The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1950-5

School and Workshop on Dynamical Systems

30 June - 18 July, 2008

Ergodic theory, combinatorics, diophantine approximation

V. Bergelson
Ohio State University, USA

A. Gorodnik
University of Bristol, UK

TRIESTE LECTURES ON ERGODIC THEOREMS, RECURRENCE, AND APPLICATIONS

VITALY BERGELSON AND ALEXANDER GORODNIK

Preliminary version

Contents

1. Introduction

The standard set-up of ergodic theory consists of a measurable space $(X, \mathcal{B})$ equipped with a probability measure μ and a measure-preserving transformation $T : X \to X$, that is, $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. In ergodic theory, one is interested in statistical properties of the orbits $\{T^n x\}_{n\geq 0}$. Origins of this subject can be traced back to the works of Boltzmann, Gibbs and Poincare, and the word "ergodic" comes from so-called ergodic hypothesis in statistical physics. A function $f : X \to \mathbb{R}$ is considered as an observable, which we can sample along the trajectory $\{T^n x\}_{n\geq 0}$ to compute its time average $\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$. The ergodic hypothesis of Boltzmann roughly stated that if the space does not split into invariant pieces, then the time average should converge to the space average $\int_X f\, d\mu$. A rigorous version of the ergodic hypothesis is the mean ergodic theorem, which was proved by von Neumann in 40s. We formulate this result in a more general setting of operators on a Hilbert space:

Theorem 1.1 (von Neumann). *Let U be a contraction (i.e., $\|U\| \leq 1$) operator on a Hilbert space $\mathcal{H}$ and P_U denotes the orthogonal projection* *on the space U-invariant vectors. Then for every $v \in \mathcal{H}$*

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n v \xrightarrow{\|\cdot\|} P_U v \quad \textit{as } n \to \infty, \tag{1.1}$$

where P denotes the orthogonal projection on the space of U-invariant vectors.

Proof. We will use the following decomposition:

$$\mathcal{H} = \ker(U^* - I) \perp \overline{\operatorname{Im}(U - I)}, \tag{1.2}$$

which is valid for an operator which is not necessarily unitary. To prove this, we observe that for $v, w \in \mathcal{H}$,

$$\langle v, (U - I)w\rangle = \langle (U^* - I)v, w\rangle .$$

So if v is orthogonal to $\operatorname{Im}(U - I)$, it follows that $(U^* - I)v = 0$. Conversely, if $v \in \ker(U^* - I)$, then v is orthogonal to $\operatorname{Im}(U - I)$.

Now we claim that $\ker(U - I) = \ker(U^* - I)$. For $v \in \ker(U - I)$, we have $\langle v, U^*v\rangle = \langle Uv, v\rangle = \|v\|^2$. Hence, by the equality case of the Cauchy-Schwartz inequality, $U^*v = v$. The other inclusion is proved similarly.

For $v \in \ker(U - I)$, (1.1) is obvious, and for $v = Uw - w$, we have

$$\left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n v\right\| = \left\|\frac{1}{N}(U^N w - w)\right\| \le \frac{2\|w\|}{N} \to 0 \quad \text{as } N \to \infty.$$

Hence, it follows from (1.2) that (1.1) holds for a dense family of vectors. The convergence for general vectors is easy to deduce with a help of the triangle inequality. □

To apply Theorem 1.1 to the study of the dynamical systems, we observe that a measure-preserving transformation $T : X \to X$ defines a contraction operator U_T on the Hilbert space $L^2(X)$:

$$(U_T f)(x) = f(Tx), \quad f \in L^2(X). \tag{1.3}$$

This simple observation provides a fundamental connection between ergodic theory and harmonic analysis, and it will be crucial for our purposes.

The transformation T is called *ergodic* if $L^2(X)$ contains no nonconstant U_T-invariant functions. Note that in this case, the projection map P_{U_T} is given by $P_{U_T} f = \int_X f\, d\mu$, and we have the following corollary:

Corollary 1.2 (von Neumann mean ergodic theorem)**.** *Let T be an ergodic measure-preserving transformation of a probability space (X,μ). Then for every $f\in L^2(X)$,*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \xrightarrow{L^2} \int_X f\,d\mu \quad \text{as } n\to\infty$$

in L^2-norm.

Here we consider some generalizations of this classical mean ergodic theorem and applications in number theory, Diophantine approximation, and combinatorics.

2. Notation and preliminaries

In these lectures G denotes a topological group. We always assume that

$$\boxed{G \text{ is locally compact and compactly generated.}}$$

A reader might think about some concrete examples: $\mathbb{R}^d$, the group of upper triangular matrices, $\mathrm{SL}_d(\mathbb{R})$, etc.

The group G supports a *right invariant* (regular) Borel measure m, which is called *Haar measure*:

$$m(Bg)=m(B) \text{ for every } g\in G \text{ and Borel } B\subset G.$$

Moreover, this measure is unique up to a scalar multiple. It follows from uniqueness that there exists a function $\Delta: G\to\mathbb{R}^+$, which is called the *modular function*, such that

$$m(gB)=\Delta(g)m(B) \text{ for every Borel } B\subset G.$$

This measure satisfies

$$\begin{aligned} m(U)&>0 \quad \text{for open } U\subset G,\\ m(K)&<\infty \quad \text{for compact } K\subset G.\end{aligned}$$

We note that in most examples such measure can be given explicitly. It is easy to construct for Lie groups by integrating a nonzero invariant differential form of top degree.

