NORM FORM EQUATIONS IV: RATIONAL FUNCTIONS

R. C. MASON

§1. *Introduction.* This paper is devoted to the establishment of explicit bounds on the rational function solutions of a general class of equations in several variables. The first general result on Diophantine equations over function fields was discovered in 1930 [**1**], as an analogy on the work of Thue on number fields: it was shown that the degrees of polynomial solutions X, Y in $k[z]$ of $f(X, Y) = c$ are bound for each c: here f denotes an irreducible binary form over $k[z]$ of degree at least three, and k is an algebraically closed field of characteristic zero. The Manin-Grauert theorem [**7**] extended this conclusion to the rational function solutions X, Y in $k(z)$ of any equation in two variables over $k(z)$, provided that the curve corresponding to the equation has genus two or more: this is the analogue for function fields of the Mordell conjecture for number fields, proved by Faltings in 1983. Parsin [**6**] made the Manin-Grauert theorem effective by furnishing explicit bounds on the degrees of X and Y: this approach followed Grauert and Shafarevitch in relying heavily on algebraic geometry. In 1976 a different attack was made by Schmidt [**8**] using the theory of algebraic differential equations, first developed by Kolchin and Osgood. This method produced very good bounds for equations such as the Thue equation discussed above: for example, Schmidt provides the bound

$$\max \{\deg X, \deg Y\} \leqslant 89 \deg_z f, \tag{1}$$

for any rational function solution X, Y in $k(z)$ of $f(X, Y) = 1$, where f denotes a binary form over $k(z)$ of degree in X and Y at least five, and which factorises into distinct linear factors. Here $\deg_z f$ is the degree of f in z. In 1983 [**2**] the power of Diophantine approximation provided another approach to attack Diophantine equations over function fields. By first establishing an inequality on solutions of the unit equation in two variables, it was shown in [**3**, p. 122] that the bound (1) could be improved to $36 \deg_z f$, and the condition on the degree in X and Y could be relaxed to include quartic forms. In fact the principal purpose of this new approach using Diophantine approximation had been to provide algorithms for the effective construction of all the polynomial solutions of various equations: it was a surprise to find that not only were these algorithms extremely efficient, but also that the technique extended to include rational function solutions. A full account of this approach and its consequences is given in a recent tome [**3**].

All the results discussed thus far deal with equations in two variables. In 1984 [**4**] the result in Diophantine approximation mentioned above on the unit equation in two variables was extended to an arbitrary number of variables. This led to an effective resolution of the norm form equation

$$\mathrm{Norm}_{K/k(z)}(x) = c$$

[MATHEMATIKA, **33** (1986), 204–211]

over a $k[z]$-module: K is a finite extension of $k(z)$ and c is a non-zero element of $k(z)$; there are also some necessary assumptions made on the $k[z]$-module. The object of the present paper is to establish bounds on the solutions of the general norm form equation in a vector space V over $k(z)$: thus if we write $x = \sum_{i=1}^{n} \alpha_i x_i$, where $x_1, \ldots, x_n$ is a basis of V, then we are discussing the solutions of an equation in the n variables $\alpha_1, \ldots, \alpha_n$ in $k(z)$; we wish to establish bounds on the degrees of the rational function solutions $\alpha_1, \ldots, \alpha_n$.

The actual result to be established here refers to a rather more general situation which we now describe. Let L denote a finite extension of $k(z)$, and K a finite extension of L. We shall suppose that, if necessary, K has been enlarged so that K/L is Galois, with Galois group $\mathcal{G}$ and degree d, say. We denote by V a vector space over L contained in K, of dimension n, and we shall consider the solutions x in V of the equation

