The generalized Fermat equation

Frits Beukers

January 20, 2006

Abstract

This article will be devoted to generalisations of Fermat's equation $x^n + y^n = z^n$. Very soon after the Wiles and Taylor proof of Fermat's Last Theorem, it was wondered what would happen if the exponents in the three term equation would be chosen differently. Or if coefficients other than 1 would be chosen. We discuss the reduction of the resolution of such equations to the determination of rational points on finite sets of algebraic curves (over $\mathbb{Q}$ if possible) and explain the full resolution of the particular equation with exponents $2, 3, 5$.

1 Introduction

Let $A, B, C \in \mathbb{Z}$ be non-zero and $p, q, r \in \mathbb{Z}_{\geq 2}$. Consider the diophantine equation

$$Ax^p + By^q = Cz^r, \qquad \gcd(x, y, z) = 1$$

in the unknown integers x, y, z. The gcd-condition is really there to avoid trivialities. For example, from $a + b = c$ it would follow, after multiplication by $a^{21}b^{14}c^6$, that

$$(a^{11}b^7c^3)^2 + (a^7b^5c^2)^3 = (a^3b^2c)^7$$

thus providing us with infinitely many trivial solutions of $x^2 + y^3 = z^7$. There are three cases to be distinguished.

1. The hyperbolic case
$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1.$$
In this case the number of solutions is at most finite, as shown in [DG, Theorem 2].

2. The euclidean case
$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1.$$
A simple calculation shows that the set $\{p, q, r\}$ equals one of $\{3, 3, 3\}, \{2, 4, 4\}, \{2, 3, 6\}$. In this case the solution of the equation comes down to the determination of rational points on twists of genus 1 curves over $\mathbb{Q}$ with $j = 0, 1728$.

3. The spherical case
$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1.$$
A simple calculation shown that the set $\{p, q, r\}$ equals one of the following: $\{2, 2, k\}$ with $k \geq 2$ or $\{2, 3, m\}$ with $m = 3, 4, 5$. In this case there are either no solutions or infinitely many. In the latter case the solutions are given by a finite set of polynomial parametrisations of the equation, see [Beu]

A special case of interest is when $A = B = C = 1$. In many such cases the solution set has been found. Below we list the exponent triples (p, q, r) of solved equations together with the non-trivial solutions ($xyz \neq 0$). We exclude the generic solution $1^k + 2^3 = 3^2$ from our listing. If no solutions are mentioned it is proven that no other solutions exist. The notation $\{p, q, r\}$ implies that all permutations of the ordered triple (p, q, r) are taken into account. This is important in the case of two even exponents.
We start with the *hyperbolic cases*. The first case $\{n, n, n\}$ is of course Wiles's proof of Fermat's Last Theorem. As is well-known this proof is based on the proof of the Shimura-Taniyama-Weil conjecture for stable elliptic curves. Later Breuil, Conrad, Diamond and Taylor proved the full conjecture for any elliptic curve in [BCDT]. In the following list the cases with variable n are all solved using Wiles's modular form approach, with possibly a few exceptions which are resolved using Chabauty's method. The isolated cases in this table are all solved using a Chabauty approach.

$\{n, n, n\}$ and $n \geq 4$. Wiles and Taylor [W],[TW] (formerly Fermat's Last Theorem).

$\{n, n, 2\}$ Darmon and Merel [DM] (for n prime ≥ 7), and Poonen for $n = 5, 6, 9$.

$\{n, n, 3\}$ Darmon and Merel [DM] (for n prime ≥ 7), Lucas (19th century) for $n = 4$ and Poonen for $n = 5$.

$\{3, 3, n\}$ Kraus [Kr1] (for $17 \leq n \leq 10000$) and Bruin [Br2,3] for $n = 4, 5$.

$(2, n, 4)$ Application of [BS], includes $(4, n, 4)$ by Darmon [D].

$(2, 4, n)$ Ellenberg [El] (for prime $n \geq 211$) and Ghioca for $n = 7$ (email, see[PSS]).

$\{2n, 2n, 5\}$ Bennett [Ben] (for $n \geq 7$ and $n = 2$) Bruin [Br3] for $n = 3$ and $n = 5$ follows from Fermat's last theorem.

$(2, 2n, 3)$ Chen [Ch] (for $7 < n < 1000$, $n \neq 31$, n prime)

($\{2, 4, 6\}$ Bruin [Br1].

$\{2, 4, 5\}$ $2^5 + 7^2 = 3^4$, $3^5 + 11^4 = 122^2$, Bruin [Br2].

$\{2,3,9\}$ $13^2 + 7^3 = 2^9$, Bruin [Br4]

$\{2,3,8\}$ $1^8 + 2^3 = 3^2$, $43^8 + 96222^3 = 30042907^2$, $33^8 + 1549034^2 = 15613^3$, Bruin [Br1, Br2].

$\{2,3,7\}$ $1^7 + 2^3 = 3^2$, $2^7 + 17^3 = 71^2$, $17^7 + 76271^3 = 21063928^2$, $9262^3 + 15312283^2 = 113^7$, Poonen, Schaefer, Stoll [PSS].

Presumably the solutions listed above are the only solutions in the hyperbolic case. Note that in all cases one of the exponents equals 2. This led Tijdeman and Zagier (in 1994) to the following conjecture.

Conjecture 1.1 *The diophantine equation*

$$x^p + y^q = z^r$$

in $x, y, z \in \mathbb{Z}$ *with* $\gcd(x,y,z) = 1$, $xyz \neq 0$ *and* $p, q, r \in \mathbb{Z}_{\geq 3}$ *has no solutions.*

Nowadays this conjecture is also known as Beal's conjecture or the Fermat-Catalan conjecture.
In the *euclidean case* it is well-known that the only non-trivial solutions arise from the equality $1^6 + 2^3 = 2^2$, as the elliptic curves $x^3 + y^3 = 1, y^2 = x^4 + 1, y^2 = x^3 \pm 1$ contain only finitely many obvious rational points.
In the *spherical cases* the solution set is infinite. In the case $\{2,2,k\}$ this is an exercise in number theory. The case $\{2,3,3\}$ was solved by Mordell, $\{2,3,4\}$ by Zagier and $\{2,3,5\}$ by J.Edwards [Ed] in 2004. The families of solutions are listed in Appendix A (please read the explanation in the beginning of Appendix A).

2 A sample solution

To illustrate the phenomena we encounter when solving the generalized Fermat equation, we give a partial solution of $x^2 + y^8 = z^3$. This equation lends itself very well to a stepwise descent method.
First we solve $x^2 + u^2 = z^3$. By factorisation on both sides over $\mathbb{Z}[i]$ we quickly see that $x + iu$ should be the cube of a gaussian integer, $(a+bi)^3$. By comparison of real and imaginary parts we get $x = a^3 - 3ab^2$, $u = b(3a^2 - b^2)$. Note that a, b should be relatively prime in order to ensure $\gcd(x,u,z) = 1$.
Next we partly solve $x^2 + v^4 = z^3$. This can be done by requiring that u, as found in the previous equation should be a square, e.g. $v^2 = b(3a^2 - b^2)$. The two factors on the right should be squares up to some factors $\pm 1, \pm 3$, since their product is a square and a, b are relatively prime. We should explore all possibilities, but in this partial solution we only continue with the possibility $b = -v_1^2$, $3a^2 - b^2 = -v_2^2$. The latter equation can be rewritten as $3a^2 = b^2 - v_2^2$. The right hand side factors as $(b - v_2)(b + v_2)$ and hence each factor is a square up to a finite number of factors. Here several possibilities present themselves again and we choose one, namely $b - v_2 = -6a_1^2$, $b + v_2 = -2a_2^2$ (and of

course $a = 2a_1a_2$). Summation of the two equalities and use of $b = -v_1^2$ gives us $v_1^2 - a_2^2 = 3a_1^2$. Now the left hand side factors and we choose the possibility $v_1 - a_2 = 6t^2,\ v_1 + a_2 = 2s^2$ (and of course $a_1 = 2st$). Solving for v_1 and a_2 gives $v_1 = s^2 + 3t^2$ and $a_2 = s^2 - 3t^2$. Hence $a = 4st(s^2 - 3t^2)$ and $b = -(s^2 + 3t^2)^2$. Further straightforward computation gives us

$$\begin{aligned} v &= (s^2 + 3t^2)(s^4 - 18s^2t^2 + 9t^4) \\ x &= 4st(s^2 - 3t^2)(3s^4 + 2s^2t^2 + 3t^4)(s^4 + 6s^2t^2 + 81t^4) \\ z &= (s^4 - 2s^2t^2 + 9t^4)(s^4 + 30s^2t^2 + 9t^4) \end{aligned}$$

A might be clear now, this gives us an infinite set of integer solutions to the equation $x^2 + v^4 = z^3$. Had we followed all possibilities we would have found more parametrised solutions to recover the full solution set in integers. For a full list see Appendix A, or Henri Cohen's recent book [Co], where one finds a complete derivation of the above type.
Finally we consider $x^2 + y^8 = z^3$. Continuing with our choices we must solve

$$y^2 = (s^2 + 3t^2)(s^4 - 18s^2t^2 + 9t^4).$$

After division by t^6 and putting $\xi = s/t,\ \eta = y/t^3$ we get

$$\eta^2 = (\xi^2 + 3)(\xi^4 - 18\xi^2 + 9),$$

i.e. we must determine the rational points on a genus two curve. To solve the equation completely we must determine the rational points on several genus two curves, namely those arising from the different parametrising solutions above. To cut things short now, we can easily calculate that

$$\frac{z^3}{y^8} = \frac{(\xi^4 - 2\xi^2 + 9)^3(\xi^4 + 30\xi^2 + 9)^3}{\eta^8}.$$

Thus, any point z^3/y^8 coming from a solution of $x^2 + y^8 = z^3$ is the image of a rational point (ξ, η) on our genus two curve under the map just given. This map is an example of a Galois cover map.
Had we followed all possibilities of the above argument, we would have obtained a number of covering maps from a genus 2 curve to $\mathbb{P}^1$ which would have covered the full set of values z^3/y^8 corresponding to all solutions of $x^2 + y^8 = z^3$ in co-prime integers x, y, z.
In this example the curves arose naturally as a result of a descent procedure. In many cases, like $x^3 + y^5 = z^7$, this descent is not so obvious any more and we have to start by constructing covers of $\mathbb{P}^1$ by curves which have a suitable ramification behaviour.

3 Galois covers of $\mathbb{P}^1$

In all approaches to the solution of the (generalised) Fermat equations one uses Galois covers in one form or another.

First we recall a few facts from the theory of algebraic curves and their function fields. For a more complete introduction we recommend Chapter II of Silverman's book [Si]. Let K be a field of characteristic zero and X a complete, smooth and geometrically irreducible curve X defined over K. In the function field $K(X)$ we consider a non-constant element which we denote by ϕ. Note that $K(X)$ is now a finite extension of the field $K(\phi)$. The degree of this extension is also called the degree of the map ϕ. Let $P \in X(\overline{K})$ (by $X(L)$ we denote the L-rational points of X, where L is a field extension of K). Assuming for the moment $\phi(P) \neq \infty$ we call the vanishing order of $\phi - \phi(P)$ at P the *ramification index* of ϕ at P. Notation: e_P. In case $\phi(P) = \infty$ we take for e_P the vanishing order of $1/\phi$ at P. If $e_P > 1$ we call P a *ramification point* of ϕ. The image $\phi(P)$ under ϕ of a ramification point P is called *branch point*. The set of branch points is called the *branch set* or *branch locus*. We now recall the Riemann-Hurwitz formula

Theorem 3.1 *With the notation above let N be the degree of the map ϕ and $g(X)$ the geometric genus of X. Then,*

$$2g(X) - 2 = -2N + \sum_{P \in X(\overline{K})} (e_P - 1).$$

As we have $e_P = 1$ for all points of X except finitely many, the sum on the right is in fact a finite sum.
We call the map given by ϕ a *geometric Galois cover* if the extension $\overline{K}(X)/\overline{K}(\phi)$ is a Galois extension of fields. The Galois group G is a subgroup of the automorphism group (over $\overline{K}$) of X and is called the *covering group*. Note that the extension $K(X)/K(\phi)$ need not be Galois. If it is we call the cover simply a *Galois cover*. For a geometric Galois cover the ramification indices of all points above a given branch point are the same. In particular we shall be interested in geometric Galois covers whose branch locus is $0, 1, \infty$. These are examples of so-called Belyi maps. An immediate consequence of the Riemann-Hurwitz theorem is the following.

