THE METHODS OF DUKE-RUDNICK-SARNAK AND ESKIN-MCMULLEN

TOBIAS FEIGENWINTER AND DOMINIK MOOS

1. Introduction

In the previous talk, we have found a counting result for primitive integer vectors, that is, we have calculated the density of $\{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mid \gcd(x,y)=1\}$ in $\mathbb{R}^2$. These notes generalize this results by replacing the concrete groups from last week by more abstract groups that satisfy some desired properties.

1.1. **The base spaces.** Before we can explain the problem, we need to introduce the spaces involved and compare them to the special case of last week's talk. Let G be an unimodular group, $H < G$ a closed unimodular subgroup, and let Γ be a lattice of G such that $G \cap H$ is a lattice of H. The groups G and H are equipped with some Haar measures m_G, m_H. For Γ and $\Gamma \cap H$ we use the counting measure as Haar measure.

Last week's talk on counting primitive integer vectors was a special case of this talk. In the last talk, G was $\mathrm{SL}_2(\mathbb{R})$, Γ was $\mathrm{SL}_2(\mathbb{Z})$, and H was $\{\left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right) \mid t \in \mathbb{R}\}$

1.2. **The quotient spaces.** We are also interested quotients of those spaces such as G/H and G/Γ. Since we did not require H and Γ to be *normal* subgroups of G, those quotient spaces might not be groups. Therefore we also do *not* have a operation that maps two elements $g_1H, g_2H \in G/H$ to their product $(g_1H)(g_2H) = (g_1g_2)H$ (and similarly for the other quotient spaces). We can, however, multiply elements of G with elements of the quotient space: The product of $g_1 \in G$ and $g_2H \in G/H$ would be given as $g_1(g_2H) = (g_1g_2)H \in G/H$.

For simplicity, we will sometimes use the notation $X = G/\Gamma$ and $Y = H/H \cap \Gamma$.

The quotient spaces can be be equipped with *compatible* measures. In the case of G/H, this means that for all $f \in L^1_{m_G}(G)$, it holds that

$$\int_{G/H} \int_H f(gh)dm_H(h)dm_{G/H}(gH) = \int_G f(g)dm_G(g).$$

In words, an integral over the entire space G can be calculated by integrating first over the subgroup H and then over the quotient G/H. Similar formulas hold for all other quotient spaces that we're working with.

Additionally we will assume that *the translated H-orbits $gH\Gamma$ equidistribute in* G/Γ (we will later explain this property in greater detail). We will then count the point set $g\Gamma H$ in the space G/H.

In last week's talk, the quotient space $G/H = \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ turned out to be equivalent to $\mathbb{R}^2 \setminus \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, and the set $g\Gamma H$ corresponded to the primitive integer vectors.

Here we should also note the difference between $g\Gamma H$ and $gH\Gamma$. The former is a subset of G/H. It is a Γ-orbit (i.e. the set $\{g\gamma \mid \gamma \in \Gamma\}$) in G, projected onto the quotient space ${}^{G}/_{H}$, and it is the object that we want to count. On the other hand, $gH\Gamma$ is a subset of ${}^{G}/_{\Gamma}$, and it is the object for which we require the equidistribution property.

2. Dynamical Assumption on X

In the general setting the equidistribution has to be made into an assumption. In the case of the hyperbolic plane this would correspond to the Theorem about equidistribution of the horocycle flow.

In particular we need the following *equidistribution assumption*:

the translated H-orbits $gH\Gamma$ equidistribute in $X = G/\Gamma$

as $gH \to \infty$ in G/H. Note that a sequence of elements is defined to go to infinity if it eventually escapes every compact set.

Stated more rigorously this means: For all $\alpha \in C_c(X)$ we have

$$\frac{1}{m_Y(Y)} \int_Y \alpha(gh\Gamma)\,\mathrm{dm}_Y \xrightarrow{gH\to\infty} \frac{1}{m_X(X)} \int_X \alpha(x)\,\mathrm{dm}_X(x). \tag{1}$$

One can think of this as stating that integrating along the orbit will yield the same value as integrating over the area since the orbit "paints" the area in uniform density.

This assumption will be enough to prove a weak version of out desire result which will be an important intermediate step.

