SIEVE IN DISCRETE GROUPS, ESPECIALLY SPARSE

EMMANUEL KOWALSKI

Abstract. We survey the recent applications and developments of sieve methods related to discrete groups, especially in the case of infinite index subgroups of arithmetic groups.

1. Introduction

Sieve methods appeared in number theory as a tool to try to understand the additive properties of prime numbers, and then evolved over the 20th Century into very sophisticated tools. Not only did they provide extremely strong results concerning the problems most directly relevant to their origin (such as Goldbach's conjecture, the Twin Primes conjecture, or the problem of the existence of infinitely many primes of the form $n^2 + 1$), but they also became tools of crucial important in the solution of many problems which were not so obviously related (examples are the first proof of the Erdös-Kac theorem, and more recently sieve appeared in the progress, and solution, of the Quantum Unique Ergodicity conjecture of Rudnick and Sarnak).

It is only quite recently that sieve methods have been applied to new problems, often obviously related to the historical roots of sieve, which involve complicated infinite discrete groups (of exponential growth) as basic substrate instead of the usual integers. Moreover, both "small" and "large" sieves turn out to be applicable in this context to a wide variety of very appealing questions, some of which are rather surprising. We will attempt to present this story in this survey, following the mini-course at the "Thin groups and super-strong-approximation" workshop. The basic outline is the following: in Section 2, we present a sieve framework that is general enough to describe both the classical examples and those involving discrete groups; in Section 3, we show how to implement a sieve, with emphasis on "small" sieves. In Section 5, we take up the "large" sieve, which we discuss in a fair amount of details since it is only briefly mentioned in [29] and has the potential to be a very useful general tool even outside of number-theoretic contexts. Finally, we conclude with a sampling of problems and further questions in Section 6.

We include a general version of the Erdös-Kac Theorem in the context of affine sieve (Theorem 4.12), which follows easily from the method of Granville and Soundararajan [19] (it generalizes a result of Djanković [7] for Apollonian circle packings.)

Apart from this, the writing will follow fairly closely the notes for the course at MSRI, and in particular there will be relatively few details and no attempts at the greatest known generality. The final section had no parallel in the actual lectures, for reasons of time. More information can be gathered from the author's Bourbaki lecture [29], or from Salehi-Golsefidy's paper in these Proceedings [50], and of course from the original papers. Overall, we have tried to emphasize general principles and some specific applications, rather than to repeat the more comprehensive survey of known results found in [29].

Notation. We recall here some basic notation.

Key words and phrases. Expander graphs, Cayley graphs, sieve methods, prime numbers, thin sets, random walks on groups, large sieve.

– The letters p will always refer to a prime number; for a prime p, we write $\mathbf{F}_p$ for the finite field $\mathbf{Z}/p\mathbf{Z}$. For a set X, $|X|$ is its cardinality, a non-negative integer or $+\infty$.
– The Landau and Vinogradov notation $f = O(g)$ and $f \ll g$ are synonymous, and $f(x) = O(g(x))$ for all $x \in D$ means that there exists an "implied" constant $C \geqslant 0$ (which may be a function of other parameters) such that $|f(x)| \leqslant Cg(x)$ for all $x \in D$. This definition *differs* from that of N. Bourbaki [1, Chap. V] since the latter is of topological nature. We write $f \asymp g$ if $f \ll g$ and $g \ll f$. On the other hand, the notation $f(x) \sim g(x)$ and $f = o(g)$ are used with the asymptotic meaning of loc. cit.

Reference. As a general reference on sieve in general, the best book available today is the masterful work of Friedlander and Iwaniec [10]. Concerning the large sieve, the author's book [28] contains very general results. We also recommend Sarnak's lectures on the affine sieve [53]. Another survey of sieve in discrete groups, with a particular emphasis on small sieves, is the Bourbaki seminar of the author [29], and Salehi-Golsefidy's paper [50] in these Proceedings gives an account of the most general version of the affine sieve, due to him and Sarnak [51].

Acknowledgments. Thanks to A. Kontorovich and A. Salehi-Golsefidy for pointing out imprecisions and minor mistakes in the first drafts of this paper, and to H. Oh for suggestions that helped clarify the text.

2. The setting for sieve in discrete groups

Sieve methods attempt to obtain estimates on the size of sets constructed using local-global and inclusion-exclusion principles. We start by describing a fairly general framework for this type of questions, tailored to applications to discrete groups (there are also other settings of great interest, e.g., concerning the distribution of Frobenius conjugacy classes related to families of algebraic varieties over finite fields, see [28, Ch. 8]).

We will consider a group Γ, viewed as a discrete group, which will usually be finitely generated, and which is given either as a subgroup $\Gamma \subset \mathrm{GL}_r(\mathbf{Z})$ for some $r \geqslant 1$, or more generally is given with a homomorphism

$$\phi : \Gamma \longrightarrow \mathrm{GL}_r(\mathbf{Z}),$$

which may not be injective (and of course is typically not surjective). Here are three examples.

Example 2.1. (1) We can take $\Gamma = \mathbf{Z}$, embedded in $\mathrm{GL}_2(\mathbf{Z})$ for instance, using the map

$$\phi(n) = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

This case is of course the most classical.

(2) Consider a finite symmetric set $S \subset \mathrm{SL}_r(\mathbf{Z})$, and let $\Gamma = \langle S \rangle \subset \mathrm{GL}_r(\mathbf{Z})$. Of particular interest for us is the case when Γ is "large" in the sense that it is Zariski-dense in SL_r. Recall that this means that there exist no polynomial relations among all elements $g \in \Gamma$ except for those which are consequence of the equation $\det(g) = 1$. A concrete example is as follows: for $k \geqslant 1$, let

$$S_k = \left\{ \begin{pmatrix} 1 & \pm k \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm k & 1 \end{pmatrix} \right\}$$

and let $\Gamma^{(k)}$ be the subgroup of $\mathrm{SL}_2(\mathbf{Z})$ generated by S_k. It is well-known that for $k \geqslant 1$, this is a Zariski-dense subgroup of SL_2.

We are especially interested in situations where Γ is nevertheless "small", in the sense that the index of Γ in the arithmetic group $\mathrm{SL}_r(\mathbf{Z})$ is *infinite*. We will call this the *sparse* case (though the terminology *thin* is also commonly used, we will wish to speak later of thin subsets of SL_r, as defined by Serre, and Γ is not thin in this sense).

In the example above, the groups $\Gamma^{(1)} = \mathrm{SL}_2(\mathbf{Z})$ and $\Gamma^{(2)}$ are of finite index in $\mathrm{SL}_2(\mathbf{Z})$ (the latter is the kernel of the reduction map modulo 2), but $\Gamma^{(k)}$ is sparse for all $k \geqslant 3$. In particular, the subgroup $\Gamma^{(3)}$ is sometimes known as the Lubotzky group.

(3) Here is an example where the group Γ is not given as a subgroup of a linear group: for an integer $g \geqslant 1$, let Γ be the mapping class group of a closed surface Σ_g of genus g, and let

$$\phi : \Gamma \longrightarrow \mathrm{Sp}_{2g}(\mathbf{Z}) \subset \mathrm{GL}_{2g}(\mathbf{Z})$$

be the map giving the action of Γ on the first homology group $H_1(\Sigma_g, \mathbf{Z}) \simeq \mathbf{Z}^{2g}$, which is symplectic with respect to the intersection pairing on $H_1(\Sigma_g, \mathbf{Z})$. Here it is known (for instance, through the use of specific generators of Γ mapping to elementary matrices in $\mathrm{Sp}_{2g}(\mathbf{Z})$) that ϕ is surjective. (All facts on mapping class groups that we will use are fairly elementary and are contained in the book of Farb and Margalit [9].)

The next piece of data are surjective maps

$$\pi_p : \Gamma \longrightarrow \Gamma_p$$

where p runs over prime numbers (or possibly over a subset of them) and Γ_p are finite groups. We view each such map as giving "local" information at the prime p, typically by reduction modulo p. Indeed, in all cases in this text, the homomorphism π_p is the composition

$$\Gamma \xrightarrow{\phi} \mathrm{GL}_r(\mathbf{Z}) \longrightarrow \mathrm{GL}_r(\mathbf{F}_p)$$

of ϕ with the reduction map of matrices modulo p, and Γ_p is defined as the image of this map.

Example 2.2. (1) For $\Gamma = \mathbf{Z}$, reduction modulo p is surjective onto $\Gamma_p = \mathbf{Z}/p\mathbf{Z}$ for all primes.

(2) If Γ is Zariski-dense in SL_r, and we use reduction modulo p to define π_p, it is a consequence of general *strong approximation* statements that there exists a finite set of primes $T(\Gamma)$ such that π_p has image *equal* to $\mathrm{SL}_r(\mathbf{F}_p)$ for all $p \notin T(\Gamma)$, and in particular for all primes large enough.[1] For instance, in the case of the subgroups $\Gamma^{(k)} \subset \mathrm{SL}_2(\mathbf{Z})$, this property is visibly valid with

$$T(\Gamma^{(k)}) = \{\text{primes } p \text{ dividing } k\}.$$

We refer to the survey [47] by Rapinchuk in these Proceedings for a general account of Strong Approximation.

(3) For the mapping class group Γ of Σ_g, and ϕ given by the action on homology, the image of reduction modulo p is equal to $\mathrm{Sp}_{2g}(\mathbf{F}_p)$ for all primes p (simply because ϕ is onto, and $\mathrm{Sp}_{2g}(\mathbf{Z})$ surjects to $\mathrm{Sp}_{2g}(\mathbf{F}_p)$ for all p).

We want to combine the maps π_p, corresponding to local information, modulo many primes in order to get "global" results. This clearly only makes sense if using more than a single prime leads to an increase of information. Intuitively, this is the case when the reduction maps π_p, π_q, associated to distinct primes p and q are *independent*: knowing

[1] This is directly related to the fact that SL_r is, as a linear algebraic group, connected and simply connected.

the reduction modulo p of an element of Γ should give no information concerning the reduction modulo q. We therefore make the following assumption on the data:

Assumption 2.3 (Independence)**.** There exists a finite set of primes $T_1(\Gamma)$, sometimes called the Γ-*exceptional primes*, such that for any finite set I of primes $p \notin T_1(\Gamma)$, the simultaneous reduction map

$$\pi_I : \Gamma \longrightarrow \prod_{p \in I} \Gamma_p$$

modulo primes in I is onto.

We will write

$$\Gamma_I = \prod_{p \in I} \Gamma_p, \qquad q_I = \prod_{p \in I} p.$$

Note that q_I is a squarefree integer, coprime with $T_1(\Gamma)$.

Example 2.4. (1) For $\Gamma = \mathbf{Z}$, the Chinese Remainder Theorem shows that for any finite set of primes I, we have

$$\prod_{p \in I} \mathbf{Z}/p_i\mathbf{Z} \simeq \mathbf{Z}/q_I\mathbf{Z},$$

and hence the map π_I above can be identified with reduction modulo q_I. In particular, it is surjective, so that the assumption holds with an empty set of exceptional primes.