$$\mathrm{Norm}_{K/L}(x) = c, \tag{2}$$

where c is a non-zero element of L. The purpose is to bound the height (see below) of x, subject to a condition on the parameters n and d. However, some further assumptions must be made on V, as otherwise the field K could simply be enlarged, altering d but not n. Moreover, as in the case of integral solutions, we must exclude the case of degenerate vector spaces: if V were to contain some yJ, where y is non-zero and J is a subfield of K strictly containing L, then $x = yj$ is a solution of (2) with $c = \mathrm{Norm}_{K/L}(y)$ for every j in J with $\mathrm{Norm}_{J/L}(j) = 1$, and such a j may be chosen with arbitrarily large height. These two difficulties are solved by introducing another parameter p, defined as follows. Elements of the Galois group $\mathcal{G}$ induce maps from V to K, that is, elements of the K-vector space $\mathrm{Hom}_L(V, K)$. We denote by p the smallest integer such that every subset of $\mathcal{G}$ consisting of p elements generates the whole of this vector space. It is evident that p satisfies $n \leqslant p \leqslant d$, for $\mathrm{Hom}_L(V, K)$ has dimension n over K, and the elements of $\mathcal{G}$ are linearly independent when viewed in the larger space $\mathrm{Hom}_L(K, K)$. The condition on the equation (2) then takes the form $d \geqslant C(n, p)$, where C is some function of n and p which is given below explicitly (7).

Before stating our main theorem we need some terminology (see [3, Chapter 1]). Associated with the field K there is a projective curve $\mathcal{C}$, whose points correspond to valuations on K. These valuations will be written additively and are normalized to have value group $\mathbb{Z}$. For an element x in K we define the *height* by

$$H(x) = -\sum_{v \in \mathcal{C}} \min(0, v(x)).$$

If $\mathcal{F}$ is a finite subset of K we define $H(\mathcal{F}) = -\sum_{v \in \mathcal{C}} \min(0, v(f); f \in \mathcal{F})$, and for a vector space V we define $H(V)$ to be the smallest value of $H(x_1, \ldots, x_n)$ for any basis $x_1, \ldots, x_n$ of V. Valuations on the two fields K and L are related in the following way. For each valuation v on K there is a unique valuation w on L such that $v(f) = e_v w(f)$ for each f in L: we write $v \mid w$, and e_v is an integer, termed the *ramification index* of v over L; v is *ramified* over L if $e_v > 1$, and v_1, v_2 are said to be *conjugate* over L if $v_1 \mid w$ and $v_2 \mid w$ for some common valuation w on L. Finally we have the notion of the genus, g, of K/k [3, p. 10]. We may now state our theorem.

THEOREM. *Suppose that x is a solution of* (2) *in the L-vector space V as above. Then, provided $d \geqslant C(n, p)$, we have*

$$H(x) \leqslant d^3 \binom{d}{n} (H(V) + H(c) + g).$$

Here $\binom{d}{n}$ is the usual binomial coefficient, and $C(n, p)$ is an explicit function given below (7).

The principal weapon in attacking this theorem is our previous result [4] on the unit equation in several variables. This lemma then formed the basis for the effective construction of integral solutions of norm form equations, and the method has since been extended to decomposable form equations [5]. It is the version of the lemma given in the latter paper that we shall quote here. We say that a set $u_1, \ldots, u_m$ of elements of K forms an *irreducible* solution of $\sum_{i=1}^{m} u_i = 0$ if $\sum_{i \in I} u_i$ is non-zero for every proper non-empty subset I of $\{1, \ldots, m\}$.

LEMMA. *Suppose that $u_1, \ldots, u_m$ is an irreducible solution of $\sum_{i=1}^{m} u_i = 0$. Then*

$$H(u_2/u_1, \ldots, u_m/u_1) \leqslant 4^{m-1}(|\mathcal{V}| + 2g), \tag{4}$$

where $\mathcal{V}$ denotes the set of valuations v on K such that $v(u_i/u_1)$ is non-zero for some i, $2 \leqslant i \leqslant m$, and $|\mathcal{V}|$ denotes the size of $\mathcal{V}$.

