JOURNAL DE THÉORIE DES NOMBRES DE BORDEAUX

MICHAEL NAKAMAYE

Diophantine approximation on algebraic varieties

Journal de Théorie des Nombres de Bordeaux, tome 11, nº 2 (1999), p. 439-502

<http://www.numdam.org/item?id=JTNB_1999__11_2_439_0>

Journal de Théorie des Nombres
de Bordeaux **11** (1999), 439–502

Diophantine Approximation on Algebraic Varieties

par MICHAEL NAKAMAYE

RÉSUMÉ. Nous donnons un aperçu de progrès récents en théorie de l'approximation diophantienne. Le point de départ étant le théorème de Roth, nous nous intéressons d'abord à la conjecture de Mordell, puis ensuite à des résultats analogues en dimension supérieure, résultats dûs à Faltings-Wustholz et à Faltings.

ABSTRACT. We present an overview of recent advances in diophantine approximation. Beginning with Roth's theorem, we discuss the Mordell conjecture and then pass on to recent higher dimensional results due to Faltings-Wustholz and to Faltings respectively.

0. INTRODUCTION

The theory of diophantine approximation often begins with a finite collection of polynomials $f_1(X), ..., f_n(X) \in \mathbf{Q}[X_1, \ldots, X_m]$ with rational coefficients. There are then two distinct types of questions commonly asked. First, one can look for rational points $(p_1/q_1, \cdots, p_m/q_m)$ which lie on the common set of zeroes of the polynomials, i.e. $f_i(p_1/q_1, \cdots, p_m/q_m) = 0$ for all i. This line of inquiry leads to the Mordell conjecture (which deals with the case of a single polynomial in two variables) and higher dimensional generalizations due to Faltings and Vojta. Alternatively, one can look for rational points $(p_1/q_1, \cdots, p_m/q_m)$ which are 'approximate solutions' or in other words lie very close the set of common zeroes of the f_i's. This question is perhaps older than the first and is aptly called diophantine approximation because one is looking not for solutions to diophantine equations but for approximate solutions. In this direction, the most striking results are Roth's theorem, dealing with the case of a single irreducible polynomial in one variable, and the Schmidt subspace theorem which allows for several variables but deals with very specific (linear) polynomials.

One of the simplest results in diophantine approximation is Liouville's theorem. For this, we consider a single irreducible polynomial $f(X) \in \mathbf{Q}[X]$

Manuscrit reçu le 9 octobre 1998.

of degree $d \geq 2$ with a root $\alpha \in \mathbf{R}$. Since f is irreducible, $\alpha \notin \mathbf{Q}$ and it makes sense to approximate this zero of $f(X)$ by rational numbers. Liouville's theorem states that for *any* rational p/q one always has an inequality of the form

$$|p/q - \alpha| > c(\alpha)/q^d \tag{0.1}$$

where $c(\alpha)$ is an explicity determined constant depending only on α. It turns out, however, that the exponent d in the denominator is not best possible except for quadratic $f(X)$. In fact, Roth proved that 0.1 can be improved by replacing d with $2 + \epsilon$ for any $\epsilon > 0$. The disadvantage to Roth's theorem is that the constant $c(\alpha)$ is no longer explicitly determined.

Roth's theorem was generalized fifteen years later by Schmidt who deals with the case of n independent linear forms $L_i(X) \in k[X_1, \dots, X_n]$ where $\mathbf{Q}$ has now been replaced by a finite extension $k/\mathbf{Q}$. The Schmidt subspace theorem (see §5 for an explicit statement) states that the set of rational points $(p_1/q_1, \dots, p_n/q_n)$ which are close to the n linear subspaces $L_i(X) = 0$ are degenerate; here degenerate means contained in a finite union of proper linear subspaces. In the one variable case, choosing $L(X) = X - \alpha$, one recovers Roth's theorem since a degenerate linear subspace of $\mathbf{Q}$ is just a point. Schmidt's subspace theorem has recently been extended by Faltings and Wüstholz to deal with the case when the L_i are no longer linear.

Returning to the first question posed in diophantine approximation we take the subvariety $X \subset \mathbf{P}^n$ defined over $\mathbf{Q}$ as the common zero locus of our collection of polynomials. As noted above, in this case one is not 'approximating' a geometric object with rational points but rather looking for actual rational solutions to the system of equations defining X. The connection between the problem of finding actual as opposed to approximate solutions to polynomial equations was developed by Vojta in his proof of the Mordell conjecture [V2]. The Mordell conjecture states that if X is a smooth projective curve of genus at least 2, defined over a number field k, then X has only *finitely many* rational points. Like Roth's theorem, the result is ineffective in that it does not bound the size of the solutions. Building upon Vojta's ideas, Faltings [F1, F2] was able to obtain similar results for higher dimensional varieties under certain restrictions.

The goal of these notes is to give an introduction to the ideas and arguments occurring in diophantine approximation, beginning with the simplest result, Liouville's theorem, and culminating with the difficult results of Faltings and Vojta on the one hand and Faltings–Wüstholz on the other. The rough structure of these notes is as follows: the first section begins with Liouville's theorem and procedes to analyse the difficulties encountered when trying to sharpen Liouville's result to obtain Roth's theorem. The second section then deals with these difficulties, concentrating on the geometry behind Roth's theorem, specifically the product theorem and Dyson's lemma. The third lecture moves from Roth's theorem to Vojta's proof of the Mordell conjecture, developing the necessary theory of height functions and the corresponding metrics on line bundles along the way. We will once more emphasize the geometric aspects of the proof and give a rough idea of the difficult arithmetic machinery involved. Lectures four and five will be devoted to higher dimensional problems, one to higher dimensional Mordell type theorems and one to the Schmidt subspace theorem respectively. The proofs presented here will not, in general, be complete but we hope to emphasize all of the important points so that the interested reader can then consult the literature for rigorous proofs and hopefully the account here will facilitate this process.

There are several expositions of the material covered in these lectures. For Roth's theorem and the Schmidt subspace theorem one can consult Schmidt's two collections of lecture notes [S1, S2]. For Dyson's lemma, in addition to Dyson's original paper [D], one can consult [B1, EV, N1, N3]. Vojta has two proofs of the Mordell conjecture, [V2, V3], the second of which is closer to the exposition given here. He also has a beautiful paper [V4] giving a full proof of the higher dimensional results of Faltings [F1, F2]. For surveys of the material behind Faltings' work, one can consult [H] for a beautiful overview or [EE] for a more thorough treatment.

These notes form the basis of a series of five lectures delivered at the Isaac Newton Institute from March 23 through March 27, 1998. It is a pleasure to thank the organizers of the workshop, Christophe Soulé, Jean-Louis Colliot-Thélène, and Jan Nekovář, for inviting me to present this material. I would also like to thank all of those who attended the lectures and whose comments have helped to remove many errors from these notes as well as to add better and more complete explanations. I also thank those who attended courses I gave on part of this material at Harvard in the spring of 1995 and at Bayreuth in the winter of 1996: it is only after trying to teach the material several times that I have reached a full appreciation of the subtleties involved. The spirit of these notes is hopefully the same as that of the lectures, namely informal but aiming to give a reasonably complete picture of the difficult techniques involved in these questions of arithmetic geometry.

1. From Liouville to Roth

We begin by recalling the quick and elegant proof of Liouville's theorem.

Theorem 1.1 (Liouville). *Suppose $\alpha \in \mathbf{R}$ is an algebraic irrational number of degree d over $\mathbf{Q}$. Then there exists an effectively computable constant*

$c(\alpha)$ such that for all $p/q \in \mathbf{Q}$

$$|p/q - \alpha| > c(\alpha)/q^d.$$

The proof can be formally divided into three stages which will reappear in all of our arguments but which are particularly transparent in this case. We begin with the irreducible polynomial $f(X) \in \mathbf{Q}[X]$ for α over $\mathbf{Q}$. To specify f uniquely one can take $f \in \mathbf{Z}[X]$ with relatively prime coefficients. The outline of the argument is as follows:

Step 1: $f(p/q) \neq 0$ for any $p/q \in \mathbf{Q}$ since otherwise f would not be irreducible over $\mathbf{Q}$.

Step 2: $|f(p/q)| \geq 1/q^d$ since q^d is a common denominator for the terms in $f(p/q)$.

Step 3: $|f(p/q)| \leq b(\alpha)|p/q - \alpha|$ for an explicit constant $b(\alpha)$.

Liouville's theorem follows, with $c(\alpha) = 1/2b(\alpha)$, by comparing the bounds in Steps 2 and 3. Only the upper bound in Step 3 requires further comment. Suppose we take the Taylor series expansion of $f(X)$ about α. Since $f(\alpha) = 0$ the first term is zero:

$$f(X) = \sum_{i=0}^{d} a_i(X - \alpha)^i = \sum_{i=1}^{d} a_i(X - \alpha)^i.$$

Thus

$$|f(X)| \leq |X - \alpha| \sum_{i=1}^{d} |a_i||X - \alpha|^{i-1}.$$

For $|p/q - \alpha| \leq 1$ we obtain

$$|f(p/q)| \leq |p/q - \alpha| \sum_{i=1}^{d} |a_i|.$$

This establishes Step 3, provided $|p/q - \alpha| \leq 1$, with $b(\alpha) = \sum_{i=1}^{d} |a_i|$ and proves Liouville's theorem (taking $c(\alpha) = \min\{1, 1/2b(\alpha)\}$). As for effectivity, the numbers a_i depend only on α as they are the coefficients of the Taylor series expansion of $f(X)$ about α. Consequently, $b(\alpha)$ and $c(\alpha)$ depend only on α.

Suppose now that one tries to improve the exponent d in Liouville's theorem. One idea would be to run through the same argument with a different auxiliarly polynomial $f(X)$. Of course this risks losing effectivity because $f(X)$ was *determined uniquely* by α. Nonetheless, one could try. A review of the argument reveals that the exponent d in the theorem comes by taking the quotient of $\deg f(X)$ by $\mathrm{mult}_\alpha(f(X))$. Thus if we replace $f(X)$ with a different polynomial $g(X)$ which also vanishes at α we will get the same result with an exponent of

$$\deg g(X)/\mathrm{mult}_\alpha(g(X)).$$

But for Step 2 to work, we need to have $g(X) \in \mathbf{Q}[X]$ and thus $g(X)$ will be divisible by $f(X)^{\mathrm{mult}_\alpha(g(X))}$. Thus we always will have

$\deg g(X)/\mathrm{mult}_\alpha(g(X)) \geq d$

and there is no improvement.

Thus some new idea is needed in order to improve the exponent d of Liouville's theorem. At the beginning of the century, Thue had the idea of running the same argument with an auxiliary polynomial in *two* variables. Thus in order to bound how close a single rational number can be to α, Thue considers a pair of rational numbers $p_1/q_1, p_2/q_2$, both of which are close to α. The reason why an extra approximating rational point helps is that now the auxiliary polynomial will be in two variables and this gives additional degrees of freedom allowing for an improvement in the important ratio

$$\frac{\deg f(X,Y)}{\mathrm{mult}_{(\alpha,\alpha)} f(X,Y)}$$

which occurs as the exponent in Liouville's theorem. There are of course immediate complications in the argument, particularly with Step 1 as with two variables there will not be a well–defined choice of auxiliary polynomial with big multiplicity at (α, α). But without a canonical choice of $f(X,Y)$ there is no longer any reason why $f(p_1/q_1, p_2/q_2)$ should be non–zero, without which the argument runs aground. Thus the simplest step in the proof of Liouville's theorem, Step 1, becomes the most difficult when one considers an auxiliary function in more than one variable.

Of course, once one is willing to consider $f(X,Y)$ there is no reason to stop there and Roth was finally able to obtain the best result possible by considering an auxiliary polynomial with an arbitrary number of variables:

Theorem 1.2 (Roth's Theorem). *Suppose $\alpha \in \mathbf{R}$ is algebraic and irrational. Then for any $\epsilon > 0$ there are only finitely many solutions to*

$$|p/q - \alpha| \leq \frac{1}{q^{2+\epsilon}}. \tag{1.3}$$

Note that the formulation of Roth's theorem is slightly different from that of Liouville's theorem. It follows from the fact that 1.3 has only finitely many solutions that there exists a constant $c(\alpha)$ such that

$$|p/q - \alpha| > \frac{c(\alpha)}{q^{2+\epsilon}}, \text{ for all } p/q \in \mathbf{Q}.$$

The problem is that the constant $c(\alpha)$ is not effective and this is why Roth's theorem is formulated in this alternative style.

The strategy for proving Roth's theorem is identical to that of Liouville's theorem except that we will use *several* good approximating points. So we assume that Roth's theorem is false and this gives a sequence of good approximating points

$$\left|\frac{p_i}{q_i} - \alpha\right| \leq \frac{1}{q_i^{2+\epsilon}}, \quad i \geq 1. \tag{1.4}$$

The strategy for the proof will be to choose m of these good approximating points, where $m \to \infty$ as $\epsilon \to 0$, and construct an auxiliary polynomial in m variables with large order vanishing at $(\alpha, \ldots, \alpha)$. Steps 2 and 3 of Liouville's theorem will then tell us that the approximations are too good, giving a contradiction. In order for Steps 2 and 3 to yield a contradiction, one needs to choose the appropriate notion for the order of vanishing of our polynomial $f(X_1, \ldots, X_m)$ at $(\alpha, \ldots, \alpha)$. We will see as we go through with the proof that the following, though at first awkward, is the apt definition:

Definition 1.5. Let $0 \neq f(X_1, \ldots, X_m) \in \mathbf{C}[X_1, \ldots, X_m]$ and let $(\alpha_1, \ldots, \alpha_m) \in \mathbf{C}^m$. Consider the Taylor series expansion of f about $(\alpha_1, \ldots, \alpha_m)$:

$$f(X_1, \ldots, X_m) = \sum_{I \geq 0} b_I (X_1 - \alpha_1)^{i_1} \cdot \ldots \cdot (X_m - \alpha_m)^{i_m}, \quad I = (i_1, \ldots, i_m).$$

Suppose f has multi–degree $(d_1, \ldots, d_m)$. The *index* of f at $(\alpha_1, \ldots, \alpha_m)$ is defined as follows:

$$\mathrm{ind}_{(\alpha_1, \ldots, \alpha_m)}(f) = \min \left\{ \sum_{j=1}^{m} \frac{i_j}{d_j} \,\middle|\, b_I \neq 0 \right\}.$$

The index of a polynomial $f(X_1, \ldots, X_m)$ at a point $x \in \mathbf{C}^m$ is a weighted multiplicity. For example, if $d_1 = \ldots = d_m = d$ then

$$\mathrm{ind}_x(f) = \frac{\mathrm{mult}_x(f)}{d}.$$

In general, each variable is weighted so that it can contribute a maximum of 1 to the index.

To see the relevance of the notion of index, we run through the three step proof of Roth's theorem which we would like to model after Liouville's theorem: we construct an auxiliary polynomial $f(X_1, \ldots, X_m)$ of multi-degree $(d_1, \ldots, d_m)$ with large index at the point $(\alpha, \ldots, \alpha)$. Then the proof should procede as follows:

Step 1: Show that $f(p_1/q_1, \ldots, p_m/q_m) \neq 0$.
Step 2: $|f(p_1/q_1, \ldots, p_m/q_m)| \geq 1/\prod q_i^{d_i}$ as in Liouville's theorem.
Step 3: $|f(p_1/q_1, \ldots, p_m/q_m)|$ is small since $|p_i/q_i - \alpha|$ is small and f has large index at $(\alpha, \ldots, \alpha)$.

As it stands, much work is needed in order to make this into a proof. Only Step 2 requires no further justification. We begin by explaining the upper bound for $|f(p_1/q_1, \ldots, p_m/q_m)|$ in Step 3 as this is what motivates the notion of index. As we did in Liouville's theorem, consider the Taylor series expansion of f about $(\alpha, \ldots, \alpha)$:

$$f(X_1, \ldots, X_m) = \sum_J a_J (X_1 - \alpha)^{j_1} \cdots (X_m - \alpha)^{j_m}. \tag{1.6}$$

Using 1.4 will give the following upper bound

$$\begin{aligned} &|f(p_1/q_1, \ldots, p_m/q_m)| \\ \leq \quad & C \cdot \max\left\{ \left|(p_1/q_1 - \alpha)^{j_1} \ldots (p_m/q_m - \alpha)^{j_m}\right| : \ a_J \neq 0 \right\} \\ \leq \quad & C \cdot \max\left\{ \frac{1}{\prod_{i=1}^m q_i^{(2+\epsilon)j_i}} : a_J \neq 0 \right\} \end{aligned} \tag{1.7}$$

for some constant C which depends on the number of terms in 1.6 and the size of the coefficients of P. We see, ignoring the constant C for the moment, that 1.7 contradicts the lower bound of Step 2 provided provided $a_J = 0$ whenever

$$\frac{1}{\prod_{i=1}^m q_i^{(2+\epsilon)j_i}} \geq \frac{1}{\prod_{i=1}^m q_i^{d_i}}. \tag{1.8}$$

Since we do not know the relative sizes of the q_i, the only reasonable way to guarantee inequality 1.8 is by choosing d_i so that $q_i^{d_i}$ are all roughly proportional, i.e. we need to choose

$$d_i \sim d/\log q_i \quad \text{for some } d \gg 0. \tag{1.9}$$

Moreover, with this choice of d_i inequality 1.8 translates into the following: we want $a_J = 0$ whenever

$$(2+\epsilon) \sum_{i=1}^m \frac{j_i}{d_i} \leq m. \tag{1.10}$$

And here the index has finally reappeared as a natural consequence of the argument: indeed, 1.10 says that we want

$$\operatorname{ind}_{(\alpha, \ldots, \alpha)} f > m/(2+\epsilon).$$

To summarize, the lower and upper bounds for $|f(p_1/q_1, \ldots, p_m/q_m)|$ given in Steps 2 and 3 respectively contradict one another, assuming we can suitably bound the constant C in 1.7, provided that $\operatorname{ind}_{(\alpha, \ldots, \alpha)} f > m/(2+\epsilon)$.

At this point there are three technical issues to deal with in order to make this sketch a rigorous proof

Problem 1: Show that $f(p_1/q_1, \dots, p_m/q_m) \neq 0$.
Problem 2: Bound the constant C occurring in 1.7.
Problem 3: Show that one can find $0 \neq f$ with index $> m/(2+\epsilon)$ at $(\alpha, \dots, \alpha)$.

We will deal with the three problems in reverse order. Problem 3 is essentially a counting problem. We want to kill leading terms a_J in the Taylor series expansion 1.6 provided J satisfies 1.10. In order to guarantee that one can force all of these a_J to vanish, it suffices to show that the number of m-tuples $(j_1, \dots, j_m)$ with $0 \leq j_i \leq d_i$ satisfying 1.10 is a *small fraction* of the total number of monomials; for if this is true then a dimension count shows that there are lots of polynomials, with coefficients in a finite extension $K/\mathbf{Q}$, having index at least $m/(2+\epsilon)$ at $(\alpha, \dots, \alpha)$ and all of its Galois conjugates. Taking the norm over $\mathbf{Q}$ of such a polynomial and clearing denominators gives the desired auxiliary polynomial. To show that all of this works in our particular setting is a special case of what Faltings and Wüstholz [FW2] refer to as the 'law of large of numbers'. Rougly speaking, this says the following: for each i, the "average" value of j_i/d_i is $1/2$ so if one randomly chooses a set $\{j_i\}_{i=1}^m$, the probability that 1.10 is violated approaches zero as m approaches infinity. For a rigorous statement and proof one can consult [FW2] Proposition 5.1. Alternatively, for the case in which we are interested, one can make an explicit computation as in [L1] pp. 170-171.

Next we deal with Problem 2. There are two separate issues here, first counting the number of terms a_J and second bounding the size of the $|a_J|$. The first of these is simple as the number of a_J is

$$\prod_{i=1}^{m}(1+d_i).$$

Bounding the size of the a_J is equivalent to bounding the size of the coefficients of the auxiliary polynomial $f(X)$ which is potentially a serious manner as f was constructed abstractly to satisfy certain vanishing conditions. The fact that the size of the coefficients of f can be controlled is a consequence of the famous Siegel lemma:

Lemma 1.11 (Siegel's Lemma). *Consider a system of m linear equations in n unknowns, with $m < n$:*

$$\begin{array}{ccccccc} a_{11}X_1 & + & \dots & + & a_{1n}X_n & = & 0 \\ \vdots & & & & & & \vdots \\ a_{m1}X_1 & + & \dots & + & a_{mn}X_n & = & 0 \end{array}$$

Suppose $a_{ij} \in \mathbf{Z}$ for all i,j and that $|a_{ij}| \leq A$ for all i,j. Then there exists a solution $(x_1, \dots, x_n) \in \mathbf{Z}^n$ to the system of equations with $|x_i| < 1 + (nA)^{m/(n-m)}$ for all i.

We apply Siegel's lemma by viewing the condition $\mathrm{ind}_{(\alpha,\dots,\alpha)}(f) > m/(2+\epsilon)$ as a system of linear equations in the *coefficients* of f, i.e. we view the coefficients of f as *variables*. Siegel's lemma cannot be applied directly in this situation because our system of linear equations will have coefficients in the number field $K = \mathbf{Q}(\alpha)$ since we are trying to impose large index at α. Choosing a basis for K as a vector space over $\mathbf{Q}$ allows one to make the reduction to Lemma 1.11 (for details, see [S2] Lemma 9A). One then obtains a polynomial $g \in \mathcal{O}_K[X_1, \dots, X_m]$ with large index along $(\alpha, \dots, \alpha)$ and its conjugates such that the coefficients have bounded size; taking the norm of g then gives the desired f.

The crucial observation to make about Siegel's lemma is that the exponent $m/(n-m)$ is small provided that the number of unknowns is a fixed multiple of the number of equations. At this point, one needs to do a lot of accounting to check that Siegel's lemma gives an auxiliary polynomial $f \in \mathbf{Z}[X_1, \dots, X_m]$ such that the constant C in 1.7 is small enough to obtain a contradiction when comparing the lower and upper bounds on $|f(p_1/q_1, \dots, p_m/q_m)|$. In practice, to get the numbers to work out one chooses m big enough to impose index $m/(2+\epsilon/2)$ at $(\alpha, \dots, \alpha)$ and the extra $\epsilon/2$ allows one to absorb the bound for C in 1.7 and still obtain a contradiction when comparing the numbers in Steps 2 and 3.

2. Interlude: Dyson's Lemma

We still have not faced the most difficult of the obstructions to proving Roth's theorem, namely the issue arising at the beginning of the argument in Step 1: how can we guarantee that

$$f(p_1/q_1, \dots, p_m/q_m) \neq 0?$$

Without this information, all of our computations to obtain a contradiction from comparing upper and lower bounds of $|f(p_1/q_1, \dots, p_m/q_m)|$ are in vain. This is by far the most difficult part of Roth's theorem as the rest of the argument essentially consists in counting. In general, there is of course no way to guarantee that $f(p_1/q_1, \dots, p_m/q_m) \neq 0$ because f has not been explicity constructed: we simply know from dimension counting that such an f exists but of course there may be several and most of them *will* vanish at the approximating point $(p_1/q_1, \dots, p_m/q_m)$. Roth originally dealt with this problem in an essentially arithmetic manner. He proved, in what is now called Roth's lemma, that provided $d_1 \gg d_2 \cdots \gg d_m$ (or, equivalently, given 1.9, $q_1 \ll q_2 \ll \cdots \ll q_m$) the polynomial $f(X)$ constructed using Siegel's lemma cannot have large order of vanishing at $(p_1/q_1, \dots, p_m/q_m)$. Taking the appropriate derivatives of f then yields a contradiction (since the number of derivatives is small, it does not affect the contradiction arrived at in Steps 2 and 3 of the argument). Roth's lemma is essentially arithmetic in nature, using the facts that f has integer coefficients and that we are interested in its order of vanishing at a rational point. For a proof, one can consult [L1] pp. 179–181.

