Return to the Siegel-Mahler-Roth Theorem

Alain Robert

Translated by Kelsea Hull

1. Introduction

The study of integer solutions to polynomial equations dates to the earliest antiquity. However, it is not until the 20th century that general results could be obtained. One such result is as follows:

Theorem A (Siegel)

Let $P \in \mathbb{Z}[X,Y]$ be a polynomial with integer coefficients such that $P(X,Y) = 0$ gives infinite solutions with integer coordinates (x,y), there exists a parameterization of the plane curve (real or complex) of the equation $P(X,Y) = 0$

$$x = a(t) \quad , \quad y = b(t)$$

with two functions, a, b, which are Laurent polynomials of t (i.e. polynomials of t and $\frac{1}{t}$).

The same conclusion subsists when one assumes that $P(X,Y) = 0$ has infinite solutions of the form (x,y) with coordinates in an algebraic ring on $\mathbb{Z}$, for example a ring such as $\mathbb{Z}\left[\frac{1}{n}\right]$ where *n* is a strictly positive integer.

The conclusion of Theorem A signifies that the projective algebraic curve associated with the linear plane curve of the equation $P(X,Y) = 0$ has at most two points at infinity and is rational (and therefore birationally isomorphic to a projective line).

In this presentation, we will give the outline of a slightly weaker version of Theorem A indicating the role that it can play in non-standard analysis (NSA). For further reading on the

specific role of NSA in the demonstration of Theorem A, refer to [4]. The theorem we will focus on here constitutes the birational portion of Theorem A and can be stated as follows.

Theorem B

Let R be a sub-ring of $\mathbb{Q}$, (of finite type on Z) finitely generated over $\mathbb{Z}$, and $P \in R[X,Y]$. Assume there exists an infinite number of algebraic points (x_i, y_i) on the curve given by $P(X,Y) = 0$ and a rational function $f = f(X,Y)$ such that $f(x_i, y_i) \in R$ for all i. Thus, there exists a parameterization of the curve in question $x = a(t), y = b(t)$ given by two rational functions of t.

In Theorem B, we do not assume that the points on the curve given by (x_i, y_i) have their coordinates in R, only in the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$. However, since all these points have the first coordinate $x_i \in R$, you can choose for the rational function f an x that satisfies the hypothesis (specifically $f = y$ such that all $y_i \in R$). However, the weaker hypothesis from Theorem B no longer allows for an upper bound on the number of points at infinity on the associated projected curve. (To be more precise would require saying that a finite number of distinct t_j parameters lead to infinity on the associated projected curve, and that the number of theses t_j values cannot be increased using only the hypotheses from Theorem B.)

2. Elementary Examples

The most interesting examples are those where the curve is defined by a polynomial of degree two. Some characteristic cases are:

a) Consider the hyperbolic equation $x^2 - 2y^2 = 1$. It contains an infinite number of points (x_i, y_i) with integer coordinates. These are obtained in the usual fashion starting from the fundamental unit $u = 3 + 2\sqrt{2}$ of the quadratic body $\mathbb{Q}(\sqrt{2})$. The norm of this unit is

$$N(u) = \left(3 + 2\sqrt{2}\right)\left(3 - \sqrt{2}\right) = 9 - 8 - 1.$$

The integer powers $u^{\mathrm{i}} = \mathrm{x_i} + \mathrm{y_i}\sqrt{2}$ of u are the integer components x_i and y_i (when i is positive, this simply means the binomial coefficients for calculating the first power of $3 + \sqrt{2}$ are integers). Further

$$x_i^2 - 2y_i^2 - N\left(x_i + y_i\sqrt{2}\right) = N\left(u^i\right) = N(u)^i = 1$$

shows that all the points (x_i, y_i) are on the considered hyperbola.

b) On the hyperbola of the equation $y^2 - x^2 - xy = 1$, there is also an infinite number of points with integer coordinates, which can be constructed as follows.

