# ON CERTAIN PLANE CURVES WITH MANY INTEGRAL POINTS

F. Rodríguez Villegas and J. F. Voloch

November 1997

**0.** In the course of another investigation we came across a sequence of polynomials $P_d \in \mathbb{Z}[x,y]$, such that $P_d$ is absolutely irreducible, of degree $d$, has low height and at least $d^2+2d+3$ integral solutions to $P_d(x,y)=0$. We know of no other family of polynomials of increasing degree with as many integral (or even rational) solutions in terms of their degree, except of course those with infinitely many rational points.

It is a consequence of Siegel's theorem [Si] that these polynomials have finitely many integral zeros, since their homogeneous part of highest degree has distinct roots. Siegel [Si, §7] speculated whether there is a bound for the number of integral zeros of a polynomial as a function of the number of non-zero coefficients, provided it has only finitely many zeros. This is still very much of an open problem but Caporaso et al. [CHM] have shown that a similar statement for rational points on curves (with the genus replacing the number of coefficients) would follow from a conjecture of Lang. Abramovich [ℵ] proved an analogue of the result of [CHM] for integral points on elliptic curves. See also [ℵV].

A polynomial in two variables and degree $d$ has $N=\binom{d+2}{2}$ coefficients, so, given points $(x_1,y_1),\dots,(x_{N-1},y_{N-1})$, one can find a non-zero polynomial that vanishes on these points. If these points have integer coordinates of absolute value at most $H$, then such a polynomial can be chosen with integer coefficients of absolute value at most $(NH^d)^N$, by a straightforward application of Siegel's lemma. We can choose $H=N/2$, for instance, and it will turn out that our polynomials $P_d$ have slightly lower height and twice as many points as this construction gives. Moreover, this construction does not ensure that the polynomial obtained is absolutely irreducible. A slightly better construction, suggested by Ed Schaffer is to take a polynomial of the shape $(x-x_1)\cdots(x-x_d)+\alpha(y-y_1)\cdots(y-y_d)$, which will vanish on the $d^2$ points $(x_i,y_j), i,j=1,\dots,d$, is irreducible for most chioces of $\alpha$ and has height at most $|\alpha|H^d$. Our polynomials $P_d$ have larger height but more points.

We have checked that $P_d=0$ defines a smooth curve for $d=1,2,\cdots,25$. We do not know whether this is true in general, though it is very likely. Also, we can

prove the existence of certain points on the curve, but numerical experimentation shows that they may contain a few more. We present the data in n° 7.

**1.** Let $T_k \in \mathbb{Z}[x,y]$ be defined recursively by

$$\textbf{(T0)} \qquad \begin{aligned} &T_0 = 1, \qquad T_1 = y, \\ &T_{k+1} = yT_k + k(x+k-1)T_{k-1}, \qquad k \in \mathbb{N}. \end{aligned}$$

The first few polynomials are

$$\begin{aligned} T_2 &= x + y^2 \\ T_3 &= 3yx + y^3 + 2y \\ T_4 &= 3x^2 + 6y^2x + 6x + y^4 + 8y^2 \\ T_5 &= 15yx^2 + 10y^3x + 50yx + y^5 + 20y^3 + 24y \\ T_6 &= 15x^3 + 45y^2x^2 + 90x^2 + 15y^4x + 210y^2x + 120x + y^6 + 40y^4 + 184y^2. \end{aligned}$$

From the recursion it follows easily that

$$\textbf{(T1)} \qquad T_k(x,-y) = (-1)^k T_k(x,y), \qquad k \in \mathbb{N}.$$

Hence for $k = 2d$ with $d \in \mathbb{N}$, $T_k(x,y) = P_d(-x, y^2)$ with $P_d \in \mathbb{Z}[x,y]$.

We will use the following notation: given a polynomial

$$H = \sum_{m,n} a_{m,n} x^m y^n \quad \in \mathbb{C}[x,y],$$

we let

$$||H||_1 = \sum_{m,n} |a_{m,n}|.$$

We will prove the following.

