Roth's Theorem: an introduction to diophantine approximation

Michael Nakamaye

June 18, 2014

0 Introduction

Fundamental to nearly all diophantine arguments is the simple fact that the smallest positive integer is one. One immediate consequence of this result is that if $P(X) \in \mathbf{Z}[X]$ is a polynomial of degree d and p/q a rational number then either $P(p/q) = 0$ or else

$$|P(p/q)| \geq \frac{1}{q^d}.$$

Indeed, $P(p/q)$ is a sum of rational numbers whose denominators are all factors of q^d: expressing this as a rational number with denominator q^d, it is either identically zero or it is at is at least $1/q^d$ in absolute value *because* one is the smallest positive integer. Since $|P(p/q)|$ is bounded below as a function of q, when it is non–zero, it follows from a continuity argument that p/q can not be too close to an irrational root of P, again as a function of q. This is the substance of the famous theorem of Liouville.

A vast refinement of this line of reasoning leads to Roth's theorem on the approximation of algebraic irrationals by rational numbers. Further advances along the same lines include the Schmidt subspace theorem dealing with, amongst other cases, the simultaneous approximation of several irrational numbers by rational numbers with fixed denominator. Moreover, the same basic observation is at the foundation of Vojta's proof of the Mordell conjecture although now the absolute value at hand is replaced with the more sophistocated machinery of height functions and metrics on line bundles.

Surely, the reader will appropriately wonder, ground–breaking results such as Roth's theorem, the Schmidt subspace theorem, the Mordell conjecture and its higher dimensional analogue, are deeper and more far–reaching than a simple observation about the smallest positive integer? Indeed, the size constraint on $|P(p/q)|$ is a starting point, a *diophantine constraint* so to speak. Coaxing a contradiction out of this constraint, however, often requires formidable skill in both geometry and arithmetic. In these lectures, we will focus more on geometry than on arithmetic, not because of a judgment about value or interest but rather because of the personal strengths and weaknesses of the author.

The fundamental difficulty to be overcome in establishing Roth's theorem is to extend the simple machinery outlined above to polynomials of several variables and, more importantly, establish the non–vanishing of the approximating polynomial at a particular rational point. Because the polynomial is constructed via an abstract existence argument, there is no reason why it might not vanish, indeed vanish to high order, at the important rational point of interest. The key non–vanishing result can be obtained in at least three ways:

1. Roth's lemma,

2. Dyson's lemma,

3. the arithmetic product theorem.

We will discuss all three approaches with an emphasis on **2** and **3**. There is a natural progression from **1** to **3** as **1** is essentially an arithmetic result, **2** is purely geometric, and **3** uses a combination of arithmetic and geometry.

The organization of these notes is as follows. In §1 we prove Liouville's theorem and study how it might be improved. This leads to consideration of a polynomial in several variables and to the key notion of the index. Using several variables also entails using several good rational approximations of the fixed algebraic irrational. The rational approximating points and the auxilliary polynomial are intimately linked and it is this connection which is studied in §1. In §2 we set up the argument to establish Roth's theorem with a focus on how the auxilliary polynomial f is constructed and what motivates the various hypotheses on f. Siegel's Lemma, a key ingredient in nearly all diophantine arguments, is discussed in detail. The proof of Roth's theorem is now reduced to showing that f does not have large index at the appropriate rational approximating points.

In §3 we discuss two methods for bounding the index of f at the rational approximating points: Roth's lemma and the arithmetic product theorem. The third approach, Dyson's lemma, is discussed in §4. A proof is sketched in the two variable case and this contains all of the important ideas of the general case. In §5, we examine how the basic techniques developed in proving Roth's theorem are fundamental to many diverse diophantine arguments which represent the main themes of the Rennes conference– the Schmidt subspace theorem, the Mordell conjecture, and Faltings' theorem on rational points of subvarieties of abelian varieties. We focus on the latter two and sketch the relationship between the proof of Roth's theorem and the proof of Faltings' theorems. Finally, we turn to connections between the techniques used to prove Roth's theorem and certain themes in higher dimensional complex algebraic geometry.

The spirit of these notes is rather different from that of [N3] which covers very similar material. The goal here is not just to present the main themes occurring in Roth's theorem but also to motivate the structure of the argument. In other words, the point of view taken is that of someone who is seeing Roth's theorem for the first time and has no clear instinct how to procede. Studying the proof in its finished form is assuredly a necessary endeavor but it is equally important to develop an intuition and understanding of why the proof is constructed the way it is. We have included some exercises at the end of each section to help

the reader better digest the subtleties of the material. Most of these are relatively routine but this does not of course imply *easy*: in any case, they are all designed to develop a deeper understanding of fundamental diophantine techniques.

It is a pleasure to thank the organizers of the Rennes conference, Antoine Chambert–Loir, Carlo Gasbarri, Lucia Di Vizio, Marc Hindry, and Hugues Randriam, for inviting me to revisit this material and try to rethink how to best present these fascinating ideas to a new generation of mathematicians. Since the 2009 conference in Rennes where these ideas were first presented, the author had an opportunity to focus on some different aspects of this story, namely the contributions of Thue and Siegel to the development of rational approximation to algebraic irrational numbers. These notes have been slightly updated to include some of the information about Thue's ideas which were presented at a conference organized by Carlo Gasbarri, Erwan Rousseau, and Steven Lu at CRM in Montréal during the summer of 2013.

1 Liouville's Theorem and Beyond

Liouville's theorem is one of the first and most important examples of the basic techniques which lie behind many arguments in diophantine approximation.

Theorem 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 $p/q \in \mathbf{Q}$*

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

As outlined in the introduction, the proof involves four key steps:

Step 1. Choose a polynomial $f(X) \in \mathbf{Z}[X]$ vanishing at α: this polynomial is unique if we ask that it be irreducible over $\mathbf{Z}$ with positive leading coefficient.

Step 2. $f(p/q) \neq 0$.

Step 3. $|f(p/q)| \geq 1/q^d$.

Step 4. $|f(p/q)| \leq b(\alpha)|p/q - \alpha|$ for an explicit constant $b(\alpha)$, provided $|p/q - \alpha| \leq 1$.

Liouville's theorem follows, with $c(\alpha) = \min\{1, 1/2b(\alpha)\}$, by comparing the bounds in Steps 3 and 4. Step 1 requires no further explanation. For Step 2, if $f(p/q) = 0$ then f is divisible, in $\mathbf{Q}[X]$, by $x - p/q$ and hence either f is not irreducible or α is not irrational, contrary to hypothesis. The lower bound in Step 3 follows from the fact that one is the smallest positive integer. The upper bound in Step 4 requires more detailed analysis. Suppose we take the Taylor series expansion of $f(X)$ about α. Since $f(\alpha) = 0$ the first term is zero:

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

Thus

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

Since $|p/q - \alpha| \leq 1$ by hypothesis,

$$|f(p/q)| \leq |p/q-\alpha| \sum_{i=1}^{d} |a_i|. \tag{2}$$

Combining (2) with the lower bound of Step 3 establishes Liouville's Theorem with $c(\alpha) = \min\{1, 1/2b(\alpha)\}$. 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.

Where, in this argument, can improvements be sought? Step 2 can not be improved. Step 3 is extremely unlikely to be sharp in any given case but it is equally unclear how to obtain any quantitative improvement that can be applied with any degree of generality. In Step 4, the only improvement possible is with the two inequalities (1) and (2), neither of which is sharp. But again, like in Step 3, without specific information about α and p/q it is hard to see where significant quantitative improvements can be found. This brings us back to Step 1, the choice of $f(X)$. Making an explicit choice, namely the irreducible polynomial of α over $\mathbf{Q}$, chosen to have relatively prime integer coefficients and positive leading coefficient, has the great advantage of making the Liouville's theorem *effective*, that is the constant $c(\alpha)$ can be explicitly computed. But what if one tries some other polynomial $g(X)$? Might it be possible to obtain a quantitative improvement at the cost, perhaps, of losing some effectivity?

Suppose then that $g(X) \in \mathbf{Q}[X]$. The diophantine method, Step 4 in particular, is based on the fact that $g(\alpha) = 0$ so we may assume this. Step 2 is not a problem as we may simply replace $g(x)$ with $g(X)/(X - p/q)^t$, where t is the order of vanishing of g at p/q. Thus we will assume that $g(p/q) \neq 0$. Step 3 requires that we choose $g(X) \in \mathbf{Z}[X]$ and, to minimize $|g(p/q)|$, it is best to choose g with relatively prime coefficients as we did for f.

In step 4, however, there is perhaps a possibility for improvement with a judicious choice of g. Suppose a is the order of vanishing of g at α. Then we will find, following the argument above, a bound of the form

$$\left| \frac{p}{q} - \alpha \right| \geq \frac{c(\alpha)}{q^{\deg(g)/a}}$$

where $c(\alpha)$ will depend on the size of the coefficients of the Taylor series expansion of g at α. Since g has rational coefficients and vanishes to order a at α, g must be divisible by f^a and in particular $\deg g \geq a \deg f$. So there is no improvement in the exponent d *regardless* of the choice of polynomial g. Is it possible, however, that maybe a clever choice of g might make the constant $c(\alpha)$ smaller? This depends on the upper bound for the coefficients of the Taylor series of f and g. This bound, in turn, depends on $|\alpha|$, over which we have no

control, and $|f|, |g|$, that is the maximum of the size of the coefficients of f and g. Since g vanishes to order a at α we have

$$g(x) = f(x)^a h(x),$$

the factorization holding over the integers since both f and g have been assumed to be primitive. At first sight, this looks promising as maybe there is some cancellation of coefficients so that g has smaller coefficients than f. Unfortunately, the appropriate generalization of Gauss's Lemma ([L] Chapter 3, Proposition 2.1) shows that no significant improvement is possible here, except possibly if f has very small coefficients. Thus, in general nothing is to be gained by making a different choice of approximating polynomial with integer coefficients and so the Liouville theorem appears to be the best possible with these techniques.

Before looking at improvements using an auxilliary function in two ore more variables, first consider work of Thue [T1] which accomplishes more than Liouville using a polynomial of one variable. Thue was able to show that diophantine equations of the form

$$ax^n - by^n = m,$$

where $n \geq 3$, a and b are whole numbers and $m \neq 0$ is an integer, have only finitely many integer solutions (x, y). To see the relationship with Liouville's theorem, dividing both sides of the equation by ay^n we find

$$\left(\frac{x}{y}\right)^n - \frac{b}{a} = \frac{m}{ay^n}.$$

Suppose we let α be the positive real number with $\alpha^n = \frac{b}{a}$. Assuming that x and y are positive, then $\frac{x}{y}$ must be close to α: this is a variant of Step 4 in Liouville's theorem. Note that α is a root of $X^n - \frac{b}{a}$ with multiplicity one as we can see, for example, by differentiating this function with respect to X. Thus we find, arguing as in Step 4 of the proof of Liouville's Theorem, that

$$\left|\frac{x}{y} - \alpha\right| = O\left(\frac{1}{y^n}\right)$$

and $\frac{x}{y}$ is a very good rational approximation of α. So in showing that the equation $ax^n - by^n = m$ only has finitely many solutions, Thue is also showing that α can not be too well approximated with rational numbers of denominator y.

Thue proved finiteness of solutions to $ax^n - by^n = m$ using an auxilliary polynomial in one variable. Formulated in terms of rational approximations, Thue's idea was to take a single good rational approximation $\frac{p}{q}$ to $\alpha = \sqrt[n]{\frac{b}{a}}$ and produce additional good rational approximations. The problem with multiple good rational approximations to a number α is simple: if $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ are both "very close" to α then they are also very close to one another. On the other hand, $\left|\frac{p_1}{q_1} - \frac{p_2}{q_2}\right| \geq \frac{1}{q_1 q_2}$ assuming these rational approximations are distinct. This argument can produce a contradiction if we have some control over the relative sizes of q_1 and q_2 and is related to what has been become well known as a "gap principle:" if $\frac{p_1}{q_1}$ and

$\frac{p_2}{q_2}$ are both good rational approximations of α then one of the denominators must be much larger than the other.

We discuss briefly how Thue created a sequence of good rational approximations from a single approximation $\frac{p}{q}$. In [T1] he uses some very explicit complex polynomials which, in addition, allow him to conclude that the sequence of rational approximations (to an n^{th} root of a rational number) are all distinct. The construction can be made effective and Thue did so ten years later. For a general irrational number α, Thue constructs polynomials $P_N(x)$ and $Q_N(x)$ (with integer coefficients) where N is a large positive integer. The goal is to take a single good rational approximation $\frac{p}{q}$ to α and look at the sequence of approximations given by $P_N(\frac{p}{q})/Q_N(\frac{p}{q})$. Without going through the details of Thue's construction, we can see what properties the polynomials P_N and Q_N should have. First, we want $\frac{P_N(x)}{Q_N(x)}$ to be close to α if x is close to α. Looking at $\frac{P_N(x)}{Q_N(x)} - \alpha$, we want this to have a zero of order N at α as this will govern how small the values of this function are for x close to α. We also need to control the size of the coefficients of P_N and Q_N as well as their degree, a theme which will become more and more important as our constructions become increasingly complex. It turns out that in order for $\frac{P_N(x)}{Q_N(x)} - \alpha$ to have a zero of order N at α the degree needs to be approximately $[\mathbf{Q}(\alpha) : \mathbf{Q}]/2$ and it precisely the 2 in the denominator that gives Thue an improvement on Liouville's theorem.

