MIXING AND LINEAR EQUATIONS OVER GROUPS IN POSITIVE CHARACTERISTIC

BY

D. W. MASSER

Mathematisches Institut, Universität Basel
Rheinsprung 21, 4051 Basel, Switzerland
e-mail: masser@math.unibas.ch

ABSTRACT

We prove a result on linear equations over multiplicative groups in positive characteristic. This is applied to settle a conjecture about higher order mixing properties of algebraic $\mathbf{Z}^d$-actions.

1. Introduction

The main purpose of the present article is to prove a result about equations over multiplicative groups in positive characteristic. This is a version of Conjecture 2.12 (p. 550) in some notes [S2] of Klaus Schmidt, who intended to provide a substitute for the classical result in zero characteristic. As observed there, the combination of both results suffices to solve Problem 2.11 (p. 550) on mixing properties of algebraic $\mathbf{Z}^d$-actions.

We begin by stating our result. Let G be a multiplicative abelian group. For a positive integer n we are interested in subsets of the Cartesian product G^n, and we say that such a subset Σ is broad if

(i) Σ is infinite,

(ii) for each g in G and each i $(1 \leq i \leq n)$ there are at most finitely many $(x_1, \ldots, x_n)$ in Σ with $x_i = g$,

(iii) if $n \geq 2$, then for each g in G and each i, j $(1 \leq i < j \leq n)$ there are at most finitely many $(x_1, \ldots, x_n)$ in Σ with $x_i/x_j = g$.

Received June 29, 2003

THEOREM: *Let K be a field of positive characteristic, and let G be a finitely generated subgroup of the set of non-zero elements of K. Suppose that there are $a_1, \ldots, a_n$ in K such that the equation*

$$a_1 X_1 + \cdots + a_n X_n = 1 \tag{1}$$

has a broad set of solutions $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ in G^n. Then there are $b_1, \ldots, b_n$ in K and $g_1, \ldots, g_n$ in G, with

$$g_i \neq 1 \quad (1 \leq i \leq n), \tag{2}$$

and if $n \geq 2$

$$g_i/g_j \neq 1 \quad (1 \leq i < j \leq n), \tag{3}$$

such that the equation

$$b_1 g_1^k + \cdots + b_n g_n^k = 1 \tag{4}$$

has infinitely many solutions in positive integers k.

In view of the strong hypothesis on (1), the conclusions (2), (3) and (4) may seem disappointingly weak, but we have deliberately chosen a formulation minimal for applications. As mentioned above, this suffices to solve Problem 2.11 of [S2], which we therefore state more affirmatively as

COROLLARY: *For a positive integer d let α be an algebraic $\mathbf{Z}^d$-action on a compact abelian group, and let $r \geq 2$ be an integer. If every subset S in $\mathbf{Z}^d$ of cardinality r is α-mixing, then α is r-mixing.*

We sketch here how the Theorem implies the Corollary, using Schmidt's book [S1] as a basic reference. Write $\mathcal{R}$ for the Laurent ring

$$\mathbf{Z}[u_1, \ldots, u_d, 1/u_1, \ldots, 1/u_d]$$

in d variables. As in Lemma 5.1 of [S1] (p. 36), the action α corresponds to a $\mathcal{R}$-module $\mathcal{M}$. By Theorem 27.2 of [S1] (p. 264) the mixing properties of α are determined by the mixing properties of the actions $\alpha^{\mathcal{R}/\mathcal{P}}$ corresponding to the prime ideals $\mathcal{P}$ of $\mathcal{R}$ associated with $\mathcal{M}$. Thus it suffices to prove the Corollary for $\alpha = \alpha^{\mathcal{R}/\mathcal{P}}$.

Let K be the quotient field of $\mathcal{R}/\mathcal{P}$, with group K^* of non-zero elements, and let G be the subgroup of K^* generated by $u_1, \ldots, u_d$. For $\mathbf{n} = (n_1, \ldots, n_d)$ in $\mathbf{Z}^d$ write for brevity $u(\mathbf{n}) = u_1^{n_1} \cdots u_d^{n_d}$. If α is not r-mixing as in the Corollary then there exist $a_1, \ldots, a_r$ in K, not all zero, together with a sequence $(\mathbf{n}_1^{(\ell)}, \ldots, \mathbf{n}_r^{(\ell)})$ $(\ell = 1, 2, \ldots)$ in $(\mathbf{Z}^d)^r$ such that

$$\mathbf{n}_i^{(\ell)} - \mathbf{n}_j^{(\ell)} \to \infty \quad (1 \leq i < j \leq r) \tag{5}$$

and

$$a_1 u(\mathbf{n}_1^{(\ell)}) + \cdots + a_r u(\mathbf{n}_r^{(\ell)}) = 0 \tag{6}$$

for every $\ell \geq 1$. See, for example, equation (27.4) of [S1] (p. 263).

