Proc. London Math. Soc. (3) 104 (2012) 1045–1083
doi:10.1112/plms/pdr040

Linear equations over multiplicative groups, recurrences, and mixing I

H. Derksen and D. Masser

Abstract

Let K be a field of positive characteristic. When V is a linear variety in K^n and G is a finitely generated subgroup of K^*, we show how to compute the set $V \cap G^n$ effectively using heights. We calculate all the estimates explicitly. A special case provides the effective solution of the S-unit equation in n variables.

1. *Introduction*

In 2004, Masser [**22**] published a paper about linear equations over multiplicative groups in positive characteristic. This was specifically aimed at an application to a problem about mixing for dynamical systems of algebraic origin, and, as a result about linear equations, it lacked some of the simplicity of the classical results in zero characteristic. A new feature was the appearance of $n-1$ independently operating Frobenius maps; here, n is the number of variables.

In 2007, Derksen published a paper [**7**] about recurrences in positive characteristic. He proved an analogue of the famous Skolem-Mahler-Lech Theorem in zero characteristic. A new feature was the appearance of integer sequences involving combinations of $d-2$ powers of the characteristic; here, d is the order of the recurrence.

It turns out that these two new features are identical. In positive characteristic the vanishing of a recurrence with d terms can be regarded as a linear equation in $d-1$ variables to be solved in a multiplicative group (so in particular $n-1=d-2$). This observation will be developed in three directions.

In this paper, we give an improved version of the result of Masser [**22**] in a form more closely related to that in zero characteristic. In fact, we shall prove some quantitative versions in which all the estimates are effective and furthermore we shall make them completely explicit. This is in sharp contrast to the situation in zero characteristic, where even in very simple circumstances there are no effective upper bounds for the solutions.

In a second paper, we shall apply these results to recover the main theorem of Derksen [**7**], which we even generalize to sums of recurrences. In zero characteristic rather little is known about such sums, and indeed there is a conjecture of Cerlienco, Mignotte and Piras [**6**] to the effect that such problems are undecidable. In positive characteristic, we will establish not only the decidability but also give completely effective algorithms to solve the problem.

In a third paper, we apply our linear equations results to give an algorithm to determine the smallest order of non-mixing of any basic action associated with a given prime ideal in a Laurent polynomial ring. From [**22**], we know that the non-mixing comes from the so-called non-mixing sets, and our work even provides a way of finding these. Again the algorithms are completely effective.

Received 9 October 2010; revised 12 August 2011; published online 29 January 2012.

2010 *Mathematics Subject Classification* 11D04, 11D72, 11G35, 11G50 (primary), 14G25 (secondary).

This research was partially supported by the NSF, grant DMS 0901298.

We begin by recalling the classical result for a linear equation in zero characteristic, for convenience in homogeneous form. For a field K, we write K^* for the multiplicative group of all non-zero elements of K. For any subgroup G of K^* and a positive integer n it makes sense to write $\mathbf{P}_n(G)$ for the set of points in projective space defined over G.

THEOREM A (Evertse [**8**], van der Poorten-Schlickewei [**25**]). *Let K be a field of zero characteristic, and for $n \geqslant 2$ let $a_0, \ldots, a_n$ be non-zero elements of K. Then for any finitely generated subgroup G of K^* the equation*

$$a_0X_0 + a_1X_1 + \ldots + a_nX_n = 0 \tag{1.1}$$

has only finitely many solutions $(X_0, X_1, \ldots, X_n)$ in $\boldsymbol{P}_n(G)$ which satisfy

$$\sum_{i \in I} a_iX_i \neq 0 \tag{1.2}$$

for every non-empty proper subset I of $\{0, 1, \ldots, n\}$.

We should point out that this remains true even when G is not finitely generated but has finite $\mathbf{Q}$-dimension. See also a recent paper of Evertse and Zannier [**10**] for an interesting function field version.

Theorem A is false in positive characteristic p; for example, in inhomogeneous form for $n = 2$ the equation

$$x + y = 1 \tag{1.3}$$

has a solution $x = t,\ y = 1 - t$ over the group G in $K = \mathbf{F}_p(t)$ generated by $t, 1 - t$; and so thanks to Frobenius infinitely many solutions

$$x = t^{p^e}, \quad y = 1 - t^{p^e} = (1 - t)^{p^e},\ \ e = 0, 1, 2, \ldots \tag{1.4}$$

which all satisfy (1.2).

We can regard Theorem A as a descent step from the hyperplane H defined by equation (1.1) to proper linear subvarieties defined by the vanishing of the left-hand sides in (1.2). We can iterate this descent by introducing special varieties T defined solely by binary equations of the shape $X_i = aX_j$ $(i \neq j, a \neq 0)$. For example, T could be a single point or, when there are no equations at all, the full $\mathbf{P}_n$. We could call such varieties linear cosets or just cosets. This word has a group-theoretical connotation, and indeed T above is a translate of a group subvariety of the multiplicative group $\mathbf{G}_{\mathrm{m}}^n$ in $\mathbf{P}_n$. Conversely, it is not difficult to see that every linear translate of a group subvariety of $\mathbf{G}_{\mathrm{m}}^n$ is a coset in our sense (see, for example, [**4**, Lemma 9.4, p. 76]). But we will in this paper make no use of these remarks or indeed hardly any further reference to group varieties.

Anyway, it is easily seen that the complete descent yields a finite collection of cosets T, each contained in the original H, such that the full solution set $H(G) = H \cap \mathbf{P}_n(G)$ coincides with the union of all $T(G) = T \cap \mathbf{P}_n(G)$. This is a little closer to the more general context of Mordell–Lang (see below). No further descent from $T(G)$ in terms of proper subvarieties is possible; by way of compensation it is very simple to describe $T(G)$ explicitly (see, for example, the discussion towards the end of Section 12).

In positive characteristic, we can establish a descent step similar to Theorem A, but it may involve Frobenius as in (1.4). This less simple situation makes the iteration more problematic, and for this reason it is clearer to present our result as a descent now from an arbitrary linear variety V to proper linear subvarieties.

However, the Frobenius does not always generate infinitely many solutions. It does above for $x+y=1$, and also for

$$t^m x + y = 1 \tag{1.5}$$

by taking a new variable $t^m x$; this is because t lies in G. The situation is slightly more subtle for (1.5) over the group G_l generated by t^l and $1-t$; the above solution of (1.3) certainly leads to solutions

$$x = t^{-m} t^{p^e}, \quad y = (1-t)^{p^e}, \ \ e = 0,1,2,\dots, \tag{1.6}$$

but these will not be over G_l unless $p^e \equiv m \mod l$. This can however happen for infinitely many e but not necessarily all e in (1.6). This time t may not lie in G_l but some positive power does. Finally, the equation $(1+t)x+y=1$ has a solution $x=1-t, y=t^2$ over G, but the use of Frobenius will bring in an extra $1+t$, no positive power of which is in G (provided $p \neq 2$).

These considerations lead naturally to the radical $\sqrt{G} = \sqrt[K]{G}$ for general G in general K^*. For us this remains in K; thus, it is the set of γ in K for which there exists a positive integer s such that γ^s lies in G. Usually, K will be finitely generated over its prime field, and then it is well known that the finite generation of G is equivalent to that of $\sqrt{G}$. We also see the need for some concept of isotriviality, already present in diophantine geometry at least since Néron's 1952 proof of the relative Mordell–Weil Theorem and Manin's 1963 proof of the relative Mordell Conjecture. In our linear context, the appropriate refinement is G-isotriviality, introduced by Voloch [**29**] for $n=2$.

Namely, let K be a field of positive characteristic p, and for $n \geqslant 2$ let V be a linear variety in $\mathbf{P}_n$ defined over K. We say that V is G-isotrivial if there is an automorphism ψ of $\mathbf{P}_n(K)$, defined by

$$\psi(X_0,\dots,X_n) = (g_0X_0,\dots,g_nX_n) \tag{1.7}$$

with $g_0,\dots,g_n$ in G, such that $\psi(V)$ is defined over the algebraic closure $\overline{\mathbf{F}_p}$. Such a ψ could be called a G-automorphism. Let us write $\mathbf{F}_K$ for $\overline{\mathbf{F}_p} \cap K$; then of course $\psi(V)$ is defined over $\mathbf{F}_K$. So $\psi(V)$ is defined over some $\mathbf{F}_q$; and now a point w on V defined over G gives $\psi(w)$ on $\psi(V)$ which by Frobenius leads to points $\psi(w)^{q^e}$ $(e=0,1,2,\dots)$ on $\psi(V)$ and so

$$\psi^{-1}(\psi(w)^{q^e}), \quad e = 0,1,2,\dots \tag{1.8}$$

on V, all still defined over G.

Of course points over G are nothing other than zero-dimensional G-isotrivial varieties.

Here is a preliminary version of our main descent step on linear equations. For V as above write $V(G) = V \cap \mathbf{P}_n(G)$ for the set of points of V defined over G. But it is clearer first to consider points over the radical $\sqrt{G}$.

Descent Step over $\sqrt{G}$. Let K be a field of positive characteristic, and suppose that the positive-dimensional linear variety V_0 defined over K is not a coset. Suppose also that $\sqrt{G}$ in K is finitely generated. Then there is an effectively computable finite collection $\mathcal{W}$ of proper $\sqrt{G}$-isotrivial linear subvarieties W of V_0, also defined over K, with the following property.

(a) If V_0 is not $\sqrt{G}$-isotrivial, then

$$V_0(\sqrt{G}) = \bigcup_{W \in \mathcal{W}} W(\sqrt{G}).$$

(b) If V_0 is $\sqrt{G}$-isotrivial and $\psi(V_0)$ is defined over $\mathbf{F}_q$, then

$$V_0(\sqrt{G}) = \psi^{-1}\left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(\sqrt{G}))^{q^e}\right).$$

Thus (a) says that the points of $V_0(\sqrt{G})$ are not Zariski-dense in V_0; and (b) says that the points on $V_0(\sqrt{G})$ such as (1.8), which can be dense, at least arise from a set of w which is not dense.

Part (a) was essentially proved for $n = 2$ as Theorem 1 by Voloch [**29**, p. 196], and his Theorem 2 (p. 198) even covers the more general case of finite $\mathbf{Q}$-dimension; here, one obtains the finiteness of the solution set. A forerunner of part (b) for $n = 2$ can be seen in Mason [**21**, pp. 107, 108]. The main result of Masser [**22**] is restricted to a single equation (1.1) and is expressed in terms of a concept of 'broad' set; as we do not need this result here (or even the concept) we refrain from quoting it. However, these authors do not discuss the effectivity in our sense (see the discussion below).

A simple example of (b) in inhomogeneous form is (1.3); this represents a line L, clearly isotrivial and even trivial in that we can take ψ as the identity automorphism. When G is generated by t and $1-t$ in $K = \mathbf{F}_p(t)$, then $\sqrt{G}$ is obtained by adding the elements of $\mathbf{F}_p^*$ as generators. Leitner [**20**] has found that for $p \geqslant 3$ there are $p+4$ points W, six of which are like $w = (t, 1-t)$ in (1.4) and the remaining $p-2$ are the $w = (x, 1-x)$ for $x = 2, 3, \ldots, p-1$.

So much for $V_0(\sqrt{G})$. In the analogous characterization of $V_0(G)$ there is no longer a clear separation of cases. In fact it can happen in case (b) above that the actions of Frobenius through q^e can get truncated, so that each e remains bounded; but then it is easy to reduce this to case (a). A simple example is (1.5) for $m = 1$ in the group $G = G_l$ above for $l = p$, when the solutions (1.6) are over G only when $e = 0$. Here is a general statement.

Descent Step over $\sqrt{G}$. Let K be a field of positive characteristic, and suppose that the positive-dimensional linear variety V_0 defined over K is not a coset. Suppose also that $\sqrt{G}$ in K is finitely generated. Then there is an effectively computable finite collection $\mathcal{W}$ of proper $\sqrt{G}$-isotrivial linear subvarieties W of V_0, also defined over K, such that either

$$V_0(G) = \bigcup_{W \in \mathcal{W}} W(G)$$

or

$$V_0(G) = \psi^{-1}\left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(G))^{q^e}\right)$$

for some q and some $\sqrt{G}$-automorphism ψ with $\psi(V_0)$ defined over $\mathbf{F}_q$.

It may be instructive here to consider the inhomogeneous example

$$x + y - z = 1 \tag{1.9}$$

still over the group G in $K = \mathbf{F}_p(t)$ generated by $t, 1-t$. Now (1.9) represents a plane P, also isotrivial and even trivial. Leitner [**20**] has found that for $p \geqslant 5$ there are 22 lines W and 8 points W. For example, the line defined by

$$tx + y = 1, \quad z = (1-t)x \tag{1.10}$$

is one of these. So is the coset line defined by $x = z,\ y = 1$. And so is the point

$$x = t, \quad y = \frac{1-t}{t}, \quad z = \frac{(1-t)^2}{t}.$$

We can easily iterate the descent from (1.10). This is isotrivial via the automorphism ψ taking x, y and z to $\tilde{x} = tx, \tilde{y} = y$ and $\tilde{z} = (t/(1-t))z$, when the equations become $\tilde{x} + \tilde{y} = 1$ and $\tilde{z} = \tilde{x}$. Now (1.4) (with e replaced by f) on (1.3) lead to the points $w = (x, y, z)$ of $W(G)$ with

$$x = t^{p^f - 1}, \quad y = (1-t)^{p^f}, \quad z = t^{p^f - 1}(1-t), \;\; f = 0, 1, 2, \ldots .$$

Then from (1.8) (with $q = p$ and the identity automorphism) we obtain the points

$$x = t^{(q-1)r}, \quad y = (1-t)^{qr}, \quad z = t^{(q-1)r}(1-t)^r \tag{1.11}$$

of $P(G)$; here, $q = p^f$ and $r = p^e$ now indicate independently varying powers of p. This is precisely the example in [**22**, p. 202].

With the help of a suitable notation we can after all do the complete descent, also for linear varieties that are cosets; then the latter arise solely as obstacles. Denote by $\varphi = \varphi_q$ the Frobenius with $\varphi(x) = x^q$. Let $\psi_1, \ldots, \psi_h$ be projective automorphisms. Then we imitate commutator brackets by defining the operator

$$[\psi_1, \ldots, \psi_h] = [\psi_1, \ldots, \psi_h]_q = \bigcup_{e_1=0}^{\infty} \cdots \bigcup_{e_h=0}^{\infty} (\psi_1^{-1}\varphi^{e_1}\psi_1) \ldots (\psi_h^{-1}\varphi^{e_h}\psi_h), \tag{1.12}$$

with of course the identity interpretation if $h = 0$. This formally resembles [**7**, Definition 7.7, p. 208].

Theorem 1. *Let K be a field of positive characteristic p, let V be an arbitrary linear variety defined over K, and suppose that $\sqrt{G}$ in K is finitely generated. Then there is a power q of p such that $V(G)$ is an effectively computable finite union of sets $[\psi_1, \ldots, \psi_h]_q T(G)$ with $\sqrt{G}$-automorphisms $\psi_1, \ldots, \psi_h$ $(0 \leqslant h \leqslant n-1)$, and cosets T contained in V.*

Here, we see quite clearly the $n-1$ Frobenius operators mentioned in the first paragraph of Section 1. In general, they act independently because they are separated by automorphisms. The example

$$x_1 + x_2 - x_3 - \ldots - x_n = 1$$

generalizes (1.3) and (1.9), and it can be used to show that the upper bound $n-1$ in Theorem 1 cannot always be improved. This we carry out in Section 13 on limitation results. The same can also be seen indirectly through the applications to recurrences, where we will see that the analogous upper bound $d-2$ cannot always be improved.

Taking $e_1 = 1$ in (1.12) and all other zero, we see that ψ_1^{q-1} is a G-automorphism. Similarly for $\psi_1^{q-1}, \ldots, \psi_h^{q-1}$. However, it may not always be possible to choose $\psi_1, \ldots, \psi_h$ as G-automorphisms. This we also prove in Section 13.

We can also symmetrize the sets in Theorem 1. We explain this with the points (1.11) on P defined by (1.9). They can be written as

$$x = t^{s-r}, \quad y = (1-t)^s, \quad z = t^{s-r}(1-t)^r \tag{1.13}$$

with $s = qr$. Here, there is asymmetry because apparently r divides s. However (1.13) has a meaning for any independent positive powers r, s of p; and it is easily checked that the resulting points remain on P.

To formulate this in general we introduce another bracket notation more related to the group law. For points $\pi_0, \pi_1, \ldots, \pi_h$, we define the set

$$(\pi_0, \pi_1, \ldots, \pi_h) = (\pi_0, \pi_1, \ldots, \pi_h)_q = \pi_0 \bigcup_{l_1=0}^{\infty} \cdots \bigcup_{l_h=0}^{\infty} (\varphi^{l_1}\pi_1) \ldots (\varphi^{l_h}\pi_h), \tag{1.14}$$

with of course the interpretation π_0 itself if $h = 0$. We introduce more special varieties S defined solely by binary equations of the shape $X_i = X_j$. For example, S could be the single point with all coordinates equal or the full $\mathbf{P}_n$. We could call such varieties linear subgroups or just subgroups. As above it is not difficult to see that they are precisely the linear group subvarieties of $\mathbf{G}_{\mathrm{m}}^n$, but again we do not need to know this.

Theorem 2. *Let K be a field of positive characteristic p, let V be an arbitrary linear variety defined over K, and suppose that $\sqrt{G}$ in K is finitely generated. Then there is a power*

q of p such that $V(G)$ is an effectively computable finite union of sets $(\pi_0, \pi_1, \ldots, \pi_h)_q S(G)$ with points $\pi_0, \pi_1, \ldots, \pi_h$ $(0 \leqslant h \leqslant n-1)$ defined over $\sqrt{G}$ and subgroups S.

As in Theorem 1, the upper bound $n-1$ in Theorem 2 cannot always be improved. We shall verify this in Section 13. Also one can easily see that $\pi_0^{q-1}, \pi_1^{q-1}, \ldots, \pi_h^{q-1}$ (as well as the product $\pi_0\pi_1 \ldots \pi_h$) are defined over G. However, this may not always be true of $\pi_0, \pi_1, \ldots, \pi_h$, as we shall also prove in Section 13.

When V is a hyperplane defined by (1.1) we can even descend to points, provided we restrict to (1.2) in the style of Theorem A.

THEOREM 3. *Let K be a field of positive characteristic p, let H be defined by*

$$a_0X_0 + a_1X_1 + \ldots + a_nX_n = 0$$

for non-zero $a_0, a_1, \ldots, a_n$ in K, and write $H^*(G)$ for the set of points in $\boldsymbol{P}_n(G)$ satisfying

$$\sum_{i \in I} a_iX_i \neq 0$$

for every non-empty proper subset I of $\{0, 1, \ldots, n\}$. Suppose that $\sqrt{G}$ in K is finitely generated. Then there is a power q of p such that $H^*(G)$ is contained both in (1) an effectively computable finite union of sets $[\psi_1, \ldots, \psi_h]_q\{\tau\}$ in $H(G)$ with $\sqrt{G}$-automorphisms $\psi_1, \ldots, \psi_h$ $(0 \leqslant h \leqslant n-1)$ and points τ, and in (2) an effectively computable finite union of sets $(\pi_0, \pi_1, \ldots, \pi_h)_q$ in $H(G)$ with points $\pi_0, \pi_1, \ldots, \pi_h$ $(0 \leqslant h \leqslant n-1)$.

We do not prove it here, but in this situation $H^*(G)$ is precisely a finite union of $[\psi_1, \ldots, \psi_h]_q\{\tau\}$. However, there seems to be a strange asymmetry between the asymmetric part (1) and the symmetric part (2). Namely it seems improbable that $H^*(G)$ is precisely a finite union of $(\pi_0, \pi_1, \ldots, \pi_h)_q$. For example, the point (1.13) on H defined by (1.9) is in $H^*(G)$ except for $r = s$, which disturbs the independence of r and s.

