DISTRIBUTION MODULO ONE AND RATNER'S THEOREM

JENS MARKLOF

Contents

1. Introduction 1
2. Randomness of point sequences mod 1 2
2.1. Distribution of gaps 4
2.2. Independent random variables 6
3. $m\alpha$ mod one 7
3.1. Geometry of $\Gamma\backslash G$ 9
3.2. Dynamics on $\Gamma\backslash G$ 10
3.3. Mixing and uniform distribution 12
4. $\sqrt{m\alpha}$ mod one 14
4.1. The case $\alpha = 1$ 15
4.2. Some heuristics in the case $\alpha = \sqrt{2}$ 16
5. Ratner's theorem 19
5.1. Limit distributions of translates 19
5.2. Equidistribution, unbounded test functions and diophantine conditions 20
References 23

1. Introduction

Measure rigidity is a branch of ergodic theory that has recently contributed to the solution of some fundamental problems in number theory and mathematical physics. Examples are proofs of quantitative versions of the Oppenheim conjecture [3], related questions on the spacings between the values of quadratic forms [4, 8, 9], a proof of quantum unique ergodicity for certain classes of hyperbolic surfaces [6], and an approach to the Littlewood conjecture on the nonexistence of multiplicatively badly approximable numbers [1].

In these lectures we discuss a few simple applications of one of the central results in measure rigidity: Ratner's theorem. We shall investigate the statistical properties of certain number theoretic sequences, specifically the fractional parts of $m\alpha$, $m = 1, 2, 3, \ldots$, (a classical, well understood problem) and of $\sqrt{m\alpha}$ (as recently studied by Elkies and McMullen [2]). By exploiting equidistribution results on a certain homogeneous space $\Gamma\backslash G$, we will show that the statistical properties of these sequences can exhibit significant deviations from those of independent random variables. The "randomness" of other, more generic sequences such as $m^2\alpha$ and $2^m\alpha$ mod 1 has been studied extensively.

Date: J. Marklof, Distribution modulo one and Ratner's theorem, in: Equidistribution in Number Theory, An Introduction, eds. A. Granville and Z. Rudnick, Springer 2007, pp. 217-244.

We refer the interested reader to the review [10], and recommend the papers [12, 13] as a first read.

These notes are based on lectures presented at the Institute Henri Poincaré Paris, June 2005, and at the summer school 'Equidistribution in number theory', CRM Montréal, July 2005. The author gratefully acknowledges support by an EPSRC Advanced Research Fellowship.

2. Randomness of point sequences mod 1

Consider an infinite triangular array of numbers on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (which we represent as the unit interval $[0,1)$ with its endpoints identified),

$$\begin{array}{lllll} \xi_{11} & & & & \\ \xi_{21} & \xi_{22} & & & \\ \vdots & \vdots & \ddots & & \\ \xi_{N1} & \xi_{N2} & \dots & \xi_{NN} & \\ \vdots & \vdots & & & \ddots \end{array} \tag{2.1}$$

We assume that each row is ordered, i.e., $\xi_{Nj} \le \xi_{N(j+1)}$, and are interested in quantifying statistical properties of the Nth row as $N \to \infty$. To simplify notation we will from now on drop the index N, and simply write ξ_j instead of ξ_{Nj}.

As we shall see later, many interesting statistical properties of a sequence on $\mathbb{T}$ can be derived from the knowledge of the number of elements in small subintervals of $\mathbb{T}$. Let χ denote the characteristic function of the interval $[-\frac{1}{2}, \frac{1}{2}) \subset \mathbb{R}$. That is, $\chi(x) = 1$ if $-\frac{1}{2} \le x < \frac{1}{2}$ and $= 0$ otherwise. The characteristic function of the interval $[x_0 - \frac{\ell}{2}, x_0 + \frac{\ell}{2}) + \mathbb{Z} \subset \mathbb{T}$ ($\ell \le 1$) can be represented as

$$\chi_\ell(x) = \sum_{n\in\mathbb{Z}} \chi\left(\frac{x - x_0 + n}{\ell}\right). \tag{2.2}$$

The sum over n makes sure χ_ℓ is periodic. The number of elements ξ_j in the interval are therefore

$$S_N(\ell) = \sum_{j=1}^{N} \chi_\ell(\xi_j). \tag{2.3}$$

We will always assume that the rows in our triangular array become uniformly distributed mod one. This means that for every x_0, ℓ,

$$\lim_{N\to\infty} \frac{1}{N} S_N(\ell) = \ell, \tag{2.4}$$

i.e., the proportion of elements in any given interval is asymptotic to the interval length ℓ.[1]

[1] If a sequence $\{\xi_j\}$ fails to be uniformly distributed but still has a resonable limiting density ρ, we may rescale the ξ_j to obtain a uniformly distributed sequence. This is done as follows. Suppose for every x_0, ℓ $\lim_{N\to\infty} \frac{1}{N} S_N(\ell) = \int_{x_0-\ell/2}^{x_0+\ell/2} \rho(x)dx$, where the integrated density $N(x) = \int_0^x \rho(x')dx'$ is continuous and strictly increasing. We rescale the sequence $\{\xi_j\}$ by setting $\tilde{\xi}_j := N(\xi_j)$. Note that $N(\xi_j) \in [0,1)$ for $\xi_j \in [0,1)$. The new sequence $\{\tilde{\xi}_j\}$ is indeed uniformly distributed modulo one (exercise).

The aim is now to characterize the different degrees of "randomness" of the deterministic sequence $\{\xi_j\}$ in terms of their distribution in very small intervals with random center x_0. A convenient length scale is the average spacing between elements, which is $1/N$. We set

$$L = N\ell. \tag{2.5}$$

We assume x_0 is a random variable uniformly distributed on $\mathbb{T}$ with respect to Lebesgue measure dx_0. We will denote expectation values by

$$\langle \ldots \rangle = \int_0^1 \ldots dx_0. \tag{2.6}$$

It is easy to work out the expectation value for the number of elements in a random interval of size ℓ,

$$\langle S_N(\ell) \rangle = L. \tag{2.7}$$

The variance is much less trivial. Let us begin by deriving a convenient representation in terms of the pair correlation density. We have for the mean square (the "number variance")

$$\Sigma_N^2(L) := \langle [S_N(\ell) - L]^2 \rangle = \langle S_N(\ell)^2 \rangle - L^2 \tag{2.8}$$

and

$$\begin{aligned} \langle S_N(\ell)^2 \rangle &= \sum_{i,j=1}^N \sum_{m,n\in\mathbb{Z}} \int_0^1 \chi\left(\frac{\xi_i - x_0 + m}{\ell}\right) \chi\left(\frac{\xi_j - x_0 + n}{\ell}\right) dx_0 \\ &= \sum_{i,j=1}^N \sum_{m\in\mathbb{Z}} \int_{\mathbb{R}} \chi\left(\frac{\xi_i - x_0 + m}{\ell}\right) \chi\left(\frac{\xi_j - x_0}{\ell}\right) dx_0 \\ &= \ell \sum_{i,j=1}^N \sum_{m\in\mathbb{Z}} \Delta\left(\frac{\xi_i - \xi_j + m}{\ell}\right) \end{aligned} \tag{2.9}$$

where

$$\Delta(x) = \int_{\mathbb{R}} \chi(x - x_0)\chi(x_0) dx_0 = \max\{1 - |x|, 0\}. \tag{2.10}$$

Now the diagonal terms $i = j$ in the above double sum can be easily evaluated. We have

$$\ell \sum_{i=j=1}^N \sum_{m\in\mathbb{Z}} \Delta\left(\frac{m}{\ell}\right) = \ell\Delta(0) = \ell \tag{2.11}$$

for $\ell < 1$.

The *pair correlation function* (also called *two-point correlation function*) for the sequence $\{\xi_j\}$ is defined by

$$R_N^2(L, \psi) = \frac{1}{N} \sum_{i\neq j=1}^N \sum_{m\in\mathbb{Z}} \psi\left(\frac{\xi_i - \xi_j + m}{\ell}\right), \tag{2.12}$$

where ψ is taken from a class of sufficiently nice test functions (e.g. continuous with compact support such as Δ). With the above calculation we therefore have the identity

$$\Sigma_N^2(L) = L - L^2 + LR_N^2(L,\Delta). \tag{2.13}$$

This says that the asymptotic analysis of the pair correlation density will give us information on the number variance.

Note that by the Poisson summation formula

$$\sum_{m\in\mathbb{Z}} f(m) = \sum_{n\in\mathbb{Z}} \widehat{f}(n), \tag{2.14}$$

where

$$\widehat{f}(y) = \int_{\mathbb{R}} f(x)e(xy)dy, \qquad e(x) := \exp 2\pi \mathrm{i} x, \tag{2.15}$$

we have

$$R_N^2(L,\psi) = \frac{L}{N^2}\sum_{i\neq j=1}^{N}\sum_{n\in\mathbb{Z}} \widehat{\psi}\Big(\frac{Ln}{N}\Big) e\big(n(\xi_i-\xi_j)\big). \tag{2.16}$$

Here ψ can be any function with absolutely convergent Fourier series (e.g. Δ).

