ACTA ARITHMETICA
LXXIX.4 (1997)

The unit equation and the cluster principle

by

E. Bombieri (Princeton, N.J.), J. Mueller (Bronx, N.Y.) and
M. Poe (Cambridge, Mass.)

Dedicated to Professor J. W. S. Cassels

1. Introduction. Let k be a number field of degree d and let Γ be a finitely generated subgroup of $(k^\times)^2 = \mathbb{G}_m^2(k)$. The unit equation for Γ is the equation

$$a_1x_1 + a_2x_2 = 1, \quad (x_1, x_2) \in \Gamma \tag{1.1}$$

with coefficients $a_1, a_2 \in k^\times$.

In this paper we shall exploit an idea introduced in the paper [P] by the third author, to obtain what may be called a *cluster principle* for solutions of (1.1).

Roughly speaking, it asserts that solutions of (1.1) up to height H can be subdivided into a bounded number of subsets, the bound depending only on the rank of Γ, such that after rescaling each subset by an element of Γ the rescaled subset has height proportional to $\log H$. These subsets therefore may be regarded as forming "clusters" of solutions of (1.1).

The principle of formation of clusters can be extended to analyze the clusters themselves, which now split into a bounded number of clusters of size $\log\log H$, and so on. After very few steps, the size of the clusters so obtained becomes very small while their number remains controlled in terms of the rank of Γ alone.

In order to apply this principle to all solutions of (1.1) we need bounds for the heights of solutions of a unit equation. Baker's theory of linear forms in logarithms provides such a bound, in the form of iterated exponentials of arguments depending on the degree and discriminant of the field k and the heights of generators and rank of Γ. The presence of iterated exponentials causes little harm here, because repeated application of the cluster principle

1991 *Mathematics Subject Classification*: 11D57, 11D61.

brings in iterated logarithms. Thus after few steps we need only deal with solutions of unit equations of rather small height.

An interesting feature of the cluster principle is that it becomes more efficient if the group Γ has always "large" generators. The upshot is that in the end we deal only with solutions with height bounded only in terms of the rank of Γ and possibly the degree of the field k, thus showing that the number of solutions of (1.1) is bounded in terms of the rank of Γ and the degree d of the number field k. We shall prove by this method

THEOREM. *Let k be a number field of degree d and let Γ be a finitely generated subgroup of rank s of $\mathbb{G}_m^2(k)$. Then the generalized unit equation*

$$a_1x_1 + a_2x_2 = 1, \qquad (x_1, x_2) \in \Gamma,$$

admits at most $d^{9s}c_1^{s^2}$ solutions, for some absolute constant c_1.

Remark. A calculation shows that $c_1 = e^{86}$ is admissible here.

Although this result falls short of the remarkable bound 256^{s+1} recently obtained by F. Beukers and H.-P. Schlickewei [BS], our method is entirely different and has the potential to be extendable to other situations, such as the study of rational points on curves in an abelian variety.

The principle behind the method has its roots in the remark that if we have a congruence $(1 + pa)^m \equiv 1 \pmod{p^r}$ with $p > 2$ a prime then $p^{r-1-\mathrm{ord}_p(a)} \mid m$ (if $p = 2$, the result holds with $r - 2$ in place of $r - 1$). This was used by C. Størmer in 1898 (see Ribenboim [R], part C, §9) to solve effectively special unit equations, such as

$$AM_1^{e_1} \ldots M_m^{e_m} - BN_1^{f_1} \ldots N_n^{f_n} = 1 \text{ or } 2$$

where A, B, M_i, e_i, N_j, f_j are positive integers, and in 1960 J. W. S. Cassels [C1] used related methods to solve effectively the above equation with any constant C in place of Størmer's 1 or 2.

In dealing with a general unit equation we encounter more general congruences

$$\prod_{i=1}^{s}(1 + pa_i)^{u_i} \equiv 1 \pmod{p^r}$$

and the point is that it is still possible to extract some information from this. The basic idea in [P] is that if r is large enough this congruence implies that there are two distinct indices i and j such that $\mathrm{ord}_p(u_ia_i) = \mathrm{ord}_p(u_ja_j)$. This is nontrivial information, lying at the basis of the clustering phenomenon.

The organization of this paper is as follows. In Section 2 we introduce basic definitions and the notion of the mass of a subset of a finitely generated abelian group G. Section 3 deals with the cluster principle at a single place. The next two sections define general regulators and study their properties. Section 6 states and proves the global version of the cluster principle. The next section deals with some consequences of Baker's theory, and the final section contains the proof of our theorem as a consequence of the cluster principle and induction on the rank of Γ.

In order not to obscure the main ideas involved here, we have not tried in this paper to obtain sharpest possible results, and significant improvements can be obtained in our formulation of the cluster principle as well as in the way of applying it.

Acknowledgements. During the period this research was done the second author was partially supported by NSA Grant MDA 904-94-H-2004.

2. Some basic definitions. We denote by M_k the set of all places of k and by $|\ |_v$ the associated normalized absolute values satisfying the product formula. The normalization we shall use is the following. If v extends the place $v_0 \in M_{\mathbb{Q}}$, then $\|\ \|_{v_0}$ denotes the usual p-adic or real absolute value. In general, $\|\ \|_v$ is normalized by means of

$$\log \|x\|_v = \log \|x\|_{v_0} \quad \text{for } x \in \mathbb{Q}^\times. \tag{2.1}$$

If v is finite, lying over the rational prime p, the associated additive valuation $\mathrm{ord}_v(\ )$ is normalized by means of $\mathrm{ord}_v(p) = 1$ and we have

$$\log \|x\|_v = -(\log p)\,\mathrm{ord}_v(x). \tag{2.2}$$

There is another normalization $|\ |_v$ defined by

$$\log |x|_v = \frac{d_v}{d} \log \|x\|_v$$

where d_v is the local degree $[k_v : \mathbb{Q}_v]$. For each finite v we also write $d_v = e_v f_v$ where e_v is the ramification index and f_v is the residue class degree.

With this normalization we have the product formula

$$\sum_{v \in M_k} \log |x|_v = 0$$

for $x \in k^\times$. If k' is a finite extension of k we also have the extension formula

$$\sum_{w \in M_{k'},\, w|v} \log |x|_w = \log |x|_v$$

for $x \in k^\times$.

The absolute logarithmic height of $x \in k^\times$ is given by

$$h(x) = \sum_{v \in M_k} \log^+ |x|_v$$

where $\log^+ t = \max(\log t, 0)$. The absolute height $H(x)$ is given by

$$H(x) = e^{h(x)}.$$

If G is a finitely generated abelian group, the rank $\mathrm{rk}(G)$ of G is the rank of the free abelian group G/tors. If E is a subset of G, the rank $\mathrm{rk}(E)$ of E is the rank of the subgroup $\langle E\rangle$ generated by all elements of E.

In what follows, it is convenient to use a modified notion of rank of a set.

Definition 1. Let T be an indeterminate which commutes with G and let $E\subset G$. The *augmented set* $\widetilde{E}$ is the set $\widetilde{E}=\{Te : e\in E\}$.

It is immediate that $\mathrm{rk}(\widetilde{E})=\mathrm{rk}(E)$ or $\mathrm{rk}(E)+1$.

Let $s=\mathrm{rk}(G)$ and let $(g_1,\ldots,g_s)$ be generators of G up to torsion. Every element of G can be written uniquely as

$$g=\varepsilon\prod_{\sigma=1}^{s} g_\sigma^{n_\sigma(g)} \tag{2.3}$$

with $\varepsilon\in\mathrm{tors}(G)$ and $n_\sigma(g)\in\mathbb{Z}$.

Definition 2. Let $E\subset G$ be a nonempty subset of G and let $s=\mathrm{rk}(G)$. If $s\ge 1$ the *mass* of E with respect to G is

$$m(E,G)=\sup|\det(n_\sigma(e_j))_{\substack{\sigma=1,\ldots,s\\ j=1,\ldots,s}}|$$

where the supremum is over all s-tuples $(e_1,\ldots,e_s)$ of elements of E. The mass so defined is independent of a choice of generators of G.

The *absolute mass* of E is

$$m(E)=m(E,\langle E\rangle)$$

provided $\mathrm{rk}(E)\ge 1$, and is undefined otherwise.

It is clear that the mass is independent of a choice of generators of G, because changing generators changes $\mathbf{n}=\{n_1(g),\ldots,n_s(g)\}$ into $\mathbf{B}\cdot\mathbf{n}$ with $\det(\mathbf{B})=\pm 1$.

The following facts are worth recording.

Lemma 1. *The mass $m(E,G)$ is 0 if and only if $\mathrm{rk}(E)<\mathrm{rk}(G)$. More generally, if G' has finite index in G and if $\langle E\rangle\subset G'$, we have*

$$m(E,G)=[G/\mathrm{tors}:G'/\mathrm{tors}]\,m(E,G'). \tag{2.4}$$

If $E'\subset E$, then

$$m(E')\le m(E).$$

We also have

$$m(\widetilde{E})\le(\mathrm{rk}(E)+1)m(E).$$

Proof. The first two statements are clear.

The third statement is proved as follows. We may assume that G is torsion free. Let $G'=\langle E'\rangle$, $G=\langle E\rangle$ be of rank s', s. We take another $s-s'$

independent elements $e_1, \ldots, e_{s-s'}$ of E and set $E^* = \{E', e_1, \ldots, e_{s-s'}\}$, $G^* = \langle E^* \rangle$. By construction, G^* has the same rank s as G. Generators for G^* are obtained by taking a set of generators of G' together with $e_1, \ldots, e_{s-s'}$. From this construction and the definition of mass we see, using this set of generators for G^*, that $m(E') = m(E', G') = m(E^*, G^*)$. Finally,

$$m(E^*, G^*) = m(E, G)/[G/\mathrm{tors} : G^*/\mathrm{tors}] \le m(E, G) = m(E),$$

proving what we want.

For the last statement, we argue as follows. We may suppose that $G = \langle E \rangle$ is torsion free. Hence let G have rank $s \ge 1$. We have two cases. Suppose first that $\mathrm{rk}(\widetilde{E}) = \mathrm{rk}(E) + 1 = s + 1$. Then $\langle \widetilde{E} \rangle$ has finite index in $\widetilde{G} = \langle T, G \rangle$, therefore by (2.4) we have

$$m(\widetilde{E}) \le m(\widetilde{E}, \widetilde{G}).$$

Since $\widetilde{G} = \langle T, G \rangle$ we see that

$$m(\widetilde{E}, \widetilde{G}) = \sup \left| \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ n_1(e_1) & n_1(e_2) & \ldots & n_1(e_{s+1}) \\ \ldots & \ldots & \ldots & \ldots \\ n_s(e_1) & n_s(e_2) & \ldots & n_s(e_{s+1}) \end{pmatrix} \right| .$$

By Laplace's rule, this determinant does not exceed $(s+1)m(E)$, proving our assertion in this case.