Once the sequence of good approximations $\frac{a_N}{b_N} = \frac{P_N(p/q)}{Q_N(p/q)}$ has been constructed, Thue focused on showing that $a_N b_{N+1} \neq a_{N+1} b_N$, that is the rational approximations are distinct. To this end, suppose we write

$$S_N(x) = P_N(x)Q'_N(x) - P'_N(x)Q_N(x).$$

After checking that $S_N(x)$ is not identically zero, Thue shows that $\frac{p}{q}$ is not a root of $S_N(x)$ with large multiplicity. To this end, Thue shows that $S_N(x)$ has large order of vanishing at α. Since $S_N(x) \in \mathbf{Z}[x]$ and the degree of $S_N(x)$ is controlled this means that S_N cannot have large order of vanishing at $\frac{p}{q}$. So after taking a small order derivative of S_N we will get $S_N\left(\frac{p}{q}\right) \neq 0$. The desired pair of good rational approximations of α is now found by using the two small order derivatives of P_N and Q_N which are now guaranteed to give different approximations, i.e. approximations satisfying $a_N b_{N+1} \neq a_{N+1} b_N$.

Thue's progress (obtaining an approximation exponent of $\frac{d}{2} + 1 + \epsilon$) outlined in the previous paragraphs can all be formulated in terms of functions of one variable. It can also be thought of as using an auxilliary polynomial $f(X, Y) \in \mathbf{Z}[X, Y]$ in *two* variables which is linear in Y. We would now like to examine in greater detail how using a polynomial in two variables might help to improve the exponent. The natural extension of Loiuville's method to a polynomial in two variables would ask for f to vanish at (α, α) or $(\alpha, 2\alpha)$ or some other point whose coordinates are determined in a simple way by α. We will then find that $|f(p/q, p/q)|$ or $|f(p/q, 2p/q)|$ respectively is very small, Step 4 in Liouville's theorem. Step 3 of the argument will still be valid although it will involve the degree of f in both X and Y. So in principle this argument can still be successful provided Step 2 can be verified.

Ignoring Step 2 for the moment, let us examine what type of quantitative improvement in the Liouville exponent d might be obtained with this new method. For a polynomial $f(X,Y) \in \mathbf{Z}[X,Y]$ of bi–degree d_1, d_2, what is the largest order of vanishing one can find at (α,α)? This is, unfortunately, not an easy question to answer although one can readily find a *lower* bound on the order of vanishing. Since $f(X,Y) \in \mathbf{Z}[X,Y]$, f must vanish at $(\sigma(\alpha),\sigma(\alpha))$ for all embeddings $\sigma : \mathbf{Q}(\alpha) \to \mathbf{C}$. Taking $r = [\mathbf{Q}(\alpha) : \mathbf{Q}]$ we see that the space of polynomials of degree at most (d_1, d_2) has dimension roughly $d_1 d_2$ while a singularity of multiplicity m at r points costs, roughly, $rm^2/2$. Thus one can expect to find a polynomial f, with integer coefficients, of degree at most (d_1, d_2) and order of vanishing roughly

$$\sqrt{\frac{2d_1 d_2}{r}}$$

at (α,α).

Step 4 of the argument will show that $|f(p/q, p/q)|$ grows like $|p/q - \alpha|^{\sqrt{\frac{2d_1 d_2}{r}}}$. *Assuming* we can show that $f(p/q, p/q) \neq 0$ in Step 2, we will find

$$|f(p/q, p/q)| \geq \frac{1}{q^{d_1+d_2}}.$$

Thus we can expect, comparing the bounds from Steps 2 and 4, a diophantine approximation exponent of roughly

$$\sqrt{r}\frac{d_1 + d_2}{\sqrt{2d_1 d_2}}.$$

Keeping in mind that $r = d$ in Liouville's theorem, this bound is quadratic in d but also depends on the extraneous factor $\frac{d_1+d_2}{\sqrt{2d_1 d_2}}$. If d_1 and d_2 are approximately equal then this factor is approximately $\sqrt{2}$ but if $d_1 \gg d_2$ or $d_2 \gg d_1$ then this factor becomes very large. We see that Siegel's exponent of $2\sqrt{d}$, obtained in the 1920's using an approximating polynomial of two variables, is very strong, essentially as strong as can be produced using this method of an approximating polynomial of two variables. Thue's bound, obtained in 1909, of $|p/q - \alpha| < \frac{c(\alpha)}{q^{\frac{d}{2}+1}}$ is not nearly as sharp as Siegel's and is obtained with a very particular auxilliary polynomial of the form $X_2 Q(X_1) - P(X_1)$. In any case, something more subtle is clearly needed here in order to progress toward the Roth exponent of $2 + \epsilon$.

Let us return to Step 2. The problem here is that there is one very simple polynomial, with very small coefficients, which vanishes at (α,α) and its conjugates, namely $X - Y$. Of course nothing interesting will be established by using this simple polynomial but in order to avoid this choice further restrictions are clearly required. One way to avoid a polynomial like $X - Y$ is to specify "unbalanced" degrees d_1, d_2, that is assume that $d_1 \gg d_2$ or $d_1 \ll d_2$. Such a polynomial may still vanish along the diagonal but not to large order and so the bulk of the vanishing at (α,α) will *not* come from factors of the form $X - Y$. Suppose we choose $d_1 \gg d_2$. Then we have just seen in the preceding paragraph that a different method will be needed as the diophantine exponent becomes less sharp when the degrees of the polynomials are sharply different.

One of the keys to reformulating the Liouville method for polynomials of several variables is to consider a modified order of vanishing, which avoids problems with the diagonal, called the *index*. This is essential in order to move beyond the impasse encoutered in the previous two paragraphs.

Definition 2 *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* $(\alpha_1, \ldots, \alpha_m)$*:*

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

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

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

For now, we will be interested in the case $m = 2$. Note that if $d_1 = d_2 = d$ then $\mathrm{ind}_x(f) = \frac{\mathrm{mult}_x(f)}{d}$ so it is nothing other than a scaled version of the multiplicity. Similarly, if d_1 and d_2 are roughly similar in size then the index will be a small perturbation of the scaled multiplicity. If, however, $d_1 \gg d_2$ then the index and the multiplicity measure the local behavior of f very differently. Consider, for example, the case where $(d_1, d_2) = (100, 1)$. The polynomials $P(X) \cdot Y$ and $X^{100} \cdot Q(Y)$, where $P(X)$ is a degree 100 polynomial not vanishing at 0 and $Q(Y)$ is a linear polynomial not vanishing at 0, both have index 1 at $(0, 0)$ while the first has multiplicity 100 and the second multiplicity 1.

What is gained by searching for a polynomial with large *index* at (α, α) as opposed to large multiplicity? The main advantage, from our point of view, has to do with the geometric properties of the index. Suppose a polynomial $f(X, Y) \in \mathbf{Z}[X, Y]$, of degree $(100, 1)$, has index $\frac{1}{10}$ at the point (α, α). Then not only is $f(\alpha, \alpha) = 0$ but

$$\left(\frac{\partial}{\partial X} \right)^k f(\alpha, \alpha) = 0, \;\; 0 \leq k \leq 9.$$

This is a lot of information which one can use to impose strong constraints on what f might look like. On the other hand, it is relatively "cheap" in the sense that it is only 10 linear conditions, in a space of dimension 100, on the polynomial f. Thus when (d_1, d_2) is highly unbalanced, severe geometric constraints can be imposed relatively easily on the auxilliary polynomial f, hopefully generating a contradiction when juxtaposed with the arithmetic constraints.

There is one final issue to be addressed before sketching the basic steps in improving Liouville's theorem by using an auxilliary polynomial in two or more variables. If $f(X, Y)$ has index t at $(p/q, p/q)$ and $f(p/q, p/q) \neq 0$ then what can be said about $|f(p/q, p/q)|$? If we take a Taylor series expansion of f about (α, α) we will find

$$f(X, Y) = \sum_{\frac{i}{d_1} + \frac{j}{d_2} \geq t} a_{ij} (X - \alpha)^i (Y - \alpha)^j \tag{3}$$

The problem here is that for each term in (3) with non–zero coefficient a_{ij}, the size of the corresponding monomial $(X-\alpha)^i(Y-\alpha)^j$, evaluated at $(p/q, p/q)$, will be of the form $|(p/q-\alpha)|^{i+j}$. But this quantity varies substantially for monomials $(X-\alpha)^i(Y-\alpha)^j$ with the same *index*. Consequently, the hypothesis that the index of f at (α,α) is at least t does not provide a good upper bound on the size of $|f(p/q,p/q)|$ even if the size of the coefficients a_{ij} can be controlled. One natural way around this problem would be to choose two *different* good rational aproximations $p_1/q_1, p_2/q_2$ to α. If q_1 and q_2 are chosen so that $q_1^{d_1}$ and $q_2^{d_2}$ are of similar size, then a quick computation shows that all monomials of index exactly t will share a similar upper bound when evaluated at $(p_1/q_1, p_2/q_2)$.

We can now summarize the conclusions of our analysis of Liouville's theorem. The logic leading to the fundamental notion of the index and to the use of several good rational approximating points is as follows:

1. Lioville's Theorem can not be improved, in general, working with a polynomial of one variable; thus we try an approximating polynomial of two variables.

2. We must avoid simple auxilliary polynomials, vanishing at (α,α), such as $X-Y$.

3. Choose an auxilliary polynomial $f(X,Y)$ with unbalanced degree $d_1 \gg d_2$.

4. Use the index to measure order of vanishing in order to take advantage of strong geometrical cosntraints which occur when $d_1 \gg d_2$.

5. Use different approximating points $p_1/q_1, p_2/q_2$, chosen *specifically* so as to translate the index of a polynomial f at (α,α) into a strong upper bound for $|f(p_1/q_1, p_2/q_2)|$, provided of course that $f(p_1/q_1, p_2/q_2) \neq 0$.

As we will see, Roth's theorem requires an auxilliary polynomial in an arbitrary number of variables but the logical steps remain the same. There is one very important point to be made about the structure of this argument. The choice of the *index*, instead of the more classical multiplicity, is far more than a convenience. As we will see later, the geometric properties of the index allow for very sharp estimates for independence of conditions being imposed at the different conjugate points $(\sigma(\alpha),\sigma(\alpha))$. Similar results, in the context of multiplicity, are unknown and are closely related to the famous Nagata conjecture which we will discuss later.

To conclude this section, we present one classical result which indicates that Roth's theorem has the best possible exponent. Construct a sequence of integers inductively as follows: let $a_1 = b_1 = 1$ and $a_{n+1} = a_n + 2b_n$, $b_{n+1} = a_n + b_n$. Then it is a straightforward exercise to show that

$$\begin{aligned} |a_n^2 - 2b_n^2| &= 1, \text{ for all } n, \\ \left|\frac{a_n}{b_n} - \sqrt{2}\right| &< \frac{1}{2b_n^2}, \text{ for all } n. \end{aligned}$$

The rational numbers a_n/b_n are all distinct and they all satisfy the inequality

$$\left|\frac{p}{q} - \sqrt{2}\right| \leq \frac{1}{2q^2} :$$

thus the ϵ in Roth's theorem cannot be taken to be zero. A classic result of Dirichlet states that if α is *any* real quadratic irrationality then $|p/q - \alpha| < \frac{1}{q^2}$ has infinitely many solutions.

EXERCISES

1. Prove Dirichlet's result: if α is *any* real quadratic irrational number then $|p/q - \alpha| < \frac{1}{q^2}$ has infinitely many distinct rational solutions. What if α is a rational number?

2. Suppose for all polynomials in $f(X) \in \mathbf{Z}[X]$ we measure the size of f by

$$|f| = \max_{0 \leq i \leq \deg(f)} \{|a_i|\}$$

where a_i is the coefficient of X^i in $f(X)$.

a. Show that $|\cdot|$ is *not* a norm on $\mathbf{Z}[X]$.

b. Can you quantify how badly $|\cdot|$ fails to be a norm? That is, can you produce constants $c, C \in \mathbf{R}$ so that

$$c|f||g| \leq |fg| \leq C|f||g|?$$

c. Can **b** be used to produce an effective algorithm for factoring polynomials in $\mathbf{Z}[X]$?

3. Here we address the limitations of the multiplicity for measuring the order of vanishing of the auxilliary polynomial $f(X_1, \ldots, X_m)$ in Roth's theorem. Although these limitations with multiplicity do not "prove" the necessity of using the index in Roth's theorem, they point out important differences inherent to other measures of the order of vanishing.

a. The index treats the contribution of each variable equally, giving rise to the expected value of $m/2$ for the index of the "average monomial" in m variables. If instead of the index one measures the order of vanishing of f using multiplicity, what is the corresponding expectation value?

b. With E denoting the expectation value for the multiplicity let V_ϵ be the vector space of all polynomials of degree $(d_1, \ldots, d_m)$ with multiplicity at least $E - \epsilon$ at a point $(x_1, \ldots, x_m)$. Can

$$\frac{\dim(V_\epsilon)}{\prod_{i=1}^m d_i}$$

always be made small, by choosing m sufficiently large, as in the case of the index?

c. If several approximating points p_i/q_i of α are used, then there is no control over the denominators q_i. What impact does this have on the degrees d_i? Is it possible to find an argument (establishing Roth's theorem) using an auxilliary polynomial with big multiplicity at $(\alpha, \ldots, \alpha)$? What goes wrong?

d. Suppose for the auxilliary polynomial in Roth's theorem one chooses f with large multiplicity at $(\alpha, \ldots, \alpha)$ and with degrees d_i all of similar size. If instead of evaluating f at $(p_1/q_1, \ldots, p_m/q_m)$ one uses only one rational approximation and the point $(p/q, \ldots, p/q)$ what goes wrong with the argument?

