Astérisque

ROGER HEATH-BROWN

Counting rational points on cubic surfaces

Astérisque, tome 251 (1998), p. 13-30

<http://www.numdam.org/item?id=AST_1998__251__13_0>

Astérisque **251**, 1998, p. 13–29

COUNTING RATIONAL POINTS ON CUBIC SURFACES

by

Roger Heath-Brown

Abstract. — Let $F[W,X,Y,Z]$ be a rational cubic form, and let $N^{(0)}(R)$ be the number of rational zeros of F of height at most R, which do not lie on any rational line in the surface $F=0$. We show that

$$N^{(0)}(R) \ll_{\varepsilon,F} R^{4/3+\varepsilon}$$

for any fixed $\varepsilon > 0$, subject to a suitable hypothesis on the size of the rank of elliptic curves. For the proof one counts points on the cubic curves obtained from hyperplane sections of the surface $F=0$.

1. Introduction

For any cubic form $F(W,X,Y,Z) \in \mathbb{Z}[W,X,Y,Z]$ let

$$N_F(R) = N(R) = \#\{\mathbf{x} \in \mathbb{Z}^4 : \ F(\mathbf{x}) = 0, \ |\mathbf{x}| \leqslant R, \ \mathbf{x} \text{ primitive}\},$$

where $|\mathbf{x}|$ is the Euclidean length of $\mathbf{x}$, and an integer vector $\mathbf{x} = (x_1, \dots, x_n)$ is defined to be primitive if $\mathbf{x} \neq \mathbf{0}$ and $x_1, \dots, x_n$ have no common factor. We are concerned here with the size of $N(R)$ as R tends to infinity. If the surface $F=0$ contains a rational line, then there will be $cR^2 + O_\varepsilon(R^{1+\varepsilon})$ primitive points on that line, counted by $N(R)$, for any positive ε and an appropriate constant $c > 0$. One would expect that such 'trivial' points greatly outnumber the remaining 'non-trivial' points and we therefore define $N^{(0)}(R)$ to be the number of points $\mathbf{x}$ counted by $N(R)$, such that $\mathbf{x}$ does not lie on any rational line in the surface $F=0$. There are some very precise conjectures about the size of $N^{(0)}(R)$, (see Franke, Manin, and Tschinkel [2], for example). However very little has been proved in general. Indeed, as far as the author is aware the following question is still open.

1991 ***Mathematics Subject Classification.*** — Primary 11G35; Secondary 11D41, 11E76, 11G05.

Key words and phrases. — Cubic Surface, Rational Points, Height, Upper Bound, Elliptic Curve, Rank.

Question 1. — *Is it true that* $N_F(R) \ll_{\varepsilon,F} R^{2+\varepsilon}$, *for any irreducible* F *and any* $\varepsilon > 0$ *?*

The notation $\ll_{\varepsilon,F}$ means that the implied constant may depend on both ε and F. A very general result has been given by Pila [8] which shows in particular that

$$N_F(R) \ll_{\varepsilon} R^{7/3+\varepsilon},$$

for any $\varepsilon > 0$, uniformly for all absolutely irreducible cubic forms F.

One might indeed be more ambitious and ask for a positive answer to the following.

Question 2. — *Is there a constant* $\theta < 2$ *such that*

$$N^{(0)}(R) \ll_F R^{\theta},$$

for every F*?*

This would demonstrate that points on rational lines really do dominate the rate of growth of $N_F(R)$. Progress has been made in certain special cases, and the author has recently shown [3] that

$$N^{(0)}(R) \ll_{\varepsilon,F} R^{4/3+\varepsilon} \tag{1}$$

for any $\varepsilon > 0$, and any non-singular F such that the surface $F = 0$ contains three coplanar rational lines. In particular (1) holds for

$$F(W, X, Y, Z) = W^3 + X^3 + Y^3 + Z^3.$$

Ideally however one would hope for an affirmative answer to the following question.

Question 3. — *Is it true that* $N^{(0)}(R) \ll_{\varepsilon,F} R^{1+\varepsilon}$, *for any* F *and any* $\varepsilon > 0$ *?*

This is only known to be true in rather trivial cases, such as those in which $N^{(0)}(R)$ is equal to 0, or forms of the shape $W^3 - XYZ$, for example.

We shall be concerned with Question 2, and it is our goal to describe an approach which yields a satisfactory answer, subject to the following natural hypothesis about elliptic curves.

Rank Hypothesis. — *For any rational elliptic curve* E *let* C_E *denote the conductor and let* r_E *denote the rank. Then we have*

$$r_E = o(\log C_E) \quad as \quad C_E \to \infty.$$

To put this into context, we observe that

$$r_E = O(\log C_E) \tag{2}$$

for all rational elliptic curves, and

$$r_E \ll \frac{\log C_E}{\log\log C_E} = o(\log C_E) \tag{3}$$

for any rational elliptic curve with at least one rational point of order 2. These assertions ought to be well-known. However we include proofs in §6, for the sake of completeness. We remark that Mestre [7] has shown, subject to the conjecture of Birch and Swinnerton-Dyer, that (3) holds for every modular rational elliptic curve E, providing the associated L-function satisfies the Riemann Hypothesis.

We can now state our principal result.

Theorem 1. — *If the Rank Hypothesis holds then*

$$N^{(0)}(R) \ll_{\varepsilon,F} R^{4/3+\varepsilon}$$

for any non-singular form F and any $\varepsilon > 0$.

The proof of Theorem 1 involves taking hyperplane sections through the surface $F = 0$, to produce a number of cubic curves, most of which will be non-singular. We then estimate the number of points on each of these curves. We remark that this line of attack can be considerably generalized. Thus one can take a completely arbitrary variety, and attempt to count how many points may lie on each of its plane sections. However we shall not explore this possibility here.

To count points on our cubic curves we introduce the following definitions. Let $G(X, Y, Z) \in \mathbb{Z}[X, Y, Z]$ be a cubic form, and let $||G||$ denote the maximum modulus of the coefficients of G. Now define $N(G, R)$ to be the number of primitive points $\mathbf{x} \in \mathbb{Z}^3$ in the sphere $|\mathbf{x}| \leqslant R$ for which $G(\mathbf{x}) = 0$, with the proviso that if $L(\mathbf{x})$ is a rational linear factor of G, then any points on $L(\mathbf{x}) = 0$ are to be excluded. Of course this latter case can only arise when G is singular.