(2) If $\Gamma \subset \mathrm{GL}_r(\mathbf{Z})$ has Zariski closure SL_r, then the Independence Assumption holds for the same set of primes $T_1(\Gamma) = T(\Gamma)$ such that π_p is surjective onto $\mathrm{SL}_r(\mathbf{F}_p)$ for $p \notin T(\Gamma)$, simply for group-theoretic reasons: any subgroup of a finite product

$$\prod_{p \in I} \mathrm{SL}_r(\mathbf{F}_p)$$

which surjects to each factor $\mathrm{SL}_r(\mathbf{F}_p)$ is equal to the whole product (this type of result is known as Goursat's Lemma, see, e.g., [6, Prop. 5.1] or as Hall's Lemma [8, Lemma 3.7]). Again a similar property holds if the Zariski closure of Γ is an almost simple, connected, simply-connected algebraic group.

(3) In particular, the Independence Assumption holds with $T_1(\Gamma) = \emptyset$ for the mapping class group of Σ_g acting on the homology of the surface, because Goursat's Lemma applies to the finite groups $\mathrm{Sp}_{2g}(\mathbf{F}_p)$.

(4) The Independence Assumption may fail, for instance in the context of orthogonal groups, when there is a global invariant which can be read off any reduction. The simplest example of such an invariant is the determinant: if $\Gamma \subset \mathrm{GL}_r(\mathbf{Z})$ is not contained in $\mathrm{SL}_r(\mathbf{Z})$, the compatibility condition

$$\det(\pi_p(g)) = \det(g) \in \{\pm 1\} \subset \mathbf{F}_p^\times$$

valid for all p and $g \in \mathrm{GL}_r(\mathbf{Z})$ shows that the image of π_I is always contained in the proper subgroup

$$\{(g_p) \in \Gamma_I \mid \det(g_p) = \det(g_q) \text{ for all } p,\, q \in I\}$$

(identifying all copies of $\{\pm 1\}$). This issue appears, concretely, in the example of the Apollonian group and Apollonian circle packings, since the latter is a subgroup of an indefinite orthogonal group intersecting both cosets of the special orthogonal group, see [11, 13] for a precise analysis of this case, and [12] for a survey.

It should be emphasized that this failure of the Independence Assumption is not dramatic: one can replace Γ by $\Gamma \cap \mathrm{SL}_r$ for instance, or by the other coset of the determinant (with some adaptation since this is not a group).

We can now define the sifted sets $\mathcal{S} \subset \Gamma$ constructed by inclusion-exclusion using local information: given a set $\mathcal{P}$ of primes (usually finite), and subsets

$$\Omega_p \subset \Gamma_p$$

for $p \in \mathcal{P}$, we let

$$\mathcal{S} = \mathcal{S}(\mathcal{P};\Omega) = \{g \in \Gamma \mid \pi_p(g) \notin \Omega_p \text{ for all } p \in \mathcal{P}\} = \bigcap_{p\in\mathcal{P}} (\Gamma - \pi_p^{-1}(\Omega_p)).$$

We want to know something about the size, or maybe more ambitiously the structure, of such sifted sets. In fact, quite often, we wish to study sets which are not exactly of this shape, but are closely related.

Frequently, we have an integer parameter $Q \geqslant 1$, and we take $\mathcal{P} = \{p \leqslant Q\}$, the set of primes up to Q. In that case, we will often denote $\mathcal{S}(Q;\Omega) = \mathcal{S}(\mathcal{P};\Omega)$, and we may even sometimes simplify this to $\mathcal{S}(Q)$ if it is clear that Ω is fixed.

Example 2.5. (1) Let $\Gamma = \mathbf{Z}$, and let $\Omega_p = \{0, -2\} \subset \mathbf{F}_p$ for all primes $p \leqslant Q$, where $Q \geqslant 2$ is some parameter. Taking $\mathcal{P} = \{p \leqslant Q\}$, we have by definition

$$\mathcal{S}(Q) = \mathcal{S}(\mathcal{P};\Omega) = \{n \in \mathbf{Z} \mid \text{neither } n \text{ nor } n+2 \text{ has a prime factor } \leqslant Q\}.$$

In particular, for $N \geqslant 1$, the initial segment $\mathcal{S}(Q) \cap \{1,\ldots,N\}$ contains all "twin primes" n between Q and N, i.e., all primes p with $Q < p \leqslant N$ such that $p+2$ is also prime. Hence an upper-bound on the size of this initial segment will be an upper-bound for the number of twin primes in this range. This is valid independently of the value of Q. Furthermore, if $Q \geqslant \sqrt{N+2}$, we have in fact equality: an integer $n \in \mathcal{S}(\sqrt{N+2}) \cap \{1,\ldots,N\}$ must be prime, as well as $n+2$, since both integers only have prime factors larger than their square-root. More generally, if $Q = N^\beta$ for some $\beta > 0$, we see that $\mathcal{S}(Q) \cap \{1,\ldots,N\}$ contains only integers n such that both n and $n+2$ have less than $1/\beta$ prime factors.

(2) The first example is the prototypical example showing how sieve methods are used to study prime patterns of various type. Bourgain, Gamburd and Sarnak [3] extended this type of questions to discrete subgroups of $\mathrm{GL}_r(\mathbf{Z})$. We present here a special case of what is called the *affine linear sieve* or the *sieve in orbits*. There will be a few other examples below, and we refer to the original paper or to [29] for a more general approach.

We assume for simplicity, as before, that Γ is Zariski-dense in SL_r. Let

$$f : \mathrm{SL}_r(\mathbf{Z}) \longrightarrow \mathbf{Z}$$

be a non-constant polynomial function, for instance the product of the coordinates. We want to study the multiplicative properties of the integers $f(g)$ when g runs over Γ. Consider

$$\Omega_p = \{g \in \Gamma_p \mid f(g) \equiv 0 \pmod{p}\} \subset \Gamma_p \subset \mathrm{SL}_r(\mathbf{F}_p), \tag{2.1}$$

for $p \leqslant Q$. Then $\mathcal{S}(Q;\Omega)$ (recall that this is the sifted set for $\mathcal{P} = \{p \leqslant Q\}$) is the set of $g \in \Gamma$ such that $f(g)$ has no prime factor $\leqslant Q$. In particular, for any $\Delta > 0$, the intersection

$$\mathcal{S}(Q;\Omega) \cap \{g \in \Gamma \mid |f(g)| \leqslant Q^\Delta\}$$

consists of elements where $f(g)$ has $< \Delta$ prime factors. For instance, when f is the product of coordinates, this set contains elements $g \in \Gamma$ where all coordinates have less than Δ prime factors.

(3) For our last example, consider the mapping class group Γ of Σ_g. Let $\mathcal{H}_g$ be a handlebody with boundary Σ_g. For a mapping class $\phi \in \Gamma$, we denote by $\mathcal{M}_\phi$ the compact 3-manifold obtained by Heegaard splitting using $\mathcal{H}_g$ and ϕ, i.e., it is the union of two

copies of $\mathcal{H}_g$ where the boundaries are identified using (a representative of) ϕ (see [8] for more about this construction).

The image J of $H_1(\mathcal{H}_g, \mathbf{Z}) \simeq \mathbf{Z}^g$ in $H_1(\Sigma_g, \mathbf{Z}) \simeq \mathbf{Z}^{2g}$ is a lagrangian subspace (i.e., a subgroup of rank g such that the intersection pairing is identically zero on J). We denote by $J_p \subset \mathbf{F}_p^{2g}$ its reduction modulo p. It follows from algebraic topology that

$$H_1(\mathcal{M}_\phi, \mathbf{Z}) \simeq H_1(\Sigma_g, \mathbf{Z})/\langle J, \phi \cdot J\rangle,$$
$$H_1(\mathcal{M}_\phi, \mathbf{F}_p) \simeq H_1(\mathcal{M}_\phi, \mathbf{Z}) \otimes \mathbf{F}_p \simeq H_1(\Sigma_g, \mathbf{F}_p)/\langle J_p, \phi \cdot J_p\rangle.$$

Thus if we let

$$\Omega_p = \{\gamma \in \mathrm{Sp}_{2g}(\mathbf{F}_p) \mid \gamma \cdot J_p \cap J_p = \emptyset\} = \{\gamma \in \mathrm{Sp}_{2g}(\mathbf{F}_p) \mid \langle J_p, \gamma \cdot J_p\rangle = \mathbf{F}_p^{2g}\}, \tag{2.2}$$

we see that any sifted set $\mathcal{S}(\mathcal{P}; \Omega)$ contains all mapping classes such that $\mathcal{M}_\phi$ has first rational Betti number positive.

We will discuss this example further in Section 5. The reader who is not familiar with sieve is however encouraged to try to find the answer to the following question: What is the great difference that exists between this example and the previous ones?

3. Conditions for sieving

Having defined sifted sets and seen that they contain information of great potential interest, we want to say something about them. The basic question is "How large is a sifted set $\mathcal{S}$?" In order to make this precise, some truncation of $\mathcal{S}$ is needed, since in general this is (or is expected to be) an infinite set. In fact, we saw in the simplest examples (e.g., twin primes) that this truncation (in that case, the consideration of an initial segment of a sifted set) is crucially linked to deriving interesting information from $\mathcal{S}$, as one needs usually to handle a truncation which is correlated with the size of the primes in the set $\mathcal{P}$ defining the sieve conditions.

When sieving in the generality we consider, it is a striking fact that there are different ways to truncate the sifted sets, or indeed to measure subsets of Γ in general (although those we describe below seem, ultimately, to be closely related.) We will speak of "counting methods" below to refer to these various truncation techniques.

Method 1. [Archimedean balls] Fix a norm $\|\cdot\|$ (or some other metric) on the ambient Lie group $\mathrm{GL}_r(\mathbf{R})$ (for instance the operator norm as linear maps on euclidean space, but other choices are possible) and consider

$$\mathcal{S} \cap \{g \in \Gamma \mid \|g\| \leqslant T\}$$

for some parameter $T \geqslant 1$. This is a finite set, and one can try to estimate (from above or below, or both) its cardinality.

Example 3.1. Let Γ be a Zariski-dense subgroup of $\mathrm{SL}_r(\mathbf{Z})$ and f a non-constant polynomial function on $\mathrm{SL}_r(\mathbf{Z})$. For some $d \geqslant 1$, we have

$$|f(g)| \ll \|g\|^d$$

for all $g \in \Gamma$. Hence if we consider the sifted set (2.1) for $Q = T^\beta$, the elements in

$$\mathcal{S}(Q) \cap \{g \in \Gamma \mid \|g\| \leqslant T\}$$

are such that $f(g)$ has at most d/β prime factors.

Counting in archimedean balls in subgroups of arithmetic groups, even without involving sieve, is a delicate matter, especially in the sparse case, which involves deep ideas from spectral theory, harmonic analysis and ergodic theory. We refer to the book of Gorodnik and Nevo [16] for the case of arithmetic groups, and to Oh's surveys [42] and [43] for the sparse case, as well as to the recent paper of Mohamadi and Oh [40] concerning geometrically finite subgroups of isometries of hyperbolic spaces.

Method 2. [Combinatorial balls] Since the groups Γ of interest are most often finitely generated, and indeed sometimes given with a set of generators, one can replace the archimedean metric of the first method with a combinatorial one. Thus if $S = S^{-1}$ is a generating set of Γ, we denote by $\ell_S(g)$ the word-length metric on Γ defined using S. The sets

$$\mathcal{S} \cap \{g \in \Gamma \mid \ell_S(g) \leqslant T\}, \quad \text{or} \quad \mathcal{S} \cap \{g \in \Gamma \mid \ell_S(g) = T\},$$

are again finite, and one can attempt to estimate their size.

