

A height gap theorem for finite subsets of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ and nonamenable subgroups

By EMMANUEL BREUILLARD

Abstract

We introduce a conjugation invariant normalized height $\widehat{h}(F)$ on finite subsets of matrices F in $\mathrm{GL}_d(\overline{\mathbb{Q}})$ and describe its properties. In particular, we prove an analogue of the Lehmer problem for this height by showing that $\widehat{h}(F) > \varepsilon$ whenever F generates a nonvirtually solvable subgroup of $\mathrm{GL}_d(\overline{\mathbb{Q}})$, where $\varepsilon = \varepsilon(d) > 0$ is an absolute constant. This can be seen as a global adelic analog of the classical Margulis Lemma from hyperbolic geometry. As an application we prove a uniform version of the classical Burnside-Schur theorem on torsion linear groups. In a companion paper we will apply these results to prove a strong uniform version of the Tits alternative.

Contents

1. Introduction	1057
2. Minimal height and displacement	1064
3. Statement of the results	1074
4. Preliminary reductions	1077
5. Local estimates on Chevalley groups	1087
6. Global bounds on arithmetic heights	1093
7. Proof of the statements of Section 3	1099
References	1107

1. Introduction

According to the Lehmer conjecture, the absolute Weil height times the degree of an algebraic number $x \in \overline{\mathbb{Q}}$ which is not a root of unity ought to be bounded below by an absolute constant. Various generalizations and extensions of this problem have been recently studied by a variety of authors, in particular in the setting of abelian varieties (e.g. [36], [44]) and also in connection with the dynamics of iterated polynomial maps (e.g. [22], [18], [4], [30]). In the present paper, we will introduce yet another height function $\widehat{h}(F)$ which is well suited to the study of the geometric and arithmetic behavior of power sets $F^n = F \cdots F$ for $n \in \mathbb{N}$, where F is a finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$.

We will investigate its properties, in particular describe when it might become small and then prove a statement analogous to the Lehmer conjecture in this setting. In fact, we will prove that if \mathbb{G} is the Zariski closure of the subgroup generated by F , then $\widehat{h}(F)$ is always bounded away from zero by a positive constant $\varepsilon = \varepsilon(d) > 0$ unless the connected component of the identity \mathbb{G}^0 is solvable. While if \mathbb{G}^0 is solvable, proving a lower bound on $\widehat{h}(F)$ boils down to the original Lehmer conjecture. Before we explain our motivations for studying this object, and present the main results of the paper, let us first define it.

Definitions. Let $d \geq 1$ be an integer, $\overline{\mathbb{Q}}$ be the field of algebraic numbers, and $K \leq \overline{\mathbb{Q}}$ a number field. We let V_K be the set of equivalence classes of absolute values on K and $n_v = [K_v : \mathbb{Q}_p]$ the degree of the completion K_v of K over the closure \mathbb{Q}_p of \mathbb{Q} in K_v . We normalize the absolute value $|\cdot|_v$ on K_v so that its restriction to \mathbb{Q}_p is the standard absolute value, i.e., $|p|_v = \frac{1}{p}$. To any finite subset F of square matrices in $M_d(K)$ we associate the following *height*

$$(1) \quad h(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ \|F\|_v,$$

where $\log^+ = \max\{0, \log\}$ and $\|F\|_v = \max\{\|f\|_v, f \in F\}$. Here $\|f\|_v$ is the operator norm on $M_d(K_v)$ associated to the *standard norm* on K_v^d . We define the standard norm for $x \in K_v^d$ to be the sup norm $\|x\|_v = \max_{1 \leq i \leq d} |x_i|_v$ if v is ultrametric and the Euclidean norm $\|x\|_v = \sqrt{\sum_{i=1}^d |x_i|_v^2}$ otherwise. If $d = 1$, then this notion coincides with the (absolute, logarithmic) Weil height of an algebraic number (see e.g. [7]).

We can now define the *normalized height* $\widehat{h}(F)$ as

$$\widehat{h}(F) = \lim_{n \rightarrow +\infty} \frac{1}{n} h(F^n).$$

This limit exists by subadditivity. Unlike $h(F)$, $\widehat{h}(F)$ is independent of the choice of basis of K_v^d used to define the norms $\|x\|_v$.

Another way to describe $\widehat{h}(F)$ is in terms of spectral radius (see §2.2 below); for instance if $F = \{A\}$ is a singleton, then $\widehat{h}(F) = h([1, \lambda_1, \dots, \lambda_d])$, where $(\lambda_1, \dots, \lambda_d)$ are the eigenvalues of A and $h([1, \lambda_1, \dots, \lambda_d])$ the standard Weil height of the point $[1, \lambda_1, \dots, \lambda_d]$ in the projective space $\mathbb{P}^d(\overline{\mathbb{Q}})$ as defined in [7, §1.5.]. This connection was first described by V. Talamanca in [37], where a closely related definition of the height and normalized height of a single matrix is given (see Remark 2.20 below).

The normalized height is an invariant of the diagonal action by conjugation of GL_d on GL_d^k , where $k = \mathrm{Card}(F)$, and it is a measure of the combined *spectral radius* of F (i.e., the rate of exponential growth of $\|F^n\|_v$) at all places v , where v varies among all possible equivalence classes of nontrivial absolute values on the number field generated by the matrix coefficients of F .

Basic properties and height gap. Here are a few sample properties which are satisfied by the normalized height. We have: $\widehat{h}(F^n) = n \cdot \widehat{h}(F)$ for $n \in \mathbb{N}$. A finite set F satisfies $\widehat{h}(F) = 0$ if and only if F generates a quasi-unipotent subgroup, i.e., a group all of whose elements have only roots of unity as eigenvalues (Proposition 3.2). Moreover, the following holds:

PROPOSITION 1.1. *There is a constant $C = C(d) > 0$ such that if F is a finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ generating a subgroup whose Zariski closure is semisimple, then*

$$\widehat{h}(F) \leq \inf_{g \in \mathrm{GL}_d(\overline{\mathbb{Q}})} h(gFg^{-1}) \leq C \cdot \widehat{h}(F).$$

In other words, F can always be conjugated back in a good position where its height is comparable to its normalized height. Also \widehat{h} has the Northcott property (cf. [7]) in the sense that a subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ whose cardinality and normalized height are bounded, which generates a subgroup with semisimple Zariski closure, and which has all its matrix coefficients of bounded degree over \mathbb{Q} , must belong to a bounded finite family of conjugacy classes of such sets.

The main result of this paper establishes the existence of a uniform gap for the normalized height of subsets F generating a nonamenable subgroup of $\mathrm{GL}_d(\overline{\mathbb{Q}})$. We have:

THEOREM 1.2. *There is a constant $\varepsilon = \varepsilon(d) > 0$ such that if F is a finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ generating a nonamenable subgroup that acts strongly irreducibly, then $\widehat{h}(F) > \varepsilon$.*

The constant $\varepsilon(d)$ can be made explicit in principle, although we make no attempt here to give a lower bound (see Remark 2.5).

Recall that, as follows for instance from the Tits alternative ([39]), amenable subgroups of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ are precisely the virtually solvable subgroups, i.e., those subgroups which contain a solvable subgroup of finite index.

Note that if $d = 1$, then \widehat{h} coincides with the classical Weil height of a nonzero algebraic number. Of course $\mathrm{GL}_1(\overline{\mathbb{Q}})$ is solvable (it is a torus), and no uniform lower bound on the height can exist there. However, the Lehmer conjecture states that one ought to have $h(x) \geq \frac{c}{\deg(x)}$ for some absolute constant $c > 0$ whenever x is not a root of unity. We refer the reader to [34] for a recent survey on this conjecture (see also [7]) and to [2], [1] and for recent progress. Theorem 1.2 can thus be seen as a positive solution to a Lehmer type problem in semisimple algebraic groups as opposed to tori.

As it turns out, for each integer $k \geq 2$, the set of k -tuples F in $\mathrm{GL}_d(\overline{\mathbb{Q}})$ which generate a virtually solvable subgroup forms a closed algebraic subvariety of $\mathrm{GL}_d(\overline{\mathbb{Q}})^k$. Therefore Theorem 1.2 implies that the set of points with small normalized height in $\mathrm{GL}_d(\overline{\mathbb{Q}})^k$ is not Zariski-dense. This is reminiscent of the

Bogomolov conjecture proved by Ullmo and Zhang (see [40], [44], [36]), which asserts that, given an abelian variety, the set of points with small Néron-Tate height on an algebraic subvariety which is not a finite union of torsion cosets of abelian subvarieties is not Zariski-dense. In fact the toric version of the Bogomolov conjecture, proved by Zhang in [43], will be a key ingredient of the proof of Theorem 1.2.

Remark 1.3. A competing definition of the normalized height $\widehat{h}(F)$ consists in replacing \log^+ by \log in (1). The two definitions coincide if $F \subset \mathrm{SL}_d$, but may differ otherwise. However the difference is minor and we found it more convenient to work with \log^+ , because all terms are then nonnegative, although many results, such as Theorem 1.2, also hold for this other definition of the height (see the discussion in Remark 3.8).

Motivation and consequences. In [10] we established a connection between the Lehmer conjecture and the uniform exponential growth problem for linear solvable groups. More precisely, we showed that proving uniform exponential growth over all solvable subgroups of $\mathrm{GL}_2(\mathbb{C})$, that is showing the existence of an absolute constant $c > 0$ such that $\lim_{n \rightarrow +\infty} |F^n|^{\frac{1}{n}} > c$ whenever F generates a solvable nonvirtually nilpotent subgroup of $\mathrm{GL}_2(\mathbb{C})$ would imply the Lehmer conjecture.

We have not settled the issue of whether or whether not the Lehmer conjecture is in fact equivalent to the uniform exponential growth of solvable subgroups of $\mathrm{GL}_2(\mathbb{C})$. However in our companion paper [11], we make use of Theorem 1.2 (height gap theorem) and Proposition 1.1 above to establish the following strengthening of the classical Tits alternative, which among other things implies the existence of a constant $c = c(d) > 0$ such that $\lim_{n \rightarrow +\infty} |F^n|^{\frac{1}{n}} > c$ whenever F generates a nonvirtually solvable subgroup of $\mathrm{GL}_d(\mathbb{C})$.

THEOREM 1.4 (Uniform Tits alternative, [11]). *There is $N = N(d) \in \mathbb{N}$, such that if K is a field and F a finite symmetric subset of $\mathrm{GL}_d(K)$ containing 1 which generates a nonvirtually solvable subgroup, then F^N contains two elements a, b which generate a non-abelian free subgroup.*

In the same vein, but in a more straightforward way, one obtains the following corollary, which answers a question from [5] and is a strengthening of a well-known theorem of Burnside and Schur (see [19]) asserting that finitely generated linear torsion groups are finite.

COROLLARY 1.5 (Effective Schur). *There is an integer $N = N(d) \in \mathbb{N}$ such that if K is a field and F is a finite subset of $\mathrm{GL}_d(K)$ which generates an infinite subgroup, then $(F \cup F^{-1})^N$ contains an element of infinite order.*

The positive characteristic case of the above corollary is easy to prove, while the characteristic zero case relies on our Theorem 1.2.

The interpretation of $\hat{h}(F)$ in terms of spectral radius allows us to derive the following:

COROLLARY 1.6. *There are constants $N_1 = N_1(d) \in \mathbb{N}$, $C = C(d) \in \mathbb{N}$ such that if F is any finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ containing 1, there is some $a \in F^{N_1}$ and some eigenvalue λ of a such that $h(\lambda) \geq \frac{1}{|F|^C} \cdot \hat{h}(F)$.*

COROLLARY 1.7. *There are constants $N_1 = N_1(d) \in \mathbb{N}$, $\varepsilon = \varepsilon(d) > 0$ such that if F is any finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ containing 1 and generating a nonvirtually solvable subgroup, then we may find $a \in F^{N_1}$ and an eigenvalue λ of a such that $h(\lambda) > \varepsilon$, for some fixed $\varepsilon = \varepsilon(d) > 0$.*

This follows easily from Corollary 1.6, Theorem 1.2 and the following fact that we prove along the way to the proof of Theorem 1.2 (see Proposition 4.1):

PROPOSITION 1.8. *Let \mathbb{G} be a connected semisimple algebraic group over an algebraically closed field of characteristic 0. There is a constant $c = c(d) \in \mathbb{N}$, where $d = \dim \mathbb{G}$, such that the following holds. Let F be a finite subset of \mathbb{G} containing 1 and generating a Zariski-dense subgroup. Then $F^{c(d)}$ contains two elements a and b which generate a Zariski dense subgroup of \mathbb{G} .*

N.B. This proposition also holds in positive characteristic, but the proof, given in our companion paper [11], is more involved. See Remark 3.7 for more on positive characteristic.

Corollaries 1.6 and 1.7 allow us to construct a short (positive) word w with letters in F which has an eigenvalue of large height. The length of the word is bounded by an absolute constant $N_1 = N_1(d)$. This type of result is crucial in order to build the so-called proximal elements which are needed in various situations, in particular in the applications to the Tits alternative given in [11].

In the same vein we have:

COROLLARY 1.9. *There is a constant $N_2 = N_2(d) \in \mathbb{N}$, such that if F is a finite subset of $\mathrm{GL}_d(\mathbb{C})$ containing 1 which generates a nonvirtually solvable subgroup, then there is a matrix $w \in F^{N_2}$ with an eigenvalue λ such that: either there exists an ultrametric absolute value $|\cdot|_v$ on $\mathbb{Q}(\lambda)$ such that $|\lambda|_v > 1$, or there is a field homomorphism $\sigma : \mathbb{Q}(\lambda) \hookrightarrow \mathbb{C}$ such that $|\sigma(\lambda)| \geq 2$.*

In particular, if \mathcal{O} is the ring of all algebraic integers, then there is an integer $N_1 = N_1(d) \in \mathbb{N}$ such that if F is a finite set of $\mathrm{SL}_d(\mathcal{O})$ containing 1, either F generates a virtually solvable subgroup, or there is an Archimedean absolute value v on $\overline{\mathbb{Q}}$ extending the canonical absolute value on \mathbb{Q} and a matrix $f \in F^{N_1}$ with at least one eigenvalue of v -absolute value ≥ 2 . Observe that this

fails for arbitrary finite subsets of $\mathrm{SL}_d(\overline{\mathbb{Q}})$. For instance, $\mathrm{SL}_3(\mathbb{Q}) \cap \mathrm{SO}(3, \mathbb{R})$ is dense in $\mathrm{SO}(3, \mathbb{R})$ and contains a finitely generated dense subgroup.

Geometric interpretation and the Margulis Lemma. Theorem 1.2 also has the following geometric interpretation. Recall that the classical Margulis Lemma (see [38]) asserts that if $S = \mathbb{H}^n$ is the hyperbolic n -space, or more generally any real symmetric space of noncompact type endowed with its Riemannian metric d , then there is a positive constant $\varepsilon = \varepsilon(S) > 0$ such that the following holds: suppose F is a finite set of isometries of S such that $\max_{f \in F} d(f \cdot x, x) < \varepsilon$ for some point $x \in S$ and suppose F lies in a discrete subgroup of isometries of S ; then F generates a virtually nilpotent subgroup. This lemma has several important consequences for the geometry and topology of hyperbolic manifolds and locally symmetric spaces, such as the structure of cusps and the thick-thin decomposition ([38]), or lower bounds for the covolume of lattices in semisimple Lie groups (see [41], [26], [23]).

What happens if one removes the discreteness assumption on the group generated by F and assumes instead that F consists of elements which are rational over some number field K ? Of course the Margulis Lemma no longer holds as such, in particular because $\varepsilon(S)$ tends to 0 as $\dim S$ tends to infinity. However Theorem 1.2 gives a kind of substitute. As will be shown below (see §2.2) the normalized height $\hat{h}(F)$ is always bounded above by the quantity $e(F)$, which we call minimal height, and which encodes, as a weighted sum over all places $v \in V_K$, the minimal displacement of F on each symmetric space or Bruhat-Tits building X_v associated to $\mathrm{SL}_d(K_v)$. In particular the height gap $\hat{h}(F) > \varepsilon$ obtained in Theorem 1.2 implies that there always is a natural space X_v (symmetric space or Bruhat-Tits building of SL_d) where F acts with a large displacement. More precisely:

COROLLARY 1.10. *Let $d \in \mathbb{N}$ and for a local field k let us denote by X_k the symmetric space or Bruhat-Tits building of $\mathrm{PGL}_d(k)$. We let $d(\cdot, \cdot)$ be a left invariant Riemannian metric on $X_{\mathbb{C}}$. There is a constant $\varepsilon = \varepsilon(d) > 0$ with the following property. Let K be a number field and F a finite subset of $\mathrm{SL}_d(K)$ which generates a nonvirtually solvable subgroup Γ , then either for some finite place v of K , the subgroup Γ acts (simplicially) without global fixed point on the Bruhat-Tits building X_{K_v} , or for some embedding $\sigma : K \hookrightarrow \mathbb{C}$*

$$\inf_{x \in X_{\mathbb{C}}} \max_{f \in F} d(\sigma(f) \cdot x, x) > \varepsilon.$$

The crucial point here of course is that ε is independent of the number field K . Thus Theorem 1.2 can be seen as a uniform Margulis Lemma for all S -arithmetic lattices of a given Lie type. For example, it is uniform over all $\mathrm{SL}_2(\mathcal{O}_K)$ where K can vary among all number fields, even though those groups can be lattices of arbitrarily large rank.

Outline of the proof of Theorem 1.2. The first part of the proof consists in reducing to the situation when F is a 2-element set $F = \{A, B\}$, where A and B are two regular semisimple elements in an absolutely almost simple algebraic group \mathbb{G} of adjoint type and F generates a Zariski-dense subgroup of \mathbb{G} . It is not hard to see that the existence of a gap for $\widehat{h}(F)$ when computed in the adjoint representation of \mathbb{G} implies the existence of a gap for $\widehat{h}(F)$ when computed in any finite dimensional linear representation of \mathbb{G} . We thus reduce to the adjoint representation of \mathbb{G} . The reduction from an arbitrary finite set F to a 2-element set makes use of a lemma due to Eskin-Mozes-Oh [21] (“escaping subvarieties” Lemma 4.2), which, given any nontrivial algebraic relation between pairs $\{x, y\}$ of elements in \mathbb{G} , produces two short words in $\{x, y\}$ which no longer satisfy this relation. This lemma is also used later on and is an essential tool here.

As we mentioned above, one may interpret $\widehat{h}(F)$ in terms of the combined minimal displacement $e(F)$ of F on all symmetric spaces and Bruhat-Tits buildings that arise through the various completions of the number field. The quantity $e(F)$ is defined as the weighted sum of the logarithm of the minimal norms $E_v(F) = \inf\{\|gFg^{-1}\|_v, g \in \mathrm{GL}_d(\overline{K_v})\}$. Crucial to this correspondence is a spectral radius formula for sets of matrices (Lemma 2.1 below), which compares the minimal displacement of F (or equivalently $E_v(F)$) with the minimal displacement of each individual matrix in the power set F^{d^2} (or equivalently its maximal eigenvalue). As a consequence, $\widehat{h}(F)$ is small if and only if $e(F)$ is small.

In the second part of the proof, we fix a place v and work in $\mathbb{G}(K_v)$. Given A, B in $\mathbb{G}(K_v)$, with A in a maximal torus T of $\mathbb{G}(K_v)$, we obtain local estimates for the minimal displacement of the action of B restricted to the maximal flat associated to T . These estimates are obtained via the Iwasawa decomposition working our way through all positive roots of A starting from the maximal one. At the end we get an upper bound for $\inf_{t_v \in T} \|t_v B t_v^{-1}\|_v$, which involves $E_v(F)$ on the one hand and the gap $|1 - \alpha(A)|_v$ between the roots of $\alpha(A)$ and 1 on the other hand.

In the last part of the proof, we put all our local estimates together and make crucial use of the product formula, so as to obtain an upper bound for the weighted sum of all $\inf_{t_v \in T} \log \|t_v B t_v^{-1}\|_v$ in terms of $e(F)$ and the average of the $\log |1 - \alpha(A)|_v$ over all Archimedean places v , for each root α . When $e(F)$ is small this upper bound becomes also small. Indeed, since the height of each $\alpha(A)$ is small, we can invoke Bilu’s equidistribution theorem: the Galois conjugates of $\alpha(A)$ equidistribute on the unit circle ([6]). Hence the average of the $\log |1 - \alpha(A)|_v$ ’s gives a negligible contribution.

Finally, considering a suitably chosen regular map f on \mathbb{G} which is invariant under conjugation by the elements of T (a suitable matrix coefficient of B

will do), we use the above upper bound to show that the height of $f(B)$ as well as $f(B^i)$ for larger and larger $i \in \mathbb{N}$, becomes small when $e(F)$ is small. However, by a theorem of Zhang [43] on small points of algebraic tori, this must force a nontrivial algebraic relation between the $f(B^i)$'s. Finally the Eskin-Mozes-Oh lemma quoted above provides the desired contradiction, as we may have chosen $F = \{A, B\}$ to avoid this relation to begin with.

The reader can also consult [12], where we gave the full details of the proof in the special case of GL_2 .

Outline of the paper. Section 2 is devoted to the definition of the normalized and minimal heights and the derivation of their most basic properties. The main results of this section are the spectral radius formula for several matrices (Lemma 2.1 below) and Proposition 2.9, which gives a lower bound on the displacement of the power set F^n . These facts will enable us to compare the normalized height with the minimal height and to reinterpret the normalized height in terms of adelic displacement.

In Section 3, we state our main results in full detail. Their proof occupies the remainder of the paper. Section 4 gives the main reduction step from an arbitrary subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ to a subset consisting of two elements $F = \{A, B\}$ which generates a Zariski-dense subgroup of a simple algebraic group \mathbb{G} . We also prove there the comparison statement between different linear representations (Proposition 3.3). The geometric interpretation in terms of displacement is also made precise at the end of Section 4.

In Section 5, we pick a Chevalley basis for the adjoint representation of \mathbb{G} and we prove local estimates whose aim is to obtain good upper bounds for the size of the matrix coefficients of a conjugate of $F = \{A, B\}$ which almost realizes the infimum $E_v(F) = \inf_{g \in \mathrm{GL}_d(\overline{K_v})} \|gFg^{-1}\|_v$ in terms of $E_v(F)$ and the simple roots $\alpha(A)$. These local bounds are then used and put together in Section 6 in order to get a global bound on the height of matrix coefficients of A and B (Proposition 6.1).

Section 7 is devoted to completing the proofs of the results stated in Section 3. In particular, we make use of the global bound proved in Section 6 to prove Theorem 1.2 (height gap) and the local estimates of Section 5 are used again to give a proof of Proposition 1.1 (good position). Finally we also derive the corollaries stated in this introduction.

2. Minimal height and displacement

2.1. Local notions of minimal norm, spectral radius and minimal displacement. Let k be a local field of characteristic 0. Let $\|\cdot\|_k$ be the standard norm on k^d , that is the canonical Euclidean (resp. Hermitian) norm if $k = \mathbb{R}$ (resp. \mathbb{C}) and the sup norm ($\|x\|_k = \max_i |x_i|_k$) if k is non-Archimedean. We will also denote by $\|\cdot\|_k$ the operator norm induced on the space of d by d matrices $M_d(k)$

by the standard norm $\|\cdot\|_k$ on k^d . Let Q be a bounded subset of matrices in $M_d(k)$. We set

$$\|Q\|_k = \sup_{g \in Q} \|g\|_k$$

and call it the *norm of Q* . Let \bar{k} be an algebraic closure of k . It is well known (see Lang's Algebra [28, XII.2, Prop. 2.5.]) that the absolute value on k extends to a unique absolute value on \bar{k} ; hence the norm $\|\cdot\|_k$ also extends in a natural way to \bar{k}^d and to $M_d(\bar{k})$. This allows us to define the *minimal norm* of a bounded subset Q of $M_d(k)$ as

$$E_k(Q) = \inf_{x \in \mathrm{GL}_d(\bar{k})} \|xQx^{-1}\|_k.$$

We will also need to consider the *maximal eigenvalue of Q* , namely

$$\Lambda_k(Q) = \max\{|\lambda|_k, \lambda \in \mathrm{spec}(q), q \in Q\},$$

where $\mathrm{spec}(q)$ denotes the set of eigenvalues (the spectrum) of q in \bar{k} . We also set $Q^n = Q \times \cdots \times Q$ to be the set of all products of n elements from Q . Finally, we introduce the *spectral radius* of Q ; that is,

$$R_k(Q) = \lim_{n \rightarrow +\infty} \|Q^n\|_k^{\frac{1}{n}},$$

in which the limit exists (and coincides with $\inf_{n \in \mathbb{N}} \|Q^n\|_k^{\frac{1}{n}}$) because the sequence $\{\|Q^n\|_k\}_n$ is sub-multiplicative.

