GROWTH IN $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$

H. A. HELFGOTT

Abstract. Let $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$, p a prime. Let A be a set of generators of G. Then A grows under the group operation.

To be precise: denote by $|S|$ the number of elements of a finite set S. Assume $|A| < |G|^{1-\delta}$ for some $\delta > 0$. Then $|A \cdot A \cdot A| > |A|^{1+\epsilon}$, where $\epsilon > 0$ depends only on δ.

We will also study subsets $A \subset G$ that do not generate G. Other results on growth and generation follow.

Contents

1. Introduction 2
1.1. Growth in groups and graphs 2
1.2. Main result 3
1.3. Consequences 4
1.4. Outline 5
1.5. Acknowledgements 7
2. Notation and preliminaries 7
2.1. General notation 7
2.2. Boundedness 7
2.3. Arithmetic combinatorics 8
2.4. Groups and generation 9
2.5. Varieties 9
3. Growth in rings and Borel subgroups 13
3.1. Growth under commuting actions 13
3.2. Growth in unipotent groups under the action of the diagonal 18
3.3. A sum-product theorem in $(\mathbb{Z}/p\mathbb{Z})^n$ 19
3.4. Linear relations over rings 21
4. Escape, non-singularity and their conditions 25
4.1. Escape from subvarieties 25
4.2. Non-singularity and almost-injectivity 27
4.3. Sticking subgroups in generic directions 30
4.4. Examining subspaces at the origin 35
4.5. Subgroups of unipotent subgroups and tori 36
5. Tori and conjugacy classes 38
5.1. The intersection with a maximal torus: an upper bound 39

2000 *Mathematics Subject Classification.* 05C25, 20G40, 20D60, 20F65, 11B75.
Key words and phrases. Cayley graphs, finite groups, generation, diameter, expander graphs.
The author was supported in part by EPSRC grant EP-E054919/1 and NSF grant DMS-0635607.

5.2. A lower bound on the number of conjugacy classes 40
5.3. From conjugacy classes to a maximal torus 44
5.4. An upper bound on the number of conjugacy classes. 46
5.5. Intersections with non-maximal tori 47
5.6. Special tuples of coefficients of characteristic polynomials 49
6. Growth of small and large sets in SL_2 and SL_3 51
6.1. Small sets in SL_2 51
6.2. Small and fairly large sets in SL_3 54
7. Subgroups and solvable groups 57
7.1. Lemmas on growth and subgroups 57
7.2. Lemmas for solvable groups 59
7.3. Examples: growth in Borel subgroups of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ and $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ 63
8. Growth in proper subgroups of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ 64
8.1. Growth in subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$ and $\mathrm{PGL}_2(\mathbb{F}_p) \sim SO_3(\mathbb{F}_p)$ 67
8.2. Parabolic subgroups of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$: general setup 69
8.3. Parabolic subgroups: passage to SL_2, or, the case $\langle\pi_1(\pi_+(E))\rangle = \mathrm{SL}_2(K)$ 71
8.4. Parabolic subgroups: solvable groups, or, the case $\langle\pi_1(\pi_+(E))\rangle = \mathrm{SL}_2(K)$ 75
8.5. Conclusions 76
9. Growth of medium-sized and large sets 79
9.1. Sticking subgroups of SL_3 in different directions 79
9.2. Conclusions 80
10. General conclusions and final remarks 82
10.1. The main theorem, related results and their consequences 82
10.2. Work to do: other groups 85
References 87

1. Introduction

1.1. **Growth in groups and graphs.** "Growth" can mean one of many things.

(a) *Growth in graphs.* Let Γ be a graph. How many vertices can be reached from a given vertex in a given amount of time?

(b) *Growth in infinite groups.* Let A be a set of generators of an infinite group G. Let $B(t)$ be the number of elements that can be expressed as products of at most t elements of A. How does $B(t)$ grow as $t \to \infty$?

(c) *Random walks in groups.* Let A be a set of generators of a finite group G. Start with $x = 1$, and, at each step, multiply x by a random element of A. After how many steps is x close to being equidistributed in G?

(d) *More on growth in graphs: the spectral gap.* Let Γ be a graph. Consider its adjacency matrix. What lower bounds can one give for the difference between its two largest eigenvalues?

(e) *Growth in arithmetic combinatorics.* Let G be an abelian group. Let $A \subset G$. How large is $A + A$ compared to A, and why? In general, let G be a group. Let $A \subset G$. How large[1] is $A \cdot A \cdot A$ compared to A, and why?

Question (e) has been extensively studied in the abelian setting. Some time ago, I started studying it for non-abelian groups, and proved [He] that every set of generators A of $G = \mathrm{SL}_2(\mathbb{F}_p)$ grows: $|A \cdot A \cdot A| > |A|^{1+\delta}$, $\delta > 0$, provided that $|A| < |G|^{1-\epsilon}$, $\epsilon > 0$. (Here $|S|$ is the number of elements of a set S.) This answered question (a) (on growth in graphs) immediately in the case of the Cayley graph of $\mathrm{SL}_2(\mathbb{F}_p)$; the bounds obtained were strong enough to constitute the first proved case of a standard conjecture (Babai's). Questions (c) and (d) (on random walks and spectral gaps) are closely related to each other, and somewhat more indirectly to (a) and (e); the result in [He] gave non-trivial bounds for (c) and (d). These bounds were greatly improved by Bourgain and Gamburd ([BG]), who showed how to use a technique of Sarnak and Xue's [SX] to derive from the results in [He] bounds for (c) and (d) that are qualitatively optimal (sufficient to amount to an *expander graph property* for all sets of generators A of G such that (G, A) has the *large girth* property).

1.2. **Main result.** It remained to be seen whether the result in [He] on growth in $\mathrm{SL}_2(\mathbb{F}_p)$ could be generalised to other groups. Much of the work in [He] was specific to $\mathrm{SL}_2(\mathbb{F}_p)$. In [BG2], the result was generalised (in a suitably strong form) to $\mathrm{SU}_2(\mathbb{C})$; there is also a recent generalisation by O. Dinai [Di] to $\mathrm{SL}_2(\mathbb{F}_q)$, as well as results [Bo] on $\mathrm{SL}_2(\mathbb{Z}/d\mathbb{Z})$. From the point of view of the Lie algebra, all of these groups are very closely related to $\mathrm{SL}_2(\mathbb{F}_p)$. Thus, the matter of the extent to which the methods in [He] were truly flexible remained open.

The point of the present paper is to prove growth for $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$. Part of the proof (§5) is ultimately derived from that in [He], and is likely to be valid for all semisimple groups of Lie type; part of the proof is essentially new.

Main Theorem. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Suppose $|A| < |G(K)|^{1-\epsilon}$, $\epsilon > 0$. Then

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{1.1}$$

where $\delta > 0$ and the implied constant depend only on ϵ.

We could, as in [He], write *let A be a subset of $G(K)$ not contained in a proper subgroup of $G(K)$* instead of *let A be a set of generators of $G(K)$*; the two statements are equivalent.

The condition that A generate $G(K)$ is easy to satisfy in applications (see, e.g., [BG], where the analogous result ([He]) on $\mathrm{SL}_2(\mathbb{F}_p)$ was applied).

Quite separately, it can be argued that the condition that A generate $G(K)$ is a natural one. If A does not generate $G(K)$, what we have is no longer a statement about $G(K)$, but, rather, a statement about the group $\langle A \rangle$ generated by A; the set A cannot know that elements outside $\langle A \rangle$ exist.

[1] In the non-abelian case, there are technical reasons why it makes more sense to consider $A \cdot A \cdot A$ rather than $A \cdot A$. The product $A \cdot A$ could be small "by accident".

We will, nevertheless, study all subsets A of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$, whether they generate the group or not.

Theorem 1.1. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$.*
Then, for every $\epsilon > 0$, either

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{1.2}$$

where $\delta > 0$ and the implied constant depend only on ϵ, or there are subgroups $H_1 \triangleleft H_2 < \langle A \rangle$ such that

(a) *H_2/H_1 is nilpotent,*
(b) *A_k contains H_1, where k depends only on ϵ, and*
(c) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2.*

It is tempting to guess that a statement of this sort should be true in general for subsets A of arbitrary groups G. The constants δ and k would then have to depend on n, where n is the smallest integer such that G is isomorphic to a subgroup of $\mathrm{SL}_n(\mathbb{F}_{p^\alpha})$ for some prime power p^α.

1.3. **Consequences.**

1.3.1. *Diameters.* By a result of Gowers, Nikolov and Pyber[2] [NP, Cor. 1 and Prop. 2],

$$A \cdot A \cdot A = \mathrm{SL}_n(K) \tag{1.3}$$

for $A \subset G$, $|A| > 2|G|^{1-\frac{1}{3(n+1)}}$, where $G = \mathrm{SL}_n(K)$ and $K = \mathbb{Z}/p\mathbb{Z}$.

Together with (1.3), the main theorem implies results on diameters. The *diameter* of a graph Γ is

$$\max_{v_1,v_2\in V}(\text{shortest distance between } v_1 \text{ and } v_2),$$

where V is the vertex set of Γ. We are especially interested in the diameters of *Cayley graphs*. The *Cayley graph* $\Gamma(G, A)$ of a pair (G, A) (where G is a group and $A \subset G$) is defined to be the graph that has G as its set of vertices and $\{(g, ag) : g \in G, a \in A\}$ as its set of edges. It is easy to see that the diameter $\mathrm{diam}(\Gamma(G, A))$ of a Cayley graph $\Gamma(G, A)$ is the least integer k such that

$$G = \{I\} \cup A \cup (A \cdot A) \cup \cdots \cup (\underbrace{A \cdot A \cdots A}_{k \text{ times}}).$$

If A is a set of generators of G, then, by definition, every element of G can be expressed as a product of elements of $A \cap A^{-1}$; when G is finite, this implies that every element of G can be expressed as a product of elements of A, i.e., the diameter $\mathrm{diam}(\Gamma(G, A))$ of the Cayley graph $\Gamma(G, A)$ is finite. The question remains: how large can the diameter $\mathrm{diam}(\Gamma(G, A))$ be in terms of G and A?

The following statement is known as *Babai's conjecture.*

[2]Gowers [Gow] proved a statement from which (1.3) quickly follows, as was pointed out by Nikolov and Pyber; see [NP]. The results in [Gow] and [NP] are of a general nature; with the aid of standard lower bounds on the dimensions of complex representations of SL_n, the special cases SL_2 and PSL_n were worked out in [Gow] and [NP], respectively. More general statements can be found in [BNP]. A weaker version of (1.3) for $n = 2$ was proven in [He, Key proposition, part (b)].

Conjecture ([BS])**.** *For every non-abelian finite simple group G and any set of generators A of G,*

$$\mathrm{diam}(\Gamma(G,A)) \ll (\log|G|)^c, \tag{1.4}$$

where c is some absolute constant and $|G|$ is the number of elements of G.

Until recently, there was no infinite family of groups G for which the conjecture was known for all A. In [He], I proved Babai's conjecture for $G = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ and all A. I shall now prove the conjecture for $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$.

Corollary 1.2 (to the main theorem and (1.3))**.** *Let p be a prime. Let $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$. Let A be a set of generators of G. Then*

$$\mathrm{diam}(\Gamma(G,A)) \ll (\log|G|)^c, \tag{1.5}$$

where c and the implied constant are absolute.

As we shall see (§10), given the main theorem, the proof of this corollary is quite simple.

It is clear that the corollary, as stated, implies that (1.5) holds for $G = \mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$ as well. (I bother to say this because $\mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$ is always simple, while $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ is not simple for some p.)

If $A \subset G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ is such that $\Gamma(G,A)$ has *girth* $\gg \log|G|$ (i.e., if it has no non-trivial cycles of length less than a constant times $\log|G|$), it is easy to see that the main theorem implies that the diameter of $\Gamma(G,A)$ is in fact $\ll \log|G|$, not simply $\ll (\log|G|)^c$ (see §10). As we are about to discuss, it is likely that even stronger statements can be made in this situation.

1.3.2. *Spectral gaps and expander graphs.* Soon after [He], Bourgain and Gamburd ([BG]) showed that, for $G = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ and A any set of generators with girth $\gg \log|G|$, the adjacency matrix of the Cayley graph $\Gamma(G,A)$ has a *spectral gap* of size $\epsilon > 0$, i.e., the difference between its largest and second largest eigenvalues is bounded below by a constant. (This implies that the endpoint of a random walk on $\Gamma(G,A)$ of length $C\cdot\log|G|$, C large, is close to being equidistributed.)

The starting point was the Key Proposition in [He], viz., the statement $|A\cdot A\cdot A| \geq |A|^{1+\epsilon}$ for $A \subset \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$; Bourgain and Gamburd succeeded in extracting a spectral gap $\epsilon > 0$ therefrom thanks to their use of a technique of Sarnak and Xue's [SX]. (In [SX], as in the work of Gowers et al., the main ingredient is the fact that $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ (or $\mathrm{SL}_n(\mathbb{F}_q)$, for that matter) has no small-dimensional complex representations.)

It is very likely that it will be possible to adapt Bourgain and Gamburd's procedure so as to prove a spectral gap $\lambda_1 - \lambda_2 > \epsilon$, $\epsilon > 0$ for $(\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z}), A)$ with large girth starting from the main theorem in the present paper. However, this is not immediate: what is needed, other than a straightforward translation of [BG] into SL_3, is a bound ruling out the possibility that the random walks on a Cayley graph of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ with large girth be highly concentrated on a subgroup early on.

1.4. **Outline.** Some basic background information will be given in §2. Sections 3 and 4 will be devoted to preparatory results in arithmetic combinatorics and growth in algebraic groups, respectively. The behaviour of a (hypothetical) non-growing set A in relation to maximal tori will be treated in §5; we will also examine the number of conjugacy classes occupied by such a set. The main result will finally be proven – for most of the possible range of $|A|$ – in §6. A case will remain; its treatment will necessitate some detailed work involving the subgroup structure of SL_3 (§9).

Section 4 treats algebraic groups in general. Most of the work in §5 will be done for SL_n. Sections 6 and 9 are quite specifically about SL_3.

1.4.1. *Plan of proof.* Let $G = \mathrm{SL}_3$, $K = \mathbb{Z}/p\mathbb{Z}$. Suppose there is a subset $A \subset G(K)$ violating the main theorem, i.e., a set A such that (a) A is substantially smaller than G ($|A| < |G|^{1-\epsilon}$, $\epsilon > 0$) and (b) A fails to grow ($|A \cdot A \cdot A| \ll |A|^{1+\delta}$, δ positive and very small). Then, as we shall show in §5, the set A must be in some sense very regular. For example, the number of conjugacy classes $\mathrm{Cl}_G(g)$ occupied by elements g of A will have to be almost precisely what one would expect out of dimensional reasons.

Perhaps more surprisingly, A will have to have a large intersection with some maximal torus T; in other words, A has many simultaneously diagonalisable elements. Our aim will be to use A to construct (§6) a set of tuples of elements of $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ satisfying too many linear relations too often. This will stand in contradiction to a bound on linearity (Cor. 3.8) that follows from a sum-product theorem (§3.3–3.4).

The above argument has a blind spot ($p^{4-\epsilon} < |A| < p^{4+\epsilon}$) resulting from the fact that sum-product theorems for $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ do have exceptions – all of size about p. For sets A of size in the blind spot, it becomes necessary to pass to a maximal parabolic subgroup and then use the fact that we already know that the main theorem holds for SL_2. If the intersection $A^{-1}A \cap M$ with a maximal parabolic subgroup M fails to generate a quotient of M isomorphic to $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, then $A^{-1}A \cap M$ must (in essence) lie in a Borel subgroup. We will see how sets grow in Borel subgroups by means of a general result (Prop. 3.1) of which the sum-product theorem is but a shadow (Lem. 3.4).

1.4.2. *Tools.* The tools used are elementary in nature – in contrast to the analytical tools sometimes used to study arithmetic groups.

The reader may wonder why the main theorem is a statement on $A \cdot A \cdot A$, as opposed to one on $A \cdot A$ or on the product of A with itself ten times. The statement $|A \cdot A| > |A|^{1+\delta}$ is not always true: let $A = H \cup \{g\}$, where H is a (non-normal) subgroup of G and $g \notin H$, for example. As for a statement on ten or twenty copies of A: we shall, in fact, be proving such a statement; a result essentially due to Ruzsa then tells us that, if $|A \cdot A^{-1} \cdot A \cdot A \cdot A \cdot A| > |A|^{1+\delta}$ (say), then $|A \cdot A \cdot A| > |A|^{1+\delta'}$ (with $\delta' > 0$ depending only on $\delta > 0$).

Additive combinatorics appears again in the guise of the Balog-Szeméredi-Gowers theorem. This is a very useful result, if somewhat rigid in its requirements; Bourgain showed in [BG2] how to remove it from the proof in [He], and it is likely that it will have to be replaced in the proof given here as well when the proof is generalised to SL_n, $n > 3$.

There is a rich literature on growth in infinite groups, based on the works of Gromov, Tits et al. There seems to be now at least one point of intersection with it: the escape argument of [EMO] will be used time and again in the course of this paper. In essence, it tells us that we may avoid any non-generic situation, such as, for example, that of matrices with repeated eigenvalues.

A truly crucial role is played by a sum-product theorem (first proven over finite fields by Bourgain, Katz and Tao [BKT] and Konyagin [Ko]). The result we need will be derived here from a more general statement (Prop. 3.1) on growth in groups under commuting actions without fixed points.

1.5. **Acknowledgements.** Starting on September 2007, I was supported by the EPSRC grant EP-E054919/1. My stay at the Institute for Advanced Study (Princeton) was supported by funds from the NSF grant DMS-0635607. Thanks are also due to the Tata Institute (Mumbai), the Institute for Mathematical Sciences (Chennai), Université Paris-Sud 11 (Orsay), École Polytechnique (Paris), the Rényi institute (Budapest) and Universidad de la Habana, for their hospitality and their support during my visits.

Nick Gill gave my paper several careful readings and made many helpful comments. Emmanuel Breuillard answered several of my questions, starting well before anything was written down. Thanks are also due to J. Bourgain, Y. Benoist, K. Buzzard, B. Conrad, O. Dinai, T. Ekedahl, G. Harcos, R. Hill, V. Meldrew, J. Pila, A. Silberstein, A. Skorobogatov, T. Szamuely and T. Wooley, for their help, and to the entire *groupe de travail* at the École Polytechnique and Chevaleret (Paris VI/VII), for hearing me out.

2. Notation and preliminaries

2.1. **General notation.** As is customary, we denote by $\mathbb{F}_{p^\alpha}$ the finite field of order p^α. Given a set A, we write $|A|$ for its number of elements. By $A+B$ (resp. $A\cdot B$), we shall always mean $\{x+y : x\in A, y\in B\}$ (resp. $\{x\cdot y : x\in A, y\in B\}$), By $A+\xi$ and $\xi\cdot A$ we mean $\{x+\xi : x\in A\}$ and $\{\xi\cdot x : x\in A\}$, respectively.

Given a positive integer r and a subset A of a group G, we define A_r to be the set of all products of at most r elements of $A\cup A^{-1}$:

$$A_r = \{g_1\cdot g_2\cdots g_r : g_i\in A\cup A^{-1}\cup\{1\}\}. \tag{2.1}$$

For us, A^r means $\{x^r : x\in A\}$; in general, if f is a function on A, we take $f(A)$ to mean $\{f(x) : x\in A\}$. If Υ is a set of maps from X to Z, and A and Y are subsets of X and Υ, respectively, then

$$Y(A) = \{y(a) : y\in Y,\ a\in A\}.$$

We write $Y(a)$ for $Y(\{a\})$ and $y(A)$ for $\{y\}(A) = \{y(a) : a\in A\}$.

2.2. **Boundedness.** We say "$a\ll b$, where the implied constant is absolute" or "$a = O(b)$, where the implied constant is absolute" when we mean that the non-negative real number a (or the absolute value of the arbitrary real number a) is at most the real number b multiplied by an absolute constant. We write $a\ll_{c_1,c_2,\dots,c_n} b$ or $a = O_{c_1,c_2,\dots,c_n}(b)$ when we mean that the non-negative real number a (or the absolute value of the arbitrary real number a) is at most the real number b multiplied by a constant depending only on $c_1,c_2,\dots,c_n$. We write $a\gg_{c_1,c_2,\dots,c_n} b$ to mean that a is larger than a positive constant depending only on $c_1,c_2,\dots,c_n$.

In particular, $a\ll_{c_1,c_2,\dots,c_n} 1$ (or $a = O_{c_1,c_2,\dots,c_n}(1)$) will mean that a is bounded in terms of $c_1,c_2,\dots,c_n$ alone. We will use this latter notation even when a is not a real number, provided that we have defined what it means for a to be bounded (in terms of other variables).

For example, when we say that a vector

$$\vec{d} = (d_0,d_1,d_2,\dots,d_n,0,0,\dots) \qquad (d_i \text{ non-negative})$$

is bounded in terms of a quantity ℓ alone, we mean that both n and $d_0, d_1, \ldots, d_n$ are bounded in terms of ℓ alone. We can then write this as follows: $\vec{d} \ll_\ell 1$. The quantity ℓ may itself be a vector: we may write, for example, $\vec{d} \ll_{\vec{d'}} 1$ – meaning that n and $d_0, d_1, \ldots, d_n$ are bounded in terms of a vector $\vec{d'}$ alone – or, for that matter, $a \ll_{\vec{d}} 1$ – meaning that a number a is bounded in terms of $\vec{d}$ alone.

2.3. **Arithmetic combinatorics.** We start with a very simple and standard lemma.

Lemma 2.1. *Let G be a finite group. Let $A \subset G$. Suppose $|A| > \frac{1}{2}|G|$. Then $A \cdot A = G$.*

Proof. Suppose there is a $g \in G$ not in $A \cdot A$. Then, for every $x \in G$, either x or gx^{-1} is not in A. As x goes over all elements of G, we see that no more than one out of every two elements of G can lie in A. In other words, $|A| \leq \frac{1}{2}|G|$. Contradiction. □

The following result is based on ideas of Ruzsa's, and, in particular, on his triangle inequality ([He, Lem. 2.1]).

Lemma 2.2 (Tripling lemma)**.** *Let $k > 2$ be an integer. Let A be a finite subset of a group G. Suppose that*

$$|A_k| \geq C|A|.$$

for some $C \geq 1$. Then

$$|A \cdot A \cdot A| \geq C^\delta |A|$$

where $\delta > 0$ depends only on k.

The dependence of *delta* on k is, in fact, inverse linear ($1/\delta = O(k)$).

Proof. See [T, Lem. 3.4] or [He, Lem. 2.2]. □

In the present paper, we shall almost always use the tripling lemma in the following form: if $|A_k| \geq c|A|^{1+\epsilon}$ with $c, \epsilon > 0$, then $|A \cdot A \cdot A| \gg_{c,\epsilon,k} |A|^{1+\epsilon'}$, where $\epsilon' > 0$ depends only on c, ϵ and k. This is simply a special case of the lemma: set $C = c|A|^\epsilon$. (The proof in [He, Lem. 2.2] is stated for this special case, but works in general.)

The Balog-Szemerédi-Gowers theorem is known in several different forms. We derive the one we need from one of the most common formulations. We make no effort to optimise the constants involved.

Proposition 2.3 (Balog-Szemerédi-Gowers)**.** *Let $A_1, A_2, \ldots, A_n$ be finite subsets of an abelian group Z. Let $m = \min_j |A_j|$ and $M = \max_j |A_j|$. Let $S \subset A_1 \times A_2 \times \cdots \times A_n$ be such that*

$$|S| \geq cM^n \quad \textit{and} \quad \left|\left\{ \sum_{1 \leq j \leq n} a_j : (a_1, a_2, \ldots, a_n) \in S \right\}\right| \leq \frac{1}{c} m \tag{2.2}$$

for some constant $c \in (0, 1)$.

Then there is a subset $A' \subset A_1$ such that

$$|A'| \gg c^C |A| \quad \textit{and} \quad |A' + A'| \ll \frac{1}{c^C}|A'|,$$

where $C > 0$ and the implied constants are absolute.

Note that condition (2.2) can hold only if $\min |A_j| \geq c \max |A_j|$. The proof below actually shows that $|A'| \gg c|A|$, which is stronger than the result in the statement.

Proof. Choose the tuple $(a_3, a_4, \ldots, a_n) \in A_3 \times A_4 \times \cdots \times A_n$ such that the number of elements of the set

$$G_{a_3,a_4,\ldots,a_n} = \{(a_1, a_2) \in A_1 \times A_2 : (a_1, a_2, a_3, \ldots, a_n) \in S\}$$

is maximal. We apply the Balog-Szemerédi-Gowers theorem as given in [TV], Thm. 2.29 with $A = A_1$, $B = A_2$ and $G = G_{a_3,a_4,\ldots,a_n}$, and obtain that there are sets $A' \subset A_1$, $B' \subset A_2$ with $|A'| \gg c|A_1|$, $|B'| \gg c|A_2|$ and

$$|A' + B'| \ll c^{-7}|A_1|^{1/2}|A_2|^{1/2} \ll c^{-8}|A'|^{1/2}|B'|^{1/2} \ll c^{-9}|A'|.$$

We apply the Plünnecke-Ruzsa estimates ([TV], Cor. 6.29) and obtain that $|A' - A'| \ll c^{-18}|A'|$. $\square$

There are non-commutative versions of Balog-Szemerédi-Gowers (see [T]); we shall not need them, however.

2.4. **Groups and generation.** By $\langle g \rangle$ we mean the group generated by an element g of a group G. By $\langle A \rangle$ we mean the group generated by a subset A of a group G. By $H < G$ we mean that H is a subgroup of the group G.

We write $\mathrm{Cl}_G(g)$ for the conjugacy class of an element $g \in G$ in G.

2.5. **Varieties.** Let us speak concretely. An (affine) *variety* V is given by a finite set of polynomial equations $F(x_1, x_2, \ldots, x_n) = 0$ in n variables with coefficients in a field. (We will usually work in an affine space (denoted by $\mathbb{A}^n$), rather than in projective space $\mathbb{P}^n$; if we work in projective space, our polynomials F must all be homogeneous.) If the coefficients all lie in a field K, we say that V is *defined over* K, or simply write V/K. If L is another field – containing, contained in, or equal to K – then we write $V(L)$ for the set of L-*valued points of* V, i.e., the set of solutions in L^n to our set of equations.

A *subvariety* $W \subset V$ is a variety that can be defined by a set of equations that contains a set of equations defining V. By a *proper subvariety* $W \subset V$ we mean simply a subvariety with $W \neq V$. (There is a very different algebraic-geometrical notion of *properness*; we shall not use it.) Clearly, if two varieties V, W defined over K satisfy $W \subset V$, then $W(L) \subset V(L)$ for every extension L of K.

A *Zariski-open set* Σ in a variety V is the complement of a variety $W \subset V$; its set of points $\Sigma(L)$ is defined to be $V(L) \setminus W(L)$. A Zariski-open set is not, in general, a variety.

All or nearly all the algebraic geometry we need can be found in [Da], for instance.

2.5.1. *Algebraic groups.* If we speak of an (affine) *algebraic group* defined over a field K, we mean an affine variety G/K with a group law such that the multiplication map $\mu : G \times G \mapsto G$ and the inverse map $\iota : G \to G$ are regular and defined over K. (Between affine varieties, a *regular map* is simply a map given by polynomials.) Thus, strictly speaking, an algebraic group G is not a group; rather, its set of points $G(L)$ will be a group for every field L containing K. The set of points $G(L)$ for L contained in K may also be a group, if it is closed under the group operation.

The following are typical examples. We may speak of the algebraic group $G = \mathrm{SL}_2$ (or $G = \mathrm{SL}_n$, or, for that matter, $G = \mathrm{SO}_n$ or $G = \mathrm{Sp}_{2n}$). This is a variety defined over $\mathbb{Z}$, and thus over an arbitrary field: it is given by the equation $g_{11}g_{22} - g_{12}g_{21} = 1$ in the four variables g_{11}, g_{12}, g_{21}, g_{22} (or, for $G = \mathrm{SL}_n$ and general n, the equation $\det(g) = 1$; note that the determinant is a polynomial in the n^2 variables g_{ij}). Its group law is given by matrix multiplication. For any field K, the set $G(K)$ is the set $\mathrm{SL}_n(K)$ of all n-by-n matrices with entries in K and determinant 1; this set is a group under the group law just given, i.e., matrix multiplication. A *maximal torus* T in $G = \mathrm{SL}_n$ is a group consisting of all diagonal matrices for some choice of basis, i.e., a group that can be made into the group of diagonal matrices by conjugation. If T can be thus diagonalised by conjugation by a matrix in $G(K)$, then T is defined over K; otherwise, T is defined over $\overline{K}$ but not over K. Even in the latter case, we may still speak of the group $T(K)$. For example, if $K = \mathbb{R}$, and we consider the matrices

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \theta \in \mathbb{R},$$

we can see that they are the points over $\mathbb{R}$ of a maximal torus T, in that they can all be diagonalised simultaneously; this torus T is defined over $\mathbb{C}$, but cannot be defined over $\mathbb{R}$.

In general, algebraic groups behave a great deal like Lie groups, even over finite fields; in particular, they have maximal tori, roots, etc. Every (affine) algebraic group is a closed algebraic subgroup of GL_n for some $n \geq 1$ ([Hum, §8.6]). For an introduction to algebraic groups, see [Bor] or [Hum].

2.5.2. *Degree and dimension.* The *dimension* $\dim(X)$ of an irreducible variety X is the length k of the longest chain $\{x\} = X_0 \subset X_1 \subset \cdots \subset X_k = X$ of irreducible subvarieties of X; this corresponds to the intuitive notion of dimension. If the irreducible components of a variety V all have the same dimension, we say V is *pure dimensional*, and define the dimension $\dim(V)$ of V to be that of any of its irreducible components. (An *irreducible component* of a variety V is an irreducible subvariety of V not contained in any other irreducible subvariety of V.)

The *degree* $\deg(V)$ of a pure-dimensional variety V of dimension r in n-dimensional affine or projective space is its number of intersection points with a generic linear variety of dimension $n - r$. (Thus, for example, the degree of an irreducible plane curve is its number of intersection points with a generic line.)

Let us first see what the degree of a variety has to do with the familiar notion of the degree of a polynomial. Let F be an irreducible polynomial in n variables with coefficients in a field K. Then the equation

$$F(x_1, x_2, \ldots, x_n) = 0$$

defines an irreducible variety V of codimension 1, i.e., of dimension $n-1$ in n-dimensional affine space $\mathbb{A}^n$. (The irreducibility of V turns out to be an easy consequence of the irreducibility of F and the fact that $K[x_1, x_2, \ldots, x_n]$ is a *unique factorisation domain*.)

Now, it is not hard to see that the degree of V will be equal to the degree of F: if we let $x_1 = a_1 + b_1 t, x_2 = a_2 + b_2 t, \ldots, x_n = a_n + b_n t$ for some constants $a_1, a_2, \ldots, a_n \in \overline{K}$, $b_1, b_2, \ldots, b_n \in \overline{K}^*$, the equation $F(a_1 + b_1 t, a_2 + b_2 t, \ldots, a_n + b_n t) = 0$ will be an equation on t of degree at most $\deg(F)$, and, for $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$ sufficiently "generic", of degree exactly $\deg(F)$. That equation on t will hence have $\deg(F)$ roots (all distinct for a_i, b_i sufficiently generic). In other words, V and the line given by $x_1 = a_1 + b_1 t$,

$x_2 = a_2 + b_2 t$, ..., $x_n = a_n + b_n t$ have $\deg(F)$ intersection points. (It should be clear now that we mean intersection points whose coordinates lie in the algebraic closure $\overline{K}$, and not necessarily in K.) We have thus sketched how to show that $\deg(V) = \deg(F)$.

All of the above can be made precise, in that all the statements above remain true when "generic" is given what we shall see as its precise meaning: namely, "outside a variety of positive codimension". Thus, for example, the degree of $F(a_1+b_1t, a_2+b_2t, \ldots, a_n+b_nt) = 0$ is exactly $\deg(F)$ provided that $(a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n)$ lies outside a variety of codimension 1 in $\mathbb{A}^{2n}$, viz., the variety given by the equation

$$\text{leading coefficient} = 0.$$

Similarly, the roots $t_1, t_2, \ldots$ of $f_{\vec{a},\vec{b}}(t) = F(a_1 + b_1t, a_2 + b_2t, \ldots, a_n + b_nt) = 0$ are all distinct if $(a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n)$ lies outside the variety given by the equation

$$\text{discriminant}(f_{\vec{a},\vec{b}}) = 0,$$

or, alternatively, if the line given by $(a_1 + b_1t, \ldots, a_n + b_nt)$ is not tangent to the surface $F(x_1, x_2, \ldots, x_n) = 0$ at any point. (Showing that the discriminant is not identically 0 may not be immediately obvious.)

The degree of a variety is a yardstick of complexity that behaves well under intersections. For the proof, we shall need the following general version of Bezout's theorem.

Lemma 2.4 (Bezout's theorem, generalised)**.** *Let $X_1, X_2, \ldots, X_k$ be pure-dimensional varieties in $\mathbb{P}^n$, and let $Z_1, Z_2, \ldots, Z_l$ be the irreducible components of the intersection $X_1 \cap X_2 \cap \cdots \cap X_k$. Then*

$$\sum_{j=1}^{l} \deg(Z_j) \leq \prod_{i=1}^{k} \deg(X_i),$$

where $X_1, X_2, \ldots, X_k$ are the irreducible components of X_j.

As is stated in [Da], the form of the statement goes back to Fulton and MacPherson.

Proof. See [Da], p. 251. □

It remains to see how to define the dimension and the degree of a variety V when V is not irreducible. We may simply define $\dim V$ to be the dimension of the irreducible subvariety of V of largest dimension. As for the degree, it will be best to see it as a vector: we define the *degree* $\overrightarrow{\deg}(V)$ of an arbitrary variety V to be

$$(d_0, d_1, \ldots, d_k, 0, 0, 0, \ldots),$$

where $k = \dim(V)$ and d_j is the degree of the union of the irreducible components of V of dimension j.

It is easy to see that Bezout's theorem implies that, for any varieties $V_1, V_2, \ldots, V_k$, the degree $\overrightarrow{\deg}(W)$ of the intersection $W = V_1 \cap V_2 \cap \cdots \cap V_k$ is bounded in terms of $\overrightarrow{\deg}(V_1)$, $\overrightarrow{\deg}(V_2)$,..., $\overrightarrow{\deg}(V_k)$ alone. (See §2.2 for an explanation of what we mean by $\overrightarrow{\deg}(W)$ being bounded in terms of such and such; we mean that both $\dim(W)$ and the degree d_j of the union of the irreducible components of V of dimension j are bounded in terms of such and such.) We can also see easily (without the use of Bezout's theorem) that the degree of the variety $V = V_1 \cup V_2 \cup \cdots \cup V_k$ is bounded in terms of $\overrightarrow{\deg}(V_1)$, $\overrightarrow{\deg}(V_2)$, ..., $\overrightarrow{\deg}(V_k)$; in fact, if the V_i's have no components in common, we will have $\overrightarrow{\deg}(V) = \sum_i \overrightarrow{\deg}(V_i)$.

A very concrete consequence of what we have said so far is the following: if a variety V is defined by equations

$$\begin{aligned} F_1(x_1, x_2, \ldots, x_n) &= 0 \\ F_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\cdots\cdots \\ F_k(x_1, x_2, \ldots, x_n) &= 0, \end{aligned}$$

then its degree $\overrightarrow{\deg}(V)$ is bounded in terms of n and $\deg(F_1)$, $\deg(F_2)$, ..., $\deg(F_n)$ alone.

If a regular map $\phi : V \mapsto W$ between two varieties $V \subset \mathbb{A}^m$, $W \subset \mathbb{A}^n$ is defined by polynomials $\phi_1, \phi_2, \ldots, \phi_n$ on the variables $x_1, x_2, \ldots, x_m$, we define $\deg_{\mathrm{pol}}(\phi)$ to be $\max_j \deg(\phi_j)$. (If several representations of ϕ by polynomials $\phi_1, \phi_2, \ldots, \phi_n$ are possible, we choose – for the purposes of defining $\deg_{\mathrm{pol}}$ – the one that gives us the least value of $\deg_{\mathrm{pol}}$.) What we have just seen amounts to stating that, if a subvariety V' of V is given by $\phi(x) = y$ for some $y \in W(\overline{K})$, then $\overrightarrow{\deg}(V')$ can be bounded in terms of $\deg_{\mathrm{pol}}(\phi)$ and n (where $W \subset \mathbb{A}^n$).

2.5.3. *Fibres and counting.* Let V be a subvariety of $X \times Y$, where X and Y are varieties. The *fibre* $V_{x=x_0}$ (or $V_{y=y_0}$) is the subvariety of Y (or X) consisting of the points y such that (x_0, y) lies on V (or of the points x such that (x, y_0) lies on V). It is an immediate consequence of Bezout's theorem that $\overrightarrow{\deg}(V_{x=x_0})$ and $\overrightarrow{\deg}(V_{y=y_0})$ are bounded in terms of $\overrightarrow{\deg}(V)$ alone.

Let V be a proper subvariety of $X \times Y$, where X and Y are varieties. Then there is a proper subvariety W of X such that, for every x_0 lying on $X \setminus W$, the fibre $V_{x=x_0}$ is a proper subvariety of Y; moreover, $\overrightarrow{\deg}(W)$ is bounded in terms of $\overrightarrow{\deg}(V)$ alone. This is easy to show: since V is a proper subvariety of $X \times Y$, there is a point of (x_0, y_0) of $X \times Y$ not on V; then the fibre $V_{y=y_0}$ is a proper subvariety of X, and, for every x_0' lying on $X \setminus V_{y=y_0}$, the fibre $V_{x=x_0'}$ does not contain the point y_0, and hence is a proper subvariety of Y. Set, then, $W = V_{y=y_0}$.

By the same argument, there is also a proper subvariety W' of Y such that, for every y_0 lying on $Y \setminus W'$, the fibre $V_{y=y_0}$ is a proper subvariety of X.

Let K be a finite field. Let $V/\overline{K}$ be a subvariety of $\mathbb{A}^n$ such that all of its irreducible components have dimension $\leq m$. Then

$$|V(K)| \ll_{\overrightarrow{\deg}(V),n} |K|^m. \tag{2.3}$$

This crude bound can be proven as follows.

We will proceed by induction on n. We can assume without loss of generality that V is irreducible of dimension m. (The number of components of a variety V is $\ll_{\overrightarrow{\deg}(V)} 1$.) See $\mathbb{A}^n$ as the product of affine varieties $\mathbb{A}^1 \times \mathbb{A}^{n-1}$. Suppose that there is a point $t \in \mathbb{A}^1$ such that the fibre $V_{x_1=t}$ has components of dimension m. Then $V = \{t\} \times V_{x_1=t}$, as otherwise V would have dimension $> m$ (by the definition of dimension). We then obtain (2.3) by the inductive assumption for $n-1$.

Suppose now that there is no point $t \in \mathbb{A}^1$ such that the fibre $V_{x_1=t}$ has components of dimension m. By the inductive assumption, (2.3) holds for $n-1$, and so $|V_{x_1=t}(K)| \ll_{\overrightarrow{\deg}(V_{x_1=t}),n-1} |K|^{m-1}$ for every t. Since there are $|K|$ possible values of t, and since $\overrightarrow{\deg}(V_{x_1=t}) \ll_1 \overrightarrow{\deg}(V)$, we conclude that

$$|V(K)| \ll_{\overrightarrow{\deg}(V),n} |K| \cdot |K|^{m-1} = |K|^m,$$

as we wished to show.

Bounds much more precise than (2.3) are known: take, for instance, the Lang-Weil theorem [LW]. (We shall not need the later and very deep results of Deligne and others.)

2.5.4. *Abuse of language.* Given a variety V defined over a field K, and a subvariety $W/\overline{K}$ defined over the algebraic completion $\overline{K}$ of K, we will write $W(K)$ for $W(\overline{K}) \cap V(K)$. (We will even speak of the points of W over K, meaning $W(K) := W(\overline{K}) \cap V(K)$.)

2.5.5. *Independence.* Let $V_1, V_2, \dots, V_k$ be linear subspaces of an affine space $\mathbb{A}^n$; let them be defined over a field K. We say that $V_1, V_2, \dots, V_k$ are *linearly independent* if there is no choice of points $v_1 \in V_1(\overline{K})$, $v_2 \in V_2(\overline{K})$, $\dots$, $v_k \in V_k(\overline{K})$, not all of them 0, such that $v_1 + v_2 + \cdots + v_k = 0$.

