Diophantine equations

F.Beukers

Spring 2011

2 Mordell's equation

2.1 Introduction

Let $d \in \mathbb{Z}$ with $d \neq 0$ and consider the equation $y^2 + d = x^3$ in $x, y \in \mathbb{Z}$. This equation is known as *Mordell's equation*. We shall prove the following Theorem.

Theorem 2.1.1 (Mordell, 1922) *For given $d \neq 0$, the equation $y^2 + d = x^3$ in $x, y \in \mathbb{Z}$ has at most finitely many solutions.*

Actually Mordell proved a more general theorem, but we will come back to that later. It should be emphasized that Mordell's proof is only a finiteness result, no algorithm is provided to actually solve the equation. Nowadays we also have methods to solve the equation explicitly. The first results in this direction are based on A.Baker's technique of *linear forms in logarithms* starting in 1966. This work earned Baker the Fields medal. Here is a theorem based on Baker's methods.

Theorem 2.1.2 (Sprindzuk, 1982) *There exists an effectively computable number $C > 0$ such that any solution $x, y \in \mathbb{Z}$ of $y^2 + d = x^3$ with $d \neq 0$ satisfies*

$$|x|, |y| \leq \exp\left(C|d|(\log|d| + 1)^6\right).$$

Note that the bound for x, y is roughly exponential in $|d|$ with a very large constant C. One expects that a much sharper bound holds. This is based on the following conjecture.

Conjecture 2.1.3 (Hall, 1971) . *To every $\epsilon > 0$ there is a positive real number $c(\epsilon)$ such that*

$$|y^2 - x^3| > C(\epsilon)x^{1/2-\epsilon}$$

for any $x, y \in \mathbb{Z}_{>0}$ with $y^2 \neq x^3$.

Actually Hall conjectured the lower bound $Cx^{1/2}$ for some $C > 0$, but this is generally believed not te be true.
As a consequence of Hall's conjecture we see that $|x|, |y| \leq c_1(\epsilon)|d|^{2+\epsilon}$. In other words, the expected upper bounds for x, y are polynomial in $|d|$.
Nowadays there are explicit algorithms to solve Mordell's equation. In [GPZ] all equations with $|d| \leq 10000$ are solved. A particularly spectacular example is $y^2 - 17 = x^3$. In 1930 T.Nagell showed that the complete set of solutions with $y > 0$ reads,

$$(x, y) = (-2, 3), (-1, 4), (2, 5), (4, 9), (8, 23), (43, 282), (52, 375), (5234, 378661)$$

References:

[M] L.J.Mordell, *Diophantine Equations*, Academic Press 1969, Chapter 26

[LF] H.London, R.Finkelstein, *On Mordell's equation* $y^2 - k = x^3$, Bowling Green State University Press 1973.

[GPZ] J.Gebel, A.Pethö, H.G. Zimmer, *On Mordell's equation* Compositio Math 110 (1998), 335-367.

2.2 Special cases

Using methods from elementary algebraic number theory we can deal with certain sets of Mordell equations.

Proposition 2.2.1 *Let $d > 0$, d square-free, $d \not\equiv -1 (\mathrm{mod}\ 4)$ and $h(\mathbb{Q}(\sqrt{-d}))$ is not divisble by 3. Suppose $y^2 + d = x^3$ has a solution. Then $d = 3a^2 \pm 1$ for some $a \geq 0$ and some choice of $\pm$ sign. The solution set consists of $(x, y) = (4a^2 - 1, \pm(8a^3 - 3a))$ if $d = 3a^2 - 1$ and $(x, y) = (4a^2 + 1, \pm(8a^3 + 3a))$ if $d = 3a^2 + 1$.*

This proposition implies for example that $y^2 + 1 = x^3$ has no non-trivial solutions and that $y^2 + 2 = x^3$ has $(x, y) = (3, \pm 5)$ as solution set. The latter fact was stated already by Fermat, but not proved.
Here is a proof of our Proposition. Suppose $y^2 + d = x^3$. First we note that $\gcd(x, 2d) = 1$. For if an odd prime p divides both d and x we see that p should divide y as well. But since p^2 divides both x^3 and y^2 we find that p^2 divides d, contradicting the fact that d is square-free. If x were even, then $y^2 \equiv -d (\mathrm{mod}\ 8)$. But since $-d \not\equiv 0, 1 (\mathrm{mod}\ 4)$ we get a contradiction again. We now assume that $\gcd(x, 2d) = 1$.
We obtain the following factorisation

$$(y + \sqrt{-d})(y - \sqrt{-d}) = x^3.$$

Since d is square-free and $d \not\equiv 1 (\mathrm{mod}\ 4)$ the ring of integers in $\mathbb{Q}(\sqrt{-d})$ is $\mathbb{Z}[\sqrt{-d}]$. Let $\wp$ be a prime ideal divisor of $(y + \sqrt{-d}, y - \sqrt{-d})$. Then it also divides x and $2\sqrt{-d}$. Hence its divides x and $2d$. This contradicts $\gcd(x, 2d) = 1$ and we conclude that the principal ideals

$(y+\sqrt{-d})$ and $(y-\sqrt{-d})$ are relatively prime. Their product is a cube and so we conclude that $(y+\sqrt{-d})$ itself is the cube of an ideal, which we call I. So we get $(y+\sqrt{-d}) = I^3$. Note that I^3 is a principal ideal. Hence its order in the ideal class group is either 1 or 3. But we are given that the class number is not divisible by 3. So the order of I is the ideal class group is 1, hence I is principal. There exist $a, b \in \mathbb{Z}$ such that $I = (a + b\sqrt{-d})$. Hence

$$y + \sqrt{-d} = \epsilon(a + b\sqrt{-d})^3$$

where ϵ is a unit in the ring of integers. When $d > 1$ we have $\mathbb{Z}[\sqrt{-d}]^* = \{\pm 1\}$ and $\mathbb{Z}[\sqrt{-1}]^* = \{\pm 1, \pm i\}$. In both cases the unit group has order relatively prime to 3, hence every unit can be considered as the cube of another unit. After redefining a, b if necessary, we get

$$y + \sqrt{-d} = (a + b\sqrt{-d})^3$$

Comparison of the coefficients before $\sqrt{-d}$ gives $1 = 3a^2b - db^3 = b(3a^2 - db^2)$. We see that $b = \pm 1$ and $3a^2 - db^2 = \pm 1$. If $b = 1$, then $3a^2 - d = 1$ and so, $d = 3a^2 - 1$. The value of y is $a^3 - 3ab^2d = a^3 - 3a(3a^2 - 1) = -(8a^3 - 3a)$. The value of x is $a^2 + db^2 = a^2 + 3a^2 - 1 = 4a^2 - 1$. When $b = -1$ we proceed similarly. **qed**