This method is particularly interesting when S is a set of free generators of Γ (and their inverses), because one knows precisely the size of the balls for the combinatorial metric in that case. And even if this is not the case, one can often find a subgroup of Γ which is free of rank $\geqslant 2$, and use this subgroup instead of the original Γ (this technique is used in [3]; in that case, the necessary free subgroup is found using the Tits Alternative, a very specific case of which says that if Γ is Zariski-dense in SL_r, then it contains a free subgroup of rank 2.)

Method 3. [Random walks] Instead of trying to reduce to free groups using a subgroup, one can replace Γ by the free group $F(S)$ generated by S and use the obvious homomorphisms

$$\phi : F(S) \longrightarrow \Gamma \longrightarrow \mathrm{GL}_r(\mathbf{Z})$$

and

$$F(S) \longrightarrow \Gamma \longrightarrow \Gamma_p$$

to define sieve problems and sifted sets. An alternative to this description is to use the generating set S and count elements in balls for the word-length metric ℓ_S *with multiplicity*, the multiplicity being the number of representations of $g \in \Gamma$ by a word of given (or bounded) length. This means one measures the size of a set $X \subset \Gamma$ truncated to the sphere of radius $N \geqslant 1$ around the origin by its density

$$\mu_N(X) = \frac{1}{|S|^N} |\{(s_1, \ldots, s_N) \in S^N \mid s_1 \cdots s_N \in X\}|$$

and therefore one tries to measure the density of the sifted set $\mu_N(\mathcal{S})$, as a way of measuring its size within a given ball. If one wishes to measure balls instead of spheres, a simple expedient is to replace S by $S_1 = S \cup \{1\}$ (since the sphere of radius N for ℓ_{S_1} is the ball of radius N for ℓ_S).

It is often convenient to think of this in terms of a random walk: one assumes given, on a probability space Ω, a sequence of independent S-valued random variables ξ_n, and one defines a random walk (γ_n) on Γ by

$$\gamma_0 = 1, \qquad \gamma_{n+1} = \gamma_n \xi_{n+1} \text{ for } n \geqslant 0.$$

If all steps ξ_n are uniformly distributed on S, it follows that

$$\mu_N(X) = \mathbf{P}(\gamma_N \in X),$$

or in other words, the density μ_N is the probability distribution of the N-th step of this random walk.

Example 3.2. The analogue (for Methods 2 and 3) of the argument in Example 3.1 is the following: given a function f as in that example, there exists $C \geqslant 1$ such that, for all $g \in \Gamma$, we have

$$|f(g)| \leqslant C^{\ell_S(g)}$$

(simply because the operator norm of g is submultiplicative and hence grows at most exponentially with the word-length metric). Thus elements which have word-length at most N and belong to a sifted set $\mathcal{S}(Q;\Omega)$ with Q of size A^N, for some $A > 1$, have at most $(\log A)/(\log C)$ prime factors.

Example 3.3 (Dunfield–Thurston random manifolds)**.** This third counting method is the least familiar to classical analytic number theory. This random walk approach was however already considered by Dunfield–Thurston [8] as a way of studying random 3-manifolds, using the Heegard-splitting construction based on mapping class groups as in Example 2.5, (3): given an integer $g \geqslant 1$, they consider a finite generating set S of the mapping class group Γ of Σ_g and the associated random walk (ϕ_n). The 3-manifolds $\mathcal{M}_{\phi_n}$ are then "random 3-manifolds" and some of their properties can be studied using sieve methods.

It is of course useful to have a way of considering these three methods in parallel. This can be done by assuming that one has a sequence (μ_N) of finite measures on Γ, and by considering the problem of estimating $\mu_N(\mathcal{S})$, the measure of the sifted set. In Method 1, these measures would be the uniform counting measure on the intersection of Γ with the balls of radius N in $\mathrm{GL}_r(\mathbf{R})$, in Method 2, the uniform counting measure on the combinatorial ball of radius N, and in Method 3, the probability law of the N-th step of the random walk.

4. Implementing sieve with expanders

We will now explain how all this relates to expanders. The one-line summary is that the expander condition will allow us to apply classical results of sieve theory to settings of discrete groups "with exponential growth" (one might prefer to say, "in non-amenable settings"). We can motivate this convincingly as follows.

The simplest possible sieve problem occurs when the set $\mathcal{P}$ of conditions is restricted to a single prime, and one is asking for

$$\mu_N(\{g \in \Gamma \mid \pi_p(g) = g_0\})$$

for a fixed prime p and a fixed $g_0 \in \Gamma_p$. One sees that, assuming p is fixed, this elementary-looking question concerns the distribution of the image of the sequence $\pi_{p,*}\mu_N$ of measures on the finite group Γ_p. This may well be expected to have a good answer.

Example 4.1. Consider (one last time) the classical case $\Gamma = \mathbf{Z}$. If we truncate by considering initial segments $\{1, \ldots, N\}$, we are asking here about the number of positive integers $\leqslant N$ congruent to a given a modulo p. The proportion of these converges of course to $1/p$, and this is usually so self-evident that one never mentions it specifically. (But, still in classical cases, note that if one starts the sieve from the set of primes instead of $\mathbf{Z}$, then this basic question is resolved by Dirichlet's Theorem on primes in arithmetic progressions, and the uniformity in this question is basically the issue of the Generalized Riemann Hypothesis.)

On intuitive grounds as well as theoretically, one can expect that the "probability" that g reduces modulo p to g_0 should be about $1/|\Gamma_p|$. This amounts to expecting that the probability measures $\pi_{p,*}(\mu)/\mu_N(\Gamma)$ converge weakly to the uniform (Haar) probability measure on this finite group. It is when considering uniformity of such convergence that expander graphs enter the picture.

We can already deduce from this intuition the following heuristic concerning the size of a sifted set $\mathcal{S}(\mathcal{P};\Omega)$: each condition $\pi_p(g) \notin \Omega_p$ should hold with "probability" approximately

$$1 - \frac{|\Omega_p|}{|\Gamma_p|},$$

and these sieving conditions, for distinct primes, should be independent. Hence one may expect that

$$\mu_N(\mathcal{S}(\mathcal{P};\Omega)) \approx \mu_N(\Gamma) \prod_{p\in\mathcal{P}} \Big(1 - \frac{|\Omega_p|}{|\Gamma_p|}\Big) \tag{4.1}$$

(where the symbol $\approx$ here only means that the right-hand side is a first guess for the left-hand side...)

The simplest counting method to explain this is Method 3, where the argument is very transparent. We therefore assume in the remainder of this section that μ_N is the probability law of the N-th step of a random walk on Γ as above.

It is then an immediate corollary of the theory of finite Markov chains (applied to the random walk on the Cayley graph of Γ_p induced by that on Γ) that, if $1 \in S$ (or more generally if this Cayley graph is not bipartite, i.e., if there exists no surjective homomorphism $\Gamma_p \longrightarrow \{\pm 1\}$ such that each generator $s \in S$ maps to -1), we have exponentially-fast convergence to the probability Haar measure. Precisely, let M_p be the Markov operator acting on functions on Γ_p by

$$(M_p\varphi)(x) = \frac{1}{|S|}\sum_{s\in S} \varphi(xs).$$

This operator also acts on functions of mean 0, i.e., on the space $L^2_0(\Gamma_p)$ of functions such that

$$\sum_{g\in\Gamma_p} \varphi(g) = 0,$$

and has real eigenvalues. Let $\varrho_p < 1$ be its spectral radius (it is < 1 because the eigenvalue 1 is removed by restricting to L^2_0, while -1 is not an eigenvalue because the graph is not bipartite). We then have

$$\Big|\mu_N(\pi_p(g) = g_0) - \frac{1}{|\Gamma_p|}\Big| \leqslant \varrho_p^N$$

for all $N \geqslant 1$.

More generally, under the Independence Assumption 2.3, if I is a finite set of primes not in $T_1(\Gamma)$, the same argument applied to the quotient

$$\Gamma \longrightarrow \Gamma_I = \prod_{p\in I} \Gamma_p$$

shows that that for any $(g_p) \in \Gamma_I$, we have

$$\Big|\mu_N\Big(\pi_p(g) = g_p \text{ for } p \in I\Big) - \prod_{p\in I} \frac{1}{|\Gamma_p|}\Big| \leqslant \varrho_I^N \tag{4.2}$$

where $\varrho_I < 1$ is the corresponding spectral radius for Γ_I. It follows by summing over $x = (g_p) \in \Gamma_I$ that we have a quantitative equidistribution

$$\int \varphi((\pi_p(g))_{p\in I}) d\mu_N(g) = \frac{1}{|\Gamma_I|}\sum_{x\in\Gamma_I} \varphi(x) + O(|\Gamma_I|\|\varphi\|_\infty \varrho_I^N) \tag{4.3}$$

(with an absolute implied constant) for any function φ on Γ_I.

In particular, we see that if $\mathcal{P}$ is a fixed set of primes (not in $T_1(\Gamma)$), then as $N \to +\infty$, the basic heuristic (4.1) is valid asymptotically:

$$\lim_{N\to+\infty} \mu_N(\mathcal{S}(\mathcal{P};\Omega)) = \lim_{N\to+\infty} \mathbf{P}(\gamma_N \in \mathcal{S}(\mathcal{P};\Omega)) = \prod_{p\in\mathcal{P}} \Bigl(1 - \frac{|\Omega_p|}{|\Gamma_p|}\Bigr) \tag{4.4}$$

(we will call this a "bounded sieve" statement).

The difficulty (and fun!) of sieve methods is that the sifted sets of most interest are such that the primes involved in $\mathcal{P}$ are *not* fixed as $N \to +\infty$: they are in ranges increasing with the size of the elements being considered (as shown already by the example of the twin primes). It is clear that in order to handle such sifted sets, we need a uniform control of the equidistribution properties modulo primes, and modulo finite sets of primes simultaneously. The best we can hope for is that (4.2) hold with the spectral radius bounded away from one *independently of I*. This is, of course, exactly the conditions under which the family of Cayley graphs of Γ_I with respect to the generators S is a family of (absolute) expander graphs.

Remark 4.2. We have discussed the example of the random walk counting method. It is a fact that analogues of (4.2) hold in all cases where sieve methods have been successfully applied. Moreover, these analogues hold uniformly with respect to I, and ultimately, the source is always equivalent to the expansion property of the Cayley graphs, although the proofs and the equivalence might be much more involved than the transparent argument that exists in the case of random walks.