2.1. **Distribution of gaps.** A popular statistical measure is the distribution of gaps

$$s_j = N(\xi_{j+1}-\xi_j) \qquad (j=1,\dots,N,\ \xi_{N+1} := \xi_1+1) \tag{2.17}$$

between consecutive elements (recall the ξ_j form an ordered sequence on $\mathbb{T}$). We have multiplied the actual gap $\xi_{j+1}-\xi_j$ by N, which means we are measuring spacings in units of the average gap $1/N$.

The gap distribution of the sequence $\xi_1,\dots,\xi_N$ is defined as

$$P_N(s) = \frac{1}{N}\sum_{j=1}^{N}\delta(s-s_j) \tag{2.18}$$

where δ is a Dirac mass at the origin. The question we will investigate is whether $P_N(s)$ has a limiting distribution $P(s)$. That is, does there exist a probability density $P(s)$ such that for every bounded continuous function $g:\mathbb{R}\to\mathbb{R}$,

$$\lim_{N\to\infty}\int_0^\infty g(s)P_N(s)ds = \int_0^\infty g(s)P(s)ds. \tag{2.19}$$

The first question in convergence of probability measures is the problem of tightness.

Lemma 2.1. *The sequence of probability measures $\{P_N(s)ds\}$ is tight on $\mathbb{R}$. That is, for every $\epsilon>0$ there is a $K>0$ such that for all N*

$$\int_{|s|>K} P_N(s)ds < \epsilon. \tag{2.20}$$

Proof. We have

$$\begin{aligned}\int_{|s|>K} P_N(s)ds &= \frac{1}{N}\#\{j \le N : s_j \ge K\} \\ &\le \frac{1}{N}\sum_{j=1}^{N}\frac{s_j}{K}\chi_{[K,\infty)}(s_j) \le \frac{1}{N}\sum_{j=1}^{N}\frac{s_j}{K} \\ &= \frac{1}{K}\sum_{j=1}^{N}(\xi_{j+1}-\xi_j) = \frac{1}{K}.\end{aligned} \tag{2.21}$$

□

Denote by $E_N(k,L)$ the probability of finding k elements in the randomly shifted interval $[x_0, x_0+\frac{L}{N})$, i.e.,

$$E_N(k,L) := \text{meas}\left\{x_0 \in \mathbb{T} : S_N(\ell) = k\right\}. \tag{2.22}$$

The following theorem explains the relation between $P(s)$ and the probability $E(0,L)$.

Theorem 2.2. *Given a probability density $P(s)$, the following statements are equivalent.*

(i) $P_N(s) \to_w P(s)$.

(ii) $\lim_{N\to\infty} E_N(0,L) = E(0,L)$ *for all* $L>0$, *where* $E(0,L)$ *is defined by*

$$\frac{d^2E(0,L)}{dL^2} = P(L), \qquad \lim_{L\to 0} E(0,L) = 1, \qquad \lim_{L\to\infty}\frac{dE(0,L)}{dL} = 0. \tag{2.23}$$

Proof. We have

$$\begin{aligned}E_N(0,L) &= \text{meas}\left\{x_0 \in \mathbb{T}^2 : \#\{j : \xi_j \in [x_0, x_0+\tfrac{L}{N}) + \mathbb{Z}\} = 0\right\} \\ &= \sum_{j=1}^{N}\text{meas}\left\{x_0 \in [\xi_j,\xi_{j+1}) : \#\{j : \xi_j \in [x_0, x_0+\tfrac{L}{N}) + \mathbb{Z}\} = 0\right\} \\ &= \sum_{j=1}^{N}\left(\xi_{j+1}-\xi_j-\frac{L}{N}\right)\chi_{[L,\infty)}(N(\xi_{j+1}-\xi_j)) \\ &= \sum_{j=1}^{N}(\xi_{j+1}-\xi_j) - \sum_{j=1}^{N}(\xi_{j+1}-\xi_j)\,\chi_{[0,L)}(N(\xi_{j+1}-\xi_j)) \\ &\quad - \frac{L}{N}\sum_{j=1}^{N}\chi_{[L,\infty)}(N(\xi_{j+1}-\xi_j)) \\ &= 1 - \frac{1}{N}\sum_{j=1}^{N} g(s_j)\end{aligned} \tag{2.24}$$

where

$$g(x) = \max\{0, x, L\} \tag{2.25}$$

is a bounded continuous function.

"(i)⇒(ii)." With the above choice of test function g, (i) implies

$$\lim_{N\to\infty} E_N(0,L) = F(L) := 1 - \int_0^L sP(s)ds - L\int_L^\infty P(s)ds. \tag{2.26}$$

Now

$$\frac{dF(L)}{dL} = -\int_L^\infty P(s)ds, \qquad \frac{d^2F(L)}{dL^2} = P(L), \tag{2.27}$$

and

$$\lim_{L\to 0} F(L) = 1, \qquad \lim_{L\to\infty}\frac{dF(L)}{dL} = 0. \tag{2.28}$$

"(ii)⇒(i)." Since the sequence of probability measures $P_N(s)$ is tight, it is relatively compact by the Helly-Prokhorov Theorem (also often called Helly's Theorem). That is, every subsequence of N contains a convergent subsequence N_i for which $P_{N_i}(s) \to_w P(s)$ as $i\to\infty$. This implies (recall the first part of the proof) that $E_{N_i}(0,L)\to E(0,L)$ for all $L>0$. Hence every convergent subsequence has the limit $E(0,L)$, and thus every subsequence convergences. □

2.2. **Independent random variables.** In order to understand which statistical behaviour we should expect for the deterministic sequences we will study later, let us assume the vector $\xi = (\xi_1,\dots,\xi_N)$ is a uniformly distributed random vector on $\mathbb{T}^N$ with respect to Lebesgue measure $d\xi = d\xi_1\cdots d\xi_N$. (This means the ξ_j are independent uniformly distributed random variables.) We can ignore the issue of ordering the ξ_j here because of the symmetry of the measure dx under permutation of coordinates. Expectation values and associated probabilities of a random variable $X = X(\xi)$ will be defined as

$$\mathbb{E}X = \int_{\mathbb{T}^N} X d\xi, \tag{2.29}$$

$$\mathrm{Prob}(X>R) = \mathrm{meas}\{\xi\in\mathbb{T}^N : X>R\}. \tag{2.30}$$

Theorem 2.3. *There is a constant $C>0$ such that, for all $\epsilon>0$, N, L,*

$$\mathrm{Prob}(|R_N^2(L,\psi) - L\widehat{\psi}(0)| > \epsilon) \le C\frac{L}{\epsilon^2 N}. \tag{2.31}$$

Proof. First of all, we have for the expectation (the $n=0$ term in (2.16))

$$\mathbb{E}R_N^2(L,\psi) = \frac{L(N-1)}{N}\widehat{\psi}(0) = L\widehat{\psi}(0)(1+O(N^{-1})). \tag{2.32}$$

Secondly, for the variance of $R_N^2(L,\psi)$,

$$\begin{aligned}\mathbb{E}|R_N^2(L,\psi) - \mathbb{E}R_N^2(L,\psi)|^2 \\ = \frac{L^2}{N^4}\sum_{\substack{i\neq j\\ i'\neq j'}}\sum_{n,n'\in\mathbb{Z}} \widehat{\psi}\Big(\frac{Ln}{N}\Big)\widehat{\psi}\Big(\frac{Ln'}{N}\Big)\mathbb{E}\big[e\big(n(\xi_i-\xi_j) - n'(\xi_{i'}-\xi_{j'})\big)\big].\end{aligned} \tag{2.33}$$

Now

$$\mathbb{E}\big[e\big(n(\xi_i-\xi_j)-n'(\xi_{i'}-\xi_{j'})\big)\big]=\begin{cases}1 & \text{if } n=n',\ i=i',\ j=j'\\ & \text{or if } n=-n',\ i=j',\ j=i'\\ 0 & \text{otherwise.}\end{cases} \tag{2.34}$$

This implies that

$$\mathbb{E}|R_N^2(L,\psi)-\mathbb{E}R_N^2(L,\psi)|^2=\frac{L^2}{N^4}O(N^3/L)=O\big(\frac{L}{N}\big). \tag{2.35}$$

□

The above theorem implies that for a "generic" choice of the triangular array (2.1), we have

$$R_N^2(L,\psi)=L\widehat{\psi}(0)+o(1) \tag{2.36}$$

in the limit $N\to\infty$, $\ell=L/N\to 0$. This implies for the variance

$$\Sigma_N^2(L)=L+o(L) \tag{2.37}$$

almost surely in the above limit.

