The standard example is the Hodge bundle associated to a variation of Hodge structure (VHS). More generally, for a VHS $\mathbb{V}$ on $Y \setminus D$, with quasi-unipotent monodromy along the components of D, the Deligne extension of the VHS across D with eigenvalues in $[0, 1)$ is a logarithmic VHS, i.e., the extension of the flat bundle is locally free with a flat logarithmic connection, and the extension of the filtration is a filtration by subbundles; see [Deligne 1970, Proposition I.5.4] and [Saito 1990, (3.10.5)]; see also [Kollár 1986, 2.5]. Hence its generalized Hodge bundle $(\mathscr{E}_\bullet, \theta_\bullet)$ is a logarithmic Higgs bundle.

We denote by $\mathscr{T}_{X/Y}(-\log D)$ the sheaf of relative vector fields with logarithmic zeros along D, and consider its symmetric algebra

$$\mathscr{A}_{X/Y}(-\log D) := \operatorname{Sym} \mathscr{T}_{X/Y}(-\log D)$$

(or $\mathscr{A}^\bullet_{X/Y}(-\log D)$ if we want to emphasize its grading). When $D = 0$ and f is trivial, we have $\mathscr{A}_X = \operatorname{gr}^F_\bullet \mathscr{D}_X$, where $\mathscr{D}_X$ is the sheaf of holomorphic differential operators with the order filtration. We have inclusions of graded $\mathscr{O}_X$-algebras

$$\mathscr{A}_{X/Y}(-\log D) \hookrightarrow \mathscr{A}_{X/Y} \hookrightarrow \mathscr{A}_X \quad \text{and} \quad \mathscr{A}_{X/Y}(-\log D) \hookrightarrow \mathscr{A}_X(-\log D) \hookrightarrow \mathscr{A}_X.$$

We will consider graded modules over these sheaves of rings. For instance, the associated graded of a filtered $\mathscr{D}_X$-module (resp. of a filtered vector bundle with flat connection with log poles along D) is an $\mathscr{A}^\bullet_X$ (resp. $\mathscr{A}^\bullet_X(-\log D)$)-module. The following reinterpretation of the definitions allows us to use relative Higgs sheaves and graded $\mathscr{A}_{X/Y}(-\log D)$-modules interchangeably.

Lemma 2.1. *The data of a relative Higgs sheaf $(\mathscr{E}_\bullet, \theta_\bullet)$ with log poles along D is equivalent to that of a graded $\mathscr{A}^\bullet_{X/Y}(-\log D)$-module structure on $\mathscr{E}_\bullet$, extending the $\mathscr{O}_X$-module structure.*

The Higgs field $\theta_\bullet$ induces a complex of graded $\mathscr{O}_X$-modules, de Rham complex

$$\mathrm{DR}^D_{X/Y}(\mathscr{E}_\bullet) := [\mathscr{E}_\bullet \to \Omega^1_{X/Y}(\log D) \otimes \mathscr{E}_{\bullet+1} \to \cdots \to \Omega^d_{X/Y}(\log D) \otimes \mathscr{E}_{\bullet+d}]$$

and we have

$$\mathrm{DR}^D_{X/Y}(\mathscr{E}_\bullet) \simeq \mathrm{DR}^D_{X/Y}(\mathscr{A}^\bullet_{X/Y}(-\log D)) \otimes_{\mathscr{A}^\bullet_{X/Y}(-\log D)} \mathscr{E}_\bullet.$$

Definition 2.2 (pull-back of Higgs bundles). Let $\mathscr{E}_\bullet$ be a relative Higgs bundle on X, and $\gamma\colon B \to X$ a holomorphic map from a complex manifold B, such that the support E of $\gamma^{-1}(D)$ is relative normal crossing over Y with respect to the induced map $B \to Y$. Then the natural $\mathscr{O}_X$-linear morphism $\mathscr{T}_{B/Y}(-\log E) \to \gamma^*\mathscr{T}_{X/Y}(-\log D)$ induces a morphism

$$\mathscr{A}_{B/Y}(-\log E) \to \gamma^*\mathscr{A}_{X/Y}(-\log D)$$

of graded $\mathscr{O}_B$-algebras. Therefore, $\gamma^*\mathscr{E}_\bullet$ is a graded $\mathscr{A}_B(-\log E)$-module, and in particular a relative Higgs bundle on B with Higgs field induced by that of $\mathscr{E}_\bullet$.

2B. ***Hodge modules for rank 1 unitary representations on quasi-projective varieties.*** We discuss Hodge modules for rank 1 unitary representations, needed in what follows. We fix a line bundle $\mathscr{B}$ on a smooth quasi-projective variety X, and assume that

$$\mathscr{B}^m \simeq \mathscr{O}_X(E),$$

for some $m \in \mathbb{N}$ and an effective divisor $E = \sum_i a_i D_i$ with simple normal crossing support. We denote $D = E_{\mathrm{red}}$. It is well-known that, for every $0 < i < m$ and every divisor E' supported on D, the line bundle $\mathscr{B}^{-i}(E')$ admits a flat connection with logarithmic poles along D. As in [Esnault and Viehweg 1992, Section 3], we set

$$\mathscr{B}^{(-i)} = \mathscr{B}^{-i}\Big(\sum_i \lfloor \tfrac{a_i}{m} \rfloor \cdot D_i\Big)$$

the Deligne canonical extension of $\mathscr{B}^{-i}|_{X\setminus D}$, which is a flat unitary line bundle on $X \setminus D$ coming from a unitary representation of the fundamental group. We also use the notation

$$\mathscr{B}^{-i}(*D) = \bigcup_{k\geq 0} \mathscr{B}^{-i}(kD)$$

for the sheaf of sections of $\mathscr{B}^{-i}$ with poles of arbitrary order along D. We define filtrations on $\mathscr{B}^{(-i)}$, $\mathscr{B}^{(-i)}(D)$ and $\mathscr{B}^{-i}(*D)$ by

$$F_p\mathscr{B}^{(-i)}(C) = \begin{cases} 0 & \text{if } p < 0, \\ \mathscr{B}^{(-i)}(C) & \text{if } p \geq 0, \end{cases}$$

where C is either 0 or D, and

$$F_p\mathscr{B}^{-i}(*D) = \begin{cases} 0 & \text{if } p < 0, \\ \mathscr{B}^{(-i)}((p+1)D) & \text{if } p \geq 0. \end{cases} \tag{2.2.1}$$

With these filtrations, $\mathscr{B}^{(-i)}(D)$ is a filtered line bundle with a flat connection with log poles along D, and $\mathscr{B}^{-i}(*D)$ is a filtered $\mathscr{D}_X$-module. Note that in particular we will always consider $\mathscr{O}_X$ with the trivial filtration $F_k\mathscr{O}_X = \mathscr{O}_X$ for $k \geq 0$, and 0 otherwise, so that $\mathrm{gr}^F_\bullet\, \mathscr{O}_X \simeq \mathscr{O}_X$.

By [Saito 1990, (3.10.3) and (3.10.8)], we know that $(\mathscr{B}^{-i}(*D), F_\bullet)$ is a direct summand of the filtered $\mathscr{D}_X$-module underlying $\pi_*\mathbb{Q}^H_Z[\dim Z]$, the direct image of the trivial Hodge module on Z, where $\pi : Z \to X$ is the m-th cyclic cover branched along the divisor E.

Note that $\mathrm{gr}^F_\bullet \mathscr{B}^{(-i)}(D)$ is a graded $\mathscr{A}^\bullet_X(-\log D)$-module, while $\mathrm{gr}^F_\bullet \mathscr{B}^{-i}(*D)$ is a graded $\mathscr{A}^\bullet_X$-module. Moreover, the natural inclusions

$$\mathrm{gr}^F_\bullet \mathscr{B}^{(-i)} \hookrightarrow \mathrm{gr}^F_\bullet (\mathscr{B}^{(-i)}(D)) \hookrightarrow \mathrm{gr}^F_\bullet \mathscr{B}^{-i}(*D)$$

preserve the Higgs structure. We have the following comparison result:

Proposition 2.3. *Assume that $f : X \to Y$ is a smooth projective morphism of relative dimension d between smooth quasi-projective varieties, and D is a divisor on X which is relatively normal crossing over Y. Then the natural morphism*

$$\mathrm{DR}^D_{X/Y}(\mathrm{gr}^F_\bullet \mathscr{B}^{(-i)}) \to \mathrm{DR}_{X/Y}(\mathrm{gr}^F_\bullet \mathscr{B}^{(-i)}(*D))$$

is a quasi-isomorphism of complexes of graded $\mathscr{O}_X$-modules.

Proof. The absolute case was proved in [Saito 1990, Section 3.b] in a more general setting. The relative case is similar; we sketch the proof for completeness.

We define graded sheaves $\mathscr{C}_\bullet$ and $\mathscr{N}_\bullet$ by

$$\mathscr{C}_\bullet := \mathscr{A}^\bullet_{X/Y} \otimes_{\mathrm{gr}^F_\bullet \mathscr{O}_X} \mathrm{gr}^F_\bullet(\mathscr{B}^{(-i)}(D)) \quad \text{and} \quad \mathscr{N}_\bullet := \mathscr{A}^\bullet_{X/Y} \otimes_{\mathscr{A}^\bullet_{X/Y}(-\log D)} \mathrm{gr}^F_\bullet(\mathscr{B}^{(-i)}(D)).$$

By definition $\mathscr{C}_\bullet$ is a graded $(\mathscr{A}^\bullet_{X/Y}\text{-}\mathscr{A}^\bullet_{X/Y}(-\log D))$-bimodule. (The $\mathscr{A}_{X/Y}(-\log D)$-module structure is induced by the product rule; that is, locally $x_i\partial_{x_i}\cdot(\nu\otimes l) = x_i\partial_{x_i}\cdot\nu\otimes l - \nu\otimes x_i\partial_{x_i}\cdot l$, if $\nu\otimes l$ is a section of $\mathscr{C}_\bullet$.) Assume now that $\mathscr{T}_{X/Y}(-\log D)$ is freely generated locally by

$$\partial_{x_1}, \ldots, \partial_{x_i}, x_{i+1}\partial_{x_{i+1}}, \ldots, x_d\partial_{x_d}.$$

The sequence of actions of these elements on $\mathscr{C}_\bullet$ (via the $\mathscr{A}_{X/Y}(-\log D)$-module structure described above) gives rise to a Koszul-type complex. Written in a coordinate free way, this is a complex of $\mathscr{A}^\bullet_{X/Y}$-modules

$$\mathcal{B}^\bullet_\bullet = \left[\mathscr{C}_{\bullet-d} \otimes \bigwedge^{d} \mathscr{T}_{X/Y}(-\log D) \to \mathscr{C}_{\bullet-d+1} \otimes \bigwedge^{d-1} \mathscr{T}_{X/Y}(-\log D) \to \cdots \to \mathscr{C}_\bullet\right].$$

Using the fact that $\mathrm{gr}^F_\bullet(\mathscr{L}^{(-i)}(D))$ is locally free of rank 1 over $\mathrm{gr}^F_\bullet \mathscr{O}_X$, one can check that this sequence is regular; therefore, the natural morphism

$$\mathcal{B}^\bullet_\bullet \to H^0\mathcal{B}^\bullet_\bullet = \frac{\mathscr{C}_\bullet}{\sum_{j=1}^{i}\partial_{x_j}\mathscr{C}_\bullet + \sum_{j=i+1}^{d} x_j\partial_{x_j}\mathscr{C}_\bullet} \simeq \mathscr{N}_\bullet$$

is a quasi-isomorphism of complexes of graded $\mathscr{A}^\bullet_{X/Y}$-modules. The exactness of the de Rham functor implies that the induced morphism

$$\mathrm{DR}_{X/Y}(\mathcal{B}^\bullet_\bullet) \to \mathrm{DR}_{X/Y}(\mathscr{N}_\bullet)$$

is a quasi-isomorphism as well. Moreover, one also sees that the natural morphism

$$\mathrm{DR}_{X/Y}\left(\mathscr{C}_{\bullet-d+p} \otimes \bigwedge^{d-p} \mathscr{T}_{X/Y}(-\log D)\right) \to \mathrm{gr}^F_{\bullet+p}\mathscr{B}^{(-i)} \otimes \Omega^p_{X/Y}(\log\ D)$$

is a quasi-isomorphism, thanks to the natural isomorphism given by contraction

$$\omega_{X/Y}(D) \otimes \bigwedge^{d-p} \mathscr{T}_{X/Y}(-\log D) \simeq \Omega^p_{X/Y}(\log\ D),$$

and the fact that $\mathrm{DR}_{X/Y}(\mathscr{A}^\bullet_{X/Y})$ is quasi-isomorphic to $\omega_{X/Y}$. Therefore, we find that $\mathrm{DR}_{X/Y}(\mathcal{B}^\bullet_\bullet)$ and $\mathrm{DR}^D_{X/Y}(\mathrm{gr}^F_\bullet\mathscr{B}^{(-i)})$ are quasi-isomorphic. We now conclude by noting that there is an isomorphism of $\mathscr{A}^\bullet_{X/Y}$-modules

$$\mathscr{N}_\bullet \simeq \mathrm{gr}^F_\bullet \mathscr{L}^{-i}(*D);$$

see for instance [Björk 1993, Proposition 4.2.18] (where it is stated locally, for more general $\mathscr{D}$-modules).

$\square$

2C. ***Hodge modules and branched coverings.*** This section is essentially a review of the constructions in [Popa and Schnell 2017, Sections 2.3 and 2.4], but with a twist which is important for the applications in this paper. We assume that we have a morphism of smooth projective varieties $f\colon X \to Y$, with connected fibers, and with $\dim Y = n$ and $\dim X = n + d$. Let $\mathscr{A}$ be a line bundle on Y, and define

$$\mathscr{B} := \omega_{X/Y} \otimes f^*\mathscr{A}^{-1}.$$

We make the following assumption:

$$\text{there exists } 0 \neq s \in H^0(X, \mathscr{B}^m) \text{ for some } m > 0. \tag{2.3.1}$$

The section s defines a branched cover $\psi\colon X_m \to X$ of degree m. Let $\delta\colon Z \to X_m$ be a desingularization of the normalization of X_m, which is irreducible if m is chosen to be minimal, and set $\pi = \psi \circ \delta$ and $h = f \circ \pi$, as in the diagram

$$\begin{tikzcd} Z \arrow[r, "\delta"] \arrow[rr, bend left, "\pi"] \arrow[drr, "h"'] & X_m \arrow[r, "\psi"] \arrow[dr] & X \arrow[d, "f"] \\ & & Y \end{tikzcd}$$

Let $\mathscr{A}_Y = \operatorname{Sym} \mathscr{T}_Y$, with the natural grading, and similarly for $\mathscr{A}_X$. A morphism of graded $\mathscr{A}_Y$-modules

$$\boldsymbol{R} f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y) \to \boldsymbol{R} h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y). \tag{2.3.2}$$

is constructed in [Popa and Schnell 2017, Section 2.4]. (We use the notation $\mathscr{B}^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y$ as shorthand for $\mathscr{B}^{-1} \otimes_{\mathscr{O}_X} \operatorname{gr}^F_\bullet \mathscr{O}_X \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y$, where we make use of the $\mathscr{A}_X$-module structure on $\operatorname{gr}^F_\bullet \mathscr{O}_X \simeq \mathscr{O}_X$.) Here we will construct a similar but slightly different morphism, that a priori coincides with the one in (2.3.2) only generically. The reason for this different construction will become clear in Section 2F, where we need to compare Hodge sheaves constructed out of branched coverings with others that are naturally related to Kodaira–Spencer maps.