3. Averaged Counting Result

Let $\{B_t \subset G/H \mid t \in \mathbb{R}\}$ be a collection of subsets each with finite Haar-measure such that $m_{G/H}(N_t) \to \infty$ as $t \to \infty$. We will call these subsets *balls*, but note that this term is unrelated to the balls we know from metric spaces.

We define a modified orbit counting function $F_t : X \to \mathbb{R}_{\geq 0}$ that measures the density of translated lattice points inside the balls $B_t \subseteq G/H$.

$$F_t(g\Gamma) = \frac{1}{m_{G/H}(B_t)} |g\Gamma H \cap B_t|$$

We want to find the asymptotic behavior of of F_t as $t \to \infty$.

For now we will prove weak* convergence. In particular this means that for $\alpha \in C_c(X)$ the following hold.

$$\int_X F_t(x)\alpha(x)\,\mathrm{dm}_X \xrightarrow{t\to\infty} \frac{m_X(X)}{m_Y(Y)} \int_X \alpha(x)\,\mathrm{dm}_X$$

The core idea of the proof is to apply the folding/unfolding trick to integral. We want to traverse the groups as follows.

$$G/\Gamma \cdot \Gamma/(H \cap \Gamma) \nearrow G/(\Gamma \cap H) \searrow G/H \cdot H/\Gamma \cap H$$

Let's start by writing down the integral.

$$\begin{aligned}\int_X F_t(x)\alpha(x)\,\mathrm{dm}_X &= \int_X F_t(g\Gamma)\alpha(g\Gamma)\,\mathrm{dm}_X(g\Gamma)\\ &= \frac{1}{m_{G/H}(B_t)}\int_X |g\Gamma H\cap B_t|\alpha(g\Gamma)\,\mathrm{dm}_X\end{aligned}$$

We can now rewrite the carnality of this (finite) set as a sum over the indicator function of B_t. The sum then is simply the integral with respect to the counting measure.

$$\begin{aligned}&= \frac{1}{m_{G/H}(B_t)}\int_{G/\Gamma}\sum_{\gamma\in\Gamma/(\Gamma\cap H)}[\mathbb{1}_{B_t}(g\gamma H)]\alpha(g\Gamma)\,\mathrm{dm}_X(g\Gamma)\\ &= \frac{1}{m_{G/H}(B_t)}\int_{G/\Gamma}\int_{\Gamma/(\Gamma\cap H)}\mathbb{1}_{B_t}(g\gamma H)\alpha(g\Gamma)\,\mathrm{dm}_{\Gamma/(\Gamma\cap H)}(\gamma)\,\mathrm{dm}_X(g\Gamma)\end{aligned}$$

Now we apply disintegration and do the folding/unfolding trick.

$$\begin{aligned}&= \frac{1}{m_{G/H}(B_t)}\int_{G/(\Gamma\cap H)}\mathbb{1}_{B_t}(gH)\alpha(g\Gamma)\,\mathrm{dm}_{G/(\Gamma\cap H)}(g(\Gamma\cap H))\\ &= \frac{1}{m_{G/H}(B_t)}\int_{G/H}\mathbb{1}_{B_t}(gH)\int_{H/(\Gamma\cap H)}\alpha(gh\Gamma)\,\mathrm{dm}_Y(h\Gamma)\,\mathrm{dm}_{G/H}(gH)\end{aligned}$$

Finally we simply change the integral with an indicator function to be simply an integral over B_t.

$$= \frac{1}{m_{G/H}(B_t)}\int_{B_t}\int_Y \alpha(gh\Gamma)\,\mathrm{dm}_Y(h\Gamma)\,\mathrm{dm}_{G/H}(gH)$$

We now notice that this looks almost like the equidistribution assumption we made earlier. The only difference is that here we have an additional averaging over B_t.