Fibonacci numbers are defined inductively

$$f_0 = 0, f_1 = 1, \dots, f_{n+1} = f_n + f_{n-1}.$$

Thus, for this example

$$f_2 = 1, f_3 = 2, f_4 = 3, f_5 = 5, f_6 = 8, f_7 = 13, \dots.$$

The reader can verify that the points are all on the aforementioned hyperbola.

$$(x_i, y_i) = (f_{2i}, f_{2i+1}) \qquad i = 0,1, \dots.$$

However, the points (f_{2i+1}, f_{2i+2}) are on the hyperbola of the equation $y^2 - x^2 - xy = -1$. These two hyperbolae in fact have for shared asymptotes like the lines $y = mx$ of slope m which satisfies $m^2 - m - 1 = 0$, knowing $m = \frac{1}{2}(1 \pm \sqrt{5}\,)$. The quotients f_{i+1}/f_1 tend towards $\frac{1}{2}(1 + \sqrt{5})$ while $i \to \infty$ (similarly, they extend towards $\frac{1}{2}(1 - \sqrt{5})$ while $i \to -\infty$).

In both examples a) and b), it is easy to find a parameterization of the hyperbola with the Laurent polynomials of t.

c) On the hyperbolic equation $xy = 1$, visibly, there is only a finite number of points with integer coordinates, but taking a finite ring of the form $R = \mathbb{Z}[\frac{1}{p}]$ (where p is a prime number for example), the points

$$(x_i, y_i) = (2^i, \frac{1}{2^i})$$

have their coordinates in R and Theorem A (formula with R in place of $\mathbb{Z}$) is still applicable.

Let us now give some examples of degree higher than two.

d) On the curves of equation type $y = a_n x^n + \cdots + a_0$ there are an infinite number of points with integer coordinates (assume that the coefficients a_i are integers, of the sort that all integer values of x give integer values of y). An obvious polynomial parameterization of these curves would be:

$$x = t, y = a_n t^n + \cdots + a_0.$$

e) Finally let's take N with n_j distinct integers and consider the curve of the equation

$$y \prod_{1 \leq j \leq N} (x - n_j) = 1.$$

This curve can be parameterized by

$$x = t, y = \prod_{1 \leq j \leq N} (t - n_j)^{-1}.$$

The integer values of t give integer values of x and there are thus an infinite number of points having prime integer coordinates on this curve. Theorem B is applicable, but the $N + 1$ values $t = t_j$ and $t = \infty$ should be excluded and theorem A would only be applicable if $N + 1 \leq 2$, i.e. $N \leq 1$ (this case illustrates the precision mentioned at the end of the introduction).

f) For the cubic equation $x^3 - y^2 = 2$ there are only two points $(x, y) = (3, \pm 5)$ with integer coordinates (this result was already shown by Fermat), but it is evident that there are an infinite number of points with rational coordinates. The set of points both real and complex on this curve cannot be parameterized with rational functions (the curve is not "unicursal"). On the

contrary, this curve can be parameterized with Weierstrass transcendental functions. It is of the genus 1: an elliptic curve. Faltings' Theorem (formerly Mordell's conjecture) shows that the algebraic curve having an infinite number of rational points is of genus inferior or equal to 1. Genus 1, where there are only a finite number of integer points, but where there could be an infinite number of rational points, is crucial. This study is not finished.

3. Principles of Non-Standard Analysis

To demonstrate a classic property like the Siegel theorem, it suffices to establish it for when all givens are standard. The set of integer points on a standard algebraic curve is standard, so we use the following general principle.

(3.1) Principle. If E is a standard set, E is infinite if and only if E contains a non-standard element.

The principle allows us to translate Siegel's Theorem to Non-Standard Analysis (NSA) and leads to the study of non-standard integer points on the algebraic curves (in which case can these exist?). More generally, if we wish to handle points having coordinates in a finite ring $R \subset \mathbb{Q}$, it is necessary to define the term 'standard' for these rings.

(3.2) Proposition. Let $a \in \mathbb{Q}$ be a rational number and $R = \mathbb{Z}[a]$. Then the ring R is standard exactly when the only prime divisors of the denominator of a are all standard.

Let us continue our list of principles of NSA with the following statement.

(3.3) Principle. Let $f: E \to F$ be a standard map between sets (standard). If $x \in E$ is a standard element, then $y = f(x)$ is also a standard element of F.

The following proposition results from the stated principles.

(3.4) Proposition. Let $f: E \to F$ be a standard map between sets (standard). Let us take $x \in E$ and fix $y = f(x)$. Then:

- y non-standard $\Rightarrow$ x non-standard,
- x non-standard and $f^{-1}(y)$ finite $\Rightarrow$ y non-standard.

