APPROXIMATE SUBGROUPS AND SUPER-STRONG APPROXIMATION

EMMANUEL BREUILLARD

Laboratoire de Mathématiques, Bâtiment 425, Université Paris Sud 11, 91405 Orsay, France
Email: emmanuel.breuillard@math.u-psud.fr

Abstract

Surveying some of the recent developments on approximate subgroups and super-strong approximation for thin groups, we describe the Bourgain-Gamburd method for establishing spectral gaps for finite groups and the proof of the classification of approximate subgroups of semisimple algebraic groups over finite fields. We then give a proof of the super-strong approximation for mod p quotients via random matrix products and a quantitative version of strong approximation. Some applications to the group sieve are also presented. These notes are based on a series of lectures given at the 2013 Groups St Andrews meeting.

1 Introduction

In the early 1980's Matthews-Vaserstein-Weisfeiler [69], and then Nori [72] and Weisfeiler [101] (independently) proved the following theorem:

Theorem 1.1 (Strong-approximation theorem) *Suppose $\mathbb{G}$ is a connected, simply connected, semisimple algebraic group defined over $\mathbb{Q}$, and let $\Gamma \leqslant \mathbb{G}(\mathbb{Q})$ be a finitely generated Zariski-dense subgroup. Then for all sufficiently large prime numbers p, the reduction Γ_p of Γ is equal to $\mathbb{G}_p(\mathbb{F}_p)$.*

For example, if $\Gamma \leqslant \mathrm{SL}_n(\mathbb{Z})$ is a finitely generated Zariski dense subgroup, then $\Gamma_p = \mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})$ for all large enough prime numbers p. When p is large enough, the algebraic group $\mathbb{G}$ (viewed as a closed subgroup of some GL_n) admits a smooth reduction defined over $\mathbb{F}_p$, which we denote by $\mathbb{G}_p$. Since Γ is finitely generated, there are finitely many primes $p_1, \ldots, p_k$ (appearing in the denominators of the matrix entries of S) such that Γ belongs to $\mathbb{G}(\mathbb{Z}[1/p_1, \ldots, 1/p_k]) := \mathbb{G} \cap \mathrm{GL}_n(\mathbb{Z}[1/p_1, \ldots, 1/p_k])$, and the reduction modulo p map is well-defined on this subgroup if p is large enough.

The result fails if $\mathbb{G}$ is not simply connected (e.g., the image of $\mathrm{SL}_2(\mathbb{Z})$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ has index 2 when $p > 2$). However every connected absolutely almost simple algebraic group admits a simply connected finite cover to which we can lift Γ and apply the theorem. This yields that $[\mathbb{G}_p(\mathbb{F}_p) : \Gamma_p]$ is nevertheless always bounded (for p large) by a constant depending only on $\mathbb{G}$ (one can take $1 + \mathrm{rank}(\mathbb{G})$, see [72, Remark 3.6]).

A similar result holds for groups defined over number fields instead of $\mathbb{Q}$. Its proof reduces to the case of $\mathbb{Q}$ by suitable restriction of scalars. See Remark 6.4 below (see also [101]).

That the result holds when Γ is an S-arithmetic group $\Gamma = \mathbb{G}(\mathbb{Z}[1/p_1, \ldots, 1/p_m])$ was known much earlier by work of Kneser [49] and Platonov [74] in particular. See [75, Chapter 7]) and [82].

Theorem 1.1 is then of particular interest when the group Γ is not a full S-arithmetic subgroup of $\mathbb{G}$ but has infinite index in one of them, while still remaining Zariski dense in $\mathbb{G}$ (S-arithmetic subgroups are Zariski dense by the Borel density theorem). Such a group is called a *thin subgroup* of $\mathbb{G}$ in recent terminology due to Peter Sarnak [91].

What we call *super-strong approximation* is the fact stated in Theorem 1.2 below that Γ not only surjects onto $\mathbb{G}_p(\mathbb{F}_p)$ for p large but that the associated Cayley graphs of $\mathbb{G}_p(\mathbb{F}_p)$ form a *family of expanders*. The goal of these notes is to give a proof of this fact, give some applications, and introduce the reader to the various techniques used in the proof.