2 Roth's Theorem: an Overview

Using an auxilliary polynomial in arbitrarily many variables, Roth [R] obtained, in light of Dirichlet's result, the best possible exponent for approximation of algebraic irrationals by rational numbers:

Theorem 3 (Roth's Theorem) *Suppose $\alpha \in \mathbf{R}$ is algebraic and irrational. Then for any $\epsilon > 0$ there are only finitely many solutions to*

$$|p/q - \alpha| \leq \frac{1}{q^{2+\epsilon}}. \tag{4}$$

The basic steps in the proof of Roth's theorem are the same as those in Liouville's theorem, adapted appropriately to the new complexities of the higher dimensional situation:

1. Choose $\epsilon > 0$ and good rational approximations $p_1/q_1, \ldots, p_m/q_m$ for α, that is

$$|p_i/q_i - \alpha| < 1/q_i^{2+\epsilon} \text{ for all } i.$$

Choose a polynomial $f \in \mathbf{Z}[X_1, \ldots, X_m]$ with large index at $(\alpha, \ldots, \alpha)$.

2. Show that $f(p_1/q_1, \ldots, p_m/q_m) \neq 0$.

3. Conclude that $|f(p_1/q_1, \ldots, p_m/q_m)| \geq \frac{1}{\prod_{i=1}^m q_i^{d_i}}$.

4. Find an upper bound $|f(p_1/q_1, \ldots, p_m/q_m)| \leq C \cdot d(p_1/q_1, \ldots, p_m/q_m)$ for some constant C which will depend on the coefficients and the degree of f and here d represents some function of the distance of the rational numbers p_i/q_i to α.

If the upper bound in Step 4 is good enough, then it will violate the lower bound of Step 3 and produce the contradiction establishing Roth's theorem. There are many serious difficulties to be overcome in order to make this sketch rigorous, the most glaring one, and the one which will receive most of our attention, being Step 2. Here is a list of smaller issues which also must be settled before moving forward:

i. What properties are to be imposed on the points p_i/q_i and the polynomial f in Step 1?

ii. How and why does the m of Step 1 depend on the choice of ϵ in Roth's Theorem?

iii. How does one make the upper bound in Step 4 sharp enough to contradict the lower bound of Step 3?

In this section, we will address **i**, **ii**, and **iii** and then in the next two sections we will tackle the very difficult Step 2.

Motivated by the desire, discussed in §1, to choose the polynomial f with rapidly decreasing degrees $d_1 \gg d_2 \gg \ldots \gg d_m$ in $X_1, \ldots, X_m$, and also by the desire for $q_i^{d_i}$ to be roughly proportional (in order to find a good upper bound in Step 4), we must choose the degrees so that they are proportional to the logarithms of the denominators of the rational approximations:

$$d_i \sim \frac{N}{\log q_i}$$

where N is some large positive number and $q_1 \ll q_2 \ll \ldots \ll q_m$. We will, somewhat abusively, write q for the quantities $q_i^{d_i}$ which, though not equal, will be chosen to be close in value to one another. In order to calculate the upper bound in Step 4, we need to know the index of f at $(\alpha, \ldots, \alpha)$. Suppose we call this index t_m and leave for a moment the calculation of this number. Let us also fix an $\epsilon > 0$ for which we will try to establish Roth's Theorem. If $I = (i_1, \ldots, i_m)$ is a multi–index and

$$M_I = (x_1 - \alpha)^{i_1} \cdot \ldots \cdot (x_m - \alpha)^{i_m}$$

is a monomial of index at least t_m at $(\alpha, \ldots, \alpha)$ and $|p_i/q_i - \alpha| < 1/q_i^{2+\epsilon}$ for all i then a quick calculation shows that

$$|M_I(p_1/q_1, \ldots, p_m/q_m)| \leq \frac{1}{q^{(2+\epsilon)t_m}}.$$

Write

$$f = \sum_{0 \leq I \leq d} a_I M_I$$

where d is the multi–degree of f. We may also assume without loss of generality that the coefficients of f are relatively prime. The total number of monomials in this expansion is $(d_1+1)(d_2+1)\ldots(d_m+1)$. If we let $H(f)$ denote the maximum of the size of the coefficients $|a_I|$ then we have the upper bound

$$|f(p_1/q_1, \ldots, p_m/q_m)| \leq C_1 H(f) \frac{1}{q^{(2+\epsilon)t_m}} : \tag{5}$$

here C_1 is a bound for the number of monomials with non–zero coefficeint a_I and hence $C_1 \leq \prod_{i=1}^m (d_i + 1)$. On the other hand, Step 3 will give a lower bound, recalling that $q \sim q_i^{d_i}$ for each i,

$$|f(p_1/q_1, \ldots, p_m/q_m)| \geq \frac{1}{q^m}. \tag{6}$$

In order to find a contradiction when comparing (5) and (6) we clearly will need

$$t_m \geq \frac{m}{2+\epsilon},$$

and then in addition we will need to show that C_1 and $H(f)$ are not so large as to destroy the inequality.

Both the lower bound on t_m and the upper bound on $H(f)$ are counting problems, though the latter leads to some subtle and interesting questions. For t_m the most elegant approach is that of Faltings and Wüstholz [FW] who resolve this computation with the law of large numbers. The idea is the following: a monomial M_I is an m–tuple $(i_1, \ldots, i_m)$ with $0 \leq i_j \leq d_j$. For each j, roughly half of the possible i_j's satisfy $\frac{i_j}{d_j} \geq \frac{1}{2}$ while the other half satisfy $\frac{i_j}{d_j} < \frac{1}{2}$. Moreover, the possible values of i_j/d_j, namely $\frac{0}{d_j}, \frac{1}{d_j}, \ldots, \frac{d_j}{d_j}$, are evenly distributed in the interval $[0,1]$. Thus for a "randomly chosen" m–tuple $(i_1, \ldots, i_m)$ the expected value of

$$\frac{i_1}{d_1} + \ldots + \frac{i_m}{d_m}$$

should be $\frac{m}{2}$. The law of large numbers says, in this situation, that as m grows, the values $\frac{i_1}{d_1} + \ldots + \frac{i_m}{d_m}$ are normally distributed about this expectation value. In particular, given $\delta, \epsilon > 0$ there exists $m > 0$ so that

$$\left| \left\{ (i_1, \ldots, i_m) : \frac{i_1}{d_1} + \ldots \frac{i_m}{d_m} < \frac{m(1-\epsilon)}{2} \right\} \right| < \delta \prod_{i=1}^{m} (d_i + 1).$$

In other words, for m sufficiently large, depnding on ϵ and δ, there will be plenty of polynomials of degree $(d_1, \ldots, d_m)$ with index at least $m(1-\epsilon)/2$ at $(\alpha, \ldots, \alpha)$ and its conjugates. Note that as ϵ approaches zero, m will approach infinity. Note too that it is not possible to take $\epsilon = 0$ as the cost of imposing index $m/2$ at a point is always high (essentially half of the total number of conditions available) *regardless* of how large m is.

We next turn to bounding $H(f)$ which requires Siegel's lemma which we will first state in its classical form applied to homogeneous systems of linear equations with integer coefficients.

Lemma 4 (Siegel's Lemma) *Consider a system of m linear equations in n unknowns, with $m < n$:*

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

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 $(x_1, \ldots, x_n) \in \mathbf{Z}^n$ to the system of equations with $|x_i| < 1 + (nA)^{m/(n-m)}$ for all i.

The proof of Lemma 4 is elementary, using the pigeon–hole principle. For fixed $C > 0$ let

$$B_n(C) = \{(b_1, \ldots, b_n) \in \mathbf{Z}^n : |b_i| \leq C \text{ for all } i\}.$$

Let $L_i(X) = a_{i1}X_1 + \ldots a_{in}X_n$. For each $y \in B_n(C)$ we have $|L_i(y)| \leq AC$. The vector $(L_1(y), \ldots, L_m(y))$ is in $\mathbf{Z}^m$ by hypothesis and also lies in $B_m(AC)$. Comparing the total number of integer points in $B_n(C)$ with the number of integer points in $B_m(AC)$ and using the hypothesis that $n > m$, it follows that for C sufficiently large, there must be points $y_1 \neq y_2 \in B_n(C)$ satisfying $L_i(y_1) = L_i(y_2)$ for all i. The integer point $y_1 - y_2$ is the desired point and Siegel's Lemma is established by finding a value of C large enough so that $|B_n(C)| > |B_m(AC)|$.

In our setting we wish to apply Siegel's Lemma to

$$f(X_1, \ldots, X_m) = \sum_{0 \leq I \leq d} a_I X^I$$

viewing the a_I as *variables*, where the equations will come from the hypothesis that f has large index at $(\alpha, \ldots, \alpha)$: this condition is equivalent to asking for

$$D(f)(\alpha, \ldots, \alpha) = 0$$

where $D = \frac{\partial^{i_1}}{\partial X^{i_1}} \ldots \frac{\partial^{im}}{\partial X_m^{im}}$ and $\frac{i_1}{d_1} + \ldots + \frac{i_m}{d_m} < M(1-\epsilon)/2$. The problem which arises here is that this system of equations in the variables a_I does not have integer coefficients but rather coefficients in $K = \mathbf{Q}(\alpha)$. Suppose we write the set of equations $D(f)(\alpha, \ldots, \alpha) = 0$ as

$$\sum_I x_{ID} a_I = 0, \;\; x_{ID} \in K, \;\; \text{index of } D < M(1-\epsilon)/2; \tag{7}$$

here the a_I are the coefficients of the general polynomial of degree $(d_1, \ldots, d_m)$ in m variables and D represents a differential operator. Letting $d = [K : \mathbf{Q}]$, one can then choose a basis $v_1, \ldots, v_d$ for K over $\mathbf{Q}$ and the system of equations (7) becomes

$$\sum_I b_{IDl} v_l a_I = 0, \;\; b_{IDl} \in \mathbf{Q}, \;\; \text{index of } D < M(1-\epsilon)/2, \;\; 1 \leq l \leq d. \tag{8}$$

Since we are looking for an integer solution $\{a_I\}$ to (7) and since the v_l are linearly independent over $\mathbf{Q}$ this is equivalent to solving

$$\sum_I b_{IDl} a_I = 0, \;\; \text{index of } D < M(1-\epsilon)/2, \;\; 1 \leq l \leq d. \tag{9}$$

Siegel's Lemma, applied to the system (9), will yield a small integer solution to our original set of equations (7). The only added complication is that we now must deal with rational coefficients instead of integer coefficients since the b_{IDl} are not necessarily integers. Moreover, the solution vector $\{a_I\}$ to (7) can be scaled by a non–zero constant. Therefore the solution vector we are looking for is appropriately viewed as a point in projective space. Thus we need to define the size of a point in projective space and this is accomplished via the simplest instance of the theory of heights.

A thorough discussion with proofs can be found in Chapters 3 and 4 of [L]. Suppose

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

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

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

With these definitions in place we can state the version of Siegel's lemma which we will use (see [HS] Lemma D.4.2):

Lemma 5 *Suppose $\{L_i(X) = 0\}_{i=1}^m$ is a system of linear equations in n unknowns, with coefficients in a number field K. Let v be the vector of coefficients of the equations $\{L_i(X) = 0\}$, viewed as a point in projective space, and let $H(v) = A$. Suppose in addition that $n - dm > 0$. There is a non–zero solution x to the system of equations $L_i(X) = 0$ satisfying*

$$H(x) \leq (nA)^{\frac{dm}{n-dm}}.$$

In our situation, it remains to put together our estimates in Steps 3 and 4 to see if we find the desired contradiction. Using the law of large numbers, choose $m > 0$ so that

$$[K : \mathbf{Q}] \left| \left\{ (i_1, \ldots, i_m) : \frac{i_1}{d_1} + \ldots \frac{i_m}{d_m} < \frac{m(1-\epsilon/2)}{2} \right\} \right| < \frac{1}{2} \prod_{i=1}^m (d_i + 1) : \tag{10}$$

note that the choice of $\frac{1}{2}$ here on the right hand side is arbitrary: it is essential to have some constant less than one. Setting

$$t_m = \frac{m(1-\epsilon/2)}{2}$$

there will exist, by Lemma 5, a polynomial $f(X) \in \mathbf{Z}[X]$ with index at least t_m at $(\alpha, \ldots, \alpha)$ and height satisfying

$$H(f) \leq nA. \tag{11}$$

Note that by hypothesis $[K : \mathbf{Q}]m/(n - [K : \mathbf{Q}]m) < 1$ so that the conclusion of Lemma 5 is particularly simple. Returning then to our basic estimates, (5) and (6), we have to show, using the assumption that $t_m \geq \frac{m(1-\epsilon/2)}{2}$ and (11), that

$$\left(\prod_{i=1}^m (d_i + 1) \right) nA \leq q^{m\epsilon/4}.$$

Recalling that $q \sim q_i^{d_i}$ the first factor $\prod_{i=1}^m (d_i + 1)$ is clearly an order of magnitude smaller than $q^{m\epsilon/4}$, provided that the q_i are not too small. This also takes care of the second factor

n. As for A, we need to bound the coefficients of the system of equations (9). This is routine (the coefficients of $Df(\alpha, \ldots, \alpha)$ will depend only on the size of α and the size of the coefficients introduced when differentiating and there is an additional constant which will depend on the choice of basis for K over $\mathbf{Q}$) and it is shown in [HS] Proposition D.3 that there exists a constant B, depending only on K, so that one can take A in Lemma 5 to be $B^{d_1+\ldots+d_m}$. This can certainly be made less than $q^{m\epsilon/4}$ for d_i suitably large and this concludes the proof of Roth's Theorem, *assuming* of course that our polynomial f of small height guarantied by Lemma 5 does *not* vanish at $(\alpha, \ldots, \alpha)$.

