Notes on thin matrix groups

PETER SARNAK

Dedicated to the memory of Frances Wroblewski

We give a brief overview of the developments in the theory, especially the fundamental expansion theorem. Applications to diophantine problems on orbits of integer matrix groups, the affine sieve, group theory, gonality of curves and Heegaard genus of hyperbolic three manifolds, are given. We also discuss the ubiquity of thin matrix groups in various contexts, and in particular that of monodromy groups.

1. The fundamental expansion theorem

The Chinese remainder theorem for $\text{SL}_n(\mathbb{Z})$ asserts, among other things, that for $q \geq 1$, the reduction $\pi_q : \text{SL}_n(\mathbb{Z}) \rightarrow \text{SL}_n(\mathbb{Z}/q\mathbb{Z})$ is onto. Far less elementary is the extension of this feature to $G(\mathbb{Z})$ where $G$ is a suitable matrix algebraic group defined over $\mathbb{Q}$. The general form of this phenomenon for arithmetic groups is known as strong approximation and it is well understood [Platonov and Rapinchuk 1994].

There is a quantification of the above that is not as well known as it should be, as it turns out to be very powerful in many contexts. We call this “superstrong” approximation and it asserts that if we choose a finite symmetric ($s \in S$ if $s^{-1} \in S$) generating set $S$ of $\text{SL}_n(\mathbb{Z})$, then the congruence Cayley graphs $(X_q, S)$ form an expander family as $q$ goes to infinity (see [Hoory et al. 2006] for the definition and properties of expanders). Here the vertices $x$ of the $|S|$-regular connected graph $(X_q, S)$ are the elements of $\text{SL}_n(\mathbb{Z}/q\mathbb{Z})$ and the edges run from $x$ to $sx$, $s \in S$. The proof of this expansion property for $\text{SL}_2(\mathbb{Z})$ has its roots in Selberg’s lower bound of $\frac{3}{16}$ for the first eigenvalue $\lambda_1$ of the Laplacian on the hyperbolic surface $\Gamma \backslash \mathbb{H}$, $\Gamma$ a congruence subgroup of $\text{SL}_2(\mathbb{Z})$ [Selberg 1965]. This bound is an approximation to the Ramanujan/Selberg conjecture for automorphic forms on $\text{GL}_2/\mathbb{Q}$. The generalizations of the expansion property to $G(\mathbb{Z})$ where $G$ is say a semisimple matrix group defined over $\mathbb{Q}$ is also known thanks to developments towards the general Ramanujan conjectures that have been established [Burger and Sarnak 1991; Clozel 2003; Sarnak 2005]. This general expansion for these $G(\mathbb{Z})$ also goes by the name “property $\tau$” for congruence subgroups [Lubotzky 2005].
Let $\Gamma$ be a finitely generated subgroup of $\text{GL}_n(\mathbb{Z})$ (more generally later on we allow it to be in $\text{GL}_n(K)$ where $K$ a number field) and denote its Zariski closure: $\text{Zcl}(\Gamma)$, by $G$. If $\Gamma$ is of finite index in $G(\mathbb{Z})$, then the discussion above of strong and superstrong approximation can be applied. However, if $\Gamma$ is of infinite index in $G(\mathbb{Z})$, then $\text{Vol}(\Gamma \backslash G(\mathbb{R})) = \infty$ and the techniques used to prove both of these properties don’t apply. In this case we call $\Gamma$ “thin”. It is remarkable that under suitable natural hypotheses, strong approximation continues to hold in this thin context. The first result in this direction is [Matthews et al. 1984], and Weisfeiler extended it much further. More recent and effective treatments of this can be found in [Nori 1987] and [Larsen and Pink 2011]. An example of the statement of strong approximation in this context is: suppose that $\text{Zcl}(\Gamma) = \text{SL}_n$, then there is a $q_0 = q_0(\Gamma)$ such that $\pi_q : \Gamma \rightarrow \text{SL}_n(\mathbb{Z}/q\mathbb{Z})$ is onto whenever $(q, q_0) = 1$.

That the expansion property might continue to hold for thin groups was first suggested by Lubotzky [1993]. Thanks to a number of major developments by many people [Sarnak and Xue 1991; Gamburd 2002; Helfgott 2008; Bourgain and Gamburd 2008b; Bourgain et al. 2010; Pyber and Szabó 2010; Breuillard et al. 2011; Varjú 2012], the general expansion property is now known. The almost final version (almost because of the restriction that $q$ be squarefree) is due to Salehi and Varjú [2012].

**Theorem** (the fundamental expansion). Let $\Gamma \leq \text{SL}_n(\mathbb{Q})$ be a finitely generated group with a symmetric generating set $S$. Then the congruence graphs $(\pi_q(\Gamma), S)$, for $q$ squarefree and coprime to a finite set of primes (which depend on $\Gamma$), are an expander family if and only if $G^0$, the identity component of $G := \text{Zcl}(\Gamma)$, is perfect (i.e., $[G^0, G^0] = G^0$). Moreover the determination of the expansion constant is in principle effective, if not feasible.\(^1\)