In a very similar way we can show that

Proposition 2.2.2 *Let $d > 0$, d square-free, $d \equiv -5 (\mathrm{mod}\ 8)$ and $h(\mathbb{Q}(\sqrt{-d}))$ not divisible by 3. Suppose that $y^2 + d = x^3$ has a solution. Then one of the following cases holds,*

1. *There exist $a \in \mathbb{Z}$ and $\epsilon \in \{\pm 1\}$ such that $d = 3a^2 + \epsilon$. The solutions read $(x, y) = (4a^2 + \epsilon, \pm(8a^3 + 3\epsilon a))$.*

2. *There exist $a \in \mathbb{Z}$ and $\epsilon \in \{\pm 1\}$ such that $d = 3a^2 + 8\epsilon$. The solutions read $(x, y) = (a^2 + 2\epsilon, \pm(a^3 + 3\epsilon a))$.*

We can apply this Proposition to $y^2 + 11 = x^3$. Note that $d = 11$ satisfies all of our conditions. Moreover, $11 = 3 \cdot 1^2 + 8$, which gives rise to the solutions $(x, y) = (3, \pm 4)$. In addition, $11 = 3 \cdot 2^2 - 1$ giving rise to $(x, y) = (15, \pm 58)$.
So far we dealt with $d > 0$ in our equation. Let us consider an example with $d < 0$, namely $y^2 - 17 = x^3$. This known to have the solution set

$$(x, y) = (-2, \pm 3), (-1, \pm 4), (2, \pm 5), (4, \pm 9), (8, \pm 23), (43, \pm 282), (52, \pm 375), (5234, \pm 378661)$$

We make a beginning with its solution. We factor as

$$(y + \sqrt{17})(y - \sqrt{17}) = x^3$$

in the field $K = \mathbb{Q}(\sqrt{17})$. In K we have unique factorization into irreducibles and its group of units is generated by -1 and $4 + \sqrt{17}$. A further remark is that $(5 \pm \sqrt{17})/2$ are the irreducible divisors of 2. The ring of integers is $\mathbb{Z}[(1 + \sqrt{17})/2]$. Let π be a common prime divisor of $y + \sqrt{17}$ and $y - \sqrt{17}$. Then π divides both $2y$ and $2\sqrt{17}$. If $\pi = \sqrt{17}$ then y is divisible by 17, as well as x. The difference $17 = x^3 - y^2$ would then be divisible by 17^2 which

is not possible. We conclude that π divides 2. But then x is even. We separate according to the cases x even or odd.
Suppose x is odd. Then, by the above, $y+\sqrt{17}$ and $y-\sqrt{17}$ have no common prime divisor and hence, by unique factorization, there exists an integer $\alpha \in \mathcal{O}_K$ and a unit η such that

$$y+\sqrt{17} = \eta\alpha^3.$$

The units η are of the form $\pm(4+\sqrt{17})^k$. Of course -1 is a cube and the unit η is a cube times $1, 4+\sqrt{17}$ or $4-\sqrt{17}$. Hence there exists an integer $\alpha \in \mathcal{O}_K$ such that $y+\sqrt{17}$ equals one of the following

$$\alpha^3, \quad (4+\sqrt{17})\alpha^3, \quad (4-\sqrt{17})\alpha^3.$$

Let us write $\alpha = (a+b\sqrt{17})/2$ with $a, b \in \mathbb{Z}$ having the same parity. Then, in the case $y+\sqrt{17} = \alpha^3$ comparison of the coefficients of $\sqrt{17}$ gives

$$8 = 3a^2b + 17b^3.$$

Since $b \neq 0$ implies $3a^2 + 17b^2 \geq 17$ we see we get a contradiction. Comparison of the coefficients of $\sqrt{17}$ in $y+\sqrt{17} = (4+\sqrt{17})\alpha^3$ yields

$$8 = a^3 + 12a^2b + 51ab^2 + 68b^3.$$

Replace a by $a-4b$ to get

$$8 = a^3 + 3ab^2 - 8b^3.$$

Hence $a(a^2+3b^2) \equiv 0 (\text{mod } 8)$, which implies that a, b should both be even. So replace a, b by $2a, 2b$ to get

$$1 = a^3 + 3ab^2 - 8b^3$$

We do not solve this equation, but note that the solutions $(a,b) = (1,0), (-3,-2)$ give rise to the solutions $(x,y) = (-1,4), (43,282)$ of the Mordell equation. The case $y+\sqrt{17} = (4-\sqrt{17})\alpha^3$ runs similarly.
Suppose now that x is even. The y is odd and $y \pm \sqrt{17}$ divisible by 2. Hence, upon replacing x by $2x$,

$$\frac{y+\sqrt{17}}{2}\ \frac{y-\sqrt{17}}{2} = 2x^3$$

From this we deduce the following possibilities

$$\frac{y+\sqrt{17}}{2} = \frac{5\pm\sqrt{17}}{2}\alpha^3, \frac{5\pm\sqrt{17}}{2}(4\pm\sqrt{17})\alpha^3.$$

for choice of $\pm$ signs and some algebraic integer $\alpha = (a+b\sqrt{17})/2$.
The first case with $+$ sign gives

$$8 = a^3 + 15a^2b + 51ab^2 + 85b^3$$

Replace a by $a - 5b$ to get

$$8 = a^3 - 24ab^2 + 80b^3$$

Hence a is even. Replace a by $2a$ to get

$$1 = a^3 - 6ab^2 + 10b^3$$

we have the small solutions $(a, b) = (1, 0), (-3, 1)$. They give rise to the solutions $(x, y) = (2, 5), (52, -375)$ of Mordell's equation.
In a similar way the other cases also give rise to diophantine equations of the form $f(x, y) = 1$ for cubic homogeneous polynomials $f \in \mathbb{Z}[x, y]$.

