THE INVERSE SIEVE PROBLEM FOR ALGEBRAIC VARIETIES OVER GLOBAL FIELDS

JUAN MANUEL MENCONI [1,2], MARCELO PAREDES [2] and ROMÁN SASYK [1,2]

Abstract. Let K be a global field and let Z be a geometrically irreducible algebraic variety defined over K. We show that if a big set $S \subseteq Z$ of rational points of bounded height occupies few residue classes modulo $\mathfrak{p}$ for many prime ideals $\mathfrak{p}$, then a positive proportion of S must lie in the zero set of a polynomial of low degree that does not vanish at Z. This generalizes the main result of Walsh in [Duke Math. J., vol.161, (2012), 2001-2022].

1. Introduction

Let S be a random set of integers. In arithmetic combinatorics, it is usual to establish "inverse theorems", in the sense that if S posses some specific arithmetic property, then S belongs to a certain family of subsets of integers; hence providing a classification for such kind of S'. For the purpose of this article, the arithmetic property in question to be studied is the equidistribution of the set S. Here, by an equidistributed subset of integers S we mean that S is well-distributed modulo p for many primes p (note that this is weaker than being well-distributed modulo m for many moduli m). We expect that a random set S is fairly well-distributed. Thus, an "inverse problem" here would be to understand whether a set that occupies few residue classes modulo p for many primes p has some specific structure. In this generality, this has been stated as follows.

Inverse Sieve Problem (see [CL07, HV09]). *Suppose that a set $S \subseteq \{0,\dots,N\}^d$ occupies very few residue classes mod p for many primes p. Then, either S is small, or it possesses some strong algebraic structure.*

In order to give a concrete example that motivated this sort of problem, consider a subset $S \subseteq \{0,\dots,N\}$ satisfying that $S_p := \{x \bmod p : x \in S\}$ has at most αp elements for many primes in the interval $[1,N]$, with $0<\alpha<1$. Gallagher's sieve [Gal71] implies that $|S| \leq c(\alpha)N^{\alpha}$. Let us further suppose that $|S_p| \leq \frac{p-1}{2}$ for all $p \leq N^{\frac{1}{2}}$. The large sieve implies (see [Mon68])

$$|S| \leq CN^{\frac{1}{2}}, \tag{1.1}$$

where C is an absolute constant. The bound (1.1) is essentially sharp, since if we consider S to be the set of the squares lying in $\{0,\dots,N\}$, we see that S occupies at most $\frac{1}{2}(p-1)$ residue classes for all primes p and $|S| \sim N^{\frac{1}{2}}$. More generally, if S is the image of a quadratic polynomial $aX^2+bX+c \in \mathbb{Z}[X]$, we also have that S occupies at most $\frac{p-1}{2}$ residue classes for all primes p not dividing a and $|S| \sim N^{\frac{1}{2}}$. Thus, we may ask if there are any other examples for which the bound (1.1) is almost optimal. This discussion on the large sieve together with the example of the squares, led Helfgott and Venkatesh [HV09] and independently Croot and Elsholtz [CL07] to conjecture that any badly distributed set S of size close to $N^{\frac{1}{2}}$ should "essentially" be the image of a quadratic polynomial. More precisely, they posed the following conjecture.

Conjecture 1.1 ([CL07, Problem 7.4], [HV09, Guess]). *Let $S \subseteq \{0,\dots,N\}$ of size $|S| \geq N^{\varepsilon}$ occupying less than αp residue classes for some $0<\alpha<1$ and every prime p. Then all but $O(N^{o(1)})$ elements of S are contained in the set of values of a polynomial $p \in \mathbb{Z}[X]$ with coefficients and degree bounded in terms of α and ε.*

2010 *Mathematics Subject Classification.* 11G50, 11G35, 11P70.

We remark that Conjecture 1.1 is known as the inverse problem for the large sieve (see [Gre08, Conjecture 1.4]). The name is due to the fact that Conjecture 1.1 classifies all sets of integers S that are obtained after sieving $\frac{p-1}{2}$ residue classes modulo p over the primes $p \leq N^{\frac{1}{2}}$, and that have size close to $N^{\frac{1}{2}}$. In [HV09, § 4.2] Helfgott and Venkatesh remarked that Conjecture 1.1 implies that there are $O(N^{\varepsilon})$ points on an irrational curve, which is considered a very hard problem. In the same article [HV09], the authors proved the following "higher dimensional" variant of Conjecture 1.1.

Theorem 1.2 ([HV09, Theorem 1.1])**.** *Let $\varepsilon, \alpha > 0$. There exist constants $c_1 = c_1(\alpha,\varepsilon)$ and $c_2 = c_2(\alpha,\varepsilon)$ such that the following holds. Let $S \subseteq \{0,\dots,N\}^2$ be a subset such that the number of residues $\{(x,y) \bmod p : (x,y) \in S\}$ is at most αp for every prime p. Then, at least one of the following holds:*

- *$|S| \leq c_1 N^{\varepsilon}$, or*
- *there exists a non-zero polynomial $P \in \mathbb{Z}[X,Y]$ of degree c_2 such that P vanishes in at least $(1-\varepsilon)|S|$ points of S.*

The argument of the proof of Theorem 1.2 is based on the larger sieve of Gallagher. Specifically, they adapted the determinant method of Bombieri-Pila (see [BP89]) to give a two dimensional version of Gallagher's sieve. Furthermore their methods gave another proof of the estimates of Bombieri-Pila for plane curves (see [BP89, Theorem 5]). While the proof of Theorem 1.2 is closely linked to the original proof in [BP89], it is remarkable that the method in [HV09] uses "local data", i.e. the size of the residue classes of the points on a curve, instead of analytic data as in [BP89], which is useful in other contexts (see, [Sed17], where an analogue of [BP89, Theorem 5] is obtained for function fields of genus 0).

Helfgott and Venkatesh conjectured that a similar result to Theorem 1.2 should hold for subsets of integers lying in $\mathbb{Z}^d$ for $d \geq 3$. Their methods, however, seem to handle only the case when S occupies very few residue classes, specifically at most αp residue classes for all primes p. Note that these sets are not what we would expect for a general badly distributed set, where the number of residues classes can be at most $O(p^{d-1})$. Using a subtle inductive argument, together with the larger sieve and the polynomial method, in [Wal12], Walsh solved this conjecture by proving the following theorem.

Theorem 1.3 ([Wal12, Theorem 1.1])**.** *Let $0 \leq k < d$ be integers and let $\varepsilon, \alpha, \eta > 0$ be positive real numbers. Then, there exists a constant C depending only on the above parameters, such that for any set $S \subseteq \{0,\dots,N\}^d$ occupying less than αp^k residue classes for every prime p at least one of the following holds:*

- *(S is small) $|S| \lesssim_{d,k,\varepsilon,\alpha,\eta} N^{k-1+\varepsilon}$;*
- *(S is strongly algebraic) there exists a non-zero polynomial $f \in \mathbb{Z}[X_1,\dots,X_d]$ of degree at most C and coefficients bounded by N^C vanishing at more than $(1-\eta)|S|$ points of S.*

Let us emphasize the important point that Theorem 1.3 means that there exist constants c, C depending on the parameters $d, k, \varepsilon, \alpha, \eta$ such that for any set S satisfying the hypothesis of the theorem, we have $|S| \leq cN^{k-1+\varepsilon}$ or there exists a non-zero polynomial $f \in \mathbb{Z}[X_1,\dots,X_d]$ of degree at most C and coefficients bounded by N^C vanishing at more than $(1-\eta)|S|$ points of S.

In this article we are interested in investigating the Inverse Sieve Problem in the context of global fields. To that end, let K be a global field and denote by $\mathcal{O}_K$ its ring of integers. A natural generalization to global fields of subsets lying in $\{0,\dots N\}^d$ is to consider the subsets $S \subseteq \{\mathbf{x} \in \mathcal{O}_K^d : H(\mathbf{x}) \leq N\}$ where H is a height function, in the sense of diophantine geometry. If such a subset S is small, or if it lies in the zero set of $\mathcal{O}_K$-points of an affine variety $Z \subseteq \mathbb{A}^d$ of dimension $l < d$ defined over K, then by classical bounds we have that for all primes $\mathfrak{p}$ in K, $Z(\mathcal{O}_K/\mathfrak{p})$ has at most $\lesssim_Z |\mathcal{O}_K/\mathfrak{p}|^l$ points. Thus, one may ask if a similar principle as in the Inverse Sieve Problem holds in the context of global fields. In this article we adapt the proof of Walsh to show that this is indeed the case. More precisely, we prove the following result.

Theorem 1.4. *Let $0 \le k < d$ be integers and let $\varepsilon, \alpha, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K. For a given $x \in K$ let $H(x)$ be the absolute multiplicative height of x. Then there exists a constant C depending only on the above parameters, such that for any set $S \subseteq \{x \in \mathcal{O}_K : H(x) \le N\}^d$ occupying less than $\alpha|\mathcal{O}_K/\mathfrak{p}|^k$ residue classes for every prime $\mathfrak{p}$ at least one of the following holds:*

- *(S is small) $|S| \lesssim_{d,k,\varepsilon,\alpha,K} N^{d_K(k-1)+\varepsilon}$;*
- *(S is strongly algebraic) there exists a non-zero polynomial $f \in \mathcal{O}_K[X_1, \dots, X_d]$ of degree at most C and coefficients bounded by N^C vanishing at more than $(1-\eta)|S|$ points of S.*

The reason why the exponent d_K appears in the first case in Theorem 1.4 is because we are counting $\mathcal{O}_K$-points with K a possibly non-trivial extension of $\mathbb{Q}$ and one expects to have a d_K-power of the usual quantities. For instance, the line $x = y$ has $\sim_{K,\varepsilon} N^{d_K+\varepsilon}$ $\mathcal{O}_K$-points of absolute height at most N.

In some situations, it is possible to have additional information about S, for instance, that S already lies in an affine variety Z defined over K, say geometrically irreducible. In this case, the statement of Theorem 1.4 is trivial, since its second condition already holds. However, it may happen that S occupies even fewer residue classes than Z. In this case, we can prove a sharper result than Theorem 1.4, as it can be seen in the next theorem.

Theorem 1.5. *Let $0 \le k < d$ be integers, $D, M > 0$ a positive integer, and let $\varepsilon, \alpha, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K. For a given $x \in K$ let $H(x)$ be the absolute multiplicative height of x. Then there exists a constant C depending only on the above parameters, such that for any set $S \subseteq \{x \in \mathcal{O}_K : H(x) \le N\}^M$ occupying less than $\alpha|\mathcal{O}_K/\mathfrak{p}|^k$ residue classes for every prime $\mathfrak{p}$ for every prime $\mathfrak{p}$, and that lies in an affine variety $Z \subseteq \mathbb{A}^M$ defined over K, geometrically irreducible of dimension d and degree D, at least one of the following holds:*

- *(S is small) $|S| \lesssim_{d,k,\varepsilon,\alpha,K,D,M} N^{d_K(k-1)+\varepsilon}$;*
- *(S is strongly algebraic) there exists a polynomial $f \in \mathcal{O}_K[X_1, \dots, X_M]$ of degree at most C vanishing at more than $(1-\eta)|S|$ points of S, that does not vanish at Z.*

In diophantine applications, one is usually interested in the rational points of some algebraic variety. Thus, one may ask if Theorem 1.5 admits a "projective version". Here we also prove such version. Furthermore, as in [Wal12], we only require that our sets occupy few residue classes for a "dense" subset of the primes. All this is summarized in the following theorem.

Theorem 1.6. *Let $0 \le k < d$ be integers, $D, M > 0$ positive integers, and let $\varepsilon, \alpha, \kappa, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K and $\mathcal{O}_K$ be its ring of integers. Set $Q = N^{\frac{\varepsilon}{2(d+1)}}$ and let*

$$P \subseteq \mathcal{P}(Q) := \{\mathfrak{p} \text{ prime of } \mathcal{O}_K : |\mathcal{O}_K/\mathfrak{p}| \le Q\}$$

be a subset of primes satisfying

$$w(P) := \sum_{\mathfrak{p} \in P} \frac{\log(|\mathcal{O}_K/\mathfrak{p}|)}{|\mathcal{O}_K/\mathfrak{p}|} \ge \kappa w(\mathcal{P}(Q)).$$

For a given $\boldsymbol{x} \in \mathbb{P}^M(K)$ let $H(\boldsymbol{x})$ be its absolute multiplicative height. Then there exists a constant C depending only on the above parameters, such that for any set $S \subseteq \{\boldsymbol{x} \in \mathbb{P}^M(K) : H(\boldsymbol{x}) \le N\}$ occupying less than $\alpha|\mathcal{O}_K/\mathfrak{p}|^k$ residue classes for every prime $\mathfrak{p}$ (i.e. the image of S in $\mathbb{P}^M(\mathcal{O}_K/\mathfrak{p})$ has at most $\alpha|\mathcal{O}_K/\mathfrak{p}|^k$ elements), and that lies in a projective variety $Z \subseteq \mathbb{P}^M$ defined over K, geometrically irreducible of dimension d and degree D, at least one of the following holds:

- *(S is small) $|S| \lesssim_{d,k,\varepsilon,\alpha,K,D,M} N^{d_K(k-1)+\varepsilon}$;*

- *(S is strongly algebraic) there exists an homogeneous polynomial $f \in \mathcal{O}_K[X_0, \ldots, X_M]$ of degree at most C vanishing at more than $(1-\eta)|S|$ points of S, that does not vanish at Z.*

As in [Wal12], it can be shown that Theorem 1.6 is sharp, and that ε can not be taken to be equal to zero.

The proofs that we present here follow the general strategy developed by Walsh in [Wal12]. However, given the nature of the statements of our theorems, several new difficulties arise. First we need to adapt two kinds of estimates over $\mathbb{Z}$ to the corresponding estimates over global fields: those concerning the behavior of heights and those concerning the distribution of primes. To that end, in section §2 we start by defining the height function that will be used throughout this article. While the theory of heights on number fields is very well documented, arguably this is not so in the case of function fields. Specifically, here we prove two statements for heights in function fields that we could not find in the literature. These are Proposition 2.1, that states that points in $\mathbb{P}^n$ of height $\lesssim 1$ are lifted to points in $\mathbb{A}^{n+1}$ of height $\lesssim_K 1$, and Proposition 2.2, which gives an upper bound for the number of points of bounded height in the ring of S-units in a function field. Both results are well known for number fields (see, for instance, [Ser89, §13.4] and [Lan83, §3] respectively). Concerning the distribution of primes, in section §3, after recalling Landau Prime Ideal Theorem for number fields and the Riemann hypothesis over function fields, we extend the larger sieve of Gallagher, as it was presented in [Wal12], to global fields.

The second kind of difficulty is that we work with sets lying in algebraic varieties, and that the bounds in our theorems are uniform in the degree and dimension of such algebraic varieties. We overcome it in section §4 in essentially two steps. First we use a standard argument to reduce Theorem 1.6 to a statement concerning affine varieties. Then, we use Noether's normalization theorem to make a change of variables and reduce Theorem 1.6 to a statement concerning badly distributed sets in an affine space. Because of the uniformity in our bounds, we need to have a nice control in the change of variables. Thus, we are led to prove in Theorem 4.2 an effective Noether's normalization theorem, which may be of interest in its own right.

Finally, in section §5 we prove Theorem 1.6. Unlike in the paper of Walsh [Wal12], we follow the dependence of the parameters in the proofs, which brings the last technical difficulty of this article. We believe that having made explicit the dependence of the constants may be useful for some diophantine applications.

2. Heights in global fields

The purpose of this section is twofold. First we establish a normalization of the absolute values of a global field. We use this to define the height function that will be used in this article and recall some basic properties of it. Secondly, we state two propositions concerning estimates for the number of points in the ring of S-units of a global field, which are well known for number fields, but for which we could not find a reference for function fields. The presentation in this section has been influenced by the standard references [BG06, HS00, Lan83, Ser89].

2.1. **Absolute values and relative height.**

Throughout this paper, K denotes a global field, i.e. a finite separable extension of $\mathbb{Q}$ or $\mathbb{F}_q(T)$, in which case we further assume that the field of constants is $\mathbb{F}_q$. We will denote by d_K the degree extension $[K : \Bbbk]$, where $\Bbbk$ indistinctively denotes the base fields $\mathbb{Q}$ or $\mathbb{F}_q(T)$.

Let K be a number field. Then each embedding $\sigma : K \hookrightarrow \mathbb{C}$ induces a place v, by means of the equation

$$||x||_v := |\sigma(x)|_\infty^{\frac{n_v}{d_K}},$$

where $|\cdot|_\infty$ denotes the absolute value of $\mathbb{R}$ or $\mathbb{C}$ and $n_v = 1$ or 2 respectively. Such places will be called the places at infinity, and denoted by $M_{K,\infty}$. Note that $\sum_{v \in M_{K,\infty}} n_v = d_K$. They are all the archimedean places of K. Since the complex embeddings come in pairs that differ by complex conjugation, we have $|M_{K,\infty}| \leq d_K$.

Now let $\mathfrak{p}$ be a non-zero prime ideal of K, and denote by $\operatorname{ord}_{\mathfrak{p}}$ the usual $\mathfrak{p}$-adic valuation. Associated to $\mathfrak{p}$, we have the place v in K given by the equation

$$||x||_v := |x|_{\mathfrak{p}} := \mathcal{N}_K(\mathfrak{p})^{-\frac{\operatorname{ord}_{\mathfrak{p}}(x)}{d_K}},$$

where $\mathcal{N}_K(\mathfrak{p})$ denotes the cardinal of the finite quotient $\mathcal{O}_K/\mathfrak{p}$. We will also denote $\mathcal{O}_{\mathfrak{p}}$ for the localization at $\mathfrak{p}$ of the ring $\mathcal{O}_K$. Such places will be called the finite places, and denoted by $M_{K,\mathrm{fin}}$. They are all the non-archimedean places of K.

The set of places of K is then the union $M_{K,\infty} \cup M_{K,\mathrm{fin}}$, and we denote it by M_K. For any finite subset $S \subseteq M_K$ containing the infinite places $M_{K,\infty}$, we define the ring of S-integers of K to be the set

$$\mathcal{O}_{K,S} := \{x \in K : ||x||_v \leq 1 \text{ for all } v \in M_K, v \notin S\}.$$

The norm of a non-zero ideal $I \subseteq \mathcal{O}_{K,S}$, denoted by $\mathcal{N}_{K,S}(I)$, is just the cardinal of the finite quotient $\mathcal{O}_{K,S}/I$. The prime ideals of $\mathcal{O}_{K,S}$ correspond to the prime ideals $\mathfrak{p}\mathcal{O}_{K,S}$ where $\mathfrak{p}$ is a prime ideal of $\mathcal{O}_K$ not lying in S.