I will not review the techniques leading to the proof of this theorem (they have been discussed in many places including Kowalski and Tao’s blogs) other than to point out that it involves three steps, the opening, the middlegame and the endgame. The endgame establishes the expansion by combining sufficiently strong (but still quite crude) upper bounds for the number of closed circuits in these graphs with largeness properties of the dimensions of the irreducible representations of the finite groups $G(\mathbb{Z}/q\mathbb{Z})$. In some cases (indeed all for which reasonable bounds for the expansion are known) the proof involves the endgame only [Sarnak and Xue 1991; Gamburd 2002]. In the general case, the upper bounds for the number of closed circuits is derived combinatorially. The opening and middlegame involve showing that smaller subsets of $G(\mathbb{Z}/p\mathbb{Z})$ grow substantially when multiplied by themselves at least three times (see [Helfgott

\(^1\)It remains an open problem how to match to some extent in this general setting the quality of expansion that is known when $\Gamma$ is arithmetic.
2008] and the extensions [Pyber and Szabó 2010] and [Breuillard et al. 2011]). A critical ingredient in the early treatments was the “sum-product” theorem [Bourgain et al. 2004] in finite fields. The middlegame is concerned with moderately large sets and is further handled by the crucial “flattening lemma” [Bourgain and Gamburd 2008b]. The latter also has its roots in combinatorics appealing to the Balog–Szemeredi theorem [Balog and Szemerédi 1994; Gowers 2001]. When \( q \) is not prime, the analysis and combinatorics is far more complicated and difficult due to the many subgroups of \( G(\mathbb{Z}/q\mathbb{Z}) \). It is handled in [Bourgain et al. 2010] for \( SL_2 \) and in [Varjú 2012] in general.

2. Applications

2.1. The affine sieve and diophantine analysis. The impetus for developing the expansion property for thin groups arose in connection with diophantine problems (in particular sieve problems for values of polynomials) on orbits of such thin groups [Bourgain et al. 2010]. Both strong approximation and superstrong approximation are crucial ingredients in executing a Brun combinatorial sieve in this setting. The theory is by now quite advanced and in particular the basic theorem of the affine sieve has been established in all cases where it is expected to hold [Salehi Golsefidy and Sarnak 2011].

For various special examples, such as for integral Apollonian packings, which has turned out to be one of the gems of the theory [Sarnak 2011], much more can be said thanks to special features. Firstly, in this case one can develop an archimedean count for the number of points in an orbit in a large region. This is done by combining spectral methods (using techniques which when \( \Gamma^\prime \) is a geometrically finite subgroup of \( O(n - 1, 1)(\mathbb{R}) \) go back to [Patterson 1976; Sullivan 1979; Lax and Phillips 1982]) with ergodic theoretic methods [Kontorovich and Oh 2011; Oh and Shah 2013; Lee and Oh 2013; Vinogradov 2012]. For the diophantine applications, one needs an archimedean spectral gap for the induced congruence groups, rather than the combinatorial expansion. [Bourgain et al. 2011] establishes the transfer of this information from the combinatorial to archimedean setting in this infinite volume case.

Two recent highlights of these developments are the “almost all” local-to-global results of [Bourgain and Kontorovich 2013a; 2013b]. The first concerns integral Apollonian packings and the question is which numbers are curvatures? The expected local-to-global conjecture [Graham et al. 2003; Fuchs and Sanden 2011] is proven for all but a zero density set of integers (the conjecture asserts that there are only a finite number of exceptions). Prior to that Bourgain and Fuchs [2011] had shown that the number of integers that are achieved is of positive density. The second development concerns the Zaremba problem, which
asserts that if $A \geq 5$, the set of integers $q \geq 1$ for which there is a $1 \leq b \leq q$, $(b, q) = 1$ and for which the coefficients of the continued fraction of $b/q$ are bounded by $A$, consists of all of $\mathbb{N}$. In [Bourgain and Kontorovich 2013b] the theory of thin subgroups of $\text{SL}_2(\mathbb{Z})$ is extended to thin semi(sub)groups (one has to abandon direct spectral methods and replace them by dynamical ones [Lalley 1989; Bourgain et al. 2011]). In [Bourgain and Kontorovich 2013b] it is shown that for $A \geq 50$, the set of exceptions to the Zaremba conjecture is of zero density in $\mathbb{N}$.

2.2. Random elements in $\Gamma$. It is well known that for any reasonable notion of randomness, the random $f \in \mathbb{Q}[x]$ is irreducible and has Galois group the full symmetric group on the degree of $f$ symbols. In [Rivin 2008] the study of such questions for the characteristic polynomial $f_\gamma$ of a random element $\gamma$ in $\text{Sp}(2g, \mathbb{Z})$ and more general $\Gamma$'s, was initiated. The random element in $\text{Sp}(2g, \mathbb{Z})$ is generated by running a symmetric random walk with respect to a measure $\mu$ whose support generates $\text{Sp}(2g, \mathbb{Z})$. The expansion property is used via a sieving argument to show that the probability that $f_\gamma$ is reducible is exponentially small. This and some generalizations are then coupled with the theory of the mapping class group $\mathcal{M}$ to show that the random element in $\mathcal{M}$ is pseudo-Anosov. These irreducibility questions and much more,\footnote{For example to showing that it is very unlikely that a random three dimensional manifold in the Dunfield–Thurston model [2006], has a positive first Betti number.} are extended and refined, especially in terms of the sieves that are applied, in [Kowalski 2008; Jouve et al. 2013; Lubotzky and Rosenzweig 2012]. Again, strong and superstrong approximation plays a central role.

In a different direction, Lubotzky and Meiri [2012] examine some group theoretic questions for linear groups using a random walk and a sieve. An example of what they show is: Let $\Gamma$ be a finitely generated subgroup of $\text{GL}_n(\mathbb{C})$ which is not virtually solvable, then the set of proper powers $P := \bigcup_{m=2}^{\infty} \{\gamma^m : \gamma \in \Gamma\}$, is exponentially small (in terms of hitting $P$ in a long random walk). In particular, this resolved an open question as to whether finitely many translates of $P$ can cover $\Gamma$, the answer being no.

