GROWTH AND EXPANSION IN GROUPS OF LIE TYPE

HARALD ANDRÉS HELFGOTT

Notes for the Arizona Winter School 2016
Preliminary version: March 3, 2016

Contents

1. Introduction

These notes are meant to serve as a brief introduction to the study of growth in groups of Lie type, with SL_2 and its subgroups as the key examples. They are based in part on the survey [Hel15] and in part on my notes for courses I gave on the subject in Cusco (AGRA II school, 2015) and Göttingen.

However, given the format of the Arizona Winter School, the emphasis here is on reaching the frontiers of current research as soon as possible, and not so much on giving a comprehensive overview of the field. For that the reader is referred to [Hel15] and its bibliography, or to [Kow13] and [Tao15]. At the same time – again motivated by the school's demands – we will be able to take a brief look at several arithmetical applications at the end.

There are two essentially equivalent ways to deal with some intermediate results: one assumes what you can find in the first chapter of Mumford's Red Book [Mum99], and the other one presupposes that you have some basic notions on groups of Lie type (such as $\mathrm{SL}_2(K)$, K a finite field) and Lie algebras (such as $\mathfrak{sl}_2$) – or at least some notions over $\mathbb{R}$, and the willingness to believe that matters work out in much the same way over finite fields. We will assume very relaxed versions of these requirements, and take whichever of the two perspectives gives a clearer view at any given point.

The purpose of these notes is expository, not historical, though I have tried to give key references. The origins of several ideas are traced in greater detail in [Hel15].

1.1. **What do we mean by "growth"?** Let A be a finite subset of a group G. Consider the sets

$$\begin{aligned} A, \\ A \cdot A = \{x \cdot y : x, y \in A\}, \\ A \cdot A \cdot A = \{x \cdot y \cdot z : x, y, z \in A\}, \\ \dots \\ A^k = \{x_1 x_2 \dots x_k : x_i \in A\}. \end{aligned}$$

Write $|S|$ for the *size* of a finite set S, meaning simply the number of elements of S. A question arises naturally: how does $|A^k|$ grow as k grows?

This kind of question has been studied from the perspective of additive combinatorics (for G abelian) and geometric group theory (G infinite, $k \to \infty$). There are also some interrelated crucial concepts coming from other fields: *diameters*, *expanders*, etc.

Let A be a set of generators of G. When G is finite, rather than asking ourselves how $|A^k|$ behaves for $k \to \infty$ – it obviously becomes constant as soon as $A^k = \langle A \rangle$, where $\langle A \rangle$ is the subgroup of G generated by A – we ask what is the least k such that $A^k = G$. This value of k is called the *diameter* of G with respect to A.

The term *diameter* comes from geometry; what we have is not just an analogy – we can actually put our basic terms in a geometrical framework, as geometric group theory does. A *Cayley graph* $\Gamma(G, A)$ is the graph having $V = G$ as its set of vertices and $E = \{(g, ag) : g \in G, a \in A\}$ as its set of edges. Define the length of a path in the graph as the number of edges in it, and the distance $d(v, w)$ between two vertices v, w in the graph as the length of the shortest path between them. The *diameter* of a graph is the maximum of $d(v, w)$ over all vertices v, w. It is easy to see that the diameter of G with respect to A, as we defined it above, equals the diameter of the graph $\Gamma(G, A)$.

It is clear, then, that showing that A^k grows rapidly is a natural route towards bounds on the diameter.

Note that, if $A = A^{-1}$ (where $A^{-1} := \{g^{-1} : g \in A\}$), the graph $\Gamma(G, A)$ is symmetric, i.e., (v, w) is an edge if and only if (w, v) is an edge. Given a graph with vertices V and edges E, the the *adjacency operator* $\mathscr{A}$ is defined to be the linear operator that maps a function $f : V \to \mathbb{C}$ to the function $\mathscr{A}f : V \to \mathbb{C}$ whose value at $v \in V$ is the average of $f(w)$ on the neighbors w of v. (A vertex w is a *neighbor* of v if there is an edge (v, w) from v to w.) If a graph is symmetric, then its adjacency operator is a symmetric operator, and so it has full real spectrum. One can then study the spectrum of $\Gamma(G, A)$, and ask oneself: how large is the gap between the largest eigenvalue – namely, 1 – and all others? If the gap is at least ϵ, we say the graph is an ϵ-expander graph. Expander graphs are called in this way because of strong growth effects within them (*vertex expansion*). Conversely, thanks to [BG08b] and other works in the same direction, we know that growth in the sense above – namely, rapid growth of $|A^k|$ – can be used to prove expansion in several contexts. We will not study expansion in detail here, but it is important to mention it, since it is a subject of great interest in its own right, and many applications of growth go through it.

* * *

We will focus on the case of G non-abelian, and, in particular, on the case of G a group of Lie type, such as SL_2 over an arbitrary field K. The case of K finite can be particularly hard, in that we cannot be helped by the topology of $\mathbb{R}$ or $\mathbb{C}$, say.

Up to about 11 years ago, some of the main techniques to study this case came from modular forms. This link remains fruitful and appealing. The new progress made starting in 2005 has been more combinatorial in nature, with some ideas actually coming from additive combinatorics. In this light, it is tantalizing that algorithmic questions remain very much open, for the most part.

The same is true for some other non-abelian groups: we now have good bounds for the diameter of the symmetric group on n elements – meaning that, given a permutation puzzle with n pieces that has a solution and satisfying a very weak condition (*transitivity*[1]), we know that a very short solution exists. However, in general, we have no idea of how to find it!

1.2. **Overview: the nilpotent, solvable and simple cases.** Growth in linear algebraic groups turns out to have a very different feel to it when the group is nilpotent, when it is solvable, and – on what might be called the other end of things – when it is simple.

Let us first review some basic terms from group theory. A *normal series* of a group G is a sequence of subgroups

$$\{e\} = H_0 \lhd H_1 \lhd H_2 \lhd \cdots \lhd H_k = G, \tag{1.1}$$

i.e., H_i is normal in H_{i+1} for every $0 \leq i < k$. We say that G is *solvable* if it has a normal series with H_{i+1}/H_i abelian for every i. Being nilpotent is a stronger condition: a group G is nilpotent if it has a normal series such that, for every i, H_i

[1]Rubik's cube is a permutation puzzle, but it is not transitive: one cannot move a corner piece to an edge, or to the center of a face, even in many moves.

is normal in G and H_{i+1}/H_i lies in the center of G/H_i. A nilpotent group can often be thought of as being "almost abelian"; the present context is no exception.

Finally, a group is simple if it has no normal subgroups other than itself and $\{e\}$. In a certain sense, simple groups are to groups as the primes are to the integers: it is not hard to see that every finite group has a normal series with H_{i+1}/H_i simple for every $0 \leq i < k$, and the Jordan-Hölder theorem tells us that that series is in essence unique – the quotients H_{i+1}/H_i are determined by the group G, and only their order can change.

Let us see some examples of nilpotent, solvable and simple groups consisting of 2-by-2 and 3-by-3 matrices.

Let K be a field – say $K = \mathbb{Z}/p\mathbb{Z}$, for concreteness. The group

$$\left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in K \right\} \tag{1.2}$$

is isomorphic to $\mathbb{Z}/p\mathbb{Z}$, hence abelian, hence nilpotent. The group

$$\left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in K \right\} \qquad \text{(Heisenberg group)} \tag{1.3}$$

is also nilpotent; indeed, it is a popular example of a nilpotent group that is not abelian.

The groups

$$\left\{ \begin{pmatrix} r & x \\ 0 & r^{-1} \end{pmatrix} : r \in K^*, x \in K \right\}, \tag{1.4}$$

$$\left\{ \begin{pmatrix} r & x \\ 0 & 1 \end{pmatrix} : r \in K^*, x \in K \right\}, \tag{1.5}$$

$$\left\{ \begin{pmatrix} r & x & y \\ 0 & s & z \\ 0 & 0 & (rs)^{-1} \end{pmatrix} : r, s \in K^*, x, y, z \in K \right\} \tag{1.6}$$

are all solvable. The first and the third groups here are examples of *Borel subgroups*, i.e., maximal solvable subgroups (of SL_2 and SL_3, respectively).

Finally, while

$$\mathrm{SL}_2(K) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in K, ad - bc = 1 \right\}$$

is not quite simple, its quotient $\mathrm{SL}_2(K)/\{\pm 1\}$ is; this quotient is called $\mathrm{PSL}_2(K)$. It does not much matter, in practice, whether we work with $\mathrm{PSL}_2(K)$ or $\mathrm{SL}_2(K)$.

We will take (1.1), (1.4) and $\mathrm{SL}_2(K)$ as our main examples. We will actually discuss (1.1) only very briefly; it well serve us to make clear that abelian groups aren't always "easier" than non-abelian groups, at least not in every sense.

We will give a full treatment of (1.4). No algebraic geometry is involved there, but the procedure ("pivoting") is non-trivial, and the main result is essentially equivalent to an important result (the sum-product theorem). The main computational questions seem to be still partly open.

Lastly, we will study SL_2. It is representative of simple (and, in general, non-solvable) groups of Lie type; much more is known on them than was the case before 2005. We still now very little about computational issues in such groups, however.

Now, by the classification of finite simple groups, every finite simple group is either (a) a simple group of Lie type, (b) the alternating group $\mathrm{Alt}(n)$ for some $n \geq 5$, or (c) one of a finite list of exceptions (culminating with the "monster group". Since we are aiming at asymptotic statements, we need not concern ourselves with (c). As for (b), the best general bound for the diameter known to date is quasipolynomial [HS14], and thus not quite as good, qualitatively speaking, as the bound we will prove for SL_2. Further developments in (b) may hold the key to good bounds on the diameter for Cayley graphs of SL_n when $n \to \infty$. Perhaps surprisingly, we do have essentially algorithmic results that work for most sets of generators [BBS04], [HSZ15]. The main reason may be that stochastic arguments play a much larger role in the study of permutation groups than in the study of groups of Lie type to date, perhaps precisely because, in the case of permutation groups, there does not seem much else to use: it seems hard to state problems in permutation groups in terms of algebraic geometry.

For the same reason, alternating groups, and permutation groups in general, lie outside the purview of these notes. The reader is referred to the last part of the survey [Hel15]. Let us just finish by saying that several of the ideas here – in particular, those having to do with orbits and induction – are also useful in that context.

1.3. **Notation.** By $f(n) \ll g(n)$, $g(n) \gg f(n)$ and $f(n) = O(g(n))$ we mean the same thing, namely, that there are $N > 0$, $C > 0$ such that $|f(n)| \leq C \cdot g(n)$ for all $n \geq N$. We write $\ll_a$, $\gg_a$, O_a if N and C depend on a (say).

As usual, $f(n) = o(g(n))$ means that $|f(n)|/g(n)$ tends to 0 as $n \to \infty$. We write $O^*(x)$ to mean any quantity at most x in absolute value. Thus, if $f(n) = O^*(g(n))$, then $f(n) = O(g(n))$ (with $N = 1$ and $C = 1$).

Given a subset $A \subset X$, we let $1_A : G \to \mathbb{C}$ be the characteristic function of A:

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

2. Elementary tools

2.1. **Additive combinatorics.** Some of additive combinatorics can be described as the study of *sets that grow slowly*. In abelian groups, results are often stated so as to classify sets A such that $|A^2|$ is not much larger than $|A|$; in non-abelian groups, works starting with [Hel08] classify sets A such that $|A^3|$ is not much larger than $|A|$. Why?

In an abelian group, if $|A^2| < K|A|$, then $|A^k| < K^{O(k)}|A|$ – i.e., if a set does not grow after one multiplication with itself, it will not grow under several. This is

a result of Plünnecke [Plü70] and Ruzsa [Ruz89]. (Petridis [Pet12] recently gave a purely additive-combinatorial proof.)

In a non-abelian group G, there can be sets A breaking this rule.

Exercise 1. *Let G be a group. Let $H < G$, $g \in G \setminus H$ and $A = H \cup \{g\}$. Then $|A^2| < 3|A|$, but $A^3 \supset HgH$, and HgH may be much larger than A. Give an example with $G = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$.* Hint: let H is the subgroup of G consisting of the elements $g \in G$ leaving the basis vector $e_1 = (1, 0)$ fixed.

However, Ruzsa's ideas do carry over to the non-abelian case, as was pointed out in [Hel08] and [Tao08]; in fact, [RT85] carries over without change, since the assumption that G is abelian is never really used. We must assume that $|A^3|$ is small, not just $|A^2|$, and then it does follow that $|A^k|$ is small.

Lemma 2 (Ruzsa triangle inequality)**.** *Let A, B and C be finite subsets of a group G. Then*

$$|AC^{-1}||B| \leq |AB^{-1}||BC^{-1}|. \tag{2.1}$$

Commutativity is not needed. In fact, what is being used is in some sense more basic than a group structure; as shown in [GHR15], the same argument works naturally in any abstract projective plane endowed with the little Desargues axiom.

Proof. We will construct an injection $\iota : AC^{-1} \times B \hookrightarrow AB^{-1} \times BC^{-1}$. For every $d \in AC^{-1}$, choose $(f_1(d), f_2(d)) = (a, c) \in A \times C$ such that $d = ac^{-1}$. Define $\iota(d, b) = (f_1(d)b^{-1}, b(f_2(d))^{-1})$. We can recover $d = f_1(d)(f_2(d))^{-1}$ from $\iota(d, b)$; hence we can recover $(f_1, f_2)(d) = (a, c)$, and thus b as well. Therefore, ι is an injection. □

Exercise 3. *Let G be a group. Prove that*

$$\frac{|(A \cup A^{-1} \cup \{e\})^3|}{|A|} \leq \left(3\frac{|A^3|}{|A|}\right)^3 \tag{2.2}$$

for every finite subset A of G. Show as well that, if $A = A^{-1}$ (i.e., if $g^{-1} \in A$ for every $g \in A$), then

$$\frac{|A^k|}{|A|} \leq \left(\frac{|A^3|}{|A|}\right)^{k-2}. \tag{2.3}$$

for every $k \geq 3$. Conclude that

$$\frac{|A^k|}{|A|} \leq 3^{k-2}\left(\frac{|A^3|}{|A|}\right)^{3(k-2)} \tag{2.4}$$

for every $A \subset G$ and every $k \geq 3$.