Now, let us suppose that K is a function field over $\mathbb{F}_q$, such that $\mathbb{F}_q$ is algebraically closed in K (in other words, the constant field of K is $\mathbb{F}_q$). A prime in K is, by definition, a discrete valuation ring $\mathcal{O}$ with maximal ideal $\mathfrak{p}$ such that $\mathbb{F}_q \subseteq \mathfrak{p}$ and the quotient field of $\mathcal{O}$ equal to K. By abuse of notation, when we refer to a prime in K, we will refer to the maximal ideal $\mathfrak{p}$. We will also denote $\mathcal{O}_{\mathfrak{p}}$ to the corresponding discrete valuation ring. Associated to $\mathfrak{p}$, we have the usual $\mathfrak{p}$-adic valuation, that we will denote by $\operatorname{ord}_{\mathfrak{p}}$. The degree of $\mathfrak{p}$, denoted by $\deg(\mathfrak{p})$ will be the dimension of $\mathcal{O}_{\mathfrak{p}}/\mathfrak{p}$ as an $\mathbb{F}_q$-vectorial space, which is finite. Then the norm of $\mathfrak{p}$ is defined as $\mathcal{N}_K(\mathfrak{p}) = q^{\deg(\mathfrak{p})}$. Any prime $\mathfrak{p}$ of K induces a place v in K by the equation

$$||x||_v := |x|_{\mathfrak{p}} := \mathcal{N}_K(\mathfrak{p})^{-\frac{\operatorname{ord}_{\mathfrak{p}}(x)}{d_K}}.$$

They are all the places in K. The set of all places in K is denoted by M_K. As in the case of number fields, for any non-empty finite subset $S \subseteq M_K$, we define the ring of S-integers of K to be the set

$$\mathcal{O}_{K,S} := \{x \in K : ||x||_v \leq 1 \text{ for all } v \in M_K, v \notin S\}.$$

Given $x \in \mathcal{O}_{K,S}$ we define $\mathcal{N}_{K,S}(x) := \prod_{\mathfrak{p}\notin S} \mathcal{N}_K(\mathfrak{p})^{\operatorname{ord}_{\mathfrak{p}}(x)}$. By definition, $\operatorname{ord}_{\mathfrak{p}}(x) \geq 0$ for all $\mathfrak{p} \notin S$, so that $\mathcal{N}_{K,S}(x)$ is a positive integer. A prime in $\mathcal{O}_{K,S}$ will be any prime $\mathfrak{p} \in K$ not in S. When $S = \{v\}$, we will usually denote $\mathcal{O}_{K,S} = \mathcal{O}_K$. If $w \in M_{\Bbbk}$ is the place below v, we will denote $M_{K,\infty} := \{v' \in M_K : v'|w\}$. Note that $|M_{K,\infty}| \leq d_K$.

Now, given a global field K, we define the relative multiplicative projective height of K of a point $\mathbf{x} = (x_0 : \ldots : x_n) \in \mathbb{P}^n(K)$, to be the function

$$H_K(\mathbf{x}) := \prod_{v \in M_K} \max_i \{||x_i||_v\},$$

and the absolute multiplicative projective height by

$$H(\mathbf{x}) := H_K(\mathbf{x})^{\frac{1}{d_K}}.$$

If $x \in K$, $H_K(x)$ will always denote the projective height $H_K(1 : x)$. The next inequalities follow immediately from the definition of the height

$$H_K(x \cdot y) \leq H_K(x) \cdot H_K(y), \tag{2.1}$$

$$H_K(x + y) \leq 2^{d_K} H_K(x) H_K(y). \tag{2.2}$$

Also, from the product formula it follows that for all $x \in K^*$,

$$H_K(x) = H_K(x^{-1}). \tag{2.3}$$

For our purposes, it will be necessary to understand how the affine height of a point behaves under the action of a polynomial. It is easy to show (see [HS00, Proposition B.2.5. (a)]) that if $P(T_1,\dots,T_n) = \sum_{(i_1,\dots,i_n)} c_{i_1,\dots,i_n} T_1^{i_1}\cdots T_n^{i_n}$, $\mathbf{c} = (c_{i_1,\dots,i_n})_{i_1,\dots,i_n}$ and R is the number of $(i_1,\dots,i_n)$ with $c_{i_1,\dots,i_n} \neq 0$, we have

$$H_K(P(\mathbf{x})) \leq R^{d_K} H_K(1:\mathbf{c}) H_K(1:\mathbf{x})^{\deg(P)}. \tag{2.4}$$

Given a set of places S and $\mathbf{x} = (x_1,\dots,x_n) \in \mathcal{O}_{K,S}^n$, we have the bound

$$H_K(x_1:\ldots:x_n) \leq H_K(1:x_1:\ldots:x_n) \leq \max_i \{H_K(x_i)\}^h, \tag{2.5}$$

where $h = \operatorname{rank}(\mathcal{O}_{K,S}^\times)+1$, $\operatorname{rank}(\mathcal{O}_{K,S}^\times)$ being the rank as an abelian group of the units of the ring of S-integers. Also, for any $x \in \mathcal{O}_{K,S}\setminus\{0\}$, it holds

$$\mathcal{N}_K(x) \leq H_K(x). \tag{2.6}$$

In section §4 we will require to lift a bounded set in projective space to a set in affine space. The next proposition states that this can be done in a controlled manner.

Proposition 2.1. *Let K be a global field, let S be a finite set of places, with the additional condition that $M_{K,\infty} \subseteq S$ if K is a number field, and let $d \geq 1$ be an integer. There exists $c = c(K,S,d)$ such that for all $\boldsymbol{x} \in \mathbb{P}^d(K)$ there are coordinates $(y_0,\dots,y_d) \in \mathcal{O}_{K,S}^{d+1}$ such that*

$$H_K(1:y_0:\ldots:y_d) \leq cH_K(\boldsymbol{x}).$$

Thus, a subset $\mathcal{S} \subseteq [N]_{\mathbb{P}^d(K)}$ can be lifted to a subset $\overline{\mathcal{S}} \subseteq [cN]_{\mathbb{A}^{d+1}(\mathcal{O}_K)}$.

This is proved in [Ser89, §13.4] when K is a number field and $S = M_{K,\infty}$. For the sake of completeness, we include the proof of this more general statement in section §6.

In section §5 we will need estimates of the numbers of points in $\mathcal{O}_K$ of a given height. This is addressed by the following proposition.

Proposition 2.2 (S-integer points of bounded height)**.** *Let K be a global field, and let $S \subseteq M_K$ be a non-empty finite subset of places of K (which we require that contains the infinite places when K is a number field, and the place v fixed to define $\mathcal{O}_K$ when K is a function field). Then*

$$|\{x \in \mathcal{O}_{K,S} : H_K(x) \leq N\}| \leq c''(K) N (\log(N))^{|S|}.$$

When K is a number field, sharper estimates than Proposition 2.2 hold; for instance, see [Lan83, Theorem 5.2] for the case $S = M_{K,\infty}$ and [Bar15, Theorem 1.1] for arbitrary S. Moreover, [Bar15, Theorem 1.1] gives effective estimates. Since we could not find a reference for Proposition 2.2 over function fields we provide a proof in the appendix, section §6, of this article.

In the remaining of the paper we will use the following notation,

$$[N]_{\mathcal{O}_K}^n := \{\mathbf{x} = (x_1,\dots,x_n) \in \mathcal{O}_K^n : \max_i\{H_K(x_i)\} \leq N\}, \tag{2.7}$$

$$[N]_{\mathbb{A}^n(\mathcal{O}_K)} := \{\mathbf{x} = (x_1,\dots,x_n) \in \mathcal{O}_K^n : H_K(1:x_1:\ldots:x_n) \leq N\}. \tag{2.8}$$

$$[N]_{\mathbb{P}^n(K)} := \{\mathbf{x} = (x_0:\ldots:x_n) \in \mathbb{P}^n(K) : H_K(\mathbf{x}) \leq N\}. \tag{2.9}$$

Note that since $\max_i\{H_K(x_i)\} \leq H_K(1:x_1:\ldots:x_n)$ for all $\mathbf{x} = (x_1,\dots,x_n) \in K^n$ we have

$$[N]_{\mathbb{A}^n(\mathcal{O}_k)} \subseteq [N]_{\mathcal{O}_K}^n. \tag{2.10}$$

3. The larger sieve over global fields

In this section, we extend the larger sieve of Gallagher as it was presented in [Wal12], to global fields. To this end, we first recall some basic inequalities concerning the distribution of primes in global fields.

3.1. **Distribution of primes in global fields.**

Let K be a global field. For any prime $\mathfrak{p}$ of $\mathcal{O}_K$ we have a reduction map $\pi_\mathfrak{p} : \mathbb{P}^n(K) \to \mathbb{P}^n(\mathcal{O}_K/\mathfrak{p})$. If $\mathbf{x}, \mathbf{y} \in \mathbb{P}^n(K)$ by $\mathbf{x} \equiv \mathbf{y} (\mathrm{mod}\ \mathfrak{p})$ we will mean $\pi_\mathfrak{p}(\mathbf{x}) = \pi_\mathfrak{p}(\mathbf{y})$. Note that if $P \in \mathcal{O}_K[X_0, \ldots, X_n]$ is an homogeneous polynomial such that $P(\mathbf{x}) = 0$, then $\mathbf{x} \equiv \mathbf{y} (\mathrm{mod}\ \mathfrak{p})$ implies $P(\mathbf{y}) \equiv P(\mathbf{x}) \equiv 0 (\mathrm{mod}\ \mathfrak{p})$. Likewise we have a reduction map $\pi_\mathfrak{p} : \mathcal{O}_K^n \to (\mathcal{O}_K/\mathfrak{p})^n$, and for $\mathbf{x}, \mathbf{y} \in \mathcal{O}_K^n$ we will denote $\mathbf{x} \equiv \mathbf{y} (\mathrm{mod}\ \mathfrak{p})$ if $\pi_\mathfrak{p}(\mathbf{x}) = \pi_\mathfrak{p}(\mathbf{y})$.

If $S \subseteq [N]_{\mathbb{A}^n(\mathcal{O}_K)}, [N]^n_{\mathcal{O}_K}$, or $[N]_{\mathbb{P}^n(K)}$ and $\mathfrak{p}$ is a prime of $\mathcal{O}_K$ we will use the notation $[S]_\mathfrak{p} := \pi_\mathfrak{p}(S)$ where $\pi_\mathfrak{p}$ is the corresponding reduction map. For any $Q > 0$, let us denote:

$$\mathcal{P} := \{\mathfrak{p} \text{ prime in } \mathcal{O}_K\},$$

$$\mathcal{P}(Q) := \{\mathfrak{p} \in \mathcal{P} : \mathcal{N}_K(\mathfrak{p}) \leq Q\}.$$

If $P \subseteq \mathcal{P}(Q)$, we denote

$$w(P) := \sum_{\mathfrak{p} \in P} \frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})}.$$

If K is a global field, there exist constants $c_{1,K}, c_{2,K}, c_{3,K}$ and $c_{4,K}$ such that for all $Q > 0$ it holds that

$$c_{1,K} \log(Q) \leq w(\mathcal{P}(Q)) \leq c_{2,K} \log(Q), \tag{3.1}$$

and

$$c_{3,K} Q \leq \sum_{\mathfrak{p} \in \mathcal{P}(Q)} \log(\mathcal{N}_K(\mathfrak{p})) \leq c_{4,K} Q. \tag{3.2}$$

Indeed, if K is a number field, (3.1) and (3.2) follow from Landau Prime Ideal Theorem (see [IK04, Theorem 5.33]). Meanwhile, if K is a function field over $\mathbb{F}_q$ of genus g, this follows from the Riemann Hypothesis over function fields (see [Ros02, Theorem 5.12]). Note that in this case, the constants will also depend on the (degree of the) prime v that we choose to define $\mathcal{O}_K$.

3.2. **Larger sieve over global fields.**

Let $S \subseteq [N] \subseteq \mathbb{Z}$ and let $Q > 0$ be a parameter. The larger sieve of Gallagher, [Gal71], consists on counting in two different ways the number of pairs $x, y \in S$ and positive primes p with $p \leq Q$ such that $x \equiv y (\mathrm{mod}\ p)$. In our context, $S \subseteq [N]_{\mathcal{O}_K}$, and we want to count the number of pairs $x, y \in S$ and primes $\mathfrak{p} \in \mathcal{P}(Q)$ such that $x \equiv y (\mathrm{mod}\ \mathfrak{p})$. Thus, we have

$$\sum_{\substack{x,y \in S \\ x \neq y}} \sum_{\mathfrak{p} \in \mathcal{P}(Q)} 1_{x \equiv y (\mathrm{mod}\ \mathfrak{p})} \log(\mathcal{N}_K(\mathfrak{p})) = \sum_{\substack{x,y \in S \\ x \neq y}} \log\left(\prod_{\substack{\mathfrak{p} | x-y \\ \mathfrak{p} \in \mathcal{P}(Q)}} \mathcal{N}_K(\mathfrak{p}) \right) \leq \sum_{\substack{x,y \in S \\ x \neq y}} \log(\mathcal{N}_K(x-y)). \tag{3.3}$$

Using (2.6) and (2.2) we obtain $\mathcal{N}_K(x-y) \leq 2^{d_K} H_K(x) H_K(y)$, that is smaller than N^3 if $N > 2^{d_K}$. Hence

$$\sum_{\substack{x,y \in S \\ x \neq y}} \sum_{\mathfrak{p} \in \mathcal{P}(Q)} 1_{x \equiv y (\mathrm{mod}\ \mathfrak{p})} \log(\mathcal{N}_K(\mathfrak{p})) \leq 3|S|^2 \log(N). \tag{3.4}$$

On the other hand, if $S(a, \mathfrak{p}) := \{x \in S : x \equiv a (\mathrm{mod}\ \mathfrak{p})\}$, the left hand side of (3.4) is equal to

$$\sum_{\mathfrak{p} \in \mathcal{P}(Q)} \sum_{a (\mathrm{mod}\ \mathfrak{p})} |S(a, \mathfrak{p})|^2 \log(\mathcal{N}_K(\mathfrak{p})) - |S| \sum_{\mathfrak{p} \in \mathcal{P}(Q)} \log(\mathcal{N}_K(\mathfrak{p})). \tag{3.5}$$

Thus, (3.4) and (3.5) imply

$$\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\sum_{a(\bmod\ \mathfrak{p})}|S(a,\mathfrak{p})|^2\log(\mathcal{N}_K(\mathfrak{p}))-|S|\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\log(\mathcal{N}_K(\mathfrak{p}))\leq 3|S|^2\log(N). \tag{3.6}$$

Note that the above argument also works if $S\subseteq[N]^d_{\mathcal{O}_K}$. Indeed, if $\pi_1: K^n\to K$ denotes the projection on the first coordinate, then $\mathbf{x}\equiv\mathbf{y}(\bmod\ \mathfrak{p})$ implies $\mathfrak{p}|\pi_1(\mathbf{x})-\pi_1(\mathbf{y})$, thus the argument to prove (3.6) still holds.

Lemma 3.1 (Compare to [Wal12, Lemma 3.1])**.** *Let $X\subseteq[N]_{\mathcal{O}_K}$, $Q=N^\gamma$, $\gamma>0$. Let κ,μ be positive real numbers. Suppose that there is a set of primes $P\subseteq\mathcal{P}(Q)$ with $w(P)\geq\kappa w(\mathcal{P}(Q))$ such that for any prime $\mathfrak{p}\in P$ there are at least $\mu|X|$ elements of X in at most $\alpha\mathcal{N}(\mathfrak{p})$ residue classes for some $\alpha>0$ independent of $\mathfrak{p}$. Then, there exists $C_1=C_1(\kappa,\mu,\gamma,K)$ such that if $\alpha\leq C_1$, it must be $|X|<Q$.*

Lemma 3.2 (Compare to [Wal12, Lemma 3.2])**.** *Let $Q=N^\gamma$ for some $\gamma>0$ and let $P\subseteq\mathcal{P}(Q)$ be a set of primes with $w(P)\geq w(\mathcal{P}(Q))$ for some $\kappa>0$. Let $S\subseteq[N]^d_{\mathcal{O}_K}$ occupying less than α residue classes modulo $\mathfrak{p}$ for every prime $\mathfrak{p}\in P$ and some constant α, independent of $\mathfrak{p}$. Then there exists a constant $C_2=C_2(\alpha,\kappa,\gamma,K)$ such that $|S|\leq C_2$.*

Lemma 3.1 and Lemma 3.2 can we proved using (3.6). The proofs are analogous to the corresponding proofs in [Wal12], so we will not include them here. Also we remark that the constants C_1 and C_2 may be taken to be

$$C_1(\kappa,\mu,\gamma,K)=\frac{\kappa\mu^2\gamma}{c_{5,K}},\ c_{5,K}:=\frac{2(c_{4,K}+3)}{c_{1,K}}, \tag{3.7}$$

$$C_2(\alpha,\kappa,\gamma,K)=\max\left\{2\alpha,2\left(\frac{12\alpha}{c^2_{1,K}\gamma\kappa}\right)^{2^{\frac{c_{3,K}}{\gamma\kappa}}}\right\}. \tag{3.8}$$

4. An effective change of variables

The main result of this section is that Theorem 1.6 follows from a similar result concerning affine spaces. In order to do that, we are going to make two reductions. The first step is to reduce our problem to that of studying subsets of affine varieties. We note that this is standard, for instance, see [Ser89, Chapter 14], where an upper bound for the number of rational points on a thin set lying in a projective space is obtained as a consequence of an analogous bound for thin sets lying in an affine space. The second step is to further reduce this problem by passing from an affine variety of dimension d to the affine space $\mathbb{A}^d$. In order to do this, we are going to make a change of variables. This will be achieved by means of Noether's normalization. Since Theorem 1.6 is uniform in the degree and dimension of the variety, we will require some effectiveness in the normalization. To that end, we are going to provide an effective version of the Noether's normalization theorem, with quite elementary methods, which is interesting in its own right.

4.1. Reduction to the affine case.

In order to carry on the first step in the discussion above, we proceed to state the following variant of Theorem 1.6 for affine varieties.

Theorem 4.1 (Affine case of Theorem 1.6)**.** *Let $0\leq k<d$ be integers, D,M be positive integers, and let $\varepsilon,\alpha,\kappa,\eta>0$ be positive real numbers. Let K be a global field of degree d_K and $\mathcal{O}_K$ be its ring of integers. Set $Q=N^{\frac{\varepsilon}{2(d+1)}}$ and let $P\subseteq\mathcal{P}(Q)$ be a subset satisfying $w(P)\geq\kappa w(\mathcal{P}(Q))$. Then there exists a constant C depending only on the above parameters, such that for any set $S\subseteq[N]_{\mathbb{A}^{M+1}(\mathcal{O}_K)}$ occupying less than $\alpha\mathcal{N}_K(\mathfrak{p})^k$ residue classes for every prime $\mathfrak{p}$, and that lies in an affine variety $Z\subseteq\mathbb{A}^{M+1}$ defined over K, geometrically irreducible of dimension $d+1$ and degree D, at least one of the following holds:*

- *(S is small)* $|S| \lesssim_{d,k,\varepsilon,K,D,M} N^{k-1+\varepsilon}$;
- *(S is strongly algebraic) there exists a homogeneous polynomial $f \in \mathcal{O}_K[X_0,\dots,X_M]$ of degree at most C vanishing at more than $(1-\eta)|S|$ points of S, that does not vanish at Z.*

There are two remarks to be made about the statement in Theorem 4.1. First, in the case when S is small, d_K does not appear in the exponent of N. This is due to the fact that here we are using the relative height to K instead of the absolute height as in Theorem 1.6. The reason for this is because it will simplify some of the cumbersome notation in our proofs, and also because it reflects more accurately the nature of the problem we are studying, which is relative to the global field K. The second remark is about the parameters k, d and M; the choice of $M+1$ instead of M and Z of dimension $d+1$ instead of dimension d are because we will lift a subset $S \subseteq Z \subseteq \mathbb{P}^M$ occupying less than $\lesssim \mathcal{N}_K(\mathfrak{p})^k$ residue classes for every prime $\mathfrak{p}$, with Z a projective variety. Thus, the lifted set $\overline{S}$ will lie in the affine cone $C(Z) \subseteq \mathbb{A}^{M+1}$ which is a geometrically irreducible variety of dimension $d+1$ and degree D.