Using standard techniques from probability theory, one can extend these results on the variance to the full distribution of a generic realization of the random sequence in a small randomly shifted interval. There are two scaling regimes.

Regime I (Central Limit Theorem): In the limit $L\to\infty$, $N\to\infty$, $\ell=L/N\to 0$ we have

$$\text{meas}\left\{x_0\in\mathbb{T}:\frac{S_N(\ell)-L}{\sqrt{\Sigma_N^2(L)}}>R\right\}\to\frac{1}{\sqrt{2\pi}}\int_R^\infty \mathrm{e}^{-t^2/2}dt \tag{2.38}$$

almost surely.

Regime II (Poisson Limit Theorem): For L fixed, $N\to\infty$, we have

$$E_N(k,L)\to\frac{L^k}{k!}\mathrm{e}^{-L}. \tag{2.39}$$

almost surely.

3. $m\alpha$ MOD ONE

We will now consider the statistical properties of the sequence given by the fractional parts of $m\alpha$, $m=1,2,3,\dots$ for some α. This problem was studied by Berry-Tabor, Pandey et al., Bleher, Mazel-Sinai and Greenman using continued fractions (see [7] for detailed references). In particular, it is a classical result that there are at most three distinct values for the gaps occurring in $m\alpha$ mod 1 which already indicates a rather non-generic behavior of the sequence, see e.g. [15].

Here we will use the approach introduced in [7] that has the advantage of avoiding continued fractions and thus allowing higher-dimensional generalizations, such as the analysis of the distribution of linear forms modulo one. It is also very close to the work of Elkies and McMullen on $\sqrt{m}$ mod 1 which we will discuss in the next section.

We will be interested in the regime where $L = N\ell$ is fixed (Poisson scaling regime). The number (2.3) of elements in an interval of size ℓ and centered at x_0 is then

$$\begin{aligned} S_N(\ell) &= \sum_{m=1}^{N} \sum_{n\in\mathbb{Z}} \chi\Big(\frac{N}{L}(m\alpha + n - x_0)\Big) \\ &= \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\Big(\frac{m}{N}\Big) \chi_{[-L/2,L/2]}\big(N(m\alpha + n - x_0)\big) \\ &= \sum_{(m,n)\in\mathbb{Z}^2} \psi\Bigg((m, n - x_0)\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}\begin{pmatrix} N^{-1} & 0 \\ 0 & N \end{pmatrix}\Bigg) \end{aligned} \tag{3.1}$$

where χ_I denotes the characteristic function of the interval $I \subset \mathbb{R}$ and

$$\psi(x, y) = \chi_{(0,1]}(x)\chi_{[-L/2,L/2]}(y) \tag{3.2}$$

is the characteristic function of a rectangle.

Define the Lie Group G by the semidirect product $\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ with multiplication law

$$(M, \xi)(M', \xi') = (MM', \xi M' + \xi'), \tag{3.3}$$

where $\xi, \xi' \in \mathbb{R}^2$ are viewed as row vectors. This group has the matrix representation

$$(M, \xi) \mapsto \begin{pmatrix} M & 0 \\ \xi & 1 \end{pmatrix} \in \mathrm{SL}(3,\mathbb{R}). \tag{3.4}$$

The function

$$F(M, \xi) = \sum_{m\in\mathbb{Z}^2} \psi(mM + \xi) \tag{3.5}$$

defines a function on G. Note that, with ψ as above, the sum in (3.5) is always finite, and hence F is a piecewise constant function. Furthermore,

$$S_N(\ell) = F(M, \xi) \tag{3.6}$$

for the special choice

$$M = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}\begin{pmatrix} N^{-1} & 0 \\ 0 & N \end{pmatrix}, \qquad \xi = (0, -x_0)M. \tag{3.7}$$

The crucial observation is now that F is left-invariant under the discrete subgroup $\Gamma = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$, and hence F may be viewed as a piecewise constant function on the homogeneous space $\Gamma\backslash G$.

Proposition 3.1. *$F(\hat\gamma g) = F(g)$ for all $\hat\gamma \in \Gamma$.*

Proof. We have the decomposition

$$\hat\gamma = (\gamma, n) = (\gamma, 0)(1, n) \tag{3.8}$$

for some $\gamma \in \mathrm{SL}(2,\mathbb{Z})$, $n \in \mathbb{Z}^2$. It is therefore sufficient to check the statement for elements of the form $(\gamma, 0)$ and $(1, n)$ separately. We have

$$\begin{aligned} F((1,n)(M,\xi)) &= F(M, nM+\xi) \\ &= \sum_{m\in\mathbb{Z}^2} \psi((m+n)M+\xi) \\ &= \sum_{m\in\mathbb{Z}^2} \psi(mM+\xi) \\ &= F(M,\xi) \end{aligned} \tag{3.9}$$

which proves one case, and

$$\begin{aligned} F((\gamma,0)(M,\xi)) &= F(\gamma M, \xi) \\ &= \sum_{m\in\mathbb{Z}^2} \psi(m\gamma M+\xi) \\ &= \sum_{m\in\mathbb{Z}^2} \psi(mM+\xi) \\ &= F(M,\xi) \end{aligned} \tag{3.10}$$

since $\gamma\mathbb{Z}^2 = \mathbb{Z}^2$. □

Alternatively, F may be expressed as

$$F(g) = \sum_{\hat{\gamma}\in\pi(\Gamma)\backslash\Gamma} \psi(\pi(\hat{\gamma}g)) \tag{3.11}$$

with the projection

$$\begin{array}{rccc} \pi: & G & \to & \mathbb{R}^2 \\ & (M,\xi) & \mapsto & \xi. \end{array} \tag{3.12}$$

From (3.11) the invariance under Γ is directly evident.

3.1. **Geometry of $\Gamma\backslash G$.** The aim is to find a good coordinate system for G. Since parametrizing $\mathbb{R}^2$ is obvious, we need to mainly worry about $\mathrm{SL}(2,\mathbb{R})$. The Iwasawa decomposition of an element $M \in \mathrm{SL}(2,\mathbb{R})$ is

$$M = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v^{1/2} & 0 \\ 0 & v^{-1/2} \end{pmatrix} \begin{pmatrix} \cos(\phi/2) & \sin(\phi/2) \\ -\sin(\phi/2) & \cos(\phi/2) \end{pmatrix} \tag{3.13}$$

where $\tau = u + \mathrm{i}v \in \mathbb{H} := \{\tau \in \mathbb{C} : \operatorname{Im}\tau > 0\}$ (the complex upper halfplane) and $\phi \in [0, 4\pi)$. This yields a 1-1 map $\mathrm{SL}(2,\mathbb{R}) \to \mathbb{H} \times [0, 4\pi)$. Left-multiplication becomes now an action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H} \times [0, 4\pi)$ given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \left(\frac{a\tau+b}{c\tau+d}, \phi - 2\arg(c\tau+d) \right) \tag{3.14}$$

(this can be checked by a straightforward calculation). The fractional linear transformation of the τ component defines an (orientation preserving) isometry with respect to the Riemannian line element

$$ds^2 = \frac{du^2+dv^2}{v^2} \tag{3.15}$$

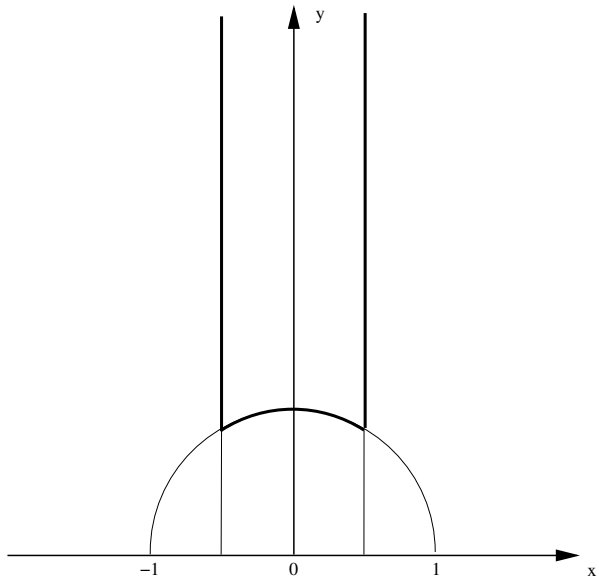

FIGURE 1. Fundamental domain of the modular group $\mathrm{SL}(2,\mathbb{Z})$ in the complex upper half plane.

and the transformation property of ϕ is identical to the direction of a tangent vector at $\tau \in \mathbb{H}$. Thus the group $\mathrm{PSL}(2,\mathbb{R}) := \mathrm{SL}(2,\mathbb{R})/\{\pm 1\} \simeq \mathbb{H} \times [0,2\pi)$ can be identified with the unit tangent bundle $\mathrm{T}^1\,\mathbb{H}$ of $\mathbb{H}$. Similarly, $\mathrm{SL}(2,\mathbb{Z})\backslash\,\mathrm{SL}(2,\mathbb{R}) \simeq \mathrm{PSL}(2,\mathbb{Z})\backslash\,\mathrm{PSL}(2,\mathbb{R})$ can be identified with the unit tangent bundle of the modular surface $\mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$. A fundamental domain $\mathcal{F}$ for the action of $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{H}$ is shown in Figure 1. We have

$$\begin{aligned}\mathcal{F} = \{\tau \in \mathbb{H} : |\tau| > 1,\ |\operatorname{Re}\tau| < 1/2\}\\ \cup \{\tau \in \mathbb{H} : |\tau| \geq 1,\ \operatorname{Re}\tau = -1/2\}\\ \cup \{\tau \in \mathbb{H} : |\tau| = 1,\ -1/2 \leq \operatorname{Re}\tau \leq 0\}.\end{aligned} \tag{3.16}$$

Note that the modular surface is not compact, there is one cusp at $\mathrm{i}\infty$. It has however finite measure with respect to the Riemannian volume $v^{-2}du\,dv$.