Corollary 3.2 . *Let $X \to \mathbb{P}^1$ be a geometric Galois cover whose branch locus is contained in the set $0, 1, \infty$. Suppose that above these points the ramification indices are p, q, r. Suppose the degree of the cover is N. Then*

$$2g(X) - 2 = N\left(1 - \frac{1}{p} - \frac{1}{q} - \frac{1}{r}\right).$$

In particular we see that if $1/p + 1/q + 1/r > 1$, then $g(X) = 0$ and when $1/p + 1/q + 1/r < 1$ we have $g(X) \geq 2$.
Here we list a series of geometric Galois covers that will occur in the sequel. We start with $X = \mathbb{P}^1$. The finite subgroups of $\mathrm{Aut}_{\overline{\mathbb{Q}}}(\mathbb{P}^1)$ have been classified by Felix Klein. Up to conjugation they are given by

1. The cyclic group of order N

2. The dihedral group of order $2N$
3. The tetrahedral group of order 12
4. The octahedral group of order 24
5. The icosahedral group of order 60

When we consider $\mathbb{P}^1$ as a sphere, each of these examples corresponds to a finite rotation group of the sphere. Here we describe them in some more detail, where z denotes a standard coordinate on $\mathbb{P}^1$. We cannot go into all the fascinating details of the Klein groups. For an extensive discussion we recommend Chapter I of Klein's original book [Kl].

Cyclic group. This group is generated by $z \mapsto \zeta_N z$ where ζ_N is a primitive N-th root of unity. The corresponding cover is given by $z \mapsto z^N$.

Dihedral group. This is generated by the cyclic group given above and $z \mapsto 1/z$. The cover is given by

$$z \mapsto \frac{1}{2}\left(z^N + \frac{1}{z^N}\right).$$

Tetrahedral group. Let ω be a primitive cube root of unity. Consider the subgroup Γ_3 of $SL(2,\mathbb{C})$ generated by

$$\frac{1}{\sqrt{-3}}\begin{pmatrix} 1 & 2\omega^j \\ \omega^{-j} & -1 \end{pmatrix} \quad (j = 0,1,2) \quad \text{and} \quad \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix}.$$

Then the tetrahedral group is the subgroup of $PSL(2,\mathbb{C})$ given by $\Gamma_3/\pm 1$. The covering map is given by

$$z \mapsto \left(\frac{4(z^3-1)}{z^4+8z}\right)^3.$$

F.Klein's (semi)-invariants of Γ_3 are

$$f = -4y(x^3 - y^3)$$

$$H = -x^4 - 8xy^3$$

$$t = -x^6 + 20x^3y^3 + 8y^6$$

with fundamental relation $t^2 + H^3 = f^3$.

Octahedral group. Consider the group Γ_4 generated by

$$\begin{pmatrix} \zeta_8 & 0 \\ 0 & \zeta_8^{-1} \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix} \zeta_8 & -\zeta_8^{-1} \\ \zeta_8 & \zeta_8^{-1} \end{pmatrix}.$$

Then the octahedral group is the subgroup of $PSL(2,\mathbb{C})$ given by $\Gamma_4/\pm 1$. The cover is given by

$$z \mapsto \frac{(z^8+14z^4+1)^3}{108(z(z^4-1))^4}.$$

F.Klein's (semi)-invariants are

$$f = 36xy(x^4 - y^4)$$

$$H = -36(x^8 + y^8 + 14x^4y^4)$$

$$t = 216(x^{12} + y^{12} - 33(x^4y^8 + x^8y^4)$$

with fundamental relation $t^2 + H^3 = -3f^4$.
Icosahedral group. Consider the group Γ_5 generated by

$$-\mathrm{Id}, \quad \begin{pmatrix} \zeta_5 & 0 \\ 0 & \zeta_5^{-1} \end{pmatrix}, \quad \frac{1}{\sqrt{5}} \begin{pmatrix} \zeta_5 - \zeta_5^4 & -\zeta_5^2 + \zeta_5^3 \\ -\zeta_5^2 + \zeta_5^3 & -\zeta_5 + \zeta_5^4 \end{pmatrix}.$$

Then the icosahedral group is the subgroup of $PSL(2, \mathbb{C})$ given by $\Gamma_5/\pm 1$. The cover is given by

$$z \mapsto \frac{(-z^{20} + 228z^{15} - 494z^{10} - 228z^5 - 1)^3}{1728z^5(z^{10} + 11z^5 - 1)^5}.$$

F.Klein's (semi)-invariants are

$$f = 12^3xy(x^{10} + 11x^5y^5 - y^{10})$$

$$H = 12^4(-x^{20} - y^{20} + 228(x^{15}y^5 - x^5y^{10}) - 494x^{10}y^{10})$$

$$t = 12^6(x^{30} + y^{30} + 522(x^{25}y^5 - x^5y^{25}) - 10005x^{20}y^{10} - x^{10}y^{20}))$$

with the fundamental relation $t^2 + H^3 = f^5$.
In the last three examples the forms f, H, t have the additional property that

$$H = \frac{1}{k^2(k-1)^2} \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix}, \quad t = \frac{1}{2k(k-2)} \begin{vmatrix} f_x & f_y \\ H_x & H_y \end{vmatrix},$$

where k is the degree of f. These relations will become important later on. Furthermore in all three examples the branch locus is given by the points $0, 1, \infty \in \mathbb{P}^1$. The ramification indices above these points are $3, r, 2$ where $r = 3, 4$ or 5 depending on the group Γ_r
Now we turn to the case when the genus of X is at least 2 and list a number of examples.

1. $X: \ x^n + y^n = z^n$ and covering map $(x : y : z) \mapsto (x/z)^n$. This map has degree n^2 and the group is given by al elements $(x : y : z) \mapsto (\zeta x : \zeta' y : z)$ where ζ, ζ' are n-th roots of unity. The branch locus is given by $0, 1, \infty$ with ramification indices n, n, n.

2. Let p and q be integers ≥ 3 and let X be given by the projective equations

$$\sum_{i=0}^{p-1} \zeta_p^{ik} x_i^q = 0 \quad (k = 1, 2, \ldots, p-2).$$

Consider the covering map

$$(x_0 : x_1 : \ldots : x_{p-1}) \mapsto \frac{(\sum_{i=0}^{p-1} x_i^q)^p}{\prod_{i=0}^{p-1} x_i^q}.$$

This has Galois group of order pq^{p-1} generated by multiplication of the coordinates x_i by a q-th root of unity and the cyclic permutation of the coordinates $(x_0 : x_1 : \ldots : x_{p-1}) \mapsto (x_1 : x_2 : \ldots : x_{p-1} : x_0)$. Notice also that for points on X we have the relation

$$(\sum_{i=0}^{p-1} x_i^q)^p + (\sum_{i=0}^{p-1} \zeta_p^{-i} x_i^q)^p = (\prod_{i=0}^{p-1} x_i)^q.$$

The map has branch locus $0, 1, \infty$ and ramification indices p, p, q.

3. Let $n \geq 2$ and let X be the complete modular curve $X(n)$. We consider the natural map $X(n) \to X(1) = \mathbb{P}^1$ using the J-function on $X(n)$. More explicitly, consider the modular J-function on the complex upper half plane $\mathcal{H}$. This map gives us the quotient map $J : \mathcal{H} \to \mathbb{C}$ with respect to the group $PSL(2, \mathbb{Z})$. It ramifies above the points $J = 0, 1$ with ramification indices 3 and 2 respectively. Let

$$\Gamma(n) = \{M \in SL(2, \mathbb{Z}) \mid M \equiv \mathrm{Id} \pmod{n}\}.$$

Then $\Gamma(n)$ is a normal subgroup of $SL(2, \mathbb{Z})$ and the quotient of $\mathcal{H}$ by $\Gamma(n)$ is denoted by $Y(n)$. Since $\Gamma(n)$ contains no elliptic elements, the cover $\mathcal{H} \to Y(n)$ is unramified. Furthermore J factors over $Y(n)$ to a finite map $J : Y(n) \to \mathbb{C}$. If we now complete the curves by adding the cusps to $Y(n)$ and ∞ to $\mathbb{C}$, we get $J : X(n) \to \mathbb{P}^1$ where $X(n)$ is the completion of $Y(n)$. This map ramifies of order n above ∞. So the ramification indices above $0, 1, \infty$ are $3, 2, n$. The covering group is $PSL(2, \mathbb{Z}/n\mathbb{Z})$. When $n = 3, 4, 5$ we recover the tetrahedral, octahedral and icosahedral covering again.

4. Let n be odd, $X = X(2n)$ and consider the natural map to $X(2) = \mathbb{P}^1$. This has ramification indices n, n, n above $0, 1, \infty$ and no others. The covering group is $PSL(2, \mathbb{Z}/n\mathbb{Z})$.

5. Let n be odd and let X be the completed modular curve corresponding to the modular group $\Gamma(n) \cap \Gamma_0(2)$. Then the natural map $X \to X_0(2) = \mathbb{P}^1$ is a geometric Galois cover ramified above $0, 1, \infty$ with ramification indices $n, n, 2$. The covering group is again $PSL(2, \mathbb{Z}/n\mathbb{Z})$

6. Similarly, when n is not divisible by 3 we consider the modular group $\Gamma(n) \cap \Gamma_0(3)$ and take for X the associated complete modular curve. Then $X \to X_0(3)$ gives us a geometric Galois cover ramified above $0, 1, \infty$ with ramification indices $n, n, 3$. The covering group is $PSL(2, \mathbb{Z}/n\mathbb{Z})$.

4 Lifting points

Let $\phi : X \to \mathbb{P}^1$ be a geometric Galois cover defined over a number field K and whose degree is N. For any point $a \in \mathbb{P}^1(K)$ the points in the inverse image $\phi^{-1}(a)$ generate a finite Galois extension L of K of degree at most N. In the following we explicitly determine the set of primes of K that ramify in L.
Let π be any finite prime of K. We extend it to a valuation of $\overline{K}$. We represent points of $\mathbb{P}^1(\overline{K})$ as points in $\overline{K} \cup \infty$. We define the π-adic intersection number on $\mathbb{P}^1$ by

$$I_\pi(a,b) = \begin{cases} \mathrm{ord}_\pi(a-b) & \text{if } \mathrm{ord}_\pi(a), \mathrm{ord}_\pi(b) \geq 0 \\ \mathrm{ord}_\pi(1/a - 1/b) & \text{if } \mathrm{ord}_\pi(1/a), \mathrm{ord}_\pi(1/b) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We say that a and b meet π-adically if $I_\pi(a,b) > 0$. The following theorem is a weakened version of a theorem proved in [Bec].

Theorem 4.1 (S.Beckmann) *Let $\phi : X \to \mathbb{P}^1$ be a Galois cover defined over a number field K and with covering group G. Let $a_1, \ldots, a_r \in \overline{K} \cup \infty$ be the set of branch points. There is finite set of primes, which we denote by S_{bad}, with the following properties. For any point $q \in K$ not equal to any a_i we have*

1. *the finite primes of K that ramify in $K(\phi^{-1}(q))$ are contained in the set $S = S_{\mathrm{bad}} \cup S_q$, where S_q is the set of primes π at which q meets a branch point a_i π-adically.*
2. *if $\pi \notin S_{\mathrm{bad}}$ and q meets the branchpoint a_i π-adically, then π ramifies of order e where e is the denominator of $I_\pi(q,a_i)/e_i$ and where e_i is the ramification index above a_i.*

In [Bec] we find a stronger statement which explicitly gives us S_{bad}. If the group G is simple or if the covering is given by a good model, then S_{bad} is the union of the primes dividing the order of G and the primes π for which at least two distinct branch points meet π-adically.
We are now able to give a proof of the following result.

