Acta Math., 193 (2004), 73–104

Counting congruence subgroups

by

DORIAN GOLDFELD, ALEXANDER LUBOTZKY and LÁSZLÓ PYBER

Columbia University
New York, U.S.A.

Hebrew University
Jerusalem, Israel

Hungarian Academy of Sciences
Budapest, Hungary

0. Introduction

Let k be an algebraic number field, $\mathcal{O}$ its ring of integers, S a finite set of valuations of k (containing all the archimedean ones), and $\mathcal{O}_S=\{x\in k \mid v(x)\geqslant 0 \text{ for all } v\notin S\}$. Let G be a semisimple, simply-connected, connected algebraic group defined over k with a fixed embedding into GL_d. Let $\Gamma=G(\mathcal{O}_S)=G\cap \mathrm{GL}_d(\mathcal{O}_S)$ be the corresponding S-arithmetic group. We assume that Γ is an infinite group (equivalently, $\prod_{\nu\in S} G(k_\nu)$ is not compact).

For every non-zero ideal I of $\mathcal{O}_S$ let

$$\Gamma(I)=\mathrm{Ker}(\Gamma\to \mathrm{GL}_d(\mathcal{O}_S/I)).$$

A subgroup of Γ is called a congruence subgroup if it contains $\Gamma(I)$ for some I.

The topic of counting congruence subgroups has a long history. Classically, congruence subgroups of the modular group were counted as a function of the genus of the associated Riemann surface. It was conjectured by Rademacher that there are only finitely many congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$ of genus zero. Petersson [Pe] proved that the number of all subgroups of index n and fixed genus goes to infinity exponentially as $n\to\infty$. Dennin [De] proved that there are only finitely many congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$ of given fixed genus and solved Rademacher's conjecture. A quantitative result was proved by Thompson [T] and Cox–Parry [CP] who showed (among other interesting results) that

$$\lim \frac{\mathrm{genus}(\Lambda)}{[\mathrm{SL}_2(\mathbf{Z}):\Lambda]}=\frac{1}{12},$$

where the limit goes over congruence subgroups Λ of $\mathrm{SL}_2(\mathbf{Z})$ with index going to ∞. It does not seem possible, however, to accurately count all congruence subgroups of index at most r in $\mathrm{SL}_2(\mathbf{Z})$ by using the theory of Riemann surfaces of fixed genus.

The first two authors' research is supported in part by the NSF. The third author's research is supported in part by OTKA T 034878. All three authors would like to thank Yale University for its hospitality.

Following [Lu], we count congruence subgroups as a function of the index. For $n>0$, define

$$C_n(\Gamma) = \#\{\text{congruence subgroups of } \Gamma \text{ of index at most } n\}.$$

THEOREM 1. *There exist two positive real numbers $\alpha_-=\alpha_-(\Gamma)$ and $\alpha_+=\alpha_+(\Gamma)$ such that for all sufficiently large positive integers n,*

$$n^{(\log n/\log\log n)\alpha_-} \leqslant C_n(\Gamma) \leqslant n^{(\log n/\log\log n)\alpha_+}.$$

This theorem is proved in [Lu], although the proof of the lower bound presented there requires the prime number theorem on arithmetic progressions in an interval where its validity depends on the GRH (generalized Riemann hypothesis for Dirichlet L-functions). By a slight modification of the proof and by appealing to a theorem of Linnik [Li1], [Li2] on the least prime in an arithmetic progression, the proof can be made unconditional. Such an approach gives, however, poor estimates for the constants.

Following [Lu] we define

$$\alpha_+(\Gamma) = \varlimsup_{n\to\infty} \frac{\log C_n(\Gamma)}{\lambda(n)} \quad \text{and} \quad \alpha_-(\Gamma) = \varliminf_{n\to\infty} \frac{\log C_n(\Gamma)}{\lambda(n)},$$

where $\lambda(n)=(\log n)^2/\log\log n$.

It is not difficult to see that α_+ and α_- are independent both of the choice of the representation of G as a matrix group and of the choice of S. Hence $\alpha_\pm$ depend only on G and k. The question whether $\alpha_+(\Gamma)=\alpha_-(\Gamma)$ and the challenge to evaluate them for $\Gamma=\mathrm{SL}_2(\mathbf{Z})$ and other groups were presented in [Lu]. Here we prove:

THEOREM 2. *We have $\alpha_+(\mathrm{SL}_2(\mathbf{Z}))=\alpha_-(\mathrm{SL}_2(\mathbf{Z}))=\frac{1}{4}(3-2\sqrt{2})=0.0428932\ldots$.*

The proof of the lower bound in Theorem 2 is based on the Bombieri–Vinogradov theorem [Bo], [Da], [V], i.e., *the Riemann hypothesis on the average.* The upper bound, on the other hand, is proved by first reducing the problem to a counting problem for subgroups of abelian groups and then solving that extremal counting problem.

In the case of a number field, we will, in fact, show a more remarkable result: the answer is independent of $\mathcal{O}$! Here, we require the GRH (generalized Riemann hypothesis) [W] for Hecke and Artin L-functions, which states that all non-trivial zeros of such L-functions lie on the critical line.

THEOREM 3. *Let k be a number field with ring of integers $\mathcal{O}$. Let S be a finite set of primes, and $\mathcal{O}_S$ as above. Assume the GRH for k and all cyclotomic extensions $k(\zeta_l)$ with l a rational prime and ζ_l a primitive l-th root of unity. Then*

$$\alpha_+(\mathrm{SL}_2(\mathcal{O}_S)) = \alpha_-(\mathrm{SL}_2(\mathcal{O}_S)) = \tfrac{1}{4}(3-2\sqrt{2}).$$

The GRH is needed only for establishing the lower bound. It can be dropped in many cases by appealing to a theorem of Murty and Murty [MM] which generalizes the Bombieri–Vinogradov theorem cited earlier.

THEOREM 4. *Theorem 3 holds unconditionally if the field k is contained in a Galois extension K such that either*

(a) $\mathfrak{g}=\mathrm{Gal}(K/\mathbf{Q})$ *has an abelian subgroup of index at most* 4 (*in particular, if k is an abelian extension*), *or*

(b) $[K:\mathbf{Q}]<42$.

The proof of the upper bound is very different from the proof of the lower bound. For a group A, we denote by $s_r(A)$ the number of subgroups of A of index at most r. A somewhat involved reduction process is applied to show that the problem of finding the upper bound is actually equivalent to an extremal counting problem of subgroups of finite abelian groups (see §5) which is given in Theorem 5. A sharp upper bound for that counting problem follows from the case $R=1$ of the following theorem.

THEOREM 5. *Let $R\geqslant 1$ be a real number and let d be a fixed integer $\geqslant 1$. Suppose that $A=C_{x_1}\times C_{x_2}\times\ldots\times C_{x_t}$ is an abelian group such that the orders $x_1,x_2,\ldots,x_t$ of its cyclic factors do not repeat more than d times each. Suppose that $r|A|^R\leqslant n$ for some positive integers r and n. Then as n tends to infinity, we have*

$$s_r(A)\leqslant n^{(\gamma+o(1))l(n)},$$

where $\gamma=\big(\sqrt{R(R+1)}-R\big)^2/4R^2$.

In an earlier version of this paper, Theorem 5 was proved in a similar manner, but only for $R=1$. The more general case was proved in an early version of [LuN]. We thank the authors of [LuN] for allowing us to include the general version here.

The above results suggest that for every Chevalley group scheme G, the upper and lower limiting constants $\alpha_\pm(G(\mathcal{O}_S))$ are equal to each other, and depend only on G and not on $\mathcal{O}$. In fact, we can make a precise conjecture, for which we need to introduce some additional notation. Let G be a Chevalley group scheme of dimension $d=\dim G$ and rank $l=\mathrm{rk}(G)$. Let $\varkappa=|\Phi^+|$ denote the number of positive roots in the root system of G, and let $R=R(G)=(d-l)/2l=\varkappa/l$. We see that if G is of type A_l (resp. $B_l,C_l,D_l,G_2,F_4,E_6,E_7,E_8$) then $R=\frac{1}{2}(l+1)$ (resp. $l,l,l-1,3,6,6,9,15$).

CONJECTURE. *Let k, $\mathcal{O}$ and S be as in Theorem 3, and suppose that G is a simple Chevalley group scheme. Then*

$$\alpha_+(G(\mathcal{O}_S))=\alpha_-(G(\mathcal{O}_S))=\frac{\big(\sqrt{R(R+1)}-R\big)^2}{4R^2}.$$

The conjecture reflects the belief that "most" subgroups of $H=G(\mathbf{Z}/m\mathbf{Z})$ lie between the Borel subgroup B of H and the unipotent radical of B. We prove here the lower bound of the general conjecture (under the same assumptions as in Theorems 3 and 4). In our earlier version this was done only for Galois extensions, but it was observed in an earlier version of [LuN] that a small modification of the argument works in the general case. We thank the authors of [LuN] for allowing us to make these small modifications here.

This paper gives a complete proof of the upper bound for the case of SL_2, based on the known detailed classification of subgroups of $\mathrm{SL}_2(\mathbf{F}_q)$ for finite fields $\mathbf{F}_q$ of order q. We also give a partial result towards the upper bound in the general case. The upper bound is proved in full for every field k in [LuN]. The reader is also referred to a more general version there when G is not assumed to be split.

THEOREM 6. *Let k, $\mathcal{O}$ and S be as in Theorem 3. Let G be a simple Chevalley group scheme of dimension d and rank l, and $R=R(G)=(d-l)/2l$. Then*

(a) *assuming the GRH or the assumptions of Theorem 4,*

$$\alpha_-(G(\mathcal{O}_S)) \geqslant \frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2} \sim \frac{1}{16R^2};$$

(b) *there exists an absolute constant C such that*

$$\alpha_+(G(\mathcal{O}_S)) \leqslant C\,\frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2}.$$

Remark. As the upper bound is proved in full in [LuN] (i.e., $C=1$ in part (b)), we omit in this paper the proof of part (b) of Theorem 6.

COROLLARY 7. *There exists an absolute constant C such that for $d=2,3,\ldots$,*

$$(1-o(1))\,\frac{1}{4d^2} \leqslant \alpha_-(\mathrm{SL}_d(\mathbf{Z})) \leqslant \alpha_+(\mathrm{SL}_d(\mathbf{Z})) \leqslant C\,\frac{1}{d^2}.$$

This greatly improves the upper bound $\alpha_+(\mathrm{SL}_d(\mathbf{Z}))<\frac{5}{4}d^2$ implicit in [Lu] and settles a question asked there.

As a byproduct of the proof of Theorem 5 in §6 we obtain the following result.

COROLLARY 8. *The subgroup growth of $\mathrm{SL}_d(\mathbf{Z}_p)$ is at least n^c, i.e.,*

$$\varlimsup_{n\to\infty} \frac{\log s_n(\mathrm{SL}_d(\mathbf{Z}_p))}{\log n} \geqslant c,$$

where

$$c=(3-2\sqrt{2})d^2-2(2-\sqrt{2}),$$

and where $\mathbf{Z}_p$ *denotes the ring of* p*-adic integers.*

The counting techniques in this paper can be applied to solve a novel extremal problem in multiplicative number theory involving the greatest common divisors of pairs $(p-1,p'-1)$, where p and p' are prime numbers. The solution of this problem does not appear amenable to the standard techniques used in analytic number theory. Considering this problem first was crucial for obtaining Theorem 5.

THEOREM 9. *For* $n\to\infty$, *let*

$$M(n)=\max\Big\{\prod_{p,p'\in\mathcal{P}}\gcd(p-1,p'-1)\;\Big|\;\mathcal{P}\textit{ is a set of distinct primes where }\prod_{p\in\mathcal{P}}p\leqslant n\Big\}.$$

Then we have

$$\lim_{n\to\infty}\frac{\log M(n)}{\lambda(n)}=\frac{1}{4},$$

where $\lambda(n)=(\log n)^2/\log\log n$.

The paper is organized as follows.

In §1, we present some required preliminaries and notation.