Our principal results on $N(G, R)$ are then the following.

Theorem 2. — *Let $G(X, Y, Z) \in \mathbb{Z}[X, Y, Z]$ be an absolutely irreducible singular cubic form. Then*

$$N(G, R) \ll_{\varepsilon} R^{2/3+\varepsilon}||G||^{\varepsilon}, \tag{4}$$

for any $\varepsilon > 0$.

Theorem 3. — *Let $G(X, Y, Z) \in \mathbb{Z}[X, Y, Z]$ be a non-singular cubic form. Then if the Rank Hypothesis holds we will have*

$$N(G, R) \ll_{\varepsilon} R^{\varepsilon}||G||^{\varepsilon},$$

for any $\varepsilon > 0$.

In Theorem 2 the exponent $2/3$ is best possible, as the example $G(X, Y, Z) = XY^2 - Z^3$ shows. This vanishes at (a^3, b^3, ab^2), which takes at least $\gg R^{2/3}$ primitive values in the sphere of radius R.

For Theorem 3, it is easy to show that

$$N(G,R) \ll_{\varepsilon,G} R^{\varepsilon},$$

and the real difficulty lies in establishing a good dependence on $||G||$. We shall also have to handle the case in which G is reducible, but the estimates we shall obtain (in §5) depend on the coefficients of G in a more delicate manner than in the cases above. It would be interesting to obtain unconditional bounds of the type given by Theorem 3, with R^{ε} replaced by a term with a larger exponent. The best such result currently available seems to be the estimate

$$N(G,R) \ll_{\varepsilon} R^{4/3+\varepsilon},$$

due to Pila [8]. This holds uniformly in G, for any fixed $\varepsilon > 0$.

2. Proof of Theorem 2

We begin our treatment of Theorem 2 by observing that G has exactly one singular point, which is therefore rational. We take the singular point to be the primitive integer vector $\mathbf{x}_0$. Elimination theory shows that $|\mathbf{x}_0| \ll ||G||^A$ for a suitable absolute constant A.

At this point it is convenient to introduce a convention concerning the large number of absolute constants which will occur in what follows. All such constants will be denoted by the same letter A, which therefore has a potentially different meaning at each occurence. This notation allows us to write $|\mathbf{x}_0|^A \ll ||G||^A$, for example, given that $|\mathbf{x}_0| \ll ||G||^A$. Such a convention needs to be treated with caution, but it avoids the notational complications of introducing A, A', A''' etc.

Proceeding with our argument, there will be an invertible integer change of variables, T say, which sends $\mathbf{x}_0$ to the point $(1,0,0)$ and with coefficients which are $O(||G||^A)$. This takes G to a form G' with $||G'|| \ll ||G||^A$. The form $G'(X,Y,Z)$ may now be written as $XQ(Y,Z) + C(Y,Z)$ where Q and C are quadratic and cubic forms respectively, and $||Q||, ||C|| \ll ||G||^A$. If $G(\mathbf{x}) = 0$ for some primitive vector $\mathbf{x}$, the corresponding triple $T\mathbf{x} = (X,Y,Z)$ will also be primitive. Thus if $Y = Z = 0$, then $X = \pm 1$. Otherwise we may set $Y = ry$, $Z = rz$ with y, z coprime, so that

$$XQ(y,z) + rC(y,z) = 0. \tag{5}$$

The polynomials Q and C are coprime, since the original cubic G is supposed to be absolutely irreducible. It follows from an application of the Euclidean algorithm that there are a quadratic form Q', a linear form L', and a non-zero constant K_1, all of which are integral, which satisfy

$$Q(Y,Z)Q'(Y,Z) + C(Y,Z)L'(Y,Z) = K_1 Y^4$$

identically. Here K_1 may depend on the original form G, and on the linear transformation T. It is clear, when one examines the algorithm, that the constant K_1

will satisfy $K_1 \ll \max\{||Q||, ||C||\}^A$. Since $Q(y,z)|rC(y,z)$ we conclude that $Q(y,z)|rK_1y^4$, with $K_1 \ll ||G||^A$.

In a precisely analogous way one finds that $Q(y,z)|rK_2z^4$, for some non-zero integer constant $K_2 \ll ||G||^A$. Thus $Q(y,z)$ divides both $rK_1K_2y^4$ and $rK_1K_2z^4$, and since y and z are coprime it follows that $Q(y,z)$ must be a factor of rK_1K_2. On the other hand r and X must be coprime, since our original vector $\mathbf{x}$ was assumed to be primitive, and hence (5) shows that $r|Q(y,z)$. We therefore conclude that $Q(y,z) = rK$ for some divisor K of K_1K_2. The number of factors of K_1K_2 is $O_\varepsilon(||G||^\varepsilon)$, by virtue of the well known estimate for the divisor function and our bounds for K_1 and K_2. Since K_1K_2 is independent of our original vector $\mathbf{x}$, we conclude that K only takes $O_\varepsilon(||G||^\varepsilon)$ values. Equation (5) now produces $KX = -C(y,z)$, $KY = yQ(y,z)$, and $KZ = zQ(y,z)$. We can therefore reverse the linear transformation T to obtain

$$K\mathbf{x} = (C_1(y,z), C_2(y,z), C_3(y,z)),$$

where the cubic forms C_i are obtained from $-C(y,z)$, $yQ(y,z)$ and $zQ(y,z)$ by T^{-1}.

It remains to estimate, for each of $O_\varepsilon(||G||^\varepsilon)$ values of K, how many primitive vectors $\mathbf{x}$, formed as above, lie in the sphere $|\mathbf{x}| \leqslant R$. Our principal tool in doing this is the following result.

Lemma 1. — *Let $f(y,z) \in \mathbb{Z}[y,z]$ be an irreducible cubic form. Then for any $\varepsilon > 0$ we have*

$$\#\{\mathbf{x} \in \mathbb{Z}^2 : 0 < |f(\mathbf{x})| \leqslant \kappa\} \ll_\varepsilon \kappa^{2/3+\varepsilon}$$

uniformly in f, for $\kappa \geqslant 0$.