Proof that Theorem 4.1 implies Theorem 1.6. Let S be as in the statement of Theorem 1.6. In particular, $|[S]_{\mathfrak{p}}| \leq \alpha\mathcal{N}_K(\mathfrak{p})^k$ for all prime $\mathfrak{p}$ and some positive α and $0 \leq k < \dim(Z)$. By Proposition 2.1, we can lift S to a subset $\overline{S} \subseteq [cN]_{\mathbb{A}^{M+1}(\mathcal{O}_K)}$ lying in the affine cone $C(Z) \subseteq \mathbb{A}^{M+1}$, where c is some positive constant depending only on K. Note that $C(Z)$ has dimension $\dim(Z)+1$ and degree $\deg(Z)$. Now, let $\mathfrak{p} \in P$. Let us bound $[\overline{S}]_{\mathfrak{p}}$. Given $\mathbf{x}' = (x_0,\dots,x_{M+1}) \in \overline{S}$, there are two possibilities: $\mathfrak{p}|x_i$ for all i, or there exists i_0 such that $\mathfrak{p} \nmid x_{i_0}$. In the second case, $(x_0 (\mathrm{mod}\ \mathfrak{p}),\dots,x_{M+1}(\mathrm{mod}\ \mathfrak{p}))$ defines a point in $\mathbb{P}^M(\mathcal{O}_K/\mathfrak{p})$ which coincides with the reduction modulo $\mathfrak{p}$ of $\mathbf{x} = (x_0 : \dots : x_{M+1}) \in S$. Thus, such points are bounded by $|[S]_{\mathfrak{p}}| \leq \alpha\mathcal{N}_K(\mathfrak{p})^k$. Meanwhile, in the first case, we just have that the points reduce to the 0 class in $(\mathcal{O}_K/\mathfrak{p})^{M+1}$. We conclude that $|[\overline{S}]_{\mathfrak{p}}| \leq \alpha\mathcal{N}_K(\mathfrak{p})^k + 1 \leq \max\{\alpha, 1\}\mathcal{N}_K(\mathfrak{p})^k = \alpha'\mathcal{N}_K(\mathfrak{p})^k$. Hence, $\overline{S}$ satisfies the hypothesis of Theorem 4.1 and at least one of the following holds:

- $|S| = |\overline{S}| \lesssim_{d,k,\varepsilon,K,D,M} (cN)^{k-1+\varepsilon} \lesssim_{d,k,\varepsilon,K,D,M} N^{k-1+\varepsilon}$;
- there exists an homogeneous polynomial $f \in \mathcal{O}_K[X_0,\dots,X_M]$ of degree at most $C = O_{d,k,\varepsilon,K,D,M,\eta}(1)$ and coefficients bounded by N^C vanishing at more than $(1-\eta)|\overline{S}| = (1-\eta)|S|$ points of $\overline{S}$, that does not vanish at $C(Z)$.

From this we deduce Theorem 1.6. □

4.2. **An effective Noether's normalization.** At this stage we have reduced Theorem 1.6 to a problem about affine varieties. In order to carry on the second step discussed in the introduction of section §4, here we prove the following version of Noether's normalization theorem.

Theorem 4.2 (Effective Noether's normalization theorem)**.** *Let $V \subseteq \mathbb{P}^m$ be an irreducible projective variety defined over a global field K. Then there exists a finite map $\varphi : V \to \mathbb{P}^{\dim(V)}$, defined over K, such that*

$$\varphi(\boldsymbol{x}) = (L_0(\boldsymbol{x}) : \dots : L_{\dim(V)}(\boldsymbol{x})),$$

with L_i linear forms with coefficients on $\Bbbk$ of height bounded by $\lesssim_{\Bbbk,m} (\deg(V))^{m-\dim(V)}$ where the implicit constant is effectively computable. Moreover, each fibre of φ has at most $\deg(V)$ elements. In particular, the same statement holds for an affine variety $Z \subseteq \mathbb{A}^m$.

Proof. Let V be an irreducible variety as in the statement of the theorem. If $V = \mathbb{P}^m$, we take φ the identity map and we are done. If $V \subsetneq \mathbb{P}^m$, then there exists $\mathbf{x} \in \mathbb{P}^m(\overline{\Bbbk}) \backslash V(\overline{\Bbbk})$ (here we are using that K is infinite). Furthermore, let us see that we may choose $\mathbf{x}$ with coordinates in K, and of small height. If $\dim(V) = m-1$, V is a hipersurface. If not, we reduce to the hypersurface case by means of a standard geometrical idea (see, for instance, [Mum70, Theorem 1], or its reprint in [Mar10]). Indeed, for that, choose a generic projective subspace W of $\mathbb{P}^m(\overline{\Bbbk})$ of dimension

$m - \dim(V) - 2$, and consider the cone $C(W, V)$ formed by taking the union of all the lines joining a point in W with a point in V. It is generically a projective subvariety of dimension $m-1$, i.e. an hypersurface. Furthermore, it has degree $\deg(V)$, so $C(W, V)$ is defined by a polynomial with coefficients on K, of degree $\deg(V)$. In any case, we then have that V is contained in an hypersurface $\mathcal{Z}(P) \subseteq \mathbb{P}^m(\overline{\mathbb{k}})$, where $P \in \overline{\mathbb{k}}[T_0, \ldots, T_m]$ is a non-zero homogeneous polynomial of degree $\deg(V)$.

Let us suppose that P has a non-zero coefficient at $T_0^{d_0} \cdots T_m^{d_m}$. Consider the set $[\deg(V)]_{\mathcal{O}_{\mathbb{k}}} = \{x \in \mathcal{O}_{\mathbb{k}} : H(x) \leq \deg(V)\}$ (note that here we are using the absolute height instead of the relative height to K). It has strictly more than $\deg(V) \geq d_i$ elements. By the Combinatorial Nullstellensatz [TV10, Theorem 9.2], there exists $x_0, \ldots, x_n \in [\deg(V)]_{\mathcal{O}_{\mathbb{k}}}$ such that $P(x_0, \ldots, x_m) \neq 0$. In particular $(x_0, \ldots, x_m) \neq 0$. Let $\mathbf{x}_1 \in \mathbb{P}^m$ be the point with projective coordinates $(x_0 : \ldots : x_m)$. By construction, $\mathbf{x}_1 \in \mathbb{P}^m(\mathbb{k}) \backslash V(\mathbb{k})$. Now, construct linear forms $L_{1,1}(T_0, \ldots, T_m), \ldots, L_{1,m}(T_0, \ldots, T_m) \in \mathbb{k}[T_0, \ldots, T_m]$ such that $\mathcal{Z}(L_{1,1}, \ldots, L_{1,m}) = \{\mathbf{x}_1\}$. Note that the coefficients of such linear forms are a basis of the vector space $V_1 := \langle (x_0, \ldots, x_m) \rangle^{\perp}$ defined over $\mathbb{k}$. Thus, if we want to construct the linear forms $L_{1,i}$'s with coefficients of small height, it is enough to find a basis of small height of V_1. For this, we use the generalizations of Siegel's lemma for number fields and function fields in [BV83, Thu95, Fuk10] to find a basis $\mathbf{y}_1, \ldots, \mathbf{y}_m \in K^{m+1}$ of V_1 such that

$$\prod_{i=1}^{m} H(1 : \mathbf{y}_i) \lesssim_{\mathbb{k},m} H(V_1), \tag{4.1}$$

where $H(V_1)$ is the height of V_1. By the duality theorem [Thu93, Duality Theorem], it holds that $H(V_1)$ coincides with the height of the linear subspace generated by the $\mathbb{k}$-vector $(x_0, \ldots, x_m)$. Moreover this height coincides with the projective height $H(x_0 : \ldots : x_m)$. Thus

$$\prod_{i=1}^{m} H(1 : \mathbf{y}_i) \lesssim_{\mathbb{k},m} H(x_0 : \ldots : x_m) \lesssim_{\mathbb{k},m} \deg(V). \tag{4.2}$$

We define $\varphi_1 : V \to \mathbb{P}^{m-1}$ as the projection away from $\mathbf{x}_1$, that is

$$\varphi_1(\mathbf{x}) = (L_{1,1}(\mathbf{x}) : \ldots : L_{1,m}(\mathbf{x})). \tag{4.3}$$

Thus φ_1 is a finite morphism (see [Sha74, Chapter 1, § 5.3, Theorem 7]), with $L_{1,1}, \ldots, L_{1,m}$ linear forms with coefficients of height bounded by $\lesssim_{\mathbb{k},m} \deg(V)$. If $\varphi_1(V) = \mathbb{P}^{m-1}$, we are done. Now suppose otherwise. Then for a generic linear space $L \subseteq \mathbb{P}^{m-1}$ of codimension $\dim(V)$, it holds that $L \cap \varphi_1(V)$ is finite. Furthermore, the pre-image of $L \cap \varphi_1(V)$ by φ_1 is finite (because φ_1 is a finite morphism, hence it has finite fibers), and this intersection is equal to the intersection of V by some linear subspace $L' \subseteq \mathbb{P}^m$ of codimension $\dim(V)$ (a finite morphism preserves the dimension). Thus $|L \cap \varphi_1(V)| \leq |L' \cap V| \leq \deg(V)$, from where we conclude that the degree of $\varphi_1(V)$ is at most $\deg(V)$.

In conclusion, the projective irreducible variety $\varphi_1(V)$ has dimension $\dim(V)$ and $\deg(\varphi_1(V)) \leq \deg(V)$. Hence we can repeat the same argument as above and obtain a sequence of finite maps $\varphi_{i+1} : \varphi_i(V) \to \mathbb{P}^{m-i+1}$, defined by

$$\varphi_{i+1}(\mathbf{x}) = (L_{i+1,1}(\mathbf{x}) : \ldots : L_{i+1,m-i+1}(\mathbf{x})), \tag{4.4}$$

with $L_{i+1,1}, \ldots, L_{i+1,m-i+1}$ linear forms with coefficients in $\mathbb{k}$ of height bounded by $\lesssim_{\mathbb{k},m} \deg(V)$. Since the sequence $\varphi_1, \varphi_2, \ldots$ ends with $i = m - \dim(V)$, we conclude that there exists a finite morphism $\varphi : V \to \mathbb{P}^{\dim(V)}$ such that

$$\varphi(\mathbf{x}) = (L_0(\mathbf{x}) : \ldots : L_{\dim(V)}(\mathbf{x})), \tag{4.5}$$

with $L_0, \ldots, L_m$ linear forms with coefficients in $\mathbb{k}$ of height bounded by $\lesssim_{\mathbb{k},m} \deg(V)^{m-\dim(V)}$.

Finally, note that the morphism $\varphi : V \to \mathbb{P}^{\dim(V)}$ that we constructed can be interpreted geometrically as the projection away from a generic linear subspace $L \subseteq \mathbb{P}^m$ of codimension $\dim(V) + 1$. Thus, given $\mathbf{z} \in \mathbb{P}^{\dim(V)}$ the points in the fibre $\varphi^{-1}(\mathbf{z})$ correspond to the points lying in the intersection of the affine cone $C(V) \subseteq \mathbb{A}^{m+1}$ of V with respect to an affine linear subspace $L' \subseteq \mathbb{A}^{m+1}$. Since the affine cone $C(V)$ has the same degree of V, we conclude that $|\varphi^{-1}(\mathbf{z})| \leq |C(Z) \cap : L'| \leq \deg(V)$. ◻

Remark 4.3. Professor Sombra informed us that in the literature there are other effective versions of the Noether's normalization theorem for number fields. For instance, [KPS01, Lemma 2.14 and Proposition 4.5] imply that for an affine variety V there exist linear forms in Noether position with height bounded by $2\deg(V)^2$. We wish to thank Prof. Sombra and Prof. Dickenstein for this.

4.3. **Reduction to the affine plane case.** Now we are in condition to make the last reduction of Theorem 1.6. Specifically, by means of an effective change of variables, we will reduce the proof of Theorem 4.1 to proving the next statement.

Theorem 4.4 (Case $Z = \mathbb{A}^{d+1}$)**.** *Let $0 \leq k < d$ be integers, and let $\varepsilon, \alpha, \kappa, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K and $\mathcal{O}_K$ be its ring of integers. Set $Q = N^{\frac{\varepsilon}{2(d+1)}}$ and let $P \subseteq \mathcal{P}(Q)$ be a subset satisfying $w(P) \geq \kappa w(\mathcal{P}(Q))$. Then there exists a constant C depending only on the above parameters, such that for any set $S \subseteq [N]^{d+1}_{\mathcal{O}_K}$ occupying less than $\alpha \mathcal{N}_K(\mathfrak{p})^k$ residue classes for every prime $\mathfrak{p}$, at least one of the following holds:*

- *(S is small) $|S| \lesssim_{d,k,\varepsilon,K} N^{k-1+\varepsilon}$;*
- *(S is strongly algebraic) there exists a non-zero homogeneous polynomial $f \in \mathcal{O}_K[X_0, \dots, X_d]$ of degree at most C vanishing at more than $(1-\eta)|S|$ points of S.*

Proof that Theorem 4.4 implies Theorem 4.1. Let $S \subseteq Z \subseteq \mathbb{A}^{M+1}$ be as in Theorem 4.1. By Theorem 4.2 there exists a finite map $\mathbf{F} = (F_1, \dots, F_{d+1}) : Z \to \mathbb{A}^{d+1}$ such that each for all i, $F_i \in \mathcal{O}_{\Bbbk}[X_0, \dots, X_{M+1}]$ is a linear form with coefficients of height bounded by $\lesssim_{\Bbbk,M,d,D} 1$. Let $\overline{S} = \mathbf{F}(S)$. Note that by (2.4) and (2.10) it holds $\overline{S} \subseteq [cN]_{\mathbb{A}^{d+1}(\mathcal{O}_K)} \subseteq [cN]^{d+1}_{\mathcal{O}_K}$ for some $c = O_{K,M,d,D}(1)$. Since $\mathbf{F}$ preserves congruences, for any prime $\mathfrak{p} \in P$ it holds $|[\overline{S}]_{\mathfrak{p}}| \leq |[S]_{\mathfrak{p}}| \leq \alpha \mathcal{N}_K(\mathfrak{p})^k$ with $k < d$. Hence, $\overline{S}$ is in the conditions of Theorem 4.4, so apply the theorem to $\overline{S}$ with $\eta = \frac{1}{2}$ to conclude that:

- $|\overline{S}| \lesssim_{d,k,\varepsilon,K} N^{k-1+\varepsilon}$; or
- there exists an homogeneous polynomial $g \in \mathcal{O}_K[Y_0, \dots, Y_d]$ of degree at most C vanishing at more than $\frac{1}{2}|\overline{S}|$ points of $\overline{S}$.

Suppose that the first possibility occurs. Since $\mathbf{F}$ has degree at most $\deg(Z) = D$, we have $|\overline{S}| \geq \frac{|S|}{\deg(Z)}$, so we deduce $|S| \lesssim_{d,k,\varepsilon,K,D} N^{k-1+\varepsilon}$. If the second possibility occurs, using again that $\mathbf{F}$ has degree at most $\deg(Z) = D$, we conclude that $g \in \mathcal{O}_K[Y_0, \dots, Y_d]$ is an homogeneous polynomial of degree at most C vanishing at more than $\frac{1}{2}|\overline{S}| \geq \frac{1}{2D}|S|$ points of S. Let $f(X_0, \dots, X_{M+1}) := g(\mathbf{F}(X_0, \dots, X_{M+1}))$. Since $g \neq 0$ and $\mathbf{F}$ is surjective, we conclude that f is an homogeneous polynomial of degree at most C, vanishing on at least $\frac{1}{2D}|S|$ points of S, that does not vanish at Z. Thus, we conclude Theorem 4.1 for $\eta_0 = \frac{1}{2D}$. We conclude Theorem 4.1 for any η by a simple partition argument. ◻

5. The inverse sieve problem in projective varieties

In this section we are going to prove Theorem 4.4. More precisely, we prove the following stronger version.

Theorem 5.1. *Let d, h be positive integers and let $\varepsilon, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K and $\mathcal{O}_K$ be its ring of integers. Set $Q = N^{\frac{\varepsilon}{2(d+1)}}$ and let $P \subseteq \mathcal{P}(Q)$ satisfying $w(P) \geq \kappa w(\mathcal{P}(Q)))$ for some $\kappa > 0$. Suppose that $S \subseteq [N]^{d+1}_{\mathcal{O}_K}$ is a set of size $|S| \geq cN^{d-h-1+\varepsilon}$ occupying at most $\alpha \mathcal{N}_K(\mathfrak{p})^{d-h}$ residue classes* *modulo* $\mathfrak{p}$ *for every prime* $\mathfrak{p} \in P$ *and some* $\alpha > 0$*. Then if* N *is sufficiently large there exists a non-zero homogeneous polynomial* $f \in \mathcal{O}_K[X_0, \ldots, X_d]$ *of degree* $O_{d,h,\varepsilon,\eta,\kappa,K}(1)$ *and coefficients bounded by* $N^{O_{d,h,\varepsilon,\eta,\kappa,K}(1)}$ *which vanishes at more that* $(1-\eta)|S|$ *points of* S*.*

In order to prove this theorem, we follow the proof of [Wal12, Theorem 2.4]. The idea of the proof is to construct a small "characteristic set" $A \subseteq S$ with the property that any "small" polynomial that vanishes at A also vanishes at a positive proportion of S. This will be done in Proposition 5.10, which is adapted from [Wal12, Proposition 2.2]. Then, by means of a variant of Siegel's lemma, we construct such a small polynomial, which will exist because of the small size of A. Albeit the proof presented here follows the same steps of the proof of Walsh, some new technical difficulties arise, that where already discussed in the introduction of this article. In order to aid the reader, we closely follow the notations and definitions used in [Wal12].