There are now two cases. If $\mathcal{P}$ meets $\mathbf{Z}$ only trivially then K has zero characteristic. In this case (6) can be handled using a deep result due to Evertse, van der Poorten and Schlickewei as in [SW]. For example, these arguments establish Theorem 27.3 of [S1] (p. 265) and it follows that α is not mixing; i.e. not 2-mixing. Then Theorem 6.5 of [S1] (p. 47) shows that G is not torsion-free. Now an equation $u(\mathbf{n}) = 1$ for $\mathbf{n} \neq \mathbf{0} = (0, \ldots, 0)$ implies $u(k\mathbf{n}) = 1$ for every $k \geq 1$, providing a set $S_2 = \{\mathbf{n}, \mathbf{0}\}$ which is not α-mixing. Thus any S of cardinality $r \geq 2$ containing S_2 is also not α-mixing, in contradiction to the hypothesis of the Corollary.

The remaining case is when $\mathcal{P}$ meets $\mathbf{Z}$ non-trivially. Now K has positive characteristic and so the result of Evertse, van der Poorten and Schlickewei does not apply. And as far as I know there is no direct analogue in the literature for positive characteristic except in Voloch's recent paper [V2], which gives a lot of information about (1) for $n = 2$. See also Lemma 10 of Mason's book [Maso] (p. 97). Such an analogue would be complicated by the existence of the Frobenius operation, which can lead very easily to an infinite set of solutions.

Anyway, if $r \geq 3$ then (6) can be rewritten in the form (1) with $n = r - 1 \geq 2$. We could even assume $a_r = -1$. If G is not torsion-free we can argue as above in zero characteristic to get a contradiction. If G is torsion-free (so in particular infinite) the condition (5) implies that the solution set of all

$$(u(\mathbf{n}_1^{(\ell)} - \mathbf{n}_r^{(\ell)}), \ldots, u(\mathbf{n}_{r-1}^{(\ell)} - \mathbf{n}_r^{(\ell)})) \quad (\ell = 1, 2, \ldots)$$

is broad. Now the Theorem leads to $b_1, \ldots, b_n$ in K and $g_1 = u(\mathbf{m}_1), \ldots, g_n = u(\mathbf{m}_n)$ in G satisfying (4) or

$$b_1 u(k\mathbf{m}_1) + \cdots + b_n u(k\mathbf{m}_n) = 1$$

for infinitely many positive integers k. But this implies — see, for example, equation (27.5) of [S1] (p. 263) — that the set $S = \{\mathbf{m}_1, \ldots, \mathbf{m}_n, \mathbf{0}\}$ is not α-mixing, again contradicting the hypothesis of the Corollary. Note that (2) and

(3) show that S has cardinality exactly $n+1=r$. This completes our sketch of how the Theorem implies the Corollary.

Sections 2, 3 and 4 of this paper are devoted to a proof of our Theorem. In Section 2 we record some preliminary observations about pth powers in characteristic p, which we apply in Section 3 to questions of linear dependence over the field of such pth powers. The main argument is in Section 4. Then in Section 5 we sketch how our methods lead to an alternative proof of Theorem 28.1 of [S1] (p. 269), which also has applications to mixing problems; further, we comment on some possible directions for future research.

ACKNOWLEDGEMENT: I am indebted to the Erwin Schrödinger Institute for Mathematical Physics in Vienna, where the work for this article was done, for friendly and unbureaucratic hospitality. The work would not have been done at all without the continuous encouragement of Klaus Schmidt. In addition I thank him, as well as Hans Peter Schlickewei, for several conversations about mixing and linear equations. And after my return to Basel, Manfred Einsiedler and Thomas Ward were kind enough to send me their preprint [EW] from which I learned about Voloch's paper [V2]. In [EW] they develop some Newton polygon techniques and prove among other things the Corollary when $d=2$ and $r=3$.

2. On pth powers in characteristic p

Let p be a rational prime, and write $F=\mathbf{F}_p$ for the field with p elements. We need two preliminary results.

LEMMA 1: *Suppose K is finitely generated over F. Then there is a height function h from K^* to $\{0,1,2,\ldots\}$ with the following properties:*

(i) $h(xy)\leq h(x)+h(y)$,

(ii) $h(x^p)=ph(x)$,

(iii) *for any real H, there are at most finitely many x in K^* with $h(x)\leq H$.*

Proof: If $\mathbf{F}_q$ is the algebraic closure of F in K, we can write K as the function field $\mathbf{F}_q(V)$ of some absolutely irreducible projective variety V defined over $\mathbf{F}_q$. By Theorem 3 of [L1] (p. 131) we can assume that V is normal. Thus by Theorem 3 of [Sh] (p. 111) it is non-singular in codimension 1. Now x in K gives rise to $\xi=(1,x)$ in $\mathbf{P}_1(K)$, and the height $H(\xi)$ can be defined using valuations associated with subvarieties W of codimension 1 as in [L2] (p. 62), with choice of constant $c=\exp(-1)$. Defining $h(x)=\log H(\xi)$, we find that

$$h(x)=\sum_W \max\{0,-\operatorname{ord}_W x\}\deg W \tag{7}$$

summed over all W, of degree $\deg W$, defined over $\mathbf{F}_q$. Thus $h(x)$ is a nonnegative rational integer. Now properties (i) and (ii) are easy; see [L2] (p. 51). A simple example would be $K = F(t)$ for a transcendental t over F; then $V = \mathbf{P}_1$ and one finds for $x = A(t)/B(t)$ that $h(x) = \max\{\deg A, \deg B\}$ if the polynomials A, B in $F[t]$ are coprime.