In §2, we introduce the notion of a Bombieri set, which is the crucial ingredient needed in the proof of the lower bounds. We then use it in §3 and §4 to prove the lower bounds of Theorems 2, 3, 4 and 6. We then turn to the proof of the upper bounds. In §5, we show how the counting problem of congruence subgroups in $\mathrm{SL}_2(\mathbf{Z})$ can be completely reduced to an extremal counting problem of subgroups of finite abelian groups; the problem is actually, as one may expect, a number-theoretic extremal problem—see §6 and §7, where this extremal problem is solved. The upper bounds of Theorems 2, 3, and 4 are then deduced in §8. Finally, in §9 we prove Theorem 9.

The results of this paper are announced in [GLNP].

The authors would like to thank J.-P. Serre and the referees for the many comments which helped to improve the exposition of this paper.

1. Preliminaries and notation

Throughout this paper we let

$$l(n)=\frac{\log n}{\log\log n}\quad\text{and}\quad\lambda(n)=\frac{(\log n)^2}{\log\log n}.$$

All logarithms in this paper are to base e. If f and g are functions of n, we will say that *f is small with respect to g* if

$$\lim_{n\to\infty}\frac{\log f(n)}{\log g(n)}=0.$$

We say that f is *small* if f is small with respect to $n^{l(n)}$. Note that if f is small, then multiplying $C_n(\Gamma)$ by f will have no effect on the estimates of $\alpha_+(\Gamma)$ or $\alpha_-(\Gamma)$. We may, and we will, ignore factors which are small.

Note also that if $\varepsilon(n)$ is a function of n (bounded away from 0) which is small with respect to n, i.e., $\log\varepsilon(n)=o(\log n)$, then

$$\varlimsup_{n\to\infty}\frac{\log C_{n\varepsilon(n)}(\Gamma)}{\lambda(n)}=\alpha_+(\Gamma) \tag{1.1}$$

and

$$\varliminf_{n\to\infty}\frac{\log C_{n\varepsilon(n)}(\Gamma)}{\lambda(n)}=\alpha_-(\Gamma). \tag{1.2}$$

The proof of (1.1) follows immediately from the inequalities

$$\begin{aligned}\alpha_+(\Gamma)=\varlimsup_{n\to\infty}\frac{\log C_n(\Gamma)}{\lambda(n)}&\leqslant\varlimsup_{n\to\infty}\frac{\log C_{n\varepsilon(n)}(\Gamma)}{\lambda(n)}\\&=\varlimsup_{n\to\infty}\frac{\log C_{n\varepsilon(n)}(\Gamma)}{\lambda(n\varepsilon(n))}\,\frac{\lambda(n\varepsilon(n))}{\lambda(n)}\leqslant\alpha_+(\Gamma)\cdot 1=\alpha_+(\Gamma).\end{aligned}$$

Here, we have used the fact that

$$\varlimsup_{n\to\infty}\frac{\lambda(n\varepsilon(n))}{\lambda(n)}=1,$$

which is an immediate consequence of the assumption that $\varepsilon(n)$ is small with respect to n. A similar argument proves (1.2).

It follows that we can, and we will sometimes indeed, enlarge n a bit when evaluating $C_n(\Gamma)$, again without influencing α_+ or α_-. Similar remarks apply if we divide n by $\varepsilon(n)$.

The following lemma is proved in [Lu] in a slightly weaker form, and in its current form is proved in [LuS, Proposition 5.1.1].

LEMMA 1.1. ("Level versus index") *Let Γ be as before. Then there exists a constant $c>0$ such that if H is a congruence subgroup of Γ of index at most n, then H contains $\Gamma(m)$ for some $m\leqslant cn$, where $m\in\mathbf{Z}$, and by $\Gamma(m)$ we mean $\Gamma(m\mathcal{O}_S)$.*

COROLLARY 1.2. *Let* $\gamma_n(\Gamma)=\sum_{m=1}^{n} s_n(G(\mathcal{O}_S/m\mathcal{O}_S))$, *where for a group* H, $s_n(H)$ *denotes the number of subgroups of* H *of index at most* n. *Then we have*

$$\alpha_+(\Gamma)=\overline{\lim_{n\to\infty}}\frac{\log\gamma_n(\Gamma)}{\lambda(n)} \quad \text{and} \quad \alpha_-(\Gamma)=\underline{\lim_{n\to\infty}}\frac{\log\gamma_n(\Gamma)}{\lambda(n)}.$$

Proof. By Lemma 1.1, $C_n(\Gamma)\leqslant\gamma_{cn}(\Gamma)$ for some $c>0$. It is also clear that $\gamma_n(\Gamma)\leqslant nC_n(\Gamma)$. Since c is small with respect to n, Corollary 1.2 follows by arguments of the type we have given above. □

The number of elements in a finite set X is denoted by $\#X$ or $|X|$. The set of subgroups of a group G is denoted by $\mathrm{Sub}(G)$.

2. Bombieri sets

We introduce some additional notation. Let a and q be relatively prime integers with $q>0$. For $x>0$, let $\mathcal{P}(x;q,a)$ be the set of primes p with $p\leqslant x$ and $p\equiv a \pmod q$. For $a=1$, we set $\mathcal{P}(x,q)=\mathcal{P}(x;q,1)$. We also define

$$\vartheta(x;q,a)=\sum_{p\in\mathcal{P}(x;q,a)}\log p.$$

If $f(x)$ and $g(x)$ are arbitrary functions of a real variable x, we say that $f(x)\sim g(x)$ as $x\to\infty$ if

$$\lim_{x\to\infty}\frac{f(x)}{g(x)}=1.$$

Define the error term

$$E(x;q,a)=\vartheta(x;q,a)-\frac{x}{\phi(q)},$$

where $\phi(q)$ is Euler's function. Then Bombieri proved the following deep theorem [Bo], [Da].

THEOREM 2.1. (Bombieri) *Let* $A>0$ *be fixed. Then there exists a constant* $c(A)>0$ *such that*

$$\sum_{q\leqslant\sqrt{x}/(\log x)^A}\max_{y\leqslant x}\max_{(a,q)=1}|E(y;q,a)|\leqslant c(A)\frac{x}{(\log x)^{A-5}}$$

for $x>3$.

This theorem shows that the error terms $\max_{(a,q)=1}E(x;q,a)$ behave as if they satisfy the Riemann hypothesis in an averaged sense.

Definition 2.2. Let $x>3$. A *Bombieri prime* (relative to x) is a prime $q\leqslant\sqrt{x}$ such that the set $\mathcal{P}(x,q)$ of primes $p\leqslant x$ with $p\equiv 1 \pmod q$ satisfies

$$\max_{y\leqslant x}|E(y;q,1)|\leqslant \frac{x}{\phi(q)(\log x)^2}.$$

We say that $\mathcal{P}(x,q)$ is a *Bombieri set* (relative to x).

Remark. In all the applications in this paper, we do not really need q to be prime, though it makes the calculations somewhat easier. We could work with q being a "Bombieri number".

LEMMA 2.3. *Fix $0<\varrho<\frac{1}{2}$. Then for x sufficiently large, there exists at least one Bombieri prime q (relative to x) in the interval*

$$\frac{x^{\varrho}}{\log x}\leqslant q\leqslant x^{\varrho}.$$

Proof. Assume that

$$\max_{y\leqslant x}|E(y;q,1)|> \frac{x}{\phi(q)(\log x)^2}$$

for all primes $x^{\varrho}/\log x\leqslant q\leqslant x^{\varrho}$, i.e., that there are no such Bombieri primes in the interval. In view of the trivial inequality, $\phi(q)=q-1<q$, it immediately follows that

$$\sum_{x^{\varrho}/\log x\leqslant q\leqslant x^{\varrho}}\max_{y\leqslant x}|E(y;q,1)|>\frac{x}{(\log x)^2}\sum_{x^{\varrho}/\log x\leqslant q\leqslant x^{\varrho}}\frac{1}{q}>\frac{x\log\log x}{2\varrho(\log x)^3},$$

say, for sufficiently large x. This follows from the well-known asymptotic formula [Ld] for the partial sum of the reciprocal of the primes

$$\sum_{q\leqslant Y}\frac{1}{q}=\log\log Y+b+O\left(\frac{1}{\log Y}\right)$$

as $Y\to\infty$. Here b is an absolute constant. This contradicts Theorem 2.1 with $A\geqslant 8$ provided x is sufficiently large. □

LEMMA 2.4. *Let $\mathcal{P}(x,q)$ be a Bombieri set. Then for x sufficiently large,*

$$\left|\#\mathcal{P}(x,q)-\frac{x}{\phi(q)\log x}\right|\leqslant 3\frac{x}{\phi(q)(\log x)^2}.$$

Proof. For a real number θ, define $\lfloor\theta\rfloor$ to be the largest integer t such that $t\leqslant\theta$. We have

$$\begin{aligned}\sum_{p\in\mathcal{P}(x,q)} 1 &= \sum_{n=2}^{x}\frac{\vartheta(n;q,1)-\vartheta(n-1;q,1)}{\log n}\\ &= \sum_{n=2}^{x}\vartheta(n;q,1)\left(\frac{1}{\log n}-\frac{1}{\log(n+1)}\right)+\frac{\vartheta(x;q,1)}{\log(\lfloor x\rfloor+1)}\\ &= \sum_{n=2}^{x}\vartheta(n;q,1)\frac{\log(1+1/n)}{\log n\log(n+1)}+\frac{\vartheta(x;q,1)}{\log x}-\vartheta(x;q,1)\left(\frac{1}{\log x}-\frac{1}{\log(\lfloor x\rfloor+1)}\right).\end{aligned}$$

It easily follows that

$$\left|\sum_{p\in\mathcal{P}(x,q)} 1-\frac{\vartheta(x;q,1)}{\log x}\right|\leqslant\sum_{n=2}^{x}\vartheta(n;q,1)\frac{1}{n(\log n)^2}+\vartheta(x;q,1)\left(\frac{1}{\log x}-\frac{1}{\log(x+1)}\right).$$

By the property of a Bombieri set, we have the estimate

$$\left|\vartheta(n;q,1)-\frac{n}{\phi(q)}\right|\leqslant\frac{x}{\phi(q)(\log x)^2}$$

for $n\leqslant x$. Since

$$\left(\frac{1}{\log x}-\frac{1}{\log(x+1)}\right)=\frac{\log(1+1/x)}{\log x\log(x+1)}=O\left(\frac{1}{x(\log x)^2}\right),$$

the second expression on the right-hand side of the above equation is very small and can be ignored. It remains to estimate the sum

$$\sum_{n=2}^{x}\vartheta(n;q,1)\frac{1}{n(\log n)^2}.$$

This sum can be broken into two parts, the first of which corresponds to $n\leqslant x/(\log x)^3$, which is easily seen to be very small, and so can be ignored. We estimate

$$\begin{aligned}\sum_{x/(\log x)^3\leqslant n\leqslant x}\vartheta(n;q,1)\frac{1}{n(\log n)^2} &= \sum_{x/(\log x)^3\leqslant n\leqslant x}\frac{n}{\phi(q)}\,\frac{1}{n(\log n)^2}\\ &\quad+O\left(\sum_{x/(\log x)^3\leqslant n\leqslant x}\frac{x}{\phi(q)(\log x)^2}\,\frac{1}{n(\log n)^2}\right)\\ &= \sum_{x/(\log x)^3\leqslant n\leqslant x}\frac{1}{\phi(q)(\log n)^2}+O\left(\frac{x}{\phi(q)(\log x)^3}\right)\\ &\leqslant\frac{3}{2}\,\frac{x}{\phi(q)(\log x)^2},\end{aligned}$$

which holds for x sufficiently large and where the constant $\frac{3}{2}$ is not optimal. Hence

$$\left|\sum_{p\in\mathcal{P}(x,q)} 1-\frac{\vartheta(x;q,1)}{\log x}\right| \leqslant \frac{7}{4}\frac{x}{\phi(q)(\log x)^2},$$

say. Since

$$\left|\vartheta(x;q,1)-\frac{x}{\phi(q)}\right| \leqslant \frac{x}{\phi(q)(\log x)^2},$$

Lemma 2.4 immediately follows. □

3. Proof of the lower bound over Q

In this section we consider the case of $k=\mathbf{Q}$ and $\mathcal{O}=\mathbf{Z}$.