3. Growth in rings and Borel subgroups

3.1. **Growth under commuting actions.** Ever since the sum-product theorem was proven by Bourgain, Katz and Tao ([BKT]), it has been subject to a series of refinements and variations. Of these, one of the most interesting is a result of Glibichuk and Konyagin ([GK], Lemma 3.2–Corollary 3.5), both because it applies to pairs of sets of completely arbitrary sizes, and because of its rather simple proof.

It will become apparent that the natural setting of "sum-product theorems" is a much broader one than the one in [BKT], [GK] or the related literature. It is not really a result about subsets of the field $\mathbb{Z}/p\mathbb{Z}$, but, rather, a result about groups (abelian or non-abelian) and commuting automorphisms thereof. We shall show that the sum-product theorem over $\mathbb{Z}/p\mathbb{Z}$ (say) is a consequence of a special case of the general result below. Before that, we shall also see how this general result has useful implications on the action of maximal tori in $\mathrm{SL}_n(K)$ on unipotent subgroups. (We will need to study unipotent subgroups of SL_n when we examine large subsets of $\mathrm{SL}_n(K)$.)

Proposition 3.1. *Let G be a group and Υ an abelian group of automorphisms of G. Let $Y \subset \Upsilon$ be a non-empty set such that*

$$\text{if } y(g) = g \text{ for } y \in Y^{-1}Y,\ g \in G, \text{ then either } y = e \text{ or } g = e. \tag{3.1}$$

Then, for any non-empty $A \subset G$ and any $Y_0 \subset \Upsilon$, $A_0 \subset G$, either

$$|A \cdot Y(a_1)| \geq |A| \cdot |Y| \tag{3.2}$$

or

$$|\{y_2(a) \cdot y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y(y_1(a_0^{-1})) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y|. \tag{3.3}$$

for some $a_0 \in A_0$, $a_1, a_2 \in A$, $y_1, y_2 \in Y$, or

$$|\{y_2(a) \cdot y_0(y(a_2^{-1}a_1)) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y| \tag{3.4}$$

for some $y_0 \in Y_0$, $a_1, a_2 \in A$, $y_1, y_2 \in Y$, *or*

$$(3.5)\quad |\{y_2(a) \cdot y(a_2^{-1}a_1) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| > \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|} \geq \frac{1}{2}\min(|A||Y|, |\mathscr{O}|),$$

where $a_1, a_2 \in A$, $y_1, y_2 \in Y$, *and* $\mathscr{O}$ *is the union of the orbits of the elements of* A *under the operations* $a \mapsto a_0 \cdot a$ *(for all* $a_0 \in A_0$*) and* $a \mapsto y_0(a)$ *(for all* $y_0 \in Y_0$*).*

It should be easy to see that the inequalities (3.2)–(3.4) must all be equalities; we phrase them as inequalities simply because we are interested in lower bounds on growth.

If we take $A_0 = A$ and $Y_0 = Y \cup Y^{-1}$, Proposition 3.1 acquires a particularly simple form:

Corollary 3.2. *For any group* G *and any abelian group* Υ *of automorphisms of* G*. Then, for any* $A \subset G$ *and any* $Y \subset \Upsilon$ *satisfying (3.1),*

$$|(Y_2(A))_6| > \frac{1}{2}\min(|A||Y|, |R|),$$

where $R = \langle\langle Y\rangle(\langle A\rangle)\rangle$ *is the set of all products of elements of the form* $y(a)$ *with* $a \in \langle A\rangle$ *and* $y \in \langle Y\rangle$*.*

Proof of Corollary 3.2. Set $A_0 = A \cup A^{-1}$, $Y_0 = Y \cup Y^{-1}$ and apply Proposition 3.1. It remains only to prove that the union $\mathscr{O}$ of the orbits of the elements of A under the action of $x \mapsto a \cdot x$ ($a \in A$) and $x \mapsto y(x)$ ($y \in Y$) is equal to the set R described in the statement. It is clear that $\mathscr{O} \subset R$.

To prove $R \subset \mathscr{O}$, we proceed by induction: let $R(n)$ be the set of all products of at most n elements of the form $y(a)$, $a \in A \cup A^{-1}$, $y \in \langle Y\rangle$. Assume $R(n) \subset \mathscr{O}$. (This is certainly true for $n = 0$, since the identity element $e = a \cdot a^{-1}$ is in $\mathscr{O}$.) We wish to prove $R(n+1) \subset \mathscr{O}$. Any $g \in R(n+1)$ can be written in the form $y(a) \cdot h$, where $y \in \langle Y\rangle$ and $a \in A \cup A^{-1}$. Now $y(a) \cdot h = y(a \cdot y^{-1}(h))$. Because $h \in R(n)$, and because y is a homomorphism, $y^{-1}(h)$ is also in $R(n)$. Since $R(n) \subset \mathscr{O}$, $y^{-1}(h)$ must be in $\mathscr{O}$. Then $y(a \cdot y^{-1}(h))$ must also be in $\mathscr{O}$. Thus every element of $R(n+1)$ is in $\mathscr{O}$. □

Examples. Before we prove Proposition 3.1, let us see two of its consequences; we shall examine them in more detail later.

(a) Let $G = \mathbb{F}_p$ (as an additive group), $\Upsilon = \mathbb{F}_p^*$ (acting on G by multiplication), $A_0 = \{1\}$, $G_0 = e$. Then condition (3.1) is easily seen to be satisfied: it just says that, in a field, if $y \cdot g = g$, then either $y = 1$ or $g = 0$. (The same is true in any ring without zero divisors.) Thus we may apply Proposition 3.1, and we obtain that, for any $A \subset \mathbb{F}_p$ and any $Y \subset \mathbb{F}_p^*$,

$$(3.6)\qquad |Y \cdot A + Y \cdot A - Y \cdot A - Y \cdot A + Y^2 - Y^2| > \frac{1}{2}\min(|A||Y|, p).$$

(This is the result of Glibichuk and Konyagin's mentioned before; see [GK, §3].) We may set $Y = A$, and then a few applications of the Plünnecke-Ruzsa estimates ([TV], Cor. 6.29) suffice to derive from (3.6) the conclusion that

$$|A \cdot A + A \cdot A| \geq |A| \cdot (\frac{1}{2}\min(|A|, p/|A|))^{1/6}$$

for every subset A of $\mathbb{F}_p^*$. An application of the Katz-Tao lemma ([TV, Lem. 2.53]; see also [B], [Ga]) then suffices to show that, for every $A \subset \mathbb{F}_p^*$ with $|A| < p^{1-\delta}$, $\delta > 0$, we have either $|A + A| > |A|^{1+\epsilon}$ or $|A \cdot A| > |A|^{1+\epsilon}$, where $\epsilon > 0$ depends only on $\delta > 0$. This is the well-known *sum-product theorem* of Bourgain, Katz and Tao ([BKT]), as extended by Konyagin. We shall not use this theorem; instead, we shall use a sum-product theorem on the ring $\mathbb{F}_p \times \mathbb{F}_p$, after proving it by proceeding much as we just did.

(b) Let G be the group of upper-triangular matrices in $\mathrm{SL}_n(K)$ with 1's on the diagonal. Let Υ be the group of diagonal matrices, acting on G by *conjugation* (not multiplication). Let $Y \subset \Upsilon$ be a set of matrices such that the map $g \mapsto g_{ii} g_{jj}^{-1}$ (i.e., a *root of* $\mathrm{SL}_n(K)$ *relative to* Υ, in the parlance of groups of Lie type) is injective on Y for all $1 \leq i, j \leq n$ distinct.

Then (3.1) is satisfied, and so, by Corollary 3.2,

$$|(Y_2(A))_6| \geq \frac{1}{2} \min(|A||Y|, |R|),$$

where $R = \langle\langle Y \rangle(\langle A \rangle)\rangle$. We shall look into this issue with more care in §3.2; see Proposition 3.3.

We will now see the proof of Proposition 3.1. It is quite close to that of [B2, Lemma 1], whose proof is in turn based closely on the argument in [GK, §3]. Our version is self-contained.

Proof of Proposition 3.1. The idea is to use a "pivot" ξ, meaning an element ξ of G such that the map ϕ_ξ from $A \times Y$ to G given by $(g, y) \mapsto (g \cdot y(\xi))$ is injective. If there is such a pivot, the injectivity of ϕ_ξ gives us that $|A \cdot Y(\xi)|$ is large: $|A \cdot Y(\xi)| \geq |A| \cdot |Y|$. Then one finishes by showing that one can construct ξ in a bounded number of steps starting from A and Y. If there is no pivot, then the set of non-pivots must be rather large. We use this fact itself to prove growth.

Saying that ϕ_ξ is injective is the same as saying that $\xi \notin \delta_{y_1,y_2}^{-1}(\{a_2^{-1} \cdot a_1\})$ for all $a_1, a_2 \in A$ and all distinct $y_1, y_2 \in Y$, where $\delta_{y_1,y_2} : G \to G$ is the map $\gamma \mapsto y_2(\gamma) \cdot (y_1(\gamma))^{-1}$. Now, if $\delta_{y_1,y_2}(\gamma_1) = y_2(\gamma_1) \cdot (y_1(\gamma_1))^{-1}$ equals $\delta_{y_1,y_2}(\gamma_2) = y_2(\gamma_2) \cdot (y_1(\gamma_2))^{-1}$, then $y_2(\gamma_1^{-1}\gamma_2) = y_1(\gamma_1^{-1}\gamma_2)$, and so $y_1^{-1}(y_2(\gamma_1^{-1}\gamma_2)) = \gamma_1^{-1}\gamma_2$. Since $y_1, y_2 \in Y$ are distinct and $\gamma_1, \gamma_2 \in G$ are distinct, this contradicts assumption (3.1). Hence $\delta_{y_1,y_2} : G \to G$ is injective for all pairs (y_1, y_2) of distinct elements of Y. This shall be crucial later.

We face two cases, depending on whether or not the set

$$S = \bigcup_{\substack{a_1, a_2 \in A \\ y_1, y_2 \in Y \\ y_1 \neq y_2}} \delta_{y_1,y_2}^{-1}(a_2^{-1} \cdot a_1) \tag{3.7}$$

contains the orbit $\mathscr{O}$. The set $\mathscr{O}$ contains all "easily constructible" elements; if $\mathscr{O}$ is not contained in S, we can construct an element not in S, i.e., a valid pivot.

Case 1: $\mathscr{O} \not\subset S$. (Read: there is a pivot.)

The set $\mathscr{O}$ is the union of orbits of the elements of A under certain actions. Hence, if $\mathscr{O} \not\subset S$, we have that either $A \not\subset S$ or there is an element s of S that is taken out of S by one of the actions: that is, either $a_0 \cdot s \notin S$ for some $a_0 \in A_0$ or $y_0(s) \notin S$ for some

$y_0 \in Y_0$. Call these three cases (a), (b) and (c). In case (a), we let ξ be any element of A not in S; in case (b), we let $\xi = a_0 \cdot s$; finally, in case (c), we let $\xi = y_0(s)$.

Now we are almost done. We have a map

$$\phi_\xi : (g, y) \mapsto g \cdot y(\xi) \tag{3.8}$$

from $A \times Y \to G$. Because $\xi \notin S$, the map is injective. The map has been constructed in a finite number of steps from the elements of A and Y, since ξ was defined that way.

Let us work out the meaning and implications of this last statement case by case.

Case 1(a): $A \not\subset S$; ξ an element of A not in S. Since ϕ_ξ is injective,

$$|A \cdot Y(\xi)| \geq |A| \cdot |Y|.$$

We have proven (3.2).

Case 1(b): $\xi = a_0 \cdot s$. Since ϕ_ξ is injective,

$$|A \cdot Y(\xi)| \geq |A| \cdot |Y|. \tag{3.9}$$

Now we must do a little work: ξ is defined in terms of s, and the definition of s involves the map δ_{y_1,y_2}^{-1}, which we must now somehow remove. Because δ_{y_1,y_2} is injective, (3.9) implies

$$|\delta_{y_1,y_2}(A \cdot Y(\xi))| \geq |A| \cdot |Y|. \tag{3.10}$$

Now, for any $a \in A$, $y \in Y$,

$$\begin{aligned} \delta_{y_1,y_2}(a \cdot y(\xi)) &= y_2(a \cdot y(\xi)) \cdot (y_1(a \cdot y(\xi)))^{-1} \\ &= y_2(a) \cdot y_2(y(\xi)) \cdot (y_1(y(\xi)))^{-1} \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(y_2(\xi)(y_1(\xi))^{-1}) \cdot (y_1(a))^{-1}. \end{aligned} \tag{3.11}$$

(It is here that the fact that Υ is abelian is finally used.) Recall that the definition of δ_{y_1,y_2} is $\delta_{y_1,y_2}(\xi) = y_2(\xi)(y_1(\xi))^{-1}$.

Because we are in case 1(b), there are $a_0 \in A_0$, $s \in S$ such that $\xi = a_0 \cdot s$. By the definition (3.7) of S, there are $y_1, y_2 \in Y$ distinct and $a_1, a_2 \in A$ such that $\delta_{y_1,y_2}(s) = a_2^{-1} \cdot a_1$. Then

$$\begin{aligned} y(y_2(\xi) \cdot y_1(\xi)^{-1}) &= y(y_2(a_0) \cdot y_2(s) \cdot (y_1(s))^{-1} \cdot (y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(y_2(s)(y_1(s))^{-1}) \cdot y((y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(\delta_{y_1,y_2}(s)) \cdot y((y_1(a_0))^{-1}) \\ &= y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y((y_1(a_0))^{-1}). \end{aligned} \tag{3.12}$$

Thus

$$\delta_{y_1,y_2}(a \cdot y(\xi)) = y_2(a) \cdot y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y((y_1(a_0))^{-1}) \cdot (y_1(a))^{-1}.$$

We conclude that

$$|\{y_2(a) \cdot y(y_2(a_0)) \cdot y(a_2^{-1} \cdot a_1) \cdot y(y_1(a_0^{-1})) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y|.$$

That is, the conclusion (3.3) is true.

Case 1(c): $\xi = y_0(s)$. We start as in case 1(b): (3.10) and (3.11) still hold. By the definition (3.7) of S, there are $y_1, y_2 \in Y$ distinct and $a_1, a_2 \in A$ such that $\delta_{y_1,y_2}(s) =$

$a_2^{-1} \cdot a_1$. Now, because we are in case 1(c) and not in case 1(b), we have $\xi = y_0(s)$ instead of $\xi = a_0 \cdot s$. We replace (3.12) by the following calculation:

$$\begin{aligned} y(y_2(\xi) \cdot y_1(\xi)^{-1}) &= y(y_2(y_0(s)) \cdot (y_1(y_0(s)))^{-1}) = y(y_2(y_0(s))) \cdot y(y_1(y_0(s^{-1}))) \\ &= y_0(y(y_2(s))) \cdot y_0(y(y_1(s^{-1}))) = y_0(y(y_2(s) \cdot y_1(s^{-1}))) \\ &= y_0(y(y_2(s) \cdot (y_1(s))^{-1})) = y_0(y(\delta_{y_1,y_2}(s))) = y_0(y(a_2^{-1}a_1)). \end{aligned}$$

(It is here that the fact that Υ is abelian is used for the second time.) Thus

$$\delta_{y_1,y_2}(a \cdot y(\xi)) = y_2(a) \cdot y_0(y(a_2^{-1}a_1)) \cdot (y_1(a))^{-1}.$$

We conclude that

$$|\{y_2(a) \cdot y_0(y(a_2^{-1}a_1)) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| \geq |A| \cdot |Y|.$$

In other words, (3.4) holds.

Case 2: $\mathscr{O} \subset S$. (Read: there is no pivot.)

Then S must be rather large. From the definition (3.7), it becomes clear that either Y or A must be rather large. It is then no surprise that some crude techniques appropriate for large sets shall be sufficient for our task.

Since δ_{y_1,y_2} is injective for $y_1 \neq y_2$, the sets

$$R_\xi = \{(a_1, a_2, y_1, y_2) \in A \times A \times Y \times Y : y_1 \neq y_2,\ a_1 \cdot y_1(\xi) = a_2 \cdot y_2(\xi)\}$$

are disjoint as ξ ranges in G. Choose $\xi_0 \in S$ such that $|R_{\xi_0}|$ is minimal. Then

$$|R_{\xi_0}| \leq \frac{|A|^2|Y|(|Y|-1)}{|S|} < \frac{|A|^2|Y|^2}{|S|} \leq \frac{|A|^2|Y|^2}{|\mathscr{O}|}$$

and so

$$|\{(a_1, a_2, y_1, y_2) \in A \times A \times Y \times Y : a_1 \cdot y_1(\xi_0) = a_2 \cdot y_2(\xi_0)\}| < \frac{|A|^2|Y|^2}{|\mathscr{O}|} + |A| \cdot |Y|.$$

Hence

$$|A \cdot Y(\xi_0)| > \frac{|A|^2|Y|^2}{\frac{|A|^2|Y|^2}{|\mathscr{O}|} + |A| \cdot |Y|} = \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|}. \tag{3.13}$$

As before, we must somehow remove δ_{y_1,y_2}^{-1} from ξ_0. By the injectivity of δ_{y_1,y_2}, (3.13) implies

$$|\delta_{y_1,y_2}(A \cdot Y(\xi_0))| > \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|}.$$

Equation (3.11) is still valid. Since $\xi \in S$, we know that $\delta_{y_1,y_2}(\xi_0) = a_2^{-1}a_1$ for some $a_1, a_2 \in A$, $y_1, y_2 \in Y$ distinct. Thus, for $a \in A$, $y \in Y$,

$$\begin{aligned} \delta_{y_1,y_2}(a \cdot y(\xi_0)) &= y_2(a \cdot y(\xi_0)) \cdot (y_1(a \cdot y(\xi_0)))^{-1} \\ &= y_2(a) \cdot y_2(y(\xi_0)) \cdot (y_1(y(\xi_0)))^{-1} \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(y_2(\xi_0) \cdot (y_1(\xi_0))^{-1}) \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(\delta_{y_1,y_2}(\xi_0)) \cdot (y_1(a))^{-1} \\ &= y_2(a) \cdot y(a_2^{-1}a_1) \cdot (y_1(a))^{-1}. \end{aligned}$$

(It is here that the fact that Υ is abelian is used for the third and last time.) Hence

$$|\{y_2(a) \cdot y(a_2^{-1}a_1) \cdot y_1(a^{-1}) : a \in A, y \in Y\}| > \frac{|A||Y||\mathscr{O}|}{|A||Y| + |\mathscr{O}|}.$$

The inequality $\frac{ab}{a+b} \geq \frac{1}{2}\min(a,b)$ is easy and true for all positive a, b. Hence we have proven (3.5). □

3.2. **Growth in unipotent groups under the action of the diagonal.** Proposition 3.1 does not require the group G to be abelian. The following is a natural application in which G is non-abelian.

Proposition 3.3. *Let G be any semisimple group of Lie type. Let B be a Borel subgroup of G defined over a field K. Let T be a maximal torus of G contained in B, and let U be the maximal unipotent subgroup of B. Assume that the exponential map* $\exp : \mathfrak{u} \to U$ *from the Lie algebra $\mathfrak{u}$ of U to U itself is injective.*

Let $D \subset T(\overline{K})$ be a finite set such that, for every root α of G relative to T, the restriction $\alpha|_D$ is injective. Then, for any finite set $A \subset U(\overline{K})$,

$$|(A \cup D)_{20} \cap U(\overline{K})| > \frac{|A||D||\mathscr{O}|}{|A||D| + |\mathscr{O}|}, \tag{3.14}$$

where $\mathscr{O}$ is the subgroup of $U(\overline{K})$ generated by $\{tut^{-1} : t \in \langle D\rangle, u \in \langle A\rangle\}$.

If $\mathscr{O}$ is infinite and A, D are finite, (3.14) reads as follows: $|(A\cup D)_{20} \cap U(\overline{K})| \geq |A||D|$. (We shall work only with finite fields K, and thus $\mathscr{O}$, A and D will always be finite; we only mention the case of infinite sets in passing.)

If G is a subgroup of $\mathrm{GL}_n(K)$, K a field of characteristic $= 0$ or $\geq n$, then the exponential map $\exp : \mathfrak{u} \to U$ is invertible and, in particular, injective. (The Taylor series for $\exp x$ and $\log x$ terminate at x^{n-1}, and the denominators of the coefficients of the terms up to x^{n-1} in either series are not divisible by any primes $\geq n$.)

Proof. We will apply Proposition 3.1 with $G = U(\overline{K})$ and Υ equal to the group $\Upsilon = \{y_t : t \in T\}$ of automorphisms of $U(\overline{K})$, where

$$y_t : u \mapsto tut^{-1}.$$

The set A will be as given, the set Y will be $\{y_t : t \in D\}$, and, finally, $A_0 = A \cup A^{-1}$ and $Y_0 = Y \cup Y^{-1}$.

We need only check condition (3.1). Let t be an element of $D^{-1}D$ other than the identity. Because $\alpha|_D$ is injective for every root α, we know that $\alpha(t) \neq 1$ for every root α. We need to show that, if $tgt^{-1} = g$ for some $g \in U(K)$, then g is the identity.

We may write[3] $g = \exp(\vec{v})$, where $\vec{v}$ lies on the Lie algebra $\mathfrak{u}$ of U. We have $tgt^{-1} = t\exp(\vec{v})t^{-1} = \exp(\mathrm{Ad}_t(\vec{v}))$. Because the exponential map $\exp : \mathfrak{u} \to U$ is injective, we shall have $tgt^{-1} = g$ if and only if $\mathrm{Ad}_t(\vec{v}) = \vec{v}$.

[3]For $G = \mathrm{SL}_n$, what follows amounts to the following prosaic observation: if t is a diagonal matrix with distinct eigenvalues, and g is an upper-triangular matrix with 1's on the diagonal, then $tgt^{-1} = g$ can be true only if g is the identity.

We may write $\vec{v}$ as a sum $\sum_\alpha \vec{v}_\alpha$ of elements $\vec{v}_\alpha$ of the root spaces corresponding to the positive roots α. Then

$$\mathrm{Ad}_t(\vec{v}) = \mathrm{Ad}_t(\sum_\alpha \vec{v}_\alpha) = \sum_\alpha \mathrm{Ad}_t(\vec{v}_\alpha) = \sum_\alpha \alpha(t) \cdot \vec{v}_\alpha.$$

Since $\alpha(t) \neq 1$ for every root α, we conclude that $\mathrm{Ad}_t(\vec{v}) = \vec{v}$ implies $\vec{v}_\alpha = 0$ for every α, i.e., $\vec{v} = 0$. Hence $g = \exp(\vec{v})$ is the identity. □

Say that we want to apply Prop. 3.3 to the study of growth in Borel subgroups. An obvious question arises: how do we obtain a large set of diagonal elements $D \subset T(\overline{K})$ and a large set of unipotent elements $A \subset U(\overline{K})$? Unipotent elements can generally be got by means of an easy pigeonhole argument, as, for any two matrices g_1, g_2 having distinct eigenvalues and lying in the same conjugacy class in B, the quotient $g_1^{-1} g_2$ is unipotent. Obtaining diagonal elements is a harder problem, but we will need to study it at any rate; we will solve it in §5.3. Once we have enough diagonal elements, we will usually be able to obtain a large subset D of them on which every root is injective by means of a covering argument.

An exception occurs when we can obtain many commuting elements inside the kernel of a root. We can still use Prop. 3.1; the set D or Y being used need not lie in a torus - it can lie in any abelian subgroup. Some case work is needed, however. In §7.3, we will study the matter in detail for the special case $G = \mathrm{SL}_3$.

3.3. **A sum-product theorem in $(\mathbb{Z}/p\mathbb{Z})^n$.** We will prove a sum-product theorem of $(\mathbb{Z}/p\mathbb{Z})^n$. As a matter of fact, we shall use only the case $n = 2$; a theorem close to the one we need was already proven for $n = 2$ by Bourgain ([B]). However, [B] requires the assumption that $|A| > p^\epsilon$. We cannot assume $|A| > p^\epsilon$ in our applications. We thus need to prove a sum-product theorem ourselves without that restriction. (Our statement will be less precise than that in [B] in another respect.)

We may as well start by reproving the sum-product theorem for $\mathbb{Z}/p\mathbb{Z}$ using Proposition 3.1. In this we are simply following upon the steps of [GK] or [B2]. The matter will take only a few lines.

Lemma 3.4. *Let p be a prime. Let $A \subset \mathbb{Z}/p\mathbb{Z}$. Assume $|A| < p^{1-\delta}$, $\delta > 0$. Then*

$$\text{either} \quad |A \cdot A| \gg |A|^{1+\epsilon} \quad \text{or} \quad |A + A| \gg |A|^{1+\epsilon}, \tag{3.15}$$

where $\epsilon > 0$ and the implied constants depend only on δ.

Proof. Suppose (3.15) does not hold with implied constants equal to 1. Then, by the Katz-Tao Lemma ([TV], Lemma 2.53), there is a subset $A' \subset A$ with $|A'| \geq \frac{1}{2}|A|^{1-\epsilon} - 1$ and

$$|A' \cdot A' - A' \cdot A'| \ll |A|^{1+O(\epsilon)}, \tag{3.16}$$

where the implied constant is absolute. We have to show that this is impossible.

Let $G = \mathbb{Z}/p\mathbb{Z}$, $\Upsilon = (\mathbb{Z}/p\mathbb{Z})^*$ (acting on G by multiplication), $A_0 = \{1\}$, $Y_0 = \emptyset$, and set both A and Y in the statement of Prop. 3.1 equal to A'. Since there are no zero divisors in $\mathbb{Z}/p\mathbb{Z}$, condition (3.1) is satisfied. Thus, we may apply Prop. 3.1, and obtain that

$$|A' \cdot A' + A' \cdot A - A' \cdot A' + A' \cdot A' - A' \cdot A' + A' \cdot A'| \geq \frac{1}{2}\min(|A|^2, p) \gg |A|^{1+\delta}.$$

Hence, by the Plünnecke-Ruzsa estimates ([TV], Cor. 6.29),

$$|A' \cdot A' - A' \cdot A'| \gg |A'|^{1+\frac{\delta}{6}}.$$

For any $\epsilon < \delta/6$, this is in contradiction to (3.16) provided that $|A|$ is larger than a constant depending only on δ and ϵ. We may in fact assume that $|A|$ is larger than a constant, as otherwise (3.15) is trivial. We have reached a contradiction. ☐

Proposition 3.5. *Let p be a prime. Let $A \subset (\mathbb{Z}/p\mathbb{Z})^n$, $n \geq 1$. Assume that either $|A| < p^{1-\delta}$, $\delta > 0$, or $p^{k+\delta} < |A| < p^{k+1-\delta}$, $\delta > 0$, $1 \leq k < n$. Then*

$$\text{either} \quad |A \cdot A| \gg |A|^{1+\epsilon} \qquad \text{or} \qquad |A + A| \gg |A|^{1+\epsilon},$$

where $\epsilon > 0$ and the implied constant depend only on n and δ.

Stronger statements are possible. Doing away with the conditions $|A| < p^{1-\delta}$, $p^{k+\delta} < |A| < p^{k+1-\delta}$ would take some detailed case work and a catalogue of counterexamples: consider $A = (\mathbb{Z}/p\mathbb{Z}) \times \{0\} \times \cdots \times \{0\}$, for example.

Proof. We proceed by induction on n. For $n = 1$, the statement is true by Lemma 3.4. Let $\pi_j : (\mathbb{Z}/p\mathbb{Z})^n \to \mathbb{Z}/p\mathbb{Z}$ be the projection map to the jth coordinate; let $\pi_{\backslash j} : (\mathbb{Z}/p\mathbb{Z})^n \to (\mathbb{Z}/p\mathbb{Z})^{n-1}$ be the projection map to all coordinates save the jth one. We may assume that A is a subset of $((\mathbb{Z}/p\mathbb{Z})^*)^n$: if at least half of A lies in $((\mathbb{Z}/p\mathbb{Z})^*)^n$, then we may work with $A \cap ((\mathbb{Z}/p\mathbb{Z})^*)^n$ instead of A, and if more than half of A lies outside $((\mathbb{Z}/p\mathbb{Z})^*)^n$, then $|A \cap \pi_j^{-1}(0)| > \frac{1}{2n}|A|$ for some j, and we may pass to $\pi_{\backslash j}(A \cap \pi_j^{-1}(0))$ and apply the inductive hypothesis.

Assume, then, that $n > 1$ and $A \subset ((\mathbb{Z}/p\mathbb{Z})^*)^n$. Suppose first that $\min(|A|^{1/n}, p^{\delta/n}) \leq |\pi_1(A)| \leq p^{1-\delta/n}$. Then, by Lemma 3.4, either

$$\begin{aligned} |\pi_1(A \cdot A)| &= |\pi_1(A) \cdot \pi_1(A)| \gg |\pi_1(A)|^{1+\epsilon} \\ &\geq \min\left(p^{\frac{\delta\epsilon}{n}}, |A|^{\epsilon/n}\right) \cdot |\pi_1(A)| \geq |A|^{\delta e/n^2} \cdot |\pi_1(A)| \end{aligned} \tag{3.17}$$

or

$$\begin{aligned} |\pi_1(A + A)| &= |\pi_1(A) + \pi_1(A)| \gg |\pi_1(A)|^{1+\epsilon} \\ &\geq \min\left(p^{\frac{\delta\epsilon}{n}}, |A|^{\epsilon/n}\right) \cdot |\pi_1(A)| \geq |A|^{\delta\epsilon/n^2} \cdot |\pi_1(A)|. \end{aligned} \tag{3.18}$$

Let $x \in \mathbb{Z}/p\mathbb{Z}$ be such that the number of elements of $S_x = A \cap (\pi_1^{-1}(\{x\})$ is maximal. Then $|S_x| \geq \frac{|A|}{|\pi_1(A)|}$. Let us examine the consequences of (3.17) and (3.18).

If (3.17) holds, then

$$\begin{aligned} |A \cdot A \cdot A| \geq |A \cdot A \cdot S_x| \geq |\pi_1(A \cdot A)||S_x| &\geq |A|^{\delta\epsilon/n^2}|\pi_1(A)||S_x| \\ \geq |A|^{\delta\epsilon/n^2}|\pi_1(A)| \cdot \frac{|A|}{|\pi_1(A)|} &= |A|^{1+\delta\epsilon/n^2}. \end{aligned}$$

Since $(\mathbb{Z}/p\mathbb{Z})^*$ is abelian, we may use the Plünnecke's inequality ([TV], Cor. 6.28) to obtain

$$|A \cdot A| \gg |A|^{1+\frac{\delta\epsilon}{3n^2}}. \tag{3.19}$$

If (3.18) holds, then one shows that

$$|A+A| \gg |A|^{1+\frac{\delta\epsilon}{3n^2}}$$

in exactly the same way that we showed (3.19). Thus we are done with the case $p^{\delta/n} \leq |\pi_1(A)| \leq p^{1-\delta/n}$.

Suppose now that either $|\pi(A)| < \min(|A|^{1/n}, p^{\delta/n})$ or $p^{1-\delta/n} < |\pi(A)| \leq p$. Choose $x \in \mathbb{Z}/p\mathbb{Z}$ such that the number of elements of $S_x = A \cap (\pi_1^{-1}(\{x\}))$ is maximal. If $|A| < p^{1-\delta}$, then $|A|^{\frac{n-1}{n}} < |S_x| < p^{1-\delta}$; if $p^{k+\delta} < |A| < p^{k+1-\delta}$, $k \geq 1$, then either $p^{k+\frac{n-1}{n}\delta} < |S_x| < p^{k+1-\delta}$ or $p^{k-1+\delta} < |S_x| < p^{k-\frac{n-1}{n}\delta}$. In all of these cases, we may apply the inductive hypothesis (with $\frac{n-1}{n}\delta$ instead of δ), and, moreover, $|S_x| > |A|^{\frac{\delta}{n}}$. Hence either

$$|S_x \cdot S_x| \gg |S_x|^{1+\epsilon} \geq |A|^{\frac{\delta\epsilon}{n}} \cdot |S_x| \tag{3.20}$$

or

$$|S_x + S_x| \gg |S_x|^{1+\epsilon} \geq |A|^{\frac{\delta\epsilon}{n}} \cdot |S_x|. \tag{3.21}$$

If (3.20) holds, then

$$\begin{aligned}|A \cdot A \cdot A| &\geq |A \cdot S_x \cdot S_x| \geq |\pi_1(A)| \cdot |S_x \cdot S_x| \\ &\gg |\pi_1(A)| \cdot |A|^{\frac{\delta\epsilon}{n}} \cdot |S_x| \geq |A|^{1+\frac{\delta\epsilon}{n}}\end{aligned}$$

and so, by Plünnecke's inequality,

$$|A \cdot A| \gg |A|^{1+\frac{\delta\epsilon}{3n}}.$$

If (3.21) holds instead, we obtain in exactly the same way that

$$|A+A| \gg |A|^{1+\frac{\delta\epsilon}{3n}}.$$

□

Remark. There is now an alternative route to the one taken in this subsection. Instead of proceeding as above, one may derive Prop. 3.5 from [T2, Thm. 5.4].

3.4. **Linear relations over rings.** The consequences of the sum-product theorem we are about to derive are closely related to incidence theorems. Such theorems have been linked to sum-product phenomena ever since Elekes's brief and elegant proof [E] of the sum-product theorem over $\mathbb{R}$ (originally due to Erdös and Szemerédi [ES]) by means of an incidence theorem over $\mathbb{R}$ (first proven by Szemerédi and Trotter [ST]). Over finite fields, the topological arguments that can be used to prove incidence theorems over $\mathbb{R}$ do not work: over $\mathbb{Z}/p\mathbb{Z}$, a line does not divide the plane into two halves. Thus, it seems necessary to prove incidence theorems using sum-product results, rather than the other way around. This is exactly what was done in [BKT, §6]: Bourgain, Katz and Tao proved an incidence theorem over $\mathbb{Z}/p\mathbb{Z}$ using their sum-product theorem.

We shall now prove – over $(\mathbb{Z}/p\mathbb{Z})^n$, not over $\mathbb{Z}/p\mathbb{Z}$ – some results that are not quite the same as incidence theorems, but are akin to them. The basic idea is the same: we are to show that there cannot be too many linear relations among too few objects.

We will need a very simple counting lemma.

Lemma 3.6. *Let A, B be finite sets. Let $S \subset A \times B$. For every $a \in A$, let $B_a = \{b \in B : (a,b) \in S\}$.*

Then there is an $a_0 \in A$ such that

$$\sum_a |B_{a_0} \cap B_a| \geq \frac{|S|^2}{|A||B|}.$$

Proof. For every $s \in S$, let $A_b = \{a \in A : (a,b) \in S\}$. Then

$$\begin{aligned}\sum_{b\in B} |A_b|^2 &= \sum_{a_1\in A}\sum_{a_2\in A} |\{b \in B : a_1, a_2 \in A_b\}| \\ &= \sum_{a_1\in A}\sum_{a_2\in A} |B_{a_1} \cap B_{a_2}|.\end{aligned}$$

Thus, if we let a_0 be such that $\sum_{a\in A} |B_{a_0} \cap B_a|$ is maximal,

$$\sum_{a\in A} |B_{a_0} \cap B_a| \geq \frac{1}{|A|} \sum_{b\in B} |A_b|^2.$$

At the same time, by Cauchy's inequality,

$$\sum_{b\in B} |A_b|^2 \geq \frac{1}{|B|} \left(\sum_{b\in B} |A_b|\right)^2.$$

Finally,

$$\sum_{b\in B} |A_b| = |S|$$

and so we are done. □

The proposition we are about to prove can be summarised as follows. Let X be a subset of a field (or a ring). Suppose that there are many linear relations satisfied by many $(n+1)$-tuples of elements of X. Then X has a large subset that grows neither under addition nor under multiplication.

Proposition 3.7. *Let R be a ring. Let $X \subset R$, $Y \subset (R^*)^n$, $n \geq 2$. Assume that the projection $\pi_1 : (R^*)^n \to R^*$ given by $(y_1, y_2, \ldots, y_n) \mapsto y_1$ is injective on Y. Assume as well that $|Y| > c|X|$, $0 < c < 1$.*

For each $\vec{y}$, let $X_{\vec{y}}$ be a subset of X^n with $|X_{\vec{y}}| > c|X|^n$. Suppose

$$\vec{y} \cdot X_{\vec{y}} = \{y_1x_1 + \ldots + y_nx_n : \vec{x} \in X_{\vec{y}}\} \tag{3.22}$$

is contained in X for every $\vec{y} \in Y$.

Then there is a subset $X_0 \subset X$ such that $|X_0| \gg c^C|X|$ and

$$|X_0 + X_0| \ll \frac{1}{c^C}|X_0|, \qquad |X_0 \cdot X_0| \ll \frac{1}{c^C}|X_0|,$$

where C is a positive absolute constant and the implied constants are absolute.

It would be desirable to replace both the assumption that $\pi_1|_Y$ is injective and the assumption that $|Y| > c|X|$ by much weaker postulates. The statement as it stands will do for SL_3, but probably not for SL_n, $n > 3$. Weakening the assumptions would probably involve using the techniques in [TV], §2.7, instead of the Balog-Gowers-Szemerédi theorem. (The following proof is, incidentally, the only place where the Balog-Gowers-Szemerédi theorem is used in this paper.)

Proof. Let X' be the set of all $x \in X$ such that

$$|\{((x_2, x_3, \ldots, x_n), \vec{y}) \in X^{n-1} \times Y : (x_1, x_2, \ldots, x_n) \in X_{\vec{y}}\}| > \frac{1}{2} c |X|^{n-1} |Y|.$$

We have

$$\sum_{\vec{y} \in Y} |\{(x_1, x_2, \ldots, x_n) \in X_{\vec{y}} : x_1 \in X'\}| > \frac{1}{2} c |X|^n |Y| \tag{3.23}$$

as $\sum_{\vec{y} \in Y} |\{(x_1, x_2, \ldots, x_n) \in X_{\vec{y}}\}| > \sum_{\vec{y} \in Y} c|X|^n = c|X|^n|Y|$ and the contribution of the terms with $x_1 \notin X'$ is clearly $\leq \frac{1}{2} c|X|^n|Y|$. Immediately from (3.23), $|X'| > \frac{1}{2}c|X|$.

Define $X'_{\vec{y}} = \{(x_1, x_2, \ldots, x_n) \in X_{\vec{y}} : x_1 \in X'\}$. Let $\vec{y}_0 \in Y$ be such that $|X'_{\vec{y}_0}|$ is maximal. By (3.23) and the pigeonhole principle, $|X'_{\vec{y}_0}| > \frac{1}{2}c|X|^n$. Now, by (3.22),

$$|\{y_{0,1}x_1 + y_{0,2}x_2 + \ldots + y_{0,n}x_n : (x_1, x_2, \ldots, x_n) \in X_{\vec{y}_0}\}'| \leq |X|.$$

We apply the Balog-Szemerédi-Gowers theorem (2.3) with

$$A_1 = y_{0,1} X'_{\vec{y}_0},\ A_2 = y_{0,2} X_{\vec{y}_0},\ \ldots,\ A_n = y_{0,n} X_{\vec{y}_0}$$

and

$$S = \{(y_{0,1}x_1, \ldots, y_{0,n}x_n) : (x_1, x_2, \ldots, x_n) \in X'_{\vec{y}_0}\},$$

and obtain that there is a subset $X'' \subset X'$ such that

$$|X''| \gg c^{C_1} |X'| \quad \text{and} \quad |y_{0,1}X'' + y_{0,1}X''| \ll \frac{1}{c^{C_1}} |X''|, \tag{3.24}$$

where C_1 is a positive absolute constant and the implied constants are also absolute. Obviously, $|X'' + X''| = |y_{0,1}X'' + y_{0,1}X''|$, and so $|X'' + X''| \ll \frac{1}{c^{C_1}} |X''|$.