Inequalities (2.2)–(2.4) go back to Ruzsa (or Ruzsa-Turjányi [RT85]), at least for G abelian.

This means that, from now on, we can generally focus on studying when $|A^3|$ is or isn't much larger than $|A|$. Thanks to (2.2), we can also assume in many contexts that $e \in A$ and $A = A^{-1}$ without loss of generality.

2.2. **The orbit-stabilizer theorem for sets.** A theme recurs in work on growth in groups: results on subgroups can often be generalized to subsets. This is especially the case if the proofs are quantitative, constructive, or, as we shall later see, probabilistic.

The *orbit-stabilizer theorem for sets* is a good example, both because of its simplicity (it should really be called a lemma) and because it underlies a surprising number of other results on growth. It also helps to put forward a case for seeing group actions, rather than groups themselves, as the main object of study.

We recall that an *action* $G \curvearrowright X$ is a homomorphism from a group G to the group of automorphisms of a set X. (The automorphisms of a set X are just the bijections from X to X; we will see actions on objects with richer structures later.) For $A \subset G$ and $x \in X$, the *orbit* Ax is the set $Ax = \{g \cdot x : g \in A\}$. The *stabilizer* $\mathrm{Stab}(x) \subset G$ is given by $\mathrm{Stab}(x) = \{g \in G : g \cdot x = x\}$.

The statement we are about to give is as in [HS14, §3.1].

Lemma 4 (Orbit-stabilizer theorem for sets)**.** *Let G be a group acting on a set X. Let $x \in X$, and let $A \subseteq G$ be non-empty. Then*

$$|(A^{-1}A) \cap \mathrm{Stab}(x)| \geq \frac{|A|}{|Ax|}. \tag{2.5}$$

Moreover, for every $B \subseteq G$,

$$|BA| \geq |A \cap \mathrm{Stab}(x)||Bx|. \tag{2.6}$$

The usual orbit-stabilizer theorem – usually taught as part of a first course in group theory – states that, for H a subgroup of G,

$$|H \cap \mathrm{Stab}(x)| = \frac{|H|}{|Hx|}.$$

This the special case $A = B = H$ of the Lemma we (or rather you) are about to prove.

Exercise 5. *Prove Lemma 4. Suggestion: for (2.5), use the pigeonhole principle.*

If we try to apply Lemma 4 to the action of the group G on itself by left multiplication

$$g \mapsto (h \mapsto g \cdot h)$$

or by right multiplication

$$g \mapsto (h \mapsto h \cdot g^{-1}),$$

we do not get anything interesting: the stabilizer of any element is trivial. However, we also have the action by conjugation

$$g \mapsto (h \mapsto ghg^{-1}).$$

The stabilizer of a point $h \in G$ is its *centralizer*

$$C(h) = C_G(h) = \{g \in G : gh = hg\};$$

the orbit of a point $h \in G$ under the action of the group G is the *conjugacy class*

$$\mathrm{Cl}(h) = \{ghg^{-1} : g \in G\}.$$

Thus, we obtain the following result, which will show itself to be crucial later. Its importance resides in making upper bounds on intersections with $\mathrm{Cl}(g)$ imply lower bounds on intersections with $C(g)$. In other words, we will first show that there are not too many elements of a special form, and then we will be able to use that to show that we do have plenty of elements of another special form. This will be very useful.

Lemma 6. *Let $A \subset G$ be a non-empty set. Then, for every $g \in A^l$, $l \geq 1$,*

$$|A^{-1}A \cap C(g)| \geq \frac{|A|}{|A^{l+2} \cap \mathrm{Cl}(g)|}.$$

Proof. Let $G \curvearrowright G$ be the action of G on itself by conjugation. Apply (2.5) with $x = g$; the orbit of g under conjugation by A is contained in $A^{l+2} \cap \mathrm{Cl}(g)$ □

It is instructive to see some other consequences of (2.5). The following tells us that, if we show that the intersection of A with a subgroup H grows rapidly, then we know that A itself grows rapidly.

Exercise 7. *Let G be a group and H a subgroup thereof. Let $A \subset G$ be a non-empty set with $A = A^{-1}$. Prove that, for any $k > 0$,*

$$|A^{k+1}| \geq \frac{|A^k \cap H|}{|A^{-1}A \cap H|}|A|. \tag{2.7}$$

Hint: *Consider the action $G \curvearrowright X = G/H$ by left multiplication, that is, $g \mapsto (aH \mapsto gaH)$.*

Exercise 8. *Let G be a group and H a subgroup thereof. Let $A \subset G$ be a non-empty set. Then*

$$|A^{-1}A \cap H| \geq \frac{|A|}{r}, \tag{2.8}$$

where r is the number of cosets of H intersecting A.

3. Growth in a solvable group

3.1. **Remarks on abelian groups.** Let G be an abelian group and A be a finite subset of G. This is the classical setup for what nowadays is called *additive combinatorics* – a field that may be said to have started to split off from additive number theory with Roth [Rot53] and Freiman [Fre73].

In general, for G abelian, $A \subset G$ may be such that $|A + A|$ is barely larger than $|A|$, and that is the case even if we assume that A generates G. For instance, take A to be a segment of an arithmetic progression: $A = \{2, 5, 8, \dots, 3m - 1\}$. Then $|A| = m$ and $|A + A| = 2m - 1 < 2|A|$.

Freiman's theorem [Fre73] (also called Freiman-Ruzsa, especially over arbitrary abelian groups [Ruz99]) tells us that, in a very general sense, this is the only kind of set that grows slowly. We have to start by giving a generalization of what we just called a segment of an arithmetic progression.

Definition 1. *Let G be a group. A* centered convex progression *of dimension d is a set $P \subset G$ such that there exist*

(a) *a convex subset $Q \subset \mathbb{R}^d$ that is also symmetric ($Q = -Q$),*
(b) *a homomorphism $\phi : \mathbb{Z}^d \to G$,*

for which $\phi(\mathbb{Z}^d \cap Q) = P$. We say P is proper *if $\phi|_{\mathbb{Z}^d \cap Q}$ is injective.*

Proposition 9 (Freiman-Ruzsa)**.** *Let G be an abelian group. Let $A \subset G$ be finite. Assume that $|A + A| \leq K|A|$ for some K. Then A is contained in at most $f(K)$ copies of $P+H$ for some proper, centered convex progression P of dimension $\leq g(K)$ and some finite subgroup $H < G$ such that $|P + H| \ll \exp(g(K))|A|$.*

The best known bounds are essentially those of Sanders [San12], as improved by Konyagin (see [San13]): $f(K), g(K) \ll (\log K)^{3+o(1)}$.

This is a broad field into which we will not venture further. Notice just that, in spite of more than forty years of progress, we do not yet have what is conjectured to be the optimal result, namely, the above with $f(K), g(K) \ll \log K$ (the "polynomial Freiman-Ruzsa conjecture"). Thus the state of our knowledge here is in some sense less satisfactory than in the case of simple groups, as will later become clear.

The situation for nilpotent groups is much like the situation for abelian groups: there is a generalization of the Freiman-Ruzsa theorem to the nilpotent case, due to Tointon [Toi14], based on groundwork laid by Fisher-Katz-Peng [FKP10] and Tao [Tao10].

Let us now consider the question of growth from a slightly different angle. Say we start from a set $A \subset G$ of bounded size and that we study how $|A^2|, |A^3|, \dots, |A|^k, \cdots$ grows as k grows. For G abelian, it is clear that $|A|^k$ is at most $\binom{|A|+k-1}{k} = \binom{|A|+k-1}{|A|-1}$ (why?) and thus, for $|A|$ fixed, $|A^k|$ grows polynomially on k; in fact, $|A|^k$ is bounded by a polynomial of degree $|A| - 1$. This implies immediately that $\mathrm{diam}(\Gamma(G, A)) \gg |G|^{1/(|A|-1)}$.

In $\mathbb{Z}$, the growth of $|A^k|$ in an abelian group is linear: to see this, note that, if n is the element of A with largest absolute value, then every element of A^k as absolute value at most nk, and thus $|A^k| \leq 2nk + 1$.

The situation for nilpotent groups is similar enough to that for abelian groups. Assume G is infinite, for simplicity, so that it is clear that asymptotic results are meaningful. Let A generate G. It is not hard to show that, if G is nilpotent, then $|A^k|$ grows polynomially in k. Giving a converse statement is considerably harder. There is a series of classical results in [Wol68], [Mil68], [Bas72], [Gui73]; in summary, if a set of generators A of a solvable group G has polynomial growth (i.e. $|A^k| \ll |A|^{O(1)}$), then G has a nilpotent subgroup of finite index. Tits later showed [Tit72], if G is assumed to be a linear group, but not necessarily solvable, polynomial growth still implies that G has a nilpotent subgroup of finite index; this is a consequence of the "Tits alternative", which has many other applications. Finally, Gromov proved the same statement in full generality, i.e., without assuming that G is linear; this is a deep and celebrated result [Gro81].

Given a set of generators A of $\mathbb{Z}$, it is trivial to give a very fast algorithm that expresses any given $m \in \mathbb{Z}$ as a word (i.e., a product of elements of A and their inverses) of length $O(|n| + |m|/|n|)$, where n is the element of A whose absolute

value is largest. The general case does not seem much harder – essentially because one can use induction on the normal series (1.1).

Exercise 10. *Let $A = \{a_1, a_2\}$ or $A = \{a_1, a_2, a_3\}$ be a set of generators of the Heisenberg group $H(K)$ (1.3) with $K = \mathbb{Z}/p\mathbb{Z}$. Our task, given any element g of $H(K)$, is to find a word of length $O(p^{3/2}) = O(\sqrt{H(K)})$ on A equal to g. Show that this can be done in time polynomial on $\log p$. (Note that inverting an element of $(\mathbb{Z}/p\mathbb{Z})^*$ takes time linear on $\log p$, by the Euclidean algorithm.)*

3.2. **The affine group.** The *affine group* G over a field K is the group we saw in (1.5):

$$\left\{\begin{pmatrix} r & x \\ 0 & 1 \end{pmatrix} : r \in K^*, x \in K\right\}. \tag{3.1}$$

(If we were to insist on using language in exactly the same way as later, we would say that the affine group is an algebraic group G (a variety with morphisms defining the group operations) and that (3.1) describes the group $G(K)$ consisting of its rational points. For the sake of simplicity, we avoid this sort of distinction here. We will go over most of these terms once the time to use them has come.)

Consider the following subgroups of G:

$$U = \left\{\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in K\right\}, \qquad T = \left\{\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} : r \in K^*\right\}. \tag{3.2}$$

These are simple examples of a *solvable* group G, of a maximal *unipotent* subgroup U and of a maximal torus T. In general, in SL_n, a maximal torus is just the group of matrices that are diagonal with respect to some fixed basis of $\overline{K}^n$, or, what is the same, the centralizer of any element that has n distinct eigenvalues. Here, in our group G, the centralizer $C(g)$ of any element g of G not in $\pm U$ is a maximal torus.

When we are looking at what elements of the group G do to each other by the group operation, we are actually looking at two actions: that of U on itself (by the group operation) and that of T on U (by conjugation; U is a normal subgroup of G). They turn out to correspond to addition and multiplication in K, respectively:

$$\begin{pmatrix} 1 & a_1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & a_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a_1 + a_2 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & ra \\ 0 & 1 \end{pmatrix}.$$

Thus, we see that growth in U under the actions of U and T is tightly linked to growth in K under addition and multiplication. This can be seen as motivation for studying growth in the affine group G. Perhaps we need no such motivation: we are studying growth in general, through a series of examples, and the affine group is arguably the simplest interesting example of a solvable group.

At the same time, the study of growth in a field under addition and multiplication was historically important in the passage from the study of problems in commutative groups (additive combinatorics) to the study of problems in noncommutative groups by related tools. (Growth in noncommutative groups had of course been studied

before, but from very different perspectives, e.g., that of geometric group theory.) Some of the ideas we were about to see in the context of groups come ultimately from [BKT04] and [GK07], which are about finite fields, not about groups.

Of course, the way we choose to develop matters emphasizes what the approach to the affine group has in common with the approach to other, not necessarily solvable groups. The idea of *pivoting* will appear again when we study SL_2.

Lemma 11. *Let G be the affine group over $\mathbb{F}_p$. Let U be the maximal unipotent subgroup of G, and $\pi : G \to G/U$ the quotient map.*

Let $A \subset G$, $A = A^{-1}$. Assume $A \not\subset U$; let x be an element of A not in U. Then

$$|A^2 \cap U| \geq \frac{|A|}{|\pi(A)|}, \qquad |A^2 \cap T| \geq \frac{|A|}{|A^5|}|\pi(A)| \tag{3.3}$$

for $T = C(x)$.

Recall U is given by (3.2). Since $x \notin U$, its centralizer $T = C(x)$ is a maximal torus.

Proof. By (2.8), $A_u := A^2 \cap U$ has at least $|A|/|\pi(A)|$ elements. Consider the action of G on itself by conjugation. Then, by Lemma 4, $|A^2 \cap \mathrm{Stab}(x)| \geq |A|/|A(x)|$. (Here $A(x)$ is the orbit of x under the action of A by conjugation, and $\mathrm{Stab}(x) = C(g)$ is the stabilizer of g under conjugation.) We set $A_t := (A^{-1}A) \cap \mathrm{Stab}(x) \subset T$. Clearly, $|A(x)| = |A(x)x^{-1}|$ and $(Ax)x^{-1} \subset A^4 \cap U$, and so $|A(x)| \leq |A^4 \cap U|$. At the same time, by (2.6) applied to the action $G \curvearrowright G/U$ by left multiplication, $|A^5| = |A^4 A| \geq |A^4 \cap U| \cdot |\pi(A)|$. Hence

$$|A_t| \geq \frac{|A|}{|A^4 \cap U|} \geq \frac{|A|}{|A^5|}|\pi(A)|.$$

□

The proof of the following proposition will proceed essentially by induction. This may be a little unexpected, since we are in a group G, not in, say, $\mathbb{Z}$, which has a natural ordering. However, as the proof will make clear, one can do induction on a group with a finite set of generators, even in the absence of an ordering.