This follows from Theorem 1C of Schmidt [9; Chapter III].

We will require a corresponding result when f is reducible, but does not have a repeated factor.

Lemma 2. — *Let $f(y,z) \in \mathbb{Z}[y,z]$ be a reducible cubic form, with no repeated factor. Then for any $\varepsilon > 0$ we have*

$$\#\{\mathbf{x} \in \mathbb{Z}^2 : 0 < |f(\mathbf{x})| \leqslant \kappa\} \ll_\varepsilon \kappa^{2/3+\varepsilon}$$

uniformly in f, for $\kappa \geqslant 0$.

This will be proved in the next section.

We may now apply Lemmas 1 and 2 to our situation. Suppose that C' and C'' are rational cubic forms, neither of which has 3 distinct linear factors. Thus, in particular, neither C' nor C'' can vanish. Then either they both contain the same repeated factor, or they may be written in the shape $C'(x,y) = x^2(ax+by)$ and $C''(x,y) = y^2(cx+$

dy) after a suitable change of variable. In the latter case the condition for the form $\lambda C' + \mu C''$ to have a repeated factor is that the discriminant

$$-27(\lambda a)^2(\mu d)^2 + 18(\lambda a)(\lambda b)(\mu c)(\mu d) + (\lambda b)^2(\mu c)^2 - 4(\lambda a)(\mu c)^3 - 4(\lambda b)^3(\mu d)$$

is non-zero. Thus we get a form with no repeated factor from at least one of

$$(\lambda, \mu) = (1,1),\ (1,-1),\ (1,2)$$

unless C or C'' vanishes, a case already excluded.

Since the forms $C(y,z)$ and $Q(y,z)$ must be coprime, there can be no repeated factor common to $C(y,z), yQ(y,z), zQ(y,z)$. The above argument then shows that there is some combination $\lambda C_i + \mu C_j$, with $\lambda, \mu \ll 1$, which has no repeated factor. We shall choose such a combination and denote it $C_0(y,z)$. We therefore need to know how many primitive pairs (y,z) have $K|C_0(y,z)$ and $|K^{-1}C_0(y,z)| \ll R$. In order to remove the factor K from the form C_0 we may follow the procedure used by Schmidt [9; Chapter III, §6]. (It should be noted that Schmidt's presentation contains the misprint '$p|k$' both in the statement of Propsition 6B, and in equation (6.3). In each case this should read $p \nmid k$.) Let K' be a positive integer, all of whose prime factors divide K, and write

$$\mathcal{N}(C_0, m) = \#\{(y,z) \in \mathbb{Z}^2 : C_0(y,z) = m,\ \text{h.c.f.}(y,z) = 1\}.$$

Schmidt's argument shows that there are at most $3^{\omega(K)}$ integral forms $C^{(i)}(y,z)$ say, each of which is equivalent over the rationals to a rational multiple of C_0, such that

$$\mathcal{N}(C_0, KK'm) \leqslant \sum_i \mathcal{N}(C^{(i)}, m)$$

for any m coprime to K.

If $C_0(y,z) = Kn$ with $0 < |n| \ll R$, we may write n as $K'm$ with m coprime to K and satisfying $0 < |m| \ll R/K'$. Now, since none of the forms $C^{(i)}$ above has a repeated factor, Lemmas 1 and 2 yield

$$\#\{(y,z) \in \mathbb{Z}^2 : C_0(y,z) = Kn,\ \text{h.c.f.}(y,z) = 1,\ 0 < |n| \ll R\} \ll_\varepsilon 3^{\omega(K)} \sum_{K'} (\frac{R}{K'})^{2/3+\varepsilon}.$$

However, since K' runs over those integers composed solely of primes which divide K, we have

$$\sum_{K'} (\frac{R}{K'})^\sigma = R^\sigma \prod_{p|K} (1 - p^{-\sigma})^{-1} \ll R^\sigma 2^{\omega(K)}$$

for any fixed $\sigma > 0$.

Since $C_0(y,z) = 0$ has $O(1)$ solutions with (y,z) primitive, these considerations therefore produce $O_\varepsilon(K^\varepsilon R^{2/3+\varepsilon})$ pairs (y,z), for each of $O_\varepsilon(||G||^\varepsilon)$ values of $K \ll ||G||^A$. Theorem 2 now follows on redefining ε.

3. The proof of Lemma 2

The argument we shall use to establish Lemma 2 is related to, but simpler than, that given by Schmidt [9; Chapter III, §§2& 5]. We observe at the outset that it suffices to establish the result for primitive points $\mathbf{x}$, since the stated form of the lemma follows trivially from the corresponding version for primitive solutions.

After a suitable change of variable we may write f in the shape $xq(x,y)$, where

$$q(x,y) = Ax^2 + Bxy + Cy^2$$

is a non-singular integral quadratic form, with $C \neq 0$. We shall write $D(\neq 0)$ for the discriminant. Without loss of generality we may assume that x and C are positive. We shall also assume throughout the proof that κ is sufficiently large. This is clearly permissable.

We now set $Q = 8\kappa^{2/3}$ and we write N for the number of primitive solutions of $0 < x|q(x,y)| \leqslant \kappa$. For any prime $p \in (Q, 2Q]$ we write $N_1(p)$ for the number of solutions for which $p|x$, and $N_2(p)$ for the remainder. Then

$$\{\pi(2Q) - \pi(Q)\}N = \sum_{Q<p\leqslant 2Q} N_1(p) + \sum_{Q<p\leqslant 2Q} N_2(p).$$

Since x is clearly at most κ the first sum on the right may be written in the form

$$\sum_{x,y} \#\{p \in (Q,2Q] : \ p|x\} \leqslant N \log \kappa,$$

the summation on the left being over pairs (x,y) counted by N. It follows that

$$\sum_{Q<p\leqslant 2Q} N_1(p) \leqslant \frac{1}{2}\{\pi(2Q) - \pi(Q)\}N,$$

and hence that

$$\{\pi(2Q) - \pi(Q)\}N \leqslant 2 \sum_{Q<p\leqslant 2Q} N_2(p).$$

We may therefore fix a prime $p \in (Q, 2Q]$ such that $N \leqslant 2N_2(p)$. For each pair (x,y) counted by $N_2(p)$ there is an integer a in the range $1 \leqslant a \leqslant p$ such that $y \equiv ax \pmod p$. Substituting $y = ax + pz$ we find that $0 < x|q_0(x,z)| \leqslant \kappa$, for an appropriate form $q_0(x,z) = A_0x^2 + B_0xz + C_0z^2$, with discriminant $D_0 = p^2D$ and with $C_0 = p^2C$. It follows that there is some value of a such that

$$N \leqslant 2N_2(p) \leqslant 2pN_0 \leqslant 32\kappa^{2/3}N_0, \tag{6}$$

where N_0 counts primitive solutions of $0 < x|q_0(x,z)| \leqslant \kappa$.