Fix a real number $0<\varrho_0<\frac{1}{2}$. It follows from Lemma 2.3 that for $x\to\infty$ there exists a real number ϱ which converges to ϱ_0, and a prime number $q\sim x^\varrho$ such that $\mathcal{P}(x,q)$ is a Bombieri set.

Define

$$P=\prod_{p\in\mathcal{P}(x,q)} p.$$

It is clear from the definition of a Bombieri set that

$$\log P\sim\frac{x}{\phi(q)}\sim x^{1-\varrho}$$

and from Lemma 2.4 that

$$L=\#\mathcal{P}(x,q)\sim\frac{x}{\phi(q)\log x}\sim\frac{x^{1-\varrho}}{\log x}.$$

Consider $\Gamma(P)=\ker(G(\mathbf{Z})\to G(\mathbf{Z}/P\mathbf{Z}))$, which is of index at most $P^{\dim G}$ in Γ. Note that for every subgroup $H/\Gamma(P)$ in $\Gamma/\Gamma(P)$ there corresponds a subgroup H in Γ of index at most $P^{\dim G}$ in Γ.

By strong approximation

$$\Gamma/\Gamma(P)=G(\mathbf{Z}/P\mathbf{Z})\cong\prod_{p\in\mathcal{P}(x,q)} G(\mathbf{F}_p).$$

Let $B(p)$ denote the Borel subgroup in $G(\mathbf{F}_p)$. Then

$$\log\#B(p)\sim\tfrac{1}{2}(\dim G+\mathrm{rk}(G))\log p,$$

where $\mathrm{rk}(G)$ denotes the rank of G as an algebraic group. But

$$\log \#G(\mathbf{F}_p) \sim \dim(G) \log p.$$

It immediately follows that (for $p \to \infty$)

$$\log [G(\mathbf{F}_p) : B(p)] \sim \tfrac{1}{2}(\dim G - \mathrm{rk}(G)) \log p,$$

and, therefore,

$$\log [G(\mathbf{Z}/P\mathbf{Z}) : B(P)] \sim \tfrac{1}{2}(\dim G - \mathrm{rk}(G)) \log P,$$

where $B(P) \leqslant G(\mathbf{Z}/P\mathbf{Z})$ is

$$B(P) = \prod_{p \in \mathcal{P}(x,q)} B(\mathbf{F}_p).$$

Now $B(p)$ is mapped onto $(\mathbf{F}_p^\times)^{\mathrm{rk}(G)}$ and, hence, is also mapped onto $(\mathbf{Z}/q\mathbf{Z})^{\mathrm{rk}(G)}$ since $\#\mathbf{F}_p^\times = p-1$ and $p \equiv 1 \pmod{q}$. So $B(P)$ is mapped onto

$$(\mathbf{Z}/q\mathbf{Z})^{\mathrm{rk}(G)L},$$

where

$$L = \#\mathcal{P}(x,q) \sim \frac{x}{\phi(q) \log x} \sim \frac{x^{1-\varrho}}{\log x}.$$

For a real number θ, define $\lceil \theta \rceil$ to be the smallest integer t such that $\theta \leqslant t$. Let $0 \leqslant \sigma \leqslant 1$.

We will now use Proposition 6.1, a basic result on counting subspaces of finite vector spaces. It follows that $B(P)$ has at least

$$q^{\sigma(1-\sigma)\mathrm{rk}(G)^2 L^2 + O(\mathrm{rk}(G)L)}$$

subgroups of index equal to

$$\iota = q^{\lceil \sigma \mathrm{rk}(G) L \rceil} [G(\mathbf{Z}/P\mathbf{Z}) : B(P)].$$

Hence, for $x \to \infty$,

$$\begin{aligned}
\log \#\{\text{subgroups}\} &= (\sigma(1-\sigma)\,\mathrm{rk}(G)^2 L^2 + O(\mathrm{rk}(G)L)) \log q \\
&\sim \sigma(1-\sigma)\,\mathrm{rk}(G)^2 \frac{x^{2-2\varrho}}{(\log x)^2}\, \varrho \log x,
\end{aligned}$$

while

$$\begin{aligned}
\log \iota &= \lceil \sigma\, \mathrm{rk}(G) L \rceil \log q + \tfrac{1}{2}(\dim G - \mathrm{rk}(G)) \log P \\
&\sim \mathrm{rk}(G) \sigma \frac{x^{1-\varrho}}{\log x} \varrho \log x + \tfrac{1}{2}(\dim G - \mathrm{rk}(G)) x^{1-\varrho} \\
&= \big(\sigma \varrho\, \mathrm{rk}(G) + \tfrac{1}{2}(\dim G - \mathrm{rk}(G))\big) x^{1-\varrho}
\end{aligned}$$

and

$$\log\log\iota \sim (1-\varrho)\log x.$$

It is clear from the estimate for $\log\iota$ above that given any index $n \gg 0$ we can choose x such that $\log\iota \sim \log n$. We compute

$$\frac{\log\#\{\text{subgroups}\}}{(\log(\text{index}))^2/\log\log(\text{index})} \sim \frac{\sigma(1-\sigma)\operatorname{rk}(G)^2\varrho x^{2-2\varrho}/\log x}{\left((\sigma\varrho\operatorname{rk}(G)+\frac{1}{2}(\dim G-\operatorname{rk}(G)))\,x^{1-\varrho}\right)^2/(1-\varrho)\log x}$$
$$\sim \frac{\sigma(1-\sigma)\varrho(1-\varrho)\operatorname{rk}(G)^2}{\left((\sigma\varrho-\frac{1}{2})\operatorname{rk}(G)+\frac{1}{2}\dim G\right)^2}$$

as $x\to\infty$.

We may rewrite

$$\frac{\sigma(1-\sigma)\varrho(1-\varrho)\operatorname{rk}(G)^2}{\left((\sigma\varrho-\frac{1}{2})\operatorname{rk}(G)+\frac{1}{2}\dim G\right)^2} = \frac{\sigma(1-\sigma)\varrho(1-\varrho)}{(\sigma\varrho+R)^2},$$

where

$$R = \frac{\dim G - \operatorname{rk}(G)}{2\operatorname{rk}(G)}.$$

Now, for fixed R, it is enough to choose σ and ϱ so that

$$\frac{\sigma(1-\sigma)\varrho(1-\varrho)}{(\sigma\varrho+R)^2}$$

is maximized. This occurs when

$$\varrho = \sigma = \sqrt{R(R+1)} - R,$$

in which case we get

$$\frac{\sigma(1-\sigma)\varrho(1-\varrho)}{(\sigma\varrho+R)^2} = \frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2}.$$

Actually, we choose ϱ_0 to be $\sqrt{R(R+1)} - R$. Then we can take ϱ to be asymptotic to ϱ_0 as x tends to infinity. Note that

$$\frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2} < \frac{1}{16R^2}$$

holds for all $R>0$. This follows from the easy inequality $\sqrt{R(R+1)} - R \leqslant \frac{1}{2}$. It is also straightforward to see that $\sqrt{R(R+1)} - R$ converges to $\frac{1}{2}$ as $R\to\infty$. Hence

$$\frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2} \sim \frac{1}{16R^2}.$$

In the special case when $R=1$, we obtain the lower bound of Theorem 2. For a simple Chevalley group scheme over $\mathbf{Q}$, this gives the lower bound in Theorem 6.

4. Proof of the lower bound for a general number field

To prove the lower bounds over a general number field we need an extension of the Bombieri–Vinogradov theorem to these fields, as was obtained by Murty and Murty [MM].

Let us first fix some notation:

Let k be a finite extension of degree f over $\mathbf{Q}$, K its Galois closure of degree d, $\mathfrak{g}=\mathrm{Gal}(K/\mathbf{Q})$, and $\mathcal{O}_k$ the ring of integers in k. For a rational prime q and $x\in\mathbf{R}$, we will denote by $\widetilde{\mathcal{P}}_K(x,q)$ the set of rational primes $p\equiv 1 \pmod q$, where p splits completely in K and $p\leqslant x$. Let

$$\tilde{\pi}_K(x,q)=\#\widetilde{\mathcal{P}}_K(x,q),\quad \tilde{\nu}_K(x,q)=\sum_{p\in\widetilde{\mathcal{P}}_K(x,q)}\log p$$

and

$$\widetilde{E}_K(x,q)=\tilde{\nu}_K(x,q)-\frac{x}{d\phi(q)}.$$

We shall show that the following theorems follow from Murty and Murty [MM].

THEOREM 4.1. *Let K be a fixed finite Galois extension of $\mathbf{Q}$. Assume the GRH (generalized Riemann hypothesis) for K and all cyclotomic extensions $K(\zeta_l)$ with l a rational prime and ζ_l a primitive l-th root of unity. Then for every $0<\varrho<\frac{1}{2}$ there exists a number $X=X(K,\varrho)$ (X depends on K and ϱ) such that if $x>X$, one can find a rational prime q satisfying*

(a) $\displaystyle\frac{x^{\varrho}}{\log x}\leqslant q\leqslant x^{\varrho}$;

(b) $\displaystyle\left|\tilde{\pi}_K(x,q)-\frac{x}{d'\phi(q)\log x}\right|\leqslant 3\frac{x}{d'\phi(q)(\log x)^2}$;

(c) $\displaystyle\max_{y\leqslant x}|\widetilde{E}_K(y,q)|\leqslant\frac{x}{d'\phi(q)(\log x)^2}$,

where $d'=[K:\mathbf{Q}]/t$ and t denotes the degree of the intersection of K and the cyclotomic field $\mathbf{Q}(\zeta_q)$ over $\mathbf{Q}$.

Remark. In fact, the GRH gives a stronger result than what is stated in Theorem 4.1. For example, it can be shown that for every prime $q<x^{1/2}$ the error terms in parts (b) and (c) take the form $\mathcal{O}(x^{1/2}\log(qx))$ (see [MMS] for a more precise bound). Theorem 4.1 is stated in this special form because it can be proved unconditionally in some cases.

THEOREM 4.2. *Theorem 4.1 can be proved unconditionally for K if either*

(a) $\mathfrak{g}=\mathrm{Gal}(K/\mathbf{Q})$ *has an abelian subgroup of index at most* 4 (*this is true, for example, if k is an abelian extension*), *or*

(b) $[K:\mathbf{Q}]<42$.

THEOREM 4.3. *Theorem* 4.1 *is valid unconditionally for every* K *with the additional assumption that* $0<\varrho<1/\eta$, *where* η *is the maximum of* 2 *and* d^*-2, *and where* d^* *is the index of the largest possible abelian subgroup of* $\mathfrak{g}=\mathrm{Gal}(K/\mathbf{Q})$. *In particular, we may take* $\eta=d^*-2$ *if* $d^*\geqslant 4$, *and* $\eta=2$ *if* $d^*\leqslant 4$.

Proof of Theorems 4.1–4.3. For any $\varepsilon>0$ and $A>0$, under the assumptions of Theorem 4.1 or Theorem 4.2 (a), Murty and Murty [MM] prove the following Bombieri theorem:

$$\sum_{q\leqslant x^{1/2-\varepsilon}} \max_{(a,q)=1} \max_{y\leqslant x} \left| \pi_C(y,q,a) - \frac{|C|}{|G|}\frac{1}{\phi(q)}\pi(y) \right| \ll \frac{x}{(\log x)^A}. \tag{4.1}$$

Here C denotes a conjugacy class in $\mathfrak{g}$, $\pi(y)=\sum_{p\leqslant y} 1$,

$$\pi_C(x,q,a) = \sum_{\substack{p\leqslant x \\ (p,K/\mathbf{Q})=C \\ p\equiv a \ (\mathrm{mod}\ q) \\ p \text{ unramified in } K}} 1$$

and $(p,K/\mathbf{Q})$ denotes the Artin symbol.