We freely apply this proposition to the following framework: E and F are the standard algebraic curves (or the set of points of algebraic coordinates on these curves) and f is a standard, non-constant regular map between these curves. Therefore f has finite degree, its fibers are finite sets. The proposition shows that an algebraic point P on the first curve is non-standard iff the image of the point on the second curve it non-standard. The particular case that gives this result which is obtained by taking F to be the right projection is sufficiently important to merit a special mention.

(3.5) Theorem. Let C be a standard algebraic curve defined on a field of numbers (standard). Let us choose and fix a point P non-standard on C with algebraic coordinates: $P \in \mathbb{C}(\overline{\mathbb{Q}})$. Let F denote the field $\overline{\mathbb{Q}}(C)$ of rational functions $C \to \mathbb{P}^1$ defined on $\overline{\mathbb{Q}}$. Thus

- For all $f \in F$,

$$f \text{ standard and } f \neq 0 \Rightarrow 0 \neq f(P) \neq \infty,$$

- For all $f \in F$,

$$f \text{ standard and non-constant} \Rightarrow f(P) \text{ non-standard},$$

- For two functions $f, g \in F$

$$f \text{ and } g \text{ standard}, f \neq g \Rightarrow f(P) \neq g(P).$$

Some commentary is necessary. Since the set $C(\overline{\mathbb{Q}})$ is always infinite, there exists a point P non-standard in this set. One such point plays the role of "generic point" at least if it is restricted to standard rational functions. The third property above shows the precise fashion in which to transition from standard functions to the values they take in P. The algebraic and geometric properties of the curve C can therefore be translated to arithmetic properties with

values $f(P)$ of standard rational functions on C. Furthermore, the coordinates of point P create a field of numbers, *k*. If possible, we will choose P in such a manner that this field of numbers is standard. Then it is only possible (according to Faltings' result) that the genus of the curve C is ≤ 1. Likewise, once the curve C is linear, say given by a polynomial equation with integer coefficients, one can try to choose the point P in a manner such that the coordinates create a standard ring R (necessarily of a finite type). It is only possible (according to Siegel's result) that the curve C is a rational curve. It is evident how the hypothesis that C has an infinite number of integer points or coordinates in a standard ring $R \subset \mathbb{Q}$ and of finite type on $\mathbb{Z}$ translates to the possibility of choosing the point P suitably in the theorem above.

4. Projective Heights

The height of a rational number written in reduced form $a = m/n$ (i.e. $n, m \in \mathbb{Z}$ and are relatively prime, and $n \geq 1$) is defined as the positive integer $H(a) = Max(|m|, n)$. It is evident then that

$H(a)$ entirely ≥ 1

$H\left(\frac{1}{a}\right) = H(a)$ if $a \neq 0$,

for every constant c, $\{a \in \mathbb{Q}: H(a) \leq c\}$ is finite.

When the constant c is standard, the set $\{a \in \mathbb{Q}: H(a) \leq c\}$ is standard and finite, only containing any standard elements. Consequently,

a rational and non-standard $\Rightarrow H(a)$ is unlimited.

The notion of height extends to the projective space $\mathbb{P}^n$ as follows.

If $a = [a_0: a_1: \ldots : a_n] \in \mathbb{P}^n(\mathbb{Q})$, we can assume that the chosen representation of a is such that $a_i \in \mathbb{Z}$ are relatively prime: this fixes the choice from representation to the exact sign. We pose

$$H(a) = Max\{|a_i|\} \text{ entirely} \geq 1.$$

When $n = 1$, we return to the preceding definition by immersing $\mathbb{Q}$ in $\mathbb{P}^1(\mathbb{Q})$ in the usual fashion $a = \frac{m}{n} \to [m:n]$. As above, we again have $a \in \mathbb{P}^n(\mathbb{Q})$ and a non-standard $\Rightarrow H(a)$ unlimited at least once the integer n is standard, which is sometimes implicitly assumed in such statements.