Theorem 4.2 *Let $\phi : X \to \mathbb{P}^1$ be a geometric Galois cover which ramifies of order p, q, r above the points $0, 1, \infty$ respectively, and which has no further ramification. Suppose that the cover is defined over the number field K. Then there exists a finite extension L of K such that $\phi^{-1}(Aa^p/Cc^r) \subset X(L)$ for every triple (a,b,c) that satisfies*

$$Aa^p + Bb^q = Cc^r, \quad \gcd(a,b,c) = 1.$$

Here $X(L)$ denotes the set of L-rational points on X.

Proof. If necessary we replace K by a finite extension so that ϕ becomes a Galois cover. Consider the field M generated over K by the coordinates of the

points in $\phi^{-1}(Aa^p/Cc^r)$. We now apply Beckmann's Theorem. We let S_{ABC} be the set of primes dividing ABC. Let π be a prime of K not dividing abc and not in $S_{\rm bad} \cup S_{ABC}$. Then the point Aa^p/Cc^r doesn't reduce to $0, 1$ or ∞ modulo π. To see that it does not reduce to 1 notice that $\frac{Aa^p}{Cc^r} - 1 = -\frac{Bb^q}{Cc^r}$ where b, c, B, C are π-adic units. . Hence π is unramified in M/K. Suppose now that $\pi \notin S_{\rm bad} \cup S_{ABC}$ and π divides a. Then the intersection number $I_\pi(Aa^p/Cc^r, 0)$ is a positive multiple of p. This is a consequence of the fact that $\gcd(a, b, c) = 1$. Since the cover ramifies of order p above zero, part 2 of Beckmann's theorem implies that π has ramification order 1, i.e. no ramification. Similarly, if π divides b or c and is not in $S_{\rm bad} \cup S_{ABC}$, then π is unramified in M/K. So we find that the coordinates of a point in $\phi^{-1}(Aa^p/Cc^r)$ are in a number field of degree at most N, the degree of the cover, and a fixed set of ramified primes. There are only finitely many such fields and for L we can take their compositum.

qed

We can now prove Theorem 2 in [DG]

Theorem 4.3 (Darmon-Granville) *Suppose* $1/p+1/q+1/r < 1$ *and* $A, B, C \in \mathbb{Z}$ *with* $ABC \neq 0$. *Then the number of solutions to*

$$Ax^p + By^q = Cz^r, \quad \gcd(x, y, z) = 1$$

is finite.

Proof. We begin by the construction of a curve X and a Galois cover $X \to \mathbb{P}^1$ of Belyi-type, i.e it ramifies only above the points $0, 1, \infty$ with ramification orders p, q, r respectively. This can be done for example by the construction in Proposition 4.4. It is well-known that Belyi-maps can be defined over $\overline{\mathbb{Q}}$.
By the Riemann-Hurwitz theorem we know that $1/p + 1/q + 1/r < 1$ implies $g(X) \geq 2$. Beckmann's theorem implies that there is a number field L such that for any solution (a, b, c) we have $\phi^{-1}(Aa^p/Cc^r) \subset X(L)$. By Faltings' theorem (formerly Mordell's conjecture) we know that $X(L)$ is finite. Hence our equation has finitely many solutions.

qed

In the proof of the Darmon-Granville theorem the existence of a suitable cover is usually accounted for by application of the Riemann existence theorem. However, the Riemann covering data to apply the existence theorem are usually not provided. With the following proposition we remedy this small gap.

Proposition 4.4 *Let* p, q, r *be three integers* ≥ 2 *and such that* $1/p+1/q+1/r < 1$. *Then there exists an algebraic curve* X *and a Galois cover* $X \to \mathbb{P}^1$ *which ramifies of order* p, q, r *above the points* $0, 1, \infty$ *respectively.*

Proof We first construct a so-called triangle group in the Poincaré disc D given by $\{z \in \mathbb{C} | \ |z| < 1\}$. We start with a hyperbolic triangle with angles $\pi/p, \pi/q, \pi/r$. Denote the hyperbolic reflection in the side opposite to the angle π/p by s_p. Similarly we define s_q, s_r. Notice that $(s_p s_q)^r = (s_p s_r)^q = (s_q s_r)^p = \mathrm{id}$. Let Δ be the group of isometries of D consisting of even-length words in

s_p, s_q, s_r. Then Δ is a group of fractional linear transformations of D which we call the (p,q,r)-triangle group. This triangle group acts discretely on D, the quotient D/Δ is $\mathbb{P}^1$ and the quotient map $D \to \mathbb{P}^1$ ramifies of order p, q, r above three points which we can choose to be $0, 1, \infty$.
To prove our Proposition it suffices to construct a normal subgroup H of Δ, of finite index, whose non-trivial elements act fixpoint-free on D. In that case the quotient map $D \to D/\Delta$ factors as $D \to D/H \to D/\Delta$, where $D \to D/H$ is unramified. Moreover, $D/H \to D/\Delta$ is a finite map with the required ramification properties. Hence $X = D/H$.
Up to conjugation the traingle group is uniquely determined. Consider now the matrices

$$A = \begin{pmatrix} 0 & \zeta_{2p}^{-1}\zeta_{2q}^{-1} \\ -\zeta_{2p}\zeta_{2q} & \zeta_{2r} + \zeta_{2r}^{-1} \end{pmatrix} \qquad B = \begin{pmatrix} 0 & -\zeta_{2p}^{-1} \\ \zeta_{2p} & \zeta_{2p} + \zeta_{2p}^{-1} \end{pmatrix}$$

in $SL(2,\mathbb{C})$. Here $\zeta_n = \exp(2\pi i/n)$. Notice that A, considered as element in $PSL(2,\mathbb{C})$ has precise order r, B has order p and AB^{-1} has order q. The entries of the elements of the group generated by A, B are all contained in the ring of integers $R = \mathbb{Z}[\zeta_{2p}, \zeta_{2q}, \zeta_{2r}]$. Furthermore, the elliptic elements in Δ are all conjugate to one of A, B, AB^{-1} or one of their powers. Choose a prime ideal π in R which does not contain any of the numbers $\zeta_{2n}^k - 1, (k = 1, \ldots, n-1)$ for $n = p, q, r$. Then the subgroup H defined by

$$H = \{g \in \Delta \mid g \equiv \mathrm{Id} \bmod \pi\}$$

is a normal subgroup of finite index in Δ without elliptic elements. This is an example of the group we were looking for.

qed

5 Galois cocycles

Let K be a number field and L a finite Galois extension. Let G be a finite group with a $\mathrm{Gal}(L/K)$ Galois action $\mathrm{Gal}(L/K) \to \mathrm{Aut}(G)$. A 1-*cocycle* is a map $\xi : \mathrm{Gal}(L/K) \to G$, mapping $\sigma \mapsto \xi_\sigma$, such that

$$\xi_{\sigma\tau} = \xi_\sigma \sigma(\xi_\tau)$$

for all $\sigma, \tau \in \mathrm{Gal}(L/K)$. Two cocycles ξ, ζ are called *cohomologous* if there exists $h \in G$ such that

$$\zeta_\sigma = h^{-1}\xi_\sigma \sigma(h).$$

The set of cocycles modulo this equivalence relation is called the first Galois cohomology set of $\mathrm{Gal}(L/K)$ in G. Notation $H^1(\mathrm{Gal}(L/K), G)$.
An important use of the first cohomology is the description of twists of algebraic varieties V, when $G = \mathrm{Aut}(V)$. To fix ideas, let X be a smooth connected algebraic curve defined over K. Any curve X' defined over K together with an isomorphism $\psi : X \to X'$, which is defined over $\overline{K}$, is called a *twist* of X. In

particular, when the twist map ψ is defined over a finite galois extension L of K, we call our twist an L-twist. Let $\psi : X \to X'$ be such an L-twist. Then, for any $\sigma \in \mathrm{Gal}(L/K)$ the composite map $\psi^{-1}\sigma(\psi)$ is an automorphism of X defined over L. One easily checks that

$$\sigma \mapsto \psi^{-1}\sigma(\psi)$$

is a Galois cocycle in $H^1(\mathrm{Gal}(L/K), \mathrm{Aut}_L(X))$. Namely,

$$\psi^{-1}\sigma\tau(\psi) = \psi^{-1}\sigma(\psi)\sigma(\psi^{-1}\tau(\psi)).$$

Two L-twists $\psi_1 : X \to X'$ and $\psi_2 : X \to X''$ are called equivalent is there exist $h \in \mathrm{Aut}_L(X)$ and an isomorphism $g : X' \to X''$ defined over K such that $\psi_2 = g \circ \psi_1 \circ h$. Denote the set of classes of L-twists by $\mathrm{Twist}(X, L/K)$. Then we have

Theorem 5.1 *The map $\psi \mapsto (\sigma \mapsto \psi^{-1}\sigma(\psi))$ gives a well-defined map from $\mathrm{Twist}(X, L/K)$ to $H^1(\mathrm{Gal}(L/K), \mathrm{Aut}_L(X))$. Moreover, this map is a bijection.*

More explicitly, if we have a 1-cocycle $\xi : \mathrm{Gal}(L/K) \to \mathrm{Aut}_L(X)$, then it is possible to find an L-twist $\psi : X \to X'$ such that $\xi_\sigma = \psi^{-1}\sigma(\psi)$ for all $\sigma \in \mathrm{Gal}(L/K)$. We now apply this to our diophantine equation.

Theorem 5.2 *Let A, B, C, p, q, r be as in the introduction. By* Sol *we denote the set of numbers Aa^p/Cc^r for all a, c belonging to triples of integers (a, b, c) that satisfy*

$$Aa^p + Bb^q = Cc^r, \quad \gcd(a, b, c) = 1, \quad abc \neq 0$$

Let $\phi : X \to \mathbb{P}^1$ be a geometric Galois cover of Belyi-type which ramifies above $0, 1, \infty$ of order p, q, r respectively. Suppose it is defined over a number field K. Then there exist finitely many twists $\psi_i : X \to X_i$, $i = 1, 2, \ldots, r$, defined over K, such that

1. *each map $\phi \circ \psi^{-1} : X_i \to \mathbb{P}^1$ is defined over K.*
2. *$\mathrm{Sol} \subset \cup_{i=1}^r \phi \circ \psi_i^{-1})(X_i(K))$.*
3. *The sets $\phi \circ \psi_i^{-1}(X_i(K))$ intersect in a subset of $0, 1, \infty$.*

Proof. According to Theorem 4.2 there is a finite Galois extension L such that $\phi^{-1}(\mathrm{Sol}) \subset X(L)$. We assume that the covering group G is also defined over L. Take any point $Q \in \mathrm{Sol}$ and let $P \in X(L)$ be such that $\phi(P) = Q$. Since ϕ is a geometric Galois cover, for any $\sigma \in \mathrm{Gal}(L/K)$ there exists a unique $g_\sigma \in G$ such that $\sigma(P) = g_\sigma(P)$. Notice that

$$g_{\sigma\tau}(P) = \sigma(\tau(P)) = \sigma(g_\tau(P)) = \sigma(g_\tau)(\sigma(P)) = \sigma(g_\tau)g_\sigma(P).$$

Hence $g_{\sigma\tau} = \sigma(g_\tau)g_\sigma$ and so we see that

$$\sigma \mapsto g_\sigma^{-1}$$

is a $\mathrm{Gal}(L/K)$ cocycle in $H^1(\mathrm{Gal}(L/K), G)$. Consider the twist $\psi : X \to X'$ that corresponds to this cocycle. This means that $g_\sigma^{-1} = \psi^{-1}\sigma(\psi)$ for all $\sigma \in \mathrm{Gal}(L/K)$. Hence

$$\sigma(\psi(P)) = \sigma(\psi)(\sigma(P)) = \psi g_\sigma^{-1} g_\sigma(P) = \psi(P).$$

In other words $\psi(P)$ is fixed under $\mathrm{Gal}(L/K)$ and hence $\psi(P) \in X'(K)$. Furthermore, for any $\sigma \in \mathrm{Gal}(L/K)$ we have

$$\sigma(\phi \circ \psi^{-1}) = \phi \circ \sigma(\psi)^{-1} = \phi \circ g_\sigma \circ \psi^{-1} = \phi \circ \psi^{-1}.$$