2.3. Gonality and Heegaard genus. A compact Riemann surface of genus $g$ can be realized as a covering of the plane of degree at most $g + 1$ (Riemann–Roch). The gonality $d(X)$ of $X$ is the minimal degree of such a realization. Unlike $g(X)$, $d(X)$ is a subtle conformal invariant. In [Zograf 1984] (see also [Yau 1996; Abramovich 1996]) the differential geometric inequality of [Yang and Yau 1980] is extended to the setting of $X = \Gamma \backslash \mathbb{H}$, a finite area quotient (orbifold) of the hyperbolic plane. If $A(X)$ is its area and $\lambda_1(X)$ its first Laplace eigenvalue,
then
\[ d(X) \geq \frac{\lambda_1(X) A(X)}{8\pi}. \] (2-1)

This together with the known bounds towards the Ramanujan/Selberg conjectures for congruence (arithmetic) Xs (see [Blomer and Brumley 2011] for the best bounds for GL\(_2/\mathbb{K}\), where \(\mathbb{K}\) is a number field which is what is relevant here) imply that for these Xs, the ratio of any two of \(d(X), A(X)\) and \((g(X) + 1)\) is bounded universally from above and below.

There is a generalization of (2-1) to finite volume quotients \(X = \Gamma \backslash \mathbb{H}^m\) (orbifolds) of hyperbolic \(m\)-space [Agol et al. 2008]. This is stated in terms of Li and Yau’s notion [1982] of conformal volume. It gives an inequality between \(\text{Vol}(X), \lambda_1(X)\) and the conformal volume of a piecewise conformal map of \(X\) into \(S^m\). Again, this, together with the known universal lower bounds for \(\lambda_1(X)\) when \(X\) is congruence arithmetic [Burger and Sarnak 1991; Clozel 2003], gives a linear in the volume, lower bound for the conformal volume of a conformal map of \(X\) to \(S^m\). This has a nice application to reflection groups. A discrete group of motions of \(\mathbb{H}^m\) is called a reflection group if it is generated by reflections (a reflection of \(\mathbb{H}^m\) is a nontrivial isometry which fixes an \(m - 1\) dimensional hyperplane). Using the inequalities mentioned above, one shows (see [Long et al. 2006] for \(m = 2\) and [Agol et al. 2008] for \(m > 2\)) that the set of maximal arithmetic reflection groups is finite for each \(m\). Now Vinberg [1984] and Prokhorov [1986] have shown that for \(m \geq 1000\), a reflection group can never be a lattice. Thus the totality of all maximal arithmetic reflection groups is finite.

Equation (2-1) has interesting applications to diophantine equations. As observed in [Abramovich 1991; Frey 1994], Faltings’ finiteness theorem [1991] for rational points on subvarieties of abelian varieties can be used to prove finiteness of rational points on curves, whose coordinates lie in the union of all number fields of a bounded degree, as long as one can show the gonality of the curve is large enough. For example, if \(X_0(N)/\mathbb{Q}\) is the familiar modular curve of level \(N\) and if \(D\) is given, then for \(N \geq 230D\) (this value following from (2-1) and explicit Ramanujan bounds), the set of points on \(X_0(N)\) with coordinates in the union of all number of fields of degree at most \(D\), is finite! Recently Ellenberg, Hall and Kowalski [Ellenberg et al. 2012] have applied similar reasoning to a diophantine problem on a tower of curves. It arises from questions of reducibility and symmetry of specializations of members of a 1-parameter family of varieties. The curves that arise (as the parameter) are determined by the monodromy group \(\Gamma\) of the family (see below), and it lies in \(\text{Sp}(2g, \mathbb{Z})\) and is assumed to be Zariski dense in \(\text{Sp}(2g)\). In order to show that the gonality of the curves in question increase quickly enough, they use the combinatorial expansion that is provided
by the fundamental expansion theorem. Typically it is not known if $\Gamma$ is thin or not in this context (see Section 3), but the beauty of the fundamental theorem is that one does not need to know!

There is an inequality similar to (2-1) for the Heegaard genus of a hyperbolic 3-manifold $X$. It is known that such an $X$ can be decomposed into two handle bodies with common boundary a surface of genus $h$ (called a Heegaard splitting). The minimal genus of such a surface in a splitting is called the Heegaard genus of $X$ which we denote by $g(X)$. Like the gonality, it is a much more subtle (this time topological) invariant of $X$ than its volume. In [Lackenby 2006] (see Theorem 4.1 and [Buser 1982]) it is shown that for complete $X$ of finite volume

$$g(X) \geq \min[\lambda_1(X), 1] \cdot \text{Vol}(X) \cdot \frac{32\pi}{\lambda_1(X)}.$$  

This together with the universal lower bounds for $\lambda_1$ for congruence arithmetic $X$’s shows that the Heegaard genus of a congruence hyperbolic three manifold is in order of magnitude, a linear function of its volume. In particular, any arithmetic 3-manifold has an infinite tower (by congruence subgroups) of coverings whose Heegaard genus grows linearly with the volume. One can ask if the same is true for any hyperbolic 3-manifold and the answer is yes as was shown in [Long et al. 2008]. Using local rigidity of lattices in $\text{SL}_2(\mathbb{C})$ one can realize $\Gamma$ where $X = \Gamma \backslash \mathbb{H}^3$, as a finitely generated subgroup of $\text{SL}_2(K)$, where $K$ is some number field. If $\Gamma$ is not arithmetic then $\Gamma$ is thin (in $\text{SL}_2(O_K)$ perhaps allowing denominators at finitely many places), since its projection on the identity embedding of $K$ into $\mathbb{C}$ is discrete. Using the fundamental expansion theorem gives a lower bound on $\lambda_1$ for a “congruence tower” of $\Gamma$ and one then applies (2-2).