Apply Lemma 3.6 with $A = X''$, $B = Y \times X^{n-1}$, and

$$S = \{(x_1, (\vec{y}, (x_2, x_3, \ldots, x_n))) \in A \times B : (x_1, x_2, \ldots, x_n) \in X_{\vec{y}}\}.$$

We obtain that there is a $x_0 \in X''$ such that

$$\sum_{x \in X''} |B_{x_0} \cap B_x| \geq \frac{|S|^2}{|X''| \cdot |Y| |X|^{n-1}},$$

where $B_x = \{(y, (x_2, \ldots, x_n)) \in Y \times X^{n-1} : (x, x_2, \ldots, x_n) \in X_{\vec{y}}\}$. Now, since $X'' \subset X'$, we obtain from the definition of X' that

$$|S| > |X''| \cdot \frac{1}{2} c |X|^{n-1} |Y|.$$

Thus

$$\sum_{x \in X''} |B_{x_0} \cap B_x| > \frac{1}{4} c^2 |X''| |X|^{n-1} |Y|. \tag{3.25}$$

Define the map $f : Y \times X^{n-1} \to R$ by

$$f(\vec{y}, (x_2, x_3, \ldots, x_n)) = x_0 y_1 + x_2 y_2 + x_3 y_3 + \cdots + x_n y_n.$$

Since (3.22) is a subset of X, the set $f(B_{x_0})$ is a subset of X. Let $r_0 \in f(B_{x_0})$ be such that

$$\sum_{x\in X''} |\{b \in B_{x_0} \cap B_x : f(b) = r_0\}|$$

is maximal. By (3.25), $\sum_{x\in X''} |\{b \in B_{x_0} \cap B_x : f(b) = r_0\}|$ is then $\geq \frac{1}{4}c^2|X''||X|^{n-2}|Y|$. For $x \in X''$ and $\vec{y} \in Y$ given, there are at most $|X|^{n-2}$ elements of B_x such that $f(b) = r_0$. (This is so because, if x_2 varies and $y, x_3, x_4, \ldots, x_n$ are held fixed, then $f(b)$ varies with x_2.) Thus, there are at least $\frac{1}{4}c^2|X''||Y|$ pairs $(x_1, \vec{y}) \in X'' \times Y$ such that there is at least one tuple $(x_2, x_3, \ldots, x_n) \in X^{n-1}$ for which

$$(\vec{y}, (x_1, x_2, \ldots, x_n)) \in B_{x_0} \cap B_x$$

and

$$x_0 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + \cdots + x_n \cdot y_n = r.$$

Let $S' \subset X'' \times Y$ be the set of all such pairs $(x_1, \vec{y})$. For any $(x_1, \vec{y}) \in S'$, there are $x_2, x_3, \ldots, x_n \in X$ such that

$$\begin{aligned} x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n y_n &= (x_1 - x_0) \cdot y_1 + x_0 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n \\ &= (x_1 - x_0) \cdot y_1 + r \end{aligned}$$

and

$$x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n \in X.$$

Thus

$$\{(x_1 - x_0) \cdot y_1 : (x_1, \vec{y}) \in S'\} \subset X - r.$$

Hence

$$|\{x \cdot y : (x, y) \in S''\}| \leq |X|,$$

where $S'' = \{(x, y) \in (X'' - x_0) \times \pi_1(Y) : (x + x_0, \pi_1^{-1}(y)) \in S'\}$. (Recall that $\pi_1 : (y_1, y_2, \ldots, y_n) \mapsto y_1$ is injective on Y.) Clearly $|S''| = |S'|$, and so $|S''| \geq \frac{1}{4}c^2|X''||Y|$.

We now apply the Balog-Szemerédi-Gowers theorem (Prop. 2.3) again, this time with multiplication, not addition, as the operation, and the following inputs: $n = 2$, $A_1 = X'' - x_0$, $A_2 = \pi(Y)$, $S = S''$. We obtain that there is a subset $X''' \subset (X'' - x_0)$ with

$$|X'''| \gg c^{C_2}|X''| \qquad \text{and} \qquad |X''' \cdot X'''| \ll \frac{1}{c^{C_2}}|X'''|.$$

At the same time, because of (3.24),

$$|X''' + X'''| = |(X''' - x_0) + (X''' - x_0)| \ll \frac{1}{c^{C_1}}|X''| \ll \frac{1}{c^{C_1+C_2}}|X'''|.$$

We let $X_0 = X'''$ and are done. □

We can finally state and prove what we worked for in this subsection.

Corollary 3.8. *Let $R = (\mathbb{Z}/p\mathbb{Z})^m$, $m \geq 1$. Let $X \subset R$, $Y \subset (R^*)^n$, $n \geq 2$. Assume that, for some $j \in \{1, 2, \ldots, n\}$, the projection $\pi_j : (R^*)^n \to R^*$ given by $(y_1, y_2, \ldots, y_n) \mapsto y_j$ is injective on Y. Assume that either $|X| \leq p^{1-\delta}$, $\delta > 0$, or $p^{k+\delta} \leq |X| \leq p^{k+1-\delta}$ for some $k \geq 1$, $\delta > 0$.*

For each $\vec{y}$, let $X_{\vec{y}}$ be a subset of X^n such that

$$\vec{y} \cdot X_{\vec{y}} = \{y_1 x_1 + \ldots + y_n x_n : \vec{x} \in X_{\vec{y}}\}$$

is contained in X*. Then either*

$$|Y| \ll |X|^{1-\eta} \quad \text{or} \quad |X_{\vec{y}}| \ll |X|^{n-\eta} \text{ for some } \vec{y} \in Y, \tag{3.26}$$

where $\eta > 0$ *and the implied constants depend only on* δ *and* m*.*

We could explain this as follows, leaving a few conditions aside. Let X be a subset of a ring R. Consider an n-dimensional box X^n. Let there be many ($\geq |X|^{1-\eta}$, η small) linear forms f such that for each form f there are many ($\geq |X|^{n-\eta}$) elements of the box on which the form f takes values in X. Corollary 3.8 shows that the situation just described *cannot* happen.

Proof. Immediate from Prop. 3.7 and Prop. 3.5. (If $j \neq 1$, permute the first and jth coordinates of $(R^*)^n$ before applying Prop. 3.7.) □

4. Escape, non-singularity and their conditions

Much of our work will consist in showing that certain statements are generically true in an effective sense – that is to say, they are true when their parameters lie outside a variety of positive codimension *and bounded degree.* We will then obtain quantitative bounds from these effective results by means of the technique of *escape from subvarieties.*

The following will be a typical situation. Say we are able to show that a map $f : G \to V$ from an algebraic group G/K to a variety V/K is non-singular almost everywhere in an effective sense, meaning that there is a variety $X_G \subset G$ of positive codimension in G and bounded degree such that, for every point y in the image of the restriction $g := f|_{G\setminus X_G}$, the preimage $g^{-1}(y)$ of y consists of a bounded number of points (i.e., it is the union of a bounded number of irreducible zero-dimensional varieties). This is a useful situation to be in, as then, for any finite subset $E \subset G(K) \setminus X_G(K)$, the image $f(E)$ satisfies $|f(E)| \gg |E|$; since we are investigating growth, we are certainly interested in maps that do not make sets smaller.

Suppose we are simply given a set $E \subset G(K)$. Then, under a very broad set of circumstances, escape from subvarieties will give us that there are $\gg |E|$ elements of E_k lying in $G(K) \setminus X_G(K)$, where k is bounded by a constant. Call the set of such elements E'. Then, by what we said before, $|f(E')| \gg |E'|$, and so $|f(E_k)| \geq |f(E')| \gg |E'| \gg |E|$, which is a conclusion we will often desire.

4.1. Escape from subvarieties.

Eskin, Mozes and Oh [EMO] have shown how to escape from varieties by means of a group action. While their result was formulated over $\mathbb{C}$, it carries over easily to other fields. The following proposition is based closely on [EMO, Prop. 3.2].

Proposition 4.1. *Let* G *be a group. Consider a linear representation of* G *on a vector space* $\mathbb{A}^n(K)$ *over a field* K*. Let* V *be an affine subvariety of* $\mathbb{A}^n$*.*

Let A *be a subset of* G*; let* $\mathscr{O}$ *be an* $\langle A\rangle$*-orbit in* $\mathbb{A}^n(K)$ *not contained in* V*. Then there are constants* $\eta > 0$ *and* m *depending only on* $\overrightarrow{deg}(V)$ *such that, for every* $x \in \mathscr{O}$*, there are at least* $\max(1, \eta|A|)$ *elements* $g \in A_m$ *such that* $gx \notin V$*.*

This may be phrased as follows: one can escape from V by the action of the elements of A.

Proof. Let us begin by showing that there are elements $g_1, \dots, g_l \in A_r$ such that, for every $x \in \mathscr{O}$, at least one of the $g_i \cdot x$'s is not in V. (Here l and r are bounded in terms of $\overrightarrow{\deg} V$ alone.) We will proceed by descent (that is, induction) on $\overrightarrow{\deg} V$, paying special attention to the number s_V of irreducible components of V of maximal dimension $\dim V$. (Notice that s_V is bounded in terms of $\overrightarrow{\deg} V$: in fact, $s_V \leq (\overrightarrow{\deg} V)_{\dim V}$.)

We shall always pass from V to a variety V' with either (a) $\dim V' < \dim V$ or (b) $\dim V' = \dim V$ and $s_{V'} < s_V$. Moreover, $\overrightarrow{\deg} V'$ will be bounded in terms of $\overrightarrow{\deg} V$ alone. We will iterate until we arrive at a variety V' of dimension 0 with $s_{V'} = 0$, i.e., an empty variety. It is clear that this process terminates in a number of steps bounded in terms of $\overrightarrow{\deg} V$ alone.

Let V_+ be the union of all irreducible components of V of maximal dimension (i.e., dimension $\dim V$). If V_+ and $\mathscr{O}$ are disjoint, we set $V' = V \setminus V_+$ and are done. Suppose otherwise. Since $\mathscr{O}$ is not contained in V_+, we can find $x_0 \in V_+ \cap \mathscr{O}$, $g \in A \cup A^{-1}$ such that $g x_0 \notin V_+$, i.e., $x_0 \notin g^{-1} V_+$. Hence the set of components of maximal dimension $\dim V$ in V is not the same as the set of components of maximal dimension $\dim g^{-1} V = \dim V$ in $g^{-1} V$. It follows that $V' = g^{-1} V \cap V$ does not contain V_+, and thus has fewer components of dimension $\dim V$ than V has.

We have thus passed from V to V', where either (a) $\dim V' < \dim V$ or (b) $\dim V' = \dim V$ and $s'_V < s_V$. Bezout's theorem assures us that $\overrightarrow{\deg} V'$ is bounded in terms of $\overrightarrow{\deg} V$ alone. By the inductive hypothesis, we already know that there are $g'_1, \dots, g'_{l'} \in A_{r'}$ such that, for every $x \in \mathscr{O}$, at least one of the $g'_i \cdot x$'s is not in V'. (Here l' and r' are bounded in terms of $\overrightarrow{\deg} V'$ alone.) Since at least one of the $g'_i \cdot x$'s is not in $V' = g^{-1} V \cap V$, either one of the $g'_i \cdot x$'s is not in V or one of the $g'_i \cdot x$'s is not in $g^{-1} V$, i.e., one of the $g g'_i \cdot x$'s is not in V. Set

$$g_1 = g'_1,\ g_2 = g'_2,\ \dots,\ g_{l'} = g'_{l'}$$
$$g_{l'+1} = g g'_1,\ g_{l'+2} = g g'_2,\ \dots,\ g_{2l'} = g g'_{l'}, \qquad l = 2l'.$$

(As can be seen, $g_i \in A_r$, where $r = r' + 1$.) We conclude that, for every $x \in \mathscr{O}$, at least one of the $g_i \cdot x$'s is not in V.

The rest is easy: for each $x \in \mathscr{O}$ and each $g \in A$, at least one of the elements $g_i g \cdot x$, $1 \leq i \leq l$ ($g_i \in A_r$) will not be in V. Each possible $g_i g$ can occur for at most l different elements $g \in A$; thus, there are at least $\min(1, |A|/l)$ elements $h = g_i g$ of A_{r+1} such that $hx \notin V$. $\square$

Many statements can be proven by the same kind of induction that one uses to prove escape.

Proposition 4.2. *Let K be a field. Let G/K be an algebraic subgroup of GL_n / K. Let S be a subgroup of $G(\overline{K})$ contained in a subvariety V of G of positive codimension.*

Then S is contained in an algebraic subgroup H of G of positive codimension and degree bounded in terms of $\overrightarrow{deg}(V)$ alone.

Proof. We shall show that S is contained in the stabiliser of a subvariety of G, and that this stabiliser satisfies the conditions required of H in the statement. We will proceed by induction on $\overrightarrow{\deg} V$, focusing on $\dim(V)$ and s_V (defined as in the proof of 4.1), which it encodes. We shall always pass from V to a variety V' with either (a) $\dim(V') < \dim(V)$ or (b) $\dim(V') = \dim(V)$ and $s_{V'} < s_V$. Moreover, $\overrightarrow{\deg} V'$ will be bounded in terms of $\overrightarrow{\deg} V$ alone. We will iterate until we either find an algebraic group containing S or arrive at a variety V' with dimension 0 and $s_{V'} = 0$ (i.e., the empty variety).

Let V_+ be the union of irreducible components of V of dimension $\dim(V)$. If $V_+ \neq V_+^{-1}$, we set $V' = V \cap V^{-1}$; we shall have either (a) $\dim(V') < \dim(V)$ or (b) $\dim(V') = \dim(V)$ and $s_{V'} < s_V$, and, by Bezout's theorem, the degree of V' is bounded in terms of the degree of V. Since S is a group, $S = S^{-1} \subset V^{-1}(\overline{K})$, and so $S \subset (V \cap V^{-1})(\overline{K}) = V'(\overline{K})$. We then use the inductive hypothesis and are done. We may thus assume from here on that we are in the other case, viz., $V_+ = V_+^{-1}$.

Suppose first that there is a pair $(g, x) \in (S, V_+(\overline{K}))$ such that $g \cdot x$ lies outside $V_+(\overline{K})$. Then $V' = gV \cap V$ has either (a) $\dim(V') < \dim(V)$ or (b) $\dim(V') = \dim(V)$ and $s_{V'} < s_V$, and, by Bezout's theorem, the degree of V' is bounded in terms of the degree of V. Since S is a group, $S = gS \subset gV(\overline{K})$, and so $S \subset (V \cap gV)(\overline{K})$. We then use the inductive hypothesis and are done. We may thus assume that there is no pair $(g, x) \in (S, V_+(\overline{K}))$ such that $g \cdot x$ lies outside $V_+(\overline{K})$.

Suppose now that there is a pair $y, z \in V_+(\overline{K})$ such that $y \cdot z^{-1} \notin V_+(\overline{K})$. Then $V' = Vz^{-1} \cap V$ has either (a) $\dim(V') < \dim(V)$ or (b) $\dim(V') = \dim(V)$ and $s_{V'} < s_V$, etc. At the same time, by our previous assumption, there is no $g \in S$ such that gz lies outside $V_+(\overline{K})$; hence $S \subset Vz^{-1}$. Since $S \subset V$, we conclude that $S \subset Vz^{-1} \cap V = V'$. We use the inductive hypothesis and are done.

We are left with the case where $V_+ = V_+^{-1}$ and there is no pair $y, z \in V_+(\overline{K})$ such that $y \cdot z^{-1} \notin V_+(\overline{K})$. Then V_+ is an algebraic group. We are assuming that there is no pair $(g, x) \in (S, V_+(\overline{K}))$ such that $g \cdot x$ lies outside $V_+(\overline{K})$; since $V_+(\overline{K})$ is a group, it contains the identity, and thus we have that there is no $g \in S$ such that $g \cdot e = g$ lies outside $V_+(\overline{K})$, i.e., we have $S \subset V_+(\overline{K})$. We set $H = V_+$ and are done. □

Remark. In the above, we have implicitly used the fact that multiplication in a linear algebraic group does not change the degree of the varieties therein: $\overrightarrow{\deg}(gV) = \overrightarrow{\deg}(V)$ (and, in particular, $\deg(gV) = \deg(V)$ for pure-dimensional varieties V). This is the only sense in which we have used "linearity" (i.e., the assumption in Prop. 4.1 that we are working with a linear representation, and the condition in Prop. 4.2 that G be a subgroup of GL_n).

4.2. **Non-singularity and almost-injectivity.** If a map f is injective, then, for every finite subset E of the domain, $|f(E)| = |E|$. If f is such that the preimage $f^{-1}(\{x\})$ of every point x consists of at most k points, then $|f(E)| \leq \frac{1}{k}|E|$. This simple fact lies at the root of several of our arguments.

Remark. Injectivity already played a role in section §3. The idea both there and in the applications we shall later give to the results about to be given here is the following: if f is a map from a product $A \times B$ to a set C, and f is "almost injective" in the sense just described, then, for any $E_1 \subset A$, $E_2 \subset B$, the image $f(A, B)$ has $\leq \frac{1}{k}|A||B|$ elements. In other words, we have obtained a rather strong kind of growth, provided that f can be defined by means of "allowable" operations, e.g., group operations involving only already accessible quantities.

First, let us see how non-singularity gives us "almost injectivity". (A regular map $f : X \mapsto Y$ is said to be *non-singular* at a point $x = x_0$ if its derivative $Df|_{x=x_0}$ at $x = x_0$ is a non-singular linear map from $(TX)_{x=x_0}$ to $(TY)_{y=f(x_0)}$.)

Lemma 4.3. *Let $X \subset \mathbb{A}^{m_1}$ and $Y \subset \mathbb{A}^{m_2}$ be affine varieties defined over a field K. Let $f : X \to Y$ be a regular map. Let V be a subvariety of X such that the derivative $Df|_{x=x_0}$ of f at $x = x_0$ is a nonsingular linear map for all x_0 on X outside V.*

Let $S \subset X(\overline{K}) \setminus V(\overline{K})$. Then

$$|f(S)| \gg_{\overrightarrow{deg}(X),\deg_{\mathrm{pol}}(f)} |S|.$$

Proof. It will be enough to show that the intersection of $X(\overline{K}) \setminus V(\overline{K})$ with the preimage $Z = f^{-1}(y_0)$ of any point y_0 on Y consists of a number of irreducible zero-dimensional varieties (that is, points) bounded in terms of $\deg(V)$ and the degree of the polynomials defining f. Now Z is the intersection $X \cap \bigcap_j X_j$, where X_j, $1 \leq j \leq n$, is the variety in $\mathbb{A}^m$ defined by by $(f(x))_j = (y_0)_j$, where we denote by y_j the jth coordinate of an element y of $\mathbb{A}^n$. Thus, by Bezout's theorem (Lem. 2.4), the degree $\overrightarrow{\deg}(Z)$ of Z is $\ll_{\overrightarrow{\deg}(X),\overrightarrow{\deg}(X_1),\dots,\overrightarrow{\deg}(X_n)} 1$. The degree $\overrightarrow{\deg}(X_j)$ of the hypersurface X_j is bounded in terms of the degree of the polynomial $(f(x))_j$, and so

$$\overrightarrow{\deg}(Z) \ll_{\overrightarrow{\deg}(X),\deg((f(x))_1),\dots,\deg((f(x))_n)} 1.$$

Thus, it remains only to show that any point x_0 on Z not lying on V lies on a component of Z of dimension 0.

Suppose it were not so. Then there would be a direction $\vec{v} \neq 0$ such that

$$Df|_{x=x_0}(\vec{v}) = 0;$$

any direction $\vec{v} \neq 0$ on the tangent space to Z at $x = x_0$ would do. Then $Df_{x=x_0}$ would have to be singular. However, this would mean that x_0 would have to lie on V. Contradiction. □

We can avoid a subvariety in an algebraic group by escape from subvarieties.

Lemma 4.4. *Let $G \subset \mathrm{GL}_n$ be an algebraic group defined over a field K. Let V be a subvariety of G such that $V(K)$ is a proper subset of $G(K)$. Let $E \subset G(K)$ be a set of generators of $G(K)$.*

Then

$$|E_k \cap (G(K) \setminus V(K))| \gg_{\overrightarrow{deg}(V)} |E|,$$

where $k \ll_{\overrightarrow{deg}(V)} 1$.

Proof. By escape from subvarieties (Prop. 4.1) with $A = E$, V as given, $x = 1$, and G and $\mathscr{O}$ both equal to $G(K)$. (We are implicitly using the fact that G is contained in an affine space, viz., $\mathbb{A}^{n^2}$.) □

Corollary 4.5. *Let $G \subset \mathrm{GL}_n$ be an algebraic group and $Y \subset \mathbb{A}^m$ an affine variety, both defined over a field K. Let $f : G \to Y$ be a regular map. Let V be a subvariety of G such that $V(K)$ is a proper subset of $G(K)$. Assume that the derivative $Df|_x$ of f at x is a nonsingular linear map for all x on G outside V.*

Let $E \subset G(K)$ be a set of generators of $G(K)$. Then

$$|f(E_k \cap (G(K) \setminus V(K)))| \gg_{\overrightarrow{deg}(G),\overrightarrow{deg}(V),\deg_{\mathrm{pol}}(f)} |E|,$$

where $k \ll_{\overrightarrow{deg}(V)} 1$.

Proof. Immediate from Lemma 4.4 and Lemma 4.3 – the latter with $m_1 = n^2$, $m_2 = m$ and $S = E_k \cap (G(K) \setminus V(K))$. $\square$

Lemma 4.4 has as one of its assumptions that $V(K)$ be a proper subset of $G(K)$. In practice, we will often want to assume instead that V is a proper subvariety of G. Let us see how to obtain the former assumption using the latter one.

In the statement below, *perfect* and *reductive* are standard technical terms (from abstract algebra and the theory of algebraic groups, respectively). The group SL_n (defined over any field K) is reductive, and a product of reductive groups is reductive as well. This is all we will need to know when applying Lem. 4.6 in the present paper.

Lemma 4.6. *Let $G \subset \mathrm{GL}_n$ be an irreducible algebraic group defined over a field K. Assume either that K is perfect or that G is reductive. Let $V/\overline{K}$ be a proper subvariety of G. Then*

$$V(K) \subsetneq G(K)$$

provided that $|K|$ is larger than a constant depending only on n, $\overrightarrow{deg}(V)$ and $\overrightarrow{deg}(G)$.

The assumption that K is perfect or G is reductive will be used only in the case of K infinite. When K is finite, we will use a counting argument that does not require the assumption. (The assumption would be fulfilled in any case, as every finite field is perfect.) When K is infinite, we do not need to assume that $|K|$ is larger than a constant depending only on n, $\overrightarrow{\deg}(V)$ or $\overrightarrow{\deg}(G)$. (Of course, when K is infinite, such an assumption is satisfied immediately anyhow, since $|K| = \infty$.)

Proof. Case 1: K finite. Since G is irreducible and $V \subset G$ is a proper subvariety of G, the maximal dimension m of the components of V is $\leq \dim(G) - 1$. Hence, by (2.3) and the fact that $V \subset \mathrm{GL}_n \subset \mathbb{A}_n^2$,

$$|V(K)| \ll_{\overrightarrow{\deg}(V),n} |K|^{\dim(G)-1}.$$

At the same time, by the Lang-Weil theorem [LW, Thm. 1], the projective closure $\overline{G}$ of G satisfies

$$|\overline{G}(K)| - |K|^{\dim(G)} = O_{\overrightarrow{\deg}(G),n}\left(|K|^{\dim(G)-\frac{1}{2}}\right).$$

Since G is irreducible, so is $\overline{G}$, and hence the intersection of $\overline{G}$ with the hyperplane at infinity (i.e., the part of projective space $\mathbb{P}^{n^2}$ that is not in affine space $\mathbb{A}^{n^2}$) has dimension $< \dim(G)$. We can use either the Lang-Weil theorem or an estimate such as (2.3) again, and obtain

$$|(\overline{G} \setminus G)(K)| \ll_{\overrightarrow{\deg}(G),n} |K|^{\dim(G)-1}.$$

Hence

$$|G(K) \setminus V(K)| = |G(K)| - |V(K)| \gg_{\overrightarrow{\deg}(G),n} |K|^{\dim(G)} - O_{\overrightarrow{\deg}(V),n}(|K|^{\dim(G)-1/2}),$$

which is positive for $|K|$ greater than a constant depending only on $\overrightarrow{\deg}(G)$, $\overrightarrow{\deg}(V)$ and n.

Case 2: K infinite. By [Bor, Cor. V.18.3], $G(K)$ is Zariski-dense in G, i.e., it is not contained in any variety. In particular, $G(K)$ is not contained in V. ☐

It may have seemed odd at first sight that Lem. 4.3 required a map to be non-singular outside a variety. In fact, this is a natural condition; for example, a map between two spaces of the same dimension is non-invertible precisely when the determinant δ of its derivative does not vanish, and we can certainly see that $\delta = 0$ defines a variety.

The following lemma is in the spirit of what was just said. The lemma could be stated in much more general terms; the fact that G will be an algebraic group is helpful but not essential.

Lemma 4.7. *Let $G \subset \mathrm{GL}_n$ be an algebraic group defined over a field K. Let X/K and Y/K be affine varieties such that $\dim(G) = \dim(Y)$. Let $f : X \times G \to Y$ be a regular map. Let $f_x : G \to Y$ be defined by $f_x(g) = f(x, g)$.*

Then there is a subvariety $Z_{X\times G} \subset X \times G$ such that, for all $(x, g_0) \in (X \times G)(\overline{K})$, the derivative

$$(Df_x)|_{g=g_0} : (TG)|_{g=g_0} \to (TY)|_{f(x,g_0)}$$

is non-singular if and only if (x, g_0) does not lie on $Z_{X\times G}$. Moreover,

$$\overrightarrow{deg}(Z_{X\times G}) \ll_{\overrightarrow{deg}(X\times G),\deg_{\mathrm{pol}}(f),n} 1. \tag{4.1}$$

Proof. For $g_0 \in G(\overline{K})$, consider the map

$$g \mapsto f_x(g_0 g). \tag{4.2}$$

Its derivative at $g = I$ is nonsingular precisely when the derivative of f_x at $g = g_0$ is nonsingular. Now, the derivative of (4.2) at $g = I$ is nonsingular precisely when a $\dim(G)$-by-$\dim(G)$ determinant D is non-zero. The entries of the determinant D are polynomials on the entries of g and x; hence, $D = 0$ defines a variety $Z_{X\times G}$. The degree of D (as a polynomial) is bounded in terms of n and $\deg_{\mathrm{pol}}(f)$; thus, $D = 0$ defines a variety of degree $\ll_{n,\deg_{\mathrm{pol}}(f)}$, and so (4.1) follows by Bezout's theorem. ☐

4.3. **Sticking subgroups in generic directions.** Let $H_1, H_2, \dots H_k$ be algebraic subgroups of an algebraic group G/K. Say the tangent spaces $\mathfrak{h}_j \subset \mathfrak{g}$ to $H_j \subset G$ at the origin are such that the dimension of their sum equals the sum of their dimension. Then we might possibly like to conclude that, for any finite sets $E_j \subset H_j(K)$,

$$|E_1 \cdot E_2 \cdots E_k| \gg |E_1| \cdot |E_2| \cdots |E_k|. \tag{4.3}$$

Unfortunately, matters are not so simple. By escape and a few simple arguments, we would indeed be able to obtain such a conclusion, provided that we assumed that E_j generates $H_j(K)$. We will not, however, be able to assume as much in the applications that will come up later: we will be provided with a generating set A of $G(K)$, but not with generating sets of $H_j(K)$. The solution is to multiply *conjugates* of the subgroups $H_j(K)$, rather than the subgroups themselves. Because A generates $G(K)$, we will be able – by escape – to take conjugates of $H_j(K)$ by generic elements of $G(K)$. As we shall see, this is good enough to obtain conclusions much like (4.3) – except for the fact that they will involve conjugates of E_j by elements of A_k, rather than the sets E_j themselves.

We recall that every algebraic group G/K acts on its Lie algebra (i.e., its tangent space $\mathfrak{g}$ at the origin) by conjugation; the *adjoint map* $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is the action of an element $g \in G(\overline{K})$. Recall as well the definition of *linear independence* of subspaces given in §2.5.5.

Lemma 4.8. *Let G be an algebraic group defined over a field K. Let $V_1, V_2, \ldots, V_k$ be linear subspaces of $\mathfrak{g}(\overline{K})$, where $\mathfrak{g}$ is the tangent space to G at the origin. Suppose that there are $g_1, g_2, \ldots, g_k \in G(\overline{K})$ such that the linear spaces*

$$\mathrm{Ad}_{g_1}(V_1), \mathrm{Ad}_{g_2}(V_2), \ldots, \mathrm{Ad}_{g_k}(V_k) \tag{4.4}$$

are linearly independent.

Then there is a proper subvariety $X \subset G^k$ such that, for all $g = (g_1, g_2, \ldots, g_k) \in G^k(\overline{K}) \setminus X(\overline{K})$, the spaces (4.4) are linearly independent. Moreover, $\overrightarrow{\deg}(X) \ll_{\dim(G)} 1$.

Proof. Let $v_{r,1}, v_{r,2}, \ldots, v_{r,l_r}$ be a basis for V_r, $1 \leq r \leq k$. For $g = (g_1, g_2, \ldots, g_k) \in G^k(\overline{K})$, let $w_1(g) = \mathrm{Ad}_{g_1}(v_{1,1})$, $w_2(g) = \mathrm{Ad}_{g_1}(v_{1,2})$,..., $w_{l_1}(g) = \mathrm{Ad}_{g_1}(v_{1,l_1})$, $w_{l_1+1}(g) = \mathrm{Ad}_{g_2}(v_{2,1})$, ..., $w_m(g) = \mathrm{Ad}_{g_k}(v_{k,l_k})$, where $m = \sum_{1\leq r\leq k} l_r$. We are told that the spaces $\mathrm{Ad}_{g_1'}(V_1), \mathrm{Ad}_{g_2'}(V_2), \ldots, \mathrm{Ad}_{g_k'}(V_k)$ are linearly independent for *some* $g_1', g_2', \ldots, g_k' \in G(\overline{K})$; this is the same as saying that the vectors $w_1(g'), w_2(g'), \ldots, w_m(g')$ are linearly independent for *some* $g' \in G^k(\overline{K})$.

Let $n = \dim(G)$. Let $v_{m+1}, v_{m+2}, \ldots, v_n$ be $n-m$ vectors in $\mathfrak{g}(\overline{K})$ such that

$$w_1(g'), w_2(g'), \ldots, w_m(g'), v_{m+1}, v_{m+2} \ldots, v_n$$

are linearly independent. Then the determinant $\delta(g)$ of the n-by-n matrix having

$$w_1(g), w_2(g), \ldots, w_m(g), v_{m+1}, v_{m+2} \ldots, v_n$$

as its rows is non-zero for $g = g'$. Thus, the subvariety X of G^k defined by $\delta(g) = 0$ is a proper subvariety of G^k. For all $g \in G^k(\overline{K})$ not on X, the determinant $\delta(g)$ is non-zero, and thus $w_1(g), w_2(g), \ldots, w_m(g), v_{m+1}, v_{m+2} \ldots, v_n$ are linearly independent; in particular, $w_1(g), w_2(g), \ldots, w_m(g)$ are linearly independent. This is the same as saying that the linear spaces (4.4) are linearly independent for all $g \in G^k(\overline{K})$ not on X. ☐

Proposition 4.9. *Let G, H and F be algebraic groups defined over a field K. Let $\phi : G \times H \to F$, $\psi : G \times H \to G$ be regular maps satisfying*

$$\phi(g, h_1h_2) = \phi(g, h_1) \cdot \phi(\psi(g, h_1), h_2) \tag{4.5}$$

for all $g \in G$, $h_1, h_2 \in H$, and

$$\psi(\psi(g, h), h^{-1}) = g \tag{4.6}$$

for all $g \in G$, $h \in H$.

Define $\phi_g : H \to F$ by $\phi_g(h) = \phi(g, h)$. For all $g_0 \in G(\overline{K})$, $h_0 \in H(\overline{K})$, write $(D\phi_{g_0})|_{h=h_0}$ for the linear map from $TH|_{h=h_0}$ to $TF|_{f=\phi(g_0,h_0)}$ given by

$$(D\phi_{g_0})|_{h=h_0} := \left(\frac{\partial}{\partial h}\phi_{g_0}(h)|_{h=h_0}\right)(v). \tag{4.7}$$

Assume that $(D\phi_{g_0})|_{h=e}$ *is non-singular for all* $g_0 \in G(\overline{K})$ *outside a proper subvariety* X_G *of* G. *Then*

(a) $(D\phi_{g_0})|_{h=h_0}$ *is non-singular exactly when* $(g_0, h_0) \in (G \times H)(\overline{K})$ *lies outside a proper subvariety* $Y_{G\times H}$ *of* $G \times H$,

(b) $\deg(Y_{G\times H}) \ll_{\deg_{\mathrm{pol}}(\phi),\, \deg_{\mathrm{pol}}(\psi),\, \dim(H)} 1$,

(c) *the fibre* $(Y_{G\times H})_{g=g_0}$ *is a proper subvariety of* H *for all* $g_0 \in G(\overline{K})$ *not on* X_G, *and*

(d) *the fibre* $(Y_{G\times H})_{h=h_0}$ *is a proper subvariety of* G *for all* $h_0 \in H(\overline{K})$.

We will need to use conditions (4.5) and (4.6) in order to prove conclusion (d), and only for that purpose. The said conditions tell us that every point on H is in some sense like every other point. If we did not have them, (b) and (c) would still hold.

Before we prove Prop. 4.9, let us see why we should care: for a map ϕ that we are rather interested in, there is a ψ such that (4.5) and (4.6) hold.

Lemma 4.10. *Let* G *be an algebraic group defined over a field* K*; let* H_0/K, $H_1/K, \ldots, H_\ell/K$ *be subgroups thereof. Let* $\phi : G' \times H' \to G$, $\psi : G' \times H' \to G'$ *(where* $G' = G^{\ell+1}$ *and* $H' = H_0 \times H_1 \times \cdots \times H_\ell$, ℓ *arbitrary) be given by*

$$\phi((g_0, g_1, \ldots, g_\ell), (h_0, h_1, \ldots, h_\ell)) = g_0 h_0 g_0^{-1} \cdot g_1 h_1 g_1^{-1} \cdots g_\ell h_\ell g_\ell^{-1}.$$

and

$$\psi((g_0, g_1, \ldots, g_\ell), (h_0, h_1, \ldots, h_\ell)) = (g_j')_{0\le j\le \ell},$$

where $g_\ell' = g_\ell$ *and* $g_j' = g_{j+1}' h_{j+1}^{-1} g_{j+1}^{-1} g_j$ *for* $0 \le j \le \ell - 1$.

Then ψ *and* ϕ *satisfy (4.5) and (4.6).*

Proof. Equation (4.6) follows easily from the definition of g_j'.

By the definition of g_j',

$$\begin{aligned} g_j' &= g_\ell h_\ell^{-1} g_\ell^{-1} \cdot g_{\ell-1} h_{\ell-1}^{-1} g_{\ell-1}^{-1} \cdots g_{j+1} h_{j+1}^{-1} g_{j+1}^{-1} \cdot g_j \\ &= \phi((g_{j+1}, \ldots, g_\ell), (h_j, h_{j+1}, \ldots, h_\ell))^{-1} \cdot g_j \end{aligned} \tag{4.8}$$

for all $0 \le j \le \ell$. Hence

$$\phi((g_j, g_{j+1}, \ldots, g_\ell), (h_j, h_{j+1}, \ldots, h_\ell)) g_j' h_j' = g_j h_j g_j^{-1} \cdot g_j \cdot h_j' = g_j h_j h_j' \tag{4.9}$$

for h_j' arbitrary. Applying (4.9) and then (4.8), we conclude that

$$\phi((g_j, g_{j+1}, \ldots, g_\ell), (h_j, h_{j+1}, \ldots, h_\ell)) g_j' h_j' g_j'^{-1}$$

equals

$$g_j h_j h_j' g_j^{-1} \cdot \phi((g_{j+1}, g_{j+2}, \ldots, g_\ell), (h_{j+1}, h_{j+2}, \ldots, h_\ell)).$$

Using this last equality in turn for $j = 0, 1, \ldots, \ell$, we obtain that

$$\phi((g_0, g_1, \ldots, g_\ell), (h_0, h_1, \ldots, h_\ell)) g_0' h_0' g_0'^{-1} g_1' h_1' g_1'^{-1} \cdots g_\ell' h_\ell' g_\ell'^{-1}$$

equals

$$g_0 h_0 h_0' g_0^{-1} \cdot g_1 h_1 h_1' g_1^{-1} \cdots g_\ell h_\ell h_\ell' g_\ell^{-1};$$

this is the same as saying that (4.5) holds. $\square$

Proof of Proposition 4.9. Define

$$\begin{aligned}\rho_{g_0,h_0}(v) &:= \left(\frac{\partial}{\partial h}((\phi(g_0,h_0))^{-1}\phi(g_0,h_0h))\right)|_{h=e}(v)\\ &= \left(\frac{\partial}{\partial h}\phi(\psi(g_0,h_0),h)\right)|_{h=e}(v).\end{aligned} \tag{4.10}$$

It is clear that $(D\phi_{g_0})|_{h=h_0}$ is non-singular if and only if ρ_{g_0,h_0} is non-singular. Now ρ_{g_0,h_0} is a linear map from the vector space $V = (TH)|_{h=e}$ to the vector space $W = (TF)|_{f=e}$; both V and W are independent of g_0 and h_0. Hence ρ_{g_0,h_0} is non-singular exactly when a $\dim(V)$-by-$\dim(V)$ determinant δ equals 0. The entries of δ are polynomials on the coordinates of g_0 and h_0. Let $Y_{G\times H}$ be the subvariety of $G\times H$ defined by $\delta = 0$. Then conclusion (a) holds by definition. It is clear that $\deg(Y_{G\times H}) \ll_{\deg(\delta)} 1$; since $\deg(\delta) \ll_{\deg_{\mathrm{pol}}(\phi),\deg_{\mathrm{pol}}(\psi),\dim(H)} 1$, it follows that $\deg(Y_{G\times H}) \ll_{\deg_{\mathrm{pol}}(\phi),\deg_{\mathrm{pol}}(\psi),\dim(H)} 1$. Thus conclusion (b) holds.

By the assumptions of the proposition, $(D\phi_{g_0})|_{h=e}$ is non-singular for all $g_0 \in G(\overline{K})$ outside X_G. This is the same as saying that (g_0, e) lies outside $Y_{G\times H}$, and so $(Y_{G\times H})_{g=g_0}$ is a subvariety of H not containing e; in particular, $(Y_{G\times H})_{g=g_0}$ is a proper subvariety of H, i.e., conclusion (c) holds.

Now, by (4.10), ρ_{g_0,h_0} is non-singular exactly when $(D\phi_{\psi(g_0,h_0)})|_{h=e}$ is non-singular, i.e., exactly when $\psi(g_0,h_0)$ lies outside X_G. By (4.6), $r_{h_0} : g \mapsto \psi(g,h_0)$ is a regular map with a regular map as its inverse; hence, $r_{h_0}^{-1}(X_G)$ is a proper subvariety of G. By what we just said, $(Y_{G\times H})_{h=h_0} = r_{h_0}^{-1}(X_G)$, and so we have obtained conclusion (d). □

We shall now see how to escape from a variety $Y_{G\times H}$ such as the one given by Prop. 4.9, even if we are not given a set of generators of $H(K)$.

Lemma 4.11. *Let $G \subset GL_n$ be an irreducible algebraic group defined over a finite field K. Assume either that K is perfect or that G is reductive. Let $H/\overline{K}$ be an algebraic subgroup of G. Let $Y_{G\times H}$ be a proper subvariety of $G\times H$ such that the fibre $(Y_{G\times H})_{h=h_0}$ is a proper subvariety of G for all $h_0 \in H(\overline{K})$.*

Let $A \subset G(K)$ be a set of generators of $G(K)$, and let E be a subset of $H(K)$.

Then there is a $g_E \in A_k$, $k \ll_{\overrightarrow{deg}(Y_{G\times H})} 1$, such that at least

$$\gg_{\overrightarrow{deg}(Y_{G\times H})} |E|$$

elements of E lie outside $(Y_{G\times H})|_{g=g_E}$, provided that $|K|$ is larger than a constant depending only on n and $\overrightarrow{deg}(Y_{G\times H})$.

Proof. Let $h_0 \in H(\overline{K})$ be arbitrary. By one of the assumptions, the fibre $(Y_{G\times H})_{h=h_0}$ is a proper subvariety of G. Hence, by escape in G (Lemma 4.4, together with Lemma 4.6), there is a $g_1 \in A_{k_1}$, $k_1 \ll_{\overrightarrow{\deg}(Y_{G\times H})} 1$, such that g_1 does not lie on $(Y_{G\times H})_{h=h_0}$. Let $Y_1 \subset H$ be the union of all connected components of $(Y_{G\times H})_{g=g_1}$ that contain elements of E; since (by the definition of g_1 and X_G) the fibre $(Y_{G\times H})_{g=g_1}$ is a proper subvariety of H, clearly Y_1 is a proper subvariety of H as well. Moreover, $\overrightarrow{\deg}(Y_1) \ll_{\overrightarrow{\deg}(Y_{G\times H})} 1$.