EXERCISES

1. For those who have a familiarity with probability theory, state and prove the desired consequence of the law of large numbers which justifies (10).

2. Assuming that we can construct an auxilliary polynomial $f(X) \in \mathbf{Z}[X]$ of suitably small height which does not vanish at $(p_1/q_1, \ldots, p_m/q_m)$ state clearly how the different variables in the proof of Roth's theorem need to be chosen: note that the only given for Roth's theorem is ϵ while the quantities which are chosen are $d_1, \ldots, d_m$, $p_1/q_1, \ldots, p_m/q_m$, and $f(X)$. Indicate the order in which the chosen quantities are determined and the mathematical constraints imposed upon them in order to lead to a contradiction.

3. Suppose $t \leq 1$ and f is a general polynomial in m variables of multi–degree d and index t at ζ. What is the multiplicity of f at ζ? What if $t > 1$? If g is a general polynomial in m variables of multi–degree d and multiplicity m what is the index of g at ζ?

3 Controlling the index: Roth's Lemma

The first, and in some sense simplest, method of establishing Roth's Theorem is via Roth's Lemma. It turns out that if the right hypotheses are put on the rational approximating points p_i/q_i, then f can NOT have large index at $(p_1/q_1, \ldots, p_m/q_m)$. Since, when $m \gg 0$, imposing index $m(1-\epsilon)/2$ at $(\alpha, \ldots, \alpha)$ and its conjugates uses up less than $1/2$ of the conditions available on the space of polynomials of degree $d_1, \ldots, d_m$, there certainly are many polynomials which have large index at $(p_1/q_1, \ldots, p_m/q_m)$ as well as $(\alpha, \ldots, \alpha)$. Thus Roth's argument establishing that the index of f at $(p_1/q_1, \ldots, p_m/q_m)$ is small cannot be geometric in nature. Indeed, as we shall see, it is an arithmetic argument.

There are two key ideas behind Roth's Lemma. To see the first, suppose $f(X) \in \mathbf{Z}[X]$ is primitive and vanishes at p/q. The Gauss Lemma says that $f(X)$ is divisible by $pX - q$ and in particular the leading term of $f(X)$ is divisible by p while the constant term is divisible by q. Hence we have

$$H(f) \geq \max\{p, q\}.$$

Thus the minimum height possible for all polynomials vanishing at p/q is exactly $H(p/q, 1)$, the height of the point p/q viewed as a point on the projective line. The basic idea behind

Roth's lemma is to insist that the auxilliary polynomial $f \in \mathbf{Z}[X_1, \ldots, X_m]$ have small height. In one variable, this polynomial would not be able to vanish at p/q, for q sufficiently large. In several variables, the situation is more complicated but Roth found a very clever way to procede by induction on the number of variables.

Let us consider for the moment the two variable case. Suppose $f(X,Y) \in \mathbf{Z}[X,Y]$ is a polynomial with "small" coefficients. We wish to show that the index at $(p_1/q_1, p_2/q_2)$ must be small. Consider

$$f(X,Y), \frac{\partial}{\partial X} f(X,Y), \ldots, \frac{\partial^{d_2}}{\partial X^{d_2}} f(X,Y) \tag{12}$$

and let I be the ideal of polynomials generated by these derivatives of f. Suppose $Z(I) \subset \mathbf{C}^2$ is the zero set of I. Then $\dim(Z(I)) \leq 1$. If $(p_1/q_1, p_2/q_2)$ is not contained in $Z(I)$ this means that the index of f at $(p_1/q_1, p_2/q_2)$ is small which is what we hope to show. Thus we may suppose $(p_1/q_1, p_2/q_2) \in Z(I)$. If there is a curve $C \subset Z(I)$ we claim that C is either of the form $\{P_1\} \times \mathbf{C}$ or of the form $\mathbf{C} \times \{P_2\}$. If not, then the irreducible equation $g(X,Y)$ defining C has degree at least one in both X and Y. Consequently, the hypothesis entails that $g(X,Y)^{d_2+1}$ divides $f(X,Y)$ which is impossible. Thus if $(p_1/q_1, p_2/q_2) \in Z(I)$ it must be an *isolated* point of $Z(I)$: if $f(X,Y)$ were divisible by $X - p_1/q_1$ or $Y - p_2/q_2$, we could simply remove or replace these extraneous factors which do not contribute toward the index of f at (α, α). The ideal I is generated by the polynomial f and small order derivatives of f. Consequently, if f has small coefficients then all of the generators of I have small coefficients. Thus we need some type of "arithmetic" Bézout theorem in order to impose a bound on the height of a point contained in an intersection of polynomials of small height. With this, we could hope to obtain a contradiction and conclude, as desired, that $(p_1/q_1, p_2/q_2)$ is not in $Z(I)$. This arithmetic Bézout approach is exactly that of Faltings' arithmetic product theorem which we will come to in a moment.

Roth bounds the index of the auxilliary polynomial f at $(p_1/q_1, \ldots, p_m/q_m)$ by a very ingenious use of Wronskian determinants. These should be familiar to many from the elementary theory of differential equations where they play a similar role. In particular it is well known that if $f_1, \ldots, f_n$ are rational functions of a single variable x then the *Wronskian determinant*

$$M = \begin{bmatrix} f_1 & \cdots & f_n \\ \frac{df_1}{dx} & \cdots & \frac{df_n}{dx} \\ \vdots & & \vdots \\ \frac{d^{n-1}f_1}{dx^{n-1}} & \cdots & \frac{d^{n-1}f_n}{dx^{n-1}} \end{bmatrix}.$$

is non–zero if and only if the functions $f_1, \ldots, f_n$ are linearly independent. Roth uses a generalization of this result to functions of several variables:

Lemma 6 *Suppose k is a field of characteristic zero and $f_1, \ldots, f_n \in k(x_1, \ldots, x_r)$. Then $f_1, \ldots, f_n$ are linearly independent over k if and only if there exist differential operators D_i in the variables $x_1, \ldots, x_r$, of order at most $i-1$, so that*

$$\det[D_i(f_j)]_{1 \leq i,j \leq n} \neq 0.$$

The proof of Lemma 6 can be found in [HS] Lemma D.6.1 or [L] Chapter 7, Proposition 6.1.

Lemma 6 is used to find an upper bound on the index of f at $(p_1/q_1, \ldots, p_m/q_m)$ in the following fashion. Write

$$f(X_1, \ldots, X_m) = \sum_{i=1}^{r} g_i(X_1, \ldots, X_{m-1}) h_i(X_m)$$

where r is *minimal* for all such possible expressions: this hypothesis forces the $\{f_i\}$ and $\{g_i\}$ to be linearly independent over k so that Lemma 6 applies. Note that $r \leq d_m + 1$ since one can take $h_i(X_m) = X_m^i$ for $0 \leq i \leq d_m$. By Lemma 6, there are differential operators D_i in $X_1, \ldots, X_{m-1}$ of order at most $i-1$, for $1 \leq i \leq r$, and a differential operator $E_j = \partial^{a_j}/\partial X_m^{a_j}$ of order $a_j \leq j-1$ for $1 \leq j \leq r$ so that

$$\begin{aligned} \det[D_i(g_j)]_{1 \leq i,j \leq r} &\neq 0, \\ \det[E_j(h_k)]_{1 \leq j,k \leq r} &\neq 0. \end{aligned}$$

In fact, the differential operator E_j can be replaced by $\frac{1}{j!}E_j$ so that the coefficients of $E_j(h_k)$ do not grow as fast. We claim that

$$\det[D_i(g_j)]_{1 \leq i,j \leq r} \cdot \det[E_j(h_k)]_{1 \leq j,k \leq r} = \det[D_i E_j(f)]_{1 \leq i,j \leq r}. \tag{13}$$

To see why (13) holds, note that $D_i(h_k) = 0$ for all i, k and similarly $E_j(g_i) = 0$ for all i, j. Thus

$$D_i E_j(f) = \sum_{k=1}^{r} D_i(g_k) E_j(h_k).$$

Let $g(X) = \det[D_i E_j(f)]$. We know, from (13), that $g(X)$ factors into a polynomial in X_m and a polynomial in $X_1, \ldots, X_{m-1}$. This factorization then allows one to use induction on the number of variables in order to control the index. It remains, of course, to *compare* the height and index of f at $(p_1/q_1, \ldots, p_m/q_m)$ with the height and index of g at $(p_1/q_1, \ldots, p_m/q_m)$. For the height, this is routine because g has been defined in terms of f via the elementary operations of addition, multipliciation, and differentiation. More serious is the problem of estimating the index of g at $(p_1/q_1, \ldots, p_m/q_m)$. In [L] Lemma 8.1, it is established that the index of f at $(p_1/q_1, \ldots, p_m/q_m)$ is bounded above in terms of the index of g at $(p_1/q_1, \ldots, p_m/q_m)$ and the rate at which the d_i are decreasing:

Lemma 7 (Roth's Lemma) *Suppose that $d_i/d_{i+1} \geq 10N$ for $1 \leq i \leq m-1$ and that $d_m \geq 10N$ where d_i is the degree of f in X_i and $N > 1$. Then*

$$\min\left\{\mathrm{ind}_{(p_1/q_1, \ldots, p_m/q_m)}(f), (\mathrm{ind}_{(p_1/q_1, \ldots, p_m/q_m)}(f))^2\right\} \leq \frac{4}{r}\mathrm{ind}(g) + \delta \sum_{i=1}^{m} d_i.$$

Note that the index on the right hand side of the inequality in Lemma 7 is computed relative to the degrees $(d_1, \ldots, d_m)$ of f; this fact explains the r in the denominator of the $\mathrm{ind}(g)$ term because g has been obtained by taking a sum of products of r terms involving derivatives of f. Lemma 7 is one version of *Roth's Lemma*, taken from [L]. The proof of Lemma 7 involves basic calculation and some clever manipulation of the terms in the determinant defining $g(X)$. Since $g(X)$ is a sum of terms, one uses the basic fact that the index of a sum is *at least* the minimum of the indices of the summands: this is of course why the index of g appears on the larger side of the inequality in Lemma 7. Each individual term in the determinant defining $g(X)$ is a product of derivatives of f: these derivatives in general will decrease the index. There are two types of derivatives which occur; first, those in $X_1, \ldots, X_{m-1}$ which have very little impact on the index of f since by hypothesis the derivitives are of order at most $d_m + 1$ and $d_m \ll d_i$ for all $i < m$. These derivatives contribute the $\delta \sum d_i$ term in Lemma 7. Next there are the derivatives in X_m and these potentially have a large impact on the index and it is because of these that the calculations involved in proving Lemma 7 are involved.

We now turn to Faltings' product theorem which gives a much cleaner proof of Roth's Lemma but at the cost of developing some arithmetic intersection theory. For Roth's theorem, the technically difficult parts of the theory are not necessary in order to *state* the arithmetic product theorem since in this situation only the heights of points on the projective line need to be defined. We begin by stating the geometric version the product theorem, [F1] Theorem 3.1:

Theorem 8 (Geometric Product Theorem) *Let $n_1, \ldots, n_m$ be positive integers and 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 $\sigma \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. Let $Z_\alpha(\sigma)$ denote the set of points in $\mathbf{P}$ where the index of σ, measured relative to the multi–degree d, is at least σ. Given $\delta > 0$ there exists $r > 0$ so that if

$$\frac{d_1}{d_2}, \ldots, \frac{d_{m-1}}{d_m} \geq r$$

then for any $\alpha \geq 0$ any irreducible component $Z \subset Z_\alpha(\sigma)$ which is also an irreducible component of $Z_{\alpha+\delta}(\sigma)$ is a product $Z_1 \times \ldots \times Z_m \subset \mathbf{P}$.