Hence $\phi \circ \psi^{-1}$ is defined over K. Since $Q = \phi(P)$, we see that Q is contained in $\phi \circ \psi^{-1}(X'(K))$.
To every class in $H^1(\mathrm{Gal}(L/K), G)$ we choose a twist and since $H^1(\mathrm{Gal}(L/K), G)$ is finite, we get a finite number of twists $\psi_i : X \to X_i$ with $i = 1, 2, \ldots, r$. Part one of our Theorem follows.
To see the disjointness, suppose $\phi \circ \psi_1^{-1}(X_1(K))$ and $\phi \circ \psi_2^{-1}(X_2(K))$ have a point $Q \in \mathbb{P}^1(K)$, $Q \neq 0, 1, \infty$ in common. For $i = 1, 2$ choose a point $P_i \in X_i(K)$ such that $Q = \phi \circ \psi_i^{-1}(P_i)$. Then there exists $k \in G$ such that $\psi_1^{-1}(P_1) = k \circ \psi_2^{-1}(P_2)$. Let ξ_i be the cocycle to which we associated ψ_i. Then application of any $\sigma \in \mathrm{Gal}(L/K)$ yields

$$\xi_{1,\sigma}^{-1} \circ \psi_1^{-1}(P_1) = \sigma(k) \circ \xi_{2,\sigma}^{-1} \psi_2^{-1}(P_2).$$

Replacing the right hand side,

$$\xi_{1,\sigma}^{-1} \circ \psi_1^{-1}(P_1) = \sigma(k) \circ \xi_{2,\sigma}^{-1} \circ k^{-1} \psi_1^{-1}(P_1).$$

Since $\psi^{-1}(P_1)$ has trivial stabilizer in G we conclude that

$$\xi_{1,\sigma} = k^{-1} \circ \xi_{2,\sigma} \circ \sigma(k)$$

for all $\sigma \in \mathrm{Gal}(L/K)$. Hence ξ_1, ξ_2 are cohomologous and the twists X_1, X_2 are equivalent.

qed

So to solve a generalised Fermat equation in the hyperbolic case it suffices to determine the K-rational points on a finite set of curves of genus ≥ 2. It would be nice if one could have $K = \mathbb{Q}$. In fact this is how the equations with exponent triples $\{2, 3, 7\}, \{2, 3, 8\}$ and $A = B = C = 1$ were solved in [PSS] and [B1],[B2]. In the spherical cases $\{p, q, r\} = \{2, 3, 3\}, \{2, 3, 4\}, \{2, 3, 5\}$ we have the Klein covers of degree $12, 24, 60$ respectively and $X = \mathbb{P}^1$. Hence the above theorem implies that the solution set of a generalised Fermat equation in the spherical case is given by a finite (possibly empty) set of rational functions $\mathbb{P}^1 \to \mathbb{P}^1$ defined over $\mathbb{Q}$.
In the following we shall determine these rational functions in detail for the spherical case $(2, 3, 5)$. We use the approach of Johnny Edwards, who found that classical invariant theory provides a convenient language to carry out the computations.

6 Invariant theory of binary forms

Here we give a very quick introduction following Hilbert's lectures from 1897. See [H]. In particular our approach will be very classical. The only difference between Hilbert's and our representation is that we use k instead of n for the degree of the base form.

6.1 Definition and first examples

Let K be an algebraically closed field of characteristic zero. Consider a form $f \in K[\mathbf{a}, \mathbf{x}]$ of the shape

$$f(\mathbf{a}, \mathbf{x}) = \sum_{i=0}^{k} \binom{k}{i} a_i x_1^{k-i} x_2^i,$$

which we call the *base form*. We have two sets of polynomial variables, $\mathbf{x} = (x_1, x_2)$ and $\mathbf{a} = (a_0, \ldots, a_k)$. For historical reasons the number k is called the order of f. The group $GL(2, K)$ acts on polynomials in x_1, x_2 as follows. For any $g \in GL(2, K)$ we replace the column vector $\mathbf{x} = (x_1, x_2)^t$ by the components of the column vector $g \cdot \mathbf{x}$. When we replace the variables x_1, x_2 in a polynomial h in this way, we denote the new polynomial by $h \circ g$.
Let $C \in K[\mathbf{a}, \mathbf{x}]$. We denote its dependence on $\mathbf{a}, \mathbf{x}$ by writing it as $C(f)$. The polynomial $C(f)$ is called a *covariant* of f if there exists an integer $p \geq 0$ such that

$$C(f \circ g) = \det(g)^p C(f) \circ g$$

for all $g \in GL(2, K)$. We call p the *weight* of the covariant. A covariant which depends only on the a_j is called an *invariant*. I.e $I(\mathbf{a}) \in K[\mathbf{a}]$ is called an invariant of weight p if

$$I(f \circ g) = \det(g)^p I(f)$$

for all $g \in GL(2, \mathbb{C})$.
Since the action of g does not change degrees in the a_i and x_j we can restrict our attention to covariants which are homogeneous in the a_j and homogeneous in the x_i. When $C(f)$ is such a bihomogeneous covariant, we call $\deg_{\mathbf{a}}(C)$ the *degree* of C and $\deg_{\mathbf{x}}(C)$ the *order* of C. Notice that f itself is a covariant of weight 0, order k and degree 1.
Here are two of our most important examples of covariants. First there is the Hessian covariant $H(f)$ defined by

$$H(f) = \frac{1}{k^2(k-1)^2} \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix}$$

where f_{ij} stands for partial differentiation with respect to x_i and x_j. It is a matter of straightforward calculus to see that this is a covariant. Its weight is 2, the order is $2k - 4$ and the degree is 2.

The other important covariant is the Jacobian determinant $t(f)$ defined by

$$t(f) = \frac{1}{k-2}\begin{vmatrix} f_1 & f_2 \\ H_1 & H_2 \end{vmatrix}.$$

Again it is straightforward to check that this is a covariant. Its weight is 3, its order $3k-6$ and degree 3.

Remark 6.2 *Let C be a covariant of f. When we specialise the variables $a_0, \ldots, a_k$ to values in some ring R and we do this both in f and $C(f)$ we will still call the specialisation of $C(f)$ a covariant of the specialised f.*

6.3 Structure of covariants

Suppose we are given a bihomogeneous polynomial

$$C(f) = \sum_{j=0}^{m} \binom{m}{i} C_j(\mathbf{a}) x_1^{m-j} x_2^j$$

We give necessary and sufficient condition for a form to be a covariant. Suppose it is a covariant. Since $GL(2,K)$ is generated by the matrices

$$\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \nu \\ 0 & 1 \end{pmatrix}$$

where $\lambda \in K^*, \nu \in K$, it suffices to verify the covariant property of C only for these matrices. First we take g to be the diagonal matrix with entries $\lambda, 1$. Then

$$g(x_1) = \lambda x_1,\ g(x_2) = x_2.$$

Let $Aa_0^{r_0} \cdots a_k^{r_k} x_1^{m-j} x_2^j$ be a non-trivial term in C. In shorthand notation: $A\mathbf{a}^{\mathbf{r}} x_1^{m-j} x_2^j$. The covariant property now implies that

$$\lambda^{kr_0+(k-1)r_1+\cdots+r_{k-1}} \mathbf{a}^{\mathbf{r}} x_1^{m-j} x_2^j = \lambda^{p+m-j} \mathbf{a}^{\mathbf{r}} x_1^{m-j} x_2^j.$$

Hence

$$kr_0 + (k-1)r_1 + \cdots + r_{k-1} = p + m - j.$$

The covariant property with respect to $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ implies if $Aa_0^{r_0} \cdots a_k^{r_k} x_1^{m-j} x_2^j$ occurs as a non-trivial term, then so does $(-1)^p A a_k^{r_0} \cdots a_0^{r_k} x_2^{m-j} x_1^j$. In particular, this observations together with previous one, leads to

$$r_1 + 2r_2 + \cdots k r_k = p + j$$

for any monomial. Addition of the two equalities gives us

$$k(r_0 + r_1 + \cdots + r_k) = 2p + m$$

Letting g be the degree (in $\mathbf{a}$) of C we get

$$kg = 2p + m.$$

Finally we need to implement the covariant property with respect to $\begin{pmatrix} 1 & \nu \\ 0 & 1 \end{pmatrix}$. It is a straightforward but slightly tedious job to show that we get

$$DC = x_2 \frac{\partial C}{\partial x_1}$$

where D is the differential operator

$$D = a_0 \frac{\partial}{\partial a_1} + 2a_1 \frac{\partial}{\partial a_2} + 3a_2 \frac{\partial}{\partial a_3} + \cdots + na_{n-1} \frac{\partial}{\partial a_n}.$$

By the symmetry $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ we also get

$$\Delta C = x_1 \frac{\partial C}{\partial x_2}$$

where Δ is the differential operator

$$\Delta = a_n \frac{\partial}{\partial a_{n-1}} + 2a_{n-1} \frac{\partial}{\partial a_{n-2}} + \cdots + na_1 \frac{\partial}{\partial a_0}.$$

A particular consequence of the first equation is that

$$D(C_0) = 0. \tag{1}$$

The second equation implies that

$$C_1 = \frac{1}{m}\Delta C_0,\ C_2 = \frac{1}{m(m-1)}\Delta^2 C_0,\ \ldots\ C_m = \frac{1}{m!}\Delta^m C_0. \tag{2}$$

In fact, these conditions turn out to be both necessary and sufficient. In the following statement an *isobaric polynomial* in the a_j is a polynomial such that for all terms $Aa_0^{r_0} \cdots a_k^{r_k}$ the sum $r_1 + 2r_2 + 3r_3 + \cdots + kr_k$ has the same value.

Theorem 6.4 *The bihomogeneous polynomial*

$$C(f) = \sum_{j=0}^{m} \binom{m}{i} C_j(\mathbf{a}) x_1^{m-j} x_2^j$$

is a covariant of weight p if and only if C_0 is homogeneous of degree g, isobaric of weight p, such that $m = kg - 2p$, and such that equations (1) and (2) are satisfied.

In particular we have a very nice corollary characterising invariants.

Corollary 6.5 *A homogeneous polynomial $C(\mathbf{a})$ is an invariant of weight p if and only if it has degree g and is isobaric of weight p such that $kg = 2p$ and such that the equation $DC(\mathbf{a}) = 0$ is satisfied.*

6.6 Further examples

First we give some examples of invariants and covariants for small k.

The case $k = 2$, $f = a_0x_1^2 + 2a_1x_1x_2 + a_2x_2^2$.

The Hessian of f equals $a_0a_2 - a_1^2$, the discriminant of f. It turns out that all invariants are powers of the discriminant.

The case $k = 3$, $f = a_0x_1^3 + 3a_1x_1^2x_2 + 3a_2x_1x_2^2 + a_3x_2^2$.

The Hessian now reads

$$H(f) = (a_0a_2 - a_1^2)x_1^2 + (a_0a_3 - a_1a_2)x_1x_2 + (a_1a_3 - a_2^2)x_2^2.$$

There is also the Jacobian covariant

$$t(f) = (a_0^2a_3 - 3a_0a_1a_2 + 2a_1^3)x_1^3 + \cdots$$

The discriminant of f is an invariant,

$$D(f) = a_0^2a_3^2 - 3a_1^2a_2^2 + 4a_1^3a_3 + 4a_0a_2^3 - 6a_0a_1a_2a_3.$$

The powers of D form a full system of invariants. We have the classical relation

$$4H^3 + t^2 = Df^2.$$

The case $k = 4$, $f = a_0x_1^4 + 4a_1x_1^3x_2 + 6a_2x_1^2x_2^2 + 4a_3x_1x_2^3 + a_4x_2^4$.

We have the Hessian and Jacobian covariants $H(f), t(f)$ as before. The ring of invariants is generated by

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

We have the classical relation

$$t(f)^2 = -4H(f)^3 + I_2H(f)f^2 - I_3f^3.$$

A general way to produce new covariants from old ones is the *transvectant* construction. Letting C_1, C_2 be two covariants and $r \in \mathbb{Z}_{\geq 1}$ we define

$$(C_1, C_2)_r = \left(\frac{(k-r)!}{k!}\right)^2 \Omega^r(C_1(x_1, x_2)C_2(x_1', x_2'))|_{x_1'=x_1, x_2'=x_2}$$

where

$$\Omega = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2'} - \frac{\partial}{\partial x_1'}\frac{\partial}{\partial x_2}.$$

The transvectants of f are defined by

$$\tau_{2m} = \frac{1}{2}(f, f)_{2m}, \quad \tau_{2m+1}(f) = (f, \tau_{2m}(f))_1.$$

This is the sequence of transvectants we find in [H, Ch I.8]. They are covariants of degrees 2 and 3 respectively with weights equal to the index n in τ_n. One notes that $H(f) = \tau_2(f)$, $t(f) = \tau_3(f)$ and

$$\begin{aligned} H(f) &= (a_0a_2 - a_1^2)x_1^{2k-4} + \cdots \\ t(f) &= (a_0^2a_3 - 3a_0a_1a_2 + 2a_1^3)x_1^{3k-6} + \cdots \\ \tau_4(f) &= (a_0a_4 - 4a_1a_3 + 3a_2^2)x_1^{2k-8} + \cdots \\ \tau_6(f) &= (a_0a_6 - 6a_1a_5 + 15a_2a_4 - 10a_3^2)x_1^{2k-12} + \cdots \end{aligned}$$

The following theorem will be crucial to us.