Proposition 12. *Let G be the affine group over $\mathbb{F}_p$, U the maximal unipotent subgroup of G, and T a maximal torus. Let $A_u \subset U$, $A_t \subset T$. Assume $A_u = A_u^{-1}$, $e \in A_t, A_u$ and $A_u \not\subseteq \{e\}$. Then*

$$|(A_t^2(A_u))^6| \geq \frac{1}{2}\min(|A_u||A_t|, p). \tag{3.4}$$

To be clear: here

$$A_t^2(A_u) = \{t_1(u_1) : t_1 \in A_t^2, u_1 \in A_u\},$$

where $t(u) = tut^{-1}$, since T acts on U by conjugation.

Proof. Call $a \in U$ a *pivot* if the function $\phi_a : A_u \times A_t \to U$ given by

$$(u, t) \mapsto ut(a) = utat^{-1}$$

is injective.

Case (a): There is a pivot a in A_u. Then $|\phi_a(A_u, A_t)| = |A_u||A_t|$, and so

$$|A_u A_t(a)| \geq |\phi_a(A_u, A_t)| = |A_u||A_t|.$$

This is the motivation for the name "pivot": the element a is the pivot on which we build an injection ϕ_a, giving us the growth we want.

Case (b): There are no pivots in U. As we are about to see, this case can arise only if either A_u or A_t is large with respect to p. Say that (u_1, t_1), (u_2, t_2) *collide* for $a \in U$ if $\phi_a(u_1, t_1) = \phi_a(u_2, t_2)$. Saying that there are no pivots in U is the same as saying that, for every $a \in U$, there are at least two distinct (u_1, t_1), (u_2, t_2) that collide for a. Now, two distinct (u_1, t_1), (u_2, t_2) can collide for at most one $a \in U \setminus \{e\}$. (Why? A collision corresponds to a solution to a non-trivial linear equation, which can have at most one solution.) Hence, if there are no pivots, $|A_u|^2|A_t|^2 \geq |U \setminus \{e\}| = p - 1$, i.e., $|A_u| \cdot |A_t|$ is large with respect to p. This already hints that this case will not be hard; it will yield to Cauchy-Schwarz and the like.

Choose the most "pivot-like" $a \in U$, meaning an element $a \in U$ such that the number of collisions

$$\kappa_a = |\{u_1, u_2 \in A_u, t_1, t_2 \in A_t : \phi_a(u_1, t_1) = \phi_a(u_2, t_2)\}|$$

is minimal. As we were saying, two distinct (u_1, t_1), (u_2, t_2) collide for at most one $a \in U \setminus e$. Hence the total number of collisions $\sum_{a \in U \setminus \{e\}} \kappa_a$ is $\leq |A_u||A_t|(p-1) + |A_u|^2|A_t|^2$, and so

$$\kappa_a \leq \frac{|A_u||A_t|(p-1) + |A_u|^2|A_t|^2}{p-1} \leq |A_u||A_t| + \frac{|A_u|^2|A_t|^2}{p}.$$

Cauchy-Schwarz implies that $|\phi_a(A_u, A_t)| \geq |A_u|^2|A_t|^2/\kappa_a$, and so

$$|\phi_a(A_u, A_t)| \geq \frac{|A_u|^2|A_t|^2}{|A_u||A_t| + \frac{|A_u|^2|A_t|^2}{p}} = \frac{1}{\frac{1}{|A_u||A_t|} + \frac{1}{p}} \geq \frac{1}{2} \min(|A_u||A_t|, p).$$

We are not quite done, since a may not be in A. Since a is *not* a pivot (as there are none), there exist distinct (u_1, t_1), (u_2, t_2) such that $\phi_a(u_1, t_1) = \phi_a(u_2, t_2)$. Then $t_1 \neq t_2$ (why?), and so the map $\psi_{t_1,t_2} : U \to U$ given by $u \mapsto t_1(u)(t_2(u))^{-1}$ is injective. The idea is that the very non-injectivity of ϕ_a gives an implicit definition of it, much like a line that passes through two distinct points is defined by them.

What follows may be thought of as the "unfolding" step, in that we wish to remove an element a from an expression, and we do so by applying to the expression a map that will send a to something known. We will be using the commutativity of T here.

For any $u \in U$, $t \in T$, since T is abelian,

$$\begin{aligned} \psi_{t_1,t_2}(\phi_a(u,t)) &= t_1(ut(a))(t_2(ut(a)))^{-1} = t_1(u)t(t_1(a)(t_2(a))^{-1})(t_2(u))^{-1} \\ &= t_1(u)t(\psi_{t_1,t_2}(a))(t_2(u))^{-1} = t_1(u)t(u_1^{-1}u_2)(t_2(u))^{-1}, \end{aligned} \tag{3.5}$$

(Note that a has just disappeared.) Hence,

$$\psi_{t_1,t_2}(\phi_a(A_u, A_t)) \subset A_t(A_u)A_t(A_u^2)A_t(A_u) \subset (A_t(A_u))^4.$$

Since ψ_{t_1,t_2} is injective, we conclude that

$$|(A_t(A_u))^4| \geq |\psi_{t_1,t_2}(\phi_a(A_u, A_t))| = |\phi_a(A_u, A_t)| \geq \frac{1}{2}\min(|A_u||A_t|, p).$$

There is an idea here that we are about to see again: any element a that is not a pivot can, by this very fact, be given in terms of some $u_1, u_2 \in A_u$, $t_1, t_2 \in A_t$, and so an expression involving a can often be transformed into one involving only elements of A_u and A_t.

Case (c): There are pivots and non-pivots in U. This is what we can think of as the inductive step. Since $A_u \not\subseteq \{e\}$, A_u generates U. This implies that there is a non-pivot $a \in U$ and a $g \in A_u$ such that ga is a pivot. Then $\phi_{ag} : A_u \times A_t \to U$ is injective. Much as in (3.5), we unfold:

$$\begin{aligned}\psi_{t_1,t_2}(\phi_{ga}(u,t)) &= t_1(ut(g)t(a))(t_2(ut(g)t(a)))^{-1} \\ &= t_1(ut(g))t(u_1^{-1}u_2)(t_2(ut(g)))^{-1},\end{aligned} \tag{3.6}$$

where (u_1, t_1), (u_2, t_2) are distinct pairs such that $\phi_a(u_1, t_1) = \phi_a(u_2, t_2)$. Just as before, ψ_{t_1,t_2} is injective. Hence

$$|A_t(A_u)A_t^2(A_u)A_t(A_u^2)A_t^2(A_u)A_t(A_u)| \geq |\psi_{t_1,t_2}(\phi_{ga}(u,t))| = |A_u||A_t|.$$

The idea to recall here is that, if S is a subset of an orbit $\mathcal{O} = \langle A\rangle x$ such that $S \neq \emptyset$ and $S \neq \mathcal{O}$, then there is an $s \in S$ and a $g \in A$ such that $gs \notin S$. In other words, we use the point at which we escape from S. □

We are using the fact that G is the affine group over $\mathbb{F}_p$ (and not over some other field) only at the beginning of case (c), when we say that, for $A_u \subset U$, $A_u \not\subseteq \{e\}$ implies $\langle A_u\rangle = U$.

Proposition 13. *Let G be the affine group over $\mathbb{F}_p$. Let U be the maximal unipotent subgroup of G, and $\pi : G \to G/U$ the quotient map.*

Let $A \subset G$, $A = A^{-1}$, $e \in A$. Assume A is not contained in any maximal torus. Then either

$$|A^{57}| \geq \frac{1}{2}\sqrt{|\pi(A)|} \cdot |A| \tag{3.7}$$

or

$$|A^{57}| \geq \frac{1}{2}|\pi(A)|p \quad \text{and} \quad U \subset A^{112}. \tag{3.8}$$

Proof. We can assume $A \not\subset \pm U$, as otherwise what we are trying to prove is trivial. Let g be an element of A not in $\pm U$; its centralizer $C(g)$ is a maximal torus T. By assumption, there is an element h of A not in T. Then $hgh^{-1}g^{-1} \neq e$. At the same time, it does lie in $A^4 \cap U$, and so $A^4 \cap U$ is not $\{e\}$.

Let $A_u = A^4 \cap U$, $A_t = A^2 \cap T$; their size is bounded from below by (3.3). Applying Prop. 12, we obtain

$$|A^{56} \cap U| \geq \frac{1}{2}\min(|A_u||A_t|, p) \geq \frac{1}{2}\min\left(\frac{|A|}{|A^5|} \cdot |A|, p\right).$$

By (2.6), $|A^{57}| \geq |A^{56} \cap U| \cdot |\pi(A)|$. Clearly, if $|A|/|A^5| < 1/\sqrt{|\pi(A)|}$, then $|A^{57}| \geq |A^5| > \sqrt{|\pi(A)|} \cdot |A|$. □

The exponent 57 in (3.7) is not optimal, but, qualitatively speaking, Prop. 13 is as good a result as one can aim to for now: the assumption $A \not\subset T$ is necessary, the bound $\gg |\pi(A)| \cdot p$ can be tight when $U \subset A$. For $A \subset U$, getting a better-than-trivial bound amounts to Freiman's theorem in $\mathbb{F}_p$, and getting a growth factor of a power $|A|^\delta$ (rather than $\sqrt{|\pi(A)|}$) would involve proving a version of Freiman's theorem of polynomial strength. As we discussed before, that is a difficult open problem.

We can see Prop. 13 as a very simple result of the "classification of approximate subgroups" kind. If a set A grows slowly ($|A^k| \leq |A|^{1+\delta}$, $k = 57$, δ small) then either

- A is contained in a subgroup, namely, a maximal torus, or
- A is almost contained in a subgroup (U, with "almost contained" meaning that $|\pi(A)| \leq |A|^\delta$), or
- A^k contains a subgroup ($H = U$) such that $\langle A \rangle / H$ is nilpotent (here, in fact, abelian).

What we have just done, then, is to prove the simplest case of what [BGT12] calls the "Helfgott-Lindenstrauss conjecture". That conjecture states, in essence, that one can give a classification of slowly growing A like the one above when A is a subset of any linear group. A qualitative version of the conjecture was proven in [BGT12]; this means, in practice, that one can say something about the case in which $|A^k|$ is at most a constant times $|A|$, but we cannot quite yet say something in the full general case when $|A^k|$ is just assumed to be at most $|A|^{1+\delta}$. For a proof for linear groups over $\mathbb{F}_p$, see [GH14]. There is clearly work that remains to be done here.

Proposition 14. *Give examples of subsets A of the affine group over $\mathbb{F}_p$ that fail to grow for each of the reasons above: a set contained in a maximal torus, a set almost contained in U, and a set containing U, or such that a power A^k, k bounded, contains U.*

What is also interesting is that the result on growth in the affine linear group we have proved can be interpreted as a *sum-product theorem*.

Exercise 15. *Let $X \subset \mathbb{F}_p$, $Y \subset \mathbb{F}_p^*$ be given with $X = -X$, $0 \in X$, $1 \in Y$. Using Prop. 12, show that*

$$|6Y^2 X| \geq \frac{1}{2} \min(|X||Y|, p). \tag{3.9}$$

This is almost exactly [GK07], Corollary 3.5], say.

Using 3.9, or any estimate like it, one can prove the following.

Theorem 16 (Sum-product theorem [BKT04], [BGK06]; see also [EM03])**.** *For any $A \subset \mathbb{F}_p^*$ with $C < |A| < p^{1-\epsilon}$, $\epsilon > 0$, we have*

$$\max(|A \cdot A|, |A + A|) > |A|^{1+\delta},$$

where $C > 0$ and $\delta > 0$ depend only on ϵ.

In fact, the proof we have given of Prop. 12 takes its ideas from proofs of the sum-product theorem. In particular, the idea of *pivoting* is already present in them. We will later see how to apply it in a broader context.

3.3. **Diameters and navigation.** We have proved that growth occurs in SL_2 under some weak conditions. This leaves open the question of what happens with A^k, k unbounded, for A not obeying those conditions.

One things that is certainly relevant here is that there is no *vertex expansion* in the affine group, and thus no expansion. Instead of speaking of the spectrum of the adjacency operator in a graph, let us state matters in elementary terms.

Proposition 17. *For any $\lambda_1, \dots, \lambda_k \in \mathbb{Z}$, and any $\epsilon > 0$, there is a constant C such that, for every prime $p > C$, there is a set $S \cap \mathbb{F}_p$, $0 < |S| \leq p/2$, such that*

$$|S \cup (S+1) \cup \lambda_1 S \cup \dots \cup \lambda_k S| \leq (1+\epsilon)|S|. \tag{3.10}$$

Exercise 18. *Prove Proposition 17.* Hints: *prove this for $k = 1$ first; you can assume $\lambda = \lambda_1$ is ≥ 2. Here is a plan. We want to show that $|S \cap (S+1) \cap \lambda S| \leq (1+\epsilon)|S|$. For $|S \cap (S+1)|$ to be $\leq (1+\epsilon/2)|S|$, it is enough that S be a union of intervals of length $> 2/\epsilon$. (By an* interval *we mean the image of an interval $[a, b] \cap \mathbb{Z}$ under the map $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \sim \mathbb{F}_p$.) We also want $|S \cap \lambda S| \leq (1+\epsilon)|S|$; this will be the case if S is the union of disjoint sets of the form V, $\lambda^{-1}V$, ..., $\lambda^{-r}V$, $r \geq \epsilon/2$. Now, in $\mathbb{F}_p$, if I is an interval of length ℓ, then $\lambda^{-1}I$ is the union of λ intervals (why? of what length?). Choose V so that $V, \lambda^{-1}V, \dots, \lambda^{-r}V$ are disjoint. Let S be the union of these sets; verify that it fulfills (3.10).*

The following exercise shows that Prop. 17 is closely connected to the fact that a certain group is *amenable*.

Exercise 19. *Let $\lambda \geq 2$ be an integer. Define the* Baumslag-Solitar group $\mathrm{BS}(1, \lambda)$ *by*

$$\mathrm{BS}(1, \lambda) = \langle a_1, a_2 | a_1 a_2 a_1^{-1} = a_2^{\lambda} \rangle.$$

(a) *A group G with generators $a_1, \dots, a_\ell$ is called* amenable *if, for every $\epsilon > 0$, there is a finite $S \subset G$ such that*

$$|F \cup a_1 F \cup \dots \cup a_\ell F| \leq (1+\epsilon)|F|.$$

Show that $\mathrm{BS}(1, \lambda)$ *is amenable.* Hint: *to construct F, take your inspiration from Exercise 18.*

(b) *Express the subgroup of the affine group over $\mathbb{F}_p$ generated by the set*

$$A = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\} \tag{3.11}$$

as a quotient of $\mathrm{BS}(1, \lambda)$*, i.e., as the image of a homomorphism π_p defined on* $\mathrm{BS}(1, \lambda)$.