It is of course not the purpose of this survey to give a complete introduction to expander graphs and for that matter we refer the reader to the many sources on the subject starting with Lubotzky's monograph [61] and survey [63] (see also [38] and [51, 96, 10]). Let us simply recall that to every finite k-regular graph $\mathcal{G}$ is associated a combinatorial Laplace operator acting on the (finite dimensional) space of functions on the vertices of the graph. It is defined by the formula

$$\Delta f(x) = f(x) - \frac{1}{k}\sum_{y\sim x} f(y),$$

where $y \sim x$ is a vertex connected to x by an edge. This operator is symmetric and non-negative. Its eigenvalues are real and non-negative. The eigenvalue 0 comes with multiplicity one if the graph is connected and the first nonzero eigenvalue is denoted by $\lambda_1(\mathcal{G})$ and satisfies:

$$\lambda_1(\mathcal{G}) = \inf\{\langle \Delta f, f\rangle, \|f\|_2 = 1, \sum_x f(x) = 0\}. \tag{1.1}$$

An infinite family of k-regular graphs $(\mathcal{G}_n)_{n\geqslant 1}$ is said to be a *family of expanders* if there is $\varepsilon > 0$ such that for all $n \geqslant 1$,

$$\lambda_1(\mathcal{G}_n) > \varepsilon.$$

We are now in a position to state the following strengthening of Theorem 1.1.

Theorem 1.2 (Super-strong approximation) *Suppose $\mathbb{G}$ is a connected, simply connected, semi-simple algebraic group defined over $\mathbb{Q}$, and let $\Gamma \leqslant \mathbb{G}(\mathbb{Q})$ be a Zariski-dense subgroup generated by a finite set S. Then there is $\varepsilon = \varepsilon(S) > 0$ such that for all large enough prime numbers p, the reduction Γ_p of Γ is equal to $\mathbb{G}_p(\mathbb{F}_p)$ and the associated Cayley graph $\mathrm{Cay}(\mathbb{G}_p(\mathbb{F}_p), S_p)$ is an ε-expander.*

Here S_p is the image of S by reduction modulo p. As before, the result also holds if $\mathbb{G}$ is not assumed to be simply connected, but Γ_p may then only be a subgroup of $\mathbb{G}_p(\mathbb{F}_p)$ whose index is nevertheless bounded independently of p, while $\mathrm{Cay}(\Gamma_p, S_p)$ remains an ε-expander.

This theorem is a special case of a result due to Salehi-Golsefidy and Varjú [87], which asserts that the conclusion also holds for quotient modulo a square free integer and even when the connected algebraic group $\mathbb{G}$ is only assumed to be perfect. Their proof follows the so-called Bourgain-Gamburd expansion machine, which can

be implemented in this context in part thanks to the recent results on approximate subgroups of linear groups due to Pyber-Szabó [80] and Breuillard-Green-Tao [19].

In these notes we describe the Bourgain-Gamburd method as well as the above mentioned results on approximate subgroups and finally give a complete proof of Theorem 1.2 (i.e., of super-strong approximation for mod p quotients) following a somewhat alternate route than in [87] by use of random matrix products [15].

1.1 The Lubotzky alternative and its expander version

One can formulate a version of the strong approximation theorem, which is valid for every finitely generated subgroup of $\mathrm{GL}_d(k)$, where k is an arbitrary field of characteristic zero (one can also deal with the positive characteristic case thanks to the work of Pink [73], however no super-strong version is known in positive characteristic thus far). When the group $\Gamma = \langle S \rangle$ we start with is non virtually solvable, one can show that there is a non trivial connected and simply connected semisimple algebraic group G defined over $\mathbb{Q}$ and a group homomorphism from a finite index subgroup of Γ into $\mathbb{G}(\mathbb{Q})$ with a Zariski-dense image (see [68, Prop. 16.4.13] and the discussion that follows). This allows to then apply the strong-approximation theorem 1.1 and deduce that Γ_0 admits $\mathbb{G}_p(\mathbb{F}_p)$ as a quotient for almost all p.

This information was used in a key way by Lubotzky and Mann in their work on subgroup growth [64]. For this version of strong approximation, called *the Lubotzky alternative*, we refer the reader to the notes devoted to it and its various refinements in the book by Lubotzky and Segal on subgroup growth ([68, 16.4.12], see also [48]). Strengthened by the super-strong approximation theorem, this gives the following statement:

Theorem 1.3 (Lubotzky super-alternative) *Let S be a finite symmetric subset of $\mathrm{GL}_d(k)$, where k is a field of characteristic zero. Then the subgroup $\Gamma = \langle S \rangle$ generated by S contains a subgroup Γ_0 whose index m in Γ is finite and bounded in terms of d only, such that*