Example 4.3. The first case beyond the classical examples (or the case of arithmetic groups, where Property (T) or (τ) can be used,[2] although this also had not been done before) where sieve in discrete groups was implemented is due to Bourgain, Gamburd and Sarnak [3], who (based on earlier work of Helfgott [21] and Bourgain–Gamburd [2]) proved that if Γ is a finitely-generated Zariski-dense subgroup of $\mathrm{SL}_2(\mathbf{Z})$ (or even of $\mathrm{SL}_2(\mathcal{O})$, where $\mathcal{O}$ is the ring of integers in a number field), the Cayley graphs of Γ_I, where I runs over finite subsets of $T_1(\Gamma)$, form a family of (absolute) expanders. The problem of generalizing this to SL_r, or to Zariski-dense subgroups of other algebraic groups, was one of the motivations for the recent developments of this result, and of the basic "growth" theorem of Helfgott, to more general groups. We now know an essentially best possible result (see [52, 56], and the surveys [50, 5, 46] of Salehi-Golsefidy, Breuillard and Pyber–Szabó in these Proceedings for introductions to this area):

Theorem 4.4 (Salehi-Golsefidy–Varjú)**.** *Let $\Gamma \subset \mathrm{GL}_r(\mathbf{Z})$ be finitely generated by $S = S^{-1}$, with Zariski-closure $\mathbf{G}$. For p prime, let Γ_p be the image of Γ under reduction modulo p, and for a finite set of primes I, let Γ_I be the image of Γ in*

$$\prod_{p\in I} \Gamma_p,$$

under the simultaneous reduction homomorphism.

If the connected component of the identity in $\mathbf{G}$ is a perfect group, then there exists a finite set of primes $T_1(\Gamma)$ such that the family of Cayley graphs of Γ_I, for $I \cap T_1(\Gamma) = \emptyset$, is an expander family.

We can now describe what is the implication of some classical sieve results in the context of discrete groups. We assume formally the following:

[2] See the works of Gorodnik and Nevo [17] for the best known in this direction.

Assumption 4.5 (Expansion)**.** There exists a finite set of primes $T_2(\Gamma)$ such that Γ satisfies the Independence Assumption 4.5 for primes not in $T_2(\Gamma)$, and furthermore the family of Cayley graphs of Γ_I, for $I \cap T_2(\Gamma) = \emptyset$, is an expander family, i.e., there exists $\varrho < 1$, such that for any finite set I of primes $p \notin T_2(\Gamma)$, the spectral radius for the Markov operator on Γ_I satisfies

$$\varrho_I \leqslant \varrho.$$

By (4.2), this assumption implies that the asymptotic formula

$$\mathbf{P}\Big(\pi_p(\gamma_n) = g_p \text{ for } p \in I\Big) \sim \prod_{p\in I} \frac{1}{|\Gamma_p|} \tag{4.5}$$

holds uniformly for $n \geqslant 1$ and sets I such that $|\Gamma_I| \leqslant \tilde{\varrho}^{-n}$, for any $\tilde{\varrho} > \varrho$. If we assume that

$$|\Gamma_p| \leqslant p^B \tag{4.6}$$

for some fixed $B \geqslant 1$, this means that we can control simultaneously and uniformly all reductions of the N-th step as long as $q_I \leqslant \tilde{\varrho}^{-n/B}$. Note that (4.6) is not very restrictive: it holds (with $B = r^2$) if π_p is just the reduction modulo p on $\mathrm{GL}_r(\mathbf{Z})$, which is the case in all our applications.

The most classical types of sieve are those when the sieving conditions determined by Ω_p hold with probability approximately κ/p, at least on average, were κ is a fixed real number traditionally called the *dimension* of the sieve. Precisely, we say that (Ω_p) is of dimension κ if we have[3]

$$\sum_{p\leqslant X} \frac{|\Omega_p|}{|\Gamma_p|} \log p = \kappa \log X + O(1) \tag{4.7}$$

for $X \geqslant 2$. Note that this is certainly true, by the Prime Number Theorem, if

$$\frac{|\Omega_p|}{|\Gamma_p|} = \frac{\kappa}{p} + O\Big(\frac{1}{p^{1+\delta}}\Big)$$

for some $\delta > 0$ and all p prime.

We then have the following basic result:

Theorem 4.6 (Small sieve in discrete groups)**.** *Let Γ be a discrete group finitely generated by $S = S^{-1}$, given with $\phi : \Gamma \longrightarrow \mathrm{GL}_r(\mathbf{Z})$ and surjective homomorphisms π_p to finite groups Γ_p as above, in particular with (4.6) for some fixed $B \geqslant 1$. Assume that Γ satisfies the Independence Assumption 2.3 and the Expansion Assumption 4.5. Let (γ_n) denote a random walk on Γ using steps from S, and let μ_n denote the probability law of the n-th step. Let $\Omega_p \subset \Gamma_p$ be finite sets such that (4.7) holds for some $\kappa > 0$.*

There exists $A > 0$ such that, for all $n \geqslant 1$, if we let $Q = A^n$ and take $\mathcal{P}$ to be the set of primes $p \leqslant Q$ with $p \notin T_2(\Gamma)$, then we have

$$\mathbf{P}(\gamma_n \in \mathcal{S}(\mathcal{P}; \Omega)) \asymp \frac{1}{n^\kappa}$$

for all N large enough.

This is essentially a direct consequence of the standard Brun-type sieve, building on the Independence and Expansion assumptions. The mechanism is explained in [29], and to avoid repetition, we will not give further details here. We simply add a few remarks. First, this result confirms the heuristic (4.1) as far as the order of magnitude is concerned,

[3] There are other weaker conditions that are enough to allow an efficient sieve, but we refer only to [10, §5.5] for a discussion of these aspects.

i.e., up to multiplicative constants. Indeed, the right-hand side of (4.1) is, in this case, given by

$$\prod_{\substack{p\leqslant A^n\\ p\notin T_2(\Gamma)}} \Bigl(1-\frac{|\Omega_p|}{|\Gamma_p|}\Bigr)\sim \prod_{p\leqslant A^n}\Bigl(1-\frac{\kappa}{p}\Bigr)\asymp n^{-\kappa}$$

as $n\rightarrow +\infty$, by (4.7) and the Mertens Formula (or the Prime Number Theorem.) Secondly, the result is best possible in the sense that one cannot replace the inequalities up to multiplicative constants by an asymptotic formula in this generality (this is also seen from the Mertens Formula and the Prime Number Theorem). Finally, the result is by no means an easy consequence of (4.2) and the uniformity afforded by expansion.

Example 4.7 (Sieve in orbits)**.** We illustrate the above result by deriving, as a corollary, a special case of the sieve in orbits (or affine linear sieve) of [3].

Let Γ be Zariski-dense in $\mathrm{SL}_r(\mathbf{Z})$ with $r\geqslant 2$, and generated by the finite set $S=S^{-1}$. We take for π_p the reduction maps. Let

$$f\,:\,\mathrm{SL}_r(\mathbf{Z})\longrightarrow \mathbf{Z}$$

be a non-constant polynomial map and let $\Omega_p\subset \mathrm{SL}_r(\mathbf{F}_p)$ be the set of zeros of f. Since f is non-constant, the algebraic subvariety Z_f of SL_r defined by the equation $f=0$ is a hypersurface in SL_r. Relatively elementary considerations of algebraic geometry, together with the Lang-Weil estimates for the number of points on algebraic varieties over finite fields, show that we have

$$\frac{|\Omega_p|}{|\Gamma_p|}=\frac{\kappa_p}{p}+O(p^{-3/2}) \tag{4.8}$$

for some $\kappa_p\geqslant 0$ which depends on the splitting of p in the field of definition of the geometrically irreducible components of Z_f (if all geometrically irreducible components of Z_f are defined over $\mathbf{Q}$, then κ_p is the number of these irreducible components, as is well-known; the general case is carefully explained by Salehi-Golsefidy and Sarnak [51, Prop. 15, Cor. 17]). A further application of the Chebotarev density theorem (see [51, Lemma 21]) shows that

$$\sum_{p\leqslant X}\kappa_p=\kappa\pi(X)+O(X/(\log X)^2)$$

where κ is the number of $\mathbf{Q}$-irreducible components of Z_f (if all geometrically irreducible components are defined over $\mathbf{Q}$, we have $\kappa_p=\kappa$ for all but finitely many p).

Example 4.8. (1) Consider the function

$$f\Bigl(\begin{pmatrix} a & b\\ c & d\end{pmatrix}\Bigr)=a^2+d^2$$

on SL_2. For $p\equiv 3\,(\mathrm{mod}\,4)$, we have $\kappa_p=0$ (since $a^2+d^2=0\in\mathbf{F}_p$ implies $a=d=0$ in this case), while for $p\equiv 1\,(\mathrm{mod}\,4)$, we have $\kappa_p=2$, reflecting the fact that Z_f has then, over $\mathbf{F}_p$, the two geometrically irreducible components defined by $a+\varepsilon d=0$ and $a-\varepsilon d=0$, where $\varepsilon^2=-1$. The average of κ_p over p is then equal to $\kappa=1$.

(2) Consider the function

$$f(g_{i,j})=\prod_{i,j} g_{i,j}$$

on SL_r. Then the irreducible components are defined by $g_{i,j}=0$ for a fixed (i,j), and are all defined over $\mathbf{Q}$. Thus we have $\kappa=\kappa_p=r^2$ for all primes p.

Thus all assumptions of Theorem 4.6 hold (the Expansion Assumption coming from Theorem 4.4), and we deduce that there exists a finite set of primes T and $A > 1$ such that if $\mathcal{P}$ is the set of primes not in T and $\leqslant A^n$, we have

$$\mathbf{P}(\gamma_n \in \mathcal{S}(\mathcal{P};\Omega)) \asymp n^{-\kappa}.$$

Using Example 3.2, we therefore deduce:

Theorem 4.9 (Sieve in orbits; Bourgain–Gamburd–Sarnak)**.** *Let Γ and f be as above. There exists $\omega \geqslant 1$ such that the set $\mathcal{O}_f(\omega)$ of all $g \in \Gamma$ such that $f(g)$ has at most ω prime factors satisfies*

$$\mathbf{P}(\gamma_n \in \mathcal{O}_f(\omega)) \asymp n^{-\kappa} \tag{4.9}$$

for n large enough.

One of the insights of Bourgain–Gamburd–Sarnak was that such a statement has a more qualitative corollary which is already very interesting and doesn't require any consideration of a special counting method:

Corollary 4.10. *Let Γ and f be as above. There exists $\omega \geqslant 1$ such that the set $\mathcal{O}_f(\omega)$ is Zariski-dense in SL_r.*

Proof. It is enough to check that if a subset $X \subset \Gamma$ is not Zariski-dense, then a lower-bound

$$\mathbf{P}(\gamma_n \in X) \gg n^{-\kappa}$$

does not hold for any $\kappa > 0$, since $\mathcal{O}_f(\omega) \subset Z$ would then contradict the sieve lower bound (4.9) (note that here (γ_n) is just an auxiliary tool).

Given X, there exists a non-trivial polynomial f such that $X \subset Z_f$ (recall that this is the zero set of f). Then, for any prime p (large enough so that reduction of f modulo p makes sense) the image of X modulo p is contained in the zeros of f modulo p. But using (4.8) and summing (4.5) over the zeros of f modulo p, we have

$$\mathbf{P}(\pi_p(\gamma_n) \in Z_f \,(\mathrm{mod}\, p)) \sim \kappa_p p^{-1}$$

uniformly for $p \leqslant A^n$ for some $A > 1$. Taking p of size A^n, we deduce

$$\mathbf{P}(\gamma_n \in X) \leqslant \mathbf{P}(\pi_p(\gamma_n) \in Z_f \,(\mathrm{mod}\, p)) \ll A^{-n}$$

for n large enough. Thus the probability to be a zero of a given function f is in fact exponentially small for a long walk, and this contradicts the lower bounds for $\mathcal{O}_f(\omega)$. □

In fact, as noted in [29] and as we will see in the next section, this has a natural refinement where Zariski-dense is replaced by "not thin" in the sense of Serre.