Apart from the work [**29**] already mentioned, there are other results of this kind, now in the more general context of Mordell–Lang for arbitrary varieties V inside arbitrary semiabelian varieties $\mathbf{S}$. Typically, here one intersects V with a finitely generated subgroup Γ of $\mathbf{S}$; however, in this paper with $\mathbf{S} = \mathbf{G}_{\mathrm{m}}^n$ we have for simplicity restricted Γ to a Cartesian product G^n.

Thus, the main result Theorem A (p. 104, see also p. 109) of Abramovich and Voloch [**1**] almost implies part (a) of our Descent Step over $\sqrt{G}$, except that they assume that V is not K^*-isotrivial and they have no information about W which would ensure linearity in our situation. The main result Theorem 1.1 (p. 667) of Hrushovski's well-known paper [**16**] gives a similar implication. The restriction to our (a) corresponds to their restriction to the non-isotrivial case. Again these authors do not discuss the effectivity in our sense.

After the earlier work by Scanlon, the isotrivial case was treated by Moosa and Scanlon. Their Theorem B [**24**, p. 477] implies that our $V(G)$ is what they call an F-set (see also [**23**]). Indeed in our situation and notation an F-set is nothing but a finite union of $(\pi_0, \pi_1, \ldots, \pi_h)_q A(G)$ with $\pi_0\pi_1 \ldots \pi_h$ and $\pi_0^{q-1}, \pi_1^{q-1}, \ldots, \pi_h^{q-1}$ defined over G and an algebraic subgroup A. They do not mention the bound $h \leqslant n-1$ and that A is linear when V is; however, a referee kindly pointed out that both facts follow from the arguments in their Theorem 7.8 (p. 512). Their ideas were developed by Ghioca [**11**], who in addition extended the results to Drinfeld modules. See also the work of Ghioca and Moosa [**12**] on division groups. Again there is no mention of effectivity.

Now let us discuss this effectivity, a key aspect of this paper.

It is well known that Theorem A (in zero characteristic) is semieffective in the sense that effective and even explicit upper bounds for the number of solutions of (1.1) subject to (1.2) can be found. However, it is not fully effective in the sense that no upper bounds are known for the size of the solutions, even in very simple cases such as $K = \mathbf{Q}$ and G generated by 3,5,7; and it is even unknown how to find all the finitely many non-negative integers a, b and c satisfying an equation such as

$$3^a + 5^b - 7^c = 1.$$

Out of the works in positive characteristic quoted above, only two discuss effectivity, and then only semieffectivity in the sense above. Voloch [**29**] in the theorems mentioned above gives explicit upper bounds for the cardinality of $V(G)$ for $n = 2$ in case (a) of Theorem 1; these are uniform in the sense that they are independent of V and further they depend on G only with regard to its rank. A similarly uniform bound is given as Theorem 6.1 (p. 687) by Hrushovski [**16**] for V in an abelian variety; however, as it stands it is not completely explicit due to the use of non-standard analysis. These bounds are in line with the well-known estimates in zero characteristic - see for example [**9**, Theorem 1.1, p. 808].

By contrast our results above are fully effective. This should be no surprise; for example, it is rather easy by differentiating to find all non-negative integers a, b and c with

$$(3+t)^a + (5+t)^b - (7+t)^c = 1$$

in any fixed $K = \mathbf{F}_p(t)$. We shall work out explicit bounds, at first for the Descent Step over $\sqrt{G}$, where the exponents appearing can reasonably be small; and then for the Descent Step over G and Theorems 1–3. See especially (12.1) and (12.10) later. It would then be a straightforward matter to deduce bounds for the various cardinalities involved; but more work may be needed to make these uniform in the sense above.

In fact the size bounds cannot be uniform in this sense. For example, from the non-isotrivial equation $x + ay = 1$ with $a = (1-t^m)/(1-t)^m$ $(m \neq p^e)$ over the group generated by t and $1-t$ in $\mathbf{F}_p(t)$, with solution $x = t^m,\ y = (1-t)^m$, we can easily show that the size of solutions for fixed G must depend on V. Similarly, the isotrivial equation $x + y = 1$ over the group generated by t^m and $(1-t)^m$ in $\mathbf{F}_p(t)$, with the same solution, demonstrates that the size of solutions for fixed V must depend on more than just the rank of G.

Because all our varieties are linear, we can measure them in a traditional way in terms of certain heights on the Grassmannian. We will show, for example, in the Descent Step over $\sqrt{G}$ that

$$h(W) \leqslant Ch(V_0)^{2n} \tag{1.15}$$

if W is no longer required to be $\sqrt{G}$-isotrivial, where C depends only on K, n and G. If we insist on W being $\sqrt{G}$-isotrivial, then the exponent is not so small. The well-known Northcott Property of heights often implies that the set of W in (1.15) is finite and easily effectively computable.

Perhaps since the results in zero characteristic are not effective, there is no tradition about measuring the groups Γ, even in $\mathbf{S} = \mathbf{G}_{\mathrm{m}}^n$. Because our $\Gamma = G^n$, it is here possible to use a basis-free notion of regulator $R(G)$. We will show that the bounds, at least when $G = \sqrt{G}$, are all of polynomial growth in $R(G)$. For example in (1.15) we obtain

$$C \leqslant cR(G)^{6n+2}$$

again if W is no longer required to be $\sqrt{G}$-isotrivial, where c now depends only on K, n and the rank r of G. In fact here

$$c = 8n^2 d(10n^3(n+r)^{3(n+r)})^{2n+1}$$

with d depending only mildly on K; for example, $d = 1$ if K is a field of rational functions in several independent variables over a finite field.

However, we did find it a small surprise to discover that when $G \neq \sqrt{G}$ the smallest bounds can be exponential in $R(G)$. A hint of this can be seen from the above discussion of (1.5) and G_l. For example, the simplest solution of the equation

$$t^{42}x + y = 1$$

with x and y in the group generated by t^{83} and $1 - t$ in $\mathbf{F}_2(t)$ is

$$x = (t^{83})^{2913074264131636565570}, \quad y = (1-t)^{2417851639229258349412352}; \tag{1.16}$$

while the regulator is only $83\sqrt{3}$. For an explanation see the end of Section 11.

In Section 12, we estimate the heights (in a natural sense) of all the quantities occurring in our theorems. The bounds are polynomial in $h(V)$ and $R(G)$ if $G = \sqrt{G}$; but otherwise they may involve an extra, possibly unavoidable, exponential dependence on $R(G)$. Here too there is a Northcott Property to ensure effectivity.

At first sight it may seem that the methods of Derksen [**7**] and Masser [**22**] are unrelated. But there are close connections, and we give some hints of this in our exposition. Here, we mention just that Masser [**22**] works with derivations and Derksen [**7**] works with p-automata and 'free Frobenius splitting'. For example, over $\mathbf{F}_p(t)$, [**22**, p. 196] has $\delta_i = (d/dt)^i$ $(i = 0, \ldots, p-1)$ while [**7**, p. 198] splits $\mathbf{F}_p(t)$ into a direct sum of one-dimensional $\mathbf{F}_p(t^p)$-subspaces V_i $(i = 0, \ldots, p-1)$ and considers the associated projections λ_i. In the natural case $V_i = t^i\mathbf{F}_p(t^p)$ one checks easily that the vectors $(\delta_0, t\delta_1 \ldots, t^{p-1}\delta_{p-1})$ and $(\lambda_0, \lambda_1, \ldots, \lambda_{p-1})$ are connected via an invertible matrix over $\mathbf{F}_p$. So in some sense differentiating is equivalent to projecting. We can also quote Hrushovski [**16**, p. 669] 'Distinguishing a basis for K/K^p has the effect of fixing also a stack of Hasse derivations.' We will follow Masser [**22**] with derivations, but as a matter of fact we do not need Hasse derivations in this paper (see the remarks at the end of Section 5). Neither do we use Model Theory as in [**16**, **23**, **24**].

Here is a brief section-by-section account of what follows.

We begin in Section 2 by explaining heights. Then in Section 3 we introduce derivations, and we use all these to give preliminary effective versions of the two main technical results of [**22**] about dependence over the field of differential constants.

In Section 4, we explain regulators, and in Section 5 we use these to refine the work of Section 3.

Then Section 6 contains a technical result which enables us to identify isotriviality, and in Section 7 we record some observations about automorphisms and heights of varieties V.

We are now in a position, in Section 8, to make effective the main argument of Masser [**22**] yielding the subvarieties W, at least for points over $\sqrt{G}$ and when V is either a hyperplane or trivial. We treat general V in Section 9 but omitting the isotriviality of the W. This omission is then remedied in Section 10 with a simple inductive argument, and in Section 11 we show how to treat points over G. We can then in Section 12 prove effective versions of our Descent Steps and Theorems.

Finally in Section 13, as already mentioned, we show that various aspects of our results cannot be further improved.

We would also like to draw attention to a very recent manuscript of Adamczewski and Bell [**2**] for further work in the context of p-automata; in particular, this covers also equations (1.1) and recurrences.

2. *Heights*

The theorems above for arbitrary fields can easily be reduced to the case when the field is finitely generated over its ground field $\mathbf{F}_p$ (see Section 12). In general, let K be finitely generated over a subfield k in any characteristic. We shall define heights on K relative to k; thus, we suppose that K is a transcendental extension of k. Here, we do not know any basis-free notion of height, and thus we choose a transcendence basis $\mathcal{B}$ of K over k with elements $t_1, \ldots, t_b$ regarded as independent variables over k. The height $\tilde{h}(a) = \tilde{h}_{\mathcal{B}}(a)$ of an element $a \neq 0$ of $k[\mathcal{B}] = k[t_1, \ldots, t_b]$ will be its total degree $\deg a$ regarded as a polynomial; also $\tilde{h}(0) = 0$. The height can be extended to an element x of the quotient field $k(\mathcal{B}) = k(t_1, \ldots, t_b)$ by writing $x = a_1/a_0$ for coprime polynomials a_0, a_1 in $k[\mathcal{B}]$ and defining

$$\tilde{h}(x) = \tilde{h}_{\mathcal{B}}(x) = \max\{\deg a_0, \deg a_1\}. \tag{2.1}$$

That suffices for most examples, but for mixing problems we have to extend further to all of K. This is a standard matter using valuations.

There is a valuation on $k[\mathcal{B}]$ corresponding to total degree and defined by $|a|_\infty = \exp(\deg a)$ $(a \neq 0)$; and of course $|0|_\infty = 0$. This extends at once to $k(\mathcal{B})$ by multiplicativity. And for every irreducible p in $k[\mathcal{B}]$ there is a valuation defined on $k[\mathcal{B}]$ by $|a|_p = \exp(-\omega_p(a) \deg p)$ $(a \neq 0)$, where $\omega_p(a)$ is the exact power of p dividing a; and again $|0|_\infty = 0$. And it too extends to $k(\mathcal{B})$ by multiplicativity. Using v to run over ∞ and all the p, we have the product formula $\prod_v |x|_v = 1$ $(x \neq 0)$ and the height formula $\tilde{h}(x) = \log \prod_v \max\{1, |x|_v\}$.

Now K is a finite extension of $k(\mathcal{B})$, say of degree d. Thus, each valuation v has finitely many extensions w to K, written $w|v$. In fact

$$|x|_w = |N(x)|_v^{1/d_w}, \tag{2.2}$$

where the norm is from the completion K_w to the completion $k(\mathcal{B})_v$ and d_w is the relative degree. We also have $\sum_{w|v} d_w = d$. Now the product formula

$$\prod_w |x|_w^{d_w} = 1 \quad (x \neq 0)$$

holds. Further, the formula

$$\tilde{h}(x) = \frac{1}{d} \log \prod_w \max\{1, |x|_w^{d_w}\}$$

extends the height $\tilde{h} = \tilde{h}_{\mathcal{B}}$ to an absolute height on K. For all these, see [**19**, pp. 1–19] or [**3**, pp. 1–10].

Actually for convenience in estimating we will use from now on the relative height

$$h(x) = h_{\mathcal{B}}(x) = d\tilde{h}(x) \geqslant 1.$$

This can be calculated directly from the minimum polynomial in the following extension of (2.1).

Lemma 2.1. *Suppose x in K satisfies an equation $A(x) = 0$ with $A(t) = a_0 t^e + \ldots + a_e$ for $a_0, \ldots, a_e$ in $k[\mathcal{B}]$ with $a_0 \neq 0$ and $A(t)$ irreducible over $k[\mathcal{B}]$. Then $eh(x) = d \max\{\deg a_0, \ldots, \deg a_e\}$.*

Proof. Over a splitting field L we have $A(t) = a_0(t - x_1) \ldots (t - x_e)$, and we can extend, keeping the same notation, all the valuations to L. Then Gauss's Lemma gives

$$\max\{|a_0|_w, \ldots, |a_e|_w\} = |a_0|_w \max\{1, |x_1|_w\} \ldots \max\{1, |x_e|_w\}.$$

If w does not divide ∞ then the left-hand side is 1 because $a_0,\ldots,a_e$ are coprime; otherwise they are all $\max\{|a_0|_\infty,\ldots,|a_e|_\infty\}$. Taking the product with exponents d_w and then taking logarithms gives on the left-hand side $d\max\{\deg a_0,\ldots,\deg a_e\}$ and on the right-hand side $h(x_1)+\ldots+h(x_e)$. This last is just $eh(x)$ because $x_1,\ldots,x_e$ are conjugate over $k(\mathcal{B})$. □

An immediate consequence of Lemma 2.1 is the Northcott Property; namely that for any H there are at most finitely many x in K with $h(x)\leqslant H$.

We will also need the standard extensions to vectors. So for $x_1,\ldots,x_l$ in K we define

$$h(x_1,\ldots,x_l)=\log\prod_w \max\{1,|x_1|_w^{d_w},\ldots,|x_l|_w^{d_w}\}.$$

For example, $h(a_0,\ldots,a_e)$ in the situation of Lemma 2.1 is just $d\max\{\deg a_0,\ldots,\deg a_e\}$. The Northcott Property extends at once to K^l.

3. *Dependence with heights*

Given K finitely generated and transcendental over k, there is always a separable transcendence basis $\mathcal{B}=(t_1,\ldots,t_b)$; this means that K is separable over $k(\mathcal{B})$. As above write $d=[K:k(\mathcal{B})]$. On $k[\mathcal{B}]$, we have the standard derivations $\partial/\partial t_1,\ldots,\partial/\partial t_b$, which extend in the obvious way to $k(\mathcal{B})$. And by separability they extend uniquely to K. For all these, see [**18**, pp. 183–184]. For an integer $i\geqslant 0$, we define $\mathcal{D}(i)$ as the set of operators

$$D=\left(\frac{\partial}{\partial t_1}\right)^{i_1}\cdots\left(\frac{\partial}{\partial t_b}\right)^{i_b}$$

as $i_1,\ldots,i_b$ run over all non-negative integers with $i_1+\ldots+i_b\leqslant i$. This is not quite the same as [**22**, p. 196], where we had $i\geqslant 1$ and $i_1+\ldots+i_b<i$.

It will be convenient for later calculations to define a quantity $h(x;i)$ as follows. We order in some way the operators $D_1,\ldots,D_l$ of $\mathcal{D}(i)$, and we define for $x\neq 0$

$$h(x;i)=h_{\mathcal{B}}(x;i)=h\left(\frac{D_1x}{x},\ldots,\frac{D_lx}{x}\right)$$

of course independent of the ordering.

The next result is an explicit version of [**22**, Lemma 3, p. 195] however without reference to any group G. We write C for the field of differential constants in K. For zero characteristic this is k, but for positive characteristic p it is the set of pth powers of elements of K.

Lemma 3.1. *For $m\geqslant 2$ suppose $c_1,\ldots,c_m$ are in C and $x_1,\ldots x_m$ are in K^* with*

$$c_1x_1+\ldots+c_mx_m=1. \tag{3.1}$$

Then either

(a) $h(c_1x_1,\ldots,c_mx_m)\leqslant(m+1)(h(x_1;m-1)+\ldots+h(x_m;m-1))$
or
(b) *$x_1,\ldots,x_m$ are linearly dependent over C.*

Proof. If (b) does not hold, then the theory of the generalized Wronskian (see for example [**19**, p. 174]) shows that we may find operators D_i in $\mathcal{D}(i)$ $(i=0,\ldots,m-1)$ such that the matrix with entries D_ix_j $(i=0,\ldots,m-1;\ j=1,\ldots,m)$ is non-singular. Applying them to (3.1) we obtain

$$\sum_{j=1}^m \frac{D_ix_j}{x_j}(c_jx_j)=D_i(1),\quad (i=0,\ldots,m-1).$$

These can be solved by Cramer's Rule to obtain $c_jx_j = (w_j/w_0)$ $(j = 1,\ldots,m)$, where $w_0 \neq 0$ is the determinant of the matrix with entries (D_ix_j/x_j) $(i = 0,\ldots,m-1;\ j = 1,\ldots,m)$. Noting that this determinant is multilinear in the columns, we find that $h(w_0) \leqslant h(x_1; m-1) + \ldots + h(x_m; m-1)$. The same bound holds for the $h(w_j)$ $(j = 1,\ldots,m)$. We conclude that $h(c_1x_1,\ldots,c_mx_m) = h(w_1/w_0,\ldots,w_m/w_0)$ is at most

$$h(w_0) + h(w_1) + \ldots + h(w_m) \leqslant (m+1)(h(x_1; m-1) + \ldots + h(x_m; m-1))$$

as required. □

We deduce an explicit version of [**22**, Lemma 4, p. 197], also without G.

LEMMA 3.2. *For $m \geqslant 2$ suppose $x_0, x_1, \ldots x_m$ are in K^* and linearly dependent over C but $x_1, \ldots x_m$ are linearly independent over C. Then there is a relation*

$$c_1x_1 + \ldots + c_mx_m = x_0 \tag{3.2}$$

with $c_1,\ldots,c_m$ in C and

$$h\left(\frac{c_1x_1}{x_0},\ldots,\frac{c_mx_m}{x_0}\right) \leqslant (m+1)\left(h\left(\frac{x_1}{x_0}; m-1\right) + \ldots + h\left(\frac{x_m}{x_0}; m-1\right)\right).$$

Proof. There is certainly a relation (3.2) with $c_1,\ldots,c_m$ in C, and we apply Lemma 3.1 to the quotients $x_1/x_0,\ldots,x_m/x_0$. As $x_1,\ldots x_m$ are linearly independent over C, the conclusion (b) cannot hold. Now conclusion (a) is just what we need, and this completes the proof. □

In Section 5, we shall prove versions of Lemmas 3.1 and 3.2 that are uniform for $x_0, x_1, \ldots, x_m$ in a finitely generated group G as in [**22**]. By way of preparation, the next result illustrates the logarithmic nature of the quantities $h(\,; i)$.

LEMMA 3.3. *For any $x \neq 0$ and $y \neq 0$ in K and any integers $i \geqslant 0$ and $e \geqslant 0$, we have $h(xy; i) \leqslant h(x; i) + h(y; i)$ and $h(x^e; i) \leqslant ih(x; i)$.*

Proof. Let D be in $\mathcal{D}(i)$. By distributing operators over the factors of xy as in Leibniz, we see that $D(xy)/xy$ is a sum with generalized binomial coefficients of products $(E(x)/x)(F(y)/y)$ with operators E and F also in $\mathcal{D}(i)$. Taking $D = D_1,\ldots,D_l$ as in the definition of $h(xy; i)$, we deduce the first inequality of the present lemma by standard height calculations.