An alternative geometric argument to establish non–vanishing of $f(p_1/q_1, \ldots, p_m/q_m)$ was developed thirty years later by Esnault and Viehweg [EV] building upon previous work of Dyson [D], Bombieri [B1], and Viola [Vi]. Esnault and Viehweg approach the problem from the following point of view. Suppose that whichever $f(X)$ we choose with large index at $(\alpha, \ldots, \alpha)$ we always find that $f(X)$ also has large index at $(p_1/q_1, \ldots, p_m/q_m)$. This says that certain linear conditions on the space of all polynomials of degree $(d_1, \ldots, d_m)$ fail to be independent. Thus if one could establish this independence then this would also yield a contradiction. The advantage to this viewpoint is that it is entirely geometric; the fact that the points in which we are interested are all algebraic becomes unimportant.

To state the powerful result which Esnault and Viehweg were able to prove, we introduce some new notation. Let

$$\mathbf{P} = \mathbf{P}^1 \times \ldots \times \mathbf{P}^1$$

be a product of m projective lines defined over k, an algebraically closed field of characteristic zero. We will assume for simplicity that $k = \mathbf{C}$, the field of complex numbers although the argument remains valid in the general case. Let $\pi_i : \mathbf{P} \to \mathbf{P}^1$ be the projection to the i^{th} factor. For positive integers $d_1, \ldots, d_m$ write $d = (d_1, \ldots, d_m)$ and

$$\mathcal{O}_{\mathbf{P}}(d) \simeq \pi_1^* \mathcal{O}_{\mathbf{P}^1}(d_1) \otimes \ldots \otimes \pi_m^* \mathcal{O}_{\mathbf{P}^1}(d_m).$$

For fixed d let $0 \neq s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. The *index* of s at a closed point ζ of $\mathbf{P}$ is defined by locally identifying s with a polynomial and applying Definition 1.5. We need to define certain volumes as in [EV] Definition 0.2:

Definition 2.1. Let $I^m = \{\xi = (\xi_1, \ldots, \xi_m) \in \mathbf{R}^m \,|\, 0 \leq \xi_i \leq 1 \text{ for all } i\}$ and let $\mathrm{Vol}(t)$ denote the volume of

$$\left\{ \xi \in I^m \,|\, \sum_{\nu=1}^{m} \xi_\nu \leq t \right\}.$$

Of course $1 - \mathrm{Vol}(t)$ measures, asymtotically in d, the proportion of sections of $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ with index $\geq t$ at a point ζ. Dyson's lemma gives conditions on a set of points $\{\zeta_i\} \subset \mathbf{P}$ and a line bundle $\mathcal{O}_{\mathbf{P}}(d)$ so that requiring index t_i at ζ_i imposes *almost* independent conditions on global sections of $\mathcal{O}_{\mathbf{P}}(d)$. The following is an alternative, slightly more transparent formulation of Esnault and Viehweg's Dyson lemma:

Theorem 2.2. *Suppose $0 \neq s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ and $\zeta_1, \ldots, \zeta_M \subset \mathbf{P}$ so that no two ζ_i are contained in a proper product subvariety, i.e. the points*

$\{\pi_j(\zeta_i)\}_{i=1}^M$ *are distinct for all* j. *Let* $t_i = \mathrm{ind}_{\zeta_i}(s)$ *and let*

$$\delta = \max\{d_{i+1}/d_i,\ 1 \le i \le m-1\}.$$

Then

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i - m\delta) \le 1.$$

Note that if $t < 0$ *then* $\mathrm{Vol}(t) = 0$.

Theorem 2.2 says that the linear conditions for imposing index $t_i - m\delta$ at ζ_i are independent inside $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. The key observation about Theorem 2.2 is that if $d_1 \gg d_2 \gg \ldots \gg d_m$ then $m\delta$ is small so in this case Theorem 2.2 says that the conditions to impose index t_i at ζ_i are nearly independent inside $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. This will then allow us to derive the desired contradiction and prove Roth's theorem, provided we can arrange for $d_1 \gg d_2 \gg \cdots \gg d_m$. But by 1.9 this is equivalent to choosing rational approximations with denominators satisfying $q_1 \ll q_2 \ll \cdots \ll q_m$ and we can achieve this since we have assumed that there are *infinitely* many good rational approximating points. It should be noted that there is still some arguing left because Dyson's lemma only shows that there exists $g(X) \in \mathbf{C}[X_1, \ldots, X_m]$ with large index at $(\alpha, \ldots, \alpha)$ and its conjugates which has small order of vanishing at $(p_1/q_1, \ldots, p_m/q_m)$. From here one can construct $f(X) \in \mathbf{Z}[X_1, \ldots, X_m]$ with large index at $(\alpha, \ldots, \alpha)$ and small index at $(p_1/q_1, \ldots, p_m/q_m)$ but a priori we have no control on the size of the coefficients of f. Thus more calculations of dimensions and of Siegel type are necessary (see [EV] §10 for details).

We now turn to the proof of Theorem 2.2. In the statement of Theorem 2.2, the volumes which occur have been perturbed by $m\delta$. The reason for this is that suitable derivatives of the non–zero section $s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ with index $\ge t_i$ at ζ_i still have index $\ge t_i - m\delta$ at ζ_i. The product theorem will tell us that the common zero locus of these sections is contained in a finite union of proper product subvarieties. Let $\mathbf{P}' \subset \mathbf{P}$ be one of these product subvarieties and let $\pi : \mathbf{P} \dashrightarrow \mathbf{P}'$ denote the natural projection. Taking $\zeta_i' = \pi(\zeta_i)$, one then shows, except in certain degenerate cases, that on $\mathbf{P}'$ one can produce a new non–zero section $s' \in H^0(\mathbf{P}', \mathcal{O}_{\mathbf{P}}(d))$ with suitable indices at the ζ_i'. This procedure is then iterated to produce a zero cycle representing $c_1(\mathcal{O}_{\mathbf{P}}(d))^m$ and the part of the class associated to ζ_i measures the cost of imposing index $t_i - m\delta$ at ζ_i. These classes have disjoint supports and this implies the desired conclusion: the cost of imposing index $t_i - m\delta$ at ζ_i for all i is the sum of the individual costs, i.e. the conditions are *independent.*

The details of the proof of Theorem 2.2 are as follows. We are given a non-zero section $s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ with $t_i = \mathrm{ind}_{\zeta_i}(s)$. Given a point $\zeta \in \mathbf{P}$, an affine open subset $U = \operatorname{Spec} A \subset \mathbf{P}$, and a local trivialization of $\mathcal{O}_{\mathbf{P}}(d)$ on U, the sections $\sigma \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ such that $\operatorname{ind}_\zeta(\sigma) \geq t$ generate an ideal $I_{\zeta,d,t} \subset A$. This ideal does not depend on the trivialization and as U varies over an affine cover of $\mathbf{P}$ this defines an ideal sheaf $\mathcal{I}_{\zeta,d,t} \subset \mathcal{O}_{\mathbf{P}}$. Let

$$Z_i = \operatorname{supp}(\mathcal{O}_{\mathbf{P}}/\mathcal{I}_{\zeta_i,d,t_i}).$$

Since we are trying to show that the various index conditions are independent, as a first approximation one should expect the supports of these ideal sheaves to be disjoint. This is not difficult to show because one can write down explicit monomial generators for $\mathcal{I}_{\zeta,d,t}$ and hence determine its support (see [EV] Lemma 2.5). It then follows readily ([EV] Lemma 2.8), as desired, that

$$Z_i \cap Z_j = \emptyset \quad \text{for } i \neq j. \tag{2.3}$$

As discussed above, we would like to be able to produce sections satisfying the relevant index conditions on proper product subvarieties. To this end, let $\mathcal{I}_d(s) = \bigcap_{j=1}^{M} \mathcal{I}_{\zeta_j,d,t_j}$ and for $\mathbf{P}' \subset \mathbf{P}$ a product subvariety let $\mathcal{O}_{\mathbf{P}'}(d) = \mathcal{O}_{\mathbf{P}}(d)|\mathbf{P}'$. If one is unlucky and $\mathbf{P}' \subset Z_i$ for some i then $H^0(\mathbf{P}', \mathcal{O}_{\mathbf{P}'}(kd) \otimes \mathcal{I}_{kd}(s)) = 0$ for all $k > 0$ so there will be no hope of producing a new section. On the other hand, [EV] 2.9 (ii) implies that whenever $\mathbf{P}' \not\subset Z_i$ for any i then for k sufficiently divisible

$$H^0\left(\mathbf{P}', \mathcal{O}_{\mathbf{P}'}(kd) \otimes \mathcal{I}_{kd}(s)\right) \neq 0. \tag{2.4}$$

There is one more crucial ingredient in the proof of Theorem 2.2 as it allows us to use the sections guaranteed by 2.4 to construct an intersection class representing $c_1(\mathcal{O}_{\mathbf{P}}(d))^m$. This is the so–called product theorem which establishes a certain degeneracy in the locus where the non–zero section $s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ can have large index. Initially inspired by the product theorem of Faltings [F1], the result presented here strengthens Theorem 3.1 of [N1].

Theorem 2.5 (Product Theorem). *Let $f \in \mathbf{C}[X_1, \dots, X_m]$ be of multi–degree $d_1, \dots, d_m$ and for an $m-1$-tuple of non-negative integers $\alpha = (\alpha_1, \dots, \alpha_{m-1})$ let $D^\alpha = \frac{\partial^{\alpha_1}}{\partial X_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_{m-1}}}{\partial X_{m-1}^{\alpha_{m-1}}}$. Let*

$$X(f) = \Big\{(x_1, \dots, x_m) \in \mathbf{C}^m \;:\; D^\alpha f(x_1, \dots, x_m) = 0, \\ \forall\, 0 \leq \alpha_i \leq \max_{i+1 \leq j \leq m} \{d_j\}\Big\}$$

Let $W \subset X(f)$ be an irreducible component. Then W is contained in a proper product subvariety.

Proof of Theorem 2.5. We will show that either $\pi_m(W)$ is a point or else $W = W' \times \mathbf{C}$ for some $W' \subset \mathbf{C}^{m-1}$. In the former case, W is contained in a proper product subvariety and we are done. In the latter case, choose a general $\lambda \in \mathbf{C}$ and replace f with

$$f_\lambda(X_1,\dots,X_{m-1}) = f(X_1,\dots,X_{m-1},\lambda).$$

Since $\partial/\partial X_1,\dots,\partial/\partial X_{m-2}$ commute with evaluation at λ we have

$$\frac{\partial^{\alpha_1}}{\partial X_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_{m-2}}}{\partial X_{m-2}^{\alpha_{m-2}}} f_\lambda\,(X_1,\dots,X_{m-1}) \tag{2.5.1}$$

$$= \frac{\partial^{\alpha_1}}{\partial X_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_{m-2}}}{\partial X_{m-2}^{\alpha_{m-2}}}\, f\,(X_1,\dots,X_{m-1},X_m)\Bigg|_{X_m=\lambda}.$$

But $\frac{\partial^{\alpha_1}}{\partial X_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_{m-2}}}{\partial X_{m-2}^{\alpha_{m-2}}}\, f\,(X_1,\dots,X_m)$ vanishes along $W'\times\mathbf{C}$ for $\alpha_i \le \max_{i+1\le j\le m}\{d_j\}$ by assumption. Hence by 2.5.1

$$\frac{\partial^{\alpha_1}}{\partial X_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_{m-2}}}{\partial X_{m-2}^{\alpha_{m-2}}}\, f_\lambda\,\Bigg|\, W' = 0 \quad \text{for } 0 \le \alpha_i \le \max_{i+1\le j\le m}\{d_j\}.$$

Thus, using the notation in the statement of Theorem 2.5, $W' \subset X(f_\lambda)$ and we can conclude by induction on m that W', and hence W, must be contained in a proper product subvariety.

We now proceed to establish that if $\pi_m(W)$ is not a point then $W = W'\times\mathbf{C}$, concluding the proof of Theorem 2.5. Choose a general point $\eta \in W$ so that

$$\eta \text{ is a smooth point of } W \tag{2.5.2}$$

$$\operatorname{mult}_\eta(f) = \operatorname{mult}_W(f) \tag{2.5.3}$$

Let $T_\eta(W)$ denote the tangent space to W at η. Since we are assuming that $\pi_m : W \to \mathbf{C}$ is surjective it follows, choosing η still more general if necessary, that we may assume the induced map on tangent spaces $T_\eta(W) \to T_{\pi_m(\eta)}(\mathbf{C})$ is surjective. Consequently, viewing $T_\eta(W)$, after translation to the origin, as a subspace of $T_0(\mathbf{C}^m)$,

$$T_\eta(W) + \langle \partial/\partial X_1,\dots,\partial/\partial X_{m-1}\rangle = T_0(\mathbf{C}^m). \tag{2.5.4}$$

We claim that for some $1 \le i \le m-1$

$$\operatorname{mult}_\eta(\partial/\partial X_i(f)) = \operatorname{mult}_\eta(f) - 1. \tag{2.5.5}$$

Observe first that if D^α is any differential operator of order $\le \operatorname{mult}_\eta(f) - 1$ and if $\partial/\partial X \in T_\eta(W)$ then $D^\alpha(\partial/\partial X(f))|_\eta = \partial/\partial X(D^\alpha(f))|_\eta = 0$ since $D^\alpha(f)$ vanishes on W and $\partial/\partial X \in T_\eta(W)$. Thus

$$\operatorname{mult}_\eta(\partial/\partial X(f)) \ge \operatorname{mult}_\eta(f), \quad \text{for all } \partial/\partial X \in T_\eta(W). \tag{2.5.6}$$

Suppose, contrary to 2.5.5, that for all $1 \le i \le m-1$ we have

$$\operatorname{mult}_\eta(\partial/\partial X_i(f)) \ge \operatorname{mult}_\eta(f).$$

By 2.5.4 and 2.5.6 we also have $\operatorname{mult}_\eta(\partial/\partial X_m(f)) \geq \operatorname{mult}_\eta(f)$. But this would imply that $\operatorname{mult}_\eta(f) \geq \operatorname{mult}_\eta(f) + 1$, a contradiction. Thus 2.5.5 holds. Applying the same argument inductively and using 2.5.3 shows that for some non-negative $i_1, \dots, i_{m-1}$

$$\frac{\partial^{i_1}}{\partial X_1^{i_1}} \cdots \frac{\partial^{i_{m-1}}}{\partial X_{m-1}^{i_{m-1}}} f(\eta) \neq 0, \text{ with } \sum_{j=1}^{m-1} i_j = \operatorname{mult}_W(f).$$

But by the definition of $X(f)$, $D^\alpha(f)|W = 0$ for any $0 \leq \alpha_i \leq d_m$. Thus $\operatorname{mult}_W(f) \geq d_m + 1$. Since f has degree d_m in X_m this is only possible if W is of the form $W' \times \mathbf{C}$. This concludes the proof of the product theorem.

Observe that if we replace d with Nd for N sufficiently divisible then we can assume without loss of generality that if $Z \not\subset Z_i$ for any i then there exists

$$(2.6) \qquad 0 \neq s_Z \in H^0\left(Z, \mathcal{O}_{\mathbf{P}}(d) \otimes \cap_{i=1}^M \mathcal{I}_{\zeta_i, d, t_i - m\delta}\right).$$

Indeed, applying the product theorem to the section s of Theorem 2.2 gives sections of $H^0\left(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d) \otimes \cap_{i=1}^M \mathcal{I}_{\zeta_i, d, t_i - m\delta}\right)$ which generate off a finite union of proper product subvarieties. If $\mathbf{P}'$ is one of these proper product subvarieties, not contained in Z_i for any i, then 2.4 gives a non–zero section

$$s_{\mathbf{P}'} \in H^0\left(\mathbf{P}', \mathcal{O}_{\mathbf{P}}(nd) \otimes \cap_{i=1}^M \mathcal{I}_{\zeta_i, nd, t_i}\right)$$

for some positive integer n. Since 2.4 only needs to be applied finitely many times in order to obtain sections on *any* subvariety $Z \not\subset Z_i$, we can choose N sufficiently divisible to give all of the sections s_Z at once.

We will use 2.6 in order to construct an effective cycle representing $c_1(\mathcal{O}_{\mathbf{P}}(d))^m$. This is done inductively as follows. With $s_{\mathbf{P}}$ as in 2.6, $Z(s_{\mathbf{P}})$ represents $c_1(\mathcal{O}_{\mathbf{P}}(d))$. Let Z be an irreducible component of $Z(s_{\mathbf{P}})$. If $Z \subset Z_i$ for some i, then choose some $0 \neq s_Z \in H^0(Z, \mathcal{O}_{\mathbf{P}}(d))$ and $Z(s_Z)$ is an effective representative for $c_1(\mathcal{O}_{\mathbf{P}}(d)) \cap Z$. Suppose next that $Z \not\subset Z_i$ for any i. Then by 2.6 there exists $0 \neq s_Z \in H^0\left(Z, \mathcal{O}_{\mathbf{P}}(d) \otimes \cap_{i=1}^M \mathcal{I}_{\zeta_i, d, t_i - m\delta}\right)$ and $Z(s_Z)$ represents $c_1(\mathcal{O}_{\mathbf{P}}(d)) \cap Z$. Repeating this procedure inductively gives an effective representative for $c_1(\mathcal{O}_{\mathbf{P}}(d))^m$. Moreover, since by 2.3 the Z_i are mutually disjoint, we can write

$$c_1(\mathcal{O}_{\mathbf{P}}(d))^m = R + \sum_{j=1}^M R_j$$

where R_j is the part of the intersection supported on Z_j. Since R is effective, the result of this construction is that

$$(2.7) \qquad \sum_{j=1}^M \deg R_j \;=\; \deg c_1(\mathcal{O}_{\mathbf{P}}(d))^m - \deg(R) \;\leq\; \deg c_1(\mathcal{O}_{\mathbf{P}}(d))^m.$$

We have $\deg c_1(\mathcal{O}_{\mathbf{P}}(d))^m = m!\prod_{i=1}^m d_i$. Thus to conclude the proof of Theorem 2.2 it suffices to prove that for $1 \le j \le M$ we have

$$\deg(R_j) \ge m!\left(\prod_{i=1}^{m} d_i\right)\mathrm{Vol}(t_j - m\delta) \tag{2.8}$$

Combining 2.8 with 2.7 yields Theorem 2.2.

A rigorous proof of 2.8 involves blowing up the ideal sheaf $\mathcal{I}_{\zeta_j,d,t_j-m\delta}$ but intuitively it has an easy explanation. Suppose $C_1, C_2 \subset \mathbf{P}^2$ are two distinct, irreducible curves meeting in a point P. Suppose moreover that both C_1 and C_2 have multiplicity m at P. Let $i_P(C_1 \cdot C_2)$ denote the part of the intersection class $C_1 \cdot C_2$ supported at P. Then one has

$$i_P(C_1 \cdot C_2) \ge m^2.$$

But of course m^2 measures precisely the cost of imposing multiplicity m at P. In the situation of 2.8, things are more complicated because the condition we are imposing will force vanishing not only at points but also along higher dimensional subvarieties. The fundamental principle, however, is the same: the part of the intersection class we have produced supported on Z_j measures by definition the cost of imposing index $t_j - m\delta$ at ζ_j. The right hand side of 2.8 *is* the cost of imposing $t_j - m\delta$ at ζ_j whence the inequality.

More precisely, suppose $\pi : Y \to \mathbf{P}$ is the blow-up of $\mathbf{P}$ along $\mathcal{I}_{\zeta_j,d,t_j-m\delta}$. The ideal sheaf $\mathcal{I}_{\zeta_j,d,t_j-m\delta}$ has been defined so that it is generated locally by global sections of $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. Hence if E denotes the exceptional divisor of π, it follows that $\pi^*\mathcal{O}_{\mathbf{P}}(d)(-E)$ is numerically effective. Lifting the construction of $c_1(\mathcal{O}_{\mathbf{P}}(d))^m$ to Y one verifies that

$$\deg R_j \ge \deg\left(c_1(\pi^*\mathcal{O}_{\mathbf{P}}(d))^m - c_1(\pi^*\mathcal{O}_{\mathbf{P}}(d)(-E))^m\right). \tag{2.9}$$

Since $\pi^*\mathcal{O}_{\mathbf{P}}(d)(-E)$ is numerically effective $c_1(\pi^*\mathcal{O}_{\mathbf{P}}(d)(-E))^m$ is determined by the growth of $h^0(\pi^*\mathcal{O}_{\mathbf{P}}(nd)(-nE))$ for large n. Pushing back down to $\mathbf{P}$, a direct computation (see [N3]) then derives 2.8 from 2.9.

We now pass to Vojta's version of Dyson's lemma on products of curves of any genus. This is a simple and beautiful theorem which initially inspired Vojta's diophantine proof of the Mordell conjecture [V2] and serves as a key ingredient in that proof. We begin by fixing some notation. Let C_1 and C_2 be smooth projective curves over an algebraically closed field of characteristic zero. Suppose L is a line bundle on $C_1 \times C_2$. Let $F_1 = P_1 \times C_2$ be a fibre of the first projection and $F_2 = C_1 \times P_2$ a fibre of the second projection for arbitrary points $P_1 \in C_1$ and $P_2 \in C_2$. Let $d_1 = L \cdot F_2$ and $d_2 = L \cdot F_1$ so d_1 and d_2 measure the 'degrees' of the line bundle L in analogy to the multi-degree of our auxiliary polynomial in Roth's theorem. In fact, the two notions of degree correspond when C_1 and C_2 are both isomorphic to the projective line. If s is a non–zero section of L it makes sense to talk about the index of s at a point $x \in C_1 \times C_2$: Definition 1.5 applies in this situation, with weights d_1, d_2, as it is a local definition. We let the genus of C_1 be g_1 and the genus of C_2 will be denoted g_2.

Theorem 2.10 (Vojta). *Suppose there is a non–zero section $s \in H^0(C_1 \times C_2, L)$. Suppose moreover that $\{\zeta_i\}_{i=1}^M$ are points in $C_1 \times C_2$ such that no two are contained in a proper product subvariety and let $\mathrm{ind}_{\zeta_i}(s) = t_i$. Write $Z(s) = \sum_{i=1}^r a_i Z_i$ where Z_i is irreducible and let*

$$e = \max\{a_i : \ Z_i \textit{ is not a fibre of either projection}\}.$$

Then

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i) \leq \frac{c_1(L)^2 + ed_2 \max\{2g_1 - 2 + M, 0\}}{2d_1 d_2}. \tag{2.11}$$

Several remarks are in order to explain the strength of Theorem 2.10. If, as will be the case in the Mordell conjecture, the genus of both C_1 and C_2 is at least 2, then $2g_1 - 2 + M \geq 0$ and one does not need to take the maximum. In these circumstances, inequality 2.11 reads

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i) \leq \frac{c_1(L)^2}{2d_1 d_2} + \frac{e(2g_1 - 2 + M)}{2d_1}. \tag{2.12}$$

Moreover, we know that $e \leq \min\{d_1, d_2\}$ since any curve Z_i which is not a fibre of either projection satisfies $Z_i.F_1 \geq 1$ and $Z_i.F_2 \geq 1$. If we assume that $d_1 \gg d_2$, as we did in Roth's theorem, this will kill the second term of 2.12 leaving us with $c_1(L)^2/2d_1d_2$. In the case we dealt with previously where both C_1 and C_2 had genus zero, we have $c_1(L)^2 = 2d_1d_2$ and this gives the 1 on the right hand side of our previous Dyson lemma. On higher genus curves, however, there are more line bundles because the diagonal inside $C \times C$ is only linearly equivalent to fibres of the two projections when the genus of C is zero. Thus if one can choose a line bundle with $c_1(L)^2 \ll 2d_1d_2$ then Vojta's version of Dyson's lemma says that *no* section of L can have big index at *any* point of $C \times C$. It is exactly in this fashion that Dyson's lemma will be applied to prove the Mordell conjecture.