Note that Salehi-Golsefidy and Sarnak [51] have extended the basic small sieve statement to much more general groups, not necessarily reductive, using the full power of Theorem 4.4 together with special considerations to handle unipotent groups.

Example 4.11. Theorem 4.6 also applies in the context of Dunfield–Thurston manifolds, as in Example 3.3. Indeed, the Expansion Assumption 4.5 is here a consequence of Property (T) for $\mathrm{Sp}_{2g}(\mathbf{Z})$. As observed in [29], a consequence of Theorem 4.6 which is similar in spirit to the affine linear sieve is that there exists $\omega \geqslant 1$ such that

$$\mathbf{P}(H_1(\mathcal{M}_{\phi_n}, \mathbf{Z}) \text{ is finite and has order divisible by} \leqslant \omega \text{ primes}) \asymp n^{-1}$$

for n large enough. (We recall that the genus g defining the Heegard splitting is fixed).

One can certainly use the sieve setting for many other purposes. As one further example, we show how the method of Granville and Soundararajan [19, Prop. 3] gives a version of the Erdös-Kac Theorem for discrete groups. For simplicity, we only state the result for the affine sieve, and give one further example afterwards (a version of this for curvatures of Apollonian circle packings was proved by Djanković [7]).

Theorem 4.12 (Erdös-Kac central limit theorem for affine sieve)**.** *Assume that* $\Gamma \subset \mathrm{SL}_r(\mathbf{Z})$ *is Zariski-dense in* SL_r *and* f *is a non-constant polynomial function satisfying the assumptions of Theorem 4.9. For a random walk* (γ_n) *on* Γ*, let* $\omega_f(\gamma_n) = \omega(f(\gamma_n))$ *if* $f(\gamma_n) \neq 0$*, and* $\omega_f(\gamma_n) = 0$ *otherwise. Then the random variables*

$$\frac{\omega_f(\gamma_n) - \kappa \log n}{\sqrt{\kappa \log n}}$$

converge in law to the standard normal random variable as $n \to +\infty$*.*

Proof. We proceed exactly as in [19], leaving some details to the reader. This uses the method of moments to prove convergence in law to the normal distribution: classical probability results imply that it is enough to prove that for all integers $k \geqslant 0$, we have

$$\mathbf{E}\Bigl(\Bigl(\frac{\omega_f(\gamma_n) - \kappa \log n}{\sqrt{\kappa \log n}}\Bigr)^k\Bigr) \longrightarrow c_k$$

as $n \to +\infty$, where $c_k = \mathbf{E}(\mathcal{N}(0,1)^k)$ is the k-th moment of a standard normal random variable.

We first deal with the possibility that $f(\gamma_n) \neq 0$. By bounding

$$\mathbf{P}(f(\gamma_n) = 0) \leqslant \mathbf{P}(f(\pi_p(\gamma)) = 0)$$

for any prime p large enough, and arguing as in the proof of Corollary 4.10, we get

$$\mathbf{P}(f(\gamma_n) = 0) \ll c^{-n}$$

for some $c > 1$. Thus the expectation above, restricted to the set $f(\gamma_n) = 0$, is

$$\ll (\kappa \log n)^{k/2} c^{-n} \longrightarrow 0$$

as $n \to +\infty$.

Below we use the notation $\tilde{\mathbf{E}}$ to denote expectation restricted to $f(\gamma_n) \neq 0$. We fix some integer $k \geqslant 0$, and fix some auxiliary $A > 1$. We will compare

$$M_k = \tilde{\mathbf{E}}((\omega_f(\gamma_n) - \kappa \log n)^k)$$

with the moment of "truncated" count of primes dividing $f(\gamma_n)$ defined by

$$N_k = \tilde{\mathbf{E}}\Bigl(\Bigl(\sum_{\substack{p \leqslant A^n \\ \pi_p(\gamma_n) \in \Omega_p}} 1 - \sum_{p \leqslant A^n} \delta_p\Bigr)^k\Bigr),$$

where

$$\delta_p = \frac{|\Omega_p|}{|\Gamma_p|},$$

and then estimate asymptotically this second moment when $A > 1$ is small enough with respect to k.

For the first step, we note that when $f(\gamma_n) \neq 0$, we have

$$\omega_f(\gamma_n) - \kappa \log n = A_1 + A_2 + A_3$$

where

$$A_1 = \sum_{\substack{p \leqslant A^n \\ \pi_p(\gamma_n) \in \Omega_p}} 1 - \sum_{p \leqslant A^n} \delta_p$$

$$A_2 = \omega(f(\gamma_n)) - \sum_{\substack{p \leqslant A^n \\ \pi_p(\gamma_n) \in \Omega_p}} 1$$

$$A_3 = \sum_{p \leqslant A^n} \delta_p - \kappa \log n$$

If $C > 1$ is such that $|f(g)| \leqslant C^{\ell_S(g)}$, then we get

$$0 \leqslant A_2 \leqslant \frac{\log C}{\log A},$$

while, by (4.7), we have

$$A_3 = \sum_{p \leqslant A^n} \delta_p - \kappa \log n = \sum_{p \leqslant A^n} \frac{|\Omega_p|}{|\Gamma_p|} - \kappa \log n = O(1),$$

so that $A_2 + A_3$ is uniformly bounded for a fixed choice of A. Using the multinomial theorem, it follows that

$$M_k = N_k + O(\max_{j \leqslant k-1} \tilde{N}_j),$$

where

$$\tilde{N}_j = \tilde{\mathbf{E}}\Big(\Big| \sum_{\substack{p \leqslant A^n \\ \pi_p(\gamma_n) \in \Omega_p}} 1 - \sum_{p \leqslant A^n} \delta_p \Big|^k \Big).$$

We have $\tilde{N}_j = N_j$ is even and if j is odd, we get

$$\tilde{N}_j \leqslant \sqrt{\tilde{N}_{j-1} \tilde{N}_{j+1}} = \sqrt{N_{j-1} N_{j+1}}$$

by the Cauchy-Schwarz inequality, showing that good understanding of N_j for $j \leqslant k$ will suffice to estimate M_k.

For the second step, we write

$$X_p = \mathbf{1}_{\pi_p(\cdot) \in \Omega_p} - \delta_p$$

for $p \leqslant A^n$, sum over p, and open the k-th power defining N_k. Note that $|X_p| \leqslant 1$. Exchanging the multiple sum over primes and the expectation, we get

$$N_k = \sum \cdots \sum_{p_1, \ldots p_k \leqslant A^n} \tilde{\mathbf{E}}\Big(\prod_{j=1}^{k} X_{p_j} \Big).$$

For any fixed $(p_1, \ldots, p_k)$, we note that

$$\tilde{\mathbf{E}}\Big(\prod_{j=1}^{k} X_{p_j} \Big) = \mathbf{E}\Big(\prod_{j=1}^{k} X_{p_j} \Big) - \mathbf{E}\Big(\prod_{j=1}^{k} X_{p_j} \mathbf{1}_{f(\gamma_n)=0} \Big)$$

and the second term is bounded by $\mathbf{P}(f(\gamma_n) = 0) \ll c^{-n}$ since $0 \leqslant X_{p_j} \leqslant 1$. Thus the total change in replacing $\tilde{\mathbf{E}}$ by $\mathbf{E}$ in the formula above for N_k is $\ll A^{nk} c^{-n}$, which is negligible if A is chosen small enough.

Having written

$$N_k = \sum_{p_1,\dots p_k \leqslant A^n} \cdots \sum \tilde{\mathbf{E}}\Bigl(\prod_{j=1}^{k} X_{p_j}\Bigr) = \sum_{p_1,\dots p_k \leqslant A^n} \cdots \sum \mathbf{E}\Bigl(\prod_{j=1}^{k} X_{p_j}\Bigr) + O(A^{nk} c^{-n}),$$

we can apply the equidistribution (4.3) to each expectation term, obtaining a main term which we will discuss in a moment and a total error term E which is bounded by

$$E \ll A^{nk(1+B)} \varrho^n + A^{nk} c^{-n}$$

(where B is as in (4.6)). Therefore E tends to 0 as $n \to +\infty$ if A is chosen small enough (in terms of k), which we assume to be done.

There remains the main term. However, the latter is, by the Independence Assumption 2.3 and by retracing our steps, almost tautologically the same as

$$\mathbf{E}\Bigl(\Bigl(\sum_{p \leqslant A^n} Y_p - \delta_p\Bigr)^k\Bigr)$$

where the (Y_p) are independent Bernoulli random variables with expectation $\mathbf{E}(Y_p) = \delta_p = |\Omega_p|/|\Gamma_p|$. It is a basic probabilistic fact that the sum

$$\sum_{p \leqslant A^n} Y_p$$

satisfies the Central Limit Theorem, with mean $\kappa \log n$ and variance $\kappa \log n$ (because of (4.7) again). Therefore this sum has the right k-th moment for all $k \geqslant 0$, and this easily concludes the proof (or see [19] for a direct analysis of this type of main terms to see the combinatorics from which the normal moments explicitly appear). ☐

Example 4.13 (Erdös-Kac theorem for random 3-manifolds)**.** It is clear that the argument can be applied in greater generality (including for other counting methods, provided the analogue of quantitative and suitably uniform equidistribution is known). For instance, one sees that, for Dunfield–Thurston random 3-manifolds, the number $\omega(\mathcal{M}_{\phi_n})$ of primes p such that $H_1(\mathcal{M}_{\phi_n}, \mathbf{F}_p) \neq 0$ is such that

$$\frac{\omega(\mathcal{M}_{\phi_n}) - \log n}{\sqrt{\log n}}$$

converges to a standard normal random variable, with the convention $\omega(\mathcal{M}_{\phi_n}) = 0$ if $H_1(\mathcal{M}_{\phi_n}, \mathbf{Q}) \neq 0$.

5. The large sieve

We begin with a motivating example.

Example 5.1. Consider Corollary 4.10. Although the Zariski topology contains a fair amount of information (see [3] for examples of distinction it makes concerning the sieve in orbits), it is not very arithmetic. By itself, the fact that $\mathcal{O}_f(\omega)$ is Zariski-dense in SL_r does not exclude the possibility that this set is contained, for instance, in the subset X of $\mathrm{SL}_r(\mathbf{Z})$ of matrices where the top-left coefficient is a perfect square (since X is Zariski-dense in SL_r.) It is natural to try to study this and similar possibilities. The following definition is relevant (see [55, Chapter 3]):

Definition 5.2 (Thin set)**.** A subset $X \subset \mathrm{SL}_r(\mathbf{Q})$ is *thin* if there exists an algebraic variety $W/\mathbf{Q}$ with $\dim(W) \leqslant r^2 - 1$ and a morphism $W \xrightarrow{\pi} \mathrm{SL}_r$ such that (1) π has no rational section; (2) we have $X \subset \pi(W(\mathbf{Q}))$.