**Theorem.** *Let $d \in \mathbb{N}$ and $P_d$ be the polynomial defined above. Then*

*a) $P_d$ has degree $d$;*

*b) $P_d$ is absolutely irreducible;*

*c) the coefficients of $P_d(-x,y)$ are relatively prime non-negative integers;*

*d) $||P_d||_1 = (2d)!$ ; and*

*e) $P_d$ vanishes at the $d^2 + 2d + 3$ integral points:*

$$\begin{aligned} I: &\quad (n,0),\ (n,2^2),\ (n,4^2), \cdots,\ (n,n^2), \qquad 0 \le n \le 2d-1, \qquad n \text{ even} \\ II: &\quad (n,1^2),\ (n,3^2),\ (n,5^2), \cdots,\ (n,n^2), \qquad 0 \le n \le 2d-1, \qquad n \text{ odd} \\ III: &\quad (4d,2^2),\ (4d,6^2),\ (4d,10^2), \cdots,\ (4d,4(2d-1)^2) \\ IV: &\quad (8d+1,3^2),\ (2d-4,-6d+4),\ (2d-3,-2d+1) \end{aligned}$$

Note that $P_d, P_{d+1}$ intersect in exactly $d(d+1)$ of the above points.

**2.** Fix $x, y$ and consider the generating function

$$F(\lambda) = \sum_{k=0}^{\infty} \frac{T_k}{(x)_k} \frac{\lambda^k}{k!},$$

where

$$(z)_0 = 1, \qquad (z)_k = z(z+1)\cdots(z+k-1), \qquad k \in \mathbb{N}.$$

It is not hard to see that the recursion defining $T_k$ implies that $F$ satisfies the differential equation

$$\lambda \frac{d^2F}{d\lambda^2} + x\frac{dF}{d\lambda} - (\lambda + y)F = 0.$$

In order to get a formula for $T_k$ we consider $G(\lambda) = e^{\lambda}F(\lambda)$. A calculation shows that $G$ satisfies the differential equation

$$\lambda \frac{d^2G}{d\lambda^2} + (x - 2\lambda)\frac{dG}{d\lambda} - (x+y)G = 0.$$

It follows that

$$G(\lambda) = \Phi(\tfrac{1}{2}(x+y), x, 2\lambda),$$

where $\Phi$ is the confluent hypergeometric function (see for example, [Le §9.9]).

If we write

$$G(\lambda) = \sum_{k=0}^{\infty} \frac{S_k}{(x)_k} \frac{\lambda^k}{k!},$$

the differential equation implies that

$$S_{k+1} = (y + x + 2k)S_k, \qquad k \in \mathbb{N}.$$

Therefore,

$$S_k = (y+x)(y+x+2)\cdots(y+x+2k-2),$$

from which we obtain

$$\textbf{(T2)} \quad T_k = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (x+y)(x+y+2)\cdots(x+y+2j-2)$$
$$(x+j)(x+j+1)\cdots(x+k-1).$$

We now may see why $P_d$ vanishes at the points $I$ and $II$ of the theorem. The principle is based on the following self-proving lemma; we leave the details to the reader.

**Lemma.** *Let $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ be two sets of $n$ elements of a field $K$. Let*

$$\phi_0 = 1, \qquad \phi_\nu(x) = (x - x_1)(x - x_2)\cdots(x - x_\nu), \in K[x] \qquad 1 \le \nu \le n$$
$$\psi_0 = 1, \qquad \psi_\nu(y) = (y - y_1)(y - y_2)\cdots(y - y_\nu), \in K[y] \qquad 1 \le \nu \le n,$$

*with $x, y$ indeterminates. Then any linear combination*

$$\sum_{\nu=0}^{n} \alpha_\nu \ \phi_\nu(x) \ \psi_{n-\nu}(y) \in K[x,y], \qquad \alpha_\nu \in K,$$

*has degree at most $n$ and vanishes at the points $(x_\mu, y_\nu)$ for all $1 \leq \mu \leq \nu \leq n$.*

**3.** It is clear from the recursion **T0** that $T_k$ has degree $k$, that the coefficients of $T_k$ are non-negative integers and that the coefficient of $y^k$ is 1. This proves parts a) and c) of the theorem. To prove part d), let $c_k = T_k(1,1)$. Note that $c_k = ||T_k||_1$ since the coefficients of $T_k$ are non-negative. From the recursion

$$\begin{aligned} &c_0 = 1, \qquad c_1 = 1, \\ &c_{k+1} = c_k + k^2 c_{k-1}, \qquad k \in \mathbb{N}. \end{aligned}$$