A related application of the expansion is to some questions in knot theory. Answering a question of Gromov, Pardon [2011] recently showed that there are isotopy classes of knots in $S^3$ which have arbitrary large distortion. In fact he shows that torus knots have this property. In [Gromov and Guth 2012] a large family of knots with large distortion is constructed using hyperbolic 3-manifolds $X$. Such an $X$ can be realized as a degree 3 cover of $S^3$ branched over a knot $K$ [Hilden 1976; Montesinos 1976]. Gromov and Guth [2012] show that the distortion $\delta(K)$ of this $K$ satisfies $\delta(K) \gg \text{Vol}(X)\lambda_1(X)$, (the implied constant being universal). From this and the lower bound for $\lambda_1$ when $X$ varies over congruence arithmetic 3-manifolds (or a congruence thin tower and using the fundamental expansion theorem) one concludes that $K$ and all knots isotopic to it has arbitrarily large distortion by choosing such $X$ of large volume.

2.4. Rotation groups. Let $\Gamma = \langle \sigma_1, \sigma_2, \ldots, \sigma_t \rangle$ be a finitely generated subgroup of the group $\text{SO}_3(\mathbb{R})$. There is an archimedean analogue of the expander property
for the congruence graphs in this setting and which likewise has many applications
[Lubotzky 1994; Sarnak 1990]. Define \( T_\sigma \) to be the averaging operator on
functions on the two sphere \( S^2 \) by:

\[
T_\sigma f(x) = \sum_{j=1}^{t} [f(\sigma_j x) + f(\sigma^{-1} j x)].
\]  

(2-3)

\( T_\sigma \) is self adjoint on \( L^2(S^2, dA) \), where \( dA \) is the rotation invariant area element
on \( S^2 \), and its spectrum is contained in \([-2t, 2t]\). The spectral gap property
is that \( 2t \) (which is an eigenvalue with eigenvector the constant function) is a
simple and isolated point of the spectrum. It is not hard to see that this property
depends only on \( \Gamma \) and not on the generators. It is conjectured that \( \Gamma \) has such
a spectral gap if and only if \( \text{Zcl}(\Gamma) = \text{SO}_3 \) (which in this case is equivalent
to the topological closure of \( \Gamma \) being \( \text{SO}_3(\mathbb{R}) \)). A lot is known towards this
conjecture. The first example of a \( \Gamma \) with a spectral gap was given by Drinfeld
[1984] and this provided the final step in the solution of the Ruziewicz problem;
that the only finitely additive rotationally invariant measure defined on Lebesgue
measurable subsets of \( S^2 \), is a multiple of \( dA \). His proof of the spectral gap
makes use of an arithmetic such \( \Gamma \) together with the full force of automorphic
forms and the solution of the Ramanujan conjectures for holomorphic cusp forms
on the upper half plane. In [Gamburd et al. 1999] many thin \( \Gamma \)s are shown to
have a spectral gap. The best result known is the analogue of the fundamental
expansion theorem in this context [Bourgain and Gamburd 2008a; 2010], and it
suffices for most applications. It asserts that if the matrix elements of members
of \( \Gamma \) are algebraic, then the conjecture is true for \( \Gamma \). Like the very thin cases
of the fundamental expansion theorem, part of the proof here relies on additive
combinatorics. This time one needs the full force of the proof of the local
Erdős–Volkmann ring conjecture [Edgar and Miller 2003; Bourgain 2003] — that
a subset of \( \mathbb{R} \) which is closed under addition and multiplication has Hausdorff
dimension zero or one. As far as some concrete applications of the spectral
gap for these groups, we mention the speed of equidistribution of directions
associated with general quaquaversal tilings of three dimensional space [Draco
et al. 2000; Radin and Sadun 1998] and constructions of quantum gates in the
theory of quantum computation (the Solovay–Kitaev theorem; see [Harrow et al.
2002]).

3. Ubiquity of thin groups

Given a finitely generated group \( \Gamma \) in \( \text{GL}_n(\mathbb{Z}) \), one can usually compute \( G = \text{Zcl}(\Gamma) \) without too much difficulty. On the other hand, deciding if \( \Gamma \) is thin
can be formidable. In fact one is flirting here with questions that have no
decision procedures (I thank Rivin for alerting me to these pitfalls that are close by). For example, if \( \Gamma = \text{SL}_2(\mathbb{Z}) \times \text{SL}_2(\mathbb{Z}) \) then there is no decision procedure to determine if an element \( A \in \Gamma \) is in the group generated by a general set of say seven elements [Mikhailova 1958]. Even for Gromov hyperbolic groups, the question of whether a finitely generated subgroup generates a finite index subgroup, has no decision procedure [Rips 1982; Baumslag et al. 1994]. Mercifully strong and superstrong approximation only ask about \( \text{Zcl}(\Gamma) \). Still one is curious about thinness when applying these theorems and sometimes for good reason. For example, in the affine sieve setting, the quality of the expansion impacts the results dramatically (see [Nevo and Sarnak 2010] for the cases when \( \Gamma \) is a lattice) while the diophantine orbit problems become more standard ones of integer points on homogeneous varieties, when \( \Gamma \) is a lattice. Whether the typical \( \Gamma \) is thin or not is not so clear, and may depend on how \( \Gamma \) arises.