Property (iii) is a bit subtler and depends on the finiteness of the ground field F; for example, that the number of polynomials in $F[t]$ with bounded degree is finite. Generally it follows from (7); the fact that there are at most finitely many W over $\mathbf{F}_q$ with bounded degree can be proved using Chow forms. See, for example, [Sh] (pp. 66–68) for the theory over algebraically closed fields. In fact the Chow form Φ_W of W can be normalized to lie over $\mathbf{F}_q$. It involves a bounded number of variables and has bounded degree; thus there are at most finitely many possibilities for Φ_W. But Φ_W determines W. Therefore, if $h(x)$ is bounded there are at most finitely many possibilities for the poles W with $\mathrm{ord}_W\, x < 0$ and then for the $\mathrm{ord}_W\, x$ themselves. By considering $1/x$ we get similar information about the zeros. Finally, according to Proposition 4 of [L1] (p. 157) the poles and zeros with multiplicities determine x up to a multiple in $\mathbf{F}_q$.

This completes the proof of the present lemma. Note that it would become false without the hypothesis of finite generation. For example, with

$$K = F(t, t', t'', \ldots) \tag{8}$$

and $t = t'^p = t''^{p^2} = \cdots$ we could deduce from (ii)

$$h(t) = ph(t') = p^2 h(t'') = \cdots.$$

As h takes only integer values it follows that

$$0 = h(t) = h(t') = h(t'') = \cdots$$

contradicting (iii).

The height function of Lemma 1 will be used only to prove the next lemma. If G is a subgroup of K^*, we write $\sqrt{G}$ for its radical in K, the group of all x in K for which there is a positive integer s with x^s in G.

Lemma 2: *Suppose K is finitely generated over F, and let G be a finitely generated subgroup of K^*. If a in K^* is such that there exists an infinite sequence $a', a'', \ldots$ in K^* with $a'^p/a, a''^p/a', \ldots$ in G, then a lies in $\sqrt{G}$.*

Proof: We use the fundamental fact, not for the last time, that G modulo its subgroup $G^{[p]}$ of pth powers g^p (g in G) is finite. This enables us to assume that the $a'^p/a, a''^p/a', \ldots$ all lie in a fixed finite set. Namely, let $\mathcal{F}$ in G be a finite set of representatives for the cosets of $G^{[p]}$ in G, and define $b_0 = a$. Now $a'^p/a = a'^p/b_0$ has the form $f_1 g_1^p$ for f_1 in $\mathcal{F}$ and g_1 in G. Define $b_1 = a'/g_1$; then $b_1^p = f_1 b_0$. Next $a''^p/b_1 = g_1 a''^p/a'$ still lies in G, so has the form $f_2 g_2^p$ for f_2 in $\mathcal{F}$ and g_2 in G. Then $b_2^p = f_2 b_1$ for $b_2 = a''/g_2$. And so on. We end up with $b_0, b_1, b_2, \ldots$ in K^* satisfying

$$b_{i+1}^p = f_{i+1} b_i \quad (i = 0, 1, 2, \ldots). \tag{9}$$

We now prove that the heights $h_i = h(b_i)$ are bounded independently of i. For (9) and Lemma 1(i),(ii) yield

$$p h_{i+1} \le \phi + h_i \quad (i = 0, 1, 2, \ldots)$$

with $\phi = \max_{f \in \mathcal{F}} h(f)$. It follows by induction that

$$h_i \le H = \max\{h_0, \phi/(p-1)\} \quad (i = 0, 1, 2, \ldots)$$

(compare [L2] p. 67). For clearly $h_0 \le H$, and if $h_i \le H$ then

$$h_{i+1} \le \phi/p + h_i/p \le \phi/p + H/p \le H$$

because $H \ge \phi/(p-1)$.

So indeed $h(b_0), h(b_1), h(b_2), \ldots$ are bounded. By Lemma 1(iii) the set $\{b_0, b_1, b_2, \ldots\}$ must be finite. So there are i, j with $0 \le i < j$ and $b_i = b_j$. From (9), b_{i+1}^p/b_i is in G and forward iteration shows that b_j^q/b_i is in G for $q = p^{j-1} > 1$. Thus b_i^{q-1} is in G and b_i is in $\sqrt{G}$. Now backward iteration from (9) shows that $b_0 = a$ is in $\sqrt{G}$. This completes the proof of the present lemma. ∎

The lemma can also be deduced quickly from Lemma 3 of [V2] (p. 197). The proof given in [V2] is based on the fact that certain relative unit groups are finitely generated.

Klaus Schmidt has shown me another proof of Lemma 3 of [V2] using standard facts about Pontryagin duality.