If instead $\mathrm{rk}(\widetilde{E}) = \mathrm{rk}(E) = s$ we see that elements of $\langle \widetilde{E} \rangle$ are written uniquely as $T^{n(g)}g$ with $g \in \langle E \rangle$, for some homomorphism $n : \langle E \rangle \to \mathbb{Z}$. The map $T^{n(g)}g \mapsto g$ gives an isomorphism between $\langle \widetilde{E} \rangle$ and $\langle E \rangle$, so that in this case we have $m(\widetilde{E}) = m(E)$. This completes the proof of the lemma.

3. The Local Cluster Principle. Notation is as in the preceding section. The following result shows that for each place v the points $\log \|x_j\|_v$, with $\mathbf{x}$ a solution of the unit equation, tend to cluster in small intervals.

Local Cluster Principle (finite places). *Let v be a finite place of k lying over the rational prime p. Let $j = 1$ or 2 and let $\mathcal{X}$ be a finite set of solutions of the unit equation $a_1x_1 + a_2x_2 = 1$ with $\mathbf{x} \in \mathit{\Gamma}$, such that* $\mathrm{ord}_v(a_jx_j)$ *has fixed sign for every $\mathbf{x} \in \mathcal{X}$. Suppose* $\mathrm{rk}(\mathcal{X}) \ge 1$. *Then there is a decomposition*

$$\mathcal{X} = \bigcup_{i=-1}^{s} \mathcal{X}_i$$

with the following two properties.

(i) *Clustering property*: *If $\mathcal{X}_i \ne \emptyset$ then for $\mathbf{x}, \mathbf{x}' \in \mathcal{X}_i$ we have*

$$|\log \|x_j'/x_j\|_v| \le \log((s+1)m(\mathcal{X}));$$

(ii) *Rank property*: *Let*

$$\mathcal{X}'_i = \bigcup_{h\ge i} \mathcal{X}_h.$$

Suppose that $i \ge 0$ *and that* $\mathcal{X}'_i \ne \emptyset$. *Then*

$$\mathrm{rk}(\widetilde{\mathcal{X}}'_i) > \mathrm{rk}(\widetilde{\mathcal{X}}'_{i+1}).$$

Proof. It suffices to deal with the case $j = 1$. For notational simplicity, we write a, b, x, y for a_1, a_2, x_1, x_2 (so as to avoid double indices) and we also write

$$M = \log(m(\widetilde{\mathcal{X}}))/\log p.$$

Consider first the case in which $\mathrm{ord}_v(ax) \ge 0$ for every $\mathbf{x} \in \mathcal{X}$. We define $\mathcal{X}_i$ inductively for $i \ge -1$ as follows.

$$\mathcal{X}_{-1} = \{\mathbf{x} \in \mathcal{X} : 0 \le \mathrm{ord}_v(ax) \le 1/(p-1)\}. \tag{3.1}$$

Once $\mathcal{X}_{-1}, \ldots, \mathcal{X}_i$ have been defined, we pick, if possible, an element

$$\mathbf{x}_0 \in \mathcal{X} - \bigcup_{h\le i} \mathcal{X}_h$$

such that $\mathrm{ord}_v(ax_0)$ is a minimum and define

$$\mathcal{X}_{i+1} = \{\mathbf{x} \in \mathcal{X} : \mathrm{ord}_v(ax_0) \le \mathrm{ord}_v(ax) \le \mathrm{ord}_v(ax_0) + M\}; \tag{3.2}$$

otherwise $\mathcal{X}_h = \emptyset$ for $h > i$.

The proof of (i) is a consequence of definitions (3.1) and (3.2), which imply that the range of $\mathrm{ord}_v(ax)$ for $x \in \mathcal{X}_i$ is contained in an interval of length M. In fact, let $\mathbf{x}, \mathbf{x}' \in \mathcal{X}_i$. Then we have, recalling (2.2) and Lemma 1,

$$\begin{aligned}|\log \|x'/x\|_v| = |\log \|(ax')/(ax)\|_v| &= (\log p)\cdot |\mathrm{ord}_v(ax') - \mathrm{ord}_v(ax)| \\ &\le (\log p)M \le \log((s+1)m(\mathcal{X}))\end{aligned}$$

proving (i).

Suppose that $\mathcal{X}'_i \ne \emptyset$. The inequality $\mathrm{rk}(\widetilde{\mathcal{X}}'_i) \ge \mathrm{rk}(\widetilde{\mathcal{X}}'_{i+1})$ for $i \ge 0$ is obvious, and our claim is that this inequality is strict:

$$\mathrm{rk}(\widetilde{\mathcal{X}}'_i) > \mathrm{rk}(\widetilde{\mathcal{X}}'_{i+1}). \tag{3.3}$$

Note that since $\mathrm{rk}(\widetilde{\mathcal{X}}'_1) \le \mathrm{rk}(\widetilde{\mathcal{X}}) \le s+1$ this implies that $\mathrm{rk}(\widetilde{\mathcal{X}}'_i) \le s+1-i$ and therefore $i \le s$ (note that if $\mathcal{Y} \ne \emptyset$ then $\mathrm{rk}(\widetilde{\mathcal{Y}}) \ge 1$). This implies that the filtration $\{\mathcal{X}'_i\}$ stops at $i = s+1$ with the empty set, whence

$$\mathcal{X} = \bigcup_{i=-1}^{s} \mathcal{X}_i.$$

Thus we need only prove (3.3).

The following argument embodies the new idea in [P]. Suppose $\mathrm{rk}(\widetilde{\mathcal{X}}'_{i+1}) = \mathrm{rk}(\widetilde{\mathcal{X}}'_i) = r$ and $i \geq 0$. Let T be an indeterminate which commutes with Γ. Then there are r elements $\mathbf{x}_h \in \mathcal{X}'_{i+1}$, $h = 1, \ldots, r$, such that $T\mathbf{x}_h$, $h = 1, \ldots, r$, are multiplicatively independent.

Let $\mathbf{x}_0 \in \mathcal{X}_i$ be an element of $\mathcal{X}_i$ for which $\mathrm{ord}_v(ax_0)$ is a minimum. Then the $r+1$ elements $T\mathbf{x}_h$, $h = 0, \ldots, r$, are multiplicatively dependent because $\mathrm{rk}(\widetilde{\mathcal{X}}'_i) = r$ by hypothesis. Therefore we have a relation

$$\prod_{h=0}^{r} (T\mathbf{x}_h)^{u_h} \in \mathrm{tors}(\Gamma) \tag{3.4}$$

for certain integers u_h, with $u_0 \neq 0$.

Such a relation is equivalent to solving in integers u_h, not all 0, the linear system of equations given by

$$\begin{aligned} &\sum_{h=0}^{r} n_1(T\mathbf{x}_h)u_h = 0, \\ &\sum_{h=0}^{r} n_2(T\mathbf{x}_h)u_h = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\sum_{h=0}^{r} n_r(T\mathbf{x}_h)u_h = 0, \end{aligned} \tag{3.5}$$

where the coefficients $n_\sigma(T\mathbf{x}_h)$ are determined as in (2.3) by a choice of generators of the group $\langle \widetilde{\mathcal{X}} \rangle$.

By construction, the two matrices

$$A_0 = \begin{pmatrix} n_1(T\mathbf{x}_0) & n_1(T\mathbf{x}_1) & n_1(T\mathbf{x}_2) & \ldots & n_1(T\mathbf{x}_r) \\ n_2(T\mathbf{x}_0) & n_2(T\mathbf{x}_1) & n_2(T\mathbf{x}_2) & \ldots & n_2(T\mathbf{x}_r) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ n_r(T\mathbf{x}_0) & n_r(T\mathbf{x}_1) & n_r(T\mathbf{x}_2) & \ldots & n_r(T\mathbf{x}_r) \end{pmatrix}$$

and

$$A_1 = \begin{pmatrix} n_1(T\mathbf{x}_1) & n_1(T\mathbf{x}_2) & \ldots & n_1(T\mathbf{x}_r) \\ n_2(T\mathbf{x}_1) & n_2(T\mathbf{x}_2) & \ldots & n_2(T\mathbf{x}_r) \\ \cdots & \cdots & \cdots & \cdots \\ n_r(T\mathbf{x}_1) & n_r(T\mathbf{x}_2) & \ldots & n_r(T\mathbf{x}_r) \end{pmatrix}$$

have the same rank r. We may solve (3.5) using Cramer's rule obtaining a solution u_h in which each u_h equals $(-1)^h$ times the determinant of the $r \times r$ minor of A_0 obtained by deleting the $(h+1)$th column. Note that we have $u_0 = \det(A_1) \neq 0$ because A_1 has maximal rank r.

It follows that

$$|u_0| = |\det(A_1)| \leq m(\widetilde{\mathcal{X}}'_{i+1}) \leq m(\widetilde{\mathcal{X}}).$$

This implies

$$0 \le \operatorname{ord}_v(u_0) \le M. \tag{3.6}$$

By (3.4) and the equation $ax + by = 1$ we get, specializing T to b,

$$\prod_{h=0}^{r} (1 - ax_h)^{u_h} \in \operatorname{tors}(k_v^\times). \tag{3.7}$$

The logarithm function

$$\log(1+x) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{1}{m} x^m$$

is well defined in the maximal ideal

$$\mathfrak{m}_v = \{x : \|x\|_v < 1\}$$

of the ring of integers of k_v, and provides a homomorphism

$$\log : 1 + \mathfrak{m}_v \to k_v^+$$

of the subgroup $1 + \mathfrak{m}_v \subset k_v^\times$ into the additive group k_v^+. The kernel of this homomorphism consists of the roots of unity in $1 + \mathfrak{m}_v$. We take the logarithm of (3.7), thus killing the torsion because v is a finite place. We obtain

$$\sum_{h=0}^{r} u_h \sum_{m=1}^{\infty} \frac{1}{m} (ax_h)^m = 0, \tag{3.8}$$

where now ax_h is understood as an element of k_v.

Since the points $\mathbf{x}_h$, $h = 0, \ldots, r$, satisfy

$$\operatorname{ord}_v(ax_h) > \frac{1}{p-1}$$

we get ([1]):

For $h = 0, \ldots, r$ the term ax_h is the unique term of lowest order in the series $\sum \frac{1}{m}(ax_h)^m$.

Since the elements in the sum (3.8) add up to 0, we see that there are two distinct terms with the same lowest order. Hence there are two distinct indices $h_0 < h_1$, with $u_{h_0} \neq 0$ and $u_{h_1} \neq 0$, such that $u_{h_0} ax_{h_0}$ and $u_{h_1} ax_{h_1}$ have the same lowest order, that is,

$$\operatorname{ord}_v(u_{h_0} ax_{h_0}) = \operatorname{ord}_v(u_{h_1} ax_{h_1}) = \min_h \operatorname{ord}_v(u_h ax_h); \tag{3.9}$$

([1]) A more refined argument shows that the condition $\operatorname{ord}_v(ax_h) \neq (p^l(p-1))^{-1}$ for $l = 0, 1, \ldots$ suffices for the validity of the arguments which follow.

moreover, since $u_0 \neq 0$, (3.6) implies

$$\min_h \operatorname{ord}_v(u_h a x_h) \le \operatorname{ord}_v(u_0 a x_0) \le \operatorname{ord}_v(a x_0) + M. \tag{3.10}$$

On the other hand, $\mathbf{x}_{h_1} \in \mathcal{X}'_{i+1}$ because $h_1 > h_0$, hence $h_1 \ge 1$. It follows that

$$\begin{aligned} \operatorname{ord}_v(a x_0) + M &< \operatorname{ord}_v(u_{h_1} a x_{h_1}) && \text{by (3.2)} \\ &\le \operatorname{ord}_v(u_0 a x_0) && \text{by (3.9)} \\ &\le \operatorname{ord}_v(a x_0) + M && \text{by (3.10)}. \end{aligned}$$

This is a contradiction, proving our claim.