3.1. Schottky, ping-pong. Schottky groups in which the generators play ping-pong [Tits 1972; Breuillard and Gelander 2003] are one of the few classes of discrete groups whose group theoretic structure is very simple. If one chooses \( A_1, A_2, \ldots, A_\ell \) independently and at random in \( \text{SL}_n(\mathbb{Z})(n \geq 2) \), then with high probability \( \Gamma = \langle A_1, \ldots, A_\ell \rangle \) will be free on these generators, Zariski dense in \( \text{SL}_n \) and thin. If the \( A_j \) are chosen at the \( m \)-th step of a \( \mu \)-random walk \( (m \to \infty) \) and support \( (\mu) \) generates \( \text{SL}_n(\mathbb{Z}) \), then this was proved in [Aoun 2011]. A more geometric version is proven in [Fuchs and Rivin \geq 2012] where the \( A \)s are chosen independently and uniformly by taking them from the set of \( B \)s with \( \max(\|B\|, \|B^{-1}\|) < X \). Here \( \| \| \) is any euclidean norm on the space of matrices and \( X \to \infty \). Not only is \( \Gamma \) thin but it is very thin in the sense that the Hausdorff dimension of the limit set of \( \Gamma \) acting on \( \mathbb{P}^{n-1}(\mathbb{R}) \) is arbitrarily small.

3.2. Nonarithmetic lattices. If \( \Gamma \leq G \) with \( G \neq \text{SL}_2(\mathbb{R}) \), is an irreducible nonarithmetic lattice in a semisimple real group \( G \), then \( \Gamma \) is naturally thin in the appropriate product by its conjugates. The argument is the same as the one in Section 2.3 using local rigidity. The certificate of being thin is that \( \Gamma \) is discrete in the factor corresponding to \( G \). Examples of this kind which come from monodromy of hypergeometric differential equations in several variables are given in [Deligne and Mostow 1986] and in one variable in [Cohen and Wolfart 1990]. It appears that these were the first examples of thin monodromy groups (see Section 3.5 below). Other examples of thin monodromy groups in products of \( \text{SL}_2 \)s are given in [Nori 1986] and these examples aren’t even finitely presented. Teichmüller curves in the moduli space \( M_2 \) of curves of genus 2, give via Abel–Jacobi, curves in \( A_2 \) whose monodromies (inclusion of fundamental
3.3. Reflection groups in hyperbolic space. Let $f$ be an integral quadratic form in $n$-variables and of signature $(n-1, 1)$. For $n \geq 3$, $O_f(\mathbb{Z})$ the group of integral automorphs of $f$ is a lattice in $G = O_f(\mathbb{R})$. The reflective subgroup $R_f$ is the subgroup of $O_f(\mathbb{Z})$ which is generated by all the hyperbolic reflections which are in $O_f(\mathbb{Z})$. $R_f$ is a normal subgroup of $O_f(\mathbb{Z})$ and if it is nontrivial, then $\text{Zcl}(R_f) = O_f$. Vinberg [1984] and Nikulin [1987] have examined the question of when $R_f$ is of finite index in $O_f(\mathbb{Z})$ (they call such an $f$ reflective). In particular, in [Nikulin 1987] it is shown that there are only finitely many $f$'s (up to integral equivalence) which are reflective. Thus for all but finitely many $f$'s, $R_f$, if it is nontrivial, is a thin group in $\text{GL}_n(\mathbb{Z})$ (albeit infinitely generated). Note that Nikulin’s theorem fails for $n = 2$. If $f$ is a binary quadratic form, then $f$ is reflective if and only if it is ambiguous in the sense of Gauss (see [Sarnak 2007]) and Gauss determined the ambiguous forms in his study of genus theory.

3.4. Rotation groups. An interesting family of rotation groups are the groups $\Gamma(m, n)$, $m \geq 3$, $n \geq 3$ generated by $\sigma_m$ and $\tau_n$ where

$$\sigma_m = \begin{bmatrix} \cos 2\pi/m & \sin 2\pi/m & 0 \\ -\sin 2\pi/m & \cos 2\pi/m & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \tau_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/n & \sin 2\pi/n \\ 0 & -\sin 2\pi/n & \cos 2\pi/n \end{bmatrix}.$$ 

That is, $\Gamma(m, n)$ is a subgroup of $\text{SO}_f(\mathbb{R})$, $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$, generated by two rotations about orthogonal axes and of orders $m$ and $n$ respectively. These arise in the theory of quaquaversal tilings of 3-space and their generalizations [Conway and Radin 1998; Radin and Sadun 1998].