- *either the subgroup Γ_0 is solvable,*
- *or there is a connected, simply connected, semisimple algebraic group $\mathbb{G}$ defined over $\mathbb{Q}$, such that for all large enough primes $p \in \mathbb{N}$, there is a surjective group homomorphism ρ_p from Γ_0 to $\mathbb{G}_p(\mathbb{F}_p)$ such that the Cayley graph $\mathrm{Cay}(\mathbb{G}_p(\mathbb{F}_p), \rho_p(S_0))$ is an ε-expander, for some $\varepsilon > 0$ independent of p, where S_0 is a subset of S^{2m} generating Γ_0.*

Note that given a group Γ generated by a symmetric set S, then every subgroup of finite index Γ_0 is finitely generated by a symmetric subset contained in S^{2m-1}, if m is the index of Γ_0 in Γ (e.g., see [19, Lemma C.1]).

A version of Theorem 1.3 for a bounded number of primes is also true: given large enough distinct primes $p_1, \ldots, p_k$, the Cayley graphs $\mathrm{Cay}(G(\mathbb{F}_{p_1}) \times \ldots \times G(\mathbb{F}_{p_k}), (\rho_{p_1} \times \ldots \times \rho_{p_k})(S))$ are ε-expanders for a uniform $\varepsilon > 0$ independent of the number of primes k. We will prove this stronger version only with an ε depending on k (but not on the choice of k primes). See Theorem 6.3 below. One needs the works of Varjú [100] and Salehi-Golsefidy-Varjú [87] to get this uniformity in the number of primes, but the proof is rather more involved. Note that at any case ε depends on S and it is an

open question whether this dependence can be removed (see [16] for partial results in this direction).

1.2 The group sieve method

Knowing that the finite quotients Cayley graphs are expanders is a very useful information for a number of applications to group theory and number theory, in particular it is the basis of the so-called Group Sieve, pioneered by Kowalski [52, 53], Rivin [83], and Lubotzky-Meiri [65, 66] and of the Affine Sieve of Bourgain-Gamburd-Sarnak [7]. See [50] and [55] for two nice expositions.

Roughly speaking, the expander property allows one to give very good bounds on the various error terms that appear when sieving modulo primes. In these notes, we will give a general statement, *the group sieve lemma* (Lemma 7.3 below), due to Lubotzky and Meiri, which allows to show that a subset Z of a given finitely generated linear group is exponentially small, provided its reduction modulo p does not occupy too large a subset of the quotient group for many primes p. For this version of the group sieve, expansion for pairs of primes is sufficient (i.e., we need that $G(\mathbb{F}_{p_1})\times G(\mathbb{F}_{p_2})$ expands for $p_1 \neq p_2$), so our version of the Lubotzky super-alternative above will be enough. Expansion for all square free moduli is necessary however, and sometimes crucial, in other situations, such as in the Affine Sieve pioneered by Bourgain-Gamburd-Sarnak [7] and further developed by Salehi-Golsefidy-Sarnak [86], Bourgain and Kontorovich [9] and others.

The conclusion of the super-strong approximation theorem (Theorem 1.2) can be reformulated in the following way: there is $\varepsilon > 0$ depending only on the generating set S such that for every real valued function f on the group $\mathbb{G}_p(\mathbb{F}_p)$, such that $\sum_{x\in\mathbb{G}_p(\mathbb{F}_p)} f(x) = 0$ and $\|f\|^2_{\ell^2} = \sum_{x\in\mathbb{G}_p(\mathbb{F}_p)} |f(x)|^2 = 1$,

$$\langle \Delta f, f\rangle > \varepsilon,$$

where

$$\langle \Delta f, f\rangle = \frac{1}{2k}\sum_{s\in S} \|s\cdot f - f\|^2_{\ell^2} = \frac{1}{2k}\sum_{s\in S}\sum_{x\in\mathbb{G}_p(\mathbb{F}_p)} |f(s^{-1}x) - f(x)|^2.$$

Let $S_p = \{s_1, \dots, s_k\}$ be the image of S under the reduction modulo p map and μ_{S_p} be the uniform probability measure on S_p, assigning equal mass $1/k$ $(= 1/|S|$ for p large enough) to each element of S_p.

$$\mu_{S_p} := \frac{1}{k}(\delta_{s_1} + \dots + \delta_{s_k})$$

Note that $\mu_{S_p} = Id - \Delta$ as operators on $\ell^2(\mathbb{G}_p(\mathbb{F}_p))$, and hence its operator norm on $\ell^2_0(\mathbb{G}_p(\mathbb{F}_p))$, the orthogonal of constants, satisfies:

$$\|\mu_{S_p}|_{\ell^2_0}\| < 1 - \varepsilon$$

It is in this form that the theorem is used in its applications to the group sieve method. For example it allows Lubotzky and Meiri [65] to establish the following result about the scarcity of proper powers in non virtually solvable linear groups. A group element is called a proper power if it is of the form g^n for some integer $n \geqslant 2$ and some other group element g (from the same group).