If $|E\cap Y_1(K)| < \frac{1}{2}|E|$, we have $|E\setminus(E\cap Y_1(K))| \geq \frac{1}{2}|E|$ and we are done. Assume otherwise, and let $E_1 = E\cap Y_1(K)$. Choose a point $h_1 \in E_1$ lying on a component of

Y_1 of maximal dimension. By escape in G (Lemmas 4.4 and 4.6), there is a $g_2 \in A_{k_2}$, $k_2 \ll_{\overrightarrow{\deg}(Y_{G\times H})} 1$, such that g_2 does not lie in $(Y_{G\times H})_{h=h_1}$. Let Y_2 be the union of all connected components of $Y_1 \cap (Y_{G\times H})_{g=g_2}$ containing elements of E. Since Y_2 does not contain h_1, it does not contain all components of Y_1 of maximal dimension. Hence either (a) $\dim(Y_2) < \dim(Y_1)$ or (b) $s_{Y_2} < s_{Y_1}$, where, for a variety V, we write s_V for the number of components of maximal dimension. Moreover, by Bezout's theorem (Lem. 2.4), $\overrightarrow{\deg}(Y_2) \ll_{\overrightarrow{\deg}(Y_1),\overrightarrow{\deg}(Y_{G\times H})} 1$.

Starting with $j = 2$, we recur, doing what we just did: if $|E_{j-1} \cap Y_j(K)| < \frac{1}{2}|E_{j-1}|$, we have $|E_{j-1} \setminus (E_{j-1} \cap Y_j(K))| \geq \frac{1}{2}|E_{j-1}| \geq \frac{1}{2^j}|E|$, and we stop; otherwise, we let $E_j = E_{j-1} \cap Y_j(K)$, we choose a point $h_j \in E_j$ lying on a component of Y_j of maximal dimension, we find a $g_{j+1} \in A_{k_{j+1}}$, $k_{j+1} \ll_{\overrightarrow{\deg}(Y_{G\times H})} 1$, such that g_{j+1} does not lie on $(Y_{G\times H})_{h=h_j}$, we let Y_{j+1} be the union of connected components of $Y_j \cap (Y_{G\times H})_{g=g_{j+1}}$ containing elements of E, etc. Thanks to Bezout's theorem, we reach $Y_j = \emptyset$ (and thus we stop) after a number of steps $\ll_{\overrightarrow{\deg}(Y_{G\times H})} 1$. Hence $|E_{j-1} \setminus (E_{j-1} \cap Y_j(K))| \geq \frac{1}{2^j}|E|$ (where j is the index j we are at when we stop) implies $|E_{j-1} \setminus (E_{j-1} \cap Y_j(K))| \gg_{\overrightarrow{\deg}(Y_{G\times H})} 1$. □

It is time to put together what we have proven in this subsection.

Proposition 4.12. *Let $G \subset \mathrm{GL}_n$ be an irreducible algebraic group defined over a field K; let $H_0/\overline{K}$, $H_1/\overline{K}, \ldots, H_\ell/\overline{K}$ be algebraic subgroups thereof. Assume either that K is perfect or that G is reductive. Write $\mathfrak{g}$ for the Lie algebra of G, and $\mathfrak{h}_j$ for the Lie algebra of H_j, $0 \leq j \leq \ell$. Assume there are $g_0, g_1, \ldots, g_\ell \in G(\overline{K})$ such that*

$$\mathrm{Ad}_{g_0}(\mathfrak{h}_0), \mathrm{Ad}_{g_1}(\mathfrak{h}_1), \ldots, \mathrm{Ad}_{g_\ell}(\mathfrak{h}_\ell) \tag{4.11}$$

are linearly independent.

Let $A \subset G(K)$ be a set of generators of $G(K)$. Then there are $g_0, g_1, \ldots, g_\ell \in A_k$, $k \ll_n 1$, such that

$$|g_0 E_0 g_0^{-1} \cdot g_1 E_1 g_1^{-1} \cdot \cdots \cdot g_\ell E_\ell g_\ell^{-1}| \gg_{n,\deg(H_0),\deg(H_1),\ldots,\deg(H_\ell)} |E_0||E_1|\cdots|E_\ell|$$

for any non-empty subsets $E_j \subset H_j(K)$, $0 \leq j \leq \ell$.

In the present paper, we will always use Prop. 4.12 with $G = \mathrm{SL}_n$, which is semisimple and hence reductive.

Proof. We will apply Prop. 4.9 with $\overline{K}$ instead of K, $G' = G^{\ell+1}$ instead of G, $H = H_0 \times H_1 \times \cdots \times H_\ell$, $F = G$ (our G, that is, not G') and ϕ and ψ as in Lem. 4.10. The derivative (4.7) is non-singular for $h_0 = e$ whenever the linear spaces (4.11) are linearly independent; by Lemma 4.8 and the assumption on (4.11) (namely, that the spaces are independent for some $(g_0, g_1, \ldots, g_\ell) \in G'(\overline{K})$), the spaces are independent for all $(g_0, g_1, \ldots, g_\ell) \in G'(\overline{K})$ outside a proper subvariety $X_{G'}$ of G' (with $\overrightarrow{\deg}(X_{G'}) \ll_n 1$). The conditions of Prop. 4.9 are thus fulfilled, and we obtain a variety $Y_{G'\times H}$ (with $\overrightarrow{\deg}(Y_{G'\times H}) \ll_n 1$) as in its statement.

We now apply Lemma 4.11 to G', H and $Y_{G'\times H}$. (We may use Lemma 4.11 because we may assume that $|K|$ is larger than a constant depending only on n, as otherwise the statement we seek to prove is trivially true. Notice also that, if G is reductive, then $G' = G^{\ell+1}$ is reductive.) We then apply Lemma 4.3 with $X = H$ and are done. □

4.4. **Examining subspaces at the origin.** Recall the definition of *linear independence* of subspaces given in §2.5.5.

Proposition 4.13. *Let $G \subset \mathrm{GL}_n$ be an algebraic group defined over a field K; write $\mathfrak{g}$ for its Lie algebra. Let $\mathfrak{h}$ be a subspace of $\mathfrak{g}$. Suppose that there are elements $\vec{g}_1, \vec{g}_2, \dots, \vec{g}_\ell$ of $\mathfrak{g}(\overline{K})$ such that the spaces*

$$\mathfrak{h}, [\vec{g}_1, \mathfrak{h}], [\vec{g}_2, \mathfrak{h}], \dots, [\vec{g}_\ell, \mathfrak{h}] \tag{4.12}$$

are linearly independent and of dimension $\dim(\mathfrak{h})$*. Suppose that the characteristic* $\mathrm{char}(K)$ *of K is either 0 or greater than k, where $k = \dim(\mathfrak{h})$.*

Then there is a proper subvariety X of G^ℓ such that, for all $(g_1, g_2, \dots, g_\ell) \in G^\ell(\overline{K})$ not on X, the spaces

$$\mathfrak{h}, \mathrm{Ad}_{g_1}(\mathfrak{h}), \mathrm{Ad}_{g_2}(\mathfrak{h}), \dots, \mathrm{Ad}_{g_\ell}(\mathfrak{h})$$

are linearly independent and of dimension $\dim(\mathfrak{h})$*. Moreover, $\overrightarrow{deg}(X) \ll_n 1$.*

Proof. Let $e_1, e_2, \dots, e_k$ be a basis of $\mathfrak{h}$. Write

$$\theta = e_1 \wedge e_2 \wedge \cdots \wedge e_k \quad \text{and} \quad \theta_j = [\vec{g}_j, [\vec{g}_j, [\cdots [\vec{g}_j, \theta] \cdots]]] \ \ (k \text{ times}) \quad \text{for } 1 \leq j \leq \ell.$$

Let us examine the wedge product

$$\Theta = \theta \wedge \theta_1 \wedge \theta_2 \wedge \cdots \wedge \theta_\ell.$$

Any term of θ_j containing a term of the form $\cdots \wedge e_r \wedge \cdots$ will be lost, as its wedge product with θ will be 0. The only terms of θ_j remaining are $k!$ identical terms of the form

$$\omega_j = [\vec{g}_j, e_1] \wedge [\vec{g}_j, e_2] \wedge \cdots \wedge [\vec{g}_j, e_k].$$

We thus have

$$\Theta = (k!)^\ell \cdot (\theta \wedge \omega_1 \wedge \omega_2 \wedge \cdots \wedge \omega_\ell).$$

By the condition stating that the spaces (4.12) are linearly independent, we have $\Theta \neq 0$ (provided that, as we are assuming, $\mathrm{char}(K)$ does not divide $k!$).

Now, θ_j is a derivative, viz., the kth order derivative at the origin of

$$\mathrm{Ad}_{g_{j,1} g_{j,2} \cdots g_{j,k}}(\theta) = \mathrm{Ad}_{g_{j,1}}(\mathrm{Ad}_{g_{j,2}}(\cdots (\mathrm{Ad}_{g_{j,k}}(\theta)) \cdots))$$

taken with respect to the variables $g_{j,1}$, $g_{j,2}, \dots$, $g_{j,k}$ one time each, always in the same direction $\vec{g}_j$. Hence Θ is itself a $(k \cdot \ell)$th order derivative (at the origin) of

$$\theta \wedge \bigwedge_{1 \leq j \leq \ell} \mathrm{Ad}_{g_{j,1} g_{j,2} \cdots_{j,k}}(\theta). \tag{4.13}$$

Since Θ is non-zero, it follows that (4.13) is not identically zero as the $g_{j,i}$ vary within $G(\overline{K})$. Setting $g_j = g_{j,1} g_{j,2} \cdots g_{j,k}$, we see that there are $g_1, g_2, \dots, g_\ell \in G(\overline{K})$ such that

$$\theta \wedge \mathrm{Ad}_{g_1}(\theta) \wedge \mathrm{Ad}_{g_2}(\theta) \wedge \ldots \wedge \mathrm{Ad}_{g_l}(\theta) = 0 \tag{4.14}$$

does not hold. Define the variety X by the equation (4.14). □

4.5. **Subgroups of unipotent subgroups and tori.** We will later want to know what kinds of subgroups a torus can have. The following lemma will actually be enough.

We recall that a *torus* is an algebraic group isomorphic to $(\mathrm{GL}_1)^m$ for some $m \geq 1$. (This should be clear by now, though we are used to speaking of maximal or non-maximal tori *of* a group G, i.e., subgroups of G that happen to be tori.) Just as we often see algebraic groups G as (algebraic) subgroups of $\mathrm{GL}_n \subset \mathbb{A}^{n^2}$, it makes sense to consider tori T (isomorphic to $(\mathrm{GL}_1)^m$) given as algebraic subgroups of $(\mathrm{GL}_1)^n \subset \mathbb{A}^n$.

A *character* $\alpha : T \to \mathrm{GL}_1$ of a torus $T \subset (\mathrm{GL}_1)^n$ is a map of the form $(x_1, x_2, \dots, x_n) \to x_1^{a_1} x_2^{a_2} \dots x_n^{a_n}$ for some $a_1, a_2, \dots, a_n \in \mathbb{Z}$ (called the *exponents* a_j of α).

Lemma 4.14. *Let K be a field. Let $T/\overline{K} \subset (\mathrm{GL}_1)^n$ be a torus. Let $V/\overline{K}$ be a proper algebraic subgroup of T.*

Then V is contained in the kernel of a non-trivial character $\alpha : T \to \mathbb{A}^1$ whose exponents are bounded in terms of n and $\overrightarrow{\deg}(V)$ alone.

Proof. Let H be the identity component of V; by the definition of the degree of a variety (§2.5.2), the degree $\deg(H)$ of the irreducible variety H is bounded in terms of $\overrightarrow{\deg}(V)$ alone. Now [BoG, Prop. 3.3.9(c)] (applied with $X = H$) states that H must be of the form $\phi_A(\widetilde{X(H)} \times (\mathrm{GL}_1)^r)$, where $r = \dim(H)$, $\widetilde{X(H)}$ is a closed subvariety of $(\mathrm{GL}_1)^{n-r}$ and $\phi_A : (\mathrm{GL}_1)^n \to (\mathrm{GL}_1)^n$ is an (invertible) *monoidal transformation* ([BoG, Def. 3.2.4]) given by a matrix $A \in \mathrm{SL}_n(\mathbb{Z})$. Since $\dim(H) = \dim(\mathrm{GL}_1^r)$ and ϕ_A is invertible, $\widetilde{X(H)}$ must be 0-dimensional; since H is connected and ϕ_A is invertible, $\widetilde{X(H)}$ must be consist of a single point; since H is a group and ϕ_A is an isomorphism of algebraic groups, that single point must be the identity. In other words, $H = \phi_A(\{e\} \times (\mathrm{GL}_1)^r)$, where e is the identity in $(\mathrm{GL}_1)^{n-r}$.

Thus H is in the kernel of the character $g \to ((\phi_A)^{-1}(g))_j = (\phi_{A^{-1}}(g))_j$ for every $1 \leq j \leq n - r$. If T were in the kernel of every such character, its dimension would be r, i.e., the same as the dimension of H; since T is irreducible and H is a proper subgroup of T, this cannot be the case. Let, then, $\alpha_0 : T \to \mathbb{A}^1$ be the restriction to T of the character $g \to (\phi_{A^{-1}}(g))_j$ for some j for which such a restriction is not trivial. The exponents of g are entries of A^{-1}; by [BoG, Remark 3.3.10], the entries of A^{-1} are $\ll_{n,\delta(H)}$, where $\delta(H)$ is the "essential degree" of H (as defined in [BoG, §3.3.1]). Now, by [BoG, Prop. 3.3.2], $\delta(H) \leq \deg(H)$. Hence the exponents of g are $\ll_{n,\deg(H)} 1$.

The number m of connected components of V is bounded by $\overrightarrow{\deg}(V)$. Now V must consist of m cosets of the form xH, where $x \in G(\overline{K})$ is such that x^m lies on H. We define $\alpha : T \to \mathbb{A}^1$ to be the character such that $\alpha(g) = \alpha_0(g^m)$, and are done. $\square$

Let U be a unipotent subgroup of SL_3. We need to classify the subgroups of $U(\mathbb{Z}/p\mathbb{Z})$. This turns out to be an easy task.

Lemma 4.15. *Let $K = \mathbb{Z}/p\mathbb{Z}$. Let $G = \mathrm{SL}_3$, and let B/K be a Borel subgroup thereof; let U/K be the subgroup of unipotent matrices of B.*

Then every subgroup H of $U(K)$ is conjugate in $B(\overline{K})$ to one of the following subgroups:

$$H = \{I\}, \text{or} \tag{4.15}$$

$$\textit{or} \quad H = U(K), \tag{4.16}$$

$$H = \left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : x, y \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.17}$$

$$H = \left\{ \begin{pmatrix} 1 & 0 & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : y, z \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.18}$$

$$H = \left\{ \begin{pmatrix} 1 & 0 & y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : y \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.19}$$

$$H = \left\{ \begin{pmatrix} 1 & x & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : x \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.20}$$

$$H = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : z \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.21}$$

$$H = \left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : x, y \in \mathbb{Z}/p\mathbb{Z} \right\}, \tag{4.22}$$

$$H = \left\{ \begin{pmatrix} 1 & x & \frac{x^2}{2} \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : x \in \mathbb{Z}/p\mathbb{Z} \right\} \quad \textit{(if } p > 2\textit{)}. \tag{4.23}$$

Proof. Let N be the normal subgroup of U consisting of the matrices of the form

$$\begin{pmatrix} 1 & 0 & y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We may identify $U(K)/N(K)$ with $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ by the bijection

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} N(K) \mapsto (x, z).$$

Consider $H' = H/(H \cap N(K))$, which can be seen as a subgroup of $U(K)/N(K) \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ by the inclusion $H \subset U(K)$. If $H' = \{(0,0)\}$, then either $H = \{I\}$ or H is as in (4.19).

Suppose $H' = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. We may then choose two matrices

$$g = \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \in H, \quad g' = \begin{pmatrix} 1 & x' & y' \\ 0 & 1 & z' \\ 0 & 0 & 1 \end{pmatrix} \in H$$

with $xz' \neq x'z$ (as we can specify x, x', z, z' arbitrarily). The two matrices g, g' do not commute. Hence $gg'g^{-1}g'^{-1} \neq I$. Because $U(K)/N(K)$ is abelian, $gg'g^{-1}g'^{-1}$ must lie in $N(K)$; since $N(K) \simeq \mathbb{Z}/p\mathbb{Z}$ and $gg'g^{-1}g'^{-1} \neq I$, we see that $gg'g^{-1}g'^{-1}$ must generate $N(K)$. Hence $N(K) \subset H$, and so, since $H' = H/(H \cap N(K))$ is all of $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we conclude that H is all of $U(K)$.

Suppose $H' = \mathbb{Z}/p\mathbb{Z} \times \{0\}$ or $H' = \{0\} \times \mathbb{Z}/p\mathbb{Z}$. Then it is easy to show that we are either in cases (4.17) or (4.20) (if $H' = \mathbb{Z}/p\mathbb{Z} \times \{0\}$) or cases (4.18) or (4.21) (if $H' = \{0\} \times \mathbb{Z}/p\mathbb{Z}$). (We initially obtain

$$\begin{pmatrix} 1 & x & rx \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

instead of (4.20), but this is conjugate to (4.20) in $B(K)$ by an element of $U(K)$. The same happens for (4.21).)

Suppose, finally, that H' is of the form $\{(x, rx) : x \in \mathbb{Z}/p\mathbb{Z}\}$ for some $r \in \mathbb{Z}/p\mathbb{Z}$, $r \neq 0$. If H contains a non-trivial element of N, we obtain (4.22) after conjugation by an element of $B(\overline{K})$. Suppose H contains no non-trivial element of $N(K)$. Then, for every $x \in \mathbb{Z}/p\mathbb{Z}$, there is exactly one element $y = y(x)$ of $\mathbb{Z}/p\mathbb{Z}$ such that

$$\begin{pmatrix} 1 & x & y(x) \\ 0 & 1 & rx \\ 0 & 0 & 1 \end{pmatrix}$$

is in H. Thus, for every $m \in \mathbb{Z}$,

$$\begin{pmatrix} 1 & 1 & y(1) \\ 0 & 1 & r \\ 0 & 0 & 1 \end{pmatrix}^m = \begin{pmatrix} 1 & m & m \cdot y(1) + (1 + 2 + \cdots + (m-1)) \cdot r \\ 0 & 1 & rm \\ 0 & 0 & 1 \end{pmatrix}$$

must be equal (mod p) to

$$\begin{pmatrix} 1 & m & y(m) \\ 0 & 1 & rm \\ 0 & 0 & 1 \end{pmatrix} \in H.$$

If $p = 2$, we set $m = 2$ and obtain a contradiction to our assumption that H contains no non-trivial element of $N(K)$. Assume, then, that $p > 2$. Then we obtain $y(x) = xy(1) + \frac{x(x-1)}{2} r = x \cdot (y(1) - r/2) + \frac{x^2}{2} r$. Then

$$\begin{pmatrix} 1 & m & y(m) \\ 0 & 1 & rm \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \rho & \rho^{-2} c & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho^{-2} \end{pmatrix} \cdot \begin{pmatrix} 1 & x & \frac{x^2}{2} \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \rho & \rho^{-2} c & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho^{-2} \end{pmatrix}^{-1}$$

where $c = y(1) - r/2$ and $\rho \in \overline{K}$ is any cube root of r. This means that H is a conjugate of (4.23) by an element of $B(\overline{K})$, and so we are done. $\square$

5. Tori and conjugacy classes

Let $A \subset \mathrm{SL}_n(K)$, K any field. We mean to show that, if A grows slowly under multiplication, then (a) many elements of A lie on a torus, and (b) there are not many more conjugacy classes intersecting A than there are elements on the torus. Somewhat counter-intuitively, we shall begin by giving an *upper* bound on the number of elements of A that can lie on a torus.

The methods in this section seem to be robust as far as the group type and the ground field are concerned. We shall work – by and large – on $\mathrm{SL}_n(K)$, n arbitrary, rather than only on $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$. A few lemmas will be proven for all classical Chevalley groups.

5.1. **The intersection with a maximal torus: an upper bound.** Let A be a set of generators of $G = \mathrm{SL}_n(K)$. We shall show that, given any torus T, the intersection of A with T is not too large.

By a *classical Lie algebra* over a field K we mean $\mathfrak{sl}_n$, $\mathfrak{so}_n$ or $\mathfrak{sp}_{2n}$ ($n \geq 1$). By a *classical Chevalley group* over K we mean SL_n, SO_n or Sp_n. We shall see SL_n, SO_n and Sp_n as subvarieties of the affine space of matrices M_n. If $\mathfrak{g}$ is a Lie algebra defined over a field K, we denote by $\mathfrak{g}^*$ the K-linear space of K-linear functions on $\mathfrak{g}$.

Lemma 5.1. *Let $\mathfrak{g}$ be a classical Lie algebra over a field K with* $\mathrm{char}(K) > 2$*. Let $\mathfrak{t}$ be a Cartan subalgebra of $\mathfrak{g}$. Let Φ be its set of roots and let $V = \mathfrak{t}^*$. Then there is a partition $\Phi = \Phi_1 \cup \Phi_2 \cdots \cup \Phi_\ell$ such that each Φ_j, $1 \leq j \leq \ell$, is a basis of V.*

Proof. Let us consider each of the classical root systems individually. We shall see them as abstract root systems, i.e., as subsets of V, which can be seen simply as a linear space over K with no further structure.

We look first at A_n. Then V can be identified with the subspace of K^{n+1} for which the coordinates sum to 0, and the set of roots Φ with the set of vectors in V having one coordinate equal to 1, one coordinate equal to -1, and all other coordinates equal to 0. Define Φ_j ($1 \leq j \leq n+1$) to be the set of roots $v_j - v_i$, $i \neq j$. Then every Φ_j is a basis of V.

Now look at B_n. Then V can be identified with K^n, and Φ with the set of vectors having at most two coordinates in $\{-1, 1\}$, and all other coordinates equal to 0. We let Φ_j ($1 \leq j \leq n$) be the set of roots $v_j - v_i$, $i \neq j$, together with the root v_j; let Φ_{n+j} ($1 \leq j \leq n$) be the set of roots $v_j + v_i$, $i > j$, together with $-v_j - v_i$, $i < j$, and $-v_j$.

Let us now consider C_n. Then V can be identified with K^n, and Φ with the set of vectors having two coordinates in $\{-1, 1\}$ and all other coordinates equal to 0, together with the vectors having one coordinate in $\{-2, 2\}$ and all other coordinates equal to 0. The choice of Φ_j is almost as for B_n: we let Φ_j ($1 \leq j \leq n$) be the set of roots $v_j - v_i$, $i \neq j$, together with the root $2v_j$; let Φ_{n+j} ($1 \leq j \leq n$) be the set of roots $v_j + v_i$, $i > j$, together with $-v_j - v_i$, $i < j$, and $-2v_j$.

Finally, we consider D_n. Then $V = K^n$, and Φ can be identified with the set of vectors having two coordinates in $\{-1, 1\}$ and all other coordinates equal to 0. Then let Φ_j ($1 \leq j \leq n-1$) be the set of roots $v_j - v_i$, $i \neq j$, together with the root $v_j + v_n$; let Φ_{j+n-1} ($1 \leq j \leq n-1$) be the set of roots $v_j + v_i$, $i > j$, together with the roots $-(v_j + v_i)$, $i < j$, and the root $-(v_j + v_n)$. $\square$

Lemma 5.2. *Let $\mathfrak{g}$ be a classical Lie algebra over a field K with* $\mathrm{char}(K) \neq 2$*. Let $\mathfrak{t}$ be a Cartan subalgebra of $\mathfrak{g}$. Let $\ell = \frac{\dim(G)}{\dim(T)} - 1$. Then there are elements $\vec{g}_1, \vec{g}_2, \ldots, \vec{g}_\ell \in \mathfrak{g}$ such that the spaces*

$$\mathfrak{t}, [\vec{g}_1, \mathfrak{t}], \cdots, [\vec{g}_\ell, \mathfrak{t}] \tag{5.1}$$

are linearly independent and of dimension $\dim(\mathfrak{t})$.

Proof. Let $\Phi = \Phi_1 \cup \Phi_2 \cup \dots \cup \Phi_\ell$ be a partition as in Lemma 5.1. For $1 \leq k \leq \ell$, choose one non-zero element $v_{k,j}$ in the root space corresponding to each element $\alpha_{k,j}$ of Φ_k; denote the set of such elements for given k by $\{v_{k,j}\}_{1\leq j\leq m}$, where $m = |\Phi_k| = \dim(\mathfrak{t}^*)$. Let $\vec{g}_k = \sum_{1\leq j\leq m} v_{k,j}$. Then, for every $t \in \mathfrak{t}(K)$, we have $[t, \vec{g}_k] = \sum_j \alpha_{k,j}(t) \cdot v_{k,j}$.

Now let $e_1, e_2, \dots, e_m$ be a basis for $\mathfrak{t}(K)$. (Since $\dim(\mathfrak{t}(K)) = \dim(\mathfrak{t}^*)$, a basis of $\mathfrak{t}(K)$ has m elements.) Then the linear map $f_k : v \to [\vec{g}_k, v]$ from $\mathfrak{t}(K)$ to the span V_k of the root spaces $\{\alpha_{k,j}\}_{1\leq j\leq m}$ is given by a square m-by-m matrix with entries $\{\alpha_{k,j}(e_i)\}_{1\leq i,j\leq m}$. Since (by Lem. 5.1) the roots in Φ_k form a basis of $\mathfrak{g}^*$, they are linearly independent, and so the matrix is non-singular. Thus, the image of f_k is all of V_k. In other words, $[\mathfrak{t}, \vec{g}_j]$ equals the span of the root spaces $\{\alpha_{k,j}\}_{1\leq j\leq m}$.

By [Hum, §26.2, Cor. B], the Lie algebra $\mathfrak{g}$ is the direct sum of $\mathfrak{t}$ and the root spaces. Since $[\mathfrak{t}, \vec{g}_j] = -[\vec{g}_j, \mathfrak{t}] = [\vec{g}_j, \mathfrak{t}]$, we are done. □

If G is a classical Chevalley group, then both G and its maximal torus are irreducible varieties over any field K (see, e.g., [Bor], §1.2, §8.5(2) and §8.7).

Proposition 5.3. *Let* $G \subset \mathrm{GL}_n$ *be a classical Chevalley group defined over a field* K *with* $\mathrm{char}(K) \neq 2$. *Let* T *be a maximal torus of* G *defined over* $\overline{K}$.

Let $A \subset G(K)$ *be a set of generators of* $G(K)$, *and let* E *be a subset of* $T(K)$. *Let* $\ell = \frac{\dim(G)}{\dim(T)} - 1$. *Then there are* $g_0, g_1, \dots, g_\ell \in A_k$, $k \ll_n 1$, *such that*

$$|g_0 E g_0^{-1} \cdot g_1 E g_1^{-1} \cdot \dots \cdot g_\ell E g_\ell^{-1}| \gg_n |E|^{\ell+1}.$$

Proof. By Proposition 4.12 with $H_j = T$ and $E_j = E$ for $0 \leq j \leq \ell$. (The condition on (4.11) is fulfilled by Lemma 5.2 and Proposition 4.13.) □

Corollary 5.4. *Let* $G \subset \mathrm{GL}_n$ *be a classical Chevalley group. Let* K *be a field with* $\mathrm{char}(K) \neq 2$. *Let* T *be a maximal torus of* G *defined over* $\overline{K}$. *Let* $A \subset G(K)$ *be a set of generators of* $G(K)$.

Then

$$|A \cap T(K)| \ll_n |A_k|^{\frac{\dim(T)}{\dim(G)}},$$

where $k \ll_n 1$.

Proof. Immediate by Prop. 5.3 (with $E = A \cap T(K)$). □

5.2. **A lower bound on the number of conjugacy classes.** Let A be a set of generators of $\mathrm{SL}_n(K)$. We shall show that there are many conjugacy classes represented by elements of A – or, at any rate, by elements of A_k.

Given a matrix g in $\mathrm{SL}_n(K)$, we define $\kappa(g) \in \mathbb{A}^{n-1}(K)$ to be the tuple

$$(a_{n-1}, a_{n-2}, \dots, a_1)$$

of coefficients of

$$\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \dots + a_1\lambda + (-1)^n = \det(\lambda I - g) \in K[\lambda]$$

(the characteristic polynomial of g).

As is well-known, $\kappa(g) = \kappa(hgh^{-1})$ for any h, i.e., $\kappa(g)$ is invariant under conjugation. If g is a regular semisimple element of SL_n – that is, if its eigenvalues are all distinct – then $\kappa(g)$ actually determines the conjugacy class $\mathrm{Cl}_G(g)$ of g.

Lemma 5.5. *Let $G = \mathrm{SL}_n$. Let K be a field. For $h_0, h_1, \dots, h_n$, define $f_{h_0,h_1,\dots,h_n}$ to be the map*

$$f_{h_0,h_1,\dots,h_n} : g \mapsto (\kappa(h_0 g), \kappa(h_1 g), \dots, \kappa(h_n g)) \tag{5.2}$$

from G to $\mathbb{A}^{(n-1)\cdot(n+1)} = \mathbb{A}^{n^2-1}$.

Let $T/\overline{K}$ be a maximal torus of G. Then there are $h_0 \in G(\overline{K})$, $h_1 \in T(\overline{K})$ and $g_0 \in G(\overline{K})$ such that the derivative of $f_{h_0,h_1,h_1^2,\dots,h_1^n}$ at $g = g_0$ is a nonsingular linear map.

Proof. We may write the elements of $G(\overline{K})$ so that the elements of $T(\overline{K}) \subset G(\overline{K})$ become diagonal matrices. Let

$$g_0 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ (-1)^{n-1} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Let $\vec{r} = (r_1, r_2, \dots, r_n)$ be a vector in $\overline{K}^n$ with $r_1 \cdot r_2 \cdots r_n = 1$. Define

$$h_1 = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & r_n \end{pmatrix} \tag{5.3}$$

for $1 \leq i \leq n$.

Let us look, then, at the derivative at $g = I$ of $g \mapsto \kappa(h_1^i g_0 g)$ for $1 \leq i \leq n$. The derivative at $g = I$ of the map taking g to the coefficient of λ^{n-1} in $\det(\lambda I - h_1^i g_0 g)$ (i.e., to (-1) times the trace of $h_1^i g_0 g$) is equal to the map taking each matrix γ in the tangent space $\mathfrak{g}$ to G at the origin to

$$(-1) \cdot (r_1^i \gamma_{2,1} + r_2^i \gamma_{3,2} + \dots + r_j^i \gamma_{j+1,j} + \dots + (-1)^{n-1} r_n^i \gamma_{1,n}),$$

where we write $\gamma_{i,j}$ for the entries of the matrix γ.

The derivative at $g = I$ of the map taking g to the coefficient of λ^{n-2} in $\det(\lambda I - h_1^i g_0 g)$ is the map taking each γ in $\mathfrak{g}$ to

$$\begin{aligned} r_1^i r_2^i \gamma_{3,1} + r_2^i r_3^i \gamma_{4,2} + \dots + r_j^i r_{j+1}^i \gamma_{j+2,j} + \dots + r_{n-3}^i r_{n-2}^i \gamma_{n-1,n-3} + r_{n-2}^i r_{n-1}^i \gamma_{n,n-2} \\ + r_{n-1}^i \cdot (-1)^{n-1} r_n^i \gamma_{1,n-1} + (-1)^{n-1} r_n^i r_1^i \gamma_{2,n}. \end{aligned}$$

In general, for $1 \leq k \leq n-1$, the derivative at $g = I$ of the map taking g to the coefficient of λ^{n-k} in $\det(\lambda I - h_1^i g_0 g)$ is the map taking γ to

$$(-1)^k \sum_{j=1}^{n-k} (r_j^i \cdot r_{\underline{j+1}}^i \cdots r_{\underline{j+k-1}}^i) \cdot \gamma_{\underline{j+k},j} + (-1)^{k+n-1} \sum_{j=n-k+1}^{n} (r_j^i \cdot r_{\underline{j+1}}^i \cdots r_{\underline{j+k-1}}^i) \cdot \gamma_{\underline{j+k},j}, \tag{5.4}$$

where by $\underline{a}$ we mean the only element of $\{1, 2, \dots, n\}$ congruent to a modulo n.

We see that the entries of γ present in (5.4) are disjoint for distinct $1 \leq k \leq n-1$ (and disjoint from $\{\gamma_{1,1}, \gamma_{2,2}, \dots, \gamma_{n,n}\}$, which would appear for $k = 0$). Now, for k fixed, (5.4) gives us a linear form on n variables $\gamma_{\underline{j+k},j}$ for each $1 \leq i \leq n$. Let us check that, for every $1 \leq k \leq n-1$, these linear forms are linearly independent, provided that $\vec{r}$ was chosen correctly.

This is the same as checking that the $n-1$ determinants

$$\left|(r_j^i \cdot r_{\underline{j+1}}^i \cdots r_{\underline{j+k-1}}^i)\right|_{1 \leq i,j \leq n} \tag{5.5}$$

for $1 \leq k \leq n-1$ are non-zero for some choice of $r_1, r_2, \dots, r_n$ with $r_1 \cdot r_2 \cdots r_n = 1$. (What we really want to check is that the determinant (5.5) is non-zero after all signs in some columns are flipped; since those flips do not affect the absolute value of the determinant, it is just as good to check that the determinant (5.5) itself is non-zero.) These are Vandermonde determinants, and thus are equal to

$$(-1)^{\lfloor n/2 \rfloor} \cdot \prod_{j_1 < j_2} (r_{j_2} \cdot r_{\underline{j_2+1}} \cdots r_{\underline{j_2+k-1}} - r_{j_1} \cdot r_{\underline{j_1+1}} \cdots r_{\underline{j_1+k-1}}).$$

For any given k, j_1, j_2 with $j_1 \neq j_2$, there are certainly $r_1, r_2, \dots, r_n \in \overline{K}$ with $r_1 r_2 \cdots r_n = 1$ such that $r_{j_1} \cdot r_{\underline{j_1+1}} \cdots r_{\underline{j_1+k-1}} \neq r_{j_2} \cdot r_{\underline{j_2+1}} \cdots r_{\underline{j_2+k-1}}$. Thus, $r_{j_1} \cdot r_{\underline{j_1+1}} \cdots r_{\underline{j_1+k-1}} = r_{j_2} \cdot r_{\underline{j_2+1}} \cdots r_{\underline{j_2+k-1}}$ defines a subvariety W_{k,j_1,j_2} of positive codimension in the (irreducible) variety $V \subset \mathbb{A}^n$ of all tuples $(r_1, r_2, \cdots, r_n) \neq 1$ with $r_1 r_2 \cdots r_n = 1$. Therefore, $W = \cup_{1 \leq k, j_1, j_2 \leq n,\, j_1 \neq j_2} W_{k,j_1,j_2}$ is a finite union of subvarieties of V of positive codimension. Take $\vec{r}$ to be any point of $V(\overline{K})$ outside $W(\overline{K})$.

It remains to choose h_0 so that the derivative of

$$g \mapsto \kappa(h_0 g)$$

at $g = I$ is a linear map of full rank on the diagonal entries $\gamma_{1,1}, \gamma_{2,2} \dots, \gamma_{n-1,n-1}$ of $\mathfrak{g}$. Let

$$h_0 = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & s_n \end{pmatrix}, \tag{5.6}$$

where $s_1, s_2, \dots, s_n \in \overline{K}$ fulfil $s_1 s_2 \cdots s_n = 1$. Then the derivative at $g = I$ of the map taking g to the coefficient of λ^{n-1} in $\det(\lambda I - h_0 g)$ (i.e., to (-1) times the trace of $h_0 g$) equals the map taking γ to

$$(-1) \cdot (s_1 \gamma_{1,1} + s_2 \gamma_{2,2} + \dots + s_n \gamma_{n,n}).$$

In general, the derivative of the map taking g to the coefficient of λ^{n-k} $(1 \leq k \leq n-1)$ in $\det(\lambda I - h_0 g)$ equals the map taking γ to

$$(-1)^k \cdot (c_{k,1} \gamma_{1,1} + c_{k,2} \gamma_{2,2} + \dots + c_{k,n} \gamma_{n,n}),$$

where $c_{k,i}$ is the sum of all monomials $s_{j_1} s_{j_2} \dots s_{j_k}$, $1 \leq j_1 < j_2 < \dots < j_k \leq n$, such that one of the indices j_l equals i. (For example, $c_{2,1} = s_1 \cdot (s_2 + s_3 + \dots + s_n)$.) Thus, our task is to find for which $s_1, s_2, \dots, s_n$ the determinant

$$|c_{i,j} - c_{i,n}|_{1 \le i,j \le n-1}$$

is non-zero. Clearly, this will happen precisely when

$$|c_{i-1,j}|_{1 \le i,j \le n} \neq 0,$$

where we adopt the (sensible) convention that $c_{0,j} = 1$ for all j.

A brief computation gives us that

$$|c_{i-1,j}|_{1 \le i,j \le n} = (-1)^{\lfloor n/2 \rfloor} \cdot |s_j^{i-1}|_{1 \le i,j \le n}.$$

This is a Vandermonde determinant; it equals $\prod_{j_1 < j_2} (s_{j_2} - s_{j_1})$. The equation $s_{j_2} = s_{j_1}$ defines a subvariety of positive codimension in the variety $V \subset \mathbb{A}^n$ of all $s_1, s_2, \dots, s_n$ with $s_1 s_2 \cdots s_n = 1$. Thus, we may choose $s_1, s_2, \dots, s_n$ such that $s_1 s_2 \cdots s_n = 1$ and $\prod_{j_1 < j_2} (s_{j_2} - s_{j_1}) \neq 0$. □

The proposition below can be applied with W empty. We will later need to invoke it with W equal to the variety of elements of G that are not regular semisimple.

Proposition 5.6. *Let $G = \mathrm{SL}_n$ be defined over a field K. Let $X = G^{n+1}$, $Y = \mathbb{A}^{(n-1)\cdot(n+1)} = \mathbb{A}^{n^2-1}$. Let $f : X \times G \to Y$ be the map given by*

$$f((h_0, h_1, \dots, h_n), g) = (\kappa(h_0 g), \kappa(h_1 g), \dots, \kappa(h_n g)).$$

Let W be a proper subvariety of G (which may be empty). Let $A \subset G(K)$ be a set of generators of $G(K)$.

Then there are elements $h_0, h_1, \dots, h_n \in A_k$, $k \ll_{n,\overrightarrow{deg}(W)} 1$, such that

$$|f((h_0, h_1, \dots, h_n), A_k \setminus (A_k \cap W(K)))| \gg_{n,\overrightarrow{deg}(W)} |A|.$$

Proof. Let $Z_{X \times G}$ be as in Lem. 4.7. By Lemma 5.5, at least one point of $(X \times G)(\overline{K})$ lies outside $Z_{X \times G}$; thus $Z_{X \times G}$ is a proper subvariety of $X \times G$. By the argument in §2.5.3, the points x_0 on X such that $(Z_{X \times G})_{x=x_0}$ is all of G lie on a proper subvariety Z_X of X of degree $\ll_{\overrightarrow{\deg}(Z_{X \times G})} 1$ (and so, by (4.1), $\overrightarrow{\deg}(Z_X) \ll_n 1$). By Lem. 4.6, there are points of $X(G)$ outside Z_X, provided that we assume that $|K|$ is greater than a constant depending only on n. (If $|K| \ll_n 1$, what we seek to prove is trivially true.) We can then use escape from groups (Lem. 4.4) to the group $X = G^{n-1}$ and the set of generators $E = A \times A \times \cdots \times A$ of X, and obtain that there is a tuple $\vec{h} = (h_0, h_1, \dots, h_n) \in E_k$, $k \ll_n 1$, such that $\vec{h}$ lies on $X \setminus Z_X$.

Define $V = (Z_{X \times G})_{x=\vec{h}} \cup W$. Again by Lem. 4.6, there are points of $G(K)$ outside V (assuming again, as we may, that $|K|$ is greater than a constant depending only on n). We can then use a general result on the consequences of being non-singular almost everywhere, namely, Cor. 4.5 (with $E = A$) and obtain that

$$|f_{\vec{h}}(A_{k'} \cap (G(K) \setminus V(K)))| \gg_{\overrightarrow{\deg}(G),\overrightarrow{\deg}(V),\deg_{\mathrm{pol}}(f_{\vec{h}})} |A|,$$

and so (since $\overrightarrow{\deg}(G) \ll_n 1$, $\deg_{\mathrm{pol}}(f_{\vec{h}}) \ll_n 1$, $\overrightarrow{\deg}(V) \ll_{n,\overrightarrow{\deg}(W)} 1$ and $W \subset V$)

$$|f_{\vec{h}}(A_{k'} \cap (G(K) \setminus W(K)))| \gg_{n,\overrightarrow{\deg}(W)} |A|,$$

where $k' \ll_{n,\overrightarrow{\deg}(W)} 1$. □

Recall that $\mathrm{Cl}_G(A)$ denotes the set of all conjugacy classes in G that contain at least one element of A.