When e is a positive integer, a similar argument shows that $D(x^e)/x^e$ is a sum with generalized binomial coefficients of products $E_1(x)/x \ldots E_e(x)/x$ with operators $E_1,\ldots,E_e$ also in $\mathcal{D}(i)$. Here, $E_1 \ldots E_e = D$, so that there are at most i terms not equal to 1 in this product. Thus, $D(x^e)/x^e$ is a polynomial of total degree at most i in the $E(x)/x$ for E in $\mathcal{D}(i)$. The second inequality now follows in a similar way, at least for $e \geqslant 1$. The result is trivial for $e = 0$. □

LEMMA 3.4. *For any $x \neq 0$ in K and any integer $i \geqslant 0$, we have $h(x; i) \leqslant 4idh(x)$.*

Proof. This is trivial for $i = 0$, so we assume from now on $i \geqslant 1$. We have an equation $A(x) = 0$ as in Lemma 2.1, of degree $e \leqslant d$. Denote by $A'(t)$ the derivative with respect to t. Pick any D in $\mathcal{D}(i)$. We claim that $B_i = (A'(x))^{2i-1}Dx$ is a polynomial in x and various derivatives D_0a of various coefficients a of A, with coefficients in k and of degree at most $(2i-1)(e-1)+1$ in x and of total degree at most $2i-1$ in the D_0a. We prove this by induction on i.

When $i=1$ we have for example $D=\partial/\partial t_1=\partial$ (say), and applying this to $A(x)=0$ yields $B_1=-\sum_{j=0}^{e}(\partial a_{e-j})x^j$ for which the claim is clear.

Assuming $Dx=B_i/(A'(x))^{2i-1}$ with B_i as above, we do the induction step by applying one more operator, again say $\partial/\partial t_1=\partial$. We obtain

$$(A'(x))^{2i}\partial Dx = A'(x)\partial B_i-(2i-1)B_i\partial(A'(x)).$$

Here ∂B_i involves x to degree at most $(2i-1)(e-1)+1$ and also x to degree at most $(2i-1)(e-1)$ multiplied by $\partial x=B_1/A'(x)$, together with D_0a to total degree at most $2i-1$. Similarly, $\partial(A'(x))$ involves x to degree at most $e-1$ and also x to degree at most $e-2$ (if $e\neq 1$) multiplied by $\partial x=B_1/A'(x)$, together with D_0a to total degree at most 1. Multiplying by $A'(x)$ we obtain $(A'(x))^{2i+1}\partial Dx$ involving x to degree at most

$$e-1+\max\{(2i-1)(e-1)+1+(e-1),\ (2i-1)(e-1)+e\}=(2(i+1)-1)(e-1)+1,$$

and the degree in D_0a is at most $(2i-1)+1+1=2(i+1)-1$. This proves the claim in general.

There follows at once the estimate

$$\log|B_i|_w\leqslant((2i-1)(e-1)+1)\log\max\{1,|x|_w\}$$

for any w not dividing ∞; otherwise, we obtain an extra term $(2i-1)\max\{\deg a_0,\ldots,\deg a_e\}$. The same estimates also hold for $\log|C|_w$ where $C=x(A'(x))^{2i-1}$.

Now write B_{ij} for the B_i corresponding to the operators D_j $(j=1,\ldots,l)$ of $\mathcal{D}(i)$, so that $D_jx/x=B_{ij}/C$. Then

$$h\left(\frac{D_1x}{x},\ldots,\frac{D_lx}{x}\right)=\sum_w d_w\max\{\log|B_{i1}|_w,\ldots,\log|B_{il}|_w,\log|C|_w\}$$

which is at most

$$((2i-1)(e-1)+1)h(x)+(2i-1)d\max\{\deg a_0,\ldots,\deg a_e\}.$$

Finally by Lemma 2.1 this is at most

$$((2i-1)(e-1)+1)h(x)+(2i-1)eh(x)\leqslant 4ieh(x)\leqslant 4idh(x)$$

as required. This completes the proof of the present lemma. □

In view of our consistent use of the relative height (as opposed to the absolute height), the factor d here looks like a normalization error. However it cannot be avoided, as the example $x=((t+1)/t)^{1/d}$ $(t=t_1)$ in $K=k(t)(x)=k(x)$ shows. One finds that the rational function $(1/x)(\partial^i x/\partial t^i)$ has denominator $(t(t+1))^i$. So its height is at least $2id=2idh(x)$, which also shows that our dependence on i is not too bad. Perhaps even the factor 4 essentially cannot be avoided.

4. *Regulators*

Let K be finitely generated and transcendental over k as in the preceding section, and let $\mathcal{B}$ be a transcendence basis. Let G be a subgroup of K^* finitely generated modulo k^*; that is, $G/(G\cap k^*)$ is finitely generated. We show here how to define a regulator $R(G)=R_{\mathcal{B}}(G)$.

For all w except finitely many we have $|g|_w=1$ for every g in G. Pick a set of $N\geqslant 1$ valuations containing these exceptions. We order the set to produce a map $\mathcal{L}$ from G into $\mathbf{R}^N$ whose typical coordinate is $d_w\log|g|_w$. In fact by (2.2) $\mathcal{L}(G)$ lies in $\mathbf{Z}^N$ and is therefore discrete. Thus, it is a (full) lattice in the real subspace it generates, whose dimension is the rank r of $G/(G \cap k^*)$. If $r \geqslant 1$, then we define the regulator just as the determinant

$$R(G) = R_{\mathcal{B}}(G) = \det \mathcal{L}(G) \geqslant 1;$$

clearly independent of the set above or its ordering, and if $r = 0$, then we define $R(G) = 1$. This does not quite coincide with the standard definition for the unit group in algebraic number theory, because the latter is obtained by a projection to one dimension lower. But they are equal up to a constant factor.

The following example will be quoted later. With $K = \mathbf{F}_p(t)$ (and the obvious $\mathcal{B}$) and G_l generated by t^l and $1 - t$ we have $N = 3$ corresponding to valuations at $t = 0, 1, \infty$; and so vectors $(l, 0, l)$ and $(0, 1, 1)$ giving $R_{\mathcal{B}}(G_l) = l\sqrt{3}$.

Lemma 4.1. *Let G and G' in K^* be finitely generated modulo k^* with G of finite index in G'. Then*

$$R(G) = \frac{[G' : G]}{[G' \cap k^* : G \cap k^*]} R(G') = [G'/(G' \cap k^*) : G/(G \cap k^*)] R(G'),$$

where we identify $G/(G \cap k^*)$ as a subgroup of $G'/(G' \cap k^*)$.

Proof. The quotients $G/(G \cap k^*)$ and $G'/(G' \cap k^*)$ are torsion-free, both with the same rank, say r. If $r = 0$, then the lemma is trivial. Otherwise, using elementary divisors we can find generators $\gamma_1, \ldots, \gamma_r$ of $G'/(G' \cap k^*)$ and positive integers $d_1, \ldots, d_r$ such that $\gamma_1^{d_1}, \ldots, \gamma_r^{d_r}$ generate $G/(G \cap k^*)$. Then the relationship between $\mathcal{L}(G')$ and $\mathcal{L}(G)$ is clear, and the lemma follows. □

Lemma 4.2. *Let G in K^* be finitely generated modulo k^*, let x be in K^*, and let G' be the group generated by x and the elements of G. Then $R(G') \leqslant 2h(x)R(G)$.*

Proof. It is geometrically clear that if Λ is any lattice in euclidean space, then $\det(\Lambda + \mathbf{Z}\mathbf{v}) \leqslant \det(\Lambda)|\mathbf{v}|$ for the length, at least if $\mathbf{v}$ is not in the space spanned by Λ. But this continues to hold for all $\mathbf{v}$ provided only $|\mathbf{v}| \geqslant 1$ and $\Lambda + \mathbf{Z}\mathbf{v}$ remains discrete. In particular, it holds for $\Lambda = \mathcal{L}(G)$ and $\mathbf{v} = \mathcal{L}(x)$. We conclude $R(G') \leqslant |\mathcal{L}(x)|R(G)$. Finally, we have by definition and the product formula

$$h(x) = \sum_w \max\{0, m_w\} = \frac{1}{2} \sum_w |m_w| \tag{4.1}$$

for $m_w = d_w \log |x|_w$. And

$$|\mathcal{L}(x)|^2 = \sum_w m_w^2 \leqslant \left(\sum_w |m_w| \right)^2 = 4(h(x))^2.$$

The lemma follows. □

We can recover a basis from the regulator as follows.

Lemma 4.3. *Let G be a subgroup of K^* finitely generated modulo k^* with $G/(G \cap k^*)$ of rank $r \geqslant 1$. Then there are $g_1, \ldots, g_r$ in G generating $G/(G \cap k^*)$, with*

$$h(g_1) \ldots h(g_r) \leqslant \frac{1}{r} \delta(r) R(G)^2$$

for $\delta(r) = r^{3r}$.

Proof. By Minkowski's Second Theorem (see for example [**5**, Theorem V, p. 218]) there are $\tilde{g}_1, \ldots, \tilde{g}_r$ in G multiplicatively independent modulo k^*, with

$$|\mathcal{L}(\tilde{g}_1)| \ldots |\mathcal{L}(\tilde{g}_r)| \leqslant \frac{2^r}{V(r)} \det \mathcal{L}(G) = \frac{2^r}{V(r)} R(G) \tag{4.2}$$

for the Euclidean norms and the volume $V(r)$ of the unit ball in $\mathbf{R}^r$. By geometry $V(r) \geqslant (2/\sqrt{r})^r$. We obtain a basis in the standard way using the argument of Mahler–Weyl (see for example [**5**, Lemma 8, p. 135]); there results

$$|\mathcal{L}(g_i)| \leqslant \max\left\{1, \frac{i}{2}\right\} |\mathcal{L}(\tilde{g}_i)|, \quad i = 1, \ldots, r,$$

and so $2^r/V(r)$ in (4.2) gets replaced by $(r!/2^{r-1})r^{r/2} \leqslant r^{3r/2}/2^{r-1}$. Now (4.1) gives

$$h(g) = \sum_w \max\{0, m_w\} = \frac{1}{2} \sum_w |m_w|$$

for $m_w = d_w \log |g|_w$ in $\mathbf{Z}$. And $|m| \leqslant m^2$ for any m in $\mathbf{Z}$, so we obtain

$$h(g) \leqslant \frac{1}{2} \sum_w m_w^2 = \frac{1}{2} |\mathcal{L}(g)|^2.$$

Therefore

$$h(g_1) \ldots h(g_r) \leqslant \frac{4r^{3r}}{2^{3r}} R(G)^2 < \frac{1}{r} \delta(r) R(G)^2$$

as desired. □

In view of (4.2) it seems a pity that the square of the regulator appears in Lemma 4.3. But it cannot be avoided. For example, let $\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_l$ be different constants in k, and consider G generated by $g = (t - \alpha_1) \ldots (t - \alpha_l)/(t - \beta_1) \ldots (t - \beta_l)$ in $K = k(t)$. Then $R(G) = \sqrt{2l}$. The only possibilities for g_1 are $\gamma g^{\pm 1}$ with γ constant. But then $h(g_1) = l$, so any bound $h(g_1) \leqslant \delta(1) R(G)$ is impossible.

This leads to the following uniform version of Lemma 3.4 when x lies in G. Write G_k for the group generated by the elements of G and k^*.

LEMMA 4.4. *Let G be a subgroup of K^* finitely generated modulo k^* with $G/(G \cap k^*)$ of rank $r \geqslant 1$. Then for any g in G we have $h(g; i) \leqslant 4i^2 d\delta(r) R(G)^2$. Further for any positive integer l there is g_0 in G_k and g' in G with $g = g_0 g'^l$ and $h(g_0) \leqslant l\delta(r) R(G)^2$.*

Proof. Choose basis elements $g_1, \ldots, g_r$ according to Lemma 4.3, and write $g = c g_1^{e_1} \ldots g_r^{e_r}$ for rational integers $e_1, \ldots, e_r$ and c in k^*. Replacing some of the g_j by their inverses, we can assume that all $e_j \geqslant 0$; this does not affect the estimate in Lemma 4.3. Then by Lemma 3.3

$$h(g; i) = h(g_1^{e_1} \ldots g_r^{e_r}; i) \leqslant h(g_1^{e_1}; i) + \ldots + h(g_r^{e_r}; i) \leqslant i(h(g_1; i) + \ldots + h(g_r; i)).$$

This in turn by Lemma 3.4 is at most

$$4i^2 d(h(g_1) + \ldots + h(g_r)) \leqslant 4i^2 d r h(g_1) \ldots h(g_r) \leqslant 4i^2 d\delta(r) R(G)^2 \tag{4.3}$$

as required in the first part of the present lemma. And the second part follows by writing $e_j = f_j + l e_j'$ with $0 \leqslant f_j < l$ $(j = 1, \ldots, r)$ (compare also [**7**, p. 197]), taking $g_0 = c g_1^{f_1} \ldots g_r^{f_r}, g' = g_1^{e_1'} \ldots g_r^{e_r'}$ and using the inequality in (4.3). □

The final result of this section will lead easily to a quantitative version of [**22**, Lemma 2, p. 193], such as those mentioned in [**22**, pp. 194, 195]. However, it involves better constants and is no longer restricted to positive characteristic. It is here, by the way, that the radical $\sqrt{G}$ makes its essential appearance in the whole story.

LEMMA 4.5. *Suppose that x and y are in K^* with x not in $\sqrt{G_k}$ and y^q/x in G for some positive integer q. Then $q \leqslant 2h(x)R(G)$.*

Proof. Let G' be the group generated by x and the elements of G, and let G'' be the group generated by y and the elements of G, so that G' lies in G''. Since x is not in $\sqrt{G}$, it is easy to see that the index $[G'' : G'] = q$. Since x is not even in $\sqrt{G_k}$, it is even easier to see that $G \cap k^* = G' \cap k^* = G'' \cap k^*$. Thus, by Lemma 4.1, we have $R(G') = qR(G'') \geqslant q$. On the other hand, $R(G') \leqslant 2h(x)R(G)$ by Lemma 4.2, and the result follows. □

5. *Dependence with regulators*

Let K be finitely generated and transcendental over k as in the preceding sections, and let $\mathcal{B}$ be a transcendence basis, now assumed separable, with elements $t_1, \ldots, t_b$. We continue to abbreviate the height $h_{\mathcal{B}}$ as h, and again we write C for the field of differential constants of K.

The following result eliminates the height functions $h(x, m-1)$ from Lemma 3.1, thereby providing a more useful explicit version of Masser [**22**, Lemma 3].

LEMMA 5.1. *Let G in K^* be finitely generated of rank $r \geqslant 1$ modulo k^*, and for $m \geqslant 2$ suppose $c_1, \ldots, c_m$ are in C and $g_1, \ldots g_m$ are in G with*

$$c_1 g_1 + \ldots + c_m g_m = 1.$$

Then either

(a) $h(c_1 g_1, \ldots, c_m g_m) \leqslant 4m^4 d\delta(r)R(G)^2$
or
(b) *$g_1, \ldots, g_m$ are linearly dependent over C.*

Proof. Just use Lemma 3.1 together with the inequalities

$$h(g; m-1) \leqslant 4(m-1)^2 d\delta(r)R(G)^2 \tag{5.1}$$

from Lemma 4.4, with $g = g_1, \ldots, g_m$. □

Similarly, we deduce a more useful explicit version of Masser [**22**, Lemma 4].

LEMMA 5.2. *Let G in K^* be finitely generated of rank $r \geqslant 1$ modulo k^*, and for $m \geqslant 2$ suppose $g_0, g_1, \ldots g_m$ are in G and linearly dependent over C but $g_1, \ldots g_m$ are linearly independent over C. Then there is a relation*

$$c_1 g_1 + \ldots + c_m g_m = g_0$$

with $c_1, \ldots, c_m$ in C and

$$h\left(\frac{c_1 g_1}{g_0}, \ldots, \frac{c_m g_m}{g_0}\right) \leqslant 4m^4 d\delta(r)R(G)^2.$$

Proof. Just use Lemma 3.2 and (5.1), this time with $g = g_1/g_0, \ldots, g_m/g_0$. □

We have followed the proof in [**22**] quite closely. It would have been nice to see the well-known number $m(m-1)/2$ in place of $4m^4$, and also some notion of genus and S-units as in various formulations of *abc* matters over function fields. But despite the considerations of Bombieri and Gubler [3, Chapter 14] in zero characteristic and those of Hsia and Wang [17] for arbitrary characteristic we have been unable to supply a satisfactory version. The results of Hsia and Wang [17] are especially interesting in their use of divided derivatives or hyperderivations, which, for example, in characteristic p leads to linear dependence over the field of p^eth powers, not just over C with $e = 1$. If this could be done in our situation, then it would probably lead to simplifications in the rest of our proof, and possibly to the elimination of the proposition in Section 8. But it seems that the results of Hsia and Wang [17] cannot be directly applied to our Lemma 5.1, due to the presence of $c_1, \ldots, c_m$ whose heights cannot be controlled.

6. *Isotriviality*

We take a well-earned break from estimating. From now on K will have positive characteristic p (actually this assumption is not really needed until Section 8), and, as in Section 1, we write $\mathbf{F}_K$ for $\overline{\mathbf{F}_p} \cap K$. This field plays the role of k in Sections 2–5.

Suppose $n \geqslant m \geqslant 1$. For $a(i, j)$ in K the normalized equations

$$X_i = a(i, 0)X_0 + \ldots + a(i, m-1)X_{m-1} = \sum_{j=0}^{m-1} a(i, j)X_j, \quad i = m, m+1, \ldots, n \tag{6.1}$$

define in $\mathbf{P}_n$ a linear variety V of dimension $m - 1$. When G is a subgroup of K^*, we need some conditions which ensure that V is G-isotrivial.

Now any G-automorphism taking $(X_0, \ldots, X_n)$ to $(g_0X_0, \ldots, g_nX_n)$ leads after renormalization to new coefficients $(g_i/g_j)a(i, j)$. If the new forms are defined over $\mathbf{F}_K$, then every non-zero $a(i, j)$ has the shape $(g_j/g_i)\alpha(i, j)$ for non-zero $\alpha(i, j)$ in $\mathbf{F}_K$. In particular, each equation in (6.1) defines a G-isotrivial variety. But also each quotient

$$\frac{a(i_1, j_1)a(i_2, j_2)a(i_3, j_3) \ldots a(i_{k-1}, j_{k-1})a(i_k, j_k)}{a(i_1, j_2)a(i_2, j_3)a(i_3, j_4) \ldots a(i_{k-1}, j_k)a(i_k, j_1)}, \quad k = 2, \ldots, n+1, \tag{6.2}$$

with everything in the numerator and denominator non-zero, lies in $\mathbf{F}_K$. The following result gives a converse statement which guarantees that the equations (6.1) become defined over $\mathbf{F}_K$ after applying a suitable G-automorphism and renormalizing. In particular, it guarantees that V is G-isotrivial; but without the need to recombine the equations.

LEMMA 6.1. *Suppose that each equation in* (6.1) *defines a G-isotrivial variety, and that each quotient* (6.2) *lies in $\mathbf{F}_K$ provided everything in the numerator and denominator is non-zero. Then V is G-isotrivial.*

Proof. We argue by induction on the number $n - m + 1 \geqslant 1$ of equations. If $n - m + 1 = 1$, then the result is obvious without using (6.2). Suppose we have done it for $n - m \geqslant 1$ equations, namely the first $n - m$ in (6.1), and let us add another equation, namely the last one in (6.1).

Restricting to $i < n$ and the appropriate j in (6.2), we see from the induction hypothesis that a suitable G-automorphism trivializes the first $n - m$ equations, without bothering about X_n. This means that we can assume that all $a(i, j) \neq 0$ $(i < n)$ are in $\mathbf{F}_K$; while the isotriviality of the last equation means that all $a(n, j) \neq 0$ are in G. We now want to trivialize the last equation.

How can we trivialize a given coefficient $a(n, j) \neq 0$ in the last equation? If all $a(i, j) = 0$ $(i < n)$, so that the first $n - m$ equations did not involve X_j, then we can simply replace X_j by $a(n, j)X_j$ and this will not change the first $n - m$ equations. We do this for all such j.

If there is only a single j with some $a(i,j) \neq 0$ $(i < n)$, then we can still replace X_j by $a(n,j)X_j$; but we then have to correct the new coefficients $a(i,j)/a(n,j) \neq 0$ of X_j in the ith equation by replacing X_i by $a(n,j)X_i$ $(i = m, \ldots, n-1)$. Things are less easy when there is more than one such j. Call these 'bad'.