5.1. **Genericity.** First we extend the notion of "generic set" of [Wal12], and then prove that any set satisfying the conditions of Theorem 5.1 contains large generic subsets. For this, we introduce the following notation. Given $S \subseteq [N]^d_{\mathcal{O}_K}$, $\mathbf{a} \in (\mathcal{O}_K/\mathfrak{p})^d$, $a \in \mathcal{O}_K/\mathfrak{p}$ and $x \in [N]_{\mathcal{O}_K}$,

$$S(\mathbf{a}, \mathfrak{p}) := \{\mathbf{x} = (x_1, \ldots, x_d) \in S : \pi_{\mathfrak{p}}(\mathbf{x}) = \mathbf{a}\},$$

$$S(a, \mathfrak{p}) := \{\mathbf{x} = (x_1, \ldots, x_d) \in S : x_1 \equiv a (\mathrm{mod}\ \mathfrak{p})\},$$

$$S_x := S \cap \pi_1^{-1}(x)$$

Definition 5.2 (Genericity)**.** Given a real number $B > 0$ and some integer $l \geq 0$ we say that a set $S \subseteq [N]^d_{\mathcal{O}_K}$ is (B, l)-generic modulo $\mathfrak{p}$ if

$$\frac{|S(\mathbf{a}, \mathfrak{p})|}{|S|} < \frac{B}{\mathcal{N}_K(\mathfrak{p})^l},$$

for every residue class $\mathbf{a}$ mod $\mathfrak{p}$.

Lemma 5.3 (Compare to [Wal12, Lemma 3.4])**.** *Let* $d, h \geq 1$ *be arbitrary integers and let* $\varepsilon > 0$ *be some positive real number. Set* $Q = N^{\frac{\varepsilon}{2d}}$ *and let* $P \subseteq \mathcal{P}(Q)$ *satisfying* $w(P) \geq \kappa w(\mathcal{P}(Q))$ *for some* $\kappa > 0$*. Suppose that* $S \subseteq [N]^d_{\mathcal{O}_K}$ *is a subset of size at least* $cN^{d-h-1+\varepsilon}$ *occupying at most* $\alpha\mathcal{N}(\mathfrak{p})^{d-h}$ *residue classes modulo* $\mathfrak{p}$ *for all prime* $\mathfrak{p} \in P$ *and some* $\alpha > 0$*. Then if* N *is sufficiently large, there exist constants* $B = B(d, h, \varepsilon, \kappa, \alpha, c)$*,* $\kappa_1 = \kappa_1(d, h, \varepsilon, \kappa, \alpha, c)$*,* $c_1 = c_1(d, h, \varepsilon, \kappa, \alpha, c)$ *such that there exists a subset of primes* $P' \subseteq P$ *with* $w(P') \geq \kappa_1 w(P)$ *such that for each* $\mathfrak{p} \in P'$ *there is some* $\mathcal{G}_{\mathfrak{p}}(S) \subseteq S$ *with* $|\mathcal{G}_{\mathfrak{p}}(S)| \geq c_1|S|$*, which is* $(B, d-h)$*-generic modulo* $\mathfrak{p}$*.*

Proof. From now on let us fix an integer $h \geq 1$. If $d \leq h$, then we may take $B = 2$, $\mathcal{G}_{\mathfrak{p}}(S) = S$ and $P' = P$. Thus

$$B = B(d, h, \varepsilon, \kappa, \alpha, c) = 2, \tag{5.1}$$

$$\kappa_1 = \kappa_1(d, h, \varepsilon, \kappa, \alpha, c) = 1, \tag{5.2}$$

$$c_1 = c_1(d, h, \varepsilon, \kappa, \alpha, c) = 1. \tag{5.3}$$

We now proceed by induction on d. Let us suppose that $d \geq h+1$ is an integer and let us assume that Lemma 5.3 holds for every smaller dimension. Let S and P be as in the statement. For each $1 \leq i \leq d$, define $\pi_i : K^d \to K$ as the projection in the i-th coordinate.

Claim 5.4. *There exists a constant* $C_3 = C_3(d, h, \varepsilon, \kappa, \alpha, c)$ *such that if* $N \geq C_3$ *then there exist* $1 \leq i \leq d$ *and a subset* $S' \subseteq S$ *with* $|S'| \geq \frac{|S|}{2^d}$ *such that for any* $A \subseteq S'$ *with* $|A| \geq \frac{|S'|}{2}$ *we have* $|\pi_i(A)| \geq Q$*.*

Proof of Claim 5.4. Let us suppose that the claim is false for $S' = S$ and $i = 1$. Hence there exists $A_1 \subseteq S$ with $|A_1| \geq \frac{|S|}{2}$ and $|\pi_1(A_1)| < Q$. Then, if the claim fails for $S' = A_1$ and $i = 2$, there exists $A_2 \subseteq A_1$ with $|A_2| \geq \frac{|A_1|}{2} \geq \frac{|S|}{2^2}$ and $|\pi_1(A_2)|, |\pi_2(A_2)| < Q$. Iterating this process d times, either we get the claim or end up with a subset $A_d \subseteq S$ with $|A_d| \geq \frac{|S|}{2^d}$. Thus,

$$cN^{d-h-1+\varepsilon} \leq |S| \leq 2^d|A_d| \leq 2^d \prod_{i=1}^{d} |\pi_i(A_d)| < 2^d Q^d = 2^d N^{\frac{\varepsilon}{2}}. \tag{5.4}$$

Now, (5.4) is absurd if $N \geq \left(\frac{2^d}{c}\right)^{\frac{1}{d-h-1+\frac{\varepsilon}{2}}}$. □

By claim (5.4), at the cost of passing to a subset of of S of density at least equal to $\frac{1}{2^d}$ and permuting the variables, from now on we assume that S satisfies

$$|\pi_1(A)| \geq Q \text{ for all } A \subseteq S \text{ with } |A| \geq \frac{|S|}{2}. \tag{5.5}$$

First we will find a dense subset of S which is in condition to apply the inductive hypothesis over the fibers, then, the generic set of the fibers will be glued together to obtain the desired generic set. In order to do so we will first eliminate some problematic fibers. Let $\mathfrak{p}$ a prime in P. Recall that for any $a \in \mathcal{O}_K/\mathfrak{p}$, $S(a, \mathfrak{p})$ denotes the elements $\mathbf{x}$ of S for which $\pi_1(\mathbf{x}) \equiv a (\text{mod } \mathfrak{p})$. Let B_1 be a constant sufficiently large to be chosen later. Since $|[S]_\mathfrak{p}| \leq \alpha \mathcal{N}_K(\mathfrak{p})^{d-h}$, it is clear that there are at most $\frac{\alpha}{B_1}\mathcal{N}_K(\mathfrak{p})$ residue classes $a \in [\pi_1(S)]_\mathfrak{p} \subseteq \mathcal{O}_K/\mathfrak{p}$ for which $|[S(a, \mathfrak{p})]_\mathfrak{p}| \geq B_1\mathcal{N}_K(\mathfrak{p})^{d-h-1}$. Let us denote

$$\mathcal{E}_1(\mathfrak{p}) := \left\{a \in [\pi_1(S)]_\mathfrak{p} : |[S(a, \mathfrak{p})]_\mathfrak{p}| \geq B_1\mathcal{N}_K(\mathfrak{p})^{d-h-1}\right\}. \tag{5.6}$$

We will also write

$$\mathcal{E}_2(\mathfrak{p}) := \left\{a \in [\pi_1(S)]_\mathfrak{p} : |S(a, \mathfrak{p})| \geq \frac{B_1}{\alpha \mathcal{N}_K(\mathfrak{p})}|S|\right\}. \tag{5.7}$$

From the identity $\sum_{a\in\mathcal{O}_K/\mathfrak{p}} |S(a, \mathfrak{p})| = |S|$ it follows that $|\mathcal{E}_2(\mathfrak{p})| \leq \frac{\alpha}{B_1}\mathcal{N}_K(\mathfrak{p})$, hence if $\mathcal{E}(\mathfrak{p}) = \mathcal{E}_1(\mathfrak{p}) \cup \mathcal{E}_2(\mathfrak{p})$ is the set of these exceptional residue classes, we have $|\mathcal{E}(\mathfrak{p})| \leq \frac{2\alpha}{B_1}\mathcal{N}_K(\mathfrak{p})$. We will use the larger sieve (Lemma 3.1) to prove that few $x \in [N]_{\mathcal{O}_K}$ lie in $\mathcal{E}(\mathfrak{p})$ for many $\mathfrak{p} \in P$. Indeed, let us consider the set

$$X := \left\{x \in [N]_{\mathcal{O}_K} : \sum_{\mathfrak{p}\in P} 1_{x (\text{mod } \mathfrak{p}) \in \mathcal{E}(\mathfrak{p})} \frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} \geq \frac{1}{2}w(P)\right\}.$$

Let $P_1 \subseteq P$ the set of primes such that at least $\frac{1}{4}|X|$ elements of X lie in the exceptional set of residue classes $\mathcal{E}(\mathfrak{p})$, namely, $P_1 := \left\{p \in P : \left|\bigcup_{a\in\mathcal{E}(\mathfrak{p})} X(a, \mathfrak{p})\right| \geq \frac{1}{4}|X|\right\}$.

Claim 5.5. $w(P_1) \geq \frac{1}{4}w(P)$.

Proof of Claim 5.5. Observe that

$$\sum_{a\in\mathcal{E}(\mathfrak{p})} |X(a, \mathfrak{p})| = \left|\bigcup_{a\in\mathcal{E}(\mathfrak{p})} X(a, \mathfrak{p})\right| = \sum_{x\in X} 1_{x(\text{mod } \mathfrak{p})\in\mathcal{E}(\mathfrak{p})}.$$

On the one hand we have

$$\sum_{\mathfrak{p}\in P}\sum_{x\in X} 1_{x(\text{mod } \mathfrak{p})\in\mathcal{E}(\mathfrak{p})} \frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})} = \sum_{x\in X}\left(\sum_{\mathfrak{p}\in P} 1_{x(\text{mod } \mathfrak{p})\in\mathcal{E}(\mathfrak{p})} \frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})}\right) \geq \sum_{x\in X} \frac{1}{2}w(P) = \frac{1}{2}w(P)|X|. \tag{5.8}$$

On the other hand,

$$\begin{aligned}\sum_{\mathfrak{p}\in P}\sum_{x\in X}1_{x(\text{mod }\mathfrak{p})\in\mathcal{E}(\mathfrak{p})}\frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})} &= \sum_{\mathfrak{p}\notin P_1}\left(\sum_{a\in\mathcal{E}(\mathfrak{p})}|X(a,\mathfrak{p})|\right)\frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})} + \sum_{\mathfrak{p}\in P_1}\left(\sum_{a\in\mathcal{E}(\mathfrak{p})}|X(a,\mathfrak{p})|\right)\frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})}\\ &< \sum_{\mathfrak{p}\notin P_1}\frac{1}{4}|X|\frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})} + \sum_{\mathfrak{p}\in P_1}|X|\frac{\log(\mathcal{N}(\mathfrak{p}))}{\mathcal{N}(\mathfrak{p})} = \frac{1}{4}|X|w(P\backslash P_1) + |X|w(P_1) \\ &= \frac{1}{4}|X|w(P) - \frac{1}{4}|X|w(P_1) + |X|w(P_1).\end{aligned} \tag{5.9}$$

Comparing (5.8) and (5.9), we conclude that $w(P_1) \geq \frac{1}{4}w(P)$. □

In Lemma 3.1, setting the constants $\gamma := \frac{\varepsilon}{2d}$, $\kappa := \frac{\kappa}{4}$, $\mu := \frac{1}{4}$, $\alpha := \frac{2\alpha}{B_1}$, it follows that if we chose

$$B_1 := \frac{2\alpha}{C_1\left(\frac{\kappa}{4},\frac{1}{4},\frac{\varepsilon}{2d},K\right)} = \frac{2^8\alpha c_{5,K}d}{\kappa\varepsilon} \tag{5.10}$$

then $|X| < Q$. From (5.5) we deduce that $|S\backslash\pi_1^{-1}(X)| \geq \frac{1}{2}|S|$.

Claim 5.6. *Let $0<\nu<1$. There exists a subset $S'\subseteq S$ with $|S'|\geq\frac{1}{4}|S|$ which does not intersect $\pi_1^{-1}(X)$ and such that $S'_x := \pi_1^{-1}(x)\cap S'$ satisfies $|S'_x| \geq c'(K,\nu)N^{d-h-2+\nu\varepsilon}$ for all $x\in\pi_1(S')$, where $c'(K,\nu)$ is a positive constant dependent on K,ν.*

Proof of Claim 5.6. The proof will be by contradiction. Let us assume that for all $S'\subseteq S\backslash\pi_1^{-1}(X)$ with $|S'|\geq\frac{1}{4}|S|$ satisfies $|S'_x| < c'(K,\nu)N^{d-h-2+\nu\varepsilon}$ for some $x\in\pi_1(S')$. Let

$$\overline{S} := \left\{s\in S\backslash\pi_1^{-1}(X) : |(S\backslash\pi_1^{-1}(X))_{\pi_1(s)}| < c'(K,\nu)N^{d-h-2+\nu\varepsilon}\right\}.$$

Note that $\overline{S}$ and $S\backslash\pi_1^{-1}(X)$ have the same sections (whenever the section of $\overline{S}$ is non-empty). Thus, from the assumption we deduce that $|\overline{S}| \geq \frac{1}{2}|S\backslash\pi_1^{-1}(X)|$ and moreover $|\overline{S}_x| \leq c'(K,\nu)N^{d-h-2+\nu\varepsilon}$ for all $x\in\pi_1(\overline{S})$. Hence,

$$\frac{1}{4}cN^{d-h-1+\varepsilon} \leq \frac{1}{4}|S| \leq |\overline{S}| = \left|\bigcup_{x\in\pi_1(\overline{S})}\overline{S}_x\right| < c'(K,\nu)N^{d-h-2+\nu\varepsilon}|\pi_1(\overline{S})| \leq c'(K,\nu)N^{d-h-2+\nu\varepsilon}|[N]_{\mathcal{O}_K}| \tag{5.11}$$

Using Proposition 2.2 in (5.11), and using the bound $\log(N) \leq \frac{|M_{K,\infty}|}{(1-\nu)\varepsilon}N^{\frac{(1-\nu)\varepsilon}{|M_{K,\infty}|}}$, we arrive at the inequality

$$\frac{1}{4}cN^{d-h-1+\varepsilon} \leq c'(K,\nu)N^{d-h-1+\nu\varepsilon}c''(K)(\log(N))^{|M_{K,\infty}|} \leq c'(K,\nu)c''(K)\left(\frac{d_K}{(1-\nu)\varepsilon}\right)^{d_K}N^{d-h-1+\varepsilon}.$$

But this inequality does not hold for $c'(K,\nu) = \dfrac{c}{8c''(K)}\left(\dfrac{(1-\nu)\varepsilon}{d_K}\right)^{d_K}$. □

Take S' as in Claim 5.6. Every $x\in\pi_1(S')$ lies outside of X, thus it has associated a subset of primes $P_x\subseteq P\subset\mathcal{P}(Q)$ such that $x(\text{mod }\mathfrak{p})\notin\mathcal{E}(\mathfrak{p})$. Furthermore, $w(P_x)\geq\frac{1}{2}w(P)\geq\frac{1}{2}\kappa w(\mathcal{P}(Q))$. Since $\mathcal{E}_1(\mathfrak{p})\subseteq\mathcal{E}(\mathfrak{p})$, we have that for each $x\in\pi_1(S')$, $|[S'_x]_{\mathfrak{p}}|\leq B_1\mathcal{N}_K(\mathfrak{p})^{d-h-1}$ and $|S'_x|\geq c'(K,\nu)N^{d-h-2+\nu\varepsilon}$. Since we are doing induction on d, at the step $d-1$, the parameter Q should be changed to $N^{\frac{\nu\varepsilon}{2(d-1)}}$. Hence if we take $\nu=\frac{d-1}{d}<1$, the parameter Q remains unchanged throughout the proof.

We are in condition to apply the inductive hypothesis to each S'_x, $x \in \pi_1(S')$ to affirm that there exists a subset $P'_x \subseteq P_x$ with $w(P'_x) \geq \kappa_1 w(P_x)$, and constants c_1, B independent of x, such that for each $p \in P'_x$ there is a $(B, d-h-1)$-generic modulo $\mathfrak{p}$ set $\mathcal{G}_\mathfrak{p}(S'_x) \subseteq S'_x$, containing at least $c_1|S'_x|$ elements. Here, the dependence of the constants are:

$$B = B\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, c'(K,\nu)\right), \tag{5.12}$$

$$\kappa_1 = \kappa_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, c'(K,\nu)\right), \tag{5.13}$$

$$c_1 = c_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, c'(K,\nu)\right), \tag{5.14}$$

$$\nu = \frac{d-1}{d}, \tag{5.15}$$

where B_1 was determined in equation (5.10).

Each fiber S'_x has its own set of primes P'_x, with density $w(P'_x) \geq \kappa_1 w(P_x)$. Since κ_1 is independent of x, we will next find a set of primes $P' \subset P$, $w(P') \gtrsim w(P)$ and then for each $\mathfrak{p} \in P'$ we will construct a generic set, $\mathcal{G}_\mathfrak{p}(S)$, by gluing the generic sets of the fibers S'_x.

Claim 5.7. *There exist a positive constant $\beta > 0$ and a set of primes $P' \subseteq P$ with $w(P') \geq \frac{\kappa\kappa_1}{4} w(P)$ such that for each $\mathfrak{p} \in P'$ there are at least $\beta|S'|$ elements $s \in S'$ for which $\mathfrak{p} \in P'_{\pi_1(s)}$.*

Proof of Claim 5.7. Let $\beta > 0$ and consider the set $P' := \left\{\mathfrak{p} \in P : \sum_{s\in S'} 1_{\mathfrak{p}\in P'_{\pi_1(s)}} \geq \beta|S'|\right\}$. Then,

$$\begin{aligned}
\sum_{s\in S'}\sum_{\mathfrak{p}\in P} 1_{\mathfrak{p}\in P'_{\pi_1(s)}} \frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} &= \sum_{\mathfrak{p}\in P}\left(\sum_{s\in S'} 1_{\mathfrak{p}\in P'_{\pi_1(s)}}\right)\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} \\
&\leq \sum_{\mathfrak{p}\notin P'} \beta|S'|\frac{\log(\mathcal{N}_K(p))}{\mathcal{N}_K(\mathfrak{p})} + \sum_{\mathfrak{p}\in P'}|S'|\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} \\
&= \beta|S'|w(P\backslash P') + |S'|w(P') \\
&= \beta|S'|w(P) - \beta|S'|w(P') + |S'|w(P').
\end{aligned}$$

On the other hand, recalling that $w(P'_x) \geq \kappa_1 w(P_x) \geq \dfrac{\kappa_1\kappa}{2} w(P)$, we have

$$\sum_{s\in S'}\sum_{\mathfrak{p}\in P} 1_{\mathfrak{p}\in P'_{\pi_1(s)}} \frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} = \sum_{s\in S'} w(P'_{\pi_1(s)}) \geq \sum_{s\in S'} \frac{\kappa_1\kappa}{2} w(P) = \frac{\kappa_1\kappa}{2} w(P)|S'|.$$

We conclude that $\left(\dfrac{\kappa_1\kappa}{2} - \beta\right) w(P) \leq (1-\beta)\, w(P') < w(P')$. It is enough to set

$$\beta = \frac{\kappa_1\kappa}{4}. \tag{5.16}$$

$\square$

To finish the proof, for each $\mathfrak{p} \in P'$ we will construct a generic set $\mathcal{G}_\mathfrak{p}(S)$ that fulfills the conditions of Lemma (5.3). For that, take

$$\mathcal{G}_\mathfrak{p}(S) := \bigcupdot_{x:\mathfrak{p}\in P'_x} \mathcal{G}_\mathfrak{p}(S'_x). \tag{5.17}$$

Observe that

$$|\mathcal{G}_{\mathfrak{p}}(S)| = \left|\bigcupdot_{x:\mathfrak{p}\in P'_x} \mathcal{G}_{\mathfrak{p}}(S'_x)\right| = \sum_{x:\mathfrak{p}\in P'_x} \left|\mathcal{G}_{\mathfrak{p}}(S'_x)\right| \geq c_1 \sum_{x:\mathfrak{p}\in P'_x} \left|S'_x\right| = c_1 \left|\left\{s \in S' : \mathfrak{p} \in P'_{\pi_1(s)}\right\}\right| > c_1\beta|S'| \geq \frac{c_1\beta}{2^2}|S| \tag{5.18}$$

and $\mathcal{G}_{\mathfrak{p}}(S) \cap \pi_1^{-1}(x) = \mathcal{G}_{\mathfrak{p}}(S'_x)$ is a $(B, d-h-1)$-generic set for all $x \in \pi_1(\mathcal{G}_{\mathfrak{p}}(S))$.

