Manin's Proof of the Hasse Inequality Revisited

Jasbir S. Chahal
Department of Mathematics
Brigham Young University
Provo, Utah 84602, USA

Vierde serie Deel 13 No. 2 juli 1995, pp. 219 – 232 Nieuw Archief voor Wiskunde

THEOREM 3. (Hasse 1936). *Let $q = p^m$ (the prime $p \neq 2, 3$) and $a, b \in \mathbb{F}_q$ with $\Delta = 4a^3 + 27b^2 \neq 0$. If N_q denote the number of solutions, in $\mathbb{F}_q$, of the equation*

$$Y^2 = X^3 + aX + b, \tag{22}$$

then

$$|N_q - q| \leq 2\sqrt{q}. \tag{23}$$

In order to prove Theorem 3, we first gather the necessary ingredients for the proof.

We let E denote the elliptic curve defined by (22). For the polynomial

$$\lambda(t) = t^3 + at + b$$

in $\mathbb{F}_q[t]$, we associate to E another elliptic curve E_λ, defined over $K = \mathbb{F}_q(t) =$ the function field in one variable t over $\mathbb{F}_q$, by

$$\lambda Y^2 = X^3 + aX + b. \tag{24}$$

The equation (24) of E_λ is not exactly in the standard Weierstrass form, so there is the following modification of the addition formulae (19) that is not hard to work out:

(i) If $P_j = (X_j, Y_j) \in E_\lambda(K)$, $P_1 \neq \pm P_2$; P_1, P_2 and O are all distinct, then

$$x(P_1 + P_2) = \lambda \left(\frac{Y_1 - Y_2}{X_1 - X_2} \right)^2 - (X_1 + X_2). \tag{25}$$

(ii) If $P = (X, Y) \neq O$, and $Y \neq 0$ then

$$x(2P) = \frac{(3X^2 + a)^2}{4(X^3 + aX + b)} - 2X. \tag{26}$$

Clearly, $(t, 1)$ and $-(t, 1) = (t, -1) \in E_\lambda(K)$. Further, using the properties

$$(\alpha + \beta)^q = \alpha^q + \beta^q$$

and

$$(\alpha\beta)^q = \alpha^q \beta^q$$

of the *Frobenius map*

$$K \ni \alpha \mapsto \alpha^q \in K = \mathbb{F}_q(t),$$

it is easy to see that $P_0 = (X_0, Y_0) = (t^q, (t^3 + at + b)^{(q-1)/2}) \in E_\lambda(K)$. We put

$$P_n = (X_0, Y_0) + n(t, 1), \qquad n \in \mathbb{Z}.$$

If $P_n = (X_n, Y_n) \neq O$, we will show (Lemma 2) that $X_n \neq 0$. Writing $X_n = f_n/g_n$, in the lowest form, with f_n, g_n in $\mathbb{F}_q[t]$, we get a well-defined function

$$d : \mathbb{Z} \to \{0, 1, 2, 3, \ldots\}$$

given by

$$d(n) = d_n = \begin{cases} 0 & \text{if } P_n = O \\ \deg f_n & \text{otherwise.} \end{cases}$$

The function $d(n)$ satisfies the following

BASIC IDENTITY:

$$d_{n-1} + d_{n+1} = 2d_n + 2. \tag{27}$$

The connection between the function $d(n)$ and Hasse's theorem is the following relation between d_n and N_q:

$$d_{-1} - d_0 - 1 = N_q - q. \tag{28}$$

To prove identity (28), we put X_{-1} in the lowest form. First by the addition formula (25), we have

$$X_{-1} = \frac{(t^3+at+b)[(t^3+at+b)^{(q-1)/2}+1]^2}{(t^q-t)^2} - (t^q+t)$$
$$= \frac{t^{2q+1} + \text{ a polynomial of degree at most} 2q}{(t^q-t)^2}.$$

To put the rational function X_{-1} in the lowest form f_{-1}/g_{-1}, first note that

$$t^q - t = \prod_{\alpha \in \mathbf{F}_q} (t-\alpha).$$

The only factors, over $\mathbb{F}_q$, to cancel from the denominator are either (i) $(t-\alpha)^2$ with $(\alpha^3+a\alpha+b)^{(q-1)/2} = -1$, or (ii) $t-\alpha$ with $\alpha^3+a\alpha+b=0$ (note that t^3+at+b has no repeated roots). Let

m = the number of factors of the first kind,

n = the number of factors of the second kind.