Now we say for different j and j' in the set $\{0, \ldots, m-1\}$ that $j \sim j'$ if there is $i < n$ with

$$a(i,j)a(i,j') \neq 0 \tag{6.3}$$

(in particular then j and j' are both bad). This relation is symmetric but probably not transitive. We can extend it via reflexivity and transitivity to a genuine equivalence relation on the bad elements of $\{0, \ldots, m-1\}$, which we then denote by $\simeq$.

We assume for the moment that there is a single equivalence class: any two j and j' are related.

Let j and j' be different bad elements, so that $a(i,j) \neq 0$, $a(i',j') \neq 0$ for some $i, i' < n$. From our equivalence class assumption $j \simeq j'$. Suppose that

$$j = j_1 \sim j_2 \sim \ldots \sim j_{k-1} \sim j_k = j',$$

where of course we can take $2 \leqslant k \leqslant n+1$. Then we obtain from (6.3)

$$a(i_1,j_1)a(i_1,j_2) \neq 0,\ a(i_2,j_2)a(i_2,j_3) \neq 0, \ldots,\ a(i_{k-1},j_{k-1})a(i_{k-1},j_k) \neq 0$$

for some $i_1, i_2, \ldots, i_{k-1} < n$. We use (6.2) with $i_k = n$ to see that

$$\frac{a(i_1,j_1)a(i_2,j_2)a(i_3,j_3)\ldots a(i_{k-1},j_{k-1})a(n,j')}{a(i_1,j_2)a(i_2,j_3)a(i_3,j_4)\ldots a(i_{k-1},j_k)a(n,j)}$$

lies in $\mathbf{F}_K$. However, the first $k-1$ terms in both numerator and denominator already lie in $\mathbf{F}_K$, because we trivialized the first $n-m$ equations. Consequently, $a(n,j')/a(n,j)$ lies in $\mathbf{F}_K$.

Thus we have shown that all $a(n,j)$ for bad j are multiples of a single one, call it g, by elements of $\mathbf{F}_K$. Now they can be simultaneously trivialized on replacing X_j by gX_j. Again we must correct the new coefficients $a(i,j)/g \neq 0$ of X_j in the ith equation by replacing X_i by gX_i $(i = m, \ldots, n-1)$.

What happens if there is more than a single equivalence class on the bad elements of $\{0, \ldots, m-1\}$? Say there are $h \geqslant 2$ classes $J_1 \ldots, J_h$. Let I_1 be the set of i in $\{m, \ldots, n-1\}$ for which there is j in J_1 with $a(i,j) \neq 0$; and similarly for $I_2, \ldots, I_h$. Then $I_1, I_2, \ldots, I_h$ are disjoint, because for example with any j_1 in J_1 and any j_2 in J_2 there can be no i with $a(i,j_1)a(i,j_2) \neq 0$, else by (6.3) we would have $j_1 \sim j_2$. (If one wishes, then one can convert the matrix of the first $n-m$ equations into a block matrix using row and column permutations.) The argument above, using $i_1, \ldots, i_{k-1}$ in I_1, shows that all non-zero $a(n,j)$ $(j \in J_1)$ are multiples of a single one, call it g_1, by elements of $\mathbf{F}_K$. Similarly we obtain $g_2, \ldots, g_h$. Now we can trivialize the last row as follows. We replace the X_j $(j \in J_1)$ by g_1X_j and we correct the effect by replacing X_i by g_1X_i $(i \in I_1)$. Similarly, using $g_2, \ldots, g_h$ we trivialize the remaining coefficients. This completes the proof. □

7. *Automorphisms*

As above let K be a field, finitely generated and transcendental over $\mathbf{F}_p$, with G a subgroup of K^*. Suppose a linear variety in $\mathbf{P}_n$ is defined over K and G-isotrivial. Then by definition there is a G-automorphism ψ taking it to something defined over $\mathbf{F}_K = \overline{\mathbf{F}_p} \cap K$. To make our Theorems 1–3 fully effective we have to estimate this ψ; indeed, after doing the whole descent to single points using Theorem 1, for example, it is mainly G-automorphisms that are left.

Now it is convenient to use the projective height $h^{\mathbf{P}} = h^{\mathbf{P}}_{\mathcal{B}}$ defined on $\mathbf{P}_{l-1}(K)$ by

$$h^{\mathbf{P}}(x_1, \ldots, x_l) = \log \prod_w \max\{|x_1|_w^{d_w}, \ldots, |x_l|_w^{d_w}\}.$$

This yields at once a height $h(\psi)$ of a G-automorphism ψ, defined by (1.7), as

$$h(\psi) = h^{\mathbf{P}}(g_0, \ldots, g_n).$$

Also if V is linear in $\mathbf{P}_n$ defined over K, then it yields a height $h(V)$ in the standard way via the Grassmannian coordinates of V; see for example [**26**, p. 28], which however is in the context of number fields with euclidean norms at the archimedean valuations. Here, we have no archimedean valuations, so the norm problem is irrelevant. If $m-1 \geqslant 0$ is the dimension of V, then its Grassmannians $A(I)$ correspond to subsets I of $\{0, \ldots, n\}$ with cardinality $n-m+1 \leqslant n$. The Northcott Property extends at once to this height. Also for ψ in (1.7) the Grassmannians of $\psi(V)$ are the $A(I)/g(I)$, where $g(I) = \prod_{i \in I} g_i$. It follows easily that

$$h(\psi(V)) \leqslant h(V) + nh(\psi), \quad h(\psi^{-1}) \leqslant nh(\psi). \tag{7.1}$$

Less obvious is the following, which involves a second linear variety W also over K.

Lemma 7.1. *If $V \cap W$ is non-empty, then we have $h(V \cap W) \leqslant h(V) + h(W)$. If further $X_{n-1} \neq 0$ on V and the equations of V do not involve X_n, and W is defined by $X_n = aX_{n-1}$, then $h(V \cap W) \geqslant \max\{h(V), h(W)\}$.*

Proof. The upper bound may be compared with the inequality $h(V \cap W) + h(V + W) \leqslant h(V) + h(W)$ due independently to Struppeck–Vaaler [**27**, Theorem 1, p. 493] and Schmidt [**26**, Lemma 8A, p. 28]. These are proved over number fields; however, it is easily checked that the proof in [**26**] remains valid with trivial modifications. Already a special case was noted by Thunder [**28**] whose Lemma 5 (p. 157) implies $h(V + W) \leqslant h(V) + h(W)$ over function fields of a single variable provided $V \cap W$ is empty.

Regarding the lower bound, let $A(I)$ be the Grassmannians of V. Then it is easy to verify that the Grassmannians of $V \cap W$ consist of the $A(I)$ together with the $aA(J)$ for J not containing $n-1$. There follows $h(V \cap W) \geqslant h(V)$ at once. Also $X_{n-1} \neq 0$ on V means that at least one $A = A(J)$ is non-zero (see for example [**15**, Theorem 1, p. 298]), so we also obtain $h(V \cap W) \geqslant h^{\mathbf{P}}(A, aA) = h(a) = h(W)$. This completes the proof. □

It is the following result which enables ψ to be estimated in the Descent Steps.

Lemma 7.2. *Suppose that V is defined over K and is G-isotrivial. Then there is a G-automorphism ψ with $\psi(V)$ defined over $\mathbf{F}_K$ and $h(\psi) \leqslant n! h(V)$.*

Proof. Suppose $\dim V = m-1$ with Grassmannians $A(I)$; then as noted above the Grassmannians of $\psi(V)$ are the $A(I)/g(I)$, where $g(I) = \prod_{i \in I} g_i$. If $\psi(V)$ is defined over $\mathbf{F}_K$, then these have the shape $\lambda\alpha(I)$ for λ in K^* and $\alpha(I)$ in $\mathbf{F}_K$. Thus, we have $A(I) = \lambda\alpha(I)g(I)$ for all I; but we can restrict to the set $\mathcal{I}$ of all I with $A(I) \neq 0$ (and so $\alpha(I) \neq 0$). We can eliminate the λ by fixing I_0 in $\mathcal{I}$; this gives

$$\frac{g(I)}{g(I_0)} = \frac{A(I)}{A(I_0)} \frac{\alpha(I_0)}{\alpha(I)} \quad (I \in \mathcal{I}). \tag{7.2}$$

Conversely (7.2) implies that $\psi(V)$ is defined over $\mathbf{F}_K$.

To solve (7.2) for $g_0, \ldots, g_n$, we divide the numerator and denominator of the left-hand side by g_0^{n-m+1} and write it as $(g_1/g_0)^{a(I,1)} \ldots (g_n/g_0)^{a(I,n)}$ for integers $a(I, i)$ which are $0, 1$ and -1. If the vectors $\mathbf{a}(I)$ $(I \in \mathcal{I})$ with coordinates $a(I, i)$ $(i = 1 \ldots, n)$ have full rank n, then we can solve (7.2) by choosing $\mathbf{a}(I_1), \ldots, \mathbf{a}(I_n)$ linearly independent and then solving the subset of (7.2) with $I = I_1, \dots, I_n$. A multiplicative form of Cramer's Rule gives

$$\left(\frac{g_i}{g_0}\right)^b = Q_1^{b_{i1}} \dots Q_n^{b_{in}}, \quad Q_j = \frac{A(I_j)}{A(I_0)} \frac{\alpha(I_0)}{\alpha(I_j)}, \quad j = 1, \dots, n$$

with integers $b \neq 0$ and b_{ij}. These b_{ij} are minors of a matrix with entries $0, 1, -1$ and so $|b_{ij}| \leqslant (n-1)!$.

Now taking heights leads to

$$|b| h\left(\frac{g_1}{g_0}, \dots, \frac{g_n}{g_0}\right) \leqslant \max_{i=1,\dots,n} \{|b_{i1}| + \dots + |b_{in}|\} h(Q_1, \dots, Q_n).$$

The height on the left is $h(\psi)$ and that on the right is at most $h(V)$. The result follows at once, at least under our assumption that the $\mathbf{a}(I)$ $(I \in \mathcal{I})$ have full rank n.

If this assumption does not hold, then we simply increase the rank by successively adjoining unit vectors $\mathbf{e}_k$ until the rank becomes n; this amounts to the addition of equations $g_k/g_0 = 1$. Now we take a subset of n independent equations and solve again with Cramer. The resulting estimates are certainly no larger than before, and this completes the proof. □

8. *A proposition*

This, the main result of this section, is a first step in the proof of the Descent Step over $\sqrt{G}$, with V in $\mathbf{P}_n$ $(n \geqslant 2)$ either a hyperplane or defined over a finite field. We continue with our assumption that K is finitely generated over $\mathbf{F}_p$; thus, $\mathbf{F}_K = \overline{\mathbf{F}_p} \cap K$ is a finite field. Let G in K^* be finitely generated of rank $r \geqslant 1$ modulo $\mathbf{F}_K^*$; now we may write without confusion simply that G is finitely generated. It is known that the radical $\sqrt{G}$, which by definition lies still in K, is also finitely generated (see for example [**22**, p. 195]), also clearly of rank r over $\mathbf{F}_K^*$. For the moment, we work exclusively with this radical. We further assume that K is transcendental over $\mathbf{F}_p$ and we choose any separable transcendence basis $\mathcal{B}$; then we are free to apply the results of Sections 3–5 about heights $h = h_{\mathcal{B}}$ and regulators $R = R_{\mathcal{B}}$.

We say that V is transversal if every coordinate X_i $(i = 0, \dots, n)$ actually occurs in the defining equations. This property is independent of the choice of equations. Its purpose is to prevent 'free variables' as in (1.1) with $a_i \neq 0$.

Transversality is a harmless restriction because we could overcome it simply by working in lower dimensions. Clearly, every linear subvariety of a transversal variety is also transversal. Also a transversal variety must be proper (that is, not the full $\mathbf{P}_n$).

We recall the function δ from Lemma 4.3.

PROPOSITION. *Let V be a transversal linear subvariety of $\boldsymbol{P}_n$ defined over K, and suppose either that V has dimension $n-1$ or that V is defined over some $\mathbf{F}_q$. Suppose also that V is not contained in any coset $T \neq \boldsymbol{P}_n$. Let π be any point of $V(\sqrt{G})$.*

If V has dimension $n-1$, then either

(i) *there is a proper linear subvariety W of V, also defined over K, with*

$$h(W) \leqslant 8n^5 4^n d\delta(n+r) h(V)^{2n} R(\sqrt{G})^2,$$

such that π lies in $W(\sqrt{G})$,

or

(ii) *there is a $\sqrt{G}$-automorphism ψ with*

$$h(\psi) \leqslant np\delta(n+r) R(\sqrt{G})^2,$$

a point π' and a linear subvariety V' of $\boldsymbol{P}_n$ such that $\pi = \psi(\pi'^p)$ and $V = \psi(V'^p)$.

If V is defined over $\mathbf{F}_q$, then either

(i) *there is a proper linear subvariety W of V, also defined over K, with*

$$h(W) \leqslant 8n^5 4^n d\delta(n+r)R(\sqrt{G})^2,$$

such that π lies in $W(\sqrt{G})$,

or

(iii) *there is a point π' in $\mathbf{P}_n(\sqrt{G})$ with $\pi = \pi'^p$.*

Proof. Suppose first that V has dimension $n-1$. Then we just have to follow the arguments of the proof of Masser [**22**, Lemma 5 (p. 197)]. Because these arguments are expressed in terms of 'broad sets' and this notion is no longer appropriate, we write out all the details.

Because V is transversal, we may work affinely with a point $\pi = (x_1, \ldots, x_n)$ satisfying a single equation

$$a_1x_1 + \ldots + a_nx_n = 1 \tag{8.1}$$

with non-zero coefficients. As in Section 3, write C for the field of pth powers in K, and consider

$$s = \dim_C(Ca_1x_1 + \ldots + Ca_nx_n),$$

so that $1 \leqslant s \leqslant n$.

First suppose that $s = n$. Then we apply Lemma 5.1 with $k = \mathbf{F}_K$, $m = n$ and $c_1 = \ldots = c_m = 1$ and $g_1 = a_1x_1, \ldots, g_m = a_mx_m$. So the group must be enlarged by adjoining $a_1, \ldots, a_n$ to $\sqrt{G}$, becoming of rank at most $n+r$. The enlarged regulator R can be estimated by Lemma 4.2, and we find

$$R \leqslant 2^n h(a_1)\ldots h(a_n)R(\sqrt{G}) \leqslant 2^n h(V)^n R(\sqrt{G}). \tag{8.2}$$

The conclusion (b) of Lemma 5.1 is ruled out by $s = n$; and the conclusion (a) shows that

$$h(a_1x_1, \ldots, a_nx_n) \leqslant 4n^4 d\delta(n+r)R^2.$$

It follows that $h(\pi) = h(x_1, \ldots, x_n)$ is at most

$$4n^4 d\delta(n+r)R^2 + h(a_1^{-1}, \ldots, a_n^{-1}) \leqslant 4n^4 d\delta(n+r)R^2 + nh(V)$$

and so from (8.2) we deduce

$$h(\pi) \leqslant 4n^4 4^n d\delta(n+r)h(V)^{2n}R(\sqrt{G})^2 + nh(V) \leqslant 8n^4 4^n d\delta(n+r)h(V)^{2n}R(\sqrt{G})^2. \tag{8.3}$$

So this gives $W = \{\pi\}$ for (i) of the proposition; and for these $h(W) = h(\pi)$ is bounded as in (8.3).

Next suppose that $1 < s < n$. By means of a permutation we can assume that $g_1 = a_1x_1, \ldots, g_s = a_sx_s$ are linearly independent over C. Take any k with $s+1 \leqslant k \leqslant n$; then we can apply Lemma 5.2 with $m = s$ and $g_0 = a_kx_k$, $\sqrt{G}$ being enlarged as above. We find relations

$$\sum_{j=1}^{s} c_{kj}a_jx_j = a_kx_k, \quad k = s+1, \ldots, n \tag{8.4}$$

with c_{kj} in C and the quotients

$$f_{kj} = c_{kj}\frac{a_jx_j}{a_kx_k}, \quad j = 1, \ldots, s; \ k = s+1, \ldots, n \tag{8.5}$$

satisfying

$$h(f_{k1}, \ldots, f_{ks}) \leqslant 4s^4 d\delta(n+r)R^2, \quad k = s+1, \ldots, n \tag{8.6}$$

We use (8.4) to eliminate the a_kx_k $(k = s+1, \ldots, n)$ in (8.1). We find

$$c_1a_1x_1 + \ldots + c_sa_sx_s = 1 \tag{8.7}$$

with

$$c_j = 1 + \sum_{k=s+1}^{n} c_{kj}, \quad j = 1, \ldots, s \tag{8.8}$$

also in C.

Next apply Lemma 5.1 with $m = s$ to (8.7) and $g_j = a_j x_j$ $(j = 1, \ldots, s)$ also in the enlarged $\sqrt{G}$. Again conclusion (b) is impossible. It follows that the

$$f_j = c_j a_j x_j, \quad j = 1, \ldots, s \tag{8.9}$$

satisfy

$$h(f_1, \ldots, f_s) \leqslant 4s^4 d\delta(n+r)R^2. \tag{8.10}$$

So in (8.5) certain quotients x_j/x_k are bounded modulo C whereas in (8.9) certain x_j themselves are bounded modulo C. We can eliminate C by substituting (8.8) into (8.9) and using (8.5) to obtain

$$f_j = a_j x_j + \sum_{k=s+1}^{n} f_{kj} a_k x_k, \quad j = 1, \ldots, s. \tag{8.11}$$

Since $a_j \neq 0$ $(j = 1, \ldots, s)$ these express the fact that $\pi = (x_1, \ldots, x_n)$ lies on a linear variety V' of dimension $n - s$; and because $s \neq 1$ this dimension is strictly less than the dimension $n - 1$ of V. So we can take W as the intersection of V' with V. This is in fact V' because if we add up all the above equations (8.11) and use (8.4), (8.5), (8.7) and (8.9), then we end up with (8.1).

Now we have to estimate the height of (8.11). In the corresponding matrix, every column has by (8.6) and (8.10) height at most $4s^4 d\delta(n+r)R^2 + h(V)$, which as above in (8.3) we can estimate by $B = 8n^4 4^n d\delta(n+r)h(V)^{2n} R(\sqrt{G})^2$. It follows that

$$h(W) \leqslant sB \leqslant 8n^5 4^n d\delta(n+r)h(V)^{2n} R(\sqrt{G})^2.$$

This too settles (i) of the proposition.

Finally suppose $s = 1$. This means that $a_1 x_1, \ldots, a_n x_n$ are in C. By Lemma 4.4 with $l = p$ we can write $x_j = g_j x_j'^p$ with g_j, x_j' in $\sqrt{G}$ $(j = 1, \ldots, n)$ and

$$h(g_j) \leqslant p\delta(r)R(\sqrt{G})^2 \leqslant p\delta(n+r)R(\sqrt{G})^2, \quad j = 1, \ldots, n.$$

Then $a_j g_j$ is in C so has the form $a_j'^p$ $(j = 1, \ldots, n)$. Finally,

$$1 = a_1 x_1 + \ldots + a_n x_n = a_1'^p x_1'^p + \ldots + a_n'^p x_n'^p = (a_1' x_1' + \ldots + a_n' x_n')^p,$$

and this gives part (ii) of the proposition, with ψ as in (1.7) above for $g_0 = 1$, $\pi' = (x_1', \ldots, x_n')$, and V' defined by (8.1) above with the new coefficients $a_1', \ldots, a_n'$.