As abstract groups, these are free products of two cyclic (or dihedral) groups amalgamated over a similar such group (except for $\Gamma(4, 4)$ which is finite and which we avoid); see [Radin and Sadun 1999]. This description can be used to decide the question of whether $\Gamma(m, n)$ is thin or not and also to show that thin is the rule rather than the exception. If

$$K = \mathbb{Q}(\cos 2\pi/m, \sin 2\pi/m, \cos 2\pi/n, \sin 2\pi/n),$$

then $K$ is a totally real Galois extension of $\mathbb{Q}$ with abelian Galois group $G_{m,n}$. It is plain that $\Gamma(m, n)$ is a subgroup of $\text{SO}_f(O[\frac{1}{2}])$, where $O$ is the ring of integers of $K$. Moreover, since $\Gamma(m, n)$ is infinite, the powers of 2 in the denominators of the matrix entries of $\Gamma(m, n)$ must be unbounded (otherwise $\Gamma(m, n)$ would be a discrete subgroup of the compact group $\prod_{\nu|\infty} \text{SO}_f(K_\nu)$). Hence the smallest $S$-arithmetic group to contain a subgroup commensurable with $\Gamma(m, n)$ is $\text{SO}_f(O_S)$ where $O_S$ are the $S$-integers of $K$, and $S$ consists of the places of $K$ dividing...
2. Our thinness question is whether $\Gamma(m, n)$ is of finite or infinite index in the
latter. If $|S| \geq 2$, then any finite index subgroup of $\text{SO}_f(O_S)$ is a lattice in the
higher rank group, $\prod_{v \mid (2)} \text{SO}_f(K_v)$. By well known rigidity properties of such
lattices [Margulis 1991] (or one can argue with vanishing of first cohomology
groups) and the description of $\Gamma(m, n)$ mentioned above, it follows that $\Gamma(m, n)$
cannot be such a lattice. That is if $|S| = 1$, then $\Gamma(m, n)$ may be arithmetic and it is so in some special cases.\(^3\)
Perhaps the most interesting cases where $|S| = 1$ are when $m = 4$ and $n = 2^\nu$, $\nu \geq 3$, for
which 2 is totally ramified. These have been investigated in [Robinson 2006; Serre 2009].
Serre shows that for $\nu = 3$ and 4, $\Gamma(4, 2^\nu)$ is arithmetic (in fact,
$\Gamma(4, 2^\nu) = \text{SO}_f(O[1/2])$) while for $\nu \geq 5$, it is thin. The thinness is proven by
comparing the Euler characteristics $\chi(\text{SO}_f(O[1/2]))$ and $\chi(\Gamma(4, 2^\nu))$, the first
using a Tamagawa number computation and the second from the abstract group
description of $\Gamma(4, 2^\nu)$.

3.5. Monodromy groups. The oldest and perhaps most natural source of finitely
generated linear groups comes from monodromy in all of its guises. These include
the very classical case of monodromy of the hypergeometric differential equation
which we discuss further below, as well as that of a family of varieties varying
over a base with its monodromy action on cohomology. For large families, and
in cases where the monodromy has been computed, it appears almost always
to be arithmetic. The question as to whether such monodromy groups are
arithmetic was first raised in [Griffiths and Schmid 1975]. For example for the
universal family of smooth projective hypersurfaces of degree $d$ and dimension
$n$ in projective space, the monodromy representation on $H^n(X_0, \mathbb{Z})$, $X_0$ a base
hypersurface, is an arithmetic subgroup of $\text{GL}(H^n(X_0))(\mathbb{Z})$; see [Beauville 1986]
where the exact level in $G(\mathbb{Z})$ is determined. For smaller families such as cyclic
covers of $\mathbb{P}^1$, which have recently been studied in [McMullen 2013] in connection
with the thinness question, the story is similar. More precisely, consider the
family of curves (in affine coordinates) given by
$$C_a : y^d = (x - a_1)(x - a_2) \cdots (x - a_{n+1}), \quad (3-1)$$
where the parameters $a$ vary so that $a_i \neq a_j$, for $i \neq j$. The fundamental group of
the space of $as$ is the pure braid group and it has a monodromy representation on
$H_1(C, \mathbb{Z}) \cong \mathbb{Z}^{2g}$, $g$ the genus of $C_a$, and again $C$ is a fixed base curve. Answering
a question in [McMullen 2013], Venkataramana [2012] shows that if $n \geq 2d$, then

\(^3\)The quaquaversal tiling [Conway and Radin 1998] has symmetry group $\Gamma(3, 6)$, which is
arithmetic [Serre 2009], while the dite/kart tiling [Radin and Sadun 1998] has symmetry $\Gamma(10, 4)$,
for which $K = \mathbb{Q}(\cos \pi/10)$. $G_{10,4} = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $|S| = 2$; hence the latter is thin.
the image of the monodromy representation of the braid group in $GL(H_1(C))(\mathbb{Z})$ is arithmetic. This generalizes a result of [A’Campo 1979] for $d = 2$. The proof is based on another result of Venkataramana [1994] which asserts that for $\mathbb{Q}$ rank two or higher arithmetic groups, a Zariski dense subgroup which contains enough elements from opposite horospherical subgroups is necessarily arithmetic. If $n < 2d$, then as observed in [McMullen 2013], there are examples based on the nonarithmetic lattices of [Deligne and Mostow 1986] in $SU(2, 1)$ which are thin (one such is $n = 3$ and $d = 18$).

The thinness story for monodromy groups of one parameter families is less clear. We discuss in some detail the very rich examples of the classical hypergeometric equation. Let $\alpha, \beta \in \mathbb{Q}^n$ and consider the $n F_{n-1}$ algebraic hypergeometric equation:

$$Du = 0,$$

where $D = (\theta + \beta_1 - 1)(\theta + \beta_2 - 1) \cdots (\theta - \beta_n - 1) - z(\theta + \alpha_1) \cdots (\theta + \alpha_n)$ and $\theta = z d/dz$.