In the following section we prove the following theorem.

Theorem 2.2.3 *For any $k \in \mathbb{Z} \neq 0$ the solution of the diophantine equation $y^2 + k = x^3$ in $x, y \in \mathbb{Z}$ can be reduced to the solution of a finite set of diophantine equation of the form $f(x, y) = 1$ in $x, y \in \mathbb{Z}$ where f is a binary cubic form with integer coefficients. Moreover, the set of forms f can be computed explicitly.*

2.3 Binary forms

A binary form is a polynomial in two variables. The general shape of a binary form of degree n reads

$$a_n X^n + a_{n-1} X^{n-1} Y + a_{n-2} X^{n-2} Y^2 + \cdots + a_1 X Y^{n-1} + a_0 Y^n$$

Two binary forms $f(X, Y), g(X, Y)$ are called equivalent if there exist p, q, r, s with $ps - qr = 1$ such that $g(X, Y) = f(pX + qY, rX + sY)$.
In *invariant theory* one looks for polynomials in the coefficients $a_0, a_1, \ldots, a_n$ which are invariant under equivalence transformations. The most familiar one is the *discriminant* of a form $f(X, Y)$, defined by

$$D = a_n^{2n-2} \prod_{i<j} (\alpha_i - \alpha_j)^2$$

where $\alpha_1, \ldots, \alpha_n$ are the zeros of the polynomial $f(X, 1)$. One can show that $D \in \mathbb{Z}[a_0, a_1, \ldots, a_n]$.
Here are two examples,
Binary quadratic forms $aX^2 + 2bXY + cY^2$ with discriminant

$$D = 4(b^2 - ac).$$

Binary cubic forms $aX^3 + 3bX^2Y + 3cXY^2 + dY^3$ with discriminant

$$D = 27(-a^2d^2 + 6abcd + 3b^2c^2 - 4ac^3 - 4db^3).$$

For quadratic and cubic forms the discriminant D and polynomials in D are the only invariants. For quartic forms $a_4X^4 + 4a_3X^3Y + 6a_2X^2Y^2 + 4a_1XY^3 + a_0Y^4$ there are two independent invariants namely

$$\begin{aligned} I_2 &= a_0a_4 - 4a_1a_3 + 3a_2^2 \\ I_3 &= a_0a_2a_4 - a_0a_3^2 - a_1^2a_4 + 2a_1a_2a_3 - a_2^3 \end{aligned}$$

The ring of invariants is the polynomial ring generated by I_2 and I_3. In particular, $D = 27(I_2^2 - 27I_3^3)$.
We shall now concentrate on binary forms with $a_0, a_1, \ldots, a_n \in \mathbb{Z}$ and call them integral binary forms. In particular the discriminant is an integer. Two integral forms $f(X,Y), g(X,Y)$ will be called $SL(2,\mathbb{Z})$-equivalent, or simply equivalent, if there exist $p,q,r,s \in \mathbb{Z}$ with $ps-qr=1$ such that $g(X,Y) = \pm f(pX+qY, rX+sY)$. We have the following Theorem.

Theorem 2.3.1 *The number of equivalence classes of binary integral forms of given degree and given discriminant is finite.*

For quadratic forms there is a very explicit reduction procedure from which finiteness of the number of equivalence classes of discriminant D follows. Let us start with an arbitrary quadratic form $aX^2+2bXY+cY^2$ which we abbreviate by $[a,b,c]$. Note that we have chosen the coefficient of XY to be even. Such quadratic forms are called *even*. Although one could also consider odd quadratic forms we concentrate here only on th even ones.
We keep on repeating the following steps. If $|b| > |a|/2$ we choose k such that $|b+ka| \leq |a|/2$. Replace X by $X+kY$ to get the new form $[a,b,c] := [a, b+ka, c+2bk+ak^2]$. If $|c| < |a|$ we make the substitution $(X,Y) \to (-Y,X)$ which changes our form into $[a,b,c] := [c,-b,a]$. Repeating this procedure we end up with an equivalent form $[a,b,c]$ which satisfies the inequalities $|2b| \leq |a| \leq |c|$. From this we derive that $|D| \geq |ac| - b^2 \geq 3b^2$. Hence $|b|$ is bounded by $\sqrt{|D|/3}$. This gives a finite number of values of b and through $b^2 - ac = D$ we get a finite number of values of a, c.
Example. We determine all equivalence classes of even quadratic forms $aX^2+2bXY+cY^2$ with $b^2-ac=17$. According to the above reduction procedure we can restrict ourselves to a,b,c satisfying $|b| \leq \sqrt{17/3}$. So $b = 0, \pm 1, \pm 2$. The corresponding a, c follow from $b^2-ac=17$ and $|c| \geq |a| \geq |2b|$. We get the following list of possibilities with $a>0$,

$$X^2 - 17Y^2$$
$$2X^2 \pm 2XY - 9Y^2$$
$$3X^2 \pm 2XY - 6Y^2$$

We now turn to cubic forms $f(X,Y) = aX^3 + 3bX^2Y + 3cXY^2 + dY^3$. We construct the Hessian form

$$H(X,Y) = \frac{-1}{36}\begin{vmatrix} f_{XX} & f_{XY} \\ f_{XY} & f_{YY} \end{vmatrix}$$

which turns out to be $H(X,Y) = (b^2-ac)X^2 + (bc-ad)XY + (c^2-bd)Y^2$. We also define the cubic form

$$G(X,Y) = \frac{1}{3}\begin{vmatrix} f_X & f_Y \\ H_X & H_Y \end{vmatrix}.$$

The forms H, G are called the covariants of f of degrees 2 and 3. The discriminant of f equals $27D_1$ where $D_1 = -a^2d^2 + 6abcd + 3b^2c^2 - 4ac^3 - 4db^3$. The discriminant of H equals $-D_1$.