Since any factor of the first kind is coprime to a factor of the second kind,

$$d_{-1} = \deg f_{-1} = 2q+1-2m-n.$$

But $d_0 = q$, so that

$$d_{-1} - d_0 = q+1-2m-n. \tag{29}$$

Any α in $\mathbb{F}_q$ with $\alpha^3+a\alpha+b$ equal to a non-zero square in $\mathbb{F}_q$ will give two solutions of (22), whereas there is only one solution of (22) corresponding to $\alpha^3+a\alpha+b=0$. Hence

$$N_q = 2(q-m)-n$$

and (28) follows from (29).

LEMMA 1. *The function $d(n)$ is a polynomial of degree 2 in n. In fact,*

$$d(n) = n^2 - (d_{-1}-d_0-1)n + d_0.$$

PROOF. The lemma is obvious for $n=-1,0$. Suppose it is true for $n-1$ and n ($n \geq 0$). By the Basic Identity (27),

$$\begin{aligned} d_{n+1} &= 2d_n - d_{n-1} + 2 \\ &= 2[n^2 - (d_{-1}-d_0-1)n + d_0] \\ &\quad -[(n-1)^2 - (d_{-1}-d_0-1)(n-1) + d_0] + 2 \\ &= (n+1)^2 - (d_{-1}-d_0-1)(n+1) + d_0 \end{aligned}$$

proving the lemma for $n+1$. By induction, the lemma follows for all $n \geq -1$. Similarly, it holds for all $n \leq 0$. □

LEMMA 2. *If $P_n = (X_n, Y_n) \neq O$, then $X_n \neq 0$. If $X_n(t) = f_n(t)/g_n(t)$ with f_n, g_n in $\mathbb{F}_q[t]$, then deg $f_n >$ deg g_n.*

COROLLARY. If $P_n \neq O$, then $d_n > 0$.

PROOF OF LEMMA 2. To prove that the degree of the numerator of a rational function $R(t)$ in $\mathbb{F}_q(t)$ is larger than that of the denominator, we formally evaluate $R(t)$ at $t = \infty$ and show that $R(t)|_\infty = \infty$.

The Lemma is obviously true for $n = 0$ and for both n and $n+1$ if $P_{n-1} = O$. Suppose the Lemma is true for a particular $n \geq 0$ for which $P_{n-1} \neq O$. The proof for all $n \geq 0$ will follow by induction if we show the Lemma to hold for $n+1$ also. Let $P_{n+1} \neq O$ and suppose to the contrary that either $X_{n+1} = 0$, or deg $f_{n+1} \leq$ deg g_{n+1}. Then it follows from

$$Y_{n+1}^2 = \frac{X_{n+1}^3 + aX_{n+1} + b}{t^3 + at + b}$$

that

$$Y_{n+1}\Big|_\infty = 0\,. \tag{30}$$

Because $(X_{n+1}, -Y_{n+1}) + (X_n, Y_n) + (t, 1) = O$, the three points $(X_{n+1}, -Y_{n+1})$, (X_n, Y_n) and $(t, 1)$ are collinear. Therefore, comparing the slopes, we get

$$Y_{n+1} = \frac{1 - Y_n}{t - X_n}(t - X_{n+1}) - 1\,,$$

and by (30),

$$0 = Y_{n+1}\Big|_\infty = \left\{\frac{1 - Y_n}{1 - X_n/t}(1 - X_{n+1}/t) - 1\right\}\Big|_\infty\,. \tag{31}$$