In order to prove Roth's theorem, we will only need to work on a product of projective lines and we will sketch a proof of the Geometric Product Theorem in this setting in the next section. For this section, it is the arithmetic version of the product theorem, [F1] Theorem 3.3, that we need and this can be stated, in our simple setting, as follows

Theorem 9 (Arithmetic Product Theorem) *With hypotheses and notation as in the geometric product theorem assume in addition that $n_1 = \ldots = n_m = 1$ and that σ is defined over $\mathbf{Q}$. Then the product subvariety Z of Theorem 8 can be selected so that*

$$\sum_{\dim(Z_i)=0} d_i h(Z_i) \leq c_1 h(\sigma) + c_2 \sum_{i=1}^{m} d_i$$

where c_1 and c_2 are constants which depend only on δ. Here $h(Z_i) = \log H(Z_i)$ is the logarithmic Weil height of the point $Z_i \in \mathbf{P}^1$ and $h(\sigma)$ is the logarithmic height of the projective vector of coefficients of σ.

The arithmetic product theorem allows for a very simple proof of Roth's lemma. Let $\epsilon > 0$ be given for Roth's theorem. Choose $m > 0$ so that Lemma 5 applies, giving us a poynomial $f(X_1, \ldots, X_m) \in \mathbf{Z}[X_1, \ldots, X_m]$ with index at least $M(1 - \epsilon/2)/2$ at $(\alpha, \ldots, \alpha)$ and $H(f) \leq B^{d_1 + \ldots + d_m}$. Note that B here depends only on α as long as m is chosen sufficiently large so that (10) holds. We wish to establish a version of Roth's lemma, bounding the index of f at $(p_1/q_1, \ldots, p_m/q_m)$. Suppose $\delta > 0$ and suppose we wish to show that

$$\mathrm{ind}_{(p_1/q_1, \ldots, p_m/q_m)}(f) < \delta.$$

Choose r in the product theorem, depending on δ, sufficiently large to guarantee that if $\mathrm{ind}_{(p_1/q_1, \ldots, p_m/q_m)}(f) \geq \delta$ then there is a proper product subvariety $Z \subset \mathbf{P}$ so that Z contains $(p_1/q_1, \ldots, p_m/q_m)$ and is contained in the zero locus of f. Choosing i so that that $Z_i \subset \mathbf{P}^1$ is proper and hence $Z_i = \{p_i/q_i\}$, we have, from the arithmetic product theorem,

$$d_i h(p_i/q_i) \leq c_1 h(\sigma) + c_2 \sum_{i=1}^{m} d_i. \tag{14}$$

Recall that d_i has been chosen to be approximately $N/\log(q_i)$ for some large positive number N. Thus the left hand side of (14) is of size roughly N. Since c_2 depends only on the choice of δ, as long as p_m/q_m has sufficiently large height relative to c_2, the second term on the right hand side of (14) which is roughly $c_2 \sum_{i=1}^{m} N/\log(q_i)$ will be smaller than the left hand side of (14). We have already seen that the other term on the right hand side of (14), $c_1 h(\sigma)$, is of size $c_1 \log B \sum d_i$, by Lemma 5: since c_1 and B depend only on the choice of α and δ, $d_i h(p_i/q_i) \gg c_1 \log B \sum d_i$, again provided $q_m \gg 0$. Thus for any $\delta > 0$ as long as

$$q_1 \gg q_2 \gg \ldots \gg q_m \gg 0$$

with implied constants depending only on δ, the index of the polynomial f at $(p_1/q_1, \ldots, p_m/q_m)$ given in Lemma 5 is less than δ.

We now turn to proving Roth's theorem using the auxilliary polynomial f with index less than δ at $(p_1/q_1, \ldots, p_m/q_m)$. Choose a differential operator D of order less than δ so that $Df(p_1/q_1, \ldots, p_m/q_m) \neq 0$. Suppose the differential operator D increases the coefficients of f by a factor of $c_1(\delta)$ and decreases the index of f at $(\alpha, \ldots, \alpha)$ by $c_2(\delta) \leq \delta$. Thus (11), becomes, after differentiation,

$$h(D(f)) \leq c_1(\delta) n A \tag{15}$$

and for the index of $D(f)$ we have

$$\mathrm{ind}_{(\alpha, \ldots, \alpha)}(D(f)) \geq \frac{m(1 - \epsilon/2)}{2} - \delta. \tag{16}$$

The argument given after (11) to prove Roth's theorem under the assumption that

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

can now be applied to $D(f)$ for which we have guaranteed the non–vanishing condition. At this point, if we choose $\delta < m\epsilon/8$ then in order to derive a contradiction from (15) and (16) we need to show that

$$c_1(\delta) < q^{m\epsilon/8}.$$

It is shown in [HS] Lemma D.3.1 that one can take $c_1(\delta) < 2^{d_1+\ldots d_m}$ provided the differential operator D is appropriately scaled so that the coefficients do not increase too rapidly, that is

$$D = \frac{1}{i_1! \ldots i_m!} \frac{\partial^{i_1+\ldots+i_m}}{\partial x_1^{i_1} \ldots \partial x_m^{i_m}}.$$

This differential operator preserves the integer coefficients of f while slowing down the growth of coefficients. As long as the q_i are all sufficiently large, $2^{d_1+\ldots+d_m}$ is certainly less than $q^{m\epsilon/8}$ and so this concludes the proof of Roth's Theorem.

EXERCISES

1. For this problem we will study the height of polynomials and their vanishing at the point $(1, 100) \in \mathbf{A}^1 \times \mathbf{A}^1$.

a. Find the linear polynomials of smallest height which generate the ideal of all polynomials vanishing at $(1, 100)$. What is the relationship between this and the height of the point $(1, 100)$?

b. Can $(1, 100)$ be cut out set theoretically by polynomials of small height, say height one?

c. What happens to the answer to **b.** if the degrees of the polynomials are suitably bounded?

2. This problem and the next seek to generalize the product theorem to situations where there is not necessarily an inherent product structure. Suppose A is an abelian surface, $0 \neq v \in T_0(A)$ a tangent vector, and D_v the derivation of $\mathcal{O}_A$ along the translation invariant vector field associated to v. Let L be a line bundle on A and $0 \neq s \in H^0(A, L)$. Suppose $x \in A$ and define the order of vanishing of s along v at x to be the *smallest* integer k such that $D_v^{(k)}(s)(x) \neq 0$: if no such k exists then the order of vanishing is infinite.

a. Show that s vanishes to infinite order along v at x if and only if $Z(s)$ contains a translate of an elliptic curve E through x so that $T_0(E)$ is generated by v.

b. Can you give an effective upper bound, in terms of L and A, for the order of vanishing of s along v at x *assuming* that this order is not infinite?

3. Problem **2** can be generalized in many directions. The interested reader is invited to formulate and prove results in the following directions:

a. The abelian variety A can be replaced by an equivariant compactification X of a commutative group variety G and v with a subspace of the tangent space $T_0(G)$. This leads to the multiplicity estimates which abound in transcendence theory.

b. If X is a smooth projective variety and $V \subset T_x(X)$ a subspace of the tangent space at a point x, one can again ask to bound the order of vanishing of a section $\sigma \in H^0(X, L)$ along V at x. In particular, one can seek a criterion for when this order can be infinite.

4 Controlling the index: Dyson's Lemma

Both Roth's original proof of Roth's lemma and the newer proof using the arithmetic product theorem exploit the fact that the point $x = (p_1/q_1, \ldots, p_m/q_m)$ is a rational point. This puts significant constraints on the order of vanishing of the polynomial $f(X_1, \ldots, X_m)$ at x. Essentially, this is for the simple reason that the *arithmetic complexity* of x, as measured by the height $h(x)$, forces f to either have large height or to cut out some very particular subvariety of $\mathbf{C}^m$ containing $(p_1/q_1, \ldots, p_m/q_m)$. Once one takes into account that the same argument can be applied to small order derivatives of f the conclusion ends up beingthat it is not possible to find a polynomial f with rapidly decreasing degrees, small coefficients, and large index at x. In this argument, the ultimate contradiction is an arithmetic one: if f does exist with small height and large index at x then it is possible to cut out one of the coordinates of x with polynomials whose height is too small relative to the coordinates of x.

The Dyson Lemma approach to Roth's theorem is more powerful in the sense that it is irrelevant that the point x where the polynomial f is to have small index has rational coordinates. It is also more powerful in the sense that the argument using Roth's Lemma requires approximating points with denominator q very large. For Dyson's Lemma, this is of no importance and hence this approach can give interesting information as soon as there is a single very good rational approximation to the irrational α. In particular, for effective diophantine approximation Dyson's Lemma is the preferred approach [B1, B2, BC, BPV].

To state Dyson's lemma, we need a few preliminary definitions.

Definition 10 *Let* $I^m = \{(\xi_\nu) \in \mathbf{R}^m : 0 \leq \xi_\nu \leq 1\}$ *and let* $\mathrm{Vol}(t)$ *denote the volume of*

$$\left\{(\xi_\nu) \in I^m \ : \ \sum_{\nu=1}^{m} \xi_\nu \leq t\right\}.$$

Thus $\mathrm{Vol}(t)$ represents the approximate cost of imposing index t at a point on polynomials of multi–degree $(d_1, \ldots, d_m)$. If the degrees d_i are replaced by nd_i and n is allowed to approach infinity then one can remove the qualifier "approximate." Recall that

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

with m factors. We again set $d = (d_1, \ldots, d_m)$ for an m–tuple of positive integers d_i.

Theorem 11 *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 *(here* π *is the projection to the* j^{th} *factor). Let* $t_i = \text{ind}_{\zeta_i}(s)$ *and let*

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

Then

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

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

If $d_1 \gg d_2 \gg \ldots \gg d_m$, as we saw was the case in our previous proof of Roth's theorem, then $m\delta$ is small so in this case Theorem 11 says that the conditions to impose index t_i at ζ_i are nearly independent inside $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$. Indeed if they were far from being independent then one would be able to impose much larger index than expected by a dimension count. Theorem 11 says that one will only be able to excede the index guaranteed from dimension counting by $m\delta$.

Before addressing the proof of Theorem 11 we first discuss how to derive Roth's theorem. There are two methods, the first following [EV] and the second following [F1]. We begin by outlining the argument of [EV] §9. Suppose rational approximating points p_i/q_i are chosen so that $q_1 \ll q_2 \ll \ldots \ll q_m$. Thus δ in Theorem 11, which measures the maximum of the q_i/q_{i+1}, is small. One then chooses $m \gg 0$ and an index $M(1-\beta)/2$ so that

i. There is a polynomial $P \in \mathbf{Z}[X_1, \ldots, X_m]$ with index $M(1-\beta)/2$ at $(\alpha, \ldots, \alpha)$; here β is chosen small so that

$$[\mathbf{Q}(\alpha) : \mathbf{Q}]\text{Vol}(\beta) = 1 - \gamma$$

for some very small positive real number γ.

ii. The height of P can be bounded in terms of γ.

According to Theorem 11, $\text{ind}_{(p_1/q_1, \ldots, p_m/q_m)}(P)$ is bounded above by some constant $c(\gamma, \delta)$. At this point, in order to prove Roth's theorem, what is necessary is some accounting work. In particular, β can be made very small as long as m is chosen sufficiently large and the δ of Theorem 11 can also be chosen very small as long as the q_i are sufficiently rapidly increasing. The place where one has to be careful is in the dependence of $H(P)$ on γ because of course $H(P)$ will grow as γ approaches zero. The proof of Roth's theorem then follows the sketch outlined in §3. The only difference is in the construction of the auxilliary polynomial. In §3, the polynomial f could have, in principle, had large index at $(p_1/q_1, \ldots, p_m/q_m)$ whereas the polynomial P constructed with Dyson's lemma is known to have small index at $(p_1/q_1, \ldots, p_m/q_m)$.

The second method for deriving Roth's theorem from Dyson's lemma follows Faltings [F1]. Since there is, by the argument of [EV], a polynomial with large index at $(\alpha, \ldots, \alpha)$ and small index at $(p_1/q_1, \ldots, p_m/q_m)$ then of course one can decrease the index at $(\alpha, \ldots, \alpha)$ a little bit

and still find such a polynomial. The advantage to decreasing the index by a small amount is that this removes the fact that $[\mathbf{Q}(\alpha):\mathbf{Q}]\mathrm{Vol}(\beta)$ is close to one and hence there will be no problem of finding a polynomial of small height with this smaller index. The new problem which is introduced, however, is that this polynomial of small height might vanish to large order at $(p_1/q_1,\dots,p_m/q_m)$. In order to avoid this, Faltings' idea is the following: instead of applying Siegel's Lemma to the space of *all* polynomials $f \in \mathbf{Z}[X_1,\dots,X_m]$ with large order of vanishing at $(\alpha,\dots,\alpha)$, he instead applies Siegel's Lemma to a smaller subspace V_a consisting of those polynomials with large index at $(\alpha,\dots,\alpha)$ *and* index at most a at $(p_1/q_1,\dots,p_m/q_m)$. Dyson's lemma, or rather its proof, bounds the dimension of V_a below as a function of a; essentially it shows that $\dim(V_a)$ is maximal. Thus one constructs *directly* a polynomial with all of the desired properties:

i. large index at $(\alpha,\dots,\alpha)$,

ii. small index, namely at most a, at $(p_1/q_1,\dots,p_m/q_m)$,

iii. small coefficients.

The smaller one chooses a the larger the coefficients become, just as in the approach of [EV], but at least one constructs directly a polynomial with all of the desired properties to prove Roth's theorem.