In order to understand the geometry of all of $\Gamma\backslash G$, write

$$g = (1,\xi)(M,0) \tag{3.17}$$

which gives a particular parametrization in terms of $\mathbb{R}^2$ and $\mathrm{SL}(2,\mathbb{R})$. Since Γ contains the subgroup $1 \ltimes \mathbb{Z}^2$, ξ can be parametrized by $\mathbb{T}^2 = \mathbb{Z}^2\backslash\mathbb{R}^2 \simeq [0,1)^2$. This concludes our analysis: we have found a 1-1 parametrization of G in terms of

$$\mathrm{T}^1(\mathrm{SL}(2,\mathbb{R})\backslash\mathbb{H}) \times \mathbb{T}^2 \simeq \mathcal{F} \times [0,2\pi) \times [0,1)^2. \tag{3.18}$$

That is, $\Gamma\backslash G$ is a (non-trivial) bundle over $\mathrm{T}^1(\mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H})$ with fibre $\mathbb{T}^2$.

3.2. **Dynamics on $\Gamma\backslash G$.** Consider the one-parameter subgroup $\Phi^{\mathbb{R}} := \{\Phi^t\}_{t\in\mathbb{R}}$ where

$$\Phi^t = \left(\begin{pmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2}\end{pmatrix}, 0\right). \tag{3.19}$$

$\Phi^{\mathbb{R}}$ defines a flow on $\Gamma\backslash G$ by right multiplication,

$$\Gamma g \mapsto \Gamma g \Phi^t. \tag{3.20}$$

The remarkable observation is that our object of interest, $S_N(\ell)$, is related to a function F on $\Gamma\backslash G$ evaluated along an orbit of this flow:

$$S_N(\ell) = F(g_0 \Phi^t) \tag{3.21}$$

with $t = 2\log N$ and initial condition

$$g_0 = \left(\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, (0, -x_0) \right). \tag{3.22}$$

Let us define

$$n_-(\alpha, y) = \left(\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, (0, y) \right). \tag{3.23}$$

The subgroup $H = \{n_-(\alpha, y)\}_{(\alpha,y)\in\mathbb{R}^2}$ is abelian and isomorphic to $\mathbb{R}^2$. Notice that

$$\Gamma \cap H = \{n_-(\alpha, y)\}_{(\alpha,y)\in\mathbb{Z}^2} \tag{3.24}$$

is a subgroup of H isomorphic to $\mathbb{Z}^2$. Therefore, for every fixed t, the set

$$\Gamma\backslash\Gamma H \Phi^t \tag{3.25}$$

describes a torus $\simeq \mathbb{T}^2$ embedded in $\Gamma\backslash G$; t parametrizes a continuous family of such tori.

We will now show that H parametrizes the unstable directions of the flow Φ^t. We employ the following parametrization of G. Write

$$g = n_-(\alpha, y)\Phi^s n_+(\beta, x), \tag{3.26}$$

where

$$n_+(\beta, x) = \left(\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}, (x, 0) \right). \tag{3.27}$$

We will write for short $g = (\alpha, y, s, \beta, x)$. The advantage of these coordinates is that the time evolution under Φ^t can be worked out very simply. We have the relation

$$(\alpha, y, s, \beta, x)\Phi^t = \Phi^t(\mathrm{e}^t\alpha, \mathrm{e}^{t/2}y, s, \mathrm{e}^{-t}\beta, \mathrm{e}^{-t/2}x). \tag{3.28}$$

Distances on $\Gamma\backslash G$ are measured by a left-G-invariant (since Γ acts on the left) Riemannian metric $d(g, g')$ on G. If $g = (\alpha, y, 0, 0, 0)$ and $g' = (\alpha', y', 0, 0, 0)$ are two initially close points, we have under the flow Φ^t (use the above formula and left-invariance of the metric)

$$\begin{aligned} d(g\Phi^t, g'\Phi^t) &= d((\mathrm{e}^t(\alpha - \alpha'), \mathrm{e}^{t/2}(y - y'), 0, 0, 0), (0, 0, 0, 0, 0)) \\ &\approx (\mathrm{e}^{2t}|\alpha - \alpha'|^2 + \mathrm{e}^t|y - y'|^2)^{1/2}. \end{aligned} \tag{3.29}$$

Hence (α, y) describe exponentially unstable directions of the flow, and by the same argument it is easy to see that (β, y) are the exponentially stable directions and s is of course the neutral flow direction. In particular we have the bound

$$d((\alpha, y, s, \beta, x)\Phi^t, (\alpha, y, 0, 0, 0)\Phi^t) = O(|s| + |\beta|\mathrm{e}^{-t} + |x|\mathrm{e}^{-t/2}) \tag{3.30}$$

for s, β, x bounded and $t > 0$. This follows directly from (3.28).

3.3. **Mixing and uniform distribution.** Recall that we are interested in the behaviour of the distribution of $S_N(\ell)$ for x_0 random and N large. At this point it will be convenient to also take α to be random, say, uniformly distributed in the interval $[a,b]$. We will see later that for fixed α there is no universal limiting distribution (an observation that is well known and related to the three gap theorem [15]).

We will use equidistribution on $\Gamma\backslash G$ to prove the following limit theorem, which asserts a limiting distribution different from Poissonian, cf. (2.39).

We will use the notation $\overline{g} = \Gamma g$.

Theorem 3.2. *For any $L > 0$,*

$$\lim_{N\to\infty} \frac{1}{b-a} \operatorname{meas}\{(\alpha, x_0) \in [a,b] \times [0,1] : S_N(\ell) = k\} = E(k,L), \tag{3.31}$$

where

$$E(k,L) = \frac{1}{\mu(\Gamma\backslash G)} \mu(\overline{g} \in \Gamma\backslash G : F(g) = k). \tag{3.32}$$

Here F is the function defined in (3.5), and μ the Haar measure on G. An explicit formula for $d\mu$ in the Iwasawa coordinates is

$$d\mu = \frac{du\, dv}{v^2}\, d\phi\, dx\, dy. \tag{3.33}$$

It is possible to derive more explicit formulas for $E(k,L)$ from (3.32), but this requires some involved calculations which we will not pursue her. See [16], Section 8, for details.

The key to the proof is the following equidistribution theorem.

Theorem 3.3. *For any bounded, piecewise continuous*[2] $f : \Gamma\backslash G \to \mathbb{R}$

$$\lim_{t\to\infty} \frac{1}{b-a} \int_a^b \int_0^1 f(n_-(\alpha,y)\Phi^t) d\alpha\, dy = \frac{1}{\mu(\Gamma\backslash G)} \int_{\Gamma\backslash G} f d\mu. \tag{3.34}$$

Proof. It is well known that the flow Φ^t is mixing,[3] that is for any $f, h \in \mathrm{L}^2(\Gamma\backslash G)$

$$\lim_{t\to\infty} \int_{\Gamma\backslash G} f(g\Phi^t) h(g) d\mu = \frac{1}{\mu(\Gamma\backslash G)} \int_{\Gamma\backslash G} f d\mu \int_{\Gamma\backslash G} h d\mu. \tag{3.35}$$

Take f to be continuous an of compact support, and h the characteristic function of the set

$$S_\epsilon = \Gamma\{(\alpha, y, s, \beta, y) : \alpha \in [a,b],\ y \in [0,1], s, \beta, x \in [-\epsilon, \epsilon]\}, \tag{3.36}$$

which forms an ϵ-neighbourhood of the embedded closed torus S_0. By the uniform continuity of f and (3.30), given any $\delta > 0$ there is an $\epsilon > 0$ such that

$$\sup_{\substack{g\in S_\epsilon \\ t>0}} |f(g\Phi^t) - f(n_-(\alpha,y)\Phi^t)| < \delta. \tag{3.37}$$

Haar measure in the local coordinates (α, y, s, β, y) reads (up to normalization)

$$d\mu = \mathrm{e}^{3s/2} ds\, d\alpha\, d\beta\, dx\, dy. \tag{3.38}$$

[2] i.e. the discontinuities are contained in a set of μ measure zero.