Theorem 6.7 (Gordan, 1887) *The fourth transvectant $\tau_4(f)$ of a non-trivial form f with $k \geq 4$ is identically zero if and only if f is $GL(2,K)$-equivalent to one of the following forms*

1. x_1^k *or* $x_1^{k-1}x_2$ *(degenerate case)*
2. $x_2(x_1^3 + x_2^3)$ *(tetrahedral case)*
3. $x_1x_2(x_1^4 + x_2^4)$ *(octahedral case)*
4. $x_1x_2(x_1^{10} - 11x_1^5x_2^5 - x_2^{10})$ *(icosahedral case)*

So the vanishing of $\tau_4(f)$ forces f to be one of the Klein forms if f is not degenerate.
Because of its importance we give a proof of this theorem. First of all a straightforward computation shows that $\tau_4(f)$ vanishes for all forms in the list. We now show the converse statement. Let f be a form with $\tau_4(f) = 0$. Very explicitly we have

$$\tau_4(f) = \sum_{r=0}^{2n-8} D_r x_1^{2n-r}x_2^r$$

where

$$D_r = \sum_{i+j=r} \binom{n-4}{i}\binom{n-4}{j}(a_ia_{j+4} - 4a_{i+1}a_{j+3} + 3a_{i+2}a_{j+2}).$$

We use the equations $D_0 = 0, D_1 = 0, D_2 = 0, \ldots$ to recursively determine the coefficients a_j. Suppose our f is not equivalent to x_1^k. Then f should have a zero of order $\leq k/2$. By application of a $GL(2,K)$ substitution, we can see to it that this zero becomes $x_2 = 0$. In particular, $a_0 = 0$.
First suppose that $a_1 = 0$. Choose $t > 1$ minimal so that $a_t \neq 0$. We have that $t \leq k/2$ because $x_2 = 0$ is a zero of order $\leq k/2$. Now note that for all $t \leq k-2$,

$$D_{2t-4} = 3\binom{k-4}{t-2}^2 a_t^2 + \cdots$$

where the omitted terms all contain a factor a_i with $i < t$. Since $a_i = 0$ for all $i < t$ it follows from $D_{2t-4} = 0$ that $a_t = 0$, a contradiction. So a_1 cannot be zero.

Now suppose, after normalisation if necessary, that $a_1 = 1$. By application of a shift $x_1 \to x_1 + \nu x_2, x_2 \to x_2$ we can see to it that $a_2 = 0$. We now determine the remaning a_i recursively using the equations

$$\begin{aligned} D_0 &= a_0a_4 - 4a_1a_3 + a_2^2 = 0 \\ D_r &= \cdots + \frac{k}{r}\frac{k-4}{r-1}\left(r - 4 + \frac{12}{k}\right)a_1a_{r+3} + \cdots = 0 \qquad (r \geq 1) \end{aligned}$$

where the omitted terms all contain a_0 or an a_i with $2 \leq i \leq r+2$. If the factor $r - 4 + 12/k$ does not vanish for any r we get that $a_3 = a_4 = \ldots = a_k = 0$ and we are in the case $x_1^{k-1}x_2$. So we need that k divides 12 and $4 > 12/k$. Hence $k = 4, 6$ or 12. Take $k = 12$, the other cases being similar. We get that $a_2 = a_3 = \ldots = a_5 = 0$ and choose $a_6 \neq 0$. By scaling we can see to it that $a_6 = -11$. Recursive solution of $D_4 = D_5 = \ldots = D_9 = 0$ shows that $a_7 = \ldots = a_{10} = a_{12} = 0$ and $a_{11} = -1$. Hence $f = x_1x_2(x_1^{10} - 11x_1^5x_2^5 - x_2^{10})$.

7 Mordell's approach

As an example of the use of invariant theory in solving diophantine equations we present Mordell's method to solve the equation

$$x^2 = -y^3 + A_2yz^2 + A_3z^3 \tag{3}$$

in integers x, y, z with $\gcd(x, y, z) = 1$. Mordell's idea is to exploit the relation

$$t(f)^2 = -4H(f)^3 + I_2H(f)f^2 - I_3f^3$$

for quartic forms f. Given a solution x, y, z with $z \neq 0$ he constructs a quartic form f with invariants $I_2(f) = 4A_2, I_3(f) = 4A_3$ and such that $f(1, 0) = z, H(1, 0) = y, t(1, 0) = 2x$. When we write f in our standard form, this amounts to solving

$$\begin{aligned} \text{(i) } z &= a_0 \\ \text{(ii) } y &= a_0a_2 - a_1^2 \\ \text{(iii) } 2x &= a_0^2a_3 - 3a_0a_1a_2 + 2a_1^3 \\ \text{(iv) } 4A_2 &= a_0a_4 - 4a_1a_3 + 3a_2^2 \\ \text{(v) } 4A_3 &= a_0a_2a_4 - a_0a_3^2 - a_1^2a_4 + 2a_1a_2a_3 - a_2^3 \end{aligned}$$

We start by setting $a_0 = z$. From (3) it follows that $x^2 \equiv -y^3 (\mathrm{mod} z^2)$. Hence $-(xy^{-1})^2 \equiv y (\mathrm{mod} z^2)$. Now choose a_1 integral so that $a_1 \equiv -xy^{-1} (\mathrm{mod} z^2)$. Then $y + a_1^2$ is divisible by $z = a_0$ and we can determine a_2 from equation (ii). Rewrite the equation (iii) as

$$z^2a_3 = 2x + 3a_1y + a_1^3.$$

To solve this, the right hand side should be divisible by z^2. This is indeed the case as follows from

$$2x + 3a_1y + a_1^3 \equiv 2x + 3(-xy^{-1})y + (-xy^{-1})^3 \equiv -x(y^3 + x^2)y^{-3} (\mathrm{mod} z^2)$$

and from equation (3).
We now determine a_4 from equation (iv). With this value of a_4, equation (v) is automatically satisfied because of (3). We now see from equations (iv) and (v) that both $a_0a_4 = za_4$ and $(a_0a_2 - a_1^2)a_4 = ya_4$ are integer. Since z and y are relatively prime this implies that a_4 is an integer.
Thus we know that to any solution of (3) we have a quartic form f with prescribed invariants $4A_1, 4A_2$ such that $f(1,0) = z, H(1,0) = y, t(1,0) = 2x$. Of course other specialisations of f, H, t will provide us with an infinity of solutions to 3. Since the number of $SL(2,\mathbb{Z})$-classes of such forms is finite, we get a finite number of parametrising solutions of (3) that give the complete solution set.
Notice that I_2 is the fouth transvectant of f. If this vanishes and if $I_3 = 4$ we get the identity $t^2 = -4H^3 - 4f^3$. This is exactly the case for which Mordell provides a full solution set in [Mo, Chapter 25].

8 Edwards's approach

The main idea in Edwards's paper [Ed] is to mimick Mordell's technique to solve the diophantine equation

$$x^2 + y^3 = dz^5 \tag{4}$$

in coprime integers x, y, z. Here d is a given non-zero integer. Let

$$\tilde{f}(x_1, x_2) = 12^3 x_1 x_2 (x_1^{10} - 11x_1^5 x_2^5 - x_2^{10})$$

be the icosahedral form of F.Klein. Letting $\tilde{H}$ and $\tilde{t}$ be its Hessian and Jacobian covariants, we get

$$(\tilde{t}/2)^2 + \tilde{H}^3 = \tilde{f}^5. \tag{5}$$

Definition 8.1 *Let d be a non-zero integer. By $C_5(d)$ we denote the set of $GL(2,\overline{\mathbb{Q}})$-transforms of $\tilde{f}$ which are of the form*

$$f(x_1, x_2) = \sum_{i=0}^{12} \binom{12}{i} a_i x_1^{12-i} x_2^i,$$

such that

1. *$a_0, \ldots, a_5, 7a_6, a_7, \ldots, a_{12} \in \mathbb{Z}$ for all i.*
2. $$(t(f)/2)^2 + H(f)^3 = df^5. \tag{6}$$

where $H(f)$ and $t(f)$ are the Hessian and Jacobian covariants of f.

Notice that a_6 is preceded by a 7 in this definition (and in all formulas to come). It turns out that the space of dodecahedral forms with $a_0, \ldots, a_5, 7a_6, a_7, \ldots, a_{12} \in \mathbb{Z}$ is stable under $SL(2,\mathbb{Z})$. From now on, when we speak of integer solutions, we will mean these variables to be integral.

Because of the covariant property it follows from (5) that for any $g \in GL(2,\mathbb{Q})$ we have for $f := \tilde{f} \circ g$ the identity

$$(t(f)/2)^2 + H(f)^3 = \det(g)^6 f^5.$$

So by taking $\det(g)^6 = d$ we can see to it that we get parametrisations of $x^2 + y^3 = dz^5$.
Our first goal is to prove the following theorem.

Theorem 8.2 *Let d be a non-zero integer. Let $x, y, z \in \mathbb{Z}$ be a coprime solution of $x^2 + y^3 = dz^5$. Then there exists a form $f \in C_5(d)$ such that*

$$f(1,0) = z, \quad H(f)(1,0) = y, \quad t(f)(1,0) = 2x. \tag{7}$$

Proof . In what follows we shall write a form

$$\sum_{i=0}^{12} \binom{12}{i} a_i x_1^{12-i} x_2^i$$

in the shape

$$[a_0, a_1, \ldots, a_{12}].$$

When $z = 0$, we have $x = \pm 1$ and $y = -1$. We can immediately write down the corresponding forms f. They read

$$[0, \pm 1, 0, 0, 0, 0, -144d/7, 0, 0, 0, 0, \mp(144d)^2, 0].$$

So from now on we can assume $z \neq 0$. We first prove our theorem without the rationality properties of the a_i. Determine $\alpha, \beta \in \overline{\mathbb{Q}}$ such that $\tilde{f}(\alpha, \beta) = z/d$ and $\tilde{H}(\alpha, \beta) = y/d^2$. Determine $\gamma, \delta \in \overline{\mathbb{Q}}$ such that $\alpha\delta - \beta\gamma = 1$. Define the dodecahedral form f by $f = d\tilde{f} \circ g$, where $g = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$. Then, because $H(f) = H(d\tilde{f} \circ g) = d^2 H(\tilde{f}) \circ g$ and $t(f) = t(d\tilde{f} \circ g) = d^3 t(\tilde{f}) \circ g$ we find that (6) is satisfied for our choice of f. Moreover, $f(1,0) = d\tilde{f}(\alpha, \beta) = z$ and similarly $H(f)(1,0) = y$. From $x^2 + y^3 = dz^5$ and (6) it follows that $t(f)(1,0) = \pm 2x$. In case $t(f)(1,0) = -2x$ we take a new f equal to the old $f(ix_1, ix_2)$. This does not change f, H but it does change t by a minus sign. We have found a solution f for the equations (6) and (7). Notice that if $f(x_1, x_2)$ is a solution, then so is $f(x_1 + \lambda x_2, x_2)$ for any $\lambda \in \overline{\mathbb{Q}}$. So we still have some freedom in the choice of f. Thus far everything has been done over $\overline{\mathbb{Q}}$. Our claim is that we can choose λ in such a way that the coefficients a_i satisfy the rationality and integrality properties of the a_i required by f being in $C_5(d)$.
Equations (7) gives us the following equations in a_i

$$\begin{aligned} z &= a_0 \\ y &= a_0 a_2 - a_1^2 \\ 2x &= a_0^2 a_3 - 3a_0 a_1 a_2 + 2a_1^3 \end{aligned}$$

precisely the same as in Mordell. We also need explicitly given necessary conditions on the a_i for f to be equivalent to $\tilde{f}$. These are given by the vanishing of the fourth transvectant according to Gordan's theorem. So we get the $D_i = 0$ where the D_i, $i = 0, 1, \ldots, 12$ are the coefficients of $\tau_4(f)$. In Appendix B, at the very end we have reproduced the explicit equations $D_i = 0$ for $i = 0, \ldots, 9$. Then we must take $a_0 = z$. For a_1 we have complete freedom because of the freedom in λ above. We set a_1 equal to a number in the residue class $-xy^{-1} (\mathrm{mod} z^5)$. From $H(1,0) = y$ and $t(1,0) = 2x$ it follows that

$$a_0 a_2 \equiv y + (xy^{-1})^2 \equiv -dz^5 y^{-2} \equiv 0 (\mathrm{mod} z^5)$$
$$a_0^2 a_3 \equiv -x(x^2 + y^3) y^{-3} \equiv -dx z^5 y^{-3} \equiv 0 (\mathrm{mod} z^5)$$