Before starting the construction, recall from [Saito 1990, Section 3.b] that $\mathscr{B}$ uniquely determines a filtered $\mathscr{D}_X$-module $(\mathscr{B}_*^{-1}, F_\bullet)$ with strict support X, which extends $(\mathscr{B}^{-1}|_{X\setminus \operatorname{div}(s)}, F_\bullet)$, where the filtration on the latter is the trivial filtration; notice that the filtered $\mathscr{D}_X$-module is exactly $(\mathscr{B}^{-1}(*D), F_\bullet)$, when $D = \operatorname{div}(s)_{\mathrm{red}}$ is normal crossing. Moreover, $(\mathscr{B}_*^{-1}, F_\bullet)$ is a direct summand of the filtered $\mathscr{D}_X$-module $\mathcal{H}^0\pi_+(\mathscr{O}_Z, F_\bullet)$.

Lemma 2.4. *We have a natural inclusion*

$$\mathscr{B}^{-1} \hookrightarrow F_0\, \mathscr{B}_*^{-1}.$$

Proof. Let $\mu\colon X' \to X$ be a log resolution of the divisor $\operatorname{div}(s)$ which is an isomorphism on its complement. Define $D' = (\mu^* \operatorname{div}(s))_{\mathrm{red}}$ and $\mathscr{B}' = \mu^*\mathscr{B}$. Then, according to the discussion in Section 2B,

$(\mathscr{B}'^{-1}(*D'), F_\bullet)$ defined as in (2.2.1) is a direct summand of a $\mathscr{D}_{X'}$-module underlying a Hodge module. By the strictness of the direct image functor for Hodge modules, we have

$$\mu_+(\mathscr{B}'^{-1}(*D'), F_\bullet) = (\mathscr{B}_*^{-1}, F_\bullet) \quad \text{and} \quad \mu_* F_0 \mathscr{B}'^{-1}(*D') = F_0 \mathscr{B}_*^{-1}.$$

On the other hand, by construction we have the injection

$$\mathscr{B}'^{-1} \subset F_0 \mathscr{B}'^{-1}(*D'),$$

and so the statement follows from the projection formula. □

We now proceed with our construction. The inclusion in Lemma 2.4 induces a morphism of graded $\mathscr{A}_X$-modules

$$\mathscr{B}^{-1} \to \operatorname{gr}_\bullet^F \mathscr{B}_*^{-1},$$

with the trivial graded $\mathscr{A}_X$-module structure on $\mathscr{B}^{-1}$. This in turn induces a morphism

$$\boldsymbol{R} f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y) \to \boldsymbol{R} f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \operatorname{gr}_\bullet^F \mathscr{B}_*^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y).$$

Now the right-hand side is a direct summand of the object $\boldsymbol{R}h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y)$; indeed, using [Popa and Schnell 2013, Theorem 2.9], we have an isomorphism

$$\operatorname{gr}_\bullet^F h_+(\mathscr{O}_Z, F_\bullet) \simeq \boldsymbol{R}h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y),$$

and we combine this with the filtered direct summand inclusion of $(\mathscr{B}_*^{-1}, F_\bullet)$ in $\mathcal{H}^0\pi_+(\mathscr{O}_Z, F_\bullet)$. Therefore we get an induced morphism

$$\boldsymbol{R} f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y) \to \boldsymbol{R}h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y), \tag{2.4.1}$$

which factors through $\boldsymbol{R} f_*(\omega_{X/Y} \otimes \operatorname{gr}_\bullet^F \mathscr{B}_*^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y)$. One can check that the morphisms (2.3.2) and (2.4.1) coincide over the locus where h is smooth; they are however not necessarily the same globally.

Let now $(\mathcal{M}, F_\bullet)$ be the filtered $\mathscr{D}_Y$-module underlying the Tate twist $M(d)$ of the pure polarizable Hodge module M which is the direct summand of $\mathcal{H}^0 h_* \boldsymbol{Q}_Z^H[n+d]$ strictly supported on Y. By [Popa and Schnell 2017, Proposition 2.4], we then have that $\operatorname{gr}_\bullet^F \mathcal{M}$ is a direct summand of $R^0 h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y)$.

Definition 2.5. We define a graded $\mathscr{A}_Y$-module $\mathscr{G}_\bullet$ as the image of the composition

$$R^0 f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y) \to R^0 h_*(\omega_{Z/Y} \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_Z} h^*\mathscr{A}_Y) \to \operatorname{gr}_\bullet^F \mathcal{M},$$

where the second morphism is given by projection.

Recall that D_f denotes the singular locus of f. We gather the constructions above and further properties in the following result, which is essentially [Popa and Schnell 2017, Theorem 2.2]; although as pointed out above the new morphism (2.4.1) is constructed slightly differently, the proof is identical.

Theorem 2.6. *With the above notation, assuming (2.3.1), the coherent graded $\mathscr{A}_Y$-module $\mathscr{G}_\bullet$ satisfies the following properties*:

(2.6.1) *There is an isomorphism $\mathscr{G}_0 \simeq \mathscr{A}$.*

(2.6.2) *Each $\mathscr{G}_k$ is torsion-free on $X \setminus D_f$.*

(2.6.3) *There is an inclusion of graded $\mathscr{A}_Y$-modules $\mathscr{G}_\bullet \subseteq \mathrm{gr}_\bullet^F \mathcal{M}$.*

2D. *Basic set-up.* We consider a smooth family $f_U : U \to V$ of projective varieties, whose geometric generic fiber admits a good minimal model. (This includes for instance families of varieties of general type, or of varieties whose canonical bundle is semiample.) We assume that the family has maximal variation; following the strategy in [Viehweg and Zuo 2003], together with the technical extensions in [Popa and Schnell 2017], our aim in the next two sections is to endow entire curves inside (a birational model of) V with Hodge theoretic objects that will be later used in order to conclude hyperbolicity.

In order to accomplish this, we will use a technical statement about the existence of sections (or the generic global generation) for suitable line bundles on a modification of the family f_U. This is proved in the Appendix in Propositions A.1 and A.4. The idea and most of the details can be found in [Viehweg and Zuo 2003]; for a detailed discussion please see the Appendix.

2E. *Main construction on* $\mathbb{C}$. In the set-up of Section 2D, our aim here is to use the constructions in the previous sections in order to produce interesting Hodge-theoretic sheaves on $\mathbb{C}$, assuming the existence of a holomorphic mapping $\gamma : \mathbb{C} \to V$.

Assumption. All VHS appearing in this paper are assumed to be polarizable, and all local monodromies to be quasi-unipotent; see for instance [Schmid 1973] for the definitions. This is of course the case for any geometric VHS, i.e., the Gauss–Manin connection of a smooth family of projective manifolds, thanks to the monodromy theorem (see for instance [Schmid 1973, Lemma 4.5]). In general, fixing a polarization induces the Hodge metric on the associated Higgs bundle, its singularities at the boundary will play a crucial role in Section 3B.

We start with the key output of the Hodge theoretic constructions above, following arguments in [Popa and Schnell 2017]. According to the strategy in [Viehweg and Zuo 2003], it will later be combined with analytic arguments in order to conclude the nonexistence of dense entire curves.

Proposition 2.7. *Let $f_U : U \to V$ be a smooth family of projective varieties, with maximal variation, and whose geometric generic fiber has a good minimal model. Then, after possibly replacing V by a birational model, there exists a smooth projective compactification Y of V, with $D = Y \setminus V$ a simple normal crossing divisor, together with a big and nef line bundle $\mathscr{L}$ and an inclusion of graded $\mathscr{A}_Y^\bullet(-\log D)$-modules*

$$(\mathscr{F}_\bullet, \theta_\bullet) \subseteq (\mathscr{E}_\bullet, \theta_\bullet),$$

on Y, that verify the following properties:

(2.7.1) $(\mathscr{E}_\bullet, \theta_\bullet)$ *is the Higgs bundle underlying the Deligne extension with eigenvalues in* $[0, 1)$ *of a VHS defined outside of a simple normal crossing divisor* $D + S$.

(2.7.2) $\mathscr{F}_0$ *is a line bundle, and we have an inclusion* $\mathscr{L} \subseteq \mathscr{F}_0$ *which is an isomorphism on* V.

(2.7.3) *If* $\gamma\colon \mathbb{C} \to V \subseteq Y$ *is a holomorphic map, then for each* $k \geq 0$ *there exists a morphism*

$$\tau_{(\gamma,k)} : \mathscr{T}_{\mathbb{C}}^{\otimes k} \to \gamma^*\left(\bigotimes^k \mathscr{T}_Y(-\log D)\right) \to \gamma^*({\mathscr{F}_0}^{-1} \otimes \mathscr{E}_k) \to \gamma^*(\mathscr{L}^{-1} \otimes \mathscr{E}_k).$$

Proof. We consider $f\colon X \to Y$ as in Proposition A.1 in the Appendix.[2] Thus there exist an integer $m > 0$ and a line bundle $\mathscr{A}$ on Y, of the form $\mathscr{A} = \mathscr{L}(D_Y)$ with $\mathscr{L}$ ample and $D_Y \geq D$, such that

$$H^0(X, (\omega_{X/Y} \otimes f^*\mathscr{A}^{-1})^m) \neq 0.$$

This means that we can apply the constructions in Section 2C; we set

$$\mathscr{B} = \omega_{X/Y} \otimes f^*\mathscr{A}^{-1}$$

and pick $0 \neq s \in H^0(X, \mathscr{B}^m)$. Associated to this section, by applying Theorem 2.6, we obtain a Hodge sheaf $\mathscr{G}_\bullet$ and a Hodge module M on Y.

For the next construction, we would like to assume that there is an effective divisor S on Y such that the singular locus of M is (contained in) $D + S$, and that $D + S$ has simple normal crossings. In fact, and this is sufficient, we can only accomplish this outside of a closed subset of codimension at least 2, as follows. We consider a further birational model $\tilde{f}\colon \tilde{X} \to \tilde{Y}$ as in Proposition A.4, imposing that the singular locus S of M contain the branch locus Δ_τ in that statement. Using the notation $\mu\colon \tilde{Y} \to Y$ for the birational map on the base, we obtain that there exists a closed subset T in $\tilde{Y}$, of codimension at least 2, such that s induces a new section

$$\tilde{s} \in H^0(\tilde{X}_0, (\omega_{\tilde{X}/\tilde{Y}} \otimes \tilde{f}^*\mu^*\mathscr{A}^{-1})^m),$$

with $\tilde{Y}_0 = \tilde{Y} \setminus T$ and $\tilde{X}_0 = \tilde{f}^{-1}(\tilde{Y}_0)$, such that $\tilde{s}$ and s coincide on the fibers over points away from S, where μ is the identity map.

We again record the conclusion of Theorem 2.6 for the new family, over $\tilde{Y}_0$ only. Note that since the sections coincide away from S, the new pure Hodge module is the unique extension with strict support of the same VHS as M, defined on the complement of $D + S$. We can therefore revert to the original notation, and assume that we have a Hodge module M and a Hodge subsheaf $\mathscr{G}_\bullet$ on Y, such that on an open subset Y_0 with complement of codimension at least 2 they coincide with those constructed as above from the section s, and in addition the divisor $D + S$ (and in particular the singular locus of M) has simple normal crossings. Note that because of the birational modification, $\mathscr{L}$ is now only a big and nef line bundle.

[2]Unlike in the Appendix, here we denote the original family $U \to V$, and we keep this notation after passing to a birational model, since there is no danger of confusion.

We now take $(\mathscr{E}_\bullet, \theta_\bullet)$ to be the Higgs bundle on Y underlying the Deligne extension with eigenvalues in $[0, 1)$ of the VHS that coincides with M outside of $D+S$. Following [Popa and Schnell 2017, Sections 2.7 and 2.8], on the open set Y_0 we define a subsheaf $(\mathscr{F}_\bullet, \theta_\bullet)$ of $(\mathscr{E}_\bullet, \theta_\bullet)$ by

$$\mathscr{F}_\bullet = (\mathscr{G}_\bullet \cap \mathscr{E}_\bullet)^{\vee\vee}.$$

Note that the intersection makes sense, since both $\mathscr{G}_\bullet$ and $\mathscr{E}_\bullet$ are contained in $\mathrm{gr}_\bullet^F \mathcal{M}$. Precisely as in [Popa and Schnell 2017, Propositions 2.14 and 2.15], on Y_0 one has the following properties for $\mathscr{F}_\bullet$:

(2.7.4) We have $\mathscr{A}(-D) \subseteq \mathscr{F}_0 \subseteq \mathscr{A}$, for some integer $l > 0$.

(2.7.5) The Higgs field θ maps $\mathscr{F}_k$ into $\Omega_Y^1(\log D) \otimes \mathscr{F}_{k+1}$.

Now since the complement of Y_0 has codimension at least 2, the sheaves $\mathscr{F}_k$ have a unique reflexive extension to the entire Y. As all the other sheaves appearing in them are locally free, the maps in (2.7.4) and (2.7.5) extend uniquely as well, and hence both properties continue to hold on Y. This realizes the global construction.

Note that $\mathscr{F}_0$ is a reflexive sheaf of rank 1 on the smooth variety Y, and hence is a line bundle. Thus (2.7.4) shows (2.7.2), while (2.7.5) leads to (2.7.3) by the following construction. Note that (2.7.5) means $\mathscr{F}_\bullet$ is an $\mathscr{A}_Y(-\log D)$-module. The $\mathscr{A}_Y(-\log D)$-module structure induces a map

$$\rho_k : \bigotimes^k \mathscr{T}_Y(-\log D) \to \mathrm{Sym}^k \mathscr{T}_Y(-\log D) \to \mathscr{F}_0{}^{-1} \otimes \mathscr{F}_k \to \mathscr{F}_0{}^{-1} \otimes \mathscr{E}_k.$$

By composing ρ_k with the k-th tensor power of the differential

$$d\gamma : \mathscr{T}_\mathbb{C} \to \gamma^* \mathscr{T}_Y(-\log D),$$

we obtain

$$\tau_{(\gamma,k)} : \mathscr{T}_\mathbb{C}^{\otimes k} \xrightarrow{d\gamma^{\otimes k}} \gamma^* \Big(\bigotimes^k \mathscr{T}_Y(-\log D) \Big) \xrightarrow{\gamma^* \rho_k} \gamma^* (\mathscr{F}_0{}^{-1} \otimes \mathscr{E}_k) \hookrightarrow \gamma^* (\mathscr{L}^{-1} \otimes \mathscr{E}_k),$$

where the last morphism is induced by the inclusion of $\mathscr{L}$ into $\mathscr{F}_0$. □

Remark 2.8. If $f_U : U \to V$ has fibers with semiample canonical bundle, then by Proposition A.1 we may also assume that $\mathscr{B}^m$ is globally generated over $f^{-1}(V)$ in the result above. This will be used in the next section.

Finally, we record a fact that will be of use later on.

Lemma 2.9. *In the notation of Proposition 2.7, the Higgs map*

$$\theta_0 : \mathscr{F}_0 \to \mathscr{F}_1 \otimes \Omega_Y^1(\log D)$$

is injective.