§2. *Proof of Theorem.* The space $\mathrm{Hom}_L(V, K)$ of linear maps from V to K has dimension n over K, so any $n+1$ elements of $\mathcal{G}$ are linearly dependent. Hence there exists a minimal linear relation over K,

$$\sum_{\sigma \in \mathcal{S}} A_\sigma \sigma x = 0 \qquad (x \in V), \tag{5}$$

where $\mathcal{S}$ is a subset of $\mathcal{G}$ with

$$|\mathcal{S}| \leqslant n + 1. \tag{6}$$

Since (5) is minimal, each A_σ is non-zero. We wish to apply our lemma to this equation; a consideration of the various terms in the inequality (4) will yield the desired result. For the most part we shall consider a single minimal linear relation (5), but later it will be necessary to consider the totality of such. In a previous work [4], when we considered solutions in a module rather than a vector space, it was possible to apply the lemma above to (5) and to calculate a bound on $|\mathcal{V}|$ in (4) quite independent of the particular solution under consideration. This is not possible for a vector space; instead we shall construct a set $\mathcal{W}$ consisting of valuations on L, depending on x, but show that whilst each such valuation makes a contribution of at least $d - p + 1$ to the height of x, we may use repeated applications of the lemma to minimal linear relations of the form (5) to show that

$$H(x) \leqslant (1/d)(np - n - p + d)(n+1)(p-1)4^n |\mathcal{W}| + a$$

for some a independent of x. We therefore achieve a bound on $H(x)$ provided that

$$d-p+1>(1/d)(np-n-p+d)(n+1)(p-1)4^n, \tag{7}$$

or, rewriting, $d \geqslant C(n,p)$ as required. We now proceed to construct the set $\mathscr{W}$ and to establish that it satisfies the inequalities we have presaged.

The definition of p states that for every subset $\mathscr{T}$ of $\mathscr{G}$ with p elements, the elements of $\mathscr{T}$ generate $\mathrm{Hom}_L(V,K)$: hence there exists a subset $\mathscr{R}$ of $\mathscr{T}$ constituting a basis for $\mathrm{Hom}_L(V,K)$, that is, there are elements $A_{\sigma\tau\mathscr{R}}$ in K for σ in $\mathscr{G}$, τ in $\mathscr{R}$ with

$$\sigma x=\sum_{\tau\in\mathscr{R}} A_{\sigma\tau\mathscr{R}}\tau x \qquad (x\in V).$$

Each of these equations is equivalent to a system of n equations

$$\sigma x_i=\sum_{\tau\in\mathscr{R}} A_{\sigma\tau\mathscr{R}}\tau x_i \qquad (1\leqslant i\leqslant n),$$

which may be solved by Cramer's rule: hence, individually,

$$H(A_{\sigma\tau\mathscr{R}})\leqslant 2nH(V),$$

and collectively,

$$H(A_{\sigma\tau\mathscr{R}};\ \sigma\in\mathscr{G},\ \tau\in\mathscr{R})\leqslant(d+n)H(V).$$

Let $\mathscr{V}_1$ denote the set of valuations v at which $v(A_{\sigma\tau\mathscr{R}})$ is negative for some triple σ, τ, $\mathscr{R}$: the last inequality yields

$$|\mathscr{V}_1|\leqslant\binom{d}{n}(d+n)H(V). \tag{8}$$

Let us also denote by $\mathscr{V}_2$ the set of valuations v on K which are conjugate to a valuation which ramifies over L (see §1): by (7) in [3, p. 11], which relates the genera of K and L, we obtain

$$|\mathscr{V}_2|\leqslant 2(d-1)(g+d). \tag{9}$$

Finally we denote by $\mathscr{V}_3$ the set of poles and zeros of c, that is, the valuations v on K at which $v(c)$ is non-zero, so

$$|\mathscr{V}_3|\leqslant 2H(c). \tag{10}$$

It should be observed that each of the set $\mathscr{V}_1$, $\mathscr{V}_2$, $\mathscr{V}_3$ consists of complete classes of conjugate valuations. We are now in a position to define the set $\mathscr{W}$ of valuations on L: it consists of those w such that there is some valuation v on K, outside $\mathscr{V}_1\cup\mathscr{V}_2\cup\mathscr{V}_3$, such that $v\,|\,w$ and $v(x)$ is non-zero. For each w in $\mathscr{W}$ we select a valuation v_w on K with $v_w(x)$ minimal amongst those valuations $v\,|\,w$. We also define, for each w in $\mathscr{W}$, the subset of $\mathscr{G}$