Theorem 1.4 (Lubotzky-Meiri [65]) *Let $\Gamma \leqslant \mathrm{GL}_d(\mathbb{C})$ be a finitely generated subgroup and let μ_S be the uniform probability measure on a finite symmetric generating S. Assume that Γ is not virtually solvable. Then the set $\mathcal{P}_\Gamma$ of proper powers in Γ is exponentially small in the sense that there is $c = c(S) > 0$ such that for every $n \in \mathbb{N}$,*

$$\mu_S^n(\mathcal{P}_\Gamma) \leqslant e^{-cn}.$$

Here μ_S^n is the n-th convolution power of the probability measure μ_S on Γ. Equivalently, it is the distribution at time n of the simple random walk starting at the identity on the associated Cayley graph $\mathrm{Cay}(\Gamma, S)$. Or more explicitly:

$$\mu_S^n(\mathcal{P}_\Gamma) = \mathbb{P}_{w \in W_{n,k}}(\mathcal{P}_\Gamma) := \frac{|\{w, |w| = n, \overline{w} \in \mathcal{P}_\Gamma\}|}{|\{w, |w| = n\}|},$$

where $W_{n,k}$ is the set of (non reduced!) words w of length $|w| = n$ in the formal alphabet made of letters from the set S, and $\overline{w}$ its value as a group element when computed inside Γ. One can analogously count reduced words of length n in the free group and get the same result, but we note in passing that obtaining a result of this kind for the average with respect to the word metric on Γ induced by S seems out of reach at the moment, because little is known about the balls for the word metric on a group of exponential growth.

1.3 On the proof of the super-strong approximation theorem

Theorem 1.2 was first proved in the special case of subgroups of $\mathrm{SL}_2(\mathbb{Z})$ in a remarkable breakthrough by Bourgain and Gamburd [5]. They deduced the expansion by showing that the simple random walk on the finite quotient $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ must equidistribute very fast, indeed after only $O(\log p)$ steps. In doing so they reversed the traditional way of looking at things: traditionally spectral gaps estimates were proven by other methods (e.g., representation theory, property (T), etc.) and were then used to prove fast equidistribution of random walks. Bourgain and Gamburd reversed this order, first proving equidistribution and then deducing the gap (see Proposition 3.3 below for the equivalence between spectral gap and fast equidistribution).

This idea can be traced back to the seminal work of Sarnak and Xue [92], which gave a new, softer, approach toward Selberg's 3/16 theorem (i.e., the first eigenvalue of the Laplace operator on quotients of the hyperbolic plane by congruence subgroups of $\mathrm{SL}(2, \mathbb{Z})$ is at least 3/16, see [93]). They exploited, via the trace formula, the high multiplicity of the spectrum coming from the $(p-1)/2$ lower bound on the dimension of the smallest non trivial complex representation of $\mathrm{SL}_2(\mathbb{F}_p)$ (this bound goes back to Frobenius) and a soft combinatorial upper bound on the number of lattice points in a ball of radius roughly $\log p$. We refer the reader to the expository papers of P. Sarnak [90, 89], where this method and its history (in particular the role of Bernstein and Kazhdan) is described.

In his thesis [29] Gamburd pursued this method and established the first spectral gap result valid for thin groups: he showed that if a finitely generated subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is large enough in the sense that the Hausdorff dimension of its limit set on $\mathbb{P}^1(\mathbb{R})$ is at least $\frac{5}{6}$, then the spectrum of the associated (infinite volume) quotients of the hyperbolic plane modulo the congruence subgroups $\Gamma_p := \Gamma \cap \ker(\mathrm{SL}_2(\mathbb{Z}) \to$

$\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$) admits a uniform lower bound independent of p. In turn the resulting Cayley graphs of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ are expander graphs.