From this we observe that a_2 and a_3 are integers divisible by z^4 and z^3 respectively. We can now determine $a_4, a_5, \ldots$ recursively. Start with

$$0 = D_0/1 = a_0 a_4 - 4a_1 a_3 + 3a_2^2.$$

Hence $a_0 a_4$ is an integer divisible by z^3. Hence a_4 is an integer divisble by z^2. Similarly it follows from $D_1 = 0$ that a_5 is an integer divisible by z and from $D_2 = 0$ it follows that $7a_6 \in \mathbb{Z}$. In

$$D_3/56 = 0 = a_0 a_7 - 6a_2 a_5 + 5a_3 a_4$$

a small miracle happens. There is no term $a_1 a_6$ and we can now see that $a_0 a_7$ is an integer divisble by z^5. Hence a_7 is divisible by z^4. The equation

$$D_4/14 = 0 = 5a_0 a_8 + 12a_1 a_7 - 6a_2(7a_6) - 20a_3 a_5 + 45a_4^2$$

poses a small problem because of the coefficient 5 in front of $a_0 a_8$. However, by elimination of a_6, a_7 from $D_4 = D_3 = D_2 = 0$ we obtain

$$a_0^2 a_8 = 12a_4 a_3 a_1 + 18a_4 a_2^2 - 24a_3^2 a_2 + 4a_5 a_3 a_0 - 9a_4^2 a_0.$$

Now it follows that a_8 is an integer divisible by z^3. Continuing with $D_5 = D_6 = D_7 = 0$ we find that a_9, a_{10}, a_{11} are integers as well. From $D_8 = D_9 = 0$ we see that $a_0 a_{12}$ and $a_1 a_{12}$ are integers. Because a_0, a_1 are coprime, we conclude that a_{12} is integral.

qed

Up to a shift $x_1 \to x_1 + ax_2$, $x_2 \to x_2$ the form f found in Theorem 8.2 is unique.

Theorem 8.3 *Let d, x, y, z be as in Theorem 8.2. Let $f_1, f_2 \in C_5(d)$ be such that*

$$f_1(1,0) = f_2(1,0) = z, \; H_1(1,0) = H_2(1,0) = y, \; t_1(1,0) = t_2(1,0) = 2x.$$

Then there exists an integer q such that $f_1(x_1, x_2) = f_2(x_1 + qx_2, x_2)$.

Proof. Notice that if $f(x_1, x_2)$ has coefficients $a_0, a_1, a_2, \ldots$, then for any number q the form $f(x_1 + qx_2, x_2)$ has coefficients $a_0, a_1 + qa_0, a_2 + 2qa_1 + q^2a_0, \ldots$. We distinguish two cases. First of all suppose that $z = 0$. Then, automatically, $y = -1, x = \pm 1$. From the proof of Theorem 8.2 it follows that $a_0 = 0, a_1 = \mp 1$. From $D_4 = a_0a_4 - 4a_1a_3 + 3a_2^2 = 0$ we see that a_2 is even. Hence by a substition of the form $(x_1, x_2) \to (x_1 + qx_2, x_2)$ we can see to it that $a_2 = 0$. The remaining a_i are now uniquely determined from the equations $D_i = 0$ and the extra equation $R_1 = 0$ (see Appendix B). This latter equation arises from the identity $\tau_6(f) = 360df$ and it fixes the proper normalisation of a_6.
Now suppose that $z \neq 0$. We should have $a_0 = z$. From $D_4 = a_0a_4 - 4a_1a_3 + 3a_2^2 = 0$ it follows that a_2 is even if a_0 is even. We can now dedude from the equations $H(1,0) = y, t(1,0) = 2x$ that $a_1 \equiv -xy^{-1} (\mathrm{mod} z)$. So by a substitution $(x_1, x_2) \to (x_1 + qx_2, x_2)$ we can see to it that $0 \leq a_1 < |z|$. This determines a_1 uniquely. The remaining a_i are now determined uniquely as well by the equations $H(1,0) = y, t(1,0) = 2x$ and $D_i = 0$.

qed

Corollary 8.4 *Let d, x, y, z be as in Theorem 8.2. Suppose we have $f_1, f_2 \in C_5(d)$ and integers a_1, b_1, a_2, b_2 such that*

$$\begin{aligned} z &= f_1(a_1, b_1) = f_2(a_2, b_2) \\ y &= H_1(a_1, b_1) = H_2(a_2, b_2) \\ 2x &= t_1(a_1, b_1) = t_2(a_2, b_2) \end{aligned}$$

Then f_1 and f_2 are $SL(2,\mathbb{Z})$-equivalent. Moreover, if the last equation reads

$$t_1(a_1, b_1) = 2x \quad t_2(a_2, b_2) = -2x$$

then f_1 and f_2 are $GL(2,\mathbb{Z})$-equivalent.

Proof. Choose $c_1, d_1 \in \mathbb{Z}$ such that $a_1d_1 - b_1c_1 = 1$ and put $g_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$. Then $f_1 \circ g_1$ is a form in $C_5(d)$ which specialises together with its covariants at the point $(1,0)$ to the solution x, y, z. We can choose $g_2 \in SL(2,\mathbb{Z})$ similarly. According to Theorem 8.2 the forms $f_1 \circ g_1$ and $f_2 \circ g_2$ are $SL(2,\mathbb{Z})$ equivalent. This shows the first part of our Corollary.
To show the second part, choose a $g \in GL(2,\mathbb{Z})$ with determinant -1. Let $f' = f \circ g$. Then $H(f') = H(f) \circ g$ and $t(f') = -t(f) \circ g$ because H has even weight and t has odd weight. According to the first part of our Corollary, f_2' and f_1 are $SL(2,\mathbb{Z})$-equivalent.

qed

We have now seen that all coprime solutions to $x^2 + y^3 = dz^5$ arise from parametrisations using forms from $C_5(d)$ and their covariants. It remains to show that $C_5(d)$ consists of a finite number of $SL(2,\mathbb{Z})$-orbits and, if possible, compute these orbits.

9 Reduction of binary forms

Also in this section we follow the approach in [Ed], but with a few simplications. Consider a form $f \in \mathbb{R}[x_1, x_2]$ of degree $k \geq 3$ in x_1, x_2. We assume once and for all that it has distinct zeros. Choose a factorisation over $\mathbb{C}$,

$$f = \prod_{i=1}^{k} (\nu_i x_1 - \mu_i x_2).$$

There is some ambiguity in the normalisation of the linear factors for the moment, but this will be cleared. For any $t_1, \ldots, t_k \in \mathbb{R}_{>0}$ define $\phi = \phi(f, \mathbf{t})$ by

$$\phi(f, \mathbf{t}) = \sum_{i=1}^{k} t_i^2 (\nu_i x_1 - \mu_i x_2)(\overline{\nu}_i x_1 - \overline{\mu}_i x_2).$$

This is a real quadratic form which is positive definite since its values for real $(x_1, x_2) \neq (0, 0)$ are strictly positive. Strictly speaking ϕ also depends on the particular factorisation of f we have chosen. Let us write $\phi(f, \mathbf{t}) = Px_1^2 - 2Qx_1x_2 + Rx_2^2$ and let $\delta(f, \mathbf{t}) = PR - Q^2$ be its determinant.

Lemma 9.1 *For any $g \in GL(2, \mathbb{R})$ we have*

$$\phi(f \circ g, \mathbf{t}) = \phi(f, \mathbf{t}) \circ g \quad \text{and} \quad \delta(f \circ g, \mathbf{t}) = \det(g)^2 \delta(f, \mathbf{t}).$$

Proof. Note that the second is a consequence of the first, while the first is immediate from the definitions.

qed

We define the *Hermite determinant* of f as

$$\Theta(f) := \min_{\mathbf{t}:\ \prod_i t_i = 1} \delta(f, \mathbf{t})^{k/2}.$$

Note that this minimum does not depend on the particular normalisation in the factorisation in f. In [CS, Lemma 4.2] it is shown that the minimum is assumed at a uniquely determined point, which we denote by $\mathbf{t}_0$. The *representative point* of f is the point $z_0 \in \mathcal{H}$ such that $\phi(f, \mathbf{t}_0)(z_0, 1) = 0$. Note also that this representative point is independent of the normalisation of the μ_i, ν_i. If the representative point of f is in the standard fundamental domain $|z| \geq 1, -1/2 \leq \Re(z) \leq 1/2$ we call f *Hermite reduced*.

Theorem 9.2 *Let f be a real form of degree $k \geq 3$ and distinct roots. Then, for any $g \in GL(2, \mathbb{R})$ we have*

1. $\Theta(f \circ g) = \det(g)^k \Theta(f)$.
2. *If z_0 is the representative point of f and $z_1 = g^{-1}(z_0)$ (fractional linear transform) then the representative point of $f \circ g$ is given by z_1 if $\det(g) > 0$ and $\overline{z}_1$ if $\det(g) < 0$.*

Proof. From $\delta(f \circ g, \mathbf{t}) = \det(g)^2 \delta(f, \mathbf{t})$ it follows that

$$\begin{aligned} \Theta(f \circ g) &= \min_{\prod t_i = 1} \delta(f \circ g, \mathbf{t})^{k/2} \\ &= |\det(g)|^k \min_{\prod t_i = 1} \delta(f, \mathbf{t})^{k/2} \\ &= |\det(g)|^k \Theta(f) \end{aligned}$$

Let $\mathbf{t}_0$ be the point $\mathbf{t}$ where the minimum is attained. Then from $\phi(f \circ g, \mathbf{t}_0) = \phi(f, \mathbf{t}_0) \circ g$ it follows that

$$\begin{aligned} \phi(f \circ g, \mathbf{t}_0)(z_1, 1) &= (\phi(f, \mathbf{t}_0) \circ g)(z_1, 1) \\ &= |\gamma z_1 + \delta|^2 \phi(f, \mathbf{t}_0)(z_0, 1) \end{aligned}$$

where $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. Hence z_1 is a zero of the quadratic form $\phi(f \circ g, \mathbf{t}_0)$. When $\det(g) > 0$ this lies in the upper half plane, so it is the representing point of $f \circ g$. When $\det(g) < 0$ however, the conjugate zero $\overline{z}_1$ lies in $\mathcal{H}$.

qed

Theorem 9.3 *Let f be a real from of degree $k \geq 3$ and distinct roots with factorisation $f = \prod_i (\nu_i x - \mu_i y)$. Let $z_0 = x + iy$ its representative point. Then,*

$$\Theta(f) = \left(\frac{k}{2y}\right)^k \prod_{i=1}^{k} (|\nu_i x - \mu_i|^2 + |\nu_i y|^2).$$

This Theorem allows us to compute the Hermite determinant of the form $\tilde{f}(x_1, x_2) = 12^3 x_1 x_2 (x_1^{10} - x_1^5 x_2^5 - x_2^{10})$. Notice that $\tilde{f}(x_1, x_2) = \tilde{f}(x_2, -x_1)$. Let z_0 be the representing point of $\tilde{f}(x_1, x_2)$. Then, by covariance, the representing point of $\tilde{f}(x_2, -x_1)$ is $-1/z_0$. But by the invariance of the form $\tilde{f}$ we should have $z_0 = -1/z_0$. Thus we conclude that $z_0 = i$. Using our Theorem it is straightforward to verify that $\Theta(\tilde{f}) = 2^{24} 3^{18} 5^5$.