$$\mathscr{A}_w=\{\sigma\in\mathscr{G}: v_w(\sigma x)>v_w(x)\},$$

and we assert that each $\mathscr{A}_w$ has fewer than p elements. For suppose not, then, by the definition of p, $\mathscr{A}_w$ has a subset $\mathscr{R}$ consisting of a basis of $\mathrm{Hom}_L(V,K)$,

and then

$$x = \sum_{\tau \in \mathscr{R}} A_{1\tau\mathscr{R}} \tau x,$$

and since $v_w(A_{1\tau\mathscr{R}}) \geqslant 0$ for each τ we obtain a contradiction. Hence $|\mathscr{A}_w| < p$ as required. Now x satisfies $\mathrm{Norm}_{K/L}(x) = c$, and thus if w lies in $\mathscr{W}$ we have $w(c) = 0$ and hence $v_w(x) < 0$. Since $v_w(x)$ is an integer we in fact obtain $v_w(x) \leqslant -1$, and so the contribution to $H(x)$ from the d distinct valuations conjugate to v_w is at least $d - |\mathscr{A}_w|$: hence

$$H(x) \geqslant (d - p + 1)|\mathscr{W}|, \tag{11}$$

which constitutes the first part of our claim concerning the set $\mathscr{W}$.

The second part of the claim asserts an upper bound on $H(x)$ in terms of $\mathscr{W}$, and in order to establish this we must apply our lemma to the linear relation (5). We fix initial attention on a particular relation, so that there is no necessity to index further our constructs. Equivalent to (5) is the system of n equations

$$\sum_{\sigma \in \mathscr{S}} A_\sigma \sigma x_i = 0 \qquad (1 \leqslant i \leqslant n).$$

Since the relation (5) is minimal over K, we may choose each A_σ to be a certain determinant of size $|\mathscr{S}| - 1$ selected from the matrix with entries σx_i, $\sigma \in \mathscr{S}$, $1 \leqslant i \leqslant n$. We deduce that, individually,

$$H(A_\sigma) \leqslant (|\mathscr{S}| - 1)H(V),$$

and collectively

$$H(A_\sigma; \sigma \in \mathscr{S}) \leqslant |\mathscr{S}| H(V).$$

The consequence is that if we denote by $\mathscr{X}$ the set of valuations v on K for which $v(A_\sigma)$ is non-zero for some σ in $\mathscr{S}$, then we obtain

$$|\mathscr{X}| \leqslant |\mathscr{S}|^2 H(V). \tag{12}$$

In order to apply our lemma to the equation (5), we need an upper bound on the size of $\mathscr{V}$, the set of valuations v on K at which $v(A_\sigma \sigma x / A_\tau \tau x)$ is non-zero for some σ, τ in $\mathscr{S}$. Let us suppose that v lies in $\mathscr{V}$, but not in any of $\mathscr{X}$, $\mathscr{V}_1$, $\mathscr{V}_2$ or $\mathscr{V}_3$, and then denote by w the valuation on L with $v \mid w$. We have $v(\sigma x) \neq v(\tau x)$ for some σ, τ in $\mathscr{S}$, and so at least one of these is non-zero: hence w lies in $\mathscr{W}$. Moreover, if w is an element of $\mathscr{W}$, then $v = v_w \mu$ for some unique μ in $\mathscr{G}$, and then $v(\sigma x) \neq v(\tau x)$ implies that at least one of $\mu\sigma$ and $\mu\tau$ lies in $\mathscr{A}_w$. Hence μ lies in $\mathscr{A}_w \mathscr{S}^{-1}$, and so each valuation in $\mathscr{W}$ contributes at most $|\mathscr{A}_w||\mathscr{S}|$ to the size of $\mathscr{V}$: consequently

$$|\mathscr{V}| \leqslant |\mathscr{A}_w||\mathscr{S}||\mathscr{W}| + |\mathscr{X}| + |\mathscr{V}_1| + |\mathscr{V}_2| + |\mathscr{V}_3|. \tag{13}$$