By our assumption, $X_{n+1}/t\Big|_\infty = 0$. Therefore by (31),

$$\frac{1 - Y_n}{1 - X_n/t}\Big|_\infty = 1\,. \tag{32}$$

By addition formula (25), i.e.,

$$X_{n+1} = \left(\frac{1 - Y_n}{t - X_n}\right)^2 (t^3 + at + b) - t - X_n\,,$$

we get

$$\frac{X_{n+1}}{t} = \left(\frac{1 - Y_n}{1 - X_n/t}\right)^2 \left(1 + \frac{a}{t^2} + \frac{b}{t^3}\right) - 1 - \frac{X_n}{t}\,.$$

Hence by (32) and the induction hypothesis on X_n,

$$0 = \frac{X_{n+1}}{t}\Big|_\infty = \left\{\left(\frac{1 - Y_n}{1 - X_n/t}\right)^2 \left(1 + \frac{a}{t^2} + \frac{b}{t^3}\right) - 1 - \frac{X_n}{t}\right\}\Big|_\infty$$
$$= -\frac{X_n}{t}\Big|_\infty \neq 0.$$

This contradiction proves the lemma for $n \geq 0$. Induction for $n \leq 0$ is carried out similarly. □

PROOF OF THEOREM 3. The quadratic polynomial

$$d(x) = x^2 - (d_{-1} - d_0 - 1)x + d_0$$

takes only non-negative values for all $n \in \mathbb{Z}$. Hence its discriminant (which by (28) is)

$$D = (N_q - q)^2 - 4q$$

cannot be positive, because, otherwise $d(x)$ will have two real roots x_1, x_2 such that for some n in $\mathbb{Z}$,

$$n \leq x_1 < x_2 \leq n + 1.$$

Moreover, both the equalities cannot hold, because by Corollary to Lemma 2, $d(n)$ cannot be zero for two successive integers. This gives a contradiction, because $(x_1 - x_2)^2 = D \in \mathbb{Z}$. Thus

$$(N_q - q)^2 - 4q \leq 0\,,$$

which proves the estimate (23).

PROOF OF THE BASIC IDENTITY

Now we prove the Basic Identity. If one of P_{n-1}, P_n and $P_{n+1} = O$, this identity is trivial. In fact, if $P_n = O$, then $X_{n-1} = X_{n+1} = t$ and $d_n = 0$, $d_{n-1} = d_{n+1} = 1$, so there is nothing to prove. If $P_{n-1} = O$, then $(X_n, Y_n) = (t, 1)$ and by addition formula (26),

$$X_{n+1} = \frac{t^4 - 2at^2 - 8bt + a^2}{4(t^3 + at + b)}\,.$$

Clearly, $d_{n-1} = 0$, $d_n = 1$. It can be easily checked that the above expression for X_{n+1} is in the lowest form, so that $d_{n+1} = 4$ and the Basic Identity is trivial again. The last possibility can be disposed of in a similar way. So without loss of generality, assume that none of P_{n-1}, P_n, P_n is O. In the addition formula (25), we bring all the terms to a common denominator to get

$$\begin{aligned} X_{n-1} &= \frac{-(tg_n + f_n)(tg_n - f_n)^2 + (1 + Y_n)^2(t^3 + at + b)g_n^3}{g_n(tg_n - f_n)^2} \\ &= \frac{(tg_n + f_n)(tf_n + ag_n) + 2bg_n^2 + 2Y_n(t^3 + at + b)g_n^2}{(tg_n - f_n)^2} \qquad (33) \\ &= \frac{R}{(tg_n - f_n)^2}\,, \end{aligned}$$

say. Similarly,

$$\begin{aligned} X_{n+1} &= \frac{-(tg_n + f_n)(tg_n - f_n)^2 + (1 - Y_n)(t^3 + at + b)g_n^3}{g_n(tg_n - f_n)^2} \\ &= \frac{(tg_n + f_n)(tf_n + ag_n) + 2bg_n^2 - 2Y_n(t^3 + at + b)g_n^2}{(tg_n - f_n)^2} \\ &= \frac{S}{(tg_n - f_n)^2}, \end{aligned} \tag{34}$$

say. Note that by (24), $Y_n(t^3 + at + b)g_n^2$ is a polynomial, hence $R, S \in \mathbb{F}_q[t]$.