It is again useful to expand the notion of height to points with algebraic coordinates. For that purpose, observe first that in $\mathbb{Q}$, if $a = m/n$ is a reduced expression,

$$H(a) = Max(|m|, n) = \prod_v Max(|a|_v, 1)$$

where v runs over the set of places of $\mathbb{Q}$: these are the prime numbers p and the Archimedean place v giving way to the usual absolute value. In effect, when p runs over the set of prime numbers, the product of absolute values $|a|_p > 1$ reconstitutes the denominator n of a. The product defining $H(a)$ will therefore be equal to n if $|a| = |a|_v < 1$ and equal to $n \cdot |a|_v = n\left|\frac{m}{n}\right| = |m|$ if $|a| > 1$ as desired. When $a = [a_0: a_1: \ldots : a_n] \in \mathbb{P}^n(\overline{\mathbb{Q}})$ has algebraic coordinates, let us say a fixed field of numbers, k, we pose

$$H(a)^{[k:\mathbb{Q}]} = \prod_v Max(|a_0|_v, |a_1|_v, \ldots, |a_n|_v)$$

where v runs over the set of places of k. The formula for the product shows that the choice for a particular representation of a does not influence the result of the maximum calculation. Furthermore, the exponent is chosen in such a manner that the result does not depend on the choice of the field of numbers, k, in which we find a representation of a.

When an algebraic curve is immersed or embedded in a projective space, it inherits a notion of height simply by being transported. One should study how the choice of submersion or immersion influences the behavior of the resulting height.

Recall that an algebraic map of degree d between projective spaces $\mathbb{P}^n$ and $\mathbb{P}^m$ is simply a map

$$f\colon [a_0\colon a_1\colon \dots\colon a_n] \to [b_0\colon b_1\colon \dots\colon b_m]$$

which can be expressed with components

$$b_j = f_j(a_0, a_1, \dots, a_n)$$

using homogenous forms f_j having all the same degree d. That is to say, the first fundamental result is as follows:

(4.1) Theorem. Let f be an algebraic map from $\mathbb{P}^n \to \mathbb{P}^m$ of degree d. Then there exists a constant $c > 0$ such that

$$c^{-1}H(P)^d \leq H\big(f(P)\big) \leq cH(P)^d$$

for all points $P \in \mathbb{P}^n(\overline{\mathbb{Q}})$.

The above inequalities are frequently shortened using the notation $H\big(f(P)\big) \asymp H(P)^d$. Using the logarithm of the height

$h(P) = log\ H(P)$,

the conclusion of the theorem can be rewritten

$h\big(f(P)\big) - d\ h(P)$ is limited for $P \in \mathbb{P}^n(\overline{\mathbb{Q}})$.

Assuming now that f is standard (therefore n, m and d are standard), we can find a standard upper bound of the function $h\big(f(P)\big) - dh(P)$. If P is non-standard, $H(P)$ and $h(P)$ are unlimited and

$\frac{h(f(P))}{h(P)} = d +$ infinitesimal,

That is to say

$$st \quad \frac{h(f(P))}{h(P)} = d.$$

But let us immediately observe that this equality is weaker than the conclusion of the theorem since it does not allow us to recover the inequalities from the theorem within ϵ in the following sense. For all $\epsilon > 0$, there exists a constant $c_\epsilon > 0$ such that

$$c_\epsilon^{-1} H(P)^{d-\epsilon} \leq H(f(P)) \leq c_\epsilon H(P)^{d+\epsilon}.$$

It is precisely this genre of inequalities which had been established by A. Weil in a more general setting (cf. [9]). In the case of the curves, we can therefore reformulate the results of Weil as follows using NSA.

(4.2) Theorem. Let C be an algebraic curve and f, g two immersions of C in projective spaces. We assume all givens are standard and defined on $\overline{\mathbb{Q}}$. Then if $P \in C(\overline{\mathbb{Q}})$ is a non-standard point,

$$st \ \frac{h(f(P))}{h(g(P))} = \frac{\deg(f)}{\deg(g)}$$

The degree of a non-constant map $f: C \to \mathbb{P}^n$ can be seen as the "generic" number of intersection points of $f(C)$ with a hyperplane of $\mathbb{P}^n$.

5. Valuations

Let C be an (irreducible) standard curve defined on $\mathbb{Q}$ and F the field of rational functions. Suppose there exists a non-standard point P of $C(\mathbb{Q})$ and we take a standard rational function $f \in F - \mathbb{Q}$ non-constant. There exists a place v of $\mathbb{Q}$ such that $|f(P)|_v$ is unlimited: We

will assume that it is possible to choose v standard with this property. With these conventions and notations, we get the following result.

