On Korobov bound concerning Zaremba's conjecture

N.G. Moshchevitin, B. Murphy and I.D. Shkredov

À Jean Bourgain
avec admiration et tristesse.

Annotation.

We prove in particular that for any sufficiently large prime p there is $1 \leqslant a < p$ such that all partial quotients of a/p are bounded by $O(\log p / \log \log p)$. For composite denominators a similar result is obtained. This improves the well–known Korobov bound concerning Zaremba's conjecture from the theory of continued fractions.

1 Introduction

Let a and q be two positive coprime integers, $0 < a < q$. By the Euclidean algorithm, a rational a/q can be uniquely represented as a regular continued fraction

$$\frac{a}{q} = [0; c_1, \ldots, c_s] = \cfrac{1}{c_1 + \cfrac{1}{c_2 + \cfrac{1}{c_3 + \cdots + \cfrac{1}{c_s}}}}, \qquad c_s \geqslant 2. \tag{1}$$

Assuming q is known, we use $c_j(a)$, $j = 1, \ldots, s = s(a)$ to denote the partial quotients of a/q; that is,

$$\frac{a}{q} := [0; c_1(a), \ldots, c_s(a)]. \tag{2}$$

Zaremba's famous conjecture [45] posits that there is an absolute constant $\mathfrak{k}$ with the following property: for any positive integer q there exists a coprime to q such that in the continued fraction expansion (1) all partial quotients are bounded:

$$c_j(a) \leqslant \mathfrak{k}, \qquad 1 \leqslant j \leqslant s = s(a).$$

In fact, Zaremba conjectured that $\mathfrak{k} = 5$. For large prime q, even $\mathfrak{k} = 2$ should be enough, as conjectured by Hensley [19], [20]. This theme is rather popular especially at the last time, see, e.g., papers [8]–[20], [23], [28], [31], [32], [39] and many others. The history of the question can

be found, e.g., in [25], [29], [30]. We just notice here a remarkable progress of Bourgain and Kontorovich [8], [9] who proved Zaramba's conjecture for "almost all" denominators q.

Zaremba's conjecture is connected with some questions of numerical integration. It was showed in [44] that if Zaremba's conjecture is true, then the two–dimensional winding of the torus

$$X = X(a, q) = \left\{ \left(\frac{j}{q}, \frac{aj}{q} \right) \right\}_{j=1}^{q} \subseteq [0, 1]^2$$

would have the least discrepancy (up to some absolute constants). Here we assume that the fraction a/q enjoys $c_j(a) = O(1)$. In this direction, using some exponential sums, Korobov [26] in 1963 proved that for any prime q there is a, $(a, q) = 1$, such that

$$\max_{\nu} c_{\nu}(a) \ll \log q \,. \tag{3}$$

The same result takes place for composite q, see [37].

In this paper we improve Korobov's bound (3). The proof is not purely analytical and uses rather well–known methods connected with the Bourgain–Gamburd machine [3] as well as an exact result from [28], see Lemma 3 below.

Theorem 1 *Let q be a positive sufficiently large integer with sufficiently large prime factors. Then there is a positive integer a, $(a, q) = 1$ and*

$$M = O(\log q / \log \log q) \tag{4}$$

such that

$$\frac{a}{q} = [0; c_1, \ldots, c_s] \,, \qquad c_j \leqslant M \,, \qquad \forall j \in [s] \,. \tag{5}$$

Also, if q is a sufficiently large square–free number, then (4), (5) take place.
Finally, if $q = p^n$, p is an arbitrary prime, then (4), (5) hold for sufficiently large n.

Our paper is organized as follows. In Section 2 we obtain Theorem 1 for sufficiently large prime q and in the next Subsection 3.1 we prove this for *all* sufficiently large square–free numbers. The last Subsection 3.2 contains some discussions of the difficulties, which do not allow to obtain Theorem 1 following the standard Bourgain–Varjú [7] variant of the Bourgain–Gamburd machine for general q. Also, we separately consider the case $q = p^n$ here (n is a sufficiently large number and p is a prime) and show that Theorem 1 remains to be true for such q. Using the specific of our problem, we combine the approach of [4], [7] with a more simple and more direct two–dimensional method from [36] to obtain Theorem 1 for general q. We should say that all sections are dependent and the complexity increases from part to part. In the appendix we obtain some results on large deviations for continued fractions with bounded partial quotients. Our Theorem 14 from the appendix is required in the previous Subsection 3.2 (as a particular two–dimensional case) and maybe it is interesting in its own right as it improves some results of Rogers [35].

The signs $\ll$ and $\gg$ are the usual Vinogradov symbols. Let us denote by $[n]$ the set $\{1, 2, \ldots, n\}$. All logarithms are to base 2.

2 The prime case

In this section we obtain our main Theorem 1 in the case of prime q although all results excluding our driving Lemma 4 take place for an arbitrary number q. The required generalization of Lemma 4 for general q is discussed in Section 3.

We start with a well–known lemma, see [26, Lemma 5, pages 25–27] or [28, Section 9]. It says that, basically, the partial quotients of a rational number are controlled via the hyperbola $x|y| = q/M$.

Lemma 2 *Let a be coprime with q and $a/q = [0; c_1, \ldots, c_s]$. Consider the equation*

$$ax \equiv y \pmod{q}, \qquad 1 \leqslant x < q, \quad 1 \leqslant |y| < q. \tag{6}$$

If for all solutions (x, y) of the equation above one has $x|y| \geqslant q/M$, then $c_j \leqslant M$, $j \in [s]$. On the other hand, if for all $j \in [s]$ the following holds $c_j \leqslant M$, then all solutions (x, y) of (6) satisfy $x|y| \geqslant q/4M$.

Let $1 \leqslant t \leqslant \sqrt{q}$ be a real number. Having a rational number $\frac{a}{q} = [0; c_1, \ldots, c_s] = \frac{p_s}{q_s}$, we write $\frac{p_\nu}{q_\nu}$ for its ν-th convergent. Define

$$Z_M(t) = \left\{ \frac{a}{q} = [0; c_1, \ldots, c_s] \ : \ c_j \leqslant M, \ \forall j \in [\nu], \ q_\nu < t \right\}. \tag{7}$$

Also, put

$$Q_M(t) = \left\{ \frac{u}{v} = [0; c_1, \ldots, c_s] \ : \ c_j \leqslant M, \ \forall j \in [s], \ v < t \right\},$$

and

$$\overline{Q_M(t)} = \left\{ \frac{u}{v} = [0; c_1, \ldots, c_s] \in Q_M(t) \ : \ K(c_1, \ldots, c_s, 1) \geqslant t \right\},$$

where by $K(d_1, \ldots, d_k)$ we have denoted the correspondent continuant, see [21]. The sets $\overline{Q_M(t)}$ and $Z_M(t)$ are closely connected to each other, see [28].

To formulate further results we need a definition from the real setting. Let $M \geqslant 1$ be an integer. Consider the set of *real* numbers F_M, having all partial quotients bounded by M. It is well–known [21], that for any M the Lebesgue measure of the set F_M is zero and its Hausdorff dimension $w_M := \mathcal{HD}(F_M)$ is $w_M = 1 - O(1/M)$, as $M \to \infty$. Good bounds and asymptotic formulae on w_M are contained in papers [16]—[18]. The following result is a combination of Lemma 2 and Lemma 3 of [28], as well as [16, Theorem 2]. With some abuse of the notation we denote by the same letter $Z_M(t)$ the set of the *numerators* $a \in [q]$, $(a, q) = 1$ from (7).

Lemma 3 *Let $t \leqslant \sqrt{q}$. Then for some absolute constants $c_1, c_2 > 0$ one has*

$$Z_M(t) = B_1 \bigsqcup \cdots \bigsqcup B_T, \qquad c_1 t^{2w_M} \leqslant T \leqslant c_2 t^{2w_M},$$

where B_j are some disjoint intervals and for all $j \in [T]$ the following holds $[q/t^2] \leqslant |B_j|$.

4

P r o o f. In [28] it was proved in particular, that $T = |\overline{Q_M(t)}|$ and $[q/t^2] \leqslant |B_j|$. Thus it remains to estimate the size of the set $\overline{Q_M(t)}$.

By [16, Theorem 2] we know that there exist absolute positive constants C_1, C_2 such that

$$C_1 t^{2w_M} \leqslant |Q_M(t)| \leqslant C_2 t^{2w_M} \tag{8}$$

for any $t \geqslant 2$. Clearly, every $u/v \in Q_M(t)$ can be written as a continued fraction

$$\frac{u}{v} = [0; A_1, .., A_l] \quad \text{with} \quad A_l \geqslant 2. \tag{9}$$

The upper bound is obvious from the inclusion of $\overline{Q_M(t)} \subset Q_M(t)$. To prove the lower bound put

$$k = \left(\frac{2C_2}{C_1} \right)^{\frac{1}{2w_M}}$$

and consider the set

$$\mathcal{W} = Q_M(t) \setminus Q_M(t/k).$$

By (8) we see that

$$|\mathcal{W}| \geqslant \frac{C_1}{2} t^{2w_M}.$$

Any $u/v \in \mathcal{W}$ can be written in the form (9) but we need another representation

$$\frac{u}{v} = [0; A_1, \dots, A_l - 1, 1]. \tag{10}$$

Recall that

$$v = K(A_1, \dots, A_l - 1, 1) = K(A_1, \dots, A_l) < t. \tag{11}$$

We define $\nu \geqslant 1$ from the condition

$$K(A_1, \dots, A_l - 1, \underbrace{1, \dots, 1}_{\nu+1}) < t \quad \text{but} \quad K(A_1, \dots, A_l - 1, \underbrace{1, \dots, 1}_{\nu+2}) \geqslant t.$$

As $K(uw) > K(u) \cdot K(w)$ and $t/k \leqslant v < t$ we have

$$K(\underbrace{1, \dots, 1}_{\nu}) < \frac{t}{v} \leqslant k,$$

and so

$$\nu \leqslant C_4 \log k. \tag{12}$$

It is clear that

$$[0; A_1, \dots, A_l - 1, \underbrace{1, \dots, 1}_{\nu+1}] \in \overline{Q_M(t)}. \tag{13}$$

Each element $u/v \in \overline{Q_M(t)}$, which can be written in the form (13) with continued fraction (10) satisfying (11), by (12) can be written in such a form not more than in $C_4 \log k$ ways. So we have the bound

$$|\overline{Q_M(t)}| \geqslant \frac{|\mathcal{W}|}{C_4 \log k} \geqslant C_5 t^{2w_M} \quad \text{with} \quad C_5 = \frac{C_1}{2C_4 \log k}$$

as required. □

The last result is actually contained in [30, Proposition 7].