The original proof of Theorem 11 by Esnault and Viehweg [EV] is long and technical. The proof in [N1] is slightly less technical but still challenging and similar to [EV]. A largely elementary proof is given in [N2] and summarized in [N3]. Therefore, in an effort to provide something new, we have decided to focus on the case where $m = 2$. The result is already interesting in this case and many of the elements of the full result are already on display here.

We first claim, assuming $m = 2$, that there is at most one index i such that $t_i > 1$. Suppose a and b are two distinct points of $\mathbf{P}^1$. Then it is possible to find a projective linear automorphism $\phi : \mathbf{P}^1 \to \mathbf{P}^1$ so that $\phi(a) = (0,1)$ and $\phi(b) = (1,0)$. If ζ_i and ζ_j are two distinct points from Theorem 11 then by hypothesis they have distinct coordinates on both factors of $\mathbf{P}^1$ and hence, applying two projective linear automorphisms, we may assume that

$$\zeta_i = (0,1)\times(0,1),$$
$$\zeta_j = (1,0)\times(1,0).$$

The section $\sigma \in H^0(\mathbf{P},\mathcal{O}_{\mathbf{P}}(d))$ can be viewed as a bi–homogeneous polynomial of bi–degree (d_1,d_2) in the projective variables (X_1,Y_1) and (X_2,Y_2). Thus σ is a sum of monomials with coefficients, each monomial being of the form

$$M = X_1^{i_1}Y_1^{i_2}\times X_2^{j_1}Y_2^{j_2},\ \ i_1+i_2=d_1,\ j_i+j_2=d_2.$$

A simple calculation shows that $\mathrm{ind}_{(0,1)\times(0,1)}(M) + \mathrm{ind}_{(1,0)\times(1,0)}(M) = 2$. Since $\mathrm{ind}_{\zeta_i}(\sigma)$ is the *minimum* of $\mathrm{ind}_{\zeta_i}(M)$ as M varies over all monomials with non–zero coefficient in the

expansion of σ it follows that $\mathrm{ind}_{\zeta_i}(\sigma) + \mathrm{ind}_{\zeta_j}(\sigma) \leq 2$ and, moreover, equality holds if and only if σ is a monomial.

Let V_i be the support of the collection of sections of $\mathcal{O}_{\mathbf{P}}(d)$ with index at leat t_i at ζ_i. If $t_i \leq 1$ we have $V_i = \zeta_i$ and if $t_i > 1$ then

$$V_i = \mathbf{P}^1 \times \{\pi_2(\zeta_i)\} \cup \{\pi_1(\zeta_i)\} \times \mathbf{P}^1.$$

From what we have shown above, V_i has dimension one for at most one index i and, more importantly, $V_i \cap V_j$ is empty for all $i \neq j$ since, by hypothesis, the ζ_i have distinct coordinates on both factors. The fact that V_i and V_j are disjoint when $i \neq j$ is a very weak form of independence. In particular, if V_i were to *meet* V_j this would mean that the conditions imposed at ζ_i and those imposed at ζ_j are *not* independent. Of course, they could still fail to be independent even though $V_i \cap V_j = \emptyset$.

Next we make an intersection theoretic construction which necessitates producing additional sections of $\mathcal{O}_{\mathbf{P}}(d)$ with similar index to σ at the points ζ_i. This is done by differentiating σ. Since $d_1 \gg d_2$ the derivatives along the first $\mathbf{P}^1$ lower the index by at most $1/d_1$ while the derivatives along the second $\mathbf{P}^1$ possibly lower the index by $1/d_2$. Thus the "cheap" direction in which to take derivatives is with respect to the first variable. The critical point to notice here is that one does not need a second section $\tau \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ so that the zero set $Z(\tau)$ of τ meets the entire zero set $Z(\sigma)$ properly. One can treat each irreducible component of $Z(\sigma)$ separately.

Let C be an irreducible component of $Z(\sigma)$. There are three cases to treat:

i. C does not contain any of the points ζ_i,

ii. C contains at least one ζ_i and C is a fibre of one of the two projections,

iii. C contains at least one ζ_i and is not a fibre of either projection.

Let $\mathcal{I}_\zeta(t) \subset \mathcal{O}_{\mathbf{P}}$ be the ideal sheaf generated by those sections of $H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$ with index at least t at ζ and let

$$\mathcal{J} = \cap_{i=1}^{M} \mathcal{I}_{\zeta_i}(t_i - \delta).$$

In case **i**, the support of $\mathcal{I}_{\zeta_i}(t_i)$ is disjoint from C if $t_i \leq 1$ because here the support of $\mathcal{I}_{\zeta_i}(t_i)$ is ζ_i. In particular $\mathcal{J}$ restricted to C is either $\mathcal{O}_C$, in case $t_i - \delta \leq 1$ for all i, or else $\mathcal{O}_C(-(t_i - \delta - 1)(d_1F_1 + d_2F_2))$ where F_1 and F_2 are the fibres of the first and second projections through the unique point ζ_i with $t_i - \delta > 1$; here we have acted as if $d_1(t_i - 1 - \delta)$ and $d_2(t_i - 1 - \delta)$ are integers. If they are not then one chooses instead the smallest positive integer greater than $d_1(t_i - 1 - \delta)$ and $d_2(t_i - 1 - \delta)$ respectively. If $\mathcal{J}|C = \mathcal{O}_C$, then we may choose $\tau_C \in H^0(C, \mathcal{O}_{\mathbf{P}}(d))$ general while in the second case we choose $\tau_C \in H^0(C, \mathcal{O}_{\mathbf{P}}(d))$ general so that $Z(\tau_C) - (t_i - \delta - 1)(d_1F_1 + d_2F_2)$ is effective: we know this is possible since $t_i - \delta - 1 < 1$.

For case **ii**, we know that the ζ_i contained in C is unique by hypothesis. When the ideal $\mathcal{J}$ is restricted to C all but at most two points go away, namely ζ_i and then the unique ζ_j, if it exists, such that $t_j > 1$. From the considerations above we know that $t_i + t_j \leq 2$

and hence there is no problem finding a section $\tau_C \in H^0(C, \mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{J})$ and we choose τ_C general. Finally, in case **iii**, we use (12) which says, locally identifying σ with a polynomial f, that there is a derivative D of order at most d_2 in X_1 so that $D(f)$ does not vanish on C. Choosing D of minimal order t and adding t general fibres of the first projection to preserve the degree gives a non–zero section τ_C of $H^0(C, \mathcal{O}_C(d_1, d_2) \otimes \mathcal{J})$.

Suppose we write $Z(\sigma) = \sum_{i=1}^k a_i C_i$ so that the cycle $\sum_{i=1}^k a_i C_i$ represents $c_1(\mathcal{O}_{\mathbf{P}}(d))$. For each curve C_i we have chosen a section $\tau_{C_i} \in H^0(C_i, \mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{J})$. If Z_i denotes the zero cycle of τ_{C_i} then $Z = \sum_{i=1}^k a_i Z_i$ represents $c_1(\mathcal{O}_{\mathbf{P}}(d))^2 = 2d_1 d_2$. Let $V_i = \text{support}(\mathcal{O}_{\mathbf{P}}/\mathcal{I}_{\zeta_i}(t_i - \delta))$ and let B_i be the part of Z supported on V_i. Then we have

$$2d_1 d_2 = \deg(E) + \sum_{i=1}^{M} \deg(B_i),$$

where E is the part of Z which is not supported on any of the V_i. Note that E is effective because each Z_i is effective. Hence we have by construction

$$2d_1 d_2 \geq \sum_{i=1}^{M} \deg(B_i), \tag{17}$$

and (17) will imply Dyson's lemma for $m = 2$ if we can show that

$$\deg(B_i) \geq 2\text{Vol}(t_i - \delta). \tag{18}$$

Fix an index i for (18). All sections, σ and each τ_C, which have gone into the construction of B_i lie in $\mathcal{J}$ and in particular in $\mathcal{I}_{\zeta_i}(t_i - \delta)$. The amount of the intersection supported on V_i cannot increase if instead one intersects *general* sections of $H^0\left(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{I}_{\zeta_i}(t_i - \delta)\right)$ and performing this operation gives a lower bound of $2\text{Vol}(t_i - \delta)$ for $\deg(B_i)$, which is precisely (18). Unfortunately, though this non–increase of intersection numbers under passing to general members of a linear system is intuitively clear it is not so easy to state in a precise way adapted to this situation. The method employed in [N2] to prove (18) is to consider the amount of Z supported *away* from V_i. The cycle Z is constructed by intersecting (non–general) members of $H^0\left(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{I}_{\zeta_i}(t_i - \delta)\right)$. The sheaf sheaf $\mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{I}_{\zeta_i}(t_i - \delta)$ is generated by global sections by definition and, blowing–up this sheaf, a simple calculation bounds from above the part of Z supported away from B_i and establishes (18). Note that equality holds in this bound exactly when the sections in the construction are "general."

The proof of Theorem 11 for general m follows the same outline given above for $m = 2$: namely an intersection product is constructed by successively constructing sections of $H^0(Y, \mathcal{O}_{\mathbf{P}}(d) \otimes \mathcal{J})$ on different subvarieties Y. In order to find these sections, the arguments for the case $m = 2$ need to be generalized. First one must show that $t_i + t_j \leq m$ whenever $i \neq j$. This is no different from the argument when $m = 2$. More seriously, one must show that decreasing the index by $m\delta$ allows enough derivatives to produce the desired sections for different subvarieties Y. At this point, (12) is replaced by a form of Faltings' product theorem which we state here:

Theorem 12 (Product Theorem) *Let $f \in \mathbf{C}[X_1, \ldots, X_m]$ be of multi-degree $d_1, \ldots, d_m$ and for an $m-1$-tuple of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_{m-1})$ 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*

$$X(f) = \left\{ (x_1, \ldots, x_m) \in \mathbf{C}^m \,\middle|\, D^\alpha f(x_1, \ldots, x_m) = 0, \quad \forall\ 0 \leq \alpha_i \leq \max_{i+1 \leq j \leq m} \{d_j\} \right\}.$$

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

The Product Theorem is applied in the following fashion. If $Y \subset \mathbf{P}$ is not contained in a proper product subvariety then Theorem 12 guarantees that there will be a derivative D, of small index, such that $D(\sigma)|Y$ is non–zero: this is the section τ_Y. If Y is contained in a proper product subvariety then another method is needed to produce τ_Y. This is done in [EV] Lemma 2.9. It would be interesting perhaps to find a conceptual proof of this result as it is a major hang–up, in general, in many diophantine arguments.

We close this section by relating Dyson's lemma to a very famous conjecture in complex algebraic geometry. Suppose $\pi : X \to \mathbf{P}^2$ is the blow–up of $\mathbf{P}^2$ at r very general points: very general here means that the points are chosen to avoid countably many proper Zariski closed subsets of the variety $(\mathbf{P}^2)^r$ parametrizing collections of r points in $\mathbf{P}^2$. Let E denote the exceptional divisor of X so E consists of r disjoint divisors lying over the r very general points of $\mathbf{P}^2$.

Conjecture 13 (Nagata) *If $r \geq 9$ then the line bundle $\pi^*(\mathcal{O}_{\mathbf{P}}(1)) \left(-\frac{1}{\sqrt{r}} E\right)$ is nef on X.*

The Nagata conjecture, like Dyson's lemma, is a statement about independence of conditions. Roughly speaking it says that as soon as $r \geq 9$ there will only exist a polynomial of degree d with multiplicity at least m at very general points $P_1, \ldots, P_r$ provided a simple dimension count predicts the existence of such a polynomial. As soon as the system of equations is overdetermined, there is no solution. This of course has precisely the same flavor as Dyson's lemma except that Dyson's lemma allows room for error. The pointwise conditions at the ζ_i in Dyson's lemma are not required to be independent but only *nearly* independent. In other words, the analogue of Dyson's lemma in the setting of the Nagata conjecture would be to give an effective number $\alpha < \frac{1}{\sqrt{r}}$ such that $\pi^*(\mathcal{O}_{\mathbf{P}}(1))(-\alpha E)$ is nef.

Not surprisingly, the methods used to study the Nagata conjecture often employ *degeneration* techniques where the points P_i are allowed to vary. Dyson's lemma is also of course proved using a degeneration technique but the method is a little more brutal as the actual polynomial is "deformed" by taking the derivative while the points ζ_i do not move. Since the points ζ_i are algebraic in the application they cannot be deformed continuously and thus this type of approach is ruled out. It is worth asking, perhaps, when the analogue of the Nagata conjecture holds in the Dyson lemma setting, that is when is it true that

$$\sum_{i=1}^{M} \mathrm{Vol}(t_i) \leq 1.$$

for all sections $\sigma \in H^0(\mathbf{P}, \mathcal{O}_{\mathbf{P}}(d))$? For those interested in learning more about the Nagata conjecture there is a vast literature on the topic [CM, CK, E, X].