The equation is regular outside $\{0, 1, \infty\}$ and the fundamental group

$$\pi_1(\mathbb{P}^1 - \{0, 1, \infty\})$$

has a representation in $GL_n$ gotten by analytic continuation of a basis of solutions to (3-2) along curves in the thrice punctured sphere. Its image in $GL_n$ is denoted by $H(\alpha, \beta)$ and is the monodromy group in question (defined up to conjugation in $GL_n$). $H(\alpha, \beta)$ is generated by the local monodromies $A, B, C$ ($C = A^{-1}B$) gotten from loops about 0, $\infty$ and 1, respectively; see [Beukers and Heckman 1989] for a detailed description. We restrict to $H$s which can be conjugated into $GL_n(\mathbb{Z})$, which is equivalent to the characteristic polynomials of $A$ and $B$ being products of cyclotomic polynomials.\footnote{We assume further that $(\alpha, \beta)$ are primitive in the sense of [Beukers and Heckman 1989].} Such $H(\alpha, \beta)$ are self-dual and according to [Beukers and Heckman 1989], their Zariski closures $G(\alpha, \beta)$ are either finite, $O_n$ or $Sp_n$, and they determine which it is explicitly in terms of $\alpha$ and $\beta$. Our interest is whether $H(\alpha, \beta)$ is of finite or infinite index in $G(\alpha, \beta)(\mathbb{Z})$. Other than the cases where $H(\alpha, \beta)$ (or equivalently $G(\alpha, \beta)$) are finite, all of which are listed in [Beukers and Heckman 1989], there are few cases where $H(\alpha, \beta)$ itself is known.

Recently Venkatamarana [2012] has shown that for $n$ even and

$$\alpha = \left( \frac{1}{2} + \frac{1}{n+1}, \frac{1}{2} + \frac{2}{n+1}, \cdots, \frac{1}{2} + \frac{n}{n+1} \right),$$

$$\beta = \left( 0, \frac{1}{2} + \frac{1}{n}, \frac{1}{2} + \frac{2}{n}, \cdots, \frac{1}{2} + \frac{n-1}{n} \right),$$

(3-3)
$H(\alpha, \beta)$ is arithmetic (here $G(\alpha, \beta) = \text{Sp}(n)$). He deduces this by showing that for these exact parameters, the monodromy representation of $\pi_1(\mathbb{P}^1 - \{0, 1, \infty\})$ factors through a representation of the braid group on (3-1) with $d = 2$. In particular the arithmeticity follows from the arithmeticity of the latter.

The very fruitful Dwork family (see [Katz 2009; Harris et al. 2010]) $n \geq 4$ even, and

$$\alpha = (0, 0, \ldots, 0),$$

$$\beta = \left(\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}\right).$$

(3-4)

is apparently different. Again $G(\alpha, \beta) = \text{Sp}(n)$ and for $n = 4$, the local monodromies are

$$A = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

and

$$C = \begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$ 

(3-5)

Very recently, Brav and Thomas [2012] have shown that $A$ and $C$ in (3-5) play generalized ping-pong on certain subsets of $\mathbb{P}^3$, from which it follows that $H(\alpha, \beta) \cong \mathbb{Z}/5\mathbb{Z} \ast \mathbb{Z}$. From rigidity, or the first cohomology properties of finite index subgroups of $\text{Sp}(4, \mathbb{Z})$, it follows that $H(\alpha, \beta)$ must be thin. It seems likely that $H(\alpha, \beta)$ is thin for the whole Dwork family, that is, $n \geq 4$, but other than showing that the corresponding $A$ and $C$s play ping-pong, there appear to be no known means of proving this and no infinite family of thin $H(\alpha, \beta)$s with $G(\alpha, \beta)$ symplectic is known. For $n = 4$ there are 112 such $H(\alpha, \beta)$s in $\text{Sp}(4, \mathbb{Z})$ [Singh and Venkataramana 2012]. Using extensions of the technique in [Venkataramana 2012] it is shown in [Singh and Venkataramana 2012] that of these, 63 are arithmetic. Of these, three correspond to the 14 hypergeometrics associated with certain Calabi–Yau three folds (see [Doran and Morgan 2006; Chen et al. 2008]). Of the other 11, seven are shown to be thin in [Brav and Thomas 2012], again by finding ping-pong sets in $\mathbb{P}^3$. This leaves four of these Calabi–Yau’s for which the thinness question is open. It would be interesting to understand the geometric significance, if there is one, for $H(\alpha, \beta)$ being thin or not in these families.

What is lacking above is a certificate for $H(\alpha, \beta)$ being thin that can be applied for example to families (i.e., $n \to \infty$). A robust such certificate has been provided in the case that $G(\alpha, \beta)(\mathbb{R})$ is of rank one and $n > 3$ [Fuchs et al. 2013]. In these cases $G(\alpha, \beta)$, as a group defined over $\mathbb{Q}$ is $O_f$, where $f$ is a rational

Remarkably, all of these are realized geometrically [Doran and Morgan 2006, Theorem 2.12].