Theorem 9.4 *Let $f \in C_5(d)$. Then*

$$\Theta(f) = 2^{24} 3^{18} 5^5 |d|^2.$$

Proof. There exists an element $g \in GL(2, \mathbb{C})$ such that $f = \tilde{f} \circ g$. In [Ed] it is shown that we can assume $g \in GL(2, \mathbb{R})$. From the covariance of the representing point we have

$$\Theta(f) = |\det(g)|^{12} \Theta(\tilde{f}).$$

(Using the $SL(2, \mathbb{C})$-reduction theory developed in [CS] one deduces that this follows also without the assumption $g \in GL(2, \mathbb{R})$). We also have that $|\det(g)|^6 = d(f)/d(\tilde{f})$ and we know that $d(\tilde{f}) = 1$. Hence we conclude

$$\Theta(f) = |d|^2 \Theta(\tilde{f})$$

and our Theorem follows.

qed

The next theorem gives us upper bounds for the coeffcients of Hermite reduced forms.

Theorem 9.5 *Let*

$$f = \sum_{i=1}^{k} \binom{k}{i} a_i x_1^{k-i} x_2^i$$

be a real, Hermite reduced form of degree k. Then for all $i + j \leq k$ we have

$$|a_i a_j| \leq \left(\frac{4}{3k^2}\right)^{k/2} \Theta(f).$$

Proof. Let $z_0 \in \mathcal{H}$ be the representing point of f and write $z_0 = x + iy$. Let $t_1, \ldots, t_k$ be the components of the vector $\mathbf{t}$ that minimizes $\delta(f, \mathbf{t})$. We shall show that for all r,

$$|a_r|^2 \leq \frac{|z_0|^{2r}}{(ky)^k} \Theta(f).$$

Recalling that $y \geq \frac{\sqrt{3}}{2} \max(|z_0|, 1)$ when z_0 is in the standard fundamental domain of $SL(2, \mathbb{Z})$, the proof of our Theorem then follows from this inequality. We abbreviate $\Theta(f)$ by Θ. Let, as before, $f = \prod_{i=1}^{k} (\nu_i x_1 - \mu_i x_2)$. We know that there exist $\delta > 0$ and $t_i > 0$ such that

$$f = \frac{\sqrt{\Theta}}{\delta^{k/4}} \prod (t_i \nu_i x_1 - t_i \mu_i x_2)$$

and δ is the determinant $PR - Q^2$ of the quadratic form

$$Px_1^2 - 2Qx_1x_2 + Rx_2^2 = \sum_{i=1}^{k} t_i^2 (\nu_i x_1 - \mu_i x_2)(\overline{\nu}_i x_1 - \overline{\mu}_i x_2).$$

Note that

$$P = \sum t_i^2 |\nu_i|^2, \quad R = \sum t_i^2 |\mu_i|^2.$$

This form also equals $P(x_1 - zx_2)(x_1 - \overline{z}x_2)$. Hence, when we write $z = x + iy$,

$$Q = xP, \qquad R = P|z|^2, \qquad \delta = P^2 y^2.$$

Choose $b_i, c_i \in \mathbb{C}$ such that $\sqrt{P} b_i = \nu_i t_i$ and $\sqrt{R} c_i = -\mu_i t_i$. Then $\sum |b_i|^2 = \sum |c_i|^2 = 1$ and also

$$f = \sqrt{\Theta} \frac{1}{y^{k/2}} \prod (b_i x_1 + c_i |z_0| x_2)(\overline{b}_i x_1 + \overline{c}_i |z_0| x_2).$$

Comparison of the r-th coefficients yields

$$\binom{k}{r} a_r = \left(\frac{|z_0|^r}{y^{k/2}}\right) \Big(\sum_{\#S = k-r} b_S c_{S'} \Big) \sqrt{\Theta}.$$

Here the summation is over all subsets S of $1, \ldots, k$ of cardinality $k-r$, and S' is the complement of S. Furthermore b_S denotes the product of all $b_i,\ i \in S$. We first use Schwarz's inequality

$$\left(\sum_{\#S=k-r} b_S c_{S'} \right)^2 \leq \left(\sum_{\#S=k-r} |b_S|^2 \right) \left(\sum_{\#S=k-r} |c_{S'}|^2 \right).$$

Finally use the generalised AM/GM inequality to obtain

$$\sum_{\#S=k-r} |b_S|^2 \leq \binom{k}{k-r} \left(\frac{1}{k} \sum_i |b_i|^2 \right)^{k-r} = \binom{k}{r} \frac{1}{k^{k-r}}$$

and similarly

$$\sum_{\#S=k-r} |c_{S'}|^2 \leq \binom{k}{r} \frac{1}{k^r}.$$

Combining all inequalities yields the desired estimate for $|a_r|$.

qed

Using the estimate of $\Theta(f)$ for any Hermite reduced $f \in C_5(d)$ we obtain the following consequence.

Corollary 9.6 *Let $f \in C_5(d)$ and suppose f is Hermite reduced. Let $a_0, \ldots, a_{12}$ be its coefficients. Then, for every i, j with $i + j \leq 12$ we have*

$$|a_i a_j| \leq 2^{12} 5^5 |d|^2.$$

In particular, $|a_i| \leq 1600\sqrt{5}|d|$ for every $i \leq 6$.

10 An algorithm to solve $x^2 + y^3 = dz^5$

Let d be any non-zero integer. We have seen in the previous two sections that all coprime solutions x, y, z to $x^2 + y^3 = dz^5$ arise as specialisation to integers of a form $f \in C_5(d)$ and its Hessian and Jacobian covariant. To determine the set $C_5(d)$ it suffices to determine the $SL(2,\mathbb{Z})$-orbits within $C_5(d)$. More particularly, it suffices to determine the Hermite reduced forms in $C_5(d)$.
Here is an algorithm to find the Hermite reduced forms with $a_0 \neq 0$.

1. Let $B = 1600\sqrt{5}|d|$.
2. For all $a_0, a_1, a_2 \in \mathbb{Z}$ with $|a_i| \leq B$ and $a_0 \neq 0$ we do the following.
   (a) Let $Z = a_0,\ Y = a_0 a_2 - a_1^2$.
   (b) Determine the at most two solutions a_3 of $X = \pm\sqrt{-Y^3 - dZ^5}$ and $a_0^2 a_3 - 3a_0 a_1 a_2 + 2a_1^3 = 2X$
   (c) Compute $a_4, \ldots, a_{12}$ from the equations defining $C_5(d)$.

(d) If all $a_3, \ldots, 7a_6, \ldots, a_{12}$ are integers and if they satisfy the bounds of Corollary 9.6 then we output the form $[a_0, \ldots, a_{12}]$.

When a_0 we follow a similar procedure, but now we can assume $a_1 \neq 0$. The values of $a_3, a_4, \ldots$ follow from the equations $D_4 = 0, D_5 = 0, \ldots$.
We have now a finite set $\mathcal{F}$ of forms in $C_5(d)$. We like to keep only the Hermite reduced ones. For that we determine the representing point $z(f) \in \mathcal{H}$ for each $f \in \mathcal{H}$. This can be a tedious computation, but we use the following observation. Every form $f \in C_5(d)$ is $GL(2,\mathbb{R})$-equivalent to $x_1x_2(x_1^{10} - 11x_1^5x_2^5 - x_2^{10})$. The latter form has four real roots, hence any form in $C_5(d)$ has four real roots. Let f_1 be the factor of f consisting of the four real linear factors of f. Then, by standard arguments as explained in [CS], it turns out that the representing point of f is the same as that of f_1. For the latter there are standard formulas. We delete from $\mathcal{F}$ the non-Hermite reduced forms. We are now left with a full set of representatives of the $SL(2,\mathbb{Z})$-orbits in $C_5(d)$.
In the final listing it saves space to look at $GL(2,\mathbb{Z})$-orbits in $C_5(d)$. Suppose we have a form f which, together with its covariants $H(f), t(f)/2$ represents a set S of solutions to $x^2 + y^3 = dz^5$. Let $g \in GL(2,\mathbb{Z})$ and $\det(g)$. Then, by the covariant property we have $H(f \circ g) = H(f)$ and $t(f \circ g) = -t(f)$. So the form $f \circ g$ represents the set $\{(-x, y, z) | (x, y, z) \in S\}$ of solutions.
Of course we also delete those f from $\mathcal{F}$ that do not give rise to coprime solutions.

11 Appendix A: Parametrizing $X^2 + Y^3 \pm Z^r = 0$

This section has been taken directly from Johnny Edwards's paper [Ed]. It gives complete parametrizations to $X^2 + Y^3 \pm Z^r = 0$ for $r = 3, 4, 5$. In the tables we list the forms

$$f = \sum_{i=0}^{k} \binom{k}{i} a_i x_1^{k-i} x_2^i$$

by the corresponding vector

$$[a_0, a_1, \ldots, a_k]$$

where $k = 4, 6, 12$ if $r = 3, 4, 5$ respectively. From this form we can compute the covariant forms

$$H = \frac{1}{k^2(k-1)^2} \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix}, \quad g = \frac{1}{2k(k-2)} \begin{vmatrix} f_1 & f_2 \\ H_1 & H_2 \end{vmatrix}.$$

Here f_{ij} means $\frac{\partial^2 f}{\partial x_i \partial x_j}$ etc. The forms then satisfy $g^2 + H^3 \pm f^r = 0$ and each give infinitely many integer primitive solutions of the corresponding diophantine equation by specialisation of the polynomial variables. Moreover, solution sets given by different parametrisations are disjoint, and their union is the full solution set. To keep the lists as short as possible, we identify the parametrizations identifying $\pm X$. If the corresponding $GL(2,\mathbb{Z})$ class of f breaks into two $SL(2,\mathbb{Z})$ classes these are really 2 distinct parametrizations.

The case $r = 3$ was already done by Mordell in [Mo], Chapter 25 using a syzygy from invariant theory. The cases $r = 4$ were done by Zagier and quoted in [Beu], appendix A. The $r = 5$ case is new and presented in [Ed].

Complete Parametrization of $X^2 + Y^3 + Z^3 = 0$

$$\begin{aligned} A1 &= [0,1,0,0,-4] \\ A2 &= [-1,0,0,2,0] \\ B1 &= [-2,-1,0,-1,-2] \\ B2 &= [-1,1,1,1,-1] \\ C1 &= [-1,0,-1,0,3] \\ C2 &= [1,0,-1,0,-3] \end{aligned}$$

In Mordell's book [Mo] he further shortens the list by assuming that Z is odd. This means that $A1, B1$ can be omitted. However, Mordell gives 5 parametrizations: $A2, B2, C1, C2$ and $f = [-1,-2,-4,-6,0]$ According to [Ed] the 5th should be superfluous. It turns out that $f(x_1 - 2x_2, x_2)$ is $A2$
In [Beu], on page 78, parametrizations obtained by interchanging Y and Z are identified.

Complete Parametrization of $X^2 + Y^3 \pm Z^4 = 0$

These two equations were solved by Zagier and quoted in [Beu]. In [Co] there is a complete solution according to classical lines and the lines followed by Zagier. To keep the lists short we identify $\pm X$ and $\pm Z$. This means every parametrization in the list is shorthand for $\pm f(x_1, \pm x_2)$. The first $\pm$ is the $\pm Z$.