In fact, under the assumption of the GRH, equation (4.1) holds, but without assuming the GRH they showed that (4.1) holds when the sum is over $q<x^{1/\eta-\varepsilon}$, where η is defined as follows: Let

$$d^* = \min_H \max_w \,[\mathfrak{g}:H]\,w(1). \tag{4.2}$$

The minimum here is over all subgroups H of $\mathrm{Gal}(K/\mathbf{Q})$ satisfying

(i) $H\cap C\neq\varnothing$;

(ii) for every irreducible character w of H and any non-trivial Dirichlet character χ, the Artin L-series $L(s,w\otimes\chi)$ is entire.

Then the maximum in (4.2) is over the irreducible characters of such H's.

Now

$$\eta = \begin{cases} d^*-2, & \text{if } d^*\geqslant 4, \\ 2, & \text{if } d^*\leqslant 4. \end{cases}$$

We need their result for the special case when C is the identity conjugacy class. In this case, $|C|/|\mathfrak{g}|=1/d'$ and $\pi_C(y,q,1)=\tilde{\pi}_k(y,q)$. So for proving Theorem 4.3 we can take for H an abelian subgroup of smallest index, and then H satisfies assumption (i) and (ii). (Recall that abelian groups satisfy (AC)—the Artin conjecture, i.e., $L(s,w\otimes\chi)$ are entire—see [H, §3]).

For Theorem 4.2 (a), again take H to be the abelian subgroup of index at most 4. It satisfies (i) and (ii), and this time $\eta=2$.

For Theorem 4.2 (b), going case by case over all possible numbers $d<42$, one can deduce by elementary group-theoretic arguments that every finite group $\mathfrak{g}$ of order $d<42$ has an abelian subgroup of index at most 4, unless $d=24$ and $\mathfrak{g}$ is isomorphic to the symmetric group S_4. But for this group, Artin [H, §3] proved Artin's conjecture in 1925. Moreover, every irreducible character of S_4 is of degree at most 4. Thus for $\mathfrak{g}=S_4$ we have $d^*=4$, and so $\eta=2$.

The proofs of Theorems 4.1–4.3 follow now in the same manner as in §2. □

Using Theorems 4.1–4.3, we can now prove the lower bounds of Theorems 3 and 4 just as in §3. Note that for every prime $p\in\widetilde{\mathcal{P}}_K(x,q)$ we may take an ideal $\pi=\pi(p)$ in $\mathcal{O}_k$ with $[\mathcal{O}_k:\pi]=p$ and $\pi\cap\mathbf{Z}=p\mathbf{Z}$. Let

$$P=\prod_{p\in\widetilde{\mathcal{P}}_K(x,q)}\pi(p).$$

Then, since $x\to\infty$, we may choose q and ϱ (using Theorem 4.1) so that

$$\log\,[\mathcal{O}:P]\sim\frac{x}{d\phi(q)}\sim\frac{x^{1-\varrho}}{d},\quad L:=|\mathcal{P}_K(x,q)|\sim\frac{x}{d\phi(q)\log x}\sim\frac{x^{1-\varrho}}{d\log x}$$

and

$$G(\mathcal{O}/P)=\prod_{p\in\mathcal{P}_K(x,q)}G(\mathcal{O}/\pi(p))\simeq\prod_{p\in\widetilde{\mathcal{P}}_k(x,q)}G(\mathbf{Z}/p\mathbf{Z}).$$

We can now take for every rational prime $p\in\widetilde{\mathcal{P}}_k(x,q)$, the Borel subgroup $B(p)$ as in §3 and define

$$B(P)=\prod_{p\in\widetilde{\mathcal{P}}_k(x,q)}B(p).$$

Then $B(P)$ is mapped onto $(\mathbf{Z}/q\mathbf{Z})^{\mathrm{rk}(G)L}$ and

$$\log\,[G(\mathcal{O}/P):B(P)]\sim\tfrac{1}{2}(\dim G-\mathrm{rk}(G))\log\,[\mathcal{O}:P].$$

Thus, by a computation similar to the one in §3 (note that the d's cancel in this computation), we can show that

$$\alpha_-(G(\mathcal{O}))\geqslant\frac{\left(\sqrt{R(R+1)}-R\right)^2}{4R^2}.$$

The lower bounds of Theorems 3, 4 and 6 are now also proved. We now turn to the proof of the upper bounds.

5. From SL_2 to abelian groups

In this section we show how to reduce the estimation of $\alpha_+(\mathrm{SL}_2(\mathbf{Z}))$ to a problem on abelian groups.

Corollary 1.2 shows us that in order to give an upper bound on $\alpha_+(\Gamma)$ it suffices to bound $s_n(G(\mathbf{Z}/m\mathbf{Z}))$ when $m\leqslant n$. Our first goal is to show that we can further assume that m is a product of different primes. To this end let $\overline{m}=\prod p$, where p runs through all the primes dividing m.

We have an exact sequence

$$1\longrightarrow K\longrightarrow G(\mathbf{Z}/m\mathbf{Z})\xrightarrow{\pi} G(\mathbf{Z}/\overline{m}\mathbf{Z})\longrightarrow 1,$$

where K is a nilpotent group of rank at most $\dim G$. Here, the rank of a finite group G is defined to be the smallest integer r such that every subgroup of G is generated by r elements (see [LuS, Window 5, §2]).

LEMMA 5.1. *Let $1\rightarrow K\rightarrow U\xrightarrow{\pi} L\rightarrow 1$ be an exact sequence of finite groups, where K is a solvable group of derived length l and of rank at most r. Then the number of supplements to K in U (i.e., of subgroups H of U for which $\pi(H)=L$) is bounded by $|U|^{3r^2+lr}$.*

Proof. See [LuS, Corollary 1.3.5]. □

COROLLARY 5.2. $s_n(G(\mathbf{Z}/m\mathbf{Z}))\leqslant m^{f'(\dim G)\log\log m}s_n(G(\mathbf{Z}/\overline{m}\mathbf{Z}))$, *where $f'(\dim G)$ depends only on* $\dim G$.

Proof. Let H be a subgroup of index at most n in $G(\mathbf{Z}/m\mathbf{Z})$ and let $L=\pi(H)\leqslant G(\mathbf{Z}/\overline{m}\mathbf{Z})$. Thus L is of index at most n in $G(\mathbf{Z}/\overline{m}\mathbf{Z})$. Let $U=\pi^{-1}(L)$, so every subgroup H of $G(\mathbf{Z}/m\mathbf{Z})$ with $\pi(H)=L$ is a subgroup of U. Given L (and hence also U) we have the exact sequence $1\rightarrow K\rightarrow U\xrightarrow{\pi} L\rightarrow 1$, and by Lemma 5.1, the number of H in U with $\pi(H)=L$ is at most $|U|^{lf(r)}$, where l is the derived length of K, $r\leqslant\dim G$ is the rank of K and $f(r)\leqslant f(\dim G)$, where f is some function depending on r and independent of m (say $f(r)=3r^2+r$). Now $|U|\leqslant m^{\dim G}$, and K being nilpotent is of derived length $O(\log\log|K|)$. We can, therefore, deduce that

$$s_n(G(\mathbf{Z}/m\mathbf{Z}))\leqslant m^{c\dim(G)f(\dim G)(\log\log m+\log\dim G)}s_n(G(\mathbf{Z}/\overline{m}\mathbf{Z}))$$

for some constant c, which proves our claim. □

Corollary 1.2 shows us that in order to estimate $\alpha_+(G(\mathbf{Z}))$ one should concentrate on $s_n(G(\mathbf{Z}/m\mathbf{Z}))$ with $m\leqslant n$. Corollary 5.2 implies that we can further assume that m is a product of different primes. So let us now assume that $m=\prod_{i=1}^t q_i$, where the q_i are different primes, and so $G(\mathbf{Z}/m\mathbf{Z})\simeq\prod_{i=1}^t G(\mathbf{Z}/q_i\mathbf{Z})$ and $t\leqslant(1+o(1))\log m/\log\log m$.

We can further assume that we are counting only fully proper subgroups of $G(\mathbf{Z}/m\mathbf{Z})$, i.e., subgroups H which do not contain $G(\mathbf{Z}/q_i\mathbf{Z})$ for any $1\leqslant i\leqslant t$, or equivalently, the image of H under the projection to $G(\mathbf{Z}/q_i\mathbf{Z})$ is a proper subgroup (see [Lu]). Thus H is contained in $\prod_{i=1}^t M_i$, where M_i is a maximal subgroup of $G(\mathbf{Z}/q_i\mathbf{Z})$.

Let us now specialize to the case $G=\mathrm{SL}_2$, and let q be a prime.

Maximal subgroups of $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ are conjugate to one of the following three types of subgroups (see [Lg, Theorems 2.2 and 2.3, pp. 183–185]).

(1) $B=B_q$—the Borel subgroup of all upper triangular matrices in SL_2.

(2) $D=D_q^-, D_q^+$—subgroups of dihedral type of order $2(q-1)$ or $2(q+1)$ (inverse images of dihedral maximal subgroups of $\mathrm{PSL}_2(\mathbf{Z}/q\mathbf{Z})$). To simplify our notation we denote these groups by D_q. The group D_q is equal to $N(T_q)$, the normaliser of a split or non-split torus T_q (a cyclic group of order $q-1$ or $q+1$). The group T_q is either the diagonal subgroup, or is obtained as follows: Let $\mathbf{F}_{q^2}$ be the field of order q^2, $\mathbf{F}_{q^2}^\times$ acts on $\mathbf{F}_{q^2}$ by multiplication. The latter is a 2-dimensional vector space over $\mathbf{F}_q$. The elements of norm 1 in $\mathbf{F}_{q^2}^\times$ induce the subgroup T_q of $\mathrm{SL}_2(\mathbf{F}_q)$.

(3) $A=A_q$—a subgroup of $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ which is of order at most 120.

There are only boundedly many conjugacy classes of each type. Also, the number of conjugates of every subgroup is small, so it suffices to count only subgroups of $\mathrm{SL}_2(\mathbf{Z}/m\mathbf{Z})$ whose projection to $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ (for $q|m$) is inside either B, D or A.

Let $S\subseteq\{q_1,...,q_t\}$ be the subset of the prime divisors of m for which the projection of H is in A_{q_i}, and let $\bar S$ be the complement to S. Let $\overline{m}=\prod_{q\in\bar S} q$ and $\overline{H}$ be the projection of H to $\mathrm{SL}_2(\mathbf{Z}/\overline{m}\mathbf{Z})$. Thus $\overline{H}$ is a subgroup of index at most n in $\mathrm{SL}_2(\mathbf{Z}/\overline{m}\mathbf{Z})$, and the kernel N from $H\to\overline{H}$ is inside a product of $|S|$ groups of type A. As every subgroup of $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ is generated by two elements, H is generated by at most $2\log m/\log\log m\leqslant 2\log n/\log\log n$ generators. Set $k=\lfloor 2\log n/\log\log n+1\rfloor$ and choose k generators for $\overline{H}$. By a lemma of Gaschütz (cf. [FJ, Lemma 15.30]) these k generators can be lifted up to give k generators for H. Each generator can be lifted up in at most $|N|$ ways, and N is a group of order at most $120^{|S|}\leqslant 120^t\leqslant 120^{\log n/\log\log n}$. We, therefore, conclude that given $\overline{H}$, the number of possibilities for H is at most $120^{2(\log n)^2/(\log\log n)^2}$, which is small with respect to $n^{l(n)}$.

We can, therefore, assume that $S=\phi$ and all the projections of H are either into groups of type B or type D.

Now, B_q, the Borel subgroup of $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$, has a normal unipotent cyclic subgroup U_q of order q. Let now S be the subset of $\{q_1,...,q_t\}$ for which the projection is in B, and let $\bar S$ be the complement. Then

$$H\leqslant\prod_{q\in S} B_q\times\prod_{q\in\bar S} D_q.$$

Let $\bar{H}$ be the projection of H to $\prod_{q\in S}(B_q/U_q)\times\prod_{q\in\bar{S}} D_q$. The kernel is a subgroup of the cyclic group $U=\prod_{q\in S} U_q$. By Lemma 5.1 we know that given $\bar{H}$, there are only a few possibilities for H. We are, therefore, led to counting subgroups in

$$L=\prod_{q\in S}(B_q/U_q)\times\prod_{q\in\bar{S}} D_q.$$

Let E now be the product

$$\prod_{q\in S}(B_q/U_q)\times\prod_{q\in\bar{S}} T_q,$$

and for a subgroup H of L we denote $H\cap E$ by $\bar{\bar{H}}$.