If instead $\operatorname{ord}_v(ax) \le 0$ for every $\mathbf{x} \in \mathcal{X}$ we argue as follows. Consider the transformation of the unit equation $ax + by = 1$ into

$$\frac{1}{a} \cdot \frac{1}{x} + \left(-\frac{b}{a}\right) \frac{y}{x} = 1.$$

This transformation $(x, y) \mapsto (1/x, y/x)$, $(a, b) \mapsto (1/a, -b/a)$ maps Γ into an isomorphic group Γ', preserves ranks and mass, changes $\operatorname{ord}_v(ax)$ into $-\operatorname{ord}_v(ax)$ and leaves $|\log \|x'/x\|_v|$ invariant. Then we conclude by the same argument.

Local Cluster Principle (infinite places). *Let v be an infinite place of k. Let $j = 1$ or 2 and let $\mathcal{X}$ be a finite set of solutions of the unit equation $a_1 x_1 + a_2 x_2 = 1$ with $\mathbf{x} \in \Gamma$ such that $\log \|a_j x_j\|_v$ has fixed sign for every $\mathbf{x} \in \mathcal{X}$. Suppose $\operatorname{rk}(\mathcal{X}) \ge 1$. Then there is a decomposition*

$$\mathcal{X} = \bigcup_{i=-1}^{s} \mathcal{X}_i$$

with the following two properties. Let $\tau = |\operatorname{tors}(\Gamma)|$.

(i) *Clustering property: If $\mathcal{X}_i \neq \emptyset$ then for $\mathbf{x}, \mathbf{x}' \in \mathcal{X}_i$ we have*

$$|\log \|x'_j/x_j\|_v| \le \log(3s(s+1)\tau m(\mathcal{X}));$$

(ii) *Rank property: Let*

$$\mathcal{X}'_i = \bigcup_{h \ge i} \mathcal{X}_h.$$

Suppose that $i \ge 0$ and $\mathcal{X}'_i \neq \emptyset$. Then

$$\operatorname{rk}(\widetilde{\mathcal{X}}'_i) > \operatorname{rk}(\widetilde{\mathcal{X}}'_{i+1}).$$

Proof. It suffices to deal with the case $j = 1$. For notational simplicity, we write a, b, x, y for a_1, a_2, x_1, x_2 (so as to avoid double indices) and we also write

$$M = 3s(s+1)\tau m(\mathcal{X}).$$

We deal first with the case in which $\log \|ax\|_v \le 0$ for $\mathbf{x} \in \mathcal{X}$. We define $\mathcal{X}_i$ inductively for $i \ge -1$ as follows.

$$\mathcal{X}_{-1} = \{\mathbf{x} \in \mathcal{X} : M^{-1} \le \|ax\|_v \le 1\}. \tag{3.11}$$

Once $\mathcal{X}_{-1}, \ldots, \mathcal{X}_i$ have been defined, we pick, if possible, an element

$$\mathbf{x}_0 \in \mathcal{X} - \bigcup_{h \le i} \mathcal{X}_h$$

such that $\|ax_0\|_v$ is a maximum and define

$$\mathcal{X}_{i+1} = \{\mathbf{x} \in \mathcal{X} : M^{-1}\|ax_0\|_v \le \|ax\|_v \le \|ax_0\|_v\}; \tag{3.12}$$

otherwise $\mathcal{X}_h = \emptyset$ for $h > i$.

The proof of (i) is as in the finite case, and as in the finite case it remains only to prove the inequality $\mathrm{rk}(\widetilde{\mathcal{X}}_i') > \mathrm{rk}(\widetilde{\mathcal{X}}_{i+1}')$ for $i \ge 0$ and $\mathcal{X}_i' \ne \emptyset$.

Again, suppose $\mathrm{rk}(\widetilde{\mathcal{X}}_{i+1}') = \mathrm{rk}(\widetilde{\mathcal{X}}_i') = r$. Let $\mathbf{x}_0 \in \mathcal{X}_i$ be an element of $\mathcal{X}_i$ for which $\|ax_0\|_v$ is a maximum. Then the $r+1$ elements $T\mathbf{x}_h$, $h = 0, \ldots, r$, are multiplicatively dependent because $\mathrm{rk}(\widetilde{\mathcal{X}}_i') = r$ by hypothesis. As in the finite case, we get a relation

$$\prod_{h=0}^{r} (T\mathbf{x}_h)^{u_h} \in \mathrm{tors}(\Gamma)$$

for certain integers u_h, with $u_0 \ne 0$ and

$$\max |u_h| \le m(\widetilde{\mathcal{X}}) \le (3s\tau)^{-1}M. \tag{3.13}$$

As before, this implies

$$\prod_{h=0}^{r} (1 - ax_h)^{\tau u_h} = 1. \tag{3.14}$$

We take the logarithm of (3.14) and find

$$\sum_{h=0}^{r} u_h \sum_{m=1}^{\infty} \frac{1}{m}(ax_h)^m \in \frac{2\pi i}{\tau} \cdot \mathbb{Z}. \tag{3.15}$$

For simplicity we write $|\ |$ instead of $\|\ \|_v$. We have

$$|ax_0| < 1/M \quad \text{and} \quad |ax_h| \le |ax_0|/M < 1/M^2 \quad \text{for } h \ge 1. \tag{3.16}$$

Then we estimate

$$\begin{aligned}\Big|\sum_{h=0}^{r} u_h \sum_{m=1}^{\infty} \frac{1}{m}(ax_h)^m\Big| &\le \sum_{h=0}^{r} |u_h| \sum_{m=1}^{\infty} \frac{1}{m}|ax_h|^m \le (3s\tau)^{-1}M \sum_{h=0}^{r} \frac{|ax_h|}{1 - |ax_h|} \\ &\le (3s\tau)^{-1}M\Big(\frac{1}{M-1} + \frac{r}{M^2-1}\Big) \\ &< (3s\tau)^{-1}\frac{2M}{M-1} < \frac{2\pi}{\tau}.\end{aligned}$$

In view of (3.15) we conclude

$$\sum_{h=0}^{r} u_h \sum_{m=1}^{\infty} \frac{1}{m}(ax_h)^m = 0. \tag{3.17}$$

On the other hand, we have, using (3.12), (3.16), $u_0 \neq 0$ and $M \geq 6$,

$$\begin{aligned}\Big|\sum_{h=0}^{r} u_h \sum_{m=1}^{\infty} \frac{1}{m}(ax_h)^m\Big| &> |ax_0|\Big(1 - \frac{1}{2}\cdot\frac{|ax_0|}{1-|ax_0|}\Big) - (3s\tau)^{-1}M\sum_{h=1}^{r}\frac{|ax_0|}{M-1} \\ &> |ax_0|\Big(1 - \frac{1}{2(M-1)} - (3s\tau)^{-1}\frac{(s+1)M}{M-1}\Big) \\ &\geq |ax_0|\big(1 - \tfrac{1}{10} - \tfrac{4}{5}\big) > 0.\end{aligned}$$

This contradicts (3.17), and completes the proof of our claim.

The proof for the case in which $\|ax\|_v \geq 1$ is the same as in the finite case.

Our next goal is to obtain a global version of the cluster principle, as well as its generalization to sets of solutions where we may have $\|ax\|_v \geq 1$.

4. Regulators. Let Γ be a finitely generated group $\Gamma \subset \mathbb{G}_m^n(k)$ of rank s, and let $\boldsymbol{\gamma}_\sigma$, $\sigma = 1,\ldots,s$, be a set of generators of Γ up to torsion. We use vector notation, so $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_n)$.

DEFINITION 3. Let S be a subset of cardinality s of $M_k \times \{1,\ldots,n\}$. The *$S$-regulator* $R_S(\Gamma)$ of Γ is by definition

$$R_S(\Gamma) = |\det(\log\|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\ldots,s\\ (v,j)\in S}}|.$$

The set S is said to be *nondegenerate* with respect to Γ if $R_S(\Gamma) \neq 0$.

The S-regulator does not depend on the choice of generators of Γ.

LEMMA 2. *A nondegenerate set S for Γ always exists.*

Proof. If we had $R_S(\Gamma) = 0$ for every S we would have

$$\operatorname{rank}(\log\|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\ldots,s\\ (v,j)\in M_k\times\{1,\ldots,n\}}} < s. \tag{4.1}$$

Then there would be a relation of linear dependence among the rows of this matrix:

$$\sum_{\sigma=1}^{s} a_\sigma \log\|\gamma_{\sigma j}\|_v = 0 \tag{4.2}$$

for $(v,j) \in M_k \times \{1,\ldots,n\}$, with not all $a_\sigma = 0$. Moreover, we see that we may assume that this relation has coefficients in $\mathbb{Z}$ rather than in $\mathbb{R}$. In fact, let h be the class number of k. Let $\eta_1,\ldots,\eta_t$ be a basis of the units of k modulo torsion and, for each finite place w, let $\mathfrak{p}_w$ be the associated prime ideal and π_w be a generator of the principal ideal $\mathfrak{p}_w^h = (\pi_w)$. We can write, with obvious vector notation,

$$\boldsymbol{\gamma}_\sigma^h = \boldsymbol{\varepsilon}_\sigma \prod_{i=1}^{t} \eta_i^{\mathbf{m}_{i\sigma}} \prod_{w \text{ finite}} \pi_w^{\mathbf{n}_{w\sigma}}$$

with $\boldsymbol{\varepsilon}_\sigma$ a torsion element and $\mathbf{n}_{w\sigma} = 0$ for almost all w. Thus (4.2) becomes

$$\sum_{\sigma=1}^{s} \Big(\sum_{i=1}^{t} \mathbf{m}_{i\sigma} \log \|\eta_i\|_v + \sum_{w \text{ finite}} \mathbf{n}_{w\sigma} \log \|\pi_w\|_v \Big) a_\sigma = 0. \tag{4.3}$$

We have $\log \|\eta_i\|_v = 0$ for every i and every finite v, and also $\log \|\pi_w\|_v = 0$ if v is finite and $v \neq w$, while $\log \|\pi_v\|_v \neq 0$. Hence, for all finite v, (4.3) is equivalent to

$$\sum_{\sigma=1}^{s} \mathbf{n}_{v\sigma} a_\sigma = 0. \tag{4.4}$$

Now we can use (4.4) to simplify and rewrite (4.3) as

$$\sum_{i=1}^{t} \log \|\eta_i\|_v \Big(\sum_{\sigma=1}^{s} \mathbf{m}_{i\sigma} a_\sigma \Big) = 0 \tag{4.5}$$

for every infinite place v.