Lemma 4 *Let p be a prime number, $A, B \subseteq \mathbb{F}_p$ be sets, and $J = [N]$ be an interval. Then there is an absolute constant $\kappa > 0$ such that*

$$\left| \{ (a+c)(b+c) = 1 \; : \; a \in A, \, b \in B, \, c \in 2 \cdot J \} \right| - \frac{N|A||B|}{p} \ll \sqrt{|A||B|} N^{1-\kappa} \,. \tag{14}$$

Lemma 4 can be deduced from [30, Proposition 7] directly. The proof of [30, Proposition 7] itself is just an application of the Bourgain–Gamburd machine [3] based on Helfgott's expansion result [22]. This method is rather well–known. However we prefer to recall the main ideas and crucial steps of the argument because we use them in the next Section 3. So we are giving a sketched proof below.

Sketch of the proof of Lemma 4. We use the notation $S(x)$ for the characteristic function of a set S. Also, write any $c \in 2 \cdot J$ as $c = 2j$, $j \in [N]$. Then clearly, the equation from the left–hand side of (14) is equivalent to $a = g_j b$, $j \in [N]$, where $a \in A, b \in B$

$$g_j = \begin{pmatrix} -2j & 1 - 4j^2 \\ 1 & 2j \end{pmatrix}, \qquad j \in [N] \tag{15}$$

with $\det(g_j) = -1$. In [30, Lemma 13] we considered the set of matrices

$$G = \left\{ \begin{pmatrix} 1 & -2j \\ 2j & 1 - 4j^2 \end{pmatrix} : 1 \le j \le N \right\} \subset \mathrm{SL}_2(\mathbb{F}_p), \tag{16}$$

and proved that the girth of the Cayley graph of G (e.g., see the definition of the Cayley graph in Section 3 below) is at least $\tau \log_N p$, $\tau = 1/5$ for all sufficiently large p. The proof uses the well–known fact that $\mathrm{SL}_2(\mathbb{Z})$ contains the free subgroup, generated by

$$u = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad v = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

Then $G = \{ v^j u^{-j} \; : \; j \in [N] \}$ and it is easy to check that G generates a free subgroup of $\mathrm{SL}_2(\mathbb{F}_p)$ of rank N. For any set $S \subseteq \mathrm{SL}_2(\mathbb{F}_p)$ write $r_{S,2m}(x)$ for the number solutions to the equation

$$r_{S,2m}(x) := \left| \{ (s_1, \ldots, s_{2m}) \in S^{2m} \; : \; s_1 s_2^{-1} s_3 \ldots s_{2m}^{-1} = x \} \right| =$$

$$= \sum_{x_1 x_2^{-1} x_3 \ldots x_{2m}^{-1} = x} S(x_1) S(x_2) S(x_3) \cdots S(x_{2m}) \,.$$

The same sum

$$\sum_{x_1 x_2^{-1} x_3 \ldots x_{2m}^{-1} = x} f(x_1) f(x_2) f(x_3) \cdots f(x_{2m})$$

can be defined for any function $f : \mathrm{SL}_2(\mathbb{F}_p) \to \mathbb{R}$. Also, let $\mathsf{T}_{2m}(S) = \sum_x r_{S,2m}^2(x)$, see the discussion concerning these important quantities in [42] and in [40, Sections 5, 6]. After that one can apply the first stage of the Bourgain–Gamburd machine [3] to the set G, see [30, Lemma 12], which asserts that for any $g \in \mathrm{SL}_2(\mathbb{F}_p)$ and an arbitrary proper subgroup $\Gamma < \mathrm{SL}_2(\mathbb{F}_p)$ one has

$$\sum_{x \in g\Gamma} r_{G,2m}(x) \leqslant \frac{|G|^{2m}}{K(G)}, \tag{17}$$

where $m = \tau/4 \cdot \log_N p$ and $K(G) = p^{\tau/6}$. The quantity $K(G) \geqslant 1$ can be defined as the maximal one such that bound (17) takes place (again it is possible to consider $K(f)$ for any non–negative function f). Here one can use the symmetrization of G, considering $G \cup G^{-1}$ instead of G as the authors did in [3] and in [30], or apply the argument directly as was done in [40, Section 6, see Theorem 49, Corollary 50]. Further several applications of Hölder inequality (see [30, Lemma 11]) or [40, Lemma 32] (here the author considered a non–symmetric case but this is not important for further results) give us for an arbitrary function $f : \mathrm{SL}_2(\mathbb{F}_p) \to \mathbb{R}$, a positive integer l, and any sets $A, B \subseteq \mathbb{F}_p$ that

$$\left| \sum_s \sum_{x \in B} f(s)A(sx) - \frac{|A||B|}{p} \sum_s f(s) \right| \leqslant \sqrt{|A||B|} \cdot \left(|B|^{-1} \sum_s r_{f,2^l}(s) \sum_{x \in B} B(sx) \right)^{1/2^l}. \tag{18}$$

More importantly, Helfgott's expansion result [22] (see [30, Propositions 5, 7]) allows us to estimate the quantity $\mathsf{T}_{2^k}(f)$ (for any sufficiently large k) and hence the right–hand side of (18) (it corresponds to the second and to the third stages of the Bourgain–Gamburd machine). More precisely, it gives us that for any function $F : \mathrm{SL}_2(\mathbb{F}_p) \to \mathbb{R}$ and a set $B \subseteq \mathbb{F}_p$ the following holds

$$\sum_s F(s) \sum_{x \in B} B(sx) \ll |B| \|F\|_1 p^{-\delta}, \tag{19}$$

where $\delta = 1/2^{k+2}$ and $k \ll \frac{\log p}{\log K(f)}$, see details in [30] and in [40, Section 6, Theorem 49] (actually, one needs to use the balanced functions in formulae (18), (19)).

To prove our lemma we apply the first bound (18) with $f(x) = G(x)$ and the maximal l such that $2^l \leqslant 2m$. After that we use the second estimate (19) with $F(x) = r_{f,2^l}(x)$. Thanks to (17) we know that $K(F) = K(r_{f,2^l}) \geqslant p^{\tau/6}$. Hence recalling that $m = \tau/4 \cdot \log_N p$, and putting $\delta = \delta(\tau) = \exp(-C/\tau)$, where $C > 0$ is an absolute constant, we derive

$$\sum_s \sum_{x \in B} G(s)A(sx) - \frac{|A||B||G|}{p} \ll \sqrt{|A||B|} |G| p^{-\delta/24m} \ll \sqrt{|A||B|} N^{1-\kappa},$$

where $\kappa > 0$ is another absolute constant. Thus we have obtained bound (14) for the set G. As for our initial family of maps (15), then, of course the multiplication of G by any element of $\mathrm{GL}_2(\mathbb{F}_p)$ does not change the energy T_k and hence everything remains to be true for the set defined in (15). An alternative (but essentially equivalent) way to obtain the required result is to show that all non–trivial representations of the non–commutative Fourier transform of the characteristic function of G enjoy an exponential saving, see [40, Corollary 50]. This completes the scheme of the proof of our lemma. $\qquad \square$

Now we are ready to prove Theorem 1 in the case of prime q. Take a parameter $\varepsilon \in (0, 1/2]$, which we will choose later and let $t = q^{1/2-\varepsilon}$. We assume that $t = o(\sqrt{q})$, $q \to \infty$ and hence we have the condition

$$\varepsilon \gg \frac{1}{\log q} \,. \tag{20}$$

Let $\mathcal{B} = \{0, 1, \ldots, cq/t^2 - 1\} = [0, 1, \ldots, cq^{2\varepsilon} - 1]$, where $c = \min\{c_1/(4c_2), 1/4\}$. Then for a certain set of shifts $\mathcal{A}$ and a set Ω, $|\Omega| \leqslant |\mathcal{B}| T \leqslant cc_2 q^{2\varepsilon} t^{2w_M}$ one has

$$Z_M := Z_M(t) = (\mathcal{B} + (\mathcal{B} \dotplus \mathcal{A})) \bigsqcup \Omega = (\mathcal{B} + Q) \bigsqcup \Omega = \tilde{Z}_M \bigsqcup \Omega \,. \tag{21}$$

We have $|Z_M| \geqslant c_1 q^{2\varepsilon} t^{2w_M}/2$ and hence $|\tilde{Z}_M| \geqslant |Z_M|/2$. Let J be the maximal interval such that $2 \cdot J \subset \mathcal{B}$. Thus $N := |J| \geqslant |\mathcal{B}|/4$. Using Lemma 4 (recall once again that q is a prime number and thus one can apply this lemma) with $A = B = Q = \mathcal{B} \dotplus \mathcal{A}$ and $J = J$, we obtain for a certain absolute constant $C > 0$ that

$$|\{(a+i)(b+i) = 1 \ : \ a, b \in Q, \ i \in 2 \cdot J\}| \geqslant \frac{N|Q|^2}{q} - C|Q|N^{1-\kappa} \geqslant \frac{N|Q|^2}{2q} > 0 \,. \tag{22}$$

To satisfy the last inequality, we need the condition $|Q|N^\kappa \gg q$. In other words, we must have

$$q^{2\varepsilon(1+\kappa-w_M)} \gg q^{1-w_M} \tag{23}$$

or, equivalently, (recall that $1 - w_M \sim 1/M$)

$$\varepsilon \gg \frac{1}{M} \,. \tag{24}$$

Returning to (22) and using decomposition (21), we see that there are $z_1, z_2 \in \tilde{Z}_M \subseteq Z_M$ with $z_1 z_2 \equiv 1 \pmod{q}$. Put $a = z_1$. In view of Lemma 2 we have that for all $x \leqslant t$ and $1 \leqslant |y| < q$ with $ax \equiv y \pmod{q}$ one has $x|y| \geqslant q/4M$. Now we recall a well–known fact that the continued fractions are connected with the question of finding the inverse a^{-1} modulo q, see [21]. More precisely, we have

$$\frac{a^{-1}}{q} = [0; c_s, c_{s-1} \ldots, c_1] \qquad \text{if } s \text{ is even} \tag{25}$$

$$\frac{a^{-1}}{q} = [0; 1, c_s - 1, c_{s-1} \ldots, c_1] \quad \text{if } s \text{ is odd.} \tag{26}$$

Thus in view of formulae (25), (26) for any $x \leqslant t$ and $1 \leqslant |y| < q$ with $a^{-1}x \equiv y \pmod{q}$ one has $x|y| \geqslant q/4M$. The last modular equation is equivalent to $x \equiv ya \pmod{q}$ and hence any solution of (6) satisfy

$$x|y| \geqslant \frac{q}{4M} \qquad \text{for} \qquad x \in [t] \qquad \text{and} \qquad x \in \left[\frac{q}{4Mt}, q\right) \,.$$

Putting $t = \sqrt{q/4M}$ we see by Lemma 2 that all partial quotients of a/q are bounded by $4M$. Since $t = q^{1/2-\varepsilon}$, it follows that $2M^{1/2} = q^\varepsilon$ or, equivalently, $\varepsilon \sim \log M/\log q$. We need to satisfy conditions (20) and (24). Hence it is enough to have

$$M \log M \gg \log q$$

as required. □

Let us make one more remark. In [41, Theorem 3] it was proved

Theorem 5 *Let p be a prime number, $\delta \in (0,1]$, $N \geqslant 1$ be a sufficiently large integer, $N \leqslant p^{c\delta}$ for an absolute constant $c > 0$, $A, B \subseteq \mathbb{F}_p$ be sets, and $g \in \mathrm{SL}_2(\mathbb{F}_p)$ be a non–linear map. Suppose that S is a set, $S \subseteq [N] \times [N]$, $|S| \geqslant N^{1+\delta}$. Then there is a constant $\kappa = \kappa(\delta) > 0$ such that*

$$\left| \{ g(\alpha + a) = \beta + b \ : \ (\alpha, \beta) \in S, \ a \in A, \ b \in B \} \right| - \frac{|S||A||B|}{p} \ll_g \sqrt{|A||B|} |S|^{1-\kappa} . \qquad (27)$$

Taking $S = [N] \times [N]$, $\delta = 1$ and $gx = 1/x$, we get an analogue of Lemma 4 for the correspondent two–dimensional family of modular transformations. This more flexible method gives an alternative way to obtain our main Theorem 1 in the prime case.