Proof. It suffices to show that θ_0 is not the zero map, since $\mathscr{F}_0$ is a line bundle and $\mathscr{F}_1$ is torsion-free. By (2.7.2), we know that $\mathscr{F}_0$ is a big line bundle. On the other hand, if θ_0 were identically zero, then we would have that $\mathscr{F}_0 \subseteq K_0$, where

$$K_0 := \ker(\theta_0 \colon \mathscr{E}_0 \to \mathscr{E}_1 \otimes \Omega^1_Y(\log D + S)).$$

Now $K_0^\vee$ is a weakly positive sheaf by [Popa and Wu 2016, Theorem 4.8] (an easy consequence of the results of [2000] and [2018] in the unipotent case), so this would imply that $\mathscr{F}_0^{-1}$ is also a pseudoeffective line bundle, a contradiction. $\square$

2F. ***Further refinements for families of minimal manifolds of general type.*** In the current section, assuming that the members of the family are minimal and of general type, we will establish a connection between the sheaf $(\mathscr{F}, \theta)$ defined in Proposition 2.7 and the Kodaira–Spencer map of f. In the canonically polarized case treated in [Viehweg and Zuo 2003], an analogous statement is proved as an application of the Akizuki–Nakano vanishing theorem, which in the present context is not available any more; we will be able to achieve this using a different argument based on transversality and a more restrictive vanishing theorem due to Bogomolov and Sommese.

We continue to be in the set-up of Section 2D, and we fix the morphism $f \colon X \to Y$ as in the proof of Proposition 2.7. We define a new graded $\mathscr{A}_Y$-module $\widetilde{\mathscr{F}}_\bullet$ by

$$\widetilde{\mathscr{F}}_\bullet = R^0 f_*(\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \otimes \mathrm{gr}^F_\bullet \mathscr{O}_X \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y), \tag{2.9.1}$$

i.e., the left-hand side of (2.4.1), where the $\mathscr{A}_Y$-module structure is induced by the $f^*\mathscr{A}_Y$-module structure on $\omega_{X/Y} \otimes_{\mathscr{O}_X} \mathscr{B}^{-1} \otimes \mathrm{gr}^F_\bullet \mathscr{O}_X \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y$. This structure induces a morphism

$$\mathscr{T}_Y \to \widetilde{\mathscr{F}}_0^{-1} \otimes \widetilde{\mathscr{F}}_1. \tag{2.9.2}$$

Also, by the projection formula, we have $\widetilde{\mathscr{F}}_0 = \mathscr{A}$.

On the other hand, over the locus where f is smooth, using the fact that the natural morphism

$$\left[\mathscr{A}_X^{\bullet-d} \otimes \bigwedge^d \mathscr{T}_{X/Y} \to \mathscr{A}_X^{\bullet-d+1} \otimes \bigwedge^{d-1} \mathscr{T}_{X/Y} \to \cdots \to \mathscr{A}_X^\bullet\right] \to f^*\mathscr{A}_Y^\bullet \tag{2.9.3}$$

induced by the natural mapping $\mathscr{T}_X \to f^*\mathscr{T}_Y$ is a quasi-isomorphism of complexes of graded $\mathscr{A}_X$-modules (see for example [Pham 1979, Lemma 14.3.5]),[3] we know that

$$\widetilde{\mathscr{F}}_\bullet \simeq R^0 f_*(f^*\mathscr{A} \otimes \mathrm{gr}^F_\bullet \mathscr{O}_X \otimes_{\mathscr{A}_X} \mathrm{DR}_{X/Y}(\mathscr{A}_X^\bullet)).$$

In particular, over this locus we have

$$\widetilde{\mathscr{F}}_1 \simeq \mathscr{A} \otimes R^1 f_* \mathscr{T}_{X/Y}.$$

Therefore, by construction we obtain:

[3] In [loc. cit.] it is stated for $\mathscr{D}_X$ and $\mathscr{D}_Y$ respectively, as opposed to their associated graded objects.

Lemma 2.10. *Over $V = Y \setminus D$, the morphism (2.9.2) is precisely the Kodaira–Spencer map*

$$\mathscr{T}_Y \to R^1 f_* \mathscr{T}_{X/Y}.$$

Consequently, in order to establish a connection between the Kodaira–Spencer map and $(\mathscr{F}_\bullet, \theta_\bullet)$ in Proposition 2.7, it suffices to establish one between $(\widetilde{\mathscr{F}}_\bullet, \tilde{\theta}_\bullet)$ and $(\mathscr{F}_\bullet, \theta_\bullet)$. This follows immediately from the next result.

Proposition 2.11. *For $k \le 1$, the natural morphism*

$$\widetilde{\mathscr{F}}_k \to \mathscr{G}_k$$

is generically an isomorphism.

Proof. For $k = 0$, the statement follows from by Theorem 2.6(2.6.1). We now focus on the $k = 1$ case. By the basepoint-free theorem, the fibers have semiample canonical bundle, hence the second part of Proposition A.1 applies and so $\mathscr{B}^m$ is generated by global sections over $f^{-1}(V)$. Replacing Y by V, after shrinking it further if necessary, and X by $\mu^{-1}(V)$, we can assume that $f|_H \colon H \to Y$, f and h are smooth and $\mathscr{B}^m \simeq \mathscr{O}_X(H)$ is globally generated, where H is a smooth divisor transversal to the fibers. Here h is the morphism defined in Section 2C by the resolution of the branched covering associated to the global section defining H. Since h is smooth, we have $\mathcal{H}^0 h_* \boldsymbol{Q}^H_Z[n+d] = M(-d)$ and so it is enough to show that the morphism

$$\widetilde{\mathscr{F}}_1 \to R^0 f_*(\omega_{X/Y} \otimes \mathrm{gr}^F_\bullet \mathscr{B}^{-1}_* \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y)_1 \tag{2.11.1}$$

defined in Section 2C is injective.

On the other hand, as f is smooth, as we have seen above we have

$$\widetilde{\mathscr{F}}_\bullet \simeq R^0 f_*(\mathscr{B}^{-1} \otimes \mathrm{gr}^F_\bullet \mathscr{O}_X \otimes_{\mathscr{A}_X} \mathrm{DR}_{X/Y}(\mathscr{A}^\bullet_X))$$

In particular, since $\mathscr{B}^{-1} = \mathscr{B}^{(-1)}$ we have

$$\widetilde{\mathscr{F}}_1 \simeq R^0 f_*(\mathscr{B}^{-1} \otimes \mathrm{gr}^F_\bullet \mathscr{O}_X \otimes_{\mathscr{A}_X} \mathrm{DR}_{X/Y}(\mathscr{A}^\bullet_X))_1 \simeq R^1 f_*(\mathscr{B}^{-1} \otimes \Omega^{d-1}_{X/Y}).$$

Moreover, since H is smooth (so that $\mathscr{B}^{-1}_*$ is the same as $\mathscr{B}^{-1}(*D)$) and transversal to the fibers, according to (2.9.3) and Proposition 2.3, we also have

$$R^0 f_*(\omega_{X/Y} \otimes \mathrm{gr}^F_\bullet \mathscr{B}^{-1}_* \overset{\boldsymbol{L}}{\otimes}_{\mathscr{A}_X} f^*\mathscr{A}_Y)_1 \simeq R^1 f_*(\mathscr{B}^{-1} \otimes \Omega^{d-1}_{X/Y}(\log H)).$$

It follows that the morphism in (2.11.1) is induced by the first map of the following short exact sequence

$$0 \to \Omega^{d-1}_{X/Y} \to \Omega^{d-1}_{X/Y}(\log H) \to \Omega^{d-2}_{H/Y} \to 0.$$

Notice that

$$\mathscr{B}|_F \simeq \omega_F$$

on each fiber F of f. Since ω_F is big and nef, by calculating the top self-intersection number we see that $\mathscr{B}|_{F_H}$ is big on the general fiber F_H of $f|_H$ for general $H \in |\mathscr{B}^m|$. Then, according to the Bogomolov–Sommese vanishing theorem (see for instance [Esnault and Viehweg 1992, Corollary 6.9]), we know that

$$f|_{H*}(\mathscr{B}^{-1} \otimes \Omega^{d-2}_{H/Y}) = 0$$

generically, and hence everywhere since it is torsion-free. Therefore, we get the desired injectivity for the morphism in (2.11.1), and this finishes the proof of the proposition. □

Corollary 2.12. *In the situation of Proposition 2.7, if we further assume that the fibers of f_U are minimal and of general type, then the natural morphism induced by the $\mathscr{A}_Y(-\log D)$-module structure*

$$\mathscr{T}_Y(-\log D) \to \mathscr{F}_0{}^{-1} \otimes \mathscr{F}_1$$

coincides with the Kodaira–Spencer map of f over a Zariski open subset of V.

Proof. Thanks to Proposition 2.11, we know that the sheaves $\widetilde{\mathscr{F}}_k$ and $\mathscr{G}_k$ are generically isomorphic for $k = 0, 1$. On the other hand, $\mathscr{F}_\bullet$ and $\mathscr{G}_\bullet$ are generically the same by construction. Therefore, $\mathscr{F}_k$ and $\widetilde{\mathscr{F}}_k$ are generically isomorphic for $k = 0, 1$. But Lemma 2.10 says that the morphism $\mathscr{T}_Y \to \widetilde{\mathscr{F}}_0^{-1} \otimes \widetilde{\mathscr{F}}_1$ coincides with the Kodaira–Spencer map of f over V, which proves the claim. □

3. Hyperbolicity properties of base spaces of families

In this final part we establish the two main results of this paper, Theorem 1.2 (and implicitly Theorem 1.1) and Theorem 1.4. Besides Proposition 2.7 and Corollary 2.12, the main ingredient in the proofs of these theorems is Proposition 3.5 below.

3A. ***Preliminaries on singular metrics on line bundles, and on Hodge metrics.*** We start with a construction and analysis of particular singular metrics on line bundles that will be of use later on. This follows very closely the material in [Viehweg and Zuo 2003, pages 136–139]. Nevertheless we include the details for later reference, and we also make a distinction between the boundary divisors D and S, as the perturbation along S will later allow us to bypass monodromy arguments in [Viehweg and Zuo 2003] in order to extend the range of applicability.

We note to begin with that a priori by a singular metric on a line bundle $\mathscr{L}$ we mean, as in [Hacon et al. 2018, Section 13], a metric h given by a weight function $e^{-\varphi}$, where φ is taken to only be a measurable function with values in $[-\infty, \infty]$. In this way, the notion is compatible with that of a singular metric on a vector bundle, in the sense of Berndtsson, Păun and Takayama (see e.g., [Hacon et al. 2018, Section 17]), which will also make an appearance later on. In the line bundle case, usually it is also required that φ be locally integrable, in which case one can talk about its curvature form as a $(1, 1)$-current; for this we use the standard notation

$$F(\mathscr{L}, h) = \frac{\sqrt{-1}}{\pi}\,\partial\bar{\partial}\varphi = -\frac{\sqrt{-1}}{2\pi}\,\partial\bar{\partial}\log\|e\|_h^2,$$

where e is a holomorphic section which trivializes $\mathscr{L}$ locally.

Let $(Y, D+S)$ be a pair consisting of a smooth projective variety Y and simple normal crossings divisors $D = D_1 + \cdots + D_k$ and $S = S_1 + \cdots + S_\ell$. For $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \ell\}$ pick

$$f_{D_i} \in H^0(Y, \mathscr{O}_Y(D_i)), \quad f_{S_j} \in H^0(Y, \mathscr{O}_Y(S_j))$$

such that $D_i = (f_{D_i} = 0)$ and $S_j = (f_{S_j} = 0)$. For each i, j, let g_{D_i}, g_{S_j} be smooth metrics on $\mathscr{O}_Y(D_i)$ and $\mathscr{O}_Y(S_i)$, respectively; after rescaling, we may assume $\| f_{D_i} \|_{g_{D_i}} < 1$ and $\| f_{S_j} \|_{g_{S_j}} < 1$.

Now, for each i and j, set

$$r_{D_i} = -\log \| f_{D_i} \|^2_{g_{D_i}}, \quad r_{S_j} = -\log \| f_{S_j} \|^2_{g_{S_j}},$$

and define

$$r_D := \prod_i r_{D_i} \quad \text{and} \quad r_S := \prod_j r_{S_j}.$$

The functions r_D^α and $\log r_D$ (resp. r_S^α and $\log r_S$) are locally L^1 on Y for all $\alpha \in \mathbb{Z}$. Indeed, if we write locally $f_{D_i} = z_i \cdot \tilde{s}_i$ (resp. $f_{S_j} = z_j \cdot \tilde{s}_j$), where $\tilde{s}_i$ (resp. $\tilde{s}_j$) trivializes $\mathscr{O}_Y(D_i)$ (resp. $\mathscr{O}_Y(S_j)$) and z_i (resp. z_j) is a coordinate, then locally

$$r_{D_i} = -\log |z_i|^2 - \log \| \tilde{s}_i \|^2_{g_{D_i}},$$

and similarly for r_{S_j}. When $\alpha < 0$, r_D^α and r_S^α are bounded and hence continuous on Y (and smooth outside of D, resp. S).

We now fix an ample line bundle $\mathscr{L}$ on Y with a smooth hermitian metric g, so that its curvature $F(\mathscr{L}, g)$ is positive. The metric g induces a hermitian metric g^{-1} on $\mathscr{L}^{-1}$. For $\alpha \in \mathbb{N}$, we define

$$g_\alpha = g \cdot (r_D \cdot r_S)^\alpha$$

to be a singular metric on $\mathscr{L}$. There is an induced singular metric $g_\alpha^{-1} = g^{-1} \cdot (r_D \cdot r_S)^{-\alpha}$ on $\mathscr{L}^{-1}$. With this notation, we have

$$\begin{aligned} F(\mathscr{L}, g_\alpha) = F(\mathscr{L}, g) - \alpha \cdot \sum_i r_{D_i}^{-1} \cdot F(\mathscr{O}_Y(D_i), g_{D_i}) - \alpha \cdot \sum_j r_{S_j}^{-1} \cdot F(\mathscr{O}_Y(S_j), g_{S_j}) \\ + \alpha \frac{\sqrt{-1}}{2\pi} \sum_i r_{D_i}^{-2} \cdot \partial r_{D_i} \wedge \bar{\partial} r_{D_i} + \alpha \frac{\sqrt{-1}}{2\pi} \sum_j r_{S_j}^{-2} \cdot \partial r_{S_j} \wedge \bar{\partial} r_{S_j}. \quad (3.0.1) \end{aligned}$$

Next we define a continuous $(1, 1)$-form η_α on Y by the formula

$$\eta_\alpha := F(\mathscr{L}, g) - \alpha \cdot \sum_i r_{D_i}^{-1} \cdot F(\mathscr{O}_Y(D_i), g_{D_i}) - \alpha \cdot \sum_j r_{S_j}^{-1} \cdot F(\mathscr{O}_Y(S_j), g_{S_j}),$$

where we use the fact that r_D^{-1} and r_S^{-1} are continuous on Y. As Y is compact, after rescaling f_{D_i} and f_{S_j}, we can arrange for the contributions of the last two terms in η_α to be sufficiently small for η_α to be a continuous and positive definite $(1, 1)$-form on Y. On the other hand, one can easily check that

$$\frac{\sqrt{-1}}{2\pi} \partial r_{D_i} \wedge \bar{\partial} r_{D_i} \quad \text{and} \quad \frac{\sqrt{-1}}{2\pi} \partial r_{S_j} \wedge \bar{\partial} r_{S_j}$$

are smooth and semi-positive $(1, 1)$-forms on $Y \setminus D_i$ (resp. $Y \setminus S_j$) for all i and j.