We now observe that for each value of x the range for z is given by

$$|(z + \frac{B_0}{2C_0}x)^2 - \frac{D_0x^2}{4C_0^2}| \leqslant \frac{\kappa}{xC_0}.$$

This inequality specifies at most two possible intervals in which z must lie, each having length $O(\sqrt{\kappa/xC_0})$. Since

$$C_0 = p^2C \geqslant Q^2C = 64\kappa^{4/3}C \geqslant \kappa$$

the intervals have length $O(1)$, and it follows that there are $O(1)$ values of z for each possible x.

We shall factor $q_0(x,z)$ as $C_0(z-\theta x)(z-\phi x)$ where

$$\theta = \frac{-B_0 + \sqrt{D_0}}{2C_0}, \qquad \phi = \frac{-B_0 - \sqrt{D_0}}{2C_0}.$$

It follows that

$$|(z-\theta x)(z-\phi x)| \leqslant \frac{\kappa}{xC_0} \tag{7}$$

We proceed to count those solutions for which $|z-\theta x| \leqslant |z-\phi x|$, the alternative case being treated in an exactly analogous manner. For such solutions we have

$$|z-\theta x| \leqslant |(z-\theta x)(z-\phi x)|^{1/2} \leqslant \{\frac{\kappa}{xC_0}\}^{1/2}. \tag{8}$$

We claim that one also has

$$|z-\theta x| \leqslant \frac{2\kappa}{x^2\sqrt{|D_0|}}. \tag{9}$$

This follows from (8) unless

$$\frac{2\kappa}{x^2\sqrt{|D_0|}} \leqslant \{\frac{\kappa}{xC_0}\}^{1/2},$$

as we now assume. We then have

$$\frac{1}{2}|\theta-\phi|x = \frac{|D_0|^{1/2}x}{2C_0} \geqslant \{\frac{\kappa}{xC_0}\}^{1/2},$$

whence (8) implies that

$$|z-\phi x| \geqslant |\theta-\phi|x - |z-\theta x| \geqslant |\theta-\phi|x - \{\frac{\kappa}{xC_0}\}^{1/2}$$

$$\geqslant \frac{1}{2}|\theta-\phi|x = \frac{|D_0|^{1/2}x}{2C_0}.$$

We therefore see that (9) is a consequence of (7). This completes the proof of (9).

Now suppose we have two distinct primitive solutions (x_1, z_1) and (x_2, z_2) counted by N_0, both with $|z - \theta x| \leqslant |z - \phi x|$. We assume further that $x_1 \leqslant x_2$. Then $x_1 z_2 \neq x_2 z_1$, whence

$$1 \leqslant |x_1 z_2 - x_2 z_1| = |x_1(z_2 - \theta x_2) - x_2(z_1 - \theta x_1)|.$$

It follows from (8) that

$$1 \leqslant x_1 \{\frac{\kappa}{x_2 C_0}\}^{1/2} + x_2 \{\frac{\kappa}{x_1 C_0}\}^{1/2} \leqslant 2x_2 \{\frac{\kappa}{x_1 C_0}\}^{1/2}.$$

Thus

$$x_1 \leqslant 4x_2^2 \kappa C_0^{-1} \leqslant \frac{1}{2} x_2,$$

providing that

$$x_2 \leqslant \frac{C_0}{8\kappa}. \tag{10}$$

Similarly, using (9) we find that

$$1 \leqslant 2x_2 \frac{2\kappa}{x_1^2 \sqrt{|D_0|}},$$

whence

$$x_1 \leqslant 2(x_2 \kappa)^{1/2} |D_0|^{-1/4} \leqslant \frac{1}{2} x_2,$$

providing that

$$x_2 \geqslant \frac{16\kappa}{\sqrt{|D_0|}}. \tag{11}$$

However

$$\frac{16\kappa}{\sqrt{|D_0|}} \leqslant \frac{C_0}{8\kappa}$$

since our initial choice of p and Q ensures that

$$C_0 \sqrt{|D_0|} = p^3 C \sqrt{|D|} \geqslant p^3 \geqslant Q^3 = 512\kappa^2.$$

It follows that the ranges (10) and (11) cover all possibilities for x_2, so that $x_1 \leqslant \frac{1}{2} x_2$ for any two solutions of the type under consideration. Now, since we always have $x \leqslant \kappa$, we may divide the available range into $O(\log \kappa)$ subintervals of the form $\kappa_0 < x \leqslant 2\kappa_0$, each of which can contain at most one relevent value for x. In this way we see that there can be at most $O(\log \kappa)$ possible values for x, and, as has previously been observed, there are $O(1)$ corresponding values of z for each of these. Thus $N_0 \ll \log \kappa$, and (6) yields $N \ll \kappa^{2/3} \log \kappa$, which proves the lemma.