3 The general case

We need some definitions, which will be used in this section. By π_n denote the canonical projection modulo n. Having a matrix

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = (\alpha\beta | \gamma\delta) \in \mathrm{Mat}_2(\mathbb{R})$$

we write $\|g\|$ for $\sqrt{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}$. The same can be defined for $\mathrm{Mat}_d(\mathbb{R})$, $d > 2$. Recall that given an arbitrary set $A \subseteq \mathbf{G}$ in a group $\mathbf{G}$ one can define the *Cayley graph* $\mathrm{Cay}(\mathbf{G}, A)$ with the vertex set $\mathbf{G}$ and a pair $(x, y) \in \mathbf{G} \times \mathbf{G}$ forms an edge iff $yx^{-1} \in A$. Having a probability measure ν on $\mathrm{SL}_d(\mathbb{R})$ (that is, a non–negative function with $\sum_x \nu(x) = 1$), let us define the *top Lyapunov exponent*

$$\lambda_1(\nu) = \lim_{n \to \infty} \frac{1}{n} \int \log \|g\| \, dr_{\nu,n}(g) , \qquad (28)$$

where we have assumed that $\int \log \|g\| \, d\nu(g) < \infty$, say (below our measures ν are supported onto a finite number of matrices and hence this condition trivially takes place). Basically, we are working in SL_2 and hence we do not need higher Lyapunov exponents (obviously, the second one is $-\lambda_1(\nu)$).

Now to consider the general case of an arbitrary composite q we naturally require a theory of the growth in $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ or, even more generally, in $\mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$ with $d > 2$ due to we want to obtain an appropriate generalization of Lemma 4. The question on the growth was considered in [4], [7], [27] and also in [10], [34]. For example, let us formulate an application of this technique, see [7].

Theorem 6 *Let $S \subset \mathrm{SL}_d(\mathbb{Z})$ be a finite and symmetric set. Assume that S generates a subgroup $G < \mathrm{SL}_d(\mathbb{Z})$ which is Zariski dense in SL_d.*
Then $\mathrm{Cay}(\pi_q(G), \pi_q(A))$ form a family of expanders, when S is fixed and q runs through the integers. Moreover, there is an integer q_0 such that $\pi_q(G) = \mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$ if q is coprime to q_0.

It is well–known [43] that if S generates a subgroup G which is Zariski dense in SL_d, then G contains a subgroup Γ, which is free on two generators and is Zariski dense in SL_d. All calculations in [4], [5], [7] concern this smaller free group Γ. Roughly speaking, in our proofs we check that these calculations remain to be true for the set G from (16), which generates a free subgroup of rank N. For simplicity, we start with the case of square–free q where the required theory of the growth in $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ is more concrete. The general case will be considered in Subsection 3.2 and our discussion follows paper [7] (notice that, actually, the proof in [7] even does not suppose that the number of generators is exactly two), as well as [4] and [36]. Finally, notice that the condition of Theorem 6 that q coprime to q_0 says, basically, that all prime divisors of q must be sufficiently large.

3.1   The square–free case

In this subsection let q be a sufficiently large square-free number and we want to obtain an analogue of Theorem 1, that is we want to find a positive a such that $(a, q) = 1$ and

$$\frac{a}{q} = [0; c_1, \ldots, c_s], \qquad c_j \leqslant M, \qquad \forall j \in [s],$$

where

$$M = O(\log q / \log \log q).$$

In this case the general scheme of the proof remains the same (of course one should replace q in (22), (23) by $q^{1+o(1)}$ because we consider $\mathbb{Z}_q^*$ but not just $\mathbb{Z}_q$, anyway condition (24) does not change) and to prove the required analogue of Lemma 4 for square–free q we need the crucial result of paper [5, Proposition 4.3].

Theorem 7 *Let q be a square–free number, $q = \prod_{p \in \mathcal{P}} p$. Also, let $A \subset \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ be a set, $\kappa_0, \kappa_1 > 0$ be constants such that $q^{\kappa_0} < |A| < q^{3-\kappa_0}$, further*

$$|\pi_{q_1}(A)| > q_1^{\kappa_1}, \qquad \forall q_1 | q, \qquad q_1 > q^{\kappa_0/40}, \tag{29}$$

and for all $t \in \mathbb{Z}/q\mathbb{Z}$, for any $b \in \mathrm{Mat}_2(q)$ with $\pi_p(b) \neq 0$, $\forall p \in \mathcal{P}$ we have

$$|\{x \in A \ : \ \gcd(q, \mathrm{Tr}(bx) - t) > q^{\kappa_2}\}| = o(|A|), \tag{30}$$

where $\kappa_2 = \kappa_2(\kappa_0, \kappa_1) > 0$. Then

$$|A^3| > q^{\kappa}|A| \tag{31}$$

with $\kappa = \kappa(\kappa_0, \kappa_1) > 0$.

One of the pleasant features of Theorem 7 is that it does not require the knowledge of the subgroup lattice of $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ (which is rather complex for square–free numbers q although, of course $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z}) \simeq \prod_{p \in \mathcal{P}} \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ by the Chinese remainder theorem).

Now to obtain Lemma 4 for square–free numbers we apply the usual Bourgain–Gamburd machine as in the previous section and we use the notation of it as well. The only thing we need to check is that for any $z \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ the product zP_* of the set

$$P_* = \{x \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z}) \ : \ \Delta < r_{G,2l}(x) \leqslant 2\Delta\} \tag{32}$$

satisfies all conditions of Theorem 7, see the proof of [40, Theorem 49] or Theorem 9 below. Here $l \geqslant m = \tau/4 \cdot \log_N q$ (see Section 2) and Δ is a positive number such that

$$\Delta |P_*| \geqslant \frac{|G|^{2l}}{K_*}, \tag{33}$$

where $K_* = q^\varepsilon$ for a certain small $\varepsilon > 0$. Notice that P_* is a symmetric set (although it is not really important for us).

To check all conditions of Theorem 7 we, basically, repeat the calculations from [5, pages 595–599]. Condition (29) follows rather quickly. Indeed, take any $q_1 | q$ such that $q_1 > q^{\kappa_0/40}$ and choose $m_1 \leqslant m$ with $(5N^2)^{10m_1} \sim q_1$. Also, notice that $\max_{g \in G} \|g\| \leqslant 5N^2$. Then $\pi_{q_1} : G^{2m_1} \to \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ is one–to–one. In view of (33), we obtain

$$\frac{|G|^{2l}}{K_*} \leqslant \Delta |P_*| \leqslant \sum_{x \in P_*} r_{G,2l}(x) \leqslant |G|^{2l-2m_1} \max_{w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})} \sum_{x \in wP_*} r_{G,2m_1}(x), \tag{34}$$

and hence by the well–known Kesten result [24] on random walks, we have for $w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ maximizing (34) that

$$|wP_* \cap \mathsf{supp}(G^{2m_1})| \geqslant \frac{|G|^{2m_1}}{K_*} \cdot (2|G| - 1)^{-m_1}. \tag{35}$$

Using the last bound, we get

$$|\pi_{q_1}(zP_*)| = |\pi_{q_1}(wP_*)| \geqslant |wP_* \cap \mathsf{supp}(G^{2m_1})| \geqslant \frac{|G|^{m_1}}{2^{m_1} K_*} \gg K_*^{-1} q_1^{1/40} = K_*^{-1} q^{\kappa_0/1600} \tag{36}$$

as required (let $\varepsilon \leqslant \kappa_0/3200$ and $\kappa_1 = \kappa_0/5000$, say).

Further notice that we can easily assume that $q^{\kappa_0} < |P_*| = |zP_*| < q^{3-\kappa_0}$. Indeed, if $|P_*| \geqslant q^{3-\kappa_0}$ for sufficiently small κ_0 (actually, the bound $|P_*| \geqslant q^{2+\zeta}$ for any $\zeta > 0$ in enough), then one can apply a suitable variant of the Frobenius Theorem [13] (an appropriate adaptation to the square–free case can be found in [5, pages 587–588] or in [4, Lemma 7.1], also see [40, Theorem 49]). The inequality $|P_*| > q^{\kappa_0}$ is also almost immediate. Indeed, as $l \geqslant m$ we have by the Kesten bound as above in (35)

$$\frac{|G|^{2l}}{K_*} \leqslant \Delta |P_*| \leqslant \sum_{x \in P_*} r_{G,2l}(x) \leqslant |P_*|(2|G|)^m |G|^{2l-2m}$$

and hence $|P_*| \geqslant (|G|/2)^m K_*^{-1} \geqslant q^{\tau/4} 2^{-m} K_*^{-1} \gg q^{1/80} K_*^{-1}$ and choosing sufficiently small ε one can take $\kappa_0 = 1/100$, say.