It can be checked, by multiplying these expressions for X_{n-1} and X_{n+1}, that

$$\frac{f_{n-1}f_{n+1}}{g_{n-1}g_{n+1}} = \frac{RS}{(tg_n - f_n)^4} = \frac{(tf_n - ag_n)^2 - 4bg_n(tg_n + f_n)}{(tg_n - f_n)^2}. \tag{35}$$

If we show that, up to a non-zero constant,

$$g_{n-1}g_{n+1} = (tg_n - f_n)^2, \tag{36}$$

then

$$f_{n-1}f_{n+1} = (tf_n - ag_n)^2 - 4bg_n(tg_n + f_n)$$

and by Lemma 2,

$$\begin{aligned} d_{n-1} + d_{n+1} &= \deg(f_{n-1}f_{n+1}) \\ &= \deg(t^2 f_n^2) \\ &= 2d_n + 2. \end{aligned}$$

It follows from (35) that $(tg_n - f_n)^2 \mid RS$. Write $(tg_n - f_n)^2 = R_1S_1$ with R_1, S_1 in $\mathbb{F}_q[t]$ where $R_1 \mid R$ and $S_1 \mid S$. Since

$$X_{n-1} = \frac{R}{(tg_n - f_n)^2} = \frac{R/R_1}{S_1}$$

and f_{n-1}/g_{n-1} is X_{n-1} written in the lowest form, $g_{n-1} \mid S_1$. Similarly, $g_{n+1} \mid R_1$, so that $g_{n-1}g_{n+1} \mid (tg_n - f_n)^2$. Thus equality (36) will follow if we show that

$$(tg_n - f_n)^2 \mid g_{n-1}g_{n+1}. \tag{37}$$

Suppose not. Then for an irreducible factor f of $tg_n - f_n$, its power $v_f((tg_n - f_n)^2)$ in the factorization of $(tg_n - f_n)^2$ is $> v_f(g_{n-1}g_{n+1})$. Therefore, it follows from (35) that

$$f(t) \mid (tf_n - ag_n)^2 - 4bg_n(tg_n + f_n) = T,$$

say. If we show that f divides both R and S, then $f(t)$ divides the polynomials $(1 - Y_n)(t^3 + at + b)g_n^2$ and $(1 + Y_n)(t^3 + at + b)g_n^2$. But f does not divide g_n, otherwise it would be a common factor of f_n and g_n. This proves that $f(t) \mid t^3 + at + b$.

Now suppose f divides R but not S. Since $\frac{f_{n+1}}{g_{n+1}}$ is X_{n+1} in the lowest form, it follows from (34) that

$$v_f(g_{n+1}) = v_f(tg_n - f_n)^2 > 0. \tag{38}$$

In particular,

$$v_f(f_{n+1}) = 0.$$

Hence from (35), we get

$$0 < v_f(T) = v_f(f_{n-1}) - v_f(g_{n-1}). \tag{39}$$

The polynomials f_{n-1}, g_{n-1} are coprime, so that $v_f(f_{n-1})$ and $v_f(g_{n-1})$ are both ≥ 0, and at most one of them is > 0. Thus (39) is possible only if

$$v_f(g_{n-1}) = 0. \tag{40}$$

From (38) and (40), we get

$$v_f(g_{n-1}g_{n+1}) = v_f(tg_n - f_n)^2,$$

a contradiction. Thus f divides both R and S.

On long division of T by $tg_n - f_n$, we can write

$$\begin{aligned} T = \; & -(tg_n - f_n)[tf_n^2 + (t^3 - 2at - 4b)g_n] \\ & +(t^4 - 2at^2 - 8bt + a^2)g_n^2 \end{aligned}$$

which shows that $f \mid t^4 - 2at^2 - 8bt + a^2$. This together with $f \mid t^3 + at + b$ and

$$(3t^3 - 5at - 27b)(t^3 + at + b) - (3t^2 + 4a)(t^4 - 2at^2 - 8bt + a^2) = \Delta \neq 0$$

implies that $f(t)$ divides the non-zero constant Δ. This contradiction proves (37). □