Our next goal will be to show that given $\bar{\bar{H}}$ in E, the number of possibilities for H is small. To this end we first formulate two easy lemmas, which will be used in the proof of Proposition 5.6 below. This proposition will complete the main reduction.

LEMMA 5.3. *Let H be a subgroup of $U=U_1\times U_2$. For $i=1,2$, set $H_i=\pi_i(H)$, where π_i is the projection from U to U_i, and $H_i^0=H\cap U_i$. Then*

(i) *H_i^0 is normal in H_i and $H_1/H_1^0\simeq H_2/H_2^0$ with an isomorphism φ induced by the inclusion of $H/(H_1^0\times H_2^0)$ as a subdirect product of H_1/H_1^0 and H_2/H_2^0;*

(ii) *H is determined by*

(a) *H_i for $i=1,2$;*

(b) *H_i^0 for $i=1,2$;*

(c) *the isomorphism φ from H_1/H_1^0 to H_2/H_2^0.*

Proof. See [Su, p. 141]. □

Definition 5.4. Let U be a group and V a subnormal subgroup of U. We say that V is copolycyclic in U of colength l if there is a sequence $V=V_0\lhd V_1\lhd\ldots\lhd V_l=U$ such that V_i/V_{i-1} is cyclic for every $i=1,\ldots,l$.

LEMMA 5.5. *Let U be a group and F a subgroup of U. The number of subnormal copolycyclic subgroups V of U containing F and of colength l is at most $|U:F|^l$.*

Proof. For $l=1$, V contains $[U,U]F$, and so it suffices to prove the lemma for the abelian group $\bar{U}=U/[U,U]F$ and $\bar{F}=\{e\}$. For an abelian group $\bar{U}$, the number of subgroups V with $\bar{U}/V$ cyclic is equal, by Pontrjagin duality, to the number of cyclic subgroups. This is clearly bounded by $|\bar{U}|\leqslant|U:F|$. If $l>1$, then by induction the number of possibilities for V_1 as in Definition 5.4 is bounded by $|U:F|^{l-1}$. Given V_1, the number of possibilities for V is at most $|V_1:F|\leqslant|U:F|$ by the case $l=1$. Thus, V has at most $|U:F|^l$ possibilities. □

PROPOSITION 5.6. *Let $D=D_1\times\ldots\times D_s$, where each D_i is a finite group with a cyclic subgroup T_i of index 2. Let $T=T_1\times\ldots\times T_s$ (and thus $|D:T|=2^s$). The number of subgroups H of D whose intersection with T is a given subgroup L of T is at most $|D|^8 2^{2s^2}$.*

Proof. Set $F_i=\prod_{j\geqslant i} D_i$. We want to count the number of subgroups H of D with $H\cap T=L$. Let $L_i=\mathrm{proj}_{F_i}(L)$, i.e., the projection of L to F_i, and $\tilde{L}_{i+1}=L_i\cap F_{i+1}$, so that $\tilde{L}_{i+1}\subseteq L_{i+1}$. Let H_i be the projection of H to F_i. Given H, the sequence $(H_1=H, H_2, ..., H_s)$ is determined, and, of course, vice versa. We will actually prove that the number of possibilities for $(H_1, ..., H_s)$ is at most $|D|^8 2^{2s^2}$.

Assume now that H_{i+1} is given. What is the number of possibilities for H_i? Well, H_i is a subgroup of $F_i=D_i\times F_{i+1}$ containing L_i, whose projection to F_{i+1} is H_{i+1}, and its intersection with F_{i+1}, which we will denote by X, contains $\tilde{L}_{i+1}$. By Lemma 5.3, H_i is determined by H_{i+1}, X, Y, Z and φ, where Y is the projection of H_i to D_i, $Z=H_i\cap D_i$ and φ is an isomorphism from Y/Z to H_{i+1}/X. Now, every subgroup of the group D_i is generated by two elements, so the number of possibilities for Y and Z is at most $|D_i|^2$ each, and the number of automorphisms of Y/Z is also at most $|D_i|^2$.

Let us now look at X: X is a normal subgroup of H_{i+1} with H_{i+1}/X isomorphic to Y/Z, so it is meta-cyclic. Moreover, X contains $\tilde{L}_{i+1}$. So by Lemma 5.5, the number of possibilities for X is at most $|H_{i+1}:\tilde{L}_{i+1}|^2$.

Now $|H_{i+1}:\tilde{L}_{i+1}|\leqslant|H_{i+1}:L_{i+1}|\cdot|L_{i+1}:\tilde{L}_{i+1}|$. We know that

$$|H_{i+1}:L_{i+1}|=|\mathrm{proj}_{F_{i+1}}(H):\mathrm{proj}_{F_{i+1}}(L)|\leqslant|H:L|\leqslant 2^s$$

and

$$|L_{i+1}:\tilde{L}_{i+1}|=|\mathrm{proj}_{F_{i+1}}(L_i):F_{i+1}\cap L_i|\leqslant|D_i|.$$

So, $|H_{i+1}:\tilde{L}_{i+1}|\leqslant 2^s|D_i|$.

Altogether, given H_{i+1} (and L, and hence also the L_i's and $\tilde{L}_i$'s), the number of possibilities for H_i is at most $|D_i|^8 2^{2s}$. Arguing now by induction, we deduce that the number of possibilities for $(H_1, ..., H_s)$ is at most $|D|^8 2^{2s^2}$ as claimed. □

Let us now get back to SL_2: Proposition 5.6 implies, in the notation given before Lemma 5.3, that when counting subgroups of

$$L=\prod_{q\in S}(B_q/U_q)\times\prod_{q\in\bar{S}} D_q,$$

we can instead count the subgroups of

$$E=\prod_{q\in S}(B_q/U_q)\times\prod_{q\in\bar{S}} T_q,$$

where T_q is a torus in $\mathrm{SL}_2(\mathbf{Z}/q\mathbf{Z})$ (so that T_q is a cyclic group of order $q-1$ or $q+1$, while B_q/U_q is a cyclic group of order $q-1$).

A remark is needed here: Let H be a subgroup of index at most n in $\mathrm{SL}_2(\mathbf{Z}/m\mathbf{Z})$ which is contained in $X=\prod_{q\in S} B_q\times\prod_{q\in\bar{S}} D_q$ and contains $Y=\prod_{q\in S} U_q\times\prod_{q\in\bar{S}}\{e\}$. By our analysis in this section, these are the groups which we have to count in order to determine $\alpha_+(\mathrm{SL}_2(\mathbf{Z}))$. We proved that for counting them, it suffices for us to count subgroups of X_0/Y, where $X_0=\prod_{q\in S} B_q\times\prod_{q\in\bar{S}} T_q$. Note though that replacing H with its intersection with X_0 may enlarge the index of H in $\mathrm{SL}_2(\mathbf{Z}/m\mathbf{Z})$. But the factor is at most

$$2^{\log m/\log\log m}=m^{1/\log\log m}\leqslant n^{1/\log\log n}.$$

As $n\to\infty$, this factor is small with respect to n. By the remark made in §1, we can deduce that our original problem is now completely reduced to the following extremal problem on counting subgroups of finite abelian groups:

Let $\mathcal{P}_-=\{q_1,\ldots,q_t\}$ and $\mathcal{P}_+=\{q'_1,\ldots,q'_{t'}\}$ be two sets of (different) primes and let $\mathcal{P}=\mathcal{P}_-\cup\mathcal{P}_+$. Put

$$f(n)=\sup\left\{s_r(X)\;\middle|\;X=\prod_{i=1}^{t} C_{q_i-1}\times\prod_{i=1}^{t'} C_{q'_i+1}\right\},$$

where the supremum is over all possible choices of $\mathcal{P}_-$, $\mathcal{P}_+$ and r such that

$$r\prod_{i=1}^{t} q_i\prod_{j=1}^{t'} q'_j\leqslant n,$$

and where C_m denotes the cyclic group of order m. The discussion above implies the following result:

PROPOSITION 5.7. *We have*

$$\alpha_+(\mathrm{SL}_2(\mathbf{Z}))=\varlimsup_{n\to\infty}\frac{\log f(n)}{\lambda(n)}.$$

6. Counting subgroups of p-groups

In this section we first give some general estimates for the number of subgroups of finite abelian p-groups which will be needed in §7. As an application we obtain a lower bound for the subgroup growth of uniform pro-p-groups (see definitions below).

For an abelian p-group G, we denote by $\Omega_i(G)$ the subgroup of elements of order dividing p^i. Then $\Omega_i(G)/\Omega_{i-1}(G)$ is an elementary abelian group of order, say, p^{λ_i} called the ith *layer* of G. We call the sequence $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_r$ the *layer type* of G. It is clear that this sequence is decreasing.

Denote by

$$\begin{bmatrix} \lambda \\ \nu \end{bmatrix}_p$$

the p-binomial coefficient, i.e., the number of ν-dimensional subspaces of a λ-dimensional vector space over $\mathbf{Z}/p\mathbf{Z}$.

The following result holds (see [LuS, Proposition 1.5.2]).

PROPOSITION 6.1. (i)

$$p^{\nu(\lambda-\nu)} \leqslant \begin{bmatrix} \lambda \\ \nu \end{bmatrix}_p \leqslant p^{\nu} p^{\nu(\lambda-\nu)}.$$

(ii)

$$\max \begin{bmatrix} \lambda \\ \nu \end{bmatrix}_p$$

is attained for $\nu = \lfloor \frac{1}{2}\lambda \rfloor$, *in which case*

$$\begin{bmatrix} \lambda \\ \nu \end{bmatrix}_p = p^{\lambda^2/4+O(\lambda)}$$

as $\lambda \to \infty$.

The starting point is the following well-known formula (see [Bu]).

PROPOSITION 6.2. *Let G be an abelian p-group of layer type $\lambda_1 \geqslant \lambda_2 \ldots \geqslant \lambda_r$. The number of subgroups of layer type $\nu_1 \geqslant \nu_2 \geqslant \ldots$ is*

$$\prod_{i\geqslant 1} p^{\nu_{i+1}(\lambda_i-\nu_i)} \begin{bmatrix} \lambda_i-\nu_{i+1} \\ \nu_i-\nu_{i+1} \end{bmatrix}_p .$$

(In the above expression we allow some of the ν_i to be 0.)

We need the following estimate.

PROPOSITION 6.3.

$$\prod_{i\geqslant 1} p^{\nu_i(\lambda_i-\nu_i)} \leqslant \prod_{i\geqslant 1} p^{\nu_{i+1}(\lambda_i-\nu_i)} \begin{bmatrix} \lambda_i-\nu_{i+1} \\ \nu_i-\nu_{i+1} \end{bmatrix}_p \leqslant p^{\nu_1} \prod_{i\geqslant 1} p^{\nu_i(\lambda_i-\nu_i)}.$$

Proof. By Proposition 6.1 we have

$$\prod_{i\geqslant 1} p^{\nu_{i+1}(\lambda_i-\nu_i)} \begin{bmatrix} \lambda_i-\nu_{i+1} \\ \nu_i-\nu_{i+1} \end{bmatrix}_p \leqslant \prod_{i\geqslant 1} p^{\nu_{i+1}(\lambda_i-\nu_i)} p^{(\nu_i-\nu_{i+1})((\lambda_i-\nu_{i+1})-(\nu_i-\nu_{i+1}))} p^{\nu_i-\nu_{i+1}}$$

$$= p^{\nu_1} \prod_{i\geqslant 1} p^{\nu_{i+1}(\lambda_i-\nu_i)} p^{(\nu_i-\nu_{i+1})(\lambda_i-\nu_i)} = p^{\nu_1} \prod_{i\geqslant 1} p^{\nu_i(\lambda_i-\nu_i)}.$$

The lower bound follows in a similar way. $\square$

COROLLARY 6.4. *Let G be an abelian group of order p^α and layer type $\lambda_1 \geqslant \ldots \geqslant \lambda_r$. Then $|G|^{-1}\prod_{i\geqslant 1} p^{\lambda_i^2/4} \leqslant |\mathrm{Sub}(G)| \leqslant |G|^2 \prod_{i\geqslant 1} p^{\lambda_i^2/4}$ holds.*

Proof. Considering subgroups H of layer type $\lfloor \frac{1}{2}\lambda_1 \rfloor \geqslant \lfloor \frac{1}{2}\lambda_2 \rfloor \geqslant \ldots$, we obtain that $|\mathrm{Sub}(G)| \geqslant \prod_{i\geqslant 1} p^{\lfloor \lambda_i/2 \rfloor (\lambda_i - \lfloor \lambda_i/2 \rfloor)} \geqslant p^{-r} \prod_{i\geqslant 1} p^{\lambda_i^2/4}$, which implies the lower bound.