Now it remains to check the property (30) and here we use calculations from [5, pages 597–599]. It is sufficient to show that for all $t \in \mathbb{Z}/q\mathbb{Z}$, for any $b \in \mathrm{Mat}_2(q)$, $\pi_p(b) \neq 0$, $\forall p \in \mathcal{P}$, and for all $q_2 | q$ satisfying $q_2 > q^{\kappa_2}$, we have

$$|\{x \in zP_* \ : \ \mathrm{Tr}(gx) \equiv t \pmod{q_2}\}| \leqslant q^{-\epsilon} |P_*| \tag{37}$$

for a certain $\epsilon > 0$. Let us choose m_2 such that $(5N^2)^{16m_2} \sim q_2$. Assuming that (37) fails, we derive as in (34) that for a certain $w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ one has

$$\sum_{x \in G^{2m_2} \,:\, \mathrm{Tr}(bwx) \equiv t \pmod{q_2}} r_{G,2m_2}(x) \geqslant |G|^{2m_2} K_*^{-1} q^{-\epsilon}. \tag{38}$$

Clearly, for $b' := bw$ one has $\pi_p(b') \neq 0$, $p \in \mathcal{P}$. Let $T \subseteq G^{2m_2}$ be the set of x from (38). It is easy to see that for any $x \in T$ one has $\|x\| \leqslant (5N^2)^{2m_2}$ and that the set T is a hyperspace in our four–dimensional vector space $\mathrm{Mat}_2(q)$ equipped with the standard inner product $\langle A, B \rangle := \mathrm{Tr}(AB^*)$. Then for any $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x \in T$, we derive for an arbitrary $p|q_2$ that

$$f(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x) :=$$

$$\det \begin{pmatrix} x_{11}^{(1)} - x_{11} & x_{11}^{(2)} - x_{11} & x_{11}^{(3)} - x_{11} & x_{11}^{(4)} - x_{11} \\ x_{12}^{(1)} - x_{12} & x_{12}^{(2)} - x_{12} & x_{12}^{(3)} - x_{12} & x_{12}^{(4)} - x_{12} \\ x_{21}^{(1)} - x_{21} & x_{21}^{(2)} - x_{21} & x_{21}^{(3)} - x_{21} & x_{21}^{(4)} - x_{21} \\ x_{22}^{(1)} - x_{22} & x_{22}^{(2)} - x_{22} & x_{22}^{(3)} - x_{22} & x_{22}^{(4)} - x_{22} \end{pmatrix} \equiv 0 \pmod{p}. \tag{39}$$

Clearly, the determinant above does not exceed $15 \cdot 2^{19}(5N^2)^{8m_2} < q_2$, say, and hence this determinant is just zero in $\mathbb{Z}$. Whence it is zero modulo any prime number and we choose a prime P such that

$$\log P \sim 2m_2 \cdot \log N \tag{40}$$

(in [5] the number P was just $\log P \sim 2m_2$ and this choice corresponds to the fixed number of generators, that is, $N = O(1)$ here). Notice that

$$P \geqslant \exp(\Omega(m_2 \log N)) \geqslant q_2^{\Omega(1)} \geqslant q^{\Omega(\kappa_2)}. \tag{41}$$

Let us estimate $\pi_P(T)$ from below. It will allow us to obtain a lower bound for the number of the solutions to equation (39) modulo P as $|\pi_P(T)|^5$. One the other hand, there is a universal Weil–type upper bound (even a rough estimate works) for the number of the solutions to the polynomial equation $f(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x) \equiv 0 \pmod{P}$ with variables in $\mathrm{SL}_2(\mathbb{Z}/P\mathbb{Z})$ and having the form $O(P^{14})$, see details and the required references in [5, page 599]. It will give the desired contradiction and hence the demanded bound (37) takes place.

Thus it requires to estimate $\pi_P(T)$ from below. By the previous section, that is, by the expansion result in $\mathrm{SL}_2(\mathbb{Z}/P\mathbb{Z})$ we know that in this group one has $r_{G,2m_2}(x) \ll |G|^{2m_2}/P^3$, thanks to our choice of P (and m_2). Returning to calculations in (38) and using the last bound, we get

$$|\pi_P(T)| \cdot |G|^{2m_2}/P^3 \gg |G|^{2m_2} K_*^{-1} q^{-\epsilon} \tag{42}$$

and hence $|\pi_P(T)| \gg P^3 K_*^{-1} q^{-\epsilon}$. Thanks to (41) it gives us at least $P^{15} K_*^{-5} q^{-5\epsilon} \gg P^{14}$ solutions to equation (39) modulo P (here ϵ and ε are sufficiently small numbers) and this is a contradiction. As we have seen from the proof the square–free case is reduced to the prime case, eventually.

Again an alternative way of the proof is to use the girth–free result [41, Theorem 3] and work with the two–dimensional family of modular transformations. $\qquad\square$

3.2 Discussion and completion of the proof

As we have seen in the previous subsection the result for square–free q can be derived from an appropriate version of the Helfgott growth theorem in $\mathrm{SL}_2(\mathbb{F}_p)$, see [22] and [3]. The growth result in $\mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})$ follows a similar scheme (combining with a deep but independent sum–product theorem in $\mathbb{Z}/q\mathbb{Z}$, see [2] plus some additional ideas, of course), that is, it follows from the growth result for prime P, see [4, formulae (4.2), (4.3) and Proposition 4.2]. As in (40) we chose P as $\log N \cdot 2m_2 \sim \log P \gg \log q$, where $q = p^n$ and thus the parameter $l \sim m_2$ in [4, see estimates (3.8), (3.9), (4.2) and further formulae] is now $l \sim \log_N P$ but not just $\log P$. Once again, it matches with the calculations of the previous subsection and reflects the fact that now we have N free generators instead of $O(1)$ and all of them have norm at most $5N^2$ but not $O(1)$. Hence we obtain Theorem 1 for $q = p^n$ for all sufficiently large primes p and n rather easily. On the other hand, for small p the result follows from the well–known Folding lemma [32].

Lemma 8 *Let $\tilde{q} \geqslant 2$ be an integer. Then for any positive integer n there exists $a_n, (a_n, \tilde{q}) = 1$ such that in the continued fraction expansion*

$$\frac{a_n}{\tilde{q}^n} = [0; c_1, \ldots, c_s]$$

all partial quotients are bounded by $c_j \leqslant \tilde{q}^2 - 1$, $j \in [s]$.

P r o o f. We use the argument from Niederreiter [32] based on the Folding lemma (see [29, 33]). It is clear that the result is true for $n = 1, 2$. Suppose that a positive integer Q can be represented via a continuant

$$Q = K(c_1, \ldots, c_{t-1}, c_t) = K(c_t, c_{t-1}, \ldots, c_1) = K(1, c_t - 1, c_{t-1}, \ldots, c_1), \quad \text{where} \quad c_j \geqslant 2 \quad (43)$$

with bounded elements $c_j \leqslant M$, $j \in [t]$. By the Folding lemma for any positive integers c_j and X we have the equality

$$K(c_1, \ldots, c_{t-1}, c_t, X, 1, c_t - 1, c_{t-1}, \ldots, c_1) \tag{44}$$

$$= K(c_1, \ldots, c_{t-1}, c_t) \cdot K(1, c_t - 1, c_{t-1}, \ldots, c_1)(X + 1) = Q^2(X + 1). \tag{45}$$

Let $Q = \tilde{q}^n$. Clearly, the continuant in (44) has elements bounded by $\max(M, X)$. Choosing $X = \tilde{q} - 1$ and $X = \tilde{q}^2 - 1$ and combining formulae (43) and (45), we obtain representations of $\tilde{q}^{2n+1}$ and $\tilde{q}^{2n+2}$ via continuants with elements bounded by $\max(M, \tilde{q}^2 - 1)$. Consider the sets

$$A_1 = \{1, 2\} \quad \text{and} \quad A_{n+1} = A_n \cup \{2n + 1, 2n + 2 : \ x \in A_n\} \quad \text{for} \quad n \geqslant 1.$$

Now $\bigcup_{n=1}^{\infty} A_n$ is the set of all positive integers and the result follows. $\qquad\square$

In the general case the argument [7], which allows to obtain Theorem 6, say, is different and it based (besides deep consideration of [7], of course) on very strong tools from [6]. Let us recall the driving result on the growth in $\mathrm{SL}_d(\mathbb{Z}/Q\mathbb{Z})$, see [7, Proposition 2].

Theorem 9 *Let $G \subset \mathrm{SL}_d(\mathbb{Z})$ be a symmetric finite set, G generates a group Γ which is Zariski–dense in SL_d. Then for any $\varepsilon > 0$ there is $\delta > 0$ such that the following hold. If $P' \subseteq \Gamma$ is a symmetric set and l, Q, $(Q, q_0) = 1$ are sufficiently large integers satisfying*

$$\sum_{x \in P'} r_{G,l}(x) > \frac{|G|^l}{Q^\delta}, \quad l > \delta^{-1} \log Q \quad and \quad |\pi_Q(P')| < Q^{3-\varepsilon}, \tag{46}$$

then $|(P')^3| > |P'|^{1+\delta}$. Here q_0 is a fixed positive integer (depending on G).

We need to check conditions (46) for a shift zP_* of our set P_* from (32), (33) and the set G is the same as in (16) (clearly, G generates a (semi)group Γ which is Zariski–dense in SL_d). But thanks to assumption (33) one can see that the first condition of (46) trivially takes place (with $l = 2l$ and $K_* = Q^\delta$), further as we have discussed before the third assumption follows from the Frobenius Theorem (again, an appropriate adaptation for general Q can be found in [7, Page 5] and in [4, Lemma 7.1]). Also, thanks to the Plünnecke–Ruzsa inequality [38] (or see [42]) it is easy to check that the growth of our symmetric set P', namely, $|(P')^3| > |P'|^{1+\delta}$ implies the growth of any of its shift $|(zP')^3| > |P'|^{1+c'\delta}$, where $c' > 0$ is an absolute constant (just consider $zP'(zP')^{-1}zP' = z(P')^3$). Thus we can think below that z is the identity and thus we can work with the set P_* solely. The only thing we need to check is the second condition $l > \delta^{-1} \log Q$, which must be replaced to $l > \delta^{-1} \log_N Q$. Then formula [7, estimate (3)] obviously works, as well as the proof of Proposition 3, page 9 of the same paper due to the fact that this proposition requires to consider just the square–free case, which was obtained in the previous subsection. Also, notice that the constant $C(d, L)$ from the proposition remains to be constant under this choice of l as calculations [7, page 9] show and this is important for us.

Theorem 9 follows from the combination of Proposition 3 and Proposition 6 of [7]. Thus it remains to check that the choice $l > \delta^{-1} \log_N Q$ does not change Proposition 6 in our particular case. Here the authors use a deep result from [6] and we formulate a convenient consequence of it (see [6, Theorem A] and [7, Theorem B, Lemma 7]).

Theorem 10 *Let $S \subset \mathrm{SL}_d(\mathbb{Z})$ be a symmetric set, S generates a subgroup $\Gamma < \mathrm{SL}_d(\mathbb{Z})$ which acts proximally and strongly irreducibly on $\mathbb{R}^d$. Assume further that any finite index subgroup of Γ generates the same $\mathbb{R}$–subalgebra of $\mathrm{Mat}_d(\mathbb{R})$ as Γ.*
Then there is a constant $c_0 > 0$ for any $a, b \in \mathbb{Z}^d \setminus \{0\}$, a is coprime to q we have

$$|S|^{-l} \sum_g e^{\frac{2\pi i \langle ga, b \rangle}{q}} r_{S,l}(g) \ll (q/\mathrm{lcm}(q, b))^{-1/C} \tag{47}$$

for $l \gg \max\{\lambda_1^{-1}(\nu) \cdot \log q, \log q\}$. Here the measure ν is $\nu(x) = S(x)/|S|$.