(c) *Displace or otherwise modify your sets F so that, for each of them, $\pi_p|_F$ is injective for p larger than a constant. Conclude that $S = \pi_p(F)$ satisfies (3.10).*

Amenability is not good news when we are trying to prove that a diameter is small, in that it closes a standard path towards showing that it is logarithmic in the size of the group. However, it does not imply that the diameter is not small.

Let us first be clear about what we can hope to prove. Let $\lambda \in \mathbb{F}_p^*$, and let A be as in (3.11). If λ generates $\mathbb{F}_p^*$, then A generates the affine group G over $\mathbb{F}_p$. However,

the diameter of the Cayley graph $\Gamma(G, A \cup A^{-1})$ is very large – equal to $(p-1)/2$, in fact (why?).

It is more interesting to consider instead the graph $\Gamma_{p,\lambda}$ with vertex set $\mathbb{F}_p$ and edge set

$$\{(x, x+1) : x \in \mathbb{F}_p\} \cup \{(x, \lambda x) : x \in \mathbb{F}_p\},$$

where $\lambda \in \mathbb{F}_p^*$. We are not avoiding the problem here, since the standard approach that Proposition 17 blocks is an approach to proving logarithmic diameter for these graphs $\Gamma_{p,\lambda}$.

Exercise 20. *Let $\lambda_0 \geq 2$ be an integer. Let $\lambda = \lambda_0 \bmod p$, which lies in $\mathbb{F}_p^*$ for $p > \lambda_0$. Show that the diameter of the graph $\Gamma_{p,\lambda}$ defined above is $O(\lambda_0 \log p)$.* Hint: *expansion base p.*

The proof suggested by the hint is actually constructive: given λ_0, a prime p and a vertex $x \in \mathbb{F}_p$, it constructs a path of length $O(\lambda_0 \log p)$ from the origin 0 to x. In other words, we know how to navigate in the graph.

Can we forget about λ_0, work with $\lambda \in \mathbb{F}_p^*$ arbitrary, and give a good bound that is independent of λ? This is the subject of one of the course projects.

At the time of writing of these notes, a bound of the quality $O((\log p)^{O(1)})$ is known (unpublished, based on work by Konyagin), but we have no efficient algorithm yet for navigating $\Gamma_{p,\lambda}$ in time $O((\log p)^{O(1)})$. We assume here the condition that the order of λ in $\mathbb{F}_p^*$ is $\gg \log p$; indeed, if the order of λ is $o(\log p / \log\log p)$, then the diameter of $\Gamma_{p,\lambda}$ is *not* $O((\log p)^{O(1)})$. (Why?).

4. Intersections with varieties

4.1. **Preliminaries.** We will need some basic terms from algebraic geometry. For us, a variety V will simply be an algebraic set in affine space $\mathbb{A}^n$ or projective space $\mathbb{P}^n$, i.e., the solutions to a system of polynomial equations. In particular, a variety may be reducible or irreducible. The coefficients of the equations are assumed to lie on a field K. Given a field L containing K, we write $V(L)$ for the set of solutions with coordinates in L.

Abstract algebraic varieties (as in Weil's foundations) will not really be needed, but, of course, they give a very natural way to handle a variety that parametrizes a family of varieties, among other things. For instance, we will tacitly refer to the variety of all d-dimensional planes in projective space, and, while that variety (a *Grassmanian*) can indeed be defined as an algebraic set in projective space, that is a non-obvious though standard fact.

We will not be using schemes. In particular, have no need to define the generic point of a variety in the abstract. While we will say, for instance, "property P holds for a generic point in the irreducible variety V", this simply means that there is a subvariety W of V, of dimension less than that of V, such that property P holds for every point that lies on V but not on W.

We assume all readers know what the dimension of an irreducible variety is. What is crucial for our purposes is that the dimension of a variety is a non-negative integer, since this will allow us to use the dimension as a counter, so to speak, in an inductive

process. The union of several irreducible varieties of dimension d is called a *pure-dimensional* variety of dimension d. Every variety V can be written as a finite union of irreducible varieties V_i, with $V_i \not\subset V_j$ for $i \neq j$; they are called the *irreducible components* (or simply the components) of V.

The degree of a variety V in $\mathbb{P}^n$ of dimension d is its number of points of intersection with a generic plane of dimension $n - d$. (See? We just referred tacitly to...)

Bézout's theorem, in its classical formulation, states that, for any two distinct irreducible curves C_1, C_2 in $\mathbb{A}^2$, the number of points of intersection $(C_1 \cap C_2)(\overline{K})$ is at most $d_1 d_2$. (In fact, for C_1 and C_2 generic, the number of points of intersection is exactly $d_1 d_2$; the same is true for *all* distinct C_1, C_2 if we count points of intersection with multiplicity.)

In general, if V_1 and V_2 are irreducible varieties, and we write $V_1 \cap V_2$ as a union of irreducible varieties $W_1, W_2, \dots, W_k$ with $W_i \not\subset W_j$ for $i \neq j$, a generalization of Bézout's theorem tells us that

$$\sum_{i=1}^{k} \deg(W_k) \leq \deg(V_1) \deg(V_2). \tag{4.1}$$

(See, for instance, [DS98, p.251], where Fulton and MacPherson are mentioned in connection to this and more general statements.) The classical form of Bézout's theorem is a special case of this.

This general form of Bézout's theorem implies immediately that, if a variety V is defined by at most m equations of degree at most d, then the number and degrees of the irreducible components of V are bounded in terms of m and d alone.

4.2. **Escape from subvarieties.** We are working with a finite subset A of a group G. At some points in the argument, we will need to make sure that we can find an element $g \in A^k$ (k small) that is *not* special: for example, we want to be able to use a g that is not unipotent, that does not have a given $\vec{v}$ as an eigenvector, that *is* regular semisimple (i.e., has a full set of distinct eigenvalues), etc.

It is possible to give a completely general argument of this form. Let us first set the framework. Let G be a group acting by linear transformations on n-dimensional space $\mathbb{A}^n$ over a field K. In other words, we are given a homomorphism $\phi : G \to \mathrm{GL}_n(K)$ from G to the group of invertible matrices $\mathrm{GL}_n(K)$. Let W be a subvariety of $\mathbb{A}^n$ of positive codimension, a variety whose every component has dimension $< n$. We may think of points on W as being *special*, and points outside W as being generic. We start with a point x of $\mathbb{A}^n$, and a subset A of G. The following proposition ensures us that, if, starting from x and acting on it repeatedly by A, we can eventually escape from W, then we can escape from it in a bounded number of steps, and in many ways.

The proof[2] proceeds by induction on the dimension, with the degree kept under control.

Proposition 21. *Let us be given*

[2]The statement of the proposition is as in [Hel11], based closely on [EMO05], but the idea is probably older.

- *G a group acting linearly on affine space $\mathbb{A}^n$ over a field K,*
- *$W \subsetneq \mathbb{A}^n$, a subvariety,*
- *A a set of generators of G with $A = A^{-1}$, $e \in A$,*
- *$x \in \mathbb{A}^n$ such that the orbit $G \cdot x$ of x is not contained in W.*

Then there are constants k, c depending only the number, dimension and degree of the irreducible components of $W(K)$ such that there are at least $\max(1, c|A|)$ elements $g \in A^k$ for which $gx \notin W(K)$.

Proof for a special case. Let us first do the special case of W an irreducible linear subvariety. We will proceed by induction on the dimension of W. If $\dim(W) = 0$, then W consists of a single point, and the statement is clear: since $G \cdot x \not\subset \{x\}$ and A generates G, there exists a $g \in A$ such that $gx \neq x$; if there are fewer than $|A|/2$ such elements of A, we let g_0 be one of them, and note that any product $g^{-1}g_0$ with $gx = x$ satisfies $g^{-1}g_0 x \neq x$; there are $> |A|/2$ such products.

Assume, then, that $\dim(W) > 0$, and that the statement has been proven for all W' with $\dim(W') < \dim(W)$. If $gW = W$ for all $g \in A$, then either (a) gx does not lie on W for any $g \in A$, proving the statement, or (b) gx lies on W for every $g \in G = \langle A \rangle$, contradicting the assumption. Assume that $gW \neq W$ for some $g \in A$; then $W' = gW \cap W$ is an irreducible linear variety with $\dim(W') < \dim(W)$. Thus, by the inductive hypothesis, there are at least $\max(1, c'|A|)$ elements $g' \in A^{k'}$ (c', k' depending only on $\dim(W')$) such that $g'x$ does not lie on $W' = gW \cap W$. Hence, for each such g', either $g^{-1}g'x$ or $g'x$ does not lie on W. We have thus proven the statement with $c = c'/2$, $k = k' + 1$. □

Exercise 22. *Generalize the proof so that it works without the assumptions that W be linear or irreducible. Sketch: work first towards removing the assumption of irreducibility. Let W be the union of r components, not necessarily all of the same dimension. The intersection $W' = gW \cap W$ may also have several components, but no more than r^2; this is what we meant by "keeping the degree under control". Now pay attention to d, the maximum of the dimensions of the components of a variety, and m, the number of components of maximal dimension. Show that either (1) d is lower for $W' = gW \cap W$ than for W, or (2) d is the same in both cases, but m is lower for W' than for W, or (3) x does not lie in any component of W of dimension d, and thus we may work instead with W with those components removed. Use this fact to carry out the inductive process.*

Now note that you never really used the fact that W is linear. Instead of keeping track of the number of components r, keep track of the sum of their degrees. Control that using the generalized form (4.1) of Bézout's theorem.

4.3. **Dimensional estimates.** By a *dimensional estimate* we mean a lower or upper bound on an intersection of the form $A^k \cap V$, where $A \subset G(K)$, V is a subvariety of G and G/K is an algebraic group. As you will notice, the bounds that we obtain will be meaningful when A grows relatively slowly. However, no assumption on A is made, other than that it generate $G(K)$.

Of course, Proposition 21 may already be seen as a dimensional estimate of sorts, in that it tells us that $\gg |A|$ elements of A^k, k bounded, lie outside W. We are now aiming at much stronger bounds; Proposition 21 will be a useful tool along the way.

What we aim for is estimates of the following form:

$$|A \cap V(K)| \ll |(A \cup A^{-1} \cup \{e\})^k|^{\frac{\dim V}{\dim G}}. \tag{4.2}$$

Such estimates can be traced in part to [LP11] (A a subgroup, V general) and in part to [Hel08] y [Hel11] (A an arbitrary set, but V special). We now have (4.2) as a fully general bound, thanks to [BGT11] and [PS16] (A an arbitrary set, V an arbitrary subvariety of G, and G a simple linear algebraic group, as in [LP11]). Here k is a constant that may depend on the number and degree of the components of V, and on the rank of G (e.g., $G = \mathrm{SL}_n$ has rank $n-1$), but not on the field K or on the set A.

We will show how to prove the estimate (4.2) in the case we actually need, but in a way that can be generalized to arbitrary V and arbitrary simple G. We will give a detailed outline of how to obtain the generalization.

Actually, as a first step towards the general strategy, let us study a particular V that we will not use in the end; it was crucial in earlier versions of the proof, and, more importantly, it makes several of the key ideas clear quickly. The proof is basically the same as in [Hel08, §4].

Lemma 23. *Let $G = \mathrm{SL}_2$, K a field. Let $A \subset G(K)$ be a finite set of generators of $G(K)$. Assume $A = A^{-1}$, $e \in A$. Then*

$$|A \cap T(K)| \ll |A^k|^{1/3}, \tag{4.3}$$

where k and the implied constant are absolute.

Proof. We can assume without loss of generality that $|K|$ is greater than a constant, as otherwise the statement is trivial. We can also assume without loss of generality that $A = A^{-1}$, $e \in A$, and $|A|$ is greater than a constant, simply by replacing A by $(A \cup A^{-1} \cup \{e\})^c$, c a constant, if necessary. We can also write the elements of T as diagonal matrices, by conjugation by an element of $\mathrm{SL}_2(\overline{K})$.

Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{4.4}$$

be any element of $\mathrm{SL}_2(\overline{K})$ with $abcd \neq 0$. Consider the map $\phi : T(K) \times T(K) \times T(K) \to G(K)$ given by

$$\phi(x, y, z) = x \cdot gyg^{-1} \cdot z.$$

We would like to show that this map is in some sense almost injective. (What for? If the map were injective, and we had $g \in A^\ell$, ℓ bounded by a constant, we would have

$$|A \cap T(K)|^3 = |\phi(A \cap T(K), A \cap T(K), A \cap T(K))| \leq |AA^\ell AA^{-\ell}A| = |A^{2\ell+3}|,$$

which would imply immediately the result we are trying to prove. Here we are simply using the fact that the image $\phi(D)$ of an injection ϕ has the same number of elements as the domain D.)

Multiplying matrices, we see that, for

$$x = \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix}, \ y = \begin{pmatrix} s & 0 \\ 0 & s^{-1} \end{pmatrix}, \ z = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix},$$

$\phi((x,y,z))$ equals

$$\begin{pmatrix} rt(sad - s^{-1}bc) & rt^{-1}(s^{-1} - s)ab \\ r^{-1}t(s - s^{-1})cd & r^{-1}t^{-1}(s^{-1}ad - sbc) \end{pmatrix}. \tag{4.5}$$

Let $s \in \overline{K}$ be such that $s^{-1} - s \neq 0$ and $sad - s^{-1}bc \neq 0$. A brief calculation shows then that $\phi^{-1}(\{\phi((x,y,z))\})$ has at most 16 elements: we have

$$rt^{-1}(s^{-1} - s)ab \cdot r^{-1}t(s - s^{-1})cd = -(s - s^{-1})^2 abcd,$$

and, since $abcd \neq 0$, at most 4 values of s can give the same value $-(s - s^{-1})^2 abcd$ (the product of the top right and bottom left entries of ((4.5)); for each such value of s, the product and the quotient of the upper left and upper right entries of (4.5) determine r^2 and t^2, respectively, and obviously there are at most 2 values of r and 2 values of t for r^2, t^2 given.