We have kept our own heights proof with a view to quantitative refinements (see the discussion in Section 5). For example, given a in K^* not in $\sqrt{G}$, it is possible to determine an upper bound for the length I of any finite sequence $a', a'', \ldots$ in Lemma 2. Refining the present argument would need estimates for the cardinalities in property (iii) above. But by using the Box Principle as, for example, in [Mass] (pp. 192–193) one can prove that $p^I \le (h(a) + h)^{\ell+1}$, where h

is the sum of the heights over any maximal set of ℓ multiplicatively independent elements of G. With the help of the Geometry of Numbers this can be improved to $p^I \leq (\ell+1)^{\ell+1}h(a)H$ for the product H of these heights.

As with the previous lemma, Lemma 2 also becomes false if K is not finitely generated over F; for example, $a = t, a' = t', a'' = t'', \ldots$ in (8), with trivial G.

A similar application of heights is to prove that $\sqrt{G}$ itself is finitely generated. Thus the arguments of [L2] (pp. 128, 206) yield a positive integer s such that x^s lies in G for every x in $\sqrt{G}$. From this the finite generation is immediate. One can take $s \leq (q-1)h^\ell$ or even $s \leq \ell^\ell(q-1)H$ as above.

3. Linear dependence over the field of pth powers

Given a field K of characteristic p, we write $C = K^{[p]}$ for the subfield consisting of all pth powers x^p (x in K). The main tool in the proof of our Theorem is the following result.

LEMMA 3: *Suppose K is finitely generated over F, let G be a finitely generated subgroup of K^*, and let $m \geq 2$ be an integer. Then there is a finite set $\mathcal{F}$ in K, depending only on K, G and m, with the following property. If $c_1, \ldots, c_m$ in C and $y_1, \ldots, y_m$ in G satisfy*

$$c_1y_1 + \cdots + c_my_m = 1, \tag{10}$$

then either

(a) *$c_1y_1, \ldots, c_my_m$ lie in $\mathcal{F}$,*

or

(b) *$y_1, \ldots, y_m$ are linearly dependent over C.*

Proof: Equations like (10), at least with $c_1 = \cdots = c_m = 1$, are familiar in the context of function fields in one variable — see especially [Maso], [V2] for the case $m = 2$ and [V1], [BM] for the general case in zero characteristic. There they are handled by means of logarithmic differentiation. In several variables it is known that logarithmic partial differentiation is appropriate, at least in zero characteristic — see, for example, [SS] for $\mathbf{C}(t_1, \ldots, t_e)$ and [N] for algebraic extensions of this. Here we treat algebraic extensions of $\mathbf{F}_p(t_1, \ldots, t_e)$ with a result that is independent of the pth power coefficients $c_1, \ldots, c_m$.

By Theorem 1 of [L1] (p. 53) our field K is separably generated over $F = \mathbf{F}_p$. If K is algebraic over F, the result holds trivially with $\mathcal{F} = K$ in (a). Thus we can assume that there is $e \geq 1$ together with $t_1, \ldots, t_e$ algebraically independent over F such that K is separably algebraic over $K_0 = F(t_1, \ldots, t_e)$. By Proposition 2 of [L1] (p. 186) the vector space of derivations of K over F has dimension e over K and indeed we can take

$$\partial/\partial t_1, \ldots, \partial/\partial t_e \tag{11}$$

as basis elements. We can compose these to construct higher order differential operators. I cannot find an explicit reference for the fact that these commute; but this is certainly well-known on K_0. So the Lie bracket of any two of (11), which is also a derivation, vanishes on K_0. Thus by Theorem 2 of [L1] (p. 185) it vanishes on the separable extension K, and this is the desired commutativity.

Therefore every composition of (11) has the shape

$$D = (\partial/\partial t_1)^{i_1} \cdots (\partial/\partial t_e)^{i_e} \tag{12}$$

with non-negative integers $i_1, \ldots, i_e$. These integers might not be uniquely determined by D; for example, $(\partial/\partial t_1)^p = (\partial/\partial t_1)^{p+1}$ $(= 0)$ but at any rate we can define $\mathcal{D}(i)$ as the set of operators (12) with $i_1 + \cdots + i_e < i$.

Choose any operators D_i in $\mathcal{D}(i)$ $(1 \leq i \leq m)$ and apply these to (10). Because all derivations vanish on C, we obtain

$$\sum_{j=1}^{m} (D_i y_j / y_j) c_j y_j = D_i 1 \quad (1 \leq i \leq m).$$

This can be regarded as a system of linear equations in the $c_j y_j$ $(1 \leq j \leq m)$. If the associated determinant $\Delta = \det(D_i y_j / y_j)$ is non-zero, then we can solve for the $c_j y_j$ $(1 \leq j \leq m)$ as fixed rational functions of the $D_i y_j / y_j$ $(1 \leq i, j \leq m)$. We claim that these latter expressions lie in a fixed set independent of $y_1, \ldots, y_m$ in G. As mentioned above, $G/G^{[p]}$ is finite, and so each y in G has the form fc for f in a finite set and c in $G^{[p]} \subseteq C$. Therefore $Dy/y = Df/f$ lies in a finite set as claimed. Thus if $\Delta \neq 0$ for some choice of $D_1, \ldots, D_m$ we obtain the conclusion (a) of the present lemma.