The proof of Theorem 10 based on the theory of products of random matrices [1], [11], [15] and in particular, on the large deviations for the Lyapunov exponents, see [6, Theorem 4.3]. It is easy to calculate the top Lyapunov exponent $\lambda_1(\nu)$ in our two–dimensional case, namely, $\lambda_1(\nu) \sim \log N$ (and as we said before $\lambda_2(\nu) = -\lambda_1(\nu)$) see, e.g., formula (66) below. Further one problem with [6, Theorem 4.3] is that all bounds here depend on ν (and hence on N). Again, in our two–dimensional case everything can be calculated effectively thanks to reducing the problem

to classical ergodic theorems with the Gauss shift T, see estimate (66) of Theorem 14 from the appendix. Nevertheless, the dependence on N in [6] does not allow to get the required bound for l (basically, due to the fact that the large deviations bounds do not use the circumstance that the top Lyapunov exponent $\lambda_1(\nu) \sim \log N$ is growing) and we leave the possibility of it as an open

Question. Is it possible to obtain Theorem 10 with $l \gg \log_N Q$ for our concrete set G of two–dimensional matrices? If so, it would allow to obtain another proof of Theorem 1 for all q with sufficiently large prime factors.

Anyway at the moment we cannot use a rather general technique from paper [6]. Instead of this we restrict ourselves to the case $d = 2$ and follow the scheme of the proof [36, Theorem 5], as well as [4, Section 4], which we have already discussed above.

The following simple lemma is a slight generalization of Exercise 1.1.8 in [42].

Lemma 11 *Let* $\mathbf{G}$ *be a group and* $A, B \subseteq \mathbf{G}$ *be sets. Then there exists a set* $X \subseteq ABB^{-1}$ *with*

$$|X| \ll \frac{|ABB^{-1}|}{|B|} \cdot \log |AB|$$

such that $AB \subseteq XB$.

Now let us obtain the following "escaping" result for our set P_*. Actually, it is a small modification of [4, Lemma 4.1] and we almost repeat the proof of it in the particular case of a linear function $f(g) = \mathrm{Tr}(wg)$, $w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ (also, see calculations (34), (39) of the previous subsection). As above we identify $\mathrm{Mat}_2(\mathbb{Z})$ with $\mathbb{Z}^4$, e.g., for $g_1, g_2, g_3, g_4 \in \mathrm{Mat}_2(\mathbb{Z})$ by (g_1, g_2, g_3, g_4) we denote the correspondent 4×4 matrix.

Lemma 12 *Let* q_* *be a divisor of* q, P_* *be a set as in* (32), *satisfying* (33) *and let* $r > 0$ *be an integer. Suppose that* $|P_*^3| = K|P_*|$. *Also, let* $f(g)$ *be a linear function on* $\mathrm{SL}_2(\mathbb{Z})$ *in 4 variables, which does not vanish identically on* $\mathrm{SL}_2(\mathbb{Z})$. *Then*

$$|\{g \in P_*^r \ : \ f(g) \equiv 0 \pmod{q_*}\}| \ll_f K^{2r^2} K_* \log^r q \cdot \frac{|P_*|}{q_*^c}, \tag{48}$$

where $c > 0$ *is an absolute constant.*

P r o o f. In view of Lemma 11, as well as the Plünnecke–Ruzsa inequality [38] (or see [42]) we can split the set P_*^r as XP_*, where $|X| \ll K^{2r^2} \log^r q$. Thus it is enough to obtain (48) for any set of g in zP_*, where $z \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and after that sum up all bounds. Further as in (34), (39) it is sufficient to consider the case $f(g) = 0$ (the equality in $\mathbb{Z}$) and then the case $f(g) \equiv 0 \pmod{q_*}$ will easily follow if we take $l_* = c_* \log_N q_*$, where $c_* > 0$ is a sufficiently small constant and consider just $2l_*$-th power of G, see below. Fix z and denote by $S = S_z$ the set of $g \in P_*$ with $f(zg) \equiv 0 \pmod{q_*}$. Then by the definition of the set P_* one has for a certain new $z' \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$

$$|G|^{2l-2l_*} \sum_{g \in S} r_{G,2l_*}(z'g) \geqslant \sum_{g \in S} r_{G,2l}(g) \geqslant |S|\Delta. \tag{49}$$

Here we have used that $l \geqslant m = \tau/4 \cdot \log_N q$ and thus we can assume that $l_* \leqslant l$. Now recall that f is a linear function on $\mathrm{SL}_2(\mathbb{Z})$. In other words, in the space $\mathrm{Mat}_2(\mathbb{Z})$ equipped with the inner product $\langle \cdot, \cdot \rangle$, we have for a certain $w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and $C \in \mathbb{Z}/q\mathbb{Z}$ that $f(g) = \mathrm{Tr}(wg) + C = \langle w, g^* \rangle + C$. Taking $g_1, \ldots, g_5 \in S$ which take part in the first summation from (49), we get $\mathrm{Tr}(wz'g_j) + C \equiv 0 \pmod{q_*}$, $j \in [5]$ and hence

$$\langle g_1 - g_2, (wz')^* \rangle = \cdots = \langle g_1 - g_5, (wz')^* \rangle \equiv 0 \pmod{q_*}.$$

Considering the adjoint matrix, we see that $\det(g_1 - g_2, \ldots, g_1 - g_5) \cdot (wz')^* \equiv 0 \pmod{q_*}$. But $(wz')^* \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$, further q_* is a divisor of q by our assumption and hence $\det = \det(g_1, \ldots, g_5) := \det(g_1 - g_2, \ldots, g_1 - g_5) \equiv 0 \pmod{q_*}$. Clearly, $|\det| \leqslant 4! 2^4 (5N^2)^{2l_*}$ and the last quantity can be done strictly less than q_* by our choice of the constant c_* in the definition of l_*. Thus $\det = 0$ in $\mathbb{Z}$. Choose a prime P similarly to (40), (41) such that $\log P \sim l_* \log N$. Clearly, we have $\det \equiv 0 \pmod{P}$. By a Weil–type bound as in the previous subsection we have seen that the number of the solutions to the equation is $O_f(P^{14})$. Now by the expansion result in $\mathrm{SL}_2(\mathbb{Z}/P\mathbb{Z})$ (see [22]) we know that in this group one has $r_{G,2l_*}(x) \ll |G|^{2l_*}/P^3$, thanks to our choice of P. As in (42) and in (49), we have

$$|\pi_P(S)| \cdot |G|^{2l_*}/P^3 \gg |S|\Delta|G|^{2l_*-2l}. \tag{50}$$

By our condition (33) and our choice of the parameter l_*, we have (compare with estimate (41))

$$(|S|P^3 K_*^{-1}|P_*|^{-1})^5 \leqslant (|S|\Delta P^3 |G|^{-2l})^5 \ll |\pi_P(S)|^5 \ll_f P^{14}$$

and hence

$$|S| \ll_f K_*|P_*|P^{-1/5} \ll_f K_*|P_*|q_*^{-c},$$

where $c > 0$ is an absolute constant. This completes the proof. $\qquad\square$

Now we are ready to obtain Theorem 1 and as we have discussed above it is enough to prove $|P_*^3| > |P_*|^{1+\delta}$ for the set P_* from (32), which satisfies (33). We write $K = |P_*^3|/|P_*|$ and our task is to obtain a good lower bound for K. As we said before we follow the argument of [36] (with some modifications), which is an adaptation of the general scheme from [22]. In particular, we avoid using the deep sum–product results in $\mathbb{Z}/q\mathbb{Z}$ from [2].

Let $T = T_w$ be the centralizer of an element $w \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$, which we call a maximal torus by uniformity reasons (see the notation from [10], [22], [36], for semisimple elements in SL_d there is no difference between its centralizers and maximal tori=maximum commutative subgroups). We say that T is *involved* with P_* if there are $p_1, p_2 \in P_*$ such that $g := p_1^{-1}p_2 \in T$ and $g \neq \pm I$ (I is the identity matrix). We now conjugate T with all elements of P_*, considering the union $\bigcup_{h \in P_*} hTh^{-1}$. If all maximal tori $T' = hTh^{-1}$, arising thereby, are involved with P_*, we continue conjugating each of these tori with elements of P_*. After that, once again, either we get at least one new torus, which is not involved with P_*, or all the tori, generated so far from T are involved with P_*. And so on. As we have discussed above the set P_* generates $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and since, the procedure will end in one of the two ways: either (i) there is some torus T involved with P_* and a certain $h \in P_*$, such that $T' = hTh^{-1}$ is not involved with P_*, or (ii) for all $h \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ and some (initial maximal torus) T, every torus hTh^{-1} is involved with P_*. Consider the two scenarios separately.

Case (i) – pivot case.

The maximal torus T' is not involved with P_*. However, $T = h^{-1}T'h$ is: there is a non-trivial element $g \in P_*^{-1}P_* = P_*^2$ (here we have used that $P_* = P_*^{-1}$ but it is not really important), lying in $h^{-1}T'h$, therefore $g' = hgh^{-1} \in T'$. Consider the projection

$$\varphi : P_* \to C_\tau, \quad p_* \to p_* g' p_*^{-1} \in P_*^6 \,,$$

where C_τ is the conjugacy class of g with $\mathrm{Tr}(g) = \tau$. This projection is at most two-to-one, for if h_1, h_2 have the same image, this means that $h_1^{-1}h_2 \in T'$, $h_1, h_2 \in P_*$, but T' is not involved with P_*, thus $h_1^{-1}h_2 = \pm I$. It follows that $|P_*^6 \cap C_\tau| \geq |P_*|/2$. Write $P_{**} = P_*^6 \cap C_\tau$. Our task is to find a good upper bound for P_{**} of the form $|P_{**}| \ll |P_*^6|^{1-\varepsilon_0}$, where $\varepsilon_0 > 0$ is an absolute constant. After that the required lower bound for K will follow from the Plünnecke–Ruzsa inequality.

Consider the function $\mathcal{C}(y) = |P_{**} \cap y^{-1}P_{**}|$. By the Cauchy–Schwarz inequality, we have

$$|P_{**}|^4 \leqslant \sum_y \mathcal{C}^2(y) \cdot |P_{**}P_{**}^{-1}| \,. \tag{51}$$

For any $g \in P_{**} \cap y^{-1}P_{**}$ one has $\tau = \mathrm{Tr}(g) = \mathrm{Tr}(yg)$. Applying Lemma 12 with $q_* = q$, $r = 12$ and the following non–vanishing linear function $f(y) = \mathrm{Tr}(yg) - \tau$, we have in view of (51)

$$|P_{**}|^4 \leqslant |P_{**}|^2|P_*^{12}| \cdot K^{288}K_* \log^{12} q \cdot \frac{|P_*|}{q^c} \tag{52}$$

and hence thanks to $|P_{**}| \geqslant |P_*|/2$ and the Plünnecke–Ruzsa inequality, we get

$$q^{c/2} \ll K_* K^{300} \,.$$

Recall that $K_* = q^\varepsilon$ and thus if we take $\varepsilon = c/20$, then one obtains $K \gg q^{c/1000}$, say. It is absolutely enough for our purposes due to the fact that P_* is large (see, e.g., calculations from (36)).