The proof of Theorem 2.10 is essentially identical to that of Theorem 2.2, the only difference being that taking derivatives is slightly more complicated on a curve of larger genus: this, in fact, is what creates the $2g_1 - 2$ occurring on the right hand side of 2.11. To see the similarity between Theorem 2.2 and Theorem 2.10 we consider first the special case of Theorem 2.10 where both C_1 and C_2 are projective lines. In this case, the theorems are qualitatively identical but the perturbation term involving $\delta = d_2/d_1$

appears on the left hand side of Theorem 2.2 while it has been moved to the right hand side in Theorem 2.10.

To see how our proof of Theorem 2.2 can be adapted to this new situation, suppose we make sure, when constructing our intersection product, that the sections which we intersect all have index $\geq t_i$ at ζ_i rather than settling for index $t_i - 2\delta$ as we did. In other words, after identifying s with a polynomial $f(X,Y)$ on an affine open subset and then differentiating $f(X,Y)$, we increase the index by multiplying by some other fixed polynomial. Of course, since the goal of the proof is to induct on the dimension of product subvarieties, the polynomial one multiplies by should if possible have zeroes in a proper product subvariety, i.e. one should just add fibres of one of the two projections. Thus the simplest solution is to replace the derivatives $\partial^i/\partial X_1^i f(X_1,X_2)$ by

$$\left(\frac{\partial^i}{\partial X_1^i} f(X_1,X_2)\right) \prod_{j=1}^{M} (X_1 - \pi_1(\zeta_j))^i. \tag{2.13}$$

Indeed, the polynomials in 2.13 still have index at least t_i at α_i for all i and the possible decrease in index incurred by differentiating has been remedied. The only problem is that the degree of the polynomial has been increased. To minimize the increase, note that one of the points, say ζ_1, can be taken to be (∞,∞) with respect to the affine open patch where we differentiate. Then the derivatives do not affect the index of the projectivization of f at ζ_1. Also, since each derivative decreases the degree of f, we have some additional room to twist in order to get a section of $\mathcal{O}_{\mathbf{P}}(d)$. In particular, taking $\zeta_2 = (0,0)$, we note that $X_1^i \frac{\partial^i}{\partial X_1^i}(f)$ has degree $\leq (d_1,d_2)$ and index $\geq t_2$ at ζ_2. So with this modification 2.13 gives us polynomials of bi–degree $(d_1 + (M-2)i, d_2)$. The number of derivatives which need to be taken in order for the common zeroes to be contained in a proper product subvariety is by definition the number e because $\partial/\partial X_1$ is transverse to all non–fibral C and so these partial derivatives always decrease the order of vanishing along such components of $Z(s)$. Thus the conclusion, if we run through the argument used to prove Theorem 2.2, is that

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i) \leq \frac{c_1\left(\mathcal{O}_{\mathbf{P}}(d_1 + e\cdot\max\{M-2,0\}, d_2)\right)^2}{2d_1d_2}. \tag{2.14}$$

Expanding 2.14 yields exactly 2.11 up to a factor of two on the second term: for a discussion of how to remove this factor of 2, see [EV] §10.

We now turn to proving Vojta's version of Dyson's lemma in full generality. In order to take derivatives of a section of L on $C_1 \times C_2$ we recall the method used by Faltings ([F1] p. 560 and p. 566) in a more general setting. The basic idea is this: the reason why one cannot just naively take derivatives of a section $s \in H^0(C_1 \times C_2, L)$ is first that there are no global vector fields if $g(C_1), g(C_2) \geq 2$ and second that after identifying s locally with some polynomial and differentiating this polynomial, the derivatives no longer patch together to give a section of L. To remedy this we consider the following construction. Suppose we are given a dominant map $p_1 : C_1 \to \mathbf{P}^1$. Then one can pull back the derivation on $\mathbf{P}^1$ via p_1, with poles along the ramification divisor R of p_1 and so obtain:

Definition 2.15. Let $s \in H^0(C_1 \times C_2, L)$ and suppose $Z \subset Z(s)$. Then $\partial/\partial X_1(s)|Z$ is the global section of $L(\pi_1^*(p_1^* K_{\mathbf{P}^1} + R))|Z \simeq L(\pi_1^* K_{C_1})|Z$ obtained locally by differentiating s via the pull–back of the derivation on $\mathcal{O}_{\mathbf{P}^1}$ by the composition $p_1 \cdot \pi_1$. Higher order derivatives are defined analogously so for $\alpha > 0$

$$D^\alpha(s) \text{ is a global section of } L\,(\alpha\pi_1^* K_{C_1})|\, Z$$

provided all partial derivatives $D^\beta(s)$ with $\beta < \alpha$ vanish identically along Z.

Using Definition 2.15, one can derive Theorem 2.10 in exactly the same manner in which we proved Theorem 2.2. The only difference is that when we take e derivatives of the section s we have to twist by $e\pi_1^* K_{C_1}$. Since we also have to add one fibre for each of the points ζ_i in order to preserve the index of the derivative of s at ζ_i this means that the total twist we need in order take the necessary e derivatives is

$$e(\pi_1^* K_{C_1} + MF_1)$$

and, since $\deg(K_{C_1}) = 2g_1 - 2$ this is exactly what occurs on the right and side in 2.11. Thus arguing exactly as before one finds, for $\delta = \max\{2g - 2 + M, 0\}$

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i) \leq \frac{c_1\,(L(d_1 + e\delta, d_2))^2}{2d_1 d_2},$$

which is 2.11 up to a factor of 2 on the second term. We already saw this factor of 2 appear after 2.14 and to avoid it one makes a direct intersection theoretic construction as in [V1] or [EV] §10.

Two further remarks are necessary in order to turn this sketch into a proof. First, one needs to be careful about twisting the line bundle L because this can make it difficult to compute $\deg R_j$ in 2.8; indeed the proof is cohomological in nature and requires that the sections constructed all be sections of the same line bundle. This problem is not serious, however, as we are merely twisting by fibres of the first projection and so one can twist all bundles involved without adversely affecting the construction. The second possible problem is that on higher genus curves, the derivative of s only makes sense as a section of a line bundle on a proper subvariety, not as a section of a line bundle on $C_1 \times C_2$. This too, however, plays no crucial role in the proof of 2.14 (and its analogue on $C_1 \times C_2$). Again the proof is cohomological in nature and only depends on numerical effectivity of a line bundle on a blow up of $C_1 \times C_2$ (see [N3] for details) and for this one does not need global sections but just non–zero sections on any curve $C \subset C_1 \times C_2$.

To conclude this section, it is worth pointing out that Theorem 2.10 does not hold, as stated, on a product of three or more curves of genus ≥ 1: in other words, the data of the degrees d_i of L and the top intersection number $c_1(L)^{\text{top}}$ are *not* sufficient to bound the sum of the relevant volumes. The reason for this, thinking of the proof of Theorem 2.10, is that the only proper product subvarieties X of $C_1 \times C_2$ are points or fibres of the two projections. In both cases, the degrees d_1, d_2 determine, at least up to numerical equivalence, the line bundle $L|X$. In higher dimension this is no longer the case and we take advantage of precisely this fact in producing a higher dimensional example where the direct generalization of Theorem 2.10 fails.

Let C be a smooth projective curve of genus $g \geq 1$ and let $Y = C \times C$. Let $F_1 \subset Y$ be a fibre of the first projection, $F_2 \subset Y$ a fibre of the second projection, and $\Delta' = \Delta - F_1 - F_2$. Finally, let $n = (n_1, n_2, n_3)$ be a 3-tuple of integers. We consider divisors of the form

$$V_n = n_1 F_1 + n_2 F_2 + n_3 \Delta'.$$

Since $F_1 + F_2$ is ample and $V_n \cdot (F_1 + F_2) = n_1 + n_2$, some multiple of V_n is effective provided $n_1 + n_2 > 0$ and $V_n^2 > 0$. We have

$$c_1(V_n)^2 = 2(n_1 n_2 - g n_3^2).$$

Hence given any $\epsilon > 0$ there exist n_1, n_2, n_3 such that $n_1 \gg n_2 \gg 0$, V_n is effective, and

$$c_1(V_n)^2 \leq \epsilon n_1 n_2. \tag{2.16}$$

Let $X = C \times C \times C$ and fix a point $P \in C$. Let $p : X \to Y$ denote the projection to the first two factors and let $\pi_3 : X \to C$ denote the projection to the third factor. Define

$$\mathcal{L} = p^* \mathcal{O}_Y(V_n) \otimes \pi_3^* \mathcal{O}_C(P).$$

Choose $0 \neq s \in H^0(Y, \mathcal{O}_Y(V_n))$ and let $t \in H^0(C, \mathcal{O}_C(P))$ be the section whose divisor of zeroes is P. Then $\operatorname{ind}(\zeta, p^* s \otimes \pi_3^* t) \geq 1$ for any point $\zeta \in X$ with $\pi_3(\zeta) = P$. A computation using 2.16 shows that

$$c_1(\mathcal{L})^3 = 3c_1(p^* \mathcal{O}_Y(V_n))^2 \cap c_1(\pi_3^* \mathcal{O}_C(P)) \leq 3\epsilon n_1 n_2. \tag{2.17}$$

Thus if Theorem 2.10 held for $\mathcal{L}$ one would find (up to an $o(1)$ depending on the constants implied in $n_1 \gg n_2 \gg 0$) that

$$1/6 = \mathrm{Vol}(1) \leq \frac{c_1(\mathcal{L})^3}{6n_1n_2},$$

contradicting 2.17.

It should be noted that the particular divisor V_n which we chose above will play a central role in Vojta's proof of the Mordell conjecture to be given in the next section. As was noted above, Vojta's Dyson lemma is extremely powerful when applied to a divisor like V_n as it says that no section of $H^0(C \times C, V_n)$ can have large index at any point.

3. The Mordell Conjecture

In this section we will give a sketch of Vojta's proof [V2, V3] of the famous Mordell conjecture

Theorem 3.1 (Faltings). *Suppose C is a smooth projective curve of genus ≥ 2 defined over a number field k. Then C has at most finitely many k-rational points.*

Of course, Mordell originally conjectured Theorem 3.1 only for the case where $k = \mathbf{Q}$. This is in fact the case which we will ultimately consider, not because it is intrinsically easier than the general case but because there is less notational complication as $\mathbf{Q}$ admits only one embedding in $\mathbf{C}$.

How does one use Vojta's Theorem 2.10 to prove the Mordell conjecture? Vojta's beautiful idea is to take advantage of the freedom to choose the line bundle L in Theorem 2.10. If C is a curve of genus at least one then there are more line bundles on $C \times C$ than the pull–backs of bundles from the two projections because the diagonal is not linearly equivalent to a sum of fibres. This additional parameter of freedom allows for the choice of a line bundle L (given in fact by the divisor V_n at the end of the last section) such that

$$c_1(L)^2 \ll d_1d_2.$$

As was noted after the statement of Theorem 2.10, if we also arrange for $d_1 \gg d_2$ then 2.12 tells us that *no section* $s \in H^0(C \times C, L)$ can have large index at *any* point. Thus all of the hard work we spent in Roth's theorem trying to show that the auxiliary polynomial $P(X_1, \ldots, X_m)$ has small index at $(p_1/q_1, \ldots, p_m/q_m)$ is accomplished automatically, *regardless* of our choice of section s.

This sounds too good to be true and indeed it is. It will turn out that what was easy in the case of Roth's theorem (namely to produce an arithmetic reason why P must vanish at the rational approximating point) will be extremely difficult in Mordell's conjecture and conversely the difficult part of Roth's theorem has been rendered easy by Vojta's beautiful translation of Dyson's lemma to higher genus curves.

Why is it so difficult to force a section $s \in H^0(C \times C, L)$ to vanish at a rational point $(x, y) \in C \times C$? One reason is that unlike Roth's theorem which took place on $\mathbf{P}^1 \times \cdots \times \mathbf{P}^1$ where there are explicit coordinates, we are now on a curve of higher genus and instead of constructing a polynomial we are trying to construct a section of a line bundle. In addition to the lack of coordinates there is also no way to evaluate a section at a point without fixing a trivialization. These difficulties can all be dealt with by introducing a metric on L but there is one even more fundamental difficulty: what will play the role of α in our argument? How can we force a section of some line bundle to vanish at (x, y) without actually geometrically imposing this condition?

Here again, Vojta had a beautiful idea, inspired by Mumford's argument [M] which showed that if there are infinitely many rational points on a curve of genus ≥ 2 then they are very 'sparsely' distributed in the sense that the size of their coordinates grows rapidly. Mumford's argument uses height theory and the Mordell–Weil theorem, two key ingredients in Vojta's proof of Mordell's conjecture so we will begin by briefly recalling the key results.

We begin with a brief review of height theory. A thorough discussion with proofs can be found in Chapters 3 and 4 of [L1]. Suppose

$$x = (x_0, \dots, x_n) \in \mathbf{P}^n(\mathbf{Q}).$$

Suppose that one chooses the projective coordinates x_i to be relatively prime integers. Then the Weil height of x is defined by

$$H(x) = \sup_{0 \leq i \leq n} |x_i|.$$

The logarithmic Weil height is defined by

$$h(x) = \log H(x).$$

More generally, if k is a number field and M_k denotes the set of absolute values on k, normalized so that the product formula holds, then for $x \in \mathbf{P}^n(k)$ one defines

$$H_k(x) = \prod_{v \in M_k} \sup_{0 \leq i \leq n} \| x_i \|_v .$$

These height functions unfortunately depend on the field k, but can be normalized so as to avoid this problem. In particular, if $x \in \mathbf{P}^n(k)$ then we can define the normalized height

$$H(x) = H_k(x)^{1/[k:\mathbf{Q}]}.$$

Similarly one defines

$$h(x) = \log H(x).$$

The fact that these height functions do not depend on the choice of k is shown in [L1] pages 51–52 and consequently the subscript k can be ommitted. One thus obtains height functions H and h defined on the algebraic points $\mathbf{P}^n(\mathbf{Q}^a)$. As noted above, to avoid the cumbersome notation, we will limit sometimes limit ourselves to $\mathbf{Q}$–valued points. Observe that by definition $h(x)$ is a non–negative function on $\mathbf{P}^n(\mathbf{Q}^a)$.

Suppose now that L is an ample line bundle on a smooth projective variety X and $\phi_L : X \to \mathbf{P}^n$ is an embedding such that $\phi_L^* \mathcal{O}_{\mathbf{P}}(1) \simeq L$. Then one can try to define

$$h_L(x) = h(\phi_L(x)), \; x \in X(\mathbf{Q}^a). \tag{3.2}$$

The trouble with this is of course that the map ϕ_L is not uniquely determined by L; it depends on the choice of $n+1$ sections of $H^0(X, L)$. One can show, however, that if ϕ_1, ϕ_2 are two different embeddings of X, via global sections of L, in $\mathbf{P}^{n_1}$ and $\mathbf{P}^{n_2}$ respectively, then

$$h(\phi_1(x)) - h(\phi_2(x)) \;\text{ is a bounded function on } X(\mathbf{Q}^a).$$

The result of this is that there is a map

$$\mathrm{Pic}(X) \to \text{real valued functions on } X(\mathbf{Q}^a) \text{ modulo bounded functions.}$$

If L is a line bundle then the associated 'function' (which is only well–defined up to a bounded function) is denoted by h_L. Furthermore, this association is unique if one requires that for each very ample L in $\mathrm{Pic}(X)$, the function h_L must be equivalent to the function h_L in 3.2. As an immediate consequence of the definitions we see that if L is an ample line bundle and h_L any choice of height function associated to L then

$$h_L(x) \geq -C \;\text{ for some constant } C \geq 0. \tag{3.3}$$

We should note here that 3.3 will be the basic idea used to force a section $s \in H^0(C \times C, L)$ to vanish at a point x. In particular, 3.3 can be strengthened to say that if $L^{\otimes n}$ is generated by global sections on an open set U, for some positive integer n, then

$$h_L(x) \geq -C \;\text{ for all } x \in U. \tag{3.4}$$

Thus if we can choose L and x so that $h_L(x) \ll 0$ then we can hope to force all sections of L to vanish at x. Moreover, the vanishing is being controlled by arithmetic properties of L and x as in Roth's theorem.

On an abelian variety A one can use the group structure to obtain more information about the height functions h_L on A and in fact they can be normalized so that one has honest functions rather than entire equivalence classes modulo bounded functions ([L1] Chapter 5 gives a full account of this theory). The theory of heights on abelian varieties was developed by Néron and Tate. The basic results which we will use are the following: suppose L is a line bundle on A. Let $[-1] : A \to A$ denote the morphism given by taking inverses under the group law. A line bundle L on A is said to be *symmetric* if

$$[-1]^*L \simeq L.$$

Thus, given a line bundle L, $[-1]^*L \otimes L$ is always symmetric. If L is a symmetric line bundle on an abelian variety A then

$$h_L = q_L + O(1)$$

for some quadratic function q_L; here quadratic means that $q(x+y) - q(x) - q(y)$ is a bilinear function of x and y. This means that it is possible to *uniquely determine* h_L by choosing q_L within the equivalence class of height functions modulo bounded functions. The function q_L is called the Néron–Tate height function associated to L and it is denoted $\hat{h}_L$. The Néron–Tate heights satisfy the following properties (we assume here that the abelian variety A is defined over a number field k):

1: $\hat{h}_{L_1 \otimes L_2} = \hat{h}_{L_1} + \hat{h}_{L_2}$.

2: If L is ample and symmetric then $\hat{h}_L(x) \geq 0$, $\forall x \in A$. Moreover, $\hat{h}(x) = 0$ if and only if x is a torsion point.

3: If L is ample and symmetric then $\langle x, y\rangle_L = \hat{h}_L(x+y) - \hat{h}_L(x) - \hat{h}_L(y)$ is a symmetric bilinear form on $A(k)$.

4: If L is ample and symmetric then $\hat{h}_L$ is a positive definite quadratic form on $A(k)/\text{torsion}$.

In order to fully utilize the strength of 4 above one needs

Theorem 3.5 (Mordell–Weil). *If A is an abelian variety defined over a number field k then the set of k–valued points of A, $A(k)$, is a finitely generated abelian group.*

Using Theorem 3.5 together with 3 and 4 one sees that an ample symmetric line bundle L on A gives the vector space $A(k) \otimes_{\mathbf{Z}} \mathbf{R}$ a Euclidean structure:

$$\langle x, x\rangle_L = \hat{h}_L(2x) - 2\hat{h}_L(x) = 2\hat{h}_L(x) = 2|x|^2.$$

Furthermore, with respect to this Euclidean structure, the cosine between two points $0 \neq x, y \in A(k) \otimes_{\mathbf{Z}} \mathbf{R}$ is defined by

$$\cos(x, y) = \frac{\langle x, y\rangle_L}{2|x||y|}.$$

Suppose now that C is a curve of genus at least 2 defined over a number field k. Let J denote the Jacobian of C, viewed as the divisor classes of degree zero on C modulo linear equivalence. Choose a divisor class c_0 in $\mathrm{Pic}(C)$ of degree 1 such that the embedding $C \to J$ given by $P \mapsto \mathrm{cl}(P - c_0)$ gives an ample, symmetric Θ divisor (one essentially needs for c_0 to satisfy $(2g-2)c_0 = K_C$: see [SE] pp. 74–75 for details). In general, the class c_0 will only be defined over a finite extension of k so this will require a base change. With these conventions, Mumford's Theorem can be stated as follows:

Theorem 3.6 (Mumford). *Consider the positive definite quadratic form $\langle x, y\rangle_\Theta$ on $J(k)$ and its associated norm $|\cdot|$ and cosine. There exists a constant c depending only on C such that for all non–torsion points $x, y \in J(k)$*

$$\cos(x,y) \leq \frac{1}{2g}\left(\frac{|x|}{|y|}+\frac{|y|}{|x|}\right)+\frac{c}{|x||y|}.$$

The upshot of Theorem 3.6 is that if $\cos(x, y)$ is close to 1 (i.e. the angle between x and y is small) then it must be that $|x|$ and $|y|$ differ by a multiple of at least $2g$, i.e. the size of the rational points on J must grow exponentially inside the finite number of slices of $J(k)\otimes\mathbf{R}$ where the angles are small.

Proof of Mumford's Theorem. The proof of Theorem 3.6 is elementary, using linear equivalence of divisors and functoriality of heights. To this end, let $p_1 : J\times J \to J$ and $p_2 : J\times J \to J$ denote the projections to the first and second factors respectively. Also, let $p_{12} : J \times J \to J$ denote the addition morphism. Consider the Poincaré bundle on $J \times J$:

$$P = p_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta.$$

One shows that if $i : C \to J$ denotes our embedding

$$(i\times i)^*(P) \sim c_0\times C + C\times c_0 - \Delta, \tag{3.6.1}$$

$$i^*(\Theta) \sim gc_0 \tag{3.6.2}$$

where $\Delta \subset C \times C$ denotes the diagonal and $\sim$ denotes linear equivalence. Using the functorial properties of heights, 3.6.1, and 3.6.2, we obtain for any non–torsion $x, y \in C(k)$

$$\begin{aligned}\langle x,y\rangle_\Theta &= \hat{h}_P(x,y)\\ &\equiv h_{i\times i^*P}(x,y)\\ &\equiv h_{c_0\times C}(x,y)+h_{C\times c_0}(x,y)-h_\Delta(x,y)\\ &\equiv \frac{1}{g}\left(h_{gc_0}(x)+h_{gc_0}(y)\right)-h_\Delta(x,y)\\ &\equiv \frac{1}{g}(|x|^2+|y|^2)-h_\Delta(x,y),\end{aligned} \tag{3.6.3}$$

where $\equiv$ denotes equality up to the bounded function always implied when writing standard Weil heights. It then follows from 3.6.3 that

$$\begin{aligned}\cos(x,y) &= \frac{\langle x,y\rangle_\Theta}{2|x||y|}\\ &\equiv \frac{1}{2g}\left(\frac{|x|}{|y|}+\frac{|y|}{|x|}\right)-\frac{h_\Delta(x,y)}{2|x||y|}.\end{aligned}$$

But since Δ is effective 3.4 implies that for all $x \neq y$ (i.e. for all $(x,y) \notin \Delta$) we have

$$h_\Delta(x,y) \geq -O(1).$$

Thus for all $x \neq y$ we have

$$\cos(x,y) \leq \frac{1}{2g}\left(\frac{|x|}{|y|}+\frac{|y|}{|x|}\right)+\frac{O(1)}{|x||y|}. \tag{3.6.4}$$

Of course computing the constant implied in the $O(1)$ term would involve going through each choice being made of a height function on $C \times C$ and comparing it with the corresponding Néron–Tate height function on $J \times J$.