(5.1) Theorem. There is a unique standard valuation m_v of F such that

$$m_v(g) = -st\left(\frac{\log|g(P)|_v}{\log|f(P)|_v}\right)$$

for all standard functions $g \in F - \{0\}$. This valuation is trivial on $\mathbb{Q}$.

Proof. When g is standard, it is easily shown that the quotient of the logarithms is limited and it is therefore legitimate to take the standard part. The function m_v is perfectly well defined. The addition of the logarithm provides

$$m_v(g_1 g_2) = m_v(g_1) + m_v(g_2)$$

(firstly, for the standard g_i). The map m_v is a homomorphism $F \to R^+$. It remains to verify that

$$m_v(g_1 + g_2) \geq Inf(m_v(g_1), m_v(g_2))$$

when the two arguments are defined. We can assume that g_i are standard. This inequality is clearly satisfied when the place v is non-Archimedean as in the case

$$|\, g_1(P) + g_2(P)|_v \leq Max\ (|g_1(P)|_v, |g_2(P)|_v).$$

In the case where v is Archimedean, we have

$$|a + b| \leq |a| + |b| \leq 2\ Max(|a|, |b|)$$

and $\frac{\log 2}{\log|f(P)|} = 0$ shows that the factor of 2 disappears when taking the standard part. (For the complex place defined by $|a|_{\mathbb{C}} = a\bar{a} = |a|^2$ we have $|a + b|_{\mathbb{C}} \leq 4\ Max\ (|a|_{\mathbb{C}}, |b|_{\mathbb{C}})$ and the same reasoning would apply.) Let us finally show that m_v is trivial on the constants. As the function m_v is standard, it suffices to verify that $m_v(c) = 1$ for all nontrivial standard constants c. By definition, we have

$$m_v(c) = -\frac{st\log|c|_v}{\log|f(P)|_v} = 0$$

as $|c|_v$ and $\log|c|_v$ are limited.

When the curve C is projective and therefore complete, the valuation m_v is associated to a point $P_v \in C(\overline{\mathbb{Q}})$ in the following manner. There exists a constant $r_v > 0$ such that

$$ord_{P_v}(g) = r_v m_v(g) \text{ for } g \in F^{\times}.$$

Taking $g = f$, we find $r_v = -ord_{P_v}(f)$ and

(5.2) $$\frac{ord_{P_v}(g)}{ord_{P_v}(f)} = st\left\{\frac{\log|g(P)|_v}{\log|f(P)|_v}\right\}$$

for standard g. More precisely, since the coordinates of P_v are algebraic, the valuation m_v is associated to the prime rational series (standard) $Z_v = \sum P_v^{\sigma}$ sum of the conjugates of P_v on $\mathbb{Q}$. The exponent runs over the finite set of submersions $\mathbb{Q}(P_v) \to \overline{\mathbb{Q}}$ of the field generated by the coordinates of P_v and for all $g \in F^x$, the orders of g in P_v and P_v^{σ} are the same, which allows us to define the order of g on Z_v as being the order of g in any of the P_v^{σ}.

6. Sketch of a Proof

To give an idea of the proof of Theorem B, let us use the following particular case. The curve C is projective, irreducible (non-singular) on $\mathbb{Q}$. Standard and of the genus $g \geq 1$. We assume that $C(\overline{\mathbb{Q}})$ contains a non-standard point P with a rational coordinate $f(P) \in \mathbb{Q}$ (f is therefore a rational standard non-constant function). The hypothesis for this genus implies that C admits the unramified coverings of arbitrarily large degrees. As a consequence the denominator of $f(P)$ is itself also "large". To simplify, we assume that C admits a large standard covering $\pi\colon \tilde{C} \to C$ of degree m and defined over $\mathbb{Q}$, with a rational function $\tilde{f}$ on $\tilde{C}$ of the same degree as f and also defined on $\mathbb{Q}$. Finally, assume that it is possible to find a point $\tilde{P} \in C(\overline{\mathbb{Q}})$ with $\tilde{f}(\tilde{P}) \in \mathbb{Q}$ and above $P\colon \pi(\tilde{P}) = P$. In summary, we therefore have the situation depicted in the diagram below.