This proves the proposition when V has dimension $n - 1$. Incidentally, when the coefficients in (8.1) are in some $\mathbf{F}_q$, then the argument for $s = 1$ shows that $x_1, \ldots, x_n$ are in C. So they are pth powers $x_1'^p, \ldots, x_n'^p$; and clearly $x_1', \ldots, x_n'$ are in $\sqrt{G}$. Thus, we obtain the conclusion (iii) of the proposition when V has dimension $n - 1$. And the case $s \neq 1$ leads of course to (i). So it remains only to treat V of dimension $m - 1 < n - 1$ defined over some $\mathbf{F}_q$.

This we do by expressing the affine equations of V in triangular form, which after a permutation we can suppose are

$$x_i = a_{i0} + a_{i1} x_1 + \ldots + a_{i,m-1} x_{m-1}, \quad i = m, m+1, \ldots, n \tag{8.12}$$

with the a_{ij} in $\mathbf{F}_q$. This gives $V = V_m \cap \ldots \cap V_n$ for the varieties defined individually by each equation.

Consider the first equation. There may be some zero coefficients a_{mj}, but not all are zero, because $V(\sqrt{G})$ is non-empty. In fact at least two are non-zero otherwise V would be contained in a coset $T \neq \mathbf{P}_n$ contrary to our assumption. We can thus regard V_m as a transversal variety of codimension 1 in some projective space of dimension at least 2 and at most $m < n$. Applying the proposition for the cases already proved, we obtain two possibilities (i) and (iii). If (i) holds for V_m, then we obtain a proper subvariety W_m of V_m with

$$h(W_m) \leqslant 8n^5 4^n d\delta(n+r)R(\sqrt{G})^2. \tag{8.13}$$

But it is not difficult to see that each W_m intersects the remaining intersection $U_m = \bigcap_{i \neq m} V_i$ in a proper subspace of $V = V_m \cap U_m$. For example, the triangular nature of (8.12) makes it clear that $x_{m+1}, \ldots, x_n$ are determined by $x_1, \ldots, x_{m-1}$ on U_m, and then that x_m is determined by $x_1, \ldots, x_{m-1}$ on W_m in V_m; but also some non-zero polynomial of degree at most 1 in $x_1, \ldots, x_{m-1}$ must vanish on W_m. So $W = W_m \cap U_m$ has dimension strictly less than $m-1$. By Lemma 7.1, we have $h(W) \leqslant h(W_m)$. So by (8.13) we obtain (i) of the proposition for the original V. But what happens if (iii) holds for V_m?

This means that all the x_j actually occurring in the first equation of (8.12) are pth powers, which certainly goes some way in the direction of (iii) for V. But then we can try the second equation instead. Either we obtain a W as above, or all the x_j actually occurring in the second equation of (8.12) are pth powers. And so on. In the end, we either obtain W or that all the x_j actually occurring in all the equations (8.12) are pth powers. Because V is transversal this does give the full (iii) for V; and so completes the proof of the proposition. □

9. *The main estimate*

This is a quantitative version of our Descent Step over $\sqrt{G}$ without the requirement that the subvarieties W are isotrivial. This leads to a relatively small exponent attached to the height $h(V)$. As before $n \geqslant 2$, and we continue with our assumption that K is finitely generated and transcendental over $\mathbf{F}_p$, with separable transcendence basis $\mathcal{B}$ and $\mathbf{F}_K = \overline{\mathbf{F}_p} \cap K$; further G is finitely generated of rank $r \geqslant 1$ modulo $\mathbf{F}_K^*$.

Main estimate. Let V be a positive-dimensional linear subvariety of $\mathbf{P}_n$ defined over K but not a coset.

(a) If V is not $\sqrt{G}$-isotrivial, then

$$V(\sqrt{G}) = \bigcup_{W \in \mathcal{W}} W(\sqrt{G})$$

for a finite set $\mathcal{W}$ of proper linear subvarieties W of V, also defined over K and with

$$h(W) \leqslant 8n^2 d(10n^3\delta(n+r))^{2n+1} h(V)^{2n} R(\sqrt{G})^{6n+2}.$$

(b) If V is $\sqrt{G}$-isotrivial and $\psi(V)$ is defined over $\mathbf{F}_q$, then

$$V(\sqrt{G}) = \psi^{-1}\left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(\sqrt{G}))^{q^e}\right)$$

for a finite set $\mathcal{W}$ of proper linear subvarieties W of V, also defined over K and with

$$h(\psi(W)) \leqslant 8n^5 4^n (q/p) d\delta(n+r)R(\sqrt{G})^2.$$

Proof. We prove this first when V is transversal and not contained in any coset $T \neq \mathbf{P}_n$.

We start with $\sqrt{G}$-isotrivial V. Because we estimate $h(\psi(W))$ and not $h(W)$, it clearly suffices to assume that ψ is the identity, so that V is defined over $\mathbf{F}_q$. Take arbitrary π in $V(\sqrt{G})$ not in $V(\mathbf{F}_K)$. Then either (i) or (iii) of the proposition holds.

If (i) holds, then (*b*) looks good with $e = 0$ (and ψ the identity); at least π lies in some $W(\sqrt{G})$ for a proper subvariety W of V, defined over K, with

$$h(W) \leqslant 8n^5 4^n d\delta(n+r)R(\sqrt{G})^2. \tag{9.1}$$

What if (iii) holds? Now any a in $\mathbf{F}_q$ has a unique pth root $a^{1/p}$ in $\mathbf{F}_q$, which is also a conjugate of a over $\mathbf{F}_p$. We obtain a new point π' in $V'(\sqrt{G})$, also not in $V'(\mathbf{F}_K)$, for a new variety V' in $\mathbf{P}_n$ which is a conjugate of V. The new variety has the same dimension as V, and is also defined over $\mathbf{F}_q$. So we can repeat the process, and again we obtain either (i) or (iii) of the proposition.

If (i) holds, then π' lies in some $W'(\sqrt{G})$ again with W' over K and $h(W')$ bounded as in (9.1). So π lies in $(W'(\sqrt{G}))^p$ as in (b) with $e = 1$.

Or if (iii) holds, then we obtain a new point π'' in $V''(\sqrt{G})$ for a new conjugate V'' of V in $\mathbf{P}_n$.

And so on, in a manner similar to the looping in the p-automata of [7, Section 4]. Because π was not in $V(\mathbf{F}_K)$, this procedure must eventually stop at some proper subvariety $W^{(L)}$ over K of $V^{(L)}$ (here the number L of repetitions might depend on π). Now the original point π lies in $(W^{(L)}(\sqrt{G}))^{p^L}$ with $h(W^{(L)})$ bounded as in (9.1).

Because π was arbitrary in $V(\sqrt{G})$ not in the finite set $V(\mathbf{F}_K)$, the conclusion so far is

$$V(\sqrt{G}) \subseteq \bigcup_{W \in \mathcal{W}} \bigcup_{L=0}^{\infty} (W(\sqrt{G}))^{p^L}$$

for a collection $\mathcal{W}$ of proper subvarieties W of conjugates of V defined over K and satisfying (9.1); here we may have to include single points W with $h(W) = 0$. To obtain equality, we write $q = p^f$ and $L = fe + l$ for $e \geqslant 0$ and $0 \leqslant l \leqslant f - 1$; this gives

$$V(\sqrt{G}) \subseteq \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \bigcup_{e=0}^{\infty} (\tilde{W}(\sqrt{G}))^{q^e}$$

with a new collection $\tilde{\mathcal{W}}$ of proper subvarieties $\tilde{W} = W^{p^l}$ of conjugates of V with

$$h(\tilde{W}) = p^l h(W) \leqslant 8n^5 4^n (q/p) d\delta(n+r) R(\sqrt{G})^2.$$

Finally by intersecting each $\tilde{W}$ with $V = V^q$ we can assume that each $\tilde{W}$ is a proper subvariety of V itself in the above, without increasing the height further. Because V is defined over $\mathbf{F}_q$, the $(\tilde{W}(\sqrt{G}))^{q^e}$ now lie in $(V(\sqrt{G}))^{q^e} = V(\sqrt{G})$, and so at last the two sides are equal. Now we have the desired (b); of course, the finiteness of the collection of $\tilde{W}$ follows from the Northcott Property already noted in Section 7. This settles the case of transversal $\sqrt{G}$-isotrivial V not contained in a proper coset.

Henceforth (until further notice) we will assume that V is not $\sqrt{G}$-isotrivial (and still transversal not contained in a proper coset).

Suppose first that V is a hyperplane. Take arbitrary π in $V(\sqrt{G})$. Then either (i) or (ii) of the proposition holds. We regard this dichotomy as the starting stage $l = 1$.

If (i) holds, then as before (a) of the Main Estimate looks good; at least π lies in some $W(\sqrt{G})$ for a proper subvariety W of V, defined over K, with

$$h(W) \leqslant Ch(V)^{2n} \tag{9.2}$$

for

$$C = 8n^5 4^n d\delta(n+r) R(\sqrt{G})^2. \tag{9.3}$$

What if (ii) holds? We obtain a new point π' in $V'(\sqrt{G})$ for a new variety V' in $\mathbf{P}_n$ with

$$\pi = \psi(\pi'^p), \quad V = \psi(V'^p). \tag{9.4}$$

Here, ψ is a $\sqrt{G}$-automorphism with

$$h(\psi) \leqslant pB \tag{9.5}$$

for

$$B = n\delta(n+r) R(\sqrt{G})^2. \tag{9.6}$$

This V' is also a hyperplane, and also not $\sqrt{G}$-isotrivial. So we can repeat the process, and again we obtain either (i) or (ii) of the proposition. This dichotomy is the next stage $l = 2$.

If (i) holds, then π' lies in some $W'(\sqrt{G})$. So π lies in $W(\sqrt{G})$ for $W = \psi(W'^p)$, almost as good as above, except that $h(W)$ could be larger than before. We take care of this later.

Or if (ii) holds, then we obtain a new point π'' in $V''(\sqrt{G})$ for a new variety V'' in $\mathbf{P}_n$.

And so on. At stage l, we obtain either $\pi^{(l-1)}$ in a proper subvariety $W^{(l-1)}$ of $V^{(l-1)}$ with

$$h(W^{(l-1)}) \leqslant Ch(V^{(l-1)})^{2n} \tag{9.7}$$

as in (9.2) and (9.3), or a new point $\pi^{(l)}$ in $V^{(l)}(\sqrt{G})$ for a new variety $V^{(l)}$ with

$$\pi^{(l-1)} = \psi^{(l-1)}((\pi^{(l)})^p), \quad V^{(l-1)} = \psi^{(l-1)}((V^{(l)})^p) \tag{9.8}$$

as in (9.4), for

$$h(\psi^{(l-1)}) \leqslant pB \tag{9.9}$$

as in (9.5) and (9.6).

We claim that this procedure must eventually stop because V is not $\sqrt{G}$-isotrivial, and after a certain number L of repetitions which this time is independent of π. Actually let us define the integer $L_0 \geqslant 0$ by

$$p^{L_0} \leqslant 2h(V)R(\sqrt{G}) < p^{L_0+1}. \tag{9.10}$$

From (9.8) we obtain $V = \psi_l((V^{(l)})^{p^l})$ with the $\sqrt{G}$-automorphism

$$\psi_l = \psi\psi'^p \dots (\psi^{(l-1)})^{p^{l-1}}. \tag{9.11}$$

Writing the hyperplane V in the affine form (8.1), we know that some coefficient $x = a_j \neq 0$ does not lie in $\sqrt{G}$, and $x = gy^{p^l}$ for some g in $\sqrt{G}$ and some y in K. We can now apply Lemma 4.5, because $\sqrt{G_k}$ there is just $\sqrt{G}$. We conclude that

$$p^l \leqslant 2h(x)R(\sqrt{G}) \leqslant 2h(V)R(\sqrt{G}).$$

In view of (9.10) this means that (ii) cannot hold for $l = L_0 + 1$. Thus, there is some L with $0 \leqslant L \leqslant L_0$ such that (ii) holds at stages $l = 1, \dots, L$ (at least if $L \geqslant 1$), and then (i) holds at stage $l = L + 1$. We conclude that $\pi^{(L)}$ lies in $W^{(L)}$, and from (9.7)

$$h(W^{(L)}) \leqslant Ch(V^{(L)})^{2n}. \tag{9.12}$$

Thus, $\pi = \psi_L((\pi^{(L)})^{p^L})$ lies in $W = \psi_L((W^{(L)})^{p^L})$. By (7.1) and (9.11) we obtain

$$h(W) \leqslant p^L h(W^{(L)}) + nh(\psi_L) \leqslant p^L h(W^{(L)}) + n(h(\psi) + ph(\psi') + \dots + p^{L-1}h(\psi^{(L-1)})),$$

which using (9.9) and (9.12) yields

$$h(W) \leqslant Cp^L h(V^{(L)})^{2n} + 2np^L B \leqslant C(p^L h(V^{(L)})^{2n} + 2np^L B. \tag{9.13}$$

To estimate $h(V^{(L)})$, we use (7.1), (9.8) and (9.9) to obtain

$$ph(V^{(l)}) = h((\psi^{(l-1)})^{-1}V^{(l-1)}) \leqslant h(V^{(l-1)}) + n^2 h(\psi^{(l-1)}) \leqslant h(V^{(l-1)}) + n^2 pB.$$

If $L \geqslant 1$, then we multiply this by p^{l-1} and sum from $l = 1$ to $l = L$, obtaining $p^L h(V^{(L)}) \leqslant h(V) + 2n^2 p^L B$ (which holds also if $L = 0$). Inserting this into (9.13) we obtain

$$h(W) \leqslant C(h(V) + 2n^2 p^L B)^{2n} + 2np^L B \leqslant 2C(h(V) + 2n^2 p^L B)^{2n},$$

and then using (9.6) and (9.10) with $L \leqslant L_0$ we find

$$h(W) \leqslant 2Ch(V)^{2n}(1 + 4n^3\delta(n+r)R(\sqrt{G})^3)^{2n} \leqslant 2Ch(V)^{2n}(5n^3\delta(n+r)R(\sqrt{G})^3)^{2n}$$

From (9.3), we obtain finally

$$h(W) \leqslant C'h(V)^{2n}R(\sqrt{G})^{6n+2} \tag{9.14}$$

with

$$C' = 16n^5 4^n d\delta(n+r)(5n^3\delta(n+r))^{2n} \leqslant 2n^2 d(10n^3\delta(n+r))^{2n+1}.$$

Because π was arbitrary, the conclusion so far is

$$V(\sqrt{G}) \subseteq \bigcup_{W\in\mathcal{W}} W(\sqrt{G})$$

for a finite collection $\mathcal{W}$ of proper subvarieties W of V satisfying (9.14). But then the two sides are of course equal. This settles the Main Estimate for transversal hyperplanes V that are not $\sqrt{G}$-isotrivial and not contained in a proper coset.

Next suppose that V, still not $\sqrt{G}$-isotrivial (and still transversal not contained in a proper coset), has dimension $m-1$ for some $m<n$. So after a permutation of variables it can be defined by equations (6.1). Each of these equations defines a hyperplane V_i, so that $V = V_m \cap \ldots \cap V_n$.

We claim that we can assume that all non-zero $a(i,j)$ lie in $\sqrt{G}$. Otherwise, for example, V_m is transversal and not $\sqrt{G}$-isotrivial in the projective space with coordinates X_j corresponding to $j=m$ and the j with $a(m,j)\neq 0$. Since no $X_m - aX_j$ $(m\neq j, a\neq 0)$ vanishes on V, this projective space has dimension at least 2. So then we could apply the hyperplane result (9.14) to deduce that all solutions lie in a finite union of proper subspaces W_m of this V_m with

$$h(W_m) \leqslant C' h(V_m)^{2n} R(\sqrt{G})^{6n+2}.$$

But as in the affine situation just after (8.13), it can be seen that W_m intersects the remaining intersection $U_m = \bigcap_{i\neq m} V_i$ in a proper subspace of $V = V_m \cap U_m$. For example, the triangular nature of (6.1) makes it clear that $X_{m+1},\ldots,X_n$ are determined by $X_0,\ldots,X_{m-1}$ on U_m, and then that X_m is determined by $X_0,\ldots,X_{m-1}$ on W_m in V_m; but also some non-zero linear form in $X_0,\ldots,X_{m-1}$ must vanish on W_m. Therefore, $W = W_m \cap U_m$ has dimension strictly less than $m-1$. So we are indeed in a proper subspace as required by (a) of the Main Estimate. Further, $W = W_m \cap V$ and so $h(W) \leqslant h(W_m) + h(V)$ by Lemma 7.1; moreover, $h(V_m) \leqslant h(V)$ because the $a(m,j)$ are themselves among the Grassmannian coordinates of V. We end up with (9.14) with say an extra factor 2.

So indeed from now on we can assume that all non-zero $a(i,j)$ in (6.1) lie in $\sqrt{G}$. This means that we are set up to apply Lemma 6.1. We will see that the effect is to pass to a proper subvariety of at least one of $V_m,\ldots,V_n$ despite their being separately isotrivial. As V is not $\sqrt{G}$-isotrivial by assumption, we find some quotient (6.2), say Q, not lying in $\mathbf{F}_K$. Let $\pi = (\xi_0,\ldots,\xi_n)$ be any point of $V(\sqrt{G})$. For a typical factor $a(i,j)/a(i,j')$ in Q we apply part (b) of the Main Estimate in lower dimensions to V_i, with ψ_i determined by 1 and the non-zero $a(i,j)$. So here $q=p$. We find finitely many proper subspaces W_i of V_i such that $\psi_i(V_i(\sqrt{G}))$ lies in the union of the $\bigcup_{e=0}^{\infty}(\psi_i(W_i)(\sqrt{G}))^{p^e}$, with

$$h(\psi_i(W_i)) \leqslant 8n^5 4^n d\delta(n+r) R(\sqrt{G})^2 \tag{9.15}$$

(now independent of p). In particular, writing π_i for the projection of π to the lower-dimensional space, we have equations

$$\psi_i(\pi_i) = \sigma_i^{q_i} \tag{9.16}$$

for σ_i in some $\psi_i(W_i)$ and some power q_i of p. Thus, $a(i,j)\xi_j/a(i,j')\xi_{j'} = \eta^{q_i}$ for certain $\eta = \eta(i,j,j')$ in K^*. Multiplying all these over the factors in (6.2) we find $Q = \eta_1^{q_1}\ldots\eta_k^{q_k}$ for certain $\eta_1,\ldots,\eta_k$ in K^*. Because the fixed Q is not in $\mathbf{F}_K$, this forces $q = \min\{q_1,\ldots,q_k\}$ to be bounded above by some quantity depending only on V. In fact $h(Q)\geqslant q$, but on the other hand from (6.2) we see that $h(Q) \leqslant (n+1)h(V)$. Thus,

$$q \leqslant (n+1)h(V). \tag{9.17}$$

Say this minimum is $q = q_i$. Now (9.16) says that π_i and so π lies in the variety $U = \psi_i^{-1}(\psi_i(W_i))^q$ of a dimension strictly less than the dimension of V_i. This intersects V_i in a proper subvariety W_i' of V_i. Once more this W_i' intersects the remaining intersection $\bigcap_{i' \neq i} V_{i'}$ in a proper subvariety W of V. As for heights, we have $W = W_i' \cap V$ so $h(W) \leqslant h(W_i') + h(V)$. Also $h(W_i') \leqslant h(U) + h(V_i) \leqslant h(U) + h(V)$, and also

$$h(U) \leqslant qh(\psi_i(W_i)) + nh(\psi_i^{-1}) \leqslant qh(\psi_i(W_i)) + n^2 h(V_i)$$

because of the definition of ψ_i. Putting these together and using (9.15) and (9.17), we conclude that

$$h(W) \leqslant 8n^5(n^2+n+3)4^n d\delta(n+r)h(V)R(\sqrt{G})^2.$$

This is much smaller than (9.14), and so we have completed the proof of the Main Estimate when V is transversal and not contained in a proper coset. In case (a) we have reached so far the bound $h(W) \leqslant Ah(V)^{2n}R^{6n+2}$ with $R = R(\sqrt{G})$ and $A = 4n^2 d(10n^3\delta(n+r))^{2n+1}$ due to the extra factor 2 encountered after establishing (9.14).