4. Proof of Theorem 3

If there is no point $\mathbf{x} \neq \mathbf{0}$ on the curve $G(\mathbf{x}) = 0$ lying in the sphere $|\mathbf{x}| \leqslant R$, then of course $N(G, R) = 0$. Otherwise we can use such a point $\mathbf{x}_0$ as a base point to transform the curve into Weierstrass normal form. Thus there is a birational transformation, θ say, which takes G into

$$E : x_2^2 x_3 = 4x_1^3 - g_2 x_1 x_3^2 - g_3 x_3^3.$$

By abuse of notation we shall also write E for the cubic form

$$E(\mathbf{x}) = x_2^2 x_3 - (4x_1^3 - g_2 x_1 x_3^2 - g_3 x_3^3).$$

The map θ is given by formulae in which the coefficients are rational functions involving the coefficents of the original form G and the coordinates of $\mathbf{x}_0$. We may clear the denominators in these formulae so that g_2 and g_3 are integers, and so that θ takes integer vectors to integer vectors. It follows that any integer solution $\mathbf{x}$ of $G(\mathbf{x}) = 0$ leads to a rational point $P(\mathbf{x})$ on E. Moreover there will be an absolute constant A for which

$$||P(\mathbf{x})|| \leqslant A||G||^A |\mathbf{x}_0|^A |\mathbf{x}|^A,$$

where $||P|| = |(x_1, x_2, x_3)|$ for $P = (x_1, x_2, x_3)$. (Here we should recall the convention concerning the symbol A, which was introduced in §2.) Since θ is a birational transformation, points $\mathbf{x}$ on $G = 0$ which are distinct up to projective equivalence, map to points $P(\mathbf{x})$ on E, which are also distinct up to projective equivalence. It follows that

$$N(G, R) \ll N(E, A||G||^A R^A).$$

Moreover, since $g_2, g_3 \ll ||G||^A |\mathbf{x}_0|^A$, we have

$$||E|| \ll ||G||^A R^A.$$

These considerations therefore yield the following conclusion.

Lemma 3. — *To establish Theorem 3 it suffices to prove it for the case of curves in Weierstrass normal form.*

We are now in a position to use some of the standard results on elliptic curves. We begin by translating our problem into one concerning the canonical height function $h(P)$. This is related to the 'logarithmic height' via the estimate

$$h(P) = \log ||P|| + O(\log ||E||)$$

for primitive points P. This follows from the work of Zimmer [12] for example. It follows there is an absolute constant A such that $h(P) \leqslant A \log(||E||R)$ for all points

P to be counted. If the rank of E is r and $P_1, \dots, P_r$ are generators for the Mordell-Weil group, our point P may be expressed as

$$P = T + \sum_{i=1}^{r} n_i P_i,$$

where T is a rational torsion point. Then

$$h(P) = h(n_1 P_1 + \cdots + n_r P_r) = Q(n_1, \dots, n_r),$$

say, where Q is a positive definite quadratic form. Since there are at most 16 values for T, by the theorem of Mazur [6], it follows that

$$N(E, R) \ll \#\{(n_1, \dots, n_r) \in \mathbb{Z}^r : \ Q(n_1, \dots, n_r) \leqslant A \log(||E||R)\}.$$

We therefore call on the following lemma.

Lemma 4. — *Let $Q(x_1, \dots, x_r) \in \mathbb{R}[x_1, \dots, x_r]$ be a positive definite quadratic form, and suppose that $Q(\mathbf{n}) \geqslant B_0$ for every non-zero vector $\mathbf{n} \in \mathbb{Z}^r$. Then*

$$\#\{(n_1, \dots, n_r) \in \mathbb{Z}^r; \ Q(n_1, \dots, n_r) \leqslant B\} \leqslant 1 + (9B/B_0)^{r/2}.$$

The proof of this will be given at the end of this section.

At this point we shall require an admissable value for B_0. This is given by the following corollary of a result of Hindry and Silverman [4; Theorem 0.3].

Lemma 5. — *Let E be a rational elliptic curve of conductor C_E and discriminant D_E. Then there is an absolute constant A such that the canonical height on E satisfies*

$$h(P) \geqslant (\log |D_E|) \exp\{-A \frac{\log |D_E|}{\log C_E}\}$$

for every non-torsion rational point P on E.

On comparing our various estimates we may now conclude that

$$N(E, R) \ll 1 + \left\{ \frac{A \log(||E||R)}{\log |D_E|} \exp(A \frac{\log |D_E|}{\log C_E}) \right\}^{r/2}.$$

We now set $\lambda = A(\log |D_E|)/(\log C_E)$, whence our bound becomes

$$N(E, R) \ll 1 + \{\frac{A \log(||E||R)}{\log C_E} e^{\lambda} \lambda^{-1}\}^{r/2}. \tag{12}$$

We clearly have $\log |D_E| \ll \log ||E||$, so that

$$\lambda \leqslant A \frac{\log ||E||}{\log C_E}.$$

On the other hand every prime factor of C_E occurs in $6D_E$, and the exponent to which a prime may occur in C_E is absolutely bounded. It therefore follows that $\log |D_E| \gg \log C_E$, and hence that $\lambda \gg 1$.

We now observe that the function e^λ/λ is increasing with respect to λ, for $\lambda \geqslant 1$. It then follows from our bounds for λ that

$$\frac{e^\lambda}{\lambda} \leqslant \exp\{A\frac{\log ||E||}{\log C_E}\}$$

whether $\lambda \geqslant 1$ or not. On substituting into (12) we therefore obtain

$$N(E,R) \ll 1 + \{A\mu e^{A\mu}\}^{r/2},$$

where $\mu = (\log(||E||R))/(\log C_E)$. Since $A\mu \leqslant \exp(A\mu)$ this leads to the bound

$$N(E,R) \ll \exp\{Ar\mu\}.$$

We are now ready to prove Theorem 3 for the curve E. Let $\varepsilon > 0$ be given. Then, according to the Rank Hypothesis, there is a constant $C(\varepsilon)$ such that

$$r \leqslant \frac{\varepsilon}{A}(\log C_E)$$

for $C_E \geqslant C(\varepsilon)$. Thus

$$N(E,R) \ll \exp\{\varepsilon \log(||E||R)\} = (||E||R)^\varepsilon,$$

as required, for $C_E \geqslant C_\varepsilon$. In the remaining case, when $C_E \leqslant C(\varepsilon)$, the height $||E||$ of the curve E, and also the rank r and the discriminant D_E, can all be bounded in terms of ε. This follows from the fact (Shafarevich [10]) that there are only finitely many curves E for each value of C_E. The estimate (12) therefore becomes

$$N(E,R) \ll_\varepsilon (\log R)^{r(\varepsilon)/2} \ll_\varepsilon R^\varepsilon,$$

where $r(\varepsilon)$ is an upper bound for the ranks of all curves with conductor at most $C(\varepsilon)$. This completes the proof of Theorem 3.