Now, there are at most 4 values of s such that $s^{-1} - s = 0$ or $sad - s^{-1}bc = 0$. Hence,

$$|\phi(A \cap T(K), A \cap T(K), A \cap T(K))| \geq \frac{1}{16}|A \cap T(K)|(|A \cap T(K)| - 4)|A \cap T(K)|,$$

and, at the same time, $\phi(A \cap T(K), A \cap T(K), A \cap T(K)) \subset AA^\ell AA^{-\ell}A = A^{3+2\ell}$, as we said before. If $|A \cap T(K)|$ is less than 8 (or any other constant), conclusion (4.3) is trivial. Therefore,

$$|A \cap T(K)|^3 \leq 2|A \cap T(K)|(|A \cap T(K)| - 4)|A \cap T(K)| \leq 32|A^{2\ell+3}|,$$

i.e., (4.3) holds.

It only remains to verify that there exists an element (4.4) of A^ℓ with $abcd \neq 0$. Now, $abcd = 0$ defines a subvariety W of $\mathbb{A}^4$, where $\mathbb{A}^4$ is identified with the space of 2-by-2 matrices. Moreover, for $|K| > 2$, there are elements of $G(K)$ outside that variety. Hence, the conditions of Prop. 21 hold (with $x = e$). Thus, we obtain that there is a $g \in A^\ell$ (ℓ a constant) such that $g \notin W(K)$, and that was what we needed. □

Let us abstract the essence of what we have just done, so that we can then generalize the result to an arbitrary variety V instead of working just with T. For the sake of convenience, we will do the case $\dim V = 1$, which is, at any rate, the case we will need. The strategy of the proof of Lemma 23 is to construct a morphism $\phi : V \times V \times \cdots \times V \to G$ (r copies of V, where $r = \dim(G)$) of the form

$$\phi(v_1, \dots, v_r) = v_1 g_1 v_2 g_2 \cdots v_{r-1} g_{r-1} v_r, \tag{4.6}$$

where $g_1, g_2, \dots, g_{r-1} \in A^\ell$, in such a way that, for $v = (v_1, \dots, v_r)$ generic (that is, outside a subvariety of $V \times \cdots \times V$ of positive codimension), the preimage $\phi^{-1}(\{\phi(v)\})$ has dimension 0. Actually, as we have just seen, it is enough to prove that this true for $(g_1, g_2, \dots, g_{r-1})$ a generic element of G^{r-1}; the escape argument (Prop. 21) takes care of the rest.

In order to make the argument work nicely for V general (and G general), we need to assume some background. In essence, we have the choice of either working over the Lie algebra or assuming a little more algebraic geometry. The first choice was the one taken in [Hel15], following [Hel11]; the second one is closer to [Tao15], which follows [BGT11]. It really does not much matter.

Here, we choose to assume some passing familiarity with both formalisms. This should make the picture clear quickly. Let us review what sort of background material we will use. It is more than enough if the reader is familiar with these matters over $\mathbb{R}$ and $\mathbb{C}$, and is willing to believe assurances that they work out in much the same way over a finite field. As some readers will know, the fact that these assurances are true was established in the first half of the XXth century (Zariski, Chevalley, etc.).

Some algebraic geometry. It is clear that, if $\phi : \mathbb{A}^n \to \mathbb{A}^m$ is a morphism (that is, a map $(x_1, \dots, x_n) \to (P_1(x_1, \dots, x_n), \dots, P_m(x_1, \dots, x_n))$ given by polynomials $P_1, P_2, \dots, P_m$) and $V \subset \mathbb{A}^m$ is a variety, then the preimage $\phi^{-1}(V)$ is a variety. What is not at all evident a priori is that, if ϕ is as we said and $V \subset \mathbb{A}^n$ is a variety, then $\phi(V)$ is a *constructible set*, meaning a finite union of terms of the form $W \setminus W'$, where W and $W' \subset W$ are varieties. (For instance, if $V \subset \mathbb{A}^2$ is the variety given by $x_1 x_2 = 1$ (a hyperbola), then its image under the morphism $\phi(x_1, x_2) = x_1$ is the constructible set $\mathbb{A}^1 \setminus \{0\}$.) This is a theorem of Chevalley's [Mum99, §I.8, Cor. 2]; it encapsulates some of elimination theory under its polished surface. It is an immediate corollary that, for V constructible, $\phi(V)$ is constructible.

We can always express a constructible S as a union $\cup_i (W_i \setminus W_i')$ with $\dim(W_i') < \dim(W_i)$. The *Zariski closure* $\overline{S}$ of S is nothing other than $\cup_i W_i$. If V is a variety, ϕ a morphism and $\overline{\phi(V)} = \cup_i W_i$, then $\max_i \dim(W_i) \leq \dim(V)$. It is possible to bound $\sum_i \deg(W_i)$ solely in terms of the degrees of the polynomials defining f and the number and degrees of the components of V.

Given an irreducible variety V, there is a possibly empty subvariety $W \subsetneq V$ such that, for every point x on V not on W, there is a well-defined tangent space $T_x G$ of G at x; it is a linear space of dimension equal to $\dim(V)$. The tangent space of G at x is defined over an arbitrary field K just as it is defined over $\mathbb{R}$ or $\mathbb{C}$: it is the intersection of the kernels of the derivatives $DP : K^n \to K$ of all polynomials P in the system of equations $P(x_1, \dots, x_n) = 0$ defining V. There is no problem with taking derivatives over an arbitrary field here: these are all polynomials, and thus their derivatives can be taken formally.

In general, if $f : V \to V'$ is a morphism of varieties (that is, f is given by polynomial maps), we can take the derivative Df_x of f at any point x on V; it is a linear map from the tangent space $T_x V$ to the tangent space to $T_{f(x)} V'$. Over $\mathbb{R}$ or $\mathbb{C}$, if the derivative at x of a map $f : V \to V'$ is a non-singular linear map, then f is injective when restricted to some neighborhood of x. Something close to this is true over an arbitrary field, with respect to the *Zariski topology*. In very concrete terms: for V of dimension m, the linear map Df_x is singular exactly when its m-by-m minors vanish; thus, Df_x vanishes exactly on a subvariety W of V (which may be all of V), called the variety of singular points of f. For every point y on V', the intersection of the preimage $f^{-1}(V')$ with $V \setminus W$ is a zero-dimensional variety; its

degree d, and thus its number of points, is bounded by the degree of the polynomials defining f, and by the number and degrees of the components of W. (In fact, for y generic, d is constant, and we can define the *degree* of f to be d.)

A little on groups of Lie type. The group $\mathrm{SL}_2(\mathbb{R})$ is a Lie group, i.e., a group G that is also a differentiable manifold, with the group operation and inversion being smooth. Its *Lie algebra* $\mathfrak{g}$ is its tangent space at the point that is the group identity e; this tangent space is endowed with an operation that we will describe in a moment. The tangent space at any other point x can be taken to the tangent space at e (i.e., the Lie algebra) simply by composition with multiplication by x^{-1}.

The map $ghg^{-1}h^{-1}$ from $G \times G$ to G sends the identity to the identity; thus, its derivative at the identity e is a linear map $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. We denote it by $(x, y) \mapsto [x, y]$ and call it the *Lie bracket*. The Lie bracket obeys certain rules besides bilinearity, e.g., $[x, x] = 0$. The Lie group G acts on the Lie algebra $\mathfrak{g}$ by conjugation: for $g \in G$, the linear map $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is the derivative of the map $h \mapsto ghg^{-1}$ from G to G at e. The derivative of $h \mapsto ghg^{-1}h^{-1}$ is thus $\mathrm{Ad}_g - I$. Thus, we easily see that, for $y \in \mathfrak{g}$ given, the derivative of $g \mapsto \mathrm{Ad}_g(y)$ at e is just $x \mapsto [x, y]$.

An *ideal* I of a Lie algebra $\mathfrak{g}$ is a subspace I of $\mathfrak{g}$ such that $[\mathfrak{g}, I] \subset I$. By the above, if a subspace I is invariant under Ad_g for every $g \in G$, then it is an ideal.

A Lie group G is *almost simple* if every normal subgroup of G is either finite or of finite index; for instance, $\mathrm{SL}_2(\mathbb{R})$ is simple, and has $\{-1, 1\}$ as a normal subgroup. A Lie algebra $\mathfrak{g}$ is *simple* if it has no ideals other than 0 and $\mathfrak{g}$. A Lie group G is almost simple if and only if its Lie algebra is simple. This is not hard to establish, thanks to the fact that we can pass from $\mathfrak{g}$ to G by means of an *exponentiation map*. (Of course, we can pass from G to $\mathfrak{g}$ by taking derivatives.)

A *group of Lie type* is a reductive linear algebraic group G defined over a finite field $K = \mathbb{F}_q$. (There is no need to define "reductive" here; suffice it to say that every almost-simple algebraic group is reductive.) The point here is that they behave very much like Lie groups. This is sometimes clear and sometimes far from immediate, in that we no longer have an exponentiation map for G, $\mathfrak{g}$ general. See [Spr98] and [Hum81] as general references, each of which has much more material than what we will actually need.

We define the Lie bracket just above; it still endows the tangent space at the origin with the structure of a Lie algebra over K.

A group of Lie type G is *almost simple* if every normal algebraic subgroup of G is of dimension 0 or $\dim(G)$. If G is almost simple, then its Lie algebra $\mathfrak{g}$ is simple; this is proven just as over $\mathbb{R}$ or $\mathbb{C}$. Because we no longer have a map exp defined on all of $\mathfrak{g}$, going in the other direction is harder, and in fact there are exceptions. See [Hog82]. To summarize: for $G = \mathrm{SL}_n$, the Lie algebra $\mathfrak{g} = \mathfrak{sl}_n$ is simple provided that the characteristic p of the field K does not divide n. (The problem here comes mainly from the diagonal matrix I, whose trace n equals 0 in K when $p|n$. It is a non-trivial element of the center of $\mathfrak{g} = \mathfrak{sl}_n$, and thus its multiples are an ideal of $\mathfrak{g}$.) For almost simple Lie groups G such that $\mathfrak{g}$ is not isomorphic to $\mathfrak{sl}_n$, we have that $\mathfrak{g}$ is simple provided that $\mathrm{char}(K) > 3$ [Hog82, Table 1].

In spite of this small-characteristic phenomenon, we will nevertheless work over Lie algebras whenever possible, as then matters arguably become particularly clear

and straightforward. The following lemma is the same as [Tao15, Prop. 1.5.30], which, in turn, is the same as [LP11, Lemma 4.5]. We will give a proof valid for $\mathfrak{g}$ simple.

Lemma 24. *Let $G \subset \mathrm{SL}_n$ be an almost-simple algebraic group defined over a finite field K. Let $V, V' \subsetneq G$ be irreducible subvarieties with $\dim(V) < \dim(G)$ and $\dim(V') > 0$. Then, for every $g \in G(\overline{K})$ outside a subvariety $W \subsetneq G$, some component of $\overline{VgV'}$ has dimension $> \dim(V)$.*

Moreover, the number and degrees of the irreducible components of W are bounded by a constant that depends only on n and the number and degree of components of $\deg(G)$.

Proof for $\mathfrak{g}$ simple. We can assume without loss of generality – displacing everything by multiplication, if necessary – that V and V' go through the origin, and that the origin is a non-singular point for V and V'. Let $\mathfrak{v}$ and $\mathfrak{v}'$ be the tangent spaces to V and V' at the origin. The tangent space to $\overline{VgV'}g^{-1}$ at the identity is $\mathfrak{v} + \mathrm{Ad}_g \mathfrak{v}'$. Thus, for $\overline{V'gV}$ to have a component of dimension $> \dim(V)$, it is enough that $\mathfrak{v} + \mathrm{Ad}_g \mathfrak{v}'$ have dimension greater than $\dim(\mathfrak{v})$.

Suppose that this is not the case for any g on G. Then the space $\mathfrak{w}$ spanned by all spaces $\mathrm{Ad}_g \mathfrak{v}'$, for all g, is contained in $\mathfrak{v}$. Since $\dim(V) < \dim(G)$, $\mathfrak{v} \subsetneq \mathfrak{g}$. Clearly, $\mathfrak{w}$ is non-empty and invariant under Ad_g for every g. Hence it is an ideal. However, we are assuming $\mathfrak{g}$ to be simple. Contradiction.

Thus, $\mathfrak{v} + \mathrm{Ad}_g \mathfrak{v}'$ has dimension greater than $\dim(\mathfrak{v})$ for some g. It is easy to see that the points g where that isn't the case are precisely those such that all $\dim(\mathfrak{v}) \times \dim(\mathfrak{v})$ minors of a matrix – with entries polynomial on the entries of g – vanish. Hence, the points where $\dim(\mathfrak{v} + \mathrm{Ad}_g \mathfrak{v}') > \dim(\mathfrak{v})$ are the points outside a variety $W \subsetneq V$ given by a system of equations whose degree and number is bounded in terms of n and the number and degree of components of $\deg(G)$. □

We can now generalize our proof of Lemma 23, and thus prove (4.2) for all varieties of dimension 1. Before we start, we need a basic counting lemma.

Exercise 25. *Let $W \subset \mathbb{A}^n$ be a variety defined over K such that every component of W has dimension $\leq d$. Let S be a finite subset of K. Then the number of points $(x_1, \dots, x_n) \in S \times S \times \cdots \times S$ (n times) lying on W is $\ll |S|^d$, where the implied constant depends only on n and on the number and degrees of the components of W.*
Sketch/hints: *we can assume without loss of generality that W is irreducible. We will reduce the problem either to a case with lower d, or to a case with the same d and lower n; we will work with projections, so degrees will keep themselves bounded. Let $\pi : V \to \mathbb{A}^{n-1}$ be the projection to the first $n-1$ coordinates. The derivative $D\pi$ is singular on a subvariety W of V. If $W = V$, then $\dim(W) < \dim(V)$; we have reduced the problem to one of lower d and n, and finish matters by the trivial fact that at most $|S|$ points in $S \times \cdots \times S$ (n times) can be projected to a single point by π. If $W \subsetneq V$, then, since V is irreducible, $\dim(W) < \dim(V)$. Bound separately the number of points on $W(K)$ (lower d) and the number of points on $V(K) \setminus W(K)$; the latter is at most the degree of π (bounded by the degree of W) times the number of points on $\overline{\pi(V(K))}$ (lower n).*

Proposition 26. *Let K be a finite field. Let $G \subset \mathrm{SL}_n$ be an almost-simple, irreducible algebraic group such that $|G(K)| \geq c|K|^{\dim(G)}$, $c > 0$. Let $Z \subset G$ be a variety of dimension 1. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Then

$$|A \cap Z(K)| \ll |(A \cup A^{-1} \cup \{e\})^k|^{1/\dim(G)}, \tag{4.7}$$

where k and the implied constant depend only on n, c, $\deg(G)$ and the number and degrees of the irreducible components of Z.