Proposition 2.3.2 *Let notation be as above. Then*

$$G^2 + D_1 f^2 = 4H^3$$

We can now make the following observation. Let f be a binary cubic form with $D_1 = 4k$ such that $f(x, y) = 1$ has the solution x_0, y_0. Then Mordell's equation $y^2 + k = x^3$ has the solution $y = G(x_0, y_0)/2$, $x = H(x_0, y_0)$. It turns out that the converse is also true.

Proposition 2.3.3 *Consider the equation $y^2 + k = x^3$ and suppose that we have a solution p, q. Then the cubic form $f(x, y) = x^3 - 3pxy^2 + 2qy^3$ has $D_1 = 4k$ and $p = H(1, 0), q = G(1, 0)/2$. Note in addition that $H(X, Y) = pX^2 - 2qXY + p^2Y^2$, so H is an even form. We also have $G(X, Y) = 2(-qX^3 + 3p^2X^2Y - 3pqXY^2 + (-p^3 + 2q^2)Y^3)$, i.e. $G(X, Y)/2$ is an integral form.*

The proof of this Proposition follows by direct computation.
To solve Mordell's equation it suffices to find a representing element of each equivalence class of integral cubic forms with discriminant $108k$. For each such form f we solve the equation $f(x, y) = 1$ in $x, y \in \mathbb{Z}$. The latter equation is known as a cubic Thue equation. In the next section we will see that it has finitely many solutions. Assuming this we now see that Mordell's finiteness result 2.1.1 follows.

3 Thue's equation

3.1 Introduction

Let F be an integral binary form and m a non-zero integer. The equation

$$F(x, y) = m$$

in $x, y \in \mathbb{Z}$ is called the *Thue equation*.

Theorem 3.1.1 (Thue, 1909) *Let F be an integral binary form such that $F(x, 1)$ has at least three distinct zeros. Let m be a non-zero integer. Then the equation $F(x, y) = m$ has at most finitely many solutions.*

Note that if F is reducible over $\mathbb{Z}$ then we can restrict ourselves to equations of the form $G(x, y) = m'$ where G is an irreducible factor of F and m' a divisor of m. Notice also that the requirement of at least three zeros is essential. An example of a quadratic equation would be Pell's equation $x^2 - dy^2 = 1$ which is known to have infinitely many solutions if d is a positive integer and not a square.
Using Thue's theorem and Proposition 2.3.3 we conclude that Mordell's equation has at most finitely many solutions. Thue's theorem is proved using methods from diophantine approximation. Due to the nature of this technique Thue's theorem is only a finiteness statement. It does not give a method to solve the equation. We shall come back to this. An effective method to solve Thue's equation became available through A.Baker's method on linear forms in logarithms around 1966. As an application of these methods the following was shown.

Theorem 3.1.2 (Feld'man, Baker) *Suppose that $F(x, y)$ is a form in two variables such that $F(x, 1)$ has at least three disctinct zeros. Then there exist positive, effectively computable numbers C_1, C_2, depending only on F such that any solution $x, y \in \mathbb{Z}$ of $F(x, y) = m$ (with $m \neq 0$) satisfies*

$$\log(\max(|x|, |y|) \leq C_1 |m|^{C_2}.$$

By 'effective method' we mean that the upper bound for x, y provides us with an algorithm to determine the solution set. However, due to the enormous size of this bound the algorithm is certainly not efficient. With the speed of present day computers a naive search of x, y up to the bound given above would take the life time of the universe and more. So extra ideas have to be invoked to solve the equation.
In the years before the 1980's about the only such method was *Skolem's method*, next to simple minded congrence considerations which work only in rare cases. In solving Thue's equation there is a big difference between the cases when F has a positive or a negative discriminant as we shall see
As a first example consider $f(x, y) = 1$ where $f(x, 1)$ is a monic cubic irreducible polynomial. Let K be the field $\mathbb{Q}[x]/(f(x, 1))$. Write $f(x, 1) = (x - \alpha)(x - \alpha')(x - \alpha'')$ where $\alpha \in K$ and

α', α'' are its algebraic conjugates. Then the equation $f(x, y = 1$ implies that $x - \alpha y = \beta$ where β is a unit in K. We also have the conjugate equations $x - \alpha' y = \beta'$ and $x - \alpha'' y = \beta''$. As an exercise one can verify that

$$\frac{1}{f'(\alpha)} + \frac{1}{f'(\alpha')} + \frac{1}{f'(\alpha'')} = \frac{\alpha}{f'(\alpha)} + \frac{\alpha'}{f'(\alpha')} + \frac{\alpha''}{f'(\alpha'')} = 0.$$

As a consequence we get

$$\frac{\beta}{f'(\alpha)} + \frac{\beta'}{f'(\alpha')} + \frac{\beta''}{f'(\alpha'')} = 0.$$

Now suppose that K has negative discriminant, which is equivalent with $r_1 = r_2 = 1$. By Dirichlet's unit theorem the units in K are of the form $\pm\eta^n$ where η is a fundamental unit and $n \in \mathbb{Z}$. So our equation becomes

$$\frac{\eta^n}{f'(\alpha)} + \frac{(\eta')^n}{f'(\alpha')} + \frac{(\eta'')^n}{f'(\alpha'')} = 0.$$

Note that we have turned our Thue equation into an exponential equation in the unknown exponent n. In the next section we show how to deal with this equation.
Consider the explicit example

$$x^3 - xy^2 + y^3 = 1$$

Its solution set reads $(x, y) = (1, 0), (0, 1), (1, 1), (-1, 1), (4, -3)$, which was shown by T.Nagell. The discriminant of the form $x^3 - xy^2 + y^3$ is -23. This is the minimal negative discriminant possible for an irreducible cubic form. The polynomial $X^3 - X + 1$ has a real zero and two complex ones, this is because the discriminant is negative. Let α be a zero. Then the field $\mathbb{Q}(\alpha)$ has $r_1 = r_2 = 1$. Its ring of integers is $\mathbb{Z}[\alpha]$ and the group of units $\mathbb{Z}[\alpha]^* = \{\pm\alpha^n | n \in \mathbb{Z}\}$. We compute that $1/f'(\alpha) = (4 - 9\alpha - 6\alpha^2)/23$. So our exponential equation becomes

$$\theta\alpha^n + \theta'(\alpha')^n + \theta''(\alpha'')^n = 0$$

where $\theta = -4 + 9\alpha + 6\alpha^2$ and θ', θ'' are its conjugates.