To see that this convergence also holds with the additional averaging we only need the additional assumption that $m_{G/H}(B_t) \xrightarrow{t\to\infty} \infty$. Note that compact sets have finite Haar-measure, therefore it can't be the case that all B_t are contained in a compact set since their measure is unbounded. Finally we get

$$\frac{1}{m_Y(Y)m_{G/H}(B_t)}\int_{B_t}\int_Y \alpha(gh\Gamma)\,\mathrm{dm}_Y(h\Gamma)\,\mathrm{dm}_{G/H}(gH) \xrightarrow{t\to\infty} \frac{1}{m_X(X)}\int_X \alpha(x)\,\mathrm{dm}_X(x)$$

which is exactly the desired result.

4. Interlude: Some intuition

The previous section contained a lot of calculations. They aren't particularly hard per se, but it isn't obvious on first glance what we actually did. We will now provide a more intuitive—albeit incomplete—explanation of what we did previously as well as what we will do in the next section. For this, we will often refer to the example from the previous talk where we wanted to count the primitive integer vectors in $\mathbb{R}^2$.

Remember that we want to calculate, in some sense, the "density" of $g\Gamma H$ in G/Γ. Since there is no clear meaning of the density of infinitely many points in an infinitely large space, we define this "density" as the limit of the density within some sequence of finite subsets whose size tends to infinity.

It turns out (this will be shown in the next section) that, as long as we take a somewhat "reasonable" sequence of balls, this limit of densities will be independent of the exact sequence of balls that we choose. To better understand both the previous and the next section, we will now look at two kinds of "unreasonable" sequences of balls.

Firstly, we may force an artficially high density by intentionally having our balls hit many of the points that we want to count. In the case of primitive integer vectors, such balls might look like in figure 3. Secondly, we may force an artificially low density by intentionally avoiding lattice points. This can be seen in figure 2

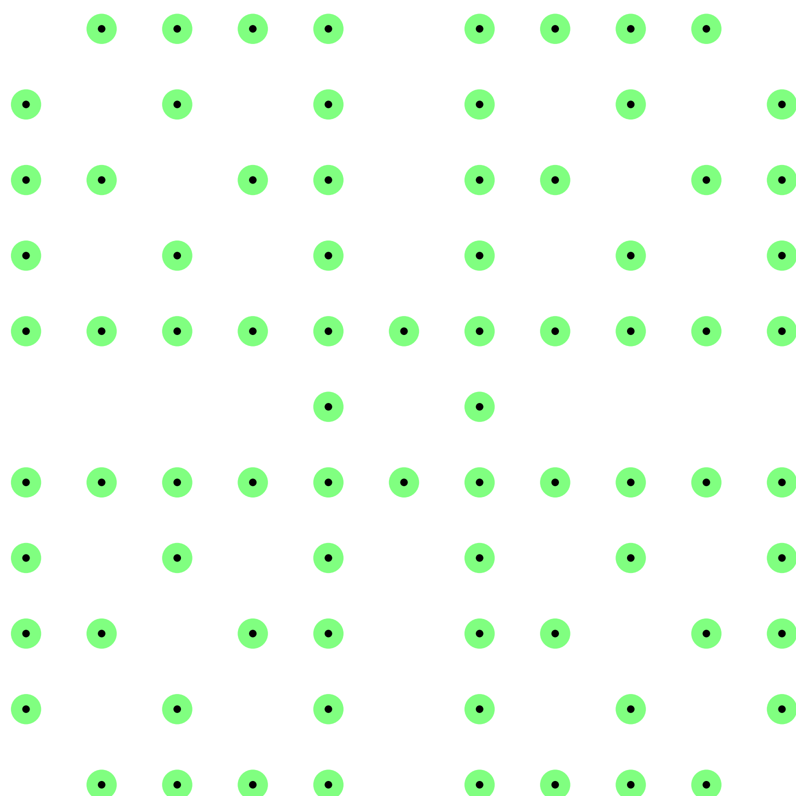

Figure 1. An ball containing an artificially high amount of lattice points. The black dots are the primitive integer vectors. Note that the entire green area is a single ball.

The previous section solved this problem by not taking the density of the balls themselves, but instead by taking a weighted average of slightly translated balls (to be precise, the theorem does not in any way require that the balls are translated by only a small amount, but this is the "interesting" or "important" case). Figure **??** shows the ball from figure 3 translated by various small elements of $G = \mathrm{SL}_2(\mathbb{R})$.