The equation $X^2 + Y^3 + Z^4 = 0$:

$$\begin{aligned} f_1 &= [0,1,0,0,0,-12,0] \\ f_2 &= [0,3,0,0,0,-4,0] \\ f_3 &= [-1,0,1,0,3,0,-27] \\ f_4 &= [-3,-4,-1,0,1,4,3] \end{aligned}$$

The equation $X^2 + Y^3 - Z^4 = 0$:

$$\begin{aligned} f_1 &= [0,1,0,0,0,12,0] \\ f_2 &= [0,3,0,0,0,4,0] \\ f_3 &= [-1,0,0,2,0,0,32] \end{aligned}$$

$$\begin{aligned} f_4 &= [-1,0,-1,0,3,0,27] \\ f_5 &= [-1,1,1,1,-1,5,17] \\ f_6 &= [-5,-1,1,3,3,3,9] \\ f_7 &= [-7,-1,2,4,4,4,8] \end{aligned}$$

Complete Parametrization of $X^2 + Y^3 + Z^5 = 0$

Beukers in [Beu] was able to produce parametrizations, though his method was unable to produce a complete set. If we identify $\pm X$, we have the following complete set:

$$\begin{aligned} f_1 &= [0,1,0,0,0,0,-144/7,0,0,0,0,-20736,0] \\ f_2 &= [-1,0,0,-2,0,0,80/7,0,0,640,0,0,-102400] \\ f_3 &= [-1,0,-1,0,3,0,45/7,0,135,0,-2025,0,-91125] \\ f_4 &= [1,0,-1,0,-3,0,45/7,0,-135,0,-2025,0,91125] \\ f_5 &= [-1,1,1,1,-1,5,-25/7,-35,-65,-215,1025,-7975,-57025] \\ f_6 &= [3,1,-2,0,-4,-4,24/7,16,-80,-48,-928,-2176,27072] \\ f_7 &= [-10,1,4,7,2,5,80/7,-5,-50,-215,-100,-625,-10150] \\ f_8 &= [-19,-5,-8,-2,8,8,80/7,16,64,64,-256,-640,-5632] \\ f_9 &= [-7,-22,-13,-6,-3,-6,-207/7,-54,-63,-54,27,1242,4293] \\ f_{10} &= [-25,0,0,-10,0,0,80/7,0,0,128,0,0,-4096] \\ f_{11} &= [6,-31,-32,-24,-16,-8,-144/7,-64,-128,-192,-256,256,3072] \\ f_{12} &= [-64,-32,-32,-32,-16,8,248/7,64,124,262,374,122,-2353] \\ f_{13} &= [-64,-64,-32,-16,-16,-32,-424/7,-76,-68,-28,134,859,2207] \\ f_{14} &= [-25,-50,-25,-10,-5,-10,-235/7,-50,-49,-34,31,614,1763] \\ f_{15} &= [55,29,-7,-3,-9,-15,-81/7,9,-9,-27,-135,-459,567] \\ f_{16} &= [-81,-27,-27,-27,-9,9,171/7,33,63,141,149,-67,-1657] \\ f_{17} &= [-125,0,-25,0,15,0,45/7,0,27,0,-81,0,-729] \\ f_{18} &= [125,0,-25,0,-15,0,45/7,0,-27,0,-81,0,729] \\ f_{19} &= [-162,-27,0,27,18,9,108/7,15,6,-51,-88,-93,-710] \\ f_{20} &= [0,81,0,0,0,0,-144/7,0,0,0,0,-256,0] \\ f_{21} &= [-185,-12,31,44,27,20,157/7,12,-17,-76,-105,-148,-701] \\ f_{22} &= [100,125,50,15,0,-15,-270/7,-45,-36,-27,-54,-297,-648] \\ f_{23} &= [192,32,-32,0,-16,-8,24/7,8,-20,-6,-58,-68,423] \\ f_{24} &= [-395,-153,-92,-26,24,40,304/7,48,64,64,0,-128,-512] \\ f_{25} &= [-537,-205,-133,-123,-89,-41,45/7,41,71,123,187,205,-57] \\ f_{26} &= [359,141,-1,-21,-33,-39,-207/7,-9,-9,-27,-81,-189,-81] \\ f_{27} &= [295,-17,-55,-25,-25,-5,31/7,-5,-25,-25,-55,-17,295] \end{aligned}$$

The $GL(2,\mathbb{Z})$ classes of the 27 forms split into 2 distinct $SL(2,\mathbb{Z})$ classes, unless $f = f_3, f_4, f_{12}, f_{17}, f_{18}, f_{27}$. This means that the above list becomes 48 parametrizations if we do not identify $\pm X$. This is a slight correction of [Ed], where the form f_{12} was omitted as giving one $SL(2,\mathbb{Z})$ class.

12 Appendix B: fourth transvectants

In this appendix, again reproduced from [Ed], we reproduce the equations satisfied by f of any form satisfying $g^2 + H^3 + df^r = 0$, where r, g, H are as in Appendix A. These equations are obtained by setting the fourth transvectant of f equal to zero and a further equation to specify scaling. The expressions D_i are the coefficients of the fourth transvectant $\tau_4(f) = \sum_{i=0}^{2k-8} D_i x_1^{r-i} x_2^i$. Note that in all cases to any such form there corresponds a solution X, Y, Z of the equation $X^2 + Y^3 + dZ^r = 0$ by evaluation f, H, g at $(1,0)$,

$$\begin{aligned} Z &= a_0 \\ Y &= a_0 a_2 - a_1^2 \\ 2X &= a_0^2 a_3 - 3a_0 a_1 a_2 + 2a_1^3 \end{aligned}$$

The tetrahedral case $r = 3$

$$\begin{aligned} 0 &= a_0 a_4 - 4a_1 a_3 + 3a_2^2 \\ -4d &= a_0 a_2 a_4 + 2a_1 a_2 a_3 - a_2^3 - a_0 a_3^2 - a_1^2 a_4 \end{aligned}$$

The octahedral case $r = 4$

$$\begin{aligned} D_0/1: \ 0 &= a_4 a_0 - 4a_3 a_1 + 3a_2^2 \\ D_1/2: \ 0 &= a_0 a_5 - 3a_1 a_4 + 2a_3 a_2 \\ D_2/1: \ 0 &= a_0 a_6 - 9a_2 a_4 + 8a_3^2 \\ D_3/2: \ 0 &= a_1 a_6 - 3a_2 a_5 + 2a_3 a_4 \\ -72d &= a_0 a_6 - 6a_1 a_5 + 15a_2 a_4 - 10a_3^2 \end{aligned}$$

The last equation is obtained from $\tau_6(f) = 72d$.

The icosahedral case $r = 5$

$$\begin{aligned} D_0/1: \ 0 &= a_0 a_4 - 4a_1 a_3 + 3a_2^2 \\ D_1/8: \ 0 &= a_0 a_5 - 3a_1 a_4 + 2a_3 a_2 \end{aligned}$$

$$\begin{aligned}
D_2/4: \ 0 &= a_0(7a_6) - 12a_1a_5 - 15a_2a_4 + 20a_3^2 \\
D_3/56: \ 0 &= a_0a_7 - 6a_2a_5 + 5a_3a_4 \\
D_4/14: \ 0 &= 5a_0a_8 + 12a_1a_7 - 6a_2(7a_6) - 20a_3a_5 + 45a_4^2 \\
D_5/56: \ 0 &= a_0a_9 + 6a_1a_8 - 6a_2a_7 - 4a_3(7a_6) + 27a_4a_5 \\
D_6/28: \ 0 &= a_0a_{10} + 12a_1a_9 + 12a_2a_8 - 76a_3a_7 - 3a_4(7a_6) + 27a_4a_5 \\
D_7/8: \ 0 &= a_0a_{11} + 24a_1a_{10} + 90a_2a_9 - 130a_3a_8 - 405a_4a_7 + 60a_5(7a_6) \\
D_8/1: \ 0 &= a_0a_{12} + 60a_1a_{11} + 534a_2a_{10} + 380a_3a_9 - 3195a_4a_8 \\
&\quad -720a_5a_7 + 60(7a_6)^2 \\
D_9/8: \ 0 &= a_1a_{12} + 24a_2a_{11} + 90a_3a_{10} - 130a_4a_9 - 405a_5a_8 + 60(7a_6)^2
\end{aligned}$$

By elimination of a_6, a_7 from $D_2 = D_3 = D_4 = 0$ we get

$$D_4^*: \ a_0^3a_8 = 12a_4a_3a_1a_0 + 18a_4a_2^2a_0 - 24a_3^2a_2a_0 + 4a_5a_3a_0^2 - 9a_4^2.$$

From $\tau_6(f) = 360df$ we get by comparison of the coefficients of x_1^{12} and $x_1^{11}x_2$,

$$\begin{aligned}
R_0/1: \ 360da_0 &= a_0(7a_6) - 42a_1a_5 + 105a_2a_4 - 70a_3^2 \\
R_1/6: \ 720da_1 &= 7a_0a_7 - 5a_1(7a_6) + 63a_2a_5 - 35a_3a_4
\end{aligned}$$

13 References

[BCDT] C.Breuil, B.Conrad, F.Diamond, R.Taylor, On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises, J.Amer.Math.Soc. 14 (2001), 843-939.

[Bec] S. Beckmann, On extensions of number fields obtained by specializing branched coverings, J.reine angew. Math. 419 (1991), 27-53.

[Ben] M.A.Bennett, The equation $x^{2n} + y^{2n} = z^5$.

[Beu] F.Beukers, The diophantine equation $Ax^p + By^q = Cz^r$, Duke Math.J. 91(1998), 61-88.

[Br1] N.Bruin, The diophantine equations $x^2 \pm y^4 = \pm z^6$ and $x^2 + y^8 = z^3$, Compositio Math. 118 (1999), 305-321.

[Br2] N.Bruin, Chabauty methods using elliptic curves, J.reine angew. Math. 562 (2003), 27-49.

[Br3] N.Bruin, On powers as sums of two cubes, ANTS IV, Leiden 2000, 169-184, Lecture Notes in Comput.Sci. 1838, Springer 2000.

[Br4] N.Bruin, Visualising Sha[2] in abelian surfaces, Math.Comp 73(2004), 1459-1476 (electronic).

[BS], M.A.Bennett, C.Skinner, Ternary diophamtine equations via Galois representations and modular forms, Canad.J.Math. 56(2004), 23-54.

[Ch] I.Chen, On the equation $s^2 + y^{2p} = \alpha^3$, preprint July 2004.

[Co] H.Cohen, *Diophantine Equations, p-adic Numbers and L-Functions*, book in preparation.

[CS], J.Cremona, M.Stoll, On the reduction theory of binary forms, J.reine angew. Math. 565(2003), 79-99.

[D], H.Darmon, The equation $x^4 - y^4 = z^p$, C.R.Math.Rep.Acad.Sci. Canada 15 (1993), 286-290.

[DG] H.Darmon, A.Granville, On the equations $z^m = F(x, y)$ and $Ax^p + By^q = Cz^r$, Bull. London Math.Soc. 27(1995), 513-543.

[DM] H.Darmon, L.Merel, Winding quotients and some variants of Fermat's Last Theorem, J.reine angew.Math. 490 (1997), 81-100.

[Ed] J.Edwards, A complete solution to $X^2 + Y^3 + Z^5 = 0$, J.reine angew. Math. 571 (2004), 213-236.

[El] J.Ellenberg, Galois representations attached to $\mathbb{Q}$-curves and the generalised Fermat equation $A^4 + B^2 = C^p$, Amer.J.Math 126 (2004), 763-787.

[H] D.Hilbert, *Theory of algebraic invariants*, Cambridge Mathematical Library, 1993.

[Kl] F.Klein, *Vorlesungen Über das Ikosaeder*, New edition with a preface by P.Slodowy, Birkhäuser-Teubner, 1993.

[Kr1] A.Kraus, Sur l'équation $a^3 + b^3 = c^p$, Experiment.Math.7 (1998), 1-13.

[Kr2] A.Kraus, On the equation $x^p + y^q = z^r$: a survey, Ramanujan J. 3 (1999), 315-333.

[Me] L.Merel, Arithmetic of elliptic curves and diophantine equations, J.Théor. Nombers Bordeaux 11(1999), 173-200.

[Mo] L.E.Mordell, *Diophantine Equations*, Academic Press, London 1969.

[P] B.Poonen, Some diophantine equations of the form $x^n + y^n = z^m$, to appear

[PSS] B.Poonen, E.Schaefer, M.Stoll, preprint, www.arxiv.org/pdf/math.NT/0508174.

[Si] J.Silverman, *The Arithmetic of Elliptic Curves*, Springer 1986.

[TW] R.Taylor, A.Wiles, Ring-theoretic properties of certain Hecke algebras, Ann. of Math. 141 (1995), 553-572.

[W] A.Wiles, Modular elliptic curves and Fermat's Last Theorem, Ann. of Math, 141 (1995), 443-551.