3.2 Skolem's method

Here is a Proposition which solves exponential equations of the type we have just seen.

Proposition 3.2.1 (Skolem's method) . *Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be a set of algebraic integers. Choose an odd rational prime p and $m \in \mathbb{N}$ such that there exist an integer $a \in \mathbb{Z}$, not divisible by p and algebraic integers $\beta_1, \ldots, \beta_k$ with the property that $\alpha_i^m = a + p\beta_i$ for $i = 1, 2, \ldots, k$. Let $\theta_1, \ldots, \theta_k$ be algebraic integers and suppose that $\theta_1 + \cdots + \theta_k = 0$ and $\theta_1\beta_1 + \cdots + \theta_k\beta_k \not\equiv 0 (\mathrm{mod}\ p)$. Then $\theta_1\alpha_1^{mn} + \cdots + \theta_k\alpha_k^{mn} = 0$ implies $n = 0$.*

The proof really depends on the use of so-called p-adic numbers, but here we give a version which avoids mentioning them.

Let us rewrite our equation by using the binomial theorem for $\alpha_i^{mn} = (a + p\beta_i)^n$. We get

$$0 = npa^{n-1}(\theta_1\beta_1 + \cdots + \theta_k\beta_k) + \sum_{r=2}^{n} \binom{n}{r} p^r a^{n-r}(\theta_1\beta_1^r + \cdots + \theta_k\beta_k^r)$$

Here we have used that $\theta_1 + \cdots + \theta_k = 0$. Now assume that n is not zero and divide on both sides by npa^{n-1}. Using $\frac{1}{n}\binom{n}{r} = \frac{1}{r}\binom{n-1}{r-1}$ we get,

$$0 = \theta_1\beta_1 + \cdots + \theta_k\beta_k + \sum_{r=2}^{n} \frac{p^{r-1}}{ra^{r-1}} \binom{n-1}{r-1} (\theta_1\beta_1^r + \cdots + \theta_k\beta_k^r).$$

Since p is odd prime the fraction $\frac{p^{r-1}}{ra^{r-1}}$ has the factor p in its numerator for every $r \geq 2$. In particular it follows form our equation that $\theta_1\beta_1 + \cdots + \theta_k\beta_k$ is divisible by p. This is impossible by our assumption. Hence we conclude that $n = 0$. **qed**

Here is an application to the explicit equation of the previous section. Let $\alpha_1, \alpha_2, \alpha_3$ be the zeros of $X^3 - X + 1$. We want to solve

$$\mu_1\alpha_1^n + \mu_2\alpha_2^n + \mu_3\alpha_3^n = 0.$$

in $n \in \mathbb{Z}$, where $\mu_i = -4 + 9\alpha + 6\alpha^2$. We note that $\alpha_i^{24} = 1 + 5\beta_i$ where $\beta_i = 30 - 53\alpha_i + 40\alpha_i^2$. We write $\mathrm{Tr}(\mu\alpha^n) = \mu_1\alpha_1^n + \mu_2\alpha_2^n + \mu_3\alpha_3^n$, which is a rational integer called the trace of the algebraic number $\mu\alpha^n$.
Using PARI we see that $\mathrm{Tr}(\mu\alpha^n) \equiv 0 (\mathrm{mod}\ 5)$ implies that $n \equiv 0, 1, 3, 11, 20 (\mathrm{mod}\ 24)$. Interestingly enough each of these congruence classes corresponds to an exact solution. For $\mathrm{Tr}(\mu\alpha^n) = 0$ we have the solutions $n = -13, -4, 0, 1, 3$ and we expect these to be the only solutions. We can check the latter expectation by solving $\mathrm{Tr}(\theta\alpha^{24n}) = 0$ for $\theta = \mu\alpha^{-13}, \mu\alpha^{-4}, \mu, \mu\alpha, \mu\alpha^3$ respectively. For all choices of θ except $\theta = \mu$ we can check that $\mathrm{Tr}(\mu\beta) \not\equiv 0 (\mathrm{mod}\ 5)$, so our Proposition applies to these four cases. We are left with the problem to solve $\mathrm{Tr}(\mu\alpha^{24n}) = 0$. Let us now put $\gamma = \alpha^{24}$. Then we note that $\gamma^2 = 1 + 7\delta$ for $\delta = 18400 - 32290\alpha + 24375\alpha^2$. We check that $\mathrm{Tr}(\mu\gamma) \not\equiv 0 (\mathrm{mod}\ 7)$, so $\mathrm{Tr}(\mu\alpha^{24n}) = 0$ implies that n is even. Therefore it remains to solve $\mathrm{Tr}(\mu\alpha^{48n}) = \mathrm{Tr}(\mu\gamma^{2n}) = 0$. We check that $\mathrm{Tr}(\mu\delta) \not\equiv 0 (\mathrm{mod}\ 7)$ and now apply our Proposition once more, where we take for α and β the numbers γ and δ and $p = 7$. We conclude that $n = 0$. In combination with the remarks at the end of the previous section we now conclude the following theorem.