It follows easily that $c_k = k!$ hence

(**T3**) $$||T_k||_1 = k! \ , \qquad k \in \mathbb{N}.$$

Let us also remark that **T2** implies the following

(**T4**) $$\frac{T_k(m,n)}{k!} \in \mathbb{Z}, \qquad \text{for all } m, n \in \mathbb{Z}.$$

**4.** Let $\widetilde{T}_k = z^k T_k(x/z^2, y/z)$. Then $\widetilde{T}_k$ is isobaric of weight $k$, if we assign $x$ weight 2, $y$ weight 1, and $z$ weight 1. These polynomials satisfy the recursion

$$\begin{aligned} &\widetilde{T}_0 = 1, \qquad \widetilde{T}_1 = y, \\ &\widetilde{T}_{k+1} = y\widetilde{T}_k + k(x + z^2(x-1))\widetilde{T}_{k-1}, \qquad k \in \mathbb{N}. \end{aligned}$$

Set now $R_k = \widetilde{T}_k(1,t,0)$, the leading terms of $\widetilde{T}_k$ at infinity. Then

$$\begin{aligned} &R_0 = 1, \qquad R_1 = t, \\ &R_{k+1} = tR_k - kR_{k-1}, \qquad k \in \mathbb{N}. \end{aligned}$$

It follows that $R_k(t) = 2^{-k/2} H_k(t/\sqrt{2})$, where $H_k$ is the classical Hermite polynomial (see for example [Le §4.9]). More precisely,

(**T5**) $$R_k(t) = z^k T(1/z^2, t/z)\big|_{z=0} = k! \sum_{j=0}^{[k/2]} \frac{(-1)^j}{j! \ (k-2j)! \ 2^j} \ t^{k-2j}$$

It is interesting that the discriminant can be computed explicitly

$$\operatorname{disc}(R_k) = \prod_{j=1}^{k} j^j,$$

but we only need to know that it is non-zero.

**Lemma.** *Let $K$ be a perfect field and $\overline{K}$ an algebraic closure of $K$. Let $P \in K[x,y,z]$ be a homogeneous polynomial of degree $d$. Suppose that $P(t,1,0) \in K[t]$ also has degree $d$, is irreducible over $K$ and $P(x,y,z) = 0$ has more than $d^2/4$ projective solutions over $K$. Then $P$ is irreducible over $\overline{K}$.*

*Proof.* Since $P(t,1,0)$ has degree $d$ and is irreducible over $K$ it follows that $P(x,y,z)$ is also irreducible over $K$. Suppose $P$ is not absolutely irreducible. Then, $P = \prod_\sigma Q^\sigma$, where $Q$ is an irreducible factor of $P$ over $\overline{K}$ of degree $e \leq d/2$ and $\sigma$ runs through the embeddings of the field of definition of $Q$ into $\overline{K}$. Any $K$-rational point of $P = 0$ is a rational point of $Q^\sigma = 0$ for every $\sigma$. Since the $Q^\sigma$'s are all distinct, Bezout's theorem implies that the number of $K$-rational points of $P = 0$ is bounded by $e^2 \leq d^2/4$, a contradiction. □

According to Schur [Sc] the polynomials $R_k$ for $k$ even and $R_k/t$ for $k$ odd are irreducible over $\mathbb{Q}$. Hence, the above lemma applies and we deduce part b) of the theorem.