Applying the lemma above from [5] to the equation (5), we conclude that if x is a solution in V of $\mathrm{Norm}_{K/L}(x) = c$, then either (5) is reducible, so that there is a proper non-empty subset $\mathscr{Q}$ of $\mathscr{S}$ with $\sum_{\sigma \in \mathscr{Q}} A_\sigma \sigma x = 0$, or

$$H(A_\tau \tau x / A_\sigma \sigma x; \tau \in \mathscr{S}) \leqslant 4^{|\mathscr{S}|-1}(|\mathscr{V}| + 2g).$$

The former possibility, when $\mathscr{Q}$ exists, will be dealt with later: here we assume that x is a solution of (2) in V such that every minimal linear relation (5) is

irreducible. We conclude that if the identity in $\mathcal{G}$ lies in $\mathcal{S}$ then

$$H(\tau x/x) \leqslant 4^{|\mathcal{S}|-1}(|\mathcal{V}|+2g)+|\mathcal{S}|H(V) \qquad (\tau \in \mathcal{S}). \tag{14}$$

The inequality (14) applies whenever there is a minimal linear relation in $\mathrm{Hom}_L(V, K)$ connecting 1 and τ: let $\mathcal{P}$ denote the subset of $\mathcal{G}$ consisting of those τ for which such a relation exists. We assert that $\mathcal{G}\backslash\mathcal{P}$ has fewer than p elements: for otherwise, by the definition of p, 1 is linearly dependent in $\mathrm{Hom}_L(V, K)$ on the elements of $\mathcal{G}\backslash\mathcal{P}$, so there is some minimal linear relation connecting 1 and some of the elements of $\mathcal{G}\backslash\mathcal{P}$, a contradiction. Hence $\mathcal{G}\backslash\mathcal{P}$ has fewer than p elements, and since by (7) we have $d > 2p$, we conclude that $\mathcal{P}$ has at least p elements. By the definition of the parameter p we may, therefore, select a subset $\mathcal{R}$ of $\mathcal{P}$ which forms a basis of $\mathrm{Hom}_L(V, K)$. For the elements of $\mathcal{R}$ the inequality (14) applies, and so, using (6), (8), (9), (10), (12) and (13) above, we obtain

$$H(\tau x/x;\ \tau \in \mathcal{R}) \leqslant n(n+1)(p-1)4^n|\mathcal{W}|+b,$$

where b is independent of $\mathcal{W}$. However, since $\mathcal{R}$ forms a basis of $\mathrm{Hom}_L(V, K)$ we have, as above,

$$\sigma x = \sum_{\tau \in \mathcal{R}} A_{\sigma\tau\mathcal{R}}\tau x \qquad (x \in V),$$

and so

$$H(\sigma x/x) \leqslant 2nH(V)+H(\tau x/x;\ \tau \in \mathcal{R}) \qquad (\sigma \in \mathcal{G}). \tag{15}$$

Finally, we have the inequality

$$H(c/x^d) \leqslant \sum_{1 \neq \tau \in \mathcal{P}} H(\tau x/x)+\sum_{\tau \in \mathcal{G}\backslash\mathcal{P}} H(\tau x/x):$$

for the first sum on the right we employ the inequality (14), and for the second (15). Since

$$dH(x) = H(x^d) \leqslant H(c/x^d)+H(c),$$

we obtain the desired inequality

$$H(x) \leqslant (1/d)(d-p+np-n)(n+1)(p-1)4^n|\mathcal{W}|+a,$$

where a is independent of $\mathcal{W}$. A combination of this upper bound with the lower bound (11) yields a bound on $|\mathcal{W}|$ and so on $H(x)$, provided that (7) is satisfied. The actual bound given can be calculated readily from the estimates above, noting as below that it may be assumed that n exceeds 1 and thus also $H(V)$ is non-zero.