Mumford's result falls short of showing finiteness of $C(k)$ although it certainly points in this direction by indicating that the rational points are sparse in $J(k)$. Vojta's idea to improve upon this result is to associate to *each pair* of points $x, y \in C(k)$ a line bundle L on $C \times C$, similar in form to Mumford's bundle but designed specifically in order to force its sections to vanish at the point $(x,y) \in C \times C$. As in Mumford's argument, Vojta's bundles are linear combinations of the diagonal and fibres of the two projections. To this end, we write $F_1 = c_0 \times C$ and $F_2 = C \times c_0$ so F_1 and F_2 are (numerically equivalent to) fibres of the first and second projections respectively. Write

$$\Delta' = \Delta - F_1 - F_2.$$

Vojta considers divisors of the form

$$V = d_1F_1 + d_2F_2 + d\Delta'.$$

By functoriality of heights and 3.6.1, 3.6.2, one can choose the following for h_V:

$$h_V(x,y) = \frac{d_1|x|^2}{g}+\frac{d_2|y|^2}{g}-d\langle x,y\rangle. \tag{3.7}$$

As explained above after 3.4, we would like to force sections of V to vanish at the point (x,y) by showing that $h_L(x,y) \ll 0$. On the other hand, this is of little use if L has no global sections so one must look for a very particular line bundle.

Vojta's choice of divisor V is

$$\begin{aligned} d_1 &= \frac{N\sqrt{g+\epsilon}}{|x|^2}, \\ d_2 &= \frac{N\sqrt{g+\epsilon}}{|y|^2}, \\ d &= \frac{N}{|x||y|}, \end{aligned} \tag{3.8}$$

where $N \gg 0$ and $\epsilon \ll 1$. It should be remarked that one needs to be slightly careful here because one wants the coefficients d_1, d_2, d to be integers. This can of course be arranged by perturbing ϵ or N slightly but the perturbations may need to be different for d_1, d_2, and d. We will ignore this as it has no practical impact on the argument so we will simply assume that d_1, d_2, and d are integers. For this particular choice of V, we see from 3.7 that

$$h_V(x,y) = N\left(\frac{2\sqrt{g+\epsilon}}{g} - 2\cos(x,y)\right). \tag{3.9}$$

Since $g \geq 2$, the right hand side of 3.9 can be made *negative* if $\cos(x,y)$ is close to 1, i.e. if the points x, y are nearly parallel in $J(k) \otimes \mathbf{R}$. This is half of what we need; it remains to show that this choice of divisor is linearly equivalent to an effective divisor. In fact, an even stronger result holds

Lemma 3.10 (Vojta). *Given $\epsilon > 0$ if $d_1 \gg d_2$, with implied constant depending on ϵ, then the divisor V with d_1, d_2, and d as in 3.8 is ample.*

Proof of Lemma 3.10. One can compute explicitly the self-intersection number $V^2 = 2d_1d_2 - 2gd^2 > 0$. On the other hand, $F_1 + F_2$ is an ample divisor with

$$(F_1 + F_2) \cdot V > 0$$

which implies that some multiple of V must be effective. Thus we may assume that V itself is effective by increasing N. Moreover, the same argument applies to $V - \frac{\epsilon}{4}(F_1 + F_2)$ so this is also an effective divisor. Taking derivatives of a non-zero section of $V - \frac{\epsilon}{4}(F_1 + F_2)$ as we did in the product theorem gives a *nef* divisor and finally one can add $F_1 + F_2$ to make the divisor ample. We have differentiated a section of $V - \frac{\epsilon}{4}(F_1 + F_2)$ so that there is room to add the necessary fibres in order to make V itself ample.

Lemma 3.10 and 3.9 form the foundation of the basic argument used to prove the Mordell conjecture. Observe first that with the choice of V in 3.8, d_1 and d_2 are exactly the degrees appearing in in Theorem 2.10. Since $c_1(V)^2/2d_1d_2$ is small by construction, this means that as long as $d_1 \gg d_2$, Theorem 2.10 will bound the index of any non-zero section $s \in H^0(C \times$

$C, V)$ at any point of $C \times C$. Suppose now that we choose $x, y \in C(k)$ with $\cos(x, y) \sim 1$: this can be done if $C(k)$ is infinite because for any $\epsilon > 0$ one can divide $J(k) \otimes \mathbf{R}$ into finitely many regions such that any two points lying in the same region have cosine at least $1 - \epsilon$. Suppose in addition that $|x| \ll |y|$: one can do this because one of the regions of $J(k) \otimes \mathbf{R}$ described above must itself contain infinitely many points and hence points of arbitrarily large height. We choose $|x| \ll |y|$ so that $d_1 \gg d_2$ by 3.8. Since the Vojta divisor V is ample by Lemma 3.10, we can choose a section $s \in H^0(C \times C, V)$ such that $s(x, y) \neq 0$. On the other hand, by 3.9 we have

$$h_V(x, y) \ll 0.$$

Thus we can try to invoke 3.4 to force s to vanish at (x, y), thereby giving a contradiction. As in the proof of Roth's theorem, however, the argument will not be quite so simple: just as we needed to choose a very specific auxiliary polynomial $P \in \mathbf{Z}[X_1, \dots, X_m]$, so in the proof of the Mordell conjecture we will need to make a special choice for the section $s \in H^0(C \times C, V)$. Consequently it will no longer be clear that $s(x, y) \neq 0$. But we know in any case by the choice of V and Theorem 2.10 that $\mathrm{ind}_{(x,y)}(s)$ is small so an iteration of the above height theoretic argument, applied to derivatives of s, will ultimately lead to a contradiction.

In order to make this argument rigorous, we need to make the constant in 3.4 explicit and then show that 3.9 contradicts 3.4. We will first deal with the constant implicit in 3.4 which is equivalent to choosing a particular height function in the equivalence class of h_V. For this, choose $n_0 > 0$ such that $n_0 c_0$ is a very ample divisor on C and fix an embedding

$$\phi : C \to \mathbf{P}^n, \quad \phi^* \mathcal{O}_{\mathbf{P}^n}(1) \simeq \mathcal{O}_C(n_0 c_0).$$

Next choose a positive integer s such that $sF_1 + sF_2 - \Delta'$ is very ample and consider a second embedding

$$\psi : C \to \mathbf{P}^m, \quad \psi^* \mathcal{O}_{\mathbf{P}^m}(1) \simeq \mathcal{O}_{C \times C}(sF_1 + sF_2 - \Delta').$$

We will write $B = sF_1 + sF_2 - \Delta'$. Taking $\delta_1 = (d_1 + sd)/n_0$ and $\delta_2 = (d_2 + sd)/n_0$ we have

$$V = d_1 F_1 + d_2 F_2 + d\Delta' = \delta_1 n_0 F_1 + \delta_2 n_0 F_2 - dB.$$

By the basic properties of height functions we can choose

$$h_V = \delta_1 h_{n_0 F_1} + \delta_2 h_{n_0 F_2} - d h_B, \tag{3.11}$$

where the heights $h_{n_0 F_1}, h_{n_0 F_2}$, and h_B come from the morphisms ϕ and ψ above. The advantage to and need for choosing height functions as in 3.11 is that in order to prove the Mordell conjecture we will need to deal with a pair of rational points (x, y) of C and allow them to vary: since the height functions are only defined up to a bounded function, there will be no hope of deriving a contradiction unless the bounded functions are chosen in a uniform way, independent of the pair (x, y). More explicitly, with the choice of height functions in 3.11 we have

$$(3.12)\quad h_V(x,y) = \frac{d_1|x|^2}{g} + \frac{d_2|y|^2}{g} - d\langle x, y\rangle + O(d_1 + d_2 + d).$$

This is because the height functions associated to the embeddings ϕ and ψ differ from the corresponding Néron–Tate heights by fixed constants and these heights are then extended linearly to compare 3.11 to its associated Néron–Tate height.

Having now chosen explicit height functions, we have a quantitative version of 3.9. Recalling the definition 3.8 of d_1, d_2, and d, we see from 3.12 that

(3.13)

$$h_V(x,y) = N\left(\frac{2\sqrt{g+\epsilon}}{g} - 2\cos(x,y) + O\left(\frac{1}{|x|^2} + \frac{1}{|y|^2} + \frac{1}{|x||y|}\right)\right).$$

Thus if we choose $|x|$ and $|y|$ sufficiently large with respect to the implied constant, then 3.13 gives an explicit negative value for a particular choice of height function. We now need to work on making 3.4 effective in order to contradict 3.13.

The problem with 3.4 as stands is that it is abstract, having nothing to do with our particular embeddings and choices of height functions. We have chosen height functions by choosing the two embeddings $\phi : C \to \mathbf{P}^n$ and $\psi : C \times C \to \mathbf{P}^m$. By 3.11 we have

$$(3.14)\quad h_V(x,y) = \delta_1 h_{\mathbf{P}^n}(\phi(x)) + \delta_2 h_{\mathbf{P}^n}(\phi(y)) - d h_{\mathbf{P}^m}(\psi(x,y)).$$

In order to compare these three height functions, let the coordinates on $\mathbf{P}^n$ be denoted by $X_0, \ldots, X_n$ and on $\mathbf{P}^m$ we take projective coordinates $Y_0, \ldots, Y_m$. Let

$$f : C \times C \to \mathbf{P}^{N'}$$

denote the embedding given by bi–homogeneous monomials of bi–degree (δ_1, δ_2) in the coordinates $X_0, \ldots, X_n$. Let $0 \neq s \in H^0(C \times C, V)$ and consider the map

$$\beta_s : C \times C \dashrightarrow \mathbf{P}^m, \quad \beta_s(x,y) = \left(Y_0(x,y)^d s(x,y), \ldots, Y_m(x,y)^d s(x,y)\right).$$

The map β_s is defined on the open set $U = \{(x,y) \in C \times C : s(x,y) \neq 0\}$. Since

$$\beta_s^* \mathcal{O}_{\mathbf{P}^m}(1) \simeq f^* \mathcal{O}_{\mathbf{P}^{N'}}(1)$$

and since the linear series $H^0\left(C \times C, f^*\mathcal{O}_{\mathbf{P}}^{N'}(1)\right)$ is complete for δ_1, δ_2 sufficiently large, it follows that there is a rational map

$$g : \mathbf{P}^{N'} \dashrightarrow \mathbf{P}^m, \quad g(z) = (g_0(z), \ldots, g_m(z))$$

such that $g \circ f = \beta_s$ on U, where the g_i are linear forms:

$$\begin{tikzcd} & \mathbf{P}^{N'} \arrow[dr, "g"] & \\ U \arrow[ur, "f"] \arrow[rr, "\beta_s"'] & & \mathbf{P}^m \end{tikzcd}$$

Now suppose $(x, y) \in U$ and hence $s(x, y) \neq 0$. Consider

$$\beta_s(x, y) = \left(Y_0(x, y)^d s(x, y), \ldots, Y_m(x, y)^d s(x, y)\right) \in \mathbf{P}^m.$$

Since $s(x, y) \neq 0$ we have

$$h_{\mathbf{P}^m}(\beta_s(x, y)) = h_{\mathbf{P}^m}\left(Y_0(x, y)^d, \ldots, Y_m(x, y)^d\right) = dh_B(x, y). \tag{3.15}$$

On the other hand, we have

$$\beta_s(x, y) = (g_0 \circ f(x, y), \ldots, g_n \circ f(x, y)).$$

And from this we derive

$$\begin{aligned} h_{\mathbf{P}^m}(\beta_s(x, y)) &\leq h(f(x, y)) + O(1) \\ &= \delta_1 h_{n_0c_0}(x) + \delta_2 h_{n_0c_0}(y) + O(1), \end{aligned} \tag{3.16}$$

where the constant $O(1)$ measures the size of the coefficients of the linear forms $g_0, \ldots, g_n$. Combining 3.14, 3.15, and 3.16 gives

$$h_V(x, y) = \delta_1 h_{n_0c_0}(x) + \delta_2 h_{n_0c_0}(y) - dh_B(x, y) \geq O(1) \tag{3.17}$$

and this is exactly what we want, an effective version of 3.4 which can contradict 3.13 if the implied constant term of 3.17 is small enough. Assuming 3.17 and 3.13 are contradictory, this means that the section s must have vanished at the point (x, y) and we have achieved our goal, namely to force s to vanish at (x, y) for arithmetic reasons.

Controlling the size of the constant $O(1)$ in 3.17 is very much analogous to Siegel's lemma bounding the size of our auxiliary polynomial in the proof of Roth's theorem. Indeed, as the section $s \in H^0(C \times C, V)$ varies, the linear forms g_i also vary and we want to choose s so that these forms have small coefficients. Siegel's lemma requires an integral structure in order to work (one looks for solutions to equations with coefficients in the ring of integers of k) and consequently we need to introduce an integral structure on $H^0(C \times C, V)$, the vector space in which we are looking for our section s. This is relatively straightforward as it involves choosing a model for C over $\operatorname{Spec} \mathcal{O}_k$. So we choose a model $p : \mathcal{C} \to \operatorname{Spec} \mathcal{O}_k$ whose generic fibre is isomorphic to C. One asks in addition that $\mathcal{C}$ be regular, reduced, projective, and flat over $\operatorname{Spec} \mathcal{O}_k$: if necessary, one can pass to a finite extension of k in order to achieve this. We then consider

$$\mathcal{X} = \mathcal{C} \times_{\operatorname{Spec} \mathcal{O}_k} \mathcal{C}.$$

The model $\mathcal{X}$ can be singular and some modification is necessary in order to assume that $\mathcal{X}$ is regular and flat over $\operatorname{Spec} \mathcal{O}_k$ (for details, see [V2] p. 514). Choose a very ample line bundle $\mathcal{F}$ on $\mathcal{C}$ whose restriction to the generic fibre is $n_0 c_0$ and a very ample line bundle $\mathcal{B}$ on $\mathcal{X}$ whose restriction to the generic fibre is B. Consider the two projections $\pi_1 : \mathcal{X} \to \mathcal{C}$ and $\pi_2 : \mathcal{X} \to \mathcal{C}$. We define $\mathcal{F}_1 = \pi_1^* \mathcal{F}$ and $\mathcal{F}_2 = \pi_2^* \mathcal{F}$. Thus we have

$$H^0(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2) = H^0(\mathcal{X}, \delta_1 \mathcal{F}_1 + \delta_2 \mathcal{F}_2) \otimes_{\mathbf{Z}} \mathbf{Q}. \tag{3.18}$$

Hence $H^0(\mathcal{X}, \delta_1 \mathcal{F}_1 + \delta_2 \mathcal{F}_2)$ is a lattice inside $H^0(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2)$ giving us the desired integral structure.

Now that we have an integral structure on $H^0(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2)$ we can try to find a section s of 'small size' by applying Siegel's lemma. There is, however, one other technical issue to deal with. In particular, we need an intrinsic notion for the size of s which is not at all clear since there is no fixed basis for $H^0(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2)$ to give us coordinates whose size we can attempt to bound. Bombieri [B2] introduces coordinates and gives a very explicit construction. We will instead follow Faltings [F1], introducing metrics on our line bundles and use these to recover the height functions we are interested in computing.

Suppose Y is a smooth complex variety and L a line bundle on Y. Let $\{U_i\}$ be an open cover of Y on which we have trivializations

$$\psi_i : L|U_i \to \mathcal{O}_{U_i}.$$

Then we have the transition functions

$$\psi_i \circ \psi_j^{-1} : \mathcal{O}_{U_i \cap U_j} \to \mathcal{O}_{U_i \cap U_j}$$

which are sheaf isomorphisms and hence can be identified with multiplication by a unit $\phi_{ij} \in \mathcal{O}^*_{U_i \cap U_j}$.

Definition 3.19. A metric on a line bundle L is a collection $\{\rho_i\}$ of smooth real-valued functions

$$\rho_i : U_i \to \mathbf{R}_{>0}$$

satisfying the compatibility relations

$$\rho_i = |\phi_{ij}|^2 \rho_j.$$

The importance of a metric ρ on a line bundle L is that it allows one to measure the 'size' of a section $s \in H^0(Y, L)$ at a point $y \in U_i$:

$$|s(y)|_\rho^2 = \frac{|s_i(y)|^2}{\rho_i(y)}, \quad s_i = \psi_i(s).$$

The metric ρ is exactly what is needed to make this value independent of the open set U_i on which one evaluates the section s.

For us, the most important example of a metric is the standard metric on $\mathcal{O}_{\mathbf{P}^n}(1)$. Suppose the projective coordinates are $X_0, \ldots, X_n$ and we take the standard open affine cover $U_i = \{P \in \mathbf{P}^n : X_i(P) \neq 0\}$. Then we define ρ by

$$\rho_i : U_i \to \mathbf{R}_{>0} \quad \text{given by } \rho_i\left(\frac{X_0}{X_i}, \ldots, \frac{X_n}{X_i}\right) = \sum_{j=0}^{n} \left|\frac{X_j}{X_i}\right|^2.$$

One verifies that this collection of $\{\rho_i\}$ does indeed define a metric on $\mathcal{O}_{\mathbf{P}^n}(1)$. Metrized line bundles, like heights, have good functorial properties. In particular if (L, ρ) and (M, σ) are metrized line bundles then $(L \otimes M, \rho \otimes \sigma)$ is also a metrized a line bundle. Similarly $(L^{\otimes -1}, \rho^{-1})$ is a metrized line bundle and if $f : X \to Y$ is a morphism and (L, ρ) a metrized line bundle on Y then $(f^*L, f^*\rho)$ is a metrized line bundle on X.

We would like to use metrics in order to recover the height functions defined in 3.11 so that now all of the information we need in order to prove the Mordell conjecture will be encoded in the pair $(\mathcal{X}, \mathcal{V})$ where $\mathcal{V}$ will be a metrized version of the Vojta divisor V exetended to our model $\mathcal{X}$. We need one more definition before putting everything together. Suppose $\mathcal{Y} \to \operatorname{Spec} \mathbf{Z}$ is regular. A *metrized line bundle* $\mathcal{L}$ on $\mathcal{Y}$ is a line bundle $\mathcal{L}$ together with a metric on $\mathcal{L}$ over the complex points $\mathcal{Y} \otimes_{\mathbf{Z}} \mathbf{C}$. Recall the line bundles $\mathcal{B}, \mathcal{F}_1$, and $\mathcal{F}_2$ on $\mathcal{X}$. These give rise to maps to projective space

$$\phi : \mathcal{C} \to \mathbf{P}^n_{\mathbf{Z}}, \quad \psi : \mathcal{X} \to \mathbf{P}^m_{\mathbf{Z}}$$

endowing $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{B}$ with metrics; this is completely analagous to the manner in which we chose height functions for F_1, F_2, and B and so we have used the same letters to denote the maps to projective space.

At this point, to recover our height functions, we will assume that $\mathcal{C}$ and $\mathcal{X}$ are actually defined over $\operatorname{Spec} \mathbf{Z}$ in order to avoid considering multiple complex models. For a full discussion of this with details in the general case, one can consult [FW1] Chapter II. Suppose $0 \neq s \in H^0\left(\mathbf{P}^n_{\mathbf{Z}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)\right)$ and suppose $P \in \mathbf{P}^n(\mathbf{Q})$ is a point on the generic fibre of $\mathbf{P}^n_{\mathbf{Z}}$. Let E_P denote the closure of P in $\mathbf{P}^n_{\mathbf{Z}}$ so that E_P is a section of $\mathbf{P}^n_{\mathbf{Z}}$ over $\operatorname{Spec} \mathbf{Z}$. Let

$$\sigma_P : \operatorname{Spec} \mathbf{Z} \to \mathbf{P}^n_{\mathbf{Z}}$$

denote the morphism whose image is E_P.

Definition 3.20. Suppose $s \in H^0\left(\mathbf{P}^n_{\mathbf{Z}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)\right)$ and suppose $\mathcal{L}$ denotes the pair consisting of $\mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)$ with the standard metric ρ on $\mathcal{O}_{\mathbf{P}^n_{\mathbf{C}}}(1)$. We define the *degree* of $\mathcal{L}$ at P by

$$\deg_{\mathcal{L}}(P) = \log\left(|\Gamma(E_P, \sigma_P^*\mathcal{L})/s \cdot \mathbf{Z}|\right) - \log|s(P)|_\rho,$$

where $s \in H^0\left(\mathbf{P}^n_{\mathbf{Z}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)\right)$ does not vanish identically along E_P and Γ denotes global sections.

One can check that the definition of $\deg_{\mathcal{L}}(P)$ does not depend on the choice of s. Indeed, suppose $s' \in H^0\left(\mathbf{P}^n_{\mathbf{Z}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)\right)$ with s' not identically zero along E_P. Then, restricted to E_P, we have $s = rs'$ for $r \in \mathbf{Q}$. Equivalently, $ps = qs'$ for $p, q \in \mathbf{Z}$. Thus it suffices to show that the definition of $\deg_{\mathcal{L}}(P)$ does not change when one replaces s by as for $a \in \mathbf{Z}$. But this is immediate as both terms in the definition of $\deg_{\mathcal{L}}(P)$ change by $\log a$ under the substitution $s \mapsto as$.

Note moreover, for future reference, that identifying E_P with $\operatorname{Spec}\mathbf{Z}$ and $\sigma_P^*\mathcal{O}_{\mathbf{P}_{\mathbf{Z}}}(1)$ with $\mathcal{O}_{\operatorname{Spec}\mathbf{Z}}$ we obtain $\sigma_P^* s \in \mathbf{Z}$. Thus if v denotes a p–adic norm of $\mathbf{Z}$ we can define

$$|s|_v = |\sigma_P^* s|_v.$$

With this convention Definition 3.20 translates into the following:

$$\deg_{\mathcal{L}}(P) = -\sum_v \log|s|_v \tag{3.21}$$

where the sum ranges over all places, both finite and infinite, of $\mathbf{Q}$. This definition of the arithmetic degree of P makes it clear that it does not depend on the choice of s and will also be useful in §5.

We now return to our model $\mathcal{X}$ over $\operatorname{Spec}\mathbf{Z}$. We have extended the Vojta divisor V to a divisor $\mathcal{V}$ on $\mathcal{X}$ which inherits a metric μ from the metrics on $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{B}$. Suppose $(x, y) \in \mathcal{X}(\mathbf{Q})$ and let $E_{x,y}$ be the corresponding section over $\operatorname{Spec}\mathbf{Z}$. We claim that

$$\deg_{\mathcal{V}}(E_{x,y}) = h_V(x, y) + O(d_1 + d_2 + d), \quad \forall\ (x, y) \in \mathcal{X}(\mathbf{Q}). \tag{3.22}$$

Using the notation of Definition 3.20, we can establish 3.22 provided that we can show

$$\deg_{\mathcal{L}}(P) = h_{\mathbf{P}^n}(P) + O(1). \tag{3.23}$$

Indeed, both sides of 3.22 are defined via fixed maps of $\mathcal{X}$ and $C \times C$ to projective space, extending by linearity. But 3.23 is routine and can be found in [FW1] Chapter 2. Observe that according to Definiton 3.20 we have

$$\deg_{\mathcal{V}}(E_{x,y}) \geq -\log|s(P)|_\mu. \tag{3.24}$$

The advantage to this formulation in terms of metrics on line bundles is that we are now coordinate free and have a quantitative formulation of

3.17: in particular, according to 3.24 in order to obtain a lower bound for $\deg_{\mathcal{V}}(E_{x,y})$ it is sufficient to find a section $s \in H^0(\mathcal{X}, \mathcal{V})$ such that $s(P) \neq 0$ and $|s(P)|_\mu$ is small. At this point, we have finally come full circle and returned to the simple terms of Roth's theorem where we sought an auxiliarly polynomial with coefficients of small size and which did not vanish at P. To bound the size of the coefficients, we invoked Siegel's lemma. We now formulate a more general version of this principle due to Faltings [F1]:

Lemma 3.25 (Faltings). *Suppose V, W are metrized vector spaces with lattices $\Lambda_V \subset V$ and $\Lambda_W \subset W$ of maximal rank. Suppose $\psi : V \to W$ and $\psi(\Lambda_V) \subset \Lambda_W$. Suppose there exists a constant $C \geq 2$ such that each non-trivial element of Λ_V or Λ_W has norm at least $1/C$. Furthermore, we assume that Λ_V is generated by elements of norm at most C and that the norm of the map ψ is at most C. Let $b = \dim(V)$ and $a = \dim(\ker(\psi))$. Then there exists $0 \neq \alpha \in \ker(\psi) \cap \Lambda_V$ with $|\alpha| \leq (C^{3b} b!)^{1/a}$.*

Before applying Lemma 3.25, we note how it generalizes the classical Siegel lemma (Lemma 1.11). Suppose we apply Lemma 3.25 to the case where $V = \mathbf{Q}^n$ and $W = \mathbf{Q}^m$ where V and W are endowed with the standard metric and we take $\mathbf{Z}^n$ and $\mathbf{Z}^m$ as our respective lattices. Then the map ψ is given by an n by m matrix M with integer entries which we suppose are bounded in absolute value by A. Then the norm of the map ψ is at most nA. One checks readily that nA satisfies the other hypotheses for the constant C in Lemma 3.25 and the conclusion is that there exists an integer vector $(x_1, \ldots, x_n) \in \ker(\psi)$ satsifying

$$|x_i| \leq (nA)^{3n/(n-m)} (n!)^{1/(n-m)}, \quad 1 \leq i \leq n.$$

This is slightly weaker than Lemma 1.11 but qualitatively similar. Finally, for future reference, we give the full result as stated by Faltings ([F1] Proposition 2.18):

Lemma 3.26 (Faltings). *With notation as in Lemma 3.25, suppose*

$$\lambda_i = \min\{\lambda > 0 : \quad \textit{there exist } i \textit{ linearly independent elements of } V \textit{ with norm } \leq \lambda\}.$$

Then $\lambda_{i+1} \leq (C^{3b} b!)^{1/(a-i)}$.