$$\begin{array}{c} \tilde{P} \in \tilde{C}(\overline{\mathbb{Q}}) \xrightarrow{\tilde{f}} P^1(\overline{\mathbb{Q}}) \ni \tilde{f}(\tilde{P}) = \tilde{a} \in \mathbb{Q} \\ \pi \downarrow \deg m \\ P \in C(\overline{\mathbb{Q}}) \xrightarrow{f} P^1(\overline{\mathbb{Q}}) \ni f(P) = a \in \mathbb{Q} \end{array}$$

The result (4.2) of Weil gives

$$st\frac{h(a)}{h(\tilde{a})} = \deg f \circ \frac{\pi}{\deg \tilde{f}} = \deg \pi = m$$

as f and $\tilde{f}$ have the same degree. Furthermore, the result (5.2) gives us an estimation of the logarithmic quotient of the absolute values $|\tilde{a}|_v$ and $|a|_v$ in which they are unlimited.

$$\text{st}\,\frac{\log|a|_v}{\log|\tilde{a}|_v} = \text{st}\frac{\log|f(P)|_v}{\log|f(\tilde{P})|_v} = \text{st}\frac{\log|f\cdot\pi(\tilde{P})|_v}{\log|f(\tilde{P})|_v} = \frac{ord_{Z_v}(f\cdot\pi)}{ord_{Z_v}(f)} = r \qquad (1 \le r \le \deg f)$$

By comparison

(6.1) $\quad \text{st}\frac{h(a)}{\log|a|_v} = \left(\frac{m}{r}\right)\text{st}\,h(\tilde{a})/\log|\tilde{a}|_v.$

The quantity $h(a)/\log|a|_v$ is interpreted as the "spread" of the denominator of a. Let us in fact take for v the place with maximal $|a|_v$. We then have

$$|a|_v \le H(a) \le |a|_v^s$$

where s designates the number of places w with $|a|_w > 1$. Through taking the logarithm, we arrive at

$$1 \le \frac{h(a)}{\log|a|_v} \le s.$$

Therefore (6.1) provides

$$s \ge st\frac{h(a)}{\log|a|_v} \ge \frac{m}{r}.$$

Since the integer m is arbitrary (but standard), we conclude that s is unlimited. This case clearly demonstrates why the first coordinate $a = f(P)$ of P would not be an integer.

7. Objections and Conclusions

In general, when we consider a coating $\pi\colon \tilde{C} \to C$, a point $\tilde{P}$ above P will have coordinates in a field of numbers (finite extension of $\mathbb{Q}$). By definition of the height of the algebraic number $\tilde{a} = \tilde{f}(\tilde{P})$, we will only have

$$|\tilde{a}|_v \leq H(\tilde{a})^{[k:\mathbb{Q}]}$$

from which

$$\frac{h(\tilde{a})}{\log|\tilde{a}|_v} \geq \frac{1}{[k:\mathbb{Q}]}.$$

This is no longer sufficient to decide the spread of the denominator of a: when the degree m of the covering goes up, the degree $[k:\mathbb{Q}]$ could also go up! The Roth Theorem supplies, on the contrary, a universal lower bound, independent of the degree of the algebraic number in consideration. It is written

$$st\ h(a)/\log|a|_v \geq \frac{1}{2}$$

for all places v of $\overline{\mathbb{Q}}$ and all non-standard algebraic numbers $a \in \overline{\mathbb{Q}}$.

Furthermore, the explicit construction of coverings of C can be carried out by submerging C in its Jacobian A and by taking for $\tilde{C}$ the reciprocal image of C by integer multiplication with r.

$$\begin{array}{ccc} \tilde{C} = [r]^{-1}(C) & \to & C \\ \downarrow & & \downarrow \\ A & \to & A \end{array}$$

$$P \mapsto P + P + \cdots + P = [r]P$$

$$(\text{of degree } m = r^2)$$

Using the weaker part of the Mordell-Weil theorem,

For all fields of k numbers and all abelian varieties A defined on k, the groups $A(k)/[r]A(k)$ are finite,

we show how the hypotheses made in the preceding section can be realized. The reader interested in the details of the case of genus 1 (elliptic curves) can reference the book by Silvermann [8].