Bourgain and Gamburd [5] pushed the method even further to implement it for all Zariski-dense subgroups of $\mathrm{SL}_2(\mathbb{Z})$ with no restriction on the limit set. The structure of their proof retained the same patterns, playing the high multiplicity lower bound against a combinatorial upper bound via the trace formula applied to convolution powers of a fixed probability measure on the generating set. Achieving this combinatorial upper bound is the gist of their work: they brought in an important graph theoretic result (the Balog-Szemerédi-Gowers lemma, a parent of the celebrated Szemerédi regularity lemma) revisited in this context by Tao [97] to show that convolution powers of probability measures decay in ℓ^2 norm (the so-called ℓ^2-flattening) unless the measure charges significantly a certain approximate subgroup. That there exists no interesting approximate subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$ was established for this purpose by Helfgott [36]. The combinatorial upper bound (on the probability of return to the identity of the simple random walk at time roughly $\log p$), and hence the spectral gap, then reduces to establishing a certain non concentration estimate on subgroups for random walks on $\mathrm{SL}_2(\mathbb{Z})$ (see Theorem 5.1), which in this case can easily be deduced from Kesten's theorem [47].

This new method became known as the *Bourgain-Gamburd expansion machine* (see, e.g., the papers [20, 22] as well as the forthcoming book [96]). Its scope goes beyond $\mathrm{SL}_2(\mathbb{F}_p)$ and, quite remarkably, it can potentially be applied to any finite group (see Proposition 3.1 for a precise formulation of the method and its ingredients). It was understood early on that the scheme of the proof in [5] was general enough that it could be made to work in the general setting of Theorem 1.2, provided one could establish each step in the right generality. The bounds on the dimension of complex representations are well-known thanks to classical work of Landazuri-Seitz [57]. The graph theoretic lemma needs no modification in the general setting. The remaining two items however require deeper consideration. The classification of approximate groups, first established by Helfgott for $\mathrm{SL}_2(\mathbb{F}_p)$ and $\mathrm{SL}_3(\mathbb{F}_p)$, was finally completed in the general case by Pyber and Szabó [80] and independently by Breuillard-Green-Tao [19]. Regarding the upper bounds on the probability of hitting a subgroup, there are two known ways to achieve them. The first is to use the theory of random matrix products, and this was done in subsequent work of Bourgain-Gamburd [6], but only in the special case of subgroups of $\mathrm{SL}_n(\mathbb{Z})$, because the estimates from the theory of random matrix products required to deal with the general case were lacking. The second consists in applying a ping-pong argument akin to the proof of the Tits alternative [99], and this was performed by Varjú in his thesis [100] and subsequently by Salehi-Golsefidy and Varjú in their joint work [87], in which they establish Theorem 1.2 in full generality.

In the remainder of these notes we will prove Theorem 1.2 following each of these steps very closely. The only novelty in our proof lies in the last step: thanks to [15], we now understand how to use random matrix products to prove in the desired generality the required upper bounds for the probability of hitting a subgroup (the non-concentration estimates). This approach is somewhat more direct than the one taken by Salehi-Golsefidy and Varjú in [87], and it is very close to what Green, Tao and I had in mind, when we announced a proof of Theorem 1.2 in [18, Theorem 7.3]

in the special case of absolutely simple groups over $\mathbb{Z}$, but never came to the point of writing it up in full.

As already mentioned Salehi-Golsefidy and Varjú [87] actually proved a strong version of Theorem 1.2 showing the expansion property also for the quotients modulo a square free integer, and assuming only that $\mathbb{G}$ is perfect (which is also a necessary condition for expansion). See Theorem 6.5 below. That strong version is crucial for certain applications to sieving in orbits (à la Bourgain-Gamburd-Sarnak [7]), but its proof is much more involved. Often it is enough to have Theorem 1.2, or its extension to two or a bounded number of primes, which is not more costly. That will be the case for the applications presented in this paper. This, I thought, was enough justification for writing a complete proof of super-strong approximation for prime moduli in one place.

1.4 Outline of the article

In Section 2 we present a proof of the strong approximation theorem of Matthews, Vassertein and Weisfeiler following Nori's proof. Our treatment yields a quantitative version in the sense that it gives a upper bound on the first p for which the surjectivity of the reduction mod p holds in terms of the height of the generating set. Section 3 is devoted to the Bourgain-Gamburd machine: we state very general conditions on the Cayley graph of an arbitrary finite group that are sufficient to establish a spectral gap. Section 4 is devoted to approximate subgroups of linear groups over finite fields. We prove there the theorem of Pyber-Szabó and Breuillard-Green-Tao. In Section 5 we discuss random matrix products and a general non-concentration on subgroups result for random walks on linear groups. Finally in Section 6 we combine the results of the preceding three sections to complete the proof of the super-strong approximation theorem in the case of mod p quotients (Theorems 1.2 and 6.3). The final section is devoted to applications to the group sieve method and results of Aoun, Jouve-Kowalski-Zywina, Lubotzky-Meiri, Lubotzky-Rosenzweig and Prasad-Rapinchuk on generic properties elements in non virtually solvable linear groups.