Example 5.3. (1) The set $X = \{g \in \mathrm{SL}_r(\mathbf{Q}) \mid g_{1,1} \text{ is a square}\}$ is thin. Indeed, we have a $\mathbf{Q}$-morphism

$$\pi : \mathbf{A}^{r^2} \longrightarrow \mathbf{A}^{r^2}$$

mapping $(g_{i,j})$ to the matrix $(h_{i,j})$ with $h_{1,1} = g_{1,1}^2$ and all other coordinates unchanged. The pull-back of this morphism to $\mathrm{SL}_r \subset \mathbf{A}^{r^2}$ gives a morphism

$$\pi : W \longrightarrow \mathrm{SL}_r,$$

where

$$W = \{g \in \mathbf{A}^{r^2} \mid \det(\pi(g)) = 1\},$$

for which we have $X \subset \pi(W(\mathbf{Q}))$ by construction (and $\dim W \leqslant \dim \mathrm{SL}_r$ is clear since π has finite fibers.)

(2) A subset X which is not Zariski-dense is thin.

We wish to prove:

Proposition 5.4. *Let Γ and f be as in Corollary 4.10. Then there exists $\omega \geqslant 1$ such that $\mathcal{O}_f(\omega)$ is not thin in SL_r.*

The natural idea to prove this is to prove that if X is a thin set, then for a random walk on Γ, the probability

$$\mathbf{P}(\gamma_n \in X)$$

is too small to be compatible with (4.9). For this, we observe, as in the proof of the Zariski-density, that if $X \subset \pi(W(\mathbf{Q}))$ for some

$$\pi : W \longrightarrow \mathrm{SL}_r$$

as in the definition, we have

$$\pi_p(X) \subset \pi(W(\mathbf{F}_p)),$$

for all primes p large enough (such that W and π make sense modulo p). Hence if $g \in X$, we have

$$\pi_p(g) \notin \Omega_p = \mathrm{SL}_r(\mathbf{F}_p) - \pi(W(\mathbf{F}_p)),$$

for all p large enough. This implies a sieve upper bound

$$X \subset \mathcal{S}(\mathcal{P}; \Omega),$$

where $\mathcal{P}$ contains all but finitely many primes.

However, the size of Ω_p is typically much larger than the number of points of an algebraic variety, as one can guess by just looking at the example of squares in $\mathbf{Q}$, where the image modulo p contains roughly half of all residue classes. Indeed, in general we have:

Lemma 5.5. *Let $\pi : W \longrightarrow \mathrm{SL}_r$ be a $\mathbf{Q}$-rational morphism with $\dim W \leqslant r^2 - 1$ and with no $\mathbf{Q}$-rational section. There exists $\delta < 1$ such that, for p large enough, we have*

$$\frac{|\pi(W(\mathbf{F}_p))|}{|\,\mathrm{SL}_r(\mathbf{F}_p)|} \leqslant \delta.$$

For the proof, see e.g. [55, Th. 3.6.2].

Example 5.6 (Homology of Dunfield–Thurston random manifolds)**.** We consider the situation of Example 2.5, (3). Here we found sifting conditions Ω_p defined in (2.2) such that, if $\mathcal{M}_{\phi_n}$ denotes the manifold obtained from the n-th step of a random walk on the mapping class group Γ (as in Example 3.3), we have

$$\mathbf{P}(H_1(\mathcal{M}_{\phi_n}, \mathbf{Q}) \neq 0) \leqslant \mathbf{P}(\phi_n \in \mathcal{S}(Q, \Omega_p))$$

for any $Q \geqslant 1$, where Q refers to using all primes $p \leqslant Q$ as sieve conditions. It is an interesting computation to show that

$$\frac{|\Omega_p|}{|\Gamma_p|} = 1 - \prod_{j=1}^{g} \frac{1}{1+p^{-j}}$$

(see [8, Th. 8.4]) so that, for fixed g, there exists $\delta_g > 0$ for which

$$\frac{|\Omega_p|}{|\Gamma_p|} \geqslant \delta_g$$

for all p.

We now revert to the general setting of a discrete group Γ with local information $\pi_p : \Gamma \longrightarrow \Gamma_p$. We have found above two natural instances of *large sieves*, a terminology which refers to sieving problems where the sets Ω_p are "large", something which most commonly means that they contain a positive proportion of Γ_p: for some $\delta > 0$, we have

$$\frac{|\Omega_p|}{|\Gamma_p|} \geqslant \delta > 0 \tag{5.1}$$

for all $p \in \mathcal{P}$. This is to be compared with the "small" sieve assumption (4.7), and this leads to an interesting remark (answering the question to the reader at the end of Example 2.5, (3)): the primes occur explicitly on both sides of (4.7), but as far as the left-hand side is concerned, they are just indices that could be replaced with any other countable set. However, on the right-hand side, the actual size of primes (and hence their distribution) is involved. This feature disappears in (5.1). This suggests that the "large" sieve could be of interest in much wider contexts outside of number theory. This is indeed the case, as was shown already partly in the book [28], and even more convincingly in the recent works of Lubotzky, Meiri and Rosenzweig that we will discuss, some of which prove general algebraic statements about linear groups using some forms of sieve methods.

To present the large sieve in the context of discrete groups, we will use here the very simple version from the paper [33] of Lubotzky and Meiri, adapted to our setting.

Theorem 5.7 (Large sieve)**.** *Let Γ be a group generated by a finite symmetric set S with $1 \in S$. Let $\Gamma \longrightarrow \Gamma_p$ be surjective homomorphisms onto finite groups for $p \geqslant p_0$. Assume that:*

(1) *For any $p \neq q$ primes $\geqslant p_0$, the induced homomorphisms*

$$\Gamma \longrightarrow \Gamma_p \times \Gamma_q = \Gamma_{p,q} \tag{5.2}$$

are onto.

(2) *The family of Cayley graphs of $\Gamma_{p,q}$ and Γ_p with respect to S is an expander family, for $p, q \geqslant p_0$.*

(3) *For some $B \geqslant 1$ we have*

$$|\Gamma_p| \leqslant p^B.$$

Let $\Omega_p \subset \Gamma_p$ be such that

$$\frac{|\Omega_p|}{|\Gamma_p|} \geqslant \delta \tag{5.3}$$

for some $\delta > 0$ independent of p.

Then there exists $A > 1$ and $c > 1$ such that for $Q = A^n$, we have

$$\mathbf{P}(\gamma_n \in \mathcal{S}(Q;\Omega)) \ll c^{-n}$$

for n large enough, where the sieving is done using primes $p_0 \leqslant p \leqslant Q$.

Note how the assumptions concerning the group and the Γ_p are slightly weaker versions of those used for the small sieve in Theorem 4.6, since expansion and independence is only required for pairs of primes instead of all squarefree integers. Thus this version of the large sieve applies whenever Theorem 4.6 is applicable.

In particular, in view of the example at the beginning of this section, we see that this theorem proves Proposition 5.4. Similarly, for the Dunfield–Thurston random manifolds of Example 5.6, this implies the following:

Proposition 5.8. *Let $g \geqslant 1$ be an integer, and let (ϕ_n) be a random walk on the mapping class group Γ of Σ_g associated to a finite generating set S. Then there exists $c > 1$ such that*

$$\mathbf{P}(H_1(\mathcal{M}_{\phi_n}, \mathbf{Q}) \neq 0) \ll c^{-n}$$

for $n \geqslant 1$.

The fact that the probability tends to zero was already proved by Dunfield and Thurston (see [8, Cor. 8.5]), and the exponential decay was obtained in [28, Prop. 7.19].

Remark 5.9. It would be unreasonable to expect lower-bounds for the size of sifted sets in the large sieve situation, unless the set $\mathcal{P}$ determining the sieving conditions is extremely small (so the situation essentially reverts to a bounded sieve (4.4)). Indeed, if we consider integers and sieve by removing the non-square residue classes modulo p for all $p \in \mathcal{P}$, which is certainly a large sieve, the right-hand side of the heuristic size of the remaining set is $(1/2)^{|\mathcal{P}|}$. If $\mathcal{P}$ is the set of primes $\leqslant Q$, then this is much smaller than the number of squares in $\{1, \ldots, N\}$, which certainly remain after the sieve, if $Q = N^{\varepsilon}$ for any fixed $\varepsilon > 0$. (See [20] for a discussion of the fascinating question of the possibility of an "inverse" large sieve statement for integers.)

We adapt the simple proof of Theorem 5.7 in [33] (due to R. Peled; it is reminiscent of some classical arguments going back to Rényi and Turán, see [28, Prop. 2.15].)

Proof. For a fixed n, let X_p denote the Bernoulli random variable equal to 1 when $\pi_p(\gamma_n) \in \Omega_p$ and 0 otherwise, and let

$$X = \sum_{p_0 \leqslant p \leqslant Q} X_p.$$

We see that $\gamma_n \in \mathcal{S}(Q; \Omega)$ is tautologically equivalent to the condition $X = 0$. We can compute easily the expectation of X, namely

$$\mathbf{E}(X) = \sum_p \mathbf{P}(\pi_p(\gamma_n) \in \Omega_p),$$

which, by Expansion for (Γ_p), satisfies

$$\mathbf{E}(X) = \sum_p \frac{|\Omega_p|}{|\Gamma_p|} + O(Q^{1+B} \varrho^n),$$

where $\varrho < 1$ is an upper-bound for the spectral radius of the expansion of the Cayley graphs. If $Q^{1+B} \ll \varrho^n$, this gives

$$\mathbf{E}(X) \gg \pi(Q) \gg \frac{Q}{\log Q}$$

using the large sieve assumption on the size of Ω_p.

We will now use the Chebychev inequality

$$\mathbf{P}(\gamma_n \in \mathcal{S}(Q; \Omega)) = \mathbf{P}(X = 0) \leqslant \frac{\mathbf{V}(X)}{\mathbf{E}(X)^2},$$

where $\mathbf{V}(X)$ is the variance of X. We compute

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \sum_{p,q} W(p,q)$$

by expanding the square, where

$$\begin{aligned} W(p,q) &= \mathbf{E}(X_p X_q) - \mathbf{E}(X_p)\,\mathbf{E}(X_q) \\ &= \mathbf{P}(\pi_p(\gamma_n) \in \Omega_p \text{ and } \pi_q(\gamma_n) \in \Omega_q) - \mathbf{P}(\pi_p(\gamma_n) \in \Omega_p)\,\mathbf{P}(\pi_q(\gamma_n) \in \Omega_q) \end{aligned}$$

(a measure of the correlation between two primes). We isolate the diagonal terms where $p = q$, for which we use the trivial bound $|W(p,p)| \leqslant 1$, and obtain

$$\mathbf{V}(X) \leqslant Q + Q^2 \max_{p \neq q} |W(p,q)|.$$

Finally, to estimate $W(p,q)$ when $p \neq q$, we can apply the assumption (5.2) and the expansion of the Cayley graphs of $\Gamma_{p,q}$ and Γ_p: we have

$$\mathbf{P}(\pi_p(\gamma_n) \in \Omega_p \text{ and } \pi_q(\gamma_n) \in \Omega_q) = \frac{|\Omega_p||\Omega_q|}{|\Gamma_p||\Gamma_q|} + O(Q^{2B}\varrho^n)$$

while, by the same argument used for computing the expectation, we have

$$\mathbf{P}(\pi_p(\gamma_n) \in \Omega_p)\,\mathbf{P}(\pi_q(\gamma_n) \in \Omega_q) = \frac{|\Omega_p||\Omega_q|}{|\Gamma_p||\Gamma_q|} + O(Q^{B}\varrho^n).$$

The main terms cancel, and therefore

$$Q^2 \max_{p \neq q} |W(p,q)| \ll Q^{2+2B}\varrho^n.$$

Take Q as large as possible so that $Q^{2+2B}\varrho^n < 1$, so that $Q \geqslant A^n$ for some $A > 1$. Then the Chebychev inequality gives

$$\mathbf{P}(\gamma_n \in \mathcal{S}(Q;\Omega)) \ll \frac{(Q + Q^{2+2B}\varrho^n)(\log Q)^2}{Q^2} \ll \frac{(\log Q)^2}{Q}$$

which is of exponential decay in terms of n. □

Remark 5.10. (1) Clearly, one can restrict the large sieve assumption (5.3) to a subset of primes with positive natural density (e.g., some arithmetic progression) without changing the conclusion, and this is often useful.