**5.** For $p > 2$ a prime number let us consider the recursion defining $T_k$ modulo $p$. It turns out to have a very simple structure. First, from **T2** it follows that

$$T_p \equiv y^p - y \bmod p, \qquad p > 2, \qquad p \text{ prime.}$$

Also,

$$T_{p+k} \equiv (y^p - y)T_k \bmod p$$
$$T_{p+k+1} \equiv yT_{p+k} + k(x+k-1)T_{p+k-1} \bmod p$$

and we conclude that

(**T6**) $T_k \equiv T_{a_0}(y^p - y)^{a_1}(y^{p^2} - y^p)^{a_2} \cdots \bmod p, \qquad k = a_0 + a_1 p + a_2 p^2 \cdots \in \mathbb{N}.$

**6.** We now prove that $P_d$ vanishes on the points $III$ of the theorem. First we need the following. For each $k \in \mathbb{N}$ consider the polynomials

$$U_k(z,w) = T_k(x,y), \qquad z = \tfrac{1}{2}(x-y), \qquad w = x - k + 1.$$

Let $\lambda$ be an indeterminate and $z, w$ two fixed integers. Then using **T2** we obtain

(**T7**) $$\sum_{k=0}^{\infty} U_k(z,w)\frac{\lambda^k}{k!} = \frac{(1+2\lambda)^z}{(1+\lambda)^w}, \qquad z,w \in \mathbb{Z}.$$

From this identity it is not hard to see that

(**T8**) $$\frac{U_k(z,w)}{k!} = \sum_{j=0}^{w-1}(-2)^j \binom{z}{j}\binom{k+w-j-1}{w-j-1}, \qquad 0 \leq z \leq w.$$

Note that the right hand side is a polynomial of degree $w-1$ in $k$. Without the hypothesis $z \leq w$ **T8** holds for all $k$ sufficiently large. In particular, given integers $z, w$ there is only finitely many polynomials $U_k$ that vanish at the point $(z, w)$, for $w \geq 0$.

It follows that $P_d$ vanishes at the points III if

$$(*) \qquad \sum_{j=0}^{m} (-2)^j \binom{m}{k} \binom{2k-j}{k} = 0, \qquad 0 \leq m \leq k, \quad m \text{ odd},$$

where $k = 2d$.

To prove this identity we start with

$$\binom{a+b}{k} = \sum_{r=0}^{a} \binom{a}{r} \binom{b}{k-r}, \qquad a, b \in \mathbb{Z}_{\geq 0},$$

which follows from the binomial theorem by comparing the $k$-th coefficients on both sides of

$$(1+\lambda)^{a+b} = (1+\lambda)^a (1+\lambda)^b.$$

Applying this to $a = m-j, b = 2k-m$ we obtain

$$\binom{2k-j}{k} = \sum_{r=0}^{m-j} \binom{m-j}{r} \binom{2k-m}{k-r}$$

and hence (*) is equivalent to

$$\sum_{j=0}^{m} \sum_{r=0}^{m-j} (-2)^j \binom{m}{j} \binom{m-j}{r} \binom{2k-m}{k-r} = 0.$$

This in turn follows from the stronger fact

$$\sum_{j=0}^{m-r} (-2)^j \binom{m}{j} \binom{m-j}{r} = (-1)^m \sum_{j=0}^{r} (-2)^j \binom{m}{j} \binom{m-j}{m-r},$$

since $\binom{2k-m}{k-r} = \binom{2k-m}{k-m+r}$, obtained by expanding

$$(\lambda - 1)^m = (\lambda + 1 - 2)^m$$

and comparing the coefficients of $\lambda^r$ and $\lambda^{m-r}$ respectively.

The fact that the points listed in $IV$ are in $P_d = 0$ will be left to the reader.

**7.** We now present the experimental data. We first discuss the cases $d = 3, 4$ in more detail, where the equations $P_d(x, y) = 0$ determine smooth projective curves of genus $1, 3$ respectively.

For $d = 3$ we have

$$P_3 = -15x^3 + 45yx^2 + 90x^2 - 15y^2x - 210yx - 120x + y^3 + 40y^2 + 184y.$$

The equation $P_3 = 0$ defines an elliptic curve and a Weierstrass equation for it (courtesy of F. Hajir) is given by:

$$y^2 = x^3 + 230940/23 * x^2 + 9286041600/529 * x + 90438421708800/12167.$$

A computer search yielded the following 25 integral solutions $(x, y)$ to $P_3(x, y) = 0$.