Case (ii) – large set case. Suppose, for any $h \in G$, all tori hTh^{-1} are involved with P_*. The number of such tori (not meeting, except at $\{\pm I\}$) will be calculated in purely algebraic Lemma 13 and (as the worst case scenario) one may assume that $P_*P_*^{-1} \setminus \{\pm I\} = P_*^2 \setminus \{\pm I\}$ is partitioned between these tori.

Thus it follows by the Helfgott orbit–stabilizer Theorem [22], [36, Lemma 11 and page 19] that

$$K|P_*| \geqslant |P_*^2| \geqslant \sum_{h \in \mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})/N(T)} |P_*^2 \cap hTh^{-1}| \gg \frac{|\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})|}{|N(T)|} \cdot \frac{|P_*|}{|P_*^4 \cap C_\tau|} \,, \tag{53}$$

where $N(T)$ is the normaliser of T. Similarly to above (see calculations in (52)) we estimate $|P_*^4 \cap C_\tau|$ as $|P_*^4 \cap C_\tau| \ll K^{150}K_*^{1/2} q^{-c/20}|P_*|$ (actually, before we have considered six products instead of four and hence one can obtain even better bound). Now suppose that we have chosen our torus T as

$$|N(T)| \ll q^{1+\zeta} \,, \tag{54}$$

where $\zeta = \zeta(c) > 0$ is a sufficiently small number. One can see that the size of the normaliser of "typical" T is $O(q)$ and hence bound (54) is close to the optimal. Then thanks to (53), (54), we obtain

$$K^{151} K_*^{1/2} |P_*| \gg q^{2+c/20-\zeta} \geqslant q^{2+c/40},$$

where we have chosen $\zeta \leqslant c/40$. Taking the parameter ε in $K_* = q^\varepsilon$ to be $\varepsilon = c/100$, we see that either $K \gg q^{c/30000}$ or $|P_*| \gg q^{2+c/160}$. In the former case we are done and the last case was discussed before and follows from the Frobenius Theorem (again, an appropriate adaptation for general q can be found in [7, Page 5] and in [4, Lemma 7.1]).

It remains to obtain an algebraic lemma to satisfy condition (54) and we use some ideas of paper [4]. Somehow we need to choose $T = T_w$ such that w is "far" from the identity $\pm I$ (clearly, $|N(\pm I)| \sim q^3$ and hence (54) fails in this case). Below we assume that all primes p (they will be divisors of q) are odd. For any $g \in \mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})$ we write

$$g = \frac{\mathrm{Tr}\, g}{2} I + p^{r(g)} \cdot (ab|c(-a)) = \frac{\mathrm{Tr}\, g}{2} I + p^{r(g)} \cdot g', \tag{55}$$

where not all a, b, c vanish modulo p. Since $\det(g) \equiv 1 \pmod{p^n}$, we have

$$\left(\frac{\mathrm{Tr}\, g}{2} \right)^2 \equiv 1 + p^{2r(g)}(a^2 + bc) \pmod{p^n}, \tag{56}$$

and hence in particular,

$$\mathrm{Tr}\, g \equiv \pm 2 \pmod{p^{s_*(g)}}, \qquad \text{where} \qquad s_*(g) = \min\{n, 2r(g)\}. \tag{57}$$

Writing $r = r(g) = r_p(g)$, we can calculate several algebraic characteristics of g in terms of r.

Lemma 13 *Let* $g \in \mathrm{SL}_2(\mathbb{Z}/p^n\mathbb{Z})$ *and* $r = r_p(g)$. *Then* $|\mathrm{Stab}(g)| \leqslant 8p^{n+2r}$. *Further* $|N(\mathrm{Stab}(g))| \leqslant 300p^{n+3r}$.

P r o o f. Taking $h \in \mathrm{Stab}(g)$ and using (55) with $t := r_p(h)$ and $h' = (\alpha\beta|\gamma(-\alpha))$, we obtain

$$p^{t+r} \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix} \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \equiv \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix} p^{t+r} \pmod{p^n}. \tag{58}$$

We assume firstly that $t + r < n$ and write $q_1 = p^{n-t-r}$ and $q_2 = p^{n-t} \geqslant q_1$. Then we obtain from (58) the following system of equations

$$\beta c \equiv \gamma b \pmod{q_1}, \qquad \alpha b \equiv a\beta \pmod{q_1}, \qquad \gamma a \equiv \alpha c \pmod{q_1}. \tag{59}$$

Since not all a, b, c vanish modulo p, it follows that there are exactly q_1 solutions to system (59). Hence we obtain

$$3p^{3r} q_1 = 3p^{3r} p^{n-t-r} = 3p^{n-t+2r} \leqslant 3p^{n+2r}$$

solutions to (58). Returning to (56), (57) for h, we find $\mathrm{Tr}\, h$ solving the quadratic equation modulo $p^{s_*(h)}$ and then modulo p^n (it gives two solutions) and hence by (55) we reconstruct h.

Now if $t+r \geqslant n$, then we can take $\alpha, \beta, \gamma \in [q_2]$ in an arbitrary way and after that we reconstruct $\operatorname{Tr} h$ as above. It gives us at most

$$2q_2^3 = 2p^{3n-3t} \leqslant 2p^{3r} \leqslant 2p^{n+2r}$$

solutions to (58).

Now let us obtain that $|N(\operatorname{Stab}(g))| \leqslant 300p^{n+3r}$. Let $\mathrm{n} = (\alpha\beta|\gamma\delta) \in N(T)$ and $h \in \operatorname{Stab}(g)$. Suppose that (other cases can be considered in a similar way) in system (59), we have $a \neq 0$ (mod p) and hence

$$h' = h_t' = \mu \cdot (1b|c(-1)) + q_1 \cdot (AB|C(-A)), \tag{60}$$

where $\mu \neq 0$ (mod p) runs over $[q_1]$, A, B, C run over $[p^r]$ and b, $c \in [q_1]$ are some new fixed elements. Having the condition

$$\mathrm{n}^{-1}h\mathrm{n} \equiv \frac{\operatorname{Tr} h_t}{2}I + p^t \cdot \mathrm{n}^{-1}h_t'\mathrm{n} \equiv \frac{\operatorname{Tr} h_{t_1}}{2}I + p^{t'} \cdot \mathrm{n}^{-1}h_{t_1}'\mathrm{n} \pmod{p^n}, \tag{61}$$

we clearly, derive $\operatorname{Tr} h_t = \operatorname{Tr} h_{t_1}$ and thanks to (60) one can see that $t = t'$. Further identity (61) holds for all t and in particular for $t = 0$. Using (60) for this choice of t (it gives us $t' = 0$ and $q_1 = p^{n-r}$), we see that

$$\begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix} \begin{pmatrix} 1 & b \\ c & -1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \equiv \lambda \begin{pmatrix} 1 & b \\ c & -1 \end{pmatrix} \pmod{q_1},$$

where $\lambda \neq 0$ (mod p) is a number. The last equation is equivalent to the system modulo q_1

$$\alpha(\delta - \beta c) + \gamma(\delta b + \beta) = \lambda, \quad 2\beta\delta - \beta^2 c + \delta^2 b = b\lambda, \quad -2\alpha\gamma + \alpha^2 c - \gamma^2 b = c\lambda. \tag{62}$$

Solving the second equation in (62), which is a non–vanishing quadratic equation, we obtain at most $2q_1$ solutions. Now combining the first equation of (62) with another linear equation in α, γ, namely, with $\alpha\delta - \beta\gamma \equiv 1 \pmod{q_1}$, we find the only solution in α, γ unless $\delta(1 - \lambda) = \beta c$, and $-\delta b = \beta(1 + \lambda)$. If the last equation has the only solution in δ, β, then we have at most $2q_1$ solutions in α, γ (it follows from $\alpha\delta - \beta\gamma \equiv 1 \pmod{q_1}$ or from the third equation of system (62)). Otherwise $bc = \lambda^2 - 1$. Here we have used the fact that either β or δ is invertible modulo p. Applying this (without loss of generality we assume that δ is invertible), as well as the third equation from (62), combining with $\alpha\delta - \beta\gamma = 1$, we derive

$$\gamma^2(c\beta^2 - b\delta^2 - 2\beta\delta) + 2\gamma(\beta c - \delta) + c - c\delta^2\lambda = 0. \tag{63}$$

If the last quadratic equation is trivial modulo p^s for a certain s, then we have $\delta = \beta c$, $c\beta^2(1 + bc) = c\beta^2\lambda^2 = 0$ and $c = c^3\beta^2\lambda = 0$. Here we have used that $bc = \lambda^2 - 1$ and $\lambda \neq 0$ (mod p). Hence $\delta = 0$ and returning to (62), we see that $\lambda = -1$, $b = \alpha = 0$. If $p^s = q_1$, then from $\alpha\delta - \beta\gamma \equiv 1 \pmod{q_1}$, we see that there are at most q_1 solutions in β, γ. If $p^s < q_1$, then there exists at most two solutions in γ of equation (63) and we reconstruct α from the third equation of (62), say, in at most two ways. Thus we have in total at most $9 \cdot 2^5 q_1$ solutions modulo q_1 and hence we obtain at most $9 \cdot 2^5 p^{n+3r}$ solutions modulo p^n. This completes the proof of the lemma. $\qquad\square$