Lemma 3.1. *In the above setting, for each $\alpha \in \mathbb{N}$, after rescaling f_{D_i} and f_{S_i} there is a continuous, positive definite, hermitian form w_α on $\mathscr{T}_Y(-\log D)$ such that*

$$F(\mathscr{L}, g_\alpha)|_{Y\setminus(D+S)} \geq r_D^{-2} \cdot w_\alpha|_{Y\setminus(D+S)}.$$

Proof. After suitable rescaling, we may assume that $r_{D_i} \geq 1$ for all i. Since η_α is positive on Y, and $\sqrt{-1}/(2\pi)\partial r_{D_i} \wedge \bar{\partial} r_{D_i}$ (resp. $\sqrt{-1}/(2\pi)\partial r_{S_j} \wedge \bar{\partial} r_{S_j}$) are semi-positive on $Y \setminus D_i$ (resp. $Y \setminus S_j$), using (3.0.1) we obtain

$$\begin{aligned} F(\mathscr{L}, g_\alpha)|_{Y\setminus(D+S)} &\geq \left(\eta_\alpha + \alpha\frac{\sqrt{-1}}{2\pi}\sum r_{D_i}^{-2}\cdot\partial r_{D_i}\wedge\bar{\partial} r_{D_i}\right)\Big|_{Y\setminus(D+S)} \\ &\geq r_D^{-2}\cdot\underbrace{\left(\eta_\alpha + \alpha\frac{\sqrt{-1}}{2\pi}\sum \partial r_{D_i}\wedge\bar{\partial} r_{D_i}\right)}_{w_\alpha}\Big|_{Y\setminus(D+S)}. \end{aligned}$$

Now, the claim that w_α is positive-definite on $\mathscr{T}_Y(-\log D)$ follows from the fact that η_α is positive on $\mathscr{T}_Y$ and that $\sum \partial r_{D_i} \wedge \bar{\partial} r_{D_i}$ is positive definite along the vector fields that are tangent to D [Viehweg and Zuo 2003, Claim 7.2]. □

We now switch our focus to Hodge metrics. Recall that we are always dealing with polarizable VHS with quasi-unipotent monodromy along simple normal crossings boundary. The following lemma translates the results of [Cattani et al. 1986] on the singularities of Hodge metrics in the unipotent case to this setting, and will be important for the proof of Proposition 3.5.

Lemma 3.2 (estimates for Hodge metrics; the quasi-unipotent case). *Suppose Δ^n is a polydisk with coordinates $(z_1, \dots, z_n)$. Let $\mathbb{V}$ be a polarized VHS on the open set $U = \Delta^n \setminus \big(\{(z_1, \dots, z_k) \mid \prod_{i=1}^k z_i = 0\}\big)$, $k \leq n$, with quasi-unipotent monodromies along each $\{z_i = 0\}$, and denote by $\mathscr{E}_\bullet$ the Higgs bundle associated to the Deligne extension of $\mathbb{V}$ with eigenvalues in $[0, 1)$. Then the Hodge metric induced by the polarization has at most logarithmic singularities along each z_i, for $i = 1, \dots, k$; that is, there exists an integer $d > 0$ such that for any section e of $\mathscr{E}_\bullet$ locally we have*

$$\|e\|_h^2 \leq C \cdot \prod_{i=1}^{k} (-\log|z_i|)^d \tag{3.2.1}$$

for some constant $C = C(e) \in \mathbb{R}_{>0}$.

Proof. Let L be the local system underlying $\mathbb{V}$, with monodromy Γ_i along z_i, for $i = 1, \dots, k$. Since the Γ_i commute pairwise, we have

$$L = \bigoplus_\alpha L_\alpha.$$

as the simultaneous (generalized) eigenspace decomposition with respect to the monodromy actions. Thus the monodromy action Γ_i on $L_{\alpha=(\alpha_1,\dots,\alpha_k)}$ has a unique eigenvalue $e^{-2\pi\sqrt{-1}\alpha_i}$. By the quasi-unipotent assumption, we can assume all α_i are rational numbers contained in $[0, 1)$. By the lower semicontinuity of rank functions of matrices, the above decomposition induces a decomposition of polarized variations of Hodge structure

$$\mathbb{V} = \bigoplus \mathbb{V}_\alpha,$$

and hence a decomposition of Higgs bundles

$$\mathscr{E}_\bullet = \bigoplus \mathscr{E}_\bullet^\alpha,$$

where $\mathscr{E}_\bullet^\alpha$ is the Higgs bundle associated to the Deligne extension of $\mathbb{V}_\alpha$ with eigenvalues in $[0, 1)$. (Note that the extension of $\mathbb{V}_\alpha$ has only one eigenvalue along each z_i.)

If $\alpha = (\alpha_1, \dots, \alpha_k) \neq 0$, then we can write $\alpha_i = p_i/q_i$ for some nonnegative integers $p_i < q_i$. Now, let $g : \Delta^n \to \Delta^n$ be the branched covering given by

$$g^* z_i = \begin{cases} w_i^{q_i} & \text{if } i = 1, \dots, k, \\ w_i & \text{otherwise,} \end{cases} \tag{3.2.2}$$

where $(w_1, \dots, w_n)$ define a coordinate system on the domain of g. It follows that the monodromies of $g^*\mathbb{V}_\alpha$ along w_i are unipotent. By comparing the eigenvalues of the residues upstairs, we have

$$g^* \mathscr{E}_\bullet^\alpha = \prod_{i=1}^k w_i^{p_i} \cdot \mathscr{E}_{g,\bullet}^\alpha,$$

where $\mathscr{E}_{g,\bullet}^\alpha$ is the Higgs bundle associated to the Deligne canonical extension of $g^*\mathbb{V}_\alpha$. Since the Hodge metric on $\mathscr{E}_{g,\bullet}^\alpha$ has logarithmic singularities (see [Cattani et al. 1986, Section 5.21]), for a section e of $\mathscr{E}_\bullet^\alpha$ we know that

$$\|e\|_h^2 \leq C \cdot \prod_{i=1}^k (|z_i|^{\alpha_i} \cdot (-\log|z_i|)^{d_i}) \leq C \cdot \prod_{\alpha_i = 0} (-\log|z_i|)^{d_i},$$

for some positive integers $d_i > 0$. The Hodge metric has logarithmic singularities along the z_i whenever $\alpha_i = 0$. In particular, we get the inequality (3.2.1) when $\alpha = (\alpha_1, \dots, \alpha_k) \neq 0$.

On the other hand, if $\alpha = (\alpha_1, \dots, \alpha_k) = 0$, then we know that the monodromies of $\mathbb{V}_\alpha$ are unipotent. Therefore, again thanks to [Cattani et al. 1986], the Hodge metric on $\mathscr{E}_\bullet^\alpha$ has logarithmic singularities along each z_i, as required. □

Remark 3.3. The above lemma implies that the Hodge metric is a singular metric on the vector bundle $\mathscr{E}_\bullet$.

Let us now return to the setting described at the beginning of this section, and suppose in addition that $\mathscr{E}_\bullet$ is the Higgs bundle associated to the Deligne extension of a VHS on $Y \setminus (D + S)$, with eigenvalues in $[0, 1)$. We define a singular metric h_g^α on the vector bundle $\mathscr{L}^{-1} \otimes \mathscr{E}_\bullet$ by

$$h_g^\alpha = g_\alpha^{-1} \otimes h,$$

where h is the Hodge metric on $\mathscr{E}_\bullet$.

Corollary 3.4. *For all $\alpha \gg 0$, the singular metric h_g^α is locally bounded.*

Proof. Assume that in local coordinates $D+S$ is given by $z_1\cdots z_{k+\ell}=0$. By construction, the singular metric g_α^{-1} degenerates to 0 at a rate proportional to

$$(r_D\cdot r_S)^{-\alpha}=\prod_{i=1}^{k}(-\log|z_i|^2-\log\|\tilde{s}_i\|^2_{g_{D_i}})^{-\alpha}\cdot\prod_{i=k+1}^{k+\ell}(-\log|z_i|^2-\log\|\tilde{s}_i\|^2_{g_{S_i}})^{-\alpha}.$$

On the other hand, the Hodge metric h on $\mathscr{E}_\bullet$ blows up to infinity along $z_i=0$ bounded by a quantity proportional to $\prod(-\log|z_i|^2)^d$, for some fixed $d>0$, thanks to Lemma 3.2. Hence, the metric h_g^α is bounded by a quantity proportional to

$$(r_D\cdot r_S)^{-\alpha+d}\cdot\prod_{i=1}^{k}\left(\frac{-\log|z_i|^2}{-\log|z_i|^2-\log\|\tilde{s}_i\|^2_{g_{D_i}}}\right)^d\cdot\prod_{j=k+1}^{k+\ell}\left(\frac{-\log|z_j|^2}{-\log|z_j|^2-\log\|\tilde{s}_j\|^2_{g_{S_j}}}\right)^d.$$

When $\alpha>d$ the above product is bounded. The compactness of Y gives the conclusion. □

3B. ***An application of the singular Ahlfors–Schwarz Lemma.*** In this section we establish the key technical ingredient. This is done by applying the tools discussed in the previous section to the base spaces of families of varieties, via the Hodge-theoretic set-up provided by the constructions in Section 2, especially those in Proposition 2.7.

Proposition 3.5. *In the situation of Proposition 2.7, the morphism*

$$\tau_{(\gamma,1)}:\mathscr{T}_{\mathbb{C}}\to\gamma^*(\mathscr{L}^{-1}\otimes\mathscr{E}_1)$$

induced by the entire curve $\gamma:\mathbb{C}\to Y\setminus D$ is identically zero.

Proof. The proof will be by contradiction. First we note that, assuming that $\tau_{(\gamma,1)}$ is nontrivial, the following claim holds.

Claim 3.6. *There exist*

(3.6.1) *integers $m>0$ and $p>0$,*

(3.6.2) *an ample line bundle $\mathscr{H}$ on Y, and*

(3.6.3) *a Higgs bundle $(\mathscr{E}'_\bullet,\theta'_\bullet)$ on Y underlying the Deligne extension with eigenvalues in $[0,1)$ of a VHS defined outside of $D+S$*

such that there is a nontrivial (hence injective) morphism $\tau_m:\mathscr{T}_{\mathbb{C}}^{\otimes m}\to\gamma^*(\mathscr{H}^{-1}\otimes\mathscr{E}'_p)$ factoring as

$$\tau_m:\mathscr{T}_{\mathbb{C}}^{\otimes m}\xrightarrow{d\gamma^{\otimes m}}\gamma^*\left(\bigotimes^m\mathscr{T}_Y(-\log D)\right)\to\gamma^*\mathscr{H}^{-1}\otimes\mathscr{N}'_{(\gamma,p)}\hookrightarrow\gamma^*(\mathscr{H}^{-1}\otimes\mathscr{E}'_p),\tag{3.6.4}$$

where $\mathscr{N}'_{(\gamma,\bullet)}=\ker\theta'_{(\gamma,\bullet)}$, with $\theta'_{(\gamma,\bullet)}$ the Higgs field of $\gamma^*\mathscr{E}'_\bullet$ (see Definition 2.2).

Proof of Claim 3.6. By construction, for all sufficiently large k we have $\tau_{(\gamma,k)}=0$. We set

$$p:=\max\{k\mid\tau_{(\gamma,k)}\neq 0\}.$$

By assumption (the injectivity of $\tau_{(\gamma,1)}$), we have $p \geq 1$. On the other hand, we know that $\tau_{(\gamma,p+1)}$ factors as

$$\tau_{(\gamma,p+1)} : \mathscr{T}_{\mathbb{C}}^{\otimes(p+1)} \xrightarrow{\mathrm{Id}\otimes\tau_{(\gamma,p)}} \mathscr{T}_{\mathbb{C}} \otimes \gamma^*\mathscr{L}^{-1} \otimes \gamma^*\mathscr{E}_p \to \gamma^*\mathscr{L}^{-1} \otimes \gamma^*\mathscr{E}_{p+1}(P),$$

where the last map is induced by the $\mathscr{A}_{\mathbb{C}}(-\log P)$-module structure on $\gamma^*\mathscr{E}_\bullet$, with $P = \gamma^{-1}(S)$. (Note that in fact its image lands in $\gamma^*\mathscr{L}^{-1} \otimes \gamma^*\mathscr{E}_{p+1}$ as required, due to the fact that in the definition, see Proposition 2.7, we factor through the Higgs field of $\mathscr{F}_\bullet$, which does not have poles along S.) Since $\tau_{(\gamma,p+1)} = 0$, we obtain that $\tau_{(\gamma,p)}$ injects $\mathscr{T}_{\mathbb{C}}^{\otimes p}$ into $\gamma^*\mathscr{L}^{-1} \otimes \mathscr{N}_{(\gamma,p)}$, where $\mathscr{N}_{(\gamma,\bullet)} = \ker\theta_{(\gamma,\bullet)}$, with $\theta_{(\gamma,\bullet)}$ the induced Higgs field of $\gamma^*\mathscr{E}_\bullet$. Thus we have a nontrivial composition of morphisms

$$\tau_{(\gamma,p)} : \mathscr{T}_{\mathbb{C}}^{\otimes p} \xrightarrow{d\gamma^{\otimes p}} \gamma^*\left(\bigotimes^{p} \mathscr{T}_Y(-\log D)\right) \to \gamma^*\mathscr{L}^{-1} \otimes \mathscr{N}_{(\gamma,p)} \hookrightarrow \gamma^*(\mathscr{L}^{-1} \otimes \mathscr{E}_p).$$

Now since $\mathscr{L}$ is big and nef, there exists $q > 0$ and an ample line bundle $\mathscr{H}$ such that $\mathscr{H} \subseteq \mathscr{L}^{\otimes q}$. Similarly to the proof of [Viehweg and Zuo 2003, Lemma 6.5], we consider the Higgs bundle $(\mathscr{E}'_\bullet, \theta'_\bullet)$ on Y given by

$$\begin{gathered}\mathscr{E}'_\bullet = \mathscr{E}_\bullet^{\otimes q} \quad \text{and} \quad \theta'_\bullet : \mathscr{E}_\bullet^{\otimes q} \to \mathscr{E}_{\bullet+1}^{\otimes q} \otimes \Omega_Y^1(D+S), \\ \theta'_\bullet = \theta_\bullet \otimes \mathrm{id}_{\mathscr{E}} \otimes \cdots \otimes \mathrm{id}_{\mathscr{E}} + \mathrm{id}_{\mathscr{E}} \otimes \theta_\bullet \otimes \cdots \otimes \mathrm{id}_{\mathscr{E}} + \cdots + \mathrm{id}_{\mathscr{E}} \otimes \cdots \otimes \mathrm{id}_{\mathscr{E}} \otimes \theta_\bullet.\end{gathered} \tag{3.6.5}$$

As noted in [loc. cit.], this Higgs bundle corresponds to the locally free extension V' to Y of the bundle coming from the VHS $\mathbb{V}^{\otimes q}$ on $Y \setminus (D+S)$, where $\mathbb{V}$ is the VHS underlying $\mathscr{E}_\bullet$. The induced connection on V' has residues with eigenvalues in $\mathbb{Q}_{\geq 0}$, and therefore V' is contained in V'', the Deligne extension with eigenvalues in [0, 1) (see [Popa and Wu 2016, Proposition 4.4]). Therefore, without loss of generality, in the paragraph below we can assume that $(\mathscr{E}'_\bullet, \theta'_\bullet)$ is in fact the Higgs bundle associated to this extension. Note moreover that when pulling back by γ, the above construction implies that we have an inclusion of logarithmic Higgs bundles on $\mathbb{C}$

$$((\gamma^*\mathscr{E}_\bullet)', \theta'_\bullet) \subseteq (\gamma^*\mathscr{E}'_\bullet, \theta'_{(\gamma,\bullet)}), \tag{3.6.6}$$

where the Higgs bundle on the left is the analogue for $\gamma^*\mathscr{E}_\bullet$ of the construction in (3.6.5).