On the other hand, for any fixed layer type $\nu_1 \geqslant \nu_2 \geqslant \ldots$, the number of subgroups H with this layer type is at most

$$p^{\nu_1} \prod_{i\geqslant 1} p^{\nu_i(\lambda_i - \nu_i)} \leqslant |G| \prod_{i\geqslant 1} p^{\lambda_i^2/4}.$$

The number of possible layer types $\nu_1 \geqslant \nu_2 \geqslant \ldots$ of subgroups of G is bounded by the number of partitions of the number α, and hence it is at most $2^\alpha \leqslant |G|$. This implies our statement. □

Let us make an amusing remark which will not be needed later.

If G is an abelian p-group of the form $G = C_{x_1} \times C_{x_2} \times \ldots \times C_{x_t}$, then it is known (see [LuS, §1.10]) that $|\mathrm{End}(G)| = \prod_{j,k\geqslant 1} \gcd(x_j, x_k)$. Noting that $\prod_{j,k\geqslant 1} \gcd(x_j, x_k) = \prod_{i\geqslant 1} p^{\lambda_i^2}$ we obtain that

$$|G|^{-1} |\mathrm{End}(G)|^{1/4} \leqslant |\,\mathrm{Sub}(G)| \leqslant |G|^2 |\mathrm{End}(G)|^{1/4}.$$

These inequalities clearly extend to arbitrary finite abelian groups G.

For the application of the above results to estimating the subgroup growth of $\mathrm{SL}_d(\mathbf{Z}_p)$ we have to introduce additional notation. For a group G let G^k denote the subgroup generated by all kth powers. For odd p a *powerful* p-group G is a p-group with the property that G/G^p is abelian. (In the rest of this section we will always assume that p is odd; the case $p=2$ requires only slight modifications.) The group G is said to be *uniformly powerful* (*uniform*, for short) if it is powerful and the indices $|G^{p^i} : G^{p^{i+1}}|$ do not depend on i as long as $i < e$, where p^e is the exponent of G.

Now let G be a uniform group of exponent p^e, where $e = 2i$, with d generators. Then G^{p^i} is a homocyclic abelian group of exponent p^i with d generators (i.e., it has layer type $d, d, \ldots, d$ with i terms) [Sh].

Consider subgroups H of G^{p^i} of layer type $\nu, \nu, \ldots, \nu$ (i terms). The number of such subgroups is at least $p^{i\nu(d-\nu)}$ by Proposition 6.3. The index n of such a subgroup H in G is $p^{di+(d-\nu)i}$. Hence the number of subgroups of index n in G is at least n^x, where $x = \nu(d-\nu)/(2d-\nu)$. Substituting $\nu = \lfloor d(2-\sqrt{2}) \rfloor$ we see that x can be as large as $(3-2\sqrt{2})d - (\sqrt{2}-1)$.

Let now U be a uniform pro-p-group of rank d, i.e., an inverse limit of d-generated finite uniform groups G. Then we see that for infinitely many n we have $s_n(G) \geqslant n^{(3-2\sqrt{2})d - (\sqrt{2}-1)}$.

Now $\mathrm{SL}_d(\mathbf{Z}_p)$ is known to have a finite-index uniform pro-p-subgroup of rank d^2-1 (see [DDMS, Theorem 5.2]). This proves the following result.

PROPOSITION 6.5. *The subgroup growth of* $\mathrm{SL}_d(\mathbf{Z}_p)$ *is at least* n^c, *i.e.*,

$$\varlimsup_{n\to\infty} \frac{\log s_n(\mathrm{SL}_d(\mathbf{Z}_p))}{\log n} \geqslant c,$$

where $c=(3-2\sqrt{2})d^2-2(2-\sqrt{2})$.

Note that for a pro-p-group P of rank d we have $\varlimsup_{n\to\infty}\log(s_n(P)/\log n)\leqslant d$, [LuS, Chapter 4.1]. Hence we have $\varlimsup_{n\to\infty}\log s_n(\mathrm{SL}_d(\mathbf{Z}_p))/\log n\leqslant d^2-1$.

B. Klopsch proved [K] that if G is a residually finite virtually soluble minimax group of Hirsch length $h(G)$, then its subgroup growth is of type at least $n^{h(G)/7}$. By using the above argument one can improve this to $n^{(3-2\sqrt{2})h(G)-(\sqrt{2}-1)}$.

7. Counting subgroups of abelian groups

The aim of this section is to solve a somewhat unusual extremal problem concerning the number of subgroups of abelian groups. The result we prove is the crucial ingredient in obtaining a sharp upper bound for the number of congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$. Actually we prove a slightly more general result, which will be used in [LuN] to obtain similar bounds for other arithmetic groups.

We will use Propositions 6.2 and 6.3 in conjunction with the following simple (but somewhat technical) observations.

PROPOSITION 7.1. *Let* $R\geqslant 1$ *and let* $C,t\in\mathbf{N}$ *be fixed. Consider pairs of sequences* $\{\lambda_i\}$ *and* $\{\nu_i\}$ *of non-negative integers such that* $\lambda_i\leqslant t$ *for all* i *and* $\sum_{i\geqslant 1}(R\lambda_i+\nu_i)\leqslant C$.

Under these conditions the maximal value of the expression

$$A(\{\lambda\},\{\nu\})=\sum_{i\geqslant 1}\nu_i(\lambda_i-\nu_i)$$

can be attained by a pair of sequences $\{\lambda_i\}$ *and* $\{\nu_i\}$, $i=1,2,\dots,r$, *such that*

(i) $\lambda_1\geqslant\lambda_2\geqslant\dots\geqslant\lambda_r$, $\nu_1\geqslant\nu_2\geqslant\dots\geqslant\nu_r\geqslant 1$ *and* $\lambda_i\geqslant\nu_i$ *for all* i;

(ii) $\lambda_1=\lambda_2=\dots=\lambda_{r-1}=t$;

(iii) *for some* $0\leqslant b\leqslant r-1$ *we have* $\nu_1=\nu_2=\dots=\nu_b=\nu_{b+1}+1=\dots=\nu_{r-1}+1$. *If* $\lambda_r=t$ *then also* $\nu_r\in\{\nu_1,\nu_1-1\}$.

Proof. Suppose that the maximum of $A(\{\lambda\},\{\nu\})$ is attained by a pair $\{\lambda_i\},\{\nu_i\}$ of sequences of non-negative integers. Deleting pairs with $\nu_j=0$ does not change the value of $A(\{\lambda\},\{\nu\})$, hence we can assume that all $\nu_i\geqslant 1$. If $\lambda_j<\nu_j$ for some j, then we can delete λ_j and ν_j from the sequences, and in this way the value of $A(\{\lambda\},\{\nu\})$ increases, a contradiction. Hence we have that $\lambda_i \geqslant \nu_i$ for all i. By relabelling the indices we can further assume that $\nu_1 \geqslant \nu_2 \geqslant \ldots \geqslant \nu_r \geqslant 1$.

Now, if π is a permutation of $\{1,2,\ldots,r\}$, it is clear that the maximum of $\sum_i \lambda_{\pi(i)}\nu_i$ (and hence of $A(\{\lambda_{\pi(i)}\},\{\nu_i\})$) is achieved for permutations π such that $\lambda_{\pi(1)} \geqslant \lambda_{\pi(2)} \geqslant \ldots \geqslant \lambda_{\pi(r)}$. By the maximality of the pair $\{\lambda_i\},\{\nu_i\}$ it now follows that $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_r$ as well, proving (i). We shall call a pair of sequences $\{\lambda_i\}$ and $\{\nu_i\}$ satisfying (i) *good*.

Let j be the smallest index such that we have $t > \lambda_j \geqslant \lambda_{j+1} \geqslant 1$ (if there is no such j then (ii) holds).

Assume that $\lambda_{j+1} = \ldots = \lambda_{j+k}$ and $\lambda_{j+k} > \lambda_{j+k+1}$ or $j+k=r$. The condition $\nu_j \geqslant \nu_{j+k}$ implies that $\nu_j((\lambda_j+1)-\nu_j)+\nu_{j+k}((\lambda_{j+k}-1)-\nu_{j+k}) \geqslant \nu_j(\lambda_j-\nu_j)+\nu_{j+k}(\lambda_{j+k}-\nu_{j+k})$. If $\lambda_{j+k}=\nu_{j+k}$ then (by deleting some terms and relabelling the rest) we can replace our sequences by another good pair for which $\sum_{i\geqslant 1}\lambda_j$ is strictly smaller and the value of $A(\{\lambda_i\},\{\nu_i\})$ is the same. Otherwise, replacing λ_j by λ_j+1 and λ_{j+k} by $\lambda_{j+k}-1$ we obtain a good pair of sequences for which $\{\lambda_i\}$ is lexicographically strictly greater and for which $A(\{\lambda_i\},\{\nu_i\})$ is at least as large (and hence maximal).

It is clear that by repeating these two types of moves we eventually obtain a good pair $\{\lambda_i\},\{\nu_i\}$ satisfying (ii) as well.

Now set $\beta = \nu_1+\nu_2+\ldots+\nu_{r-1}$. Then

$$\sum_{i\geqslant 1}\nu_i(\lambda_i-\nu_i) = t\beta - (\nu_1^2+\ldots+\nu_{r-1}^2)+\nu_r(\lambda_r-\nu_r).$$

It is clear that if the value of such an expression is maximal, then the difference of any two of the ν_j with $j \leqslant r-1$ is at most 1. Part (iii) follows. $\square$

PROPOSITION 7.2. *Let $x_1, x_2, \ldots, x_t$ be positive integers such that at most d of the x_i can be equal. Then*

$$\prod_{i=1}^{t} x_i \geqslant \left(\frac{t}{ed}\right)^t.$$

Proof. If, say, x_1 is the largest among the x_i then $x_1 \geqslant t/d$. By induction we can assume that $\prod_{i=2}^{t} x_i \geqslant ((t-1)/ed)^{t-1}$. Then

$$\begin{aligned}\prod_{i=1}^{t} x_i &\geqslant \frac{t}{d}\left(\frac{t-1}{ed}\right)^{t-1} = e\frac{t}{ed}\left(\frac{t-1}{ed}\right)^{t-1} \\ &= e\left(\frac{t}{ed}\right)^t\left(\frac{t-1}{t}\right)^{t-1} = \left(\frac{t}{ed}\right)^t \frac{e}{(1+1/(t-1))^{t-1}} \geqslant \left(\frac{t}{ed}\right)^t,\end{aligned}$$

as required. $\square$

The main result of this section is the following theorem.

THEOREM 7.3. *Let $R\geqslant 1$ be a real number and d be a fixed integer $\geqslant 1$. Let n and r be positive integers. Let G be an abelian group of the form $G=C_{x_1}\times C_{x_2}\times\ldots\times C_{x_t}$, where at most d of the x_i can be equal. Suppose that $r|G|^R\leqslant n$ holds. Then the number of subgroups of order $\leqslant r$ in G is at most $n^{(\gamma+o(1))l(n)}$, where $\gamma=\left(\sqrt{R(R+1)}-R\right)^2/4R^2$. In particular, if $R=1$ then $\gamma=\frac{1}{4}\left(3-2\sqrt{2}\right)$.*

Proof. We start the proof with several claims.

CLAIM 1. $t\leqslant(1+o(1))l(n)$.