Theorem 3.2.2 (T.Nagell) *The equation $x^3 - xy + y^3 = 1$ has the solutions*

$$(x, y) = (1, 0), (0, 1), (1, 1), (-1, 1), (4, -3).$$

Skolem's method is particularly suitable to solve cubic Thue equation with negative discriminant. A negative discriminant of a cubic is equivalent to the form having one real and two complex (non-real) solutions. To see what goes wrong in the case of a positive discriminant we take the equation

$$x^3 + x^2y - 2xy^2 - y^3 = 1$$

Baulin showed that the only solutions are

$$(x,y) = (1,0),(0,-1),(-1,1),(-1,-1),(2,-1),(-1,2),(5,4),(4,-9),(-9,5).$$

As a start we notice that the polynomial $x^3 + x^2 - 2x - 1$ has discrimant 49. As a result there are three real zeros. The cubic extension generated by one such zero α has three real embeddings, i.e. $r_2 = 3$. Using PARI we see that this cubic field has ring of integers $\mathbb{Z}[\alpha]$ and its unit group consists of the elements $\pm\alpha^k(1+\alpha)^l$, with $k, l \in \mathbb{Z}$. A solution x, y of the Thue equation above implies that $x - \alpha y$ in a unit in $\mathbb{Z}[\alpha]$, in other words $x - \alpha y = \pm\alpha^k(1+\alpha)^l$ with k, l. The difficulty now becomes clear, we have an exponential equation with two unknowns, k and l. Skolem's method does not work here and we have to go over to finite extensions of $\mathbb{Q}(\alpha)$ to be able to apply generalisations of Skolem's method. This is the reason why Baulin's paper consists of some 40 pages.

3.3 Thue's method

Let α be a fixed, real irrational number. Consider a rational approximation $\frac{p}{q}$ to α with $p, q \in \mathbb{Z}$, $q > 0$ and $\gcd(p,q) = 1$. The *quality* of this approximation is the number $M > 0$ such that

$$\left|\alpha - \frac{p}{q}\right| = \frac{1}{q^M}$$

or, if it does not exist, we take $M = 0$. As a first result we prove,

Theorem. Let α be an irrational number. Then there exist infinitely many approximations to α of quality ≥ 2.

This statement is part of the theory of continued fractions. But also without knowledge of continued fractions it is not hard to show. Fix a large positive integer Q and consider the set of numbers $\{q\alpha\}$ for $q = 0, 1, 2, \ldots, Q$, where $\{x\}$ denotes the difference between x and the largest integer $\leq x$. The set of $\{q\alpha\}$ is a set of $Q+1$ numbers in the interval $[0,1)$. So it tends to be crowded when Q gets large. In particular, there must be two values of q, say $q_1 < q_2$, such that the difference between $\{q_1\alpha\}$ and $\{q_2\alpha\}$ is less than $\frac{1}{Q}$ in absolute value. Choose integers p_1, p_2 such that $\{q_i\alpha\} = q_i\alpha - p_i$. Then, $|(q_2 - q_1)\alpha - (p_2 - p_1)| < \frac{1}{Q}$. Since clearly $0 < q_2 - q_1 \leq Q$ we see that $\frac{p_2 - p_1}{q_2 - q_1}$ is an approximation of quality at least 2. By choosing increasingly large values for Q we can produce an infinite sequence of such approximations. **qed**

Here is the quality of the two famous rational approximations to π,

$$\left|\frac{22}{7} - \pi\right| \approx \frac{1}{7^{3.429}} \quad , \quad \left|\frac{355}{113} - \pi\right| \approx \frac{1}{113^{3.201}}$$

One may wonder if an infinite number of such good quality approximations exist for π, or any other irrational we are looking at. To that end we introduce the following concept.

Definition The *irrationality measure* of an irrational number α is defined as the limsup over all qualities of all rational approximations and is denoted by $\mu(\alpha)$.

We have taken the limsup in our definition rather than the maximum since we are for example interested in the question whether π has infinitely many approximations of quality at least 3. The first two occurrences from the introduction may have been exceptional coincidences. If we assume that π behaves like most other numbers, then there is very little chance that $\mu(\pi) \geq 3$. This is shown by the following Theorem.

Theorem The set of irrational numbers with irrationality measure strictly larger than 2 has Lebesgue measure zero.

This Theorem is not hard to prove. Let us restrict ourselves to the irrational numbers in the interval $[0, 1]$. Choose $\epsilon > 0$. A number α with $\mu(\alpha) \geq 2 + 2\epsilon$ is, by definition, contained in an interval of the form

$$\left[\frac{p}{q} - \frac{1}{q^{2+\epsilon}}, \frac{p}{q} + \frac{1}{q^{2+\epsilon}}\right],$$

with $0 < p < q$ integers, infinitely many times. Let us give an upper bound for the total length of these intervals with $q > Q$, where Q is some large fixed positive integer. Such a bound can be given by

$$\sum_{q=Q+1}^{\infty} \sum_{p=1}^{q} \frac{2}{q^{2+\epsilon}}$$

The inner sum is equal to $\frac{2}{q^{1+\epsilon}}$. The sum over q can be estimated by the integral criterion,

$$\sum_{q=Q+1}^{\infty} \frac{2}{q^{1+\epsilon}} < \int_{Q}^{\infty} \frac{2dx}{x^{1+\epsilon}} = \frac{2}{\epsilon Q^{\epsilon}}.$$

When we let $Q \to \infty$ we see that the latter bound goes to zero. Hence the Lebesgue measure of the numbers in $[0, 1]$ with irrationality measure $\geq 2+2\epsilon$ is zero. The set of numbers in $[0, 1]$ with irrationality measure > 2 is the union of all sets of numbers with irrationality measure at least $2 + 2/n$ for $n = 1, 2, 3, 4, \ldots$. Since a countable union of measure zero sets has again measure zero, our result follows. **qed**

We note that numbers with irrationality measure > 2 do exist. In fact there exist irrational numbers with irrationality measure ∞. These are the so-called *Liouville numbers*. An example of such a number is given by

$$\sum_{n \geq 0} \frac{1}{2^{n!}}.$$

The reader may wish to verify as an exercise that the truncated series form a sequence of approximations whose qualities go to ∞. On the other hand, numbers like Liouville numbers are a bit artificial. They are constructed for the purpose of having large irrationality measures. It is expected that the irrationally measure for a naturally occurring number is 2. Unfortunately, there are not many instances where this is known. A classical instance is e.