It remains to establish Lemma 4. Since $Q(\mathbf{x})$ is positive definite there is a non-singular $r \times r$ real matrix M such that $Q(\mathbf{x}) = |M\mathbf{x}|^2$. Thus, on taking $C = \sqrt{B/B_0}$, it suffices to show that

$$\#\{\mathbf{n} \in \mathbb{Z}^r;\ |M\mathbf{n}| \leqslant C\} \leqslant 1 + (3C)^r, \tag{13}$$

under the assumption that $|M\mathbf{n}| \geqslant 1$ for every non-zero vector $\mathbf{n} \in \mathbb{Z}^r$. However we may observe that the ellipsoids

$$S(\mathbf{n}) = \{\mathbf{x} \in \mathbb{R}^r :\ |M(\mathbf{x}+\mathbf{n})| < \frac{1}{2}\}$$

are disjoint for distinct $\mathbf{n}$, and lie inside the ellipsoid

$$\{\mathbf{x} \in \mathbb{R}^r :\ |M(\mathbf{x})| \leqslant C + \frac{1}{2}\}$$

for the vectors $\mathbf{n}$ to be counted. Thus if

$$V_r = \text{Meas}\{\mathbf{x} \in \mathbb{R}^r :\ |M(\mathbf{x})| \leqslant 1\}$$

we see that

$$\frac{V_r}{2^r} \cdot \#\{\mathbf{n} \in \mathbb{Z}^r;\ |M\mathbf{n}| \leqslant C\} \leqslant V_r(C + \frac{1}{2})^r,$$

whence

$$\#\{\mathbf{n} \in \mathbb{Z}^r;\ |M\mathbf{n}| \leqslant C\} \leqslant (2C+1)^r.$$

Thus (13) follows if $C \geqslant 1$, since $2C + 1 \leqslant 3C$, while if $C < 1$ then only $\mathbf{n} = \mathbf{0}$ can be counted. This completes the proof of Lemma 4.

5. Deduction of Theorem 1

In order to cover our cubic surface with plane sections we use the following result, whose proof is a trivial application of the pigeonhole principle.

Lemma 6. — *Let $\mathbf{x} \in \mathbb{Z}^n$ lie in the sphere $|\mathbf{x}| \leqslant R$. Then there is a primitive vector $\mathbf{y} \in \mathbb{Z}^n$, for which $\mathbf{x} \cdot \mathbf{y} = 0$, and such that $|\mathbf{y}| \ll_n R^{1/(n-1)}$.*

In the case $n = 4$ each set $\{\mathbf{x} \in \mathbb{Z}^4 :\ \mathbf{x} \cdot \mathbf{y} = 0\}$ is a lattice of rank 3, and we can choose a basis $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ say such that if $\mathbf{x} = \lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \lambda_3\mathbf{x}_3$ then

$$|\lambda_1||\mathbf{x}_1| + |\lambda_2||\mathbf{x}_2| + |\lambda_3||\mathbf{x}_3| \ll |\mathbf{x}|. \tag{14}$$

This follows from Davenport [1; Lemma 5].

We therefore see that the points on $F(\mathbf{x}) = 0$ which also lie in the plane $\mathbf{x} \cdot \mathbf{y} = 0$ are in 1–1 correspondence with points on the curve $G_{\mathbf{y}}(\lambda_1, \lambda_2, \lambda_3) = 0$, where

$$G_{\mathbf{y}}(\lambda_1, \lambda_2, \lambda_3) = F(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \lambda_3\mathbf{x}_3).$$

Moreover primitive points on $F = 0$ correspond to primitive points on $G_{\mathbf{y}} = 0$, and vice-versa. Since (14) implies that $|\lambda_i| \ll R$ and $|\mathbf{x}_i| \ll R$ in all relevent cases, we deduce that

$$|(\lambda_1, \lambda_2, \lambda_3)| \ll R$$

and that

$$||G_{\mathbf{y}}|| \ll_F R^3. \tag{15}$$

These estimates can be improved somewhat, but they suffice for our purposes. Finally we observe that if $G_{\mathbf{y}}$ factors, with $(\lambda_1, \lambda_2, \lambda_3)$ being a zero of a rational linear factor, then the 'curve' $G_{\mathbf{y}} = 0$ is reducible, and includes a rational line, on which $(\lambda_1, \lambda_2, \lambda_3)$ lies. The corresponding point $\mathbf{x}$ of the surface $F = 0$ therefore lies on a rational line in the surface.

The above remarks show that there is an absolute constant A for which

$$N^{(0)}(R) \leqslant \sum_{\mathbf{y}} N(G_{\mathbf{y}}, AR),$$

where $\mathbf{y}$ runs over primitive integer vectors in the sphere $|\mathbf{y}| \leqslant AR^{1/3}$. Here we recall that points lying on rational lines are to be excluded from $N(G_{\mathbf{y}}, AR)$ when $G_{\mathbf{y}}$ is reducible. Theorems 2 and 3, taken in conjunction with the bound (15) therefore lead to the conclusion that

$$N^{(0)}(R) \ll_{\varepsilon,F} \{R^{4/3} + R^{2/3}N^{(1)}(AR^{1/3})\}R^{4\varepsilon} + \sum_{\mathbf{y}}{}^{*}\, N(G_{\mathbf{y}}, AR). \tag{16}$$

Here $N^{(1)}(P)$ is the number of non-zero integer vectors $\mathbf{y}$ in the sphere $|\mathbf{y}| \leqslant P$, for which $G_{\mathbf{y}}$ is singular, and Σ^* denotes a sum over primitive vectors for which $G_{\mathbf{y}}$ is reducible.