The matrix $(\log \|\eta_i\|_v)$, $i = 1, \ldots, t$, $v \mid \infty$, has rank t by Dirichlet's Unit Theorem, therefore (4.5) is equivalent to

$$\sum_{\sigma=1}^{s} \mathbf{m}_{i\sigma} a_\sigma = 0 \tag{4.6}$$

for $i = 1, \ldots, t$. Since (4.4) and (4.6) form a system with rational integral coefficients equivalent to (4.3), we see that if (4.2) has a nontrivial solution in $\mathbb{R}$ it also has a nontrivial solution in $\mathbb{Z}$.

Let $a_\sigma \in \mathbb{Z}$, $\sigma = 1, \ldots, s$, be a nontrivial solution of (4.2). Then we see that

$$\Big\| \prod_{\sigma=1}^{s} \boldsymbol{\gamma}_\sigma^{a_\sigma} \Big\|_v = \prod_{\sigma=1}^{s} \|\boldsymbol{\gamma}_\sigma\|_v^{a_\sigma} = \{1\}$$

for every $v \in M_k$. On the other hand, by a result which goes back to Kronecker, we know that any element $\alpha \in k^\times$ with $\|\alpha\|_v = 1$ for every v is a root of unity. It follows that

$$\prod_{\sigma=1}^{s} \boldsymbol{\gamma}_\sigma^{a_\sigma} \in \operatorname{tors}(\Gamma),$$

contradicting the fact that Γ has rank s and completing the proof.

Definition 4. The *support* $\mathrm{supp}(\Gamma)$ of Γ consists of all places v of k such that $\log\|\gamma_j\|_v \neq 0$ for some $j \in \{1,\ldots,n\}$ and some $\boldsymbol{\gamma} \in \Gamma$.

It is clear that $\mathrm{supp}(\Gamma)$ is a finite set because Γ is finitely generated.

Lemma 3. *Let* $\mathrm{H} \subset \Gamma$ *be a subgroup of* Γ. *We have* $\mathrm{supp}(\mathrm{H}) \subseteq \mathrm{supp}(\Gamma)$ *and equality holds if* $\mathrm{rk}(\mathrm{H}) = \mathrm{rk}(\Gamma)$. *Moreover*, *if* $\mathrm{rk}(\mathrm{H}) = \mathrm{rk}(\Gamma)$ *then*

$$R_S(\mathrm{H}) = [\Gamma/\mathrm{tors} : \mathrm{H}/\mathrm{tors}] R_S(\Gamma).$$

Proof. The inclusion $\mathrm{supp}(\mathrm{H}) \subseteq \mathrm{supp}(\Gamma)$ is clear.

The rank of H is at most the rank of Γ ([MKS], Th. 4.5, Cor. 4.5.2 and Th. 4.1, p. 146). If H and Γ have the same rank then H has finite index in Γ ([MKS], Th. 4.1, p. 146). Let $v \in \mathrm{supp}(\Gamma)$ and let $\boldsymbol{\gamma} \in \Gamma$ and j be such that $\log\|\gamma_j\|_v \neq 0$. The powers $\boldsymbol{\gamma}^m$ of $\boldsymbol{\gamma}$ fall into finitely many cosets of H in Γ, therefore there are two distinct powers in a same coset and, taking their quotient, there is a power $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^m \in \mathrm{H}$, with $m \neq 0$. Now $\log\|\gamma'_j\|_v = m \log\|\gamma_j\|_v \neq 0$ and $v \in \mathrm{supp}(\mathrm{H})$, as asserted. The final statement of the lemma is also clear.

Let $\mathcal{O}_k^\times$ be the group of units of k and define

$$\Gamma_\infty = \Gamma \cap (\mathcal{O}_k^\times)^n.$$

Then there is a free subgroup Γ_0 of Γ such that

$$\Gamma = \Gamma_\infty \Gamma_0 \quad \text{and} \quad \Gamma_0 \cap (\mathcal{O}_k^\times)^n = \{1\}.$$

To see this, let $r = \mathrm{rk}(\Gamma_\infty)$. By the structure theorem for finitely generated abelian groups (see e.g. [MKS], Cor. 4.5.2, p. 146), there are generators $\boldsymbol{\gamma}_\sigma$, $\sigma = 1,\ldots,s$, of Γ up to torsion and integers d_i such that $\boldsymbol{\gamma}_1^{d_1},\ldots,\boldsymbol{\gamma}_r^{d_r}$ are generators of Γ_∞ up to torsion, with d_i dividing d_{i+1}. On the other hand, if γ^d is a unit then γ itself is a unit, hence we must have $d_1 = \ldots = d_r = 1$. Also $\mathrm{tors}(\Gamma) \subset \Gamma_\infty$. Thus we may take

$$\Gamma_\infty = \langle \boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_r\rangle \cdot \mathrm{tors}(\Gamma) \quad \text{and} \quad \Gamma_0 = \langle \boldsymbol{\gamma}_{r+1},\ldots,\boldsymbol{\gamma}_s\rangle.$$

Although this decomposition is not canonical, the regulator behaves nicely with respect to it. We have

Lemma 4. *Let* $\Gamma = \Gamma_\infty\Gamma_0$ *be as before and let* $S \subset M_k \times \{1,\ldots,n\}$ *be nondegenerate with respect to* Γ. *Let* S_∞ *and* S_0 *be the subsets of* S *consisting of* (v,j) *with* $v \mid \infty$ *and* $v \nmid \infty$ *respectively*, *and let us say that* S *is special if their cardinality equals the ranks of* Γ_∞ *and* Γ_0. *Then*

(i) *there is a nondegenerate and special set* S *for* Γ;

(ii) *for special* S *we have* $R_S(\Gamma) = R_{S_\infty}(\Gamma_\infty) R_{S_0}(\Gamma_0)$.

Proof. If S_∞ and S_0 are nondegenerate for Γ_∞ and Γ_0 then $S = S_\infty \cup S_0$ is nondegenerate for Γ. A nondegenerate S_∞ for Γ_∞ may consist only of (v,j) with $v \mid \infty$ hence it is automatically special with respect to Γ_∞.

Therefore, to complete the proof of (i) we need only show that there is a nondegenerate special set $S_0 = \{(v,j)\}$ for Γ_0, in other words, such that $v \nmid \infty$ for $(v,j) \in S_0$.

Suppose this is not the case. Let $\{\boldsymbol{\gamma}_\sigma\}$, $\sigma = 1, \ldots, s_0$, be generators for Γ_0; then the matrix

$$(\log \|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\ldots,s_0 \\ (v,j)\in\{v\nmid\infty\}\times\{1,\ldots,n\}}}$$

would have rank strictly less than s_0. As in the proof of Lemma 2, this would give us a nontrivial element $\boldsymbol{\gamma} \in \Gamma_0$ such that $\|\gamma_j\|_v = 1$ for $(v,j) \in \{v\nmid\infty\} \times \{1, \ldots, n\}$. Hence $\boldsymbol{\gamma} \in (\mathcal{O}_k^\times)^n$, contradicting the definition of Γ_0. This proves statement (i).

For the second statement we proceed as follows. Let $\Gamma_\infty = \langle \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_r \rangle \times$ tors and $\Gamma_0 = \langle \boldsymbol{\gamma}_{r+1}, \ldots, \boldsymbol{\gamma}_s \rangle$. Since $\gamma_{\sigma j}$ is a unit for $\sigma = 1, \ldots, r$ and $j = 1, \ldots, n$ we have $\log \|\gamma_{\sigma j}\|_v = 0$ for every (v,j) with finite v and $\sigma = 1, \ldots, r$; it follows that the matrix $(\log \|\gamma_{\sigma j}\|_v)_{(v,j)\in S}$ has a block structure as

$$\begin{pmatrix} (\log \|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\ldots,r \\ (v,j)\in S_\infty}} & 0 \\ * & (\log \|\gamma_{\sigma j}\|_v)_{\substack{\sigma=r+1,\ldots,s \\ (v,j)\in S_0}} \end{pmatrix}.$$

Since we assume that S is special, the two blocks containing the diagonal are square blocks and the result follows upon taking determinants.

We conclude this section with an important definition.

Definition 5. Let $T \subset M_k \times \{1, \ldots, n\}$ and let $\mathcal{X} \neq \emptyset$ be a finite subset of Γ. Then we define

$$\nu_T(\mathcal{X}) = \max_{\mathbf{x}\in\mathcal{X}} \max_{(v,j)\in T} |\log \|x_j\|_v|.$$

5. Finding good generators for Γ. The following result gives us good generators for Γ.

Lemma 5. *Let T be a subset of $M_k \times \{1, \ldots, n\}$, of cardinality t, containing a nondegenerate subset for Γ. Then we can find generators $\boldsymbol{\gamma}_\sigma$ of Γ such that, with S denoting s-subsets of T,*

$$t^{-s} \prod_{\sigma=1}^{s} \max_{(v,j)\in T} |\log \|\gamma_{\sigma j}\|_v| \le \max_{S\subset T} R_S(\Gamma) \le s^s \max_{S\subset T} \prod_{\sigma=1}^{s} \max_{(v,j)\in S} |\log \|\gamma_{\sigma j}\|_v|.$$

Proof. Let $\boldsymbol{\gamma}_\sigma$, $\sigma = 1, \ldots, s$, be a set of generators of Γ up to torsion and let $\mathcal{T}$ be the box in $\mathbb{R}^s$ defined by

$$\mathcal{T} = \Big\{\mathbf{y} \in \mathbb{R}^s : \max_{(v,j)\in T} \Big| \sum_{\sigma=1}^{s} \log \|\gamma_{\sigma j}\|_v y_\sigma \Big| \le 1\Big\}.$$

By the Cube Slicing Theorem of Vaaler ([V], Th. 1, p. 543) the volume of $\mathcal{T}$ is bounded below by

$$\mathrm{Vol}(\mathcal{T}) \geq 2^s\{\det((\log\|\gamma_{\sigma j}\|_v)'_{\substack{\sigma=1,\dots,s\\(v,j)\in T}} \cdot (\log\|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\dots,s\\(v,j)\in T}})\}^{-1/2} \tag{5.1}$$

with $'$ denoting the transpose of the matrix $A = (\log\|\gamma_{\sigma j}\|_v)$.