These quantities are related to one another. The key property concerning them is given in the following result, which, together with its corollary below (Proposition 2.7), we call “*spectral radius formula for several matrices*” because of its parallel with the classical spectral radius formula relating the asymptotics of the powers of a matrix with its maximal eigenvalue:

LEMMA 2.1 (Spectral radius formula for Q). *Let Q be a bounded subset of $M_d(k)$.*

- (a) *If k is non-Archimedean, then there is an integer $q \in [1, d^2]$ such that $\Lambda_k(Q^q) = E_k(Q)^q$.*
- (b) *If k is Archimedean, then there is a constant $c = c(d) \in (0, 1)$ independent of Q and an integer $q \in [1, d^2]$ such that $\Lambda_k(Q^q) \geq c^q \cdot E_k(Q)^q$.*

N.B. In the work of Eskin-Mozes-Oh [21] a result of a similar nature appears between the lines inside their argument (when they consider almost algebras). A weaker version of this lemma (essentially part (b)) was already used in [14]. The equality in part (a) is new and will be crucial in our arguments.

Proof. Let K be a field. We make use of two well-known theorems. The first is a theorem of Wedderburn (see Curtis-Reiner [19, 27.27]) that if an algebra A over K has a linear basis over K consisting of nilpotent elements, then $A^m = 0$ for some integer m . The second is a theorem of Engel (see

Jacobson [25]) that if A is a subset of $M_d(K)$ such that $A^m = 0$ for some integer m , then A can be simultaneously conjugated in $\mathrm{GL}_d(K)$ inside $N_d(K)$, the subalgebra of upper triangular matrices with zeroes on and below the diagonal. Combined together, these facts yield:

LEMMA 2.2. *Let K be a field. If Q is any subset of $M_d(K)$ such that Q^q contains only nilpotent matrices for every q , $1 \leq q \leq d^2$, then there is $g \in \mathrm{GL}_d(K)$ such that $gQg^{-1} \subset N_d(K)$.*

Proof. Since $\dim_K M_d(K) \leq d^2$, the K -algebra generated by Q has a linear basis made of elements in $\cup_{1 \leq q \leq d^2} Q^q$. By Wedderburn and Engel, the result follows. \square

We first quickly prove (b). We argue by contradiction. There is a sequence Q_n with $E_k(Q_n) = 1$ while $\max_{1 \leq q \leq d^2} \Lambda_k(Q_n^q)^{\frac{1}{q}}$ tends to 0. Up to conjugating by some $g_n \in \mathrm{GL}_d(\mathbb{C})$, we may assume that $\|Q_n\|_{\mathbb{C}} \leq 1 + \frac{1}{n}$, and passing to a Hausdorff limit, we obtain a compact set Q with $E_{\mathbb{C}}(Q) = \|Q\|_{\mathbb{C}} = 1$, while $\max_{1 \leq q \leq d^2} \Lambda_k(Q^q)^{\frac{1}{q}} = 0$. But this is a contradiction with Lemma 2.2 as $E_{\mathbb{C}}(C) = 0$ for any bounded subset C of $N_d(\mathbb{C})$. This proves (b).

In order to prove (a) we first show:

LEMMA 2.3 (Small eigenvalues implies large fixed point set). *Let $d \in \mathbb{N}$. There exists an integer $N = N(d) \in \mathbb{N}$ with the following property. Let k be a non-Archimedean local field with absolute value $|\cdot|_k$ and \mathcal{O}_k its ring of integers. Let Q be a subset of $M_d(\mathcal{O}_k)$ such that for each integer $q \in [1, d^2]$ every element of Q^q has all its eigenvalues of absolute value at most $|\pi|_k^N$, where π is a uniformizer for \mathcal{O}_k . Then there is $g \in \mathrm{GL}_d(k)$ such that gQg^{-1} belongs to $\pi M_d(\mathcal{O}_k)$.*

Proof. We argue by contradiction. This means that we have a sequence of local fields k_n and subsets Q_n in $M_d(\mathcal{O}_{k_n})$ such that $\|gQ_ng^{-1}\|_{k_n} \geq 1$ for all $g \in \mathrm{GL}_d(k_n)$ and all eigenvalues of Q_n^q have absolute value at most $|\pi_n|_{k_n}^n$. Let us consider a nonprincipal ultrafilter \mathcal{U} on \mathbb{N} and form the ultraproduct ring $A = \prod_{\mathcal{U}} \mathcal{O}_{k_n}$. First let us decide that we have chosen the absolute value $|\cdot|_n$ on k_n in such a way that $|\pi_n|_n = \frac{1}{2}$ for every n , where π_n is a fixed uniformizer in \mathcal{O}_{k_n} . For every $x_n \in \mathcal{O}_{k_n}$ the quantity $|x_n|_n$ may only take values among $2^{-(\mathbb{N} \cup \{\infty\})}$. It follows that for every $x \in A$ represented by $(x_n)_{n \in \mathbb{N}}$, the quantity $|x| := \lim_{\mathcal{U}} |x_n|_n$, which is well defined, may only take values in $2^{-(\mathbb{N} \cup \{\infty\})}$. Moreover, the defining properties of the absolute values $|\cdot|_n$ are inherited by $|\cdot|$; that is, $|xy| = |x| \cdot |y|$ and $|x + y| \leq \max\{|x|, |y|\}$, except that there may be nonzero elements $x \in A$ with $|x| = 0$. We will quotient these elements out. Let $I = \{x \in A, |x| = 0\}$. Then I is clearly a prime ideal of A . We can now set $\mathcal{O} = A/I$, which is a domain on which our absolute value $|\cdot|$

descends to a well-defined absolute value, which we still denote by $|\cdot|$. On \mathcal{O} the absolute value $|\cdot|$ takes values in $2^{-(\mathbb{N} \cup \{\infty\})}$ and satisfies the standard axioms ($|xy| = |x| \cdot |y|$; $|x + y| \leq \max\{|x|, |y|\}$; $|x| = 0$ if and only if $x = 0$) which make \mathcal{O} a discrete valuation ring (see [3, Chapter 9]) with uniformizer π equal to the class of $(\pi_n)_{n \in \mathbb{N}}$ in A/I . Let K be the field of fractions of \mathcal{O} . It is a field with a non-Archimedean absolute value and $\mathcal{O} = \{x \in K, |x| \leq 1\}$. Let Q be the class of $(Q_n)_{n \in \mathbb{N}}$ in $M_d(\mathcal{O})$. Then Q^q is the class of $(Q_n^q)_{n \in \mathbb{N}}$ for each q . But by assumption $|a|_n \leq \frac{1}{2^n}$ for every nondominant coefficient a of the characteristic polynomial of any matrix in Q_n^q . It follows that Q^q is made of nilpotent matrices for each q , $1 \leq q \leq d^2$. We may thus apply Lemma 2.2 to Q in $M_d(K)$. There is a matrix $g \in \mathrm{GL}_d(K)$ such that $gQg^{-1} \subset N_d(K)$. Write $g = \pi^{-L}\bar{g}$ where $\bar{g} \in M_d(\mathcal{O})$. There is $\hat{g} \in M_d(\mathcal{O})$ such that $\bar{g}\hat{g} = \det \bar{g}$ which is the transpose of the matrix of minors. We thus have $\bar{g}Q\hat{g} \subset N_d(\mathcal{O})$. This means that there is a function $f(n)$ going to $+\infty$ with n such that $\bar{g}_n Q_n \hat{g}_n \subset N_d(\mathcal{O}_{k_n}) \bmod \pi_n^{f(n)}$ for most n 's (i.e., for a set of n 's belonging to \mathcal{U}). In particular for every $M \in \mathbb{N}$, for most n 's one may find a matrix $h_n \in \mathrm{GL}_d(k_n)$ such that $h_n \bar{g}_n Q_n \hat{g}_n h_n^{-1} \subset \pi_n^{M+1} M_d(\mathcal{O}_{k_n})$ (e.g. take h_n diagonal with coefficients $\pi_n^{-i(M+1)}$, $i = 1, \dots, d$). Finally note that $\det \bar{g} \in \mathcal{O} \setminus \{0\}$ so that if $(\bar{g}_n)_n$ is a representative of \bar{g} in $M_d(A)$, there is $M \in \mathbb{N}$ such that $|\det \bar{g}_n|_n \geq 2^{-M}$ for most $n \in \mathbb{N}$. Hence $h_n \bar{g}_n Q_n \hat{g}_n h_n^{-1} \subset \pi_n M_d(\mathcal{O}_{k_n})$ for most n 's, which is the desired contradiction. \square

We can now prove (a). Let π be a uniformizer for k and let $\delta \geq 0$ be such that $\max_{1 \leq q \leq d^2} \Lambda_k(Q^q)^{\frac{1}{q}} = |\pi|_k^\delta E_k(Q)$. Assume by contradiction that $\delta > 0$. Let $m \geq N(d)/\delta$. Let $k_1 = k(\pi_1)$ where $\pi_1^m = \pi$ and $F_{k_1}(Q) = \min_{x \in \mathrm{GL}_d(k_1)} \|xQx^{-1}\|_{k_1}$. Up to conjugating by $x \in \mathrm{GL}_d(k_1)$, we may assume that $F_{k_1}(Q) = \|Q\|_{k_1} \geq E_k(Q)$. Let $Q_0 = \frac{Q}{q_0}$ for some $q_0 \in k_1$ such that $|q_0|_{k_1} = \|Q\|_{k_1}$. Then

$$\max_{1 \leq q \leq d^2} \Lambda_k(Q_0^q)^{\frac{1}{q}} \leq |\pi_1|_{k_1}^{\delta m} \leq |\pi_1|_{k_1}^{N(d)}$$

while $F_{k_1}(Q_0) = 1$. But this obviously contradicts Lemma 2.3. This ends the proof of (a). \square

Remark 2.4. In the proof we just gave of item (a) in Lemma 2.1, we used an ultralimit argument in order to establish Lemma 2.3. Passing to ultralimits allowed us to obtain a set Q made of genuinely nilpotent (instead of almost nilpotent) matrices in the ultraproduct field K and to thereby be able to apply the theorems of Wedderburn and Engel in the field K (i.e., Lemma 2.2). Without such a limiting object at our disposal, we would have had to work much harder and prove an epsilon version of the theorems of Wedderburn and Engel, where nilpotency is replaced by ε -nilpotency (see Remark 2.5 below).

Of course the use of ultralimits has the drawback that the constant $N(d)$ we get in Lemma 2.3 is noneffective. However this noneffectiveness has no effect for our purposes (and no effect on the effectivity of the height gap $\varepsilon(d)$ from Theorem 1.2) because only the equality obtained in Lemma 2.1(a) (and not the constant $N(d)$ of Lemma 2.3) will be used later. See [12] for an alternative argument for 2-by-2 matrices.

Remark 2.5. The proof of item (b) in Lemma 2.1 was by contradiction and gave no indication about how large c is. This is, in fact, the only place in this paper (and hence in the determination of the height gap $\varepsilon(d)$ from Theorem 1.2) where we have a constant which is not explicitable in principle. However we can give another proof of (b) which is constructive and gives a lower bound of order $\exp(-d^{d^2})$ for $c(d)$. We do not include this proof here because it is much lengthier and requires us to prove an approximate version of the theorems of Wedderburn and Engel valid for a set of matrices Q such that each Q^q is made of ε -nilpotent matrices (i.e., matrices all of whose eigenvalues have modulus $\leq \varepsilon$). Details can be found in [13].

Remark 2.6. Although we will not need this in the sequel, we observe in passing and also to justify the title of Lemma 2.3 that it has the following geometric interpretation in terms of the Bruhat-Tits building $\mathcal{BT}(\mathrm{GL}_d, k)$ of $\mathrm{GL}_d(k)$. Let S be a bounded subset of $\mathrm{GL}_d(k)$. If every element of S^q , $q \in [1, d^2]$, fixes pointwise a ball of radius n in $\mathcal{BT}(\mathrm{GL}_d, k)$ for the combinatorial distance, then there is a common ball of radius $\Omega_d(n)$ which is fixed pointwise by all elements in S . This statement does not follow directly from Lemmas 2.2 and 2.3, but from a simple modification of these lemmas, where one considers the k -algebra generated by the $S^q - I_d$, $q \in [1, d^2]$ in $M_d(k)$ in place of the one generated by the Q^q as in the proof of Lemma 2.3.

With the spectral radius formula at our disposal, that is Lemma 2.1, we can now understand the relationships between the various quantities at hand, i.e., the minimal norm, spectral radius and maximal eigenvalue.

PROPOSITION 2.7. *Let Q be a bounded subset of $M_d(k)$. We have*

- (i) $\Lambda_k(Q) \leq R_k(Q) \leq E_k(Q) \leq \|Q\|_k$, and $R_k(gQg^{-1}) = R_k(Q)$ for any $g \in \mathrm{GL}_d(\overline{k})$;
- (ii) $\Lambda_k(Q^n) \geq \Lambda_k(Q)^n$, $E_k(Q^n) \leq E_k(Q)^n$ and $R_k(Q^n) = R_k(Q)^n$ for all $n \in \mathbb{N}$;
- (iii) $R_k(Q) = \lim_{n \rightarrow +\infty} E_k(Q^n)^{\frac{1}{n}} = \inf_{n \in \mathbb{N}} E_k(Q^n)^{\frac{1}{n}}$;
- (iv) $R_k(Q) = \sup_{n \in \mathbb{N}} \Lambda_k(Q^n)^{\frac{1}{n}}$;
- (v) if k is non-Archimedean, $R_k(Q) = E_k(Q)$;
- (vi) if k is Archimedean, $c \cdot E_k(Q) \leq R_k(Q) \leq E_k(Q)$, where c is the constant from Lemma 2.1(b).

Proof. Items (i) and (ii) are clear from the definitions. Let us first show (iii). We have $E_k(Q^n) \leq \|Q^n\|_k$ for every $n \in \mathbb{N}$; hence $\limsup E_k(Q^n)^{\frac{1}{n}} \leq R_k(Q)$. On the other hand, $R_k(Q) = R_k(gQg^{-1}) \leq \|gQg^{-1}\|_k$ for every $g \in \mathrm{GL}_d(\overline{k})$. Hence $R_k(Q) \leq E_k(Q)$ and for every $n \in \mathbb{N}$, $R_k(Q)^n = R_k(Q^n) \leq E_k(Q^n)$, thus $R_k(Q) \leq \liminf E_k(Q^n)^{\frac{1}{n}}$. So we have shown that $\lim E_k(Q^n)^{\frac{1}{n}}$ exists and equals $R_k(Q)$. Furthermore, for every $n, p \in \mathbb{N}$, $E_k(Q^{np})^{\frac{1}{np}} \leq E_k(Q^p)^{\frac{1}{p}}$. Letting n tend to $+\infty$, we obtain $R_k(Q) \leq E_k(Q^p)^{\frac{1}{p}}$. Hence $R_k(Q) = \inf_{n \in \mathbb{N}} E_k(Q^n)^{\frac{1}{n}}$.

Now consider (iv). It is clear that as $\Lambda_k(Q^n) \leq R_k(Q^n) = R_k(Q)^n$, we have $\sup \Lambda_k(Q^n)^{\frac{1}{n}} \leq R_k(Q)$. On the other hand, given $n \in \mathbb{N}$, there is $0 \leq q \leq d^2$ from Lemma 2.1, such that $\Lambda_k(Q^{qn})^{\frac{1}{qn}} \geq c^{\frac{1}{n}} \cdot E_k(Q^n)^{\frac{1}{n}}$ (where $c = 1$ if k is non-Archimedean) which forces $\sup \Lambda_k(Q^n)^{\frac{1}{n}} \geq \limsup E_k(Q^n)^{\frac{1}{n}} = R_k(Q)$.

Now (v). From (iii) and (iv), for any $q \in \mathbb{N}$, we clearly have $\Lambda_k(Q^q)^{\frac{1}{q}} \leq R_k(Q) \leq E_k(Q)$. If k is non-Archimedean, then this fact combined with Lemma 2.1(a) shows the desired identity. If k is Archimedean, then it gives $\Lambda_k(Q^q) \leq R_k(Q)^q$, which when combined with Lemma 2.1(b) gives $c \cdot E_k(Q) \leq R_k(Q)$. \square

Remark 2.8. It can be shown that $R_k(Q)$ coincides with the infimum of $\|Q\|$ over all possible operator norms $\|\cdot\|$ not necessarily assumed to be operators norms of Euclidean or ℓ^∞ norms (see [13]). Observe, however, that when $k = \mathbb{R}$ or \mathbb{C} , then we may have $R_k(Q) < E_k(Q)$. For instance, consider $Q = \{1, T, S\} \subset \mathrm{SL}_2(\mathbb{Z})$, where T and S are the matrices corresponding to the standard generators of $\mathrm{PGL}_2(\mathbb{Z})$, i.e., $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ acts by translation by 1 and $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ by inversion around the circle of radius 1 in the upper half-plane. Then it is easy to compute $E_k(Q) = \sqrt{2} = \|tQt^{-1}\|_k$ where t is the diagonal matrix $t = \mathrm{diag}(\frac{1}{4\sqrt{2}}, \sqrt{2})$. On the other hand, one can check that $\|tQ^2t^{-1}\|_k < 2$, and thus $R_k(Q) \leq E_k(Q^2)^{\frac{1}{2}} < E_k(Q)$.

Note that if Q belongs to $\mathrm{SL}_d(k)$, then $E_k(Q) \geq R_k(Q) \geq \Lambda_k(Q) \geq 1$. The following proposition explains what happens if these quantities are close or equal to 1.

PROPOSITION 2.9 (Growth of displacement). *Suppose k is Archimedean (i.e., $k = \mathbb{R}$ or \mathbb{C}). Then for every $n \in \mathbb{N}$ and every bounded subset Q of $\mathrm{SL}_d(k)$ containing 1, we have*

$$(2) \quad E_k(Q^n) \geq E_k(Q) \sqrt{\frac{n}{4d}}$$

and

$$(3) \quad \log R_k(Q) \geq c_1 \cdot \log E_k(Q) \cdot \min\{1, \log E_k(Q)\},$$

where $c_1 = c_1(d) > 0$ is a positive constant.

Proof. We will use nonpositive curvature of the symmetric space \mathbb{X}_k associated to $\mathrm{SL}_d(k)$. Let $d(\cdot, \cdot)$ be the left invariant Riemannian metric on \mathbb{X}_k normalized in such a way that $d(ax_0, x_0)^2 = \sum_i (\log |a_i|)^2$, if $x_0 \in \mathbb{X}_k$ is the base point corresponding to $\mathrm{SO}_d(\mathbb{R})$ (resp. $\mathrm{SU}_d(\mathbb{C})$) and a is a diagonal matrix in $\mathrm{SL}_d(k)$. We set $L_k(Q) = \inf_{x \in \mathbb{X}_k} \max_{q \in Q} d(q \cdot x, x)$. Observe that $L_k(Q) \in [1, \sqrt{d}] \log E_k(Q)$ (see also Lemma 4.14 below).

Let $\ell_n := L_k(Q^n)$ and let r_n be the infimum over $x \in \mathbb{X}_k$ of the smallest radius of a closed ball containing $Q^n x$. Note first that $r_n \leq \ell_n \leq 2r_n$. Indeed if $\ell_n < t$, then there is $x \in \mathbb{X}_k$ such that $d(qx, x) < t$ for all $q \in Q^n$, i.e., $Q^n x$ lies in the ball of radius t centered at x , so $r_n \leq t$ and thus $r_n \leq \ell_n$. Similarly if $r_n < t$, then there is $x \in \mathbb{X}_k$ such that $Q^n x$ is contained in a ball of radius t . In particular, $d(y, z) \leq 2t$ for all $y, z \in Q^n x$ and thus $d(qx, x) \leq 2t$ for all $q \in Q^n$, i.e., $\ell_n \leq 2t$, so $\ell_n \leq 2r_n$.

We now prove (2). Fix $\varepsilon > 0$ and let $x, y \in \mathbb{X}_k$ be such that $Q^{n+1}x$ is contained in a ball of radius $r_{n+1} + \varepsilon$ around y . Let $q \in Q$ be arbitrary. Since Q contains 1, we have $Q^n x \subset Q^{n+1}x$, and $qQ^n x$ lies in the two balls of radius $r_{n+1} + \varepsilon$ centered around qy and around y . By the CAT(0) inequality for the median, the intersection of the two balls is contained in the ball B of radius $t := \sqrt{(r_{n+1} + \varepsilon)^2 - d(qy, y)^2}/4$ centered around the midpoint m between y and qy . Translating by q^{-1} , we get that $Q^n x$ lies in the ball of radius t centered at $q^{-1}m$. In particular $r_n \leq t$. This means $d(qy, y)^2 \leq 4((r_{n+1} + \varepsilon)^2 - r_n^2)$. Since $q \in Q$ and $\varepsilon > 0$ were arbitrary, we obtain $\ell_1^2 \leq 4(r_{n+1}^2 - r_n^2)$. Summing over n , we get $n\ell_1^2 \leq 4r_n^2 \leq 4\ell_n^2$, hence (2).

For (3), note that for every n , by Lemma 2.1 there is $q \leq d^2$ such that $\Lambda_k(Q^{qn})^{\frac{1}{q}} \geq cE_k(Q^n) \geq cE_k(Q)\sqrt{\frac{n}{4d}}$, and hence $R_k(Q) \geq c^{\frac{1}{n}}E_k(Q)\sqrt{\frac{1}{4dn}}$. Optimizing in n we obtain a constant $c_1 = c_1(d)$ for which (3) holds. \square

Remark 2.10. The above inequality (2) is interesting only when $E_k(Q)$ is small. Indeed, a better estimate holds if $E_k(Q) > \frac{1}{c}$, where c is the constant $c \in (0, 1)$ obtained in Lemma 2.1(b)

$$E_k(Q^n) \geq \max_{q \in [1, d^2]} \Lambda_k(Q^{nq})^{\frac{1}{q}} \geq \max_{q \in [1, d^2]} \Lambda_k(Q^q)^{\frac{n}{q}} \geq (cE_k(Q))^n.$$

Remark 2.11. Observe that if $Q \subset \mathrm{SL}_d(k)$, then adding the identity to Q does not modify our quantities. Namely, $Q_1 = Q \cup \{Id\}$, then $E_k(Q_1) = E_k(Q)$, $\Lambda_k(Q_1) = \Lambda_k(Q)$ and also $R_k(Q_1) = R_k(Q)$. For the last identity, note that for all $n \in \mathbb{N}$, there is $m \leq n$ such that $\Lambda_k(Q_1^n) = \Lambda_k(Q^m) \leq R_k(Q)^m \leq R_k(Q)^n$, since $R_k(Q) \geq 1$, hence taking the supremum over n , $R_k(Q_1) \leq R_k(Q)$, while the converse inequality is clear.

2.2. Height, normalized height and minimal height. Let p be a prime number (abusing notation, we allow $p = \infty$). Fix an algebraic closure $\overline{\mathbb{Q}_p}$ of the field of p -adic numbers \mathbb{Q}_p (if $p = \infty$, set $\mathbb{Q}_p = \mathbb{R}$). We take the standard

normalization of the absolute value on \mathbb{Q}_p (i.e., $|p|_p = \frac{1}{p}$), while $|\cdot|_\infty$ is the standard absolute value on \mathbb{R} . It admits a unique extension to $\overline{\mathbb{Q}_p}$, which we again denote by $|\cdot|_p$. Let $\overline{\mathbb{Q}}$ be the field of all algebraic numbers over \mathbb{Q} and K a number field. Let V_K be the set of equivalence classes of valuations on K . For $v \in V_K$ let K_v be the corresponding completion. For each $v \in V_K$, K_v is a finite extension of \mathbb{Q}_p for some prime p . We normalize the absolute value on K_v to be the unique one which extends the standard absolute value on \mathbb{Q}_p . Namely, $|x|_v = |N_{K_v|\mathbb{Q}_p}(x)|_p^{\frac{1}{n_v}}$, where $n_v = [K_v : \mathbb{Q}_p]$. Equivalently, K_v has n_v different embeddings in $\overline{\mathbb{Q}_p}$, and each of them gives rise to the same absolute value on K_v . We identify $\overline{K_v}$, the algebraic closure of K_v with $\overline{\mathbb{Q}_p}$. Let V_f be the set of finite places and V_∞ the set of infinite places.

Let $d \in \mathbb{N}$ be an integer $d \geq 2$. For $v \in V_K$, in order not to surcharge notation, we will use the subscript v instead of K_v in the quantities $E_v(F) = E_{K_v}(F)$, $\Lambda_v(F) = \Lambda_{K_v}(F)$, etc.

Recall that if $x \in K$, then its height is by definition (see e.g. [7]) the following quantity:

$$h(x) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ |x|_v.$$

It is well defined (i.e., independent of the choice of $K \ni x$). We will make constant use of the following basic inequalities valid for every algebraic numbers x and y : $h(xy) \leq h(x) + h(y)$ and $h(x+y) \leq h(x) + h(y) + \log 2$.