Obviously, $G = \mathrm{SL}_n$ is a valid choice, since it is almost-simple and $|\mathrm{SL}_n(K)| \gg |K|^{n^2-1} = |K|^{\dim(G)}$.

Proof. We will use Lemma 24 repeatedly. When we apply it, we get a subvariety $W \subsetneq G$ such that, for every g outside W, some component of $\overline{VgV'}$ has dimension $> \dim(V)$ (where V and V' are varieties satisfying the conditions of Lemma 24). Since G is irreducible, W is of positive codimension, i.e., every component of W has dimension less than $\dim(G)$. By Exercise 25 (with $S = K$) and the assumption $|G(K)| \geq c|K|^{\dim(G)}$, there is at least one point of $G(K)$ not on W, provided that $|K|$ is larger than a constant, as we can indeed assume. Hence, we can use escape from subvarieties (Prop. 21) to show that there is a $g \in (A \cup A^{-1} \cup \{e\})^\ell$, where ℓ depends only on the number and degrees of components of W, that is to say – by Lemma 24 – only on n and $\deg(G)$.

So: first, we apply Lemma 24 with $V = V' = Z$; we obtain a variety $V_2 = \overline{Vg_1V'} = \overline{Zg_1Z}$ with $g \in (A \cup A^{-1} \cup \{e\})^\ell$ such that V_2 has at least one component of dimension 2. (We might as well assume V is irreducible from now on; then V_2 is irreducible.) We apply Lemma 24 again with $V = V_2$, $V' = Z$, and obtain a variety $V_3 = \overline{V_2g_2Z} = \overline{Zg_1Zg_2Z}$ of dimension 3. We go on and on, and get that there are $g_1, \dots, g_{m-1} \in (A \cup A^{-1} \cup \{e\})^{\ell'}$, $r = \dim(G)$, such that $\overline{Zg_1Zg_2 \dots Zg_{r-1}Z}$ has dimension r.

This means that the variety W of singular points of the map f from $Z^r = Z \times Z \times \dots \times Z$ (r times) to G given by

$$f(z_1, \dots, z_m) = z_1g_1z_2g_2 \dots z_{r-1}g_{r-1}z_r$$

cannot be all of G. Thus, since G is irreducible, W is of positive codimension in V. Again by Exercise 25 (with $S = A \cap Z(K)$), at most $O(|A \cap Z(K)|^{r-1})$ points of $(A \times Z(K)) \times \dots \times (A \times Z(K))$ (r times) on W. The number of points of $(A \times Z(K)) \times \dots \times (A \times Z(K))$ not on W is at most the degree of f times the number of points on $f(A \times Z(K), \dots, A \times Z(K))$, which is contained in A^k for $k = r + (r-1)\ell'$. Check this means we are done. □

In general, one can prove (4.2) for $\dim(V)$ arbitrary using very similar arguments, together with an induction on the dimension of the variety V in (4.2). We will demonstrate the basic procedure doing things in detail for $G = \mathrm{SL}_2$ and for the kind of variety V for which we really need to prove estimates.

We mean the variety V_t defined by

$$\det(g) = 1, \mathrm{tr}(g) = t \tag{4.8}$$

for $t \neq \pm 2$. Such varieties are of interest to us because, for any regular semisimple $g \in \mathrm{SL}_2(K)$ (meaning: any matrix in SL_2 having two distinct eigenvalues), the conjugacy class $\mathrm{Cl}(g)$ is contained in $V_{\mathrm{tr}(g)}$.

Proposition 27. *Let K be a field. Let $A \subset \mathrm{SL}_2(K)$ be a set of generators of $\mathrm{SL}_2(K)$. Let V_t be given by (4.8).*

Then, for every $t \in K$ other than ± 2,

$$|A \cap V_t(K)| \ll |(A \cup A^{-1} \cup \{e\})^k|^{\frac{2}{3}}, \tag{4.9}$$

where k and the implied constant are absolute.

Needless to say, $\dim(\mathrm{SL}_2) = 3$ y $\dim(V_t) = 2$, so this is a special case of (4.2).

Proof. Consider the map $\phi : V_t(K) \times V_t(K) \to \mathrm{SL}_2(K)$ given by

$$\phi(y_1, y_2) = y_1 y_2^{-1}.$$

It is clear that

$$\phi(A \cap V_t(K), A \cap V_t(K)) \subset A^2.$$

Thus, if ϕ were injective, we would obtain immediately that $|A \cap V_t(K)|^2 \leq |A^2|$. Now, ϕ is not injective, not even nearly so. The preimage of $\{h\}$, $h \in \mathrm{SL}_2(K)$, is

$$\phi^{-1}(\{h\}) = \{(w, h^{-1}w) : \mathrm{tr}(w) = t, tr(h^{-1}w) = t\}.$$

We should thus ask ourselves how many elements of A lie on the subvariety $Z_{t,h}$ of G defined by

$$Z_{t,h} = \{(w, hw) : \mathrm{tr}(w) = t, tr(h^{-1}w) = t\}.$$

For $h \neq \pm e$, $\dim(Z_{t,h}) = 1$, and the number and degrees of irreducible components of $Z_{t,h}$ are bounded by an absolute constant. Thus, applying Proposition 26, we get that, for $h \neq \pm e$,

$$|A \cap Z_{t,h}(K)| \ll |A^{k'}|^{1/3},$$

where k' and the implied constant are absolute.

Now, for every $y_1 \in V_t(K)$, there are at least $|V_t(K)| - 2$ elements $y_2 \in V_t(K)$ such that $y_1 y_2^{-1} \neq \pm e$. We conclude that

$$|A \cap V(K)|(|A \cap V(K)| - 2) \leq |A^2| \cdot \max_{g \neq \pm e} |A \cap Z_{t,h}(K)| \ll |A^2||A^{k'}|^{1/3}.$$

We can assume that $|A \cap V(K)| \geq 3$, as otherwise the desired conclusion is trivial. We obtain, then, that

$$|A \cap V(K)| \ll |A^k|^{2/3}$$

for $k = \max(2, k')$, as we wanted. □

Now we can finally prove the result we needed.

Corollary 28. *Let $G = \mathrm{SL}_2$, K a field. Let A be a set of generators of $G(K)$; let $g \in A^\ell$ ($\ell \geq 1$) be regular semisimple. Then*

$$|A^{-1}A \cap C(g)| \gg \frac{A}{|(A \cup A^{-1} \cup \{e\})^{k\ell}|^{2/3}}, \tag{4.10}$$

where k and the implied constant are absolute.

In particular, if $|A^3| \leq |A|^{1+\delta}$*, then*

$$|A^{-1}A \cap C(g)| \gg_\ell |A|^{1/3-O(\delta\ell)}. \tag{4.11}$$

Proof. Proposition 21 and Lemma 6 imply (4.10) immediately, and (4.11) follows readily from (4.10) via (2.4). □

Let us now see two problems whose statements we will not use; they are, however, essential if one wishes to work in SL_n for n arbitrary, or in an arbitrary almost-simple group of Lie type. The first problem is challenging, but we have already seen and applied the main ideas involved in its solution. In essence, it is a matter of setting up a recursion properly.

Exercise 29. *Generalize 26 to Z of arbitrary dimension.*

As we said before in passing, an element $g \in \mathrm{SL}_n(K)$ is *regular semisimple* if it has n distinct eigenvalues. For $G = \mathrm{SL}_n$ and $g \in G(K)$, the elements of $C(g)$ are the points $T(K)$ of an abelian algebraic subgroup T of G, called a maximal torus. Just as in the case of SL_2, it is always possible to conjugate by some element of $G(\overline{K})$ so that T becomes simply the group of diagonal matrices. It is thus clear that $\dim(T) = n - 1$; it is also easy to see that $\dim(\overline{\mathrm{Cl}(g)}) = \dim(G) - \dim(T)$, since $\overline{\mathrm{Cl}(g)}$ is just the variety consisting of the matrices in G having the same eigenvalues as g.

Exercise 30. *Generalize 28 to $G = \mathrm{SL}_n$, with g semisimple. Instead of (4.11), the conclusion should be:*

$$|A^{-1}A \cap C(g)| \gg |A|^{\frac{\dim(T)}{\dim(G)}-O(\delta)}, \tag{4.12}$$

where the implied constants depend only on n.

5. Growth in $\mathrm{SL}_2(K)$

5.1. **The case of large subsets.** Let us first see how A grows when $A \subset \mathrm{SL}_2(\mathbb{F}_q)$ is large with respect to $G = \mathrm{SL}_2(\mathbb{F}_q)$. In fact, it is not hard to show that, if $|A| \geq |G|^{1-\delta}$, $\delta > 0$ a small constant, then $(A \cup A^{-1} \cup \{e\})^k = G$, where k is an absolute constant. This I proved back in the day in [Hel08], with my bare hands. We will prove something stronger and nicer: $A^3 = G$. The proof is due to Nikolov and Pyber [NP11]; it is based on a classical idea, brought to bear to this particular context by Gowers [Gow08]. It will give us the opportunity to revisit the adjacency operator $\mathscr{A}$ and its spectrum.

Recall that a *complex representation* of a group G is just a homomorphism $\phi : G \to \mathrm{GL}_d(\mathbb{C})$; by the *dimension* of the representation we just mean d. A representation ϕ is *trivial* if $\phi(g) = e$ for every $g \in G$.

The following result is due to Frobenius (1896), at least for q prime. It can be proven simply by examining a character table, as in [Sha99]; this also gives analogues of the same result for other groups of Lie type. Alternatively, there is a very nice elementary proof for q prime, to be found, for example, in [Tao15, Lemma 1.3.3].

Proposition 31. *Let $G = \mathrm{SL}_2(\mathbb{F}_q)$, $q = p^\alpha$. Then every non-trivial complex representation of G has dimension $\geq (q-1)/2$.*

We recall that the adjacency operator $\mathscr{A}$ on a Cayley graph $\Gamma(G, A)$ is the linear operator that takes a function $f : V \to \mathbb{C}$ to the function $\mathscr{A}f : V \to \mathbb{C}$ given by

$$\mathscr{A}f(g) = \frac{1}{|A|} \sum_{a \in A} f(ag).$$

By an *eigenspace* of $\mathscr{A}$ we mean, of course, the vector space consisting of functions f such that $\mathscr{A}f = \nu f$ for some fixed eigenvalue ν. It is clear from the definition that every eigenspace of $\mathscr{A}$ is invariant under the action of G by multiplication on the right. In other words, it is a complex representation of G – and it can be trivial only if it is the one-dimensional space consisting of constant functions, i.e., the eigenspace corresponding to the eigenvalue $\nu_0 = 1$. Hence, by Prop. 31, all other eigenvalues have multiplicity Assume, as we often do, that $A = A^{-1}$; this implies that $\mathscr{A}$ is symmetric and all its eigenvalues are real:

$$\dots \leq \nu_2 \leq \nu_1 \leq \nu_0 = 1.$$

The idea now is to obtain a spectral gap, i.e., a non-trivial upper bound on ν_j, $j > 0$. It is standard to use the fact that the trace of a power $\mathscr{A}^r$ of an adjacency operator $\mathscr{A}$ can be expressed in two ways: as a the number of cycles of length r in the graph $\Gamma(G, A)$ (multiplied by $1/|A|^r$), and as the sum of the rth powers of the eigenvalues of $\mathscr{A}$. In our case, for $r = 2$, this gives us

$$\frac{|G||A|}{|A|^2} = \sum_j \nu_j^2 \geq \frac{q-1}{2} \nu_j^2, \tag{5.1}$$

for any $j \geq 1$, and hence

$$|\nu_j| \leq \sqrt{\frac{|G|/|A|}{(q-1)/2}}. \tag{5.2}$$

This is a very low upper bound when $|A|$ is large. This means that a few applications of the operator $\mathscr{A}$ are enough to render any function almost uniform, since any component orthogonal to the space of constant functions is multiplied by some ν_j, $j \geq 1$, at every step. The following proof puts in practice this observation efficiently.

Proposition 32 ([NP11]). *Let $G = \mathrm{SL}_2(\mathbb{F}_q)$, $q = p^\alpha$. Let $A \subset G$; assume $|A| \geq 2|G|^{8/9}$. Then*

$$A^3 = G.$$

Proof. We will assume $A = A^{-1}$, as usual; thanks to [Gow08], essentially the same argument works in the case $A \neq A^{-1}$.

Suppose there is a $g \in G$ such that $g \notin A^3$. Then the scalar product

$$\langle \mathscr{A} 1_A, 1_{gA} \rangle = \sum_{x \in G} (\mathscr{A} 1_A)(x) \cdot 1_{gA^{-1}}(x)$$

equals 0. We may assume that the eigenvectors v_j satisfy $\langle v_j, v_j\rangle = 1$. Then

$$\begin{aligned}\langle \mathscr{A} 1_A, 1_{gA}\rangle &= \langle \sum_{j\geq 0} \nu_j \langle 1_A, v_j\rangle v_j, 1_{gA}\rangle \\ &= \nu_0 \langle 1_A, v_0\rangle \langle v_0, 1_{gA^{-1}}\rangle + \sum_{j>0} \nu_j \langle 1_A, v_j\rangle \langle v_j, 1_{gA^{-1}}\rangle.\end{aligned}$$

Now, v_0 is a constant function satisfying $\langle v_0, v_0\rangle = 1$; thus, it equals $1/\sqrt{|G|}$ everywhere. Hence

$$\nu_0 \langle 1_A, v_0\rangle \langle v_0, 1_{gA^{-1}}\rangle = 1 \cdot \frac{|A|}{\sqrt{|G|}} \cdot \frac{|g^{-1}A|}{\sqrt{|G|}} = \frac{|A|^2}{|G|}.$$

At the same time, by (5.2) and Cauchy-Schwarz,

$$\begin{aligned}\left|\sum_{j>0} \nu_j \langle 1_A, v_j\rangle \langle v_j, 1_{gA^{-1}}\rangle\right| &\leq \sqrt{\frac{2|G|/|A|}{q-1}} \sqrt{\sum_{j\geq 1} |\langle 1_A, v_j\rangle|^2} \sqrt{\sum_{j\geq 1} |\langle v_j, 1_{gA^{-1}}\rangle|^2} \\ &\leq \sqrt{\frac{2|G|/|A|}{q-1}} |1_A|_2 |1_{gA^{-1}}|_2 = \sqrt{\frac{2|G||A|}{q-1}}.\end{aligned}$$

Since $|G| = (q^2-q)q$, we see that $|A| \geq 2|G|^{8/9}$ implies

$$\frac{|A|^2}{|G|} > \sqrt{\frac{2|G||A|}{q-1}},$$

and thus $\langle \mathscr{A} 1_A, 1_{gA^{-1}}\rangle > 0$. Contradiction. □

5.2. **Growth in $\mathrm{SL}_2(K)$, K arbitrary.** We finally come to the proof of our main result. Here we will be closer to newer treatments (in particular, [PS16]) than to what was the first proof, given in [Hel08]; these newer versions generalize more easily. We will give the proof only for SL_2, and point out the couple of places in the proof where one would has to be especially careful when generalizing matters to SL_n, $n > 2$, or other groups of Lie type.