To treat the more general situation when V is transversal and not itself a coset, we use induction on $n \geqslant 2$, and we will obtain in case (a) the slightly weaker result $h(W) \leqslant Ah(V)^{2n}R^{6n+2} + nh(V)$. This leads at once to the bound given in the Main Estimate.

If $n = 2$, then there is a single equation $a_0X_0 + a_1X_1 + a_2X_2 = 0$, and transversality implies all $a_i \neq 0$. Thus, no $X_i - aX_j$ $(i \neq j, a \neq 0)$ vanishes on V, and we are done. Thus, we can suppose that $n \geqslant 3$.

After permuting the variables, we can suppose that $X_n - aX_{n-1}$ $(a \neq 0)$ vanishes on V. In the remaining equations for V we may eliminate X_n to obtain a linear variety $\tilde{V}$ in $\mathbf{P}_{n-1}$. This $\tilde{V}$ cannot be a coset otherwise V would be. Also $\tilde{V}$ certainly involves the variables $X_0, \ldots, X_{n-2}$ and so is transversal in $\mathbf{P}_{\tilde{n}}$ for $\tilde{n} = n-2$ or $\tilde{n} = n-1$. Here $\tilde{n} \geqslant 2$ unless $n = 3$; but in that case if $\tilde{V}$ is not transversal in $\mathbf{P}_2$ then V would be defined by equations $X_3 = aX_2$ and $b_0X_0 + b_1X_1 = 0$ so would be a coset. Thus, we can assume that $\tilde{V}$ is transversal in $\mathbf{P}_{\tilde{n}}$ with $\tilde{n} \geqslant 2$.

Suppose first that V is not $\sqrt{G}$-isotrivial as in (a). Then $\tilde{V}$ cannot be $\sqrt{G}$-isotrivial otherwise we could transform X_n to make V isotrivial. Thus by induction the Main Estimate holds for $\tilde{V}$. It is now relatively straightforward to deduce the Main Estimate for V. Thus by case (a) for $\tilde{V}$ we obtain

$$\tilde{V}(\sqrt{G}) = \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \tilde{W}(\sqrt{G}) \tag{9.18}$$

for a finite set $\tilde{\mathcal{W}}$ of proper linear subvarieties $\tilde{W}$ of $\tilde{V}$, also defined over K and with $h(\tilde{W}) \leqslant Ah(\tilde{V})^{2n}R^{6n+2} + (n-1)h(\tilde{V})$. Now we will check that (a) for V follows with W defined by the equations of $\tilde{W}$ together with $X_n = aX_{n-1}$. First the upper bound of Lemma 7.1 gives

$$h(W) \leqslant h(\tilde{W}) + h(a) \leqslant Ah(\tilde{V})^{2n}R^{6n+2} + (n-1)h(\tilde{V}) + h(a). \tag{9.19}$$

We can suppose $X_{n-1} \neq 0$ on $\tilde{V}$, else (9.18) would be empty; and so the lower bound of Lemma 7.1 gives $h(V) \geqslant \max\{h(\tilde{V}), h(a)\}$. Therefore (9.19) implies

$$h(W) \leqslant Ah(V)^{2n}R^{6n+2} + nh(V)$$

as required.

And in case (b) for $\sqrt{G}$-isotrivial V (assuming as above that ψ is the identity) we see that $\tilde{V}$ is $\sqrt{G}$-isotrivial and a lies in $\mathbf{F}_q$. We obtain (b) for V from (b) for $\tilde{V}$ using the analogue $\tilde{V}(\sqrt{G}) = \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \bigcup_{e=0}^{\infty} (\tilde{W}(\sqrt{G}))^{q^e}$ of (9.18) with as above W defined by the equations of $\tilde{W}$ together with $X_n = aX_{n-1}$; now $h(W) \leqslant h(\tilde{W})$.

What if V is not transversal (and of course still not a coset)? Then it is transversal (and still not a coset) in some projective subspace of dimension $n' \leqslant n-1$. Here $n' \geqslant 2$; otherwise it would be a coset. The above cases (a) and (b) in dimension n' now lead immediately to the same cases in $\mathbf{P}_n$; we have merely ignored $n - n'$ projective variables that were never in the equations anyway.

This finally finishes the proof of the Main Estimate. $\square$

In view of the fact that the estimate in case (a) is independent of the characteristic p, it may seem a nuisance that the estimate in case (b) depends on p. But actually this is unavoidable, and there are even examples to show that the full q/p is needed. To see this, take any power $q > 1$ of p, and define $K = \mathbf{F}_q(t)$ with $G = \sqrt{G}$ generated by t and $1 - t$ and a generator ζ of $\mathbf{F}_q^*$. Here, we have $r = 2, R(\sqrt{G}) = \sqrt{3}$ and, with the obvious transcendence basis, $d = 1$. The affine equations

$$x + y = 1, \quad x + \zeta z = 1$$

give rise to a $\sqrt{G}$-isotrivial line V (with $h(V) = 0$ and ψ the identity), and an upper bound B in (b) would mean that all solutions over $\sqrt{G}$ are given by $w, w^q, w^{q^2}, \ldots$ for some w with $h(w) \leqslant B$. Thus every solution π would have either $h(\pi) \leqslant B$ or $h(\pi) \geqslant q$. But

$$\pi = (x, y, z) = \left((1-t)^{q/p}, t^{q/p}, \frac{t^{q/p}}{\zeta}\right)$$

is a solution with $h(\pi) = q/p$. It follows that $B \geqslant q/p$.

10. *Isotrivial W*

We show here how to ensure that all the subvarieties W in the Main Estimate can be made $\sqrt{G}$-isotrivial, at the expense of enlarging the exponents in the upper bounds for their heights. To simplify the various expressions, we abbreviate the factors in case (a) of the Main Estimate by

$$\Delta = \Delta(n, r, d) = 8n^2 d(10n^3 \delta(n + r))^{2n+1} \geqslant 1, \quad h = h(V), \quad R = R(\sqrt{G}), \tag{10.1}$$

and that in case (b) of the Main Estimate by

$$\Psi = \Psi(n, r, d, p, q) = 8n^5 4^n (q/p) d\delta(n + r) \geqslant 1. \tag{10.2}$$

We also define some exponents

$$\rho(m) = \rho_n(m) = \frac{(2n)^m - 1}{2n - 1}, \quad \eta(m) = \eta_n(m) = (2n)^m, \ \ m = 1, 2, \ldots$$

Main Estimate for isotrivial W. Let V be a linear subvariety of $\mathbf{P}_n$ defined over K but not a coset, with dimension $m - 1 \geqslant 1$.

(a) If V is not $\sqrt{G}$-isotrivial, then

$$V(\sqrt{G}) = \bigcup_{W \in \mathcal{W}} W(\sqrt{G})$$

for a finite set $\mathcal{W}$ of proper linear $\sqrt{G}$-isotrivial subvarieties W of V, also defined over K and with

$$h(W) \leqslant (\Delta R^{6n+2})^{\rho(m)} h^{\eta(m)}. \tag{10.3}$$

(b) If V is $\sqrt{G}$-isotrivial and $\psi(V)$ is defined over $\mathbf{F}_q$, then

$$V(\sqrt{G}) = \psi^{-1}\left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(\sqrt{G}))^{q^e}\right)$$

for a finite set $\mathcal{W}$ of proper linear $\sqrt{G}$-isotrivial subvarieties W of V, also defined over K and with

$$h(\psi(W)) \leqslant (\Delta R^{6n+2})^{\rho(m-1)} (\Psi R^2)^{\eta(m-1)}.$$

Proof. We start with case (a), and now we can write the bound as

$$h(W) \leqslant \Delta h^{2n} R^{6n+2} \tag{10.4}$$

with W not necessarily $\sqrt{G}$-isotrivial. We show by induction on the dimension $m-1 \geqslant 1$ of V that the increased bound

$$h(\tilde{W}) \leqslant (\Delta R^{6n+2})^{\rho(m)} h^{\eta(m)} \tag{10.5}$$

as in (10.3) holds where now all the $\tilde{W}$ are $\sqrt{G}$-isotrivial.

When $m=2$, then the W are points and so automatically $\sqrt{G}$-isotrivial as long as $W(\sqrt{G})$ is non-empty.

When $m \geqslant 3$, we are fine unless some W is not $\sqrt{G}$-isotrivial. We observe that such a W cannot be a coset T. For the latter is defined by finitely many $X_i = a_{ij} X_j$ ($a_{ij} \neq 0$), and if $T(\sqrt{G})$ is non-empty, then clearly each a_{ij} lies in $\sqrt{G}$. But now it is easy to see that T is $\sqrt{G}$-isotrivial after all. For example, we can rewrite the equations as $a_i X_i = a_j X_j$ with a_i, a_j in $\sqrt{G}$. Then we can set up an equivalence relation on $\{0, 1, \ldots, n\}$ characterized by the equivalence of such i and j . And now we need change only the variables in the equivalence classes of cardinality at least 2 in order to trivialize T.

So by induction each of these W satisfies

$$W(\sqrt{G}) = \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \tilde{W}(\sqrt{G})$$

with $\sqrt{G}$-isotrivial $\tilde{W}$ such that

$$h(\tilde{W}) \leqslant (\Delta R^{6n+2})^{\rho(m-1)} h(W)^{\eta(m-1)}.$$

Therefore, all we have to do is to substitute (10.4) into this. We find the upper bound (10.5) because

$$\rho(m-1) + \eta(m-1) = \rho(m), \quad 2n\eta(m-1) = \eta(m).$$

For case (b) we write the bound as

$$h(\psi(W)) \leqslant \Psi R^2 \tag{10.6}$$

with W not necessarily $\sqrt{G}$-isotrivial. If some W is not $\sqrt{G}$-isotrivial, then neither is $\psi(W)$, and we can write

$$\psi(W)(\sqrt{G}) = \bigcup_{W^* \in \mathcal{W}^*} W^*(\sqrt{G})$$

with $\sqrt{G}$-isotrivial W^* such that

$$h(W^*) \leqslant (\Delta R^{6n+2})^{\rho(m-1)} h(\psi(W))^{\eta(m-1)}. \tag{10.7}$$

Now we can see (without induction) that the bound

$$h(\psi(\tilde{W})) \leqslant (\Delta R^{6n+2})^{\rho(m-1)} (\Psi R^2)^{\eta(m-1)} \tag{10.8}$$

holds, where now all the $\tilde{W} = \psi^{-1}(W^*)$ are $\sqrt{G}$-isotrivial. In fact just as above, all we have to do is to substitute (10.6) into (10.7), and we find at once (10.8). This completes the proof. $\square$

11. *Points over G*

We show here how to replace $V(\sqrt{G})$ and $W(\sqrt{G})$ in the Main Estimate by $V(G)$ and $W(G)$ at the expense of worsening the dependence on the regulator. However, we no longer insist that the W are isotrivial. If needed, this could be secured just by repeating the arguments of the previous section. We retain the notations (10.1), (10.2) from that section. Of course $n \geqslant 2$, and we continue with our assumption that K is finitely generated over $\mathbf{F}_p$, with $\mathbf{F}_K = \overline{\mathbf{F}_p} \cap K$; further G is finitely generated of rank $r \geqslant 1$ modulo $\mathbf{F}_K^*$.

Main Estimate for points over G. There is a positive integer $f = f_K(G) \leqslant [\sqrt{G} : G]$, depending only on K and G, with the following property. Let V be a positive-dimensional linear subvariety of $\mathbf{P}_n$ defined over K but not a coset.

(a) If V is not $\sqrt{G}$-isotrivial, then

$$V(G) = \bigcup_{W \in \mathcal{W}} W(G)$$

for a finite set $\mathcal{W}$ of proper linear subvarieties W of V, also defined over K and with

$$h(W) \leqslant \Delta h^{2n} R(\sqrt{G})^{6n+2}.$$

(b) If V is $\sqrt{G}$-isotrivial and $\psi(V)$ is defined over $\mathbf{F}_q$, then either

(ba) we have

$$V(G) = \bigcup_{W \in \mathcal{W}} W(G)$$

for a finite set $\mathcal{W}$ of proper linear subvarieties W of V, also defined over K and with

$$h(\psi(W)) \leqslant |\mathbf{F}_K| \Psi R(G)^2$$

or

(bb) we have

$$V(G) = \psi^{-1}\left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(G))^{q^{fe}}\right) \tag{11.1}$$

for a finite set $\mathcal{W}$ of proper linear subvarieties W of V, also defined over K and with

$$h(\psi(W)) \leqslant q^f |\mathbf{F}_K| \Psi R(G)^2. \tag{11.2}$$

We need first a simple remark about congruences. Here ϕ is the Euler function.

LEMMA 11.1. *For a given power $Q > 1$ of a prime P consider a finite collection of congruence equations*

$$LQ^e \equiv M \bmod N \tag{11.3}$$

with N taken from a finite set $\mathcal{N}$ of positive integers and L and M taken from $\mathbf{Z}$. Suppose that the set of solutions $e \geqslant 0$ is non-empty. Then if there is some $M \neq 0$ with $\mathrm{ord}_P M < \mathrm{ord}_P N$ this set is

(a) *finite with $Q^e \leqslant \max_{N \in \mathcal{N}} N$,*
otherwise

(b) *a finite union of arithmetic progressions $e = e_0, e_0 + f, e_0 + 2f, \ldots$ with $f = \prod_{N \in \mathcal{N}} \phi(N)$ and $Q^{e_0} < Q^f \max_{N \in \mathcal{N}} N$.*

Proof. Suppose first that there is some $M \neq 0$ with $\mathrm{ord}_P M < \mathrm{ord}_P N$. Then the corresponding $L \neq 0$, and we obtain

$$e \,\mathrm{ord}_P Q \leqslant \mathrm{ord}_P LQ^e = \mathrm{ord}_P M < \mathrm{ord}_P N$$

giving case (a).

Thus we can assume that $\mathrm{ord}_P M \geqslant \mathrm{ord}_P N$ whenever $M \neq 0$. We proceed to verify case (b). Now the congruences (11.3) can be split into congruences modulo powers of P and congruences modulo powers $\tilde{P}^m$ of other primes $\tilde{P} \neq P$.

The former congruences, if any, will be satisfied as soon as e is sufficiently large. Indeed they amount to $LQ^e \equiv 0 \bmod P^{\mathrm{ord}_P N}$ and so conditions $e \geqslant \lambda$ for various real $\lambda \leqslant \mathrm{ord}_P N/\mathrm{ord}_P Q$; that is, $Q^\lambda \leqslant P^{\mathrm{ord}_P N} \leqslant N$. Thus, together, they give a single condition $e \geqslant \Lambda$ for some real Λ with $Q^\Lambda \leqslant \max_{N \in \mathcal{N}} N$.

We note that whether e satisfies the other congruences depends only on its congruence class modulo f. For if $\tilde{P}^m$ divides some N, then $\phi(\tilde{P}^m)$ divides $\phi(N)$ which divides f, and so $Q^f \equiv 1 \bmod \tilde{P}^m$.

Thus the solutions e satisfy $e \geqslant \Lambda$ and also must lie in a finite number of arithmetic progressions modulo f. If e_0 is the smallest member of one of these progressions with $e_0 \geqslant \Lambda$, then $e_0 - f < \Lambda$ and this leads to case (b), thereby completing the proof. $\square$

We can now start on the proof of the Main Estimate for points over G.

Suppose first that V is not $\sqrt{G}$-isotrivial. Then (a) of the Main Estimate gives

$$V(\sqrt{G}) = \bigcup_{W \in \mathcal{W}} W(\sqrt{G})$$

for W satisfying (10.4). Now we can descend to G simply by intersecting with $\mathbf{P}_n(G)$.

Next suppose that V is $\sqrt{G}$-isotrivial and $\psi(V)$ is defined over $\mathbf{F}_q$. Using elementary divisors we can find generators $\gamma_1, \ldots, \gamma_r$ of $\sqrt{G}$ modulo constants and positive integers $d_1, \ldots, d_r$ such that $\gamma_1^{d_1}, \ldots, \gamma_r^{d_r}$ generate G modulo constants. The constants can be taken care of with an extra γ_0 generating $\sqrt{G} \cap \mathbf{F}_K$ and $\gamma_0^{d_0}$ generating $G \cap \mathbf{F}_K$; here d_0 divides the order of γ_0 as a root of unity. Thus,

$$[\sqrt{G} : G] = d_0 d_1 \ldots d_r. \tag{11.4}$$

We write

$$\psi(X_0, \ldots, X_n) = (\psi_0 X_0, \ldots, \psi_n X_n)$$

with

$$\psi_i = \gamma_0^{a_{0i}} \gamma_1^{a_{1i}} \ldots \gamma_r^{a_{ri}}, \quad i = 0, \ldots, n \tag{11.5}$$

in $\sqrt{G}$. Now (b) of the Main Estimate gives

$$V(\sqrt{G}) = \psi^{-1} \left(\bigcup_{W \in \mathcal{W}} \bigcup_{e=0}^{\infty} (\psi(W)(\sqrt{G}))^{q^e} \right) \tag{11.6}$$

for W satisfying (10.6). But we can no longer descend to G simply by intersecting with $\mathbf{P}_n(G)$.

Consider a point $\pi = (\pi_0, \ldots, \pi_n)$ of $V(G)$. By (11.6), there is a point $\sigma = (\sigma_0, \ldots, \sigma_n)$ in some $W(\sqrt{G})$ and some $e \geqslant 0$ such that $\pi = \psi^{-1}(\psi(\sigma))^{q^e}$. As in (11.5), we write

$$\sigma_i = \gamma_0^{b_{0i}} \gamma_1^{b_{1i}} \ldots \gamma_r^{b_{ri}}, \quad (i = 0, \ldots, n), \tag{11.7}$$

however π is over G and so

$$\pi_i = \gamma_0^{c_{0i} d_0} \gamma_1^{c_{1i} d_1} \ldots \gamma_r^{c_{ri} d_r}, \quad (i = 0, \ldots, n).$$

Equating exponents we find a system of congruences

$$(a_{ji} + b_{ji}) q^e \equiv a_{ji} \bmod d_j, \quad i = 0, \ldots, n; \; j = 0, 1, \ldots, r \tag{11.8}$$

depending only on σ. We can apply Lemma 11.1, and the argument splits into two according to the conclusion. As the b_{ji} in (11.7) appear only in the coefficients L, the splitting is independent of σ.

Suppose first that Lemma 11.1(a) holds. Then

$$q^e \leqslant \max\{d_0, d_1, \ldots, d_r\} \leqslant d_0 d_1 \ldots d_r = [\sqrt{G} : G] \tag{11.9}$$

by (11.4). Now π lies in the finitely many $\tilde{W} = \psi^{-1}(\psi(W))^{q^e}$, which we can put together into a set $\tilde{\mathcal{W}}$, and then we have shown that

$$V(G) \subseteq \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \tilde{W}(\sqrt{G}).$$

Now intersecting with $\mathbf{P}_n(G)$ gives the same inclusion but with $\tilde{W}(G)$ on the right-hand side. On the other hand,

$$\tilde{W} = \psi^{-1}(\psi(W))^{q^e} \subseteq \psi^{-1}(\psi(V))^{q^e} = \psi^{-1}(\psi(V)) = V$$

because $\psi(V)$ is defined over $\mathbf{F}_q$. Thus, we conclude

$$V(G) = \bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \tilde{W}(G)$$

as in (ba) of the Main Estimate for points over G. But now from (11.9) and (10.6) the heights satisfy

$$h(\psi(\tilde{W})) = q^e h(\psi(W)) \leqslant d_0 d_1 \ldots d_r \Psi R(\sqrt{G})^2.$$

Using Lemma 4.1, we see that $R(G) = d_1 \ldots d_r R(\sqrt{G})$, and so we can absorb some terms into the regulator to obtain

$$h(\psi(\tilde{W})) \leqslant d_0 \Psi R(G)^2 \leqslant |\mathbf{F}_K| \Psi R(G)^2. \tag{11.10}$$

This completes the proof of (ba).