2 Nori's theorem and a quantitative version of strong approximation

It was Matthews, Vaserstein and Weisfeiler [69] who first proved the strong approximation theorem for Zariski-dense subgroups, i.e., Theorem 1.1, in the case when G is absolutely simple. Their proof made use of the (brand new at the time) classification of finite simple groups. Another, classification-free proof was found roughly at the same time and independently by M. Nori, yielding also the case G semisimple, as a consequence of the following general result proved in [72].

Theorem 2.1 (Nori [72]) *Let H be a subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$, and H^+ the subgroup generated by its elements of order p. If p is larger than some constant $c(n)$ depending only on n, then there is a connected algebraic subgroup $\widetilde{H}$ of GL_n defined over $\mathbb{F}_p$ such that H^+ coincides with $\widetilde{H}(\mathbb{F}_p)^+$. Moreover there is a normal abelian subgroup $A \leqslant H$ such that $[H : AH^+]$ is bounded in terms of n only.*

Observe that if $p \geqslant n$, then elements of order p in $\mathrm{GL}_n(\mathbb{F}_p)$ are precisely the unipotent matrices: indeed $x^p = 1$ is equivalent to $(x-1)^p = 0$ for $x \in \mathrm{GL}_n(\mathbb{F}_p)$ and

hence to $x = 1 + n$, where n is a nilpotent matrix. As Nori explains in [72, Remark 3.6.], the index of $\widetilde{H}(\mathbb{F}_p)^+$ in $\widetilde{H}(\mathbb{F}_p)$ is bounded by a function of n only. So the meaning of Nori's theorem is that finite subgroups of $\mathrm{GL}_n(\mathbb{F}_p)$ generated by elements of order p are essentially algebraic subgroups, if $p > c(n)$.

The key feature of Nori's theorem is that no assumption whatsoever is made on the subgroup H. Hence Nori's theorem can be seen as a description of *arbitrary* subgroups of $\mathrm{GL}_n(\mathbb{F}_p)$. It can be viewed as complementing the celebrated theorem of Camille Jordan [44] on finite subgroups of $\mathrm{GL}_n(K)$ whose order is prime to the characteristic of the field K: such a group admits an abelian subgroup whose index is bounded by some function of n only. Nori's theorem explains what happens when the characteristic divides the order of the finite group: recall that a finite group has an element of prime order p if and only if its order is a multiple of p (Cauchy's theorem).

Jordan's theorem is usually quoted for subgroups of $\mathrm{GL}_n(\mathbb{C})$, but this stronger version can be derived easily by lifting the group to $\mathbb{C}$ (see [72, Theorem C]). In fact Jordan had already proved this stronger version in his original paper: his proof is purely algebraic and applies to any finite subgroup of $\mathrm{GL}_n(K)$ all of whose elements are semisimple (or equivalently to finite subgroups without a non trivial unipotent element), where K is any algebraically closed field (see [11] for a discussion).

Textbooks presenting Jordan's theorem usually give a different, more geometric treatment, due to Frobenius, Bieberbach and Blichfeldt. Jordan's own argument seems to have been forgotten for more than a hundred years until Larsen and Pink [59] rediscovered it and generalized it considerably to obtain a classification of all finite subgroups of GL_d in every characteristic. The Larsen-Pink theorem is more general than Nori's result stated above in that it applies to finite subgroups of GL_d regardless of the field and the size of the characteristic. We will comment on the Larsen-Pink theorem further below, when we discuss approximate subgroups of linear groups. The proof of the Larsen-Pink theorem, which by the way is also independent of the classification of finite simple groups, plays a key role in the structure theorem for approximate subgroups of linear groups (see Theorem 4.5 below).

For the applications to strong and super-strong approximation, we will not need the full force of Theorem 2.1 above. Rather the following important special case will be sufficient.