Finally, it remains to choose an appropriate initial torus T, satisfying condition (54). Let $q = p_1^{\alpha_1} \dots p_s^{\alpha_s}$. Notice that, by the Chinese remainder theorem, we have $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z}) \simeq \prod_{j=1}^s \mathrm{SL}_2(\mathbb{Z}/p^{\alpha_j}\mathbb{Z})$. Consider the collection $\mathcal{C}$ of all divisors of q, having the size at least q^κ, where $\kappa = \kappa(\zeta) > 0$ is a parameter. Clearly, by the divisor function bound one has $|\mathcal{C}| \leqslant q^{o(1)}$. For any $q' \in \mathcal{C}$ we apply Lemma 12 with $q_* = q'$, $r = 2$, $f_\pm(g) = \mathrm{Tr}\,(g) \pm 2$. Thus we either find an element $w \in P_*^2$ with $\mathrm{Tr}\,w \neq \pm 2 \pmod{q'}$, where q' runs over $\mathcal{C}$ or

$$K^8 K_* \gg q^{\kappa c - o(1)}\,.$$

Suppose that the later holds. Choosing the parameter ε in $K_* = q^\varepsilon$ to be $\varepsilon = \kappa c/4$, we see that $K \gg q^{\kappa c/16}$ and we are done. Now take our element w and consider the following sets

$$G = \{j \in [s] \ : \ r_{p_j}(w) \leqslant p^{\kappa_1 \alpha_j}\}\,, \qquad B = [s] \setminus G\,,$$

where κ_1 is another parameter. By (57) for any $j \in B$ one has either $\mathrm{Tr}\,w \equiv 2 \pmod{p_j^{\kappa_1 \alpha_j}}$, or $\mathrm{Tr}\,w \equiv -2 \pmod{p_j^{\kappa_1 \alpha_j}}$. Hence by our construction of the set $\mathcal{C}$, we have

$$\prod_{j \in B} p_j^{\kappa_1 \alpha_j} \leqslant q^{2\kappa}\,. \tag{64}$$

Using Lemma 13, we derive

$$|N(T_w)| \ll \prod_{j \in G} p_j^{\alpha_j(1+3\kappa)} \cdot \prod_{j \in B} p^{3\alpha_j} \leqslant \prod_j p_j^{\alpha_j(1+3\kappa)} \cdot q^{2\kappa/\kappa_1} \leqslant q^{1+3\kappa+2\kappa\kappa_1^{-1}}\,.$$

It remains to take $\kappa = \kappa_1^2$ and, say, $\kappa_1 = \zeta/8$. Thus we have obtained the required condition (54). This completes the proof of Theorem 1 for general q. $\qquad\square$

4 Appendix

In this section we obtain the large deviations estimate for the top Lyapunov exponent of our set G defined in (16), namely, for the measure $G(x)/|G|$ (one can see that the top Lyapunov exponent is just $\lim_{n \to \infty} \frac{1}{n} \log q_n([0; c_1, -c_1, \dots, c_n, -c_n])$, where $c_j \in 2 \cdot [N]$, $j \in [n]$). Such bounds can be used in the proof of Theorem 10, see [6] in the particular case of the group $\mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$ with $d = 2$ (and for the concrete measure). For $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$ the theory of products of random matrices can be replaced by the standard considerations from the theory of continued fractions. Recall one more time that in the considered regime the parameter N tends to infinity. We hope that Theorem 14 is interesting in its own right and even in the classical case, see formula (65) below. At least we should mention that this inequality implies the identity $g_N := \lim_{n \to \infty} q_n^{1/n}([0; c_1, \dots, c_n]) = N/e + o(N)$ for a.e. $(c_1, \dots, c_n) \in [N]^n$, $N \to \infty$ and this is better than the estimates on g_N from [35, Lemma 3], also see discussion [35, pages 43–44].

In our proof we follow the method from [12]. Recall that by $T : [0, 1] \to [0, 1]$ we denote the Gauss shift, that is, $Tx = \{1/x\}$ for $x \in (0, 1)$ and $T0 = 0$.

Theorem 14 *Let n, N be positive integers, $\delta \in (0, 1]$ be a real number. Then there are absolute constants $\kappa \in (0, 1]$ and $K \geqslant 1$ such that for all $N \geqslant K\delta^{-2} \log(1/\delta)$ and $n \geqslant K\delta^{-1} \log(1/\delta) \log N$ one has*

$$N^{-n} \left| \left\{ (c_1, \ldots, c_n) \in [N]^n \; : \; \left| \frac{1}{n} \log q_n([0; c_1, \ldots, c_n]) - \frac{\log(N!)}{N} \right| \geqslant \delta \right\} \right|$$

$$\leqslant 2 \exp\left(-\frac{\kappa \delta^2 n}{\log(1/\delta)} \right) . \tag{65}$$

Similarly, under the same conditions on n and N the following holds

$$N^{-n} \left| \left\{ (c_1, \ldots, c_n) \in 2 \cdot [N]^n \; : \; \left| \frac{1}{n} \log q_n([0; c_1, -c_1, \ldots, c_n, -c_n]) - \frac{2\log(N!)}{N} \right| \geqslant \delta \right\} \right|$$

$$\leqslant 4 \exp\left(-\frac{\kappa \delta^2 n}{\log(1/\delta)} \right) . \tag{66}$$

P r o o f. Let $L = \log N$. Writing $X_j = [0; c_j, \ldots, c_n]$ and applying the well–known formula $p_j(x) = q_{j-1}(Tx)$ for any $x \in [0, 1]$, we see that

$$q_n([0; c_1, \ldots, c_n]) := q_n(x) = \frac{q_n(x)}{p_n(x)} \cdot \frac{q_{n-1}(Tx)}{p_{n-1}(Tx)} \cdots \frac{q_1(T^{n-1}x)}{p_1(T^{n-1}x)} \tag{67}$$

(we have used that $p_1(T^{n-1}x) = 1$) and hence

$$q_n([0; c_1, \ldots, c_n]) = \prod_{j=1}^{n} X_j^{-1} . \tag{68}$$

Thus it is sufficient to estimate the probability

$$\mathbb{P}_{\delta, [n]} := \mathbb{P} \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} \log X_j + \frac{\log(N!)}{N} \right| \geqslant \delta \right\} .$$

Notice that $J := \frac{\log(N!)}{N}$ is close to the expectation of the random variable $\frac{1}{n} \sum_{j=1}^{n} \log X_j$. Indeed, using the standard estimates for continuants, we have by the stationarity

$$-\frac{1}{n} \sum_{j=1}^{n} \mathbb{E} \log X_j = N^{-2} \sum_{a,b=1}^{N} \log(a + \theta_1 b^{-1}) = \frac{\log(N!)}{N} + \theta_2 \frac{\log^2 N}{N^2} , \tag{69}$$

where here and below $|\theta_j| \leqslant 1$ are some absolute constants. In (69) we have used the approximation

$$X_j^{-1}(c_j, \ldots, c_n) = c_j + \frac{\theta_3}{c_{j+1}} . \tag{70}$$

Similarly, notice that

$$X_j(c_j, \ldots, c_n) = c_j^{-1} + \frac{\theta_4}{c_j^2 c_{j+1}}. \tag{71}$$

Now by our assumption we have $N \geqslant K\delta^{-2} \log(1/\delta)$ and hence the error in (69) is at most $\delta/4$ for large K and hence it is negligible. Also, let us remark that by the Stirling formula one has

$$N^{-1} \log(c_1 N) \leqslant |J - (L-1)| \leqslant N^{-1} \log(c_2 N), \tag{72}$$

(here and below $c_j > 0$ are some absolute constants). Similarly, take any $0 < s \leqslant 1/2$ and using the Euler–Maclaurin formula (or just a direct calculation) and formulae (70), (71), we derive that

$$\log \mathbb{E}|X_1|^s \leqslant -sL - \log(1-s) + \frac{c_3 L}{N} \leqslant -sL + s + s^2 + \frac{c_3 L}{N}, \tag{73}$$

as well as

$$\log \mathbb{E}|X_1|^{-s} \leqslant sL - \log(1+s) + \frac{c_4 L}{N} \leqslant sL - s + \frac{s^2}{2} + \frac{c_4 L}{N}. \tag{74}$$

Now let $4 \leqslant M \leqslant n/4$ be an even parameter and we split $[n]$ into M arithmetic progressions of size $t := [n/M]$, namely, $Q_1, \ldots, Q_M$ having the step M. Since the union of Q_j is $[n]$ plus at most $M - 1$ points, we can assume that n is divisible by M and hence $t = n/M$. Indeed, it requires just to replace δ in $\mathbb{P}_{\delta,[n]}$ to $\delta/2$ and notice that

$$2M \max_j \|\log X_j\|_\infty \leqslant 2ML \leqslant \delta n/4,$$

where the condition $n \geqslant 8ML/\delta$ will be checked later. Now we have $t = n/M$ and use the exponential Markov inequality with a parameter $\lambda > 0$, $\lambda \leqslant 1/(2M)$ and the Hölder inequality to derive

$$\mathbb{P}_{\delta/2,[n]} \leqslant \exp(-\delta\lambda n/2 + \lambda nJ) \cdot \mathbb{E}(\prod_{i=1}^{n} |X_i|^\lambda) = \exp(-\delta\lambda n/2 + \lambda nJ) \cdot \mathbb{E}(\prod_{i=1}^{M} \prod_{j \in Q_i} |X_j|^\lambda)$$

$$\leqslant \exp(-\delta\lambda n/2 + \lambda nJ) \cdot \prod_{i=1}^{M} \left(\mathbb{E} \prod_{j \in Q_i} |X_j|^{\lambda M} \right)^{1/M}. \tag{75}$$

Here we have considered the case when $\frac{1}{n} \sum_{j=1}^{n} \log X_j - \frac{\log(N!)}{N}$ is positive and the opposite situation will be considered below in a similar way. Thus it remains to estimate $\mathbb{E} \prod_{j \in Q_i} |X_j|^{\lambda M}$ for any $i \in [M]$. Using the well–known ψ–mixing property of our shift T with $\psi(m) = C\mu^m$, where $C > 0$ and $1/2 < \mu < 1$ are some absolute constants, we get by the stationarity and the assumption $\lambda M \leqslant 1/2$ (see details in [12, Lemmas 2, 3]) that

$$\mathbb{E} \prod_{j \in Q_i} |X_j|^{\lambda M} \leqslant (1 + \psi(M/2))^t \left(\mathbb{E}|X_1|^{\lambda M} \right)^t.$$

Substituting the last bound into (75) and using estimates (72), (73), we obtain for sufficiently large N, $L/N \ll (\lambda M)^2$ that

$$\mathbb{P}_{\delta/2,[n]} \leqslant \exp(-\delta\lambda n/2 + \lambda nJ + t\psi(M/2) - \lambda MtL + \lambda Mt + 2t(\lambda M)^2)$$

$$\leqslant \exp(-\delta\lambda n/2 + nM^{-1}\psi(M/2) + 4n\lambda^2 M)\,.$$

Now we choose $\lambda = \delta/(16M) \leqslant 1/(2M)$ and after that we take the parameter M such that $M^{-1}\psi(M/2) \leqslant \delta\lambda/8 = \delta^2/(128M)$. In other words, $\psi(M/2) \leqslant \delta^2/128$ and hence we can choose $M \ll \log(1/\delta)$. It gives us

$$\mathbb{P}_{\delta/2,[n]} \leqslant \exp(-\delta\lambda n/8) = \exp(-\delta^2 n/(128M)) = \exp(-\kappa\delta^2 n/\log(1/\delta))\,,$$

where $\kappa > 0$ is an absolute constant. We need to check that $n \geqslant 8ML/\delta$ and $L/N \ll (\lambda M)^2 = 2^{-8}\delta^2$ but our assumptions $N \geqslant K\delta^{-2}\log(1/\delta)$, $n \geqslant K\delta^{-1}\log(1/\delta)\log N$ guarantee it.