If, on the other hand, $\Delta = 0$ for all choices of $D_1, \ldots, D_m$ then all generalized Wronskians $y_1 \cdots y_m \Delta = \det(D_i y_j)$ vanish. This implies that $y_1, \ldots, y_m$ are linearly dependent over the field of differential constants of K with respect to (11). Strangely enough the only characteristic-free reference I can find for this is Proposition 6.1 of [L2] (p. 174); it is here that the above commutativity is important. Finally, by Proposition 1 of [L1] (p. 185) we know that C is this constant field. So we obtain the conclusion (b); and this completes the proof of the present lemma. ∎

The generalized Wronskians could be avoided as in [V2] by using a result of Baer to the effect that there is a single derivation ∂ whose field of constants is C.

From this result we see that C-dependence plays a key role, just as in [KS] (p. 708). The next lemma helps to fix the coefficients in a C-dependence relation, provided the relation is essentially unique.

LEMMA 4: *Suppose K is finitely generated over F, let G be a finitely generated subgroup of K^*, and let $m \geq 2$ be an integer. Then there is a finite set $\mathcal{F}$ in K, depending only on K, G and m, with the following property. If $z_0, z_1, \ldots, z_m$ in G are linearly dependent over C but $z_1, \ldots, z_m$ are not, then there is a relation*

$$c_1 z_1 + \cdots + c_m z_m = z_0 \tag{13}$$

with $c_1, \ldots, c_m$ in C and $c_1 z_1/z_0, \ldots, c_m z_m/z_0$ in $\mathcal{F}$.

Proof: There is certainly a relation (13) with $c_1, \ldots, c_m$ in C, and we apply Lemma 3 with $y_j = z_j/z_0$ $(1 \leq j \leq m)$. As $z_1, \ldots, z_m$ are linearly independent over C, conclusion (b) cannot hold. Now conclusion (a) is just what we need, and this completes the proof of the present lemma. ∎

4. Proof of Theorem

Suppose K is finitely generated over F, and let G be a finitely generated subgroup of K^*. We recall the notion of a broad subset of G^n, and for future convenience we list the following easily verified properties:

(p) any non-trivial coordinate projection of a broad set is broad,

(q) as $(x_1, \ldots, x_n)$ runs through a broad set, so does $(1/x_1, x_2/x_1, \ldots, x_n/x_1)$,

(r) if the powers $(x_1^p, \ldots, x_n^p)$ of a set Σ of elements $(x_1, \ldots, x_n)$ form a broad set, then Σ is broad,

(s) any infinite subset of a broad set is broad,

(t) if $(g_1, \ldots, g_n)$ in G^n and $(x_1, \ldots, x_n)$ runs through a broad set, then the translated elements $(g_1 x_1, \ldots, g_n x_n)$ run through a broad set.

We start with the following result, which prepares the way for a proof by induction on n.

LEMMA 5: *For $n \geq 2$ suppose there are $a_1, \ldots, a_n$ in K^* such that the equation*

$$a_1 X_1 + \cdots + a_n X_n = 1 \tag{14}$$

has a broad set of solutions in G^n. *Then either*

(aa) *there is* m *with* $1 \leq m < n$ *together with* $b_1, \ldots, b_m$ *in* K *such that the equation*

$$b_1 X_1 + \cdots + b_m X_m = 1 \tag{15}$$

has a broad set of solutions in G^m,

or

(bb) *there are* $a'_1, \ldots, a'_n$ *in* K^* *such that* a'^p_j / a_j $(1 \leq j \leq n)$ *are in* G *and the equation*

$$a'_1 X_1 + \cdots + a'_n X_n = 1 \tag{16}$$

has a broad set of solutions in G^n.

Proof: Let Σ be the broad set of solutions of (14). For $\sigma = (x_1, \ldots, x_n)$ in Σ write $r(\sigma)$ for the dimension over C of the vector space generated by $a_1 x_1, \ldots, a_n x_n$ over C. Then $1 \leq r(\sigma) \leq n$. By property (s), there is an integer r with $1 \leq r \leq n$ such that $r(\sigma) = r$ for all σ in some broad subset of Σ. Making Σ smaller if necessary, we can assume that $r(\sigma) = r$ for all σ in Σ itself.

First we claim that the case $r = n$ is impossible. To see this apply Lemma 3 to (14) with $(X_1, \ldots, X_n) = \sigma$. We take $m = n$ and $c_1 = \cdots = c_m = 1$ in Lemma 3; this means that we have to enlarge G to contain $a_1, \ldots, a_n$. The conclusion (a) says that $a_1 x_1, \ldots, a_n x_n$ lie in a finite set independent of $\sigma = (x_1, \ldots, x_n)$. So outside this set conclusion (b) holds, and this means $r \neq n$ as claimed.