Theorem 2.2 (Sufficiently Zariski-dense subgroups) *There is $M = M(d)$ such that the following holds. Let $p > M$ be a prime number and $\mathbb{G}_p \leqslant \mathrm{GL}_d$ be a semisimple simply connected algebraic group defined over $\mathbb{F}_p$. If a subgroup $H \leqslant \mathbb{G}_p(\mathbb{F}_p)$ is not contained in a proper algebraic subgroup of $\mathbb{G}_p$ of complexity at most M, then it must be equal to $\mathbb{G}_p(\mathbb{F}_p)$.*

We say informally that a closed algebraic subvariety of GL_d has complexity at most M if it can be defined as the vanishing locus of a finite set of polynomials such that the sum of their degrees in each variable is at most M. See [19] for background on this notion. It is particularly useful in positive characteristic: saying that a finite subgroup of $\mathrm{GL}_d(\overline{\mathbb{F}_p})$ is algebraic is meaningless, because every finite subgroup is an algebraic subset with several (possibly many) irreducible components. However putting a bound on the complexity forces a bound on the number of irreducible components [19, Lemma A.4] and hence restricts the class of finite subgroups drastically

and leads to interesting statements, such as the above.

We now sketch Nori's proof of Theorem 2.2. A similar argument is due to Gabber, see [46, Thm 12.4.1]. Pushing this idea a bit further allows Nori to also prove Theorem 2.1.

Proof (sketch) If H had no non trivial unipotent element, it would have an abelian subgroup of bounded index by Jordan's theorem. But this would violate the assumption that H is sufficiently Zariski-dense. So H contains a unipotent element, which we may write in the form $h = \exp \xi$, for some nilpotent matrix ξ. The $\mathbb{F}_p$-span V_H of all H-conjugates of ξ is invariant under the adjoint action of H. The assumption that H is sufficiently Zariski-dense implies that V_H must be the full $\mathbb{F}_p$-Lie algebra of $\mathbb{G}_p$ in $gl_d(\mathbb{F}_p)$. Pick unipotent elements $h_1, \dots, h_d \in H$ such that the corresponding ξ_i's form a basis of $\mathrm{Lie}(\mathbb{G}_p)$.

Now consider the map $\Phi : \mathbb{F}_p^{\dim \mathbb{G}} \to \mathbb{G}_p(\mathbb{F}_p)$, $(t_1, \dots, t_d) \mapsto h_1^{t_1} \cdot \dots \cdot h_d^{t_d}$. Note that Φ is a polynomial map whose degree is bounded in terms of d only. Its image lies in H. We claim that there is a constant $c = c(d) > 0$ such that $|\operatorname{Im} \Phi| \geqslant cp^d$. Indeed, the Jacobian of Φ is not identically zero, so outside its vanishing locus (a proper subvariety, hence a subset of size $O(p^{d-1})$) the fibers of Φ are of bounded cardinality. This implies the desired bound.

Now since there are positive constants c_1, c_2 such that $c_1 p^d \leqslant |\mathbb{G}_p(\mathbb{F}_p)| \leqslant c_2 p^d$ (e.g., see [72, Lemma 3.5.]), we get that the index $[\mathbb{G}_p(\mathbb{F}_p) : H]$ is bounded. However since $\mathbb{G}$ is simply connected, $\mathbb{G}_p(\mathbb{F}_p)$ is an almost direct product of quasi-simple groups and thus has no subgroups of bounded index when p is large (Kneser-Tits for $\mathbb{F}_p$, see [75], see also Remark 3.4). Hence $H = \mathbb{G}_p(\mathbb{F}_p)$. □

Nori's proof of strong approximation (i.e., of Theorem 1.1) is based on Theorem 2.2 alone. We will explain this argument below. It turns out that this argument even yields a quantitative lower bound on the first prime number for which we can claim that $\Gamma_p = \mathbb{G}_p(\mathbb{F}_p)$ in terms of the height of the generating set of Γ. Namely:

Theorem 2.3 (Strong approximation, quantitative version)
Suppose $\mathbb{G} \leqslant \mathrm{GL}_d$ is a connected, simply connected, semisimple algebraic group defined over $\mathbb{Q}$. Then there are constants $p_0, C_0 \geqslant 1$ such that if $S \subset \mathbb{G}(\mathbb{Q})$ is a finite symmetric set generating a Zariski-dense subgroup $\Gamma = \langle S \rangle$ of $\mathbb{G}$, and M_S denotes the maximal height of an element of S, then for every prime number $p > \max\{p_0, M_S^{C_0}\}$, the reduction Γ_p of Γ is equal to $\mathbb{G}_p(\mathbb{F}_p)$.