Proof. By Proposition 7.2 we have $(t/ed)^t\leqslant n$. This easily implies the claim. □

CLAIM 2. *In proving the theorem, we may assume that $t\geqslant\gamma l(n)$.*

Proof. For otherwise every subgroup of G can be generated by $\gamma l(n)$ elements, and hence $|\mathrm{Sub}(G)|\leqslant|G|^{\gamma l(n)}\leqslant n^{\gamma l(n)}$. □

Now let $a(n)$ be a monotone increasing function which goes to infinity sufficiently slowly. For example, we may set $a(n)=\log\log\log\log n$.

Let G_p denote the Sylow p-subgroup of G and let $\lambda_1^p\geqslant\lambda_2^p\geqslant\ldots$ denote the layer type of G_p. Loosely speaking, we call any layer of some G_p a layer of G. We call such a layer *essential* if its dimension λ_i^p is at least $l(n)/a(n)$. Clearly the essential layers in G_p correspond to the layers of a certain subgroup E_p of G_p (which equals $\Omega_i(G_p)$ for the largest i such that $\lambda_i^p\geqslant l(n)/a(n)$). Let us call $E=\prod_p E_p$ the *essential subgroup* of G.

CLAIM 3. *Given $E\cap T$ we have at most $n^{o(l(n))}$ (i.e., a small number) of choices for a subgroup T of G.*

Proof. It is clear from the definitions that every subgroup of the quotient groups G_p/E_p, and hence of G/E, can be generated by less than $l(n)/a(n)$ elements. Therefore the same is true for $T/(T\cap E)$. This implies the claim. □

By Claim 3, in proving the theorem, it is sufficient to consider subgroups T of E.

Let v denote the exponent of E. Then E is the subgroup of elements of order dividing v in G. Now v is the product of the exponents of the E_p, hence the product of the exponents of the essential layers of G. It is clear from the definitions that we have $v^{l(n)/a(n)}\leqslant n$, and hence $v\leqslant(\log n)^{a(n)}$. Using well-known estimates of number theory [R] we immediately obtain the following result.

CLAIM 4. (i) *The number z of different primes dividing v is at most*

$$\frac{\log v}{\log\log v}\leqslant\frac{a(n)\log\log n}{\log\log\log n}.$$

(ii) *The total number of divisors of v is at most $v^{c/\log\log v} \leqslant (\log n)^{ca(n)/\log\log\log n}$ for some constant $c>0$.*

CLAIM 5. $|G:E| \geqslant (\log n)^{(1+o(1))t}$.

Proof. Consider the subgroup $E^i = E \cap C_{x_i}$. It follows that E^i is the subgroup of elements of order dividing v in C_{x_i}. Set $e_i = |E^i|$ and $h_i = x_i/e_i$. It is easy to see that $E = \prod_{i\geqslant 1} E^i$, and hence $|G:E| = \prod_{i\geqslant 1} h_i$. □

By Claim 4 (ii) for the number s of different values of the numbers e_i we have $s = (\log n)^{o(1)}$. We put the numbers x_i into s blocks according to the value of e_i. By our condition on the x_i it follows that at most d of the numbers h_i corresponding to a given block are equal. Hence altogether ds of the h_i can be equal. Using Proposition 7.2 we obtain that $|G:E| \geqslant \prod_{i\geqslant 1} h_i \geqslant (t/eds)^t$.

Since $sd = (\log n)^{o(1)}$ and by Claim 2, $t \geqslant \gamma(\log n/\log\log n)$, we obtain that $|G:E| \geqslant (\log n)^{(1+o(1))t}$ as required.

Let us now choose a group G and a number r as in the theorem for which the number of subgroups $T \leqslant E$ of order dividing r is maximal. To complete the proof it is clearly sufficient to show that this number is at most $n^{(\gamma+o(1))l(n)}$.

Denote the order of the corresponding essential subgroup E by f, and the index $|G:E|$ by m.

Using Propositions 6.2 and 6.3 we see that apart from an $n^{o(l(n))}$-factor (which we ignore) the number of subgroups T as above is at most

$$\prod_{p|f} \prod_{i\geqslant 1} p^{\nu_i^p(\lambda_i^p - \nu_i^p)} \tag{7.1}$$

for some ν_i^p and λ_i^p, where $\{\lambda_i^p\}, \{\nu_i^p\}$ is a pair of sequences for every p, $\prod_p \prod_{i\geqslant 1} p^{\lambda_i^p}$ divides f, and $\prod_p \prod_{i\geqslant 1} p^{\nu_i^p}$ divides r. Assuming that $f^R r$ is fixed together with the upper bound t for all the λ_i^p and μ_i^p, let us estimate the value of the expression (7.1).

By Proposition 7.1, a maximal value of an expression like (7.1) is attained for a choice of the λ_i^p and ν_i^p (for the sake of simplicity we use the same notation for the new sequences) such that for every p there are at most 3 different pairs $(p^{\lambda_i^p}, p^{\nu_i^p})$ equal to, say,

$$(p^t, p^{\mu^p+1}), \quad (p^t, p^{\mu^p}) \quad \text{and} \quad (p^{\tau^p}, p^{\mu_0^p}),$$

where $\mu_0^p \leqslant \tau^p < t$ and $\mu^p < t$ for all p.

Exchange the pairs equal to the first type for pairs equal to (p^t, p^{μ^p}). We obtain an expression like (7.1) such that the ratio of the two expressions is at most

$$\prod_p \prod_{i\geqslant 1} p^{\lambda_i^p} \leqslant n.$$

If now there are, say, α^p pairs with $(p^{\lambda_i^p}, p^{\nu_i^p})$ equal to (p^t, p^{μ^p}), then take β^p to be the largest integer with $2^{\beta^p} \leqslant p^{\alpha^p}$ and set $\beta_1^p = \lfloor \log_2 p \rfloor$. (Note that for every p there is at most one pair of the form $(p^{\tau^p}, p^{\mu_0^p})$.)

Consider the expression

$$\prod_p 2^{\beta^p \mu^p (t-\mu^p)} 2^{\beta_1^p \mu_0^p (\tau^p - \mu_0^p)}. \tag{7.2}$$

Its value may be less than that of (7.1), but in this case their ratio is bounded by $(2^{2z})^{t^2} n$ (where z is the number of primes dividing v). Hence this ratio is at most

$$2^{(2+o(1)) l(n)^2 a(n) \log\log n / \log\log\log n} \leqslant n^{(2+o(1)) l(n) a(n)/\log\log\log n} = n^{o(l(n))}.$$

To prove our theorem it is sufficient to bound the value of (7.2) by $n^{(\gamma+o(1)) l(n)}$.

It is clear that the value of (7.2) is equal to the value of

$$\prod_{k\geqslant 1} 2^{\nu_k(\lambda_k - \nu_k)} \tag{7.3}$$

for appropriate sequences $\{\lambda_k\}$ and $\{\mu_k\}$ which both have $\sum_p (\beta^p + \beta_1^p)$ terms and for which $\lambda_k, \mu_k \leqslant t$ and also

$$\prod_{k\geqslant 1} 2^{R\lambda_k + \nu_k} \leqslant r f^R, \quad \text{i.e.,} \sum_{k\geqslant 1} (R\lambda_k + \nu_k) \leqslant \log(r f^R). \tag{7.4}$$

More precisely, the sequence $\{\lambda_k\}$ has $\sum_p \beta^p$ terms equal to t and β_1^p terms equal to τ^p for every p, while $\{\mu_k\}$ consists of μ^p repeated β^p times and μ_0^p repeated β_1^p times each (in the appropriate order).

By Proposition 7.1 the expression (7.3) attains its maximal value for some sequences $\{\lambda_k\}$ and $\{\nu_k\}$ such that all but one of the λ_k, say λ_{a+1}, are equal to t and we have $\nu_1 = \nu_2 = \ldots = \nu_b = 1 + \nu_{b+1} = \ldots = 1 + \nu_a$ for some $b \leqslant a$.

Consider now the expression

$$\prod_{k\geqslant 1} 2^{\nu_k'(\lambda_k' - \nu_k')}, \tag{7.5}$$

where

$$t = \lambda_1' = \ldots = \lambda_a' \quad (\lambda_{a+1}' = 0)$$

and $\nu_a = \nu_1' = \nu_2' = \ldots = \nu_a'$ $(\nu_{a+1}' = 0)$.

It easily follows that the value of (7.3) is at most 2^{2t^2} times as large as the value of (7.5) and $2^{2t^2} = n^{o(l(n))}$. Hence it suffices to bound the value of (7.5) by $n^{(\gamma+o(n)) l(n)}$.

To obtain our final estimate let $y=2^a$, $w=m^{1/t}$ (where $m=|G:E|$), and set $x=yw$.

For some constants between 0 and 1 we have $y=x^\varrho$ and $\nu_1'=\sigma t$. Then $w=x^{1-\varrho}=y^{(1-\varrho)/\varrho}$.

Note that the condition (7.4) implies $2^{at(R+\sigma)}=y^{\sigma t}y^{Rt}\leqslant rf^R$. We have $n\geqslant r(mf)^R\geqslant y^{\sigma t}y^{Rt}w^{Rt}$, and hence $\log n\geqslant t(\log y)(R+\sigma+R(1-\varrho)/\varrho)$.

By Claim 5 we have $w\geqslant(\log n)^{1+o(1)}$. Hence

$$(1+o(1))\log\log n\leqslant \log w=\frac{1-\varrho}{\varrho}\log y.$$

Therefore

$$\begin{aligned}\frac{(\log n)^2}{\log\log n}&\geqslant\frac{t^2(\log y)^2(R+\sigma+R(1-\varrho)/\varrho)^2}{((1-\varrho)/\varrho)\log y}(1+o(1))\\&=(1+o(1))\,t^2(\log y)\left(R+\sigma+R\frac{1-\varrho}{\varrho}\right)^2\frac{\varrho}{1-\varrho}.\end{aligned}$$

The value of (7.5) is $y^{\sigma t(t-\sigma t)}$, which as we saw is an upper bound for the number of subgroups R (ignoring an $n^{o(l(n))}$-factor). Hence

$$\begin{aligned}\frac{\log\,(\text{number of subgroups } T)}{(\log n)^2/\log\log n}&\leqslant(1+o(1))\frac{t^2\sigma(1-\sigma)\log y}{t^2(\log y)(R+\sigma+R(1-\varrho)/\varrho)^2\varrho/(1-\varrho)}\\&=(1+o(1))\frac{\sigma(1-\sigma)(1-\varrho)/\varrho}{(R+\sigma+R(1-\varrho)/\varrho)^2}\\&=(1+o(1))\frac{\sigma(1-\sigma)\varrho(1-\varrho)}{(R+\varrho\sigma)^2}.\end{aligned}$$

As observed in §3, the maximum value of $\sigma(1-\sigma)\varrho(1-\varrho)/(R+\varrho\sigma)^2$ for $\sigma,\varrho\in(0,1)$ is γ. The proof of the theorem is complete. □

By using a similar but simpler argument, one can also show the following result.

PROPOSITION 7.4. *Let G be an abelian group of order n of the form*

$$G=C_{x_1}\times C_{x_2}\times\ldots\times C_{x_t},\quad \text{where } x_1>x_2>\ldots>x_t.$$

Then $|\mathrm{Sub}(G)|\leqslant n^{(1/16+o(1))l(n)}$. This bound is attained if $x_i=ti$ for all i.

Combining this result with an earlier remark, we obtain that $n^{(1/4+o(1))l(n)}$ is the maximal value of $\prod_{i,j}\gcd(x_i,x_j)$, where the x_i are different numbers whose product is at most n.

Note that $|\mathrm{Sub}(G)|$ is essentially the number of subgroups T of order $\lfloor\sqrt{|G|}\rfloor$ (see [Bu] for a strong version of this assertion). Hence Proposition 7.4 corresponds to the case $R=1$ and $r\sim n^{1/3}$ of Theorem 7.3.

8. End of the proofs of Theorems 2, 3 and 4

Theorem 2 is actually proved now: the lower bound was shown as a special case of $R=R(G)=1$ in §3. For the upper bound, we have shown in Proposition 5.7 how $\alpha_+(\mathrm{SL}_2(\mathbf{Z}))$ is equal to $\overline{\lim}_{n\to\infty}(\log f(n)/\lambda(n))$ (see Proposition 5.7 for the definition of $f(n)$). But Theorem 7.3 implies, in particular, that $f(n)$ is at most $n^{(\gamma+o(1))l(n)}$, where $\gamma=\frac{1}{4}(3-2\sqrt{2})$. This proves that $\alpha_+(\mathrm{SL}_2(\mathbf{Z}))\leqslant\gamma$ and finishes the proof.