It remains only to suppose that Lemma 11.1(b) holds. Then we know that $e = e_0 + f\tilde{e}$ with $\tilde{e} \geqslant 0$ and e_0 bounded as in (11.9) but with an extra q^f. In particular, taking $\tilde{e} = 0$, we obtain a solution of (11.8) and this means that $\tilde{\sigma} = \psi^{-1}(\psi(\sigma))^{q^{e_0}}$ is also defined over G. It lies in

$$\tilde{W} = \psi^{-1}(\psi(W))^{q^{e_0}} \tag{11.11}$$

and so in $\tilde{W}(G)$. We also have

$$\psi(\pi) = (\psi(\sigma))^{q^e} = (\psi(\tilde{\sigma}))^{\tilde{q}^{\tilde{e}}}$$

for $\tilde{q} = q^f$. Thus, we conclude

$$V(G) \subseteq \psi^{-1}\left(\bigcup_{\tilde{W} \in \tilde{\mathcal{W}}} \bigcup_{\tilde{e}=0}^{\infty} (\psi(\tilde{W})(G))^{\tilde{q}^{\tilde{e}}} \right) \tag{11.12}$$

for the finite set $\tilde{\mathcal{W}}$ of $\tilde{W}$ in (11.11). On the other hand,

$$\psi(\tilde{W})^{\tilde{q}^{\tilde{e}}} = (\psi(W))^{q^{e_0}\tilde{q}^{\tilde{e}}} \subseteq (\psi(V))^{q^{e_0}\tilde{q}^{\tilde{e}}} = \psi(V)$$

again because $\psi(V)$ is defined over $\mathbf{F}_q$. Thus, we conclude equality in (11.12).

Finally, we calculate that $h(\psi(\tilde{W})) = q^{e_0} h(\psi(W))$ is bounded above by

$$q^f \max\{d_0, d_1, \ldots, d_r\} \Psi R(\sqrt{G})^2 \leqslant q^f |\mathbf{F}_K| \Psi R(G)^2 \tag{11.13}$$

as in (11.10), and of course $f = \phi(d_0)\phi(d_1)\ldots\phi(d_r)$ depends only on K and G with

$$f \leqslant d_0 d_1 \ldots d_r = [\sqrt{G} : G].$$

This completes the proof of (bb); and so the Main Estimate for points over G is proved. □

In (11.13), the term q^f cannot be so easily absorbed into the regulator without introducing an exponential dependence on $R(G)$. Let us discuss some aspects of this.

When $G = \sqrt{G}$ then $f = 1$ in (bb) and we are more or less back to (b) of the Main Estimate. But in general we need the extra f in (11.1). The following example shows that it sometimes must be almost as large as $[\sqrt{G} : G]$.

We go back to the equation $t^m x + y = 1$ of (1.5) over $K = \mathbf{F}_p(t)$, with $n = 2$. It is to be solved in the group $G = G_l$ generated by t^l and $1 - t$, so that $r = 2$. Here, $\sqrt{G}$ is generated by t and $1 - t$ together with a generator ζ of $\mathbf{F}_p^*$. The equation defines a $\sqrt{G}$-isotrivial line V with $\psi(x, y) = (t^m x, y) = (\tilde{x}, \tilde{y})$, so that $\tilde{V} = \psi(V)$ is defined by $\tilde{x} + \tilde{y} = 1$, with $q = p$.

Now Leitner [20] has found all points on $\tilde{V}(\sqrt{G})$. If p is odd, then there are $p - 2$ constant points in $\mathbf{F}_p^2$ together with six infinite families

$$(\tilde{x}, \tilde{y}) = (\tilde{x}_0^{p^{\tilde{e}}}, \tilde{y}_0^{p^{\tilde{e}}}), \quad \tilde{e} = 0, 1, \ldots,$$

where $(\tilde{x}_0, \tilde{y}_0)$ are given by

$$(t, 1-t), \quad (1-t, t), \quad \left(\frac{1}{t}, -\frac{1-t}{t}\right), \quad \left(-\frac{1-t}{t}, \frac{1}{t}\right), \quad \left(\frac{1}{1-t}, -\frac{t}{1-t}\right), \quad \left(-\frac{t}{1-t}, \frac{1}{1-t}\right).$$

The $(x, y) = \psi^{-1}(\tilde{x}, \tilde{y}) = (t^{-m}\tilde{x}, \tilde{y})$ are all the points on $V(\sqrt{G})$. Choosing m not divisible by l, we see that none of the constant points give rise to points of $V(G)$. Similarly for the second family above. And the same is true of the last four families above, simply because of the minus signs. However, the first family gives $(t^{-m} t^{p^{\tilde{e}}}, (1-t)^{p^{\tilde{e}}})$, which is in G^2 if and only if

$$p^{\tilde{e}} \equiv m \bmod l. \tag{11.14}$$

Now Artin's Conjecture implies that given any prime p, there are infinitely many primes l for which p is a primitive root modulo l. And Heath-Brown's Corollary 2 of [14, p. 27] implies that this is true for at least one of $p = 3, 5, 7$. We can choose m with $1 \leqslant m < l$ and $p^{l-2} \equiv m$ mod l. Now (11.14) implies $\tilde{e} \equiv l - 2 \mod l - 1$ so $\tilde{e} = l - 2 + (l - 1)e$ $(e = 0, 1, \ldots)$. Thus, the surviving points on $V(G)$ are just the

$$\pi = \psi^{-1}(\psi(W))^{p^{(l-1)e}}, \quad e = 0, 1, \ldots \tag{11.15}$$

with W as the single point $(t^{-m} t^{p^{l-2}}, (1-t)^{p^{l-2}})$. This makes it clear that $f \geqslant l - 1$ in (11.1); almost as big as $[\sqrt{G} : G] = (p - 1)l$ for fixed p.

We could also see this from (11.2). For as $R(G) = l\sqrt{3}$, it implies that there would be a point π on $V(G)$ with $h(\psi(\pi)) \leqslant cp^f l^2$ for c absolute. But the point (11.15) has $y = \tilde{y} = (1-t)^{p^{l-2} p^{(l-1)e}}$ so

$$h(\psi(\pi)) \geqslant p^{l-2} p^{(l-1)e} \geqslant p^{l-2}. \tag{11.16}$$

Making $l \to \infty$, we deduce $f \geqslant l - c' \log l$, also almost as big as $[\sqrt{G} : G] = (p - 1)l$.

Less precisely, there can be no estimate

$$h(\psi(W)) \leqslant C(n, r, K)(h(V) + R(G))^\kappa$$

replacing (11.2) which is polynomial in $h(V)$ and $R(G)$ for fixed n, r, K. For this would give a point with $h(\psi(\pi)) \leqslant c''(m + l)^\kappa \leqslant c''' l^\kappa$, contradicting (11.16). Similarly, one sees that if the dependence on $h(V)$ is polynomial, then the dependence on $R(G)$ must be exponential. This explains the large solutions such as (1.16), with $p = 2, l = 83, m = 42$.

12. *Proof of descent steps and theorems*

In the Descent Steps, the variety V is certainly defined over a finitely generated transcendental extension K of $\mathbf{F}_p$, and now we can choose any separable transcendence basis to obtain a height function. Now the Descent Step over $\sqrt{G}$ follows from the Main Estimate for isotrivial W. And the Descent Step over G follows, at least without the assumption that the W are $\sqrt{G}$-isotrivial, from the Main Estimate for points over G. This assumption can be removed by induction just as in Section 10 (without bothering about estimates): any W that is not $\sqrt{G}$-isotrivial can be replaced by a finite union of $\sqrt{G}$-isotrivial varieties.

To prove Theorem 1 we may assume that V has positive dimension. We apply the Main Estimate for points over G repeatedly, taking always $q = |\mathbf{F}_K|^{f_K(G)}$ for safety. With $V_0 = V$, an arbitrary point π of $V_0(G)$ is either a point of $W(G)$ for finitely many W in V_0 with $\dim W \leqslant \dim V - 1$, or a point $\psi_1^{-1}\varphi^{e_1}\psi_1(\pi_1)$ for π_1 in $V_1(G)$ for finitely many V_1 in V_0 with $\dim V_1 \leqslant \dim V - 1$ and some $e_1 \geqslant 0$, with $\psi_1(V_0)$ defined over $\mathbf{F}_K$. Then we argue similarly with π_1; and so on. After at most $\dim V \leqslant n-1$ steps we descend to cosets $T = V_h$, and only finitely many $\psi_1, \ldots, \psi_h$ turn up on the way, leading to expressions as in (1.12) and thereby establishing Theorem 1.

For later use, we note that not just the varieties T but also the whole unions $[\psi_1, \ldots, \psi_h]T$ lie in the variety V. Why is this? Well, a typical point of the union has the shape $\pi = (\psi_1^{-1}\varphi^{e_1}\psi_1)\ldots(\psi_h^{-1}\varphi^{e_h}\psi_h)(\tau)$ for some $e_1, \ldots, e_h$ and some τ in T. The descent for Theorem 1 provides linear varieties $V = V_0, V_1, \ldots, V_h = T$. Now clearly τ lies in T inside V_{h-1}, so $\psi_h^{-1}\varphi^{e_h}\psi_h(\tau)$ lies in

$$\psi_h^{-1}\varphi^{e_h}\psi_h(V_{h-1}) = \psi_h^{-1}\psi_h(V_{h-1}) = V_{h-1}$$

inside V_{h-2}. In the same way, $(\psi_{h-1}^{-1}\varphi^{e_{h-1}}\psi_{h-1})(\psi_h^{-1}\varphi^{e_h}\psi_h)(\tau)$ lies in V_{h-2} inside V_{h-3}. Continuing backwards we see that $\pi = (\psi_1^{-1}\varphi^{e_1}\psi_1)\ldots(\psi_h^{-1}\varphi^{e_h}\psi_h)(\tau)$ lies in V.

We leave it to the reader to check, by a straightforward induction argument such as that in Section 10 and also using Lemma 7.2, that for Theorem 1 one can take

$$\max\{h(\psi_1), \ldots, h(\psi_h), h(T)\} \leqslant (2q^2\Delta R(G)^{6n+2})^{\rho(m)} h(V)^{\eta(m)} \tag{12.1}$$

in the notation of Section 10. This indeed looks polynomial in $R(G)$ and $h(V)$; however, as we noted, an exponential dependence on $R(G)$ may be hiding in $q = |\mathbf{F}_K|^{f_K(G)}$.

For the symmetrization argument in the proof of Theorem 2, we need a version of [**7**, Lemma 8.1, p. 209], partly removed from its recurrence context.

LEMMA 12.1. *For $m \geqslant 1$ and $x_1, \ldots, x_m, y_1, \ldots, y_m$ in K suppose that*

$$x_1 y_1^{q^l} + \ldots + x_m y_m^{q^l} = 0 \tag{12.2}$$

for all large l. Then this holds for all $l \geqslant 0$.

Proof. The proof will be by induction on m, the case $m = 1$ being trivial. For the induction step we can clearly assume that $x_1, \ldots, x_m$ are non-zero. Now we note that (12.2) for any m consecutive integers $l = g, g+1, \ldots, g+m-1$ implies the linear dependence of $y_1, \ldots, y_m$ over $\mathbf{F}_q$. For if we regard these as linear equations for $x_1, \ldots, x_m$, then the underlying determinant is the q^g power of that with entries $y_i^{q^{j-1}}$ $(i, j = 1, \ldots, m)$, and it is well known that the latter, a so-called Moore determinant, is up to a constant the product of the $\beta_1 y_1 + \ldots + \beta_m y_m$ taken over all $(\beta_1, \ldots, \beta_m)$ in $\mathbf{P}_{m-1}(\mathbf{F}_q)$ (see for example [**13**, Corollary 1.3.7, p. 8]). Thus, after permuting we can suppose that $y_m = \alpha_1 y_1 + \ldots + \alpha_{m-1} y_{m-1}$ for $\alpha_1, \ldots, \alpha_{m-1}$ in $\mathbf{F}_q$. Substituting into (12.2) gives

$$(x_1 + \alpha_1 x_m) y_1^{q^l} + \ldots + (x_{m-1} + \alpha_{m-1} x_m) y_{m-1}^{q^l} = 0,$$

which therefore also holds for all large l. By the induction hypothesis we conclude that this holds for all $l \geqslant 0$, which leads back to (12.2) for all $l \geqslant 0$ and thus completes the proof. □

To prove Theorem 2 consider a single $[\psi_1, \ldots, \psi_h]T(G)$ coming from Theorem 1. Fix τ_0 in $T(G)$; then $T = \tau_0 S$ for a linear subgroup S.

We argue first on the geometric level. According to (1.12) a typical point of $[\psi_1,\ldots,\psi_h]T$ has the shape

$$\psi_1^{q_1-1}\psi_2^{q_1q_2-q_1}\psi_3^{q_1q_2q_3-q_1q_2}\ldots\psi_h^{q_1\cdots q_h-q_1\cdots q_{h-1}}(\tau_0\sigma)^{q_1\cdots q_h}$$

with $q_i=q^{e_i}$ $(i=1,\ldots,h)$ and σ in S; here, we are regarding the ψ_i $(i=1,\ldots,h)$ as multiplication by points instead of automorphisms. This expression can be written as

$$\pi_0\pi_1^{q_1}\pi_2^{q_1q_2}\ldots\pi_{h-1}^{q_1\cdots q_{h-1}}\pi_h^{q_1\cdots q_h}\sigma^{q_1\cdots q_h} \tag{12.3}$$

with

$$\pi_0=\psi_1^{-1},\quad \pi_1=\psi_2^{-1}\psi_1,\ldots,\quad \pi_{h-1}=\psi_h^{-1}\psi_{h-1},\quad \pi_h=\psi_h\tau_0. \tag{12.4}$$

Now when we write $q^{l_i}=q_1\ldots q_i$ $(i=1,\ldots,h)$ we certainly obtain a point of $(\pi_0,\pi_1,\ldots,\pi_h)S$ according to (1.14); but at the moment we have asymmetry $l_1\leqslant\ldots\leqslant l_h$. We eliminate the inequalities here as in [**7**, p. 212].

Let us start with the last inequality. We can write (12.3) as $\xi\eta^{q^l}$ with ξ and η independent of $l=l_h$. We already remarked that $[\psi_1,\ldots,\psi_h]T$ lies in V, so (12.3) does. Thus, for each linear form $\mathcal{L}$ defining V we have $\mathcal{L}(\xi\eta^{q^l})=0$ for all $l_1,\ldots,l_{h-1},l$ with $0\leqslant l_1\leqslant\ldots\leqslant l_{h-1}\leqslant l$. Fixing $l_1,\ldots,l_{h-1}$, we see from Lemma 12.1 that this equation for all large l implies the same equation for all $l\geqslant 0$. Thus, the inequality $l_{h-1}\leqslant l_h$ has indeed been eliminated. Similar arguments work for the other conditions, as is clear from the arguments of [**7**, p. 212] after equation (22). For example, the next step fixes $l_1,\ldots,l_{h-2},l_h$ but not $l=l_{h-1}$.

Looking back at (12.3), we have therefore proved that all the points

$$\pi_0\pi_1^{r_1}\pi_2^{r_2}\ldots\pi_{h-1}^{r_{h-1}}\pi_h^{r_h}\sigma^{r_h} \tag{12.5}$$

lie in V, where the integers $r_i=q^{l_i}$ $(i=1,\ldots,h)$ now range independently over all positive integral powers of q. This is the required symmetrization at the geometric level.

It actually shows that the entire $(\pi_0,\pi_1,\ldots,\pi_h)S$ lies in V. For a typical point of the former has the shape

$$\pi_0\pi_1^{r_1}\pi_2^{r_2}\ldots\pi_{h-1}^{r_{h-1}}\pi_h^{r_h}\tilde{\sigma} \tag{12.6}$$

for $\tilde{\sigma}$ in S. And there is σ in S with $\sigma^{r_h}=\tilde{\sigma}$. This could be interpreted as something about the divisibility of group varieties; but for us it is just a simple consequence of the fact that S is defined by equations $X_i=X_j$. And now (12.6) and (12.5) are equal.

At the arithmetic level we claim that $(\pi_0,\pi_1,\ldots,\pi_h)S(G)$ lies in $V(G)$. In fact every point

$$\pi=\pi_0\pi_1^{r_1}\pi_2^{r_2}\ldots\pi_{h-1}^{r_{h-1}}\pi_h^{r_h} \tag{12.7}$$

with asymmetry $r_1\leqslant\ldots\leqslant r_h$ has the shape (12.3) (with all coordinates of σ equal to 1). It therefore lies in $[\psi_1,\ldots,\psi_h]T(G)$ which is in turn contained in $V(G)$. In particular π lies in $\mathbf{P}_n(G)$. But why does it continue to lie in $\mathbf{P}_n(G)$ when the asymmetry is lifted?

Well, we can take $r_1=\ldots=r_h=1$ in (12.7) to see that the product

$$\pi_0\pi_1\ldots\pi_h \tag{12.8}$$

lies in $\mathbf{P}_n(G)$. Then taking $r_1=\ldots=r_{h-1}=1, r_h=q$ we can deduce that π_h^{q-1} lies in $\mathbf{P}_n(G)$. And taking $r_1=\ldots=r_{h-2}=1, r_{h-1}=r_h=q$ we deduce that π_{h-1}^{q-1} lies in $\mathbf{P}_n(G)$. And so on, until we see that all of

$$\pi_1^{q-1},\ldots,\pi_h^{q-1} \tag{12.9}$$

lie in $\mathbf{P}_n(G)$ (this was already remarked in Section 1).

And now if $r_1,\ldots,r_h$ are arbitrary integral powers of q in (12.7), we can write

$$\pi=(\pi_0\pi_1\ldots\pi_h)\pi_1^{r_1-1}\ldots\pi_h^{r_h-1}$$

to see from (12.8) and (12.9) that indeed π lies in $\mathbf{P}_n(G)$.

Now any point of $(\pi_0, \pi_1, \ldots, \pi_h)S(G)$ by (12.5) has the form $\pi\sigma^{r_h}$ with π as above and σ in $S(G)$. It follows that $(\pi_0, \pi_1, \ldots, \pi_h)S(G)$ lies in $V(G)$ as claimed.

On the other hand, taking all coordinates of σ as 1 in (12.3) shows that $[\psi_1, \ldots, \psi_h]\{\tau_0\}$ lies in $(\pi_0, \pi_1, \ldots, \pi_h)S(G)$. As we could have fixed τ_0 arbitrarily in $T(G)$, we see that $[\psi_1, \ldots, \psi_h]T(G)$ lies in $(\pi_0, \pi_1, \ldots, \pi_h)S(G)$.

It follows that $V(G)$ is indeed the union of the $(\pi_0, \pi_1, \ldots, \pi_h)S(G)$, which completes the proof of Theorem 2. We note for later use the fact, already observed, that each $(\pi_0, \pi_1, \ldots, \pi_h)S$ is contained in V.

Here too we leave it to the reader to check using (12.1) that for Theorem 2 one can take

$$\max\{h(\pi_0), h(\pi_1), \ldots, h(\pi_h)\} \leqslant (n+1)(2q^2\Delta R(G)^{6n+2})^{\rho(m)} h(V)^{\eta(m)}. \tag{12.10}$$

This follows quickly from (12.4) and the easy fact that any $T(G)$ contains τ_0 with $h(\tau_0) \leqslant h(T)$.

To prove part (1) of Theorem 3 we start from Theorem 1 with $V = H$. We first claim that if some π in $H(G)$ lies in some $[\psi_1, \ldots, \psi_h]T(G)$ with T not a single point, then some (1.2) fails for π. To see this, note that if T is not a single point, then there is a partition of $\{0, 1, \ldots, n\}$ into proper subsets $I, J, \ldots$ such that T is defined by the proportionality of the homogeneous coordinates X_i $(i \in I)$, X_j $(j \in J)$, and so on. We may suppose that I contains 0 and that the equations corresponding to I are $g_iX_0 = g_0X_i$ for i in I. Consider the point τ_I in $\mathbf{P}_n$ whose coordinates $X_i = g_i$ for i in I but with all other coordinates zero. It also lies in T.