The proof in [Hel08] used the sum-product theorem (Thm. 16). We will not use it, but the idea of "pivoting" will reappear. It is also good to note that, just as before, there is an inductive process here, carried out on a group G, even though G does not have a natural order $(1, 2, 3, \dots)$. All we need for the induction to work is a set of generators A of G.

Proposition 33 (Helfgott [Hel08])**.** *Let K be a field. Let $A \subset \mathrm{SL}_2(K)$ generate $\mathrm{SL}_2(K)$. Assume $|A| < |\mathrm{SL}_2(K)|^{1-\epsilon}$, $\epsilon > 0$. Then*

$$|A^3| \ll |A|^{1+\delta},$$

where $\delta \gg \epsilon$ and both implied constants are absolute.

Actually, [Hel08] proved this for $K = \mathbb{F}_p$; the first generalization to a general finite field K was given by [Din11]. The proof we are about to see works for K general without any extra effort. It works, incidentally, for K infinite as well, dropping the

condition $|A| < |\mathrm{SL}_2(K)|^{1-\epsilon}$, which becomes trivially true. The case of characteristic 0 is actually easier than the case $K = \mathbb{F}_p$ the proof in [Hel08] was already valid for $K = \mathbb{R}$ or $K = \mathbb{C}$, say. However, for applications, the "right" result for $K = \mathbb{R}$ or $K = \mathbb{C}$ is not really Prop. 33, but a statement counting how many elements there can be in A and $A \cdot A \cdot A$ that are separated by a given small distance from each other; that was proven in [BG08a], adapting the techniques in [Hel08].

Proof. We may assume that $|A|$ is larger than an absolute constant, since otherwise the conclusion would be trivial. Let $G = \mathrm{SL}_2$.

Suppose that $|A^3| < |A|^{1+\delta}$, where $\delta > 0$ is a small constant to be determined later. By escape (Prop. 21), there is an element $g_0 \in A^c$ that is regular semisimple (that is, $\mathrm{tr}(g_0) \neq \pm 2$), where c is an absolute constant. (Easy exercise: show we can take $c = 2$.) Its centralizer in $G(K)$ is $C(g) = T(\overline{K}) \cap G(K)$ for some maximal torus T.

Call $\xi \in G(K)$ a *pivot* if the map $\phi_g : A \times C(g) \to G(K)$ defined by

$$(a, t) \mapsto a\xi t\xi^{-1} \tag{5.3}$$

is injective as a function from $\pm e \cdot A/\{\pm e\} \times C(g)/\{\pm e\}$ to $G(K)/\{\pm e\}$.

Case (a): There is a pivot ξ in A. By Corollary 28, there are $\gg |A|^{1/3-O(c\delta)}$ elements of $C(g)$ in $A^{-1}A$. Hence, by the injectivity of ϕ_ξ,

$$\left|\phi_\xi(A, A^{-1}A \cap C(g))\right| \geq \frac{1}{4}|A||A^{-1}A \cap C(g)| \gg |A|^{\frac{4}{3}-O(c\delta)}.$$

At the same time, $\phi_\xi(A, A^{-1}A \cap C(g)) \subset (A \cup A^{-1})^5$, and thus

$$|(A \cup A^{-1})^5| \gg |A|^{4/3-O(c\delta)}.$$

For $|A|$ larger than a constant and $\delta > 0$ less than a constant, this gives us a contradiction with $|A^3| < |A|^{1+\delta}$ (by Ruzsa (2.3)).

Case (b): There are no pivots ξ in $G(K)$. Then, for every $\xi \in G(K)$, there are $a_1, a_2 \in A$, $t_1, t_2 \in T(K)$, $(a_1, t_1) \neq (\pm a_2, \pm t_2)$ such that $a_1\xi t_1\xi^{-1} = \pm e \cdot a_2\xi t_2\xi^{-1}$, and that gives us that

$$a_2^{-1}a_1 = \pm e \cdot \xi t_2 t_1^{-1}\xi^{-1}.$$

In other words, for each $\xi \in G(K)$, $A^{-1}A$ has a non-trivial intersection with the torus $\xi T\xi^{-1}$:

$$A^{-1}A \cap \xi T(K)\xi^{-1} \neq \{\pm e\}. \tag{5.4}$$

(Note this means that case (b) never arises for K infinite. Why?)

Choose any $g \in A^{-1}A \cap \xi T(K)\xi^{-1}$ with $g \neq \pm e$. Then g is regular semisimple (note: this is peculiar to SL_2; this is the place in the proofs that requires some work when you generalize it to other groups). The centralizer $C(g)$ equals $\xi T(K)\xi^{-1}$ (why?). Hence, by Corollary 28, we obtain that there are $\geq c'|A|^{1/3-O(\delta)}$ elements of $\xi T(K)\xi^{-1}$ en in $A^{-1}A$, where c' and the implied constant are absolute.

At least $(1/2)|G(K)|/|T(K)|$ maximal tori of G are of the form $\xi T\xi^{-1}$, $\xi \in G(K)$ (check this yourself!). Every element of G that is not $\pm e$ can lie on at most one

maximal torus (again, this is peculiar to SL_2). Hence

$$|A^{-1}A| \geq \frac{1}{2}\frac{|G(K)|}{|T(K)|}(c|A|^{1/3-O(\delta)} - 2) \gg q^2|A|^{1/3-O(\delta)}.$$

Therefore, either $|A^{-1}A| > |A|^{1+2\delta}$ (say) or $|A| \geq |G|^{1-O(\delta)}$. In the first case, Ruzsa's distance inequality (Lemma 2) with $B = A^{-2}$ and $C = A$ gives us that $|A^3| > |A|^{1+\delta}$, in contradiction to what we were assuming. If we are in the second case, Proposition 32 implies that $A^3 = G$.

Case (c): There are pivots and non-pivots in $G(K)$. Since $\langle A \rangle = G(K)$, this implies that there exists a non-pivot $\xi \in G$ and an $a \in A$ such that $a\xi \in G$ is a pivot. Since ξ is not a pivot, (5.4) holds, and thus there are $|A|^{1/3-O(\delta)}$ elements of $\xi T \xi^{-1}$ in A^k.

At the same time, $a\xi$ is a pivot, i.e., the map $\phi_{a\xi}$ defined in (5.3) is injective (considered as an application from $A/\{\pm e\} \times C(g)/\{\pm e\}$ to $G(K)/\{\pm e\}$). Therefore,

$$\left|\phi_{a\xi}(A, \xi^{-1}(A^k \cap \xi T \xi^{-1})\xi)\right| \geq \frac{1}{4}|A||A^k \cap \xi T \xi^{-1}| \geq \frac{1}{4}|A|^{\frac{4}{3}-O(\delta)}.$$

Since $\phi_{a\xi}(A, \xi^{-1}(A^k \cap \xi T \xi^{-1})\xi) \subset A^{k+3}$, we obtain that

$$|A^{k+3}| \geq \frac{1}{4}|A|^{4/3-O(\delta)}. \tag{5.5}$$

Thanks again to Ruzsa (2.3), this contradicts $|A^3| \leq |A|^{1+\delta}$ for $\delta > 0$ smaller than a constant. □

Putting Prop. 32 and Prop. 33 together, we obtain our main result.

Theorem 34. *Let K be a finite field. Let $G = \mathrm{SL}_2(K)$, K a finite field. Let $A \subset G$ generate G. Then either*

$$|A^3| \geq |A|^{1+\delta}$$

or

$$A^3 = G,$$

where $\delta > 0$ is an absolute constant.

Exercise 35. *Let K be a finite field. Let $G = \mathrm{SL}_2(K)$, K a finite field. Let $A \subset G$ generate G. Using Thm. 34, prove that the diameter of $\Gamma(G, A)$ is $\ll (\log |G|)^{O(1)}$, where the implied constants are absolute.*

6. Further perspectives and open problems

6.1. **Generalizations.** Theorem 34 (with $(A \cup A^{-1} \cup \{e\})^{O(1)} = G$ instead of $A^3 = G$) and the statement of Exercise 35 were first proven for $K = \mathbb{Z}/p\mathbb{Z}$ in [Hel08]. The way to more general statements was gradual. First there was a generalization [BG08a] to the group SU(2); its Lie algebra is isomorphic to $\mathfrak{sl}_2$ over $\mathbb{C}$. This was a "strong" generalization, i.e., one of sufficient strength to be used to prove a spectral gap (see below). In effect this means that "finite field" was not just changed to "field" in the statement of Thm. 34 (this is easy, and would have given a statement similar to [EK01]) but that the maximal number of points $n_\delta(A)$ separated by δ in the usual

complex metric grows: $n_\delta(A^3) \geq n_\delta(A)^{1+\delta}$. (Over a finite field, such a statement is not needed, nor does it make sense.)

There was then a generalization to general finite K [Din11]; this is automatic in the version of the proof we have seen, but it wasn't so at the time. Generalizing the statement to groups with Lie algebra other than $\mathfrak{sl}_2$ was at first rather difficult. The generalization to $\mathrm{SL}_3(\mathbb{F}_p)$ [Hel11] contained many of the ideas that we have seen here (in particular, estimates on intersections with tori and some other varieties) but got stuck in the way to SL_n in ways that now seem odd (in particular, Corollary 28 was proven for most g, not all g). Here [GH11] was not, in the end, the best way out. Theorem 34 was finally generalized to all groups of Lie type (in particular, SL_n, SO_n, Sp_{2n}, etc.) in [BGT11] and [PS16], independently. The constant δ in $|A^3| \geq |A|^{1+\delta}$ here depends on n. This has to be so, i.e., the inequality cannot be true for A completely arbitrary and an absolute constant $\delta > 0$ independent of n.

This represents the natural reach of the methods here, rather than the ultimate generalization possible. In particular, what happens as $n \to \infty$? It is still believed that the diameter of any Cayley graph of $\mathrm{SL}_n(K)$ (say) should be $\ll (\log |G|)^{O(1)}$ (Babai's conjecture), and not just $\ll (\log |G|)^{O_n(1)}$, but how does one prove that? There is also a closely related, and older, question: what happens in the symmetric group $G = \mathrm{Sym}(n)$? There, too, the diameter of any Cayley graph is supposed to be $\ll (\log |G|)^{O(1)} \ll n^{O(1)}$, by an older folk conjecture. A growth result with a strong dependence on n would be nearly meaningless in $\mathrm{Sym}(n)$.

Here the best known result is [HS14], which states that every Cayley graph of $\mathrm{Sym}(n)$ has diameter

$$\ll e^{O((\log n)^4 \log\log n)} = e^{O_\epsilon\left((\log\log |G|)^{4+\epsilon}\right)}.$$

Any improvement here would be of interest. It is to be hoped that, once the proof of this or stronger bounds becomes sufficiently streamlined, it will help in giving better bounds for the diameter of $\mathrm{SL}_n(K)$. (In fact, some see $\mathrm{Sym}(n)$ as SL_n or PGL_n over the field with one element; no such field exists, but objects over it may.)

Another question is what happens when g_1, g_2 are random elements of a group G. For several kinds of groups (linear algebraic, $\mathrm{Sym}(n)$) it is known that, with probability tending to one, g and h generate G (or a very large subgroup thereof, such as $\mathrm{Alt}(n)$, which is of index 2 in $\mathrm{Sym}(n)$). What is the diameter of the Cayley graph of G with respect to $\{g, h\}$ likely to be? For $G = \mathrm{SL}_2(\mathbb{F}_p)$, it is known that it is $O(\log |G|)$ with probability tending to one (by [GHS+09] taken together with Thm. 34). For $\mathrm{Sym}(n)$, it is known to be $O(n^2 (\log n)^{O(1)})$ with probability tending to one [HSZ15]. Is it actually $O(n(\log n)^{O(1)})$, or even $O(n \log n)$, with probability tending to one?

Lastly, a question that is partly solved. What happens when G is a linear algebraic group that is neither almost-simple nor solvable? It turns out that it is in principle possible to give a unified statement that relies on what we know about the almost-simple and solvable cases. This is what some call the *Helfgott-Lindenstrauss conjecture*; [BGT12] proved a qualitative version of it, and [GH14] proved the original

version (using [PS16]), but only over $\mathbb{F}_p$, not over general finite fields. The problem probably requires a fresh, clean treatment for it to be solved completely.

6.2. **Expansion, random walks and the affine sieve.** Many applications of Thm. 34 go through *expander graphs*. We have already seen the adjacency operator $\mathscr{A}$ on a Cayley graph $\Gamma(G, A)$ and talked about its spectrum

$$\dots \leq \nu_2 \leq \nu_1 \leq \nu_0 = 1,$$

and spoke as well of the spectral gap $n_0 - n_1$. As we said before, a graph $\Gamma(G, A)$ is called an ϵ-*expander* if $\nu_1 - n_0 \geq \epsilon$. An infinite family of graphs $\Gamma(G_i, A_i)$ is called an *expander family* if there is an $\epsilon > 0$ such that every $\Gamma(G_i, A_i)$ is an ϵ-expander. Of particular interest are expander families with $|A_i|$ bounded.

It is a standard fact that an ϵ-expander graph $\Gamma(G, A)$ has diameter $\ll (\log |G|)/\epsilon$; in other words, being an expander is stronger than just having small diameter. Using Thm. 34 (among other tools), Bourgain and Gamburd proved the following result [BG08b].