Let us similarly define the height of a matrix $f \in M_d(K)$ by

$$h(f) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ \|f\|_v,$$

where $\|f\|_v$ is the operator norm of f . We set the height of a finite set F of matrices in $M_d(K)$ to be

$$(4) \quad h(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ \|F\|_v,$$

where $n_v = [K_v : \mathbb{Q}_v]$ and where $\|F\|_v = \max_{f \in F} \|f\|_v$. We also define the *minimal height* of F as

$$(5) \quad e(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ E_v(F)$$

and the *normalized height* of F as

$$(6) \quad \hat{h}(F) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \log^+ R_v(F).$$

For any height \mathbf{h} (i.e., h, e or \hat{h}), we also set $\mathbf{h} = \mathbf{h}_\infty + \mathbf{h}_f$, where \mathbf{h}_∞ is the infinite part of \mathbf{h} (i.e., the part of the sum over the infinite places of K) and \mathbf{h}_f

is the finite part of \mathbf{h} (i.e., the part of the sum over the finite places of K). Note that these heights are well defined independently of the number field K such that $F \subseteq M_d(K)$. We also set $h_v(F) = \log^+ \|F\|_v$ (resp. $e_v(F) = \log^+ E_v(F)$, etc.) so that $h = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v h_v$, etc.

Remark 2.12. If we choose another basis of $\overline{\mathbb{Q}}^d$, then the new height $h_{\text{new}}(F)$ differs only from the original height by a bounded additive error. Indeed there are only finitely many places where the new standard norm may differ from the original one. On the other hand, $\hat{h}(F)$ is independent of the choice of basis.

The above terminology is justified by the following facts:

PROPOSITION 2.13. *For any finite set F in $M_d(\overline{\mathbb{Q}})$, we have:*

- (a) $\hat{h}(F) = \lim_{n \rightarrow +\infty} \frac{1}{n} h(F^n) = \inf_{n \in \mathbb{N}} \frac{1}{n} h(F^n)$;
- (b) $e_f(F) = \hat{h}_f(F)$ and $e(F) + \log c \leq \hat{h}(F) \leq e(F)$, where c is the constant in Lemma 2.1(b);
- (c) $\hat{h}(F^n) = n \cdot \hat{h}(F)$ and $\hat{h}(F \cup \{\text{Id}\}) = \hat{h}(F)$;
- (d) $\hat{h}(xFx^{-1}) = \hat{h}(F)$ if $x \in \text{GL}_d(\overline{\mathbb{Q}})$.

Proof. Since F is finite, there are only finitely many places v such that $\|F\|_v > 1$. For each such place, $\frac{1}{n} \log^+ \|F^n\|_v \rightarrow \log^+ R_v(F)$; hence $\frac{1}{n} h(F^n) \rightarrow \hat{h}(F)$. By Proposition 2.7(vii) we have $E_v(F) = R_v(F)$ if $v \in V_f$; hence $e_f(F) = \hat{h}_f(F)$, while $c \cdot E_v(F) \leq R_v(F) \leq E_v(F)$ if $v \in V_\infty$, thus $e_\infty(F) + \log c \leq \hat{h}_\infty(F) \leq e_\infty(F)$. Finally, by Proposition 2.7(ii), $R_v(F^n) = R_v(F)^n$ for every $n \in \mathbb{N}$ and every place v . Hence $\hat{h}(F^n) = n \cdot \hat{h}(F)$. \square

We also record the following simple observation.

PROPOSITION 2.14. (a) $e(xFx^{-1}) = e(F)$ for all F finite in $M_d(\overline{\mathbb{Q}})$ and $x \in \text{GL}_d(\overline{\mathbb{Q}})$;

- (b) $e(F^n) \leq n \cdot e(F)$;
- (c) If λ is an eigenvalue of an element of F , then $h(\lambda) \leq \hat{h}(F) \leq e(F)$;
- (d) If $F \subset \text{GL}_d(\overline{\mathbb{Q}})$, then $e(F \cup F^{-1}) \leq (d|F| + d - 1) \cdot e(F)$ and $e(F \cup \{1\}) = e(F)$. If F is a subset of $\text{SL}_d(\overline{\mathbb{Q}})$, then $e(F \cup F^{-1}) \leq (d - 1) \cdot e(F)$.

Proof. The first three items are clear. For the last, observe that $\|x^{-1}\|_v = \frac{1}{|\det(x)|_v} \|x\|_v^{d-1}$ for any $x \in \text{GL}_d(K_v)$ as can be seen by expressing those norms in terms of the KAK decomposition of x . Hence $\|(F \cup F^{-1})\|_v \leq \|F\|_v^{d-1} \cdot \max\{\frac{1}{|\det(x)|_v}, x \in F \cup \{1\}\}$ and $E_v(F \cup F^{-1}) \leq E_v(F)^{d-1} \cdot \max\{\frac{1}{|\det(x)|_v}, x \in F \cup \{1\}\}$. So $e(F \cup F^{-1}) \leq (d - 1)e(F) + \sum_{x \in F} h(\det(x)^{-1})$ i.e., $e(F \cup F^{-1}) \leq (d|F| + d - 1) \cdot e(F)$. \square

We can also compare $e(F)$ and $\hat{h}(F)$ when $\hat{h}(F)$ is small.

PROPOSITION 2.15. *For every $\varepsilon > 0$ there is $\delta = \delta(d, \varepsilon) > 0$ such that if F is a finite subset of $\mathrm{SL}_d(\overline{\mathbb{Q}})$ containing 1 with $\widehat{h}(F) < \delta$, then $e(F) < \varepsilon$. Moreover, $\widehat{h}(F) = 0$ if and only if $e(F) = 0$.*

This follows immediately from Proposition 2.13(b) and the following proposition.

PROPOSITION 2.16. *Let c_1 be the constant from Proposition 2.9; then*

$$\widehat{h}_\infty(F) \geq \frac{c_1}{4} \cdot e_\infty(F) \cdot \min\{1, e_\infty(F)\}$$

for any finite subset F of $\mathrm{SL}_d(\overline{\mathbb{Q}})$ containing 1.

Proof. From Proposition 2.9, $\widehat{h}_v(F) \geq c_1 \cdot e_v(F) \cdot \min\{1, e_v(F)\}$ for every $v \in V_\infty$. We may write $e_\infty(F) = \alpha e^+(F) + (1 - \alpha)e^-(F)$ where e^+ is the average of the e_v greater than 1 and e^- the average of the e_v smaller than 1 (i.e., $e^+ \sum_{v \in V_\infty, e_v > 1} n_v = \sum_{v \in V_\infty, e_v > 1} n_v e_v$ and similarly for e^-). Applying Cauchy-Schwarz, we have $\widehat{h}_\infty(F) \geq c_1 \cdot (\alpha e^+ + (1 - \alpha)e^-)^2$. If $\alpha e^+(F) \geq \frac{1}{2}e_\infty(F)$, then $\widehat{h}_\infty(F) \geq \frac{c_1}{2}e_\infty(F)$, and otherwise $(1 - \alpha)e^- \geq \frac{e_\infty}{2}$, hence $\widehat{h}_\infty(F) \geq c_1(1 - \alpha)(e^-)^2 \geq \frac{c_1}{4}e_\infty^2$. At any case $\widehat{h}_\infty(F) \geq \frac{c_1}{4} \cdot e_\infty(F) \cdot \min\{1, e_\infty(F)\}$. \square

In order to use the previous proposition inside GL_d , we shall need the following:

PROPOSITION 2.17. *For every finite set F in $\mathrm{GL}_d(\overline{\mathbb{Q}})$, then*

- (i) $\widehat{h}(\mathrm{Ad}(F)) \leq d(|F| + 1) \cdot \widehat{h}(F)$,
- (ii) $e(\mathrm{Ad}(F)) \leq d(|F| + 1) \cdot e(F)$ and
- (iii) $e(F) \leq e(\mathrm{Ad}(F)) + |F| \cdot \widehat{h}(F)$.

Proof. By Lemma 2.18 below, $\log \|\mathrm{Ad}(x)\|_v \leq d \log^+ \|x\|_v + \log^+ |\det x^{-1}|_v$ for every place v and $x \in F^n$. Thus

$$\log \|\mathrm{Ad}(F^n)\|_v \leq d \log^+ \|F^n\|_v + n \max_{f \in F} \log^+ |\det f^{-1}|_v.$$

Letting n go to infinity, we get

$$\log R_v(\mathrm{Ad}(F)) \leq d \log^+ R_v(F) + \max_{f \in F} \log^+ |\det f^{-1}|_v.$$

Summing over the places we obtain $\widehat{h}(\mathrm{Ad}(F)) \leq d\widehat{h}(F) + \sum_{f \in F} h(\det f^{-1}) \leq d(1 + |F|) \cdot \widehat{h}(F)$, where the last inequality follows from Proposition 2.14(c). The other two inequalities are proven in a similar way. \square

We used:

LEMMA 2.18. *For every local field k and every $x \in \mathrm{GL}_d(k)$, $\frac{1}{|\det(x)|_k^{1/d}} \|x\|_k \leq \|\mathrm{Ad}(x)\|_k \leq \frac{1}{|\det(x)|_k} \|x\|_k^d$, where $\mathrm{Ad}(x) \in \mathrm{GL}(M_{d,d}(k))$.*

Proof. By the Cartan decomposition, we may assume that x is diagonal $x = \text{diag}(a_1, \dots, a_d)$ with $|a_1| \geq \dots \geq |a_d|$. Then $\|x\|_k = |a_1|_k$ and $\|\text{Ad}(x)\|_k = \frac{|a_1|_k}{|a_d|_k}$. On the other hand, $|\det(x)| = |a_1 \cdots a_d|$ hence $\frac{|a_1|_k}{|\det(x)|_k^{1/d}} \leq |a_1|_k / |a_d|_k \leq \frac{|a_1|_k^d}{|\det(x)|_k}$. We are done. \square

COROLLARY 2.19. *Let F be a finite subset of $\text{GL}_d(\overline{\mathbb{Q}})$. Then $\widehat{h}(F) = 0$ if and only if $e(F) = 0$.*

Proof. By Proposition 2.17, if $\widehat{h}(F) = 0$, then $\widehat{h}(\text{Ad}(F)) = 0$. Since the elements of $\text{Ad}(F)$ have determinant 1, we may apply Proposition 2.15 and obtain $e(\text{Ad}(F)) = 0$. By the last inequality in Proposition 2.17, we get $e(F) = 0$. The converse is clear from Proposition 2.13(b). \square

Remark 2.20. In [37] a variant of our height function \widehat{h} is studied in the case when F is a single matrix. Namely setting $h_0(g) := \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \log \|g\|_v$ for $g \in M_d(\overline{\mathbb{Q}})$, then it is shown in [37] among other things that if $g \in \text{GL}_d(\overline{\mathbb{Q}})$, then $h_0(g) = \sup_{x \in \overline{\mathbb{Q}}^d \setminus \{0\}} (h_0(gx) - h_0(x))$ and that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} h_0(g^n) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \log \Lambda_v(g).$$

The results of this section can be seen as a generalization of [37] to sets F with more than one matrix.

3. Statement of the results

We state here our results. The main theorem is the following:

THEOREM 3.1 (Height gap). *There exists a positive constant $\varepsilon = \varepsilon(d) > 0$ with the following property. Let F be a finite subset of $\text{GL}_d(\overline{\mathbb{Q}})$ generating a nonvirtually solvable subgroup. Then $\widehat{h}(F) \geq \varepsilon$.*

It is easy to characterize sets of zero normalized height.

PROPOSITION 3.2 (Height zero points). *If F is a finite subset of $\text{GL}_d(\overline{\mathbb{Q}})$, then $\widehat{h}(F) = 0$ if and only if the group generated by F is virtually unipotent.*

Proof. If $\widehat{h}(F) = 0$, then $e(F) = 0$ by Corollary 2.19. Now by Proposition 2.14, $e(F \cup F^{-1}) = 0$; hence $\widehat{h}((F \cup F^{-1})^n) = n\widehat{h}(F \cup F^{-1}) = 0$ for each $n \in \mathbb{N}$. Thus every element from the group $\langle F \rangle$ generated by F has only roots of unity as eigenvalues. However, according to Theorem 6.11 in [31], $\langle F \rangle$ has a finite index subgroup Γ_0 for which no element has a nontrivial root of unity as eigenvalue. Therefore every element in Γ_0 must be unipotent, i.e., Γ_0 is unipotent. Conversely, if $\langle F \rangle$ is virtually unipotent, then every element in $\langle F \rangle$

has its eigenvalues among the roots of unity. In particular, as follows from Proposition 2.7(iv), $R_v(F) = 1$ for every place v . Hence $\widehat{h}(F) = 0$. \square

The above results dealt with small values of the normalized height. The following proposition says in substance that, provided $\langle F \rangle$ has semisimple Zariski closure, the normalized height is attained up to a constant by the height of some suitable conjugate of F . We have

PROPOSITION 3.3 (Comparison between h and \widehat{h}). *If \mathbb{G} is a semisimple algebraic group over $\overline{\mathbb{Q}}$ and (ρ, V) is a finite-dimensional linear representation of \mathbb{G} , then there is $C \geq 1$ and there is a choice of a basis on V with associated height function h on $\mathrm{End}(V)$, such that if F is any finite subset of $\mathbb{G}(\overline{\mathbb{Q}})$ generating a Zariski-dense subgroup of \mathbb{G} , we have*

$$\widehat{h}(\rho(F)) \leq e(\rho(F)) \leq h(\rho(gFg^{-1})) \leq C \cdot \widehat{h}(\rho(F))$$

for some $g \in \mathbb{G}(\overline{\mathbb{Q}})$.

Recall from Remark 2.12 that if we change the basis of V , then the associated height differs from the original one only by an additive constant. This proposition subsumes Proposition 1.1 from the introduction. It is important for the applications as it allows us to conjugate F back in the “right position”. Observe that by definition $e(F)$ is equal to the infimum of $h(gFg^{-1})$ when $g = (g_v)_{v \in V_K}$ is allowed to vary among the full group of adèles $\mathrm{GL}_d(\mathbb{A})$. This proposition shows that this infimum is attained up to a multiplicative constant on principal adèles, i.e., on $\mathrm{GL}_d(\overline{\mathbb{Q}})$. The condition that the Zariski closure of the group generated by F should be semisimple is important, as easy examples show that the result of the proposition can fail if for instance F normalizes a unipotent subgroup.

The normalized height \widehat{h} was defined for an arbitrary finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$. If \mathbb{G} is an arbitrary semisimple group, then one can define the normalized height for \mathbb{G} as the one you obtain after taking some absolutely irreducible representation of \mathbb{G} which is nontrivial on each factor of \mathbb{G} . The following proposition shows that up to constants, this height is independent of the choice of the representation.

PROPOSITION 3.4 (Invariance under change of representation). *Let \mathbb{G} be a semisimple algebraic group over $\overline{\mathbb{Q}}$ and (ρ_i, V_i) for $i = 1, 2$ be two finite-dimensional linear representations of \mathbb{G} which are nontrivial on each simple factor of \mathbb{G} . Let h_i be a height function on $\mathrm{End}(V_i)$ defined as above by the choice of a basis in each V_i . Then there are constants $C_{12}, C'_{12} \geq 1$ such that for any finite subset F of $\mathbb{G}(\overline{\mathbb{Q}})$, we have*

$$\frac{1}{C_{12}} \cdot h_2(\rho_2(F)) - C'_{12} \leq h_1(\rho_1(F)) \leq C_{12} \cdot h_2(\rho_2(F)) + C'_{12}.$$

In particular,

$$\frac{1}{C_{12}} \cdot \widehat{h}_2(\rho_2(F)) \leq \widehat{h}_1(\rho_1(F)) \leq C_{12} \cdot \widehat{h}_2(\rho_2(F)).$$

Moreover, the constant C_{12} depends only on ρ_1 and ρ_2 , and is independent of the choice of basis used to define h_1 and h_2 .

Finally, we record the following consequences.

COROLLARY 3.5. *There are constants $\varepsilon = \varepsilon(d)$, $\kappa = \kappa(d) \in \mathbb{N}$ and $C = C(d) \in \mathbb{N}$ such that if F is any finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ containing 1, there is some $a \in F^\kappa$ and some eigenvalue λ of a such that*

$$h(\lambda) \geq \frac{1}{|F|^C} \cdot \widehat{h}(F).$$

As a corollary of this and the height gap theorem we obtain an effective version of Schur's classical result on torsion linear groups (see [33]).

COROLLARY 3.6 (Effective Schur: no large torsion balls). *There is an integer $N_2 = N_2(d) \in \mathbb{N}$ such that if K is a field and if F is a finite subset of $\mathrm{GL}_d(K)$ containing 1, then either it generates a finite subgroup, or $(F \cup F^{-1})^{N_2(d)}$ contains an element of infinite order. Furthermore, if F generates a nonvirtually nilpotent subgroup, then we can find the element of infinite order already in $F^{N_2(d)}$.*

The following example gives a situation showing that without the assumption on F in the last sentence of this corollary, the conclusion may fail. Consider the subgroup of $\mathrm{GL}_2(\mathbb{C})$ consisting of affine transformations of the complex line. Then, for arbitrary $N \in \mathbb{N}$ one may find a finite (nonsymmetric!) set F containing the identity such that the group generated by F is infinite and virtually abelian, while F^N consists solely of elements of finite order. For instance, take $F = \{\mathrm{id}, a_\omega, ta_\omega t^{-1}\}$ where $a_\omega = \begin{pmatrix} \omega & 0 \\ 0 & 1 \end{pmatrix}$ is multiplication by ω (a root of 1 of order $N+1$) and $t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is translation by 1. Then the commutator $[a_\omega, ta_\omega t^{-1}]$ is $\neq 1$ if $N \geq 0$ and unipotent so of infinite order, while F^N is made of homotheties of ratio ω^k with $1 \leq k \leq N$ (i.e., elements of the form $\begin{pmatrix} \omega^k & * \\ 0 & 1 \end{pmatrix}$), which are all torsion elements.

Remark 3.7. In the entire paper we work over $\overline{\mathbb{Q}}$. However a fair amount of what we do remains valid over global fields of positive characteristic, i.e., over the algebraic closure of $\mathbb{F}_p(t)$. In particular, the definition of the heights makes sense, except that all places are non-Archimedean. Also all properties of Section 2 hold in positive characteristic as well, and they even become simpler since all places are non-Archimedean and can thus be treated on an equal footing, and $e(F) = \widehat{h}(F)$ always. Proposition 3.4 remains true for irreducible representations of \mathbb{G} . Moreover the additive constant disappears. Also Proposition 3.3 remains true for irreducible representations. The same is

true for Corollary 3.5. This is key for the applications to the Tits alternative in positive characteristic proved in [11]. The proof of these propositions is word-by-word the same as in the $\overline{\mathbb{Q}}$ case, except for the proof of Proposition 3.3 which needs some mild modification if the characteristic is 2 or 3 or if \mathbb{G} is of type A (see Remark 7.3). Theorem 3.1 however has no direct analog in positive characteristic (nor does Zhang's Theorem 7.1): for a counterexample take F_n to be the two-element set in SL_2 consisting of an upper triangular and a lower triangular unipotent matrix with coefficient $t^{\frac{1}{n}}$; then F_n generates a Zariski-dense subgroup, but $\hat{h}(F_n) \rightarrow 0$. Nevertheless this is not a problem for the applications to the Tits alternative, since all places being non-Archimedean in positive characteristic, only the positivity of \hat{h} matters there. See [11] for more on positive characteristic.

Remark 3.8. Another possible definition of our height functions h, \hat{h} and e consists in replacing the \log^+ by \log in (4), (5) and (6). This new definition (let us denote it by h_0 and \hat{h}_0) is more adapted to PGL_d while ours is more adapted to SL_d , but the differences are minor. First of all, it is clear that the two notions coincide if $F \subset SL_d$, because each norm $\|F\|_v$ is then greater or equal to 1. Moreover, $h_0(F) \geq 0$ for all F (from the product formula applied to any eigenvalue of an element of F , say). Also $h_0(\lambda F) = h_0(F)$ for all $\lambda \in \overline{\mathbb{Q}}^\times$, and $h(F) = h_0(\rho(F))$ where ρ is the obvious embedding of GL_d inside GL_{d+1} in the upper-left corner. Of course $\hat{h}_0(F) \leq \hat{h}(F)$.

Moreover, Theorem 3.1 also holds for \hat{h}_0 . This follows easily from Corollary 1.7. Indeed, let $F' = \{f/(\det f)^{1/d}, f \in F\}$; then $\langle F' \rangle$ is virtually solvable if and only if $\langle F \rangle$ is. By Corollary 1.7 there is $g \in F'^{N_1(d)}$ and an eigenvalue λ of g such that $h(\lambda) > \varepsilon = \varepsilon(d) > 0$. But $h(\lambda) \leq \hat{h}_0(\{g\})$, because $g \in SL_d$, and there is $\mu \in \overline{\mathbb{Q}}^\times$ such that $\mu g \in F^{N_1(d)}$. So $h(\lambda) \leq \hat{h}_0(\{g\}) = \hat{h}_0(\{\mu g\}) \leq N_1(d)\hat{h}_0(F)$. Hence the result.

4. Preliminary reductions

The main goal of this section is to establish Proposition 4.11 below, which reduces the proof of Theorem 3.1 to the case when $F = \{a, b\}$ is a finite set of two regular semisimple elements generating a Zariski dense subgroup inside $\mathbb{G}(\overline{\mathbb{Q}})$, where \mathbb{G} is a Zariski-connected absolutely simple algebraic group of adjoint type defined over $\overline{\mathbb{Q}}$, and where the underlying vector space is the Lie algebra \mathfrak{g} of \mathbb{G} on which \mathbb{G} acts via the adjoint representation, so that $\mathbb{G} \subset SL(\mathfrak{g})$.

4.1. Escape and reduction to a 2-element set. In this section, we prove Proposition 1.8 from the introduction in the slightly stronger form given below in Proposition 4.1. The key ingredient there is a lemma due to Eskin-Mozes-Oh about escaping from algebraic subvarieties in bounded time.

First we recall some terminology. Let \mathbb{G} be a connected semisimple algebraic group over $\overline{\mathbb{Q}}$. A semisimple group element $a \in \mathbb{G}(\overline{\mathbb{Q}})$ is said to be regular if $\ker(\mathrm{Ad}(a) - 1)$ has the minimal possible dimension (namely equal to the absolute rank of \mathbb{G}). For $A_1 \in \mathbb{N}$, we will say that $a \in \mathbb{G}(\overline{\mathbb{Q}})$ is A_1 -regular if $\ker(\mathrm{Ad}(a) - \omega)$ has minimal possible dimension for every root of unity ω of order at most A_1 (namely dimension 0 if $\omega \neq 1$ and the absolute rank if $\omega = 1$). It is clear that the subset of A_1 -regular elements of \mathbb{G} is a nonempty Zariski open subset of \mathbb{G} consisting of semisimple elements.

If Z is a proper Zariski closed subset of \mathbb{G} invariant under conjugation by a maximal torus T , then we let \widehat{Z} be the Zariski-closure of $\{(gag^{-1}, gbg^{-1}) \in \mathbb{G}^2 \text{ with } g \in \mathbb{G}, a \in T \text{ and } b \in Z, \text{ or } a \in Z \text{ and } b \in T\}$. It is a proper algebraic subset of $\mathbb{G} \times \mathbb{G}$ of dimension at most $2 \dim \mathbb{G} - 1$.

PROPOSITION 4.1. *Let \mathbb{G} be a connected semisimple algebraic subgroup of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ with maximal torus T . Let Z be a proper Zariski closed subset of \mathbb{G} invariant under conjugation by T . Then there is an integer $c = c(\mathbb{G}, Z) > 0$ such that if F is a finite subset of $\mathbb{G}(\overline{\mathbb{Q}})$ generating a Zariski-dense subgroup in \mathbb{G} , then $(F \cup \{1\})^{c(d)}$ contains two elements a and b which are regular semisimple, generate a Zariski dense subgroup of \mathbb{G} , and satisfy $(a, b) \notin \widehat{Z}$. For any given integer $A_1 \in \mathbb{N}$, by allowing c to depend also on A_1 , i.e., $c = c(\mathbb{G}, Z, A_1) > 0$, we may further assume that a and b are A_1 -regular.*

The key ingredient in this proposition is the following lemma. For an algebraic variety X we will denote by $m(X)$ the sum of the degree and the dimension of each of its irreducible components.

LEMMA 4.2 (Eskin-Mozes-Oh escape lemma [21, Lemma 3.2]). *Given an integer $m \geq 1$ there is $N = N(m)$ such that for any field K , any integer $d \geq 1$, any K -algebraic subvariety X in $\mathrm{GL}_d(K)$ with $m(X) \leq m$ and any subset $F \subset \mathrm{GL}_d(K)$ which contains the identity and generates a subgroup which is not contained in $X(K)$, we have $F^N \not\subseteq X(K)$.*

This result is a consequence of the following generalized version of Bezout's theorem about the intersection of finitely many algebraic subvarieties (see Zannier's appendix in [32]).