Finally, let $m := pq$. Raising $\tau_{(\gamma,p)}$, seen as the composition of morphisms above, to the q-th tensor power, gives rise to a new nontrivial composition of morphisms:

$$\tau_m : \mathscr{T}_{\mathbb{C}}^{\otimes m} \xrightarrow{d\gamma^{\otimes m}} \gamma^*\left(\bigotimes^{m} \mathscr{T}_Y(-\log D)\right) \to \gamma^*\mathscr{H}^{-1} \otimes \mathscr{N}_{(\gamma,p)}^{\otimes q} \hookrightarrow \gamma^*\mathscr{H}^{-1} \otimes \gamma^*\mathscr{E}'_p,$$

where we used the inclusion of $\mathscr{L}^{\otimes -q}$ into $\mathscr{H}^{-1}$. In addition, the formula for the Higgs field on the left-hand side of (3.6.6) (compare (3.6.5)) implies immediately that $\mathscr{N}_{(\gamma,p)}^{\otimes q} \subseteq \mathscr{N}'_{(\gamma,p)}$, where we recall that $\mathscr{N}'_{(\gamma,p)} = \ker\theta'_{(\gamma,p)}$, so τ_m does factor as in (3.6.1). This concludes the proof of Claim 3.6. $\square$

We continue with the proof of Proposition 3.5. For simplicity, after renaming again $\mathscr{H}$ and $\mathscr{E}'_\bullet$ in Claim 3.6 by $\mathscr{L}$ and $\mathscr{E}_\bullet$, we assume from now on that $\mathscr{L}$ is ample, and that the morphism τ_m is given as

$$\tau_m\colon \mathscr{T}_{\mathbb{C}}^{\otimes m} \xrightarrow{d\gamma^{\otimes m}} \gamma^*\Big(\bigotimes^m \mathscr{T}_Y(-\log D)\Big) \to \gamma^*\mathscr{L}^{-1}\otimes\mathscr{N}_{(\gamma,p)} \hookrightarrow \gamma^*\mathscr{L}^{-1}\otimes\gamma^*\mathscr{E}_p. \tag{3.6.7}$$

Our aim is now to extract a contradiction from the existence of a nontrivial such morphism, by showing that $\mathbb{C}$ inherits a singular metric $h_{\mathbb{C}}$ satisfying the distance decreasing property for any holomorphic map $g\colon(\mathbb{D},\rho)\to(\mathbb{C},h_{\mathbb{C}})$, that is $d_{h_{\mathbb{C}}}(g(x),g(y))\le A\cdot d_\rho(x,y)$, where ρ is the Poincaré metric on the unit disk, and $A\in\mathbb{R}_{>0}$. Since the Kobayashi pseudometric is larger than any such distance function, this forces it to be nondegenerate, contradicting the fact that on $\mathbb{C}$ it is identically zero. For background on this material, see for instance [Kobayashi 1970, Chapter IV, Section 1].

Note first that, according to the Ahlfors–Schwarz lemma for (locally integrable) singular metrics over curves [Demailly 1997, Lem. 3.2] any singular metric verifying, for some $B\in\mathbb{R}_{>0}$, the inequality

$$F(\mathscr{T}_{\mathbb{C}},h_{\mathbb{C}})\le -B\cdot w_{h_{\mathbb{C}}} \tag{3.6.8}$$

in the sense of currents, satisfies the above distance decreasing property. (Here $\omega_{h_{\mathbb{C}}}=\sqrt{-1}/(2\pi)\|\partial_t\|^2_{h_{\mathbb{C}}}dt\wedge d\bar{t}$ denotes the fundamental form of the metric $h_{\mathbb{C}}$, which we have assumed to be a $(1,1)$-current, where t is the coordinate of $\mathbb{C}$.) Therefore, to conclude, it suffices to construct a metric $h_{\mathbb{C}}$ on $\mathbb{C}$ verifying the inequality (3.6.8). We next proceed to construct such a metric.

We first fix a smooth metric g on $\mathscr{L}$, so that the curvature form $F(\mathscr{L},g)$ is positive. Following the notation in Section 3A, this induces a singular metric g_α on $\mathscr{L}$, and a singular metric $h_g^\alpha=g_\alpha^{-1}\otimes h$ on $\mathscr{L}^{-1}\otimes\mathscr{E}_\bullet$, where we fix an $\alpha\gg 0$ as in Corollary 3.4. Consequently $\gamma^*h_g^\alpha$ is a singular metric on $\gamma^*(\mathscr{L}^{-1}\otimes\mathscr{E}_\bullet)$, and the m-th root of its pullback,

$$h_{\mathbb{C}}:=(\tau_m^*\gamma^*h_g^\alpha)^{1/m},$$

defines a singular metric on (the trivial line bundle) $\mathscr{T}_{\mathbb{C}}$.

Similarly, we have the continuous positive definite hermitian form ω_α on $\mathscr{T}_Y(-\log D)$ as in Lemma 3.1, and so $\gamma^*\omega_\alpha$ induces a singular metric on $\gamma^*\mathscr{T}_Y(-\log D))$, and hence also a singular metric on $\mathscr{T}_{\mathbb{C}}$ through the differential map. For the next claim, recall that $P=\gamma^{-1}(S)\subset\mathbb{C}$.

Claim 3.7. *We have $m\cdot F(\mathscr{T}_{\mathbb{C}},h_{\mathbb{C}})|_{\mathbb{C}\setminus P}\le-\gamma^*(r_D^{-2})\cdot\gamma^*\omega_\alpha|_{\mathbb{C}\setminus P}$, in the sense of currents.*

Proof of Claim 3.7. Note that $F(\mathscr{T}_{\mathbb{C}},h_{\mathbb{C}})$ makes sense as a current on $\mathbb{C}\setminus P$. The proof of the claim will also imply that it is indeed a current everywhere on $\mathbb{C}$, as we explain afterwards.

Denote by $\mathscr{B}$ the saturation of $\tau_m(\mathscr{T}_{\mathbb{C}}^{\otimes m})$ inside $\gamma^*(\mathscr{L}^{-1}\otimes\mathscr{E}_\bullet)$, so that

$$\mathscr{B}\simeq\mathscr{T}_{\mathbb{C}}^{\otimes m}(G),$$

where $G\ge 0$ is a divisor on $\mathbb{C}$. Since τ_m factors through $\gamma^*\mathscr{L}^{-1}\otimes\mathscr{N}_{(\gamma,p)}$, we know that $\theta_{(\gamma,\bullet)}(\gamma^*\mathscr{L}\otimes\mathscr{B})=0$. Recall that as a consequence of Griffiths' curvature estimates for Hodge metrics, it is well-known (see e.g., [Viehweg and Zuo 2001, Lemma 1.1] and the references therein) that the Hodge metric restricted to any subbundle inside the kernel of the Higgs field associated to a VHS has semi-negative curvature. We thus conclude that

$$F(\mathscr{B}, \gamma^* h_g^\alpha|_{\mathscr{B}})|_{\mathbb{C}\setminus P} + \gamma^* F(\mathscr{L}, g_\alpha)|_{\mathbb{C}\setminus P} \leq 0,$$

and since

$$(\mathscr{T}_{\mathbb{C}}^{\otimes m} \otimes \gamma^* \mathscr{L})(G) \simeq \mathscr{B} \otimes \gamma^* \mathscr{L},$$

this implies

$$m \cdot F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})|_{\mathbb{C}\setminus P} + \gamma^* F(\mathscr{L}, g_\alpha)|_{\mathbb{C}\setminus P} \leq F(\mathscr{B}, \gamma^* h_g^\alpha|_{\mathscr{B}})|_{\mathbb{C}\setminus P} + \gamma^* F(\mathscr{L}, g_\alpha)|_{\mathbb{C}\setminus P} \leq 0. \tag{3.7.1}$$

Now the statement follows from Lemma 3.1. □

As mentioned above, the proof of the claim also implies that $F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})$ is a current on $\mathbb{C}$. Indeed, from construction, we know $F(\mathscr{L}, g_\alpha)$ is a $(1,1)$-current and $F(\mathscr{L}, g_\alpha)|_{Y\setminus(D+S)}$ is positive. Hence, by (3.7.1), we know $F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})|_{\mathbb{C}\setminus P}$ is negative; or equivalently, $\log\|\partial_t\|^2_{h_{\mathbb{C}}}$ is subharmonic on $\mathbb{C}\setminus P$. Since $h_{\mathbb{C}}$ is locally bounded (see Corollary 3.4), $\log\|\partial_t\|^2_{h_{\mathbb{C}}}$ extends to a subharmonic function on $\mathbb{C}$ (see [Demailly 2012, Theorem 5.23]), and so $F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})$ is a negative current.

Next we fix a polydisk neighborhood $\Delta^n \subseteq Y$. The continuous metric $\|\cdot\|_{\omega_\alpha}$ on $\mathscr{T}_Y(-\log D)$ given by ω_α induces a metric on $\bigotimes^m \mathscr{T}_Y(-\log D)$. We also fix an orthonormal basis $\{\psi_1, \ldots, \psi_N\}$ of continuous sections of $\bigotimes^m \mathscr{T}_Y(-\log D)|_{\Delta^n}$ with respect to the induced metric. (By abuse of notation, we use $\bigotimes^m \mathscr{T}_Y(-\log D)|_{\Delta^n}$ even when considering the associated sheaf of continuous sections.)

We fix a holomorphic basis $\{e_1, e_2, \ldots, e_M\}$ of $\mathscr{L}^{-1} \otimes \mathscr{E}_p|_{\Delta^n}$ as well. We write

$$\tilde{\tau}_m(\psi_i) = \sum_j b_i^j \cdot e_j \tag{3.7.2}$$

for some continuous functions b_i^j on Δ^n, where

$$\tilde{\tau}_m : \bigotimes^m \mathscr{T}_Y(-\log D) \to \mathscr{L}^{-1} \otimes \mathscr{E}_p,$$

and we also write

$$d\gamma^{\otimes m}(\partial_t^m|_{\gamma^{-1}(\Delta^n)}) = \sum_i c_i \cdot \gamma^* \psi_i, \tag{3.7.3}$$

for some continuous (complex valued) functions c_i.

Claim 3.8. *We have $\gamma^*(r_D^{-2}) \cdot \gamma^* \omega_\alpha \geq B \cdot \omega_{h_{\mathbb{C}}}$ in the sense of currents on $\mathbb{C}$, for some $B > 0$.*

Proof of Claim 3.8. Since $\mathscr{T}_{\mathbb{C}}$ is trivialized by ∂_t globally, it enough to show

$$\gamma^*(r_D^{-2m}) \cdot \|d\gamma^{\otimes m}(\partial_t^m)\|_{\gamma^*\omega_\alpha} \geq B \cdot \|\tau_m(\partial_t^m)\|_{\gamma^* h_g^\alpha}.$$

By the compactness of Y, it is enough to prove the inequality locally on neighborhoods of the form $\gamma^{-1}(\Delta^n)$, with $\Delta^n \subset Y$ as above.

First, since $\{\psi_1, \dots, \psi_N\}$ is an orthonormal basis, by (3.7.3) we see that

$$\|d\gamma^{\otimes m}(\partial_t^m|_{\gamma^{-1}(\Delta^n)})\|_{\gamma^*\omega_\alpha} = \Big(\sum_i |c_i|^2\Big)^{1/2}. \tag{3.8.1}$$

By (3.7.2) and (3.7.3), we also have

$$\|\tau_m(\partial_t^m|_{\gamma^{-1}(\Delta^n)})\|_{\gamma^* h_g^\alpha} = \Big\| \sum_i c_i \sum_j \gamma^*(b_i^j \cdot e_j) \Big\|_{\gamma^* h_g^\alpha}.$$

On the other hand, by the Cauchy–Schwarz inequality, we have

$$\Big\| \sum_i c_i \sum_j \gamma^*(b_i^j \cdot e_j) \Big\|_{\gamma^* h_g^\alpha} \le \Big(\sum_i |c_i|^2\Big)^{1/2} \cdot \Big(\sum_i \gamma^* \Big\| \sum_j (b_i^j \cdot e_j) \Big\|_{h_g^\alpha}^2 \Big)^{1/2}.$$

By Corollary 3.4, we know that h_g^α is bounded over Δ^n by a quantity proportional to

$$(r_D \cdot r_S)^{-\alpha+d} \cdot \prod_{i=1}^{k} \left(\frac{-\log|z_i|^2}{-\log|z_i|^2 - \log\|\tilde{s}_i\|^2_{g_{D_i}}} \right)^d \cdot \prod_{j=k+1}^{k+\ell} \left(\frac{-\log|z_j|^2}{-\log|z_j|^2 - \log\|\tilde{s}_j\|^2_{g_{S_j}}} \right)^d,$$

for some fixed $d > 0$. Therefore, we have

$$\|\tau_m(\partial_t^m|_{\gamma^{-1}\Delta^n})\|_{\gamma^* h_g^\alpha} \le \tfrac{1}{B} \cdot \gamma^*(r_D^{(d-\alpha)/2}) \cdot \Big(\sum |c_i|^2\Big)^{1/2}$$

for some $B > 0$ and α sufficiently large (recall that r_S^γ is bounded for $\gamma < 0$). This implies the conclusion, given (3.8.1) and the fact that earlier we have chosen our scaling so that $r_D \ge 1$. □

Finally, the inequality (3.6.8) follows from Claims 3.7, 3.8, and the fact that if the inequality

$$F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})|_{(\mathbb{C}\setminus P)} \le -B \cdot (\omega_{h_{\mathbb{C}}}|_{(\mathbb{C}\setminus P)})$$

holds as currents for some $B > 0$, then we also have

$$F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}}) \le -B \cdot \omega_{h_{\mathbb{C}}},$$

as currents on $\mathbb{C}$. But this is an easy consequence of the negativity of $F(\mathscr{T}_{\mathbb{C}}, h_{\mathbb{C}})$, together with the continuity of $\omega_{h_{\mathbb{C}}}$. □

3C. *Some further background.* In this section we collect a few useful facts regarding entire maps on the one hand, and families with maximal variation on the other.

3C.1. *Algebraic degeneracy to Brody hyperbolicity.* In Section 2D we observed that the Hodge theoretic constructions of Section 2C are valid as long as we replace the initial family $f_U : U \to V$ by a birational model, compactified by the family $f : X \to Y$ in Proposition 2.7. We recall below, following [Viehweg and Zuo 2003, Section 1], that the study of the hyperbolicity properties can be reduced to investigating algebraic nondegeneracy on such models.

Lemma 3.9 [Viehweg and Zuo 2003, Lemma 1.2]. *Let $\gamma : \mathbb{C} \to V$ be an entire curve with a Zariski-dense image, and $\mu : \tilde{V} \to V$ a surjective proper birational morphism of varieties. Then the map $(\mu^{-1} \circ \gamma)$ extends to a holomorphic map $\tilde{\gamma} : \mathbb{C} \to \tilde{V}$.*

Proposition 3.10 (reduction of Brody hyperbolicity to algebraic degeneracy). *Let P_h be a coarse moduli space of polarized manifolds, as in the Introduction, and V and Y as in Proposition 2.7:*

(3.10.1) *The image of $\gamma : \mathbb{C} \to V$ is algebraically degenerate if and only if the induced morphism $\tilde{\gamma} : \mathbb{C} \to \tilde{V}$ defined in Lemma 3.9 is so.*

(3.10.2) *To prove the Brody hyperbolicity of P_h, in the sense of Theorem 1.1, it suffices to show that for every smooth quasi-projective variety V with a generically finite morphism $V \to P_h$, every entire curve $\mathbb{C} \to V$ is algebraically degenerate.*

Proof. (3.10.1) is the direct consequence of Lemma 3.9. For (3.10.2), note that given a quasi-finite morphism $W \to P_h$ from a variety W, and $\gamma : \mathbb{C} \to W$, the restriction W' of $\mathrm{Im}(\gamma)$ to the Zariski closure W' of $\mathrm{Im}(\gamma)$ is also quasi-finite. Furthermore, we can desingularize W' by $\mu : \widetilde{W}' \to W'$, and by (3.10.1), the degeneracy of the induced map $\mathbb{C} \to \widetilde{W}'$ is equivalent to the fact that γ is constant. □

Therefore, to prove Theorem 1.1 on the Brody hyperbolicity of P_h, it suffices to establish Theorem 1.2.