By the Cauchy–Binet formula, we have

$$\det(A'\cdot A) = \sum_{S\subset T}\det(\log\|\gamma_{\sigma j}\|_v)^2_{\substack{\sigma=1,\dots,s\\(v,j)\in S}} \leq \binom{t}{s}\max_{S\subset T} R_S^2(\Gamma).$$

If we combine this inequality with the preceding lower bound (5.1) for the volume we get

$$\mathrm{Vol}(\mathcal{T}) \geq 2^s\binom{t}{s}^{-1/2}(\max_{S\subset T} R_S(\Gamma))^{-1}. \tag{5.2}$$

Let $\lambda_1,\dots,\lambda_s$ be the successive minima of $\mathcal{T}$ with respect to the standard lattice $\mathbb{Z}^s$. By a theorem of Mahler (see for instance [C2], Ch. V, Lemma 8, p. 135), there is a basis $\mathbf{y}_1,\dots,\mathbf{y}_s$ of $\mathbb{Z}^s$ such that (2)

$$\max_{(v,j)\in T}\Big|\sum_{\varrho=1}^{s}\log\|\gamma_{\varrho j}\|_v\cdot y_{\sigma\varrho}\Big| \leq \sigma\lambda_\sigma. \tag{5.3}$$

This basis yields new generators $\boldsymbol{\gamma}'_\sigma$ defined by

$$\boldsymbol{\gamma}'_\sigma = \prod_{\varrho=1}^{s}\boldsymbol{\gamma}_\varrho^{y_{\sigma\varrho}}, \qquad \sigma = 1,\dots,s,$$

such that

$$\log\|\gamma'_{\sigma j}\|_v = \sum_{\varrho=1}^{s} y_{\sigma\varrho}\log\|\gamma_{\varrho j}\|_v \tag{5.4}$$

for every $v\in M_k$ and $j\in\{1,\dots,n\}$. In terms of these generators, we can rewrite (5.3) as

$$\max_{(v,j)\in T}|\log\|\gamma'_{\sigma j}\|_v| \leq \sigma\lambda_\sigma \quad \text{for } \sigma = 1,\dots,s. \tag{5.5}$$

By Minkowski's Second Theorem in the Geometry of Numbers we have

$$\lambda_1\lambda_2\dots\lambda_s\,\mathrm{Vol}(\mathcal{T}) \leq 2^s$$

and hence by (5.2) we get

$$\lambda_1\cdot(2\lambda_2)\dots(s\lambda_s) \leq s!\binom{t}{s}^{1/2}\max_{S\subset T} R_S(\Gamma) \leq t^s\max_{S\subset T} R_S(\Gamma).$$

This and (5.3) prove the lower bound in Lemma 5.

(2) One can take $\max(1,\sigma/2)\lambda_\sigma$ in place of $\sigma\lambda_\sigma$, but such improvements are irrelevant here.

Since regulators do not depend on the choice of generators we may use the generators $\boldsymbol{\gamma}'$ satisfying (5.5) to define the regulator $R_S(\Gamma)$. Now $R_S(\Gamma)$ can be estimated from above using Laplace's expansion.

A set of generators satisfying the conclusion of Lemma 5 will be referred to as a set of *good generators* relative to T.

We need lower bounds for regulators. This is provided by

LEMMA 6. *Let $G = \mathbb{G}_m^n(k)$ and let Γ be a finitely generated subgroup of G of rank s. Let $\Gamma = \Gamma_\infty \Gamma_0$ be the decomposition considered in Lemma 4 and let $S = S_\infty \cup S_0$ be a special nondegenerate subset of $M_k \times \{1, \ldots, n\}$ relative to this decomposition. Then*

$$R_S(\Gamma) \geq (9nd^5)^{-s}.$$

More generally, for every subgroup H *of Γ of positive rank r there is a subset $\Sigma \subseteq S$ which is nondegenerate and special for* H, *such that*

$$R_\Sigma(\mathrm{H}) \geq (9nd^5)^{-r}.$$

Moreover, if $\Gamma_0 \neq \{1\}$ let P be the largest prime such that there is v_0 in the support of Γ with $v_0 \mid P$. Then there is a special nondegenerate S for Γ with $v_0 \in S_0$, and we have the improved lower bound

$$R_S(\Gamma) \geq (9nd^5)^{-s}(\log P).$$

Further, suppose that for some integer $m \geq 1$ we have

$$\Gamma \subset G_m = \{\varepsilon g^m : g \in G,\ \varepsilon \in \mathrm{tors}(G)\},$$

or in other words, suppose that elements of Γ are mth powers up to torsion. Then the lower bounds for $R_S(\Gamma)$ and $R_\Sigma(\mathrm{H})$ can be improved by a factor m^s and m^r respectively.

Proof. By Lemma 4 we have

$$R_S(\Gamma) = R_{S_\infty}(\Gamma_\infty) R_{S_0}(\Gamma_0)$$

so it suffices to prove the result separately in the two cases $\Gamma = \Gamma_0$, $S = S_0$ and $\Gamma = \Gamma_\infty$, $S = S_\infty$; the statement about subgroups follows from the fact that if S is nondegenerate and special for Γ and H is a subgroup of Γ then there is a subset $\Sigma \subset S$ which is nondegenerate and special for H.

To obtain a lower bound for the case $\Gamma = \Gamma_0$, $S = S_0$ we note that for $a \in k^\times$ and finite v we have

$$\log \|a\|_v \in \frac{1}{e_v} (\log p_v) \cdot \mathbb{Z}$$

where p_v is the rational prime such that $v \mid p_v$ and e_v is the ramification index. Hence

$$R_S(\Gamma) = |\det(\log \|\gamma_{\sigma j}\|_v)_{\substack{\sigma=1,\ldots,s \\ (v,j)\in S}}| \in \Big(\prod_{(v,j)\in S} \frac{\log p_v}{e_v} \Big) \cdot \mathbb{Z}. \tag{5.6}$$

Since every v in $(v,j)\in S$ is a finite place, we get

$$R_S \ge ((\log 2)/d)^{-s}.$$

Moreover, we can choose S such that there is an element $(v_0,i)\in S$ with $v_0\,|\,P$, in which case there is a factor $(\log P)/e_{v_0} \ge (\log P)/d$ in the right-hand side of (5.6). *A fortiori*, this proves the required lower bounds if $\Gamma=\Gamma_0$, $S=S_0$.

Now consider the case $\Gamma=\Gamma_\infty$, $S=S_\infty$. In this case we need only consider the set $T=\{v\,|\,\infty\}\times\{1,\ldots,n\}$ since $\log\|\gamma_{\sigma j}\|_v=0$ if $(v,j)\notin T$.

By Lemma 5, noting that $|T|\le nd$, we get a set $S\subset T$ of cardinality s and generators $\boldsymbol{\gamma}_\sigma$ up to torsion, such that

$$(5.7)\qquad (nd)^{-s}\prod_{\sigma=1}^{s}\max_{(v,j)\in T}|\log\|\gamma_{\sigma j}\|_v| \le R_S(\Gamma).$$

By a result of Dobrowolski [D], we have $h(\gamma_{\sigma j})\ge 1/(9d^3)$ because $\gamma_{\sigma j}$ is not a root of unity. Since $\log\|\gamma_{\sigma j}\|\ne 0$ may occur only for $v|\infty$ we deduce that

$$\max_{(v,j)\in T}|\log\|\gamma_{\sigma j}\|_v| \ge (9d^4)^{-1}$$

for $\sigma=1,\ldots,s$.

This inequality and (5.7) give

$$R_S(\Gamma)\ge (9nd^5)^{-s}.$$

This completes the proof of the stated lower bounds for $R_S(\Gamma)$.

The last statement of Lemma 6 is easy to prove. Let $\boldsymbol{\gamma}_\sigma$ be generators of Γ up to torsion. If $\Gamma\subset G_m$ we can write $\boldsymbol{\gamma}_\sigma=\boldsymbol{\varepsilon}_\sigma\boldsymbol{\eta}_\sigma^m$. Let $\Gamma'=\langle\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_s\rangle\times$ tors(Γ). Then we have $[\Gamma':\Gamma]=m^s$ and there is a subgroup $\mathrm{H}'\supseteq\mathrm{H}$ of Γ' such that H has index $[\mathrm{H}':\mathrm{H}]=m^r$ in H'. By Lemma 3 we have $R_S(\Gamma)=m^sR_S(\Gamma')$ and $R_\Sigma(\mathrm{H})=m^rR_\Sigma(\mathrm{H}')$, and the result follows by applying the preceding lower bounds to Γ' and H'.

We use the results we have proved on regulators to compare the mass and height of a set $\mathcal{X}\subset\Gamma$. We begin by giving an upper bound for the mass in terms of the height.

LEMMA 7. *Let $\mathcal{X}\subset\Gamma$ and let $T\subset M_k\times\{1,\ldots,n\}$ contain a nondegenerate subset Σ for $\mathcal{X}$. Let $\mathrm{H}=\langle\mathcal{X}\rangle$ and suppose $r=\mathrm{rk}(\mathcal{X})\ge 1$. Then*

$$m(\mathcal{X})\le R_\Sigma(\mathrm{H})^{-1}(r\nu_T(\mathcal{X}))^r.$$

In particular, if $\Gamma\subset G_m$ and T contains a nondegenerate special set for Γ

$$m(\mathcal{X})\le\left(\frac{9nd^5r}{m}\nu_T(\mathcal{X})\right)^r.$$

P r o o f. Let $\mathrm{H}=\langle\mathcal{X}\rangle$ and let Σ be a nondegenerate set for H. Let $r=\mathrm{rk}(\mathcal{X})$ and let $\mathbf{x}_i=(x_{i1},\ldots,x_{in})\in\mathcal{X}$, $i=1,\ldots,r$, be r elements of $\mathcal{X}$.

Then

$$|\det(\log\|x_{ij}\|_v)_{\substack{i=1,\dots,r\\(v,j)\in\Sigma}}| = R_\Sigma(\mathrm{H})\cdot|\det(n_\sigma(\mathbf{x}_i))_{\substack{i=1,\dots,r\\ \sigma=1,\dots,r}}|. \tag{5.8}$$

It follows that the maximum of this quantity over all choices of $\mathbf{x}_i\in\mathcal{X}$, $i=1,\dots,r$, is

$$R_\Sigma(\mathrm{H})\cdot m(\mathcal{X}). \tag{5.9}$$

On the other hand, by the Laplace expansion we have

$$|\det(\log\|x_{ij}\|_v)_{\substack{i=1,\dots,r\\(v,j)\in\Sigma}}| \le (r\max_{\mathbf{x}\in\mathcal{X}}\max_{(v,j)\in\Sigma}|\log\|x_j\|_v|)^r \le (r\nu_T(\mathcal{X}))^r. \tag{5.10}$$

If we combine (5.8), (5.9) and (5.10) we deduce ([3])

$$m(\mathcal{X}) \le R_\Sigma(\mathrm{H})^{-1}(r\nu_T(\mathcal{X}))^r, \tag{5.11}$$

which is the first statement of the lemma.

For the second statement we fix a nondegenerate and special set $S\subset T$ for Γ and restrict our attention to Σ nondegenerate and special for H. Lemma 6 provides a lower bound for $R_\Sigma(\mathrm{H})$, and we are done.

The following result will be used to provide control of the height.