[3] This is guaranteed by a general theorem by Moore for semisimple Lie groups, which can be extended to the non-semisimple G considered here, cf. [5].

We conclude that

$$\liminf_{t\to\infty} \frac{1}{b-a}\int_a^b \int_0^1 f(n_-(\alpha,y)\Phi^t)d\alpha\, dy = \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu + O(\delta) \tag{3.39}$$

and

$$\limsup_{t\to\infty} \frac{1}{b-a}\int_a^b \int_0^1 f(n_-(\alpha,y)\Phi^t)d\alpha\, dy = \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu + O(\delta), \tag{3.40}$$

where the implied constants are independent of ϵ. This works for any $\delta > 0$, and hence the limit must exist and equal $\frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu$.

To extend the statement of the theorem to bounded continuous functions, we observe that it holds (trivially) for constant f, and therefore also for continuous functions f that are constant outside some compact set.

Let f be a bounded piecewise continuous function. Given any $\epsilon > 0$ we can find continuous functions $f_\pm$, constant outside some constant set, such that

$$f_- \le f \le f_+ \tag{3.41}$$

and

$$\frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} (f_+ - f_-)d\mu < \epsilon. \tag{3.42}$$

This implies

(3.43)

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{b-a}\int_a^b \int_0^1 f(n_-(\alpha,y)\Phi^t)d\alpha\, dy &\ge \liminf_{t\to\infty} \frac{1}{b-a}\int_a^b \int_0^1 f_-(n_-(\alpha,y)\Phi^t)d\alpha\, dy \\ &= \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f_- d\mu \\ &> \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu - 2\epsilon.\end{aligned}$$

The analogous argument shows

$$\limsup_{t\to\infty} \frac{1}{b-a}\int_a^b \int_0^1 f(n_-(\alpha,y)\Phi^t)d\alpha\, dy < \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu + 2\epsilon. \tag{3.44}$$

Taking $\epsilon > 0$ arbitrarily small proves the theorem. □

Remark 3.1. An alternative proof of Theorem 3.3 follows from Ratner's theorem, since the subgroup $\{n_-(\alpha,y)\}_{\alpha,y\in\mathbb{R}}$ is generated by unipotent elements. We will get back to this later.

Proof of Theorem 3.2. Apply Theorem 3.3 to the characteristic function of the set of $\overline{g} \in \Gamma\backslash G$ for which $F(g) = k$ (to make sure the characteristic function is piecewise continuous, check that the set has a boundary of μ measure zero). □

Remark 3.2. As we had mentioned earlier, there is no limiting distribution as in Theorem 3.2 if α is fixed, since there is no analog of the equidistribution result, Theorem 3.3.

One can show, however, that if α is irrational we have for any continuous, compactly supported function

$$\int_{\mathbb{T}} f(n_-(\alpha,y)\Phi^t)dy = \overline{f}\left(\begin{pmatrix}1 & \alpha\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathrm{e}^{-t/2} & 0\\ 0 & \mathrm{e}^{t/2}\end{pmatrix}\right) + o(1) \qquad (t\to\infty) \tag{3.45}$$

where $\overline{f}$ is a (non-constant!) continuous, compactly supported function on $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$ defined by

$$\overline{f}(M) = \int_{\mathbb{T}^2} f((1,\xi)(M,0))d\xi. \tag{3.46}$$

Remark 3.3. If one however fixes $y=-x_0\notin\mathbb{Q}$ and keeps α random, Ratner's Theorem implies the following equidistribution result. For any bounded piecewise continuous $f:\Gamma\backslash G\to\mathbb{R}$

$$\lim_{t\to\infty}\int_{\mathbb{T}} f(n_-(\alpha,y)\Phi^t)d\alpha = \frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu. \tag{3.47}$$

Hence the limiting distribution is universal (i.e. independent of y as long as y is irrational) and the same as for random y. Thus the probability of finding k points in the interval $x_0-\ell/2, x_0+\ell/2)$ with *fixed* center $x_0\notin\mathbb{Q}$ has the limiting distribution

$$\lim_{N\to\infty}\mathrm{meas}\{\alpha\in\mathbb{T}^2 : S_N(\ell)=k\} = E(k,L), \tag{3.48}$$

the same as for *random* center. We will prove (3.3) in Section 5.

4. $\sqrt{m\alpha}$ mod one

The problem of the statistics of $\sqrt{m\alpha}$ mod 1 has been understood by Elkies and McMullen [2] in the case $\alpha=1$ (and in principle also for all other rational α). The uniform distribution of $\sqrt{m\alpha}$ mod 1 may be shown by using the fact that $\sqrt{n+m}-\sqrt{n}\to 0$ for $n\to\infty$, m fixed (we leave this as an exercise). As in the last section, the key idea is the reduce the problem to equidistribution on a homogeneous space. Lucky for us, this homogeneous space will turn out to be $\Gamma\backslash G$ with the same G, Γ as encountered earlier.

We are as in the previous section interested in the "Poisson scaling limit", i.e. L is fixed. Now (we swap m and n in our notation)

$$S_N(\ell) = \sum_{n=1}^{N}\sum_{m\in\mathbb{Z}}\chi\left(\frac{N}{L}(\sqrt{n\alpha}-x_0+m)\right). \tag{4.1}$$

The condition imposed on the summation can be re-written as

$$\left(x_0-m-\frac{L}{2N}\right)^2 \le n\alpha < \left(x_0-m+\frac{L}{2N}\right)^2 \tag{4.2}$$

which amounts to

$$-\frac{L}{N}(x_0-m) \le n\alpha-(x_0-m)^2-\left(\frac{L}{2N}\right)^2 < \frac{L}{N}(x_0-m). \tag{4.3}$$

Notice also that

$$|\sqrt{n\alpha} - (x_0 - m)| \leq \frac{L}{2N}. \tag{4.4}$$

This yields

$$S_N(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{x_0 - m + O(L/2N)}{\sqrt{N\alpha}}\right) \chi_{[-L,L)}\left(\frac{N^{1/2}[n\alpha - (x_0-m)^2 - (L/2N)^2]}{N^{-1/2}(x_0-m)}\right). \tag{4.5}$$

A more convenient object would be

$$\widetilde{S}_{N,\epsilon,\delta}(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(-\epsilon,1+\epsilon]}\left(\frac{x_0-m}{\sqrt{N\alpha}}\right) \chi_{[-L,L)}\left(\frac{N^{1/2}[n\alpha - (x_0-m)^2] + \delta}{N^{-1/2}(x_0-m)}\right). \tag{4.6}$$

For the right choices of ϵ (positive/negative) we obtain upper/lower bounds for $S_N(\ell)$ which would eventually allow us to infer the limiting distribution of $S_N(\ell)$ from $\widetilde{S}_{N,\epsilon,\delta}(\ell)$ by taking $\delta \to 0$, $\epsilon \to \pm 0$. We will ignore this technical point here and simply take

$$S_N(\ell) \approx \widetilde{S}_{N,0,0}(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{x_0-m}{\sqrt{N\alpha}}\right) \chi_{[-L,L)}\left(\frac{N^{1/2}[n\alpha - (x_0-m)^2]}{N^{-1/2}(x_0-m)}\right). \tag{4.7}$$

The manipulations we will now perform on the r.h.s. of (4.7) can be adapted step by step for more general values of $\delta, \epsilon \neq 0$ (recommended exercise). We will use the shorthand $\widetilde{S}_N(\ell) := \widetilde{S}_{N,0,0}(\ell)$ in the following.

4.1. **The case $\alpha = 1$.** We have, after substituting $(m,n) \to (-m,-n)$,

$$\widetilde{S}_N(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{x_0+m}{\sqrt{N}}\right) \chi_{(-L,L]}\left(\frac{N^{1/2}[n + (x_0+m)^2]}{N^{-1/2}(x_0+m)}\right), \tag{4.8}$$

an thus, after substituting $n \mapsto n + m^2$ in the sum over n,

$$\widetilde{S}_N(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{x_0+m}{\sqrt{N}}\right) \chi_{(-L,L]}\left(\frac{N^{1/2}(n + x_0^2 + 2mx_0)}{N^{-1/2}(x_0+m)}\right), \tag{4.9}$$

We will now show that, in analogy with the previous section, we can find a function $F : \Gamma\backslash G \to \mathbb{R}$ of the form

$$F(M,\xi) = \sum_{m\in\mathbb{Z}^2} \psi(mM + \xi) \tag{4.10}$$

so that

$$\widetilde{S}_N(\ell) = F(g) \tag{4.11}$$

for a suitable choice of $g \in G$ and a piecewise continuous $\psi : \mathbb{R}^2 \to \mathbb{R}$ with compact support. To this end define

$$\psi(x,y) = \chi_{(0,1]}(x)\chi_{(-L,L]}\left(\frac{y}{x}\right) \tag{4.12}$$

(which indeed has compact support: it is the characteristic function of a triangle). Now consider the one parameter subgroup $\{n_1(x)\}_{x\in\mathbb{R}}$ with

$$n_1(x) = \left(\begin{pmatrix} 1 & 2x \\ 0 & 1 \end{pmatrix}, (x, x^2)\right) \tag{4.13}$$

(check that this indeed yields a one parameter group). Then the choice (set $t = \log N$, $x = x_0$)

$$\begin{aligned} (M, \xi) &= n_1(x_0)\Phi^t \\ &= \left(\begin{pmatrix} N^{-1/2} & 2x_0 N^{1/2} \\ 0 & N^{1/2} \end{pmatrix}, (N^{-1/2}x_0, N^{1/2}x_0^2)\right) \end{aligned} \tag{4.14}$$

yields

$$(m, n)M + \xi = \big(N^{-1/2}(x_0 + m), N^{1/2}(2mx_0 + n + x_0^2)\big). \tag{4.15}$$

Using this result in the definition (4.10) then confirms the desired (4.11).