THEOREM 4.3 (Generalized Bezout theorem). *Let K be a field, and let Y_1, \dots, Y_p be pure dimensional algebraic subvarieties of K^n . Denote by W_1, \dots, W_q the irreducible components of $Y_1 \cap \dots \cap Y_p$. Then $\sum_{i=1}^q \deg(W_i) \leq \prod_{j=1}^p \deg(Y_j)$.*

In order to apply the escape lemma to the proof of Proposition 4.1, we need:

PROPOSITION 4.4. *Let \mathbb{G} be a connected semisimple algebraic group over \mathbb{C} . There is a proper algebraic subvariety X of $\mathbb{G} \times \mathbb{G}$ such that any pair $(x, y) \notin X$ is made of regular semisimple elements which generate a Zariski-dense subgroup of \mathbb{G} .*

Proof. Recall the well-known:

LEMMA 4.5. *The set U of regular semisimple elements of \mathbb{G} is a nonempty Zariski-open subset of \mathbb{G} .*

Proof. The set U coincides with the set of $g \in \mathbb{G}$ such that $\ker(\mathrm{Ad}(g) - 1)$ is of minimal dimension. This is clearly a Zariski-open condition. \square

We will make use of Jordan's theorem on finite subgroups of $\mathrm{GL}_d(\mathbb{C})$ (see [19]). Recall that according to this theorem, there is a constant $C = C(d) \in \mathbb{N}$, such that if Γ is a finite subgroup of $\mathrm{GL}_d(\mathbb{C})$, then Γ contains a abelian subgroup A with $[\Gamma : A] \leq C(d)$. As the kernel of the adjoint representation coincides with the center of \mathbb{G} , it follows that the same bound apply for all finite subgroups of $\mathbb{G}(\mathbb{C})$ as long as $\dim(\mathbb{G}) \leq d$. Let $V(\mathbb{G})$ be the proper Zariski-closed subset of $\mathbb{G} \times \mathbb{G}$ consisting of all couples (x, y) such that $[x^{C!}, y^{C!}] = 1$. By Jordan's theorem, if $(x, y) \notin V$, then the subgroup generated by x and y infinite.

Let $(\mathbb{G}_i)_{1 \leq i \leq k}$ be the \mathbb{C} -simple factors of \mathbb{G} , together with their factor maps $\pi_i : \mathbb{G} \rightarrow \mathbb{G}_i$. For convenience, let us denote $\mathbb{G}_0 = \mathbb{G}$. Let X_i , for $0 \leq i \leq k$, be the subset of $\mathbb{G} \times \mathbb{G}$ consisting of couples (x, y) such that the \mathbb{C} -subalgebra of $\mathrm{End}(\mathfrak{g}_i)$ generated by $\mathrm{Ad}(\pi_i(x))$, and $\mathrm{Ad}(\pi_i(y))$ is of strictly smaller dimension than the subalgebra generated by the full of $\mathrm{Ad}(\mathbb{G}_i)$, where \mathfrak{g}_i is the Lie algebra of \mathbb{G}_i . This is a Zariski-closed subset of $\mathbb{G} \times \mathbb{G}$. According to [9, VIII.2, Ex. 8], each \mathfrak{g}_i is generated by two elements. It follows that X_i is a proper closed subvariety. Also let V_i be the set of couples $(x, y) \in \mathbb{G} \times \mathbb{G}$ such that $(\pi_i(x), \pi_i(y)) \in V(\mathbb{G}_i)$, where $V(\mathbb{G}_i)$ is the proper closed subset defined above.

Finally, let X be the proper closed subvariety $X = U^c \cup \bigcup_i X_i \cup \bigcup_i V_i$. Let us verify that X satisfies the conclusion of the proposition. Suppose $(x, y) \notin X$. Then $(x, y) \in U$ and x, y are regular semisimple. Let \mathbb{H} be the Zariski closure of the group generated by x and y . Let \mathfrak{h}_i be the Lie algebra of $\pi_i(\mathbb{H})$, which is a Lie subalgebra of \mathfrak{g}_i . As \mathfrak{h}_i is invariant under $\mathrm{Ad}(\pi_i(x))$ and $\mathrm{Ad}(\pi_i(y))$, it must be invariant $\mathrm{Ad}(\mathbb{G}_i)$, by the assumption that $(x, y) \notin X_i$. Therefore \mathfrak{h}_i is an ideal of \mathfrak{g}_i . As \mathfrak{g}_i is a simple Lie algebra, either $\mathfrak{h}_i = \{0\}$ or $\mathfrak{h}_i = \mathfrak{g}_i$. In the former case, this means that $\pi_i(\mathbb{H})$ is finite. However, by assumption $(\pi_i(x), \pi_i(y)) \notin V(\mathbb{G}_i)$, this means that the group generated by $\pi_i(x)$ and $\pi_i(y)$ is infinite. So $\pi_i(\mathbb{H})$ is not finite, $\mathfrak{h}_i = \mathfrak{g}_i$ and $\pi_i(\mathbb{H}) = \mathbb{G}_i$.

On the other hand, since $(x, y) \notin X_0$, the same argument shows that the Lie algebra of \mathbb{H} itself is an ideal in \mathfrak{g} . Hence \mathbb{H}° is a normal subgroup of \mathbb{G} , hence is the product of the simple factors of \mathbb{G} contained in it. The fact that $\pi_i(\mathbb{H}) = \mathbb{G}_i$ for each i forces $\mathbb{H} = \mathbb{G}$. \square

Proof of Proposition 4.1. This is immediate by the combination of Proposition 4.4 and Lemma 4.2 applied to $F \times F$ in $\mathbb{G} \times \mathbb{G}$. \square

4.2. *Reduction to semisimple \mathbb{G} .* This section is devoted to the proof of

PROPOSITION 4.6. *In order to prove Theorem 3.1, it is enough to prove the following assertion. There is $\varepsilon = \varepsilon(d) > 0$ such that: If $\mathbb{G} \subseteq \mathrm{SL}_d$ is a semisimple algebraic group over $\overline{\mathbb{Q}}$ acting irreducibly on $\overline{\mathbb{Q}}^d$ and $F = \{\mathrm{Id}, a, b\}$ is a subset of \mathbb{G} generating a Zariski-dense subgroup, then $e(F) > \varepsilon(d)$.*

The proof of this will rest mainly on the following proposition.

PROPOSITION 4.7. *There are constants $C = C(d) > 0$ and $m = m(d) \in \mathbb{N}$ such that if F is a finite subset of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ containing 1 and generating a nonvirtually solvable subgroup, there exists a subset $F_1 \subset F^m$, a connected semisimple algebraic group \mathbb{H} together with a faithful irreducible representation (ρ_0, V_0) of \mathbb{H} with $\dim V_0 \leq d$ and a homomorphism $\pi : \Gamma_0 \rightarrow \mathbb{H}(\overline{\mathbb{Q}})$, where Γ_0 contains F_1 and has index at most m in $\Gamma = \langle F \rangle$, such that $\pi(\Gamma_0)$ is Zariski dense in \mathbb{H} and*

$$e(\rho_0 \circ \pi(F_1)) \leq C(d) \cdot e(F).$$

The proof of this proposition will occupy the rest of this subsection. At the end we derive Proposition 4.6 from it.

We first analyse the local behavior at each place. Let K be a number field and $(e_i)_{1 \leq i \leq d}$ be the canonical basis of $V = K^d$. Let $V = \bigoplus_{1 \leq i \leq m} V_i$ be a direct sum decomposition adapted to this basis, i.e., there are indices $j_1 < \dots < j_m$ such that $V_i = \mathrm{span}\{e_{j_i}, \dots, e_{j_{i+1}-1}\}$. Let P be the group of block upper-triangular matrices determined by the corresponding flag, i.e., the parabolic subgroup of GL_d fixing the flag. Let $\rho : P \rightarrow \mathrm{GL}_d$ be the natural homomorphism that sends a matrix $A = (a_{ij})_{ij} \in P$ to the matrix $\rho(A) = (a'_{ij})_{ij}$ with $a'_{ij} = a_{ij}$, if e_i and e_j belong to the same V_k and $a'_{ij} = 0$ otherwise.

LEMMA 4.8. *Let $v \in V_K$ be a place of K . Let F be a finite subset of $\mathrm{GL}_d(K) \cap P$. Then*

$$E_v(\rho(F)) = E_v(F).$$

Proof. One needs first to observe that if $\|\cdot\|$ is any standard norm (i.e., a Euclidean norm associated to some basis of k^d when k is archimedean, a sup-norm associated to some \mathcal{O}_k lattice in k^d , say R , when k is ultrametric), then $\|\rho(x)\|_v \leq \|x\|_v$ for every $x \in P$. This fact easily follows after we check that there is a direct sum decomposition of K_v^d as $\bigoplus_{1 \leq i \leq m} W_i$ where the W_i 's are orthogonal (Archimedean case) or give rise to a direct factor decomposition $R = \bigoplus_{1 \leq i \leq m} (W_i \cap R)$ (ultrametric case) and for which x remains block upper-triangular in any basis adapted to this decomposition. From this we get the

first half of the claimed relation, i.e., $E_v(\rho(F)) \leq \inf_{g \in GL_d(\overline{\mathbb{Q}}_v)} \|g\rho(F)g^{-1}\| \leq \inf_{g \in GL_d(\overline{\mathbb{Q}}_v)} \|gFg^{-1}\|_v = E_v(F)$.

The second half follows from the remark that $\rho(F)$ can be approximated uniformly by the $\delta F\delta^{-1}$'s for some suitably chosen $\delta \in \Delta(\overline{\mathbb{Q}}_v)$, where Δ is the group of block scalar matrices associated with the V_i 's. Indeed we get

$$\begin{aligned} E_v(F) &= \inf_{g \in GL_d(\overline{\mathbb{Q}}_v)} \|gFg^{-1}\|_v = \inf_{g \in GL_d(\overline{\mathbb{Q}}_v)} \inf_{\delta \in \Delta(\overline{\mathbb{Q}}_v)} \|g\delta F\delta^{-1}g^{-1}\|_v \\ &\leq \inf_{g \in GL_d(\overline{\mathbb{Q}}_v)} \|g\rho(F)g^{-1}\|_v = E_v(\rho(F)). \end{aligned} \quad \square$$

This lemma gives that if $\overline{\mathbb{Q}}^d = \bigoplus_{1 \leq i \leq m} V_i$ is a direct sum decomposition associated to a composition series for \mathbb{G} , then $e(\rho(F)) = e(F)$. Moreover $\langle F \rangle$ is virtually solvable if and only if $\rho(\langle F \rangle)$ is virtually solvable and if and only if each $\rho_i(\langle F \rangle)$ is virtually solvable, where ρ_i is the induced action on V_i . Hence there must be one ρ_{i_0} for which $\rho_{i_0}(\langle F \rangle)$ is not virtually solvable. Note that $e(\rho_{i_0}(F)) \leq e(F)$.

Let \mathbb{H}_0 be the Zariski closure of $\rho_{i_0}(F)$ in $GL(V_{i_0})$. Note at this point that if we knew that \mathbb{H}_0 was connected semisimple, we would be done.

Clearly, the connected component \mathbb{H}_0° is a reductive group, since a nontrivial unipotent radical would have a nontrivial pointwise fixed subspace: this subspace would then be globally invariant under \mathbb{H}_0 and contradict the irreducibility of the action on V_{i_0} .

Let W_1 be a \mathbb{H}_0° -irreducible subspace of minimal dimension in V_{i_0} . As \mathbb{H}_0° is normal in \mathbb{H}_0 , and \mathbb{H}_0 acts irreducibly on V_{i_0} , we have a direct sum decomposition $V_{i_0} = \bigoplus_{1 \leq j \leq q} W_j$ into \mathbb{H}_0° -irreducible subspaces where $\mathbb{H}_0/\mathbb{H}_0^\circ$ permutes transitively the W_j 's. Since \mathbb{H}_0 is not virtually solvable, \mathbb{H}_0° is not solvable; thus its image into $GL(W_1)$ (say, all W_j are isomorphic representations of \mathbb{H}_0°) is not solvable. Observe that, since \mathbb{H}_0° is reductive and acts irreducibly on W_1 , its center must act by homotheties (by Schur's lemma), hence the semisimple part, say \mathbb{S} , of \mathbb{H}_0° also acts irreducibly.

Let \mathbb{H}_1 be the stabilizer of W_1 in \mathbb{H}_0 . Then $[\mathbb{H}_0 : \mathbb{H}_1] \leq q \leq d$. We now use:

LEMMA 4.9. *Suppose \mathbb{L} is a linear algebraic group with \mathbb{L}° reductive. Let \mathbb{S} be the semisimple part of \mathbb{L}° ($\mathbb{S} = [\mathbb{L}^\circ, \mathbb{L}^\circ]$) and \mathbb{Z} be the centralizer of \mathbb{S} in \mathbb{L} . Then $[\mathbb{L} : \mathbb{Z}\mathbb{S}] \leq c(d)$, where $c(d)$ is a constant depending only on $d = \dim(\mathbb{L})$.*

Proof. The group \mathbb{S} is normal in \mathbb{L} ; let $\sigma : \mathbb{L} \rightarrow \text{Aut}(\mathbb{S})$ be the map given by conjugation. It induces $\bar{\sigma} : \mathbb{L} \rightarrow \text{Out}(\mathbb{S})$. But $\text{Out}(\mathbb{S})$ is a finite group whose order depends only on the Dynkin diagram of \mathbb{S} ; hence it is bounded in terms of d only (see [8, 14.9]). Let \mathbb{K} be the kernel of $\bar{\sigma}$. Then $[\mathbb{L} : \mathbb{K}] \leq c(d)$ by the latter remark. On the other hand, by definition of \mathbb{K} , $\mathbb{K} = \mathbb{Z}\mathbb{S}$. \square

We apply this lemma to $\mathbb{L} = \mathbb{H}_1$. Since \mathbb{S} acts irreducibly on W_1 , \mathcal{Z} must act by homotheties (Schur's lemma). Set $\mathbb{H}_2 = \mathcal{Z}\mathbb{S}$. We have $\mathbb{H}_0^\circ \subset \mathbb{H}_2$ and $[\mathbb{H}_0 : \mathbb{H}_2] \leq dc(d)$. Also $[\Gamma : \Gamma_0] \leq d$ where $\Gamma_0 = \Gamma \cap \mathbb{H}_2$ is Zariski dense in \mathbb{H}_2 . By the (well-known) Lemma 4.10 below, we may find a finite set F_0 in $(F \cup \{1\})^{2dc(d)-1}$ containing 1 such that $\langle F_0 \rangle = \Gamma_0$. Moreover $e(F_0) \leq e(F^{2dc(d)-1}) \leq (2dc(d) - 1)e(F)$.

LEMMA 4.10. *Let F be a finite subset of a group Γ containing 1. Assume that the elements of F (together with their inverses) generate Γ . Let Γ_0 be a subgroup of index k in Γ . Then F^{2k-1} contains a generating set of Γ_0 .*

Proof. It is clear that F^{k-1} contains a set of representatives for each left coset in Γ/Γ_0 , say $\{s_1, \dots, s_k\}$. Similarly, $(F^{-1})^{k-1}$ contains a set of representatives of the left cosets, say $\{u_1, \dots, u_k\}$. Consider all elements of Γ_0 of the form $s_i f u_j^{-1}$ for $i, j \in [1, k]$ and $f \in F$. They all belong to F^{2k-1} . It is straightforward to verify that, together with their inverses, they generate Γ_0 . \square

In order to get rid of \mathcal{Z} , we now consider the action of \mathbb{H}_2 by conjugation on $\text{End}(W_1)$. The action factors through \mathbb{S} , hence the image is a connected semisimple algebraic subgroup of $\text{GL}(\text{End}(W_1))$, say \mathbb{H}_3 . Moreover, we can bound the new height in terms of the old one by making use of Proposition 2.17 above. In particular, if F_1 is any subset of $\mathbb{H}_2(\overline{\mathbb{Q}})$, then $e(\text{Ad}(F_1)) \leq d(|F_1| + 1) \cdot e(F_1)$.

By Proposition 4.1 above (or Proposition 1.8 from the introduction), we may find a pair a, b in $F_0^{c_2(d)}$ (for some constant $c_2(d)$) which generates modulo \mathcal{Z} a Zariski dense subgroup of \mathbb{H}_3 . Let $F_1 = \{1, a, b\}$. Then $e(\text{Ad}(F_1)) \leq 4d \cdot e(F_1) \leq 4d \cdot c_2(d) \cdot e(F_0)$ and $e(\text{Ad}(F_1)) \leq O_d(1) \cdot e(F)$ where $O_d(1) = 8d^2 c(d) c_2(d)$.

Now the group $\langle \text{Ad}(F_1) \rangle$ is Zariski dense in \mathbb{H}_3 , and we may apply verbatim the beginning of the proof to this group, to conclude that for some irreducible subrepresentation of \mathbb{H}_3 on $\text{End}(W_1)$, say (ρ, \overline{W}) we have $e(\rho(\text{Ad}(F_1))) \leq e(\text{Ad}(F_1)) \leq O_d(1) \cdot e(F)$. Set \mathbb{H} to be the image of \mathbb{H}_3 in $\text{GL}(\overline{W})$. Clearly Γ_0 acts on \overline{W} with Zariski closure \mathbb{H} . Thus the proof of Proposition 4.7 is complete.

Proof of Proposition 4.6. In the setting of Theorem 3.1 we first reduce to proving a gap for $e(F)$ instead of $\hat{h}(F)$. This can indeed be achieved since, with the notation of the last section, $\hat{h}(F) = \frac{1}{C_d} \hat{h}(F^{C_d}) \geq \frac{1}{C_d} \hat{h}(F_1)$ with $C_d = 2dc(d)c_2(d)$. Moreover, Proposition 2.17 also yields $\hat{h}(\text{Ad}(F_1)) \leq d(|F_1| + 1) \cdot \hat{h}(F_1) \leq O_d(1) \hat{h}(F)$. But $\text{Ad}(F_1)$ lies in matrices with determinant 1, and generates a nonvirtually solvable subgroup; hence Proposition 2.15 shows that $\hat{h}(\text{Ad}(F_1))$ is bounded away from 0 if and only if $e(\text{Ad}(F_1))$ is. But

$e(\rho(\text{Ad}(F_1))) \leq e(\text{Ad}(F_1))$ and $\rho(\text{Ad}(F_1))$ generates a Zariski dense subgroup of the semisimple algebraic group \mathbb{H} . Applying Proposition 1.8 we are done. \square

4.3. Comparison of heights under different representations. In this section we prove Proposition 3.4 and we conclude the reduction step of Theorem 3.1 by proving Proposition 4.11 below.

First let us recall some facts about representations of Chevalley groups. Let \mathbb{G} be a semisimple algebraic group over $\overline{\mathbb{Q}}$. The group \mathbb{G} is a Chevalley group and comes with an associated \mathbb{Z} structure. For general background on Chevalley groups we refer the reader to Steinberg [35] and to Bourbaki [9, Ch. 8]. We let $\mathfrak{g}_{\mathbb{Z}}$ be a Chevalley order corresponding to \mathbb{G} on the Lie algebra \mathfrak{g} of \mathbb{G} and \mathfrak{a} the associated Cartan subalgebra in \mathfrak{g} . Also let (Y_1, \dots, Y_d) be a Chevalley basis of $\mathfrak{g}_{\mathbb{Z}}$ so that the Y_i 's for $i \in [|\Phi^+| + 1, |\Phi^+| + r]$ span the admissible lattice $\mathfrak{g}_{\mathbb{Z}} \cap \mathfrak{a}$ of \mathfrak{a} (here Φ^+ is the set of positive roots and r the absolute rank of \mathbb{G}). We denote by T the maximal split torus of \mathbb{G} corresponding to \mathfrak{a} and by τ the Cartan involution.

Given a local field k , we define the “Killing norm” $\|\cdot\|_{\text{Kill},k}$ on \mathfrak{g}_k to be the one given by the Killing form $B_{\mathfrak{g}}$ when k is Archimedean (i.e., $\|X\|_{\text{Kill},k} = -B_{\mathfrak{g}}(X^{\tau}, X)$) and the one arising from the lattice $\mathfrak{g}_{\mathbb{Z}} \otimes \mathcal{O}_k = \mathfrak{g}_{\mathcal{O}_k}$ when k is ultrametric (i.e., $\|X\|_{\text{Kill},k} = \max_i |x_i|_k$ if $X = \sum x_i Y_i$). This allows us to define what we will call the “Killing height” $h_{\text{Kill}}(F)$ for $F \subseteq \mathbb{G}(\overline{\mathbb{Q}})$ by the usual formula (4) where we use the Killing norm at each place.

We denote by K_0 the stabilizer of $\|\cdot\|_{\text{Kill},k}$. It is a maximal compact subgroup of $\mathbb{G}(k)$. It is also a *good* maximal compact subgroup in the sense of [16, 3.3], that is K_0 contains a copy of the Weyl group, so that $N_{K_0}(T(k))T(k) = N_{\mathbb{G}(k)}(T(k))$.

Let V, ρ_V be a finite-dimensional linear representation of \mathbb{G} which is non-trivial on each factor of \mathbb{G} . By Steinberg [35, §2, Cor. 1], there exists an integer lattice, say $V_{\mathbb{Z}}$, of V which is invariant under $\mathbb{G}(\mathbb{Z})$ and which is spanned by a basis (Y_1, \dots, Y_D) made of weight vectors for the action of T . When k is ultrametric $V_{\mathcal{O}_k} = V_{\mathbb{Z}} \otimes \mathcal{O}_k$ defines the following norm on $V_k = V_{\mathbb{Z}} \otimes k$. We denote it by $\|X\|_{\rho_V,k} := \max_i |x_i|_k$ if $X = \sum_{i=1}^D x_i Y_i \in V_k$. When k is Archimedean, then there exists a hermitian scalar product on V_k which is invariant under K_0 and for which $\mathbb{G}(k)$ is stable under taking the adjoint (see [29]). We denote again by $\|\cdot\|_{\rho_V,k}$ the corresponding hermitian norm. Together these norms define a height function h_{ρ_V} on finite subsets of $\text{End}(V)$ defined as in (4). When V, ρ_V is the adjoint representation, the just defined norms and height coincide with the Killing norms and height.

Proof of Proposition 3.4. By complete reducibility (true in characteristic zero, in positive characteristic one has to assume irreducibility to begin with), we may assume that both representations are irreducible, with highest weight

χ_1 and χ_2 respectively. Let W be the Weyl group of \mathbb{G} . If $g \in T$, then $\|\rho_i(g)\|_{\rho_i, k} = \max_{w \in W} |\chi_i(w(g))|_k$. Since the root lattice is of finite index in the weight lattice, there exists $n_0 = n_0(\mathbb{G}) \in \mathbb{N}$ such that $n_0\chi_i$ is a linear combination $\sum_{\alpha \in \Pi} n_\alpha^{(i)}\alpha$ with nonnegative integer coefficients of the simple roots $\alpha \in \Pi$ of \mathbb{G} . Since the inverse of the Cartan matrix of an irreducible root system has no zero entry (see [9]), and since each ρ_i is nontrivial on each nontrivial factor of \mathbb{G} , the coefficients n_α are nonzero. It follows that

$$\begin{aligned} \|\rho_1(g)\|_{\rho_1, k}^{n_0} &\leq \max_{\alpha \in \Pi} \max_{w \in W} |\alpha(w(g))|_k^M \leq \max_{w \in W} |\chi_2(w(g))|_k^{Mn_0} \\ &\leq \|\rho_2(g)\|_{\rho_2, k}^{Mn_0}, \end{aligned}$$

where $M = \max_{i=1,2, \alpha \in \Pi} n_\alpha^{(i)}$. Now the Cartan decomposition implies that the above inequality holds for every $g \in \mathbb{G}(k)$. It follows that $h_{\rho_1} \leq Mh_{\rho_2}$. Finally, if we considered instead the norm built from the basis (Y_1, \dots, Y_D) of V_i over \mathbb{Z} defined above, then it would differ from $\|\cdot\|_{\rho_1, k}$ only at infinite places by a fixed multiplicative constant, say C_i . Let h_i be the associated height. Then $|h_{\rho_i} - h_i| \leq C_i$. Therefore $h_1 \leq Mh_2 + C_1 + MC_2$. Together with Remark 2.12 this ends the proof of Proposition 3.4. \square

We can now conclude this section of preliminary reductions by proving:

PROPOSITION 4.11. *In Theorem 3.1, we may assume that $F = \{\text{Id}, a, b\}$ is a subset of $\mathbb{G}(\overline{\mathbb{Q}})$, where \mathbb{G} is a Zariski-connected absolutely simple algebraic group of adjoint type defined over \mathbb{Q} , viewed via the adjoint representation as an algebraic subgroup of $\text{SL}(\mathfrak{g})$, where \mathfrak{g} is the Lie algebra of \mathbb{G} .*

Proof. According to Proposition 4.6, when proving Theorem 3.1, we may assume that F generates a Zariski-dense subgroup of a semisimple algebraic group \mathbb{G} acting irreducibly on $\overline{\mathbb{Q}}^d$. By Proposition 3.4, the normalized heights of this representation of \mathbb{G} and of the adjoint representation of \mathbb{G} are comparable. Hence proving the gap for the first amounts to proving the gap for the second. We may thus assume that $\mathbb{G} = \text{Ad}(\mathbb{G})$ is acting via the adjoint representation on its Lie algebra \mathfrak{g} . It remains to verify that we can reduce to a simple factor of \mathbb{G} . Recall that \mathbb{G} is the direct product of its simple factors. As the representation space \mathfrak{g} splits into the \mathbb{G} -invariant subspaces corresponding to the simple ideals $(\mathfrak{g}_i)_i$ of \mathfrak{g} , and as $h(\text{Ad}(F)) \geq h(\text{Ad}(F)|_{\mathfrak{g}_i})$ for each i , it is enough to prove the theorem for one of the simple factors. Finally by Proposition 1.8, we may assume that F has three elements $\{\text{Id}, a, b\}$. \square

4.4. Geometric interpretation and displacement on symmetric spaces and Bruhat-Tits buildings. In this final section of preliminary reductions, we give a geometric interpretation of the minimal norm $E_v(F)$ and prove Lemma 4.15, which will be key in the proof of the main theorem. We keep the notation of the previous section. Here again \mathbb{G} is a Chevalley group and k is a local

field. We set $\mathcal{BT}(\mathbb{G}, k)$ to be the Bruhat-Tits building (resp. the symmetric space if k is Archimedean) associated to $\mathbb{G}(k)$ as defined in [16]. We fix V, ρ_V a finite dimensional linear representation of \mathbb{G} which is nontrivial on each factor of \mathbb{G} as in Section 4.3 above. We let x_0 be the base point of $\mathcal{BT}(\mathrm{SL}_V, k)$ corresponding to the stabilizer of the norm $\|\cdot\|_{\rho_V, k}$ defined in Section 4.3. The maximal compact subgroup K_0 of $\mathbb{G}(k)$ defined in Section 4.3 coincides with the stabilizer of $\|\cdot\|_{\rho_V, k}$ inside $\mathbb{G}(k)$.