The fact that $\mu(e) = 2$ can easily be shown by using the continued fraction expansion of e which, contrary to that of π, is completely known. Although it is expected that $\mu(\pi) = 2$, it is very hard to get any results on $\mu(\pi)$. It was only in 1953 that K.Mahler was able to show for the first time that $\mu(\pi)$ is finite. Nowadays we know that $\mu(\pi) < 8.02$. The following statement is not hard to show.

Exercise 3.3.1 *Prove that any real algebraic number of degree* 2 *has irrationality measure* 2.

Exercise 3.3.2 *Let α be an algebraic number of degree $n \geq 2$. Prove that $\mu(\alpha) \leq n$.*

Let α be an algebraic number of degree n with $n > 2$. The first non-trivial result on irrationality measures is by A.Thue, who showed in 1909 that $\mu(\alpha) \leq n/2+1$. This was improved by C.L.Siegel who showed in 1929 that $\mu(\alpha) < 2\sqrt{n}$. Finally in 1955 K.F.Roth finished the problem by showing that $\mu(\alpha) = 2$. This result won him the Fields medal in mathematics.
Using Thue's upper bound for $\mu(\alpha)$ we can prove Theorem 3.1.1. Suppose that the equation $F(x, y) = m$ has infinitely many solutions $x, y \in \mathbb{Z}$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the zeros of $F(x, 1)$. Then the inequality

$$\prod_{i=1}^{n} \left| \alpha_i - \frac{x}{y} \right| \leq \frac{|m|}{y^n} \tag{1}$$

has infinitely many solutions in $x, y \in \mathbb{Z}$ and $y > 0$. Let $A = \min_{i \neq j} |\alpha_i - \alpha_j|/2$. Suppose x, y is a solution of the inequality and suppose that $y > |m|^{1/n}/A$. Then there exists an i such that $|\alpha_i - \frac{x}{y}| < A$. By the definition of A this means that $|\alpha_j - \frac{x}{y}| > A$ for all $j \neq i$. Combining this with inequality (1) again, gives us

$$\left| \alpha_i - \frac{x}{y} \right| \leq \frac{|m|}{A^{n-1} y^n}. \tag{2}$$

To every solution x, y with $y > |m|^{1/n}/A$ there corresponds an i such that the latter inequality holds. Since there are infinitely many solutions, there exists an i such that (2) has infinitely many solutions. But this implies that $\mu(\alpha_i) \geq n$. This contradicts Thue's inequality $\mu(\alpha_i) \leq n/2 + 1$ when $n \geq 3$. **qed**

3.4 Siegel's Lemma

This subsection and the next ones will be devoted to a proof of Thue's inequality $\mu(\alpha) \leq n/2 + 1$ for algebraic numbers of degree n. An important ingredient is the so-called Siegel Lemma.

Theorem 3.4.1 (Siegel's Lemma) *Let a_{ij} with $i = 1, \ldots, m$ and $j = 1, \ldots, n$ be integers, not all zero, and suppose that $A = \max_{i,j} |a_{ij}|$. Then the system of linear equations*

$$\sum_{i=1}^{n} a_{ij} x_j, \qquad i = 1, \ldots, m$$

has a non-trivial solution in the integers $x_1, x_2, \ldots, x_n$ with the property that

$$\max_j |x_j| \leq (2nA)^{m/(n-m)}.$$

A remarkable application is for example the following. Take 10 integers $a_1, a_2, \ldots, a_{10}$ of ten digits each. Suppose we want to find integers $x_1, x_2, \ldots, x_{10}$, not all zero, such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

Siegel's Lemma with $A = 10^{10}, m = 1, n = 10$ tells us that we can find such x_i of absolute value at most 18. Surprisingly small given the size of the numbers a_i.
Here is a proof of Siegel's Lemma. Choose an integer Q. Let $B(Q)$ be the box consisting of points $(x_1, \ldots, x_n)$ with $x_1, \ldots, x_n$ integers with $0 \leq x_i \leq Q$. Consider the map $\phi : B(Q) \to \mathbb{Z}^m$ given by $\phi : (x_1, \ldots, x_n) \mapsto (\sum_{j=1}^n a_{1j}x_j, \ldots, \sum_{j=1}^n a_{mj}x_j)$. The image of $B(Q)$ is contained in the box $[-nAQ, nAQ]^m$. The number of points with integral coordinates in this box is at most $(2nAQ + 1)^m$. The number of points in $B(Q)$ is precisely $(Q + 1)^n$. So if $(Q + 1)^n > (2nAQ + 1)^m$, then ϕ is not surjective and we find two integral vectors $\mathbf{x}_1, \mathbf{x}_2$ in $B(Q)$ such that $\phi(\mathbf{x}_1 - \mathbf{x}_2) = 0$. In other words, $\mathbf{x}_1 - \mathbf{x}_2$ is a solution of our system of equations. In addition, the components of this difference are all bounded by Q in absolute value. A straightforward calculation shows that $(Q + 1)^n > (2nAQ + 1)^m$ is satisfied if we choose $Q = [(2nA)^{m/(n-m)}]$. **qed**

3.5 Sketch of the proof of Thue's Theorem

Let α be an algebraic number of degree n and suppose that $M(\alpha) > n/2 + 1$. Hence there exists $\theta > 0$ such that

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{y^{n/2+1+\theta}} \tag{3}$$

has infinitely many solutions $x, y \in \mathbb{Z}$ with $y > 0$. Without loss of generality we can assume that α is an algebraic integer. Let D be a large positive integer and $\epsilon > 0$. We construct polynomials $P(x), Q(x) \in \mathbb{Z}[x]$ of degree $\leq D$ such that $P(x) - \alpha Q(x)$ vanishes of order m, where $m \approx (\frac{2}{n} - \epsilon)D$. This problem can be considered as a system of linear equations in the coefficients of P and Q as follows,

$$\begin{aligned}
0 &= P(\alpha) - \alpha Q(\alpha) \\
0 &= P'(\alpha) - \alpha Q'(\alpha) \\
0 &= \frac{1}{2!}P''(\alpha) - \frac{\alpha}{2!}Q''(\alpha) \\
&\vdots \quad\quad \vdots \\
0 &= \frac{1}{(m-1)!}P^{(m-1)}(\alpha) - \frac{\alpha}{(m-1)!}Q^{(m-1)}(\alpha)
\end{aligned}$$