If $n = 2$ then only the case $r = 1$ is left, and we will see later on that this leads to conclusion (bb) of the present lemma. Otherwise, we consider now the cases $2 \leq r \leq n - 1$. These will lead to conclusion (aa) with $m = n - r + 1 < n$.

By means of a permutation we may assume that $z_1 = a_1 x_1, \ldots, z_r = a_r x_r$ are linearly independent over C. Take any k with $r + 1 \leq k \leq n$; then we can apply Lemma 4 with $m = r$ and $z_0 = a_k x_k$, G being enlarged as above. We find relations

$$\sum_{j=1}^{r} c_{kj} a_j x_j = a_k x_k \quad (r + 1 \leq k \leq n) \tag{17}$$

with c_{kj} in C and the quotients

$$f_{kj} = c_{kj} a_j x_j / a_k x_k \quad (1 \leq j \leq r,\ r + 1 \leq k \leq n) \tag{18}$$

lying in a fixed set independent of σ. We use (17) to eliminate the $a_k x_k$ $(r+1 \leq k \leq n)$ in (14) with $(X_1, \ldots, X_n) = \sigma$. We find

$$c_1 a_1 x_1 + \cdots + c_r a_r x_r = 1 \tag{19}$$

with

$$c_j = 1 + \sum_{k=r+1}^{n} c_{kj} \quad (1 \leq j \leq r) \tag{20}$$

also in C.

Next apply Lemma 3 with $m = r$ to (19) and $y_j = a_j x_j$ $(1 \leq j \leq r)$ also in the enlarged G. Now conclusion (b) is impossible. It follows that the

$$f_j = c_j a_j x_j \quad (1 \leq j \leq r) \tag{21}$$

also lie in a finite set independent of σ.

So in (18) certain quotients x_j/x_k are fixed modulo C whereas in (21) certain x_j themselves are fixed modulo C. This is enough to get an equation (15) with many solutions. Namely, we substitute (20) into (21) and use (18) to get

$$f_j = a_j x_j + \sum_{k=r+1}^{n} f_{kj} a_k x_k \quad (1 \leq j \leq r).$$

We need just one of these relations, say with $j = 1$; then dividing by $a_1 x_1$ gives an equation

$$(f_1/a_1)X_1 - \sum_{k=r+1}^{n} (f_{k1} a_k/a_1) X_{k-r+1} = 1 \tag{22}$$

with solutions

$$(X_1, X_2, \ldots, X_{n-r+1}) = (1/x_1, x_{r+1}/x_1, \ldots, x_n/x_1)$$

in G^{n-r+1}. By properties (p) and (q) these solutions make up a broad set.

Now there are only finitely many equations (22), and so by property (s) at least one of these also has a broad set of solutions in G^{n-r+1}. This we take as (15), leading to conclusion (aa) with $m = n - r + 1$ as promised.

Finally, we examine the case $r = 1$. Now $a_1 x_1, \ldots, a_n x_n$ generate a one-dimensional space over C. By (14) with $(X_1, \ldots, X_n) = \sigma$ this space contains 1 and so must be C itself. Thus $a_1 x_1, \ldots, a_n x_n$ themselves lie in C. By property

(s) we can assume that all the σ lie in a fixed coset of $(G^{[p]})^n$ in G^n. Let $\sigma_0 = (x_{10}, \dots, x_{n0})$ be one of these; then

$$a_j x_{j0} = a_j'^p \quad (1 \le j \le n)$$

for some a_j' in K^* $(1 \le j \le n)$. In particular each $a_j'^p/a_j$ lies in G $(1 \le j \le n)$, which is part of conclusion (bb) also as promised.

To check the rest, take any other solution $\sigma = (x_1, \dots, x_n)$ of (14); then

$$a_j x_j = a_j'^p(x_j/x_{j0}) = a_j'^p x_j'^p \quad (1 \le j \le n) \tag{23}$$

for x_j' also in G $(1 \le j \le n)$, by the coset assumption above. Thus the equation (16) has solutions $(X_1, \dots, X_n) = \sigma' = (x_1', \dots, x_n')$ for each σ. As σ runs through our broad set, it follows from (23) and properties (r), (t) that σ' also runs through a broad set. This gives conclusion (bb) and completes the proof of the present lemma. ∎

We can now prove the Theorem. Clearly we may assume that K is finitely generated over F. We use induction on n, the case $n = 1$ being without content. So we assume it is proved for equations like (1) with fewer than $n \ge 2$ terms on the left-hand side, and we proceed to deal with (1) itself. In particular we can suppose $a_1, \dots, a_n$ in K^*.

We apply Lemma 5. The conclusion (aa) leads by induction to $b_1, \dots, b_m$ in K and $g_1, \dots, g_m$ in G satisfying the analogues of (2), (3) and (4). So (2), (3) and (4) themselves follow on defining $b_{m+1} = \cdots = b_n = 0$ with suitable $g_{m+1}, \dots, g_n$ (note that G must be infinite, otherwise G^n could not have any broad subset at all).