EXERCISES

1. Prove that if $M = 2$ in Dyson's lemma then the strongest possible inequality holds:

$$\mathrm{vol}(t_1) + \mathrm{vol}(t_2) \leq 1.$$

2. Suppose one tries to extend Dyson's lemma to a product of $\mathbf{P}^2$'s. The volumes will still make sense but they will be measuring the size of a suitable convex body in $\mathbf{R}^{2m}$ rather than $\mathbf{R}^m$.

a. Is the $\mathbf{P}^1$ version of Dyson's lemma still true in this setting, that is can you find conditions on a collection of points $\{\zeta_i\} \in (\mathbf{P}^2)^m$ so that the conditions for imposing index t_i at ζ_i are close to being independent?

b. What is geometrically different about the higher dimensional setting, where $\mathbf{P}^1$ is replaced by $\mathbf{P}^2$, which creates a problem? Note: it is partly for this reason that the Schmidt subspace theorem has a very different and very involved proof.

3. Would a Nagata type result, that is a result like Problem **1** above, for the index be of any additional help in arithmetic applications?

4. Prove that Nagata's conjecture is true for $\mathbf{P}^2$ blown up at one or two points. Will your methods be successful for establishing the conjecture for $\mathbf{P}^2$ blown up at three points or five points?

5 Extensions and Questions

The material treated in the first four sections has been developed further in many directions as will be clearly on display throughout this conference. On the arithmetic side are the Mordell Conjecture, the Schmidt Subspace Theorem, and extensions due to Faltings and Faltings–Wüstholz. On the geometric side, there are many interesting results which have arisen by studying the type of question posed by Dyson's Lemma: these include the indentification of the dual to the effective cone of divisors on a smooth projective variety, investigation of the Nagata Conjecture, and study of the variation of base loci. The most impressive progress in higher dimensional geometry in recent years is the successful demonstration of finite generation of the canonical ring for varieties of general type and some aspects of this story too are closely related to what has been described here.

We will first discuss further developments on the arithmetic side and then turn to geometric questions. One extremely important development is the proof of the Mordell Conjecture and its higher dimensional analogue for subvarieties of abelian varieties by Faltings [F1]:

Theorem 14 (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.*

The Mordell conjecture is a special case of Theorem 14 since a curve of genus at least 2 can be embedded in its Jacobian. In [F2] Faltings generalized this result to deal with the case where X can contain translates of non–trivial abelian subvarieties $B \subset A$, though of course not enough translates to cover X.

Theorem 15 (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.*

By induction on $\dim X$ one sees that Theorem 15 implies that if $X \subset A$ is of general type then the Zariski closure of $X(k)$ is a finite union of translated abelian subvarieties. Before sketching the basic structure of the proof of Theorem 14 we should point out that the idea of using a diophantine argument to establish the Mordell Conjecture is due to Vojta. He first [V2] produced a "diophantine" style proof in the function field case and then [V1, V3] successfully adapted this to the number field setting. Faltings' methodology [F1, F2] adopts Vojta's conceptual framework.

Comparison of proofs for Roth's Theorem and Faltings' Theorem		
	Roth's Theorem	**Faltings' Theorem**
1.	$\alpha \in \mathbf{R}$ algebraic irrational	$X \subset A$ containing no translates of non–trivial abelian subvarieties
2.	$p_1/q_1, \ldots, p_m/q_m$ with $\lvert p_i/q_i - \alpha\rvert < \frac{1}{q^{2+\epsilon}}$	$x_1, \ldots, x_m \in X(k)$
3.	$0 \ll q_1 \ll q_2 \ll \ldots \ll q_m$	$0 \ll h(x_1) \ll h(x_2) \ll \ldots \ll h(x_m)$
4.	$P(X) \in \mathbf{Z}[X_1, \ldots, X_m]$ with large index at $(\alpha, \ldots, \alpha)$ and $h(P)$ small	$\sigma \in H^0\left(X^m, L_{-\epsilon,x}\right)$ with $x = (x_1, \ldots, x_m)$ and $h(\sigma)$ small
5.	$\mathrm{ind}_{(p_1/q_1,\ldots,p_m/q_m)}(P)$ is big	$\mathrm{ind}_{(x_1,\ldots,x_m)}(\sigma)$ is big
6.	$\mathrm{ind}_{(p_1/q_1,\ldots,p_m/q_m)}(P)$ is small	$\mathrm{ind}_{(x_1,\ldots,x_m)}(\sigma)$ is small

The left hand side of the table has already been thoroughly discussed. Note that $\mathrm{ind}_{(p_1/q_1,\ldots,p_m/q_m)}(P)$ is large because Steps 2 and 4 imply that $\lvert P(p_1/q_1, \ldots, p_m/q_m)\rvert$ is too small to be non–zero, and similarly for small index derivatives of P. The upper bound for the index $\mathrm{ind}_{(p_1/q_1,\ldots,p_m/q_m)}(P)$ in Step 6 which contradicts the bound in Step 5 comes from Roth's lemma or Dyson's lemma or the arithmetic product theorem.

Steps 3 through 6 on the right hand side of the table all require significant clarification. The heights $h(x_i)$ in Step 3 are very much analogous to the heights of the rational numbers p_i/q_i which were defined by viewing a rational number as a point on the projective line. Similarly here, an embedding $i : X \hookrightarrow \mathbf{P}^N$, defined over k, is fixed. If $x \in X(k)$ then $i(x) \in \mathbf{P}^N(k)$ and we can attach a height to x by looking at the size of the coordinates of $i(x)$ with respect to the different absolute values on k. The trouble with this method of defining $h(x)$ is that it depends not only on the choice of line bundle $L = i^*\mathcal{O}_{\mathbf{P}}(1)$ but also on the specific *choice of embedding*: altering the embedding changes the heights by

a bounded amount. There are a couple of ways around this problem. One is to look for intrinsically defined heights, which can be done, for example, on an abelian variety by using the group law. This is the theory of Néron–Tate heights. An alternative is to fix several embeddings of X in different projective spaces and then use these coordinates to define all heights. Interestingly, *both* of these methods are central to the proof of Theorem 14: indeed it is the conflict between the different estimates for the height functions which fuels the contradiction establishing Theorem 14.

Step 4 requires much explanation. Here $X^m = X \times \ldots \times X$ where there are m factors. The line bundle $L_{-\epsilon,x}$ is carefully chosen so that

$$h_{L_{-\epsilon,x}}(x) \ll 0. \tag{19}$$

The height function $h_{L_{-\epsilon,x}}$ for which (19) holds is obtained by fixing heights for a few specific bundles and then extending them by linearity. In order to obtain (19) $L_{-\epsilon,x}$ is written as a difference $L_{0,x} - \epsilon A_x$ where $L_{0,x}$ is associated to a morphism of X^m to A^{m-1} and A_x is a very simple ample line bundle.

The inequality (19) uses the Mordell–Weil theorem in a very interesting way and of course requires that the points x_i be chosen carefully, just as the rational approximating points p_i/q_i to α had to be chosen carefully in proving Roth's Theorem. Finally the height $h(\sigma)$ is to be understood in the metric sense; that is, for each complex or real model Y of X, the section σ gives a section of $H^0\left(Y, L_{-\epsilon,x}\right)$ whose size is measured relative to a metric ρ on the bundle $L_{-\epsilon,x}$ over Y. This needs to be done in a uniform way: in other words, the bundles $L_{-\epsilon,x}$ *depend* on x and the metrics, like the height functions for $L_{-\epsilon,x}$, must be determined in a regular fashion. Note that when $s \in H^0\left(\mathbf{P}^n_{\mathbf{Q}}, \mathcal{O}_{\mathbf{P}}(d)\right)$ then the size of s, measured relative to the Fubini–Study metric m by taking $\sup_{x \in \mathbf{P}^n} |s(x)|_m$, will be directly proportional to the size of s given by taking the height of the projective vector of coefficients of s, viewed as a homogeneous polynomial of degree d. Thus the concept of height for sections of line bundles is a direct generalization of that we already encountered for polynomials.

For Step 5 the theory of arithmetic intersection theory is necessary and in particular the fact that the height functions $h_L(x)$ can be estimated whenever one has a section $\sigma \in H^0(X, L)$ with $\sigma(x) \neq 0$. In particular, the section σ, if it did not vanish at x and if $h(\sigma)$ were sufficiently small, could be used to find a lower bound on $h_{L_{-\epsilon,x}}$ which contradicts (19). This argument leads to a lower bound on the vanishing of σ at x. Some difficult work is necessary here, however, because to recover the height function $h_{L_{-\epsilon,x}}$ via arithmetic intersection theory not only are metrics required on all complex and real models of X, but one also needs an arithmetic model $\mathcal{X}$ of X, that is an extension of X to a variety defined over $\mathrm{Spec}(\mathcal{O}_K)$.

For step 6, the line bundle $L_{-\epsilon,x}$ is ample, provided the points x_i are chosen carefully and in particular provided

$$0 \ll h(x_1) \ll \ldots \ll h(x_m).$$

The proof of this positivity is very ingenious and, not surprisingly, resembles the proof of Dyson's lemma: both results are established by taking the derivative of a section and then

inductively applying the product theorem. There are, however, additional subtleties to be overcome in the situation of Theorem 14. First $L_{-\epsilon,x}$ does not necessarily have a section, even on X^m, and some computation and cohomology estimates are necessary here. Next, unlike $\mathbf{P}^1 \times \ldots, \times \mathbf{P}^1$ where all product subvarieties have the same structure as the original variety, on X^m there are many product subvarieties and it is critical in practice to control their degrees. Using the positivity of $L_{-\epsilon,x}$, the section σ can be chosen to have both small height and small index at x, using a natural generalization of Siegel's Lemma: this method has already been described when we discussed proving Roth's theorem with Dyson's lemma. The new version of Siegel's Lemma requires some real work as it requires a lattice structure on $H^0(X, L_{-\epsilon,x})$ which comes from the arithmetic model $\mathcal{X}$ for X.

The proof of Theorem 15 looks significantly different from that of Theorem 14, essentially because it follows the arithmetic product theorem proof of Roth's theorem rather than the Dyson lemma proof. When $X \subset A$ is of general type, the union of all translates of positive dimensional abelian subvarieties contained in X is a proper subvariety (not necessarily irreducible) $Z \subset X$. If Theorem 15 were false then there would be infinitely many rational points $x \in X(k) \backslash Z(k)$. Points $x_1, \ldots, x_m \in X(k) \backslash Z(k)$ are selected with

$$0 \ll h(x_1) \ll h(x_2) \ll \ldots \ll h(x_m)$$

and it is arranged, using the Mordell–Weil theorem, for the line bundle $L_{-\epsilon,x}$ to have a uniform choice of height function satisfying

$$h_{L_{-\epsilon,x}}(x) \ll 0.$$

The problem is that $L_{-\epsilon,x}$ is no longer ample: the subvariety $Z \subset X$, when it is positive dimensional, creates a locus inside X^m where the bundle $L_{-\epsilon,x}$ is negative. In principle, this is irrelevant provided one can still produce a section $\sigma \in H^0(X^m, L_{-\epsilon,x})$ which does not vanish at x. Unfortunately, the geometric properties of non–ample line bundles are not sufficiently well understood to guarantee this. Thus an alternative approach is used in [F2]. It is first shown that $L_{-\epsilon,x}$ is effective as long as x is chosen appropriately. Then the arithmetic product theorem is applied to reduce to a question on $Y_1 \times \ldots \times Y_m \subset X^m$, also containing x, where the degrees and heights of the Y_i are bounded by the arithmetic product theorem. One then repeats the procedure, producing a new section

$$\sigma_Y \in H^0\left(Y_1 \times \ldots \times Y_m, L_{-\epsilon,x}\right).$$

The final contradiction reached, as in the proof of Roth's theorem via the arithmetic product theorem, is that the height of one of the coordinates of x is too small.

The main reason why the arithmetic product theorem is used to prove Theorem 15 while the Dyson lemma approach was successful for Theorem 14 is, as noted above, that our geometrical knowledge is lacking. The theory for producing sections of ample line bundles is very well understood while the same theory for non–ample line bundles is not well developed. The proof in [FW] of the generalized Schmidt Subspace Theorem also uses the same

arithmetic product theorem approach, again for the same reason. The geometric problem posed repeatedly in these different diophantine settings is to limit the vanishing, at certain points, of sections of some line bundle L on a smooth projective variety X. When L is ample, it is often possible to produce such bounds. Without this global hypothesis of ampleness, however, little is known about the base locus of L, either qualitatively nor quantitatively.

One exception to this rule is for surfaces X where base loci of line bundles are well understood: in this case one can appeal to the Zariski decomposition of an effective divisor D, at least asymptotically, to determine the base locus of D. There are a couple of equivalent definitions of the Zariski decomposition, both of them technical. If D is a divisor on X then a pair of $\mathbf{Q}$–divisors P, N is called a Zariski decomposition (see [KMM] §7.3) of D if

i. D is $\mathbf{Q}$– linearly equivalent to $P+N$, that is an integral multiple of D is linearly equivalent to the corresponding multiple of $P + N$.

ii. P is nef and N is effective.

iii. If $f : Y \to X$ is a birational map with Y normal and E is effective on Y with $f^*(D) - E$ nef then $E - f^*(N)$ is effective on Y.