To estimate $N^{(1)}(P)$ we note that $G_{\mathbf{y}}$ is singular if and only if $\hat{F}(\mathbf{y}) = 0$, where $\hat{F} = 0$ is the surface dual to $F = 0$. Moreover, if $\hat{F}(\mathbf{y}) = 0$, then $\mathbf{y} = \nabla F(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{C}^4$ on the surface $F = 0$. This vector $\mathbf{x}$ will, according to Lemma 2 of Hooley [5], be a scalar multiple of $\nabla\hat{F}(\mathbf{y})$, unless $\mathbf{y}$ is a singular point of $\hat{F}$. Since the form $\hat{F}$ may be taken to have integer coefficients we conclude in the former case that $\mathbf{y} = \lambda\nabla F(\mathbf{z})$ for some primitive vector $\mathbf{z} \in \mathbb{Z}^4$ on the surface $F(\mathbf{z}) = 0$. The singular locus of $\hat{F}$ has dimension 1 at most, whence the number of integer vectors $\mathbf{y}$ in the sphere $|\mathbf{y}| \leqslant P$, for which $\nabla\hat{F}(\mathbf{y}) = \mathbf{0}$, can be at worst $O_F(P^2)$.

We proceed to investigate the alternative possibility, that $\mathbf{y} = \lambda\nabla F(\mathbf{z})$ for some primitive vector $\mathbf{z} \in \mathbb{Z}^4$. We begin by noting that $p^e|\nabla F(\mathbf{z})$ implies $p|\mathbf{z}$ if either the prime p or the exponent e is large enough, since F is non-singular. Since $\mathbf{y}$ is integral and primitive, the set of possible values of the scalar λ is finite and is determined by F. In order to bound $|\mathbf{z}|$ we observe that the function $|\nabla F(\mathbf{u})|$ is continuous on the compact set $|\mathbf{u}| = 1$, under the usual topology, and hence attains its lower bound. Since F is non-singular this lower bound must be strictly positive, whence $|\nabla F(\mathbf{u})| \gg_F 1$ for $|\mathbf{u}| = 1$. We therefore conclude in general that $|\nabla F(\mathbf{z})| \gg_F |\mathbf{z}|^2$. In the situation under consideration we have $|\nabla F(\mathbf{z})| \ll P$, whence $|\mathbf{z}| \ll_F P^{1/2}$.

We may now conclude that there are $O_F(P^2)$ possible points $\mathbf{y}$ for the case under consideration, and hence that

$$N^{(1)}(AR^{1/3}) \ll_F R^{2/3}. \tag{17}$$

It remains to consider the contribution from the sum Σ^*. Since the forms $G_{\mathbf{y}}$ are singular in this case, there will be $O_F(R^{2/3})$ possible values of $\mathbf{y}$, by (17). Suppose firstly that $G_{\mathbf{y}}$ has a factor H, say, which is not proportional to a rational form. Then if $H(\mathbf{x}) = 0$ for some $\mathbf{x} \in \mathbb{Z}^3$ we must also have $\tilde{H}(\mathbf{x}) = 0$ for any conjugate $\tilde{H}$ of H. These two equations have $O(1)$ solutions in primitive vectors $\mathbf{x}$, and this produces a total contribution $O_F(R^{2/3})$ to Σ^*.

In the alternative case $G_{\mathbf{y}}$ has a rational linear factor, so that the plane $\mathbf{x}.\mathbf{y} = 0$ must contain a rational line l, say, in the surface $F = 0$. There are at most 27 such lines, so it suffices to restrict attention to values of $\mathbf{y}$ for which the corresponding plane includes a specific line l_0, say. It is convenient to change coordinates so that l_0

is given by $x_1 = x_2 = 0$. Then the form F takes the shape

$$F(\mathbf{x}) = x_1 Q_1(x_1, \dots, x_4) + x_2 Q_2(x_1, \dots, x_4),$$

and the vector $\mathbf{y}$ takes the shape $\mathbf{y} = (\alpha, \beta, 0, 0)$, where α and β are coprime integers. Since $|\mathbf{y}| \ll R^{1/3}$ we now have $|\alpha|, |\beta| \ll R^{1/3}$. Moreover primitive vectors $\mathbf{x}$ in the plane $\mathbf{x}.\mathbf{y} = 0$ necessarily have the form $(\beta t_1, -\alpha t_1, t_2, t_3)$ with (t_1, t_2, t_3) being a primitive integer vector. Since $|\mathbf{x}| \ll R$ it follows that

$$|t_1| \ll R/\max(|\alpha|, |\beta|), \quad |t_2|, |t_3| \ll R. \tag{18}$$

Now if $F(\mathbf{x}) = 0$ then either $t_1 = 0$, in which case $\mathbf{x}$ is on the line l_0, or $Q(\mathbf{t}) = 0$, where

$$Q(\mathbf{t}) = Q(\mathbf{t}, \alpha, \beta) = \beta Q_1(\beta t_1, -\alpha t_1, t_2, t_3) - \alpha Q_2(\beta t_1, -\alpha t_1, t_2, t_3).$$

We now call on the following result on solutions of quadratic forms, see Heath-Brown [3; Theorem 3], for example.

Lemma 7. — *Let Q be a non-singular integral ternary quadratic form, and suppose that the binary form $Q(0, x_2, x_3)$ is also non-singular. Then for any integer k the equation $Q(\mathbf{x}) = 0$ has $O_\varepsilon((||Q||R)^\varepsilon)$ primitive integer solutions in the sphere $|\mathbf{x}| \leqslant R$, with $x_1 = k$.*

In our application we have $||Q|| \ll_F R$, so that there are $O_{\varepsilon,F}(R^{2\varepsilon})$ solutions t_2, t_3 for each value of t_1, providing that Q and $Q(0, t_2, t_3)$ are non-singular. If Q were singular, with two quadratic conjugate factors, then $Q(\mathbf{x}) = 0$ has $O(1)$ primitive integer solutions. On the other hand, if Q has a rational linear factor L say, then the points for which $L(\mathbf{t}) = 0$ produce a rational line on the surface $F = 0$, and hence are not counted in $N(G_\mathbf{y}, AR)$.