So we can suppose that conclusion (bb) holds. Now Lemma 5 may be applied to (16) instead of (14). The conclusion (aa) leads to the inductive situation; and the conclusion (bb) leads to a_j'' in K^* with $a_j''^p/a_j'$ in G $(1 \le j \le n)$ and a new equation in n variables.

Now we can simply keep on going. If we never fall back to the inductive situation, then we obtain infinite sequences as in Lemma 2. There may appear to be "one infinity too much" here, but any use of the Axiom of Choice can be avoided with the finite version of Lemma 2 mentioned earlier. Anyway, we deduce that $a_1, \dots, a_n$ all lie in the radical $\sqrt{G}$. This sort of conclusion for $n = 2$ also occurs in [EW], where it comes from [V2].

Let s be a positive integer such that x^s lies in G for every x in $\sqrt{G}$. Then for each solution $\sigma = (x_1, \dots, x_n)$ of (1) the point

$$\gamma = \gamma(\sigma) = (a_1^s x_1^s, \dots, a_n^s x_n^s)$$

lies in G^n, and it is easy to see that we can choose σ in our broad set such that $\gamma = (g_1, \ldots, g_n)$ satisfies the conditions (2) and (3). Of course we could secure much more than this.

Now there is a congruence class (mod s) containing infinitely many powers q of p; that is, there is r such that $q \equiv r \pmod{s}$ for these q. Writing $q = ks + r$ we deduce from (1)

$$1 = a_1^q x_1^q + \cdots + a_n^q x_n^q = b_1 g_1^k + \cdots + b_n g_n^k$$

with $b_j = (a_j x_j)^r$ $(1 \leq j \leq n)$; and this is none other than the desired equation (4). Therefore our Theorem is proved. ∎

5. Additional remarks

The last congruence trick was used in [KS] (p. 710) in the proof of Theorem 2.1 (p. 707), which is also Theorem 28.1 of [S1] (p. 269). However, we can use the techniques of the present paper to give a new proof of this result. We content ourselves here with a sketch that (1) of Theorem 2.1 implies (2) of Theorem 2.1.

Suppose $x_0, \ldots, x_n$ in K^* are such that

$$a_0 x_0^k + \cdots + a_n x_n^k = 0 \tag{24}$$

for some $a_0, \ldots, a_n$ in K, not all zero, and infinitely many positive integers k. We can suppose $a_0 \neq 0$ then $a_0 = -1$, and dividing by x_0^k gives an equation (14) with infinitely many solutions in G^n, where G is generated by $g_1 = x_1/x_0, \ldots, g_n = x_n/x_0$. But in fact the solutions are diagonally situated in the subgroup generated by $(g_1, \ldots, g_n)$. The arguments of Lemma 5 respect this situation. They yield as in (aa) a shorter equation to which induction can be applied, or as in (bb) new coefficients $a_1', \ldots, a_n'$ in K^* with $a_j'^p/a_j$ in G and a new equation having similar properties. In fact these quotients lie in the group G_j generated by g_j, and we can even ensure that $a_j = g_j^r a_j'^p$ for an integer r with $0 \leq r < p$ independent of j.

Now iteration of (bb) leads to equations

$$a_j = g_j^{r(i)} (a_j^{(i)})^{q(i)} \quad (i = 1, 2, \ldots)$$

for $r(i)$ in $\mathbf{Z}$ with $0 \leq r(i) < q(i) = p^i$ and $a_j^{(i)}$ in K^*. By the height arguments used in the proof of Lemma 2 there are at most finitely many possibilities for the vectors $\alpha(i) = (a_1^{(i)}, \ldots, a_n^{(i)})$ $(i = 1, 2, \ldots)$. Now $\alpha(i) = \alpha(i')$ for $i \neq i'$ leads to equations

$$a_j^s = g_j^\ell \quad (1 \leq j \leq n)$$

for $s \geq 1$ and ℓ in $\mathbf{Z}$ independent of j. These equations would also come directly from a suitable analogue of Lemma 2 for G^n instead of G.

Now take $y_j = x_j^{1/s}$ in the algebraic closure $\bar{K}^*$. We can choose k in (24) such that $t = ks + \ell \geq 1$, and then $(a_j g_j^k)^s = (y_j/y_0)^{ts}$. Thus $a_j g_j^k = \lambda_j (y_j/y_0)^t$ for roots of unity λ_j in $\bar{\mathbf{F}}_p^*$. Now the equation (24) implies that $y_0^t, \ldots, y_n^t$ are linearly dependent over $\bar{\mathbf{F}}_p$ as desired in (2) of Theorem 2.1. This completes the sketch of the proof. ∎