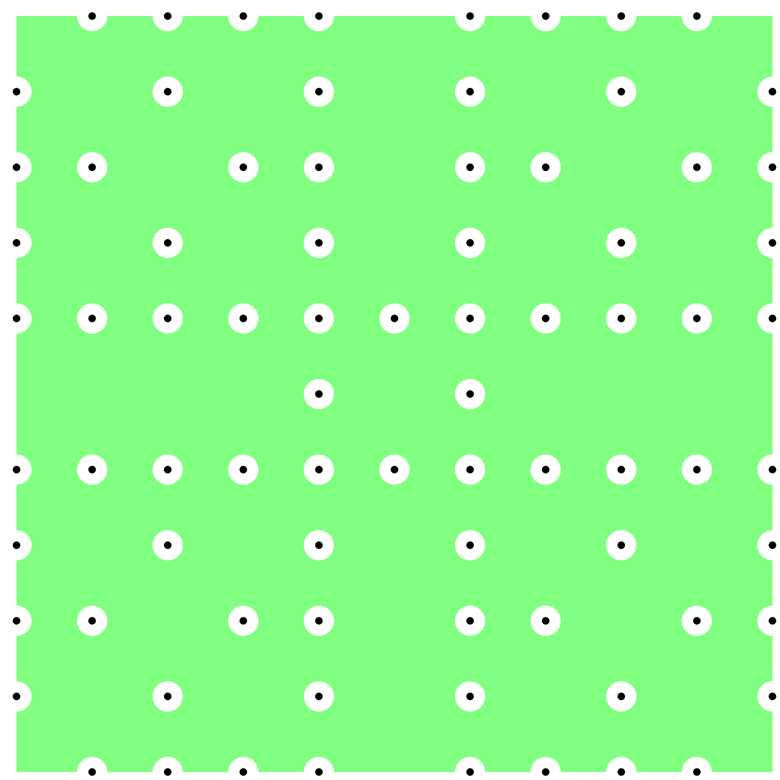

FIGURE 2. An ball containing an artificially low amount of lattice points. The black dots are the primitive integer vectors. Again, the ball is the green area.

Note that "translating" by some g just means multiplication of the points of the ball with that g and in this case does not correspond to the geometric concept of translation.

Figure **??** only shows a few discrete translations. In reality, we would have a continuum of translations, and taking the average of all those translations would be akin to blurring the edges of the ball. This is shown in 4.

In this special case of the primitive integers, we can see that points far away from the origin are more affected by this blurring. This can be used as an intuitive explanition for why, as the ball grows in size, the *blurred* ball covers an increasingly representative area of the entire space. This means that the limit of densities of the blurred balls will give us the correct density. To rigorously show this (and also to cover our more general case), we didn't use the fact that distant points are more affected by translation but instead argued with the equidistribution property. Altough slightly less intuitive, it still seems plausible that this property guarantees that the blurred balls will eventually cover a sufficiently representative area to find the correct density.

The next section relies on the same intuition. However, we will no longer blur the balls with an integral. Instead, this blurring is already encoded in our sequence