Let us now see that $\mathcal{G}_{\mathfrak{p}}(S)$ is indeed a generic set. First, note that by construction, the residue classes modulo $\mathfrak{p}$ of $\pi_1(\mathcal{G}_{\mathfrak{p}}(S))$ do not lie in $\mathcal{E}(\mathfrak{p})$. Then, recalling the definition of $\mathcal{E}_2(\mathfrak{p}) \subset \mathcal{E}(\mathfrak{p})$, from (5.18) it follows that

$$|\mathcal{G}_{\mathfrak{p}}(S)(a, \mathfrak{p})| \leq \frac{B_1}{\alpha\mathcal{N}_K(\mathfrak{p})}|S| \leq \frac{2^2 B_1}{c_1\beta\alpha}\frac{|\mathcal{G}_{\mathfrak{p}}(S)|}{\mathcal{N}_K(\mathfrak{p})}. \tag{5.19}$$

Let $\mathbf{a}$ be a residue class modulo $\mathfrak{p}$. Observe that

$$\mathcal{G}_{\mathfrak{p}}(S)(\mathbf{a}, \mathfrak{p}) = \left(\bigcupdot_{x:\mathfrak{p}\in P'_x} \mathcal{G}_{\mathfrak{p}}(S'_x)\right)(\mathbf{a}, \mathfrak{p}) = \bigcupdot_{x:\mathfrak{p}\in P'_x} \mathcal{G}_{\mathfrak{p}}(S'_x)(\mathbf{a}, \mathfrak{p}) = \bigcupdot_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} \mathcal{G}_{\mathfrak{p}}(S'_x)(\mathbf{a}, \mathfrak{p}) \tag{5.20}$$

Hence the fact that $\mathcal{G}_{\mathfrak{p}}(S'_x)$ is $(B, d-h-1)$-generic implies

$$\begin{aligned} |\mathcal{G}_{\mathfrak{p}}(S)(\mathbf{a}, \mathfrak{p})| &= \left|\bigcupdot_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} \mathcal{G}_{\mathfrak{p}}(S'_x)(\mathbf{a}, \mathfrak{p})\right| \leq \sum_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} \left|\mathcal{G}_{\mathfrak{p}}(S'_x)(\mathbf{a}, \mathfrak{p})\right| \leq \sum_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} |\mathcal{G}_{\mathfrak{p}}(S'_x)|\frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}} \\ &= \frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}} \sum_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} |\mathcal{G}_{\mathfrak{p}}(S'_x)| = \frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}} \left|\bigcupdot_{\substack{x:\mathfrak{p}\in P'_x \\ x\equiv\pi_1(\mathbf{a}) (\mathrm{mod}\ \mathfrak{p})}} \mathcal{G}_{\mathfrak{p}}(S'_x)\right|. \end{aligned} \tag{5.21}$$

Since $\bigcupdot_{x:x\equiv\pi_1(\mathbf{a})(\mathrm{mod}\ \mathfrak{p})} \mathcal{G}_{\mathfrak{p}}(S'_x) \subseteq \mathcal{G}_{\mathfrak{p}}(S)(\pi_1(\mathbf{a}), \mathfrak{p})$, combining this with (5.21) and (5.19) we get

$$|\mathcal{G}_{\mathfrak{p}}(S)(\mathbf{a}, \mathfrak{p})| \leq \frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}} |\mathcal{G}_{\mathfrak{p}}(S)(\pi_1(\mathbf{a}), \mathfrak{p})| \leq \frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}} 2^2\frac{B_1}{c_1\beta\alpha}\frac{1}{\mathcal{N}_K(\mathfrak{p})}|\mathcal{G}_{\mathfrak{p}}(S)| = 2^2\frac{BB_1}{c_1\beta\alpha}\frac{1}{\mathcal{N}_K(\mathfrak{p})^{d-h}}|\mathcal{G}_{\mathfrak{p}}(S)|. \tag{5.22}$$

□

The dependence of the constants is as follows: from Claim 5.7,

$$\kappa_1(d, h, \varepsilon, \kappa, \alpha, c) = \frac{\kappa}{4}\kappa_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right) \tag{5.23}$$

where 2^d is due to the assumption made at (5.5); from (5.16) and (5.18)

$$c_1(d-h, d, \varepsilon, \kappa, \alpha, c) = \frac{\kappa}{2^{d+4}}\kappa_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right)c_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right) \tag{5.24}$$

where 2^d is due to the assumption made at (5.5); and from (5.10), (5.22) and (3.7)

$$B(d, h, \varepsilon, \kappa, \alpha, c) = \frac{2^{12}dc_{5,K}}{\kappa^2\varepsilon}\frac{B\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right)}{\kappa_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right)c_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, \frac{2\alpha}{C_1}, \frac{c'(K,\nu)}{2^d}\right)} \tag{5.25}$$

where, by (5.10) and (3.7)

$$C_1 = C_1\left(\frac{\kappa}{4}, \frac{1}{4}, \frac{\varepsilon}{2d}, K\right) = \frac{\kappa\varepsilon}{2^7 c_{5,K} d}. \tag{5.26}$$

5.2. **Characteristic sets.** Having proved the existence of generic subsets, now we prove that there exist "small characteristic subsets". In order to make precise the notion of characteristic subset, first we define what we mean by a "small" polynomial.

Definition 5.8 (r-polynomial)**.** Let K be a global field of degree d_K. Given a parameter N and some integer $d > 0$ by an r-polynomial we mean a non-zero polynomial $f \in \mathcal{O}_K[X_1, \dots, X_d]$ such that for all $\mathbf{x} \in [N]^d_{\mathcal{O}_K}$ it holds $H_K(f(\mathbf{x})) < N^{3rd_K}$.

We remark that the reason of the exponent $3rd_K$ in the definition is because for any $f \in \mathcal{O}_K[X_1, \dots, X_d]$ with degree and coefficients of height bounded by r satisfies that for N large enough, $H_K(f(\mathbf{x})) < N^{3rd_K}$ for all $[N]^d_{\mathcal{O}_K}$.

Definition 5.9 (Characteristic subset)**.** Let $0 < \delta \leq 1$ be a real number and $r > 0$ a positive integer. We say that $A \subseteq S$ is (r, δ)-characteristic for S if there exists $A \subseteq L \subseteq S$ of size $|L| \geq \delta|S|$ such that whenever an r-polynomial vanishes at A, it also vanishes at L.

The main result of this section is that for a given $S \subseteq [N]^d_{\mathcal{O}_K}$, there exists a positive constant $\delta > 0$ such that there are always (r, δ)-characteristic subsets, provided that N is large enough.

Proposition 5.10 (Compare to Proposition 2.2 in [Wal12])**.** *Let $d, h \geq 0$ be arbitrary integers and $\varepsilon > 0$ a positive real number. Let $Q = N^{\frac{\varepsilon}{2d}}$ and $P \subseteq \mathcal{P}(Q)$ satisfying $w(P) \geq \kappa w(\mathcal{P}(Q))$ for some $\kappa > 0$. Let $r > 0$ be an integer. Suppose that $S \subseteq [N]^d_{\mathcal{O}_K}$ is a subset of size $|S| \geq cN^{d-h-1+\varepsilon}$ occupying at most $\alpha\mathcal{N}_K(\mathfrak{p})^{d-h}$ residue classes modulo $\mathfrak{p}$ for all $\mathfrak{p} \in P$ and some $\alpha > 0$. Then, if N is sufficiently large, there exists a set $A \subseteq S$ of size $|A| \leq c_2 r^{d-h}$ which is (r, δ)-characteristic for S, for some $\delta = \delta(d, h, \varepsilon, \kappa, \alpha, c, K)$ and $c_2 = c_2(d, h, \varepsilon, \kappa, \alpha, c, K)$.*

Proof of Proposition 5.10. We fix h and proceed by induction on d. When $d < h$, for N sufficiently large there exists $\mathfrak{p} \in P$ such that $\alpha\mathcal{N}_K(\mathfrak{p})^{d-h} < 1$, hence S is empty. When $d = h$, the result follows from Lemma 3.2. Indeed, in this case any subset $S \subseteq [N]^d_{\mathcal{O}_K}$ satisfying the hypothesis of Proposition 5.10 occupies at most α residue classes modulo $\mathfrak{p}$ for all $\mathfrak{p} \in P$ and some $\alpha > 0$. Thus, Lemma 3.2 implies that $|S| \leq C_2 = C_2(\alpha, \kappa, \frac{\varepsilon}{2d}, K)$. Thus, taking $A = B = S$ we see that A is $(r, 1)$-characteristic for S. In particular, we have that

$$c_2(d, h, \varepsilon, \kappa, \alpha, c, K) = C_2, \tag{5.27}$$

$$\delta(d, h, \varepsilon, \kappa, \alpha, c, K) = 1. \tag{5.28}$$

Hence, let us assume that $d \geq h + 1$ and that the result holds for smaller dimensions. First we find generic subsets inside the sections of S for many primes $\mathfrak{p}$, as we did in Lemma 5.3. Proceeding as in Claim 5.4, we pass to a subset $S_1 \subseteq S$ of size $|S_1| \geq \frac{1}{8^d}|S|$ such that

$$|\pi_1(A)| \geq Q \text{ for all } A \subseteq S_1 \text{ with } |A| \geq \frac{|S_1|}{8}. \tag{5.29}$$

We may further refine the set S_1 to have a control in the size of the sections:

Claim 5.11. *There exists a subset $S_2 \subseteq S_1$ of size $|S_2| \geq \frac{3}{4}|S_1|$, such that*

$$|(S_2)_x| \leq 2\frac{|S_2|}{Q} \text{ for all } x \in [N]_{\mathcal{O}_K}, \tag{5.30}$$

where $(S_2)_x := \pi_1^{-1}(x) \cap S_2$.

Proof of Claim 5.11. Let $W := \left\{ s \in S_1 : \left|(S_1)_{\pi_1(s)}\right| > \dfrac{|S_1|}{Q} \right\}$. Note that $|W| \le \frac{1}{4}|S_1|$, otherwise $|W| > \frac{1}{4}|S_1| > \frac{1}{8}|S_1|$ and (5.29) imply $|\pi_1(W)| \ge Q$. This entails that

$$|S_1| = \left| \bigcup_{x \in \pi_1(S_1)} (S_1)_x \right| \ge \left| \bigcup_{x \in \pi_1(W)} (S_1)_x \right| > |\pi_1(W)| \frac{|S_1|}{Q} \ge |S_1|,$$

which is a contradiction. Define $S_2 := S_1 \backslash W$. Then $|S_2| \ge \frac{3}{4}|S_1|$ and for any $x \in \pi_1(S_2)$, $(S_2)_x = (S_1)_x$, thus for such x it holds

$$|(S_2)_x| \le \frac{|S_1|}{Q} \le \frac{4}{3}\frac{|S_2|}{Q} \le 2\frac{|S_2|}{Q}.$$

□

Now, we proceed as in the proof Claim 5.6 to obtain a subset $S_3 \subseteq S_2$ of size $|S_3| \ge \frac{1}{2}|S_2|$ such that

$$|(S_3)_x| \ge c'(K,\nu) N^{d-h-2+\nu\varepsilon} \text{ for all } x \in \pi_1(S_3) \tag{5.31}$$

where, as before, $\nu = \frac{d-1}{d}$ and

$$c'(K,\nu) = \frac{3c}{2^{3d+4} c''(K)} \left(\frac{(1-\nu)\varepsilon}{d_K} \right)^{d_K}. \tag{5.32}$$

Let B_1 be a large constant. For any prime $\mathfrak{p}$ we denote $\mathcal{E}_1(\mathfrak{p})$ for the set of residue classes $a \in \mathcal{O}_K/\mathfrak{p}$ such that $|[S_3(a,\mathfrak{p})]_{\mathfrak{p}}| \ge B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}$. Since $|\mathcal{E}_1(\mathfrak{p})| \le \frac{\alpha}{B_1}\mathcal{N}_K(\mathfrak{p})$, applying Lemma 3.1 as in the proof of Lemma 5.3 we conclude that if B_1 is sufficiently large enough, namely, if

$$B_1 := \frac{\alpha}{C_1\left(\frac{\kappa}{4}, \frac{1}{4}, \frac{\varepsilon}{2d}, K\right)} = \frac{2^7 \alpha c_{5,K} d}{\kappa \varepsilon}, \tag{5.33}$$

then the set

$$X := \left\{ x \in [N]_{\mathcal{O}_K} : \sum_{\mathfrak{p} \in P} 1_{x (\mathrm{mod}\ \mathfrak{p}) \in \mathcal{E}_1(\mathfrak{p})} \frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} \ge \frac{1}{2} w(P) \right\},$$

has size $|X| < Q$. Thus, $|\pi_1^{-1}(X) \cap S_3| < \frac{1}{8}|S_1|$, since otherwise the inequality $|\pi_1^{-1}(X) \cap S_3| \ge \frac{1}{8}|S_1|$ and (5.29) imply $|\pi_1(\pi_1^{-1}(X) \cap S_3)| \ge Q$. But this can not hold, since $|\pi_1(\pi_1^{-1}(X) \cap S_3)| \le |X| < Q$. Since $|S_3| \ge \frac{1}{2}|S_2| \ge \frac{3}{8}|S_1|$, we conclude that $|S_3 \backslash \pi_1^{-1}(X)| \ge \frac{2}{3}|S_3|$.

If we set $S_4 := S_3 \backslash \pi_1^{-1}(X)$, we see that S_4 satisfies (5.30) and (5.31), and for all $x \in \pi_1(S_4)$ we have a subset of primes

$$P_x := \{ p \in P : x (\mathrm{mod}\ \mathfrak{p}) \notin \mathcal{E}_1(\mathfrak{p}) \}$$

for which $w(P_x) \ge \frac{1}{2} w(P) \ge \frac{1}{2}\kappa w(\mathcal{P}(Q))$.

Since for each non empty section $(S_4)_x$ we have that $|(S_4)_x| \ge c'(K,\nu) N^{d-h-2+\nu\varepsilon}$ and $|[(S_4)_x]_{\mathfrak{p}}| \le B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}$, for all $\mathfrak{p} \in P_x$, we are in the conditions of the inductive hypothesis for the parameter $Q' = N^{\frac{\nu\varepsilon}{2(d-1)}}$ which is equal to Q by our choice of ν. We may deduce that for each non-empty section $(S_4)_x$ there exist δ_0 and c_2, independent on x, such that $(S_4)_x$ admits an (r, δ_0)-characteristic subset A_x of size $|A_x| \le c_2 r^{d-h-1}$. The dependence of the constants is as follows,

$$c_2 = c_2\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, c'(K,\nu), K\right), \tag{5.34}$$

$$\delta_0 = \delta\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, c'(K,\nu), K\right). \tag{5.35}$$

Since each A_x is (r,δ_0)-characteristic, there exists a subset $A_x \subseteq L_x \subseteq (S_4)_x$ of size $|L_x| \geq \delta_0|(S_4)_x|$ the condition of Definition 5.9. Thus, $S' = \bigcup_{x\in\pi_1(S_4)} L_x$ satisfies

$$|S'| \geq \delta_0|S_4| \geq \delta_0\frac{2}{3}|S_3| \geq \delta_0\frac{2}{3}\frac{3}{8}|S_1| \geq \delta_0\frac{2}{3}\frac{3}{8}\frac{1}{8^d}|S| = \frac{\delta_0}{2^{3d+2}}|S|, \tag{5.36}$$

the bound (5.30), and

$$|S'_x| = |L_x| \geq \delta_0|(S_4)_x| \geq \delta_0 c'(K,\nu)N^{d-h-2+\nu\varepsilon} \text{ for all } x \in \pi_1(S'). \tag{5.37}$$

Moreover, for all $x \in \pi_1(S')$, A_x is a $(r,1)$-characteristic subset of S'_x.

Now, because of Lemma 5.3 we can construct a subset of primes $P'_x \subseteq P_x$ with $w(P'_x) \geq \kappa_1 w(P_x)$ such that for all $\mathfrak{p} \in P'_x$ there exists a $(B, d-h-1)$-generic subset $\mathcal{G}_\mathfrak{p}(S'_x) \subseteq S'_x$, of size $|\mathcal{G}_\mathfrak{p}(S'_x)| \geq c_1|S'_x|$, where the constants are independent of $\mathfrak{p}$ and x. Specifically,

$$B = B\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, \delta_0 c'(K,\nu)\right), \tag{5.38}$$

$$\kappa_1 = \kappa_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, \delta_0 c'(K,\nu)\right), \tag{5.39}$$

$$c_1 = c_1\left(d-1, h, \nu\varepsilon, \frac{\kappa}{2}, B_1, \delta_0 c'(K,\nu)\right). \tag{5.40}$$

By Claim 5.7, we can find a subset of primes $P' \subseteq P$ such that

$$w(P') \geq \frac{\kappa\kappa_1}{4}w(P), \tag{5.41}$$

and for all $\mathfrak{p} \in P'$ there are at least $\beta|S'|$ elements $s \in S'$ for which $\mathfrak{p} \in P'_{\pi_1(s)}$. Thus, for $\mathfrak{p} \in P'$

$$\mathcal{G}_\mathfrak{p} := \biguplus_{x:\mathfrak{p}\in P'_x} \mathcal{G}_\mathfrak{p}(S'_x)$$

is a subset of S of size

$$|\mathcal{G}_\mathfrak{p}| \geq \beta c_1|S'| \geq \beta c_1\frac{\delta_0}{2^{3d+2}}|S|, \tag{5.42}$$

such that each non-empty section $(\mathcal{G}_\mathfrak{p})_x$ is a $(B, d-h-1)$-generic set.