How do we apply Faltings' version of Siegel's lemma? Faltings [F1] had the insight to use a Koszul–complex. To write down this sequence, recall that $\psi : C \times C \to \mathbf{P}^m$ is an embedding via sections of B and Y_i are the projective coordinates on $\mathbf{P}^m$. Furthermore, let $\gamma_i = \psi^* Y_i$. Consider the following exact sequence:

$$
\begin{aligned}
0 \to H^0(C \times C, V) \quad &\overset{\alpha}{\to} \quad \bigoplus_{i=0}^{m} H^0\left(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2\right) \\
\text{(3.27)} \qquad\qquad &\overset{\beta}{\to} \quad \bigoplus_{i,j=0}^{m} H^0\left(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2 + dB\right),
\end{aligned}
$$

where $\qquad \alpha(s) = \left(s \otimes \gamma_0^{\otimes d}, \dots, s \otimes \gamma_m^{\otimes d}\right)$

and $\qquad \beta(s_0, \dots, s_m) = \left\{\gamma_j^{\otimes d} s_i - \gamma_i^{\otimes d} s_j\right\}_{0 \le i,j \le m}$.

The lattice structure on these vector spaces comes from the arithmetic model $\mathcal{X}$ (see 3.18). Some care is needed here (for details, consult [F2] pp. 567–8) because the sections γ_i may not extend to sections over $\operatorname{Spec}\mathbf{Z}$ and consequently the sequence 3.27 may not be defined on the respective lattices inside the cohomology groups. This is not a serious problem, however, in this case, as the denominators over $\operatorname{Spec}\mathbf{Z}$ can be cleared uniformly in d.

The vector spaces in 3.27 inherit metrics from the metrics we have put on $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{B}$. Faltings applies Lemma 3.25 to the map β where the cohomology groups are given the associated sup–norms: if L is a line bundle with metric m on a smooth projective variety X and $s \in H^0(X, L)$ then we define

$$|s|_{\sup,m} = \sup\{|s(P)|_m : P \in X\}.$$

With these norms on the vector spaces in 3.27, Faltings' Siegel lemma 3.25 gives

$$\sigma \in H^0\left(\mathcal{X}, \delta_1 \mathcal{F}_1 + \delta_2 \mathcal{F}_2\right) = \ker(\beta)$$

with $|\sigma|_{\sup,\mu}$ suitably bounded.

In order to compute the upper bound for $|\sigma|_{\sup,\mu}$, we need to evaluate the constant C in Lemma 3.25. If we choose generators of the cohomology groups $H^0(\mathcal{X}, \mathcal{F}_1), H^0(\mathcal{X}, \mathcal{F}_2)$, and $H^0(\mathcal{X}, \mathcal{B})$ then monomials in these will give generators of the relevant lattices such that all norms involved are $\le O(d_1 + d_2 + d)$. In addition we need to compare the metric with which we have endowed

$$H^0(C \times C, V)$$

with the one which it inherits as a subspace of

$$\bigoplus_{i=0}^{m} H^0\left(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2\right).$$

For the details here, one can consult [EE] Chapter XI Lemma 3.1 or [V4] 13.2. Once these technical issues have been dealt with, we obtain from 3.24 a bound of the form

$$\deg_{\mathcal{V}}(E_{x,y}) \ge -\log|s(x,y)|_\mu \ge -\log|s|_{\sup,\mu}. \tag{3.28}$$

At this point, one must do a lot accounting to check to see that 3.28 gives a better lower bound than 3.22 and 3.13. Since 3.22 only holds under the assumption that $s(x,y) \neq 0$ the conclusion, as desired, is that $s(x,y) = 0$.

One would like to iterate this argument, applying it to derivatives of s to give a lower bound for $\mathrm{ind}_{(x,y)}(s)$. Definition 2.15 gives us a means to take derivatives of s over $\mathbf{Q}$ provided that all objects involved, i.e. the projections to $\mathbf{P}^1$ and the derivation on $\mathbf{P}^1$, are defined over $\mathbf{Q}$. This may introduce poles when extending to our model $\mathcal{X}$ over $\operatorname{Spec}\mathbf{Z}$ but this can be corrected by multiplying by a fixed integer a. Moreover, since the derivative is a linear map it can only increase norms by a fixed constant. In order to make this argument rigorous, one needs to check uniformity of the constants involved, i.e. taking α derivatives of s will only increase the norm of s by $O(\alpha)$ for some fixed O(1). There is some further subtlety here because s and ∂s are not sections of the same line bundles but the respective line bundles differ by fibres and these can be metrized in a uniform way. For precise details on this construction, one can consult [F1] pp. 570–571 or [EE] pp. 103–105.

Vojta [V4] §6 argues in a slightly different fashion. In particular, given a point $P \in C \times C$, he takes a local generator s_P for $\mathcal{O}_{C\times C}(V)$ on a product open set containing P. He then defines the derivative by taking a leading term in the Taylor series expansion of s/s_P. He then verifies that if the s_P are chosen "uniformly" as (d_1, d_2, d) varies, then the derivatives do not increase $|s|_{\mathrm{sup}}$ too much (see [V4] Lemma 6.2).

We now summarize the argument used to prove the Mordell conjecture, returning to the simple model of Liouville's theorem. So we first assume that there are infinitely many rational points of C and then choose $x, y \in C(\mathbf{Q})$ such that $0 \ll |x| \ll |y|$ and $\cos(x,y) \sim 1$. How large one needs to take $|x|$ and $|y|/|x|$ is determined by 3.13 as we need $h_V(x,y) \ll 0$. In addition, one needs to make sure that $|y| \gg |x|$ so that Theorem 2.10 will apply to the associated divisor V giving a very good bound for the index of any section of $H^0(C \times C, V)$ at any point. Once all of these choices have been made, the basic argument has the same steps as Roth and Liouville :

Step 1: Choose $0 \neq s \in H^0(\mathcal{X}, \mathcal{V})$ so that $|s|_{\mathrm{sup},\mu}$ is small.
Step 2: Show that $\deg_{\mathcal{V}}(E_{x,y}) \geq -\log|s|_{\mathrm{sup},\mu}$.
Step 3: Show that $\deg_{\mathcal{V}}(E_{x,y}) \ll 0$.

The bounds in Steps 2 and 3 contradict one another, forcing s to vanish on $E_{x,y}$. Taking derivatives gives a lower bound for $\mathrm{ind}_{(x,y)}(s)$ and we finally obtain a contradiction from Theorem 2.10.

4. Rational Points on Subvarieties of Abelian Varieties

Faltings [F1] extended Vojta's arithmetic tecnhiques to prove finiteness results for some higher dimensional subvarieties of abelian varieties. The main result is the following:

Theorem 4.1 (Faltings). *Suppose $X \subset A$ is a subvariety of an abelian variety with both X and A defined over a number field k. Suppose furthermore that X contains no translate of a non-trivial abelian subvariety $B \subset A$. Then $X(k)$ is finite.*

Of course the Mordell conjecture is a special case of Theorem 4.1 since a curve of genus at least 2 can not contain an elliptic curve inside its Jacobian. In [F2] Faltings generalized this result to deal with the case where X can contain translates of non-trivial abelian subvarieties $B \subset A$.

Theorem 4.2 (Faltings). *Suppose $X \subset A$ is of general type where both X and A are defined over a number field k. Then $X(k)$ is not Zariski dense in X.*

Using Ueno's theorem (see [L2] p. 35) on the structure of subvarieties of abelian varieties, one can in fact derive from Theorem 4.2 an even stronger corollary: namely, if $X \subset A$ is a subvariety and $X(k)$ is dense in X, then X must be a translate of an abelian subvariety.

We will first sketch the proof of Theorem 4.1: the proof of Theorem 4.2 is rather different and relies more heavily on the techniques of arithmetic intersection theory. It is worth remarking from the beginning that there is no analogue of Theorem 2.10 on products of higher dimensional varieties. The fundamental difficulty, looking back at the proof of Dyson's lemma, is that the ideal sheaves $\mathcal{I}_{\zeta_i,d,t_i}$ which were introduced in the proof always have mutually disjoint support (see 2.3): this is no longer the case in higher dimension and consequently the intersection theoretic construction used to prove Dyson's lemma breaks down as it is no longer possible to count independently the contribution of each point ζ_i. We will therefore introduce the proof of Theorem 4.1 by returning first to the Mordell conjecture and showing how our proof can be modified slightly to avoid appealing to Dyson's lemma.

Vojta's Dyson lemma (Theorem 2.10) was invoked in order to establish that the 'auxiliary section' $0 \neq s \in H^0(\mathcal{X}, \mathcal{V})$ with $|s|_{\sup,\mu}$ small does not have large index at $(x, y) \in C \times C$. This contradicts the lower bound for $\mathrm{ind}_{(x,y)}(s)$ obtained using the degree estimate 3.28 in conjunction with Faltings' Siegel lemma 3.25. Faltings' method [F1] is to explicitly construct s so that it is known to have small index at (x, y). Once this is accomplished, it is no longer necessary to appeal to Theorem 2.10.

Recall that in Lemma 3.10 it was shown that the Vojta divisor

$$V = d_1F_1 + d_2F_2 + d\Delta'$$

is ample. But V has been chosen so that it is also very 'close' to the boundary of the ample cone of $C \times C$; if d were increased slightly, the new divisor would no longer be ample because it would no longer satisfy $V^2 > 0$. Moreover, Theorem 2.10 would not hold if V were not close to the boundary of the ample cone: indeed, in this case one can always find a section of $H^0(C \times C, V)$ with large index at any given point. Faltings' approach uses a different divisor which will be similar in form to the Vojta divisor V. In particular we have

$$F = d_1F_1 + d_2F_2 + d\Delta'.$$

But the choice of d_1, d_2, and d is as follows:

$$\begin{aligned} d_1 &= \frac{N(g-\epsilon)}{|x|^2}, \\ d_2 &= \frac{N(g-\epsilon)}{|y|^2}, \\ d &= \frac{N}{|x||y|}. \end{aligned} \tag{4.3}$$

Note that this is identical to Vojta's choice of divisor (see 3.8) except that d_1 and d_2 have both been increased as $\sqrt{g+\epsilon}$ has been replaced by $g-\epsilon$. It follows immediately from Lemma 3.10 that F is ample but unlike the Vojta divisor it is *not* close the boundary of the ample cone: indeed it has been obtained from V by adding the ample divisor

$$(g - \epsilon - \sqrt{g+\epsilon})(d_1F_1 + d_2F_2)$$

to V.

Suppose now that ϵ is chosen sufficiently small so that

$$g - 2\epsilon \geq \sqrt{g+\epsilon}. \tag{4.4}$$

Note that since in practice ϵ can be taken arbitrarily small 4.4 is a very weak condition; since $g \geq 2$, one could in fact even assume that $g-\epsilon-1/2 \geq \sqrt{g+\epsilon}$. The exact value of ϵ is not, however, important in the argument. By 4.4

$$F - \epsilon(d_1F_1 + d_2F_2)$$

is an ample divisor on $C \times C$. Choosing d_1 and d_2 sufficiently divisible, we can assume that $F - \epsilon(d_1F_1 + d_2F_2)$ is generated by global sections, regardless of which fibers F_1 and F_2 are chosen. In particular, given any polynomial $f(\zeta_1, \zeta_2)$ of degree $\leq (d_1, d_2)$, where ζ_1 is a local parameter at x and ζ_2 a local parameter at y, it is possible to find a section $s \in H^0(C \times C, F)$ whose truncated Taylor series expansion about (x, y) is $f(\zeta_1, \zeta_2)$. In other words, if $\mathcal{I}_{(x,y)}(\epsilon)$ denotes the ideal sheaf of functions with index $\geq \epsilon$ at (x, y) then the map

$$H^0(C \times C, F) \to H^0\left(C \times C, F \otimes \mathcal{O}_{C \times C}/\mathcal{I}_{(x,y)}(\epsilon)\right) \tag{4.5}$$

is surjective.

Now, as in the proof of the Mordell conjecture, we extend F to a line bundle $\mathcal{F}$ on the arithmetic model $\mathcal{X}$ for $C \times C$ over $\operatorname{Spec} \mathcal{O}_k$. Since F, like V, is a linear combination of fibres and the diagonal, one can metrize $\mathcal{F}$ using the metrics on $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{B}$, extending by linearity. Suppose, in analogy to the Vojta divisor V in §3, we write

$$F = \delta_1 n_0 F_1 + \delta_2 n_0 F_2 - d\Delta'.$$

Consider now the Koszul complex for F, analogous to 3.27 for the Vojta divisor V:

$$\begin{aligned} 0 \to H^0(C \times C, F) \quad &\overset{\alpha}{\to} \quad \bigoplus_{i=0}^{m} H^0\left(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2\right) \\ &\overset{\beta}{\to} \quad \bigoplus_{i,j=0}^{m} H^0\left(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2 + dB\right). \end{aligned} \tag{4.6}$$

Each vector space in 4.6 inherits a metric from the metrics on $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{B}$.

Consider the vector space

$$U = \left\{s \in H^0(C \times C, F) : \operatorname{ind}_{(x,y)}(s) \geq \epsilon\right\}.$$

By 4.5, we can estimate the codimension

$$\operatorname{codim}(U, H^0(C \times C, F)) \geq \frac{\epsilon^2 d_1 d_2}{2}. \tag{4.7}$$

On the other hand, by 4.3 and Riemann–Roch on $C \times C$,

$$h^0(C \times C, F) = d_1 d_2 - g d^2/2 + O(d_1 + d_2 + d). \tag{4.8}$$

Choosing d, d_1 and d_2 sufficiently large to kill the $O(d_1 + d_2 + d)$ term in 4.8 and combining 4.7 with 4.8, we obtain

$$\operatorname{codim}(U, H^0(C \times C, F)) \geq \frac{\epsilon^2 h^0(C \times C, F)}{3}. \tag{4.9}$$

Now we apply Lemma 3.26 with $i = \dim U$, giving $\lambda_{i+1} \leq (A^{3b} b!)^{1/a-i}$ for some constant A satisfying the hypotheses of Lemma 3.26. Thus there exist $\dim U + 1$ independent sections $s_i \in H^0(\mathcal{X}, \mathcal{F})$, each satisfying

$$|s_i|_{\sup} \leq \left(A^{3b} b!\right)^{1/(a - \dim U)},$$

where $b = (m+1) h^0(\delta_1 n_0 F_1 + \delta_2 n_0 F_2)$ and $a = h^0(C \times C, V)$. Using 4.9 we see that

$$b/(a - \dim U) = O(1/\epsilon^2)$$

and consequently we obtain independent sections $s_1, \ldots, s_{a+1}$ each satisfying

$$|s_i|_{\sup} \leq O(\exp(1/\epsilon^2) \log A). \tag{4.10}$$

By the choice of a, at least one of these sections $s = s_i$ must satisfy $\operatorname{ind}_{(x,y)}(s) \leq \epsilon$ and one then checks that the bound on the norm of s, 4.10, is good enough to force s to vanish at (x, y). Applying the same argument to derivatives of s gives a lower bound on $\operatorname{ind}_{(x,y)}(s)$ which is larger than ϵ, giving the desired contradiction.

The proof of Theorem 4.1 has the same structure as the proof of Mordell's conjecture outlined above. We will first give an overview of the strategy and key steps and then fill in some of the missing details. As in the proof of the Mordell conjecture, Theorem 4.1 is proven by contradiction so we begin by assuming that $X(k)$ is infinite.

Step 1: Fix an ample symmetric line bundle L on A giving us the corresponding positive definite quadratic form on $A(k) \otimes_{\mathbf{Z}} \mathbf{R}$. To an m–tuple of points $x = (x_1, \ldots, x_m) \in X^m(k)$ one associates a line bundle L_x on X^m. Note that in the Mordell conjecture $m = 2$ but in the higher dimensional case one can have $m > 2$. One chooses L_x, just as we did in the Mordell conjecture, so that

$$h_{L_x}(x_1, \ldots, x_m) \ll 0.$$

The line bundle L_x is, as in the proof of the Mordell Conjecture, a linear combination of two types of line bundles, Poincaré bundles and pull–backs of L via the projection maps. More precisely, let $\pi_i : X^m \to X$ denote the projection to the i^{th} factor and for $1 \leq i \leq m-1$, consider the map

$$p_i : A^m \to A, \quad p_i(x_1, \ldots, x_m) = (x_i + x_{i+1}).$$

Then, exactly as we did for $C \times C$, we consider the Poincaré bundle

$$P_{i,i+1} = p_i^* L - \pi_i^* L - \pi_{i+1}^* L. \tag{4.11}$$

The line bundle L_x is a linear combination of the bundles $\pi_i^* L|X^m$ and $P_{i,i+1}|X^m$, chosen precisely to guarantee negativity of the height at $(x_1, \ldots, x_m)$.

Step 2: Next, one chooses integral models for X and A and extends L_x to a line bundle $\mathcal{L}_x$ on the integral model $\mathcal{X}$ for X. We put metrics on all of our line bundles and then apply Faltings' Siegel lemma 3.25 to obtain a section $s \in H^0(\mathcal{X}, \mathcal{L}_x)$ of small norm and as in 3.28

$$\deg_{\mathcal{L}_x}(E_{(x_1, \ldots, x_m)}) \geq -\log |s|_{\sup}.$$

Step 3: By the choice of $(x_1, \dots, x_m)$ and of height functions we obtain

$$\deg_{\mathcal{L}_x}(E_{(x_1,\dots,x_m)}) \ll 0.$$

Steps 2 and 3 force s to have large index at $(x_1, \dots, x_m)$. Unfortunately, without Dyson's lemma this is not yet a contradiction. We can alter the argument slightly, however, by applying Lemma 3.26 in Step 2 instead of Lemma 3.25. In particular, perturbing L_x by a small amount, we can assume that we can specify any leading terms of index $< \delta$ at $(x_1, \dots, x_m)$ and find a section of L_x with this truncated Taylor series expansion. As in the revised proof of the Mordell conjecture, we let

$$U = \left\{ s \in H^0(X^m, L_x) \middle| \ \mathrm{ind}_{(x_1,\dots,x_m)}(s) \geq \delta \right\}.$$

As in 4.7, the fact that we can specify any leading terms of index $< \delta$ gives an estimate for the codimension of U

$$\mathrm{codim}(U, H^0(X^m, L_x)) \geq f(\delta).$$

for some function f of δ. The refined Siegel Lemma 3.26 then gives a section index $< \delta$ at $(x_1, \dots, x_m)$ with $|s|_{\sup}$ suitably bounded and a contradiction is then obtained establishing Theorem 4.1.

The main technical issues in this sketch which remain to be clarified are the choice of the line bundle L_x and its positivity, in Step 1, and how to take derivatives of s once the bounds in Steps 2 and 3 force s to vanish at $(x_1, \dots, x_m)$. We will deal first with the choice of L_x. So suppose we are given an m–tuple $x = (x_1, \dots, x_m)$ of rational points of X and we wish to find a line bundle L_x, preferably ample, so that $h_{L_x}(x) \ll 0$. This is a more delicate question when X is of higher dimension essentially because in this case the diagonal $\Delta \subset X \times X$ is no longer a divisor (a fact which was essential in constructing the Vojta divisor V on $C \times C$) and consequently more work needs to be done in order find the appropriate line bundle L_x. The important observation is that in the proof of Theorem 3.6, the diagonal $\Delta \subset C \times C$ occurs in a linear equivalence relation involving the Poincaré bundle.

Following Mumford's lead, Faltings' choice of line bundle on X^m, as noted above in Step 1, is a linear combination of Poincaré bundles and pull–backs from the factors. To see how the bundle is chosen, first consider how we can choose a line bundle M on X^m such that $h_M(x_1, \dots, x_m) \sim 0$. So as in §3 we fix an ample symmetric line bundle L on A and let $|\cdot|$ denote the absolute value induced by $\hat{h}_L$ on $A(k) \otimes_{\mathbf{Z}} \mathbf{R}$. As in the Mordell conjecture, we assume that $|x_i|$ is an integer for all i. Let $N = \prod_{i=1}^m |x_i|$ and set

$$s_i = N|x_i|^{-1} \tag{4.12}$$

so s_i is an integer for $1 \leq i \leq m$. Consider the morphism $\phi_s : A^m \to A^{m-1}$ given by

$$\phi_s(a_1, \dots, a_m) = (s_1a_1 - s_2a_2, \dots, s_{m-1}a_{m-1} - s_ma_m). \tag{4.13}$$

Suppose, again as in the Mordell conjecure, that we choose the points $x_1, \dots, x_m$ to be nearly parallel in $A(k) \otimes_{\mathbf{Z}} \mathbf{R}$, i.e. we assume that $\cos(x_i, x_j) \sim 1$ for all $i \neq j$. Then $x_i/|x_i| - x_{i+1}/|x_{i+1}|$ is close to 0 and hence 4.13 and the choice of s_i implies that

$$|\phi_s(x_1, \dots, x_m)| = O(N\delta) \tag{4.14}$$

for $\delta = \max_{i,j}\{1 - \cos(x_i, x_j)\} \ll 1$; the absolute value $|\cdot|$ on $A^{m-1}(k) \otimes_{\mathbf{Z}} \mathbf{R}$ is the one induced by $|\cdot|$ on each factor. For a postive integer r, we will denote by $\mathcal{O}_{A^r}(1, \dots, 1)$ the line bundle $\otimes_{i=1}^r \pi_i^* L$ where $\pi_i : A^r \to A$ is the projection to the i^{th} factor. Similarly, we let $\mathcal{O}_{X^r}(1, \dots, 1) = \mathcal{O}_{A^r}(1, \dots, 1)|X^r$. Suppose now that we choose

$$M_s = \phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1).$$

By functoriality of heights, we can choose

$$h_{M_s} = \phi_s^* h_{\mathcal{O}_{A^{m-1}}(1, \dots, 1)}$$

giving, for $x = (x_1, \dots, x_m)$,

$$\begin{aligned} h_{M_s}(x) &= h_{\mathcal{O}_{A^{m-1}}(1, \dots, 1)}(\phi_s(x)) \\ &= \sum_{i=1}^{m-1} |s_ix_i - s_{i+1}x_{i+1}|^2 \\ &= O(N^2\delta^2). \end{aligned} \tag{4.15}$$

Thus we have produced a line bundle M_s, depending on the m–tuple $x = (x_1, \dots, x_m)$, with $h_{M_s}(x)$ small. Moreover, M_s is the pull–back of an ample divisor on A^{m-1} and consequently it is *nef* on X^m. In fact, if one knew that the morphism $\phi_s : X^m \to A^{m-1}$ were finite onto its image, this would imply that M_s is ample on X^m. Once this is known, one then needs to check how close M_s is to the boundary of the ample cone of X^m and then subtract the appropriate divisor to make the height at x negative.