It remains to consider the possibility that

$$Q(0, t_2, t_3) = \beta Q_1(0, 0, t_2, t_3) - \alpha Q_2(0, 0, t_2, t_3)$$

is singular. This can happen for $O(1)$ coprime pairs of integers α, β, except in the case which $Q_1(0, 0, t_2, t_3)$ and $Q_2(0, 0, t_2, t_3)$ are both proportional to the square of the same linear form. In the latter case we find that

$$F(\mathbf{x}) = x_1(x_1 L_1(\mathbf{x}) + x_2 L_2(\mathbf{x}) + c_1 L(x_3, x_4)^2) \\ + x_2(x_1 L_3(\mathbf{x}) + x_2 L_4(\mathbf{x}) + c_2 L(x_3, x_4)^2).$$

This has a singular point when $x_1 = x_2 = L(x_3, x_4) = 0$, contradicting our original assumption.

Summarizing the above considerations, we find that the equation $Q(\mathbf{t}) = 0$ has $O_{\varepsilon,F}(R^{1+2\varepsilon}/\max(|\alpha|, |\beta|))$ primitive solutions in the region given by (18), with the possible exception of $O(1)$ pairs α, β. To cover these exceptional cases we observe that if Q is non-singular, at least one of the binary forms

$$Q(x_1, x_2, 0), \quad Q(x_1, 0, x_3), \quad Q(0, x_2, x_3), \quad Q(x_1, x_2, x_2)$$

must also be non-singular. Thus, for the exceptional pairs α, β, Lemma 7 produces $O_{\varepsilon,F}(R^{1+2\varepsilon})$ primitive solutions in the region given by (18).

We may now estimate $\Sigma^* N(G_{\mathbf{y}}, AR)$ as

$$\ll_{\varepsilon,F} R^{1+2\varepsilon} + \sum_{|\alpha|,|\beta| \ll R^{1/3}} R^{1+2\varepsilon} / \max(|\alpha|, |\beta|) \ll_{\varepsilon,F} R^{4/3+2\varepsilon}.$$

Taken in conjunction with the bounds (16) and (17), this completes the proof of Theorem 1.

6. Bounds for the rank of elliptic curves

In this section we shall sketch the proof of the estimates (2) and (3). We write our curve in the form

$$E : y^2 = x^3 - ax - b,$$

where a, b are integers, and the discriminant $D = 4a^3 - 27b^2$ is non-zero. We may assume, without loss of generality, that the equation is in global minimal form, to the extent that there is no integer $d \geqslant 2$ for which $d^4 | a$ and $d^6 | b$. We now estimate the rank of E by the familiar 2-descent process. If K is the field generated by the roots of the cubic $X^3 - aX - b$, then

$$r_E \leqslant 2\{\omega_K(|D|) + h_2(K)\} + O(1),$$

where $\omega_K(|D|)$ is the number of distinct prime ideal factors of D, and $h_2(K)$ is the 2-rank of the ideal class group of K.

Since K has degree at most 6, we see that $\omega_K(|D|) \leqslant 6\omega(|D|)$, where $\omega(|D|)$ is the number of distinct rational primes dividing $|D|$. Moreover, we have

$$\omega(|D|) = \omega(C_E) + O(1)$$

since, with the possible exceptions $p = 2$ and $p = 3$, a prime p divides D if and only if it divides C_E. We note also that

$$\omega(C_E) \ll \frac{\log C_E}{\log \log C_E} \ll \log C_E.$$

We must now consider the discriminant D_K of the field K. The curve E will have bad reduction at every prime $p \geqslant 6$ which ramifies in K. Moreover since the degree of K is at most 6, the exponent to which p occurs in D_K is absolutely bounded. It follows from these considerations that

$$\log |D_K| \ll \log C_E.$$

It is easy to bound the class number $h(K)$ of K in terms of D_K, given that the degree is at most 6. For example, an estimate of Weyl [11; page 166] gives $h(K) \ll |D_K|^{7/2}$. Since $2^{h_2(K)} \leqslant h(K)$, we conclude that

$$h_2(K) \ll \log |D_K| \ll \log C_E.$$

Taken in conjunction with our earlier estimates this last bound now shows that

$$r_E \ll \log C_E$$

as claimed in (2).

When the curve has a rational point of order 2, the field K is quadratic, or may even reduce to $\mathbb{Q}$. In this case the theory of genera shows that

$$h_2(K) \ll \omega(|D_K|) \ll \omega(C_E).$$

In this case it therefore follows that

$$r_E \ll \omega(C_E) \ll \frac{\log C_E}{\log \log C_E},$$

as claimed in (3).

7. Acknowledgement

This work was initiated while the author was visiting the Institute for Advanced Study, Princeton. The hospitality of the Institute, and its financial support, are gratefully acknowledged.

References

[1] H. Davenport, Cubic forms in 16 variables, *Proc. Roy. Soc. A,* 272 (1963), 285-303.

[2] J. Franke, Yu.I. Manin, and Yu Tschinkel, Rational points of bounded height on Fano varieties, *Invent. Math.,* 95 (1989) 421-435.

[3] D.R. Heath-Brown, The density of rational points on cubic surfaces, *Acta Arithmetica,* 272 (1997) 17-30.

[4] M. Hindry and J.H. Silverman, The canonical height and integral points on elliptic curves, *Invent. Math.,* 93 (1988), 419-450.

[5] C.Hooley, On nonary cubic forms, *J. reine angew. Math.,* 386 (1988), 32-98.

[6] B. Mazur, Modular curves and the Eisenstein ideal, *Publ. Math. IHES,* 47 (1977).

[7] J.-F. Mestre, Courbes elliptiques et formules explicites, *Séminaire de théorie des nombres, Paris 1981-82,* 179-182 Progress in Math., 38, (Birkhäuser, Boston, Mass., 1983).

[8] J. Pila, The density of integral and rational points on varieties, *Astérisque,* 228 (1995), 183-187.

[9] W.M. Schmidt, *Diophantine Approximations and Diophantine Equations,* (Springer, Berlin, 1991).

[10] I.R. Shafarevich, Algebraic number fields, *Proc. Internat. Cong. Math. Stockholm, 1982,* 163-176.

[11] H. Weyl, *Algebraic theory of numbers,* (Princeton Univ. Press, N.J., 1940).

[12] H.G. Zimmer, On the difference of the Weil height and the Néron height, *Math. Zeit.,* 147 (1976), 35-51.

Received: 3 July 1996

ROGER HEATH-BROWN, Magdalen college, Oxford • *E-mail :* math3@hermine.ox.ac.uk