In order to prove Proposition 5.10, we are going to glue some of the characteristic subsets that we found on each section of S'. In order to do this, while obtaining a small characteristic subset of S, we need to locate sections of S containing the residue class of many elements of S for many primes $\mathfrak{p}$. This is the content of the next lemma.

Lemma 5.12. *There exists a subset $\mathcal{B} \subseteq S'$ with $|\mathcal{B}| \geq c_3|S'|$, such that for each non-empty section $\mathcal{B}_x$ of $\mathcal{B}$ there exists a subset of primes $P''_x \subseteq P'$ with $w(P''_x) \geq \kappa_3 w(P')$, such that for every $\mathfrak{p} \in P''_x$*

$$\left|\left\{s \in S' : [s]_\mathfrak{p} \in [\mathcal{B}_x]_\mathfrak{p}\right\}\right| \geq c_4\frac{|S'|}{\mathcal{N}_K(\mathfrak{p})}.$$

Proof of Lemma 5.12. We begin by fixing a prime $\mathfrak{p} \in P'$ and consider some residue class $a \in [\pi_1(\mathcal{G}_\mathfrak{p})]_\mathfrak{p}$. Since $\mathfrak{p}$ is fixed we are going to denote $\mathcal{G}_\mathfrak{p}(a)$ for those elements in $\mathcal{G}_\mathfrak{p}$ with first coordinate congruent to a modulo $\mathfrak{p}$. Moreover, given a class $\mathbf{b} \in (\mathcal{O}_K/\mathfrak{p})^d$ we denote $\mathcal{G}_\mathfrak{p}(\mathbf{b})$ for those elements of $\mathcal{G}_\mathfrak{p}$ congruent to $\mathbf{b}$ modulo $\mathfrak{p}$. By the pigeonhole principle and the fact that by construction of P' it holds $|[\mathcal{G}_{\mathfrak{p}}(a)]_{\mathfrak{p}}| \leq B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}$ it follows that we can find $\mathbf{b}_1 \in [\mathcal{G}_{\mathfrak{p}}(a)]_{\mathfrak{p}} \subseteq (\mathcal{O}_K/\mathfrak{p})^d$ with

$$|\mathcal{G}_{\mathfrak{p}}(\mathbf{b}_1)| \geq \frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}}.$$

Consider now $\mathcal{B}_1 \subseteq \mathcal{G}_{\mathfrak{p}}(a)$ defined by

$$\mathcal{B}_1 := \bigcup_{s:[s]_{\mathfrak{p}}=\mathbf{b}_1} (\mathcal{G}_{\mathfrak{p}})_{\pi_1(s)}, \tag{5.43}$$

that is, $\mathcal{B}_1$ is the union of those sections of $\mathcal{G}_{\mathfrak{p}}$ containing a representative of $\mathbf{b}_1$.

Since each $(\mathcal{G}_{\mathfrak{p}})_x$ is $(B, d-h-1)$-generic, we have

$$|(\mathcal{G}_{\mathfrak{p}})_x| \geq \frac{\mathcal{N}_K(\mathfrak{p})^{d-h-1}}{B} |(\mathcal{G}_{\mathfrak{p}})_x(\mathbf{b}_1)|,$$

thus

$$|\mathcal{B}_1| \geq \frac{\mathcal{N}_K(\mathfrak{p})^{d-h-1}}{B} |\mathcal{G}_{\mathfrak{p}}(\mathbf{b}_1)| \geq \frac{1}{B_1 B} |\mathcal{G}_{\mathfrak{p}}(a)|. \tag{5.44}$$

Note that since $|\mathcal{G}_{\mathfrak{p}}(a)| \geq |\mathcal{B}_1|$ and $|[\mathcal{G}_{\mathfrak{p}}(a)]_{\mathfrak{p}}| \leq B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}$, by the first inequality of (5.44) and the pigeonhole principle we may find another residue class $\mathbf{b}_2 \in [\mathcal{G}_{\mathfrak{p}}(a)]_{\mathfrak{p}}$ with

$$|\mathcal{G}_{\mathfrak{p}}(\mathbf{b}_2)| \geq \frac{1}{B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}} |\mathcal{G}_{\mathfrak{p}}(a) \backslash \mathcal{G}_{\mathfrak{p}}(\mathbf{b}_1)| \geq \frac{1}{B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}} \left(1 - \frac{B}{\mathcal{N}_K(\mathfrak{p})^{d-h-1}}\right) |\mathcal{G}_{\mathfrak{p}}(a)|,$$

which is at least $\frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{2B_1 \mathcal{N}_K(\mathfrak{p})^{d-h-1}}$ if $\mathcal{N}_K(\mathfrak{p})^{d-h-1} > 2B$. In such case, if we define $\mathcal{B}_2$ as in (5.3), the same reasoning that gives (5.44) implies $|\mathcal{B}_2| \geq \frac{1}{2B_1 B} |\mathcal{G}_{\mathfrak{p}}(a)|$. Iterating this process we obtain a sequence $\mathbf{b} = \{\mathbf{b}_1, \dots, \mathbf{b}_q\}$ of residue classes, $q = \left\lceil \frac{\mathcal{N}_K(\mathfrak{p})^{d-h-1}}{2B} \right\rceil$, satisfying

$$|\mathcal{G}_p(\mathbf{b}_j)| \geq \frac{1}{B_1 \mathcal{N}(p)^{d-h-1}} |\mathcal{G}_p(a) \backslash \bigcup_{i=1}^{j-1} \mathcal{G}_p(\mathbf{b}_i)| \geq \frac{1}{B_1 \mathcal{N}(p)^{d-h-1}} \left(1 - \frac{(q-1)B}{\mathcal{N}(p)^{d-h-1}}\right) |\mathcal{G}_p(a)|,$$

and $|\mathcal{B}_j| \geq \frac{1}{2B_1 B} |\mathcal{G}_{\mathfrak{p}}(a)|$. In particular,

$$\sum_{j=1}^{q} |\mathcal{B}_j| \geq \frac{q}{2B_1 B} |\mathcal{G}_{\mathfrak{p}}(a)|. \tag{5.45}$$

Now, consider the set

$$\mathcal{B}[a] := \left\{ s \in \mathcal{G}_{\mathfrak{p}}(a) : \sum_{j=1}^{q} 1_{s \in \mathcal{B}_j} \geq \frac{q}{4B_1 B} \right\}.$$

Note that $\mathcal{B}[a]_x := \mathcal{B}[a] \cap \pi_1^{-1}(x)$ equals to $(\mathcal{G}_{\mathfrak{p}})_x$ whenever this intersection is non-empty. Additionally, (5.45) implies

$$\begin{aligned} \frac{q}{2B_1 B} |\mathcal{G}_p(a)| &\leq \sum_{j=1}^{q} |\mathcal{B}_j| = \sum_{j=1}^{q} \sum_{s} 1_{s \in \mathcal{B}_j} = \sum_{s} \sum_{j=1}^{q} 1_{s \in \mathcal{B}_j} = \sum_{s \in \mathcal{B}[a]} \sum_{j=1}^{q} 1_{s \in \mathcal{B}_j} + \sum_{s \notin \mathcal{B}[a]} \sum_{j=1}^{q} 1_{s \in \mathcal{B}_j} \\ &\leq \sum_{s \in \mathcal{B}[a]} q + \sum_{s \notin \mathcal{B}[a]} \frac{q}{4B_1 B} \leq q \left(|\mathcal{B}[a]| + \frac{1}{4B_1 B} |\mathcal{G}_p(a)| \right), \end{aligned} \tag{5.46}$$

from which we deduce

$$|\mathcal{B}[a]| \geq \frac{1}{4B_1B}|\mathcal{G}_{\mathfrak{p}}(a)|. \tag{5.47}$$

Claim 5.13. *For each x such that $\mathcal{B}[a]_x \neq \emptyset$, there are at least $\frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{(4B_1B)^2}$ elements $s \in \mathcal{G}_{\mathfrak{p}}(a)$ such that $s \equiv \mathbf{y}(mod\ \mathfrak{p})$ for some $\mathbf{y} \in \mathcal{B}[a]_x$.*

Proof. Indeed, let x such that $\mathcal{B}[a]_x \neq \emptyset$. Thus, $\mathcal{B}[a]_x = (\mathcal{G}_{\mathfrak{p}})_x$ and by definition of $\mathcal{B}[a]$, it holds that $(\mathcal{G}_{\mathfrak{p}})_x \subseteq \mathcal{B}_j$ for at least $\frac{q}{4B_1B}$ values of j. Now, fix any such j. Then by definition of $\mathcal{B}_j$, there exists $s_j \in \mathcal{G}_{\mathfrak{p}}(a)$ such that $(\mathcal{G}_{\mathfrak{p}})_x = (\mathcal{G}_{\mathfrak{p}})_{\pi_1(s_j)}$ and $s_j \equiv \mathbf{b}_j(\text{mod } \mathfrak{p})$. Note that $s_j \in (\mathcal{G}_{\mathfrak{p}})_x$. Hence, for any $s \in \mathcal{G}_{\mathfrak{p}}(\mathbf{b}_j)$, it holds $s \equiv \mathbf{b}_j \equiv s_j(\text{mod } \mathfrak{p})$. We deduce that there are $|\mathcal{G}_{\mathfrak{p}}(\mathbf{b}_j)| \geq \frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{2B_1\mathcal{N}_K(\mathfrak{p})^{d-h-1}}$ elements $s \in \mathcal{G}_{\mathfrak{p}}(a)$ such that $s \equiv \mathbf{y}(\text{mod } \mathfrak{p})$ for some $\mathbf{y} \in \mathcal{G}_{\mathfrak{p}}(a)$. Now, the elements we constructed have residue class $\mathbf{b}_j$ modulo $\mathfrak{p}$. Since the residue classes $\mathbf{b}_j$ are all different, we conclude that there are at least $\frac{q}{4B_1B}\frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{2B_1\mathcal{N}_K(\mathfrak{p})^{d-h-1}} \geq \frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{(4B_1B)^2}$ elements $s \in \mathcal{G}_{\mathfrak{p}}(a)$ such that $s \equiv \mathbf{y}(\text{mod } \mathfrak{p})$ for some $\mathbf{y} \in \mathcal{B}[a]_x$. □

Now, let

$$\mathcal{R} := \left\{a \in [\pi_1(S')]_{\mathfrak{p}} \subseteq \mathcal{O}_K/\mathfrak{p} : |\mathcal{G}_{\mathfrak{p}}(a)| \geq \frac{1}{2\mathcal{N}_K(\mathfrak{p})}|\mathcal{G}_{\mathfrak{p}}|\right\}$$

and let us denote

$$\mathcal{B}[\mathfrak{p}] := \left\{s \in S' : (S')_{\pi_1(s)} \cap \mathcal{B}[a] \neq \emptyset \text{ for some } a \in \mathcal{R}\right\}.$$

In other words, $\mathcal{B}[\mathfrak{p}]$ consists of those sections of S' with non-trivial intersection with $\bigcup_{a\in\mathcal{R}}\mathcal{B}[a]$. In particular, since $\mathcal{B}[\mathfrak{p}]$ contains the disjoint union $\bigcup_{a\in\mathcal{R}}\mathcal{B}[a]$, from the definition of $\mathcal{R}$ we deduce:

Claim 5.14. *The following bound holds*

$$|\mathcal{B}[\mathfrak{p}]| \geq \frac{\beta c_1}{8B_1B}|S'|.$$

Proof of Claim 5.14. First note that by the pigeonhole principle, $\mathcal{R} \neq \emptyset$. Now, if $\mathcal{R} = \{a_1, \ldots, a_h\}$,

$$|\mathcal{G}_{\mathfrak{p}}| = \sum_a |\mathcal{G}_{\mathfrak{p}}(a)| = \sum_{a\in\mathcal{R}} |\mathcal{G}_{\mathfrak{p}}(a)| + \sum_{a\notin\mathcal{R}} |\mathcal{G}_{\mathfrak{p}}(a)| < \sum_{a\in\mathcal{R}} |\mathcal{G}_{\mathfrak{p}}(a)| + \frac{\mathcal{N}_K(\mathfrak{p}) - h}{2\mathcal{N}_K(\mathfrak{p})}|\mathcal{G}_{\mathfrak{p}}|,$$

thus

$$\sum_{a\in\mathcal{R}} |\mathcal{G}_{\mathfrak{p}}(a)| > \left(1 - \frac{\mathcal{N}_K(\mathfrak{p}) - h}{2\mathcal{N}_K(\mathfrak{p})}\right)|\mathcal{G}_{\mathfrak{p}}| = \frac{\mathcal{N}_K(\mathfrak{p}) + h}{2\mathcal{N}_K(\mathfrak{p})}|\mathcal{G}_{\mathfrak{p}}| > \frac{|\mathcal{G}_{\mathfrak{p}}|}{2}.$$

Combining this together with (5.42) and (5.47) we get

$$|\mathcal{B}[\mathfrak{p}]| \geq \left|\bigcup_{a\in\mathcal{R}} \mathcal{B}[a]\right| = \sum_{a\in\mathcal{R}} |\mathcal{B}[a]| \geq \sum_{a\in\mathcal{R}} \frac{1}{4B_1B}|\mathcal{G}_{\mathfrak{p}}(a)| > \frac{1}{8B_1B}|\mathcal{G}_{\mathfrak{p}}| > \frac{\beta c_1}{8B_1B}|S'|.$$

□

For an element $s \in S'$ denote P''_s for the set of primes $\mathfrak{p} \in P'$ for which $s \in \mathcal{B}[\mathfrak{p}]$.

Claim 5.15. *There exist constants κ_3 and c_3 such that the set*

$$\mathcal{B} := \left\{s \in S' : w(P''_s) \geq \kappa_3 w(P')\right\}$$

satisfies $|\mathcal{B}| \geq c_3|S'|$.

Proof. On the one hand we have

$$\sum_{\mathfrak{p}\in P'}\sum_{s\in S'} 1_{s\in\mathcal{B}[\mathfrak{p}]}\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} = \sum_{\mathfrak{p}\in P'}|\mathcal{B}[\mathfrak{p}]|\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} \geq \frac{\beta c_1}{8B_1B}|S'|w(P').$$

On the other hand,

$$\begin{aligned}\sum_{s\in S'}\sum_{\mathfrak{p}\in P'} 1_{s\in\mathcal{B}[\mathfrak{p}]}\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} &= \sum_{s\in S'}\sum_{\mathfrak{p}\in P''_s}\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})} = \sum_{s\in S'} w(P''_s) = \sum_{s\in\mathcal{B}} w(P''_s) + \sum_{s\notin\mathcal{B}} w(P''_s)\\ &\leq |\mathcal{B}|w(P') + \kappa_3|S'|w(P'),\end{aligned}$$

because $P''_s \subseteq P'$. Hence $\left(\frac{\beta c_1}{8B_1B} - \kappa_3\right)|S'| \leq |\mathcal{B}|$. Thus, it is enough to take

$$\kappa_3 = \frac{\beta c_1}{16B_1B} = c_3. \tag{5.48}$$

□

Now, let us check that $\mathcal{B}$ satisfies the condition of lemma 5.12. Let x be such that $\mathcal{B}_x \neq \emptyset$. By construction, $\mathcal{B}[a]_x \subseteq \mathcal{B}_x$ for some $a \in \mathcal{R}$. Since there are at least $\frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{(4B_1B)^2}$ elements $s \in \mathcal{G}_{\mathfrak{p}}(a)$ such that $s \equiv \mathbf{y}(\mathrm{mod}\ \mathfrak{p})$ for some $\mathbf{y} \in \mathcal{B}[a]_x$, we conclude

$$\left|\left\{s \in S' : [s]_{\mathfrak{p}} \in [\mathcal{B}_x]_{\mathfrak{p}}\right\}\right| \geq \frac{|\mathcal{G}_{\mathfrak{p}}(a)|}{(4B_1B)^2} \geq \frac{1}{2^5(B_1B)^2}\frac{1}{\mathcal{N}_K(\mathfrak{p})}|\mathcal{G}_{\mathfrak{p}}| \geq \frac{\beta c_1}{2^5(B_1B)^2}\frac{|S'|}{\mathcal{N}_K(\mathfrak{p})} = c_4\frac{|S'|}{\mathcal{N}_K(\mathfrak{p})},$$

where

$$c_4 = \frac{\beta c_1}{2^5(B_1B)^2} \tag{5.49}$$

This finishes the proof of Lemma 5.12. □

In order to conclude the proof of Proposition 5.10 we are going to show that if an r-polynomial vanishes at the sections $\mathcal{B}_x$ for $\gtrsim_{r,d,h,\varepsilon,\kappa,K} 1$ values of x, then it should vanish at a positive proportion of S. To show this, we choose m different sections of S' having non-trivial intersection with $\mathcal{B}$, where $m = O_{r,d,h,\varepsilon,\kappa,K}(1)$ is to be specified later. Note that because of (5.30) and Lemma 5.12 this is always possible, provided that N is large enough.

Indeed, let $S'_{x_1}, \dots, S'_{x_l}$ be all the sections of S', and let $S'_{x_1}, \dots, S'_{x_m}$ be those with $S'_{x_i} \cap \mathcal{B} \neq \emptyset$. Thus

$$\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\sum_{s\in S'} 1_{\exists x_i:[s]_{\mathfrak{p}}\in[\mathcal{B}_{x_i}]_{\mathfrak{p}}}\log(\mathcal{N}_K(\mathfrak{p})) \leq \sum_{\mathfrak{p}\in\mathcal{P}(Q)}\sum_{i=1}^{m}|S'_{x_i}|\log(\mathcal{N}_K(\mathfrak{p})) \leq \frac{2|S|m}{Q}\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\log(\mathcal{N}_K(\mathfrak{p})) \leq 2|S|mc_{4,K} \tag{5.50}$$

where the first inequality is because $\mathcal{B}_{x_i} \neq \emptyset$ if and only if $1 \leq i \leq m$, the second inequality is due to 5.30 and the third inequality is due to 3.2.

On the other hand we have that

$$\begin{aligned}\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\sum_{s\in S'} 1_{\exists x_i:[s]_{\mathfrak{p}}\in[\mathcal{B}_{x_i}]_{\mathfrak{p}}}\log(\mathcal{N}_K(\mathfrak{p})) &\geq \sum_{\mathfrak{p}\in P''_{x_i}}\left|\left\{s \in S' : [s]_{\mathfrak{p}} \in [\mathcal{B}_{x_i}]_{\mathfrak{p}}\right\}\right|\log(\mathcal{N}_K(\mathfrak{p})) \geq \sum_{\mathfrak{p}\in P''_{x_i}} c_4|S'|\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})}\\ &= c_4|S'|w(P''_{x_i}) \geq c_4|S'|\kappa_3 w(P') \geq c_4\kappa_3|S'|\frac{\kappa^2\kappa_1}{4}w(\mathcal{P}(Q)).\end{aligned} \tag{5.51}$$

Where the last two inequalities are because we can chose $x_i \in \mathcal{B}$ in claim 5.15 and then we use (5.41). Since by 5.36 we have that $|S'| \geq \frac{\delta_0}{2^{3d+2}}|S|$, then

$$m \geq \frac{c_4\kappa_3\kappa^2\kappa_1\delta_0}{2^{3d+5}c_{4,K}}w(\mathcal{P}(Q)) \tag{5.52}$$

Thus, if N is large enough, m will be large enough.