In practice, showing that M_s is ample and bounding its distance from the boundary of the ample cone take place together. Therefore, we state the full result and will divide the proof up into several pieces: for a detailed discussion of the proof, see [EE] Chapter IX.

Lemma 4.16. *Let*

$$F_t = t\sum_{i=1}^{m} s_i^2 \pi_i^* L,$$

where we assume that $ts_i^2 \in \mathbf{Z}$ for all i. There exist positive integers $r, m > 0$ and a positive real number ϵ such that for any integer m–tuple $(s_1, \dots, s_m)$ satisfying $s_i/s_{i+1} \geq r$ for $1 \leq i \leq m-1$ the line bundle $M_{\epsilon,s} = \phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1) - F_\epsilon$ is ample on X^m.

Note that $M_s = M_{0,s}$ and that $M_{\epsilon,s} = M_s - F_\epsilon$. In particular once checks by 4.15 that

$$h_{M_{\epsilon,s}}(x) = h_{M_s}(x) - h_{F_\epsilon}(x) \ll 0 \tag{4.17}$$

provided δ is chosen sufficiently small with respect to ϵ, which one can arrange from the beginning since δ does not enter the statement of Lemma 4.16. Thus Lemma 4.16 produces precisely the analogue of the Faltings divisor F used above in the revised proof of the Mordell conjecture.

To establish ampleness of $M_{\epsilon,s}$ Faltings uses Kleiman's criterion for ampleness. In particular, it is enough to produce ϵ_0 such that $M_{\epsilon_0,s}$ is *nef* for all s satisfying the hypotheses of Lemma 4.16. Indeed, taking $\epsilon = \epsilon_0/2$, one sees that

$$M_{\epsilon,s} = M_{\epsilon_0,s} + \frac{1}{2} F_{\epsilon_0}$$

and the right hand side is ample by Kleiman's criterion.

Next note that Lemma 4.16 implies that $M_s = M_{0,s}$ is itself an ample line bundle under the hypotheses of the lemma. This means precisely that the map $\phi_s : X^m \to A^{m-1}$ is finite onto its image. Faltings begins the proof of Lemma 4.16 by showing that ϕ_s is finite onto its image for a *particular* choice of s. Then, using intersection theory and a higher dimensional version of the product theorem, he is able to prove Lemma 4.16.

Thus we begin by choosing particular values

$$t_i = 2^{m-i}, \quad 1 \leq i \leq m \tag{4.18}$$

and then show that the corresponding map ϕ_t is finite onto its image. This follows from [F1] Lemma 4.1 which says that for m sufficiently large the map $\psi : X^m \to A^{m-1}$ given by

$$\psi(\xi_1, \dots, \xi_m) = (2\xi_1 - \xi_2, \dots, 2\xi_{m-1} - \xi_m)$$

is finite onto its image. Indeed, consider the finite map $\phi : A^{m-1} \to A^{m-1}$ given by

$$\phi(a_1, \dots, a_{m-1}) = (2^{m-2}a_1, 2^{m-3}a_2, \dots, a_{m-1}).$$

Since $\phi_t = \phi \circ \psi$ it follows that ϕ_t is also finite onto its image. The proof that ψ is a finite map is more involved and it is for this that one needs to take m possibly quite large.

We conclude that $M_{0,t} = \phi_t^* \mathcal{O}_{A^{m-1}}(1, \dots, 1)$ is ample on X^m. It follows of course that $M_{\epsilon,t}$ is also ample for all ϵ sufficiently small. In order to quantify this, Faltings uses the following

Theorem 4.19 (Product Theorem). *Let* $\mathbf{P} = \mathbf{P}^{n_1} \times \ldots \times \mathbf{P}^{n_m}$. *Let* $d = (d_1, \ldots, d_m)$ *be an* m*-tuple of positive integers and suppose*

$$0 \neq s \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d)).$$

Finally, let

$$Z_\delta(s) = \{x \in \mathbf{P} : \mathrm{ind}_x(s) \geq \delta\}.$$

Given $\delta > 0$ *there exists* $r > 0$ *such that whenever* $d_i/d_{i+1} \geq r$ *for all* i *the set* Z_δ *is contained in a finite union of proper product subvarieties* $Y_1, \ldots, Y_k$. *Moreover, the degrees of the subvarieties* Y_i *with respect to* $\mathcal{O}_{\mathbf{P}}(1, \ldots, 1)$ *are bounded in terms of* δ.

The statement of Theorem 4.19 is identical to that of Theorem 2.5 with the additional information that the proper product subvarieties have bounded degree: of course, in Theorem 2.5 the only possibility for each Y_i is a point or $\mathbf{P}^1$, each having degree one. The proof of Theorem 4.19 is slightly more involved than the proof of Theorem 2.5 but the fundamental principle is the same. Suppose $W \subset Z_\delta(s)$ is an irreducible component which is *not* contained in a proper product subvariety; in particular, W is not a point. Let

$$r = 1 + \sum_{i=1}^{m} n_i.$$

By a dimension argument, there exists $0 \leq i \leq r - 1$ such that $Z_{i\delta/r}$ and $Z_{(i+1)\delta/r}$ have a *common* irreducible component Z with $W \subset Z$. The strategy of the proof is to show that Z is itself a proper product subvariety. To see why this is the case, let $c = \mathrm{codim}(Z, \mathbf{P})$. If $D_1(s), \ldots, D_c(s)$ denote general derivatives of s of index $\leq \frac{i\delta}{r}$ then W will be an irreducible component of $\cap_{j=1}^{c} Z(D_j(s))$. On the other hand, each of the derivatives $D_j(s)$ still has index at least $\frac{\delta}{r}$ along Z because Z is also an irreducible component of $Z_{(i+1)\delta/r}$. This gives a lower bound for the intersection multiplicity

$$i(Z : D_1(s) \cdot \ldots \cdot D_c(s))$$

and it turns out that this bound is too big if Z is not a product subvariety of $\mathbf{P}$. For details of the argument see [F1] Theorem 3.1 or [EE] p. 80.

We would like, in practice, to be able to apply the product theorem to a section $\sigma \in H^0(X, \mathcal{O}_X(d))$ where $X \subset \mathbf{P}$ is a proper product subvariety. The difficulty encountered here, as we saw in formulating Dyson's lemma for higher genus curves, is that there is no simple way to differentiate σ on X. To do so requires twisting the line bundle $\mathcal{O}_X(d)$. One can, however, project from X to a product of projective spaces and apply the product theorem there as on pp. 565–6 of [F1]. Thus we will continue to use Theorem 4.19 on proper product subvarieties.

The product theorem is applied in the following fashion: suppose $0 \neq s \in H^0(\mathbf{P}, M_{\epsilon,s})$. Provided the degrees d_i are chosen to be sufficiently rapidly decreasing, Theorem 4.19 can be applied to show that $Z_\delta(s)$ is contained in a finite union of product subvarieties of bounded degree. The conclusion is that $M_{\epsilon,s}$ is generated by global sections, namely the derivatives of s, on a large open subset. One would then like to apply the same argument inductively to the exceptional proper product subvarieties Y_i of Theorem 4.19. The inductive step is cohomological in nature and is achieved via the following result:

Lemma 4.20. *Let $s = (s_1, \ldots, s_m)$ be an arbitrary m–tuple of positive integers. Suppose $Y_1 \times \ldots \times Y_m \subset \mathbf{P}$ is a proper product subvariety with $\deg Y_i \leq N$ for all i and for some real number N. Then there exists $\epsilon(N) > 0$ such that for all s*

$$h^0\left(Y_1 \times \ldots \times Y_m, M_{\epsilon(N),s}^{\otimes n_s}\right) > 0 \quad \textit{for } n_s \textit{ sufficiently large.}$$

To see how to derive Lemma 4.16 from Lemma 4.20, choose ϵ_1 so that Lemma 4.20 applies with $Y_i = X$ for all i. Thus for any s, if n_s sufficiently large there exists a non–zero section $0 \neq \sigma_s \in H^0\left(X^m, M_{\epsilon_1,s}^{n_s}\right)$. We choose the number r in Lemma 4.16 to guarantee that Theorem 4.19 applies to the sections σ_s. Thus we conclude that each of the line bundles $M_{\epsilon_1,s}^{\otimes n_s}$ is generated by global sections outside of a finite union of proper product subvarieties $Y = Y_1 \times \ldots Y_m$; of course, the subvarieties Y may depend on the particular choice of s. Moreover, Theorem 4.19 tells us that

$$\deg(Y_i) \leq O(\epsilon_1) \quad \text{for all } i. \tag{4.21}$$

Choose $N_1 \geq O(\epsilon_1)$, where $O(\epsilon_1)$ satisfies 4.21, and choose $\epsilon_2 = \epsilon(N_1)$, with $\epsilon(N_1)$ defined as in Lemma 4.20. For each proper product subvariety Y appearing above in the first step of the induction, we apply Lemma 4.20, with $N = N_1$, giving sections of $H^0\left(Y, M_{\epsilon_2,s}^{\otimes n_s}\right)$ which generate off of a finite union of still smaller product subvarieties. We can now iterate this argument, possibly increasing the value of r at each stage when we need to apply Theorem 4.19, and eventually we find $\epsilon > 0$, namely the minimum of the ϵ_i encountered in the inductive procedure, such that $M_{\epsilon,s}$ is *nef* on X^m for all s satisfying the hypotheses of Lemma 4.16; since Theorem 4.19 needs to be applied at each stage of the induction, one must choose the largest value of r which occurs in the process. As noted above, this is sufficient to prove Lemma 4.16.

Proof of Lemma 4.20. The proof of Lemma 4.20 can be found in [V4] Proposition 11.5 or [F2] Proposition 3.4 and we sketch the important points here. There are two steps in the proof of Lemma 4.20: we first establish the result for $\epsilon(N) = 0$ and then refine the argument to obtain a positive lower bound for $\epsilon(N)$. Recall that

$$M_{0,s} = \phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1).$$

Faltings evaluates intersection products of the form

$$c_1(M_{0,s})^{\dim Y} \cap Y \tag{4.20.1}$$

for product subvarieties $Y = Y_1 \times \ldots \times Y_m \subset X^m$. The Chern class $c_1(\phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1))$ can be represented using two types of divisors, namely pull–backs $\pi_i^* L$ and the Poincaré bundles defined in 4.11. In the intersection class 4.20.1, most terms give zero contribution. Faltings shows that those terms which are non–zero are all proportional to $\prod_{i=1}^m s_i^{2\dim Y_i}$ and hence

$$c_1(M_{0,s})^{\dim Y} \cap Y = c_Y \prod_{i=1}^{m} s_i^{2\dim Y_i} \tag{4.20.2}$$

for some constant c_Y. Since we know for the particular choice $s = t$, with t as defined in 4.18, that $\phi_t^* \mathcal{O}_A(1, \dots, 1)$ is ample it follows that $c_Y > 0$ for all Y.

Since $\mathcal{O}_{A^{m-1}}(1, \dots, 1)$ is ample it follows that $\phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1))$ is *nef* for all s. Thus we conclude from 4.20.2 that $\phi_s^* \mathcal{O}_{A^{m-1}}(1, \dots, 1)|Y$ is also big for any product subvariety $Y \subset X$, i.e.

$$\lim_{n \to \infty} \frac{(\dim Y!) h^0\left(Y, \phi_s^* \mathcal{O}_{A^{m-1}}(n, \dots, n)\right)}{n^{\dim Y}} = c_Y \prod_{i=1}^m s_i^{2\dim Y_i} > 0. \tag{4.20.3}$$

Taking $Y = X^m$, 4.20.3 establishes Lemma 4.20 for $\epsilon(N) = 0$. In order to obtain a positive lower bound for $\epsilon(N)$ suppose

$$Y = Y_1 \times \ldots \times Y_m$$

is given with $\deg(Y_i) \le N$ for all i. Observe that c_Y in 4.20.3 must be an integer and hence $c_Y \ge 1$: this follows from 4.20.2 once one observes that the intersection number must be an integer.

Suppose D_n is an effective representative for $M_{0,s}^{\otimes n} - M_{\epsilon,s}^{\otimes n}$. In other words, in the notation of Lemma 4.16, D_n is an effective representative for $F_\epsilon^{\otimes n}$. Consider the short exact sequence of sheaves on X^m

$$0 \to M_{\epsilon,s}^{\otimes n} \to M_{0,s}^{\otimes n} \to M_{0,s}^{\otimes n}|D_n \to 0$$

and the beginning of its associated long exact cohomology sequence

$$0 \to H^0\left(Y, M_{\epsilon,s}^{\otimes n}\right) \to H^0\left(Y, M_{0,s}^{\otimes n}\right) \to H^0\left(D_n, M_{0,s}^{\otimes n}\right) \to \cdots \tag{4.20.4}$$

Thus we find

$$h^0\left(Y, M_{\epsilon,s}^{\otimes n}\right) \ge h^0\left(Y, M_{0,s}^{\otimes n}\right) - h^0\left(D_n, M_{0,s}^{\otimes n}\right). \tag{4.20.5}$$

Assuming without loss of generality that the fixed polarization L on A, used to define $M_{\epsilon,s}$, is very ample, we can choose

$$D_n = \sum_{i=1}^{m} \sum_{j=1}^{\epsilon n s_i^2} \pi_i^* D_{ij}$$

where D_{ij} are general effective representatives for $H^0(Y_i, L)$. Applying 4.20.5 gives

$$h^0\left(Y, M_{\epsilon,s}^{\otimes n}\right) \geq h^0\left(Y, M_{0,s}^{\otimes n}\right) - \sum_{i=1}^{m} \sum_{j=1}^{\epsilon n s_i^2} h^0(\pi_i^* D_{ij}, M_{0,s}^{\otimes n}). \tag{4.20.6}$$

Since $\deg Y_i \leq N$ for each i we find for $n \gg 0$

$$h^0\left(\pi_i^* D_{ij}, M_{0,s}^{\otimes n}\right) \leq \left(O(N) s_i^{2\dim Y_i - 2} \prod_{j \neq i} s_j^{2\dim Y_j} \right) n^{\dim Y - 1}, \tag{4.20.7}$$

for all i, j.

Combining 4.20.6 and 4.20.7 shows that

$$h^0\left(Y, M_{\epsilon,s}^{\otimes n}\right) \geq h^0\left(Y, M_{0,s}^{\otimes n}\right) - \left(m\epsilon O(N) \prod_{i=1}^{m} s_i^{2\dim Y_i} \right) n^{\dim Y}.$$

Choosing ϵ sufficiently small and applying 4.20.3 plus the observation that $c_Y \geq 1$ concludes the proof of Lemma 4.20.

This completes Step 1 of the proof of Theorem 4.1; in particular, we have associated to an m–tuple $x = (x_1, \ldots, x_m)$ of rational points of X a line bundle $M_{\epsilon,s}$ so that $h_{M_{\epsilon,s}}(x) \ll 0$ (see 4.17). Moreover, provided that $|x_1| \ll \ldots \ll |x_m|$, and hence $s_1 \gg \ldots \gg s_m$, we know from Lemma 4.16 that the line bundle $M_{\epsilon,s}$ is ample. Consequently, replacing ϵ with $\epsilon/2$, we can assume that there is a section $s \in H^0\left(X^m, M_{\epsilon,s}^{\otimes n}\right)$ with arbitrary truncated Taylor series expansion at x up to degree $(\epsilon n s_1^2, \ldots, \epsilon n s_m^2)$. In order to find a section $0 \neq \sigma \in H^0\left(X^m, M_{\epsilon,s}\right)$ of small index at x *and* small size, we wish to apply Lemma 3.26 as we did in 4.10. This will require an arithmetic model for X^m and an analogue of the Koszul complex 4.6.

In order to set up the arithmetic needed for the Koszul complex, we first choose a model $\mathcal{A}$ for A, regular and flat over $\operatorname{Spec} \mathcal{O}_k$. Then we extend the polarization L to an ample line bundle $\mathcal{L}$ on $\mathcal{A}$ and we fix a metric μ on $\mathcal{L}$, given for example by a projective embedding of $\mathcal{A}$ in $\mathbf{P}^n_{\mathcal{O}_k}$. As in the Mordell conjecture, we will try to metrize all line bundles functorially in terms of μ. Suppose $\mathcal{X}$ is the closure of X in $\mathcal{A}$, defined over $\operatorname{Spec} \mathcal{O}_k$. We would like to extend $M_{\epsilon,s}$ to a line bundle $\mathcal{M}_{\epsilon,s}$ on $\mathcal{X}^m$ and then endow $\mathcal{M}_{\epsilon,s}$ with a metric: here

$$\mathcal{X}^m = \mathcal{X} \times_{\operatorname{Spec} \mathcal{O}_k} \cdots \times_{\operatorname{Spec} \mathcal{O}_k} \mathcal{X}.$$

To extend $M_{\epsilon,s}$ to $\mathcal{X}^m$, observe that $M_{\epsilon,s}$ can be expressed as a linear combination of two types of line bundles, namely the pull–backs $\pi_i^* L$ via the projections and the Poincaré bundles $P_{i,i+1}$ defined in 4.11. Analogously to the fibres F_1 and F_2 in the Mordell conjecture, the line bundles $\pi_i^* L$ extend to $\pi_i^* \mathcal{L}$ and hence inherit a metric from the metric μ on $\mathcal{L}$. Similarly for the Poincaré bundles, one can choose the model $\mathcal{A}$ so that there is a line bundle $\mathcal{P}$ on $\mathcal{A} \times_{\operatorname{Spec} \mathcal{O}_k} \mathcal{A}$ whose restriction to $A \times A$ is P; we then fix a metric ν on $\mathcal{P}$ and $\mathcal{M}_{\epsilon,s}$ inherits a metric functorially from μ and ν.

Next, we deal with the analogue of the Koszul complex 4.6. This complex arose from a set of injections

$$f_\alpha : H^0(C \times C, F) \to H^0(C \times C, \delta_1 n_0 F_1 + \delta_2 n_0 F_2)$$

without common zeroes in the sense that for any section $s \in H^0(C \times C, F)$ the common zeroes of $\{f_\alpha(s)\}$ are contained in $Z(s)$. Working over k, we can try to implement the same technique on X^m. Returning to the definition of $M_{\epsilon,s}$, suppose we define a map $p_s : A^m \to A^{m-1}$, in analogy with 4.13, by

$$(4.22)\quad p_s(a_1, \ldots, a_m) = (s_1 a_1 + s_2 a_2, \ldots, s_{m-1} a_{m-1} + s_m a_m)$$

where as before $s = (s_1, \ldots, s_m)$. Using the theorem of the cube, one checks that

$$M_{\epsilon,s} \otimes p_s^* \mathcal{O}_{A^{m-1}}(1, \ldots, 1) \simeq \mathcal{O}_{A^m}\left(s_1^2, 2s_2^2, \ldots, 2s_{m-1}^2, s_m^2\right) \otimes F_{\epsilon,s}.$$

Choosing a set of generators $\{\gamma_i\}_{i=1}^N$ for $H^0(A^{m-1}, \mathcal{O}_{A^{m-1}}(1, \ldots, 1))$ gives sections $\beta_i = p_s^* \gamma_i$ which generate $p_s^* \mathcal{O}_{A^m}(1, \ldots, 1)$, thus giving rise to a Koszul complex for each positive integer n:

$$\begin{aligned}(4.23)\qquad & 0 \to H^0\left(X^m, M_{\epsilon,s}^{\otimes n}\right) \\ & \xrightarrow{\alpha} \bigoplus_{i=1}^{N} H^0\left(X^m, \mathcal{O}_{X^m}\left(s_1^2 - \epsilon, 2s_2^2 - \epsilon, \ldots, 2s_{m-1}^2 - \epsilon, s_m^2 - \epsilon\right)^{\otimes n}\right) \\ & \xrightarrow{\beta} \bigoplus_{i,j=1}^{N} H^0\left(X^m, \mathcal{O}_{X^m}\left(2s_1^2 - \epsilon, 4s_2^2 - \epsilon, \ldots, 4s_{m-1}^2 - \epsilon, 2s_m^2 - \epsilon\right)^{\otimes n} \otimes M_{0,s}^{\otimes -n}\right).\end{aligned}$$

Faltings and Vojta procede to embed $M_{0,s}^{\otimes -n}$ into

$$\bigoplus \mathcal{O}_X\left(s_1^2, 2s_2^2, \ldots, 2s_{m-1}^2, s_m^2\right)^{\otimes n}$$

by tensoring with sections of the appropriate line bundle, giving a new version of 4.23 where the bundles in the third direct summand are all of the simpler form

$$\mathcal{O}_{X^m}\left(3s_1^2 - \epsilon, 6s_2^2 - \epsilon, \ldots, 6s_{m-1}^2 - \epsilon, 3s_m^2 - \epsilon\right)^{\otimes n}.$$

Taking n large then allows one to choose generators for the second and third cohomology groups in 4.23 by taking polynomials of degree n in a fixed set of respective generators, precisely as we did in the case of the Mordell Conjecture. This is necessary because when we apply Lemma 3.26, we need a bound for the constant C appearing therein.

At this stage there are difficulties as we need an integral structure on the vector spaces in 4.23. The central difficulty here is that the morphism p_s does not necessarily extend to $\mathcal{A}^m$ and consequently the line bundle $p_s^*\mathcal{O}_{A^{m-1}}(1,\dots,1)$ does not necessarily extend to the arithmetic model $\mathcal{A}^m$. On the other hand, we have already chosen extensions of $\pi_i^*\mathcal{O}_A(1)$ and of the Poincaré bundles $P_{i,i+1}$ to $\mathcal{A}^m$ and hence any linear combination of such bundles also admits an extension. But (see [V4] 13.1) by the theorem of the cube, if a and b are integers and $f_{a,b} : A \times A \to A$ denotes the morphism given by $f_{a,b}(x,y) = ax + by$, then

$$f_{a,b}^*\mathcal{O}_A(1) \simeq \mathcal{O}_{A^2}(a^2, b^2) \otimes \mathcal{O}_{A^2}(abP); \tag{4.24}$$

here P denotes the Poincaré divisor on $A \times A$. Since all line bundles occurring in 4.23 can be expressed in terms of bundles of the form $f_{a,b}^*\mathcal{O}_A(1)$, we can use the isomorphisms of 4.24 to obtain an integral structure on the cohomology groups occurring in 4.23.

One needs to bound the denominators introduced by the maps α and β in 4.23 as these maps are not a priori defined over $\operatorname{Spec}\mathcal{O}_k$: this is done [V4] Corollary 13.7 or [EE] Proposition 3.3. Then, as in 4.10, one applies Siegel's lemma 3.26 to the sequence 4.23 to obtain a section $\sigma \in H^0(\mathcal{X}^m, \mathcal{M}_{\epsilon,s}^{\otimes n})$ with $|\sigma|_{\sup}$ suitably small satisfying $\operatorname{ind}_x(\sigma) \leq \epsilon$. Note that if $k \neq \mathbf{Q}$ then there are several complex models for X^m, each giving rise to a different metric on $\mathcal{M}_{\epsilon,s}^{\otimes n}$ and $|\sigma|_{\sup}$ denotes the supremum over all infinite places. Let E_x denote the fibre of $\mathcal{X}^m$ corresponding to the rational point $x = (x_1, \dots, x_m)$. We know by 3.24 that provided $\sigma(x) \neq 0$

$$\deg_{\mathcal{M}_{\epsilon,s}}(E_x) \geq -\log|\sigma|_{\sup} \tag{4.25}$$

On the other hand, 4.17 implies that

$$h_{M_{\epsilon,s}}(x) \ll 0, \tag{4.26}$$

provided one chooses $x_1, \dots, x_m$ so that $0 \ll h(x_1) \ll \dots \ll h(x_m)$. Using 3.22, one sees that 4.25 and 4.26 contradict one another, implying that σ must have vanished identically along E_x.