As mentioned in the Introduction, much more can be said about the solutions of

$$(25) \qquad a_1X_1 + \cdots + a_nX_n = 1$$

in zero characteristic. Namely, it was proved by Evertse [E] and van der Poorten and Schlickewei [PS] that if $a_1, \ldots, a_n$ are non-zero then (25) has at most finitely many "non-degenerate" solutions $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ in any G^n; this means that no proper subsum of $a_1x_1, \ldots, a_nx_n$ vanishes. Such a result is certainly false in positive characteristic p; for example, the equation

$$(26) \qquad X + Y = 1$$

in the group G generated by t and $1-t$ in $\mathbf{F}_p(t)^*$ has the solutions $(t^q, (1-t)^q)$ for all powers q of p. In other words, the existence of Frobenius leads to infinitely many solutions. If $n = 2$ a Structure Theorem taking this into account can be proved (compare Voloch [V2]); but if $n > 2$ the situation is more complicated. For example, the equation

$$X + Y - Z = 1$$

has non-degenerate solutions

$$X = t^{(q-1)q'}, \quad Y = (1-t)^{qq'}, \quad Z = t^{(q-1)q'}(1-t)^{q'}$$

in G^3 for any powers q, q' of p. In general there are similar examples showing that the solution set of (25) can involve $n-1$ independent "nested Frobeniuses", mixed up with suitable "translations". And indeed it may not be too difficult to prove a corresponding Structure Theorem using the methods of this article. Thus conclusion (bb) of Lemma 5 leads to Frobenius and conclusion (aa) leads to more of them by induction.

But in zero characteristic the study of (25) can be taken even further. For example, Evertse, Schlickewei and W. M. Schmidt in [ESS] prove that there are at most $\exp\{(6n)^{3n}(nr+1)\}$ non-degenerate solutions. Here it is not assumed that G is finitely generated, only of finite rank r. So the number of such solutions is bounded independently of the field K and the coefficients $a_1, \ldots, a_n$.

Such a uniform bound cannot hold in characteristic p. For example, the algebraic closure $\bar{\mathbf{F}}_p$ contains the group $G = \bar{\mathbf{F}}_p^*$ of rank zero. So (26) has infinitely many solutions $(X, Y) = (x, 1 - x)$ in G^2; and most of these are not related by Frobenius (which in this situation is a Galois action).

Nevertheless, it can be hoped that the methods of this paper will lead to an Effective Structure Theorem. Certainly Voloch [V2] has elegant results for $n = 2$. In the most optimistic scenario one would obtain bounds even for the solutions themselves and not just their number, a situation considered unattainable today in the case of zero characteristic, even for simple examples like $3^a + 5^b - 7^c = 1$.

References

[BM] W. D. Brownawell and D. W. Masser, *Vanishing sums in function fields,* Mathematical Proceedings of the Cambridge Philosophical Society **100** (1986), 427–434.

[EW] M. Einsiedler and T. Ward, *Asymptotic geometry of non-mixing sequences,* Ergodic Theory and Dynamical Systems **23** (2003), 75–85.

[E] J.-H. Evertse, *On sums of S-units and linear recurrences,* Compositio Mathematica **53** (1984), 225–244.

[ESS] J.-H. Evertse, H. P. Schlickewei and W. M. Schmidt, *Linear equations in variables which lie in a multiplicative group,* Annals of Mathematics **155** (2002), 807–836.

[KS] B. Kitchens and K. Schmidt, *Mixing sets and relative entropies for higher-dimensional Markov shifts,* Ergodic Theory and Dynamical Systems **13** (1993), 705–735.

[L1] S. Lang, *Introduction to Algebraic Geometry,* Addison-Wesley, Redwood City, CA, 1973.

[L2] S. Lang, *Fundamentals of Diophantine Geometry,* Springer, Berlin, 1983.

[Maso] R. C. Mason, *Diophantine Equations over Function Fields,* Cambridge University Press, 1984.

[Mass] D. W. Masser, *Multiplicative isogeny estimates,* Journal of the Australian Mathematical Society **A64** (1998), 178–194.

[N] J. Noguchi, *Nevanlinna–Cartan theory over function fields and a Diophantine equation,* Journal für die reine und angewandte Mathematik **487** (1997), 61–83.

[PS] A. J. van der Poorten and H. P. Schlickewei, *Additive relations in fields,* Journal of the Australian Mathematical Society A **51** (1991), 154–170.

[S1] K. Schmidt, *Dynamical Systems of Algebraic Origin,* Birkhäuser, Basel, 1995.

[S2] K. Schmidt, *The dynamics of algebraic* $\mathbf{Z}^d$*-actions,* in *European Congress of Mathematics,* Vol. I (Barcelona, 2000), Progress in Mathematics 201, Birkhäuser, Basel, 2001, pp. 543–553.

[SW] K. Schmidt and T. Ward, *Mixing automorphisms of compact groups and a theorem of Schlickewei,* Inventiones Mathematicae **111** (1993), 69–76.

[Sh] I. R. Shafarevich, *Basic Algebraic Geometry,* Springer, Berlin, 1974.

[SS] H. N. Shapiro and G. H. Sparer, *Extension of a theorem of Mason,* Communications on Pure and Applied Mathematics **47** (1994), 711–718.

[V1] J. F. Voloch, *Diagonal equations over function fields,* Boletim da Sociedade Brasileira de Matemática **16** (1985), 29–39.

[V2] J. F. Voloch, *The equation* $ax + by = 1$ *in characteristic* p*,* Journal of Number Theory **73** (1998), 195–200.