Finally, let $\frac{1}{n}\sum_{j=1}^{n}\log X_j + \frac{\log(N!)}{N} < 0$ and hence our exponential Markov inequality requires to estimate the probability

$$\mathbb{P}\left\{\exp\left(-\lambda\sum_{j=1}^{n}\log X_j\right) \geqslant \exp(n\lambda(J + \delta/2))\right\}\,.$$

In this case we use the same calculations, the same choice of the parameter $\lambda = \delta/(16M) \leqslant 1/2$, as well as formulae (72), (74) to get for sufficiently large N such that $L/N \ll (\lambda M)^2$

$$\mathbb{P}_{\delta/2,[n]} \leqslant \exp(-\delta\lambda n/2 - \lambda nJ + t\psi(M/2) + \lambda MtL - \lambda Mt + t(\lambda M)^2)$$

$$\leqslant \exp(-\delta\lambda n/2 + nM^{-1}\psi(M/2) + 2n\lambda^2 M) \leqslant \exp(-\kappa\delta^2 n/\log(1/\delta))\,.$$

It remains to obtain estimate (66). As in (67), (68) (recall that we assume that M and hence n are even numbers) we derive

$$-\log q_n([0; c_1, -c_1, \ldots, c_n, -c_n]) := -\log q_n(x) = \sum_{j=1}^{n}\log Y_j(x) + \sum_{j=1}^{n}\log Z_j(x)\,,$$

where $Y_j = [0; c_j, -c_j, \ldots, c_n, -c_n]$ and $Z_j = [0; c_j, -c_{j+1}, c_{j+1}, \ldots, -c_n, c_n]$. Thus it is sufficient to obtain the large deviation principle for the random variables Y_j, Z_j separately. Similarly to (70), (71), we have (recall that by the assumption $c_j \in 2 \cdot [N]$)

$$Y_j^{-1}(\omega) = c_j + \frac{2\theta_1}{c_j}\,, \qquad\qquad Y_j(\omega) = c_j^{-1} + \frac{2\theta_2}{c_j^3}\,, \qquad\qquad (76)$$

and

$$Z_j^{-1}(\omega) = c_j + \frac{2\theta_3}{c_{j+1}}\,, \qquad\qquad Z_j(\omega) = c_j^{-1} + \frac{2\theta_4}{c_j^2 c_{j+1}}\,. \qquad\qquad (77)$$

Thus we have the same asymptotic formulae for $-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\log Y_j$, $-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\log Z_j$ as in (69). Also, thanks to (76), (77), we get (73), (74) for Y_j, Z_j. After that we repeat the calculation above and obtain the required estimate (66). This completes the proof. $\square$

References

[1] P. BOUGEROL, J. LACROIX, *Products of random matrices with applications to Schrödinger operators, Progress in Probability and Statistics,* vol. 8, Birkhäuser Boston Inc., Boston, MA, 1985.

[2] J. BOURGAIN, *The sum-product theorem in $\mathbb{Z}_q$ with q arbitrary,* Journal d'Analyse Mathématique 106.1 (2008): 1–93.

[3] J. BOURGAIN, A. GAMBURD, *Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$,* Ann. of Math., 167(2):625–642, 2008.

[4] J. BOURGAIN, A. GAMBURD, *Expansion and random walks in $\mathrm{SL}_d(\mathbb{Z}/p^n\mathbb{Z})$:I,* J. Eur. Math. Soc. 10 (2008), 987–1011.

[5] J. BOURGAIN, A. GAMBURD, P. SARNAK, *Affine linear sieve, expanders, and sum–product,* Inventiones mathematicae 179.3 (2010): 559–644.

[6] J. BOURGAIN, A. FURMAN, E. LINDENSTRAUSS, S. MOZES, *Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus,* Journal of the American Mathematical Society, 24(1) (2011): 231–280.

[7] J. BOURGAIN, P.P. VARJÚ, *Expansion in $\mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$, q arbitrary,* Inventiones mathematicae 188.1 (2012): 151–173.

[8] J. BOURGAIN, A. KONTOROVICH, *On Zaremba's conjecture,* C. R. Math. Acad. Sci. Paris, 349(9–10):493–495, 2011. URL: https://doi.org/10.1016/j.crma.2011.03.023, doi:10.1016/j.crma.2011.03.023.

[9] J. BOURGAIN, A. KONTOROVICH, *On Zaremba's conjecture,* Annals of Mathematics 180(1): 137–196, 2014.

[10] E. BREUILLARD, B. GREEN, T. TAO, *Approximate subgroups of linear groups,* Geom. Funct. Anal. 21:4 (2011), 774–819.

[11] P. DUARTE, PEDRO, AND S. KLEIN, *Continuity of the Lyapunov exponents of linear cocycle,* Publicacoes Matematicas do IMPA, 31o Coloquio Brasileiro de Matematica IMPA (2017).

[12] L. FANG, M. WU, N.R. SHIEH, AND B. LI, *Random continued fractions: Lévy constant and Chernoff–type estimate,* Journal of Mathematical Analysis and Applications, 429:1 (2015), 513–531.

[13] G. FROBENIUS, *Über Gruppencharaktere, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin,* (1896), 985–1021.

[14] D.A. FROLENKOV, I.D. KAN, *A strengthening of a theorem of Bourgain-Kontorovich II,* Mosc. J. Comb. Number Theory, 4(1):78–117, 2014.

[15] H. Furstenberg, *Noncommuting random products,* Trans. Amer. Math. Soc., 108:377–428, 1963.

[16] D. Hensley, *The distribution of badly approximable numbers and continuants with bounded digits,* In *Théorie des nombres* (Quebec, PQ, 1987), pages 371–385, de Gruyter, Berlin, 1989.

[17] D. Hensley, *The distribution of badly approximable rationals and continuants with bounded digits II,* J. Number Theory, 34(3):293–334, 1990. URL: https://doi.org/10.1016/0022-314X(90)90139-I, doi:10.1016/0022-314X(90)90139-I.

[18] D. Hensley, *Continued fraction Cantor sets, Hausdorff dimension, and functional analysis,* J. Number Theory, 40(3):336–358, 1992. URL: https://doi.org/10.1016/0022-314X(92)90006-B, doi:10.1016/0022-314X(92)90006-B.

[19] D. Hensley, *The distribution mod n of fractions with bounded partial quotients,* Pacific J. Math., Vol. 166 (1):43–54, 1994.

[20] D. Hensley, *A polynomial time algorithm for the Hausdorff dimension of continued fraction Cantor sets,* J. Number Theory, 58(1):9–45, 1996.

[21] A.Ya. Hinchin, *Continued fractions,* M., Fizmatlit, 1960.

[22] H. Helfgott, *Growth and generation in* $SL_2(Z/pZ)$, Annals of Math. 167 (2008), no. 2, 601–623.

[23] I.D. Kan, *A strengthening of a theorem of Bourgain and Kontorovich. IV,* Izv. Ross. Akad. Nauk Ser. Mat., 80(6):103–126, 2016. URL: https://doi.org/10.4213/im8360, doi:10.4213/im8360.

[24] H. Kesten, *Symmetric random walks on groups,* Transactions of the American Mathematical Society 92 (1959), 336–354.

[25] A. Kontorovich, *From Apollonius to Zaremba: local-global phenomena in thin orbits,* Bulletin of the American Mathematical Society 50.2 (2013): 187–228.

[26] N.M. Korobov, *Number–theoretical methods in numerical analysis,* Moscow, 1963 (in Russian).

[27] M. Magee, H. Oh, D. Winter, *Uniform congruence counting for Schottky semigroups in* $SL_2(\mathbb{Z})$, Journal für die reine und angewandte Mathematik (Crelles Journal) 2019.753 (2019): 89–135.

[28] N.G. Moshchevitin, *Sets of the form $A + B$ and finite continued fractions,* Sbornik:Mathematics, 198(4):95–116, 2007. URL: http://stacks.iop.org/1064-5616/198/i=4/a=A05.

[29] N.G. Moshchevitin, *On some open problems in Diophantine approximation,* arXiv:1202.4539 (2012).

[30] N.G. MOSHCHEVITIN, B. MURPHY, I.D. SHKREDOV, *Popular Products and Continued Fractions,* Israel J. Math., **238** (2020) 807–835; DOI:10.1007/s11856-020-2039-3

[31] N.G. MOSHCHEVITIN, I.D. SHKREDOV, *On a modular form of Zaremba's conjecture,* Pacific J. of Math., **309**:1 (2020), 195–211; DOI 10.2140/pjm.2020.309.195

[32] H. NIEDERREITER, *Dyadic fractions with small partial quotients,* Monatsh. Math., 101(4):309–315, 1986. URL: https://doi.org/10.1007/BF01559394, doi:10.1007/BF01559394.

[33] A. J. VAN DER POORTENM J. SHALLIT, *Folded continued fractions,* J. Number Theory, 40 (1992), 237–250.

[34] L. PYBER, E. SZABÓ, *Growth in finite simple groups of Lie type of bounded rank,* Journal of the American Mathematical Society 29.1 (2016): 95–146.

[35] C.A. ROGERS, *Some sets of continued fractions,* Proceedings of the London Mathematical Society 3.1 (1964): 29–44.

[36] M. RUDNEV, I.D. SHKREDOV, *On growth rate in $SL_2(\mathbb{F}_p)$, the affine group and sum-product type implications,* Mathematika, **68**:3 (2022) 738–783; DOI: 10.1112/mtk.12120

[37] M. G. RUKAVISHNIKOVA, *Probabilistic bound for the sum of partial quotients of fractions with a fixed denominator,* Chebyshevskii Sbornik 7 (2006), 113–121.

[38] I.Z. RUZSA, *Sums of Finite Sets,* in: Chudnovsky D.V., Chudnovsky G.V., Nathanson M.B. (eds) NumberTheory: New York Seminar 1991–1995. Springer, New York, NY.

[39] I.D. SHKREDOV, *Growth in Chevalley groups relatively to parabolic subgroups and some applications,* Rev. Mat. Iberoam., accepted; DOI 10.4171/RMI/1344

[40] I.D. SHKREDOV, *Noncommutative methods in Additive Combinatorics and Number Theory,* Uspekhi Mat. Nauk, 76:6 (462) (2021): 119–180.

[41] I.D. SHKREDOV, *On a girth-free variant of the Bourgain–Gamburd machine,* arXiv:2111.05751 (2021).

[42] T. TAO, V. VU, *Additive combinatorics,* Cambridge University Press 2006.

[43] J. TITS, *Free subgroups in linear groups,* J. Algebra 20, 250– 270 (1972).

[44] S. K. ZAREMBA, *Good lattice points, discrepancy, and numerical integration,* Ann. Mat. Pura Appl. (4), 73:293–317, 1966.

[45] S. K. ZAREMBA, *La méthode des "bons treillis" pour le calcul des intégrales multiples,* Academic Press, New York, 1972.