Let ℓ be a finite extension of k . On $\mathcal{BT}(\mathbb{G}, \ell)$ we define the distance d to be the standard left invariant distance on $\mathcal{BT}(\mathbb{G}, \ell)$ with the following normalization: if $a \in A$, then $d(a \cdot x_0, x_0) = \sqrt{\sum_{i=1}^d (\log |a_i|_k)^2}$, where \log is the logarithm in base $|\pi_\ell^{-1}|_k$, with π_ℓ a uniformizer for \mathcal{O}_ℓ when k is non-Archimedean, and the standard logarithm if k is Archimedean. With this normalization, the distance between adjacent vertices on $\mathcal{BT}(\mathbb{G}, \ell)$ is of order 1 and independent of ℓ (when k is non-Archimedean).

Proposition 4.12 below, which was communicated to us by P. E. Caprace [17], shows that the symmetric space or building $\mathcal{BT}(\mathbb{G}, k) \simeq \mathbb{G}(k)/K_0$ embeds isometrically in $\mathcal{BT}(\mathrm{SL}_V, k)$ as a closed and convex subspace via the orbit map $\mathbb{G}(k)/K_0 \rightarrow \mathcal{BT}(\mathrm{SL}_V, k)$, $gK_0 \mapsto g$. The short proof given below makes use of the general theory of CAT(0) spaces (examples of which are the symmetric spaces and buildings $\mathcal{BT}(\mathrm{SL}_V, k)$ considered here). We refer the reader to the book by Bridson and Haefliger [15] for background on CAT(0) spaces. In particular, the notion of a semisimple isometry of a CAT(0) space is defined in [15, II.6.].

PROPOSITION 4.12. *As above let k be a local field and \mathbb{G} a semisimple k -split linear algebraic group, with Cartan decomposition $\mathbb{G}(k) = K_0 T(k) K_0$. Assume that $\mathbb{G}(k)$ acts properly by isometries on a complete CAT(0) space X in such a way that semisimple elements of $\mathbb{G}(k)$ act by semisimple isometries. Assume that K_0 fixes a point p in X which belongs to a flat P stabilized by $T(k)$. Then the map $gK_0 \mapsto g \cdot p$ induces (up to renormalizing the metric on X) a $\mathbb{G}(k)$ -equivariant isometric embedding f from $\mathcal{BT}(\mathbb{G}, k)$ to X .*

Proof. Let $G = \mathbb{G}(k)$, $T = T(k)$ and P_0 the T -invariant flat in $\mathcal{BT}(\mathbb{G}, k)$ containing the base point p_0 associated to K_0 . According to the Flat Torus Theorem (see [15, II.7.]), there is a unique minimal T -invariant flat containing p and its dimension is $\dim T = r = rk(\mathbb{G})$. We may thus assume that P is this minimal flat. However, the normalizer $N_G(T)$ permutes the T -invariant flats and $N_G(T)$ is generated by T and by $N_G(T) \cap K_0$. It follows that $N_G(T)$ stabilizes P . Hence $g \cdot p_0 \mapsto g \cdot p$ induces an $N_G(T)$ -equivariant map f between P_0 and P .

Note first that it is enough to show that f is a homothety from P_0 to P . Indeed up to renormalizing the metric in X , we may then assume that f is an

isometry from P_0 to P , i.e., $d(a \cdot p, p) = d(a \cdot p_0, p_0)$. But then for any $g, h \in G$, $d(f(g \cdot p_0), f(h \cdot p_0)) = d(h^{-1}g \cdot p, p) = d(a \cdot p, p) = d(g \cdot p_0, h \cdot p_0)$ if $h^{-1}g = k_1 a k_2$ is a Cartan decomposition of $h^{-1}g$.

The fact that $f : P_0 \rightarrow P$ is a homothety follows from the rigidity of Euclidean Coxeter group actions. Indeed $N_G(T)$ contains the affine Weyl group as a co-compact subgroup which acts co-compactly by isometries on both P_0 and P . But any such action is isometric to the standard Coxeter representation (cf. [9]). \square

Remark 4.13. This proposition is a special case of a theorem of Landvogt about functoriality properties of Bruhat-Tits buildings (see [27]) in the non-Archimedean case and a theorem of Karpelevich and Mostow (see [29]) in the form given by Eberlein in [20, 2.6.] in the Archimedean case.

The relation between the operator norm on $\mathrm{SL}(V_k)$ and the displacement on $\mathcal{BT}(\mathrm{SL}_V, k)$ is given by the following well-known:

LEMMA 4.14. *For any $f, g \in \mathrm{SL}(V_k)$ and $x = g^{-1} \cdot x_0 \in \mathcal{BT}(\mathrm{SL}_V, k)$, letting \log be the logarithm in base $|\pi_k^{-1}|_k$, we have*

$$\log \|gfg^{-1}\|_{\rho_{V,k}} \leq d(f \cdot x, x) \leq \sqrt{\dim V} \cdot \log \|gfg^{-1}\|_{\rho_{V,k}}.$$

Proof. Since $d(\cdot, \cdot)$ is left invariant, we may assume that $g = 1$. Then we may write $f = k_1 a k_2$ the Cartan decomposition for f . Since the norm is fixed by K_0 we can assume that $f = a$. Then the estimate is obvious from the normalization we chose for $d(\cdot, \cdot)$ above. \square

A consequence of this lemma is that the logarithm of the minimal norm of a finite set F is comparable to the minimal displacement of F on $\mathcal{BT}(\mathrm{SL}_V, k)$. As in [14], 5.4.1., we will use a projection argument and the fact that $\mathcal{BT}(\mathrm{SL}_V, k)$ is a CAT(0) space in order to show that the minimal displacement of F is attained on $\mathcal{BT}(\mathbb{G}, k)$. More precisely:

LEMMA 4.15. *For every finite set $F \in \mathbb{G}(k)$, we have*

$$E_k(\rho_V(F)) \leq \inf_{g \in \mathbb{G}(\bar{k})} \|\rho_V(gFg^{-1})\|_{\rho_{V,k}} \leq E_k(\rho_V(F))^{\sqrt{\dim V}}.$$

Proof. The left side of the inequalities is obvious from the definition of $E_k(\rho_V(F))$. For any $\varepsilon > 0$, one can find a finite extension ℓ of k such that $\inf_{g \in \mathbb{G}(\bar{\mathbb{Q}}_v)} \|\rho_V(gFg^{-1})\|_{\rho_{V,k}} \leq \inf_{g \in \mathbb{G}(\ell)} \|\rho_V(gFg^{-1})\|_{\rho_{V,k}} + \varepsilon$. By Lemma 4.14,

$$\begin{aligned} (7) \quad \inf_{g \in \mathbb{G}(\ell)} \log \|\rho_V(gFg^{-1})\|_{\rho_{V,k}} &\leq \inf_{g \in \mathbb{G}(\ell)} \max_{f \in F} d(fgx_0, gx_0) \\ &\leq \inf_{x \in \mathcal{BT}(\mathbb{G}, \ell)} \max_{f \in F} d(fx, x) + c, \end{aligned}$$

where the log is in base $|\pi_\ell^{-1}|_k$ and c is the maximal distance from any point in $\mathcal{BT}(\mathbb{G}, \ell)$ to the nearest point in the orbit $\mathbb{G}(\ell) \cdot x_0$. Note that this constant c is independent of the choice of ℓ . Since $\mathcal{BT}(\mathrm{SL}_V, \ell)$ is a CAT(0) metric space and $\mathcal{BT}(\mathbb{G}, \ell)$ a closed convex subset, for every $x \in \mathcal{BT}(\mathrm{SL}_V, \ell)$, one can define the projection $p(x)$ of x on $\mathcal{BT}(\mathbb{G}, \ell)$ to be the (unique) point that realizes the distance from x to $\mathcal{BT}(\mathbb{G}, \ell)$. The projection map is 1-Lipschitz; hence $d(fx, x) \geq d(fp(x), p(x))$ for any $x \in \mathcal{BT}(\mathrm{SL}_V, \ell)$. Therefore,

$$(8) \quad \inf_{x \in \mathcal{BT}(\mathbb{G}, \ell)} \max_{f \in F} d(fx, x) = \inf_{x \in \mathcal{BT}(\mathrm{SL}_V, \ell)} \max_{f \in F} d(fx, x).$$

Combining (7) with (8) and Lemma 4.14, we get

$$\inf_{g \in \mathbb{G}(\overline{k})} \left\| \rho_V(gFg^{-1}) \right\|_{\rho_V, k} \leq (|\pi_\ell^{-1}|_k)^c \inf_{g \in \mathrm{SL}_V(\ell)} \left\| g\rho_V(F)g^{-1} \right\|_{\rho_V, k}^{\sqrt{\dim V}} + \varepsilon.$$

But ℓ can be taken arbitrarily large, so that $|\pi_\ell^{-1}|_k$ can be taken arbitrarily close to 1, and since c was independent of ℓ and ε was arbitrary, we finally get the right-hand side of the desired inequality. \square

5. Local estimates on Chevalley groups

In this section, we work locally in a fixed local field, and prove several crucial estimates relating the minimal norm $E_k(F)$ and the matrix coefficients of the elements of F in the adjoint representation. In the next section, we will gather this local information at each place and put it together to obtain global bounds.

5.1. Notation. Recall our notation. The group \mathbb{G} is an absolutely simple algebraic group of adjoint type defined over $\overline{\mathbb{Q}}$, viewed via the adjoint representation as an algebraic subgroup of $\mathrm{GL}(\mathfrak{g})$, where \mathfrak{g} is the Lie algebra of \mathbb{G} . We let L be a number field over which \mathbb{G} splits. The set $F = \{\mathrm{Id}, a, b\}$ consists of the identity and two semisimple regular elements of $\mathbb{G}(\overline{\mathbb{Q}})$ which generate a Zariski-dense subgroup of \mathbb{G} .

Let T be the unique maximal torus of \mathbb{G} containing a . Let $\Phi = \Phi(\mathbb{G}, T)$ be the set of roots of \mathbb{G} with respect to T . Let r be the absolute rank of \mathbb{G} . Let us also choose a Borel subgroup B of \mathbb{G} containing T , thus defining the set of positive roots Φ^+ and a base Π for Φ . For $\alpha \in \Phi$, let \mathfrak{g}_α be the root subspace corresponding to α and $\mathfrak{t} = \mathfrak{g}_0$ be the Lie algebra of T , so that we have the direct sum decomposition

$$(9) \quad \mathfrak{g} = \mathfrak{t} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha.$$

Let $(\alpha_1, \dots, \alpha_r)$ be an enumeration of the base associated to the choice of B . The chosen enumeration of the elements of the base induces a total order on the set of roots, namely two roots $\alpha = \sum n_i \alpha_i$ and $\beta = \sum m_i \alpha_i$

satisfy $\alpha \geq \beta$ if and only if $(n_1, \dots, n_r) \geq (m_1, \dots, m_r)$ for the canonical lexicographical order on r -tuples. We may label the roots in decreasing order, so that $\alpha_1 > \dots > \alpha_{|\Phi^+|} > 0 > \alpha_{|\Phi^+|+r+1} > \dots > \alpha_{|\Phi^+|+r}$ is the full list of all roots. Note that $d = \dim \mathfrak{g} = |\Phi| + r$ and that $\alpha_{|\Phi^+|+r+i} = -\alpha_{|\Phi^+|+1-i}$ for $1 \leq i \leq |\Phi^+|$. Also set $\alpha_0 = 0$ and $\alpha_i = 0$ if $i \in I_r = [|\Phi^+| + 1, |\Phi^+| + r]$. Finally, for any root α , let i_α be the index such that $\alpha_{i_\alpha} = \alpha$.

For every $\alpha \in \Phi^+ \cup \{0\}$, let \mathfrak{u}_α be the subspace of \mathfrak{g} generated by the \mathfrak{g}_β 's for all roots $\beta > \alpha$.

LEMMA 5.1. *For each $\alpha \in \Phi^+$, \mathfrak{u}_α is an ideal in $\mathfrak{b} = \mathfrak{t} \oplus \bigoplus_{\alpha \in \Phi^+} \mathfrak{g}_\alpha$. Moreover the sequence of \mathfrak{u}_α 's for $\alpha \in \Phi^+$ is a decreasing (with α) sequence of nontrivial ideals in \mathfrak{b} starting with $\mathfrak{u}_0 = \bigoplus_{\alpha \in \Phi^+} \mathfrak{g}_\alpha$, each one being of codimension 1 inside the previous one.*

Proof. We have $\mathfrak{u}_\alpha = \bigoplus_{\beta > \alpha} \mathfrak{g}_\beta$. Moreover $[\mathfrak{g}_\gamma, \mathfrak{g}_\beta] \leq \mathfrak{g}_{\gamma+\beta}$ and $\gamma + \beta > \alpha$ for any $\gamma \in \Phi^+ \cup \{0\}$, and so clearly $[\mathfrak{b}, \mathfrak{u}_\alpha] \leq \mathfrak{u}_\alpha$. The second assertion follows from the fact that each \mathfrak{g}_α , $\alpha \in \Phi$, has dimension 1. \square

We also denote by U_α the unipotent algebraic subgroup of \mathbb{G} whose Lie algebra is \mathfrak{u}_α , and by U_0 the maximal unipotent subgroup, whose Lie algebra is \mathfrak{u}_0 . Furthermore, for each $\alpha \in \Phi$, we denote by $e_\alpha : \mathbb{G}_a \rightarrow \mathbb{G}$ the morphism of algebraic groups corresponding to $X_\alpha \in \mathfrak{g}_\alpha$, i.e., $e_\alpha(t) = \exp(tX_\alpha)$. Recall that $U_\alpha = \prod_{\beta > \alpha} e_\beta(\mathbb{G}_a)$, so any element in U_α can be written as a product of $e_\beta(t_\beta)$'s for $\beta > \alpha$.

Recall that since \mathfrak{g} is a simple Lie algebra, it has a Chevalley basis (canonical up to automorphisms of \mathfrak{g}) $\{H_\alpha, \alpha \in \Pi\} \cup \{X_\alpha, \alpha \in \Phi\}$ with $H_\alpha \in \mathfrak{t}$ and $X_\alpha \in \mathfrak{g}_\alpha$. Let $(\omega_\alpha)_{\alpha \in \Pi}$ be the basis of \mathfrak{t} which is dual to Π . Equivalently $\beta(\omega_\alpha) = \delta_{\alpha\beta}$. Then $\{\omega_\alpha, \alpha \in \Pi\} \cup \{X_\alpha, \alpha \in \Phi\}$ is also a basis of \mathfrak{g} and defines a \mathbb{Z} -structure $\mathfrak{g}_{\mathbb{Z}}$ on \mathfrak{g} with $[\mathfrak{g}_{\mathbb{Z}}, \mathfrak{g}_{\mathbb{Z}}] \subset \mathfrak{g}_{\mathbb{Z}}$ (see [35]). Hence for any field k , we can define $\mathfrak{g}_k = \mathfrak{g}_{\mathbb{Z}} \otimes_{\mathbb{Z}} k$. If K is a number field and v a place of K with corresponding embedding $\sigma_v : K \rightarrow K_v$ where K_v is the associated completion of K , then we will use the notation \mathfrak{g}_v to mean \mathfrak{g}_{K_v} .

Since the definition of $e(F)$ does not depend on the choice of the basis of \mathfrak{g} used to define the standard norm appearing in the quantities $E_v(F)$, we may as well fix the basis of \mathfrak{g} to be the basis $\{\omega_\alpha, \alpha \in \Pi\} \cup \{X_\alpha, \alpha \in \Phi\}$, which we denote (Y_1, \dots, Y_d) with $Y_i = X_{\alpha_i} \in \mathfrak{g}_{\alpha_i}$ if $i \notin I_r = [|\Phi^+| + 1, |\Phi^+| + r]$ and $Y_i \in \{\omega_\alpha, \alpha \in \Pi\}$ if $i \in I_r$.

Let $B(X, Y)$ be the Killing form on \mathfrak{g} . We have $B(Y_i, Y_j) \in \mathbb{Z}$ for all i, j . The Chevalley involution is the linear map $\tau : \mathfrak{g} \rightarrow \mathfrak{g}$ by $Y_i^\tau = -Y_i$ for $i \in I_r$ and $X_\alpha^\tau = -X_{-\alpha}$ for each $\alpha \in \Phi$. Then τ is an automorphism of \mathfrak{g} which preserves $\mathfrak{g}_{\mathbb{Z}}$. We set $\phi(X, Y) = -B(X^\tau, Y)$.

We now describe how to choose the norm $\|\cdot\|_v$ on \mathfrak{g}_v . First consider the case when v is Archimedean, i.e., $\overline{\mathbb{Q}}_v = \mathbb{C}$. We set $\langle X, Y \rangle_v = \phi(X, \bar{Y})$, and

thus get a positive definite scalar product on \mathfrak{g}_v and a norm $\|\cdot\|_v$ on \mathfrak{g}_v . Let $\mathbf{K}_v = \{g \in \mathbb{G}(\mathbb{C}), g^\tau = g\}$, where we denoted again by τ the automorphism of $\mathbb{G}(\mathbb{C})$ induced by the Chevalley involution τ . Then \mathbf{K}_v is a maximal compact subgroup of $\mathbb{G}(\mathbb{C})$ and this group coincides with the stabilizer of $\langle \cdot, \cdot \rangle_v$ in $\mathbb{G}(\mathbb{C})$, which in turn coincides with $\{g \in \mathbb{G}(\mathbb{C}), \|\text{Ad}(g)\|_v = 1\}$ where the norm is the operator norm associated to $\langle \cdot, \cdot \rangle_v$. Note that (Y_1, \dots, Y_d) however is not orthogonal with respect to $\langle \cdot, \cdot \rangle_v$ but the decomposition (9) is orthogonal. Finally observe that according to the Iwasawa decomposition we may write $\mathbb{G}(\mathbb{C}) = \mathbf{K}_v U_0(\mathbb{C}) T(\mathbb{C})$.

Suppose now that v is non-Archimedean. We let $\|\cdot\|_v$ be the norm induced on \mathfrak{g}_v by the basis (Y_1, \dots, Y_d) , i.e., $\|\sum y_i Y_i\|_v = \max_{1 \leq i \leq d} |y_i|_v$. Then we set \mathbf{K}_v to be the stabilizer in $\mathbb{G}(\overline{\mathbb{Q}}_v)$ of $\mathfrak{g}_{\mathcal{O}_v} = \mathfrak{g}_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathcal{O}_v$, where \mathcal{O}_v is the ring of integers in $\overline{\mathbb{Q}}_v$. In this situation, the Iwasawa decomposition (see [24]) reads $\mathbb{G}(\overline{\mathbb{Q}}_v) = \mathbf{K}_v U_0(\overline{\mathbb{Q}}_v) T(\overline{\mathbb{Q}}_v)$. Recall (see [35, §1, Lemma 6]) that for any $n \in \mathbb{N}$ and any $\alpha \in \Phi$, $\frac{\text{ad}(X_\alpha)^n}{n!}$ fixes $\mathfrak{g}_{\mathbb{Z}}$. Hence $\left\| \frac{\text{ad}(X_\alpha)^n}{n!} \right\|_v \leq 1$.

Let $c_v = \sup_{\alpha \in \Phi} \frac{\|\text{ad}(X_\alpha)\|_v}{\|X_\alpha\|_v}$ if v is Archimedean and set $c_v = 0$ if v is non-Archimedean. Then, for any place v and $x \in \overline{\mathbb{Q}}_v$, the following holds:

$$(10) \quad \|\text{Ad}(e_\alpha(x))\|_v = \left\| 1 + \text{ad}(xX_\alpha) + \frac{\text{ad}(xX_\alpha)^2}{2!} + \dots + \frac{\text{ad}(xX_\alpha)^d}{d!} \right\|_v \\ \leq e^{c_v} \cdot \max\{1, \|xX_\alpha\|_v\}^d$$

for every $\alpha \in \Phi$, where $d = \dim \mathfrak{g}$.

Finally we observe that we have:

LEMMA 5.2. *Suppose v is non-Archimedean. Then, for each root $\alpha \in \Phi$, the norm $|\alpha|_v := \sup_{Y \in \mathfrak{t}_v \setminus \{0\}} \frac{|\alpha(Y)|_v}{\|Y\|_v}$ satisfies $|\alpha|_v = 1$.*

Proof. First, note that it obviously holds when $\alpha \in \Pi$, because $\alpha(\omega_\beta) = \delta_{\alpha\beta}$. As every $\alpha \in \Phi$ is a linear combination with integer coefficients of elements from Π , we must have $|\alpha|_v \leq 1$. To show the opposite inequality, observe that $\gcd(\alpha(\omega_\beta), \beta \in \Pi) = 1$. Indeed, suppose there were a prime number p such that p divides $\gcd(\alpha(\omega_\beta), \beta \in \Pi)$. Then $\alpha = p\alpha_0$ with $\alpha_0 = \sum_{i=1}^r n_i \alpha_i$ for some $n_i \in \mathbb{Z}$ and $\Pi = \{\alpha_1, \dots, \alpha_r\}$. But since Φ is reduced, α belongs to some base of the root system say $\alpha = \alpha'_1, \dots, \alpha'_r$ ([9, VI.1.5]). Since each α_i is a linear combination with integer coefficients of some α'_i 's, we get that $\alpha_0 \in \mathbb{Z}\alpha$, a contradiction. \square

Note that when v is Archimedean, then $|\alpha|_v$ is independent of v (it is the norm of α with respect to the canonical scalar product induced on the real vector space spanned by the root system). We denote it by $|\alpha|_\infty$.

5.2. *Some local estimates.* We work locally, fixing the place v . The aim of this subsection is to record two estimates, namely Propositions (5.5) and (5.6) below.

Let now $(e_i)_{1 \leq i \leq d}$ be an orthonormal basis for $\mathfrak{g}_{\mathbb{C}}$ such that for each $1 \leq i \leq d$, $e_i \in \mathfrak{g}_{\alpha_i}$. Note that if $b \in \text{Ad}(B(\mathbb{C}))$, then the matrix of b is upper-triangular in the basis $(e_i)_i$.