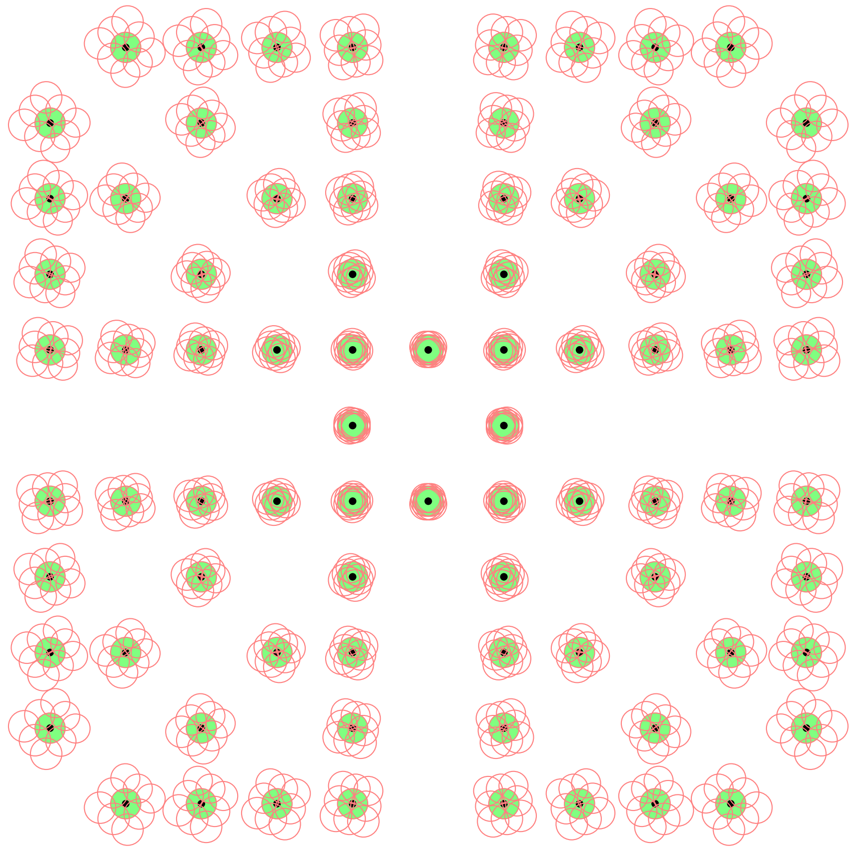

FIGURE 3. The ball from figure 3 (green), translated by various elements of $G = \mathrm{SL}_2(\mathbb{R})$. The translated balls are shown as red outlines.

of balls. Namely, we will require that if we blur the ball B_t, the result must be contained in the slightly larger ball $B_{t+\delta}$.

This condition will not quite suffice: We can still break the theorem by expanding our ball in an "unreasonable" way faster than the blurring makes it "reasonable". For example, we could still add artificially dense parts of the space (like in figure 3), we just have to add them faster than the blurring takes effect. To avoid this, we will need an additional condition that bounds the relative growth of the balls.

5. (Non-averaged) counting result

We will now implement what we've discussed at the end of the last chapter in a more rigorous manner. We have mentioned that we want some additional properties for the sequence of balls $\{B_t \mid t \in \mathbb{R}\}$. We want the sequence to be *well-rounded*, which means that

(i) $m_{G/H}(B_t) \to \infty$ as $t \to \infty$
(ii) For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $t \geq 0$ we have $\frac{m_{G/H}(B_{t+\delta})}{m_{G/H}(B_t)} < 1 + \varepsilon$.

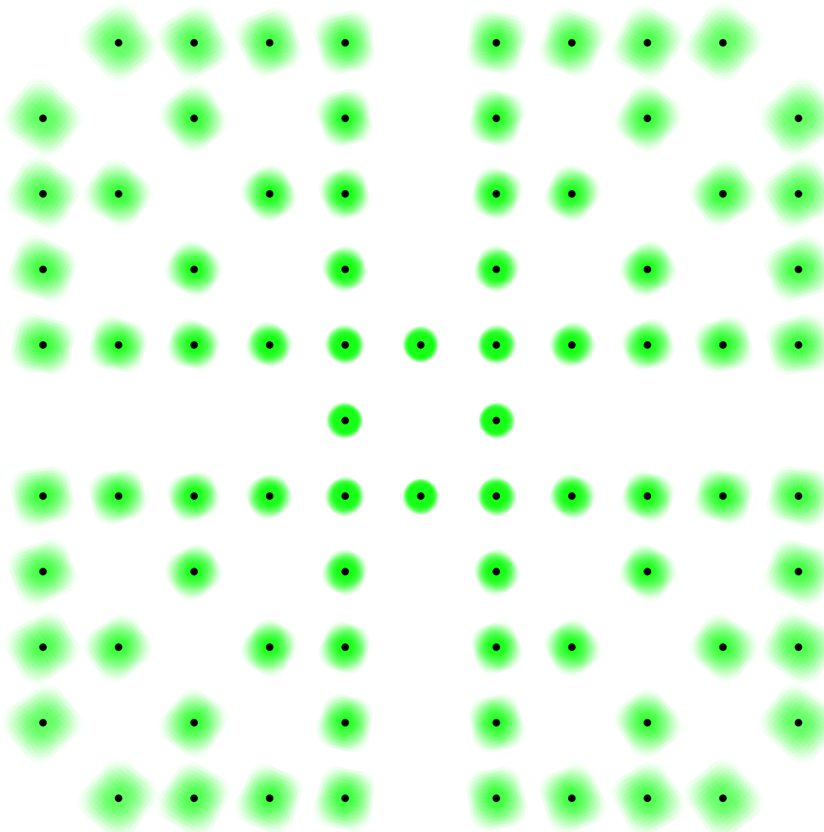

FIGURE 4. The ball from figure 3, "blurred". The color indicates how strongly each point is weighted.