LEMMA 8. *Let $T\subset M_k\times\{1,\dots,n\}$ contain a subset S^* such that $R=R_{S^*}(\Gamma)$ is a maximum. Let t be the cardinality of* $\mathrm{supp}(\Gamma)$ *and let $\mathcal{X}\subset\Gamma$ be nonempty. Then*

$$h(\mathcal{X}) \le (nst)^{s+1}\nu_T(\mathcal{X}).$$

Proof. Let $T=\mathrm{supp}(\Gamma)\times\{1,\dots,n\}$. By definition of support, for $\boldsymbol{\gamma}\in\Gamma$ we have $\log\|\gamma_j\|_v=0$ unless $(v,j)\in T$. Now Lemma 5 gives us generators $\boldsymbol{\gamma}_\sigma$ of Γ up to torsion such that

$$(nt)^{-s}\prod_{\sigma=1}^{s}\max_{\text{all }(v,j)}|\log\|\gamma_{\sigma j}\|_v| \le R. \tag{5.12}$$

For $\mathbf{x}\in\mathcal{X}$ consider the linear system of equations

$$\sum_{\sigma=1}^{s}(\log\|\gamma_{\sigma j}\|_v)\,n_\sigma(\mathbf{x}) = \log\|x_j\|_v \tag{5.13}$$

for $(v,j)\in S^*$, with unknowns $n_\sigma(\mathbf{x})$. By Cramer's rule we obtain

$$R|n_\sigma(\mathbf{x})| = \pm\det(\log\|x_j\|_v,\log\|\gamma_{\varrho j}\|_v)_{\substack{\varrho\ne\sigma\\(v,j)\in S^*}}. \tag{5.14}$$

([3]) We may replace $r\nu_T(\mathcal{X})$ by $\sqrt{r}\nu_T(\mathcal{X})$ if we use Lord Kelvin's inequality (also more widely known as Hadamard's inequality).

By the lower bound in (5.12) and the Laplace expansion of the determinant in (5.14) we deduce

$$(nt)^{-s}|n_\sigma(\mathbf{x})|\Big\{\prod_{\varrho=1}^{s}\max_{\text{all }(v,j)}|\log\|\gamma_{\varrho j}\|_v|\Big\}$$
$$\le s^s\big(\max_{(v,j)\in S^*}|\log\|x_j\|_v|\big)\Big\{\prod_{\varrho\ne\sigma}\max_{(v,j)\in S^*}|\log\|\gamma_{\varrho j}\|_v|\Big\}$$

and *a fortiori* we deduce

(5.15) $$|n_\sigma(\mathbf{x})|\max_{\text{all }(v,j)}|\log\|\gamma_{\sigma j}\|_v|\le(nst)^s\max_{(v,j)\in S^*}|\log\|x_j\|_v|.$$

We have, using (5.13) and (5.15),

$$h(\mathbf{x})=\tfrac12\sum_{v\in\operatorname{supp}(\Gamma)}\frac{d_v}{d}\max_j|\log\|x_j\|_v|\le\tfrac12 t\max_{\text{all }(v,j)}|\log\|x_j\|_v|$$
$$\le\tfrac12 t\max_{\text{all }(v,j)}\sum_{\sigma=1}^{s}|n_\sigma(\mathbf{x})|\cdot|\log\|\gamma_{\sigma j}\|_v|\le(nst)^{s+1}\max_{(v,j)\in S^*}|\log\|x_j\|_v|.$$

This proves the lemma.

6. The Global Cluster Principle. We begin with an extension of the Local Cluster Principle to several places. We need the following definition of *signature* of a solution $\mathbf{x}=(x_1,x_2)$ of $a_1x_1+a_2x_2=1$.

DEFINITION 6. Let T be a finite subset of $M_k\times\{1,2\}$. The *signature* $\boldsymbol{\varepsilon}_T(\mathbf{x})$ of $\mathbf{x}=(x_1,x_2)$ relative to T is the vector

$$\boldsymbol{\varepsilon}_T(\mathbf{x})=\{\operatorname{sign}\log\|a_jx_j\|_v:(v,j)\in T\},$$

where by convention $\operatorname{sign}0=1$.

Let us denote by $w=(v,j)$ elements of T and let $\mathbf{i}$ denote a vector $\mathbf{i}=(i_w:w\in T)$ with integer entries i_w satisfying $-1\le i_w$. We define

$$\|\mathbf{i}\|=\sum_{w\in T}\max(0,i_w).$$

We have a partial ordering on the set of such vectors, namely $\mathbf{i}\le\mathbf{i}'$ if and only if $i_w\le i'_w$ for every $w\in T$.

GLOBAL CLUSTER PRINCIPLE. *Let $\Gamma\subset\mathbb{G}_m^2(k)$ be a finitely generated group of rank $s\ge1$ and let $\mathcal{X}$ be a finite set of solutions $(x_1,x_2)\in\Gamma$ of the generalized unit equation $a_1x_1+a_2x_2=1$. Let $T=\{(v,j)\}$ be a finite subset of $M_k\times\{1,2\}$. Suppose also that the signature of elements of $\mathcal{X}$ relative to T is constant, i.e. $\boldsymbol{\varepsilon}_T(\mathbf{x})=\boldsymbol{\varepsilon}_T$ for all $\mathbf{x}\in\mathcal{X}$, for some vector*

$\boldsymbol{\varepsilon}_T = \{\varepsilon_w : w \in T\}$ *with* $\varepsilon_w = -1, 1$. *Then there is a partition*

$$\mathcal{X} = \bigcup \mathcal{X}_{\mathbf{i}}$$

indexed by vectors $\mathbf{i} = \{i_\nu\}_{\nu \in T}$, $i_\nu \geq -1$, *with the following properties.*

(i) *Clustering property*: *If* $\mathcal{X}_{\mathbf{i}} \neq \emptyset$ *then for* $\mathbf{x} \in \mathcal{X}_{\mathbf{i}}$ *we have*

$$\nu_T(\mathbf{x}^{-1}\mathcal{X}_i) \leq \log(6s^2\tau m(\mathcal{X})).$$

(ii) *Rank property*: *Let*

$$\mathcal{X}'_{\mathbf{i}} = \bigcup_{\mathbf{h} \geq \mathbf{i}} \mathcal{X}_{\mathbf{h}}.$$

Suppose that $\mathcal{X}_{\mathbf{i}} \neq \emptyset$. *Then*

$$\mathrm{rk}(\widetilde{\mathcal{X}}'_{\mathbf{i}}) \leq \mathrm{rk}(\widetilde{\mathcal{X}}) - \|\mathbf{i}\|.$$

P r o o f (by induction on the cardinality of T). We apply the Local Cluster Principle as follows.

If T is empty, there is nothing to prove.

Suppose the statement is true for T and let us prove it for $T^* = T \cup \{(v, j)\}$. We abbreviate $w = (v, j)$.

We define $\mathcal{X}'_{\mathbf{i},-1} = \mathcal{X}'_{\mathbf{i}}$ and apply the Local Cluster Principle at $w = (v, j)$ to the set $\mathcal{X}'_{\mathbf{i},-1}$. This gives a partition

$$\mathcal{X}'_{\mathbf{i},-1} = \bigcup_{h=-1}^{s} \mathcal{X}_{\mathbf{i},h}$$

and sets

$$\mathcal{X}'_{\mathbf{i},i'} = \bigcup_{h \geq i'} \mathcal{X}_{\mathbf{i},h}$$

such that (i) holds and $\mathrm{rk}(\widetilde{\mathcal{X}}'_{\mathbf{i},i'}) > \mathrm{rk}(\widetilde{\mathcal{X}}'_{\mathbf{i},i'+1})$ provided $\widetilde{\mathcal{X}}'_{\mathbf{i},i'} \neq \emptyset$.

To verify (ii), it suffices to note that, by the induction hypothesis, we have

$$\begin{aligned}\mathrm{rk}(\widetilde{\mathcal{X}}'_{\mathbf{i},i'}) &\leq \mathrm{rk}(\widetilde{\mathcal{X}}'_{\mathbf{i}}) - \max(0, i') \leq \mathrm{rk}(\widetilde{\mathcal{X}}) - \|\mathbf{i}\| - \max(0, i') \\ &= \mathrm{rk}(\widetilde{\mathcal{X}}) - \|\{\mathbf{i}, i'\}\|.\end{aligned}$$

This proves the result.

For applications, we need an upper bound for the number of sets $\mathcal{X}_{\mathbf{i}}$.

Lemma 9. *The number of sets* $\mathcal{X}_{\mathbf{i}}$ *does not exceed* $2^{s+2|T|}$.

P r o o f. Let us write $t = |T|$. If $\mathcal{X}_{\mathbf{i}}$ is not empty then by (ii) we must have $\|\mathbf{i}\| \leq \mathrm{rk}(\widetilde{\mathcal{X}})$. Since $\mathbf{i}$ has t components, the number of solutions of this inequality is exactly

$$\sum_{l=0}^{s} \binom{t}{l} \binom{\mathrm{rk}(\widetilde{\mathcal{X}}) + t - l}{t - l} < 2^{\mathrm{rk}(\widetilde{\mathcal{X}})+2t-1} \le 2^{\mathrm{rk}(\widetilde{\Gamma})+2t-1},$$

proving the stated bound because $\mathrm{rk}(\widetilde{\Gamma}) \le s+1$.

One way of applying the Global Cluster Principle is as follows. We abbreviate $G = \mathbb{G}_m^2(k)$ and define G_m as in Lemma 6, namely G_m is the subgroup of elements of G which are mth powers up to torsion. We also define

$$L(t) = 1 + \log^+(t).$$

LEMMA 10. *Let $\Gamma \subset G_m$ be a finitely generated group of rank $s \ge 1$ and let $\mathcal{X}$ be a finite set of solutions $(x_1, x_2) \in \Gamma$ of the generalized unit equation $a_1x_1 + a_2x_2 = 1$. Let $T = \{(v, j)\}$ be a finite subset of $M_k \times \{1,2\}$ containing a nondegenerate special subset for Γ and define $K = s(1+\log^+(432d^9s^3m^{-1}))$. Then there is a partition*

$$\mathcal{X} = \bigcup \mathcal{X}_i$$

of $\mathcal{X}$ into at most $2^{s+3|T|}$ subsets $\mathcal{X}_i$, with the following property. For every subset $\mathcal{X}_i \ne \emptyset$ and any choice of $\mathbf{x}_i \in \mathcal{X}_i$ we have

$$\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le KL(\nu_T(\mathcal{X})).$$

In particular, if $m \ge 432d^9s^3$ we can take $K = s$.

Proof. We first split $\mathcal{X}$ into not more than $2^{|T|}$ sets of constant signature and apply the Global Cluster Principle to each subset so obtained, say $\mathcal{Y}$.

We estimate the mass of $\mathcal{Y}$ using the second statement of Lemma 7. If $r = \mathrm{rk}(\mathcal{Y})$ we find

$$\begin{aligned}
\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) &\le \log(6s^2\tau m(\mathcal{Y})) \\
&\le \log\left(6s^2\tau\left(\frac{18d^5r}{m}\right)^r \nu_T(\mathcal{Y})^r\right) \\
&\le \log\left(6s^2\tau\left(\frac{18d^5r}{m}\right)^r \nu_T(\mathcal{X})^r\right) \\
&\le \log\left(6s^2\tau\left(\frac{18d^5r}{m}\right)^r\right) + s\log^+(\nu_T(\mathcal{X})).
\end{aligned}$$

Since $\tau \le |\mathrm{tors}(k^\times)|^2 \le (2d^2)^2$, we have

$$\begin{aligned}
\log\left(6s^2\tau\left(\frac{18d^5r}{m}\right)^r\right) &\le r\log\left(6s^2(2d^2)^2\frac{18d^5r}{m}\right) \\
&\le s\log^+(432d^9s^3m^{-1}) = K - s,
\end{aligned}$$

and we conclude that

$$\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le K - s + s\log^+(\nu_T(\mathcal{X})) \le KL(\nu_T(\mathcal{X})).$$

Finally, $K = s$ if $m \ge 432d^9s^3$.