LEMMA 5.3. *Let V be a complex vector space of dimension n endowed with a hermitian scalar product $\langle \cdot, \cdot \rangle$. Let $(e_i)_{1 \leq i \leq n}$ be an orthonormal basis of V and assume that $b \in \text{SL}(V)$ has an upper-triangular matrix in this basis. Then*

$$\sum_{i < j} |\langle be_i, e_j \rangle|^2 \leq n \cdot (\|b\|^2 - 1).$$

Proof. Let $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ be the eigenvalues of b^*b . According to Cartan's *KAK* decomposition, we have $\|b\|^2 = \lambda_1$. We have

$$\text{tr}(b^*b) = \sum \lambda_i \leq n \cdot \lambda_1 = n \cdot \|b\|^2.$$

On the other hand,

$$\text{tr}(b^*b) = \sum_{i,j} |\langle be_i, e_j \rangle|^2 = \sum_{i < j} |\langle be_i, e_j \rangle|^2 + \sum_{1 \leq i \leq n} |\mu_i|^2,$$

where μ_1, \dots, μ_n are the eigenvalues of b . But $\frac{1}{n} \sum_{1 \leq i \leq n} |\mu_i|^2 \geq (\prod |\mu_i|)^{\frac{2}{n}} = 1$ since $\det(b) = 1$. Hence

$$n \cdot \|b\|^2 \geq \text{tr}(b^*b) \geq \sum_{i < j} |\langle be_i, e_j \rangle|^2 + n. \quad \square$$

LEMMA 5.4. *Let v be any place. Let $\alpha \in \Phi^+$, $a \in T(\overline{\mathbb{Q}}_v)$ regular, $v_\alpha \in U_\alpha(\overline{\mathbb{Q}}_v)$ and $n_\alpha = e_\alpha(x)$ for some $x \in \overline{\mathbb{Q}}_v$ and let $b = \text{Ad}(n_\alpha a v_\alpha n_\alpha^{-1})$. Then if v is Archimedean,*

$$(11) \quad \|xX_\alpha\|_v \leq \frac{\sqrt{d \cdot (\|b\|_v^2 - 1)}}{|1 - \alpha(a)|_v |\alpha|_v}$$

while if v is non-Archimedean,

$$(12) \quad \|xX_\alpha\|_v \leq \frac{\|b\|_v}{|1 - \alpha(a)|_v |\alpha|_v},$$

where $|\alpha|_v$ is the norm of α viewed as linear form on \mathfrak{t}_v as in Lemma 5.2.

Proof. First observe that if $m \in U_\alpha$ and $Y \in \mathfrak{t}_v$, then $\text{Ad}(m)Y \in Y + \mathfrak{u}_\alpha$, while if $m = e_\alpha(x)$ for some x , then $\text{Ad}(m)Y = Y + x[X_\alpha, Y] = Y - \alpha(Y)xX_\alpha$. Let $Y \in \mathfrak{t}_v$ be arbitrary. We have $n_\alpha a v_\alpha n_\alpha^{-1} = a a^{-1} n_\alpha a n_\alpha^{-1} n_\alpha v_\alpha n_\alpha^{-1} = a \cdot e_\alpha((\alpha(a^{-1}) - 1)x) \cdot n''$ where $n'' \in U_\alpha$. We then compute:

$$(13) \quad bY \in Y + x(1 - \alpha(a))\alpha(Y)X_\alpha + \mathfrak{u}_\alpha.$$

Suppose first that v is Archimedean:

$$\langle bY, X_\alpha \rangle_v = x(1 - \alpha(a))\alpha(Y) \|X_\alpha\|_v^2.$$

On the other hand, $Y = \sum y_i e_i$ for some $y_i \in \overline{\mathbb{Q}}_v$ all zero except if $i \in I_r = [|\Phi^+| + 1, |\Phi^+| + r]$ (recall that we defined the vectors e_i 's in Lemma 5.3 to be any orthonormal basis for $\mathfrak{g}_{\mathbb{C}}$ such that for each $1 \leq i \leq d$, $e_i \in \mathfrak{g}_{\alpha_i}$). Using Cauchy-Schwarz, we get:

$$|\langle bY, X_\alpha \rangle_v| \leq \|X_\alpha\|_v \|Y\|_v \sqrt{\sum_{i \in I_r} |\langle be_i, e_{i_\alpha} \rangle_v|^2}.$$

But b is upper-triangular in the basis $(e_i)_i$ because $n_\alpha a v_\alpha n_\alpha^{-1}$ belongs to the Borel subgroup $B(\overline{\mathbb{Q}}_v)$. We are in a position to apply Lemma 5.3, which yields

$$|(1 - \alpha(a))\alpha(Y)|_v \cdot \|xX_\alpha\|_v \cdot \|X_\alpha\|_v \leq \|X_\alpha\|_v \cdot \|Y\|_v \cdot \sqrt{d \cdot (\|b\|_v^2 - 1)}.$$

As this is true for all $Y \in \mathfrak{t}$, we indeed obtain (11).

Now assume v is non-Archimedean. Then (13) shows that

$$\|x(1 - \alpha(a))\alpha(Y)X_\alpha\|_v \leq \|b\|_v \|Y\|_v,$$

which is what we wanted. \square

PROPOSITION 5.5. *There are explicitly computable positive constants $(C_i)_{1 \leq i \leq 3}$ depending only on $d = \dim \mathfrak{g}$ and $p = |\Phi^+|$ such that for any $a \in T(\overline{\mathbb{Q}}_v)$ regular and $u \in U_0(\overline{\mathbb{Q}}_v)$, we have*

$$(14) \quad \|\mathrm{Ad}(u)\|_v \leq C_3 \cdot \left\| \mathrm{Ad}(uau^{-1}) \right\|_v^{C_1} \cdot \left(\prod_{i=1}^p \max\{1, L_i\} \right)^{C_2},$$

where $L_i = (|1 - \alpha_i(a)|_v \cdot |\alpha_i|_v)^{-1}$. Moreover, if v is non-Archimedean, then (14) holds with $C_3 = 1$.

Proof. Recall that we may write $u = e_{\alpha_p}(x_p) \cdots e_{\alpha_1}(x_1)$, where $p = |\Phi^+|$ and $x_i \in \overline{\mathbb{Q}}_v$ for each i . We want to apply Lemma 5.4 recursively starting with $\alpha = \alpha_p$ and going up to α_1 . For each $\alpha \in \Phi^+$ let $u_\alpha = e_{\alpha_{i_\alpha-1}}(x_{i_\alpha-1}) \cdots e_{\alpha_1}(x_1)$ and $n_\alpha = e_\alpha(x_\alpha)$. For each $i \in [1, p]$ we have $u_{\alpha_{i+1}} a u_{\alpha_{i+1}}^{-1} = n_\alpha u_\alpha a u_\alpha^{-1} n_\alpha^{-1} = n_\alpha a v_\alpha n_\alpha^{-1}$, where $\alpha = \alpha_i$ $v_\alpha = a^{-1} u_\alpha a u_\alpha^{-1} \in U_\alpha$.

We set $b_{p+1} = \mathrm{Ad}(uau^{-1})$ and $b_i = \mathrm{Ad}(u_{\alpha_i} a u_{\alpha_i}^{-1})$. Lemma 5.4 gives for each $i \in [1, p]$,

$$\|x_{\alpha_i} X_{\alpha_i}\|_v \leq f_v \cdot L_i \cdot \|b_{i+1}\|_v,$$

where $f_v = \sqrt{d}$ if v is Archimedean, and $f_v = 1$ otherwise. Since $b_{i+1} = \mathrm{Ad}(n_{\alpha_i}) b_i \mathrm{Ad}(n_{\alpha_i}^{-1})$, we have

$$\|b_i\|_v \leq \|b_{i+1}\|_v \cdot e^{2c_v} \cdot \max\{1, \|x_{\alpha_i} X_{\alpha_i}\|_v\}^{2d},$$

where we have used (10). Hence combining the last two lines, we get

$$(15) \quad \|b_i\|_v \leq \mu_i \cdot \|b_{i+1}\|_v^{2d+1},$$

where $\mu_i = e^{2c_v} f_v^{2d} \max\{1, L_i\}^{2d}$.

On the other hand, $\|\mathrm{Ad}(u)\|_v \leq \prod_{\alpha \in \Phi^+} \|\mathrm{Ad}(e_\alpha(x_\alpha))\|_v$, and using (10) again we obtain

$$\begin{aligned} \|\mathrm{Ad}(u)\|_v &\leq e^{pc_v} \cdot \left(\prod_{\alpha \in \Phi^+} \max\{1, \|x_\alpha X_\alpha\|_v\} \right)^d \\ &\leq e^{pc_v} \cdot f_v^{dp} \cdot \left(\prod_{i=1}^p \max\{1, L_i\} \right)^d \cdot \left(\prod_{i=2}^{p+1} \|b_i\|_v \right)^d. \end{aligned}$$

It remains to estimate the last term in the right-hand side. Recursively from (15), we get

$$\prod_{i=2}^{p+1} \|b_i\|_v \leq \|b\|_v^{\sum_{k=0}^{p-1} (2d+1)^k} \cdot \prod_{i=2}^p \prod_{k=i}^p \mu_k^{(2d+1)^{k-i}}.$$

Hence we do indeed obtain a bound of the desired form. \square

The above proposition is useful to bound $\|\mathrm{Ad}(u)\|_v$ when $\|\mathrm{Ad}(uau^{-1})\|_v$ may be large. We now need an estimate (only when v is Archimedean) when this norm is small. Let L_i be defined as in the previous statement.

PROPOSITION 5.6. *Suppose v is Archimedean. Then there are positive constants $(D_i)_{1 \leq i \leq 3}$ depending only on $d = \dim \mathfrak{g}$ and $p = |\Phi^+|$, such that for any $u \in U_0(\overline{\mathbb{Q}}_v)$ and $a \in T(\overline{\mathbb{Q}}_v)$ regular with $\log \|\mathrm{Ad}(uau^{-1})\|_v \leq 1$, we have*

$$\log \|\mathrm{Ad}(u)\|_v \leq D_3 \cdot L_v^{D_2} \cdot \left(\log \|\mathrm{Ad}(uau^{-1})\|_v \right)^{D_1},$$

where $L_v = \prod_{i=1}^p \max\{1, L_i(a)_v\}$.

Proof. In this proof, by a constant we mean a positive number that depends only on d and p . Observe that there exists $\varepsilon_1 > 0$ such that $\sqrt{x^2 - 1} \leq 2\sqrt{\log x}$ as soon as $x \geq 1$ and $\log x \leq \varepsilon_1$. We keep the notation of the proof of the previous proposition. Applying Lemma 5.4, we thus obtain that as soon as $\ell_{i+1} \leq \varepsilon_1$

$$\|x_{\alpha_i} X_{\alpha_i}\|_v \leq 2\sqrt{d} \cdot L_i \cdot \sqrt{\ell_{i+1}},$$

where we set $\ell_i = \log \|b_i\|_v$ for each $i \in [1, p]$, and $\ell = \ell_{p+1} = \log \|\mathrm{Ad}(uau^{-1})\|_v$.

We may choose a smaller ε_1 so that

$$\|\mathrm{Ad}(e_\alpha(x))\|_v \leq 1 + 2c_v \|x X_\alpha\|_v$$

for each $\alpha \in \Phi^+$ as soon as $|x|_v \leq \varepsilon_1$, as we see from (10). Hence if $\sqrt{\ell_{i+1}} \leq \frac{\varepsilon_1}{2\sqrt{d} \cdot L}$, then

$$\|b_i\|_v \leq \|b_{i+1}\|_v \cdot \left(1 + 4\sqrt{d} \cdot L_i \cdot \sqrt{\ell_{i+1}}\right)^2$$

or

$$\begin{aligned} \ell_i &\leq \ell_{i+1} + 8\sqrt{d} \cdot L_i \cdot \sqrt{\ell_{i+1}} \\ &\leq C \cdot L \cdot \sqrt{\ell_{i+1}} \end{aligned}$$

for some constant C . Applying this recursively, we see that, as soon as L is bigger than some constant, if $\ell \leq \frac{\varepsilon_1^{2^{p+1}}}{L^{3^{p+1}}}$ then, for each $i \in [1, p]$, $\ell_i \leq \frac{\varepsilon_1^{2^{p+1}}}{L^{3^{p+1}}}$ and

$$\ell_i \leq C' \cdot L^2 \cdot \ell^{\frac{1}{2^{p+1-i}}}$$

for each $i \in [1, p]$ and some constant C' . On the other hand, $\|\text{Ad}(u)\|_v \leq \prod_{\alpha \in \Phi^+} \|\text{Ad}(e_\alpha(x_\alpha))\|_v \leq \prod_{\alpha \in \Phi^+} e^{c_v \|x_\alpha X_\alpha\|_v}$ and

$$\begin{aligned} \log \|\text{Ad}(u)\|_v &\leq c_v \cdot \sum_{i=1}^p \|x_{\alpha_i} X_{\alpha_i}\|_v \leq C'' \cdot L \cdot \sqrt{\sum_{2 \leq i \leq p+1} \ell_i} \\ &\leq C''' \cdot L^2 \cdot \ell^{\frac{1}{2^{p+1}}}. \end{aligned}$$

On the other hand, the cruder bound obtained in Proposition 5.5 shows that without a condition on ℓ ,

$$\log \|\text{Ad}(u)\|_v \leq \log C_3 + C_1 \cdot \ell + C_2 \cdot \log L;$$

hence

$$\log \|\text{Ad}(u)\|_v \leq C_0 \cdot L$$

for some constant C_0 if $\ell \leq 1$ and L larger than some constant. Take $D_1 = \frac{1}{2^{p+1}}$, $D_2 \geq 1 + \left(\frac{3}{2}\right)^{p+1}$ and $D_3 \geq \max\{\frac{C_0}{\varepsilon_1}, C'''\}$. Then if $\ell \geq \frac{\varepsilon_1^{2^{p+1}}}{L^{3^{p+1}}}$, we have $D_3 \cdot L^{D_2} \cdot \ell^{\frac{1}{2^{p+1}}} \geq D_3 \cdot L \cdot \varepsilon_1 \geq C_0 \cdot L$. Therefore, as soon as $\ell \leq 1$ and L larger than some constant say C_4 , we have

$$\log \|\text{Ad}(u)\|_v \leq D_3 \cdot L^{D_2} \cdot \ell^{D_1}.$$

Hence up to changing D_3 into $D_3 C_4^{D_2}$ if necessary, we obtain the desired result. \square

6. Global bounds on arithmetic heights

In this section we gather together all the local data obtained in the previous section and sum it up to obtain a global bound (see Proposition 6.1 below) on the height of the matrix coefficients of the finite set F .

Recall our set of notation from the last section (see §5.1). \mathbb{G} is a Chevalley group of adjoint type and T a maximal torus. We had fixed a total order on the set of all roots induced by an ordering of the simple roots; that is,

$\Phi = \{\alpha_1, \dots, \alpha_{|\Phi^+|}, \alpha_{|\Phi^+|+r+1}, \dots, \alpha_d\}$, where r is the rank of $\mathfrak{g} = \text{Lie}(\mathbb{G})$ and $I_r = [|\Phi^+|, |\Phi^+| + r]$. The Lie algebra \mathfrak{g} has a basis (Y_1, \dots, Y_d) obtained from a Chevalley basis of \mathfrak{g} , with $Y_i = X_{\alpha_i}$ if $i \notin I_r$ and $Y_i \in \{\omega_\alpha, \alpha \in \Pi\}$ if $i \in I_r$. Also $\mathbb{g}_{\mathbb{Z}}$ denotes the integer lattice generated by the basis (Y_1, \dots, Y_d) . Recall further that for $X, Y \in \mathfrak{g}$ we had set $\phi(X, Y) = -B(X^\tau, Y)$ where B is the Killing form and τ the Chevalley involution. Note that (9) is an orthogonal decomposition for the symmetric bilinear form ϕ .

We will consider the elements $A = \text{Ad}(a)$ and $B = \text{Ad}(b)$ from $F = \{\text{Id}, a, b\} \subset \mathbb{G}(\overline{\mathbb{Q}})$ with $a \in T$ as matrices in the basis (Y_1, \dots, Y_d) . Then A is diagonal and $B = (b_{ij})_{ij} \in \text{SL}_d(\overline{\mathbb{Q}})$. Consider the regular function on \mathbb{G} given by $f(g) = g_{dd}$ in this basis. The root $\alpha(d)$ is the smallest root in the above ordering. It coincides with the opposite of the highest root of Φ in the sense of [9, VI.1.8.]. Observe the following:

- for every $t \in T$, we have $f(tgt^{-1}) = f(g)$;
- for every $t \in T$, $f(t) = \alpha_d(t)$, hence f is not constant;
- $\phi(\text{Ad}(g)Y_d, Y_d) = f(g)\phi(Y_d, Y_d)$;
- for every place v we have $|f(g)|_v \leq \|\text{Ad}(g)\|_v$.

Recall that we consider \mathbb{G} as a subgroup of $\text{SL}(\mathfrak{g})$ and thus define the heights e and \widehat{h} of finite subsets of $\mathbb{G}(\overline{\mathbb{Q}})$ with respect to the adjoint representation. The goal of this section is to prove:

PROPOSITION 6.1. *For every $n \in \mathbb{N}$ and any $\alpha > 0$ there is $\eta > 0$ and $A_1 > 0$ such that if $F = \{\text{Id}, a, b\}$ is a subset of $\mathbb{G}(\overline{\mathbb{Q}})$ with $a \in T(\overline{\mathbb{Q}})$ such that $e(F) < \eta$ and $\deg(\alpha_i(a)) > A_1$ for each positive root α_i , then we have for every $i \in \mathbb{N}$, $1 \leq i \leq n$,*

$$h(f(b^i)) < \alpha,$$

where f is the function defined above.

This proposition is central to the proof of the main theorem of this paper, that is Theorem 3.1. How to proceed from it to a complete proof of Theorem 3.1 will be explained in the next section. For the moment, let us just say that given the assumptions of Theorem 3.1, if $\widehat{h}(F)$ is small, then by escape from subvarieties (see Proposition 4.1) one may find many pairs $\{a, b\}$ in a bounded power of F that satisfy the requirements of the above proposition, indeed one may find such $\{a, b\}$ outside every given subvariety of $\mathbb{G} \times \mathbb{G}$. Applying Zhang's theorem (see [43] and Theorem 7.1 below), we will see however that the height bounds imposed upon the $f(b^i)$'s by the above proposition will force $\{a, b\}$ to belong to a proper algebraic variety of \mathbb{G} , thus contradicting the choice of $\{a, b\}$.

We now begin the proof of the above proposition. It will make use of the local estimates obtained in the previous section as well as Bilu's equidistribution theorem (see below). The proof will occupy the next two subsections.

First, we collect the local estimates and see what bounds they give us. Then we use Bilu's theorem to show that the remainder terms give only a small contribution to the height.

6.1. *Preliminary upper bounds.* Recall that the $(C_i)_{1 \leq i \leq 3}$'s and $(D_i)_{1 \leq i \leq 3}$'s are the constants obtained in Propositions 5.5 and 5.6. For $A \geq 1$ and $x \in \overline{\mathbb{Q}}$ we set

$$(16) \quad h_\infty^A(x) := \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, |x|_v \geq A} n_v \cdot \log^+ |x|_v,$$

where the sum is limited to those $v \in V_\infty$ such that $|x|_v \geq A$. Recall that for $x \in \overline{\mathbb{Q}}$, $h_f(x)$ denotes the finite part of the Weil height, namely the sum over the non Archimedean places. In this section, we prove the following:

PROPOSITION 6.2. *There are positive constants A_0 , C_4 and D_4 , such that for every $\varepsilon \in (0, 1)$ and $A > A_0$, for every $j \in \mathbb{N}$, and for every choice of two regular semisimple elements $a, b \in \mathbb{G}(\overline{\mathbb{Q}})$ with $a \in T(\overline{\mathbb{Q}})$, we have the following bound:*

$$(17) \quad \frac{h(f(b^j))}{j} \leq C_4 \frac{\log A}{\varepsilon} e(\{a, b\}) + D_4 A^{D_2} \varepsilon^{D_1} + C_4 \sum_{1 \leq i \leq p} (h_f(\delta_i^{-1}) + h_\infty^A(\delta_i^{-1})),$$

where $\delta_i = 1 - \alpha_i(a)$ for each positive root $\alpha_i \in \Phi^+$ and $p = |\Phi^+|$.

As before, we set $F = \{a, b\}$. For each place v let $s_v > \log(E_v^{\mathrm{Ad}}(F))$ be some real number. According to Lemma 4.15, there exists $g_v \in \mathbb{G}(\overline{\mathbb{Q}}_v)$ such that $\|\mathrm{Ad}(g_v F g_v^{-1})\|_v \leq e^{s_v}$. Since by the Iwasawa decomposition we have $\mathbb{G}(\overline{\mathbb{Q}}_v) = \mathbf{K}_v U_0(\overline{\mathbb{Q}}_v) T(\overline{\mathbb{Q}}_v)$, and \mathbf{K}_v stabilizes the norm, we may assume that $g_v \in U_0(\overline{\mathbb{Q}}_v) T(\overline{\mathbb{Q}}_v)$, i.e., $g_v = u_v \cdot t_v$. Since t commutes with a we get

$$\begin{aligned} \|\mathrm{Ad}(u_v a u_v^{-1})\|_v &\leq e^{s_v}, \\ \|\mathrm{Ad}(u_v b^{t_v} u_v^{-1})\|_v &\leq e^{s_v}, \end{aligned}$$

where $b^{t_v} = t_v b t_v^{-1}$. Recall that $d = \dim \mathbb{G}$.

According to Proposition 5.5, we have

$$\begin{aligned} (18) \quad \|\mathrm{Ad}(b^{t_v})\|_v &\leq e^{s_v} \cdot \|\mathrm{Ad}(u_v)\|_v^d \\ &\leq e^{s_v} \cdot C_3^d \cdot \|\mathrm{Ad}(u_v a u_v^{-1})\|_v^{dC_1} \cdot \left(\prod_{i=1}^p \max\{1, L_i(a)_v\} \right)^{dC_2} \\ &\leq C_3^d \cdot \left(\prod_{i=1}^p \max\{1, L_i(a)_v\} \right)^{dC_2} \cdot e^{s_v(1+dC_1)} \end{aligned}$$

with $C_3 = 1$ if v is non-Archimedean. Let $L_v = \prod_{i=1}^p \max\{1, L_i(a)_v\}$. We get

$$(19) \quad \log \|\mathrm{Ad}(b^{t_v})\|_v \leq d \log C_3 + dC_2 \cdot \log L_v + (1 + dC_1) \cdot s_v.$$

Now assume v is Archimedean. According to Proposition 5.6 we have constants $D_i > 0$ such that if $s_v \leq 1$, then

$$(20) \quad \log \left\| \text{Ad}(b^{t_v}) \right\|_v \leq s_v + d \log \left\| \text{Ad}(u_v) \right\|_v \leq s_v + dD_3 L_v^{D_2} \cdot s_v^{D_1} \\ \leq D'_4 L_v^{D_2} \cdot s_v^{D_1},$$

where $D'_4 = dD_3 + 1$ and where we have chosen $D_1 \leq 1$ as we may so that $s_v \leq s_v^{D_1}$. Since $|f(b^j)|_v \leq \left\| \text{Ad}(b^{t_v}) \right\|_v^j$ for each $j \in [1, n]$ and v , we have

$$\frac{h(f(b^j))}{j} \leq \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \cdot \log \left\| \text{Ad}(b^{t_v}) \right\|_v.$$

In order to prove Proposition 6.2, we will decompose this sum into four parts. Let $\kappa = \min_i |\alpha_i|_\infty$ (see Lemma 5.2 and the remark following it for the definition of $|\alpha_i|_\infty$). We split the set of places v into four parts: $v \in V_\infty$, $s_v \leq \varepsilon$ and $L_v \geq A/\kappa$ (this gives H_\leq^+), $v \in V_\infty$, $s_v \leq \varepsilon$ and $L_v < A/\kappa$ (this gives H_\leq^-), $v \in V_\infty$ and $s_v > \varepsilon$ (this gives H_\geq) and finally $v \in V_f$ (this gives H_f). So

$$\frac{h(f(b^j))}{j} \leq H_\leq^- + H_\leq^+ + H_\geq + H_f.$$

Making use of the bound (19) for H_\leq^+ , H_\geq and H_f and the bound (20) for H_\leq^- respectively, we obtain the following estimates as soon as A is large enough ($\log A > \log A_0 := 1 + dC_1 + d \log C_3 + \log |\kappa|$, we also set $C_4 = 4dC_2$):

$$H_f \leq (1 + dC_1) \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_f} n_v \cdot s_v + (dC_2) \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_f} n_v \cdot \log L_v, \\ H_\geq \leq \left(\frac{d \log C_3}{\varepsilon} + (1 + dC_1) \right) \cdot \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v > \varepsilon} n_v \cdot s_v \\ + \frac{C_4}{4} \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v > \varepsilon} n_v \cdot \log L_v \\ \leq \frac{C_4 \log A}{\varepsilon} \cdot \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v > \varepsilon} n_v \cdot s_v \\ + \frac{C_4}{4} \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v > \varepsilon, L_v \geq A/\kappa} n_v \cdot \log L_v, \\ H_\leq^+ \leq (2dC_2) \cdot \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v \leq \varepsilon, L_v \geq A/\kappa} n_v \cdot \log L_v, \\ H_\leq^- \leq \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_\infty, s_v \leq \varepsilon, L_v < A/\kappa} n_v \cdot D'_4 L_v^{D_2} \cdot s_v^{D_1} \\ \leq 2 \frac{D'_4}{\kappa^{D_2}} A^{D_2} \varepsilon^{D_1} \leq D_4 A^{D_2} \varepsilon^{D_1}$$

for $D_4 = 2 \frac{D'_4}{\kappa^{D_2}}$, since $n_v \leq 2$ for $v \in V_\infty$.