(2) This very simple proof is well-suited to situations where precise information on the expansion constant of the relevant Cayley graphs is missing (as is most often the case). When such information is available, one gets from this argument an explicit constant $c > 1$, and one may wish to get it as large as possible. For this, one can use rather more precise inequalities, as discussed extensively in [28].

The point of the large sieve is really the exponential decay it provides. If one is interested in a statement of qualitative decay

$$\lim_{n \to +\infty} \mathbf{P}(\gamma_n \in X) = 0 \tag{5.4}$$

for a subset $X \subset \Gamma$ such that

$$X \subset \mathcal{S}(Q;\Omega)$$

for *all* Q large enough, where the Ω_p satisfy (5.3), then one can more easily apply the bounded sieve (4.4) to a finite set I, getting

$$\limsup_{n \to +\infty} \mathbf{P}(\gamma_n \in X) \leqslant \lim_{n \to +\infty} \mathbf{P}(\pi_p(\gamma_n) \notin \Omega_p \text{ for } p \in I) = \prod_{p \in I}\Big(1 - \frac{|\Omega_p|}{|\Gamma_p|}\Big) \leqslant (1-\delta)^{|I|}.$$

Then, letting $|I| \to +\infty$, we obtain (5.4). As an example, note that this qualitative decay is *not* sufficient to prove Proposition 5.4.

Lubotzky and Meiri introduce the following convenient definition:

Definition 5.11 (Exponentially small sets)**.** Let Γ be a finitely generated group. A subset $X \subset \Gamma$ is *exponentially small* if, for any finite symmetric generating set S containing 1, and with (γ_n) the corresponding random walk on Γ, there exists a constant $c_S > 1$ such that

$$\mathbf{P}(\gamma_n \in X) \ll c_S^{-n}$$

for $n \geqslant 1$.

Remark 5.12. Thus, we can summarize part of our previous discussion by stating that if X is a thin subset of $\mathrm{SL}_r(\mathbf{Q})$, and Γ is a finitely generated Zariski-dense subgroup of $\mathrm{SL}_r(\mathbf{Z})$, then $X \cap \Gamma$ is exponentially small in Γ, and by saying that the set of mapping classes (in a fixed mapping class group Γ of genus $g \geqslant 1$) for which the corresponding manifold obtained by Heegaard splitting has positive first rational Betti number is exponentially small.

The first inkling of the large sieve in non-abelian discrete groups is found in applications of the qualitative argument above by Dunfield–Thurston [8] and Rivin [48, 49] in geometric contexts (the second paper [49] of Rivin was the first to obtain exponential decay, though its publication was delayed by a journal with overly long backlog; we thank I. Rivin for clarifying the priority in publication here). We illustrate further the large sieve with an example from the second, and then discuss briefly two other applications.

Example 5.13 (Pseudo-Anosov elements of the mapping class group)**.** Let $g \geqslant 1$ be given and let Γ be the mapping class group of Σ_g. Thurston's celebrated theory classifies the elements $\gamma \in \Gamma$ as (1) reducible; (2) finite-order; or (3) pseudo-Anosov. To quantify the feeling that "most" elements are of the third type, Rivin used a criterion based on the action of γ on $H_1(\Sigma_g, \mathbf{Z})$, which says that *if* (but not only if) the characteristic polynomial P_γ of this action is P_γ is irreducible, and satisfies further easy conditions, then γ is pseudo-Anosov. One then notes that if P_γ is reducible, then so is its reduction modulo any prime, so $\pi_p(\gamma)$ is not in the subset Ω_p of elements of $\mathrm{Sp}_{2g}(\mathbf{F}_p)$ for which the characteristic polynomial is irreducible. A computation that goes back to Chavdarov [6, §3] shows that, for some $\delta > 0$, we have

$$\frac{|\Omega_p|}{|\mathrm{Sp}_{2g}(\mathbf{F}_p)|} \geqslant \delta > 0$$

for all p, and hence the large sieve applies. A simple further argument deals with the other necessary conditions in the pseudo-Anosov criterion, and one concludes that the set of non-pseudo-Anosov elements is exponentially small in Γ.

It should be said, however, that this proof is to some extent unsatisfactory, because it doesn't use the deeper structural and dynamical properties of pseudo-Anosov elements. For instance, using the action on homology means that one cannot argue similarly for subgroups $\tilde{\Gamma} \subset \Gamma$ for which the action on homology is small, especially subgroups of the Torelli group, which is defined precisely as the kernel of the homomorphism

$$\Gamma \longrightarrow \mathrm{Sp}_{2g}(\mathbf{Z})$$

giving this action.

However, Maher [37, 38] has shown, using more geometric methods, that non-pseudo-Anosov elements are exponentially small in any subgroup of Γ, except those for which this property is false for obvious reasons, and his work applies in particular to the Torelli subgroup.

On the other hand, Lubotzky–Meiri [34] and Malestein–Souto [39] (independently) have recently found proofs that non-pseudo-Anosov elements are exponentially small in the Torelli group using ideas similar to those above.

Example 5.14 (Powers in linear groups)**.** In [33], Lubotzky and Meiri prove the following statement using the large sieve. The reader should note that this is, on the face of it, a purely algebraic property of finitely generated linear groups.

Theorem 5.15 (Lubotzky–Meiri)**.** *Let* Γ *be a finitely generated subgroup of* $\mathrm{GL}_r(\mathbf{C})$ *for some* $r \geqslant 2$*. If* Γ *is not virtually solvable,*[4] *then the set* X *of proper powers, i.e., the set of those* $g \in \Gamma$ *such that there exists* $k \geqslant 2$ *and* $h \in \Gamma$ *with* $g = h^k$*, is exponentially small in* Γ*.*

This strenghtens considerably some earlier work of a Hrushovski, Kropholler, Lubotzky and Shalev [22]. The proof is also very instructive, in particular by showing how sieve should be considered as a *tool* among others: here, one can use the large sieve to control elements which are k-th powers for a *fixed* $k \geqslant 2$, but taking the union over all $k \geqslant 2$ cannot be done with sieve alone. So Lubotzky and Meiri use other tools to deal with large values of k, in that case based on ideas related to the work of Lubotzky, Mozes and Raghunathan comparing archimedean and word-length metrics [36].

Example 5.16 (Typical Galois groups of characteristic polynomials)**.** Our last example has been studied by Rivin [48], Jouve–Kowalski–Zywina [23], Gorodnik–Nevo [18] and most recently Lubotzky–Rosenzweig [35], who were the first to explicitly consider the case of sparse subgroups. However, the underlying idea of probabilistic Galois theory goes back to versions of Hilbert's irreducibility theorem, and especially to Gallagher's introduction of the large sieve in this context [15]. (There are also relations with works of Prasad and Rapinchuk [44, 45].)

In the (most general) version of Lubotzky–Rosenzweig, one considers a finitely generated field $K \subset \mathbf{C}$ and a finitely generated subgroup $\Gamma \subset \mathrm{GL}_r(K)$ for some $r \geqslant 2$. The basic question is: what is the "typical" behavior of the splitting field of the characteristic polynomial $\det(T - g) \in K[T]$ for some element $g \in \Gamma$?

This can be studied using the large sieve, as we explain in the simplest case when $\Gamma \subset \mathrm{SL}_r(\mathbf{Z})$. Let $\mathbf{G}$ be the Zariski-closure of Γ, and assume that $\mathbf{G}$ is connected and split over $\mathbf{Q}$, for instance $\mathbf{G} = \mathrm{SL}_r$. Let W be the Weyl group of $\mathbf{G}$: this will turn out to be the typical Galois group in this case.

To see this, the first ingredient is the existence, for any prime p large enough (such that $\mathbf{G}$ can be reduced modulo p), of a certain map

$$\varphi : \mathbf{G}(\mathbf{F}_p)^\sharp_r \longrightarrow W^\sharp$$

going back to Carter and Steinberg, where $G^\sharp$ denotes the set of conjugacy classes of a finite group and the subscript r restricts to regular semisimple elements in the finite group $\mathbf{G}(\mathbf{F}_p)$.

This map is used to detect elements in the Galois groups of elements in Γ in the following way. First, for $g \in \mathrm{SL}_r(\mathbf{Z})$, let P_g be the characteristic polynomial and let K_g be its splitting field, Gal_g its Galois group over $\mathbf{Q}$. The point is that, if g is a regular semisimple element of Γ, it is shown in [23] that there exists an injective homomorphism

$$j_g : \mathrm{Gal}_g \hookrightarrow W,$$

[4] I.e., there is no finite-index solvable subgroup of Γ.

canonical up to conjugation, such that if p is any prime unramified in K_g, the Frobenius conjugacy class at p maps under j_g to the conjugacy class $\varphi(\pi_p(g)) \in W^\sharp$. Thus one can detect whether the image of Gal_g in W intersects various conjugacy classes by seeing where the reduction modulo p of g lies with respect to φ. As it turns out, the image of φ becomes equidistributed among the conjugacy classes in W as p becomes large. Using this, it is not too hard to show that if $\alpha \in W^\sharp$ is a given conjugacy class and if Ω_p denotes the set of $g \in \mathbf{G}(\mathbf{F}_p)$ such that $\varphi(g) \notin \alpha$, then these sets satisfy a large sieve density assumption

$$\frac{|\Omega_g|}{|\mathbf{G}(\mathbf{F}_p)|} \geqslant \delta_\alpha > 0$$

for some $\delta_\alpha > 0$ and all p large enough. It follows by the large sieve that the probability that the element γ_n at the n-th step of a random walk on Γ has Galois group such that $j_g(\mathrm{Gal}_g) \cap \alpha = \emptyset$ is exponentially small. This holds for all the finitely many classes in W, and a well-known lemma of Jordan[5] allows us to conclude that the set of $g \in \Gamma$ where j_g is not onto is exponentially small.

The general case treated by Lubotzky–Rosenzweig is quite a bit more involved. In particular, new phenomena appear when $\mathbf{G}$ is *not* connected, and the different cosets of the connected component of the identity then usually have different typical Galois groups. We refer to their paper for details.

6. Problems and questions

We discuss here a few questions and problems, selected to a large extent according to the author's own interests and bias.