3C.2. *More on families with maximal variation.* We recall a few facts about families with maximal variation that were established by Kollár [1987]. Here $f : U \to V$ is a smooth projective morphism of smooth varieties, with fibers of nonnegative Kodaira dimension.

Lemma 3.11 [Kollár 1987, Corollary 2.9]. *If $\mathrm{Var}(f) = \dim V$, and if v is a very general point of V, then for any analytic arc $\gamma : \Delta \to V$ passing through v, not all fibers of f over $\gamma(\Delta)$ are birational.*

We will denote by $W \subset V$ the locus of points v satisfying the property in Lemma 3.11. In general we see that W is the complement of a countable union of closed subsets of V. The following result says that when the fibers of f are of general type, it is guaranteed to contain a Zariski open set V_0.

Lemma 3.12 [Kollár 1987, Theorem 2.5]. *If the fibers of f are of general type, then there exists an open subset $V_0 \subseteq V$ and a morphism $g : V_0 \to Z$ onto an algebraic variety, such that for $v_1, v_2 \in V_0$ the fibers U_{v_1} and U_{v_2} are birational if and only if $g(v_1) = g(v_2)$.*

Indeed, when $\mathrm{Var}(f) = \dim V$, in the lemma above we have $\dim Z = \dim V$, and the map g is generically finite. Thus there exists a, perhaps smaller, dense open subset $\tilde{V}_0 \subseteq V$, such that $\tilde{V}_0 \subseteq W$ (namely the complement of the positive dimensional fibers of g).

3D. ***Algebraic degeneracy for base spaces of families of minimal varieties of general type.*** We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. We first show that every holomorphic curve $\gamma : \mathbb{C} \to V$ is algebraically degenerate. According to Proposition 3.10(3.10.1), we can assume that $V = Y \setminus D$ as in Proposition 2.7.

Recall that the mapping appearing in Proposition 3.5 can be written as the composition

$$\tau_{(\gamma,1)}\colon \mathscr{T}_{\mathbb{C}} \to \gamma^*\mathscr{T}_Y(-\log D) \to \gamma^*({\mathscr{F}_0}^{-1}\otimes\mathscr{F}_1) \hookrightarrow \gamma^*(\mathscr{L}^{-1}\otimes\mathscr{E}_1).$$

Now by Corollary 2.12 we have a generic identification of

$$\tau_1\colon \mathscr{T}_Y(-\log D) \to {\mathscr{F}_0}^{-1}\otimes\mathscr{F}_1$$

with the Kodaira–Spencer map of the family $f\colon X\to Y$, and so by base change the composition of the first two maps in the definition of $\tau_{(\gamma,1)}$ can be identified with the Kodaira–Spencer map of the induced family over $\mathbb{C}$. If $\gamma(\mathbb{C})$ were dense, we would obtain a family with maximal variation over $\mathbb{C}$, implying that this Kodaira–Spencer map is injective; indeed, over a curve it can only be injective or 0, the latter case of course implying that the family is locally trivial. But this in turn implies that $\tau_{(\gamma,1)}$ is injective, which contradicts Proposition 3.5.

We now show the stronger statement that $\mathrm{Exc}(V)$ is a proper subset, knowing that the algebraic degeneracy statement we just proved holds for any base of a family as in the theorem. Let V_0 be the Zariski open subset in Lemma 3.12 and $\tilde{V}_0$ be the subset of V_0 over which the morphism g is finite. We claim that

$$\mathrm{Exc}(V)\subseteq V\setminus\tilde{V}_0.$$

To see this, assume that there exists an entire curve $\gamma:\mathbb{C}\to V$ with $\gamma(\mathbb{C})\cap\tilde{V}_0\neq\varnothing$, and denote by W the Zariski closure of $\gamma(\mathbb{C})$ in V. If γ is not constant, then by definition the restriction of the family f over W has maximal variation. Furthermore, using again Proposition 3.10, we can assume that W is smooth. We then obtain a contradiction with the algebraic degeneracy of all maps $\mathbb{C}\to W$. □

3E. *Algebraic degeneracy for surfaces mapping to moduli stacks of polarized varieties.* We now prove the stronger statements in the case when the base of the family is a smooth surface. We start with two basic lemmas about pulling back sheaf morphisms via γ, the first of which is immediate.

Lemma 3.13. *Let $\gamma\colon\mathbb{C}\to V$ be a holomorphic map with Zariski dense image, where V is an algebraic variety. If $\varphi\colon\mathscr{E}\to\mathscr{F}$ is an injective morphism of locally free $\mathscr{O}_V$-modules, then $\gamma^*\varphi\colon\gamma^*\mathscr{E}\to\gamma^*\mathscr{F}$ is also injective.*

Lemma 3.14. *Let $\gamma\colon\mathbb{C}\to V$ be a holomorphic map with Zariski dense image, where V is a smooth algebraic surface. Let Z be a 0-dimensional local complete intersection subscheme of V. Then we have an inclusion $\gamma^*I_Z\hookrightarrow\mathscr{O}_{\mathbb{C}}$.*

Proof. We can cover $\mathbb{C}$ with the preimages of open subsets in V on which Z is given as $f_1=f_2=0$, where f_1 and f_2 are two non-proportional functions. Denoting by D_1 and D_2 the divisors of these two functions, so that Z is the scheme theoretic intersection $D_1\cap D_2$, we can thus assume that we have a Koszul complex

$$0\to\mathscr{O}_V(-D_1-D_2)\to\mathscr{O}_V(-D_1)\oplus\mathscr{O}_V(-D_2)\to I_Z\to 0.$$

Pulling back this sequence by γ, we still have a short exact sequence, as the first map degenerates only at the points of Z. Therefore we have a commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
0 \arrow[r] & \gamma^*\mathscr{O}_V(-D_1-D_2) \arrow[r] \arrow[d] & \gamma^*\mathscr{O}_V(-D_1)\oplus\gamma^*\mathscr{O}_V(-D_2) \arrow[r] \arrow[d] & \gamma^* I_Z \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathscr{O}_{\mathbb{C}} \arrow[r] \arrow[d] & \mathscr{O}_{\mathbb{C}}\oplus\mathscr{O}_{\mathbb{C}} \arrow[r] \arrow[d] & \mathscr{O}_{\mathbb{C}} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathscr{O}_{P_1+P_2} \arrow[r] \arrow[d] & \mathscr{O}_{P_1}\oplus\mathscr{O}_{P_2} \arrow[r] \arrow[d] & \mathscr{O}_{P_1\cap P_2} \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd}$$

where $P_1 = \gamma^*D_1$ and $P_2 = \gamma^*D_2$ are divisors on $\mathbb{C}$, and we used the identification $\gamma^*\mathscr{O}_V = \mathscr{O}_{\mathbb{C}}$. Note that the two left vertical sequences are exact because of the Zariski density of the image of γ, which consequently cannot be contained in any divisor on V (a special example of Lemma 3.13 above). By the snake lemma we obtain that the map in the upper right corner is also injective. $\square$

Proof of Theorem 1.4. We first prove (1.4.1). Aiming for a contradiction, we assume that the image $\gamma(\mathbb{C})$ is Zariski dense in V. We follow the set-up and notation of Proposition 2.7. By Proposition 3.10, we may assume that $V = Y \setminus D$.

We may also assume that the morphism

$$\mathscr{T}_Y(-\log D) \xrightarrow{\psi} \mathscr{F}_0{}^{-1} \otimes \mathscr{F}_1$$

is not injective, as otherwise by Lemma 3.13 it follows that the composition of morphisms

$$\mathscr{T}_{\mathbb{C}} \to \gamma^*\mathscr{T}_Y(-\log D) \to \gamma^*(\mathscr{F}_0{}^{-1}\otimes\mathscr{F}_1) = \gamma^*(\mathscr{L}^{-1}\otimes\mathscr{F}_1) \hookrightarrow \gamma^*(\mathscr{L}^{-1}\otimes\mathscr{E}_1)$$

is also injective, contradicting Proposition 3.5. By Lemma 2.9, we also know that ψ is not the zero map. We define $\mathscr{G} := \mathrm{Im}(\psi)$, which therefore has generic rank one, and leads to a short exact sequence

$$0 \to \mathscr{K} \to \mathscr{T}_Y(-\log D) \to \mathscr{G} \to 0.$$

Since $\mathscr{G}$ injects in a torsion-free sheaf, it is torsion-free itself. Therefore $\mathscr{K}$ is reflexive, hence an invertible sheaf since we are on a smooth surface. Moreover, since it is saturated in $\mathscr{T}_Y(-\log D)$, we must have

$$\mathscr{G} \simeq \mathscr{M} \otimes \mathcal{I}_Z,$$

where $\mathscr{M}$ is a line bundle and Z is a (possibly empty) 0-dimensional subscheme of Y. It is standard that Z is a local complete intersection.

Note that since $\mathscr{L} \subseteq \mathscr{F}_0$, we have an inclusion $\mathscr{G} \subseteq \mathscr{L}^{-1} \otimes \mathscr{E}_1$. We claim that this induces an inclusion

$$\gamma^*\mathscr{G} \subseteq \gamma^*(\mathscr{L}^{-1} \otimes \mathscr{E}_1),$$

which in particular shows that $\gamma^*\mathscr{G}$ it torsion-free. To see this, note that the initial inclusion factors as a composition

$$\mathscr{M} \otimes I_Z \hookrightarrow \mathscr{M} \hookrightarrow \mathscr{L}^{-1} \otimes \mathscr{E}_1,$$

and the second map pulls back to an injective map by Lemma 3.13. It suffices then to have that the inclusion $I_Z \hookrightarrow \mathscr{O}_V$ also pulls back to an injective map, and this is precisely the content of Lemma 3.14.

Again by Lemma 3.13, the pullback sequence

$$0 \to \gamma^*\mathscr{K} \to \gamma^*\mathscr{T}_Y(-\log D) \to \gamma^*\mathscr{G} \to 0$$

is also exact. Since $\gamma^*\mathscr{G}$ it torsion-free, and the image of $\mathscr{T}_{\mathbb{C}}$ inside $\gamma^*(\mathscr{L}^{-1} \otimes \mathscr{E}_1)$ is zero by Proposition 3.5, it follows that the map $\mathscr{T}_{\mathbb{C}} \to \gamma^*\mathscr{T}_Y(-\log D)$ factors through $\gamma^*\mathscr{K}$.

Consider now the saturation $\mathscr{K}'$ of $\mathscr{K}$ in $\mathscr{T}_Y$, which defines a foliation on Y. Since the differential $\mathscr{T}_{\mathbb{C}} \to \gamma^*\mathscr{T}_Y$ clearly factors through $\gamma^*\mathscr{K}'$ as well, the image $\gamma(\mathbb{C})$ sits inside (or equivalently is tangent to) a leaf of this foliation. On the other hand, according to [Popa and Schnell 2017, Theorem A], the pair (Y, D) is of log general type. But this contradicts McQuillan's result [1998] on the degeneracy of entire curves tangent to leaves of nontrivial foliations on surfaces of general type (see also [Rousseau 2018, Theorem 3.13]), and more precisely its natural extension to the log setting as in El Goul [2003, Theorem 2.4.2]. This finishes the proof of (1.4.1).

To prove (1.4.2), just as in the proof of Theorem 1.2 let V_0 be the Zariski open subset in Lemma 3.12 and $\tilde{V}_0$ be the subset of V_0 over which the morphism g is finite. We again claim that

$$\operatorname{Exc}(V) \subseteq V \setminus \tilde{V}_0.$$

Assume on the contrary that there exists an entire curve $\gamma : \mathbb{C} \to V$ with $\gamma(\mathbb{C}) \cap \tilde{V}_0 \neq \varnothing$. Then, by definition, the pull-back of the family f via γ has maximal variation. Since $\gamma(\mathbb{C})$ cannot be Zariski dense in V by (1.4.1), it is either a point, or it is dense in a quasi-projective curve C, which by Proposition 3.10 can be assumed to be smooth. In the latter case, we thus obtain a smooth family of varieties of general type over C, with maximal variation. But then by [Viehweg and Zuo 2001, Theorem 0.1] we know that C cannot be $\mathbb{C}^*$, $\mathbb{C}$, $\mathbb{P}^1$ or an elliptic curve, which gives a contradiction. □

Proof of Corollary 1.5. According to Proposition 3.10(3.10.2), it is enough to show that there cannot be algebraically nondegenerate holomorphic maps $\gamma : \mathbb{C} \to V$, where V is a smooth quasi-projective variety of dimension 1 or 2 with a generically finite map $V \to P_h$. If $\dim V = 2$, this follows from Theorem 1.4(1.4.1). If $\dim V = 1$, it follows again from [Viehweg and Zuo 2001, Theorem 0.1], as explained at the end of the proof of Theorem 1.4. □

Remark 3.15. We note that Proposition 3.5 gives an alternative proof of [Viehweg and Zuo 2001, Theorem 0.1], since it shows that a quasi-projective variety V of dimension one is hyperbolic if it supports a birationally nonisotrivial smooth family of projective varieties whose geometric generic fiber admits a good minimal model. This is because, in this case, the map $\mathscr{T}_V \to (\mathscr{L}^{-1} \otimes \mathscr{E}_1)|_V$ induced by $\mathscr{F}_0 \to \mathscr{F}_1 \otimes \Omega^1_Y(\log D)$ as in Proposition 2.7 is an injection, as the latter map is injective by Lemma 2.9.

Appendix: Generic freeness and construction of sections

This is a technical appendix verifying that the sections needed in order to perform the Hodge module and Higgs bundle constructions in Section 2 can indeed be produced even after a birational modification ensuring that the singular locus of these Hodge theoretic objects has simple normal crossings. This is stated in [Viehweg and Zuo 2003, Lemma 5.4] when the fibers of the family have semiample canonical bundle, and in [Popa and Schnell 2017, Section 2.2] in general, but in both references the concrete details are not included. It turns out that they are somewhat technical, and therefore worth recording; however, we emphasize that all the ingredients needed for the proof can be found in [Viehweg and Zuo 2003], only one technical addition being needed when the canonical bundle of the fibers is not assumed to be semiample.

What we are aiming for is Proposition A.4 below. For its statement and proof, the starting point is the following generic freeness statement. We consider a smooth family $f_{\tilde{U}} : \tilde{U} \to \tilde{V}$ with projective fibers, whose geometric generic fiber admits a good minimal model, and with $\tilde{U}$ and $\tilde{V}$ smooth and quasi-projective. We assume that $f_{\tilde{U}}$ has maximal variation.

Proposition A.1. *With the assumptions above, there exist a smooth birational model $V \to \tilde{V}$, a smooth projective compactification Y of V with $D = Y \setminus V$ a simple normal crossings divisor, an algebraic fiber space $f : X \to Y$, smooth over V, with X smooth projective and $f^{-1}(D)$ a simple normal crossings divisor, as well as an ample line bundle $\mathscr{L}$ and an effective divisor $D_Y \geq D$ on Y, such that*

$$f_*\omega^m_{X/Y} \otimes \mathscr{L}(D_Y)^{-m}$$

is generated by global sections over V for all m sufficiently large and divisible. Moreover, if the fibers of $f_{\tilde{U}}$ have semiample canonical bundle, then

$$\omega^m_{X/Y} \otimes f^*\mathscr{L}(D_Y)^{-m}$$

is also generated by global sections over $U = f^{-1}(V)$.