Corollary 5.7. *Let $G = \mathrm{SL}_n$. Let K be a field. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Let W be a (possibly empty) proper subvariety of G. Then

$$|\mathrm{Cl}_G(A_k \setminus (A_k \cap W(K)))| \gg_{n,\overrightarrow{\deg}(W)} |A|^{\frac{1}{n+1}},$$

where $k \ll_{n,\overrightarrow{\deg}(W)} 1$. In particular,

$$|\mathrm{Cl}_G(A_k)| \gg_n |A|^{\frac{1}{n+1}},$$

where $k \ll_n 1$.

Here $\frac{1}{n+1} = \frac{n-1}{n^2-1}$ is the exponent one would expect for SL_n: the variety $G = \mathrm{SL}_n$ is of dimension $n^2 - 1$, and the characteristic polynomial of a matrix has $n - 1$ coefficients other than the leading and the constant terms, which are identically 1.

Proof. From Prop. 5.6, we have that there are at least $\gg_{n,\overrightarrow{\deg}(W)} |A|$ distinct $(n+1)$-tuples $(\kappa(g_0), \kappa(g_1), \dots, \kappa(g_n))$, where $g_j = h_j g$ is an element of A_{2k}, $k \ll_{n,\overrightarrow{\deg}(W)} 1$. Clearly, this implies that there are at least $\gg_{n,\overrightarrow{\deg}(W)} |A|^{1/(n+1)}$ distinct elements $\kappa(g)$, $g \in A_{2k}$. Two matrices g, g' in distinct conjugacy classes in G cannot have the same characteristic polynomial. The statement now follows immediately. □

5.3. **From conjugacy classes to a maximal torus.** The following lemma uses nothing, and yet the rest of the section spins around it. As before, given $A \subset G$, we denote by $\mathrm{Cl}_G(A)$ the set of conjugacy classes of G occupied by elements of A.

Proposition 5.8. *Let G be a group. Let $A, A' \subset G$. Then there is a $g \in A'$ such that*

$$|C_G(g) \cap A^{-1}A| \geq \frac{|A|}{|AA'A^{-1}|} \cdot |\mathrm{Cl}_G(A')|.$$

Proof. Write c_g for the number of elements of $A^{-1}A$ commuting with a given $g \in G$. For every g,

$$|\{hgh^{-1} : h \in A\}| \geq \frac{|A|}{c_g}.$$

(Otherwise there would be a $h_0 \in A$ such that $h_0 g h_0^{-1} = hgh^{-1}$ for more than c_g elements h of A – and, since $h_0 g h_0^{-1} = hgh^{-1}$ implies that $h^{-1}h_0 \in A^{-1}A$ commutes with g, we would have a contradiction.) At the same time, for g_1, g_2 in different conjugacy classes,

$$\{hg_1h^{-1} : h \in A\} \qquad \text{and} \qquad \{hg_2h^{-1} : h \in A\}$$

are disjoint.

Hence

$$|\{hgh^{-1} : h \in A, g \in A'\}| \geq \sum_g \frac{|A|}{c_g},$$

where the sum is over representatives $g \in A'$ of conjugacy classes intersecting A'. Therefore, there is a $g \in A'$ such that

$$\frac{|A|}{c_g} \leq \frac{1}{|\operatorname{Cl}_G(A')|} \cdot |\{hgh^{-1} : h \in A, g \in A'\}|,$$

and so

$$c_g \geq \frac{|A|}{|\{hgh^{-1} : h \in A, g \in A'\}|} \cdot |\operatorname{Cl}_G(A')| \geq \frac{|A|}{|AA'A^{-1}|} \cdot |\operatorname{Cl}_G(A')|.$$

□

Remark. It should be clear from the proof that $c_g = |C_G(g) \cap A^{-1}A|$ is large not just for one $g \in A'$, but for many $g \in A'$. We shall not need this fact.

We say that an element of an algebraic group is *regular semisimple* if its centraliser is a torus. The following is a special case of a much more general statement.

Lemma 5.9. *Let $G = \mathrm{SL}_n$. Let K be a field. Then there is a subvariety W/K of G of positive codimension and degree $\overrightarrow{deg}(W) \ll_n 1$ such that every element $g \in G(K)$ not on W is regular semisimple.*

In fact, the variety W will be defined over $\mathbb{Z}$, independently of K; we just need to check that it has positive codimension over K, i.e., that there are points in $G(\overline{K}) \setminus W(\overline{K})$.

Proof. An element of SL_n is regular semisimple if (and only if) its eigenvalues are distinct. Let W be the variety of all $g \in G$ whose characteristic polynomials have multiple roots, i.e., define W by $\mathrm{Disc}(\det(\lambda I - g)) = 0$. As we can easily find points in $G(\overline{K}) \setminus W(\overline{K})$ (say, diagonal elements with distinct entries), we are done. □

Since W is a subvariety of SL_n of positive codimension, we may escape from it.

Corollary 5.10 (to Prop. 5.8)**.** *Let $G = \mathrm{SL}_n$. Let K be a field. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Then there is a maximal torus $T/\overline{K}$ of G such that

$$|A_k \cap T(K)| \gg_n \frac{|A|}{|A_{k+2}|} \cdot |A|^{\frac{1}{n+1}}, \tag{5.7}$$

where $k \ll_n 1$.

If $|A \cdot A \cdot A| \ll |A|^{1+\epsilon}$, then (by the tripling lemma, viz., Lem. 2.2) the inequality (5.7) reads: $|A_k \cap T(K)| \gg |A|^{\frac{1}{n+1} - O_k(\epsilon)}$.

Proof. Let W be as in Lemma 5.9. By Corollary 5.7,

$$\operatorname{Cl}_G(A') \gg |A|^{\frac{1}{n+1}},$$

where $A' = A_k \setminus (A_k \cap W(K))$. At the same time, by Prop. 5.8,

$$|C_G(g) \cap A^{-1}A| \geq \frac{|A|}{|AA'A^{-1}|} \cdot |\operatorname{Cl}_G(A')|$$

for some $g \in A'$.

By the definition of W, all elements of A' are regular semisimple; in other words, the centraliser $C_G(g)$ is a torus (and, in fact, will have to be maximal, since any element of

$G(\overline{K})$ commuting with all elements of $C_G(g)$ must, in particular, commute with g, and thus must lie in $C_G(g)$ already). Hence

$$|T(K) \cap A^{-1}A| \gg_n \frac{|A|}{|AA'A^{-1}|} \cdot |A|^{\frac{1}{n+1}} \geq \frac{|A|}{|A_{k+2}|} \cdot |A|^{\frac{1}{n+1}},$$

where $T = C_G(g)$. □

5.4. **An upper bound on the number of conjugacy classes.** Consider a set $A \subset \mathrm{SL}_n(K)$ such that $|A \cdot A \cdot A| \ll |A|^{1+\epsilon}$. Using Prop. 5.8 and the fact that there are not too few conjugacy classes, we have just shown that there is a torus T such that there are not too few elements on T. Using, again, Prop. 5.8 and the fact that there are not too many elements on T, we shall now show that there are not too many conjugacy classes.

Corollary 5.11 (to Cor. 5.4 and Prop. 5.8)**.** *Let $G = \mathrm{SL}_n$. Let K be a field. Let $A \subset G(K)$ be any set of generators of $G(K)$. Assume that $|A|$ is greater than a constant depending on n.*

Then

$$|\operatorname{Cl}_G(A \cap \Sigma(K))| \ll_n \frac{|AAA^{-1}|}{|A|} |A_k|^{\frac{1}{n+1}},$$

where $k \ll_n 1$ and Σ is the Zariski-open set of regular semisimple elements of G.

Proof. Let $A' = A \cap \Sigma(K)$. By Prop. 5.8, there is a $g \in A'$ such that

$$|C_G(g) \cap A^{-1}A| \geq \frac{|A|}{|AA'A^{-1}|} \cdot |\operatorname{Cl}_G(A')|.$$

Since g is regular semisimple, its centraliser $T = C_G(g)$ is a maximal torus. By Cor. 5.4 (applied to $A^{-1}A$ rather than A),

$$|T(K) \cap A^{-1}A| \ll_n |A_s|^{\frac{\dim(T)}{\dim(G)}} = |A_s|^{\frac{1}{n+1}},$$

where $s \ll_n 1$. Thus

$$|\operatorname{Cl}_G(A')| \ll_n \frac{|AA'A^{-1}|}{|A|} |A_s|^{\frac{1}{n+1}}.$$

□

In brief: we already knew that, for any set of generators A,

$$|A \cap T(K)| \ll |A_k|^{\frac{1}{n+1}} \quad \text{and}$$
$$|A|^{\frac{1}{n+1}} \ll |\operatorname{Cl}(A_k)|.$$

We now know that, if A does not grow (i.e., $|A \cdot A \cdot A| \ll |A|^{1+\epsilon}$) then the inequalities can be reversed:

$$|A_k|^{\frac{1}{n+1}-O(\epsilon)} \ll |A_k \cap T(K)| \quad \text{and} \qquad (5.8)$$
$$|\operatorname{Cl}(A')| \ll |A_k|^{\frac{1}{n+1}+O(\epsilon)},$$

where A' is the set of regular semisimple elements of A.

Our plan in §6 will be to derive a contradiction from this tight situation. We shall eventually construct what may be seen as a counterexample to an incidence theorem: the elements of the torus shall give us the lines (i.e., the linear relations), and the conjugacy classes shall give us the points. There will be too many lines with many points on each and too few points in total.

Before we finish this section, we must do some auxiliary work on intersections with non-maximal tori.

5.5. **Intersections with non-maximal tori.** We already know that we can find a torus T such that $|A \cap T(K)|$ is large; we will now show that, for any T and for any subtorus $T' \subset T$ given as the kernel of a character of T, the intersection $|A \cap T'(K)|$ is small.

(A *character* of a maximal torus is a homomorphism from $T(\overline{K})$ to $\overline{K}^*$ given as an algebraic map defined over $\overline{K}$. If $G = \mathrm{SL}_n$ and T is a maximal torus of G given as the group of diagonal matrices, then the characters are the maps of the form $t \mapsto \prod_{1\leq j\leq n} t_{jj}^{m_j}$, where $m_j \in \mathbb{Z}$ and $\sum_j m_j = 0$. There is an analogous notion of *character* for Lie algebras; the characters of a Cartan subalgebra of sl_n (seen as the algebra of diagonal matrices) are maps of the form $t \mapsto \sum_j m_j t_{jj}$, where $m_j \in \mathbb{Z}$ and $\sum_j m_j = 0$.)

Lemma 5.12. *Let $\mathfrak{g} = \mathrm{sl}_n$ be defined over a field K. Assume $\mathrm{char}(K) \nmid n$. Let $\mathfrak{t}$ be a Cartan subalgebra of $\mathfrak{g}$. Let $\mathfrak{t}'$ be the kernel of a non-trivial character $\alpha : \mathfrak{t} \to \mathbb{A}^1$.*

Then there are elements $g_0, g_1, \ldots, g_n \in \mathfrak{g}(K)$ such that the spaces

$$\mathfrak{t}', [g_0, \mathfrak{t}'], [g_1, \mathfrak{t}'], [g_2, \mathfrak{t}'], \ldots, [g_n, \mathfrak{t}'] \tag{5.9}$$

are linearly independent and of dimension $\dim(\mathfrak{t}')$.

Proof. Write the elements of $\mathfrak{g}$ as matrices so that $\mathfrak{t}$ becomes the algebra of diagonal matrices with trace 0. Let $e_{i,j}$ be the matrix having a 1 at the (i,j)th entry and 0s at all other entries. We define

$$g_j = \sum_{\substack{i \\ i\neq j}} e_{i,j}$$

for $1 \leq j \leq n$, where the sum goes through all i from 1 to n other than j. For every $t \in \mathfrak{t}$, the matrix $[g_j, t]$ has 0s at the (j,j)th entry and throughout all columns save for the jth column. In fact, for any $g \in \mathfrak{g}(K)$ and any $t \in \mathfrak{t}(K)$, the matrix $[g, \mathfrak{t}]$ has 0s throughout the diagonal.

Thus, it remains only to find a $g_0 \in \mathfrak{g}(K)$ not in $\mathfrak{t}$ such that the linear space $V_0 = [g_0, \mathfrak{t}']$ and the linear space

$$V = [g_1, \mathfrak{t}'] + [g_2, \mathfrak{t}'] + \ldots + [g_n, \mathfrak{t}']$$

intersect only at the origin. (We can already see that each space $[g_i, \mathfrak{t}']$, $1 \leq i \leq n$, intersects the sum of all the others only at the origin, and that the space $\mathfrak{t}$ intersects the sum V of all of them only at the origin. Since g_0 will not be in $\mathfrak{t}$, $V_0 = [g_0, \mathfrak{t}']$ will have the same dimension as $\mathfrak{t}$.)

For $1 \leq i_0 \leq n$, let $s(i_0)$ be the matrix in $\mathfrak{t}$ having $(s(i_0))_{ii} = -1$ for all $i \neq i_0$ and $(s(i_0))_{i_0 i_0} = n - 1$. If $\mathrm{char}(K) \nmid n$, such matrices span $\mathfrak{t}(K)$ as a linear space; since $\mathfrak{t}' \neq \mathfrak{t}$, there must be at least one such matrix $s(i_0)$ not in $\mathfrak{t}'(K)$. Fix that i_0 from now on. Because $s(i_0) \notin \mathfrak{t}'(K)$, there is no non-zero matrix t' in $\mathfrak{t}'$ with all of its diagonal entries other than t'_{i_0,i_0} equal to each other. Again by $\mathrm{char}(K) \nmid n$, there is also no non-zero matrix t' in $\mathfrak{t}'$ with all of its diagonal entries equal to each other.

Now define

$$g_0 = \sum_{\substack{j \\ j \neq i_0}} e_{i_0,j}. \tag{5.10}$$

Suppose there is a $t \in \mathfrak{t}$ such that $[g_0, t] \in V$. Then, for every j between 1 and n, the jth column of the matrix $[g_0, t]$ equals the jth column of the matrix $[g_j, t'_j]$ for some $t'_j \in \mathfrak{t}$.

Let $1 \leq j \leq n$. If $j = i_0$, then, as can be computed easily from the definition (5.10) of g_0, the jth column of g_0 has all of its entries equal to 0. Suppose $j \neq i_0$. Then the jth column of $[g_0, t]$ has all of its entries equal to 0 save for the (i_0)th entry, which is equal to $t_{j,j} - t_{i_0,i_0}$. The jth column of $[g_j, t'_j]$ is

$$\begin{pmatrix} t'_{jj} - t'_{11} \\ t'_{jj} - t'_{22} \\ \vdots \\ t'_{jj} - t'_{nn} \end{pmatrix}.$$

If these two jth columns were equal, then all diagonal entries of t' (save possibly for t'_{j_0,j_0}) are equal to each other. As we have seen, this implies that $t' = 0$. Hence the jth column of $[g_0, t]$ has all of its entries equal to 0.

We let j vary between 1 and n and obtain that, in every column of $[g_0, t]$, all of the entries are equal to 0; in other words, $[g_0, t] = 0$. We conclude that the intersection of $[g_0, \mathfrak{t}]$ and V is $\{0\}$. □

Proposition 5.13. *Let $G = \mathrm{SL}_n$. Let K be a field. Assume that K is either finite with $\mathrm{char}(K) \nmid n$ or infinite with $\mathrm{char}(K) = 0$. Let T be a maximal torus of G defined over $\overline{K}$. Let $\alpha : T \to \mathbb{A}^1$ be a character of T, and let T' be the kernel of α.*

For every $(g_0, g_1, \ldots, g_n) \in (G(K))^{n+1}$, let $f_{g_0,g_1,\ldots,g_n} : (T')^{n+2} \to G$ be the map defined by

$$f_{g_0,g_1,\ldots,g_n}(t, t_0, t_1, \ldots, t_n) = t \cdot g_0 t_0 g_0^{-1} \cdot g_1 t_1 g_1^{-1} \cdots g_n t_n g_n^{-1}. \tag{5.11}$$

Let $A \subset G(K)$ be a set of generators of $G(K)$, and let E be a non-empty subset of $T'(K)$. Then there are $g_0, g_1, \ldots, g_n \in A_k$, $k \ll_n 1$, such that

$$|f_{g_0,g_1,\ldots,g_n}(E, E, \ldots, E)| \gg_{n,\deg(T')} |E|^{n+2}.$$

Proof. We may assume that the derivative of α does not vanish at the origin: if it does, then the characteristic $\mathrm{char}(K)$ is equal to p for some prime p, and $\alpha = \beta^p$ for some character $\beta : T \to \mathbb{A}^1$; since the Frobenius map $x \to x^p$ is an automorphism for K finite, it follows that $\ker(\alpha) = \ker(\beta)$, and so we can use β instead of α. (Repeat if needed.)

The Lie algebra $\mathfrak{t}'$ of T' lies in the kernel of the derivative α_0 of α at the origin, which is a character of the Lie algebra $\mathfrak{t}$ of T; as we have just said, α_0 is not identically zero. We may thus apply Lemma 5.12; it asserts that the assumptions of Prop. 4.13 are fulfilled (with $\ell = n+1$ and $\mathfrak{h} = \mathfrak{t}'$). The conclusions of Prop. 4.13 provide the linear-independence assumption of Prop. 4.12 (for $H_0 = H_1 = \ldots = H_\ell = T'$); we apply Prop. 4.12, and are done. □

Corollary 5.14. *Let $G = \mathrm{SL}_n$. Let K be a field. Assume that K is either finite with $\mathrm{char}(K) \nmid n$ or infinite with $\mathrm{char}(K) = 0$. Let T be a maximal torus of G defined over $\overline{K}$. Let $\alpha : T \to \mathbb{A}^1$ be a character of T, and let T' be the kernel of α.*

Let $A \subset G(K)$ be a set of generators of $G(K)$. Then

$$|A \cap T'(K)| \ll_{n,\deg(T')} |A_k|^{\frac{1}{n+2}}, \tag{5.12}$$

where $k \ll_n 1$.

Proof. Immediate from Prop. 5.13 and the definition of $f_{g_0,g_1,\dots,g_n}$. □

Since we already know from Cor. 5.10 that $|A_r \cap T(K)| \gg |A_{rk}|^{\frac{1}{n+1}-O(\epsilon)}$ for some $r \ll_n 1$ (assuming that $|A \cdot A \cdot A| \ll |A|^{1+\epsilon}$), the inequality (5.12) (applied to A_r instead of A) implies that only a very small fraction of the elements of $A_r \cap T(K)$ lie in the kernel T' of a given character α.

5.6. **Special tuples of coefficients of characteristic polynomials.** In §6, we will need to work with tuples of the form

$$(\kappa(h_0 g), \kappa(tg), \kappa(t^2 g), \dots, \kappa(t^n g)).$$

We need to show that there are many such tuples. Corollary 5.14 will make a crucial appearance towards the end.

Proposition 5.15. *Let $G = \mathrm{SL}_n$. Let K be a finite field. Let $W/\overline{K}$ be a proper subvariety of G. Let $T/\overline{K}$ be a maximal torus of G. Let $A \subset G(K)$ be a set of generators of $G(K)$. Let $E \subset T(K)$. Then, provided that $|K|$ is larger than a constant depending only on n, either*

(a) *there is an element $h_0 \in A_k$, $k \ll_n 1$, and a subset $E' \subset E_k$ with $|E'| \gg_n |E|$ such that, for each $t \in E'$, there are $\gg_{n,\overrightarrow{\deg}(W)} |A|$ distinct tuples*

$$(\kappa(h_0 g), \kappa(tg), \kappa(t^2 g), \dots, \kappa(t^n g)) \in \mathbb{A}^{n^2-1}(K)$$

with $g \in A_{k'}$, $k' \ll_{n,\overrightarrow{\deg}(W)} 1$ satisfying $h_0 g \notin W(K), g, tg, t^2 g, \dots, t^n g \notin W(K)$, or

(b) *E is contained in the kernel of a non-trivial character $\alpha : T \to \mathbb{A}^1$ whose exponents are bounded in terms of n alone.*

Proof. Let $X = G \times T$ and $Y = (\mathbb{A}^{n-1})^{n+1} = \mathbb{A}^{n^2-1}$. Let $f : X \times G \to Y$ be given by

$$f((h,t),g) = (\kappa(hg), \kappa(tg), \kappa(t^2 g), \dots, \kappa(t^n g)).$$

Let $Z_{X\times G}$ be as in Lemma 4.7; by (4.1), $\overrightarrow{\deg}(Z_{X\times G}) \ll_n 1$. Thanks to Lem. 5.5, we know $Z_{X\times G}$ is a proper subvariety of $X \times G$.

Let $Z_{G\times T\times G}$ be $Z_{X\times G}$ under the identification $G \times T \times G = X \times G$; write the elements of $Z_{G\times T\times G}$ in the form (h,t,g). By the argument in §2.5.3, there is a proper subvariety $Z_G \subset G$ (with $\overrightarrow{\deg}(Z_G) \ll_{\overrightarrow{\deg}(Z_{G\times T\times G})} 1$, and so $\overrightarrow{\deg}(Z_G) \ll_n 1$) such that, for all $h_0 \in G(\overline{K})$ not on Z_G, the fibre $(Z_{G\times T\times G})_{h=h_0}$ is a proper subvariety of $T \times G$.

By escape from groups (Lem. 4.4 and Lem. 4.6; it is here that that $|K| \gg_n 1$ is used), there is an $h_0 \in A_k$, $k \ll_n 1$, such that h_0 lies outside Z_G; thus, by the definition of Z_G, the fibre $V_{T\times G} := (Z_{G\times T\times G})_{h=h_0}$ is a proper subvariety of $T \times G$. Again by §2.5.3, there is a proper subvariety V_T with $\overrightarrow{\deg}(V_T) \ll_{\overrightarrow{\deg}(V_{T\times G})} 1$ (and so $\overrightarrow{\deg}(V_T) \ll_n 1$) such that, for all $t_0 \in T(\overline{K})$ not on V_T, the fibre $(V_{T\times G})_{t=t_0}$ is a proper subvariety of G.

Suppose first that $\langle E\rangle \not\subset V_T(K)$. We may then use escape from subvarieties (Prop. 4.1 with $A = E$, $V = V_T(K)$ and $G = \mathscr{O} = \langle E\rangle$) to obtain a subset $E' \subset E_k$ ($k \ll_n 1$) with $|E'| \gg_n |E|$ and $E' \subset T(K) \setminus V_T(K)$. Now consider any $t_0 \in E'$. The fibre $(V_{T\times G})_{t=t_0}$ is a proper subvariety of G, and, since W is a proper subvariety of G, we conclude that

$$V' = (V_{T\times G})_{t=t_0} \cup h_0^{-1}W \cup W \cup t_0^{-1}W \cup \ldots \cup t_0^{-n}W.$$

is a proper subvariety of G as well (with $\overrightarrow{\deg}(V') \ll_{n,\overrightarrow{\deg}(W)} 1$). We now recall the definition of $Z_{X\times G}$ (a variety outside which the map f is non-singular) and use the result on non-singularity (Corollary 4.5 applied to the function $f_{h_0,t_0} : G \to Y$ given by $f_{h_0,t_0}(g) = f((h_0,t_0),g)$; here Lem. 4.6 supplies the condition $V(K) \subsetneq G(K)$, which is a requirement for the application of Cor. 4.5) to obtain that

$$|f_{h_0,t_0}(A_{k'} \cap (G(K) \setminus V'(K)))| \gg_{n,\overrightarrow{\deg}(W)} |A|$$

with $k' \ll_{n,\overrightarrow{\deg}(W)} 1$. This gives us conclusion (a).

Suppose now that $\langle E\rangle \subset V_T(K)$. Then, by Prop. 4.2, $\langle E\rangle$ is contained in an algebraic subgroup H of T of positive codimension and degree $\overrightarrow{\deg}(H) \ll_{\overrightarrow{\deg}(V_T(K))} 1$ (and so $\overrightarrow{\deg}(H) \ll_n 1$). By Lemma 4.14, we obtain that H is contained in the kernel of a non-trivial character $\alpha : T \to \mathbb{A}^1$ whose exponents are $\ll_n 1$. □

As before, we write Σ for the (algebraic) set of regular semisimple elements of G; in the case of $G = \mathrm{SL}_n$, this is simply the (algebraic) set consisting of every g whose eigenvalues are all distinct. The sets of points $\Sigma(K)$ and $\Sigma(\overline{K})$ are what one would expect, viz., the sets consisting of the elements of $G(K)$ and $G(\overline{K})$ having distinct eigenvalues. For $G = \mathrm{SL}_n$, the complement of Σ is a variety W with $\overrightarrow{\deg}(W) \ll_n 1$.

Corollary 5.16. *Let $G = \mathrm{SL}_n$. Let K be a field. Assume that K is either finite with $\mathrm{char}(K) \nmid n$ or infinite with $\mathrm{char}(K) = 0$. Let $T/\overline{K}$ be a maximal torus of G. Let $A \subset G(K)$ be a set of generators of $G(K)$. Suppose that $|A|$ is greater than a constant depending only on n.*

Then there is an ϵ_0 depending only on n such that, if $|A\cdot A\cdot A| \leq |A|^{1+\epsilon}$ for some positive $\epsilon < \epsilon_0$, then there is an element $h_0 \in A_k$, $k \ll_n 1$, and a subset $E' \subset A_k \cap T(K)$ with $|E'| \gg_n |A_k \cap T(K)|$ such that, for each $t \in E'$, there are $\gg_n |A|$ distinct tuples

$$(\kappa(h_0 g), \kappa(tg), \kappa(t^2 g), \ldots, \kappa(t^n g)) \in \mathbb{A}^{n^2-1}(K)$$

with $g \in A_k$ satisfying $h_0 g \in \Sigma(K)$ and $t^\ell g \in \Sigma(K)$ for $\ell = 0,1,2,\ldots,n$, where k and the implied constants depend only on n.

Proof. By Corollary 5.10 and the tripling lemma (Lem. 2.2),

$$|A_k \cap T(K)| \gg_n |A|^{\frac{1}{n+1} - O_n(\epsilon)}, \tag{5.13}$$

where $k \ll_n 1$. Let $E = A_k \cap T(K)$; let W be the complement of Σ. Apply Prop. 5.15. If case (a) of Prop. 5.15 applies, we are done.

It remains only to rule out case (b) of Prop. 5.15. Suppose E is contained in the kernel T' of a non-trivial character $\alpha : T \to \mathbb{A}^1$ whose exponents are $\ll_n 1$. Then $\deg(T') \ll_n 1$. We now apply Cor. 5.14 (to A_k rather than A), and obtain that that

$$|E| \ll_n |A_{kk'}|^{\frac{1}{n+2}} \ll_n |A|^{\frac{1}{n+2}+O_n(\epsilon)}$$

for some $k' \ll_n 1$, in contradiction to (5.13). (Recall that $E = A_k \cap T(K)$.) □

6. Growth of small and large sets in SL_2 and SL_3

For the sake of clarity and completeness, we shall do things twice: once for SL_2 and once for SL_3. Of course, in the case of SL_2, we could refer to [He] instead; since, however, the method in this paper is somewhat different – especially in this part of the argument – we would like to work things out for both SL_2 and SL_3.

The key observation in the proofs below is the following. Consider $n+1$ diagonal matrices $t_0, t_1, \dots, t_n \in A_k$. The maps $g \mapsto \mathrm{tr}(t_0 g)$, $g \mapsto \mathrm{tr}(t_1 g)$, $\dots$, $g \mapsto \mathrm{tr}(t_n g)$ from $\mathrm{SL}_n(K)$ to K can be seen as linear forms – that is, homogeneous linear polynomials – on the n variables $g_{1,1}, \dots, g_{n,n}$ (the diagonal entries of g).

Any $n+1$ linear forms on n variables must be linearly dependent. Hence there are coefficients $c_0, c_1, \dots, c_n \in K^n$ depending on $t_1, t_2, \dots, t_n$ (but not on g) such that

$$c_0 \,\mathrm{tr}(t_0 g) + c_1 \,\mathrm{tr}(t_1 g) + \cdots + c_n \,\mathrm{tr}(t_n g) = 0 \tag{6.1}$$

for all g. Thus we have a linear relation holding for many tuples (namely, the tuples $(\mathrm{tr}(t_0 g), \mathrm{tr}(t_1 g), \dots, \mathrm{tr}(t_n g))$ for any $g \in A$) all of whose entries $\mathrm{tr}(t_j g)$ lie in a small set (viz., $\mathrm{tr}(A_{k+1})$).

As $t_0, t_1, \dots, t_n$ vary, the coefficients $c_0, c_1, \dots, c_n$ will vary as well. We will obtain too many linear relations (of the form (6.1)), and thus a contradiction to Corollary 3.8.

6.1. **Small sets in SL_2.** The treatment of SL_2 in [He] was based on the identity

$$(x + x^{-1})(y + y^{-1}) = (xy + (xy)^{-1}) + (xy^{-1} + (xy^{-1})^{-1}), \tag{6.2}$$

which is a special case of the identity

$$\mathrm{tr}(g)\,\mathrm{tr}(h) = \mathrm{tr}(gh) + \mathrm{tr}(gh^{-1}) \tag{6.3}$$

valid in SL_2 (but not in SL_n, $n > 2$). We shall now do without (6.2) and (6.3).

Proposition 6.1. *Let $G = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, p a prime. Let $A \subset G$ be a set of generators of G. Assume $|A| < p^{3-\delta}$, $\delta > 0$. Then*

$$|A \cdot A \cdot A| \gg_\delta |A|^{1+\epsilon}, \tag{6.4}$$

where $\epsilon > 0$ depends only on δ.

This is part (a) of the Key Proposition in [He].

Proof. Suppose $|A \cdot A \cdot A| \leq |A|^{1+\epsilon}$. Then, by the tripling lemma (Lem. 2.2), $|A_\ell| \leq |A|^{1+O_\ell(\epsilon)}$ for every ℓ. Starting from here, we shall arrive at a contradiction for ϵ small.

By Corollary 5.10, there is a maximal torus $T/\overline{K}$ of G such that

$$|A_k \cap T(K)| \gg \frac{|A|}{|A_{k+2}|} |A|^{1/3} \geq |A|^{\frac{1}{3}-O(\epsilon)}, \tag{6.5}$$

where k and the implied constants are absolute. We may write the elements of $T(K)$ as diagonal matrices, after conjugation by an appropriate element of $\mathrm{SL}_2(\overline{K})$. We can thus see that any 3 elements $t_0, t_1, t_2 \in A_k \cap T(K)$ are linearly dependent. (Linear dependences are invariant under conjugation.)

In particular, for $t_0 = I$, $t_1 = t$, $t_2 = t^2$ ($t \in T(K)$ given), we have

$$c_0 t_0 + c_1 t_1 + c_2 t_2 = 0 \tag{6.6}$$

for $c_0 = 1$, $c_1 = -(r + r^{-1})$ and $c_2 = 1$, where r and r^{-1} are the eigenvalues of t. The map

$$\phi : t \mapsto (c_0, c_1, c_2)$$

from T to $\mathbb{A}^3$ is almost injective: the preimage of any point $(c_0, c_1, c_2) \in K^3$ consists of at most two elements of $T(\overline{K})$. (The only thing that is particularly good about the choice $t_0 = I$, $t_1 = t$, $t_2 = t^2$ is that this almost-injectivity is easy to prove for this choice, as we have just seen.)

It follows immediately from (6.6) that, for any $g \in G$,

$$c_0 \operatorname{tr}(t_0 g) + c_1 \operatorname{tr}(t_1 g) + c_2 \operatorname{tr}(t_2 g) = 0. \tag{6.7}$$

If $g \in A$, then $t_0 g, t_1 g, t_2 g \in A_{k+2}$. (To see this, note that, if a basis is chosen for which t is diagonal and $\operatorname{tr}(t_i g)$ is then written out in full, the only entries of g appearing in $\operatorname{tr}(t_i g)$ are the diagonal entries g_{ii}; moreover, the coefficient of g_{ii} in (6.7) is $c_0(t_0)_{ii} + c_1(t_1)_{ii} + c_2(t_2)_{ii}$, which is 0 by (6.6).)

It is worthwhile to examine A_{k+2} in some more detail. By Corollary 5.11,

$$|\operatorname{tr}(A')| \ll |A_{(k+2)k'}|^{\frac{1}{3}+O(\epsilon)} \ll |A|^{\frac{1}{3}+O(\epsilon)},$$

where A' is the set of regular semisimple elements of A_{k+2}, and k' and the implied constant are absolute. (We may apply Cor. 5.11 because we may assume that $|A|$ is larger than an absolute constant: if $|A|$ is smaller than an absolute constant, the statement we seek to prove is trivial.)

In SL_2, a non-semisimple element has trace 2; thus, we may write simply

$$|\operatorname{tr}(A_{k+2})| \ll |A|^{\frac{1}{3}+O(\epsilon)} + 1 \ll |A|^{\frac{1}{3}+O(\epsilon)}, \tag{6.8}$$

where the implied constants are absolute. (We are assuming, as we may, that $|A|$ is larger than an absolute constant, and that $\epsilon > 0$ is smaller than an absolute constant.)

By escape from subvarieties (as in Lem. 4.4), there is an element $z \in A_{k''}$ (k'' absolute) of the form

$$z = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with a, b, c, d non-zero. (We are still writing elements of $\mathrm{SL}_2(K)$ as matrices in such a way that $T(\overline{K})$ is diagonal.) Then, for any diagonal $t \neq \pm I$, the map

$$g \mapsto (\operatorname{tr}(g), \operatorname{tr}(tg), \operatorname{tr}(zg)) = (g_{11} + g_{22}, t_{11}g_{11} + t_{22}g_{22}, ag_{11} + dg_{22} + bg_{21} + cg_{12})$$

is almost injective on SL_2: since we know that $g_{11}g_{22} - g_{12}g_{21} = 1$, the preimage of any point $(\operatorname{tr}(g), \operatorname{tr}(tg), \operatorname{tr}(zg))$ consists of at most two elements. Now $\operatorname{tr}(zg) \in \operatorname{tr}(A_{k+k''})$, and, as in (6.8),

$$|\operatorname{tr}(A_{k+k''})| \ll |A|^{\frac{1}{3}+O(\epsilon)},$$

where the implied constants (here and everywhere from now on) are absolute. Hence the image of A under the map

$$g \mapsto (\mathrm{tr}(g), \mathrm{tr}(tg))$$

has at least

$$\gg \frac{|A|}{|A|^{\frac{1}{3}+O(\epsilon)}} = |A|^{\frac{2}{3}-O(\epsilon)} \tag{6.9}$$

elements.

We are now in the situation covered by Cor. 3.8: we have many tuples ($\gg |A|^{\frac{2}{3}-O(\epsilon)}$) with entries (namely, $\mathrm{tr}(g)$, $\mathrm{tr}(tg)$ and $\mathrm{tr}(zg)$) in a small set ($|\mathrm{tr}(A_{k+1})| \ll |A|^{\frac{1}{3}+O(\epsilon)}$) and these tuples satisfy many linear relations (one for each element of $T(K) \cap A_k$). More formally: let $R = \mathbb{Z}/p\mathbb{Z}$, $X = \mathrm{tr}(A_{k+2})$,

$$Y = \left\{ (r + r^{-1}, -1) \in ((\mathbb{Z}/p\mathbb{Z})^*)^2 : r \neq \pm i, \begin{pmatrix} r & 0 \\ 0 & r^{-1} \end{pmatrix} \in A_k \cap T(\overline{K}) \right\}.$$

For each $\vec{y} = (r + r^{-1}, -1) \in Y$, let $t_{\vec{y}}$ be an element of $A_k \cap T(\overline{K})$ having r, r^{-1} as its eigenvalues. (There can be at most two such elements for given $\vec{y}$.) We define

$$X_{\vec{y}} = (\mathrm{tr}(t_{\vec{y}}x), \mathrm{tr}(x)).$$

Then, by (6.7), we have

$$y_0 \,\mathrm{tr}(t_{\vec{y}}x) + y_1 \,\mathrm{tr}(x) = \mathrm{tr}(t_{\vec{y}}^2 x),$$

and thus

$$\vec{y} \cdot X_{\vec{y}} \subset X.$$

At the same time,

$$|Y| \geq \frac{1}{2}(A_k \cap T(\overline{K})) - 1 \gg |A|^{\frac{1}{3}-O(\epsilon)} \gg |X|^{1-O(\epsilon)}$$

by (6.5), (6.8) and $X = \mathrm{tr}(A_{k+2})$, and

$$|X_{\vec{y}}| \gg |A|^{\frac{2}{3}-O(\epsilon)} \gg |X|^{2-O(\epsilon)}.$$

by (6.9). (All the constants are absolute.) We apply Cor. 3.8 and reach a contradiction, provided that $\epsilon > 0$ is smaller than a positive constant depending only on η and that $|A|$ is larger than a constant depending only on η, ϵ and δ. (The condition on $|A|$ is needed so that the condition $|X| < p^{1-\delta'}$, $\delta' > 0$, of Cor. 3.8 is fulfilled; we fulfil it by means of (6.8) and the assumption $|A| < p^{3-\delta}$.)

We set $\epsilon > 0$ to be smaller than the positive constant just mentioned. As is stated in Cor. 3.8, η depends only on δ. Hence, for the contradiction to happen, it is enough to assume that $|A|$ is larger than a constant depending only on δ. We can certainly assume this, as otherwise the statement (6.4) is trivially true. We have thus indeed reached a contradiction, and we are done. □

6.2. **Small and fairly large sets in** SL_3**.** The main idea is essentially the same as that in §6.1. Consider 4 diagonal matrices $t_0, t_1, t_2, t_3 \in \mathrm{SL}_3(K)$. The maps from SL_n to $K \times K$ given by

(6.10)
$$g \mapsto \begin{pmatrix} \mathrm{tr}(t_0 g) \\ \mathrm{tr}((t_0 g)^{-1}) \end{pmatrix}, \ g \mapsto \begin{pmatrix} \mathrm{tr}(t_1 g) \\ \mathrm{tr}((t_1 g)^{-1}) \end{pmatrix}, \ g \mapsto \begin{pmatrix} \mathrm{tr}(t_2 g) \\ \mathrm{tr}((t_2 g)^{-1}) \end{pmatrix}, \ g \mapsto \begin{pmatrix} \mathrm{tr}(t_3 g) \\ \mathrm{tr}((t_3 g)^{-1}) \end{pmatrix}$$

can be seen as linear forms (linear over $K \times K$, that is) on 3 variables. The 3 variables in question are
$$\begin{pmatrix} g_{11} \\ (g^{-1})_{11} \end{pmatrix}, \quad \begin{pmatrix} g_{22} \\ (g^{-1})_{22} \end{pmatrix}, \quad \begin{pmatrix} g_{33} \\ (g^{-1})_{33} \end{pmatrix},$$
which are elements of $K \times K$.

(We are interested in tuples of the form
$$\begin{pmatrix} \mathrm{tr}(h) \\ \mathrm{tr}(h^{-1}) \end{pmatrix}, \qquad h \in \mathrm{SL}_3(K),$$
because the tuple $\kappa(h) = (a_2, a_1)$ of coefficients of the characteristic polynomial $t^3 + a_2 t^2 + a_1 t - 1$ of an element h of $\mathrm{SL}_3(K)$ is $\kappa(h) = (-\mathrm{tr}(h), \mathrm{tr}(h^{-1}))$.)

Since the maps (6.10) are linear forms on 3 variables, they must be linearly dependent; that is, for each choice $t_0, t_1, t_2, t_3 \in A_k \cap T(\overline{K})$, there are[4] $c_0, c_1, c_2, c_3 \in K \times K$ such that

$$c_0 \begin{pmatrix} \mathrm{tr}(t_0 g) \\ \mathrm{tr}((t_0 g)^{-1}) \end{pmatrix} + c_1 \begin{pmatrix} \mathrm{tr}(t_1 g) \\ \mathrm{tr}((t_1 g)^{-1}) \end{pmatrix} + c_2 \begin{pmatrix} \mathrm{tr}(t_2 g) \\ \mathrm{tr}((t_2 g)^{-1}) \end{pmatrix} + c_3 \begin{pmatrix} \mathrm{tr}(t_3 g) \\ \mathrm{tr}((t_3 g)^{-1}) \end{pmatrix} = 0 \tag{6.11}$$

for all $g \in \mathrm{SL}_3(K)$.

Varying t_0, t_1, t_2, t_3 within $A_k \cap T(\overline{K})$, we will obtain many linear relations of the form (6.11), and, as in §6.1, we will obtain a contradiction to Cor. 3.8 thereby.