Denote $\mathcal{L} := \{S'_{x_1},\dots,S'_{x_m}\}$ this choice of m sections such that $S'_{x_i}\cap\mathcal{B}\neq\varnothing$ for all i. Let $P_{\mathcal{L}}$ be the set of primes $\mathfrak{p}$ in $\mathcal{P}(Q)$ for which there exist a pair of sections $S'_{x_i}\neq S'_{x_j}$ in $\mathcal{L}$ such that $[S'_{x_i}]_{\mathfrak{p}}\cap[S'_{x_j}]_{\mathfrak{p}}\neq\varnothing$. Given such a pair of sections, the fact that $[S'_{x_i}]_{\mathfrak{p}}\cap[S'_{x_j}]_{\mathfrak{p}}\neq\varnothing$ implies $x_i\equiv x_j (\mathrm{mod}\ \mathfrak{p})$. Since $x_i\neq x_j$ and $\mathcal{N}_K(x_i-x_j)\leq N^3$ when $N\geq 2^{d_K}$, then

$$\sum_{\mathfrak{p}\in P_{\mathcal{L}}}\log(\mathcal{N}_K(\mathfrak{p}))\leq\sum_{\{i,j\},i\neq j}\sum_{\mathfrak{p}|x_i-x_j}\log(\mathcal{N}_K(\mathfrak{p}))\leq\sum_{\{i,j\},i\neq j}\log(\mathcal{N}_K(x_i-x_j))\leq\binom{m}{2}3\log(N). \tag{5.53}$$

This implies $w(P_{\mathcal{L}})\lesssim_{r,d,h,\varepsilon,\kappa,K}\log(\log(N))$.

Now, we consider in S' the function

$$\Psi_{\mathcal{L}}(s) := \sum_{\mathfrak{p}\in\mathcal{P}(Q)} 1_{\exists\mathbf{x}\in\mathcal{L}:s\equiv\mathbf{x}(\mathrm{mod}\ \mathfrak{p})}\log(\mathcal{N}_K(\mathfrak{p})).$$

Note that $\Psi_{\mathcal{L}}(s)$ measures how much a residue class occupied by s contains a representative in $\mathcal{L}$. From Lemma 5.12 and (5.53) we deduce

$$\sum_{s\in S'}\Psi_{\mathcal{L}}(s)\geq\sum_{i=1}^{m}\sum_{\mathfrak{p}\in P_{x_i}\backslash P_{\mathcal{L}}}\sum_{s\in S'}1_{\exists\mathbf{x}\in S'_{x_i}:s\equiv\mathbf{x}(\mathrm{mod}\ \mathfrak{p})}\log(\mathcal{N}_K(\mathfrak{p}))\geq\sum_{i=1}^{m}\sum_{\mathfrak{p}\in P_{x_i}\backslash P_{\mathcal{L}}}c_4|S'|\frac{\log(\mathcal{N}_K(\mathfrak{p}))}{\mathcal{N}_K(\mathfrak{p})}=c_4|S'|\sum_{i=1}^{m}\left(w(P_{x_i})-w(P_{\mathcal{L}})\right).$$

Here, the first inequality is because for any prime $\mathfrak{p}\in\bigcup_i P_{x_i}\backslash P_{\mathcal{L}}$ we have $[S'_{x_i}]_{\mathfrak{p}}\cap[S'_{x_j}]_{\mathfrak{p}}=\varnothing$ for any $i\neq j$, thus the conditions $\exists\mathbf{x}\in S'_{x_i}: s\equiv\mathbf{x}(\mathrm{mod}\ \mathfrak{p})$ are pairwise disjoint. Now, since $w(P_{x_i})\geq\kappa_3 w(P')\geq\frac{\kappa_3\kappa_1\kappa}{4}w(P)\geq\frac{\kappa_3\kappa_1\kappa^2}{4}w(\mathcal{P}(Q))$ and $w(P_{\mathcal{L}})\lesssim_{r,d,h,\varepsilon,\kappa,K}\log(\log(N))$, it holds

$$\begin{aligned}\sum_{s\in S'}\Psi_{\mathcal{L}}(s)&\geq mc_4|S'|\left(\frac{\kappa_3\kappa_1\kappa^2}{4}w(\mathcal{P}(Q))+O_{r,d,h,\varepsilon,\kappa,K}(\log(\log(N)))\right)\geq\frac{1}{2}mc_4|S'|\frac{\kappa_3\kappa_1\kappa^2}{4}w(\mathcal{P}(Q))\\&\geq mc_4\frac{\delta_0}{2^{3d+2}}|S|\frac{\kappa_3\kappa_1\kappa^2}{8}w(\mathcal{P}(Q))\geq c_5 m|S|\log(N),\end{aligned} \tag{5.54}$$

for N large enough. Here,

$$c_5=\frac{\kappa^2}{2^{3d+6}}c_{2,K}\frac{\varepsilon}{d}c_4\kappa_1\kappa_3\delta_0. \tag{5.55}$$

We will now bound $\sum_{s\in S'}\Psi_{\mathcal{L}}(s)$ from above. In order to achieve that, first note that for s in $\mathcal{L}$, (5.30) implies that

$$\sum_{s\in\mathcal{L}}\Psi_{\mathcal{L}}(s)=\sum_{s\in\mathcal{L}}\sum_{\mathfrak{p}\in\mathcal{P}(Q)}1_{\exists\mathbf{x}\in\mathcal{L}:s\equiv\mathbf{x}(\mathrm{mod}\ \mathfrak{p})}\log(\mathcal{N}_K(\mathfrak{p}))=\sum_{s\in\mathcal{L}}\sum_{\mathfrak{p}\in\mathcal{P}(Q)}\log(\mathcal{N}_K(\mathfrak{p}))\leq c_{4,K}Q|\mathcal{L}|\leq c_{4,K}2m|S| \tag{5.56}$$

Also note that if $s \notin \mathcal{L}$, then

$$\begin{aligned}\Psi_{\mathcal{L}}(s) &= \sum_{\mathfrak{p}\in\mathcal{P}(Q)} 1_{\exists \mathbf{x}\in\mathcal{L}: s\equiv\mathbf{x} (\mathrm{mod}\ \mathfrak{p})} \log(\mathcal{N}_K(\mathfrak{p})) \le \sum_{\mathfrak{p}\in\mathcal{P}(Q)} \sum_{i=1}^{m} 1_{\exists \mathbf{x}\in S'_{x_i}: s\equiv\mathbf{x} (\mathrm{mod}\ \mathfrak{p})} \log(\mathcal{N}_K(\mathfrak{p})) \\ &= \sum_{i=1}^{m} \sum_{\mathfrak{p}\in\mathcal{P}(Q)} 1_{\exists \mathbf{x}\in S'_{x_i}: s\equiv\mathbf{x} (\mathrm{mod}\ \mathfrak{p})} \log(\mathcal{N}_K(\mathfrak{p})) \le \sum_{i=1}^{m} \sum_{\mathfrak{p}\mid\pi_1(s)-x_i} \log(\mathcal{N}_K(\mathfrak{p})) \le \sum_{i=1}^{m} \log(\mathcal{N}_K(\pi_1(s)-x_i)).\end{aligned} \tag{5.57}$$

By (2.2), (2.6) and the fact that $\pi_1(s), x_i \in [N]_{\mathcal{O}_K}$, we have $\mathcal{N}_K(\pi_1(s)-x_i) \le N^3$ when $N \ge 2^{d_K}$, thus from (5.57) we deduce that for $s \notin \mathcal{L}$ it holds

$$\Psi_{\mathcal{L}}(s) \le 3m\log(N). \tag{5.58}$$

Let $\delta_1 := \frac{c_5}{4}$, and suppose that the set

$$L := \{s \in S' : \Psi_{\mathcal{L}}(s) \ge \gamma\}$$

has size at most $\delta_1|S|$. Then

$$\begin{aligned}\sum_{s\in S'} \Psi_{\mathcal{L}}(s) &= \sum_{\substack{s\in S' \\ \Psi_{\mathcal{L}}(s)<\gamma}} \Psi_{\mathcal{L}}(s) + \sum_{\substack{s\in\mathcal{L} \\ \Psi_{\mathcal{L}}(s)\ge\gamma}} \Psi_{\mathcal{L}}(s) + \sum_{\substack{s\notin\mathcal{L} \\ \Psi_{\mathcal{L}}(s)\ge\gamma}} \Psi_{\mathcal{L}}(s) \\ &\le \sum_{\substack{s\in S' \\ \Psi_{\mathcal{L}}(s)<\gamma}} \Psi_{\mathcal{L}}(s) + \sum_{s\in\mathcal{L}} \Psi_{\mathcal{L}}(s) + \sum_{\substack{s\in S' \\ \gamma\le\Psi_{\mathcal{L}}(s)\le 3m\log(N)}} \Psi_{\mathcal{L}}(s) \\ &< \gamma|S| + 2c_{4,K}m|S| + \delta_1|S|3m\log(N)\end{aligned} \tag{5.59}$$

If we now set $\gamma := 3rd_K\log(N)$, combining 5.54 with 5.59 we get that

$$4\delta_1 m\log(N) < 3rd_K\log(N) + 2c_{4,K}m + 3\delta_1 m\log(N),$$

and then

$$m\left(\delta_1 - \frac{2c_{4,K}}{\log N}\right) < 3rd_K.$$

So, if we take $m := \frac{4rd_K}{\delta_1}$ we reach a contradiction when $N > \exp\left(\frac{8c_{4,K}}{\delta_1}\right)$. We conclude that for N sufficiently large, the set

$$L := \{s \in S' : \Psi_{\mathcal{L}}(s) \ge 3rd_K\log(N)\}$$

has size $|L| \ge \delta_1|S|$ for our choices of m and δ_1.
Since

$$\Psi_{\mathcal{L}}(s) = \log\left(\prod_{\substack{\mathfrak{p}\in\mathcal{P}(Q) \\ \exists\mathbf{x}\in\mathcal{L}: s\equiv\mathbf{x} (\mathrm{mod}\ \mathfrak{p})}} \mathcal{N}_K(\mathfrak{p})\right),$$

it follows that if $s \in L$ then

$$\prod_{\substack{\mathfrak{p}\in\mathcal{P}(Q) \\ \exists\mathbf{x}\in\mathcal{L}: s\equiv\mathbf{x} (\mathrm{mod}\ \mathfrak{p})}} \mathcal{N}_K(\mathfrak{p}) \ge N^{3rd_K}.$$

Now, let us see that if an r-polynomial vanishes at $\mathcal{L}$, then it must vanish at L. Indeed, let f be such a polynomial and let $s \in L$. If $\mathfrak{p} \in \mathcal{P}(Q)$ is a prime such that there exists $\mathbf{x} \in \mathcal{L}$ with $s \equiv \mathbf{x}(\text{mod } \mathfrak{p})$, the fact that $f(\mathbf{x}) = 0$ implies that $\mathfrak{p}|f(s)$.By Definition 5.8, $H_K(f(\mathbf{y})) < N^{3rd_K}$ for all $\mathbf{y} \in [N]^d_{\mathcal{O}_K}$. Hence if $f(s) \neq 0$ it holds that

$$N^{3rd_K} \leq \prod_{\substack{\mathfrak{p}\in\mathcal{P}(Q)\\ \exists \mathbf{x}\in\mathcal{L}: s\equiv\mathbf{x}(\text{mod } \mathfrak{p})}} \mathcal{N}(\mathfrak{p}) \leq \prod_{\mathfrak{p}|f(s)} \mathcal{N}_K(\mathfrak{p}) \leq \mathcal{N}_K(f(s)) \leq H_K(f(s)) < N^{3rd_K},$$

which is absurd. Thus $f(s) = 0$.

By the induction hypothesis and the construction of S' we know that each section $S'_{x_i} \in \mathcal{L}$ contains an $(r, 1)$-characteristic subset of size at most $c_2 r^{d-h-1}$. Taking the union of these m subsets we obtain an (r, δ_1)-characteristic subset for S, of size at most $\frac{m}{r} c_2 r^{d-h} = \frac{4d_K}{\delta_1} c_2 r^{d-h}$. □

5.3. **Construction of a polynomial of low complexity and conclusion of the proof.** Having constructed a characteristic subset $A \subseteq S$, the last step to prove Theorem 5.1 is to construct a small polynomial vanishing at A. This will be done by using the following variant of Siegel's lemma.

Lemma 5.16 ([Par19, Lemma 3.6])**.** *Let K be a global field of degree d_K. Let $(a_{ij})_{i,j}$, $1 \leq i \leq s$, $1 \leq j \leq t$ be elements of $\mathcal{O}_K$ with $H(a_{ij}) \leq C$ for all i, j. Let us suppose that $t > 2d_K^2 s$. Then, there exists $\mathbf{c} = (c_1, \ldots, c_t) \in \mathcal{O}_K^t \backslash \{0\}$, such that*

$$H_K(1 : \mathbf{c}) \leq c_{6,K}\,(tC)^{\frac{8d_K^2 s}{t-2d_K s}}$$

and

$$\sum_{j=1}^{t} c_j a_{ij} = 0 \text{ for all } 1 \leq i \leq s.$$

Proof of Theorem 5.1. The proof of the theorem is exactly the same as in [Wal12, Theorem 2.4], replacing the classical Siegel's lemma and [Wal12, Proposition 2.2] by Lemma 5.16 and Proposition 5.10. We sketch the details. First, it is enough to prove Theorem 5.1 for $\eta = 1 - \delta$; the general case follows by a simple partitioning argument. Thus, we have $S \subseteq [N]^{d+1}_{\mathcal{O}_K}$ of size $|S| \geq cN^{d-h-1+\varepsilon}$ occupying at most $\alpha\mathcal{N}_K(\mathfrak{p})^{d-h}$ residue classes modulo $\mathfrak{p}$ for every prime $\mathfrak{p} \in P$. By Proposition 5.10 there exists a subset $A \subseteq S$ of size $|A| \leq c_2 r^{d-h}$ which is (r, δ)-characteristic for S, provided that N is large enough. Consider the system of $|A|$-linear equations in $\binom{r+d}{d}$ variables given by

$$\sum_{\mathbf{i}=(i_0,\ldots,i_d)} \beta_{\mathbf{i}} \mathbf{a}^{\mathbf{i}} = 0 \text{ for all } \mathbf{a} \in A,$$

where we are using the multi-index notation $\mathbf{a}^{\mathbf{i}} = a_0^{i_0} \ldots a_d^{i_d}$ for $\mathbf{a} = (a_0, \ldots, a_d)$ and we are summing over the $\mathbf{i}$'s with $i_0 + \ldots + i_d = r$. Note that $H_K(\mathbf{a}^{\mathbf{i}}) \leq N^r$. Now choose r large enough such that

$$\binom{r+d}{d} > 18d_K^2|A|, \tag{5.60}$$

namely, since $\binom{r+d}{d} \geq \frac{r^d}{d!}$ and $|A| \leq c_2 r^{d-h}$ it is enough to choose

$$\frac{r^d}{d!} > 18d_K^2 c_2 r^{d-h}, \text{ i.e. } r > \left(18d_K^2 c_2 d!\right)^{\frac{1}{h}}. \tag{5.61}$$

Note that (5.60) implies

$$\frac{16d_K|A|}{\binom{r+d}{d} - 2d_K|A|} < 1. \tag{5.62}$$

By Lemma 5.16 and (5.62) there exists a solution $(\beta_{\mathbf{i}}) \in \mathcal{O}_K^{\binom{r+d}{d}} \backslash \{0\}$ such that

$$H_K(1:\beta_{\mathbf{i}}) \leq c_{6,K} \left[\binom{r+d}{d} N^r \right]^{\frac{8d_K^2|A|}{\binom{r+d}{d} - 2d_K|A|}} \leq c_{6,K} \binom{r+d}{d}^{\frac{d_K}{2}} N^{\frac{rd_K}{2}} \leq N^{rd_K}$$

provided N is large enough. Consider now the polynomial $f = \sum_{\mathbf{i}} \beta_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$. For any $\mathbf{x} \in [N]_{\mathcal{O}_K}^{d+1}$ (2.4) and (2.5) imply:

$$H_K(f(\mathbf{x})) \leq \binom{r+d}{d}^{d_K} H_K(1:(\beta_{\mathbf{i}})) H_K(1:\mathbf{x})^r < \binom{r+d}{d}^{d_K} N^{2rd_K} < N^{3rd_K}$$

for N large enough. Thus, f is an homogeneous polynomial of degree r that vanishes at A. This concludes the proof. In particular, we can choose $r = \left\lceil \left(18c_2 d_K^2 d!\right)^{\frac{1}{h}} \right\rceil$. □

Remark 5.17. Let us note that while Theorem 5.1 implies Theorem 4.4, and hence Theorem 1.6, it does not imply Theorem 1.5. The reason for this is that the statement of Theorem 1.5 is in terms of sets S lying in $[N]_{\mathcal{O}_K}^d$. Instead, if in Theorem 1.5 the set S lies in $[N]_{\mathbb{A}^d(\mathcal{O}_K)}$, then the conclusion of the theorem follows immediately from Theorem 4.4. In order to prove Theorem 1.5 as it is stated in the introduction, let us observe the basic fact that the number of monomials of degree r in variables $X_0, \ldots, X_d$ coincides with the number of monomials of degree at most r in variables $X_1, \ldots, X_d$. Then a very easy modification in the construction of the auxiliary polynomial in § 5.3 gives the following result.

Theorem 5.18. *Let d, h be positive integers and let $\varepsilon, \eta > 0$ be positive real numbers. Let K be a global field of degree d_K and $\mathcal{O}_K$ be its ring of integers. Set $Q = N^{\frac{\varepsilon}{2d}}$ and let $P \subseteq \mathcal{P}(Q)$ satisfy $w(P) \geq \kappa \log(Q)$ for some $\kappa > 0$. Suppose that $S \subseteq [N]_{\mathcal{O}_K}^d$ is a set of size $|S| \geq cN^{d-h-1+\varepsilon}$ occupying at most $\alpha \mathcal{N}_K(\mathfrak{p})^{d-h}$ residue classes modulo $\mathfrak{p}$ for every prime $\mathfrak{p} \in P$ and some $\alpha > 0$. Then if N is sufficiently large there exists a non-zero polynomial $f \in \mathcal{O}_K[X_1, \ldots, X_d]$ of degree $O_{d,h,\varepsilon,\eta,\kappa,K}(1)$ and coefficients bounded by $N^{O_{d,h,\varepsilon,\eta,\kappa,K}(1)}$ which vanishes at more than $(1-\eta)|S|$ points of S.*

Now, it is immediate that Theorem 5.18 implies an adequate version of Theorem 4.4, from where we deduce Theorem 1.5.