Theorem 36. *Let $A_0 \subset \mathrm{SL}(\mathbb{Z})$. Assume that A_0 is not contained in any proper algebraic subgroup of SL_2. Then*

$$\{\Gamma(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), A_0 \bmod p)\}_{p>C, p \text{ prime}} \tag{6.1}$$

is an expander family for some constant C.

There are by now wide-ranging generalizations of Thm. 36 (e.g., [**?**]).

A *random walk* on a graph is what it sounds like: we start at a vertex v_0, and at every step we move to one of the d neighbors of the vertex we are at – choosing any one of them with probability $1/d$. For convenience we work with a *lazy random walk*: at every step, we decide to stay where we are with probability $1/2$, and to move to a neighbor with probability $1/2d$. The *mixing time* is the number of steps it takes for ending point of a lazy random walk to become almost equidistributed (where "almost" is understood in any reasonable metric). In an ϵ-expander graph $\Gamma(G, A)$, the mixing time is $O_\epsilon(\log |G|)$, i.e., as small as it could be, qualitatively speaking. (For $|A|$ bounded, the mixing time (and even the diameter) has to be $\gg \log |G|$.)

Thus, Thm. 36 gives us small mixing times. This has made the *affine sieve* possible [BGS10]. This is an analogue of classical sieve methods; they are recast as sieves based on the natural action of $\mathbb{Z}$ on $\mathbb{Z}$, whereas a general affine sieve considers the actions of other groups, such as $\mathrm{SL}_2(\mathbb{Z})$.

Expansion had been shown before for some specific A_0. In particular, when A_0 generates $\mathrm{SL}_2(\mathbb{Z})$ (or a subgroup of finite index before) then the fact that (6.1) is an expander graph can be derived from the *Selberg spectral gap* [Sel65], i.e., the fact that the Laplacian on the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ of the upper half plane $\mathbb{H}$ has a spectral gap. Nowadays, one can go in the opposite direction: spectral gaps on more general quotients can be proven using Thm. 36 [BGS11].

Let us finish this discussion by saying that it is generally held to be plausible that the family of *all* Cayley graphs of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, for all p, is an expander family; in other words, there may be an $\epsilon > 0$ such that, for every prime p and every generator

A of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}$, the graph $\Gamma(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), A)$ is an ϵ-expander. Proving this is an open problem believed to be very hard.

6.3. **Final remarks.** A further discussion of open problems can be found in the associated project notes. Let us briefly mention here some links with other areas.

Group classification. It is by now clear that it is useful to look at a particular kind of result in group classification: the kind that was developed so as to avoid casework, and to do without the Classification of Finite Simple Groups. (The Classification is now generally accepted, but this was not always the case, and it is still sometimes felt to be better to prove something without it than with it; what we are about to see gives itself some validation to this viewpoint.) While results proven without the Classification are sometimes weaker than others, they are also more robust. Classifying subgroups of a finite group G is the same as classifying subsets $A \subset G$ such that $e \in A$ and $|AA| = |A|$. Some Classification-free classification methods can be adapted to help in classifying subsets $A \subset G$ such that $e \in A$ and $|AAA| \leq |A|^{1+\delta}$ – in other words, precisely what we are studying. It is in this way that [LP11] was useful in [BGT11], and [Bab82], [Pyb93] were useful in [HS14].

Model theory. Model theory is essentially a branch of logic with applications to algebraic structures. Hrushovski and his collaborators [HP95], [HW08], [Hru12] have used model theory to study subgroups of algebraic groups. This was influenced by Larsen-Pink [LP11], and also served to explain it. In turn, [Hru12] influenced later work, especially [BGT12].

Permutation-group algorithms. Much work on permutation groups has been algorithmic in nature. Here a standard reference is [Ser03]. A good example is a problem we mentioned before – that of bounding the diameter of $\mathrm{Sym}(n)$ with respect to a random pair of generators; the approach in [BBS04] combines probabilistic and algorithmic ideas – as does [HSZ15], which builds on [BBS04], and as, for that matter, does [HS14]. The reference [LPW09] treats several of the relevant probabilistic tools.

Geometric group theory. Here much work remains to be done. Geometric group theory, while still a relatively new field, is considerably older than the approach followed in these notes. It is clear that there is a connection, but it has not yet been fully explored. Here it is particularly worth remarking that [Hru12] gave a new proof of Gromov's theorem by means of the study of sets A that grow slowly in the sense used in these notes.

References

[Bab82] L. Babai. On the order of doubly transitive permutation groups. *Invent. math.*, 65(3):473–484, 1981/82.

[Bas72] H. Bass. The degree of polynomial growth of finitely generated nilpotent groups. *Proc. London Math. Soc. (3)*, 25:603–614, 1972.

[BBS04] L. Babai, R. Beals, and Á. Seress. On the diameter of the symmetric group: polynomial bounds. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1108–1112 (electronic), New York, 2004. ACM.

[BG08a] J. Bourgain and A. Gamburd. On the spectral gap for finitely-generated subgroups of SU(2). *Invent. math.*, 171(1):83–121, 2008.

[BG08b] J. Bourgain and A. Gamburd. Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$. *Ann. of Math. (2)*, 167(2):625–642, 2008.

[BGK06] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *J. London Math. Soc. (2)*, 73(2):380–398, 2006.

[BGS10] J. Bourgain, A. Gamburd, and P. Sarnak. Affine linear sieve, expanders, and sum-product. *Invent. math.*, 179(3):559–644, 2010.

[BGS11] J. Bourgain, A. Gamburd, and P. Sarnak. Generalization of Selberg's $\frac{3}{16}$ theorem and affine sieve. *Acta Math.*, 207(2):255–290, 2011.

[BGT11] E. Breuillard, B. Green, and T. Tao. Approximate subgroups of linear groups. *Geom. Funct. Anal.*, 21(4):774–819, 2011.

[BGT12] E. Breuillard, B. Green, and T. Tao. The structure of approximate groups. *Publications mathématiques de l'IHÉS*, 116:115–221, 2012.

[BKT04] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.

[Din11] O. Dinai. Growth in SL_2 over finite fields. *J. Group Theory*, 14(2):273–297, 2011.

[DS98] V. I. Danilov and V. V. Shokurov. *Algebraic curves, algebraic manifolds and schemes.* Springer-Verlag, Berlin, 1998. Translated from the 1988 Russian original by D. Coray and V. N. Shokurov, Translation edited and with an introduction by I. R. Shafarevich, Reprint of the original English edition from the series Encyclopaedia of Mathematical Sciences [*Algebraic geometry. I*, Encyclopaedia Math. Sci., 23, Springer, Berlin, 1994; MR1287418 (95b:14001)].

[EK01] Gy. Elekes and Z. Király. On the combinatorics of projective mappings. *J. Algebraic Combin.*, 14(3):183–197, 2001.

[EM03] G. A. Edgar and Ch. Miller. Borel subrings of the reals. *Proc. Amer. Math. Soc.*, 131(4):1121–1129 (electronic), 2003.

[EMO05] A. Eskin, Sh. Mozes, and H. Oh. On uniform exponential growth for linear groups. *Invent. math.*, 160(1):1–30, 2005.

[FKP10] D. Fisher, N. H. Katz, and I. Peng. Approximate multiplicative groups in nilpotent Lie groups. *Proc. Amer. Math. Soc.*, 138(5):1575–1580, 2010.

[Fre73] G. A. Freĭman. *Foundations of a structural theory of set addition.* American Mathematical Society, Providence, R. I., 1973. Translated from the Russian, Translations of Mathematical Monographs, Vol 37.

[GH11] N. Gill and H. A. Helfgott. Growth of small generating sets in $\mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})$. *Int. Math. Res. Not. IMRN*, (18):4226–4251, 2011.

[GH14] N. Gill and H. A. Helfgott. Growth in solvable subgroups of $\mathrm{GL}_r(\mathbb{Z}/p\mathbb{Z})$. *Math. Ann.*, 360(1-2):157–208, 2014.

[GHR15] N. Gill, H. A. Helfgott, and M. Rudnev. On growth in an abstract plane. *Proc. Amer. Math. Soc.*, 143(8):3593–3602, 2015.

[GHS+09] A. Gamburd, S. Hoory, M. Shahshahani, A. Shalev, and B. Virág. On the girth of random Cayley graphs. *Random Structures Algorithms*, 35(1):100–117, 2009.

[GK07] A. A. Glibichuk and S. V. Konyagin. Additive properties of product sets in fields of prime order. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 279–286. Amer. Math. Soc., Providence, RI, 2007.

[Gow08] W. T. Gowers. Quasirandom groups. *Combin. Probab. Comput.*, 17(3):363–387, 2008.

[Gro81] M. Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.

[Gui73] Y. Guivarc'h. Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, 101:333–379, 1973.

[Hel08] H. A. Helfgott. Growth and generation in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. of Math. (2)*, 167(2):601–623, 2008.

[Hel11] H. A. Helfgott. Growth in $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$. *J. Eur. Math. Soc. (JEMS)*, 13(3):761–851, 2011.

[Hel15] H. A. Helfgott. Growth in groups: ideas and perspectives. *Bull. Amer. Math. Soc. (N.S.)*, 52(3):357–413, 2015.

[Hog82] G.M.D. Hogeweij. Almost-classical Lie algebras. I,II. *Indag. Math.*, 44:441–452, 453–460, 1982.

[HP95] E. Hrushovski and A. Pillay. Definable subgroups of algebraic groups over finite fields. *J. Reine Angew. Math.*, 462:69–91, 1995.

[Hru12] E. Hrushovski. Stable group theory and approximate subgroups. *J. Amer. Math. Soc.*, 25(1):189–243, 2012.

[HS14] H. A. Helfgott and Á. Seress. On the diameter of permutation groups. *Ann. of Math. (2)*, 179(2):611–658, 2014.

[HSZ15] H. A. Helfgott, Á. Seress, and A. Zuk. Random generators of the symmetric group: diameter, mixing time and spectral gap. *J. Algebra*, 421:349–368, 2015.

[Hum81] James E. Humphreys. Linear algebraic groups. Corr. 2nd printing. Graduate Texts in Mathematics, 21. New York - Heidelberg - Berlin: Springer-Verlag. XVI, 253 p. DM 72.00; $ 34.30 (1981)., 1981.

[HW08] E. Hrushovski and F. Wagner. Counting and dimensions. In *Model theory with applications to algebra and analysis. Vol. 2*, volume 350 of *London Math. Soc. Lecture Note Ser.*, pages 161–176. Cambridge Univ. Press, Cambridge, 2008.

[Kow13] E. Kowalski. Explicit growth and expansion for SL_2. *Int. Math. Res. Not. IMRN*, (24):5645–5708, 2013.

[LP11] M. J. Larsen and R. Pink. Finite subgroups of algebraic groups. *J. Amer. Math. Soc.*, 24(4):1105–1158, 2011.

[LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[Mil68] J. Milnor. Growth of finitely generated solvable groups. *J. Differential Geometry*, 2:447–449, 1968.

[Mum99] D. Mumford. *The red book of varieties and schemes*, volume 1358 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, expanded edition, 1999. Includes the Michigan lectures (1974) on curves and their Jacobians, With contributions by Enrico Arbarello.

[NP11] N. Nikolov and L. Pyber. Product decompositions of quasirandom groups and a Jordan type theorem. *J. Eur. Math. Soc. (JEMS)*, 13(4):1063–1077, 2011.

[Pet12] G. Petridis. New proofs of Plünnecke-type estimates for product sets in groups. *Combinatorica*, 32(6):721–733, 2012.

[Plü70] H. Plünnecke. Eine zahlentheoretische Anwendung der Graphentheorie. *J. Reine Angew. Math.*, 243:171–183, 1970.

[PS16] L. Pyber and E. Szabó. Growth in finite simple groups of Lie type. *J. Amer. Math. Soc.*, 29(1):95–146, 2016.

[Pyb93] L. Pyber. On the orders of doubly transitive permutation groups, elementary estimates. *J. Combin. Theory Ser. A*, 62(2):361–366, 1993.

[Rot53] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[RT85] I. Z. Ruzsa and S. Turjányi. A note on additive bases of integers. *Publ. Math. Debrecen*, 32(1-2):101–104, 1985.

[Ruz89] I. Z. Ruzsa. An application of graph theory to additive number theory. *Sci. Ser. A Math. Sci. (N.S.)*, 3:97–109, 1989.

[Ruz99] I. Z. Ruzsa. An analog of Freiman's theorem in groups. *Astérisque*, (258):xv, 323–326, 1999. Structure theory of set addition.

[San12] T. Sanders. On the Bogolyubov-Ruzsa lemma. *Anal. PDE*, 5(3):627–655, 2012.

[San13] T. Sanders. The structure theory of set addition revisited. *Bull. Amer. Math. Soc. (N.S.)*, 50(1):93–127, 2013.

[Sel65] A. Selberg. On the estimation of Fourier coefficients of modular forms. In *Proc. Sympos. Pure Math., Vol. VIII*, pages 1–15. Amer. Math. Soc., Providence, R.I., 1965.

[Ser03] Á. Seress. *Permutation Group Algorithms*, volume 152 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003.

[Sha99] Y. Shalom. Expander graphs and amenable quotients. In *Emerging applications of number theory (Minneapolis, MN, 1996)*, volume 109 of *IMA Vol. Math. Appl.*, pages 571–581. Springer, New York, 1999.

[Spr98] T.A. Springer. *Linear algebraic groups. 2nd ed.* Boston, MA: Birkhäuser, 2nd ed. edition, 1998.

[Tao08] T. Tao. Product set estimates for non-commutative groups. *Combinatorica*, 28(5):547–594, 2008.

[Tao10] T. Tao. Freiman's theorem for solvable groups. *Contrib. Discrete Math.*, 5(2):137–184, 2010.

[Tao15] T. Tao. *Expansion in finite simple groups of Lie type*, volume 164 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[Tit72] J. Tits. Free subgroups in linear groups. *J. Algebra*, 20:250–270, 1972.

[Toi14] M. C. H. Tointon. Freiman's theorem in an arbitrary nilpotent group. *Proc. Lond. Math. Soc. (3)*, 109(2):318–352, 2014.

[Wol68] J. A. Wolf. Growth of finitely generated solvable groups and curvature of Riemanniann manifolds. *J. Differential Geometry*, 2:421–446, 1968.