Lemma 6.2. *Let $G = \mathrm{SL}_3$. Let K be a field. Let $T/\overline{K}$ be a maximal torus of G. Let Σ be the Zariski-open set of regular semisimple matrices in G.*

Then there is a map
$$c : (T \cap \Sigma) \to \mathbb{A}^1/\overline{K}$$
such that, for any $t \in (T \cap \Sigma)(\overline{K})$,
$$I - c(t) \cdot t + c(t^{-1}) t^2 - t^3 = 0. \tag{6.12}$$
Moreover, the preimage $\phi^{-1}(\{x\})$ of any $x \in \mathbb{A}^2$ under the map $\phi : (T \cap \Sigma) \to \mathbb{A}^2$ given by $\phi(t) = (c(t), c(t^{-1}))$ has at most 6 elements.

We recall that a matrix in SL_n is *regular semisimple* if and only if all of its eigenvalues are distinct.

[4]Here and henceforth we write elements of $K \times K$ in the form $\begin{pmatrix} a \\ b \end{pmatrix}$. The multiplication rule is
$$\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a \cdot c \\ b \cdot d \end{pmatrix}.$$

Proof. Write the elements of G so that the elements of T become diagonal matrices. Let $t = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$. We define

$$\begin{pmatrix} c_0(t) \\ c_1(t) \\ c_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \lambda_1^3 \\ \lambda_2^3 \\ \lambda_3^3 \end{pmatrix}. \tag{6.13}$$

Then $c_0(t)I + c_1(t)t + c_2(t)t^2 = t^3$. Starting from (6.13), a quick computation (using Cramer's rule, say) gives us $c_0(t) = 1$ and $c_2(t) = -c_1(t^{-1})$. Let $c(t) = -c_1(t)$. Then (6.12) holds.

Now, for any four distinct elements $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \overline{K}^*$, the determinant of the matrix

$$\begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ 1 & \lambda_3 & \lambda_3^2 & \lambda_3^3 \\ 1 & \lambda_4 & \lambda_4^2 & \lambda_4^3 \end{pmatrix} \tag{6.14}$$

is a Vandermonde determinant, and hence (since $\lambda_1, \dots, \lambda_4$ are distinct) non-zero. However, if the same relation (6.12) were satisfied by two matrices the union of whose sets of eigenvalues has at least four distinct elements $\lambda_1, \dots, \lambda_4$, then 1, λ_j, λ_j^2 and λ_j^3 would satisfy the same linear relation (6.12) for $j = 1, 2, 3, 4$. In other words, the columns of the matrix (6.14) would be linearly dependent. We have reached a contradiction. Hence $(c(t), c(t^{-1})) = (c(t'), c(t'^{-1}))$ can hold for $t, t' \in (T \cap \Sigma)(\overline{K})$ only if the set of eigenvalues of t equals the set of eigenvalues of t'. For t given, this can happen for only $3! = 6$ possible values of t'. ∎

Proposition 6.3. *Let $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$, p a prime. Let $A \subset G$ be a set of generators of G. Assume either $|A| \leq p^{4-\delta}$, $\delta > 0$, or $p^{4+\delta} \leq |A| \leq p^{8-\delta}$, $\delta > 0$. Then*

$$|A \cdot A \cdot A| \gg_\delta |A|^{1+\epsilon}, \tag{6.15}$$

where $\epsilon > 0$ depends only on δ.

Proof. Let $K = \mathbb{Z}/p\mathbb{Z}$. We can assume $\mathrm{char}(K) = p > 3$, as otherwise the result to be proven is trivial.

Suppose $|A \cdot A \cdot A| \leq |A|^{1+\epsilon}$. Then $|A_l| \leq |A|^{1+O_l(\epsilon)}$ for every positive l. We shall proceed from here and arrive at a contradiction for ϵ sufficiently small.

By Corollary 5.10, there is a maximal torus $T/\overline{K}$ of G such that

$$|A_k \cap T(K)| \gg |A|^{\frac{1}{4}-O(\epsilon)}, \tag{6.16}$$

where k and the implied constants depend only on $n = 3$, and are hence absolute. (Because $n = 3$ is fixed, all constants that would usually depend on n will be absolute.)

Let $c : T \cap \Sigma \to \mathbb{A}^1$ be as in Lemma 6.2. Then (6.12) implies that

$$\begin{pmatrix} c(t) \\ c(t^{-1}) \end{pmatrix} \cdot \begin{pmatrix} \mathrm{tr}(tg) \\ \mathrm{tr}((tg)^{-1}) \end{pmatrix} - \begin{pmatrix} c(t^{-1}) \\ c(t) \end{pmatrix} \cdot \begin{pmatrix} \mathrm{tr}(t^2g) \\ \mathrm{tr}((t^2g)^{-1}) \end{pmatrix} + \begin{pmatrix} \mathrm{tr}(t^3g) \\ \mathrm{tr}((t^3g)^{-1}) \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(g) \\ \mathrm{tr}(g^{-1}) \end{pmatrix}. \tag{6.17}$$

It is time to prepare ourselves to use Corollary 3.8. We first apply Cor. 5.16 to obtain a large subset $E' \subset A_k \cap T(K)$ (meaning a set $E' \subset A_k \cap T(K)$ with $|E'| \gg |A_k \cap T(K)| \gg$

$|A|^{\frac{1}{4}-O(\epsilon)}$, where the constants are absolute) satisfying the conclusion of Cor. 5.16. Let $R = (\mathbb{Z}/p\mathbb{Z})^2$, $X = \kappa(A_{k'} \cap \Sigma(K))$ (where we set k' equal to the value of k in Cor. 5.16 plus thrice the value of k in (6.16)),

$$Y = \left\{ \left(\begin{pmatrix} c(t) \\ c(t^{-1}) \end{pmatrix}, \begin{pmatrix} -c(t^{-1}) \\ -c(t) \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) : t \in E' \right\};$$

let $X_{\vec{y}}$ be the set of all tuples

$$(\kappa(tg), \kappa(t^2 g), \kappa(t^3 g))$$

with $g \in A_k$ satisfying $h_0 g \in \Sigma(K)$ (for some fixed $h_0 \in A_k$ given by Cor. 5.16) and $t^\ell g \in \Sigma(K)$ for $\ell = 0, 1, 2, 3$. (The conclusion of Cor. 5.16 was precisely that there are many such tuples.)

Having defined the sets to be used in our application of Cor. 3.8, we must now verify the assumptions of Cor. 3.8. (We already started to do so while defining the sets.) The projection $\pi_1 : (R^*)^3 \to R^*$ onto the first coordinate is clearly injective on Y: if we know $\begin{pmatrix} c(t) \\ c(t^{-1}) \end{pmatrix}$, we know $\begin{pmatrix} -c(t^{-1}) \\ -c(t) \end{pmatrix}$. By Corollaries 5.7 and 5.11,

$$|A|^{\frac{1}{4}} \ll |X| \ll |A|^{\frac{1}{4}+O(\epsilon)}, \tag{6.18}$$

where the implied constants are absolute. (In applying Cor. 5.11, we are assuming, as we may, that $|A|$ is larger than an absolute constant; otherwise the statement we seek to prove is trivial.)

We are assuming either $|A| \leq p^{4-\delta}$ or $p^{4+\delta} \leq |A| \leq p^{8-\delta}$. Hence, for ϵ small enough in terms of δ and p large enough in terms of δ, (6.18) implies that either

$$|X| \leq p^{1-\delta/2} \quad \text{or} \quad p^{1+\frac{\delta}{2}} \leq |X| \leq p^{2-\frac{\delta}{2}}.$$

Finally, for every $\vec{y} \in Y$ and every $\vec{x} \in X_{\vec{y}}$, (6.17) gives us that

$$\vec{y} \cdot \vec{x} = \begin{pmatrix} \mathrm{tr}(g) \\ \mathrm{tr}(g^{-1}) \end{pmatrix} \in \kappa(A_{k'}).$$

Because of the way we defined $X_{\vec{y}}$, the tuple $\vec{y} \cdot \vec{x}$ lies in $\kappa(\Sigma(K))$ as well.

Now we apply Corollary 3.8. It remains only to check that neither assertion in the conclusion (3.26) holds. We will then have obtained a contradiction. By Lemma 6.2, $|Y| \geq \frac{1}{6}|E'|$; by Corollary 5.10,

$$|E'| \gg |A|^{\frac{1}{4}-O(\epsilon)} \gg |X|^{1-O(\epsilon)},$$

where the implied constants are absolute. We conclude that the first assertion in (3.26) fails to hold for ϵ sufficiently small in terms of η.

Now, by Cor. 5.16, assuming that ϵ is less than an absolute constant ϵ_0, we have that, for every $t \in E'$, there are $\gg |A|$ distinct tuples

$$(\kappa(h_0 g), \kappa(tg), \kappa(t^2 g), \kappa(t^3 g))$$

with $g \in A_k$ satisfying $h_0 g \in \Sigma(K)$ and $t^\ell g \in \Sigma(K)$ for $\ell = 0, 1, 2, 3$. Now, by Cor. 5.11, the number of possible values taken by the first variable $\kappa(h_0 g)$ is at most $\ll |A|^{\frac{1}{n+1}+O(\epsilon)} =$

$|A|^{\frac{1}{4}+O(\epsilon)}$, where the implied constants are absolute. Thus, the number of elements of $X_{\vec{y}}$ – that is, the number of distinct tuples $(\kappa(tg), \kappa(t^2g), \kappa(t^3g))$ – is at least

$$\gg |A|^{\frac{3}{4}-O(\epsilon)} \gg |X|^{3-O(\epsilon)},$$

where the implied constants are absolute. Hence the second assertion in (3.26) fails to hold for ϵ sufficiently small in terms of η (and of n, which is a constant) and $|A|$ larger than a constant depending only on η.

By Cor. 3.8, η depends only on n and δ, and thus only on δ. We have it in the statement that we may assume that ϵ is smaller than a constant depending on δ. We may also assume that $|A|$ is larger than a constant depending on δ, as the implied constant in (6.15) may be taken to depend on δ. Hence we are done. □

7. Subgroups and solvable groups

We must examine how the existence of growth in subgroups of a group affects growth in the group itself. In particular, we want to have the tools that will allow us later to do induction on the group type by passing to subgroups.

We would also like to examine now how sets grow in solvable groups. (We already started to look into the issue in §3.2.) The growth of sets in a solvable group has a much more direct relationship to sum-product phenomena than the growth of sets that generate $\mathrm{SL}_2(K)$ or $\mathrm{SL}_3(K)$ does.

7.1. **Lemmas on growth and subgroups.** Let us start with two very simple lemmas.

Lemma 7.1. *Let G be a group and H a subgroup thereof. Let $A, B \subset G$ be finite sets. Then*

$$|A \cdot B| \geq r \cdot |B \cap H|,$$

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

We will usually apply this lemma with $A = B$.

Proof. Let $S \subset A$ be a set consisting of one coset representative $g \in A$ for every coset of H intersecting A. Since any two distinct cosets of a subgroup are disjoint, we have that (a) $|S| = r$, (b) all elements of the form $g \cdot h$ ($g \in S$, $h \in B \cap H$) are distinct. Thus there are $|S| \cdot |B \cap H| = r \cdot |B \cap H|$ of them. □

Lemma 7.2. *Let G be a group and H a subgroup thereof. Let $A \subset G$ be a non-empty finite set. Then*

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

where r is the number of cosets of H intersecting A. In particular,

$$|A^{-1}A \cap H| \geq \frac{|A|}{[G : H]}.$$

Proof. By the pigeonhole principle, there is at least one coset gH of H containing at least $|A|/r$ elements of A (and thus, in particular, at least one element of A). Choose an element $a_0 \in gH \cap A$. Then, for every $a \in gH \cap |A|$, the element $a_0^{-1}a$ lies both in H and in $A^{-1}A$. As a_0 is fixed and a varies, the elements $a_0^{-1}a$ are distinct. □

One of the reasons why we are interested in subgroups is that growth in subgroups H of G gives us growth in the group G.

Lemma 7.3. *Let G be a group and H a subgroup thereof. Let $A \subset G$ be a non-empty finite set. Then, for any $k > 0$,*

$$|A_{2k+1}| \geq \frac{|(A^{-1}A \cap H)_k|}{|A^{-1}A \cap H|}|A|.$$

Proof. Let r be the number of cosets of H intersecting A. It is clear that, for any $E \subset H$,

$$|A \cdot E| \geq r \cdot |E|.$$

In particular,

$$|A \cdot (A^{-1}A \cap H)_k| \geq r \cdot |(A^{-1}A \cap H)_k|$$

and the left side is evidently $\leq |A_{2k+1}|$. Now, by Lemma 7.2, $|A^{-1}A \cap H| \geq \frac{|A|}{r}$. Hence

$$|A_{2k+1}| \geq |A \cdot (A^{-1}A \cap H)_k| \geq r \cdot |(A^{-1}A \cap H)_k| \geq \frac{|(A^{-1}A \cap H)_k|}{|A^{-1}A \cap H|}|A|.$$

□

Growth in a quotient set also gives us growth in the group.

Lemma 7.4. *Let G be a group and H a subgroup thereof. Let G/H be the quotient set and $\pi : G \to G/H$ the quotient map. Then, for any finite non-empty subsets $A_1, A_2 \subset G$,*

$$|(A_1 \cup A_2)_4| \geq \frac{|\pi(A_1 A_2)|}{|\pi(A_1)|}|A_1|.$$

Actually, we will apply this lemma only for normal subgroups $H < G$, but it is true in general.

Proof. By Lemma 7.2,

$$|A_1^{-1}A_1 \cap H| \geq \frac{|A_1|}{\pi(A_1)}.$$

At the same time, it is clear that

$$|A_1 A_2 A_1^{-1} A_1| \geq |\pi(A_1 A_2)| \cdot |A_1^{-1}A_1 \cap H|.$$

Hence

$$|A_1 A_2 A_1^{-1} A_1| \geq \frac{|\pi(A_1 A_2)|}{|\pi(A_1)|}|A_1|.$$

□

Lemma 7.5. *Let G be a group and H a subgroup thereof. Let G/H be the quotient set and $\pi : G \to G/H$ the quotient map.*

Let $A \subset G$ be a finite set. Let A' be a subset of A. Then

$$|A' \cdot (A^{-1}A \cap H)| \geq \frac{|\pi(A')|}{|\pi(A)|}|A|.$$

Proof. By Lemma 7.2, $|A^{-1}A \cap H| \geq \frac{|A|}{|\pi(A)|}$. Since any distinct cosets of H are disjoint, it follows that

$$|A' \cdot (A^{-1}A \cap H)| \geq |\pi(A')| \cdot |A^{-1}A \cap H| \geq |\pi(A')| \cdot \frac{|A|}{|\pi(A)|}.$$

□

Let A be a finite subset of G and H a subset of G. By Lemma 7.2, either the intersection $A^{-1}A \cap H$ is large or there are many representatives in A of cosets of H. What we are about to show is that we can in effect remove the condition that H be a subgroup.

Lemma 7.6. *Let G be a group. Let $R \subset G$ be a subset with $R = R^{-1}$. Let $A \subset G$ be finite.*

Then there is a subset $A' \subset A$ with

$$|A'| \geq \frac{|A|}{|A^{-1}A \cap R|}$$

such that no element of $A'^{-1}A'$ (other than possibly the identity) lies in R.

Proof. Let $O = A^{-1}A \cap R$; since $R = R^{-1}$, we know that $O = O^{-1}$.

Let g_1 be an arbitrary element of A. If $A \subset g_1O$, let $A' = \{g_1\}$ and stop. Otherwise, let g_2 be in A but not in g_1O. If $A \subset g_1O \cup g_2O$, let $A' = \{g_1, g_2\}$ and stop. Otherwise, let g_3 be in A but not in $g_1O \cup g_2O$, etc. We eventually arrive at a covering $A \subset g_1O \cup g_2O \cup \cdots \cup g_\ell O$ such that $g_j \notin g_iO$ for all pairs (i,j), $1 \leq i < j \leq \ell$. As $O = O^{-1}$, it follows that we also have $g_i \notin g_jO$. Since $O = A^{-1}A \cap R$, this implies that $g_i^{-1}g_j \notin R$ for all $1 \leq i, j \leq \ell$, $i \neq j$.

Let $A' = \{g_1, g_2, \ldots, g_\ell\}$. What we have just shown can be restated as follows: no element of $A'^{-1}A' \setminus \{e\}$ lies in R.

Now, because $A \subset g_1O \cup g_2O \cup \cdots \cup g_\ell O$, there is a $g_i \in A' \subset A$ such that $|A \cap g_iO| \geq \frac{|A|}{\ell}$ (by the pigeonhole principle). Hence $|O| = |g_iO| \geq \frac{|A|}{\ell}$. By the definition of O, we conclude that $|A^{-1}A \cap R| \geq \frac{|A|}{\ell}$. Since $\ell = |A'|$, we obtain that $|A'| \geq \frac{|A|}{|A^{-1}A \cap R|}$. □

7.2. **Lemmas for solvable groups.** We will state the following lemmas in general, but they are especially useful for solvable groups G. We write $G^{(1)} := [G,G] = \{xyx^{-1}y^{-1} : x, y \in G\}$.

Lemma 7.7. *Let G be a group. Let $G^{(1)} = [G,G]$. Let $A \subset G$ be a finite set.*

Then, for every $\delta > 0$, either

(a) *$|AAA^{-1}| \geq |A|^{1+\delta}$, or*

(b) *there is a $g \in A$ such that*

$$|C_G(g) \cap A^{-1}A| \cdot |G^{(1)} \cap A^{-1}A| \geq |A|^{1-\delta}. \tag{7.1}$$

Note moreover that $C_G(g) \cap G^{(1)} = \{e\}$.

Proof. By Proposition 5.8, there is a $g \in A$ such that the set $C_G(g) \cap A^{-1}A$ has

$$\frac{|A|}{|AAA^{-1}|} \cdot |\operatorname{Cl}_G(A)|$$

elements.

By the pigeonhole principle, there is a conjugacy class C in G containing $\geq \frac{|A|}{|\mathrm{Cl}_G(A)|}$ elements of A. For any two $g_1, g_2 \in C$, the quotient $g_1^{-1} g_2$ lies in $G^{(1)}$: there is an $h \in G$ such that $g_2 = h g_1 h^{-1}$, and so

$$g_1^{-1} g_2 = g_1^{-1} h g_1 h^{-1} \in G^{(1)}.$$

Fixing $g_1 \in C$ and letting g_2 vary within C, we obtain that there are at least $|C|$ distinct elements in $A^{-1}A \cap G^{(1)}$.

Therefore

$$\begin{aligned}|C_G(g) \cap A^{-1}A| \cdot |G^{(1)} \cap A^{-1}A| &\geq |C_G(g) \cap A^{-1}A| \cdot |G^{(1)} \cap A^{-1}A| \\ &\geq |C_G(g) \cap A^{-1}A| \cdot |C| \\ &\geq \frac{|A|}{|AAA^{-1}|} |\mathrm{Cl}_G(A)| \cdot \frac{|A|}{|\mathrm{Cl}_G(A)|} \geq \frac{|A|}{|AAA^{-1}|} \cdot |A|.\end{aligned}$$

If $|AAA^{-1}| \geq |A|^{1+\delta}$, we have conclusion (a). Otherwise,

$$|C_G(g) \cap A^{-1}A| \cdot |G^{(1)} \cap A^{-1}A| \geq |A|^{1-\delta},$$

i.e., conclusion (b). □

Lemma 7.8. *Let G be a group. Let $H_1, \ldots, H_m < G$ be proper subgroups such that, if $g \in G$ does not lie in any H_j, $1 \leq j \leq m$, then $gxg^{-1} \neq x$ for every $x \in G^{(1)} = [G, G]$.*

Let $A \subset G$ be finite. Then, for every $\delta > 0$, either

(a) $|AAA^{-1}| \gg |A|^{1+\delta}$, *where the implied constant is absolute,*
(b) $|A_6 \cap (H_j \cdot G^{(1)})| \geq \frac{1}{2m} |A|^{1-2\delta}$ *for some $1 \leq j \leq m$, or*
(c) *there is a subset $Y \subset A^{-1}A$ with $|Y| \geq |A|^{\delta}$ such that*

$$gxg^{-1} \neq x$$

for every $x \in G^{(1)} \setminus \{e\}$ and every $g \in Y^{-1}Y \setminus \{e\}$.

Proof. If $|A \cap (H_1 \cup \ldots \cup H_m)| > \frac{1}{2}|A|$, we arrive at (a stronger version of) conclusion (b). Assume otherwise. Let $A' = A \setminus (A \cap (H_1 \cup \ldots \cup H_m))$. Apply Lemma 7.7 with A' instead of A. Case (a) of Lemma 7.7 gives us conclusion (a) here. Assume, then, that we are in case (b) of Lemma 7.7.

Apply Lemma 7.6 with $R = H_1 \cup H_2 \cup \ldots \cup H_m$ and $C_G(g) \cap A'^{-1}A'$ instead of A. We obtain a subset $Y \subset C_G(g) \cap A'^{-1}A'$ with

$$|Y| \geq \frac{|C_G(g) \cap A'^{-1}A'|}{|(C_G(g) \cap A'^{-1}A')^{-1}(C_G(g) \cap A'^{-1}A') \cap R|} \geq \frac{|C_G(g) \cap A'^{-1}A'|}{|(C_G(g) \cap A'_4) \cap R|}.$$

If $|Y| \geq |A|^{\delta}$, we have obtained conclusion (c). Assume $|Y| < |A|^{\delta}$. Then

$$|(C_G(g) \cap A'_4) \cap R| \geq |A|^{-\delta} \cdot |C_G(g) \cap A'^{-1}A'|,$$

and so

$$|(C_G(g) \cap A'_4) \cap H_j| \geq \frac{1}{m} |A|^{-\delta} \cdot |C_G(g) \cap A'^{-1}A'|$$

for some $1 \leq j \leq m$. By the fact that $C_G(g) \cap G^{(1)} = \{e\}$, it follows that

$$\begin{aligned}|A_6 \cap (H_j \cdot G^{(1)})| &\geq |(C_G(g) \cap A'_4) \cap H_j| \cdot |G^{(1)} \cap A'^{-1}A'| \\ &\geq \frac{1}{m}|A|^{-\delta} \cdot |C_G(g) \cap A'^{-1}A'| \cdot |G^{(1)} \cap A'^{-1}A'| \\ &\geq \frac{1}{m}|A|^{-\delta} \cdot |A'|^{1-\delta} \geq \frac{1}{2m}|A|^{1-2\delta},\end{aligned}$$

where we use (7.1). We have obtained conclusion (b). □

It is now that our generalised sum-product techniques come in.

Lemma 7.9. *Let G be a group. Assume that there is no chain of subgroups*

$$\{e\} \lneq G_1 \lneq G_2 \lneq \cdots \lneq G_r \lneq G^{(1)} \tag{7.2}$$

with $r \geq \ell$, where ℓ is an integer.

Let $A \subset G$ be finite. Suppose that there is a subset $Y \subset A$, $|Y| \geq |A|^{\delta}$, $\delta > 0$, such that

$$gxg^{-1} \neq x$$

for every $x \in G^{(1)} \setminus \{e\}$ and every $g \in Y^{-1}Y \setminus \{e\}$.

Then either

(a) $|A_k| \geq |A|^{1+\delta}$*, where k depends only on δ and ℓ, or*

(b) *there is a subgroup $X < G^{(1)} \cap \langle A\rangle$ such that (a) $X \triangleleft \langle A\rangle$, (b) $\langle A\rangle/X$ is abelian, (c) A_k contains X for some k depending only on δ and ℓ.*

The condition on the non-existence of long chains (7.2) can probably be relaxed; it will have to be if results uniform over α on algebraic groups over $\mathbb{F}_{p^\alpha}$ are to be obtained. (We will not attempt to do as much in this paper.)

Proof. If $\langle A\rangle$ is abelian, conclusion (b) holds with $X = \{e\}$. Assume otherwise. Then there are two elements g_1, g_2 of A (and not just two elements of $\langle A\rangle$) that do no not commute with each other. Hence $g_1 g_2 g_1^{-1} g_2^{-1} \neq e$, and so $A_4 \cap G^{(1)} \neq \{e\}$.

We now apply our generalised sum-product statement, Corollary 3.2, with Y acting on $S = A_4 \cap G^{(1)}$ by conjugation. We obtain that

$$\begin{aligned}|(Y_2(S))_6| &> \frac{1}{2}\min(|Y| \cdot |S|, |G_1|) \geq \frac{1}{2}\min(2|A|^{\delta} \cdot |S|, |G_1|) \\ &= \min(|A|^{\delta}|S|, \frac{1}{2}|G_1|),\end{aligned}$$

where $G_1 = \langle\langle Y\rangle(\langle S\rangle)\rangle$ is a subgroup of $G^{(1)}$. Since $S = \{e\}$, the group G_1 is not just $\{e\}$.

We apply Corollary 3.2 again and again - a total of $r = \lceil\frac{1}{\delta}\rceil + 1$ times – and obtain that

$$|(Y_{2r}(S))_{6^r}| > \min(|A|^{1+\delta} \cdot |S|, \frac{1}{2}|G_1|).$$

If $\min(|A|^{1+\delta} \cdot |S|, \frac{1}{2}|G_1|) = |A|^{1+\delta} \cdot |S|$, we have reached conclusion (a). Assume, then, that $\min(|A|^{1+\delta} \cdot |S|, \frac{1}{2}|G_1|) = \frac{1}{2}|G_1|$. By Lemma 2.1, it follows that

$$|(Y_{2r}(S))_{2\cdot 6^r}| = G_1.$$

Since

$$(Y_{2r}(S))_{2\cdot 6^r} \subset (A_{4r} \cdot A_2 \cdot A_{4r})_{2\cdot 6^r} \subset A_{2(8r+2)\cdot 6^r},$$

we have shown that $G_1 \subset A_k$, where $k = 2(8r+2)\cdot 6^r$ depends only on δ.

If G_1 is a normal subgroup of $\langle A\rangle$ and G/G_1 is abelian, we have obtained conclusion (b) (with $X = G_1$) and are done. Assume otherwise. If G_1 is not a normal subgroup of $\langle A\rangle$, there is necessarily a g in A itself (as opposed to just in $\langle A\rangle$) and an $h \in G_1$ such that such that $ghg^{-1} \notin G_1$. If G_1 is a normal subgroup but $\langle A\rangle/G_1$ is not abelian, there are two elements of A (and not just two elements of $\langle A\rangle$) that do not commute $\bmod G_1$, i.e., two elements $g_1, g_2 \in A$ such that $g_1 g_2 g_1^{-1} g_2^{-1} \notin G_1$. It is easy to see that, in the former case, ghg^{-1} is in $G^{(1)}$; in the latter case, $g_1 g_2 g_1^{-1} g_2^{-1}$ is in $G^{(1)}$. At any rate, there is an element g of A_4 such that $g \in G^{(1)} \setminus G_1$.

Now we apply Cor. 3.2 again and again to the set $H_1 \cup \{g\}$. After applying it a total of $r = \lceil \frac{1}{\delta} \rceil + 1$ times – say – we obtain

$$|(Y_{2r}(G_1 \cup \{g\}))_{6^r}| > \min\left(|A|^{1+\delta} \cdot |G_1 \cup \{g\}|, \frac{1}{2}|G_2|\right).$$

If $\min\left(|A|^{1+\delta} \cdot |G_1 \cup \{g\}|, \frac{1}{2}|G_2|\right) = |A|^{1+\delta} \cdot |G_1 \cup \{g\}|$, we have reached conclusion (a). Suppose, then, that $\min\left(|A|^{1+\delta} \cdot |G_1 \cup \{g\}|, \frac{1}{2}|G_2|\right) = \frac{1}{2}|G_2|$; by Lemma 2.1, it follows that

$$(Y_{2r}(G_1 \cup \{g\}))_{2\cdot 6^r} = G_2,$$

and so $G_2 \subset A_{k'}$, k' depending only on δ.

If G_2 is a normal subgroup of $\langle A\rangle$ stable under the action of J and $\langle A\rangle/G_2$ is abelian, we have obtained conclusion (b) and are done. Otherwise, we proceed as before, constructing an element g of $A_{k''} \cap G^{(1)}$ not in G_2, and applying Cor. 3.2 again and again to $G_2 \cup \{g\}$, then to $G_3 \cup \{g\}$, and so on. As there cannot be a chain

$$\{e\} \subsetneq G_1 \subsetneq G_2 \subsetneq \cdots \subsetneq G_r \subsetneq U(K)$$

with $r \geq \ell$, we reach conclusion (b) in at most ℓ steps, if we do not reach conclusion (a) first. □

Corollary 7.10 (to Lemmas 7.8 and 7.9)**.** *Let G be a group. Let $H_1, \ldots, H_m < G$ be proper subgroups such that, if $g \in G$ does not lie in any H_j, $1 \leq j \leq m$, then $gxg^{-1} \neq x$ for every $x \in G^{(1)} = [G,G]$. Assume that there is no chain of subgroups*

$$\{e\} \subsetneq G_1 \subsetneq G_2 \subsetneq \cdots \subsetneq G_r \subsetneq G^{(1)} \tag{7.3}$$

with $r \geq \ell$, where ℓ is an integer.

Let $A \subset G$ be finite. Then, for every $\delta > 0$, either

(a) *$|A_k| \gg |A|^{1+\delta}$, where the implied constant is absolute and k depends only on δ and ℓ,*
(b) *$|A_6 \cap (H_j \cdot G^{(1)})| \geq \frac{1}{2m}|A|^{1-2\delta}$ for some $1 \leq j \leq m$; moreover, A is contained in the union of at most $|A|^{3\delta}$ cosets of $H_j \cdot G^{(1)}$ for that same index j;*
(c) *there is a subgroup $X < G^{(1)} \cap \langle A\rangle$ such that (a) $X \triangleleft \langle A\rangle$, (b) $\langle A\rangle/X$ is abelian, (c) A_k contains X for some k depending only on δ and ℓ.*

Proof. Apply Lemma 7.8. If conclusion (c) of Lemma 7.8 holds, apply Lemma 7.9 with $A^{-1}A$ instead of A. (The comment in conclusion (b) on how A is contained in the union of few cosets of $H_j \cdot G^{(1)}$ follows from Lemma 7.1: if there were too many cosets intersecting A_6, conclusion (a) would follow.) $\square$

7.3. **Examples: growth in Borel subgroups of** $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ **and** $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$**.** We with to study the growth of sets in solvable subgroups of $\mathrm{SL}_2(K)$ and $\mathrm{SL}_3(K)$, where $K = \mathbb{Z}/p\mathbb{Z}$. The main case of interest is that of Borel subgroups B/K.

Proposition 7.11. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let B/K be a Borel subgroup of SL_2/K. Let U/K be the maximal unipotent subgroup of B/K.*

Let $A \subset B(K)$. Then, for every $\delta > 0$ smaller than an absolute constant, either

(a) *$|A_k| \gg |A|^{1+\delta}$, where the implied constant is absolute and k depends only on δ,*
(b) *$|A_6 \cap (\pm\{I\} \cdot U(K))| \geq \frac{1}{2}|A|^{1-2\delta}$; moreover, A is contained in the union of at most $|A|^{3\delta}$ cosets of $U(K)$;*
(c) *A is contained in some maximal torus $T/\overline{K}$,*
(d) *A_k contains $U(K)$ for some k depending only on δ.*

Proof. We apply Corollary 7.10 with $G = B(K)$, $H_1 = \{\pm I\} \cdot U(K)$, $m = 1$, $\ell = 1$. (Here $\ell = 1$ because $U(K)$ has no proper subgroups.) Cases (a) and (b) of Cor. 7.10 give us conclusions (a) and (b). Assume, then, that we are in case (c) of Cor. 7.10. If $X = \{e\}$, then either conclusion (c) holds or A is contained in $\pm\{I\} \cdot U(K)$; in the latter case, conclusion (b) holds. If $X = U(K)$, then conclusion (d) holds. $\square$

Let us now examine what happens in Borel subgroups of SL_3.

Proposition 7.12. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let B/K be a Borel subgroup of SL_2/K. Let U/K be the maximal unipotent subgroup of B/K.*

Then, for every $\epsilon > 0$, one of the following conclusions holds:

(a) *$|A_k| \gg |A|^{1+\delta}$, where the implied constant is absolute and k depends only on δ;*
(b) *$|A_k \cap U(K)| \gg |A|^{1-\epsilon}$, where k and the implied constant are absolute; moreover, A is contained in the union of at most $|A|^{2\epsilon}$ cosets of $U(K)$;*
(c) *there are subgroups $X < U(K) \cap \langle A \rangle$ and $Y < \langle A \rangle$ such that (a) $X \triangleleft Y$, (b) Y/X is abelian, (c) A_k contains X for some k depending only on ϵ, (d) A is contained in the union $\leq |A|^{\epsilon}$ cosets of Y.*

Proof. Let $\delta = \epsilon/3$. Let $\rho_{i,j} : B(K) \to K^*$ taking an element of $B(K)$ with diagonal entries r_1, r_2, and r_3 to $r_i r_j^{-1} \in K^*$. (In other words, $\rho_{i,j}$ is a root map.) Apply Cor. 7.10 with $G = B(K)$ and H_1, H_2, H_3 equal to the kernels of $\rho_{1,2}$, $\rho_{2,3}$ and $\rho_{1,3}$, respectively. (Condition (7.2) holds with $\ell = 3$.) Cases (a) and (c) in Cor. 7.10 give us conclusions (a) and (c) here. Assume, then, that we are in case (b) of Cor. 7.10 for some $j = 1, 2, 3$. Let $H = H_j$. From the definition of our H_j, we have $H \cdot G^{(1)} = H$ for every $j = 1, 2, 3$, and thus $|A_6 \cap H| \geq \frac{1}{6}|A|^{1-2\delta}$. We set $Y = \langle A_6 \cap H \rangle$; by Lemma 7.1 (applied with $B = A_6 \cap H$ and Y instead of H), either A is contained in the union of at most $|A|^{3\delta}$ cosets of Y, or conclusion (a) holds. Let us assume conclusion (a) does not hold.

Case 1: $H = \ker(\rho_{1,2})$ or $H = \ker(\rho_{2,3})$. Apply Cor. 7.10 once again, this time with $G = H$, $m = 1$, $H_1 = U(K)$ and $A' = A_6 \cap H$ instead of A. Cases (a), (b) and (c) in Cor. 7.10 give us conclusions (a), (b) and (c) here.

Case 2: $H = \ker(\rho_{1,3})$. Let $A' = A_6 \cap H$. Write $G^{(2)} = [G^{(1)}, G^{(1)}]$. If $g \in H$ is not contained in $U(K)$, then g acts without fixed points by conjugation on $U(K)/G^{(2)}$. Apply Cor. 7.10 with $G = H/G^{(2)}$, $m = 1$, $H_1 = U(K)/G^{(2)}$ and $A'' = \{h \cdot G^{(2)} : h \in (A_6 \cap H)\}$ instead of A. If case (a) of Cor. 7.10 holds, then Lemma 7.4 gives us conclusion (a), unless A'' is much smaller than A' ($|A''| < |A'|^\delta$), in which case $A'^{-1}A' \cap G^{(2)}$ must be very large ($\gg |A|^{1-2\delta}$), giving us conclusion (b). Case (b) gives us conclusion (b). It remains to examine case (c) of Cor. 7.10.

Suppose first that $X = U(K)/G^{(2)}$. A quick calculation suffices to show that, for any set $C \subset U(K)$ such that $\{cG^{(2)} : c \in C\}$ is all of $U(K)/G^{(2)}$, the set of commutators $[C, C]$ is all of $G^{(2)}$, and thus $C_5 = U(K)$. We conclude that A_k contains $U(K)$ for some k depending only on δ; we have obtained conclusion (b).

Suppose now that $X = \{e\}$. Then $\langle A'\rangle/G^{(2)}$ is abelian, and, since $G^{(2)}$ lies in the centre of $H = H_{1,3}$, the group $\langle A'\rangle$ must itself be abelian.

Suppose, lastly, that $X \neq U(K)/G^{(2)}$ and $X \neq \{e\}$. We know that $\langle A'\rangle/H'$ is abelian. This implies that either

$$X = \left\{ \begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot G^{(2)} : a \in \mathbb{Z}/p\mathbb{Z} \right\}$$

and all elements of A' are contained in the group

$$R = \left\{ \begin{pmatrix} r & a & b \\ 0 & r^{-2} & 0 \\ 0 & 0 & r \end{pmatrix} : r \in (\mathbb{Z}/p\mathbb{Z})^*, a, b \in (\mathbb{Z}/p\mathbb{Z}) \right\},$$

or

$$X = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix} \cdot G^{(2)} : a \in \mathbb{Z}/p\mathbb{Z} \right\}$$

and all elements of A' are contained in the group

$$R = \left\{ \begin{pmatrix} r & 0 & b \\ 0 & r^{-2} & a \\ 0 & 0 & r \end{pmatrix} : r \in (\mathbb{Z}/p\mathbb{Z})^*, a, b \in (\mathbb{Z}/p\mathbb{Z}) \right\}.$$

We apply Cor. 7.10 with $G = R$, $m = 1$, $H_1 = U(K) \cap R$ and A' instead of A. (We can do this because all elements of G not in H_1 act on $G^{(1)}$ without fixed points: $G^{(1)}$ is now smaller than it was when G was $B(K)$ or H.) Cases (a), (b) and (c) of Cor. 7.10 give us conclusions (a), (b) and (c) here. $\square$

8. Growth in proper subgroups of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$

Let $K = \mathbb{Z}/p\mathbb{Z}$ and $G = \mathrm{SL}_3$. Suppose $A \subset G(K)$ does not generate $G(K)$. Then A generates a proper subgroup $\langle A\rangle$ of G. Does A grow? That is: does $|A \cdot A \cdot A| > |A|^{1+\delta}$ hold?

The answer depends on which subgroup of G the group $\langle A\rangle$ happens to be. The subgroups of $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ are not particularly hard to classify.

Proposition 8.1 (Mitchell [Mi])**.** *Let $G = \mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$, p odd. The maximal subgroups of G are*

(a) *the stabiliser of a point in $\mathbb{P}^3(\mathbb{Z}/p\mathbb{Z})$,*
(b) *the stabiliser of a line in $\mathbb{P}^3$ defined over $\mathbb{Z}/p\mathbb{Z}$,*
(c) *the stabiliser of a set of three points in $\mathbb{P}^3(\overline{\mathbb{Z}/p\mathbb{Z}})$,*
(d) *the stabiliser of a conic in $\mathbb{P}^3(\mathbb{Z}/p\mathbb{Z})$,*
(e) *groups of order ≤ 360.*

Proof. This is Theorem 2.4 for q prime in the survey paper [Ki]. Cases (a) and (b) in [Ki, Thm. 2.4] correspond to cases (a) and (b) here; cases (c) and (d) correspond to case (c) here; case (e) is (d) here; cases (f)–(i) do not happen; finally, cases (j) and (k) in [Ki, Thm. 2.4] go into case (e) here. □

From this, we get the following classification.

Corollary 8.2. *Let $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$, p odd. Let H be a proper subgroup of G. Then at least one of the following statements holds:*

(a) *H is contained in the stabiliser of a point in $\mathbb{P}^3(\mathbb{Z}/p\mathbb{Z})$,*
(b) *H is contained in the stabiliser of a line in $\mathbb{P}^3$ defined over $\mathbb{Z}/p\mathbb{Z}$,*
(c) *H has an abelian subgroup of index ≤ 6,*
(d) *H is contained in a subgroup of G isomorphic to $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z})$,*
(e) *H is of order ≤ 360.*

Proof. Let M be a maximal subgroup of $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ containing H. Let $\overline{M}$ be the image of M under the natural map $\pi : \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z}) \to \mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$. If $\overline{M}$ were not a proper subgroup of $\mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$, then M would have index 3 in G. The action of G on cosets of M would induce a non-trivial homomorphism ϕ from G to the symmetric group S_3. The kernel $\ker(\phi)$ of that homomorphism would be a proper normal subgroup of G of index at most 6. Now, $G/Z(G) = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})/Z(\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z}))$ is simple, and so $\ker(\phi)$ would have to be contained in $Z(G)$. Since $Z(G)$ has at most 3 elements, it would follow that G has at most $6 \cdot 3 = 18$ elements. This is clearly false. Thus, $\overline{M}$ is a proper subgroup of $\mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$.

Moreover, $\overline{M}$ is a maximal subgroup of $\mathrm{PSL}_3(\mathbb{Z}/p\mathbb{Z})$, as otherwise M would not be maximal in $G = SL_3(\mathbb{Z}/p\mathbb{Z})$. Now apply Prop. 8.1.

If $\overline{M}$ is the stabiliser of a line, then M is contained in the stabiliser in $G = SL_3(\mathbb{Z}/p\mathbb{Z})$ of a line. (The action of G on $\mathbb{P}^3$ factors through $PSL_3(\mathbb{Z}/p\mathbb{Z})$.) If $\overline{M}$ is the stabiliser of a point, then M is contained in the stabiliser of a point. This takes care of cases (a) and (b) of Prop. 8.1.