Here the height $H(s)$ of an element $s \in \mathrm{GL}_d(\mathbb{Q})$ is defined naively as the maximum of the numerators and denominators appearing in the expressions of the matrix coefficients of s as irreducible fractions. The bound p_0 is related to the bound $c(n)$ from Nori's theorem and to p_M from Lemma 2.7 below. There is very little control on this bound in general (see [87, Appendix] for a discussion of this issue).

Several other proofs and extensions of Theorem 1.1 (to groups defined over number fields, to positive characteristic, etc.) have since been found. For those we refer the reader to the original articles, in particular [101], [72], [41], [73], and to the chapter on strong approximation in the recent book by Lubotzky and Segal [68] or in Nikolov's

lecture notes in [48, chapter II]. We also recommend reading Rapinchuk's recent survey [82], which gives a thorough overview of strong approximation.

We now pass to the derivation of Theorem 2.3 from Nori's theorem. First, we replace the naive height with another height, which is better suited for our purposes since it is sub-additive. Given $a \in \mathrm{GL}_d(\mathbb{Q})$, set

$$h(a) := \sum_{p,\infty} \log^+ \|a\|_p,$$

where the sum is over all prime numbers p as well as the infinite place ∞. Here $\log^+ := \max\{\log, 0\}$, and $\|a\|_p$ denotes $\max_{ij} |a_{ij}|_p$, the maximum p-adic absolute value of a matrix entry a_{ij} of a, while $\|a\|_\infty$ is the operator norm of a for the standard Euclidean norm on $\mathbb{R}^d$. The following is straightforward:

Lemma 2.4 (a) *The height $h(a)$ is sub-additive, i.e., for all $a, b \in \mathrm{GL}_d(\mathbb{Q})$,*

$$h(ab) \leqslant h(a) + h(b),$$

and (b) *it is comparable to the naive height $H(a)$, namely, for all a,*

$$H(a) \leqslant e^{h(a)} \leqslant d(H(a))^{d^2}.$$

We conclude that for all $a_1, \ldots, a_n \in \mathrm{GL}_d(\mathbb{Q})$,

$$H(a_1 \cdot \ldots \cdot a_n) \leqslant d^n (H(a_1) \cdot \ldots \cdot H(a_n))^{d^2} \tag{2.1}$$

Combined with the next lemma, this inequality allows us to assume, in the proof of Theorem 2.3 that Γ is generated by two elements, i.e., that $S := \{1, a^{\pm 1}, b^{\pm 1}\}$.

Lemma 2.5 (Reduction to 2 generators) *Let $\mathbb{G}$ be a semisimple algebraic group over $\mathbb{C}$. Then there is $c > 0$ such that given any finite symmetric subset $S \subset \mathbb{G}(\mathbb{C})$, with $1 \in S$, generating a Zariski dense subgroup of $\mathbb{G}$, the bounded power S^c contains two elements a, b which alone already generate a Zariski-dense subgroup.*

Proof This is Proposition 1.8. from [13]. The proof is fairly classical, and relies on Jordan's theorem and the Eskin-Mozes-Oh escape from subvarieties lemma (see, e.g., [19, Lemma 3.11]). □

Lemma 2.6 (Generating is an algebraic condition) *Let $\mathbb{G} \leqslant \mathrm{GL}_d$ be a semisimple algebraic group defined over $\mathbb{Q}$. There is a proper closed algebraic subvariety $\mathbf{X} \leqslant \mathbb{G} \times \mathbb{G}$ defined over $\mathbb{Q}$, whose points are precisely the pairs of elements in $\mathbb{G}$ which are contained in a proper algebraic subgroup of $\mathbb{G}$.*

Proof This is well-known (see, e.g., [35, Theorem 11.6]). We work over an algebraic closure of $\mathbb{Q}$ and show that $\mathbf{X}$ is a closed algebraic subset. Since $\mathbf{X}$ is invariant under Galois automorphisms, it will automatically be defined over $\mathbb{Q}$. We claim that there are finitely many absolutely irreducible finite dimensional non trivial modules of $\mathbb{G}$, say $\rho_1, \ldots, \rho_k$ such that a subgroup $\Gamma \leqslant \mathbb{G}$ is not Zariski-dense if and only if $\rho_i(\Gamma)$ fixes a line in the representation space V_i of ρ_i for some $i = 1, \ldots, k$. And this happens