The proof of Theorem 3 is similar, but several remarks should be made: The lower bound was deduced in §4. For the upper bound, one should follow the reductions made in §6. The proof can be carried out in a similar way for $\mathrm{SL}_2(\mathcal{O})$ instead of $\mathrm{SL}_2(\mathbf{Z})$, but the following points require careful consideration.

(1) One can pass to the case that m is an ideal, which is a product of different primes π_i in $\mathcal{O}$, but it is possible that $\mathcal{O}/\pi_i$ is isomorphic to $\mathcal{O}/\pi_j$. Still, each such isomorphism class of quotient fields can occur at most d times when $d=[k:\mathbf{Q}]$.

(2) The maximal subgroups of $\mathrm{SL}_2(\mathbf{F}_q)$ when $\mathbf{F}_q$ is a finite field of order q (q is a prime power, not necessarily a prime) are the same B, D and A as described in (1), (2) and (3) of §5.

The rest of the reduction can be carried out in a way similar to §5. The final outcome is not exactly as $f(n)$ at the end of §5, but can be reduced to a similar problem when $\tilde{f}(n)$ counts $s_r(X)$ when X is a product of abelian cyclic groups, with a bounded multiplicity. Theorem 7.3 covers also this case and gives a bound to $\tilde{f}(n)$, which is the same as for $f(n)$. Thus $\alpha_+(\mathrm{SL}_2(\mathcal{O}))\leqslant\gamma=\frac{1}{4}(3-2\sqrt{2})$.

We finally mention the easy fact that replacing $\mathcal{O}$ by $\mathcal{O}_S$ when S is a finite set of primes (see the introduction) does not change α_+ or α_-. To see this, one can use the fact that for every completion at a simple prime π of $\mathcal{O}$, $G(\mathcal{O}_\pi)$ has polynomial subgroup growth, and then use the well-known techniques of subgroup growth and the fact that

$$G(\widehat{\mathcal{O}})=G(\widehat{\mathcal{O}}_S)\times\prod_{\pi\in S\setminus V_\infty}G(\mathcal{O}_\pi)$$

to deduce that $\alpha(G(\widehat{\mathcal{O}}))=\alpha(G(\widehat{\mathcal{O}}_S))$.

Another way to see it, is to observe that $G(\widehat{\mathcal{O}}_S)$ is a quotient of $G(\widehat{\mathcal{O}})$, and hence $\alpha_+(G(\mathcal{O}))\geqslant\alpha_+(G(\mathcal{O}_S))$. On the other hand, the proof of the lower bound for $\alpha(G(\mathcal{O}))$ clearly works for $G(\mathcal{O}_S)$. Theorem 3 is, therefore, now proved, as well as Theorem 4 (since we have not used the GRH for the upper bounds in Theorem 3).

9. An extremal problem in elementary number theory

The counting techniques in this paper can be applied to solve the following extremal problem in multiplicative number theory.

For $n \to \infty$, let

$$M_1(n) = \max\left\{ \prod_{1 \leqslant i,j \leqslant t} \gcd(a_i, a_j) \,\middle|\, 0 < t,\ a_1 < a_2 < \ldots < a_t \in \mathbf{Z},\ \prod_{i=1}^{t} a_i \leqslant n \right\},$$

$$M_2(n) = \max\left\{ \prod_{p,p' \in \mathcal{P}} \gcd(p-1, p'-1) \,\middle|\, \mathcal{P} \text{ is a set of distinct primes where } \prod_{p \in \mathcal{P}} p \leqslant n \right\}.$$

We shall prove the following theorem, which can be considered as a baby version of Theorem 2 (compare also to Theorem 7.3). Note that Theorem 9.1 immediately implies Theorem 9.

THEOREM 9.1. *Let* $\lambda(n) = (\log n)^2 / \log\log n$. *Then*

$$\varliminf_{n\to\infty} \frac{\log M_1(n)}{\lambda(n)} = \varlimsup_{n\to\infty} \frac{\log M_2(n)}{\lambda(n)} = \frac{1}{4}.$$

Proof. Recall that if $a_1, a_2, \ldots, a_t \in \mathbf{Z}$ and $G = C_{a_1} \times C_{a_2} \times \ldots \times C_{a_t}$ is a direct product of cyclic groups, then by §7,

$$|G|^{-1} |\mathrm{End}(G)|^{1/4} \leqslant |\mathrm{Sub}(G)| \leqslant |G|^2 |\mathrm{End}(G)|^{1/4}$$

and

$$|\mathrm{End}(G)| = \prod_{1 \leqslant i,j \leqslant t} \gcd(a_i, a_j).$$

Proposition 7.4 implies that

$$\varlimsup_{n\to\infty} \frac{\log M_1(n)}{\lambda(n)} \leqslant \frac{1}{4}.$$

It is clear that $M_2(n) \leqslant M_1(n)$, so to finish the proof it is enough to obtain a lower bound for $M_2(n)$.

Now, for $x \to \infty$ and $x^\varrho / \log x \leqslant q \leqslant x^\varrho$ (with $0 < \varrho < \frac{1}{2}$) choose

$$\mathcal{P} = \mathcal{P}(x, q) = \{ p \leqslant x \mid p \equiv 1 \pmod{q} \}$$

to be a Bombieri set relative to x, where q is a prime number (a Bombieri prime). By Lemma 2.4 we have the asymptotic relation $\#\mathcal{P}(x,q) \sim x / \phi(q) \log x$. In order to satisfy the condition $\prod_{p\in\mathcal{P}} p\leqslant n$, we choose $x\sim q\log n$. Without loss of generality, we may choose $q=x^{\varrho}$ for some $0<\varrho<\frac{1}{2}$. It follows that

$$x^{1-\varrho}\sim\log n,\quad \log x\sim\frac{\log\log n}{1-\varrho}\quad\text{and}\quad \#\mathcal{P}=\#\mathcal{P}(x,q)\sim\frac{x}{\phi(q)\log x}\sim\frac{(1-\varrho)\log n}{\log\log n}.$$

Consequently,

$$\prod_{p,p'\in\mathcal{P}}\gcd(p-1,p'-1)\geqslant q^{(\#\mathcal{P})^2}\geqslant(x^{\varrho})^{(1-\varrho)^2(\log n)^2/(\log\log n)^2}\sim e^{\varrho(1-\varrho)(\log n)^2/\log\log n}.$$

Let now ϱ tend to $\frac{1}{2}$, and the theorem is proved. □

References

[Bo] Bombieri, E., On the large sieve. *Mathematika*, 12 (1965), 201–225.

[Bu] Butler, L. M., A unimodality result in the enumeration of subgroups of a finite abelian group. *Proc. Amer. Math. Soc.*, 101 (1987), 771–775.

[CP] Cox, D. A. & Parry, W. R., Genera of congruence subgroups in **Q**-quaternion algebras. *J. Reine Angew. Math.*, 351 (1984), 66–112.

[Da] Davenport, H., *Multiplicative Number Theory*, 2nd edition. Graduate Texts in Math., 74. Springer-Verlag, New York–Berlin, 1980.

[De] Dennin, J. B., Jr., The genus of subfields of $K(n)$. *Proc. Amer. Math. Soc.*, 51 (1975), 282–288.

[DDMS] Dixon, J. D., du Sautoy, M. P. F., Mann, A. & Segal, D., *Analytic Pro-p-Groups*. London Math. Soc. Lecture Note Ser., 157. Cambridge Univ. Press, Cambridge, 1991.

[FJ] Fried, M. D. & Jarden, M., *Field Arithmetic*. Ergeb. Math. Grenzgeb., 11. Springer-Verlag, Berlin, 1986.

[GLNP] Goldfeld, D., Lubotzky, A., Nikolov, N. & Pyber, L., Counting primes, groups and manifolds. *Proc. Natl. Acad. Sci. USA*, 101 (2004), 13428–13430.

[H] Heilbronn, H., Zeta-functions and L-functions, in *Algebraic Number Theory* (Brighton, 1965), pp. 204–230. Thompson, Washington, DC, 1967.

[K] Klopsch, B., Linear bounds for the degree of subgroup growth in terms of the Hirsch length. *Bull. London. Math. Soc.*, 32 (2000), 403–408.

[Ld] Landau, E., *Handbuch der Lehre von der Verteilung der Primzahlen*, 2nd edition. Chelsea, New York, 1953.

[Lg] Lang, S., *Introduction to Modular Forms*. Grundlehren Math. Wiss., 222. Springer-Verlag, Berlin–New York, 1976.

[Li1] Linnik, U. V., On the least prime in an arithmetic progression, I. The basic theorem. *Rec. Math. [Mat. Sbornik] N.S.*, 15 (57) (1944), 139–178.

[Li2] — On the least prime in an arithmetic progression, II. The Deuring–Heilbronn phenomenon. *Rec. Math. [Mat. Sbornik] N.S.*, 15 (57) (1944), 347–368.

[Lu] Lubotzky, A., Subgroup growth and congruence subgroups. *Invent. Math.*, 119 (1995), 267–295.

[LuN] Lubotzky, A. & Nikolov, N., Subgroup growth of lattices in semisimple Lie groups. *Acta Math.*, 193 (2004), 105–139.

[LuS] LUBOTZKY, A. & SEGAL, D., *Subgroup Growth*. Progr. Math., 212. Birkhäuser, Basel, 2003.

[MM] MURTY, M. R. & MURTY, V. K., A variant of the Bombieri–Vinogradov theorem, in *Number Theory* (Montreal, PQ, 1985), pp. 243–272. CMS Conf. Proc., 7. Amer. Math. Soc., Providence, RI, 1987.

[MMS] MURTY, M. R., MURTY, V. K. & SARADHA, N., Modular forms and the Chebotarev density theorem. *Amer. J. Math.*, 110 (1988), 253–281.

[Pe] PETERSSON, H., Konstruktionsprinzipien für Untergruppen der Modulgruppe mit einer oder zwei Spitzenklassen. *J. Reine Angew. Math.*, 268/269 (1974), 94–109.

[R] RAMANUJAN, S., Highly composite numbers. *Proc. London Math. Soc.* (*2*), 14 (1915), 347–409.

[Sh] SHALEV, A., On almost fixed point free automorphisms. *J. Algebra*, 157 (1993), 271–282.

[Su] SUZUKI, M., *Group Theory*, Vol. 1. Grundlehren Math. Wiss., 247. Springer-Verlag, Berlin–New York, 1982.

[T] THOMPSON, J. G., A finiteness theorem for subgroups of $\mathrm{PSL}(2, \mathbf{R})$ which are commensurable with $\mathrm{PSL}(2, \mathbf{Z})$, in *The Santa Cruz Conference on Finite Groups* (Santa Cruz, CA, 1979), pp. 533–555. Proc. Sympos. Pure Math., 37. Amer. Math. Soc., Providence, RI, 1980.

[V] VINOGRADOV A. I., On the density conjecture for Dirichlet L-series. *Izv. Akad. Nauk SSSR Ser. Mat.*, 29 (1965), 903–934 (Russian).

[W] WEIL, A., Sur les «formules explicites» de la théorie des nombres premiers. *Comm. Sém. Math. Univ. Lund*, tome supplémentaire dédié à Marcel Riesz (1952), 252–265.

DORIAN GOLDFELD
Department of Mathematics
Columbia University
New York, NY 10027
U.S.A.
goldfeld@columbia.edu

ALEXANDER LUBOTZKY
Einstein Institute of Mathematics
The Hebrew University of Jerusalem
IL-91904 Jerusalem
Israel
alexlub@math.huji.ac.il

LÁSZLÓ PYBER
A. Renyi Institute of Mathematics
Hungarian Academy of Sciences
P. O. Box 127
HU-1364 Budapest
Hungary
pyber@renyi.hu

Received December 18, 2003
Received in revised form August 12, 2004