Suppose now that we are in case (c) of Prop. 8.1. Since $\overline{M}$ is the stabiliser of a set of three points, M is contained in the stabiliser of a set of three points. The stabiliser in G of a set of three points in $\mathbb{P}^3(\overline{\mathbb{Z}/p\mathbb{Z}})$ is equal to the semidirect product of the points over $\mathbb{Z}/p\mathbb{Z}$ of a torus T in G (defined over $\overline{\mathbb{Z}/p\mathbb{Z}}$) and the elements of G that induce elements of the Weyl group of the torus. Since the Weyl group of a torus in SL_3 has index 6, we see that the group M must have an abelian subgroup of index ≤ 6, and thus H itself has an abelian subgroup of index ≤ 6. We have obtained conclusion (c).

Suppose that we are in case (d) of Prop. 8.1. The conic in question is given by an equation $Q(v) = 0$, where Q is some non-degenerate quadratic form. The group G_Q of all elements $g \in G$ such that $Q(gv) = Q(v)$ is isomorphic to $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z})$ ([KL, Prop. 2.5.4]). The group G_Q is certainly contained in the stabiliser of $Q(v) = 0$. Comparing orders (where the order of the stabiliser of a conic $Q(v) = 0$ is given by [Ki, Thm. 2.4]) we see that G_Q is actually equal to the stabiliser of $Q(v) = 0$.

Finally, case (e) of Prop. 8.1 corresponds to case (e) here, and so we are done. □

Let us see what we can say about each of the cases of Cor. 8.2.

For groups of bounded order, the statement $|A \cdot A \cdot A| \gg |A|^{1+\delta}$ is trivially true (as one may adjust δ and the implied constant if needed). Thus, we may ignore case (e). As for case (a), it reduces to case (b): the stabiliser in G of a point in $\mathbb{P}^3(\mathbb{Z}/p\mathbb{Z})$ is always conjugate (and hence isomorphic) to the subgroup

$$\left\{ g = \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix} : \det(g) = 1 \right\} \tag{8.1}$$

of G, whereas the stabiliser in G of a line in $\mathbb{P}^3$ defined over $\mathbb{Z}/p\mathbb{Z}$ is always conjugate (and hence isomorphic) to the subgroup

$$\left\{ g = \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & * \end{pmatrix} : \det(g) = 1 \right\} \tag{8.2}$$

of G. The subgroups (8.1) and (8.2) are isomorphic as groups. (They and their conjugates are called the *maximal parabolic subgroups* of G.) Thus, case (a) and case (b) are essentially the same.

We hence have three cases to study: (1) subgroups of $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z})$ (case (d) in Cor. 8.2); (2) subgroups of G having abelian subgroups of small index (case (c) in Cor. 8.2); (3) subgroups of maximal parabolic subgroups of $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ (that is, subgroups of stabilisers of points and lines, i.e., cases (a) and (b) in Cor. 8.2). Let us consider them in order.

(1) *The group* $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z}) \sim \mathrm{PGL}_2(\mathbb{Z}/p\mathbb{Z})$.
As it happens, $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z})$ is isomorphic as a group to $\mathrm{PGL}_2(\mathbb{Z}/p\mathbb{Z})$ ([Ta, Thm. 11.6]). We will conclude our study of growth in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, and then use the fact that $\mathrm{PGL}_2(\mathbb{Z}/p\mathbb{Z})$ has a subgroup of index 2 isomorphic to $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})/Z(\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}))$.

(2) *Subgroups of* $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ *having abelian subgroups of small index.*

This is a different kettle of fish. Some subsets of abelian groups grow and others do not. (This matter is the classical object of study of additive combinatorics.) A great deal has been said on this general subject, but very little is known on the question of which subsets of abelian groups grow truly rapidly ($|A \cdot A \cdot A| \gg |A|^{1+\delta}$). All we know is which sets grow very slowly ($|A \cdot A \cdot A| \ll (\log |A|)^{1/3}|A|$, say); this is Freiman's theorem, generalised to arbitrary abelian groups by Green and Ruzsa [GR]).

We will not attempt to improve on this; we will do no more than set aside the abelian case whenever we come across it.

(3) *Subgroups of maximal parabolic subgroups of $G = SL_3(\mathbb{Z}/p\mathbb{Z})$.*

These are the groups isomorphic to (8.1) and (8.2). They are the main subject of this section (§8.2 – §8.5). A subset A of a maximal parabolic subgroup of G may or may not be contained in a Borel subgroup of G. Growth in Borel subgroups is closely related to Prop. 3.1, i.e., to generalised sum-product phenomena. If a subset A of a parabolic subgroup is not contained in a Borel subgroup, the study of its growth amounts more or less to the study of growth in SL_2 plus a little cohomology.

8.1. **Growth in subgroups of** $\mathrm{SL}_2(\mathbb{F}_p)$ **and** $\mathrm{PGL}_2(\mathbb{F}_p) \sim SO_3(\mathbb{F}_p)$**.** The classification of the proper subgroups of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ is classical.

Proposition 8.3. *Let $K = \mathbb{Z}/p\mathbb{Z}$. Let $G = \mathrm{SL}_2(K)$. Let H be a proper subgroup of G with more than 120 elements. Then either*

(a) *H is contained in a Borel subgroup B of G defined over K, or*

(b) *there is a maximal torus $T/\overline{K}$ such that either*

  (i) *H is contained in $T(K)$, or*

  (ii) *H is contained in a subgroup $H' < G(K)$ with $T(K) < H'$, $[H' : T(K)] = 2$.*

Proof. See [Di], p. 286. □

We now need to do very little work given what we already did in §7.3.

Theorem 8.4. *Let $G = \mathrm{SL}_2$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$.*

Then, for every $\epsilon > 0$, either

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{8.3}$$

where $\delta > 0$ and the implied constant depend only on ϵ, or one of the following cases holds:

(a) *A generates $G(K)$ and $|A| > |G(K)|^{1-\epsilon}$, or*

(b) *there is a maximal torus $T/\overline{K}$ such that either*

  (i) *A is contained in $T(K)$, or*

  (ii) *A is contained in a subgroup $H < G(K)$ with $T(K) < H$, $[H : T(K)] = 2$;*

(c) *there is a Borel subgroup B/K such that $A \subset B(K)$, such that either*

  (i) *$|A_6 \cap (\{\pm I\} \cdot U(K))| \geq |A|^{1-\epsilon}$ (where U/K is the maximal unipotent subgroup of B) and A intersects at most $|A|^{2\epsilon}$ cosets of $U(K)$, or*

  (ii) *A_k contains $U(K)$ for some k depending only on ϵ.*

Proof. If A generates $G(K)$, then, by Proposition 6.1, either (8.3) or conclusion (a) holds. Assume, then, that A does not generate $G(K)$.

Thanks to the classification of the proper subgroups of $G(K)$ (Prop. 8.3), either conclusion (b) holds or A is contained in $B(K)$, where B/K is a Borel subgroup of G. In the latter case, we apply Prop. 7.11. If case (d) in Prop. 7.11 holds, then (8.3) follows by the tripling lemma (Lem. 2.2).

(We use the fact that we can assume that $|A|$ is larger than an absolute constant, as otherwise (8.3) holds trivially; this allows us, for example, to do without a factor of $\frac{1}{4}$ in front of $|A|^{1-\epsilon}$ when deriving conclusion c(i) from case (b) of Prop. 7.11.) □

One may ask how tight Thm. 8.4 is. There are examples of sets A falling into one of the cases b(i), b(ii), c(i), c(ii) in Thm. 8.4 and failing to grow (i.e., failing to satisfy (8.3)).

To wit –

Case b(i): Let
$$A = \left\{ \begin{pmatrix} x^n & 0 \\ 0 & x^{-n} \end{pmatrix} : 1 \leq n \leq N \right\},$$
where x is a generator of $(\mathbb{Z}/p\mathbb{Z})^*$ and $N \leq p-1$. Then $|A| = N$ and $|A \cdot A \cdot A| < 3N = 3|A|$.

Case b(ii): Let
$$A = \left\{ \begin{pmatrix} x^n & 0 \\ 0 & x^{-n} \end{pmatrix} : -N \leq n \leq N \right\} \cup \left\{ \begin{pmatrix} 0 & x^n \\ x^{-n} & 0 \end{pmatrix} : -N \leq n \leq N \right\},$$
where x is a generator of $(\mathbb{Z}/p\mathbb{Z})^*$ and $N \leq (p-1)/2$. Then $|A| = 4N+2$ and $|A \cdot A \cdot A| < 2 \cdot (6N+1) < 3|A|$.

Case c(i): Let
$$A = \left\{ \begin{pmatrix} n & m \\ 0 & n^{-1} \end{pmatrix} : 1 \leq n \leq N^\epsilon,\ 1 \leq m \leq N \right\}.$$
(Here n^{-1} stands for inverse of $n \bmod p$.) Then $|A| \sim N^{1+\epsilon}$ and $|A \cdot A \cdot A| \ll N^{1+9\epsilon}$.

Case c(ii): Let
$$A = \left\{ \begin{pmatrix} x^n & m \\ 0 & x^{-n} \end{pmatrix} : 1 \leq n \leq N,\ m \in \mathbb{Z}/p\mathbb{Z} \right\},$$
where x is a generator of $(\mathbb{Z}/p\mathbb{Z})^*$ and $N \leq p-1$. Then $|A| = pN$ and $|A \cdot A \cdot A| < 3pN = 3|A|$.

We can rewrite the conclusions of Thm. 8.4 so that it looks more like what a general statement on all groups would be likely to look like. (See the remarks after Thm. 1.1.)

Corollary 8.5 (to Theorem 8.4)**.** *Let $G = \mathrm{SL}_2$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$. Then, for every $\epsilon > 0$, either*

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{8.4}$$

where $\delta > 0$ and the implied constant depend only on ϵ, or there are normal subgroups $H_1, H_2 \lhd \langle A \rangle$, $H_1 < H_2$ such that

(a) *H_2/H_1 is abelian,*
(b) *A_k contains H_1, where k depends only on ϵ, and*
(c) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2.*

In other groups,“abelian” would be replaced by “nilpotent” (as in the statement of Theorem 1.1). We have “abelian” here simply because there is not much room for non-abelian nilpotent groups in SL_2.

Proof. Apply Thm. 8.4 (with $\epsilon/2$ instead of ϵ). Equation (8.3) in Thm. 8.4 is equation (8.4) here. By the Key Proposition (part (b)) in [He, §1], case (a) in Thm. 8.4 implies that A_k contains H_1 and is contained in H_2, where $H_1 = H_2 = G(K)$. Case (b) in Thm. 8.4 gives us that A is contained in the union of ≤ 2 cosets of $H_2 = T(K) \lhd \langle A \rangle$; since H_2

is abelian, we can set $H_1 = \{e\}$. Case c(i) in Thm. 8.4 gives us that A is contained in few subsets of the abelian group $H_2 = (\{\pm I\} \cdot U(K)) \lhd B(K)$; again, we set $H_1 = \{e\}$. Finally, case c(ii) tells us that A contains $H_1 = U(K)$ and is contained in $H_2 = B(K)$; H_1 is a normal subgroup of H_2, H_2 is a normal subgroup of $\langle A \rangle = H_2$, and H_2/H_1 is abelian. □

Remark. Note a difference between the statement of Cor. 8.5 and the statement of Thm. 1.1: here, in Cor. 8.5, we are asserting that H_1 and H_2 are normal subgroups of $\langle A \rangle$, and not just that H_1 is normal in H_2.

We could almost certainly strengthen the statement of Thm. 1.1. It is not clear, however, how to do this without either plenty of case-work or the use of the fact that all subgroups of $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ fit more or less snugly inside algebraic subgroups of bounded degree. The second strategy might work for $n > 3$ (possibly with some careful usage of different finite fields), but then it can probably be applied at the end, taking the current statement of Thm. 1.1 as a starting point.

One advantage of a statement such as that of Thm. 1.1 – namely, a statement that does not require H_1 and H_2 to be normal in G – is that it is robust when we pass to a subgroup of small index, as we are about to see.

Corollary 8.6 (to Corollary 8.5)**.** *Let $G = \mathrm{PGL}_2$ or $G = \mathrm{SO}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$.*

Then, for every $\epsilon > 0$, either

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{8.5}$$

where $\delta > 0$ and the implied constant depend only on ϵ, or there are subgroups $H_1 \lhd H_2 < \langle A \rangle$ such that

(a) *H_2/H_1 is abelian,*
(b) *A_k contains H_1, where k depends only on ϵ, and*
(c) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2.*

Proof. Since $\mathrm{SO}_3(\mathbb{Z}/p\mathbb{Z})$ and $\mathrm{PGL}_2(\mathbb{Z}/p\mathbb{Z})$ are isomorphic as groups, it is enough to prove the statement for $G = \mathrm{PGL}_2$.

Let, then, $G = \mathrm{PGL}_2$. Let $G' = \mathrm{PSL}_2$. Then $G'(K) < G(K)$ and $[G(K) : G'(K)] = 2$. Let $A' = A^{-1}A \cap G'(K)$. By Lemma 7.2, $|A'| \geq \frac{1}{2}|A|$. Now let $\pi : \mathrm{SL}_2(K) \to \mathrm{PSL}_2(K)$ be the natural projection map, and let $A'' = \pi^{-1}(A') \subset \mathrm{SL}_2(K)$.

Apply Corollary 8.5 to A''. Clearly (8.4) implies (8.5) (with the implied constant changing by a factor of at most 2). If (8.4) does not hold, then Corollary 8.5 provides subgroups $H_1' \lhd H_2' < \langle A'' \rangle$; we use them to define subgroups $H_1 = \pi(H_1')$, $H_2 = \pi(H_2')$ satisfying the properties in the statement of the corollary we are proving. (In order to get that A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2, use Lemmas 7.2 and 7.1.) □

8.2. **Parabolic subgroups of** $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$**: general setup.** Let $K = \mathbb{Z}/p\mathbb{Z}$, $G = \mathrm{SL}_3(K)$. Let $e_1, e_2, e_3 \in K^3$ be a basis of K^3. Let $P \subset G$ be the stabiliser of the subspace $Ke_1 + Ke_2$ of K^3 under the natural action of G in K^3. (This is the same as the stabiliser of $Ke_1 + Ke_2$ seen as a line in $\mathbb{P}^3(K)$; we shall be working in affine rather than projective language.) Let H_0 be the group consisting of the elements $g \in P(K)$ sending e_3 to elements of the form $a_1e_1 + a_2e_2 + a_3e_3$, with $a_1, a_2 \in K$ and $a_3 \in K^*$ a square in K^*. Let M be the subgroup of H_0 consisting of the elements $g \in P(K)$ sending e_3 to elements $a_1 e_1 + a_2 e_2 + e_3$ with $a_1, a_2 \in K$. Let

- G_+ be the subgroup of H_0 consisting of all $g \in H_0$ fixing the space Ke_3,
- G_- be the subgroup of M consisting of all $g \in M$ fixing e_3,
- A_0 be the subgroup of M consisting of all $g \in M$ fixing both e_1 and e_2, and
- $Z(G_+)$ be the center of G_+,
- $\pi_+ : H_0 \to G_+$, $\pi_- : M \to G_-$ be the natural projections.

More legibly, in matrix form (with e_1, e_2, e_3 as the basis),

$$
\begin{aligned}
H_0 &= \left\{ g = \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & s^2 \end{pmatrix} : \det(g) = 1, s \in K^* \right\}, \\
G_+ &= \left\{ g = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & s^2 \end{pmatrix} : \det(g) = 1, s \in K^* \right\}, \\
M &= \left\{ g = \begin{pmatrix} a & b & * \\ c & d & * \\ 0 & 0 & 1 \end{pmatrix} : \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 1 \right\}, \\
G_- &= \left\{ g = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} : \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 1 \right\}, \\
Z(G_+) &= \left\{ g = \begin{pmatrix} s^{-1} & 0 & 0 \\ 0 & s^{-1} & 0 \\ 0 & 0 & s^2 \end{pmatrix} : s \in (\mathbb{Z}/p\mathbb{Z})^* \right\}, \\
A_0 &= \left\{ \begin{pmatrix} 1 & 0 & e \\ 0 & 1 & f \\ 0 & 0 & 1 \end{pmatrix} \right\},
\end{aligned}
\tag{8.6}
$$

where all entries are understood to lie in K. The projections $\pi_+ : H_0 \to G_+$, $\pi_- : M \to G_-$ are given by

$$
\pi_+ \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & s^2 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s^2 \end{pmatrix}, \quad \text{and} \quad \pi_- \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{8.7}
$$

It is clear that $G_- \simeq \mathrm{SL}_2(K)$ and $A_0 \simeq K^2$. Moreover, A_0 is a normal subgroup of M; the projection $\pi_- : M \to G_-$ can be identified with the quotient homomorphism $M \to A_0 \backslash M \simeq G_-$. (Here we write $A_0 \backslash M$ for the group of right cosets of A_0 in M.) We can thus see M as a semidirect product $A_0 \rtimes G_-$ of A_0 and G_-. The action of G_- on A_0 in the semidirect product $M = A_0 \rtimes G_-$ is the natural one, as is shown by the identity

$$
\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & e \\ 0 & 1 & f \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & ae+bf \\ 0 & 1 & ce+df \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} I & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} e \\ f \end{pmatrix} \\ 0 & 1 \end{pmatrix}.
$$

In other words, we may write the elements of M as pairs (a, g), $a \in A_0$, $g \in G_-$, and then the group law of M looks as follows:

$$(a_1, g_1) \cdot (a_2, g_2) = (a_1 + g_1 \cdot a_2, g_1 g_2),$$

where $g_1 \in G_- \simeq \mathrm{SL}_2(K)$ acts on $a_2 \in A_0 \simeq K^2$ by the natural action of $\mathrm{SL}_2(K)$ on K^2.

We can also decompose G_+ as a product, namely, $G_+ \simeq \mathrm{SL}_2(K) \times \{x^2 : x \in K^*\}$. We let the projection maps $\pi_1 : G_+ \to \mathrm{SL}_2(K)$, $\pi_2 : G_+ \to K^*$ be given by

$$\pi_1 \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s^2 \end{pmatrix} = \begin{pmatrix} sa & sb \\ sc & sd \end{pmatrix} \in SL_2(K), \quad \pi_2 \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s^2 \end{pmatrix} = s^2. \tag{8.8}$$

The above setup will be somewhat familiar to some readers from the theory of automorphic forms. (The group M is of a kind called *mirabolic* by some; the decomposition $M = A_0 \rtimes G_-$ treated above is well-known in general.)

In the following, we shall examine a subset E of H_0, and determine its growth. (We call our set E rather than A so as to avoid confusion with the group A_0, which is usually called A in the literature.) Since $\pi_1(\pi_+(E))$ is a subset of $\mathrm{SL}_2(K)$, it generates a subgroup of $\mathrm{SL}_2(K)$. This subgroup can be all of $\mathrm{SL}_2(K)$, or it can lie inside one of the maximal subgroups of $\mathrm{SL}_2(K)$ (which were classified in Prop. 8.3). We treat these two cases individually.

8.3. Parabolic subgroups: passage to SL_2, or, the case $\langle \pi_1(\pi_+(E)) \rangle = \mathrm{SL}_2(K)$.

Proposition 8.7. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $G = \mathrm{SL}_3$. Let G_+, G_- and H_0 be as in (8.6); let $\pi_+ : H_0 \to G_+$ and $\pi_1 : G_+ \to \mathrm{SL}_2(K)$ be as in (8.7) and (8.8).*

Let E be a subset of H_0 such that $\pi_1(\pi_+(E))$ generates $\mathrm{SL}_2(K)$. Then either

$$|E \cdot E \cdot E| > |\pi_1(\pi_+(E))|^{\epsilon} \cdot |E| \tag{8.9}$$

or

$$\pi_1(\pi_+(E_k)) = \mathrm{SL}_2(K), \tag{8.10}$$

where $\epsilon > 0$ and k are absolute constants.

Proof. It is here that the inductive step happens; we will use what we know on SL_2. By the Key Proposition in [He, §1], there are absolute constants $\delta, \epsilon, k > 0$ such that, for $A \subset \mathrm{SL}_2(K)$ generating $\mathrm{SL}_2(K)$, we have two cases:

- If $|A| \leq |\mathrm{SL}_2(K)|^{1-\delta}$, then $|A \cdot A \cdot A| > |A|^{1+\epsilon}$.
- If $|A| > |\mathrm{SL}_2(K)|^{1-\delta}$, then $A_k = \mathrm{SL}_2(K)$.

Now define $\pi = \pi_1 \circ \pi_+$. By the statement of the lemma, $\pi(E)$ generates $\mathrm{SL}_2(K)$. If $|\pi(E)| > |\mathrm{SL}_2(K)|^{1-\delta}$, then

$$\pi(E_k) = \pi(E)_k = \mathrm{SL}_2(K),$$

and we are done. Suppose $|\pi(E)| \leq |\mathrm{SL}_2(K)|^{1-\delta}$. Then

$$|\pi(E \cdot E \cdot E)| = |\pi(E) \cdot \pi(E) \cdot \pi(E)| > |\pi(E)|^{1+\epsilon},$$

and so, by Lemma 7.4 (applied with $A_1 = E$, $A_2 = E \cdot E$),

$$|E_8| > |\pi(E)|^{\epsilon} \cdot |E|.$$

Statement (8.9) then follows by the tripling lemma (Lemma 2.2). □

If we have (8.9) and $|\pi_1(\pi_+(E))| > |E|^\delta$ for some fixed $\delta > 0$, the problem is solved. We will leave the case of $|\pi_1(\pi_+(E))| \leq |E|^\delta$ for later. We focus for now on (8.10), i.e., on the case of sets E with $\pi_1(\pi_+(E)) = \mathrm{SL}_2(K)$. (The case of sets E with $\pi_1(\pi_+(E_k)) = \mathrm{SL}_2(K)$ reduces to this after we multiply E with itself and its inverse a few times.)

We will need the following result, credited by Dickson to Galois.

Proposition 8.8 (Galois). *Let p be a prime with $p > 11$. Let G be $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ or $\mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$. Let H be a proper subgroup of G. Then $[G : H] \geq p + 1$.*

This can be derived quickly from Prop. 8.3.

Proof. See, e.g., [Di], e.g., Ch. XII, Theorem 261. □

We can now proceed.

Lemma 8.9. *Let $K = \mathbb{Z}/p\mathbb{Z}$, $p > 11$ a prime. Let G, H_0 and M be as in (8.6); let $\pi_+ : H_0 \to G_+$, $\pi_- : M \to G_-$ and $\pi_1 : G_+ \to \mathrm{SL}_2(K)$ be as in (8.7) and (8.8).*

Let E be a subset of H_0 such that $\pi_1(\pi_+(E)) = \mathrm{SL}_2(K)$. Then

$$\pi_-(E_k \cap M) = G_-,$$

where k is an absolute constant.

Proof. Since $\pi_1(\pi_+(E)) = \mathrm{SL}_2(K)$, we know that $|\pi_+(E)| \geq |\mathrm{SL}_2(K)|$. Let

$$R = (\pi_+(E)^{-1}\pi_+(E)) \cap G_-.$$

Then, by Lemma 7.2,

$$|R| \geq \frac{|\pi_+(E)|}{|[G_+ : G_-]|} \geq \frac{|\mathrm{SL}_2(K)|}{(p-1)/2} = \frac{|G_-|}{(p-1)/2}. \tag{8.11}$$

By Prop. 8.8, $G_- \simeq \mathrm{SL}_2(K)$ has no proper subgroups of index $\leq \frac{p-1}{2}$; hence R generates G_-. Now, (8.11) also gives us that $|R| > |G_-|^{2/3}$. We apply the Key Proposition in [He] (part (b)) and obtain that

$$R_k = G_-,$$

where k is an absolute constant.

Since $R = \pi_+(E^{-1}E) \cap G_- \subset \pi_+(E_2) \cap G_-$, it follows that

$$\pi_-(E_{2k} \cap M) \supseteq \pi_-((E_2 \cap M)_k) = (\pi_-(E_2 \cap M))_k = (\pi_+(E_2) \cap G_-)_k \supseteq R_k = G_-,$$

as we desired. □

We need a little lemma on cohomology.

Lemma 8.10. *Let G be a group acting on an abelian group R. Suppose the centre Z of G contains an element $z \in Z$ with $z^2 = e$ such that $zv = -v$ for all $v \in R$. Suppose furthermore that every element of R is uniquely 2-divisible, i.e., suppose that, for every $r \in R$, there is a unique $r' \in R$ such that $r = 2r'$.*

Then

$$H^1(G, R) = 0. \tag{8.12}$$

We can restate (8.12) in non-cohomological language as follows: given any map $s : G \to R$ satisfying $s(g_1g_2) = s(g_1) + g_1s(g_2)$ for all $g_1, g_2 \in G$, there is a v such that $s(g) = gv - v$ for all g.

One can show [Hi] that $H^n(G, R) = 0$, $n \geq 1$, under the same conditions we have given; we shall need only the case $n = 1$. The conditions of the lemma are clearly satisfied when $G = \mathrm{SL}_2(K)$, $R = K^2$, K a finite field of odd order: set $z = -I$.

Proof. Let $g \in G$. Because z is in the centre, z and g commute; we also have that z is an involution, i.e., $z^2 = e$. Thus

$$\begin{aligned} s(g) &= s(g \cdot z^2) = s(z \cdot g \cdot z) = s(z) + z \cdot s(g \cdot z) = s(z) - s(g \cdot z) \\ &= s(z) - (s(g) + g \cdot s(z)) = -s(g) + s(z) - g \cdot s(z). \end{aligned}$$

Thus

$$s(g) = \frac{1}{2}(s(z) - g \cdot s(z)).$$

So

$$s(g) = gv - v$$

for $v = -\frac{1}{2}s(z)$. □

Lemma 8.11. *Let $K = \mathbb{Z}/p\mathbb{Z}$, where p is an odd prime. Let M, G_- and A_0 be as in (8.6). Let $\pi_- : M \to G_-$ be as in (8.7).*

Let $E \subset M$ be such that $\pi_-(E) = G_-$. Then either

$$E_k = M,$$

where k is an absolute constant, or

$$E = gG_-g^{-1} \qquad \text{for some } g \in M.$$

Proof. Suppose first that there are two distinct $g_1, g_2 \in E$ such that $\pi_-(g_1) = \pi_-(g_2)$. Then $g_1^{-1}g_2$ is an element of A_0 other than I. Since $\mathrm{SL}_2(K)$ acts transitively on the set of non-zero elements of K^2, we have that G_- acts transitively on A_0 by conjugation. (Recall that $M \simeq A_0 \rtimes G_-$, $A_0 \simeq K^2$, $G_- \simeq \mathrm{SL}_2(K)$, and that the action of G_- on A_0 by conjugation is described by the action of $\mathrm{SL}_2(K)$ on K^2.) Since $\pi_-(E) = G_-$, it follows that $Eg_1^{-1}g_2E^{-1} \subset M$ is all of A_0, and so

$$Eg_1^{-1}g_2E^{-1}E \subset E_5$$

is equal to all of M.

Now suppose that there are no two distinct $g_1, g_2 \in E$ with $\pi_-(g_1) = \pi_-(g_2)$. Then E is of the form $\{(s(h), h) : h \in G_-\}$, where s is a map $s : G_- \to A_0$. If there are $h_1, h_2 \in G_-$ such that

$$(s(h_1), h_1) \cdot (s(h_2), h_2) \neq (s(h_1h_2), h_1h_2),$$

then the argument is as before: there are two distinct elements (namely, $(s(h_1), h_1) \cdot (s(h_2), h_2)$ and $(s(h_1h_2), h_1h_2)$) whose image h_1h_2 under π_- is the same, and so

$$E_{10} = M.$$

Suppose, then, that

$$(s(h_1), h_1) \cdot (s(h_2), h_2) = (s(h_1h_2), h_1h_2)$$

for all $h_1, h_2 \in G_-$, or what is the same,

$$s(h_1 h_2) = s(h_1) + h_1 s(h_2)$$

for all $h_1, h_2 \in G_-$. We now use Lemma 8.10, and conclude that $s(h) = hv - v$ for some $v \in A_0$. Hence

$$\begin{aligned} E = \{(s(h), h) : h \in G_-\} &= (-v, 1) \cdot \{(0, h) : h \in G_-\} \cdot (v, 1) \\ &= g \cdot G_- \cdot g^{-1}, \end{aligned}$$

where g is the element of M corresponding to $(-v, 1)$ under the isomorphism $M \simeq A_0 \rtimes G_-$. □

We can now draw certain conclusions.

Proposition 8.12. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $G = \mathrm{SL}_3$. Let $H_0 < G(K)$ be as in (8.6); let π_+ and π_1 be as in (8.7) and (8.8).*

Let $E \subset H_0$ be such that $\pi_1(\pi_+(E))$ generates $\mathrm{SL}_2(K)$. Then, for every $\epsilon > 0$, either

(a) *$|E \cdot E \cdot E| > |E|^{1+\delta}$, where $\delta > 0$ depends only on ϵ, or*
(b) *E_k contains M, where k is an absolute constant, or*
(c) *E_k contains gG_-g^{-1} and is contained in gG_+g^{-1}, where $g \in M$ and k is an absolute constant,*
(d) *$E^{-1}E$ has $\geq |E|^{1-\epsilon}$ elements in the subgroup H of H_0 given by*

$$\left\{ \begin{pmatrix} s & 0 & a \\ 0 & s & b \\ 0 & 0 & s^{-2} \end{pmatrix} : a, b \in K,\ s \in K^* \right\}; \tag{8.13}$$

moreover, E intersects at most $|E|^{2\epsilon}$ cosets of H_0.

Here (d) is in effect a reduction to one of the cases to be treated in the next subsection (Lem. 8.14). We will treat it further there.

Proof. Suppose $|\pi_1(\pi_+(E))| \leq |E|^{\epsilon}$. Then, by Lemma 7.2, there are are $\geq |E|^{1-\epsilon}$ elements of $E^{-1}E$ lying in the kernel of $\pi_1 \circ \pi_+$. The kernel of $\pi_1 \circ \pi_+$ is precisely (8.13), and so we obtain (d). (If the statement on the number of cosets of H_0 that E intersects did not hold, conclusion (a) would follow by Lemma 7.1.)

Suppose now that $|\pi_1(\pi_+(E))| > |E|^{\epsilon}$. If we have (8.9), we are done. It remains to consider what happens if we have (8.10). Applying Lemma 8.9 (with $E = E_k$, k an absolute constant), we see that $\pi_-(E_{kk'} \cap M) = G_-$ for k' an absolute constant. (We may assume that $p > 11$ (as is required by Lemma 8.9) because (a) is trivially true otherwise.) We now apply Lemma 8.11, and obtain that $(E_{kk'} \cap M)_{k''}$ (k'' an absolute constant) equals either M or a conjugate gG_-g^{-1}, $g \in M$, of G_-. If $(E_{kk'} \cap M)_{k''} = M$, we have obtained (b).

Suppose, then, that $(E_{kk'} \cap M)_{k''} = gG_-g^{-1}$. If E is contained in the group gG_+g^{-1}, we have (c) and are done. Assume, then, that there is a $g_1 \in E$ such that $g_1 \notin gG_+g^{-1}$. We can write $g_1 = gazg^{-1}g_2$, where $a \in A$, $z \in Z(G_+)$, $g_2 \in gG_-g^{-1}$, $a \neq I$. The orbit of $gazg^{-1}$ under the action of gG_-g^{-1} by conjugation is all of gA_0zg^{-1}. (This is so because the action of G_- on A_0 by conjugation can be identified with the action of $\mathrm{SL}_2(K)$ on K^2

by left multiplication; since the latter action is transitive, the former action is transitive too.) Thus $gG_-g^{-1} \cdot g_1 \cdot gG_-g^{-1}$ contains gA_0zg^{-1}, and hence

$$gG_-g^{-1} \cdot g_1 \cdot gG_-g^{-1} = gG_-g^{-1} \cdot (gG_-g^{-1} \cdot g_1 \cdot gG_-g^{-1}) \supset gG_-A_0zg^{-1} = gMzg^{-1} = Mz.$$

Therefore, $E_{2kk'+1}$ contains Mz, and so $E_{4kk'+2}$ contains M. We have obtained (b). $\square$

We have spent enough time for now studying subsets $E \subset H_0(K)$ such that $\pi_1(\pi_+(E))$ generates $\mathrm{SL}_2(K)$; let us now pass to the other cases.

8.4. **Parabolic subgroups: solvable groups, or, the case $\langle \pi_1(\pi_+(E)) \rangle = \mathrm{SL}_2(K)$.** Let $K = \mathbb{Z}/p\mathbb{Z}$, $G = \mathrm{SL}_3(K)$. Let $H_0 < G(K)$ be as in (8.6); let π_+, π_1 be as in (8.7) and (8.8). Consider a subset $E \subset H_0$ such that $\pi_1(\pi_+(E))$ does not generate $\mathrm{SL}_2(K)$.

By Prop. 8.3, either (a) $\pi_1(\pi_+(E))$ is contained in a Borel subgroup B/K of $\mathrm{SL}_2(K)$ or (b) $\pi_1(\pi_+(E))$ is contained in a subgroup $H < \mathrm{SL}_2(K)$ having a subgroup of index ≤ 2 lying within a maximal torus $T/\overline{K}$ of SL_2.

In case (a), E must be contained in a Borel subgroup B'/K of $SL_3(K)$. We have already examined this situation in §7.3 (Prop. 7.12).

In case (b), there will be (by Lemma 7.2) at least $\frac{1}{2}|E|$ elements $g \in E^{-1}E$ such that $\pi_1(\pi_+(g)) \in T_0(K)$. We will examine this in detail; the solution will be a simple application of Cor. 7.10.

Proposition 8.13. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $H_0 < G(K)$ and $A_0 < G(K)$ be as in (8.6); let π_+ and π_- be as in (8.7) and (8.8). Let $T_0/\overline{K}$ be a maximal torus of* SL_2 *not defined over K. Let $H < H_0$ be the preimage $(\pi_1 \circ \pi_+)^{-1}(T_0(K))$.*

Let $E \subset H$. Then, for every $\epsilon > 0$, either

$$|E_k| \geq |E|^{1+\delta}, \tag{8.14}$$

where k and $\delta > 0$ depend only on ϵ, or one of the following cases holds:

(a) *E is contained in at most $|E|^\epsilon$ cosets of A_0,*
(b) *E_k contains a subgroup X of $Z(G) \cdot A_0$ for some k depending only on ϵ; moreover, X is a normal subgroup of $\langle E \rangle$ and $\langle E \rangle / X$ is abelian.*

Proof. Consider any $g \in H$ with $\pi_1(\pi_+(g)) \neq \pm I$. We wish to show that g has three distinct eigenvalues. (This will simplify matters when we apply Cor. 7.10.) Since $\pi_1(\pi_+(g)) \neq \pm 1$ belongs to a torus, it must have two distinct eigenvalues $\lambda_1, \lambda_2 \in \overline{K}$. If either were in K, the other one would be in K as well (by $\lambda_1\lambda_2 = 1$), and then $\pi_1(\pi_+(g))$ would be diagonalisable over K, i.e., $T_0/\overline{K}$ would be defined over K. Since we are assuming that that cannot happen, it follows that $\lambda_1, \lambda_2 \notin K$. Thus g has one rational (that is, $\in K$) eigenvalues s^2, and two irrational and distinct eigenvalues $s^{-1}\lambda_1$, $s^{-1}\lambda_2$. In particular, g has three distinct eigenvalues.

Consider now any $g \in H$. If g has three distinct eigenvalues, then it clearly has no fixed points when acting on $H^{(1)} \subset A_0$ by conjugation. If, instead, $\pi_1(\pi_+(g)) = \pm I$, then, unless g is actually in $Z(G) \cdot A_0$, it is also easy to see that g acts without fixed points on $H^{(1)}$.

We can thus apply Cor. 7.10 with $G = H$, $m = 1$, $H_1 = Z(G) \cdot A_0$, $\ell = 2$. Case (a) of Cor. 7.10 gives us (8.14); cases (b) and (c) give us conclusions us conclusions (a) and (b). $\square$

We can now study the case of Prop. 8.12 that we left for later.

Lemma 8.14. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $H_0, A_0 < \mathrm{SL}_3(K)$ be as in (8.6); let π_+ and π_1 be as in (8.7) and (8.8).*

Let $E \subset H_0$ be such that $\pi_1(\pi_+(E))$ generates $\mathrm{SL}_2(K)$. Suppose that $E^{-1}E$ lies in the union of at most $|E|^\delta$ ($\delta > 0$) cosets of the subgroup H' of H_0 given by

$$\left\{ \begin{pmatrix} s & 0 & a \\ 0 & s & b \\ 0 & 0 & s^{-2} \end{pmatrix} : a, b \in K, s \in K^* \right\}.$$

Then either

$$|E_k| \geq |E|^{1+\delta}, \tag{8.15}$$

where k and $\delta > 0$ depend only on ϵ, or one of the following cases holds:

(a) *E is contained in at most $|E|^{4\delta}$ cosets of A_0,*
(b) *E is contained in gG_+g^{-1}, where $g \in M$; moreover, E is contained in the union of at most $|E|^\delta$ cosets of $gZ(G_+)g^{-1}$;*
(c) *E_k contains A_0 for some k depending only on δ.*

Proof. Apply Cor. 7.10 with $G = H$, $m = 1$, $H_1 = A_0 \cdot Z(G)$, $\ell = 2$, $A = H \cap E^{-1}E$ and 2δ instead of δ. Case (a) in Cor. 7.10 gives us (8.15). Case (b) (together with Lem. 7.1) gives us conclusion (a). Assume, then, that case (c) holds.

Suppose first that $X \neq \{e\}$. If $X = A_0$, we have obtained conclusion (c). Suppose X is neither $\{e\}$ nor A_0. Since $\pi_1(\pi_+(E))$ generates $\mathrm{SL}_2(K)$, X is not stabilised by the action of $\langle E \rangle$ by conjugation. Thus, there is an $h \in E$ such that hXh^{-1}, while in A_0, is not equal to X. Hence $hXh^{-1}X$ is all of A_0, and thus we have reached conclusion (c) again.

Suppose now that $X = \{e\}$. Then $\langle A \rangle$ is abelian. Unless conclusion (a) holds, this means that $\langle A \rangle$ lies in a conjugate $gZ(G_+)g^{-1}$ of $Z(G_+)$ ($g \in M$). If every element of E lies in gG_+g^{-1}, we have obtained conclusion (b). If there is an element h of E not in gG_+g^{-1}, then $h\langle A \rangle h^{-1}$, while certainly in H', is a different torus from $\langle A \rangle$, and so $hAh^{-1}A^{-1}$ contains an element of A_0 other than the identity. We apply Cor. 7.10 with $G = H$, $m = 1$, $H_1 = A_0 \cdot Z(G)$, $\ell = 2$ and $hAh^{-1}A^{-1}$ instead of A; each case works out as before, except that $X = \{e\}$ is no longer a possibility. □

8.5. **Conclusions.**

We can now give a detailed account of what happens inside a parabolic subgroup.

Proposition 8.15. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $G = \mathrm{SL}_3$. Let G_+, G_- and H_0 be as in (8.6).*

Let A be a subset of H_0. Then, for every $\epsilon > 0$, either

$$|AAA| \gg |A|^{1+\delta}, \tag{8.16}$$

where δ and the implied constant depend only on ϵ, or one of the following cases holds:

(a) *A_k is contained in the union of at most $|A|^\epsilon$ cosets of $U(K)$ for some maximal unipotent subgroup $U < G$;*

(b) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H, where $H < H_0$ is conjugate in H_0 to one of the following abelian subgroups of H_0:*

$$H_{diagonal} = \left\{ \begin{pmatrix} as^{-1} & 0 & 0 \\ 0 & bs^{-1} & 0 \\ 0 & 0 & s^2 \end{pmatrix} : a, b, s \in K^* \right\}, \tag{8.17}$$

$$H_{1,2} = \left\{ \begin{pmatrix} r & x & 0 \\ 0 & r & 0 \\ 0 & 0 & r^{-2} \end{pmatrix} : r \in K^*, x \in K \right\}, \tag{8.18}$$

$$H_{2,3}(K) = \left\{ \begin{pmatrix} r^{-2} & 0 & 0 \\ 0 & r & z \\ 0 & 0 & r \end{pmatrix} : r \in K^*, z \in K \right\} \tag{8.19}$$

$$H_{1,3}(K) = \left\{ \begin{pmatrix} r & 0 & y \\ 0 & r^{-2} & 0 \\ 0 & 0 & r \end{pmatrix} : r \in K^*, y \in K \right\}; \tag{8.20}$$

(c) *there is a Borel subgroup B/K of G containing A and a subgroup $H_1 < U(K)$ ($H_1 \neq \{e\}$) of the maximal unipotent subgroup U/K of B/K such that A_k contains H_1 for some k depending only on ϵ; moreover, there is a subgroup $H_2 < B(K)$ such that $H_1 \triangleleft H_2$, H_2/H_1 is abelian, and A is contained in the union of at most $|A|^\epsilon$ cosets of H_1;*

(d) *A_k contains M, where k is an absolute constant;*

(e) *A_k contains gG_-g^{-1} and is contained in gG_+g^{-1}, where $g \in M$ and k is an absolute constant,*

(f) *A_k is contained in the union of at most $|A|^\epsilon$ cosets of $gZ(G_+)g^{-1}$ for some $g \in M$; moreover, A is contained in gG_+g^{-1} for the same $g \in M$.*

Proof. Suppose first that A is contained in a Borel subgroup B/K of $G = \mathrm{SL}_3$. Apply Proposition 7.12. Cases (a) and (b) of Prop. 7.12) give us (8.16) and conclusion (a). (We can always adjust ϵ by a constant factor before applying Prop. 7.12, say.) Assume, then, that we are in case (c) of Prop. 7.12. If $X \neq \{e\}$, we have reached conclusion (c). Suppose $X = \{e\}$.