(1) [Effective results] A striking aspect of the results we have described is how little they use the many refinements and developments of sieve theory, as described in [10] for small sieves, and in [28] for the large sieve. This is due to the almost complete absence of explicit forms of the Expansion Assumption for sparse groups, from which it follows that one can not, for instance, give a numerical value of the integer ω guaranteed to exist in Corollary 4.10 (recall that in classical sieves, the current state of the art is very refined indeed: one knows, for instance, that the number of primes $p \leqslant x$ such that $p+2$ has at most two prime factors is of the right order of magnitude). In fact, when implementing the combinatorial counting methods (either word-length or random walks), there is *no* known explicit sieve statement, as far as the author knows[6] (whereas a few explicit bounds do exist for archimedean balls, based on spectral or ergodic methods, see, e.g., the works of Kontorovich [24], Kontorovich–Oh [27], Nevo–Sarnak [41], Liu–Sarnak [32], and Gorodnik–Nevo [17], or for random walks in a few arithmetic groups [28]). It seems clear that the current proofs of expansion for sparse groups, although they are effective, would lead to dreadful bounds on a suitable ω (see [30] for a numerical upper-bound on the spectral radius for Cayley graphs of Zariski-dense subgroups of $\mathrm{SL}_2(\mathbf{Z})$ modulo primes, which suggests, e.g., that one could not get better than ω of size at least $2^{2^{40}}$ or so for the product of coordinates function on the Lubotzky group...).

[5] In a finite group G, there is no proper subgroup H such that $H \cap \alpha \neq \emptyset$ for all conjugacy classes α in G.

[6] The remarkable results of Bourgain and Kontorovich [4] are explicit, but not directly related to the sieve as we have considered here; see [25] for a survey in these Proceedings.

(2) [Average expansion?] One possibility suggested by the classical Bombieri–Vinogradov Theorem is to attempt a proof of expansion “on average” for the relevant Cayley graphs: for many applications, it would be sufficient to prove estimates for quantities like

$$\sum_{q_I \leqslant Q} \max_{(g_p)\in\Gamma_I} \Big|\mu_N(\pi_p(g) = g_p \text{ for } p \in I) - \frac{1}{|\Gamma_I|}\Big|,$$

and such estimates could conceivably be provable without resorting to individual estimates for each q_I. They could also, optimistically, be of better quality than what is true for individual I. (Such a property is known for classical sieve, by work of Fouvry, Bombieri, Friedlander and Iwaniec).

(3) [Combinatorial balls] It would be very interesting to have equidistribution and sieve results using trunctions based on word-length balls, without resorting to random walks. Here, the hope is that one might not need to compute the asymptotics of the size of the combinatorial balls, since one is only interested in relative proportions of elements in a ball mapping to a given $g \in \Gamma_p$.

(4) [Reverse power] This question is related to (1): at least in some cases, one has very convincing conjectures for the counting function of primes arising from small sieve in orbits (see, e.g., [11, 13]). Suppose one assumes such conjectures. What does this imply for prime numbers? In other words, can one exploit information on primes represented using the sieve in orbits to derive other properties of prime numbers? Here the reference to keep in mind is the result of Gallagher (see [14] and the generalization in [31]) that shows that uniform versions of the Hardy–Littlewood k-tuples conjecture imply that the number of primes $p \leqslant x$ in intervals of length $\lambda \log x$, for fixed $\lambda > 0$, is asymptotically Poisson-distributed.

References

[1] N. Bourbaki: *Fonctions d'une variable réelle*, Paris, Hermann, 1976.

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

[3] J. Bourgain, A. Gamburd and P. Sarnak: *The affine linear sieve*, Invent. math. 179 (2010), 559–644.

[4] J. Bourgain and A. Kontorovich: *On representations of integers in thin subgroups of* $\mathrm{SL}(2,\mathbf{Z})$, GAFA 20 (2010), 1144–1174; Erratum, GAFA 20 (2010), 1548–1549.

[5] E. Breuillard: *Mini-course on approximate groups*, these Proceedings.

[6] N. Chavdarov: *The generic irreducibility of the numerator of the zeta function in a family of curves with large monodromy*, Duke Math. J. 87 (1997), 151–180.

[7] G. Djanković: *The Erdös-Kac theorem for curvatures in integral Apollonian circle packings*, Publications de l'Institut Mathématique 89 (2011), 11–17.

[8] N. Dunfield and W. Thurston: *Finite covers of random:* 3*-manifolds*, Invent. math. 166 (2006), 457–521.

[9] B. Farb and D. Margalit: *A primer on mapping class groups*, Princeton Univ. Press 2011.

[10] J. Friedlander and H. Iwaniec: *Opera de cribro*, Colloquium Publ. 57, A.M.S, 2010.

[11] E. Fuchs: *Strong approximation in the Apollonian group*, J. Number Theory 131 (2011), 2282–2302.

[12] E. Fuchs: *Counting problems in Apollonian packings*, preprint (2012).

[13] E. Fuchs and K. Sanden: *Some experiments with integral Apollonian circle packings*, Exp. Math. 20 (2011), 380–399.

[14] P.X. Gallagher, *On the distribution of primes in short intervals*, Mathematika 23 (1976), 4–9.

[15] P.X. Gallagher: *The large sieve and probabilistic Galois theory*, in Proc. Sympos. Pure Math., Vol. XXIV, Amer. Math. Soc. (1973), 91–101.

[16] A. Gorodnik and A. Nevo: *The ergodic theory of lattice subgroups*, Annals of Math. Studies 172, Princeton Univ. Press, 2009.

[17] A. Gorodnik and A. Nevo: *Lifting, restricting and sifting integral points on affine homogeneous varieties*, Compositio Math., to appear.

[18] A. Gorodnik and A. Nevo: *Splitting fields of elements in arithmetic groups*, Math. Res. Letters 18 (2011), 1281–1288.

[19] A. Granville and K. Soundararajan: *Sieving and the Erdös-Kac theorem*, in "Equidistribution in number theory, an introduction", 15–27, Springer 2007.

[20] B. Green: *On a variant of the large sieve*, preprint `arXiv:0807.5037`.

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

[22] E. Hrushovski, P.H. Kropholler, A. Lubotzky and A. Shalev: *Powers in finitely generated groups*, Trans. Amer. Math. Soc. 348 (1996), no. 1.

[23] F. Jouve, E. Kowalski and D. Zywina: *Splitting fields of characteristic polynomials of random elements in arithmetic groups*, to appear in Israel J. Math., `arXiv:1008.3662`

[24] A. Kontorovich: *The hyperbolic lattice point count in infinite volume with applications to sieves*, Duke Math J. 149 (2009), 1–36.

[25] A. Kontorovich: *The orbital circle method*, these Proceedings.

[26] A. Kontorovich and H. Oh: *Apollonian circle packings and closed horospheres on hyperbolic* 3-*manifolds*, J. of the A.M.S 24 (2011), 603–648; also `arXiv:0811.2236v4`

[27] A. Kontorovich and H. Oh: *Almost prime Pythagorean triples in thin orbits*, J. reine und angew. Math. 667 (2012), 89–131; also `arXiv:1001.0370`.

[28] E. Kowalski: *The large sieve and its applications*, Cambridge Tracts in Math. 175, Cambridge Univ. Press, 2008.

[29] E. Kowalski: *Sieve in expansion*, Séminaire Bourbaki, Exposé 1028 (November 2010).

[30] E. Kowalski: *Explicit growth and expansion for* SL_2, preprint (2012).

[31] E. Kowalski: *Averages of Euler products, distribution of singular series and the ubiquity of Poisson distribution*, Acta Arithmetica 148.2 (2011), 153–187.

[32] J. Liu and P. Sarnak: *Integral points on quadrics in three variables whose coordinates have few prime factors*, Israel J. of Math. 178 (2010), 393–426.

[33] A. Lubotzky and C. Meiri: *Sieve methods in group theory I: powers in linear groups*, Journal of the A.M.S 25 (2012), 1119–1148.

[34] A. Lubotzky and C. Meiri: *Sieve methods in group theory II: the mapping class group*, preprint `arXiv:1104.2450`.

[35] A. Lubotzky and L. Rosenzweig: *The Galois group of random elements of linear groups*, preprint `arXiv:1205.5290`.

[36] A. Lubotzky, S. Mozes and M. Ragunathan: *The word and riemannian metrics on lattices of semisimple groups*, Publications Mathématiques de l'IHÉS, 91 (2000), 5–53.

[37] J. Maher: *Random walks on the mapping class group*, Duke Math. J. 156 (2011), 429–468.

[38] J. Maher: *Exponential decay in the mapping class group*, J. London Math. Soc., to appear, `arXiv:1104.5543`.

[39] J. Malestein and J. Souto: *Genericity of pseudo-Anosovs in the mapping class group*, International Math. Res. Notices (2012).

[40] A. Mohamadi and H. Oh: *Matrix coefficients and effective counting for geometrically finite groups*, preprint (2012).

[41] A. Nevo and P. Sarnak: *Prime and almost prime integral points on principal homogeneous spaces*, Acta Math. 205 (2010), 361–402.

[42] H. Oh: *Dynamics on geometrically finite hyperbolic manifolds with applications to Apollonian circle packings and beyond*, Proc. ICM Hyderabad, India, 2010, `arXiv:1006.2590`.

[43] H. Oh: *Harmonic analysis, ergodic theory and counting for thin groups*, these Proceedings.

[44] G. Prasad and A. Rapinchuk: *Weakly commensurable arithmetic groups and isospectral locally symmetric spaces*, Publ. Math. Inst. Hautes Études Sci. 109 (2009), 113–184.

[45] G. Prasad and A. Rapinchuk: *Existence of irreducible* $\mathbf{R}$*-regular elements in Zariski-dense subgroups*, Math. Res. Letters 10 (2003), 21–32.

[46] L. Pyber and E. Szaboó: *Growth in linear groups*, these Proceedings.

[47] A. Rapinchuk: *On strong approximation for algebraic groups*, these Proceedings.

[48] I. Rivin: *Counting Reducible Matrices, Polynomials, and Surface and Free Group Automorphisms*, Duke Math. J. 142 (2008), 353–379.

[49] I. Rivin: *Walks on graphs and lattices – effective bounds and applications*, Forum Math. 21 (2009), 673–685 and `arXiv:math/0703533`.

[50] A. Salehi Golsefidy: *Affine sieve and expanders*, these Proceedings.

[51] A. Salehi Golsefidy and P. Sarnak: *Affine sieve*, preprint (2012).

[52] A. Salehi Golsefidy and P. Varjú: *Expansion in perfect groups*, preprint (2010).
[53] P. Sarnak: *Affine sieve*, slides from lectures given in June 2010, `http://www.math.princeton.edu/sarnak/Affinesievesummer2010.pdf`
[54] P. Sarnak: *Notes on thin matrix groups*, these Proceedings.
[55] J-P. Serre: *Topics in Galois theory*, Res. Notes in Math. 1, A.K. Peters, 2008.
[56] P. Varjú: *Expansion in* $\mathrm{SL}_d(O_K/I)$*, I squarefree*, preprint (2010), `arXiv:1001.3664`.

ETH Zürich – D-MATH, Rämistrasse 101, 8092 Zürich, Switzerland
E-mail address: `kowalski@math.ethz.ch`