Let us abbreviate

$$L_n(t) = \underbrace{L \circ L \circ \ldots \circ L}_{n \text{ times}}(t).$$

The iteration of Lemma 10 leads to the following result.

COROLLARY. *Assume the same hypotheses as in Lemma* 10 *and let* $n \ge 1$ *be a positive integer. Then there are a partition*

$$\mathcal{X} = \bigcup \mathcal{X}_i$$

and points $\mathbf{x}_i \in \mathcal{X}_i$ *such that*

(i) *we have* $\max_i \nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le 2KL(K)L_n(\nu_T(\mathcal{X}))$;
(ii) *the number of sets* $\mathcal{X}_i$ *does not exceed* $2^{n(s+3|T|)}$.

Proof (by induction on n). Let us denote by $\mathcal{X}^n$ a typical set $\mathbf{x}_i^{-1}\mathcal{X}_i$ obtained at stage n. We will show that there is a sequence $1 = \kappa_1 < \kappa_2 < \ldots < 2$ such that

$$\nu_T(\mathcal{X}^n) \le \kappa_n KL(K)L_n(\nu_T(\mathcal{X})).$$

The Corollary clearly follows from this statement.

If $n = 1$, this comes from the Cluster Principle. Now suppose the statement is true for n, so that

$$\nu_T(\mathcal{X}^n) \le \kappa_n KL(K)L_n(\nu_T(\mathcal{X})),$$

and apply the Cluster Principle to each set $\mathcal{X}^n$. Then we obtain

$$\nu_T(\mathcal{X}^{n+1}) \le KL(\nu_T(\mathcal{X}^n))$$

with not more than $2^{s+3|T|}$ sets $\mathcal{X}^{n+1}$ arising from each set $\mathcal{X}^n$. In particular, the total number of sets $\mathcal{X}^{n+1}$ is at most $2^{(n+1)(s+3|T|)}$.

From the last two displayed inequalities and $L(uv) \le L(u)L(v)$ we infer

$$\nu_T(\mathcal{X}^{n+1}) \le KL(\kappa_n KL(K)L_n(\nu_T(\mathcal{X}))) \le KL(\kappa_n KL(K))L_{n+1}(\nu_T(\mathcal{X})),$$

yielding

$$\nu_T(\mathcal{X}^{n+1}) \le \kappa_{n+1} KL(K)L_{n+1}(\nu_T(\mathcal{X}))$$

with

$$\kappa_{n+1} = \max_t \frac{L(\kappa_n tL(t))}{L(t)}.$$

We have $L(\kappa_n t L(t)) = \log(\kappa_n) + L(t) + \log L(t)$ for $t \geq 1$, therefore setting $u = L(t)$ we find

$$\kappa_{n+1} = 1 + \max_{u \geq 1} \frac{\log \kappa_n + \log u}{u}.$$

The maximum occurs for $u = e/\kappa_n$, whence $\kappa_{n+1} = 1 + e^{-1}\kappa_n$ and

$$\kappa_n = \frac{1 - e^{-n}}{1 - e^{-1}} < \frac{e}{e-1} < 2$$

by induction on n. This completes the proof of the Corollary.

7. Moderate growth bounds. For $t \geq 1$ we define $E(t) = \exp(t-1)$ and

$$E_n(t) = \underbrace{E \circ E \circ \ldots \circ E}_{n \text{ times}}(t).$$

The inverse function of $E(t)$ is the function $L(t) = 1 + \log^+(t)$ introduced in the preceding section, and the inverse function of $E_n(t)$ is

$$L_n(t) = \underbrace{L \circ L \circ \ldots \circ L}_{n \text{ times}}(t).$$

Let $f(z) \geq 1$ be a function of a positive argument $z \geq 1$. We say that f has *moderate growth of order* n if there is a positive integer n such that $f(z) \leq E_n(z+1)$. The property of being of moderate growth is stable by sum, product and composition.

In this section, we consider *normalized* equations $ax + (1-a)y = 1$ and obtain, as a consequence of Baker's theory of linear forms in logarithms, bounds for various quantities associated with them. All such bounds will be described by functions of moderate growth in their arguments.

We say that $ax+by = 1$ is *equivalent* to $a'x+b'y = 1$ if $(a', b') = (a\gamma_1, b\gamma_2)$ for some $\boldsymbol{\gamma} \in \Gamma$.

Definition 6. The equation $ax + (1-a)y = 1$ is said to be *reduced* if $h(a)$ is a minimum among all equivalent normalized equations.

We note the following property of a reduced equation.

Lemma 11. *Let* $\mathbf{x} = (x, y) \neq (1, 1)$ *be a solution of a normalized unit equation* $ax + (1-a)y = 1$. *Then*

$$h(a) \leq 3h(\mathbf{x}) + \log 4.$$

Proof. Since $ax + (1-a)y = 1$ we have

$$a = \frac{1-y}{x-y}.$$

Taking heights, we find

$$h(a) \le h(1-y) + h(x-y) \le h(x) + 2h(y) + 2\log 2.$$

Lemma 12. *Let Γ be a finitely generated subgroup of $\mathbb{G}_m^2(k)$ of rank s, let t be the cardinality of the support of Γ, and let R be the largest regulator associated with Γ. Also let P be the largest prime such that there is $v \in$* supp(Γ) *with $v \mid P$ if* supp(Γ) *contains at least one finite place, and $P = 1$ otherwise. Then there is a set of generators $\boldsymbol{\gamma}_\sigma$ of Γ up to torsion whose height $h(\boldsymbol{\gamma}_\sigma)$ is bounded by a function of moderate growth in d, s and R, namely*

$$h(\boldsymbol{\gamma}_\sigma) \le E_3(3\max(d,s,R)).$$

Moreover, $t \le 2dP$ *and* $\log P \le (2ds)^s R$, *and t and P are also bounded by* $E_3(3\max(d,s,R))$.

Proof. Let $T = \mathrm{supp}(\Gamma) \times \{1,2\}$ and let $\boldsymbol{\gamma}_\sigma$ be good generators relative to T. Since $\log \|\gamma_{\sigma j}\|_v = 0$ if $(v,j) \notin T$, Lemma 5 shows that

$$(2t)^{-s} \prod_{\sigma=1}^{s} \max_{\text{all } (v,j)} |\log \|\gamma_{\sigma j}\|_v| \le R. \tag{7.1}$$

We have

$$h(\boldsymbol{\gamma}_\sigma) = \tfrac{1}{2} \sum_v \frac{d_v}{d} \max_{\text{all } (v,j)} |\log \|\gamma_{\sigma j}\|_v| \le \tfrac{1}{2} t \max_{\text{all } (v,j)} |\log \|\gamma_{\sigma j}\|_v|,$$

therefore by (7.1) we obtain

$$\prod_{\sigma=1}^{s} h(\boldsymbol{\gamma}_\sigma) \le t^{2s} R. \tag{7.2}$$

A bound for the individual quantities $h(\boldsymbol{\gamma}_\sigma)$ is obtained from (7.2) and Dobrowolski's [D] lower bound $h(\boldsymbol{\gamma}_\sigma) \ge 1/(9d^3)$, giving

$$h(\boldsymbol{\gamma}_\sigma) \le (9d^3 t^2)^s R. \tag{7.3}$$

The bound $t \le dP$ is obvious, and the bound $\log P \le (18d^5)^s R$ follows from Lemma 6. In view of (7.3), this gives

$$h(\boldsymbol{\gamma}_\sigma) \le (9d^5)^s R \exp(2s(18d^5)^s R), \tag{7.4}$$

which indeed can be expressed as a function of moderate growth in the argument $3\max(d,s,R)$. We leave it to the reader to determine that

$$(9d^5)^s R \exp(2s(18d^5)^s R) \le E_3(3\max(d,s,R)).$$

This completes the proof.

Lemma 13. *There is a positive integer $n_1 \ge 3$ with the following property. Let the notation and hypotheses of Lemma 12 hold. Then the height* *of solutions of the unit equation* $ax + (1-a)y = 1$ *in* Γ *is majorized by a function of moderate growth in* d, s, R *and* $h(a)$, *namely*

$$h((ax, (1-a)y)) \le E_{n_1}(3\max(d, s, R, h(a))).$$

Remark. A calculation shows that we can take $n_1 = 6$.

Proof of Lemma 13. Bounds for solutions of the unit equations have been obtained by several authors using Baker's theory of linear forms in logarithms. For our purposes, we refer to Győry [G] and Evertse, Győry, Stewart and Tijdeman [EGST]. Lemma 7 of [EGST] states a bound for the height of solutions of the equation in question which is of moderate growth in d, P, the cardinality t of $\mathrm{supp}(\Gamma)$, and two unspecified constants c_{14} and c_{15} depending only on the degree d and the discriminant D_k of the field k (in their notation, $A \le 2\exp(h(a))$ and $s = t + u$ with $u < d$ the rank of the group of units of k). However, inspection of [G] shows that these constants are also majorized by functions of moderate growth in d and D_k (use for example Siegel's bounds [S, Satz 1] to bound the product of the class number and regulator of the field k).

Since by Lemma 12 both t and P are majorized by functions of moderate growth in d, s and R, it remains to prove that D_k can also be majorized by a function of moderate growth in d, s and R. If the equation $ax + (1-a)y = 1$ has only the trivial solution $(1, 1)$, Lemma 13 is trivial. If instead this equation admits a nontrivial solution then $a = (1-y)/(x-y)$, therefore

$$a \in \mathbb{Q}(\mathrm{tors}(k^\times), \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_s)$$

with $\boldsymbol{\gamma}_\sigma$, $\sigma = 1, \ldots, s$, a set of generators of Γ up to torsion. The group $\mathrm{tors}(k^\times)$ is cyclic of order at most $2d^2$ (and better bounds are easily provided), hence it is controlled solely in terms of d.

Replacing k by $\mathbb{Q}(\mathrm{tors}(k^\times), \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_s)$ it is easy to see that D_k admits a bound which is a function of moderate growth in the heights of the generators $\boldsymbol{\gamma}_\sigma$ of Γ and in the degree d of the field; a neat explicit bound is in [BW], Lemma 2.

Finally, Lemma 12 shows that we can find a set of generators of Γ such that $h(\boldsymbol{\gamma}_\sigma)$ is majorized by a function of moderate growth in $3\max(d, s, R)$. The required result follows.

8. Application of the Cluster Principle. In this section we prove the theorem stated in the introduction.

LEMMA 14. *Let* $ax + (1-a)y = 1$ *be a reduced equation to be solved with* $\mathbf{x} = (x, y) \in \Gamma$. *Let* $R = \max_S R_S(\Gamma)$ *be the largest regulator associated with* Γ. *Then either this equation has at most* c_2^s *solutions for a certain absolute constant* c_2 *or* $h(a)$ *is majorized by a function of moderate growth in* d, s *and* R.