What we have is that A lies in the union of $\leq |A|^\epsilon$ cosets of an abelian subgroup Y of $B(K)$ not contained in $U(K)$. If Y has an element with three distinct eigenvalues, then Y is contained in a conjugate of the group H_{diagonal} in (8.17), and so conclusion (b) holds. If Y has an element with two distinct eigenvalues, then Y is contained in a conjugate of one of the groups $H_{1,2}$, $H_{2,3}$, $H_{1,3}$ defined in (8.18)–(8.20), and so conclusion (c) holds. If Y has only elements with identical eigenvalues, then conclusion (a) holds.

Suppose now that A is not contained in any Borel subgroup B/K of SL_3. Let π_+ and π_1 be as in (8.7) and (8.8). Assume first that $\pi_1(\pi_+(A))$ generates $\mathrm{SL}_2(K)$. We apply Proposition 8.12 (with A instead of E and $\epsilon/8$ instead of ϵ). Cases (a), (b) and (c) in Prop. 8.12 give us (8.16), conclusion (d) and conclusion (e), respectively. If case (d) in Prop. 8.12 holds, apply Lemma 8.14 (with A instead of E). Equation (8.15) gives us (8.16); cases (a), (b) and (c) of Lemma 8.14 give us conclusions (b), (f) and (c), respectively.

Assume now, lastly, that (a) A is not contained in any Borel subgroup of G, and (b) $\pi_1(\pi_+(A))$ does not generate $\mathrm{SL}_2(K)$. Then, as we discussed at the beginning of §8.4, the situation reduces to that of Prop. 8.13. Equation (8.14) gives us (8.16); case (a) of Prop. 8.13) gives us conclusion (a), and case (b) of Prop. 8.13 gives us conclusion (b) or (c), depending on whether or not $X = \{e\}$. □

Corollary 8.16. *Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $G = \mathrm{SL}_3$. Let H_0 be as in (8.6).*
Let A be a subset of H_0. Then, for every $\epsilon > 0$, either

$$|AAA| \gg |A|^{1+\delta},$$

where δ and the implied constant depend only on ϵ, or there are subgroups $H_1 \lhd H_2 < \langle A\rangle$ such that

(a) *H_2/H_1 is nilpotent,*
(b) *A_k contains H_1, where k depends only on ϵ, and*
(c) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2.*

Proof. Apply Prop. 8.15. Case (a) of Prop. 8.15 gives us $H_1 = \{e\}$, $H_2 = U(K)$. Case (b) of Prop. 8.15 gives us $H_1 = \{e\}$, $H_2 = H$. Case (c) gives us sets H_1 and H_2 explicitly. Case (d) gives us $H_1 = M$ and $H_2 = H_0$. Case (e) gives us $H_1 = gG_-g^{-1}$ and $H_2 = gG_+g^{-1}$. Finally, case (f) gives us $H_1 = gZ(G_+)g^{-1}$ and $H_2 = gG_+g^{-1}$. □

Since the above is true for subsets A of H_0, and every maximal parabolic subgroup P of $\mathrm{SL}_3(K)$ is isomorphic to a group having H_0 as a subgroup of index 2, the statement of Cor. 8.16 is true for all subsets A of P just as well. (Use Lemma 7.2).

We can now prove Thm. 1.1 in the case where A does not generate G.

Proposition 8.17. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set that does not generate $G(K)$.*
Then, for every $\epsilon > 0$, either

$$|A \cdot A \cdot A| \gg |A|^{1+\delta},$$

where $\delta > 0$ and the implied constant depend only on ϵ, or there are subgroups $H_1 \lhd H_2 < \langle A\rangle$ such that

(a) *H_2/H_1 is nilpotent,*
(b) *A_k contains H_1, where k depends only on ϵ, and*
(c) *A is contained in the union of $\leq |A|^\epsilon$ cosets of H_2.*

Proof. Let $H = \langle A\rangle$. Then H satisfies one of the descriptions in Cor. 8.2, cases (a)–(e).

If case (a) of Cor. 8.2 holds, apply Cor. 8.16 to H_0 and $A' = A^{-1}A \cap H_0$, where H_0 is as defined in the beginning of §8.1. Since stabilisers of lines in $\mathbb{P}^3$ defined over $\mathbb{Z}/p\mathbb{Z}$ are isomorphic as groups to stabilisers of points in $\mathbb{P}^3$ defined over $\mathbb{Z}/p\mathbb{Z}$ (see (8.1) and (8.2)), case (b) of Cor. 8.2 reduces to case (a) of Cor. 8.2. Case (d) of Cor. 8.2 gives us desired conclusion immediately (with $H_1 = \{e\}$). If case (c) of Cor. 8.2 holds, apply Cor. 8.6. Finally, if case (e) holds, then $|A|$ is bounded by an absolute constant and so (1.2) holds trivially.

□

9. Growth of medium-sized and large sets

Let $G = \mathrm{SL}_3$, $K = \mathbb{Z}/p\mathbb{Z}$. Let A be a set of generators of $G(K)$. We must show that, if $p^{4-\delta} \leq |A| \leq p^{4+\delta}$, $\delta > 0$, then A grows. We will, in fact, be able to show something stronger: if $|A| \geq p^{3.2+\delta'}$, $\delta' > 0$, then A grows.

The key here will be to pass to a subgroup. We let H_0 be as in §8.2. The group H_0 is then a subgroup of index 2 in a maximal parabolic subgroup of $G(K)$, and so $[G(K) : H_0] = 2(p^2 + p + 1)$. Then A is a great deal larger than $[G(K) : H_0]$, and thus, by Lemma 7.2, the intersection $A^{-1}A \cap H$ will be large. We devoted most of §8 to the question of which subsets of H_0 grow. If $A^{-1}A \cap H_0 \subset H_0$ grows, then, by Lemma 7.3, A itself grows. If, instead, $A^{-1}A \cap H_0$ falls into cases (a)–(f) of Proposition 8.15, we obtain that $A^{-1}A \cap H$ either contains a fairly large subgroup of H (called H_1 in Cor. 8.16) or has most of its elements in a proper subgroup of H (called H_2 in Cor. 8.16). (If H_2 were not proper (i.e., $H_2 = G(K)$) and H_1 were not large, then H_2/H_1 would not be nilpotent, in contradiction to Cor. 8.16.) In either case, we can multiply conjugates of that subgroup ("sticking subgroups in different directions") to obtain that A grows.

9.1. Sticking subgroups of SL_3 in different directions.

Let us begin by considering abelian subgroups that are neither tori nor subgroups of unipotent subgroups.

Lemma 9.1. *Let $G = \mathrm{SL}_3$, seen as a group defined over a field K of characteristic $\neq 3$. Let H be one of the subgroups $H_{i,j} \subset G$ listed in Prop. 8.15(b). Let $\mathfrak{g}$ be the Lie algebra of G and $\mathfrak{h}$ the Lie algebra of H. Then there are $\vec{g}_1, \vec{g}_2 \in \mathfrak{g}$ such that*

$$\mathfrak{h}, [\vec{g}_1, \mathfrak{h}], [\vec{g}_2, \mathfrak{h}]$$

are linearly independent and of dimension $\dim(\mathfrak{h})$.

Proof. Since the three subgroups $H_{i,j}$ listed in Prop. 8.15(b), are conjugate over $G(K)$, we can assume without loss of generality that we have $H = H_{1,2}$ (where $H_{1,2}$ is as in (8.18)). Then $\mathfrak{h}$ is spanned by $e_{1,1} + e_{2,2} - 2e_{3,3}$ and $e_{1,2}$, where $e_{i,j}$ is the 3-by-3 matrix having a 1 at the (i,j)th entry and 0s elsewhere. Set $\vec{g}_1 = e_{3,1}$, $\vec{g}_2 = e_{2,3}$. □

Proposition 9.2. *Let $G = \mathrm{SL}_3$, seen as a group defined over a field K of characteristic* $\mathrm{char}(K) = 0$ *or* $\mathrm{char}(K) > 3$. *Let $H \subset G$ be conjugate over $G(\overline{K})$ to one of the subgroups $H_{i,j} \subset H$ listed in Prop. 8.15(b), (8.18)–(8.20).*

Let A be a set of generators of $G(K)$, and E a non-empty subset of $H(K)$. Then there are $g_0, g_1, g_2 \in A_k$, $k \ll 1$, such that

$$|g_0 E g_0^{-1} \cdot g_1 E g_1^{-1} \cdot g_2 E g_2^{-1}| \gg |E|^3,$$

where the implied constants are absolute.

Proof. By Lemma 9.1, the assumptions of Prop. 4.13 are fulfilled. The conclusions of Prop. 4.13 provide the linear-independence assumption of Prop. 4.12; we apply Prop. 4.12, and are done. The implied constants are absolute because they depend only on n, which is fixed ($n = 3$). □

Let us now look at algebraic subgroups of a unipotent subgroup of SL_3.

Lemma 9.3. *Let $G = \mathrm{SL}_3$, defined over a field K. Let H be one of the algebraic subgroups of G listed in Lemma 4.15 (equations (4.16) – (4.23)) other than $\{I\}$. Let $T \subset G$ be the subgroup of diagonal matrices of G. Write $\mathfrak{g}$ for the Lie algebra of G, $\mathfrak{h} \subset \mathfrak{g}$ for the Lie algebra of H, and $\mathfrak{t} \subset \mathfrak{g}$ for the Lie algebra of T.*

Then there are $g_0, g_1, g_2, \ldots, g_\ell \in G(\overline{K})$ such that

$$\mathrm{Ad}_{g_0}(\mathfrak{t}), \mathrm{Ad}_{g_1}(\mathfrak{h}), \mathrm{Ad}_{g_2}(\mathfrak{h}), \ldots, \mathrm{Ad}_{g_\ell}(\mathfrak{h})$$

are linearly independent. Here $\ell = (\dim(\mathfrak{g}) - \dim(\mathfrak{t}))/\dim(\mathfrak{h})$.

Strictly speaking, Lemma 4.15 actually lists the sets of points $H(K)$; it should be clear which algebraic groups H are thereby listed.

Proof. Set $g_0 = g_1 = I$. In every case, we will set $g_2, \ldots, g_\ell$ equal to permutation matrices. If H is the whole group U of upper-triangular unipotent matrices, then $\ell = 2$; set g_2 equal to the permutation matrix corresponding to the permutation (1 3) – that is,

$$g_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

If H is as in (4.17), let g_2, g_3 be the matrices corresponding to the permutations (1 2) and (1 3). For (4.18), choose g_2, g_3 corresponding to (1 3) and (2 3). For (4.18), (4.20) or (4.21), we use the entire permutation group S_3, i.e., we let $g_1, \ldots, g_6$ be the permutation matrices corresponding to each element of S_3 in turn. For (4.22), we let g_2 and g_3 be the permutation matrices corresponding to the 3-cycles in S_3. Finally, for (4.23), we use the entire permutation group S_3. $\square$

Proposition 9.4. *Let $G = \mathrm{SL}_3$, defined over a field K. Let H be conjugate to one of the algebraic subgroups of G listed in Lemma 4.15 (equations (4.16) – (4.23)) other than $\{I\}$. Let $T/\overline{K}$ be a maximal torus of G. Let $\ell = (\dim(G) - \dim(T))/\dim(H)$, where $\dim(G)$, $\dim(T)$ and $\dim(H)$ are the dimensions of G, T and H as varieties.*

Let A be a set of generators of $G(K)$, D a non-empty subset of $T(K)$, and E a non-empty subset of $H(K)$. Then there are $g_0, g_1, \ldots, g_\ell \in A_k$, $k \ll 1$, such that

$$|g_0 D g_0^{-1} \cdot g_1 E g_1^{-1} \cdots g_\ell E g_\ell^{-1}| \gg |D| \cdot |E|^\ell,$$

where the implied constants are absolute.

Proof. Lemma 9.3 states that the linear-independence assumption of Prop. 4.12 is true. We apply Prop. 4.12, and we are done. The implied constants are absolute because they depend only on n, which is fixed ($n = 3$). $\square$

9.2. **Conclusions.** The following lemma contains all of our work in this section so far.

Lemma 9.5. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $H_0 < G(K)$ be as in (8.6). Let $A \subset G(K)$ be a set of generators of $G(K)$, and let E be a non-empty subset of H_0.*

Then, for every $\epsilon > 0$, either

(a) $|A_k| \gg |A|^{1+\epsilon}$, *where k and the implied constant are absolute, or*

(b) $|E \cdot E \cdot E| \gg |E|^{1+\delta}$, *where δ and the implied constant depend only on ϵ, or*

(c) $|(A\cup E)_k| \gg |A|^{\frac{1}{4}-\epsilon}\cdot|E|^{2-2\epsilon}$, *where the implied constant is absolute and* k *depends only on* ϵ,

(d) $|(A\cup E)_k| \gg |E|^{3-3\epsilon}$, *where* k *and the implied constant are absolute,*

(e) $|(A\cup E)_k| \gg p^6\cdot|A|^{\frac{1}{4}-\epsilon}$, *where* k *and the implied constant are absolute, or*

Proof. Apply Prop. 8.15 with E instead of A. If (8.16) holds, we have conclusion (b). Assume (8.16) does not hold. Then one of cases (a)–(f) in Prop. 8.15 holds.

Suppose case (a) of Prop. 8.15 holds. Then, by Lemma 7.2, $|E_{2k}\cap U(K)| \geq |E|^{1-\epsilon}$. By Cor. 5.10, either conclusion (a) holds (with $k+2$ instead of k) or $|A_k\cap T(K)| \gg |A|^{\frac{1}{4}-\epsilon}$, where the implied constant is absolute. Suppose $|A_k\cap T(K)| \gg |A|^{\frac{1}{4}-\epsilon}$. By Prop. 9.4, it follows that

$$|(A\cup E)_{6k'+5k}| \gg |A_k\cap T(K)||E_{2k}\cap U(K)|^2 \gg |A|^{\frac{1}{4}-\epsilon}|E|^{2-2\epsilon},$$

where k' and the implied constant are absolute (k' is the constant k from Prop. 9.4). Conclusion (c) follows (with $6k'+5k$ instead of k).

Suppose case (b) of Prop. 8.15 holds. Then, by Lemma 7.2, $|E_2\cap H(K)| \geq |E|^{1-\epsilon}$, where H is one of the groups listed in (8.17)–(8.20). Suppose first that H is as in (8.17). Then, by Proposition 5.3,

$$|(A\cup E)_{8k+8}| \gg |E_2\cap H(K)|^4 \geq |E|^{4-4\epsilon},$$

where k and the implied constant are absolute. Conclusion (d) follows (with $8k+8$ instead of k; we may assume $\epsilon<1$, and so $4-4\epsilon > 3-3\epsilon$). Now suppose H is as in (8.18), (8.19) or (8.20). Then, by Proposition 9.2,

$$|(A\cup E)_{6k+3}| \gg |E_2\cap H(K)|^3 \geq |E|^{3-3\epsilon},$$

where k and the implied constant are absolute. Conclusion (d) follows again (with $6k+3$ instead of k).

Suppose case (c) of Prop. 8.15 holds. Then H_1 is one of the subgroups listed in Lem. 4.15, (4.17)–(4.23). By Cor. 5.10, either conclusion (a) holds (with $k+2$ instead of k) or $|A_k\cap T(K)| \gg |A|^{\frac{1}{4}-\epsilon}$, where the implied constant is absolute. Suppose $|A_k\cap T(K)| \gg |A|^{\frac{1}{4}-\epsilon}$. Then, by Prop. 9.4,

$$|(A\cup E)_{k+12k'+6k''}| \gg |A|^{\frac{1}{4}-\epsilon}\cdot p^6,$$

where k, k', k'' and the implied constants are absolute. Conclusion (e) follows (with $k+12k'+6k''$ instead of k).

Suppose case (d) or case (e) of Prop. 8.15 holds. Then E_k contains a conjugate of the unipotent group (4.20), and so cases (d) and (e) reduce to case (c), which we have already treated.

Suppose, lastly, that case (f) of Prop. 8.15 holds. Then, by Lemma 7.2,

$$|E_2\cap gZ(G_+)g^{-1}| \geq |E|^{1-\epsilon}.$$

Now recall that $Z(G_+)$ (defined in (8.6)) is a non-maximal torus. Thus, by Prop. 5.13,

$$|(A\cup E)_{8k+10}| \gg |E|^{5-5\epsilon},$$

where k and the implied constant are absolute. Conclusion (d) follows (with $8k+10$ instead of k; we can assume $\epsilon<1$, and so $5-5\epsilon > 3-3\epsilon$). □

Proposition 9.6. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set of generators of $G(K)$. Suppose $p^{3.2+\eta} \leq |A| \leq p^{8-\eta}$, where $\eta > 0$. Then*

$$|A \cdot A \cdot A| \gg |A|^{1+\delta}, \tag{9.1}$$

where $\delta > 0$ and the implied constant depend only on η.

Proof. Let $E = A^{-1}A \cap H_0$, where $H_0 < G(K)$ is as in (8.6). By Lemma 7.2,

$$|E| \geq \frac{|A|}{[G(K) : H_0]} = \frac{|A|}{2(p^2+p+1)} > \frac{|A|}{3p^2}.$$

Apply Lem. 9.5 with $\epsilon = \min\left(\frac{1}{32}, \frac{\eta}{3}\right)$.

If case (a) of Lem. 9.5 holds, we obtain (9.1) by the tripling lemma (Lemma 2.2). If case (b) of Lem. 9.5 holds, we obtain (9.1) by Lemma 7.3.

Suppose case (c) of Lemma 9.5 holds. Then

$$|A_{2k}| \geq |(A \cup E)_k| \gg |A|^{\frac{1}{4}-\epsilon}|E|^{2-2\epsilon} > \frac{1}{9p^4}|A|^{2+\frac{1}{4}-3\epsilon},$$

where the implied constant is absolute. Since $|A| \geq p^{3.2+\eta}$, we see that $|A|^{1+\frac{1}{4}} \geq p^{4+\frac{5}{4}\eta}$, and so

$$|A_{2k}| \gg |A|^{1+\frac{5}{4}\eta-3\epsilon} = |A|^{1+\frac{1}{4}\eta},$$

where the implied constant is absolute. We then obtain (9.1) by the tripling lemma.

Suppose case (d) of Lemma 9.5 holds. Then

$$|A_{2k}| \geq |(A \cup E)_k| \gg |E|^{3-3\epsilon} > \frac{1}{27p^6}|A|^{3-3\epsilon},$$

where the implied constant is absolute. Since $|A| \geq p^{3.2+\eta}$, we have $|A|^{2\cdot\frac{3}{3.2}} \gg p^6$, and thus

$$|A_{2k}| \gg |A|^{1+(2-2\cdot\frac{3}{3.2})-3\epsilon} = |A|^{1+\frac{1}{8}-3\epsilon} \geq |A|^{1+\frac{1}{32}},$$

where the implied constant is absolute. We obtain (9.1) by the tripling lemma.

Suppose, finally, that (e) of Lem. 9.5 holds. Then

$$|A_{2k}| \geq |(A \cup E)_k| \gg |E|^{3-3\epsilon} > p^6 \cdot |A|^{\frac{1}{4}-\epsilon}.$$

Since $|A| \leq p^{8-\eta}$, where $\eta > 0$, we have $p^6 \geq |A|^{\frac{6}{8-\eta}} \geq |A|^{\frac{3}{4}+\eta}$ (as η is certainly < 7) and so

$$|A_{2k} \geq |A|^{1+\eta-\epsilon}.$$

We obtain (9.1) by the tripling lemma. $\square$

10. General conclusions and final remarks

10.1. **The main theorem, related results and their consequences.** We must now simply put together our work on small and large sets (§3 – §6) with our work on medium-sized and large sets (§9). (As should be clear from the wording, there is an overlap; we have no use for it.)

Main Theorem. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Suppose $|A| \leq |G(K)|^{1-\delta}$, $\delta > 0$. Then

$$|A \cdot A \cdot A| \gg |A|^{1+\epsilon}, \tag{10.1}$$

where $\epsilon > 0$ and the implied constant depend only on δ.

Proof. The condition $|A| \leq |G(K)|^{1-\delta}$ implies $|A| \leq p^{8-8\delta} < p^{8-\delta}$. If $|A| \leq p^{3.5}$ (say) or $p^{4.5} \leq |A| \leq p^{8-\delta}$, use Prop. 6.3. If $p^{3.5} < |A| < p^{4.5}$, use Prop. 9.6 (with $\eta = 3.5 - 3.2 = 0.3$, say). $\square$

In the remainder, we shall need the following extremely simple lemma.

Lemma 10.1. *Let G be a group. Let A be a finite set of generators of G. Then, for every $\ell \geq 1$, either*

$$|A_{\ell+1}| \geq |A_\ell| + 1 \quad \text{or} \quad A_\ell = G.$$

Proof. Since $A_\ell \subset A_{\ell+1}$, either $|A_{\ell+1}| \geq |A_\ell| + 1$ or $A_{\ell+1} = A_\ell$ holds. If $A_{\ell+1} = A_\ell$, then A_ℓ is closed under multiplication by elements of $A \cup A^{-1}$. By the definition of A_ℓ, this implies that A_ℓ is closed under the group operation. We already know that A_ℓ is closed under inversion by the definition of A_ℓ (see (2.1)). Hence A_ℓ is a subgroup of G. Since A generates G, this means that $A_\ell = G$. $\square$

The main theorem has the following alternative statement. It looks stronger, but it isn't really; it is merely simpler to use sometimes.

Proposition 10.2. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

Suppose $|A| \leq |G(K)|^{1-\epsilon}$, $\epsilon > 0$. Then

$$|A \cdot A \cdot A| \geq |A|^{1+\delta},$$

where $\delta > 0$ and the implied constant depend only on ϵ.

This amounts simply to the following: the $\gg$ in (10.1) has been replaced by a $\geq$.

Proof. First, notice that (10.1) (that is, $|A \cdot A \cdot A| \gg |A|^{1+\delta}$) implies $|A \cdot A \cdot A| \geq |A|^{1+\delta/2}$ for $|A|$ larger than a constant C depending only on δ and on the implied constant in (10.1). Since δ and the implied constant in (10.1) depend only on ϵ, which is fixed, we conclude that the main theorem implies that

$$|A \cdot A \cdot A| \geq |A|^{1+\delta/2} \tag{10.2}$$

whenever $|A| \leq |G|^{1-\epsilon}$ and $|A| \geq C$, where C is an absolute constant.

If $|A| < C$, then (10.2) and Lemma 10.1 imply that $|A_3| \geq |A| + 2 \geq |A|^{1+2/C}$. By the tripling lemma (Lemma 2.2), it follows that $|A \cdot A \cdot A| \geq |A|^{1+\delta}$, δ depending only on C, which is an absolute constant. $\square$

For most applications, it is necessary to supplement the main theorem with a result on very large sets. The result we need was proven by Gowers [Gow] and (in great generality) by Babai, Nikolov and Pyber ([NP], [BNP]).

Lemma 10.3. *Let $G = \mathrm{SL}_3$. Let $K = \mathbb{Z}/p\mathbb{Z}$, p a prime. Let $A \subset G(K)$ be a set of generators of $G(K)$.*

There is an absolute constant $\epsilon > 0$ such that, if $|A| > |G|^{1-\epsilon}$, then

$$A \cdot A \cdot A = G(K). \tag{10.3}$$

Proof. By [NP, Cor. 1 and Prop. 2],

$$A \cdot A \cdot A = G(K) \tag{10.4}$$

provided that $|A| > 2|G|^{1-\frac{1}{3(n+1)}} = 2|G|^{1-1/12}$.

Let $\epsilon = \frac{1}{13}$. Now $|A| > |G(K)|^{1-\frac{1}{13}}$ implies $|A| > 2|G(K)|^{1-\frac{1}{12}}$, provided that p is larger than an absolute constant, and so (10.4) follows. (If p is not larger than an absolute constant, Lemma 10.1 gives us (10.3) easily.) $\square$

Proof of Thm. 1.1. If A does not generate $G(K)$, apply Proposition 8.17. Assume A generates $G(K)$. If $|A| > |G|^{1-\epsilon}$, where ϵ is as in Lemma 10.3, then, by Lemma 10.3, $A \cdot A \cdot A = G(K)$. If $|A| \leq |G|^{1-\epsilon}$, the main theorem shows that $|A \cdot A \cdot A| \gg |A|^{1+\delta}$, where δ is absolute. $\square$

We recall that Corollary 1.2 states that, for any set of generators A of $G = \mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$,

$$\mathrm{diam}(\Gamma(G, A)) \ll (\log |G|)^c, \tag{10.5}$$

where c and the implied constant are absolute.

Proof of Corollary 1.2. Let ϵ be as in Lem. 10.3. Apply Prop. 10.2 to A, then to $A' = A \cdot A \cdot A$, then to $A'' = A' \cdot A' \cdot A'$, etc. After at most

$$k = \log_{(1+\delta)} \frac{\log |G|}{\log |A|} = \frac{\log((\log |G|)/(\log |A|))}{\log(1+\delta)} \leq \frac{\log((\log |G|)/(\log 2))}{\log(1+\delta)} \ll \frac{1}{\delta} \cdot \log \log |G|$$

steps, we shall have obtained a set $A^{(k)}$ with $|A^{(k)}| > |G|^{1-\epsilon}$ elements, where ϵ is as in Lemma 10.3. We now apply Lemma 10.3 to $A^{(k)}$.

We conclude that

$$\underbrace{A \cdot A \cdot A \cdots A}_{\ell \text{ times}} = G,$$

where $\ell = 3^{k+1} = 3 \cdot e^{O\left(\frac{1}{\delta} \cdot \log \log |G|\right)} = 3 \cdot (\log |G|)^{O(1/\delta)}$. Thus, the statement (10.5) holds with $c = O(1/\delta)$, where the implied constant is absolute. Since δ depends only on ϵ, and ϵ is as in Lemma 10.3, i.e., an absolute constant, we see that δ itself is an absolute constant. $\square$

For the sake of making matters self-contained, we could replace Lemma 10.3 with a weaker result that we can prove "by hand", namely, Lemma 10.3 with

$$A_k = G(K) \tag{10.6}$$

instead of (10.3). (Here k is an absolute constant.) This weaker version of Lemma 10.3 can be proven as follows.

Sketch of proof of (10.6). Let U_1, U_2 and T be the algebraic subgroups of G consisting of unipotent upper-triangular, unipotent lower-triangular and diagonal matrices, respectively. By Lemma 7.2, there are many ($\geq p^{3-\epsilon}$) elements of $A^{-1}A$ in $U_1(K)$, many ($\geq p^{3-\epsilon}$) elements of $A^{-1}A$ in $U_2(K)$, and many ($\geq p^{2-\epsilon}$) elements of $A^{-1}A$ in $T(K)$. Assume $\epsilon < 1$. Then the set $A^{-1}A \cap U_1(K)$ is too large not to generate $U_1(K)$ (Lemma 4.15), and thus indeed generates $U_1(K)$. For the same reason, $A^{-1}A \cap U_2(K)$ generates $U_2(K)$.

Let $D_0 = A^{-1}A \cap T(K)$. There are fewer than $3p$ elements of $T(K)$ with repeated eigenvalues; hence, for every $g \in G$, there are fewer than $3p$ elements $g' \in T(K)$ such that $g^{-1}g'$ has repeated eigenvalues. We choose an element $g_1 \in D_0$, then an element $g_2 \in D_0$ such that $g_1^{-1}g_2$ does not have repeated eigenvalues, then a $g_3 \in D_0$ such that neither $g_1^{-1}g_3$ nor $g_2^{-1}g_3$ has repeated eigenvalues, etc. We stop when we cannot find a $g_{k+1} \in D_0$ such that each of $g_1^{-1}g_{k+1}$, $g_2^{-1}g_{k+1}, \dots,$ $g_k^{-1}g_{k+1}$ has distinct eigenvalues. Now, for each $1 \leq j \leq k$, the condition that $g_j^{-1}g_{k+1}$ have distinct eigenvalues rules out fewer than $3p$ possible elements g_{k+1} of D_0. Thus

$$k > \frac{|D_0|}{3p} \gg p^{1-\epsilon},$$

where the constant is absolute. Let $D = \{g_1, g_2, \dots, g_k\}$. Then every element of $D^{-1}D$ has distinct eigenvalues.

Therefore, every element of $D^{-1}D$ acts on $U_1(K)$ and $U_2(K)$ without fixed points (condition (3.1)). We now apply Cor. 3.2, once to the action of $T(K)$ on $U_1(K)$ and once to the action of $T(K)$ on $U_2(K)$. We obtain

$$U_1(K) \subset A_k \quad \text{and} \quad U_2(K) \subset A_k \tag{10.7}$$

(where k is absolute provided that ϵ is less than some fixed constant less than 1).

By direct computation, one can verify that every matrix $g \in G(K)$ that is not upper triangular can be written in the form $g = u_1 \cdot u_2 \cdot u_1'$, where $u_1, u_1' \in U_1(K)$, $u_2 \in U_2(K)$, and every matrix $g \in G(K)$ that is not lower triangular can be written in the form $g = u_2 \cdot u_1 \cdot u_2'$, where $u_1 \in U_1(K)$ and $u_2, u_2' \in U_2(K)$. It can then be easily shown that every $g \in G(K)$ can be written in the form $g = u_1 \cdot u_2 \cdot u_1' \cdot u_2'$, where $u_1, u_1' \in U_1(K)$ and $u_2, u_2' \in U_2(K)$. Thus, by (10.7), we conclude that

$$G(K) \subset A_{4k},$$

as was desired. $\square$

If, for the sake of making the paper relatively self-contained, one were to use (10.6) instead of Lem. 10.3 in the proof of Cor. 1.2, one would obtain $\mathrm{diam}(\Gamma(G, A \cup A^{-1})) \ll (\log |G|)^c$ instead of (10.5). (One could then deduce (10.5) by [Ba, Thm. 1.4].) Neither the proof nor the statement of Thm. 1.1 would require any changes.

10.2. **Work to do: other groups.** It is natural to hope for a broad generalisation. The methods in §5 are very likely to carry over to all semisimple groups of Lie type over arbitrary fields. One can thus arguably hope for a proof of the main theorem with $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$ replaced by $G(K)$, G semisimple of Lie type, K a finite field. (The methods in §5 are such that ϵ would have to depend on the Lie type of G. Some very recent results of

Pyber [P] show that this is a reality, and not just a limitation of the method – a statement such as the main theorem with ϵ independent of the rank of G would be false.)

Results such as those in §3.4 can probably be strengthened at least to the extent needed for SL_n. The main difficulties reside in generalising §6 and §9. As it stands, §6 uses the fact that, for $n = 2, 3$, the conjugacy class of an element $g \in \mathrm{SL}_3(K)$ is given by the values $\chi(g)$ of characters χ of dimension n. This is no longer the case for $n > 3$. There do seem to be somewhat involved ways to avoid this problem by the use of a single character of dimension n (such as the trace).

One of the problems in generalising §8 – which is used in §9 – lies in the fact that SL_{n-1} can have a rather complicated subgroup structure for $n > 3$. It does seem that, if one's goal is simply to prove results on medium-sized sets as in §9 – rather than to study growth in subgroups for its own sake – there are ways to limit oneself to the consideration of *algebraic* subgroups (of bounded degree) of SL_{n-1}, as opposed to *all* subgroups of $\mathrm{SL}_{n-1}(K)$. This does simplify matters. However, the problem remains that growth in some algebraic subgroups of SL_{n-1} may be harder to study than in SL_n itself. For example, right now, we are farther away from understanding growth in SO_{n-1} ($n > 5$) than in SL_n.

Thus, one must either study all groups of Lie type together (since they are all isomorphic to some algebraic subgroup of SL_{n-1} for some n) or find a way to do things so that one needs to examine only those subgroups of SL_{n-1} that are more or less isomorphic to products of copies of SL_m, $m \leq n-1$ (times something they act on). There seems to be a way to carry out the latter plan – and thus arrive at results for SL_n before the available techniques can be successfully modified to work for SO_n – but substantial technical difficulties remain.

Needless to say, what we have just discussed makes sense only if we aim at a statement like the main theorem in the present paper – that is, a statement valid for sets $A \subset G(K)$ that generate $G(K)$. If we do not require that A generate $G(K)$, then, by definition, proving growth in $G(K)$ involves proving growth in all subgroups of $G(K)$, algebraic or not. (There will be some subgroups where growth does not actually happen, namely, solvable groups; they are not the real difficulty.) Thus, for example, proving growth an analogue of Theorem 1.1 for $G(K) = \mathrm{SL}_n(\mathbb{F}_q)$ would involve proving growth in all finite groups (with bounds allowed to depend only on n, where n is the dimension of the smallest faithful representation over $\mathbb{F}_q$ of the finite group in question). This seems to be far away, and will probably be rather cumbersome once it becomes possible: it might have to involve the classification of finite simple groups.

* * *

The main theorem is still true if $\mathbb{Z}/p\mathbb{Z}$ is replaced by $\mathbb{R}$ or $\mathbb{C}$; it is easy to modify the proof slightly to show as much. (Sum-product results over $\mathbb{R}$ and $\mathbb{C}$ are older and stronger than those over finite fields.) However, this would arguably not be the right generalisation to $\mathbb{R}$ or $\mathbb{C}$. What is needed for results on expansion is a statement on convolutions of measures; in the case of $\mathbb{Z}/p\mathbb{Z}$, such statements follow from results such as the main theorem – namely, results on multiplication of sets – but, for infinite fields such as $\mathbb{C}$, one needs to start from stronger results. Over $\mathbb{C}$, one should show that $A \cdot A \cdot A$ not only has more elements than A, but, furthermore, has more elements that are at a certain distance from each other. Bourgain and Gamburd [BG2] showed how to strengthen the proof in

[He] accordingly in the case of SU(2). It remains to be seen how difficult it will be to do the same to the proof in the present paper.

We finish by remarking that a recent result of Breuillard [Br] gives some hope that certain results that depend on the assumption of large girth (such as those of Bourgain and Gamburd) may some day be proven without that assumption.

References

[Ba] L. Babai, On the diameter of Eulerian orientations of a graph, *Proc. 17th Ann. Symp. on Discr. Alg. (SODA '06)*, ACM-SIAM 2006, pp. 822–831.

[BNP] L. Babai, N. Nikolov and L. Pyber, Product growth and mixing in finite groups, *Proceedings of the nineteenth annual ACM-SIAM symposium on discrete algorithms*, SIAM, Philadelphia, PA, USA, pp. 248–257.

[BS] L. Babai and Á. Seress, On the diameter of permutation groups, *European J. Combin.* **13** (1992), pp. 231–243.

[BoG] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, Cambridge University Press, 2007.

[Bor] A. Borel, *Linear algebraic groups*, 2nd. ed., Springer, New York, 1991.

[B] J. Bourgain, Mordell's exponential sum estimate revisited, *J. Amer. Math. Soc.* **18** (2005), pp. 477–493.

[B2] J. Bourgain, Multilinear exponential sums in prime fields under optimal entropy condition on the sources, preprint.

[Bo] J. Bourgain, Sum-product theorems and exponential sum bounds in residue classes for general modulus, *C. R. Math. Acad. Sci. Paris* **344** (2007), no. 6, pp. 349–352.

[BG] J. Bourgain and A. Gamburd, Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$, *Ann. of Math.* **167** (2008), no. 2, pp. 625–642.

[BG2] J. Bourgain and A. Gamburd, On the spectral cap for finitely generated subgroups of SU(2), *Invent. Math.* **171** (2008), no. 1, pp. 83–121.

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

[Br] E. Breuillard, A strong Tits alternative, preprint, arXiv:0804.1395.

[Da] V. I. Danilov, Algebraic varieties and schemes, in V. I. Danilov and V. V. Shokurov, *Algebraic curves, algebraic manifolds and schemes*, *Encyclopaedia of mathematical sciences* **23**, Springer (1998), pp. 167–297.

[Di] Dickson, L. E., *Linear groups, with an exposition of the Galois field theory,* Teubner, Leipzig, 1901.

[Din] O. Dinai, Poly-log diameter bounds for some families of finite groups, *Proc. Amer. Math. Soc.* **134** (2006), pp. 3137–3142 (electronic).

[E] G. Elekes, On the number of sums and products, *Acta Arith.* **81** (1997), pp. 365–367.

[EMO] A. Eskin, S. Mozes, and H. Oh, On uniform exponential growth for linear groups, *Invent. Math.* **160** (2005), pp. 1–30.

[ES] P. Erdős and E. Szemerédi, On sums and products of integers, in *Studies in Pure Mathematics; To the memory of Paul Turán*, P. Erdős, L. Alpar, and G. Halasz (eds.), Akademiai Kiado – Birkhauser Verlag, Budapest, 1983, pp. 213–218.

[Ga] M. Z. Garaev, An explicit sum-product estimate in $\mathbb{F}_p$, preprint, arxiv:math/0702780.

[GK] A. A. Glibichuk and S. V. Konyagin, Additive properties of product sets in fields of prime order, preprint.

[Gow] W. T. Gowers, Quasirandom groups, preprint, arXiv:0710.3877.

[GR] B. Green and I. Z. Ruzsa, Freiman's theorem in an arbitrary abelian group, *J. London Math. Soc.* **75** (2007), pp. 163–175.

[He] H. Helfgott, Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$, *Ann. of Math.* **167** (2008), no. 2, pp. 601–623.

[Hi] R. Hill, personal communication.

[Hum] J. Humphreys, *Linear algebraic groups*, Springer, New York, 1975.

[Ki] O. H. King, The subgroup structure of finite classical groups in terms of geometric configurations, *Surveys in combinatorics 2005*, 29–56, London Math. Soc. Lecture Note Ser. **327**, Cambridge Univ. Press, Cambridge, 2005.

[KL] P. Kleidman and M. Liebeck, *The subgroup structure of the finite classical groups,* London Mathematical Society Lecture Note Series **129**, Cambridge University Press, Cambridge, 1990.

[Ko] S. V. Konyagin, A sum-product estimate in fields of prime order, arXiv:math.NT/03042147.

[LW] S. Lang and A. Weil, Number of points of varieties in finite fields, *Amer. J. Math.* **76** (1954), 819–827.

[Mi] H. Mitchell, Determination of the ordinary and modular ternary linear groups, *Trans. Amer. Math. Soc.* **12** (1911), no. 2, 207–242.

[NP] N. Nikolov and L. Pyber, Product decompositions of quasirandom groups and a Jordan-type theorem, preprint, arXiv:math/0703.5343.

[P] L. Pyber, personal communication.

[SX] P. Sarnak and X. Xue, Bounds for multiplicities of automorphic representations, *Duke Math. J.* **64** (1991), pp. 207–227.

[ST] E. Szemerédi and W. T. Trotter Jr., Extremal problems in discrete geometry, *Combinatorica* **3** (1983), pp. 381–392.

[T] T. Tao, Product set estimates for non-commutative groups, preprint, arxiv:math/0601431.

[T2] T. Tao, The sum-product phenomenon in arbitrary rings, preprint, arxiv:math/0806.2497.

[TV] T. Tao and V. Vu, *Additive Combinatorics*, *Cambridge Studies in Adv. Math.* **105**, Cambridge Univ. Press, Cambridge, 2006.

[Ta] D. Taylor, *The geometry of the classical groups*, Heldermann Verlag, Berlin, 1992.

H. A. Helfgott, School of Mathematics, University of Bristol, Bristol, BS8 1TW, United Kingdom