$$\begin{array}{cccccccc} (0,0) & (1,1) & (2,-14) & (2,0) & (2,4) & (3,-5) & (3,1) & (3,9) \\ (4,0) & (4,4) & (4,16) & (5,1) & (5,9) & (5,25) & (9,25) & (12,4) \\ (12,36) & (12,100) & (16,144) & (25,9) & (67,25) & (345,1225) & (-1,-9) & (-4,-20) \\ (-14,-56) & & & & & & & \end{array}$$

For $d = 4$

$$\begin{aligned} P_4 = {} & 105x^4 - 420x^3y - 1260x^3 + 210x^2y^2 + 4200x^2y + 4620x^2 \\ & - 28xy^3 - 1540xy^2 - 11872xy - 5040x + y^4 + 112y^3 + 2464y^2 + 8448y \end{aligned}$$

A computer search yielded the following 31 integral solutions $(x, y)$ to $P_4(x, y) = 0$.

$$\begin{array}{cccccccc} (0,0) & (2,0) & (4,0) & (6,0) & (1,1) & (3,1) & (5,1) & (7,1) \\ (3,-3) & (2,4) & (4,4) & (6,4) & (16,4) & (5,-7) & (3,9) & (5,9) \\ (7,9) & (33,9) & (4,16) & (6,16) & (4,-20) & (0,-24) & (5,25) & (7,25) \\ (3,-35) & (6,36) & (16,36) & (7,49) & (16,10) & (16,196) & (-11,-35) & \end{array}$$

For higher $d$ we have the following data, where we only present those points not given by the Theorem. We searched exhaustively for points with $|x| \leq 1000$. We haven't found any patterns in the extra points; perhaps a more attentive reader will.

| $d$ | *new points* | *total number of points* |
|---|---|---|
| 5 | $(16,144), (17,81), (25,441), (99,589)$ | 42 |
| 6 | $(1,-11), (17,121), (34,784)$ | 54 |
| 7 | $(16,16), (17,49), (25,169), (36,676), (98,16)$ | 71 |
| 8 | *none* | 85 |
| 9 | $(9,-35), (33,289)$ | 104 |
| 10 | *none* | 123 |
| 11 | $(34,784), (36,676), (41,441), (57,2601), (67,3249)$ | 160 |
| 12 | *none* | 171 |

To verify that $P_d = 0$ defines a smooth curve is enough to check that it has no affine singularities as the Hermite polynomial is separable. For this we computed, using the recursion modulo $p$, the quantity

$$\mathrm{Res}_y(\mathrm{Res}_x(P_d, \frac{\partial P_d}{\partial x}), \mathrm{Res}_x(P_d, \frac{\partial P_d}{\partial y})) \bmod p,$$

for various primes $p$, where $\mathrm{Res}_t$ stands for resultant in the variable $t$, and confirmed it is not zero for $d = 2, 3, \ldots, 25$.

Acknowledgements: We would like to thank R. Coleman, A. Granville, F. Hajir and B. Poonen for suggestions, and the NSF (FRV) and NSA (JFV) for financial support. We also acknowledge the use of the software PARI for the numerical calculations. The routines we used are available at the URL: `http://www.ma.utexas.edu/users/voloch/polynomial.html`.

## References

[ℵ] D. Abramovich, *Uniformity of stably integral points on elliptic curves*, Invent. Math. **127** (1997), 307-317.

[ℵV] D. Abramovich, J. F. Voloch, *Lang's conjectures, fibered powers, and uniformity*, New York J. Math. **2** (1996), 20-34.

[CHM] L. Caporaso, J. Harris, B. Mazur, *Uniformity of rational points*, J. Amer. Math. Soc. **10** (1997), 1-35.

[Le] N. N. Lebedev, *Special functions and their applications*, Dover Publications, 1972.

[Si] C. L. Siegel, *Uber einige Anwendungen Diophantischer Approximationen*, Gesammelte Abhandlungen I, Springer Verlag, 1966, pp. 209-266.

[Sc] I. Schur, *Affektlose Gleichungen in der Theorie der Laguerreschen und Hermiteschen Polynome*, J. Reine Angew. Math. **165** (1931), 52-58.

Department of Mathematics
University of Texas at Austin
Austin, TX 78712 USA
*E-mail address*: `villegas@math.utexas.edu, voloch@math.utexas.edu`