Remark. The proof shows that $h(a) \le E_4(3\max(d,s,R))$.

Proof of Lemma 14. If $h(a) \le \max(d,s,R)$ there is nothing to prove. Hence let us assume that $h(a) > \max(d,s,R)$.

Let $T \subset M_k \times \{1,2\}$ be given by $T = S \cup S^*$ where S is nondegenerate and special for Γ and S^* is such that $R = R_{S^*}$. Clearly $|T| \le 2s$. Also let $K = s(1+\log(432d^9s^3))$ and let $\mathcal{X}$ be the set of solutions of the unit equation in question.

We apply the Corollary to Lemma 10 to this situation obtaining a partition $\mathcal{X} = \bigcup \mathcal{X}_i$ of $\mathcal{X}$ into not more than 2^{7ns} disjoint subsets $\mathcal{X}_i$ such that for $\mathbf{x}_i \in \mathcal{X}_i$ we have

$$\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le 2KL(K)L_n(\nu_T(\mathcal{X})). \tag{8.1}$$

Since $S^* \subset T$, Lemma 8 shows that

$$h(\mathbf{x}_i^{-1}\mathcal{X}_i) \le (2st)^{s+1}\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \tag{8.2}$$

where t is the cardinality of the support of Γ.

On the other hand, for any $\mathbf{x} \in \Gamma$ we have

$$\nu_T(\mathbf{x}) \le dh(\mathbf{x}) \le d(h(a) + h(a\mathbf{x})),$$

and therefore using Lemma 13 and $h(a) \ge \max(d,s,R)$ we get

$$\nu_T(\mathcal{X}) \le d(h(a) + E_{n_1}(3h(a))) \le E_{n_1+1}(3h(a)). \tag{8.3}$$

By (8.1), (8.2) and (8.3) taking $n = n_1 + 2$ we deduce

$$\begin{aligned} h(\mathbf{x}_i^{-1}\mathcal{X}_i) &\le (2st)^{s+1}2KL(K)L_{n_1+2}(E_{n_1+1}(3h(a))) \\ &= (2st)^{s+1}2KL(K)L(3h(a)); \end{aligned} \tag{8.4}$$

the number of sets $\mathcal{X}_i$ is at most $2^{7(n_1+2)s}$.

Now suppose that there is a set $\mathcal{X}_i$ containing at least two elements, say $\mathbf{x}_i = (x_i, y_i)$ and $\mathbf{x}'_i = (x'_i, y'_i)$. Set $a' = ax_i$, $\mathbf{x}' = (x', y') = \mathbf{x}'_i/\mathbf{x}_i$. Then (8.4) yields

$$h(\mathbf{x}') \le (2st)^{s+1}2KL(K)L(3h(a)). \tag{8.5}$$

Since $\mathbf{x}'$ is a nontrivial solution of the normalized equation $a'x' + (1-a')y' = 1$, we may apply Lemma 11 and deduce from (8.5) that

$$h(a') \le 3h(\mathbf{x}') + \log 4 \le 3(2st)^{s+1}2KL(K)L(3h(a)) + \log 4. \tag{8.6}$$

Finally, $h(a) \le h(a')$ because $ax + (1-a)y = 1$ is a reduced equation. In view of (8.6), this gives

$$h(a) \le 3(2st)^{s+1}2KL(K)L(3h(a)) + \log 4. \tag{8.7}$$

By Lemma 12, $t \le E_3(3\max(d,s,R))$ and (8.7) now implies that $h(a)$ admits a bound of moderate growth in d, s and R, which is the conclusion of Lemma 14.

It remains to consider the case in which every $\mathcal{X}_i$ consists of only one element. Since the number of sets $\mathcal{X}_i$ is at most $2^{7(n_1+2)s}$, Lemma 14 follows.

P r o o f o f T h e o r e m. For the rest of this section, we assume that $h(a)$ satisfies the moderate growth bound provided by the second alternative of Lemma 14. Then Lemma 13 provides a bound

$$h(\mathcal{X}) \le E_{n_2}(3\max(d,s,R)) \tag{8.8}$$

for the set of solutions of the reduced and normalized equation $ax+(1-a)y = 1$ with $\mathbf{x} \in \Gamma$, for a certain absolute integer constant n_2, in fact with $n_2 \le 10$.

We apply again the Corollary to Lemma 10 to this situation taking $T = S \cup S^*$ as in the proof of Lemma 14 and $n = n_2 + 1$, and obtain a partition $\mathcal{X} = \bigcup \mathcal{X}_i$ of $\mathcal{X}$ into not more than $2^{7(n_2+1)s}$ disjoint subsets $\mathcal{X}_i$ such that for $\mathbf{x}_i \in \mathcal{X}_i$ we have

$$\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le 2KL(K)L_{n_2+1}(\nu_T(\mathcal{X})). \tag{8.9}$$

As noted in the proof of Lemma 14 we have $\nu_T(\mathbf{x}) \le dh(\mathbf{x})$, and therefore from (8.8) and (8.9) we easily infer that

$$\nu_T(\mathbf{x}_i^{-1}\mathcal{X}_i) \le 2KL(K)L(4\max(d,s,R)). \tag{8.10}$$

We claim that if $\Gamma \subset G_m$ and $m \ge 1800d^9s^4$ we have

$$\mathrm{rk}(\mathbf{x}_i^{-1}\mathcal{X}_i) < \mathrm{rk}(\mathcal{X}). \tag{8.11}$$

To see this, let us consider a subset $\mathbf{x}_i^{-1}\mathcal{X}_i$. Suppose that $\mathrm{rk}(\mathbf{x}_i^{-1}\mathcal{X}_i) = \mathrm{rk}(\mathcal{X}) = s$. Then $\mathrm{H} = \langle \mathbf{x}_i^{-1}\mathcal{X}_i \rangle$ would have finite index in Γ and therefore

$$R_{S^*}(\mathrm{H}) \ge R_{S^*}(\Gamma) = R. \tag{8.12}$$

Now we apply Lemma 7 to the set $\mathbf{x}_i^{-1}\mathcal{X}_i$. In view of (8.10) and (8.12) we infer

$$m(\mathbf{x}_i^{-1}\mathcal{X}_i)^{1/s} \le R^{-1/s}s2KL(K)L(4\max(d,s,R)). \tag{8.13}$$

If we take $m \ge 1800d^9s^4$ then, as noted in Lemma 10, we have $K = s$ and (8.13) simplifies, after some generous majorizations, to

$$\begin{aligned} m(\mathbf{x}_i^{-1}\mathcal{X}_i)^{1/s} &\\ &\le R^{-1/s}2s^2L(s)L(4\max(d,s,R)) \\ &= R^{-1/s}2s^2(1+\log s)(1+\log 4+\max(\log d,\log s,\log R)) \\ &\le 5ds^2(1+\log s)^2R^{-1/s}\max(1,\log R). \end{aligned} \tag{8.14}$$

Now recall that by the last part of Lemma 6 and $m \ge 1800d^9s^4$ we have

$$R^{1/s} \ge \frac{m}{18d^5} \ge 100(ds)^4,$$

while $R^{-1/s}\log R$ is decreasing in R for $R^{1/s}\ge e$. Then (8.14) gives *a fortiori* the bound

$$
\begin{aligned}
(8.15)\quad m(\mathbf{x}_i^{-1}\mathcal{X}_i)^{1/s} &\le 5ds^2(1+\log s)^2R^{-1/s}\log R\\
&\le 5ds^2(1+\log s)^2(100(ds)^4)^{-1}4s\log(\sqrt[4]{100}ds)\\
&\le 5^{-1}d^{-3}s^{-1}(1+\log s)^2\log(4ds)\\
&\le 5^{-1}\max_{z\ge 1} z^{-1}(1+\log z)^2\log(4z)<1
\end{aligned}
$$

as a simple numerical maximization shows.

On the other hand, the mass of a set of positive rank is always a positive integer, which contradicts (8.15). This proves our claim (8.11).

Let $N_m(s)$ be the maximum number of solutions of a generalized unit equation for a group $\Gamma\subset G_m$. We have shown that if $m\ge 1800d^9s^4$ we have

$$N_m(s)\le 2^{7(n_2+1)s}N_m(s-1)$$

and $N_m(s)\le 2^{7(n_2+1)s^2}N_m(0)$ follows by induction on s. Also, it is trivial that $N_m(0)\le 2$ (intersect the circle $|z|=1$ with the circle $|1-az|=|1-a|$), therefore for $m\ge 1800d^9s^4$ we have

$$(8.16)\qquad N_m(s)\le 2^{7(n_2+1)s^2+1}.$$

For a general Γ we may replace Γ by $\Gamma_m=\Gamma\cap G_m$ and note that Γ_m has index m^s in Γ. By expressing Γ as a union of m^s cosets of Γ_m we obtain m^s generalized unit equations to be solved in Γ_m. By (8.16), we conclude that

$$N_1(s)\le m^sN_m(s)\le(1800d^9s^4)^s2^{7(n_2+1)s^2+1}\le d^{9s}c_1^{s^2}$$

and with it the proof of the Theorem.

References

[BW] A. Baker and G. Wüstholz, *Logarithmic forms and group varieties*, J. Reine Angew. Math. 442 (1993), 19–62.

[BS] F. Beukers and H. P. Schlickewei, *The equation $x+y=1$ in finitely generated groups*, Acta Arith. 78 (1996), 189–199.

[C1] J. W. S. Cassels, *On a class of exponential equations*, Ark. Mat. 3 (1960), 231–233.

[C2] —, *An Introduction to the Geometry of Numbers*, Springer-Verlag, Berlin–Göttingen–Heidelberg, 1959.

[D] E. Dobrowolski, *On a question of Lehmer and the number of irreducible factors of a polynomial*, Acta Arith. 34 (1979), 391–401.

[EGST] J.-H. Evertse, K. Győry, C. L. Stewart and R. Tijdeman, *On S-unit equations in two unknowns*, Invent. Math. 92 (1988), 461–477.

[G] K. Győry, *On the number of solutions of linear equations in units of an algebraic number field*, Comment. Math. Helv. 54 (1979), 583–600.

[MKS] W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory*, 2nd ed., revised, Dover, New York, 1976.

[P] M. Poe, *On distribution of solutions of S-unit equations*, J. Number Theory 62 (1997), 221–241.

[R] P. Ribenboim, *Catalan's Conjecture*, Academic Press, Boston, 1994.

[S] C. L. Siegel, *Abschätzung von Einheiten*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. 9 (1969), 71–86.

[V] J. D. Vaaler, *A geometric inequality with applications to linear forms*, Pacific J. Math. 83 (1979), 543–553.

Institute for Advanced Study
Olden Lane
Princeton, New Jersey 08540
U.S.A.

Department of Mathematics
Fordham University
Bronx, New York 10458
U.S.A.

Department of Mathematics
Harvard University
Cambridge, Massachusetts 02138
U.S.A.

Received on 22.2.1997 (3136)