Now $\pi = (\psi_1^{-1}\varphi^{e_1}\psi_1)\ldots(\psi_h^{-1}\varphi^{e_h}\psi_h)(\tau)$ for some $e_1, \ldots, e_h$ and some τ in T. From our remark following the proof of Theorem 1, we see that $\pi_I = (\psi_1^{-1}\varphi^{e_1}\psi_1)\ldots(\psi_h^{-1}\varphi^{e_h}\psi_h)(\tau_I)$ lies in H. Now τ and τ_I have the same coordinates X_i $(i \in I)$. It follows that π and π_I have the same coordinates X_i $(i \in I)$. Since the other coordinates of π_I are zero, this means that (1.2) fails for π as claimed.

Therefore, $H^*(G)$ is contained in a finite union of sets $[\psi_1, \ldots, \psi_h]\{\tau\}$. And each of these lies in $H(G)$. This proves part (1) of Theorem 3.

Part (2) follows in a similar way with the help of the remark after the proof of Theorem 2, with $\pi = \pi_0(\varphi^{l_1}\pi_1)\ldots(\varphi^{l_h}\pi_h)\sigma$ and $\pi_I = \pi_0(\varphi^{l_1}\pi_1)\ldots(\varphi^{l_h}\pi_h)\sigma_I$ for σ_I defined by $X_i = 1$ for i in I but with all other coordinates zero. This shows that we can restrict to single points S, and the proof is finished as above. We have therefore proved all of Theorem 3.

It is easy to deduce explicit estimates for Theorem 3 as for Theorems 1 and 2. One obtains at once (12.1) (with T replaced by τ) and (12.10).

13. *Limitation results*

We show here that for each $n \geqslant 2$ the bounds $h \leqslant n-1$ in Theorems 1 and 2 cannot always be improved; and also that if $p > 2$ the $\psi_1, \ldots, \psi_h$ in Theorem 1 and the $\pi_0, \pi_1, \ldots, \pi_h$ in Theorem 2 cannot always be chosen over G.

We start with $h \leqslant n-1$. Because Theorem 1 directly implies Theorem 2 and then Theorem 3, it will suffice to prove the analogous statements for Theorem 3. Also we have seen that each $[\psi_1, \ldots, \psi_h]\{\tau\}$ in Theorem 3(1) is contained in some $(\pi_0, \pi_1, \ldots, \pi_h)$ in Theorem 3(2). So it is enough to treat Theorem 3(2).

This we do with the affine hyperplane

$$x_1 + x_2 - x_3 - \ldots - x_n = 1 \tag{13.1}$$

already mentioned.

We need a simple observation. For a prime p let $R = R_p$ be the set of points $(1, r_1, \ldots, r_{n-1})$ as the integers $r_1, \ldots, r_{n-1}$ run through all powers of p satisfying the asymmetry conditions that r_i divides r_{i+1} $(i = 1, \ldots, n-2)$ and also the extra conditions

$$r_{n-1} \neq r_{n-2}, r_{n-2} + r_{n-3}, \ldots, r_{n-2} + r_{n-3} + \ldots + r_1. \tag{13.2}$$

Lemma 13.1. *The set R does not lie in a finite union of proper subgroups of $\boldsymbol{Z}^n$.*

Proof. We can actually disregard (13.2) because their failure would just add more to the finite union of proper subgroups. Now the falsity of the lemma would lead to an equation

$$\mathcal{F}(p^{e_1}, \dots, p^{e_{n-1}}) = 0 \tag{13.3}$$

holding for all non-negative integers $e_1, \dots, e_{n-1}$, where $\mathcal{F}(y_1, \dots, y_{n-1})$ is a finite product of polynomials

$$\mathcal{A} = a_0 + a_1 y_1 + a_2 y_1 y_2 + \ldots + a_{n-1} y_1 y_2 \dots y_{n-1}$$

corresponding to the proper subgroups of $\mathbf{Z}^n$ perpendicular to $(a_0, \dots, a_{n-1}) \neq 0$. It is clear that each $\mathcal{A} \neq 0$ and so $\mathcal{F} \neq 0$. On the other hand, it is easy to see that the points in (13.3) are Zariski-dense in $\mathbf{R}^{n-1}$. This contradiction proves the lemma. □

Take as usual $K = \mathbf{F}_p(t)$ and G generated by t and $1 - t$. We proceed to exhibit many points on $H^*(G)$ with H defined by (13.1).

For integral powers $q_1, \dots, q_{n-1}$ of p define

$$r_1 = q_{n-1}, \quad r_2 = q_{n-1} q_{n-2}, \dots, \quad r_{n-1} = q_{n-1} \dots q_1$$

and

$$d_1 = r_{n-1} - r_{n-2} - \ldots - r_2 - r_1,$$

$$d_2 = r_{n-1} - r_{n-2} - \ldots - r_2,$$

down to

$$d_{n-2} = r_{n-1} - r_{n-2}$$

and

$$d_{n-1} = r_{n-1}.$$

Then

$$x_1 = t^{d_1}, \quad x_2 = 1 - t^{d_{n-1}}, \quad x_3 = t^{d_{n-2}} - t^{d_{n-1}}, \dots, x_n = t^{d_1} - t^{d_2} \tag{13.4}$$

certainly satisfy (13.1), so the point $\xi = (x_1, \dots, x_n)$ lies in H. It is in $H(G)$ because

$$x_2 = 1 - t^{r_{n-1}} = (1 - t)^{r_{n-1}},$$

$$x_3 = t^{d_{n-2}}(1 - t^{r_{n-2}}) = t^{d_{n-2}}(1 - t)^{r_{n-2}},$$

and so on.

This also leads to a multiplicative representation

$$\xi = \xi_1^{r_1} \dots \xi_{n-1}^{r_{n-1}} \tag{13.5}$$

of the point in (13.4), where

$$\xi_1 = \left(\frac{1}{t}, 1, 1, 1, 1, \dots, 1, 1, \frac{1-t}{t}\right),$$

$$\xi_2 = \left(\frac{1}{t}, 1, 1, 1, 1, \dots, 1, \frac{1-t}{t}, \frac{1}{t}\right)$$

$$\xi_3 = \left(\frac{1}{t}, 1, 1, 1, 1, \dots, \frac{1-t}{t}, \frac{1}{t}, \frac{1}{t}\right)$$

down to

$$\xi_{n-2} = \left(\frac{1}{t}, 1, \frac{1-t}{t}, \frac{1}{t}, \frac{1}{t}, \dots, \frac{1}{t}, \frac{1}{t}, \frac{1}{t}\right),$$

but

$$\xi_{n-1} = (t, 1-t, t, t, t, \dots, t, t, t).$$

We can quickly check that $\xi_1, \dots, \xi_{n-1}$ are multiplicatively independent. Namely, a relation

$$\xi_1^{a_1} \dots \xi_{n-1}^{a_{n-1}} = (1, 1, 1, 1, 1, \dots, 1, 1, 1)$$

would lead to $a_{n-1} = 0$ on examining the second components, then $a_{n-2} = 0$ from the third components, and so on down to $a_1 = 0$.

The case $n = 3$ with $q_1 = q, q_2 = r$ is of course (1.11) or (1.13).

We can see that (13.4) lies in $H^*(G)$ provided $(1, r_1, \dots, r_{n-1})$ lies in R. For the various exponents of t clearly satisfy $d_{n-1} > d_{n-2} > \dots > d_2 > d_1$. There is one more exponent 0; but $d_{n-1} \neq 0$ and from the definition of R we also have $d_{n-2} \neq 0, \dots, d_1 \neq 0$. Thus, the exponents $d_{n-1}, \dots, d_1, 0$ in (13.4) are distinct, and it is easy to see that there can be no vanishing subsum of $x_1, x_2, -x_3, \dots, -x_n$ (in fact each of $d_{n-2} = 0, \dots, d_1 = 0$ does lead to a vanishing subsum). We already remarked that (1.13) is in H^* as long as $r \neq s$, that is $q_1 \neq 1$, that is $r_2 \neq r_1$ as in (13.2).

Now we can prove as promised that $H^*(G)$ does not lie in a finite union of sets

$$\Pi = (\pi_0, \pi_1, \dots, \pi_h)_q = \bigcup_{l_1=0}^{\infty} \dots \bigcup_{l_h=0}^{\infty} \pi_0 \pi_1^{q^{l_1}} \dots \pi_h^{q^{l_h}} \tag{13.6}$$

for some q and points $\pi_0, \pi_1 \dots, \pi_h$ with $h < n-1$. The idea is to note that each Π lies in a coset of $\mathbf{G}_{\mathrm{m}}^n$ of dimension at most $h \leqslant n-2$; whereas the points (13.5) have rank $n-1$.

Accordingly, we assume that $H^*(G)$ does lie in such a finite union and we shall reach a contradiction.

Now for each element of R the corresponding (13.5) lies in $H^*(G)$ so in some Π. This provides a partition of R into a finite union of subsets R_Π. By Lemma 13.1 we will be through if we can prove that each R_Π lies in a proper subgroup of $\mathbf{Z}^n$.

Suppose for some Π we are lucky in the sense that the corresponding π_0 in (13.6) is multiplicatively independent of $\xi_1, \dots, \xi_{n-1}$. The corresponding

$$\pi_0^{-1}\xi = \pi_0^{-1}\xi_1^{r_1} \dots \xi_{n-1}^{r_{n-1}}$$

all lie in the group generated by $\pi_1, \dots, \pi_h$, and so the multiplicative rank of the various $\pi_0^{-1}\xi$ is at most $h \leqslant n-2$. Since $\pi_0^{-1}, \xi_1, \dots, \xi_{n-1}$ are independent, it follows that the set R_Π cannot contain n (or even $n-1$) independent elements. So it must indeed lie in a proper subgroup of $\mathbf{Z}^n$.

In fact we are not so likely to be that lucky, and it is more probable that there is a relation $\pi_0^a = \xi_1^{a_1} \dots \xi_{n-1}^{a_{n-1}}$ with $a \neq 0$. Now the

$$\pi_0^{-a}\xi^a = \xi_1^{ar_1 - a_1} \dots \xi_{n-1}^{ar_{n-1} - a_{n-1}}$$

still lie in a group of rank at most $n-2$. Since $\xi_1, \dots, \xi_{n-1}$ are independent, we deduce as above that the set of all $(ar_1 - a_1, \dots, ar_{n-1} - a_{n-1})$ lie in a proper subgroup of $\mathbf{Z}^{n-1}$. And this implies as above that R_Π lies in a proper subgroup of $\mathbf{Z}^n$.

That finishes the proof of the first limitation result. We could also have argued with a symmetrized version of R; then the $\mathcal{A}$ in the proof of Lemma 13.1 could be taken more simply as $a_0 + a_1 y_1 + a_2 y_2 + \dots + a_{n-1} y_{n-1}$.

We can use similar arguments to prove the second limitation result concerning non-definability over G. Because the $[\psi_1, \dots, \psi_h]T(G)$ in Theorem 1 lead to $(\pi_0, \pi_1, \dots, \pi_h)$ in Theorem 2 with (12.4) for τ_0 in $T(G)$, it will again suffice to check the matter for Theorem 3(2).

This we do with the affine line H defined by $tx + y = 1$ also over $K = \mathbf{F}_p(t)$, now with G generated by t^{p-1} and $1-t$. It is the example treated at the end of Section 11 with $m = 1$ and $l = p - 1$. We need another simple observation.

LEMMA 13.2. *For an odd prime p suppose that*

$$q_1 + q_2 + q_3 = \tilde{q}_1 + \tilde{q}_2 + \tilde{q}_3 \tag{13.7}$$

for integral powers $q_1, q_2, q_3, \tilde{q}_1, \tilde{q}_2$ and $\tilde{q}_3$ of p. Then $\tilde{q}_1, \tilde{q}_2$ and $\tilde{q}_3$ are a permutation of q_1, q_2 and q_3.

Proof. If q_1, q_2 and q_3 are all different, then the left-hand side of (13.7) has just three ones in its expansion to base p. So also the right-hand side; which means that $\tilde{q}_1, \tilde{q}_2$ and $\tilde{q}_3$ are also all different. The result in this case is now clear (even for $p = 2$). If say $q_1 \neq q_2 = q_3$, then we obtain a one and a two in the expansion because $p \neq 2$; so after a permutation $\tilde{q}_1 \neq \tilde{q}_2 = \tilde{q}_3$ too, and the result is still clear. Similarly, if $q_1 = q_2 = q_3$ as long as $p \neq 3$. This last case can also be checked directly when $p = 3$ and this proves the lemma; however, the example $1 + 1 + 4 = 2 + 2 + 2$ shows that $p = 2$ is not to be saved. $\square$

Now the analysis in Section 11 before the primitive root business shows easily that the points of $H^*(G) = H(G)$ are given by

$$x = t^{r-1}, \quad y = (1-t)^r \ \ (r = 1, p, p^2, \ldots). \tag{13.8}$$

This is $(x, y) = \xi_0 \xi_1^r$ for $\xi_0 = (t^{-1}, 1)$ and $\xi_1 = (t, 1-t)$. Assume $p \neq 2$. If $H^*(G)$ were contained in a finite union of

$$\Pi = (\pi_0, \pi_1)_q = \bigcup_{l=0}^{\infty} \pi_0 \pi_1^{q^l}$$

for some q and some π_0, π_1 over G, then one of these Π would certainly contain at least three different points (13.8). This gives equations

$$\xi_0 \xi_1^r = \pi_0 \pi_1^s, \quad \xi_0 \xi_1^{r'} = \pi_0 \pi_1^{s'}, \quad \xi_0 \xi_1^{r''} = \pi_0 \pi_1^{s''} \tag{13.9}$$

for powers $r < r' < r''$ of p and powers s, s' and s'' of q. Eliminating π_0 and π_1 leads to

$$(\xi_0 \xi_1^r)^{s'-s''} (\xi_0 \xi_1^{r'})^{s''-s} (\xi_0 \xi_1^{r''})^{s-s'} = 1;$$

that is, $\xi_1^a = 1$ for

$$a = r(s' - s'') + r'(s'' - s) + r''(s - s').$$

So $a = 0$; that is,

$$rs' + r's'' + r''s = rs'' + r's + r''s'.$$

Lemma 13.2 shows in particular that rs' is one of the terms on the right. But which one? Certainly $rs' \neq r''s'$. And $rs' \neq rs''$ else $s' = s''$ and (13.9) would imply $r' = r''$. It follows that $rs' = r's$. But now eliminating ξ_1 from the first two equations in (13.9) leads to $\xi_0^{r'-r} = \pi_0^{r'-r}$. Thus there would be α and β in $\overline{\mathbf{F}_p}$ with $(\alpha t^{-1}, \beta) = (\alpha, \beta)\xi_0 = \pi_0$; however, this is impossible because αt^{-1} is not in G if $p \neq 2$.

References

1. D. ABRAMOVICH and F. VOLOCH, 'Toward a proof of the Mordell–Lang conjecture in characteristic p', *Int. Math. Res. Notices* 5 (1992) 103–115.
2. B. ADAMCZEWSKI and J. P. BELL, 'On vanishing coefficients of algebraic power series over fields of positive characteristic', *Inventiones Math.*, to appear.
3. E. BOMBIERI and W. GUBLER, *Heights in diophantine geometry*, New Mathematical Monographs 4 (Cambridge University Press, Cambridge, 2006).
4. E. BOMBIERI, D. MASSER and U. ZANNIER, 'Intersecting a plane with algebraic subgroups of multiplicative groups', *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (5) VII (2008) 51–80.

5. J. W. S. Cassels, *An introduction to the geometry of numbers*, Classics in Mathematics (Springer, Berlin, 1971).
6. L. Cerlienco, M. Mignotte and F. Piras, 'Suites récurrentes linéaires: propriétés algébriques et arithmétiques', *L'Enseignement Mathématique* 33 (1987) 67–108.
7. H. Derksen, 'A Skolem-Mahler-Lech theorem in positive characteristic and finite automata', *Invent. Math.* 168 (2007) 175–224.
8. J.-H. Evertse, 'On sums of S-units and linear recurrences', *Compositio Math.* 53 (1984) 225–244.
9. J.-H. Evertse, H.P. Schlickewei and W.M. Schmidt, 'Linear equations in variables which lie in a multiplicative group', *Ann. of Math.* 155 (2002) 807–836.
10. J.-H. Evertse and U. Zannier, 'Linear equations with unknowns from a multiplicative group in a function field', *Acta Arith.* 133 (2008) 159–170.
11. D. Ghioca, 'The isotrivial case in the Mordell–Lang theorem', *Trans. Amer. Math. Soc.* 360 (2008) 3839–3856.
12. D. Ghioca and R. Moosa, 'Division points on subvarieties of isotrivial semiabelian varieties', *Int. Math. Res. Notices* 19 (2006) 1–23, Article ID 65437.
13. D. Goss, *Basic structures of function field arithmetic*, Ergebnisse der Math. 35 (Springer, 1996).
14. D. R. Heath-Brown, 'Artin's conjecture for primitive roots', *Quart. J. Math. Oxford Ser.* (2) 37 (1986) 27–38.
15. W. V. D. Hodge and D. Pedoe, *Methods of algebraic geometry I* (Cambridge University Press, Cambridge, 1968).
16. E. Hrushovski, 'The Mordell–Lang conjecture for function fields', *J. Amer. Math. Soc.* 9 (1996) 667–690.
17. L.-C. Hsia and J. T.-Y. Wang, 'The ABC theorem for higher-dimensional function fields', *Trans. Amer. Math. Soc.* 356 (2003) 2871–2887.
18. S. Lang, *Introduction to algebraic geometry* (Addison-Wesley, New York, 1973).
19. S. Lang, *Fundamentals of diophantine geometry* (Springer, Berlin, 1983).
20. D. Leitner, 'Linear equations over multiplicative groups in positive characteristic', *Acta Arithmetica.*, to appear.
21. R. C. Mason, *Diophantine equations over function fields*, London Mathematical Society Lecture Notes 96 (Cambridge University Press, Cambridge, 1984).
22. D. Masser, 'Mixing and linear equations over groups in positive characteristic', *Israel J. Math.* 142 (2004) 189–204.
23. R. Moosa and T. Scanlon, 'The Mordell–Lang conjecture in positive characteristic revisited', *Model theory and applications*, Quaderni di matematica 11 (eds L. Belair, Z. Chatzidakis, P. D'Aquino, D. Marker, M. Otero, F. Point and A. Wilkie; Dipartimento di Matematica Seconda Università di Napoli, Italy, 2002) 273–296.
24. R. Moosa and T. Scanlon, 'F-structures and integral points on semiabelian varieties over finite fields', *Amer. J. Math.* 126 (2004) 473–522.
25. A. J. van der Poorten and H. P. Schlickewei, 'Additive relations in fields', *J. Austral. Math. Soc.* A 51 (1991) 154–170.
26. W. M. Schmidt, *Diophantine approximations and diophantine equations*, Lecture Notes in Mathematics 1467 (Springer, Berlin, 1991).
27. T. Struppeck and J. D. Vaaler, 'Inequalities for heights of algebraic subspaces and the Thue-Siegel principle', *Analytic Number Theory (Allerton Park 1989)*, Progress in Mathematics 85 (Birkhäuser Boston, 1990) pp. 493–528.
28. J. Thunder, 'Siegel's lemma for function fields', *Michigan Math. J.* 42 (1995) 147–162.
29. J. F. Voloch, 'The equation $ax + by = 1$ in characteristic p', *J. Number Theory* 73 (1998) 195–200.

H. Derksen
Department of Mathematics
University of Michigan
East Hall 530 Church Street
Ann Arbor, MI 48109
USA

hderksen@umich.edu

D. Masser
Mathematisches Institut
Universität Basel
Rheinsprung 21
4051 Basel
Switzerland

david.masser@unibas.ch