When the fibers of the family have semiample canonical bundle, this is nothing else but [2003, Proposition 4.1 and Corollary 4.3]. In the general case the proof is identical, based on Viehweg's fiber product trick and the mild reduction of Abramovich–Karu, except in one step we need to replace the use of weak positivity by that of the following analytic extension theorem of Berndtsson, Păun and Takayama, as stated by Cao [2016, Theorem 2.10]:[4]

[4] We are stating more precisely what is the locus over which global generation holds, but this is an immediate consequence of the proof in [loc. cit.]

Theorem A.2. *Let $p : X \to Y$ be an algebraic fiber space between smooth projective varieties, and let $\mathscr{M}$ be a line bundle on X with a singular metric h such that $i\Theta_h(\mathscr{M}) \geq 0$ in the sense of currents. Let $\mathscr{B}$ be a very ample line bundle on Y such that the global sections of $\mathscr{B} \otimes \omega_Y^{-1}$ separate $2n$-jets, where n is the dimension of Y, and let $V \subseteq Y$ be a Zariski open set such that p is flat over V and $h^0(X_y, \omega_{X/Y}^k \otimes \mathscr{M}|_{X_y})$ is constant over $y \in V$, for some positive integer k. Assume also that the multiplier ideal $\mathcal{J}(h^{\frac{1}{k}}|_{X_y}) = \mathscr{O}_{X_y}$ for $y \in V$. Then*

$$f_*(\omega_{X/Y}^k \otimes \mathscr{M}) \otimes \mathscr{B}$$

is globally generated over V.

The idea of using this ingredient as a substitute for weak positivity is due to Y. Deng [2018a], whom we thank for allowing us to use it here. We give the proof including all the details from [Viehweg and Zuo 2003], as some are also necessary for the proof of Proposition A.4.

Proof of Proposition A.1. We use the theory of mild morphisms; for the definition, and a discussion of the relevant properties, please see [Viehweg and Zuo 2003, Section 2]. Using the mild reduction procedure of Abramovich–Karu (see [Viehweg and Zuo 2003, Lemma 2.3]), there exists $f_U : U \to V$, a birational model of the original family $f_{\tilde{U}} : \tilde{U} \to \tilde{V}$ with U and V smooth, which fits into a diagram

$$\begin{tikzcd}
U \arrow[r, "\subseteq"] \arrow[d, "f_U"] & W \arrow[d, "f_W"] & W' \arrow[l] \arrow[d] & Z_W \arrow[l, "\sigma_W"'] \arrow[d, "g_W"] & W'' \arrow[l, "\rho_W"'] \arrow[r, "\delta_W"] \arrow[d, "f''_W"] & Z'_W \arrow[d, "g'_W"] \\
V \arrow[r, "\subseteq"] & Y & Y' \arrow[l, "\tau"'] & Y' \arrow[l, "="'] & Y' \arrow[l, "="'] \arrow[r, "="] & Y',
\end{tikzcd}$$

where τ is a finite morphism with Y' smooth, branched over a simple normal crossing divisor Δ_τ, W' is the normalization of the main component of $W \times_Y Y'$, σ_W is a resolution of W' with centers in its singular locus, ρ_W and δ_W are birational with W'' smooth, and g'_W is mild. By taking further resolutions, we can assume that $\Delta_\tau + D$ and $f_W^{-1}(\Delta_\tau + D)$ are simple normal crossings divisors, where $D = Y \setminus V$. By possibly composing τ with a Kawamata covering, we are allowed to assume that Y' is smooth, and hence W' is normal with rational singularities (see for instance [Viehweg 1983, Lemma 2.1]).

Since $f_{\tilde{U}}$ is of maximal variation, with geometric generic fiber admitting a good minimal model, by construction so is g_W, hence by a well-known result of Kawamata [1985] we know that $[\det]g_{W*}\omega_{Z_W/Y'}^v$ is a big line bundle for some integer v sufficiently large and divisible. Here and in what follows we use $[\bullet]$ to denote the reflexive hull of the corresponding operator on sheaves.

Fix an ample line bundle $\mathscr{A}_Y$ on Y. Pick k_0 large enough so that $\mathscr{A} = \mathscr{A}_Y^{k_0}(-D)$ is also ample, $\tau^*\mathscr{A}$ is very ample, and the global sections of $\tau^*\mathscr{A} \otimes \omega_{Y'}^{-1}$ separate $2n$-jets. Then, by [Viehweg and Zuo 2003, Corollary 2.4(ix)], we have

$$([\det]g_{W*}\omega_{Z_W/Y'}^v)^{N_v} = \tau^*\mathscr{A}(D' + D)$$

for some positive integer N_v, where D' an effective divisor on Y.

Another input we need is the fact that the quantity

$$e(\omega_{W_y}^v) = \sup\left\{\frac{1}{\mathrm{lct}(B)} \mid B \in |\omega_{W_y}^v|\right\},$$

where $\mathrm{lct}(B)$ is the log canonical threshold of B, is upper semicontinuous as a function of $y \in V$. This is simple consequence of the standard lower semicontinuity of the log canonical threshold of divisors that are relative for smooth proper morphisms, combined with the invariance of plurigenera. It follows that there exists a positive integer C such that

$$e(\omega_{W_y}^v) < Cv$$

uniformly for every $y \in V$.

We now take $r = C(C+1)vN_v r_0$, where $r_0 = \mathrm{rank}(g_{W*}\omega_{Z_W/Y'}^v)$. We obtain a new family $f: X \to Y$ by taking $X = W^{(r)}$, a desingularization of the main component of the r-th fiber product $W \times_Y \cdots \times_Y W$. As always, we are allowed to assume that $f^{-1}(\Delta_\tau + D)$ is normal crossing. Completely similarly to the process for f_W, we can fit f into a reduction diagram

$$\begin{tikzcd}
X \arrow[d, "f"] & X' \arrow[l] \arrow[d] & Z \arrow[l, "\sigma"'] \arrow[d, "g"] & X'' \arrow[l, "\rho"'] \arrow[r, "\delta"] \arrow[d, "f''"] & Z' \arrow[d, "g'"] \\
Y & Y' \arrow[l, "\tau"'] & Y' \arrow[l, "="'] & Y' \arrow[l, "="'] \arrow[r, "="] & Y',
\end{tikzcd}$$

where X' is the normalization of the main component of $X \times_Y Y'$ (so that X' has rational singularities), σ is a resolution of X' with centers in the singular locus, ρ and δ are birational, with X'' smooth, and $Z' = Z'_W \times_Y \cdots \times_Y Z'_W$ with the morphism g' induced by g'_W. Since g'_W is mild, we know that g' is also mild (see [Viehweg and Zuo 2003, Lemma 2.2(ii)]).

Now by [Viehweg and Zuo 2003, Corollary 2.4(vii)] we have

$$g'_{W*}\omega_{Z'_W/Y'}^v \simeq g_{W*}\omega_{Z_W/Y'}^v,$$

and both sheaves are reflexive. Hence, by flat base change and the projection formula, since $g'_*\omega_{Z'/Y'}^v$ is also reflexive, we get

$$g'_*\omega_{Z'/Y'}^v \simeq \Big[\bigotimes^r\Big] g'_{W*}\omega_{Z'_W/Y'}^v \simeq \Big[\bigotimes^r\Big] g_{W*}\omega_{Z_W/Y'}^v.$$

Thanks again to [Viehweg and Zuo 2003, Corollary 2.4(vii)], we also have $g'_*\omega_{Z'/Y'}^v \simeq g_*\omega_{Z/Y'}^v$. On the other hand, there is a natural morphism

$$[\det] g_{W*}\omega_{Z_W/Y'}^v \to \Big[\bigotimes^{r_0}\Big] g_{W*}\omega_{Z_W/Y'}^v,$$

which splits locally over $V' = \tau^{-1}(V)$ (since g_W is smooth over V', so is g). Putting everything together, we obtain an injective morphism

$$\tau^*\mathscr{A}(D'+D)^{C(C+1)v} = ([\det] g_{W*}\omega_{Z_W/Y'}^v)^{C(C+1)vN_v} \to g_*\omega_{Z/Y'}^v,$$

which also splits locally over V'. This corresponds to an effective divisor

$$\Gamma \in |\omega^{v}_{Z/Y'} \otimes g^*\tau^*\mathscr{A}(D'+D)^{-C(C+1)v}|$$

which does not contain the fiber $Z_{y'}$ for every $y' \in V'$.

Since $Z_{y'} = W^r_y = W_y \times \cdots \times W_y$, using the bound $e(\omega^v_{W_y}) < Cv$ and [Viehweg 1995, Corollary 5.21], we have

$$\mathrm{lct}(\Gamma|_{Z_{y'}}) > \frac{1}{Cv}$$

for every $y' \in V'$, where $y = \tau(y')$.

For all $k > 0$, we can apply Theorem A.2 to the line bundle

$$\mathscr{M} = \omega^{kv}_{Z/Y'} \otimes g^*\tau^*\mathscr{A}(D'+D)^{-kC(C+1)v},$$

with the natural singular metric induced by the effective divisor $k\Gamma$, with $\mathscr{B} = \tau^*\mathscr{A}$, using the invariance of plurigenera and the fact that $\Gamma/(Cv)|_{X_{y'}}$ is klt for all $y' \in V'$. Consequently

$$g_*\omega^{k(C+1)v}_{Z/Y'} \otimes \tau^*\mathscr{A}(D'+D)^{-kC(C+1)v} \otimes \tau^*\mathscr{A}$$

is globally generated over V'.

By [Viehweg 1983, Lemma 3.2], we have a natural morphism

$$g_*\omega^{k(C+1)v}_{Z/Y'} \to \tau^* f_*\omega^{k(C+1)v}_{X/Y} \tag{A.2.1}$$

which is an isomorphism over V'. Since τ is finite, we can apply the projection formula to get a morphism

$$\bigoplus_i \tau_*\mathscr{O}_{Y'} \to f_*\omega^{k(C+1)v}_{X/Y} \otimes \tau_*\mathscr{O}_{Y'} \otimes \mathscr{A}(D+D')^{-kC(C+1)v} \otimes \mathscr{A}$$

which is surjective over V. Now we pick k sufficient large so that $\tau_*\mathscr{O}_{Y'} \otimes \mathscr{A}^{k(C+1)v-1}$ is globally generated. Therefore

$$f_*\omega^{K(C+1)v}_{X/Y} \otimes \tau_*\mathscr{O}_{Y'} \otimes \mathscr{A}^{-k(C-1)(C+1)v}(-kC(C+1)v(D+D'))$$

is generated by global sections over V, and so via the trace map so is

$$f_*\omega^{k(C+1)v}_{X/Y} \otimes \mathscr{A}^{-k(C-1)(C+1)v}(-kC(C+1)v(D+D')).$$

Setting $m = K(C+1)v$, $\mathscr{L} = \mathscr{A}^{C-1}$ and $D_Y = C(D+D')$, we obtain the first statement.

The second statement follows immediately from the first, noting that the assumption implies that the natural map

$$f^*f_*\omega^m_{X/Y} \to \omega^m_{X/Y}$$

is surjective over U. $\square$

Remark A.3. Recalling that $g'_*\omega^m_{Z'/Y'} \simeq g_*\omega^m_{Z/Y'}$, the proof of the proposition above shows more precisely that $f_*\omega^m_{X/Y} \otimes \mathscr{L}(D_Y)^{-m}$ is generated over V by sections belonging to the subspace $\mathbb{V}_m$ defined as the image of the natural map

$$H^0(Y', g'_*\omega^m_{Z'/Y'} \otimes \tau^*\mathscr{L}(D_Y)^{-m}) \to H^0(Y, f_*\omega^m_{X/Y} \otimes \mathscr{L}(D_Y)^{-m})$$

induced by (A.2.1).

Proposition A.4. *With notation as in the statement and proof of Proposition A.1 and Remark A.3, let $S \geq \Delta_\tau$ be an effective divisor and $\mu\colon \tilde{Y} \to Y$ a log resolution of $(Y, D+S)$ with centers in the singular locus of $D+S$. Then for every $s \in \mathbb{V}_m$, there exists a closed subset $T \subset \tilde{Y}$ of codimension at least 2, a birational model $\tilde{f}\colon \tilde{X} \to \tilde{Y}$ of f with $\tilde{X}$ smooth and projective, and a section*

$$\tilde{s} \in H^0(\tilde{Y}_0, \tilde{f}_*\omega^m_{\tilde{X}/\tilde{Y}} \otimes \mu^*\mathscr{L}(D_Y)^{-m}),$$

with $\tilde{Y}_0 = \tilde{Y} \setminus T$, such that

$$\tilde{s}|_{\mu^{-1}(V\setminus S)} = \mu^*(s|_{V\setminus S}).$$

Proof. By the definition of $\mathbb{V}_m$, we can lift s to a section

$$s' \in H^0(Y', g'_*\omega^m_{Z'/Y'} \otimes \tau^*\mathscr{L}(D_Y)^{-m}).$$

We set $\tilde{Y}''$ to be the normalization of the main component of $Y' \times_Y \tilde{Y}$ and $\tilde{\tau}'\colon \tilde{Y}'' \to \tilde{Y}$ the induced finite map. We compose this with a desingularization $\mu'\colon \tilde{Y}' \to \tilde{Y}''$, and get a birational map $\tilde{\mu}'\colon \tilde{Y}' \to Y'$. We then take $\tilde{X}$ to be a desingularization of the main component of $X \times_Y \tilde{Y}$, so that the induced morphism $\tilde{f}\colon \tilde{X} \to \tilde{Y}$ is a birational model of f. We obtain a mild reduction diagram for the new family $\tilde{f}$

$$\begin{tikzcd}
\tilde{X} \arrow[d, "\tilde{f}"] & \tilde{X}' \arrow[l] \arrow[d, "\tilde{f}'"] & \tilde{Z} \arrow[l, "\tilde{\sigma}"'] \arrow[d, "\tilde{g}"] & \tilde{X}''_1 \arrow[l, "\tilde{\rho}"'] \arrow[r, "\tilde{\delta}"] \arrow[d, "\tilde{f}''"] & \tilde{Z}' \arrow[d, "\tilde{g}'"] \\
\tilde{Y} & \tilde{Y}' \arrow[l, "\tilde{\tau}"'] & \tilde{Y}' \arrow[l, "="'] & \tilde{Y}' \arrow[l, "="'] \arrow[r, "="] & \tilde{Y}',
\end{tikzcd}$$

where $\tilde{\tau} = \tilde{\tau}' \circ \mu'$, a generically finite morphism, $\tilde{Z}' = Z' \times'_Y \tilde{Y}'$ and $\tilde{g}'$ is the induced mild morphism, and $\tilde{\sigma}$, $\tilde{\rho}$ and $\tilde{\delta}$ are similar to those in the mild reduction diagram for f.