We now follow the same steps as in the previous Section 3 to derive the limiting distribution for $S_N(\ell)$ from equidistribution on $\Gamma\backslash G$. We first state the limit theorem.

Theorem 4.1. *For any $L > 0$,*

$$\lim_{N\to\infty} \operatorname{meas}\{x_0 \in \mathbb{T} : S_N(\ell) = k\} = E(k, L), \tag{4.16}$$

where

$$E(k, L) = \frac{1}{\mu(\Gamma\backslash G)}\mu(\overline{g} \in \Gamma\backslash G : F(g) = k), \tag{4.17}$$

with F as defined in (4.10).

An explicit formula for $E(0, L)$ and the corresponding gap distribution $P(s)$ (recall Theorem 2.2) is worked out in [2].

The relevant equidistribution theorem needed to prove Theorem 4.1 is the following. Note that $\Gamma \cap \{n_1(x)\}_{x\in\mathbb{R}} = \{n_1(x)\}_{x\in\mathbb{Z}}$ and hence

$$\Gamma\{n_1(x)\}_{x\in\mathbb{T}}\Phi^t \tag{4.18}$$

represents a family (parametrized by t) of closed orbits embedded in $\Gamma\backslash G$.

Theorem 4.2. *For any bounded piecewise continuous $f : \Gamma\backslash G \to \mathbb{R}$*

$$\lim_{t\to\infty} \int_{\mathbb{T}} f(n_1(x)\Phi^t)dx = \frac{1}{\mu(\Gamma\backslash G)} \int_{\Gamma\backslash G} f d\mu. \tag{4.19}$$

Since $n_1(x)$ generates a unipotent flow, Ratner's theorem can be employed. We will explain the general strategy of proof for statements of this type in Section 5.

4.2. **Some heuristics in the case $\alpha = \sqrt{2}$.** We return to generic α, such as $\alpha = \sqrt{2}$, and rewrite $\widetilde{S}_N(\ell)$ as

$$\widetilde{S}_N(\ell) = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{x_0 + m}{M}\right) \chi_{(-L,L]}\left(\frac{M[\alpha^{-1}(x_0 + m)^2 + n]}{M^{-1}(x_0 + m)}\right) \tag{4.20}$$

where $M = \sqrt{N\alpha}$. For $x_0 \in [0,1]$ we can ignore terms of the form x_0/M,

$$\widetilde{S}_N(\ell) \approx \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{m}{M}\right) \chi_{(-L,L]}\left(\frac{M[\alpha^{-1}(x_0+m)^2+n]}{M^{-1}m}\right). \tag{4.21}$$

Now note that for most values of m, we have $m/M \asymp 1$, and it is natural to assume that, for random x_0, the probability of finding k elements of the set

$$\{\alpha^{-1}(x_0+m)^2 : m = 1,\dots,M\} + \mathbb{Z} \tag{4.22}$$

in an interval of size $1/M$ around the origin is given by the Poisson distribution (we must assume here that α is badly approximable by rationals, e.g. $\alpha = \sqrt{2}$ would be a good choice). Hence we may assert that the limiting distribution of $S_N(\ell)$ is the same as that of the random variable

$$X = \sum_{(m,n)\in\mathbb{Z}^2} \chi_{(0,1]}\left(\frac{m}{M}\right) \chi_{(-L,L]}\left(\frac{M(\eta_m+n)}{M^{-1}m}\right) \tag{4.23}$$

where η_m are independent uniformly distributed random variables on $[-1/2, 1/2)$. With this choice of interval the only contribution comes from the $n = 0$ term (assume $M \gg L$), so

$$X = \sum_{m=1}^{M} X_m \tag{4.24}$$

where

$$X_m = \chi_{(-L,L]}\left(\frac{M^2\eta_m}{m}\right) \tag{4.25}$$

is a sequence of independent random variables with kth moment

$$\mathbb{E}X_m^k = \int_{-1/2}^{1/2} \chi_{(-L,L]}\left(\frac{M^2\eta_m}{m}\right) d\eta_m = \frac{2Lm}{M^2}, \tag{4.26}$$

and hence

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}tX_m} - 1) = \frac{2Lm}{M^2}(\mathrm{e}^{\mathrm{i}t} - 1). \tag{4.27}$$

The characteristic function of the random variable X is therefore

$$\begin{aligned}
\mathbb{E}\mathrm{e}^{\mathrm{i}tX} &= \prod_{m=1}^{M}\left[1 + \frac{2Lm}{M^2}(\mathrm{e}^{\mathrm{i}t} - 1)\right] \\
&= \exp\left\{\sum_{m=1}^{M} \log\left[1 + \frac{2Lm}{M^2}(\mathrm{e}^{\mathrm{i}t} - 1)\right]\right\} \\
&= \exp\left\{\sum_{m=1}^{M}\left[\frac{2Lm}{M^2}(\mathrm{e}^{\mathrm{i}t} - 1) + O\left(\frac{m^2}{M^4}\right)\right]\right\} \\
&= \exp\left[L(\mathrm{e}^{\mathrm{i}t} - 1) + O\left(\frac{1}{M}\right)\right].
\end{aligned} \tag{4.28}$$

```
> alpha:=sqrt(2); N:=6001;
```

$$\alpha := \sqrt{2}$$
$$N := 6001$$

```
> L:=sort([seq(evalf[12](frac(sqrt(n*alpha))), n=1..N)]):
> alist:=seq(evalf[12](N*(L[i+1]-L[i])),i=1..N-1):
> data:=stats[transform,tallyinto['outliers']]([alist],[seq((i-1)*0.2..
> i*0.2,i=0..35)]):
> outliers;
```

[7.0547245915, 7.0674227075, 7.1105849000, 7.1693268887, 7.2093775627, 7.3219323187, 7.3381866273, 7.4195061783, 7.5000233956, 7.6451419780, 7.7497418084, 7.9388213164, 8.0221013941, 8.1512135092, 8.4582030656]

```
> data1:=stats[transform,scaleweight[1/nops([alist])]](data):
> g1:=stats[statplots,histogram](data1):
> g2:=plot(exp(-s), s=0..6):
> plots[display](g1,g2);
```

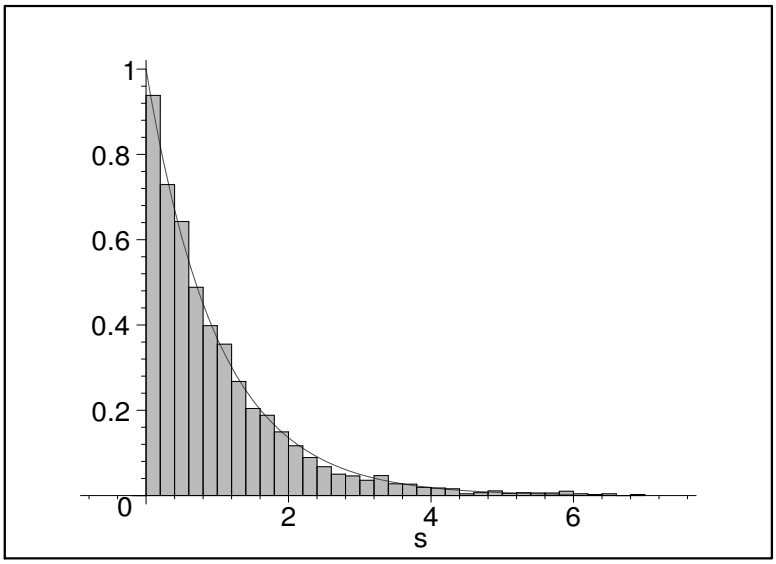

FIGURE 2. Maple worksheet for calculating the gap distribution of the fractional parts of $\sqrt{m\sqrt{2}}$, $m = 1, \ldots, 6001$.