In order to derive a contradiction, one needs to take derivatives of the section σ and check that the bounds in 4.25 and 4.26 continue to contradict one another, establishing a lower bound for $\operatorname{ind}_x(\sigma)$. But we have explicitly chosen σ in order to have index $\leq \epsilon$ at x and this gives the final contradiction establishing Theorem 4.1. Some work is of course required in order to show how the derivatives affect the norm $|\sigma|_{\sup}$; the details can be found in [EE] chapter XI §4. Vojta [V4] deals with a more complicated situation where derivatives are needed on singular subvarieties of **P** and this is not necessary for the present situation. Finally, one must check all of the numbers to see that the upper and lower bounds for $\mathrm{ind}_x(\sigma)$ genuinely give a contradiction: this is done in detail in [EE] Chapter XI §5.

The proof of Theorem 4.2 given in [F2] is rather different in flavor from the proof of Theorem 4.1 just outlined. The main problem in the more general setting is that Lemma 4.16 is false if X contains a translate of a positive dimensional abelian subvariety $B \subset A$. Indeed, consider the case where $B \subset X$ for some abelian subvariety B of dimension > 0. Then the map ϕ_s, defined in 4.13, is not finite for *any* choice of s. To see why, let $t_i = \prod_{j \neq i} s_j$ and note that

$$(t_1 b, \ldots, t_m b) \overset{\phi_s}{\mapsto} (0, \ldots, 0), \quad \forall b \in B.$$

Since ϕ_s is never finite, regardless of the choice of m or s, it follows that the line bundle $\phi_s^* \mathcal{O}_{A^{m-1}}(1, \ldots, 1)$ is never ample and Lemma 4.16 is false. On the other hand, one checks that Lemma 4.20 still holds in this more general setting provided that $\phi_s | Y_1 \times \ldots \times Y_m$ is generically finite; indeed, the proof of Lemma 4.20 is entirely cohomological and it is sufficient to know that $M_{0,s}$ is big and *nef* on $Y = Y_1 \times \ldots \times Y_m$ which is true provided $\phi_s | Y$ is generically finite.

Thus we can still find an "auxiliary" section $\sigma \in H^0\left(X^m, \mathcal{M}_{\epsilon,s}^{\otimes n}\right)$ of small size so that σ is forced to vanish to large index at x. Unfortunately, we do not have a contradiction at this point because without ampleness of $M_{\epsilon,s}$ it is impossible to ensure from the beginning that σ has small index at x. Thus in order to derive a contradiction, more work is required. Faltings [F2] ultimately derives an arithmetic contradiction in the following manner. If we apply the product theorem 4.19 to σ (or more precisely, since Theorem 4.19 was only proven on a product of projective spaces, to the norm of σ via a generically finite map to a product of projective spaces: see [F1] p. 565 for details) we find a proper product subvariety X such that

$$x \in X = X_1 \times \ldots \times X_m \tag{4.27}$$

and such that σ has appropriate index along X. This is unfortunately a vacuous statement as stands since one could simply have $X_i = x_i$ for all i. However, since X is cut out by appropriate derivatives of σ, and in fact 'cut out with large index,' one can gain some control over X. In particular, one can show that X has bounded degree. More importantly, σ and its derivatives are all defined over $\operatorname{Spec} \mathbf{Z}$ so that one can use the machinery of arithmetic intersection theory pioneered by Gillet and Soulé, already hinted at when discussing metrized line sheaves. Using this theory, Faltings is able to show, [F1] Theorem 3.3, that the subvarieties X_i in 4.27 satisfy

$$\sum_{j=1}^{m} ns_j^2 h(X_j) \leq O\left(\sum_{j=1}^{m} ns_j^2\right) + O(\log|\sigma|_{\sup}); \tag{4.28}$$

here $h(X_j)$ denotes the arithmetic height of X_j, an invariant which measures the height of the equations defining X_j, generalizing the Weil height of a point $x \in \mathbf{P}^n(\mathbf{Q})$. Strictly speaking, one needs a more general version of the arithmetic product theorem as stated in [F1] Theorem 3.3, one which applies not only on a product of projective spaces but also on proper product subvarieties. A similar version of this theorem, stated in the appropriate generality, is given in [V4] Corollary 18.3.

If $X_j = x_j$ for some index j, then inequality 4.28 gives a contradiction provided $|\sigma|_{\sup}$ is appropriately small and $h(x_j)$ is sufficiently large. Thus to conclude the proof of Theorem 4.2, we procede by induction on the dimension of product subvarieties of X^m containing x. It suffices to show that for any product subvariety $x \in Y \subset X^m$ of bounded degree satisfying 4.28 one can produce a new section

$$\sigma_Y \in H^0(Y, M_{\epsilon,s})$$

again forced for arithmetic reasons to have large index at x. As in the proof of Theorem 4.1, the numbers now have to be worked out so that when one arrives at a proper product $Z \subset X^m$ with $Z_i = x_i$ for some i the analogue of 4.28 still holds, giving the desired contradiction.

5. The Schmidt Subspace Theorem

In Roth's theorem, we examined how well rational numbers can approximate a fixed algebraic irrational number α. One possible generalization of this problem to higher dimension would start with an affine variety $V \subset \mathbf{A}^n(k)$ defined over a number field k and then examine how close a rational point $x \in \mathbf{A}^n(\mathbf{Q})$ can be to V without lying inside V. The Schmidt subspace theorem deals with the special case where V is a union of hyperplanes in general linear position. Recall that for a point $x \in \mathbf{P}^n(\mathbf{Q})$, $H(x)$ denotes the absolute Weil height and $h(x)$ the logarithmic height.

Theorem 5.1 (Schmidt Subspace Theorem). *Suppose k is a number field with a fixed embedding in $\mathbf{C}$ and $L_0, \ldots, L_n$ are linearly independent linear forms with coefficients in k. Let $\delta > 0$ be given. Then the set of points $x = (x_0, \ldots, x_n) \in \mathbf{Z}^{n+1}$ satisfying*

$$\prod_{i=0}^{n} |L_i(x)| \leq 1/H(x)^{\delta}, \quad \textit{with } x_i \in \mathbf{Z} \textit{ relatively prime} \tag{5.2}$$

is contained in a finite union of proper linear subspaces.

In the special case when $n = 1$, suppose we take $L_0 = X_0$ and $L_1 = X_0 - \alpha X_1$ for an algebraic irrational number α. Then Theorem 5.1 says that there are only finitely many integer pairs (p, q) satisfying

$$|p||p - \alpha q| \leq 1/H(p,q)^\delta.$$

Thus there are only finitely many rational solutions to

$$|p/q - \alpha| \leq 1/|p||q|H(p,q)^\delta.$$

and this is equivalent to Roth's theorem since $|p||q|H(p,q)^\delta = O(q^{2+\delta})$ for all rational p/q close to α.

There is a more intrinsic formulation of Theorem 5.1. In particular, we can view x as a rational point of $\mathbf{P}^n$ and the linear forms L_i as sections of $\mathcal{O}_{\mathbf{P}^n_k}(1)$. If ρ denotes the standard metric on $\mathcal{O}_{\mathbf{P}^n_{\mathbf{Z}}}(1)$, defined in §3, then 5.2 is equivalent, up to a constant, to

$$\prod_{i=0}^{n} |L_i(x)|_\rho \leq 1/H(x)^{n+1+\delta}. \tag{5.3}$$

This is how Faltings–Wüstholz ([FW2] Theorem 9.1) express 5.2. The discrepancy between 5.2 and 5.3 is caused by the fact that for x_i relatively prime integers

$$|L_i(x)|_\rho = \frac{|L_i(x)|}{\sqrt{|x_0|^2 + \ldots + |x_n|^2}}.$$

Note that if 5.3 is satisfied, there must exist non–negative real numbers $p_0, \cdots, p_n$ with

$$p_0 + \ldots + p_n \geq n + 1 + \delta$$

such that

$$|L_i(x)|_\rho \leq 1/H(x)^{p_i} \quad \text{for all } i. \tag{5.4}$$

We now state a special case of the theorems of Faltings–Wüstholz dealing with solutions to the equations 5.4 for specific values of p_i:

Theorem 5.5 (Faltings–Wüstholz). *Let $L_0 \ldots, L_n \in H^0\left(\mathbf{P}^n_k, \mathcal{O}_{\mathbf{P}^n_k}(1)\right)$ be linearly independent. Suppose $\{p_i\}_{i=0}^n$ are non–negative real numbers satisfying*

$$\sum_{i=0}^{n} p_i \geq n + 1 + \delta \tag{5.6}$$

for some $\delta > 0$. Then the set of rational points $x \in \mathbf{P}^n(\mathbf{Q})$ satisfying 5.4 for all i lie in finitely many linear subspaces, at most one of which is non–trivial.

It is clear that the Schmidt subspace theorem implies Theorem 5.5 *without* the added information that at most one of the exceptional subspaces is non–trivial. The converse is also true: fix $\delta > 0$ in Theorem 5.1 and for each solution x let

$$p_i(x) = \sup\{r : |L_i(x)|_\rho \leq 1/H(x)^r\}.$$

By hypothesis, $\sum_{i=0}^n p_i(x) \geq n+1+\delta$; restricting to a subset of indices $I \subset \{0,\dots,n\}$ we can assume that $p_i(x) \geq 0$ for all $i \in I$ and $\sum_{i\in I} p_i(x) \geq n+1+\delta$. Theorem 5.5 then applies to each such set of indices I and each collection of $p_i(x)$; a priori there could be infinitely many such cases but if we apply Theorem 5.5 with $\delta/2$ in place of δ the excess of $\delta/2$ allows one to partition these into finitely many cases.

The proof of Theorem 5.5 proceeds, as with Roth's theorem, by constructing an auxiliary polynomial. To fix some notation, let L be a Galois extension of $\mathbf{Q}$ containing k. Since Theorem 5.5 over L implies Theorem 5.5 over k we can assume without loss of generality that k itself is Galois over $\mathbf{Q}$ with Galois group G. Let

$$\mathbf{P} = \mathbf{P}^n \times \dots \times \mathbf{P}^n, \text{ with } m \text{ factors.}$$

If R is a ring, usually either $\mathcal{O}_k$ or k in practice, we will denote by $\mathbf{P}_R$ the product $\mathbf{P}$ taken as defined over $\operatorname{Spec} R$. Let π_i denote the projection to the i^{th} factor. For fixed positive integers $d_1,\dots,d_m$, let

$$\mathcal{O}_{\mathbf{P}_R}(d) = \bigotimes_{i=1}^m \pi_i^* \mathcal{O}_{\mathbf{P}_R^n}(d_i).$$

Suppose that Theorem 5.5 is false and choose rational points x_i satisfying 5.4 with

$$0 \ll H(x_1) \ll H(x_2) \ll \dots \ll H(x_m).$$

We need to show that all but finitely many of the x_i are contained in some fixed linear subspace. As in Roth's theorem, choose

$$d_i \sim N/h(x_i) \text{ for } N \gg 0. \tag{5.7}$$

The analogue of Step 2 in Roth's theorem can now be carried out using the language of metrized line bundles.

We will use ρ to denote the standard metric on $\mathcal{O}_{\mathbf{P}^1_{\mathbf{Z}}}(1)$ and also, abusively, for all metrics defined functorially in terms of ρ by pull–backs and tensor products. Let E_x denote the section of $\mathbf{P}_{\mathbf{Z}}$ corresponding to the rational point $(x_1,\dots,x_m)$. If $\sigma \in H^0(\mathbf{P}_{\mathbf{Z}}, \mathcal{O}_{\mathbf{P}_{\mathbf{Z}}}(d))$ and $\sigma|E_x \neq 0$, then 3.21 implies

$$\deg_{\mathcal{O}_{\mathbf{P}_{\mathbf{Z}}}(d)}(E_x) \geq -\log|\sigma(x)|_\rho. \tag{5.8}$$

On the other hand, one can use 3.23 to show that

$$\deg_{\mathcal{O}_{\mathbf{P_Z}}(d)}(E_x) = \sum_{i=1}^{m} d_i h(x_i) + O\left(\sum_{i=1}^{m} d_i\right). \tag{5.9}$$

Putting together 5.7, 5.8, and 5.9 we see that

$$\log|\sigma(x)|_\rho \geq -mN + O\left(\frac{N}{h(x_1)}\right). \tag{5.10}$$

Exactly as in Roth's theorem, an upper bound on $\log|\sigma(x)|_\rho$ can be obtained by using 5.4 and taking a "Taylor series expansion" of σ in terms of the linear forms L_i. To be more precise, since $L_0, \ldots, L_n$ are linearly independent, they form a basis of $H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(1)\right)$. Thus monomials of degree d in the L_i generate $H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(d)\right)$ and can be used to write down the analogue of the Taylor series expansion 1.6. Given a positive integer r let

$$I_r = \left\{\prod_{i=0}^{n} L_i^{k_i} \,\middle|\, k_i \geq 0 \ \forall i \ \text{and} \ \sum_{i=0}^{n} k_i = r\right\}.$$

Let $I = (I_{d_1}, \ldots, I_{d_m})$ and write

$$\sigma = \sum_I a_I \bigotimes_{j=1}^{m} \pi_j^* M_{d_j}, \quad M_{d_j} \in I_{d_j}. \tag{5.11}$$

In order to obtain an upper bound for $\log|\sigma(x)|_\rho$, we define the weight of the monomials M_{d_j}:

$$\mathrm{wt}(M_{d_j}) = \sum_{i=0}^{n} \frac{k_i p_i}{d_j}. \tag{5.12}$$

According to 5.4,

$$|M_{d_j}(x)|_\rho \leq \frac{1}{H(x_j)^{\sum_{i=0}^{n} k_i p_i}}$$

so in particular, using 5.7 and 5.12,

$$\log|M_{d_j}(x)|_\rho \leq -N\sum_{i=0}^{n} \frac{k_i p_i}{d_j} = -N\mathrm{wt}(M_{d_j}). \tag{5.13}$$

Combining 5.11 and 5.13 gives the following estimate for $\log|\sigma(x)|_\rho$, analogous to 1.7 above:

$$\log|\sigma(x)|_\rho \leq \max\left\{-N\sum_{j=1}^{m} \mathrm{wt}(M_j) : a_I \neq 0\right\} + \log C. \tag{5.14}$$

As with 1.7, C depends on the number of terms in 5.11 and the size of the coefficients a_I. Comparing 5.10 with 5.14, we conclude that in order to force σ to vanish at x we need $a_I = 0$ unless

$$\sum_{j=1}^{m} \mathrm{wt}(M_j) > m(1+\epsilon) \tag{5.15}$$

for suitable $\epsilon > 0$. As in Roth's theorem, 5.15 also admits a formulation in terms of the index of σ along certain subvarieties. In particular, if H_i is the hyperplane defined by $L_i = 0$ and if $Y_i = H_i \times \ldots \times H_i \subset \mathbf{P}_k$ then 5.15 is satisfied provided

$$\sum_{i=0}^{n} p_i\,(\mathrm{ind}_{Y_i}\sigma) > m(1+\epsilon).$$

This analysis leads naturally to the following definition, the direct generalization of the index in the one variable case:

Definition 5.16. Let $\sigma \in H^0(\mathbf{P}_k, \mathcal{O}_{\mathbf{P}_k}(d))$ with expansion as in 5.11. Let

$$\mathrm{wt}(\sigma) = \min\left\{\sum_{j=1}^{m} \mathrm{wt}(M_{d_j}) : a_I \neq 0\right\}.$$

Note that the weight depends on the choice of L_i used in 5.11 though this has been suppressed from the notation.

In the language of Definition 5.16, we want $\sigma \in H^0(\mathbf{P_Z}, \mathcal{O}_{\mathbf{P_Z}}(d))$ with $\mathrm{wt}(\sigma) > m(1+\epsilon)$. As in the one variable case, it suffices to construct σ over $\operatorname{Spec}\mathcal{O}_k$ having $\mathrm{wt}(\sigma) > m(1+\epsilon)$ with respect to $\{L_0, \ldots, L_n\}$ and its Galois conjugates. We will denote by wt_γ the weight of Definition 5.16 computed with respect to the linear forms $\{\gamma(L_i)\}$. Thus, precisely as in Roth's theorem, we need to show that the number of m-tuples of monomials violating 5.15 is small compared to the total number of monomials. Again as in Roth's theorem, this can be accomplished using probability theory: a monomial in $H^0\left(\mathbf{P}^n_k, \mathcal{O}_{\mathbf{P}^n_k}(d)\right)$ is a choice of integers $k_0, \ldots, k_n$ with $\sum_{i=0}^{n} k_i = d$ so the "expected" value of each k_i is $d/(n+1)$. Thus the law of large numbers (see [FW2] Proposition 5.1) says that the expected weight of σ in 5.11 is $m\sum_{i=0}^{n} p_i/(n+1)$. By 5.6

$$m\sum_{i=0}^{n} \frac{p_i}{n+1} \geq m + \frac{m\delta}{n+1}.$$

and so we can construct a section $\sigma \in H^0(\mathbf{P_Z}, \mathcal{O}_{\mathbf{P_Z}}(d))$ such that the bound 5.14 is good enough to force σ to vanish at x, provided of course that one can deal with the constant C appearing in 5.14 (which is done, as usual, with the appropriate version of Siegel's lemma; see [FW2] §6). A refinement of the argument shows that σ has large index at x, in fact index $\geq O(\delta)$ (see [FW2] p. 128 for details).

Unfortunately, in higher dimension there is no analogue of Dyson's lemma and consequently the fact that σ is forced to vanish at x for purely arithmetic reasons does not yield a contradiction. Instead, as in Theorem 4.2, more arithmetic and in fact the same arithmetic is needed. Faltings–Wüstholz apply the arithmetic product theorem discussed at the end of §4, or rather the appropriate generalization given in [V4] Corollary 18.3, to σ at x concluding as in 4.28 that $x \in X_1 \times \ldots X_m$ where each X_i is defined over $\mathbf{Q}$ and

$$(5.17) \qquad \sum_{j=1}^{m} nd_j h(X_j) \leq O\left(\sum_{j=1}^{m} nd_j\right) + O(\log|\sigma|_{\text{sup}}).$$

But if $X_j = x_j$ for some j then 5.17 yields a contradiction provided $h(x_1)$ is sufficiently large. Thus whenever $X = X_1 \times \ldots \times X_m$ is a proper product subvariety, defined over $\mathbf{Q}$, containing x with $\dim(X_i) \geq 1$ for all i, one needs an inductive procedure to define wt_γ on $H^0(X, \mathcal{O}_X(d))$. We then look for some $0 \neq \sigma_X \in H^0(X, \mathcal{O}_X(d))$ with $\sum_{\gamma \in G} \mathrm{wt}_\gamma(\sigma_X)$ large enough to force σ_X to vanish to high index at x. Applying the arithmetic product theorem 4.28 inductively as in the proof of Theorem 4.2 will then yield the same contradiction, establishing Theorem 5.5.

At this point, we concentrate on how to find $0 \neq \sigma_X \in H^0(X, \mathcal{O}_X(d))$ with appropriate weights for a proper product subvariety $X \subset \mathbf{P}_{\mathbf{Q}}$. Observe that there is a natural filtration on

$$W_d = H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(d)\right)$$

for all positive integers d given by the weight, namely for $\alpha \in \mathbf{R}^+$ let

$$F^\alpha(W_d) = \left(\sigma \in H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(d)\right) \,\Big|\, \mathrm{wt}(\sigma) \geq \alpha\right).$$

Since we need to impose large weight not only with respect to the linear forms $\{L_i\}$ but also with respect to their Galois conjugates, we define an analogous weight filtration for each $\gamma \in G$:

$$F_\gamma^\alpha(W_d) = \left(\sigma \in H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(d)\right) \,\Big|\, \mathrm{wt}_\gamma(\sigma) \geq \alpha\right);$$

The filtrations F_γ^α in turn define a product filtration on $H^0\left(\mathbf{P}_k, \mathcal{O}_{\mathbf{P}_k}(d)\right)$, still denoted by F_γ^α.

For any subvariety $Y \subset \mathbf{P}_{\mathbf{Q}}^n$, Serre vanishing shows that

$$H^0\left(\mathbf{P}_{\mathbf{Q}}^n, \mathcal{O}_{\mathbf{P}_{\mathbf{Q}}^n}(d)\right) \to H^0(Y, \mathcal{O}_Y(d))$$

is surjective for all $d \gg 0$ and consequently the filtrations F_γ^α induce quotient filtrations on $H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k$ for $d \gg 0$, still denoted by F_γ^α. Similarly, we obtain an induced filtration F_γ^α on $H^0(X, \mathcal{O}_X(d)) \otimes_{\mathbf{Q}} k$ for any product subvariety $X \subset \mathbf{P}_{\mathbf{Q}}$. The inductive part of the proof of Theorem 5.5 can then be stated in terms of these filtrations as follows: given non-trivial $X_1, \ldots, X_m \subset \mathbf{P}_{\mathbf{Q}}^n$ and $d_1, \ldots, d_m$ sufficiently large we wish to show that there exists $0 \neq \sigma_X \in H^0(X, \mathcal{O}_X(d)) \otimes_{\mathbf{Q}} k$ such that

$$\sum_{\gamma \in G} \mathrm{wt}_\gamma(\sigma_X) > m[k : \mathbf{Q}](1 + \epsilon), \tag{5.18}$$

for some fixed ϵ depending linearly on δ. As was the case with the finiteness results of §4, one then needs to extend this construction over $\operatorname{Spec} \mathcal{O}_k$ and apply Siegel's lemma to bound the norm of σ_X. A quick computation, using 5.10 and 5.14, then shows that 5.18 forces σ_X to vanish to large index at x (see [FW2] pp. 129–130).

Let

$$V = H^0\left(\mathbf{P}_{\mathbf{Q}}^n, \mathcal{O}_{\mathbf{P}_{\mathbf{Q}}^n}(1)\right)$$

The fundamental insight of Faltings–Wüstholz is that for all product subvarieties $X \subset \mathbf{P}_{\mathbf{Q}}$ the filtrations F_γ^α on $H^0(X, \mathcal{O}_X(d)) \otimes_{\mathbf{Q}} k$ are naturally induced by the weight filtrations wt_γ on $V \otimes_{\mathbf{Q}} k$ and hence the whole problem reduces to studying properties of a set of filtrations on a finite dimensional vector space.