The idea of a Zariski decomposition is that N is the negative or fixed part which is present in any effective representative of D while P is the moving part or that part of these effective representatives which actually varies. As far as finding the Zariski decomposition goes, there is a simple but impractical computational algorithm when $\dim(X) = 2$: beginning with the curves $C_i \subset X$ with $D \cdot C_i < 0$ one produces minimal positive rational numbers a_i so that

$$\left(D - \sum_i a_i C_i\right) \cdot C_i = 0, \ \ \forall i.$$

If $M = D - \sum a_i C_i$ is nef then the Zariski decomposition of D is

$$D = M + \sum_i a_i C_i.$$

If $D - \sum a_i C_i$ is not nef then one repeats the above procedure with $D - \sum a_i C_i$ in place of D. This process either terminates with the Zariski decomposition of D or else has an asymptotic limit which is the Zariski decomposition of D. In [BKS] Zariski decomposition is studied as the divisor D varies within the effective cone and a good understanding of the behavior is obtained.

In higher dimension, the nice picture provided by the Zariski decomposition breaks down and very little is known. As a result, in general there is no method to find the base locus of a divisor D on a variety X of dimension at least 3, neither the support nor the scheme structure. One aspect of the surface situation which generalizes directly to the higher dimensional setting, however, is the support of the base locus when the divisor in question is sufficiently close to the boundary of the ample cone. Suppose L is a line bundle on a projective variety X. We let

$$\mathrm{BS}(L) = \{x \in X : s(x) = 0 \text{ for all sections } s \in H^0(X, nL) \text{ and for all } n > 0\}.$$

We call $\mathrm{BS}(L)$ the *stable base locus* of L. Suppose L is a nef line bundle on X, that is $L \cdot C \geq 0$ for all curves $C \subset X$. Let

$$S = \left\{ P \in X : c_1(L)^{\dim V} \cap V = 0 \text{ for some } V \subset X \text{ containing } P \right\}.$$

and let $L^{\perp}$ denote the Zariski closure of S in X. It turns out that S itself is Zariski closed. The following result is shown in [N4]:

Theorem 16 *Suppose X is a smooth projective variety of dimension $d \geq 2$ and L a big and nef line bundle on X. Suppose A is an ample line bundle on X. Then there exists $\epsilon > 0$ such that*

$$L^{\perp} = \mathrm{BS}(L(-\delta A))$$

whenever $0 < \delta < \epsilon$.

This is a direct generalization of the fact, when $d = 2$, that the negative part of the Zariski decomposition of $L - \epsilon A$, for $\epsilon \ll 1$, will be exactly the curves C_i with $L \cdot C_i = 0$. Note that although Theorem 16 determines the *support* of $\mathrm{BS}(L - \delta A)$ for small δ it does *not* determine its scheme structure. In the surface case, by contrast, [BKS] determines not only the support but also the multiplicities in the negative part of the Zarsiki decomposition of $L - \delta A$.

From the point of view of Theorem 15, what is required is a *quantification* of Theorem 16. In particualar given $x = (x_1, \ldots, x_m)$ as in the proof of Theorem 15, Theorem 16 determines $\mathrm{BS}(L_{0,x} - \delta_x A_x)$ for δ_x sufficiently small but in order to establish the theorem one needs a *uniform lower bound* on δ_x and Theorem 16 is not sophistocated enough to provide this. Otherwise stated, it is not currently known how to bound the complexity of the scheme structure for the base locus of $L - \epsilon A$ as a function of ϵ.

Motivated by trying to quantify Theorem 16 and derive Theorem 15 as a corollary, the following closely related problem was studied in [N5]: given L on the boundary of the ample cone and A ample, what is the supremum of all real numbers for which $L - \alpha A$ is effective? The philosophy of [N5] is that there should be a numerical explanation for when $L - \alpha A$ reaches the boundary of the effective cone. This was quantified using an invariant called the moving Seshadri constant. A conjecture was then made to help identify this with more intrinsically defined invariants and in [BDPP] the problem was resolved.

One obvious condition satisfied by all effective divisors D is that $D \cdot C \geq 0$ provided C passes through a point x not contained in D. If C is a curve which moves in a sufficiently large family, in particular a family which covers an open set of X with only finitely many base points then $D \cdot C \geq 0$. It turns out that this necessary condition is also sufficient. More specifically, a one dimensional cycle ξ on X, with real coefficients, is said to be *mobile* if there exsits a projective, birational map $\phi : Y \to X$ and ample divisors $A_1, \ldots, A_{d-1}$ on Y so that

$$\phi_*(A_1 \cdot \ldots \cdot A_{d-1}) = \xi.$$

It follows of course that if C is a mobile curve, $V \subset X$ a proper subvariety, then there is a curve C', numerically equivalent to C, so that $V \cap C'$ is proper. Hence if D is an effective

divisor on X then $D \cdot \xi \geq 0$ for all mobile curves ξ. Let $\mathrm{Mov}(X)$ be the collection of all mobile curves, inside the real cone of curves modulo numerical equivalence $N_1(X)_{\mathbf{R}}$ and let $\overline{\mathrm{Mov}}(X)$ be its closure. Finally let $\mathrm{Eff}(X) \subset N^1(X)_{\mathbf{R}}$ be the real, effective divisors inside the cone of divisors modulo numerical equivalence. The Theorem of Boucksom, Demailly, Paun, and Peternell is the following

Theorem 17 *Under the intersection pairing, the closure of the effective cone of divisors* $\overline{\mathrm{Eff}}(X)$ *is dual to the closure of the cone of mobile curves* $\overline{\mathrm{Mov}}(X)$.

Returning to Theorem 15, what is needed in order to prove this result with the same geometric approach as that used in Theorem 14 is an effective determination of the base locus of $L - \alpha A$ for all values of α. Theorems 16 and 17 only determine $\mathrm{BS}(L - \alpha A)$ in the two extremal cases where

i. $\mathrm{BS}(L + \epsilon A) = \emptyset$ for all $\epsilon > 0$,

ii. $\mathrm{BS}(L - \epsilon A) = X$ for all $\epsilon > 0$.

For values of α in between, different invariants are introduced in [N5, ELMNP1, ELNMP2, ELNMP3] but the results established are all ineffective. As far as where to turn for new ideas, the natural candidate is the deep and beautiful work [BCHM] of Birkar, Cascini, Hacon, and McKernan and with this invitation we close these notes.

EXERCISES

1. Suppose X is a smooth surface and $C \subset X$ is an irreducible curve with $C^2 < 0$. Show that the class of C in $N_1(X)_{\mathbf{R}}$ is on the boundary of the effective cone of X.

2. Let $p : X \to \mathbf{P}^2$ be the blow up of $\mathbf{P}^2$ at two points. Find the effective cone of X and compute the Zariski decomposition of all effective divisors on X. Do the same for $\mathbf{P}^1 \times \mathbf{P}^1$ blown up at a point.

3. Prove that the procedure outlined in the text for finding the Zariski decomposition of an effective divisor D on a smooth projective surface X achieves this task. Are the coefficients in the Zariski decomposition always integers? Are they always rational? Why is it not possible to duplicate this argument when $\dim(X) > 2$?

4. Give an example of a surface X, an effective divisor D, and a curve C so that $D \cdot C > 0$ but C is in the negative part of the Zariski decomposition of D.

5. Prove Theorem 17 when $\dim(X) = 2$. Why is the case of surfaces misleadingly simpler than the general case?

6. Suppose X is a smooth surface and C an irreducible curve on X.

a. If $C^2 > 0$ show that C is mobile.

b. If $C^2 < 0$ show that C is not mobile and is not in the closure of the mobile cone.

c. What can you say when $C^2 = 0$?

7. Prove the following "piecewise linear" behavior of Zariski decompositions established in [BKS]: let D be an effective divisor with Zariski decomposition $D = P + N$. We will assume that P is non–trivial. Suppose that $\{C_i\}_{i=1}^r$ are the curves in N and suppose that *all* divisors near D have the same set of r curves in the negative part of their Zariski decomposition. Suppose A is an ample divisor on X and let $D_\alpha = D + \alpha A$. Let $P_\alpha + N_\alpha$ be the Zariski decomposition of D_α. Show that the coefficients of N_α vary linearly with α.

References

[BKS] T. Bauer, A. Küronya, T. Szemberg, *Zariski chambers, volumes, and stable base loci*, Crelle, **576**, 2004, pp. 209–233.

[BCHM] C. Birkar, P. Cascini, C. Hacon, J. McKernan, *Existence of Minimal Models for Varieties of Log General Type*, Journal of the AMS, **23**, 2010, pp. 405–468.

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

[B2] E. Bombieri, *Effective Diophantine Approximation on G_m*, Ann. Scuola Norm. Sup. Pisa Cl. Sci., (4) **20** (1993), pp. 61–89.

[BC] E. Bombieri and P. Cohen, *Effective Diophantine approximation on G_m, II*, Ann. Scuola Norm. Sup. Pisa Cl. Sci., (4) **24** (1997), pp. 205–225.

[BG] E. Bombieri and W. Gubler, *Heights in Diophantine Geometry*, Cambridge University Press, 2006.

[BPV] E. Bombieri, A Van der Poorten, and J.D. Vaaler, *Effective measures of irrationality of cubic extensions of number fields*, Ann. Scuola Norm. Sup. Pisa Cl. Sci., (4) **23** (1996), pp. 211–248.

[BDPP] S. Boucksom, J.–P. Demailly, M. Paun, and T. Peternell, *The pseudo–effective cone of a compact Kähler manifold and varieties of negative Kodaira dimension*, J. Algebraic Geom., **22** (2013), pp. 201-248

[CM] C. Ciliberto and R. Miranda, *Nagata's conjecture for a square or nearly-square number of points*, Ricerche di matematica, **55** (2006), pp. 71–78.

[CK] C. Ciliberto and A. Kouvidakis, *On the symmetric product of a curve with general moduli*, Geom. Dedicata **78** (1999), pp. 327–343.

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

[ELMNP1] L. Ein, R. Lazarsfeld, M. Mustata, M. Nakamaye, and M. Popa, *Asymptotic Invariants of Line Bundles*, Pure Appl. Math Quarterly, **1** (A. Borel issue), 2005, pp. 379–403.

[ELNMP2] L. Ein, R. Lazarsfeld, M. Mustata, M. Nakamaye, and M. Popa, *Asymptotic Invariants of Base Loci*, Annales de l'Institut Fourier, **56**, 2006, pp. 1701–1734.

[ELNMP3] L. Ein, R. Lazarsfeld, M. Mustata, M. Nakamaye, and M. Popa, *Restricted volumes and base loci of linear series*, to appear, American Journal of Mathematics, 2009.

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

[E] L. Evain, *Computing limit linear series with infinitesimal methods*, Ann. Inst. Fourier, **57** (2007), pp. 1947–1974.

[F1] G. Faltings, *Diophantine Approximation on Abelian Varieties*, Annals of Math., **133**, 1991, pp. 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, pp. 175–182.

[FW] G. Faltings and G. Wüstholz, *Diophantine approximations on projective spaces*, Inv. math., **116**, 1994, pp. 109–138.

[HS] M. Hindry and J. Silverman, *Diophantine Geometry*, Springer, 2000.

[KMM] Y. Kawamata, K. Matsuda, and K. Matsuki, *Introduction to the Minimal Model program*, In: Oda, T. (ed.) Algebraic Geometry. Proc. Symp., Sendai, 1985, (Adv. Stud. Pure Math.,**10**, pp. 283-360).

[L] S. Lang, *Fundamentals of Diophantine Geometry*, Springer Verlag, 1983.

[Laz1] R. Lazarsfeld, *Positivity in Algebraic Geometry I*, Springer, 2004.

[Laz2] R. Lazarsfeld, *Positivity in Algebraic Geometry II*, Springer, 2004.

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

[N2] M. Nakamaye, *Intersection Theory and Diophantine Approximation*, Journal of Algebraic Geometry, **8**, 1999, pp. 135–146.

[N3] M. Nakamaye, *Diophantine Approximation on Algebraic Varieties*, Journal de Théorie des Nombres de Bordeaux, **11**, 1999, pp. 439–502.

[N4] M. Nakamaye, *Stable base loci of linear series*, Mathematische Ann., **318**, 2000, pp. 837–847.

[N5] M. Nakamaye, *Base loci of linear series are numerically determined*, Transactions of the AMS, **355**, 2003, pp. 551–566.

[PS] A. N. Parshin and I. R. Shafarevich editors, *Number Theory IV: Transcendental Numbers*, Springer, Encyclopedia of Mathematical Sciences, **44**, 1998.

[R] Roth, *Rational approximations to algebraic numbers*, Mathematika, **2**, 1955, pp. 1-20.

[T1] A. Thue *Bermerkungen über gewisse Näherungsbrüche algebraischer Zahlen*, Norske Vid. Selsk. Skr., **3** (1908), pp. 1–34.

[T2] A. Thue, *Über Annäherungswerte algebraischer Zahlen*, Crelle, **135**, 1909, pp. 284–505.

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

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

[V2] P. Vojta, *Mordell's conjecture over function fields*, Inv. Math., **98**, 1989, pp. 115–138.

[V3] P. Vojta, *Siegel's theorem in the compact case*, Annals of Math., **133**, 1991, pp. 509–548.

[X] G. Xu, *Curves in* $\mathbf{P}^2$ *and symplectic packings*, Math. Ann., **299** (1994), pp. 609–613.

[Y] S. Yang, *Linear systems in* $\mathbf{P}^2$ *with base points of bounded multiplicity*, Journal of Algebraic Geometry, **16** (2007), pp. 19–48.