Note that $e(F) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \cdot s_v$, so the above bounds give

$$(21) \quad \frac{h(f(b^j))}{j} \leq C_4 \frac{\log A}{\varepsilon} e(F) + \frac{C_4}{2} \frac{1}{[K:\mathbb{Q}]} \left(\sum_{v \in V_\infty, L_v \geq A/\kappa} n_v \cdot \log L_v + \sum_{v \in V_f} n_v \cdot \log L_v \right) + D_4 A^{D_2} \varepsilon^{D_1}.$$

On the other hand, $\log L_v \leq \sum_{1 \leq i \leq p} \log^+ L_i(a)_v$ where $L_i(a)_v = |\delta_i^{-1}|_v / |\alpha_i|_v$ and $\delta_i = 1 - \alpha_i(a)$.

Clearly if $L_i(a)_v \geq A/\kappa \geq \kappa^{-2}$, then $|\delta_i^{-1}|_v \geq A$ and $L_i(a)_v \leq |\delta_i^{-1}|_v^2$. We get

$$(22) \quad \sum_{v \in V_\infty, L_v \geq A/\kappa} n_v \cdot \log L_v \leq \sum_{1 \leq i \leq p} \sum_{v \in V_\infty, L_i(a)_v \geq A/\kappa} n_v \cdot \log^+ L_i(a)_v \leq 2 \cdot \sum_{1 \leq i \leq p} \sum_{v \in V_\infty, |\delta_i|_v \leq A^{-1}} n_v \cdot \log^+ |\delta_i^{-1}|_v.$$

Now note that for $v \in V_f$ we have $|\alpha_i|_v = 1$ according to Lemma 5.2. It follows that

$$(23) \quad \sum_{v \in V_f} n_v \cdot \log^+ L_i(a)_v = \sum_{v \in V_f} n_v \cdot \log^+ |\delta_i^{-1}|_v = [K:\mathbb{Q}] \cdot h_f(\delta_i^{-1}).$$

Hence combining (21) with (22), (23) we obtain (17) and this ends the proof of Proposition 6.2.

6.2. Bilu's equidistribution theorem. We are now going to apply Bilu's equidistribution theorem to show that the last term in estimate (17) becomes very small when both A is large and $e(F)$ is small.

THEOREM 6.3 (Bilu's equidistribution of small points [6]). *Suppose that $(\lambda_n)_{n \geq 1}$ is a sequence of algebraic numbers (i.e., in $\overline{\mathbb{Q}}$) such that $h(\lambda_n) \rightarrow 0$ and $\deg(\lambda_n) \rightarrow +\infty$ as $n \rightarrow +\infty$. Let $\mathcal{O}(\lambda_n)$ be the Galois orbit of λ_n . Then we have the following weak-* convergence of probability measures on \mathbb{C} :*

$$(24) \quad \frac{1}{\#\mathcal{O}(\lambda_n)} \sum_{x \in \mathcal{O}(\lambda_n)} \delta_x \xrightarrow{n \rightarrow +\infty} d\theta,$$

where $d\theta$ is the normalized Lebesgue measure on the unit circle $\{z \in \mathbb{C}, |z| = 1\}$.

Let us first draw two consequences of this equidistribution statement.

LEMMA 6.4. *For every $\alpha > 0$ there is $A_1 > 0$, $\eta_1 > 0$ and $\varepsilon_1 > 0$ with the following property. If $\lambda \in \overline{\mathbb{Q}}$ is such that $h(\lambda) \leq \eta_1$ and $\deg(\lambda) > A_1$, then*

$$(25) \quad h_\infty^{\varepsilon_1^{-1}} \left(\frac{1}{1-\lambda} \right) \leq \alpha,$$

where $h_\infty^{\varepsilon_1^{-1}}$ was defined in (16).

Proof. We have

$$h_\infty\left(\frac{1}{1-\lambda}\right) \leq h\left(\frac{1}{1-\lambda}\right) = h(1-\lambda) \leq h_f(\lambda) + h_\infty(1-\lambda) \leq h(\lambda) + h_\infty(1-\lambda).$$

Hence

$$\frac{1}{\deg(\lambda)} \sum_{x \in \mathcal{O}(\lambda)} \log \frac{1}{|1-x|} = h_\infty\left(\frac{1}{1-\lambda}\right) - h_\infty(1-\lambda) \leq h(\lambda)$$

and

$$(26) \quad h_\infty^{\varepsilon_1^{-1}}\left(\frac{1}{1-\lambda}\right) = \frac{1}{\deg(\lambda)} \sum_{|1-x| \leq \varepsilon_1} \log \frac{1}{|1-x|} \leq h(\lambda) + \frac{1}{\deg(\lambda)} \sum_{|1-x| > \varepsilon_1} \log |1-x|.$$

Consider the function $f_{\varepsilon_1}(z) = \mathbf{1}_{|z-1| > \varepsilon_1} \log |1-z|$. It is locally bounded on \mathbb{C} . By Theorem 6.3, for every $\varepsilon_1 > 0$, there must exist $\eta_1 > 0$ and $A_1 > 0$ such that, if $h(\lambda) \leq \eta_1$, and $d = \deg(\lambda) > A_1$, then

$$\left| \frac{1}{\deg(\lambda)} \sum_x f_{\varepsilon_1}(x) - \int_0^1 f_{\varepsilon_1}(e^{2\pi i \theta}) d\theta \right| \leq \frac{\alpha}{3}.$$

On the other hand, we verify that $\int_0^1 \log |1 - e^{2\pi i \theta}| d\theta = 0$. Hence we can choose $\varepsilon_1 > 0$ small enough so that $\left| \int_0^1 f_{\varepsilon_1}(e^{2\pi i \theta}) d\theta \right| \leq \frac{\alpha}{3}$. Combining these inequalities with (26) and choosing $\eta_1 \leq \frac{\alpha}{3}$, we get (25). \square

LEMMA 6.5. *For every $\alpha > 0$ there exists $\eta > 0$ and $A_1 > 0$ such that for any $\lambda \in \overline{\mathbb{Q}}$, if $h(\lambda) \leq \eta$ and $d = \deg(\lambda) > A_1$, then*

$$\left| \frac{1}{\deg(\lambda)} \sum_{v \in V_\infty} n_v \cdot \log |1-\lambda|_v \right| \leq \alpha.$$

Proof. The previous lemma shows that the convergence (24) not only holds for compactly supported functions on \mathbb{C} , but also for functions with logarithmic singularities at 1. In particular it holds for the function $f(z) = \log |1-z|$, which is exactly what we need, since we check easily that $\int_0^1 f(e^{2\pi i \theta}) d\theta = 0$. \square

As a consequence we obtain:

LEMMA 6.6. *For every $\alpha > 0$ there exists $\eta_0 > 0$ and $A_1 > 0$ such that for any $\lambda \in \overline{\mathbb{Q}}$, if $h(\lambda) \leq \eta_0$ and $d = \deg(\lambda) > A_1$, then*

$$h_f\left(\frac{1}{1-\lambda}\right) \leq 2\alpha.$$

Proof. We apply the product formula to $\delta = 1 - \lambda$, which takes the form $h(\delta) = h(\delta^{-1})$; hence

$$h_f(\delta^{-1}) = h_\infty(\delta) - h_\infty(\delta^{-1}) + h_f(\delta).$$

But $h_f(\delta) = h_f(1 - \lambda) \leq h_f(\lambda) \leq \eta_0$ and $h_\infty(\delta) - h_\infty(\delta^{-1}) = \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_\infty} n_v \cdot \log |\delta|_v$, which is bounded by α according to Lemma 6.5. We are done. \square

The outcome of all this is that each of the terms $h_f(\delta_i^{-1}) + h_\infty^A(\delta_i^{-1})$ in (17) becomes small as soon as $e(F)$ (hence $h(\alpha_i(a))$) becomes small and A becomes large.

6.3. Proof of Proposition 6.1. Let $n \in \mathbb{N}$ and $\alpha > 0$ be arbitrary. Let $j \in [1, n]$ an integer and $F = \{a, b\} \subset \mathbb{G}(\overline{\mathbb{Q}})$ with $a \in T(\overline{\mathbb{Q}})$. Then for any $\varepsilon > 0$ and $A > 0$ large enough we obtained the upper bound (17) above. On the other hand we had $h(\alpha_i(a)) \leq e(F)$ for each positive root α_i and $\delta_i = 1 - \alpha_i(a)$. Let ε_1 , A_1 and η_0 be the quantities obtained in the previous section in Lemmas 6.4 and 6.6. Choose A so that $A^{-1} < \varepsilon_1$ and $A \geq A_0$ and consider (17). Assume that for each $i \in \{1, \dots, p\}$ $\deg(\alpha_i(a)) > A_1$. Then Lemmas 6.4 and 6.6 will hold with $\lambda = \alpha_i(a)$ as soon as $e(F) < \eta_0$. Hence for each $i = 1, \dots, p$

$$\left| h_f(\delta_i^{-1}) + h_\infty^A(\delta_i^{-1}) \right| \leq 2\alpha$$

and

$$\frac{h(f(b^j))}{j} \leq \frac{C_4 \log A}{\varepsilon} e(\{a, b\}) + D_4 A^{D_2} \varepsilon^{D_1} + 2p(4dC_2)\alpha.$$

Now choose $\varepsilon > 0$ so that $2D_4 A^{D_2} \varepsilon^{D_1} < \alpha$. Then choose $\eta > 0$ so that $C_4 \frac{\log A}{\varepsilon} \eta < \alpha$ and $\eta < \eta_0$. From (17), we then obtain that if $e(F) < \eta$ and $j \in \mathbb{N}$, then

$$\frac{1}{j} h(f(b^j)) \leq (2 + p(4dC_2))\alpha.$$

Since α was arbitrary we obtain the desired bound.

7. Proof of the statements of Section 3

7.1. Proof of Theorem 3.1. Before beginning the proof of Theorem 3.1, we recall Zhang's theorem on small points of algebraic tori. Let \mathbb{G}_m be the multiplicative group and $n \in \mathbb{N}$. On the $\overline{\mathbb{Q}}$ -points of the torus \mathbb{G}_m^n we define a notion of height in the following natural way. If $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{G}_m^n$ then $h(\mathbf{x}) := h(x_1) + \dots + h(x_n)$ where $h(x_i)$ is the standard logarithmic Weil height we have been using so far.

THEOREM 7.1 (Zhang [43]). *Let V be a proper closed algebraic subvariety of \mathbb{G}_m^n defined over $\overline{\mathbb{Q}}$. Then there is $\varepsilon > 0$ such that the Zariski closure V_ε of the set $\{\mathbf{x} \in V, h(\mathbf{x}) < \varepsilon\}$ consists of a finite union of torsion coset tori, i.e.,*

subsets of the forms ζH , where $\zeta = (\zeta_1, \dots, \zeta_n)$ is a torsion point and H is a subtorus of \mathbb{G}_m^n .

We will need the following lemma, where \mathbb{G} is a semisimple algebraic group over an algebraically closed field, T is a maximal torus together with a choice of simple roots Π , and f is the regular function defined at the beginning of the last section.

LEMMA 7.2. *For every $k \in \mathbb{N}$, the regular functions f_1, \dots, f_k defined on \mathbb{G} by $f_i(g) = f(g^i)$ are multiplicatively independent. Namely, if for each i , n_i and m_i are nonnegative integers and the f_i 's satisfy an equation of the form $\prod_1^k f_i^{n_i} = \prod_1^k f_i^{m_i}$, then $n_i = m_i$ for each i .*

Proof. To prove this lemma it is enough to show that for each i one can find a group element $g \in \mathbb{G}$ such that $f_i(g) = 0$ while all other $f_j(g)$'s are nonzero. Let H be the copy of PGL_2 corresponding to the roots $\alpha = \alpha_d$ and $-\alpha = \alpha_1$ with Lie algebra \mathfrak{h} generated by X_α , $X_{-\alpha}$ and H_α . Clearly it is enough to prove the lemma for the restriction of f to H . Therefore, without loss of generality, we may assume that $\mathbb{G} = \mathrm{PGL}_2$; hence $f(g) = a^2$ if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PGL}_2$. Let for instance $D_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \in \mathrm{PGL}_2$ and $P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Set $g_\lambda = PD_\lambda P^{-1}$. Then compute $f(g_\lambda) = 2\lambda - \lambda^{-1}$ and $f_i(g_\lambda) = f(g_{\lambda^i})$. Hence $f_i(g_\lambda) = 0$ if and only if $2\lambda^{2i} = 1$. These conditions are mutually exclusive for distinct values of i . So we are done. \square

We now conclude this subsection with the proof of Theorem 3.1. According to the reductions made in Section 4 we may assume that $F \subset \mathbb{G}(\overline{\mathbb{Q}})$ where \mathbb{G} is a connected absolutely almost simple algebraic group \mathbb{G} of adjoint type (viewed as embedded in $\mathrm{GL}(\mathfrak{g})$ via the adjoint representation) and that the group $\langle F \rangle$ is Zariski dense in \mathbb{G} . Let T be a maximal torus in \mathbb{G} and Φ be the corresponding set of roots with set of simple roots Π and let $\alpha_1 = -\alpha_d$ be the highest root. The function $f \in \overline{\mathbb{Q}}[\mathbb{G}]$ was defined at the beginning of Section 6 by $f(g) = g_{dd}$ where $\{g_{ij}\}_{1 \leq i, j \leq d}$ is the matrix of $\mathrm{Ad}(g)$ in the Chevalley basis (Y_1, \dots, Y_d) . Let $f_i(g) = f(g^i)$ and let Ω be the Zariski open subset of \mathbb{G} defined by $\{g, f_i(g) \neq 0 \text{ for each } i \leq d+1\}$. Let \mathbf{f} be the regular map $\mathbf{f}(g) := (f_1(g), \dots, f_{d+1}(g)) : \Omega \rightarrow \mathbb{G}_m^{d+1}$. Since $d = \dim \mathbb{G}$, $\mathrm{Im} \mathbf{f}$ is not Zariski dense in \mathbb{G}_m^{d+1} . Let V be its Zariski closure. According to the above theorem of Zhang, there is $\mu > 0$ such that the Zariski closure V_μ of $\{\mathbf{x} = (x_1, \dots, x_{d+1}) \in V \text{ such that } h(\mathbf{x}) < \mu\}$ is a finite union of torsion coset tori. On the other hand, Lemma 7.2 and the Zariski connectedness of \mathbb{G} shows that V cannot be equal to a finite union of torsion coset tori. Hence V_μ is a proper Zariski closed subset of V . Let $Z_\mu = \Omega^c \cup \mathbf{f}^{-1}\{V_\mu\}$. Then Z_μ is a proper Zariski-closed subset of \mathbb{G} . Note that since f is invariant under conjugation by T , Z_μ also is invariant under conjugation by T . Let \widehat{Z}_μ the

Zariski closure of the set $\{(gag^{-1}, gbg^{-1}) \in \mathbb{G}^2 \text{ with } g \in \mathbb{G}, a \in T \text{ and } b \in Z_\mu, \text{ or } a \in Z_\mu \text{ and } b \in T\}$ in $\mathbb{G} \times \mathbb{G}$. It is a proper Zariski closed subset, since $\dim \widehat{Z} \leq 2 \dim \mathbb{G} - 1$. Take $n = d + 1$ and $\alpha = \mu/n$ in Proposition 6.1, which gives us an $A_1 > 0$ and an $\eta > 0$. According to Proposition 4.1 there is a number $c = c(\mathbb{G}, Z_\mu, A_1) > 0$ such that F^c contains two elements a and b which are A_1 -regular semisimple elements, generate a Zariski-dense subgroup of \mathbb{G} and satisfy $(a, b) \notin \widehat{Z}_\mu$. Now let $\varepsilon = \eta/c$ and assume that $e(F) < \varepsilon$. Then $e(\{a, b\}) < \eta$. For some $g \in \mathbb{G}(\overline{\mathbb{Q}})$, $gag^{-1} \in T$, and since $e(\cdot)$ is invariant under conjugation by elements from $\mathbb{G}(\overline{\mathbb{Q}})$, we have $e(\{gag^{-1}, gbg^{-1}\}) < \eta$. We can now apply Proposition 6.1 to see that we must have $h(\mathbf{f}(gbg^{-1})) < \mu$, therefore $gbg^{-1} \in Z_\mu$ and hence $(gag^{-1}, gbg^{-1}) \in \widehat{Z}_\mu$ which gives the desired contradiction. Hence $e(F) > \varepsilon$ and we are done.

7.2. Proof of Proposition 3.3.

Reduction to the adjoint representation. We first reduce to proving the statement of Proposition 3.3 for the adjoint representation and the “Killing height” h_{Kill} . Changing \mathbb{G} into its image in $\mathrm{SL}(V)$ via ρ , we may assume that ρ is nontrivial on each simple factor of \mathbb{G} . Let $\mathrm{Ad}, \mathfrak{g}$ be the adjoint representation of \mathbb{G} and let h_{Kill} be the “Killing height” introduced in Section 4.3. According to Proposition 3.4 and its proof there exists a constant $C_\rho \geq 1$ and a basis of V giving rise to an associated height function h on $\mathrm{End}(V)$, such that $\frac{1}{C_\rho} \cdot h_{\mathrm{Kill}}(F) - C'_\rho \leq h(\rho(F)) \leq C_\rho \cdot h_{\mathrm{Kill}}(F) + C'_\rho$ for all F (as mentioned in §4.3, h_{Kill} and the height associated to a Chevalley basis of \mathfrak{g} only differ by an additive constant). Granting the conclusion of Proposition 3.3 for the adjoint representation, we obtain $g \in \mathbb{G}(\overline{\mathbb{Q}})$ such that $h(\rho(gFg^{-1})) \leq CC_\rho^2 \cdot \widehat{h}(\rho(F)) + C'_\rho$. But by the main Theorem 3.1, since F generates a nonvirtually solvable group, we have $C'_\rho \leq C^K \cdot \widehat{h}(\rho(F))$ and $CC_\rho^2 \leq C^K$ for some $K = K(d) \in \mathbb{N}$ independent of F . Hence $h(\rho(gFg^{-1})) \leq 2C^K \cdot \widehat{h}(\rho(F))$. The remaining inequalities are clear or follow from the basic properties of heights explained in Section 2.

Proof of Proposition 3.3 for the adjoint representation. We therefore assume that $\rho = \mathrm{Ad}$ and $h = h_{\mathrm{Kill}}$, while \mathbb{G} is semisimple of adjoint type and $\langle F \rangle$ is Zariski dense in \mathbb{G} . Again let T be a maximal torus in \mathbb{G} and pick a corresponding basis of $\mathfrak{g}_\mathbb{Z}$ made of weight vectors say (Y_1, \dots, Y_d) as in Section 5. Since \mathbb{G} is of adjoint type, it is a direct product of its simple factors. Looking at the projection of F to each simple factors, it is straightforward to verify that, when proving Proposition 3.3, we can reduce to the case when \mathbb{G} is absolutely simple. So we assume \mathbb{G} absolutely simple.

Clearly, if we prove the statement for a bounded power of F instead, then this will prove the statement for F . Hence making use of escape (i.e., applying

Proposition 4.1), and after possibly conjugating F by an element of $\mathbb{G}(\overline{\mathbb{Q}})$, we may assume that F contains two elements a, b which generate a subgroup acting irreducibly on \mathfrak{g} and such that a is a regular semisimple element in T and b is generic with respect to T , i.e., such that the matrix coefficient B_{ij} of $\text{Ad}(b)$ in the basis (Y_1, \dots, Y_d) is nonzero for any indices i, j . We thus write $F = \{a, b, b_1, \dots, b_M\}$.

Let $S \subset [1, d]$ be the set of indices corresponding to the simple roots. So $|S| = \text{rk}(\mathbb{G})$. Let $I_r \subset [1, d]$ be the set of indices corresponding to the Y_i 's that belong to $\mathfrak{t} = \text{Lie}(T)$. For each $j \in S$, let us choose some $i_j \in I_r$. We have $B_{i_j j} B_{j i_j} \neq 0$. Then one can choose a unique point $t \in T(\overline{\mathbb{Q}})$ such that $\alpha_j(t)^2 = \frac{B_{i_j j}}{B_{j i_j}}$ for each $j \in S$. As we may, we change F into $t F t^{-1}$. Then $B_{i_j j} = B_{j i_j}$ for every $j \in S$. Moreover we know from (18) that for any place v and any real number $s_v > E_v^{\text{Ad}}(F)$, there exists $t_v \in T(\overline{\mathbb{Q}}_v)$ such that

$$\|\text{Ad}(b^{t_v})\|_v \leq C_v^d \cdot \left(\prod_{k=1}^p \max\{1, L_k(a)_v\} \right)^{dC_2} \cdot e^{s_v(1+dC_1)},$$

where C_1, C_2, C_∞ are positive constants independent of v and $C_v = 1$ if v is non-Archimedean, while $C_v = C_\infty$ if v is Archimedean. Since every matrix coefficient of $\text{Ad}(b^{t_v})$ is bounded by $\|\text{Ad}(b^{t_v})\|_v$ if v is non-Archimedean and by a constant multiple of this norm if v is Archimedean, up to enlarging C_∞ if necessary, we get that the same bound holds for all matrix coefficients of $\text{Ad}(b^{t_v})$, i.e.,

(27)

$$\log^+ |\alpha_i \alpha_j^{-1}(t_v) B_{ij}|_v \leq d \log C_v + dC_2 \sum_{k=1}^p \log^+ L_k(a)_v + (1 + dC_1) s_v =: r_v(a).$$

Specializing this for $B_{ij} = B_{ji}$ when $j \in S$ and $i = i_j$ and adding, we obtain

$$2 \log^+ |B_{ij}|_v = \log^+ |B_{ij} B_{ji}|_v \leq 2r_v(a).$$

On the other hand,

$$\begin{aligned} (28) \quad \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v \cdot r_v(a) &\leq d \log C_\infty \\ &\quad + dC_2 \sum_{k=1}^p \left(h(\delta_k^{-1}) + \log^+ \frac{1}{\kappa} \right) + (1 + dC_1) e(F) \\ &\leq C'_\infty + (1 + dC_1 + dpC_2) e(F), \end{aligned}$$

where C'_∞ is another positive constant, $\delta_k = 1 - \alpha_k(a)$ for $k \in S$, $\kappa = \min_{k \in S} |\alpha_k|_\infty$ as in Section 6.1 above, and where we have used $h(\delta_k^{-1}) = h(\delta_k) \leq h(\alpha_k(a)) + \log 2 \leq e(F) + \log 2$. Hence for $j \in S$ and $i = i_j$,

$$(29) \quad h(B_{ij}) \leq C'_\infty + (1 + dC_1 + dpC_2) e(F).$$

On the other hand, since $i \in I_r$ $\alpha_i = 1$ and (27) gives

$$\begin{aligned} \log^+ |\alpha_j^{\pm 1}(t_v) B_{ij}|_v &\leq r_v(a), \\ \log^+ |\alpha_j^{\pm 1}(t_v)|_v &\leq r_v(a) + \log^+ \left| \frac{1}{B_{ij}} \right|_v. \end{aligned}$$

Taking the weighted sum over all places, we get

$$h(\alpha_j(t_v)_v), h(\alpha_j^{-1}(t_v)_v) \leq h\left(\frac{1}{B_{ij}}\right) + \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \cdot r_v(a)$$

which, as $h(B_{ij}^{-1}) = h(B_{ij})$, gives from (28) and (29)

$$(30) \quad h(\alpha_j(t_v)_v), h(\alpha_j^{-1}(t_v)_v) \leq 2C'_\infty + 2(1 + dC_1 + dpC_2)e(F).$$

Now let α be an arbitrary root, i.e., $\alpha = \prod_{j \in S} \alpha_j^{n_j}$ for some integers $n_j \in \mathbb{Z}$. Since there are only finitely many possibilities for the n_j 's given \mathbb{G} , there is a bound, say N , for the possible sums $\sum |n_j|$. Hence (30) gives

$$h(\alpha(t_v)_v) \leq 2NC'_\infty + 2N(1 + dC_1 + dpC_2)e(F)$$

for every root α . Finally, if i and j are arbitrary indices this time, from (27) and (28) we get

$$\begin{aligned} h(B_{ij}) &\leq \frac{1}{[K:\mathbb{Q}]} \sum_{v \in V_K} n_v \cdot r_v(a) + h(\alpha_i^{-1}(t_v)_v) + h(\alpha_j(t_v)_v) \\ &\leq (4N + 1)C'_\infty + (4N + 1)(1 + dC_1 + dpC_2)e(F). \end{aligned}$$

Since $A_{ij} = 0$ for $i \neq j$ while $h(A_{ii}) \leq e(F)$ by Proposition 2.14(c), we finally get $h_{\mathrm{Kill}}(A) + h_{\mathrm{Kill}}(B) \leq O_d(1) \cdot (\sum_{ij} h(A_{ij}) + h(B_{ij})) \leq C + C \cdot e(F)$.