(iii) For every $\delta > 0$ there exists a neighborhood U of the identity in G such that for every $t \geq 0$ we have

$$\bigcup_{g \in U} gB_t \subseteq B_{t+\delta}$$

We can then show the following counting result:

Lemma 2. *Assuming all the previous conditions, we have*

$$\lim_{t\to\infty} \frac{1}{m_{G/H}} |\Gamma H \cap B_t| = \frac{m_{H/\Gamma \cap H}\left(H/\Gamma \cap H\right)}{m_{G/\Gamma}(G/\Gamma)}$$

Proof. Choose some $\varepsilon > 0$, then choose δ as in (ii), then choose U as in (iii). We may assume WloG that $U = U^{-1} := \left\{u^{-1} \mid u \in U\right\}$: Otherwise, use $U \cap U^{-1}$ as our new value for U.

For any $g \in U$, we now have

$$\begin{aligned} F_{t+\delta}(g\Gamma) &\stackrel{\text{def}}{=} \frac{1}{m_{G/H}(B_{t+\delta})} |g\Gamma H \cap B_{t+\delta}| \\ &= \frac{1}{m_{G/H}(B_{t+\delta})} |\Gamma H \cap \underbrace{g^{-1}B_{t+\delta}}_{\supseteq B_t}| \\ &\geq \frac{m_{G/H}(B_t)}{m_{G/H}(B_{t+\delta})} \frac{1}{m_{G/H}(B_t)} |\Gamma H \cap B_t| \\ &\stackrel{\text{(ii)}}{\geq} \frac{1}{1+\varepsilon} \frac{1}{m_{G/H}(B_t)} |\Gamma H \cap B_t| \end{aligned}$$

Now let $\alpha \in C_c(G/\Gamma)$ be an *approximate identity at the identity coset*, meaning that α is nonnegative, its total integral is

$$\int_{G/\Gamma}^{\alpha} (g\Gamma) dm_{G/\Gamma}(g\Gamma) = 1$$

, and the support of α is contained in $U\Gamma$.

We multiply the above inequality by α, integrate, and take limits.

$$\begin{aligned} \lim_{t\to\infty} \frac{1}{m_{G/H}(B_t)} |\Gamma H \cap B_t| &= \lim_{t\to\infty} \int_{G/H} \frac{1}{m_{G/H}(B_t)} |\Gamma H \cap B_t| \alpha(g\Gamma) dm_{G/\Gamma}(g\Gamma) \\ &\leq \lim_{t\to\infty} (1+\varepsilon) \int_{G/\Gamma} F_{t+\delta}(g\Gamma)\alpha(g\Gamma) dm_{G/\Gamma}(g\Gamma) \\ &= (1+\varepsilon) \frac{m_{H/\Gamma\cap H}(H/\Gamma \cap H)}{m_{G/\Gamma(G/\Gamma)}} \end{aligned}$$

We get the first equality by taking constants out of the integral and remembering that the total integral of α is one. The inequality was taken from the previous calculation (after multiplying by α, integrating, and taking limits). Finally, the last equality is the averaged counting result.

If we let ε approach zero, we get

$$\lim_{t\to\infty} \frac{1}{m_{G/H}} |\Gamma H \cap B_t| \leq \frac{m_{H/\Gamma\cap H}\left(H/\Gamma \cap H\right)}{m_{G/\Gamma}(G/\Gamma)}.$$

By starting our calculations with $F_{t-\delta}(g\Gamma)$ instead of $F_{t+\delta}(g\Gamma)$, we get the inverse inequality, and the two inequalities together give us the equality that we wanted to show. □