We have now proved our theorem, subject to the condition that x is a solution of (2) in V such that each of the minimal linear relations (5) is irreducible. In general we may argue as follows. Consider a solution x in V of (2): some of the relations (5) may be irreducible, whilst others may not. If (5) is reducible at x then we may partition $\mathcal{S}$ into subsets $\mathcal{Q}$ such that each $\sum_{\sigma \in \mathcal{Q}} A_\sigma \sigma x = 0$ is an irreducible relation. The inequality (14) obtains whenever there is an irreducible relation connecting τx and x which forms part of one of those in (5). If we now denote by $\tilde{\mathcal{P}}$ the subset of $\mathcal{G}$ consisting of those τ for which such a relation exists, then if $\tilde{\mathcal{P}}$ were to contain at least p elements

then there would be a minimal linear relation in $\mathrm{Hom}_L(V, K)$ connecting 1 and some of the elements of $\mathcal{G}\backslash\tilde{\mathcal{P}}$. However, some irreducible constituent of this relation would then involve both x and some elements τx with τ in $\mathcal{G}\backslash\tilde{\mathcal{P}}$, a contradiction. Hence $\mathcal{G}\backslash\tilde{\mathcal{P}}$ has fewer than p elements, so $\tilde{\mathcal{P}}$ contains a basis of $\mathrm{Hom}_L(V, K)$. The remaining estimates follow as before.

It is a slight convenience when computing the bound which verifies the theorem to assume that n exceeds 1, for then also $H(V)$ is non-zero. If $n=1$, then the theorem is trivial, for $x = x_1 l$ for some l in L, and $c = l^d \mathrm{Norm}_{K/L}(x_1)$, so since

$$x^d = c \prod_{\sigma \in \mathcal{G}} (x_1/\sigma x_1),$$

we have

$$H(x) \leqslant 2H(x_1) + H(c)/d.$$

This completes the proof of our theorem.

§3. *Conclusion.* We conclude with two brief remarks, the first concerning the condition that d is not less than $C(n, p)$, and the second concerning an extension of our theorem. The principal factor in the condition (7) is 4^n which in turn derives from the lemma in [4] or [5]. As discussed there, this factor may easily be replaced by some explicit quantity asymptotic to $\frac{7}{5}4^n$, a slight improvement which is immediately reflected in (7). However, it may be conjectured that the factor 4^n might be improved to some quantity $\ll n$, which would result in a considerable relaxation of the condition (7). This would result in our theorem being applicable to a wider class of norm form equations, and thus such an improvement would constitute an important advance. (*Note added in proof.* The factor 4^n has recently been improved to $\frac{1}{2}n(n-1)$. See W. D. Brownawell and D. W. Masser, vanishing sums in function fields, to appear in *Math. Proc. Camb. Phil. Soc.*)

Finally, as for the solutions in a module of norm form equations, which apply to the more general decomposable form equation

$$F(\underline{x}) = c,$$

where F factorizes as $\prod_{i=1}^{d} L_i(\underline{x})$, and each $L_i(\underline{x})$ is a linear form $\sum_{j=1}^{n} a_{ij}x_j$ with coefficients in K, it may be conjectured that the method in this paper applies to furnish a bound on the solutions $x_1, \ldots, x_n$ in K of this equation, *mutatis mutandis.*

This paper was written while I was at Gonville and Caius College, Cambridge.

References

1. B. P. Gill. An analogue for algebraic functions of the Thue-Siegel theorem. *Ann. of Math.* (2), 31 (1930), 207–218.
2. R. C. Mason. The hyperelliptic equation over function fields. *Proc. Camb. Phil. Soc.*, 93 (1983), 219–230.

3. R. C. Mason. Diophantine equations over function fields. *Recent Advances in Transcendence Theory, London Math. Soc. Lecture Notes*, 96 (Cambridge University Press), to appear.
4. R. C. Mason. Norm form equations I. *J. Number Theory.* To appear.
5. R. C. Mason. Norm form equations II: Decomposable forms. To appear.
6. A. N. Parsin. Algebraic curves over function fields. *I. Math. USSR Izvestija*, 2 (1968), 1145–1170.
7. P. Samuel. *Lectures on old and new results on algebraic curves* (Bombay, 1966).
8. W. M. Schmidt. Thue's equation over function fields. *J. Austral. Math. Soc. Ser.* A, **25** (1978), 385–422.

Dr. R. C. Mason,
Credit Suisse First Boston, Ltd.,
22 Bishopsgate,
London.

10B16: *NUMBER THEORY*; *Diophantine equations*; *Norm form equations.*

Received on the 26th of July, 1985.