Let $Y \subset \mathbf{P}_{\mathbf{Q}}^n$ be a subvariety. To each filtration F_γ on $H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k$, Faltings–Wüstholz associate an expectation value $E_\gamma(Y)$ (see Definition 5.22 below) for which the law of large numbers applies. More precisely, let $X = X_1 \times \ldots \times X_m \subset \mathbf{P}_{\mathbf{Q}}$ be a product subvariety. Then for any $\epsilon_1, \epsilon_2 > 0$ there exists an m sufficiently large such that

$$\lim_{d \to \infty} \frac{\dim\left\{\sigma \in H^0(X, \mathcal{O}_X(d)) \otimes_{\mathbf{Q}} k \,|\, \mathrm{wt}_\gamma(\sigma) \geq \sum_{i=1}^m E_\gamma(X_i)(1 - \epsilon_1)\right\}}{h^0(X, \mathcal{O}_X(d))} \geq 1 - \epsilon_2 \tag{5.19}$$

for all $X = X_1 \times \ldots \times X_m$ with X_i non-trivial for each i. As we have seen, the assumption 5.6 guarantees that when $X_i = \mathbf{P}^n$ for all i, the expectation value $E_\gamma(\mathbf{P}_k^n)$ will be large enough to construct $\sigma \in H^0(\mathbf{P}_k, \mathcal{O}_{\mathbf{P}_k}(d))$ with

$$\sum_{\gamma \in G} \mathrm{wt}_\gamma(\sigma) > (1 + \epsilon) m [k : \mathbf{Q}]$$

with ϵ depending linearly on δ. Thus if we could show that for all subvarieties $Y \subset \mathbf{P}_{\mathbf{Q}}^n$ and for all $\gamma \in G$ we have $E_\gamma(Y) \geq E_\gamma(\mathbf{P}_k^n)$ then we would be done. In fact, it is sufficient to have

$$\sum_{\gamma \in G} E_\gamma(Y) \geq \sum_{\gamma \in G} E_\gamma(\mathbf{P}_k^n), \tag{5.20}$$

a *significantly* weaker condition. The reason why only $\sum_{\gamma \in G}$ is relevant is that whenever 5.18 is satisfied, σ_X is forced to have large index at x (provided, as always, that $|\sigma_X|_{\sup}$ is suitably bounded). It turns out that 5.20 means precisely that the filtrations F_γ on $V \otimes_{\mathbf{Q}} k$ are 'jointly semi–stable' (see Definition 5.23 below). When the filtrations are not semi–stable, all but finitely many solutions to 5.4 lie in the zero set of the maximal destabilizing subspace and thus one obtains a geometrical insight into where the solutions lie.

To make the discussion of the previous paragraph precise, we first need to define an expectation value associated to a filtration on a vector space and then examine how these filtrations behave in our particular situation when restricted to product subvarieties $X \subset \mathbf{P}$.

Definition 5.21. Let U be a finite dimensional vector space with a filtration indexed by positive integers:

$$U = F^0(U) \supset F^1(U) \supset \dots .$$

The *expectation value* associated to the filtration is defined as

$$\mu(U) = \frac{\sum_{i=0}^{\infty} i(\dim(F^i(U)/F^{i+1}(U)))}{\dim U}.$$

The invariant $\mu(U)$ denotes the maximum 'average weight' of a basis for U. In practice, our filtrations are indexed by the weight of sections of $\mathcal{O}_{\mathbf{P}_k}(d)$ which can be real valued since they depend on the p_i. One can, however, by a limit procedure ([FW2] p. 119), reduce to the case where the p_i are rational and then clear denominators in order to have an integer–valued filtration.

To see the meaning of Definition 5.21 and the reason why $\mu(U)$ is called the expectation value, consider the case where $U = H^0\left(\mathbf{P}^1_{\mathbf{C}}, \mathcal{O}_{\mathbf{P}^1_{\mathbf{C}}}(d)\right)$ and the filtration is given by order of vanishing at a fixed point $P \in \mathbf{P}^1_{\mathbf{C}}$. Then one sees that

$$\dim F^i(U)/F^{i+1}(U) = 1 \quad \text{for all } 0 \le i \le d$$

and all further steps in the filtration are trivial. Thus

$$\mu(U) = \frac{\sum_{i=0}^{d} i}{d+1} = \frac{d(d+1)}{2(d+1)} = \frac{d}{2}.$$

So the 'expected' multiplicity of a monomial of degree d at a fixed point is $d/2$ and the expected index is $1/2$: we have of course already encountered this in dealing with how much index we could impose at $(\alpha, \dots, \alpha)$ in Roth's theorem. The same analysis holds for $U_n = H^0\left(\mathbf{P}^n_{\mathbf{C}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{C}}}(d)\right)$ with the filtration given by wt_γ. Here the expected weight, asymtotically as $d \to \infty$, is $\sum_{i=0}^{n} p_i/(n+1)$.

These examples indicate the following definition for the expectation value of $Y \subset \mathbf{P}^n_{\mathbf{Q}}$:

Definition 5.22. Let $Y \subset \mathbf{P}^n_{\mathbf{Q}}$ and $\gamma \in G$. For d large, we have defined a filtration F^α_γ on $H^0(Y, \mathcal{O}_Y(d))$. Let μ_γ be the invariant in Definiton 5.21 computed with respect to the filtration F^α_γ. Let

$$E_\gamma(Y) = \lim_{d \to \infty} \frac{\mu_\gamma\left(H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k\right)}{d}.$$

Also define

$$E(Y) = \sum_{\gamma \in G} E_\gamma(Y).$$

Some work is necessary in order to show that the limit in Definition 5.22 exists and that it is an 'expectation value' to which the law of large numbers applies. This work is carried out in [FW2] §5. As observed above, we are interested in how the expectation value E behaves under taking quotients. To see how this is related to stability of filtrations, consider the following:

Definition 5.23. Let V be a finite dimensional $\mathbf{Q}$–vector space and k a finite Galois extension of $\mathbf{Q}$. Suppose $F_1, \ldots, F_r$ are filtrations of $V \otimes_{\mathbf{Q}} k$ with associated invariants μ_i as in Definition 5.21. Then $F_1, \ldots, F_r$ are said to be *jointly semi-stable* if for all $V' \subset V$

$$\sum_{i=1}^{r} \mu_i(V' \otimes_{\mathbf{Q}} k) \leq \sum_{i=1}^{r} \mu_i(V \otimes_{\mathbf{Q}} k).$$

Here $\mu_i(V' \otimes_{\mathbf{Q}} k)$ is computed with respect to the induced filtration on $V' \otimes_{\mathbf{Q}} k \subset V \otimes_{\mathbf{Q}} k$.

In practice, we will be interested in filtrations induced not on sub–spaces but rather on quotients, as $H^0(X, \mathcal{O}_X(d))$ is a quotient of $H^0\left(\mathbf{P}_{\mathbf{Q}}, \mathcal{O}_{\mathbf{P}_{\mathbf{Q}}}(d)\right)$. Definition 5.23 can be restated for this setting as follows: the filtrations $F_1, \ldots, F_r$ are jointly semi–stable if and only if for all *quotients* $V \to V'$ with the quotient filtration

$$\sum_{i=1}^{r} \mu_i(V \otimes_{\mathbf{Q}} k) \leq \sum_{i=1}^{r} \mu_i(V' \otimes_{\mathbf{Q}} k). \tag{5.24}$$

At this point we are almost prepared to prove Theorem 5.5. We apply Definition 5.23 to the vector space $U_n = H^0\left(\mathbf{P}^n_{\mathbf{Q}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Q}}}(d)\right)$ with filtrations $F_1, \ldots, F_{[k:\mathbf{Q}]}$ on $U_n \otimes_{\mathbf{Q}} k$ given by weight relative to $\{\gamma(L_i)\}$ for $\gamma \in G$. As we have already noted, for $Y \subset \mathbf{P}^n_{\mathbf{Q}}$ each filtration F_γ induces a filtration on $H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k$ provided $d \gg 0$. Note morever that the filtrations F_γ on $U_n \otimes_{\mathbf{Q}} k$ are in turn induced by the weights wt_γ on $V \otimes_{\mathbf{Q}} k$. Thus we would like to determine the joint semi-stability of the F_γ on the cohomology groups $H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k$ purely in terms of the filtrations F_γ on the vector space $V \otimes_{\mathbf{Q}} k$.

At this point in the argument Faltings–Wüstholz introduce the concept of semi-stability for vector bundles on curves and are able to resolve the semi-stability issues for the filtrations F_γ on $V \otimes_{\mathbf{Q}} k$ in these terms. To be precise, let W be a complex vector space, $V \otimes \mathbf{C}$ in practice, with filtrations $F_1, \ldots, F_r$ and associated expectation values $\mu_i(W)$. To the data $W, F_1, \ldots, F_r$, Faltings–Wüstholz ([FW2] Theorem 4.1) associate an algebraic curve defined over $\mathbf{C}$ with a vector bundle $E(W)$ satisfying

(i) $\sum_{i=1}^{r} \mu_i(W) = \mu(E(W))$ where $\mu(E(W))$ denotes the slope of $E(W)$,

(ii) W is jointly semi-stable if E(W) is slope semi-stable,

(iii) $W \mapsto E(W)$ commutes with tensor products.

The only difficult part of this is *(ii)* and this is shown in [FW2] Theorem 4.1. But, due to a result of Narasimhan and Seshadri, the tensor product of two semi-stable vector bundles on a curve is again semi-stable and so by *(ii)* and *(iii)* it follows that $\{F_\gamma\}$ are jointly semi-stable on $W^{\otimes d}$ for all $d > 0$ provided that they are jointly semi-stable on W. Putting everything together, we can now prove:

Theorem 5.25 (Faltings–Wüstholz). *Let*

$$V = H^0\left(\mathbf{P}^n_{\mathbf{Q}}, \mathcal{O}_{\mathbf{P}^n_{\mathbf{Q}}}(1)\right).$$

With assumptions and notation as in Theorem 5.5, assume in addition that the weight filtrations F^α_γ on V, induced by μ_γ, are jointly semi-stable. Then 5.4 admits only finitely many solutions.

Proof of Theorem 5.25. To begin the proof, note that by direct computation

$$\mu_\gamma(V \otimes_{\mathbf{Q}} k) = \sum_{i=0}^{n} \frac{p_i}{n+1} > 1 + \frac{\delta}{n+1}, \text{ for all } \gamma \in G.$$

On the other hand, by the fundamental properties of the association $V \mapsto E(V)$,

(5.25.1) $\mu_\gamma\left((V \otimes_{\mathbf{Q}} k)^{\otimes d}\right) = \mu_\gamma\left(E(V)^{\otimes d}\right) = d \cdot \mu_\gamma(E(V)) = d \cdot \mu_\gamma(V).$

For any $Y \subset \mathbf{P}^n_{\mathbf{Q}}$, we know that $H^0(Y, \mathcal{O}_Y(d)) \otimes_{\mathbf{Q}} k$ is a quotient of $(V \otimes_{\mathbf{Q}} k)^{\otimes d}$ for d sufficiently large. Thus 5.25.1 and joint semi-stability of the filtrations F_γ imply that

$$\begin{aligned} E(Y) = \sum_{\gamma\in G} \lim_{d\to\infty} \frac{\mu_\gamma\left(H^0(\mathcal{O}_Y(d))\otimes_{\mathbf{Q}} k\right)}{d} &\geq \sum_{\gamma\in G} \lim_{d\to\infty} \frac{\mu_\gamma\left((V\otimes_{\mathbf{Q}} k)^{\otimes d}\right)}{d} \\ &= \sum_{\gamma\in G} \mu_\gamma(V\otimes_{\mathbf{Q}} k) \\ &= [k:\mathbf{Q}]\sum_{i=0}^{n} \frac{p_i}{n+1}. \qquad (5.25.2) \end{aligned}$$

Since E is additive on product subvarieties $X_1 \times \ldots \times X_m$ we see that for any non–trivial $X_1, \ldots, X_m \subset \mathbf{P}^n(\mathbf{Q})$

$$E(X_1 \times \ldots \times X_m) \geq m[k:\mathbf{Q}]\sum_{i=0}^{n} \frac{p_i}{n+1}.$$

We know, however, from 5.6 that

$$\sum_{i=0}^{n+1} \frac{p_i}{n+1} \geq 1 + \frac{\delta}{n+1}$$

and, as noted in 5.18, this is sufficient to conclude the proof of Theorem 5.25.

To complete the proof of Theorem 5.5, it suffices to analyze the case when the filtrations μ_γ on $V\otimes_{\mathbf{Q}} k$ are not jointly semi–stable. In this case let $W \subset V$ be the first step of the Harder–Narasimhan filtration so that

$$\sum_{\gamma\in G} \mu_\gamma(W\otimes_{\mathbf{Q}} k) \geq \sum_{\gamma\in G} \mu_\gamma(V'\otimes_{\mathbf{Q}} k) \text{ for any } V' \subset W$$

and W is of maximal dimension satisfying this property. In particular W is jointly semi–stable with the induced set of filtrations $\{F_\gamma\}_{\gamma\in G}$. Moreover, being the first step in the Harder–Narasimhan filtration,

$$\sum_{\gamma\in G} \mu_\gamma(W\otimes_{\mathbf{Q}} k) \geq \sum_{\gamma\in G} \mu_\gamma(V\otimes_{\mathbf{Q}} k) \geq [k:\mathbf{Q}](1+\delta). \qquad (5.26)$$

By a linear change of coordinates we can assume that $W = (x_t, \ldots, x_n)$. Let

$$\Lambda = \{(x_0, \ldots, x_n \in \mathbf{P}^n : x_t = x_{t+1} = \ldots = x_n = 0\}.$$

and write

$$\phi : \mathbf{P}^n \backslash \Lambda \to \mathbf{P}^{t-1}$$

for the natural projection away from Λ. Theorem 5.5 follows from the following:

Proposition 5.27. *With assumptions as in Theorem 5.3, there are only finitely many points $x \in \mathbf{P}^n\backslash\Lambda(\mathbf{Q})$ satisfying 5.5.*

Proof of Proposition 5.27. Proposition 5.27 implies Theorem 5.5 with exceptional linear subspace Λ. Proposition 5.27 is proved by reducing to Theorem 5.25, using the fact that $W \otimes_{\mathbf{Q}} k$ with the induced filtrations is jointly semi-stable. Although it may happen that none of the $\{\gamma(L_i)\}$ actually lie in $W \otimes_{\mathbf{Q}} k$, it follows from 5.26 that there exist linear forms $M_1, \dots, M_l \in W \otimes_{\mathbf{Q}} k$ and non-negative real numbers $q_1, \dots q_l$ (a subset of $\{p_i\}$) such that, up to a constant c,

$$(5.27.1) \quad |L_i(x)|_\rho \leq \frac{1}{H(x)^{p_i}} \text{ for all } i \Rightarrow |M_j(x)|_\rho \leq \frac{c}{H(x)^{q_j}} \text{ for all } j,$$

$$(5.27.2) \qquad \{M_j, q_j\} \text{ give the same filtration on } W \otimes_{\mathbf{Q}} k \text{ as } \{L_i, p_i\}.$$

We now view W as $H^0\left(\mathbf{P}^{t-1}, \mathcal{O}_{\mathbf{P}^{t-1}}(1)\right)$ and apply Theorem 5.25 with the induced filtrations $F_\gamma^\alpha(W \otimes_{\mathbf{Q}} k)$. By assumption the $F_\gamma^\alpha(W \otimes_{\mathbf{Q}} k)$ are jointly semi-stable and hence it follows from Theorem 5.25, 5.26, and 5.27.2 that the system of equations

$$|M_j(x)|_\rho \leq \frac{1}{H(x)^{q_j}}$$

has only finitely many solutions. One can check that the same holds when we add the multiplicative constant c to the right hand side. But then by 5.27.1 we conclude that the solutions to

$$|L_i(x)|_\rho \leq \frac{1}{H(x)^{p_i}}, \quad 0 \leq i \leq n$$

map to a finite set of points under $\phi : \mathbf{P}^n \backslash \Lambda \to \mathbf{P}^{t-1}$. The fibres of ϕ, however, are isomorphic to affine space and if there were infinitely many solutions to $\left\{|L_i(x)|_\rho \leq \frac{1}{H(x)^{p_i}}\right\}$ contained in one of these fibres then their heights would become arbitrarily large; but the M_j are constant on the fibres of ϕ and so 5.27.1 would be violated.

At this point we can state the general result of Faltings–Wüstholz. The fact that we took a set of linearly independent forms $\{L_i\}$ and their Galois conjugates is not essential. Suppose, in fact, that we are given *any* finite set of absolute values $|\cdot|_v$, possibly including p–adic absolute values of the field k, and a system of inequalities

$$(5.28) \qquad |L_{i,v}(x)|_v \leq \frac{1}{H(x)^{p_{i,v}}}, \quad 0 \leq i \leq n_v,$$

for linear forms $L_{i,v} \in H^0\left(\mathbf{P}_k^n, \mathcal{O}_{\mathbf{P}_k^n}(1)\right)$. For each such v there is associated a filtration F_v of $V \otimes_{\mathbf{Q}} k$ where the linear form $L_{i,v}$ is assigned weight $p_{i,v}$. Let $\mu_v(V) = \mu_v(V \otimes_{\mathbf{Q}} k)$ denote the corresponding expectation value. Using

5.14 and 5.19 we can choose $\sigma \in H^0\left(\mathbf{P}_{\mathcal{O}_k}, \mathcal{O}_{\mathbf{P}_{\mathcal{O}_k}}(d)\right)$ so that for each place v we have

$$\log|\sigma(x)|_v \leq -mN(\mu_v(V) - \epsilon) + \log C_v \tag{5.29}$$

for an appropriate constant C_v; here ϵ can be chosen arbitrarily small provided that m is large enough to apply 5.19. In order to compute the constant C_v one needs an appropriate version of Siegel's lemma which can be found in [FW2] Proposition 6.1.

We also need bounds analogous to 5.29 for places w which are not amongst our finite set of places $\{v\}$. If w lies over p for some prime p then we have

$$\log|\sigma(x)|_w \leq 0, \tag{5.30}$$

since σ is defined over $\mathcal{O}_k$. For the archimedian places w one obtains a bound from the appropriate Siegel lemma of type

$$\log|\sigma(x)| \leq O(d_1 + \ldots + d_m). \tag{5.31}$$

Let E_x denote the section of $\mathbf{P}_{\mathbf{Z}}$ corresponding to the point x. Then we know from 3.21, or rather the appropriate extension to schemes over $\operatorname{Spec}\mathcal{O}_k$, 5.29, 5.30, and 5.31 that

$$\begin{aligned} \deg_{\mathcal{O}_{\mathbf{P}_{\mathcal{O}_k}}(d)}(E_x) &= \sum_w -\log|\sigma(x)|_w \\ &\geq mN\left(\sum_v(\mu_v(V) - \epsilon)\right) + O\left(\sum_{i=1}^m d_i\right); \end{aligned} \tag{5.32}$$

here w ranges over all places of k while v ranges only over the finite number we selected in 5.28. On the other hand, by 5.9, again over $\operatorname{Spec}\mathcal{O}_k$,

$$\deg_{\mathcal{O}_{\mathbf{P}_{\mathcal{O}_k}}(d)}(E_x) = [k:\mathbf{Q}]\sum_{i=1}^m d_i h(x_i) + O\left(\sum_{i=1}^m d_i\right). \tag{5.33}$$

Suppose now that

$$\sum_v \mu_v(V) > (1+\delta)[k:\mathbf{Q}] \tag{5.34}$$

for some $\delta > 0$. Combining 5.32, 5.33, 5.34 gives, for ϵ sufficiently small with respect to δ,

$$[k:\mathbf{Q}]mN + O\left(\frac{N}{h(x_1)}\right) \geq (1+\delta/2)[k:\mathbf{Q}]mN$$

which is a a contradiction provided $h(x_1) \gg 0$. We conclude, as desired, that $\sigma(x) = 0$; as usual, a refinement gives a lower bound for $\operatorname{ind}_x(\sigma)$, depending on δ. The inductive argument on a proper product subvariety

$X \subset \mathbf{P}$ is completely analogous, provided of course that the filtrations behave well when restricted to X. Thus one obtains

Theorem 5.35 (Faltings–Wüstholz). *Let μ_v be the invariant of Defninition 5.11 computed with respect to the filtration induced by $\{L_{i,v}, p_{i,v}\}$. Suppose*

$$\sum_v \mu_v(V) > [k : \mathbf{Q}].$$

Then outside of the zero set of the maximal destabilizing subspace, relative to $\sum_v \mu_v$, of V, there are only finitely many $x \in \mathbf{P}^n(\mathbf{Q})$ satisfying

$$|L_{i,v}(x)|_v \leq 1/H(x)^{p_{i,v}}, \text{ for all } i, v.$$

References

[B1] E. Bombieri, *On the Thue–Siegel–Dyson theorem.* Acta Math. **148** (1982), 255–296.

[B2] E. Bombieri, *The Mordell Conjecture revisited.* Ann. Sc. Norm. Sup. Pisa, Cl. Sci., IV, **17** (1991), 615–640.

[D] F.J. Dyson, *The approximation to algebraic numbers by rationals,* Acta Math. **9** (1947), 225–240.

[EE] B. Edixhoven & J.–H. Evertse editors, *Diophantine Approximation and Abelian Varieties.* Springer Lecture Notes **1566** (1993).

[EV] H. Esnault & E. Viehweg, *Dyson's Lemma for polynomials in several variables (and the Theorem of Roth).* Inv. Math. **78** (1984), 445–490.

[F1] G. Faltings, *Diophantine Approximation on Abelian Varieties.* Annals of Math. **133** (1991), 549–576.

[F2] G. Faltings, *The general case of S. Lang's conjecture.* in: Christante and Messing (eds.), Barsotti symposium in algebraic geometry, Academic Press, (1994), 175–182.

[FW1] G. Faltings & G. Wüstholz, editors, *Rational Points.* Vieweg, (1984).

[FW2] G. Faltings & G. Wüstholz, *Diophantine approximations on projective spaces.* Inv. math. **116** (1994), 109–138.

[H] M. Hindry, *Sur les Conjectures de Mordell et Lang.* Astérisque, **209** (1992), 39–56.

[L1] S. Lang, *Fundamentals of Diophantine Geometry.* Springer Verlag, (1983).

[L2] S. Lang (Ed.), *Number Theory III: Diophantine Geometry.* Springer Verlag, (1991).

[M] D. Mumford, *A Remark on Mordell's Conjecture.* American Journal of Math. **87, No. 4** (1965), 1007–1016.

[N1] M. Nakamaye, *Dyson's Lemma and a Theorem of Esnault and Viehweg.* Inv. Math. **121** (1995), 355–377.

[N2] M. Nakamaye, *Dyson's Lemma with Moving Parts.* Mathematische Annalen, **310** (1998), 161–168.

[N3] M. Nakamaye, *Intersection Theory and Diophantine Approximation* . to appear, Journal of Algebraic Geometry.

[S1] W. Schmidt, *Diophantine Approximation,* Springer Lecture Notes **785** (1980).

[S2] W. Schmidt, *Diophantine Approximations and Diophantine Equations.* Springer Lecture Notes **1467** (1991).

[SE] J.-P. Serre, *Lectures on the Mordell–Weil Theorem.* Vieweg, (1990).

[Vi] C. Viola, *On Dyson's lemma.* Ann. Sc. Norm. Super. Pisa, **12** (1985), 105–135.

[V1] P. Vojta, *Dyson's lemma for products of two curves of arbitrary genus.* Inv. Math. **98** (1989), 107–113.

[V2] P. Vojta, *Siegel's theorem in the compact case.* Annals of Math. **133** (1991), 509–548.

[V3] P. Vojta, *A generalization of theorems of Faltings and Thue-Siegel-Roth-Wirsing.* Journal AMS, 4 (1992), 763–804.

[V4] P. Vojta, *Some applications of arithmetic algebraic geometry to diophantine approximations.* Proceedings of the CIME Conference, Trento, (1991), LNM 1553, Springer, ((1993)).

[V5] P. Vojta, *Integral points on subvarieties of semi-abelian varieties, I.* Inv. Math. **126** (1996), 133–181.

Michael NAKAMAYE
Mathematics Department
University of New Mexico
Albuquerque, NM 87131
USA
E-mail : nakamaye@math.unm.edu