$((0, 0, 0, 0), (\frac{1}{6}, \frac{1}{6}, \frac{5}{6}, \frac{5}{6})), ((0, 0, 0, 0), (\frac{1}{6}, \frac{1}{6}, \frac{3}{6}, \frac{3}{6})), ((0, 0, 0, 0), (\frac{1}{10}, \frac{3}{10}, \frac{7}{10}, \frac{9}{10})).$
NOTES ON THIN MATRIX GROUPS 355

quadratic form in an odd number of variables and of signature \((n - 1, 1)\) (over \(\mathbb{R}\)). We call these \((\alpha, \beta)\)'s hyperbolic hypergeometrics and besides a (long) list of sporadic examples, they come in seven infinite parametric families \([\text{Fuchs et al. 2013}]\). Our conjecture for these is that thin rules, that is for all but finitely many of the hyperbolic hypergeometrics, \(H(\alpha, \beta)\) is thin. This is proved in \([\text{Fuchs et al. 2013}]\) for a number (but not all) of the seven families. For example for \(n\) odd consider the two families

\[
\alpha = \left(0, \frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n-1}{2(n+1)}, \frac{n+3}{2(n+1)}, \ldots, \frac{n}{n+1}\right),
\]

\[
\beta = \left(\frac{1}{2}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right),
\tag{3-6}
\]

and

\[
\alpha = \left(\frac{1}{2}, \frac{1}{2n-2}, \frac{3}{2n-2}, \ldots, \frac{2n-3}{2n-2}\right),
\]

\[
\beta = \left(0, 0, 0, \frac{1}{n-2}, \frac{2}{n-2}, \ldots, \frac{n-3}{n-2}\right).
\tag{3-7}
\]

Both of these families are hyperbolic hypergeometrics and for both \(H(\alpha, \beta)\) is thin for \(n \geq 5\) and is arithmetic for \(n = 3\).

The proof is based on the following principle: if \(\psi : G(\mathbb{Z}) \to K\) is a morphism onto a group \(K\) for which \(|\psi(H(\alpha, \beta))\backslash K| = \infty\), then certainly \(H(\alpha, \beta)\) is of infinite index in \(G(\mathbb{Z})\). Now in the higher rank cases there are no useful such \(\psi\)s (by the Margulis normal subgroup theorem \([1991]\) in these cases if \(K\) is infinite then \(\ker(\psi)\) is finite), however, in the rank one case such \(\psi\)s may exist and yield a certificate of thinness. Indeed in this hyperbolic case if \(R_f\) is the Vinberg reflection subgroup described in Section 3.3, then as mentioned there, except for finitely many \(f\)s, \(K_f := O_f(\mathbb{Z})/R_f\) is infinite. To use this one needs to analyze the image of \(H(\alpha, \beta)\) in \(K_f\). The key observation is that up to the finite index the hyperbolic hypergeometrics are generated by Cartan involutions.\(^7\)

These are linear reflections of \(Q^n\) which induce isometries on hyperbolic space given by geodesic inversions in a point \([\text{the hyperbolic reflections are generated by root vectors } v \in \mathbb{Z}^n \text{ outside the light cone } (f(v) > 0) \text{ while the Cartan involutions by root vectors } w \in \mathbb{Z}^n \text{ inside the light cone, in fact } f(w) = -2]\).

In order to examine the image of a group generated by such Cartan involutions in \(K_f\), consider the “minimum distance graph,” \(X_f\). Its vertices are the integral Cartan root vectors \(V_{-2}(\mathbb{Z}) = \{v \in \mathbb{Z}^n : f(v) = -2\}\), and \(v\) and \(w\) are joined if \(f(v, w) = -3\). One can show that the components of \(X_f\) consist of finitely many isomorphism types and each is the Cayley graph of a finitely generated

\(^7\)The local monodromy \(C\) about 1 is always a pseudoreflection and in these cases yields a Cartan involution.
The Coxeter group. The main lemma [Fuchs et al. 2013] asserts that if \( \Sigma \subset V_{-2}(\mathbb{Z}) \) is a connected component of \( X_f \) then the image of the group generated by the Cartan involutions \( r_v \) with roots \( v \in \Sigma \), is a finite subgroup of \( K_f \).\(^8\) This together with Vinberg and Nikulin’s theorems gives a robust certificate for the thinness of these hyperbolic hypergeometric monodromies. As far as I know (3-6) and (3-7) give the first family of thin monodromy groups in high dimensions for which \( G \) is simple.

We end with some comments about the arithmetic Ramanujan conjectures. The gonality of a congruence arithmetic surface being linear in its genus and the Heegaard genus of a congruence hyperbolic three manifold being linear in its volume, as well as the proof that there are only finitely many maximal arithmetic reflection groups, all appeal to the uniform lower bounds for \( \lambda_1 \) for all such manifolds. This follows from what is known towards the Ramanujan conjectures but it does not follow from the fundamental expansion theorem since the latter only applies to one tower at a time. As far as the general Ramanujan conjectures, some progress has been made since [Sarnak 2005]. Namely in [Arthur 2013] a precise formulation of the Ramanujan conjectures for these groups is given, and moreover it is shown (assuming forms of the fundamental lemma which themselves should be theorems before too long) that these conjectures will follow if one can prove the Ramanujan conjectures for \( GL_m \).

Acknowledgement

These brief notes cover a lot of ground. I thank my collaborators, the people whose work is quoted and the many mathematicians with whom I have discussed aspects of the theory connected with these thin groups. Thanks to the referee for pointing me to the relevant applications in [Gromov and Guth 2012].

References


\(^8\)The proof makes use of the quite special feature of the binary form \( g = x^2 + 3xy + y^2 \) of being integrally equivalent to \( -g \) (called reciprocal in [Sarnak 2007]).


[Fuchs and Rivin ≥ 2012] E. Fuchs and I. Rivin, “Finitely generated subgroups of $SL_n(\mathbb{Z})$ are generically thin”.