In particular we have a Cartesian diagram

$$\begin{tikzcd}
Z' \arrow[d, "g'"] & \tilde{Z}' \arrow[l, "\tilde{\nu}'"'] \arrow[d, "\tilde{g}'"] \\
Y' & \tilde{Y}'. \arrow[l, "\tilde{\mu}'"']
\end{tikzcd}$$

By the mildness of the vertical morphisms we have that $\tilde{Z}'$ is normal with rational singularities, and so

$$\tilde{\nu}'^*\omega^m_{Z'/Y'} \simeq \omega^m_{\tilde{Z}'/\tilde{Y}'}.$$

Setting $\tilde{\mathscr{L}} = \mu^*\mathscr{L}$ and $\tilde{D}_{\tilde{Y}} = \mu^* D_Y$, we thus conclude that s' lifts to a section

$$\tilde{s}' = \tilde{\mu}'^* s' \in H^0(\tilde{Y}', \tilde{g}'_* \omega^m_{\tilde{Z}'/\tilde{Y}'} \otimes \tilde{\tau}^* \tilde{\mathscr{L}}(\tilde{D}_{\tilde{Y}})^{-m}).$$

Since $\tilde{\tau}$ is generically finite and branched over the normal crossing divisor $\mu^{-1}(\Delta_\tau)$, there exists a subset $T \subset \tilde{Y}$ of codimension at least 2 such that $\tilde{\tau}|_{\tilde{Y}_0}$ is finite and flat, where $\tilde{Y}_0 = \tilde{Y} \setminus T$. We are also allowed to assume that $\tilde{f}^{-1}(\mu^{-1}(S))$ is a simple normal crossing divisor, and hence so is $\tilde{f}^{-1}(\mu^{-1}(\Delta_\tau))$, by taking further blow-ups if necessary. Setting $\tilde{Y}'_0 = \tilde{\tau}^{-1}(\tilde{Y}_0)$, we deduce that $\tilde{X}'_0 = \tilde{f}'^{-1} Y'_0$ is normal with rational singularities. (For all of these statements, see e.g., [Viehweg 1983, Lemma 2.1].) Therefore, thanks to [Viehweg 1983, Lemma 3.2], there is a morphism

$$\tilde{g}'_* \omega^m_{\tilde{Z}'/\tilde{Y}'}|_{\tilde{Y}'_0} \to \tilde{\tau}^* \tilde{f}_* \omega^m_{\tilde{X}/\tilde{Y}}|_{\tilde{Y}'_0}, \tag{A.4.1}$$

which is identical to (A.2.1) over $\tau^{-1}(V \setminus S)$ (as μ is the identity over $V \setminus S$). This in turn induces

$$\tilde{\tau}_* \tilde{g}'_* \omega^m_{\tilde{Z}'/\tilde{Y}'}|_{\tilde{Y}_0} \to \tilde{\tau}_* \tilde{\tau}^* \tilde{f}_* \omega^m_{\tilde{X}/\tilde{Y}}|_{\tilde{Y}_0} \to \tilde{f}_* \omega^m_{\tilde{X}/\tilde{Y}}|_{\tilde{Y}_0},$$

where the last morphism is induced by the trace map $\tilde{\tau}_* \mathscr{O}_{\tilde{Y}'_0} \to \mathscr{O}_{\tilde{Y}_0}$. Finally, we conclude the existence of a morphism

$$\eta : \tilde{\tau}_* \tilde{g}'_* \omega^m_{\tilde{Z}'/\tilde{Y}'} \otimes \tilde{\mathscr{L}}(\tilde{D}_{\tilde{Y}})^{-m}|_{\tilde{Y}_0} \to \tilde{f}_* \omega^m_{\tilde{X}/\tilde{Y}} \otimes \tilde{\mathscr{L}}(\tilde{D}_{\tilde{Y}})^{-m}|_{\tilde{Y}_0}$$

and define

$$\tilde{s} = \eta(\tilde{s}'|_{\tilde{Y}'_0}) \in H^0(\tilde{Y}_0, \tilde{f}_* \omega^l_{\tilde{X}/\tilde{Y}} \otimes \mu^* \mathscr{L}(D_Y)^{-l}).$$

Since (A.2.1) and (A.4.1) are isomorphisms over $\tau^{-1}(V \setminus S)$, we have

$$\tilde{s}|_{\mu^{-1}(V \setminus S)} = \eta(\tilde{s}')|_{\mu^{-1}(V \setminus S)} = \mu^*(s|_{V \setminus S}).$$

□

Acknowledgements

We thank Laura DeMarco, Ya Deng, Henri Guenancia, Sándor Kovács, Mihai Păun, Erwan Rousseau, Christian Schnell and Sai-Kee Yeung for answering our questions and for useful suggestions. The authors owe a special thanks to Ariyan Javanpeykar, Ruiran Sun and Kang Zuo for pointing out a mistake in an earlier version of this paper, and the referee for helpful comments and references.

References

[Ahlfors 1961] L. V. Ahlfors, "Some remarks on Teichmüller's space of Riemann surfaces", *Ann. of Math.* (2) **74**:1 (1961), 171–191. MR Zbl

[Berndtsson et al. 2017] B. Berndtsson, M. Păun, and X. Wang, "Algebraic fiber spaces and curvature of higher direct images", preprint, 2017. arXiv

[Björk 1993] J.-E. Björk, *Analytic $\mathscr{D}$-modules and applications*, Math. Appl. **247**, Kluwer, Dordrecht, Netherlands, 1993. MR Zbl

[Brunebarbe 2018] Y. Brunebarbe, "Symmetric differentials and variations of Hodge structures", *J. Reine Angew. Math.* **743** (2018), 133–161. MR Zbl

[Campana and Păun 2019] F. Campana and M. Păun, "Foliations with positive slopes and birational stability of orbifold cotangent bundles", *Publ. Math. Inst. Hautes Études Sci.* **129**:1 (2019), 1–49. MR Zbl

[Cao 2016] J. Cao, "Albanese maps of projective manifolds with nef anticanonical bundles", 2016. To appear in *Ann. Sci. École Norm. Sup.* arXiv

[Cattani et al. 1986] E. Cattani, A. Kaplan, and W. Schmid, "Degeneration of Hodge structures", *Ann. of Math.* (2) **123**:3 (1986), 457–535. MR Zbl

[Deligne 1970] P. Deligne, *Équations différentielles à points singuliers réguliers*, Lecture Notes in Math. **163**, Springer, 1970. MR Zbl

[Demailly 1997] J.-P. Demailly, "Algebraic criteria for Kobayashi hyperbolic projective varieties and jet differentials", pp. 285–360 in *Algebraic geometry* (Santa Cruz, CA 1995), edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Demailly 2012] J.-P. Demailly, "Complex analytic and differential geometry", open content book, 2012, Available at http://www-fourier.ujf-grenoble.fr/~demailly/books.html.

[Deng 2018a] Y. Deng, "Kobayashi hyperbolicity of moduli spaces of minimal projective manifolds of general type", preprint, 2018. arXiv

[Deng 2018b] Y. Deng, "Pseudo Kobayashi hyperbolicity of base spaces of families of minimal projective manifolds with maximal variation", preprint, 2018. arXiv

[El Goul 2003] J. El Goul, "Logarithmic jets and hyperbolicity", *Osaka J. Math.* **40**:2 (2003), 469–491. MR Zbl

[Esnault and Viehweg 1992] H. Esnault and E. Viehweg, *Lectures on vanishing theorems*, DMV Seminar **20**, Birkhäuser, Basel, 1992. MR Zbl

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. MR Zbl

[Gieseker 1977] D. Gieseker, "Global moduli for surfaces of general type", *Invent. Math.* **43**:3 (1977), 233–282. MR Zbl

[Green and Griffiths 1980] M. Green and P. Griffiths, "Two applications of algebraic geometry to entire holomorphic mappings", pp. 41–74 in *The Chern symposium* (Berkeley, 1979), edited by H. H. Wu et al., Springer, 1980. MR Zbl

[Hacon et al. 2018] C. Hacon, M. Popa, and C. Schnell, "Algebraic fiber spaces over abelian varieties: around a recent theorem by Cao and Păun", pp. 143–195 in *Local and global methods in algebraic geometry* (Chicago, 2016), edited by N. Budur et al., Contemp. Math. **712**, Amer. Math. Soc., Providence, RI, 2018. MR Zbl

[Javanpeykar and Loughran 2018] A. Javanpeykar and D. Loughran, "Good reduction of Fano threefolds and sextic surfaces", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **18**:2 (2018), 509–535. MR Zbl

[Kawamata 1985] Y. Kawamata, "Minimal models and the Kodaira dimension of algebraic fiber spaces", *J. Reine Angew. Math.* **363** (1985), 1–46. MR Zbl

[Kebekus and Kovács 2008a] S. Kebekus and S. J. Kovács, "Families of canonically polarized varieties over surfaces", *Invent. Math.* **172**:3 (2008), 657–682. MR Zbl

[Kebekus and Kovács 2008b] S. Kebekus and S. J. Kovács, "Families of varieties of general type over compact bases", *Adv. Math.* **218**:3 (2008), 649–652. MR Zbl

[Kebekus and Kovács 2010] S. Kebekus and S. J. Kovács, "The structure of surfaces and threefolds mapping to the moduli stack of canonically polarized varieties", *Duke Math. J.* **155**:1 (2010), 1–33. MR Zbl

[Kobayashi 1970] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*, Pure Appl. Math. **2**, Dekker, New York, 1970. MR Zbl

[Kollár 1986] J. Kollár, "Higher direct images of dualizing sheaves, I", *Ann. of Math.* (2) **123**:1 (1986), 11–42. MR Zbl

[Kollár 1987] J. Kollár, "Subadditivity of the Kodaira dimension: fibers of general type", pp. 361–398 in *Algebraic geometry* (Sendai, 1985), edited by T. Oda, Adv. Stud. Pure Math. **10**, North-Holland, Amsterdam, 1987. MR Zbl

[Kovács 2000] S. J. Kovács, "Algebraic hyperbolicity of fine moduli spaces", *J. Algebraic Geom.* **9**:1 (2000), 165–174. MR Zbl

[Kovács and Lieblich 2010] S. J. Kovács and M. Lieblich, "Boundedness of families of canonically polarized manifolds: a higher dimensional analogue of Shafarevich's conjecture", *Ann. of Math.* (2) **172**:3 (2010), 1719–1748. Correction in **173**:1 (2011), 585–617. MR Zbl

[Lang 1986] S. Lang, "Hyperbolic and Diophantine analysis", *Bull. Amer. Math. Soc.* (*N.S.*) **14**:2 (1986), 159–205. MR Zbl

[McQuillan 1998] M. McQuillan, "Diophantine approximations and foliations", *Inst. Hautes Études Sci. Publ. Math.* **87** (1998), 121–174. MR Zbl

[Migliorini 1995] L. Migliorini, "A smooth family of minimal surfaces of general type over a curve of genus at most one is trivial", *J. Algebraic Geom.* **4**:2 (1995), 353–361. MR Zbl

[Patakfalvi 2012] Z. Patakfalvi, "Viehweg's hyperbolicity conjecture is true over compact bases", *Adv. Math.* **229**:3 (2012), 1640–1642. MR Zbl

[Pham 1979] F. Pham, *Singularités des systèmes différentiels de Gauss–Manin*, Progr. Math. **2**, Birkhäuser, Boston, 1979. MR Zbl

[Popa 2018] M. Popa, "Positivity for Hodge modules and geometric applications", pp. 555–584 in *Algebraic geometry* (Salt Lake City, 2015), edited by T. de Fernex et al., Proc. Sympos. Pure Math. **97**, Amer. Math. Soc., Providence, RI, 2018. MR

[Popa and Schnell 2013] M. Popa and C. Schnell, "Generic vanishing theory via mixed Hodge modules", *Forum Math. Sigma* **1** (2013), art. id. e1. MR Zbl

[Popa and Schnell 2017] M. Popa and C. Schnell, "Viehweg's hyperbolicity conjecture for families with maximal variation", *Invent. Math.* **208**:3 (2017), 677–713. MR Zbl

[Popa and Wu 2016] M. Popa and L. Wu, "Weak positivity for Hodge modules", *Math. Res. Lett.* **23**:4 (2016), 1139–1155. MR Zbl

[Rousseau 2018] E. Rousseau, "KAWA lecture notes on complex hyperbolic geometry", *Ann. Fac. Sci. Toulouse Math.* (6) **27**:2 (2018), 421–443. MR Zbl

[Royden 1975] H. L. Royden, "Intrinsic metrics on Teichmüller space", pp. 217–221 in *Proc. Int. Congr. Math., II* (Vancouver, 1974), edited by R. D. James, 1975. MR Zbl

[Saito 1990] M. Saito, "Mixed Hodge modules", *Publ. Res. Inst. Math. Sci.* **26**:2 (1990), 221–333. MR Zbl

[Schmid 1973] W. Schmid, "Variation of Hodge structure: the singularities of the period mapping", *Invent. Math.* **22** (1973), 211–319. MR Zbl

[Schumacher 2012] G. Schumacher, "Positivity of relative canonical bundles and applications", *Invent. Math.* **190**:1 (2012), 1–56. MR Zbl

[Schumacher 2018] G. Schumacher, "Moduli of canonically polarized manifolds, higher order Kodaira–Spencer maps, and an analogy to Calabi–Yau manifolds", pp. 369–399 in *Uniformization, Riemann–Hilbert correspondence, Calabi–Yau manifolds and Picard–Fuchs equations*, edited by L. Ji and S.-T. Yau, Adv. Lect. Math. **42**, Int. Press, Somerville, MA, 2018. MR Zbl

[Siu 1986] Y. T. Siu, "Curvature of the Weil–Petersson metric in the moduli space of compact Kähler–Einstein manifolds of negative first Chern class", pp. 261–298 in *Contributions to several complex variables*, edited by A. Howard and P.-M. Wong, Aspects Math. **E9**, Vieweg, Braunschweig, Germany, 1986. MR Zbl

[To and Yeung 2015] W.-K. To and S.-K. Yeung, "Finsler metrics and Kobayashi hyperbolicity of the moduli spaces of canonically polarized manifolds", *Ann. of Math.* (2) **181**:2 (2015), 547–586. MR Zbl

[To and Yeung 2018] W.-K. To and S.-K. Yeung, "Augmented Weil–Petersson metrics on moduli spaces of polarized Ricci-flat Kähler manifolds and orbifolds", *Asian J. Math.* **22**:4 (2018), 705–727. MR Zbl

[Viehweg 1983] E. Viehweg, "Weak positivity and the additivity of the Kodaira dimension for certain fibre spaces", pp. 329–353 in *Algebraic varieties and analytic varieties* (Tokyo, 1981), edited by S. Iitaka, Adv. Stud. Pure Math. **1**, North-Holland, Amsterdam, 1983. MR Zbl

[Viehweg 1995] E. Viehweg, *Quasi-projective moduli for polarized manifolds*, Ergebnisse der Mathematik (3) **30**, Springer, 1995. MR Zbl

[Viehweg and Zuo 2001] E. Viehweg and K. Zuo, "On the isotriviality of families of projective manifolds over curves", *J. Algebraic Geom.* **10**:4 (2001), 781–799. MR Zbl

[Viehweg and Zuo 2002] E. Viehweg and K. Zuo, "Base spaces of non-isotrivial families of smooth minimal models", pp. 279–328 in *Complex geometry* (Göttingen, Germany, 2000), edited by I. Bauer et al., Springer, 2002. MR Zbl

[Viehweg and Zuo 2003] E. Viehweg and K. Zuo, "On the Brody hyperbolicity of moduli spaces for canonically polarized manifolds", *Duke Math. J.* **118**:1 (2003), 103–150. MR Zbl

[Wolpert 1986] S. A. Wolpert, "Chern forms and the Riemann tensor for the moduli space of curves", *Invent. Math.* **85**:1 (1986), 119–145. MR Zbl

[Zuo 2000] K. Zuo, "On the negativity of kernels of Kodaira–Spencer maps on Hodge bundles and applications", *Asian J. Math.* **4**:1 (2000), 279–301. MR Zbl

Communicated by Gavril Farkas
Received 2019-03-08 Revised 2019-06-22 Accepted 2019-07-10

mihnea.popa@northwestern.edu *Department of Mathematics, Northwestern University, Evanston, IL, United States*

behrouz.taji@sydney.edu.au *School of Mathematics and Statistics, The University of Sidney, NSW, Australia*

lwu@math.utah.edu *Department of Mathematics, University of Utah, Salt Lake City, UT, 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 13 No. 9 2019