Now recall that a and b were chosen so that they generate a subgroup which acts irreducibly on $\mathfrak{g}(\overline{\mathbb{Q}})$. By Burnside's theorem, this means that $\mathrm{Ad}(a)$ and $\mathrm{Ad}(b)$ generate $\mathrm{End}(\mathfrak{g})$ as an associative $\overline{\mathbb{Q}}$ -algebra. In particular, one can find d^2 elements, say u_1, \dots, u_{d^2} , in $\{\mathrm{Id}, \mathrm{Ad}(a), \mathrm{Ad}(b)\}^{d^2}$ which form a basis of $\mathrm{End}(\mathfrak{g})$ over $\overline{\mathbb{Q}}$. Clearly $h_{\mathrm{Kill}}(u_i) \leq d^2(C + Ce(F))$ for each $i = 1, \dots, d^2$. Let E_{ij} be the elementary matrices associated to our basis (Y_1, \dots, Y_d) of \mathfrak{g} . We may write u_i as a linear combination $\sum U_{kl}^{(i)} E_{kl}$ with $U_{kl}^{(i)} \in \overline{\mathbb{Q}}$. By definition of the height $h = h_{\mathrm{Kill}}$ on $\mathrm{End}(\mathfrak{g})$, it differs from the height associated to the basis (Y_1, \dots, Y_d) only by an additive constant C_∞ due to the fact that the Y_i 's are not necessarily orthogonal at infinite places. Thus each height $h(U_{kl}^{(i)})$ is at most $h(u_i) + C_\infty$. In particular, the height of the determinant of (u_1, \dots, u_{d^2}) in the basis of the E_{ij} is bounded in terms of the $h(u_i)$ hence in terms of $e(F)$ only. As a result, if we write each E_{ij} as a linear combination $\sum x_k^{(ij)} u_k$ with $x_k^{(ij)} \in \overline{\mathbb{Q}}$, then the height $h(x_k^{(ij)})$ is bounded in terms of $e(F)$ (and d) only, i.e., $\leq C''_\infty + O_d(1) \cdot e(F)$ for some other constant $C''_\infty > 0$ depending on d only.

Let c be any element of $F = \{a, b, b_1, \dots, b_M\}$. Then we may write $C = \text{Ad}(c) = \sum C_{ij} E_{ij}$ and $C_{ij} = (E_{ji} C)_{jj} = \sum x_k^{(ij)} (u_k C)_{jj}$. Now observe that we may apply (18) to the two matrices $\{\text{Ad}(a), u_k C\}$ and get as in (27) for each place v and all $j = 1, \dots, d$,

$$\log^+ |(u_k C)_{jj}|_v \leq d \log C_v + d C_2 \sum_{k=1}^p \log^+ L_k(a)_v + (1 + d C_1) s_v = r_v(a).$$

We may now estimate $\log^+ \|F\|_v$. First if v is non-Archimedean, one gets $\log^+ \|F\|_v \leq \log^+ \max_{k,j,c} |(u_k C)_{jj}|_v + \log^+ \max_{k,i,j} |x_k^{(ij)}|_v$ and

$$\log^+ \|F\|_v \leq r_v(a) + \sum_{k,i,j} \log^+ |x_k^{(ij)}|_v,$$

while if v is Archimedean, we get the same estimate plus an additive error. Summing over the places as in (29) we have

$$h_{\text{Kill}}(F) \leq C''_\infty + \sum_{k,i,j} h(x_k^{(ij)}) + \frac{1}{[K : \mathbb{Q}]} \sum_{v \in V_K} n_v r_v(a),$$

and thus $h_{\text{Kill}}(F) \leq O_d(1)(1 + e(F))$. Using Theorem 3.1, this upper bound can be replaced by $h_{\text{Kill}}(F) \leq O_d(1) \cdot e(F)$, and Proposition 3.3 is proved. \square

Remark 7.3. In positive characteristic p with p not 2 nor 3 and \mathbb{G} not of type A_n , the adjoint representation is irreducible and the above proof continues to hold verbatim without having to appeal to Theorem 3.1 at the end because no additive constant appears in the upper bound (since all places are non-Archimedean). In the cases where the adjoint representation is not irreducible, one can modify the above proof to make it work for every irreducible rational representation instead of Ad . One has to take a set of linearly independent weights χ_j in place of the simple roots in order to define the conjugating element $t \in T$, and then modify (18) accordingly. Details are left to the reader.

Proof of Proposition 1.1 from the introduction. Let \mathbb{G} be the Zariski closure of F in GL_d . Since we are in characteristic 0, \mathbb{G} is completely reducible when acting on $\overline{\mathbb{Q}}^d$. Since there are only finitely many isomorphism classes of semisimple algebraic subgroups of GL_d and finitely many isomorphism classes of irreducible representations of \mathbb{G} of dimension at most d , we may consider the maximum of all constants $C \geq 1$ appearing in Proposition 3.3 for the various semisimple groups \mathbb{G} and representations that can arise. Thus Proposition 3.3 gives a basis of V with height h_0 and $g_0 \in \mathbb{G}(\overline{\mathbb{Q}})$ such that $h_0(g_0 F g_0^{-1}) \leq C \widehat{h}_0(F)$. But there is $g \in \text{GL}_d(\overline{\mathbb{Q}})$ such that $h(\cdot) = h_0(g \cdot g^{-1})$ and $\widehat{h} = \widehat{h}_0$, so we are done. \square

7.3. *Proof of Corollaries 3.5 and 1.7.* First we assume that F generates a nonvirtually solvable group. From Lemma 2.1, we have for any set F containing 1, $\sum_{a \in F^{d^2}} e(\{a\}) \geq e(F) - |\log c|$. In particular,

$$(31) \quad \max\{e(\{a\}), a \in F^{nd^2}\} \geq \frac{1}{|F|^{nd^2}} (n\hat{h}(F) - |\log c|)$$

for every $n \in \mathbb{N}$. Now by Theorem 3.1, we have $\hat{h}(F) > \varepsilon = \varepsilon(d) > 0$. Hence for some $n_0 = n_0(d) \in \mathbb{N}$,

$$\max\{e(\{a\}), a \in F^{n_0}\} \geq \frac{d}{|F|^{n_0}} \cdot \hat{h}(F).$$

On the other hand, we clearly have $e(\{a\}) \leq \sum h(\lambda)$ where the sum is over the d eigenvalues of a . Hence the assertion of Corollary 3.5.

Now assume that F generates a virtually solvable subgroup. It is well known (see [42, 3.6 and 10.10]) that there is an integer $n_0 = n_0(d) \in \mathbb{N}$ such that any virtually solvable subgroup of $\mathrm{GL}_d(\mathbb{C})$ contains a subgroup of index at most n_0 which can be conjugated inside the upper-triangular matrices. Applying Lemma 4.10 (and its proof), we may find $F_1 \subset F^{2n_0-1}$ such that $F^n \cap B \subset (F_1 \cup F_1^{-1})^{2n}$ for all n , where $B = T_d(\mathbb{C})$ is the subgroup of upper-triangular matrices. But $F^n = \cup (F^n \cap f_i^{-1}B)$ for at most n_0 elements f_i in F^{n_0} . Hence $F^n \subset \cup f_i^{-1}(F^{n+n_0} \cap B)$ and $R_v(F) \leq \liminf \|F^n \cap B\|^{1/n} \leq R_v(F_1 \cup F_1^{-1})^2$. However, since $F_1 \subset B$, it is straightforward to observe that $R_v(F_1 \cup F_1^{-1}) = \Lambda_v(F_1 \cup F_1^{-1})$. Summing over all places, we obtain $\hat{h}(F) \leq 2 \sum_{a \in F_1} e(\{a\}) + e(\{a^{-1}\}) \leq 2|F|^{2n_0} \max\{\sum h(\lambda) + h(\lambda^{-1}), \lambda \text{ eigenvalue of } a \in F_1\}$. Since $h(\lambda) = h(\lambda^{-1})$, we get the desired result.

Now we turn to Corollary 1.7. By the remark above on the bound n_0 of the index of a triangular subgroup in any virtually solvable subgroup, it is easy to see that the set of pairs (A, B) in $\mathrm{GL}_d \times \mathrm{GL}_d$ that generate a virtually solvable subgroup is a closed subvariety. Since every connected simple algebraic group can be topologically generated (for the Zariski topology) by two elements (see Proposition 1.8), we can apply the escape from subvarieties lemma (Lemma 4.2) and conclude that there is a pair $\{A, B\}$ in $F^{c(d)}$ which generates a nonvirtually solvable subgroup of $\langle F \rangle$. Then apply Corollary 3.5 to $\{\mathrm{Id}, A, B\}$.

7.4. *Proof of Corollaries 3.6, 1.9 and 1.10.*

Proof of Corollary 3.6. Let k be the algebraic closure of K and Γ the subgroup generated by F . First assume that $\Gamma \leq \mathrm{GL}(W)$ acts absolutely irreducibly on $W = k^d$. According to Burnside's theorem the k -subalgebra generated by the elements of Γ is the full algebra $\mathrm{End}_k(W)$. Since $D = \dim \mathrm{End}_k(W) = (\dim W)^2 \leq d^2$, there exists a linear basis, say w_1, \dots, w_D of $\mathrm{End}_k(W)$ in F^{d^2} (start with $w_1 = 1$, then multiply by the elements of F one after the

other). Since $\{x \mapsto \text{tr}(zx)\}_{z \in \text{End}_k(W)}$ account for all linear forms on $\text{End}_k(W)$, the linear forms $x \mapsto \text{tr}(w_i x)$ must be linearly independent, and the matrix $\{\text{tr}(w_i w_j)\}_{1 \leq i, j \leq D}$ is invertible. Let L be the field generated by the eigenvalues of all elements of F^{2d^2+1} . Note that L contains $\text{tr}(w_i w_j)$ and $\text{tr}(f w_i w_j)$ for $f \in F$ and all i, j . We claim that $\Gamma \leq \bigoplus_{1 \leq i \leq D} L w_i \leq \text{End}_k(W)$. Indeed for each i , and each $f \in F$, write $f w_i = \sum a_{ij} w_j$ for some $a_{ij} \in k$. Then as $\{\text{tr}(w_i w_j)\}_{1 \leq i, j \leq D}$ is invertible, the a_{ij} must belong to L . Since $w_1 = 1$, we see that positive words in F lie all in $\bigoplus_{1 \leq i \leq D} L w_i$. On the other hand, the Cayley-Hamilton theorem implies that $f^{-1} \in L[f]$. Finally $\Gamma \leq \bigoplus_{1 \leq i \leq D} L w_i$ as claimed. The left regular representation of Γ on $\bigoplus_{1 \leq i \leq D} L w_i$ gives us a faithful representation of Γ in $\text{GL}_D(L)$. If F^{2d^2+1} consists only of torsion elements, the field L , is generated over its prime field by finitely many roots of unity. If $\text{char}(K) > 0$, then this already implies that L is finite and thus that Γ is finite, a contradiction. If $\text{char}(K) = 0$, then L belongs to $\overline{\mathbb{Q}}$ and we are thus reduced to the case when Γ lies in $\text{GL}_D(\overline{\mathbb{Q}})$. Then, by the combination of Corollary 3.5 with Theorem 3.1 we are done unless Γ is virtually solvable.

If Γ does not act irreducibly on k^d , let $\{0\} \leq V_1 \leq \dots \leq V_k = k^d$ be a composition series for Γ and let $W = V_{i_0}/V_{i_0+1}$ be an (irreducible) composition factor. If $\text{char}(K) > 0$, by the above, the image of Γ in $\text{GL}(W)$ is finite. It follows that Γ is virtually unipotent and hence finite, because finitely generated unipotent subgroups in positive characteristic are finite.

If $\text{char}(K) = 0$, then the image of Γ on each composition factor is virtually solvable, and hence Γ itself is virtually solvable. Recall that there is an integer $n_0 = n_0(d) \in \mathbb{N}$ such that any virtually solvable subgroup of $\text{GL}_d(\mathbb{C})$ contains a subgroup of index at most n_0 which can be conjugated inside the upper-triangular matrices (see [42, 3.6 and 10.10]). Applying Lemma 4.10, we may assume without loss of generality that F is made of upper-triangular matrices. Then for every $a, b \in F$, the commutator $[a, b]$ is a unipotent matrix in $\text{SL}_d(\mathbb{C})$, hence is either trivial or of infinite order. If one of them has infinite order, we are done. Otherwise this means that the matrices in F commute. But a finitely generated abelian group generated by torsion elements is finite. We are done.

The argument above works verbatim without the need to take inverses until the point in the last paragraph when F is assumed to consist of upper-triangular matrices. Note that if the elements of F are torsion, then their eigenvalues are roots of unity, hence the group generated by F is virtually nilpotent. This completes the proof of the corollary. \square

Proof of Corollary 1.9 from the introduction. If γ has a transcendental eigenvalue for some $\gamma \in F^{2d^2+1}$, then the second alternative obviously holds. If

no $\gamma \in F^{2d^2+1}$ has a transcendental eigenvalue, then the argument given in the proof of Corollary 3.6 shows that Γ has a faithful representation in $\mathrm{GL}_{d^2}(\overline{\mathbb{Q}})$. So we are reduced to this situation and the claim is clear by Corollary 1.9. \square

Proof of Corollary 1.10 from the introduction. If F fixes a point in the Bruhat-Tits building X_k of SL_d over a p -adic field k , then F fixes a vertex of X_k (it fixes the vertices of the smallest simplex containing the fixed point). But vertices of X_k are permuted transitively by the action of $\mathrm{GL}_d(k)$. It follows from Lemma 4.14 that $E_k(F) = 1$. Hence if F fixes a point on each X_k for k non-Archimedean, then $e_f(F) = 0$. Hence by Theorem 3.1 we must have $e_\infty(F) > \varepsilon$. Thus there exists an embedding σ of K in \mathbb{C} such that $\log E_{\mathbb{C}}(\sigma(F)) > \varepsilon$. Then by Lemma 4.14, every point of $X_{\mathbb{C}}$ must be moved by at least ε by some element of F . \square

Acknowledgments. This paper grew out of my attempts to improve the results of [21] and [14]. I wish to thank T. Gelander for our past collaboration and joint works which naturally led me to study the questions addressed in this paper.

I thank the referee for his careful reading of the paper and many comments that helped improve the exposition. I am grateful to P. Sarnak and A. Yafaev from whom I learned about Bilu's theorem. I thank E. Bombieri, J.-B. Bost and A. Chambert-Loir for their insights about diophantine geometry and P. E. Caprace for his invaluable help with buildings. I also thank G. Chenevier, E. Lindenstrauss, G. Prasad, A. Salehi-Golsefidy and V. Talamanca for stimulating conversations.

References

- [1] F. AMOROSO and S. DAVID, Le problème de Lehmer en dimension supérieure, *J. Reine Angew. Math.* **513** (1999), 145–179. MR 1713323. Zbl 1011.11045. <http://dx.doi.org/10.1515/crll.1999.058>.
- [2] F. AMOROSO and R. DVORNICICH, A lower bound for the height in abelian extensions, *J. Number Theory* **80** (2000), 260–272. MR 1740514. Zbl 0973.11092. <http://dx.doi.org/10.1006/jnth.1999.2451>.
- [3] M. F. ATIYAH and I. G. MACDONALD, *Introduction to Commutative Algebra*, Addison-Wesley Publishing Co., Reading, MA, 1969. MR 0242802. Zbl 0175.03601.
- [4] M. H. BAKER and R. RUMELY, Equidistribution of small points, rational dynamics, and potential theory, *Ann. Inst. Fourier (Grenoble)* **56** (2006), 625–688. MR 2244226. Zbl 05176555. Available at http://aif.cedram.org/item?id=AIF_2006_56_3_625_0.
- [5] L. BARTHOLDI and Y. DE CORNULIER, Infinite groups with large balls of torsion elements and small entropy, *Arch. Math. (Basel)* **87** (2006), 104–112. MR 2249644. Zbl 1136.20035. <http://dx.doi.org/10.1007/s00013-005-1684-4>.

- [6] Y. BILU, Limit distribution of small points on algebraic tori, *Duke Math. J.* **89** (1997), 465–476. MR 1470340. Zbl 0918.11035. <http://dx.doi.org/10.1215/S0012-7094-97-08921-3>.
- [7] E. BOMBIERI and W. GUBLER, *Heights in Diophantine Geometry*, *New Math. Monogr.* **4**, Cambridge Univ. Press, Cambridge, 2006. MR 2216774. Zbl 1115.11034. <http://dx.doi.org/10.1017/CBO9780511542879>.
- [8] A. BOREL, *Linear Algebraic Groups*, W. A. Benjamin, New York, 1969, notes taken by Hyman Bass. MR 0251042. Zbl 0186.33201.
- [9] N. BOURBAKI, *Groupes et Algèbres de Lie*, Hermann, chapitres 4–8.
- [10] E. BREUILLARD, On uniform exponential growth for solvable groups, *Pure Appl. Math. Q.* **3** (2007), 949–967. MR 2402591. Zbl 1147.20027.
- [11] ———, A strong Tits alternative, 2008, preprint. arXiv 0804.1395.
- [12] ———, Heights on sl_2 and free subgroups, in *Geometry, Rigidity and Group Actions, Zimmer's Festschrift* (B. FARB and D. FISHER, eds.), Chicago Univ. Press, Chicago, IL, 2011.
- [13] ———, Effective estimates for the spectral radius of a finite set of matrices, preprint.
- [14] E. BREUILLARD and T. GELANDER, Uniform independence in linear groups, *Invent. Math.* **173** (2008), 225–263. MR 2415307. Zbl 1148.20029. <http://dx.doi.org/10.1007/s00222-007-0101-y>.
- [15] M. R. BRIDSON and A. HAEFLIGER, *Metric Spaces of Non-Positive Curvature*, *Grundle. Math. Wissen.* **319**, Springer-Verlag, New York, 1999. MR 1744486. Zbl 0988.53001.
- [16] F. BRUHAT and J. TITS, Groupes réductifs sur un corps local, *Inst. Hautes Études Sci. Publ. Math.* (1972), 5–251. MR 0327923. Zbl 0254.14017. Available at http://www.numdam.org/item?id=PMIHES_1972__41__5_0.
- [17] P. E. CAPRACE, personal communication.
- [18] A. CHAMBERT-LOIR, Mesures et équidistribution sur les espaces de Berkovich, *J. Reine Angew. Math.* **595** (2006), 215–235. MR 2244803. Zbl 1112.14022. <http://dx.doi.org/10.1515/CRELLE.2006.049>.
- [19] C. W. CURTIS and I. REINER, *Representation Theory of Finite Groups and Associative Algebras*, *Pure and Applied Math.* **11**, Interscience Publishers, a division of John Wiley & Sons, New York, 1962. MR 0144979. Zbl 0131.25601.
- [20] P. B. EBERLEIN, *Geometry of Nonpositively Curved Manifolds*, *Chicago Lectures in Math.*, University of Chicago Press, Chicago, IL, 1996. MR 1441541. Zbl 0883.53003.
- [21] A. ESKIN, S. MOZES, and H. OH, On uniform exponential growth for linear groups, *Invent. Math.* **160** (2005), 1–30. MR 2129706. Zbl 1137.20024. <http://dx.doi.org/10.1007/s00222-004-0378-z>.
- [22] C. FAVRE and J. RIVERA-LETELIER, Équidistribution quantitative des points de petite hauteur sur la droite projective, *Math. Ann.* **335** (2006), 311–361. MR 2221116. Zbl 1175.11029. <http://dx.doi.org/10.1007/s00208-006-0751-x>.

- [23] T. GELANDER, Homotopy type and volume of locally symmetric manifolds, *Duke Math. J.* **124** (2004), 459–515. MR 2084613. Zbl 1076.53040. <http://dx.doi.org/10.1215/S0012-7094-04-12432-7>.
- [24] N. IWAHORI and H. MATSUMOTO, On some Bruhat decomposition and the structure of the Hecke rings of p -adic Chevalley groups, *Inst. Hautes Études Sci. Publ. Math.* (1965), 5–48. MR 0185016. Zbl 0228.20015. Available at http://www.numdam.org/item?id=PMIHES_1965__25_5_0.
- [25] N. JACOBSON, *Lie Algebras, Interscience Tracts in Pure and Applied Math.* **10**, Interscience Publishers, a division of John Wiley & Sons, New York, 1962. MR 0143793. Zbl 0121.27504.
- [26] D. A. KAZDAN and G. A. MARGULIS, A proof of Selberg’s hypothesis, *Mat. Sb.* **75** (117) (1968), 163–168. MR 0223487. Zbl 0241.22024.
- [27] E. LANDVOGT, Some functorial properties of the Bruhat-Tits building, *J. Reine Angew. Math.* **518** (2000), 213–241. MR 1739403. Zbl 0937.20026. <http://dx.doi.org/10.1515/crll.2000.006>.
- [28] S. LANG, *Algebra*, third ed., *Graduate Texts in Math.* **211**, Springer-Verlag, New York, 2002. MR 1878556. Zbl 0984.00001.
- [29] G. D. MOSTOW, Self-adjoint groups, *Ann. of Math.* **62** (1955), 44–55. MR 0069830. Zbl 0065.01404. <http://dx.doi.org/10.2307/2007099>.
- [30] J. PINEIRO, L. SZPIRO, and T. J. TUCKER, Mahler measure for dynamical systems on \mathbb{P}^1 and intersection theory on a singular arithmetic surface, in *Geometric Methods in Algebra and Number Theory, Progr. Math.* **235**, Birkhäuser, Boston, MA, 2005, pp. 219–250. MR 2166086. Zbl 1101.11020. http://dx.doi.org/10.1007/0-8176-4417-2_10.
- [31] M. S. RAGHUNATHAN, *Discrete Subgroups of Lie Groups, Ergeb. Mat. Grenz.* **68**, Springer-Verlag, New York, 1972. MR 0507234. Zbl 0254.22005.
- [32] A. SCHINZEL, *Polynomials with Special Regard to Reducibility, Encycl. Math. Appl.* **77**, Cambridge Univ. Press, Cambridge, 2000, with an appendix by Umberto Zannier. MR 1770638. Zbl 0956.12001. <http://dx.doi.org/10.1017/CBO9780511542916>.
- [33] I. SCHUR, *Über Gruppen periodischer Substitutionen*, Sitzber. Preuss. Akad. Wiss., 1911. JFM 42.0155.01.
- [34] C. SMYTH, The Mahler measure of algebraic numbers: a survey, in *Number Theory and Polynomials, London Math. Soc. Lecture Note Ser.* **352**, Cambridge Univ. Press, Cambridge, 2008, pp. 322–349. MR 2428530. <http://dx.doi.org/10.1017/CBO9780511721274.021>.
- [35] R. STEINBERG, *Lectures on Chevalley Groups*, Yale University, New Haven, Conn., 1968, notes prepared by John Faulkner and Robert Wilson. MR 0466335. Zbl 1196.22001.
- [36] L. SZPIRO, E. ULLMO, and S. ZHANG, Équirépartition des petits points, *Invent. Math.* **127** (1997), 337–347. MR 1427622. Zbl 0991.11035. <http://dx.doi.org/10.1007/s002220050123>.

- [37] V. TALAMANCA, A Gelfand-Beurling type formula for heights on endomorphism rings, *J. Number Theory* **83** (2000), 91–105. MR 1767654. Zbl 0965.16018. <http://dx.doi.org/10.1006/jnth.1999.2506>.
- [38] W. P. THURSTON, *Three-Dimensional Geometry and Topology. Vol. 1*, Princeton Math. Ser. **35**, Princeton Univ. Press, Princeton, NJ, 1997, edited by Silvio Levy. MR 1435975. Zbl 0873.57001.
- [39] J. TITS, Free subgroups in linear groups, *J. Algebra* **20** (1972), 250–270. MR 0286898. Zbl 0236.20032. [http://dx.doi.org/10.1016/0021-8693\(72\)90058-0](http://dx.doi.org/10.1016/0021-8693(72)90058-0).
- [40] E. ULLMO, Positivité et discrétion des points algébriques des courbes, *Ann. of Math.* **147** (1998), 167–179. MR 1609514. Zbl 0934.14013. <http://dx.doi.org/10.2307/120987>.
- [41] H. C. WANG, Topics on totally discontinuous groups, in *Symmetric Spaces* (Short Courses, Washington Univ., St. Louis, Mo., 1969–1970), *Pure and Appl. Math.* **8**, Dekker, New York, 1972, pp. 459–487. MR 0414787. Zbl 0232.22018.
- [42] B. A. F. WEHRFRITZ, *Infinite Linear Groups. An Account of the Group-Theoretic Properties of Infinite Groups of Matrices*, *Ergeb. Mat. Grenz.* **76**, Springer-Verlag, New York, 1973. MR 0335656. Zbl 0261.20038.
- [43] S.-W. ZHANG, Small points and adelic metrics, *J. Algebraic Geom.* **4** (1995), 281–300. MR 1311351. Zbl 0861.14019.
- [44] ———, Equidistribution of small points on abelian varieties, *Ann. of Math.* **147** (1998), 159–165. MR 1609518. Zbl 0991.11034. <http://dx.doi.org/10.2307/120986>.

(Received: April 17, 2008)

UNIVERSITÉ PARIS SUD, ORSAY, FRANCE
E-mail: emmanuel.breuillard@math.u-psud.fr