These are linear equations with coefficients in $\mathbb{Q}(\alpha)$. Each such equation can be rewritten as n equations with coefficients in $\mathbb{Z}$. So we have mn equations with coefficients in $\mathbb{Z}$. The

coefficients are bounded by C^D, where C is some number depending only on α. There are $2D+2$ unknowns is $\mathbb{Z}$. We can apply Siegel's Lemma and find polynomials $P(x), Q(x)$ with coefficients whose absolute value is at most

$$(4(D+1)C^D)^{mn/(2D+2-mn)} < (4(D+1)C^D)^{2D/(2D-(2-\epsilon n)D)} = (4(D+1)C^D)^{2/(n\epsilon)}.$$

Notice that we can find another number C_1, depending only on α such that $(4(D+1)C^D)^{2/n} < C_1^D$. Numbers depending only on α will be denoted by $C_1, C_2, \ldots$ in the sequel. We conclude that we have found non-trivial polynomials $P(x), Q(x)$ with integral coefficients bounded by $C_1^{D/\epsilon}$ such that $P(x) - \alpha Q(x)$ vanishes of order at least m in $x = \alpha$.
Let $x_1/y_1, x_2/y_2$ be two very large solutions of (3) with $y_2 >> y_1 >> 1$. The idea is now to find both upper and lower bounds for

$$\Delta = \left| P(\frac{x_1}{y_1} - \frac{x_2}{y_2} Q(\frac{x_1}{y_1}) \right|.$$

First an upper bound.

$$\begin{aligned} \Delta &\leq \left| P(\frac{x_1}{y_1}) - \alpha Q(\frac{x_1}{y_1}) \right| + \left| (\alpha - \frac{x_2}{y_2}) Q(\frac{x_1}{y_1}) \right| \\ &< C_1^{D/\epsilon} \left| \alpha - \frac{x_1}{y_1} \right|^m + C_1^{D/\epsilon} \frac{1}{y_2^{n/2+1+\theta}} \\ &< C_1^{D/\epsilon} \frac{1}{y_1^{m(n/2+1+\theta)}} + C_1^{D/\epsilon} \frac{1}{y_2^{n/2+1+\theta}} \end{aligned}$$

A lower bound for Δ can be attained if we assume that $\Delta \neq 0$. For then Δ is a non-zero rational number with denominator dividing $y_2 y_1^D$. Combining this upper and lower bound we get

$$\frac{1}{y_2 y_1^D} < C_1^{D/\epsilon} \left(\frac{1}{y_1^{m(n/2+1+\theta)}} + \frac{1}{y_2^{n/2+1+\theta}} \right)$$

where $m \approx (2/n - \epsilon)D$. Now choose ϵ such that $(2/n - \epsilon)(n/2 + \theta) = 1 + \delta$ for some $\delta > 0$. We choose D, and as a consequence m, in such a way that $y_1^m \approx y_2$. Then our inequality simplifies to

$$\frac{1}{y_2 y_1^D} < 2C_1^{D/\epsilon} \frac{1}{y_2 y_1^{(1+\delta)D}}.$$

Hence

$$1 < 2C_1^{D/\epsilon} y_1^{-\delta D}.$$

But this gives a contradiction if y_1 is large enough. Since there are infinitely many choices for y_1 we do arrive at a contradiction.
A problem arises if Δ does vanish. To that end we prove the following Lemma.

Lemma 3.5.1 *Let $P(x), Q(x)$ be two non-trivial polynomials with rational coefficients and of degree $\leq D$. Let α be an algebraic number of degree n such that $P(x) - \alpha Q(x)$ vanishes of order at least m at $x = \alpha$. Suppose $m > D/n$. Then, for any numbers β, γ, with β not a conjugate of α, the polynomial $P(x) - \gamma Q(x)$ has vanishing order at most $2D - (m-1)n$ at $x = \beta$.*

Suppose $P(x) - \gamma Q(x)$ vanishes of order μ at $x = \beta$. Taking the derivative of $P(x) - \gamma Q(x)$ we see that $P'(x) - \gamma Q'(x) = O((x-\beta)^{\mu-1})$. Elimination of γ shows that $P'(x)Q(x) - P(x)Q'(x) = O((x-\beta)^{\mu-1})$. In the same way we can show that $P'(x)Q(x) - P(x)Q'(x) = O((x-\alpha)^{m-1})$. Since P, Q have coefficients in $\mathbb{Q}$, we find that $f(x)^{m-1}$ divides $P'Q - PQ'$, where $f(x)$ is the minimal polynomial of α.
An important remark is that $P'Q - PQ'$ cannot be identically zero. If it were, then $P(x) = \lambda Q(x)$ for some constant λ. But then $P - \alpha Q = O((x-\alpha)^m)$ would imply that $Q(x)$ is divisible by $(x-\alpha)^m$ and hence by $f(x)^m$. But this is impossible by degree considerations, since $mn > D$ by assumption.
So $P'Q - PQ'$ is a non-trivial polynomial divisible by $(x-\beta)^{\mu-1}$ and by $f(x)^{m-1}$. By degree considerations we get $(m-1)n + \mu - 1 \leq 2D - 1$, which proves our Lemma. **qed**

Application of the Lemma to our situation shows that $P(x) - \frac{x_2}{y_2} Q(x)$ has vanishing order at most $2D - n(2/n - \epsilon)D = n\epsilon D$ at $x = \frac{x_1}{y_1}$. So there exists $\mu \leq n\epsilon D$ such that

$$\Delta_\mu = P^{(\mu)}(\frac{x_1}{y_1}) - \frac{x_2}{y_2} Q^{(\mu)}(\frac{x_1}{y_1}) \neq 0$$

We now carry out our argument on Δ_μ instead of Δ.