The expression $\mathrm{e}^{L(\mathrm{e}^{\mathrm{i}t}-1)}$ is the characteristic function of the Poisson law

$$E(k, L) = \frac{L^k}{k!}\mathrm{e}^{-L}. \tag{4.29}$$

Hence this should be our prediction for the limiting distribution of $S_N(\ell)$, which in turn implies that we expect the exponential distribution for gaps in $\sqrt{m\alpha}$ mod 1. This is in good agreement with our Maple experiment, Figure 2.

5. Ratner's theorem

An excellent introduction to Ratner's theory is Dave W. Morris' recent textbook [11]. Let G be a Lie group (e.g. $\mathrm{SL}(2,\mathbb{R})\times\mathbb{R}^2$) and Γ be a discrete subgroup (e.g. $\mathrm{SL}(2,\mathbb{Z})\times\mathbb{Z}^2$). It is at this point not necessary to assume that Γ is a lattice in G, i.e., that $\Gamma\backslash G$ has finite volume with respect to Haar measure μ on G. Ratner's measure classification theorem gives a complete geometric description of all measures that are invariant and ergodic under the a unipotent one parameter subgroup U (or, more generally, invariant and ergodic under a subgroup generated by unipotent subgroups) acting on $\Gamma\backslash G$ by right multiplication. Examples of unipotent subgroups that appeared in the previous sections are $\{n_-(\alpha,0)\}_{\alpha\in\mathbb{R}}$, $\{n_-(0,y)\}_{\alpha\in\mathbb{R}}$ and $\{n_1(x)\}_{x\in\mathbb{R}}$.

Theorem 5.1 (Ratner's theorem)**.** *Let ν be an ergodic, U-invariant probability measure on $\Gamma\backslash G$. Then there is a closed, connected subgroup $H\subset G$, and a point $\overline{g}\in\Gamma\backslash G$ such that*

(1) *ν is H-invariant,*
(2) *ν is supported on the orbit $\overline{g}H$.*

Remark 5.1. Let $g\in G$ be a representative of the coset $\overline{g}=\Gamma g$, and define the subgroup $\Gamma_H=(g^{-1}\Gamma g)\cap H$. Then the orbit $\overline{g}H$ may be identified with the homogeneous space $\Gamma_H\backslash H$ and ν with the Haar measure on H. Furthermore one can deduce (since ν is a probability measure) that Γ_H is a lattice in H, i.e., $\nu(\overline{g}H)<\infty$, and that the orbit $\overline{g}H$ is closed in $\Gamma\backslash G$.

In simple words, measures ν invariant and ergodic under unipotent subgroups are supported on nice embedded closed subvarieties, of which there can be only countably many (modulo translations of course). We will now discuss two corollaries of Ratner's theorem that are relevant to the equidistribution theorems discussed earlier.

5.1. Limit distributions of translates.

The following is special case of Shah's extremely useful theorem, Theorem 1.4 in [14].

Theorem 5.2. *Suppose G contains a Lie subgroup H isomorphic to $\mathrm{SL}(2,\mathbb{R})$ (we denote the corresponding embedding by $\varphi:\mathrm{SL}(2,\mathbb{R})\to G$), such that the set $\Gamma\backslash\Gamma H$ is dense in $\Gamma\backslash G$. Then, for any bounded, piecewise continuous $f:\Gamma\backslash G\to\mathbb{R}$ and any piecewise continuous $h:\mathbb{R}\to\mathbb{R}$ with compact support*

$$\lim_{t\to\infty}\int_{\mathbb{R}} f\left(\varphi\left(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathrm{e}^{-t/2} & 0\\ 0 & \mathrm{e}^{t/2}\end{pmatrix}\right)\right)h(x)dx=\frac{1}{\mu(\Gamma\backslash G)}\int_{\Gamma\backslash G} f d\mu\int_{\mathbb{R}} h(x)dx \tag{5.1}$$

where μ is the Haar measure of G.

The general strategy of proof for statements of the above type is as follows.

(1) Normalize h such that it defines a probability density.
(2) Show that the sequences of probability measures ν_t defined by

$$\nu_t(f)=\int_{\mathbb{R}} f\left(\varphi\left(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathrm{e}^{-t/2} & 0\\ 0 & \mathrm{e}^{t/2}\end{pmatrix}\right)\right)h(x)dx \tag{5.2}$$

is tight. Then, by the Helly-Prokhorov theorem, it is relatively compact, i.e., every sequence of ν_t contains a convergent subsequence with weak limit ν, say.

(3) Show that ν is invariant under a unipotent subgroup U; in the present case,

$$U = \left\{ \varphi\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \right) \right\}_{x\in\mathbb{R}}. \tag{5.3}$$

(4) Use a density argument to rule out measures concentrated on subvarieties (exploit the assumption that $\Gamma\backslash\Gamma H$ is dense in $\Gamma\backslash G$).

As an application of Shah's theorem we give a proof of the statement in Remark 3.3, in fact a slightly more general version allowing for non-constant h. Recall that here $G = \mathrm{SL}(2,\mathbb{R}) \times \mathbb{R}^2$ and $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \times \mathbb{Z}^2$.

Corollary 5.3. *Let $y \notin \mathbb{Q}$. For any bounded piecewise continuous $f : \Gamma\backslash G \to \mathbb{R}$ and piecewise continuous $h : \mathbb{R} \to \mathbb{R}$ with compact support*

$$\lim_{t\to\infty} \int_{\mathbb{R}} f(n_-(\alpha,y)\Phi^t)h(\alpha)d\alpha = \frac{1}{\mu(\Gamma\backslash G)} \int_{\Gamma\backslash G} f d\mu \int_{\mathbb{R}} h(\alpha)d\alpha. \tag{5.4}$$

Proof. We define the embedding $\varphi : \mathrm{SL}(2,\mathbb{R}) \to G$ by

$$M \mapsto (1,(0,y))(M,0)(1,(0,y))^{-1}. \tag{5.5}$$

We need to show that

$$(\gamma,n)(1,(0,y))(M,0)(1,(0,y))^{-1} \tag{5.6}$$

is dense in G as γ, n, M vary over $\mathrm{SL}(2,\mathbb{Z})$, $\mathbb{Z}^2$, $\mathrm{SL}(2,\mathbb{R})$, respectively. It is obviously sufficient to show this for

$$(\gamma,n)(1,(0,y))(M,0) = (\gamma M, (n_1,(y+n_2))M), \tag{5.7}$$

and thus for $(M,(n_1,(y+n_2))\gamma^{-1}M)$. It is however easy to see, using the irrationality of y, that $(n_1,(y+n_2))\gamma^{-1}$ is dense in $\mathbb{R}^2$ (exercise). The completes the proof of the density.

Shah's theorem says now that

$$\lim_{t\to\infty} \int_{\mathbb{R}} \tilde{f}(n_-(\alpha,y)\Phi^t n_-(0,y)^{-1})h(\alpha)d\alpha = \frac{1}{\mu(\Gamma\backslash G)} \int_{\Gamma\backslash G} \tilde{f} d\mu \int_{\mathbb{R}} h(\alpha)d\alpha. \tag{5.8}$$

for all bounded, piecewise continuous $\tilde{f}$. Choosing the test function

$$\tilde{f}(g) = f(gn_-(0,y)) \tag{5.9}$$

which is left-Γ-invariant and bounded, piecewise continuous, if f is (as assumed). This yields (5.4). □

5.2. **Equidistribution, unbounded test functions and diophantine conditions.** In some applications of Ratner's theorem, e.g., in questions of value distribution of quadratic forms [3, 4, 8, 9], the test functions f in the equidistribution theorems are no longer bounded. Under such circumstances the convergence of the integral can only be assured by assuming certain diophantine conditions. Without going into the intricate details for general $\Gamma\backslash G$, we will illustrate this phenomenon in the distribution of $m\alpha$ on $\mathbb{T}$, which indeed may be viewed as a unipotent orbit on the homogeneous space $\mathbb{Z}\backslash\mathbb{R}$. As

mentioned earlier, it is well known that for $\alpha \notin \mathbb{Q}$ the sequence is uniformly distributed mod 1. That is, for any bounded continuous function $f : \mathbb{T} \to \mathbb{R}$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} f(m\alpha) = \int_{\mathbb{T}} f(x)dx. \tag{5.10}$$

Let us know formulate the analogous statement for test functions with a possible singularity at $x = 0$.

It is convenient to identify $\mathbb{T}$ with $[-1/2, 1/2)$. For any $\beta \geq 0$ we define the class $K_\beta(\mathbb{T})$ of functions continuous on $\mathbb{T} - \{0\}$, with the property that there is a constant $C > 0$ such that

$$|f(x)| \leq C|x|^{-\beta}, \qquad \text{for all } x \in [-1/2, 1/2). \tag{5.11}$$

We say $\alpha \in \mathbb{R}$ is *diophantine of type* κ if there exists a constant $c > 0$ such that

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^\kappa}$$

for all $p, q \in \mathbb{Z}$, $q > 0$. The smallest possible value of κ is $\kappa = 2$ (achieved for quadratic surds, e.g., $\alpha = \sqrt{2}$), and it is well known that for any $\kappa > 2$ there is a set of full Lebesgue measure of α that have type κ.