As a closing remark, we mention that Theorem 1.4, Theorem 1.5, Theorem 5.1, and Theorem 5.18 are also valid when $S \subseteq [N]_{\mathcal{O}_{K,S}}^d$ where, as in section §2, $\mathcal{O}_{K,S}$ denotes the rings of S-integers, for S a finite subset of places containing $M_{K,\infty}$ when K is a number field, and containing the distinguished place v in the case when K is a function field over $\mathbb{F}_q$.

6. Appendix: effective estimates for heights over global fields

In this appendix we provide the proofs of Proposition 2.1 and 2.2.

In number fields and for $S = M_{K,\infty}$, Proposition 2.1 is proved in [Ser89, §13.4]. There, Serre also observed that the constant c can be made effective. In order to prove Proposition 2.1, we adapt Serre's proof to global fields and general rings of S-integers. Since all the results in this article are effective, we take the opportunity to fulfill Serre's observation by carrying on the extra work needed to make the dependence of c on K and S explicit.

Proof of Proposition 2.1. The statement is trivial if $K = \Bbbk = \mathbb{Q}$ and $S = \{\infty\}$ or $K = \Bbbk = \mathbb{F}_q(T)$ and S is the place corresponding to the infinite point in $\mathbb{P}^1(K)$: for any set of coordinates of $\mathbf{x}$, clean denominators and common factors. In this way we obtain coordinates $(x_0 : \ldots : x_d)$ such that $x_i \in \mathcal{O}_{\Bbbk}$, $\gcd(x_0, \ldots, x_d) = 1$ and

$$H_K(1 : x_0 : \ldots : x_d) = H_K(\mathbf{x}).$$

Now, let $\mathbf{x} \in \mathbb{P}^d(K)$ and choose coordinates $(x_0, \ldots, x_d)$ with $x_i \in \mathcal{O}_{K,S}$. If $\mathfrak{a}_{x_0,\ldots,x_d} := \sum_{i=0}^d x_i \mathcal{O}_{K,S}$ then

$$H(\mathbf{x}) = \frac{1}{\mathcal{N}_K(\mathfrak{a}_{x_0,\ldots,x_d})^{\frac{1}{d_K}}} \prod_{v \in S} \max_i \|x_i\|_v.$$

Note that the ideal $\mathfrak{a}_{x_0,\ldots,x_d}$ depends on the coordinates, but its ideal class depends only on $\mathbf{x}$. Hence, if we take integral ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_l$ representing all the ideal classes of $\mathcal{O}_{K,S}$, satisfying Minkowski's bound $\mathcal{N}_K(\mathfrak{a}_j) \lesssim_K 1$ for all j (when K is a function field, this bound may be deduced from [Lor96, Chapter V, §8.9.]), it holds that $\mathfrak{a}_{x_0,\ldots,x_d}\mathfrak{a}_j^{-1} = \alpha \mathcal{O}_{K,S}$ for some j and some $\alpha \in K^\times$. Thus, $\alpha^{-1}\mathcal{O}_{K,S} \cdot \mathfrak{a}_{x_0,\ldots,x_d} = \mathfrak{a}_j$. We conclude that $(x_0', \ldots, x_d') := (\alpha^{-1}x_0, \ldots, \alpha^{-1}x_d)$ are coordinates of $\mathbf{x}$ satisfying $\alpha^{-1}x_i \in \mathcal{O}_{K,S}$ for all i and $\mathfrak{a}_{x_0',\ldots,x_d'} = \mathfrak{a}_j$. In particular,

$$H(\mathbf{x}) = \frac{1}{\mathcal{N}_K(\mathfrak{a}_j)^{\frac{1}{d_K}}} \prod_{v \in S} \max_i \|x_i'\|_v.$$

Now, we use the following lemma, whose proof will be given after we finish the proof of Proposition 2.1

Lemma 6.1. *Let K be a global field and S a set of places which contains the infinite places when K is a number field. There exists a constant $c_{K,S}' > 1$, depending only on K, S, such that for every $\boldsymbol{x} = (x_v)_{v \in S} \in (\mathbb{R}_{>0})^{|S|}$ there exist $\varepsilon \in \mathcal{O}_{K,S}^\times$ and $t > 0$ verifying that*

$$\left(c_{K,S}'\right)^{-1} \frac{x_v}{t} \leq \|\varepsilon\|_v \leq c_{K,S}' \frac{x_v}{t} \text{ for all } v \in S. \tag{6.1}$$

Fix any $v_0 \in S$. Taking $(x_v)_{v \in S} := (\max_i \|x_i'\|_v)_{v \in S}$, Lemma 6.1 implies that there exist $t > 0$ and $\varepsilon \in \mathcal{O}_{K,S}^\times$ verifying

$$\left(c_{K,S}'\right)^{-2} \max_i \|\varepsilon^{-1}x_i\|_{v_0} \leq \left(c_{K,S}'\right)^{-1} t \leq \max_i \|\varepsilon^{-1}x_i'\|_v \leq c_{K,S}' t \leq \left(c_{K,S}'\right)^2 \max_i \|\varepsilon^{-1}x_i'\|_{v_0}, \text{ for all } v \in S. \tag{6.2}$$

In particular, if $h := \max_i \|\varepsilon^{-1}x_i'\|_{v_0}$, from (6.2) we deduce

$$H(\mathbf{x}) = \frac{1}{\mathcal{N}_K(\mathfrak{a}_j)^{\frac{1}{d_K}}} \prod_{v \in S} \max_i \|\varepsilon^{-1}x_i'\|_v \gtrsim_{K,S} \prod_{v \in S} \max_i \{\|\varepsilon^{-1}x_i'\|_{v_0}\} \gtrsim_{K,S} h^{|S|}. \tag{6.3}$$

Now, note that (6.2) and (6.3) imply that for all $v \in S$

$$\|\varepsilon^{-1}x_j'\|_v \leq \max_w \max_i \|\varepsilon^{-1}x_i'\|_w \lesssim_{K,S} h \lesssim_{K,S} H(\mathbf{x})^{\frac{1}{|S|}}.$$

Thus,

$$H(1 : \varepsilon^{-1}x_0' : \ldots : \varepsilon^{-1}x_d') \lesssim_{K,S} H(\mathbf{x}).$$

Taking $y_j := \varepsilon^{-1}x_j'$, we conclude that the coordinates $(y_0, \ldots, y_d)$ satisfy the conclusion of Proposition 2.1. □

Proof of Lemma 6.1. Note that equation (6.1) is equivalent to prove that there exist $t > 0$ and an unit ε such that

$$\log(t) + \log(\|\varepsilon\|_v) = \log(h_v) + O_K(1) \text{ for all } v \in S. \tag{6.4}$$

Denote $(1)_{v \in S}$ for the vector in $\mathbb{R}^{|S|}$ with coordinates all equal to 1 and let W be the $\mathbb{Z}$-module given by

$$W := \left\langle \{(\log(\|\varepsilon\|_v))_{v \in S} : \varepsilon \in \mathcal{O}_{K,S}^\times\}, (1)_{v \in S} \right\rangle_{\mathbb{Z}} \subseteq \mathbb{R}^{|S|}.$$

Denoting by $\|\cdot\|$ the l^∞-norm in $\mathbb{R}^{|S|}$, we see that in order to prove (6.4) it is enough to find a positive constant C_W such that for any $\mathbf{x} \in \mathbb{R}^{|S|}$ there exists $\mathbf{w} \in W$ satisfying

$$\|\mathbf{x} - \mathbf{w}\| \leq C_W. \tag{6.5}$$

Moreover, we may take $c'_{K,S} = \exp(C_W)$. In general, it is easy to see that if $W \subseteq \mathbb{R}^{|S|}$ is an additive subgroup satisfying that $\mathbb{R}^{|S|}/W$ is compact, and Ω is a bounded set containing a representative of each class of $\mathbb{R}^{|S|}/W$, then (6.5) is verified with $C_W = \sup_{\mathbf{y}\in\Omega} \|\mathbf{y}\|$. In our case, the subgroup

$$W_1 := \left\langle \{ (\log(\|\varepsilon\|_v)_{v\in S}) : \varepsilon \in \mathcal{O}_{K,S}^{\times} \} \right\rangle \subseteq \mathbb{R}^{|S|}$$

defines a lattice of rank $|S|-1$ in the hyperplane $\{(x_v)_{v\in S} \in \mathbb{R}^{|S|} : \sum_{v\in S} x_v = 0\}$. Furthermore (see [TV91, Proposition 5.4.7 (b),Theorem 5.4.9 (b)]) it holds that the volume of W_1 satisfies the bound:

$$\det(W_1) \leq \begin{cases} |S|^{\frac{1}{2}} R_K h_K \prod\limits_{v\in M_{K,\mathrm{fin}}\cap S} \log(\mathcal{N}(\mathfrak{p}_v)) \text{ if } K \text{ is a number field,} \\ |S|^{\frac{1}{2}} \left(1 + q + \dfrac{|X(\mathbb{F}_q)| - q - 1}{g_K}\right)^{g_K} \text{ if } K \text{ is the function field of a curve } X \text{ over } \mathbb{F}_q. \end{cases} \tag{6.6}$$

In (6.6), as usually, R_K and h_K denote respectively the regulator and the class number of K, and $\mathfrak{p}_v$ is the prime corresponding to the finite place v. Meanwhile, when K is a function field, g_K denotes the genus of the curve X.

Using (6.6), it can be shown that there exists a fundamental system of units $\{\varepsilon_1, \ldots, \varepsilon_{|S|-1}\}$ such that

$$\prod_{i=1}^{|S|-1} \log(H(\varepsilon_i)) \lesssim_{|S|,d_K} \det(W_1), \tag{6.7}$$

where the implicit constant is effective. A proof of (6.7) can be found in [EG15, Proposition 4.3.9 (i)] for number fields, but the same proof carries over to function fields. On the other hand, given non-zero elements $z_1, \ldots, z_m \in \mathcal{O}_{K,S}$ which are multiplicatively independent, there exist well known effective lower bounds of the form $\prod_{i=1}^m \log(H(z_i)) \gtrsim_{d_K,|S|} 1$. Indeed, for function fields this is obvious; for number fields, we can use Dobrowolski's theorem (see, for instance, [BG06, Theorem 4.4.1.]) or stronger estimates, such as the one given in [LM04, Corollary 3.1]. Thus, from (6.7) we deduce

$$\max_{1\leq i\leq |S|-1} \log(H(\varepsilon_i)) \lesssim_{|S|,d_K} \det(W_1). \tag{6.8}$$

From (6.8) we conclude that we may find a fundamental domain $\Omega_{W_1} \subseteq \mathbb{R}^{|S|}$ such that $\max_{\mathbf{y}\in\Omega_{W_1}} \|\mathbf{y}\| \lesssim_{|S|,d_K} \det(W_1)$. Moreover, we deduce that there exists a fundamental domain $\Omega_W \subseteq \mathbb{R}^{|S|}$ with $\max_{\mathbf{y}\in\Omega_W} \|\mathbf{y}\| \lesssim_{|S|,d_K} \det(W_1)$ and thus, the constant C_W is effective, and can be taken to be of the form $O_{|S|,d_K}(\det(W_1))$. Thus we may take $C'_{K,S} = \exp(C_W)$. □

Now we prove Proposition 2.2 for function fields. Our argument essentially follows the number field case as in [Lan83, Theorem 5.2], replacing the argument of geometry of numbers by estimates of the dimension of some spaces of divisors.

Proof of Proposition 2.2 for function fields. Assume that K is a function field with field of constants $\mathbb{F}_q$. We have a map

$$\varphi : [N]_{\mathcal{O}_{K,S}} \to \{D \in \mathrm{Div}(K) : D \geq 0, \deg(D) \leq \log_q(N)\} \times \{D \in \mathrm{Div}(K) : D = \sum_{v\in S} a_v \cdot v, |a_v| \leq \log_q(N)\},$$

$$\varphi(x) := \left(\sum_{v\notin S} \mathrm{ord}_v(x) \cdot v, \sum_{v\in S} \mathrm{ord}_v(x) \cdot v \right)$$

Note that, modulo constants, the map $x \mapsto \mathrm{div}(x)$ is injective. Thus, φ has fibers with q elements. It is clear that

$$\left|\left\{D \in \mathrm{Div}(K) : D = \sum_{v\in S} a_v \cdot v : |a_v| \leq \log_q(N)\right\}\right| \leq (\log(N))^{|S|}. \tag{6.9}$$

Meanwhile, by [Ros02, Lemma 5.6], the number of divisor clases of degree zero, $h := h_K$, is finite. By [Ros02, Lemma 5.8], for every integer n, there are h divisor classes of degree n. Suppose that $n \geq 0$ and that $\{\overline{A}_1, \ldots, \overline{A}_h\}$ are the divisor classes of degree n. Then the number of effective divisors of degree n, b_n, is given by $\sum_{i=1}^h \frac{q^{l(\overline{A}_i)}-1}{q-1}$, where $l(\overline{A}_i)$ is the dimension of the Riemann-Roch space associated to $\overline{A}_i$. By [Ros02, Exercise 18], $l(\overline{A}_i) \leq \deg(\overline{A}_i) + 1 = n + 1$, hence

$$b_n = \sum_{i=1}^h \frac{q^{l(\overline{A}_i)}-1}{q-1} \leq h\frac{q^{n+1}-1}{q-1} \leq 2hq^n.$$

Thus:

$$\left|\left\{D \in \mathrm{Div}(K) : D \geq 0, \deg(D) \leq \log_q(N)\right\}\right| = \sum_{i\leq \log_q(N)} b_i \leq 2h \sum_{i\leq \log_q(N)} q^i \leq 2hN. \tag{6.10}$$

By (6.9) and (6.10) we conclude

$$|[N]_{\mathcal{O}_{K,S}}| \leq 2qhN(\log(N))^{|S|}.$$

Note that by [Ros02, Proposition 5.11], $h \leq (\sqrt{q}+1)^{2g}$, where g is the genus of K. Thus,

$$|[N]_{\mathcal{O}_{K,S}}| \lesssim_{q,g} N(\log(N))^{|S|}.$$

□

Acknowledgments. J. M. Menconi and M. Paredes were supported in part by CONICET Doctoral Fellowships. R. Sasyk was partially supported by the grant PIP-CONICET 11220130100073CO. Some of this work was carried out while the second author was a visitor at the Institut Henri Poincaré participating in the Trimester Program "Reinventing Rational Points". This visit was supported by a CIMPA-Carmin grant. He wishes to thank the organizers, the hospitality of the Institut and the financial support he received there.

References

[Bar15] F. Barroero. Algebraic S-integers of fixed degree and bounded height. *Acta Arith.*, 167(1):67–90, 2015.
[BG06] E. Bombieri and W. Gubler. *Heights in Diophantine geometry*, volume 4 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2006.
[BP89] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.
[BV83] E. Bombieri and J. Vaaler. On Siegel's lemma. *Invent. Math.*, 73(1):11–32, 1983.
[CL07] E. Croot and V. F. Lev. Open problems in additive combinatorics. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 207–233. Amer. Math. Soc., Providence, RI, 2007.
[EG15] J. Evertse and K. Győry. *Unit equations in Diophantine number theory*, volume 146 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2015.
[Fuk10] L. Fukshansky. Algebraic points of small height missing a union of varieties. *J. Number Theory*, 130(10):2099–2118, 2010.
[Gal71] P. X. Gallagher. A larger sieve. *Acta Arith.*, 18:77–81, 1971.
[Gre08] B. Green. On a variant of the large sieve. *arXiv e-prints*, page arXiv:0807.5037, Jul 2008.
[HS00] M. Hindry and J. Silverman. *Diophantine geometry*, volume 201 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000. An introduction.
[HV09] H. A. Helfgott and A. Venkatesh. How small must ill-distributed sets be? In *Analytic number theory*, pages 224–234. Cambridge Univ. Press, Cambridge, 2009.
[IK04] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
[KPS01] T. Krick, L. M. Pardo, and M. Sombra. Sharp estimates for the arithmetic Nullstellensatz. *Duke Math. J.*, 109(3):521–598, 2001.

[Lan83] S. Lang. *Fundamentals of Diophantine geometry*. Springer-Verlag, New York, 1983.

[LM04] T. Loher and D. Masser. Uniformly counting points of bounded height. *Acta Arith.*, 111(3):277–297, 2004.

[Lor96] D. Lorenzini. *An invitation to arithmetic geometry*, volume 9 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[Mar10] E. Marchionna, editor. *Questions on algebraic varieties*, volume 51 of *Centro Internazionale Matematico Estivo (C.I.M.E.) Summer Schools*. Springer, Heidelberg; Fondazione C.I.M.E., Florence, 2010. Lectures from the Centro Internazionale Matematico Estivo (C.I.M.E.) Summer School held in Varenna, September 7–17, 1969, Reprint of the 1970 original [MR0271105].

[Mon68] H. L. Montgomery. A note on the large sieve. *J. London Math. Soc.*, 43:93–98, 1968.

[Mum70] D. Mumford. Varieties defined by quadratic equations. In *Questions on Algebraic Varieties (C.I.M.E., III Ciclo, Varenna, 1969)*, pages 29–100. Edizioni Cremonese, Rome, 1970.

[Par19] M. Paredes. Ill-distributed sets over global fields and exceptional sets in Diophantine Geometry. *arXiv e-prints*, page arXiv:1901.00562, Jan 2019.

[Ros02] M. Rosen. *Number theory in function fields*, volume 210 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[Sed17] A. Sedunova. On the Bombieri-Pila method over function fields. *Acta Arith.*, 181(4):321–331, 2017.

[Ser89] J. P. Serre. *Lectures on the Mordell-Weil theorem*. Aspects of Mathematics, E15. Friedr. Vieweg & Sohn, Braunschweig, 1989. Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt.

[Sha74] I. R. Shafarevich. *Basic algebraic geometry*. Springer-Verlag, New York-Heidelberg, 1974. Translated from the Russian by K. A. Hirsch, Die Grundlehren der mathematischen Wissenschaften, Band 213.

[Thu93] J. L. Thunder. Asymptotic estimates for rational points of bounded height on flag varieties. *Compositio Math.*, 88(2):155–186, 1993.

[Thu95] J. L. Thunder. Siegel's lemma for function fields. *Michigan Math. J.*, 42(1):147–162, 1995.

[TV91] M. A. Tsfasman and S. G. Vlăduţ. *Algebraic-geometric codes*, volume 58 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1991. Translated from the Russian by the authors.

[TV10] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Paperback edition [of MR2289012].

[Wal12] M. Walsh. The inverse sieve problem in high dimensions. *Duke Math. J.*, 161(10):2001–2022, 2012.

[1]Instituto Argentino de Matemáticas Alberto P. Calderón-CONICET, Saavedra 15, Piso 3 (1083), Buenos Aires, Argentina;

[2]Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina.
E-mail address: jmenconi@dm.uba.ar
E-mail address: mparedes@dm.uba.ar
E-mail address: rsasyk@dm.uba.ar