Theorem 5.4. *Let α be diophantine of type κ. Then, for any $f \in K_\beta(\mathbb{T})$ with $0 \leq \beta < 1/(\kappa - 1)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} f(m\alpha) = \int_{\mathbb{T}} f(x)dx. \tag{5.12}$$

Proof. We split $f = f_+ - f_-$ into positive and negative part, such that $f_\pm \geq 0$. Then $f_\pm \in K_\beta(\mathbb{T})$ and we may prove (5.12) for both $f_\pm$ separately. In the following we will thus assume that $f \geq 0$.

For any $\epsilon > 0$ let us define

$$f_\epsilon(x) = \begin{cases} f(x) & \text{if } |x| > \epsilon \\ \min\{f(x), f(\epsilon)\} & \text{if } |x| \leq \epsilon \end{cases} \tag{5.13}$$

and $g_\epsilon = f - f_\epsilon$. Note that $f_\epsilon \leq f$. By assumption,

$$g_\epsilon(x) \begin{cases} \leq C|x|^{-\beta} & \text{if } |x| \leq \epsilon \\ = 0 & \text{if } |x| \geq \epsilon. \end{cases} \tag{5.14}$$

The function f_ϵ is bounded continuous, and hence by uniform distribution

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} f_\epsilon(m\alpha) = \int_{\mathbb{T}} f_\epsilon(x)dx = \int_{\mathbb{T}} f(x)dx - O(\epsilon^{1-\beta}). \tag{5.15}$$

Since $f_\epsilon \leq f$, this implies the lower bound

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} f(m\alpha) \geq \int_{\mathbb{T}} f(x)dx - O(\epsilon^{1-\beta}). \tag{5.16}$$

As to the upper bound,

$$\frac{1}{N}\sum_{m=1}^{N} g_\epsilon(m\alpha) \le \frac{C}{N}\sum_{m=1}^{N} \frac{\chi_{(0,\epsilon]}(\|m\alpha\|)}{\|m\alpha\|^\beta} \tag{5.17}$$

where $\|\cdot\|$ denotes the distance to the nearest integer. Using the dyadic decomposition of the unit interval, we find

$$\begin{aligned}\frac{1}{N}\sum_{m=1}^{N} \frac{\chi_{(0,\epsilon]}(\|m\alpha\|)}{\|m\alpha\|^\beta} &= \frac{1}{N}\sum_{j=0}^{\infty}\sum_{m=1}^{N} \frac{\chi_{(\epsilon 2^{-(j+1)},\epsilon 2^{-j}]}(\|m\alpha\|)}{\|m\alpha\|^\beta} \\ &< \frac{1}{N\epsilon^\beta}\sum_{j=0}^{\infty} 2^{\beta(j+1)}\sum_{m=1}^{N}\chi_{(\epsilon 2^{-(j+1)},\epsilon 2^{-j}]}(\|m\alpha\|) \\ &\le \frac{2B}{\epsilon^\beta}\sum_{j=0}^{\infty} 2^{\beta(j+1)}(\epsilon 2^{-(j+1)})^{\frac{1}{\kappa-1}} \quad \text{(for some } B>0) \\ &= O(\epsilon^{\frac{1}{\kappa-1}-\beta}).\end{aligned} \tag{5.18}$$

The inequality before the last follows from Lemma 5.5 below. We conclude that

$$\frac{1}{N}\sum_{m=1}^{N} g_\epsilon(m\alpha) = O(\epsilon^{\frac{1}{\kappa-1}-\beta}) \tag{5.19}$$

Therefore

$$\begin{aligned}\limsup_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} f(m\alpha) &= \limsup_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}[f_\epsilon(m\alpha)+g_\epsilon(m\alpha)] \\ &\le \int_{\mathbb{T}} f(x)dx + O(\epsilon^{1-\beta}) + O(\epsilon^{\frac{1}{\kappa-1}-\beta}),\end{aligned} \tag{5.20}$$

in view of (5.15) and (5.19).

Since $\epsilon > 0$ can be arbitrarily small, the limsup and liminf must coincide. □

The following lemma is used in the preceding proof.

Lemma 5.5. *Let α be diophantine of type κ. Then there is a constant $B > 0$ such that, for any interval $[x_0, x_0+\ell]$,*

$$\#\{m = 1,\dots,N : m\alpha \in [x_0, x_0+\ell]+\mathbb{Z}\} \le \begin{cases} 0 & \text{if } N^{\kappa-1}\ell < c \\ BN\ell^{1/(\kappa-1)} & \text{otherwise.}\end{cases} \tag{5.21}$$

Proof. Define $T = 1/\ell$. Let us divide the counting into blocks of the form

$$\#\{m_0 < m \le m_0 + T^{1/(\kappa-1)} : m\alpha \in [x_0, x_0+\ell]+\mathbb{Z}\}, \tag{5.22}$$

The number of such blocks contributing to (5.21) is less than $O(NT^{-1/(\kappa-1)}+1)$.

The gaps between elements of the sequence $m\alpha$ mod 1, $m_0 < m \le m_0 + T^{1/(\kappa-1)}$, are of the form $n\alpha$ mod 1, with $|n| < 2T^{1/(\kappa-1)}$. By the diophantine condition, the gaps therefore have seize at least

$$\|n\alpha\| \ge \frac{c}{|n|^{\kappa-1}} > \frac{c}{2^{\kappa-1}T}. \tag{5.23}$$

An interval of size $\ell = 1/T$ can hence at most contain a bounded number of elements. Hence

$$\#\{m_0 < m \leq m_0 + T^{1/(\kappa-1)} : m\alpha \in [x_0, x_0 + \ell] + \mathbb{Z}\} \leq B' \tag{5.24}$$

for some constant $B' > 0$ independent of m_0, x_0, ℓ. Recall that there were at most $NT^{-1/(\kappa-1)} + 1$ such blocks, and this yields the upper bound in the second alternative.

The first alternative is easily proven since the minimum gap size for the full sequence $m = 1, \ldots, N$ is at least $c/(2N)^{\kappa-1}$. □

References

[1] M. Einsiedler, A. Katok and E. Lindenstrauss, Invariant measures and the set of exceptions to Littlewoods conjecture, to appear in *Annals of Math.*

[2] N.D. Elkies and C.T. McMullen, Gaps in $\sqrt{n}$ mod 1 and ergodic theory. *Duke Math. J.* **123** (2004) 95–139.

[3] A. Eskin, G. Margulis and S. Mozes, Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture, *Ann. of Math.* **147** (1998) 93-141.

[4] A. Eskin, G. Margulis and S. Mozes, Quadratic forms of signature (2,2) and eigenvalue spacings on rectangular 2-tori, *Ann. of Math.* **161** (2005) 679-725.

[5] D. Kleinbock, Badly approximable systems of affine forms. *J. Number Theory* **79** (1999) 83-102.

[6] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity, *Ann. of Math.* **163** (2006) 165-219.

[7] J. Marklof, The n-point correlations between values of a linear form, with an appendix by Z. Rudnick, *Ergod. Th. Dyn. Sys.* **20** (2000) 1127-1172.

[8] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms, *Ann. of Math.* **158** (2003) 419-471.

[9] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms II, *Duke Math. J.* **115** (2002) 409-434; Correction, *ibid.* **120** (2003) 227-228.

[10] J. Marklof, Energy level statistics, lattice point problems and almost modular functions, in P. Cartier; B. Julia; P. Moussa; P. Vanhove (Editors): *Frontiers in Number Theory, Physics and Geometry. Volume 1: On random matrices, zeta functions and dynamical systems*, Springer, 2006, pp. 163-181.

[11] D.W. Morris, *Ratner's theorems on unipotent flows.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2005.

[12] Z. Rudnick and P. Sarnak, The pair correlation function of fractional parts of polynomials, *Comm. Math. Phys.* **194** (1998) 61-70.

[13] Z. Rudnick and A. Zaharescu, The distribution of spacings between fractional parts of lacunary sequences. *Forum Math.* **14** (2002) 691-712.

[14] N.A. Shah, Limit distributions of expanding translates of certain orbits on homogeneous spaces, *Proc. Indian Acad. Sci., Math. Sci.* **106** (1996) 105-125.

[15] N.B. Slater, Gaps and steps for the sequence $n\theta$ mod 1, *Proc. Cambridge Philos. Soc.* **63** (1967) 1115-1123.

[16] A. Strömbergsson and A. Venkatesh, Small solutions to linear congruences and Hecke equidistribution. *Acta Arith.* **118** (2005) 41-78.

School of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.

j.marklof@bristol.ac.uk