Example 2.1. The Lebesgue measure on $\mathbb{R}^d$ is a Haar measure.

Exercise 2.2. (1) Show that left/right Haar measure on $\mathrm{GL}_d(\mathbb{R})$ is given by

$$\det(g)^{-d}\prod_{i,j} dg_{ij}$$

(2) Show that left and right Haar measure on

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} x & y \\ z & t \end{pmatrix} : xt - yz = 1 \right\}$$

is given by $\frac{1}{x} dxdydz$.

(3) Compute left and right Haar measures for the affine group

$$\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a \in \mathbb{R}^+, b \in \mathbb{R} \right\}. \tag{2.1}$$

Compute the function Δ.

(4) Prove that Δ is a continuous homomorphism.

(5) Prove that if the Haar measure is finite, then G is compact.

A group G is called *unimodular* if Haar measure is both left and right invariant

Exercise 2.3. (1) Every compact group is unimodular.

(2) $\mathrm{SL}_d(\mathbb{R})$ is unimodular.

3. Amenable groups

We consider a measure-preserving action of a group G on a probability space (X, μ) and aim to prove an ergodic theorem for such actions. As a natural generalization of the von Neumann ergodic theorem, we consider the averages

$$A_n f(x) := \frac{1}{m(B_n)} \int_{B_n} f(g^{-1}x)\, dm(g) \tag{3.1}$$

defined for $f \in L^2(X)$ and a sequence of measurable sets B_n such that $0 < m(B_n) < \infty$. Similarly to (1.3), we can introduce a unitary representation π_X of G on the space $L^2(X)$:

$$\pi_X(g) f(x) = f(g^{-1}x), \quad g \in G,\ f \in L^2(X).$$

The averaging operators (3.1) can be defined for a general unitary representation of G on a Hilbert space $\mathcal{H}$. The operators $A_n : \mathcal{H} \to \mathcal{H}$ are defined by

$$\langle A_n v, w \rangle := \frac{1}{m(B_n)} \int_{B_n} \langle \pi(g)v, w \rangle\ dm(g), \quad v, w \in \mathcal{H}. \tag{3.2}$$

After a short contemplation on the proof of Theorem 1.1, one realizes that the key ingredient was the asymptotic invariance of the intervals $[0, N-1]$ under translations. This leads to the notion of a Følner sequence. We say that a sequence of measurable sets B_n such that

$0 < m(B_n) < \infty$ is a *Følner sequence* if for a compact generating set Q of G, we have

$$\sup_{g \in Q} \frac{m(B_n \triangle B_n g)}{m(B_n)} \to 0 \quad \text{as } n \to \infty.$$

A group G is called *amenable* if such a sequence exists.

Exercise 3.1. Prove that for every Følner sequence B_n,

$$\frac{m(B_n \triangle B_n g)}{m(B_n)} \to 0 \quad \text{as } n \to \infty$$

uniformly on g in compact sets. (Hint: use the inclusion $A\triangle C \subset (A\triangle B) \cup (B \triangle C)$ and invariance of the measure.)

Exercise 3.2. (1) Show that a sequence of boxes $[a_n^{(1)}, b_n^{(1)}] \times \cdots \times [a_n^{(d)}, b_n^{(d)}]$ is a Følner sequence iff $\min_i(b_n^{(i)} - a_n^{(i)}) \to \infty$.
(2) Construct a Følner sequence for the affine group (2.1).

The original argument of von Neumann easily generalizes to prove

Theorem 3.3. *Let B_n be a Følner sequence in a group G and π a unitary representation of G on a Hilbert space $\mathcal{H}$. Then for the sequence of operators A_n defined in (3.2) and every $v \in \mathcal{H}$,*

$$A_n v \xrightarrow{\|\cdot\|} P_G v \quad \text{as } n \to \infty, \tag{3.3}$$

where P_G is the orthogonal projection on the space of G-invariant functions.

In particular, we immediately obtain the following

Corollary 3.4. *Let B_n be a Følner sequence in a group G that acts ergodically on a probability space (X, μ). Then for every $f \in L^2(X)$,*

$$\frac{1}{m(B_n)} \int_{B_n} f(g^{-1}x)\, dm(g) \xrightarrow{L^2} \int_X f\, d\mu \quad \text{as } n \to \infty. \tag{3.4}$$

Proof of Theorem 3.4. As in the proof of Theorem 1.1, one checks the orthogonal decomposition

$$\mathcal{H} = \mathcal{H}_{inv} \perp \mathcal{H}_{erg}$$

where

$$\mathcal{H}_{inv} = \{v \in \mathcal{H} : \pi(g)v = v \text{ for all } g \in G\},$$
$$\mathcal{H}_{erg} = \overline{\operatorname{span}\{\pi(g)w - w :\ g \in G,\ w \in \mathcal{H}\}}.$$

For $v \in \mathcal{H}_{inv}$, (3.4) is obvious, and for $v = \pi(g_0)w - w$,

$$\begin{aligned}\|A_n v\| &= \left| \frac{1}{m(B_n)} \int_{B_n} \pi(gg_0)w \, dm(g) - \frac{1}{m(B_n)} \int_{B_n} \pi(g)w \, dm(g) \right| \\ &= \frac{1}{m(B_n)} \left| \int_{B_n g_0} \pi(g)w \, dm(g) - \int_{B_n} \pi(g)w \, dm(g) \right| \\ &\le \frac{1}{m(B_n)} m(B_n g_0 \triangle B_n) \|w\| \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Taking linear combinations, we obtain a dense family of vectors for which (3.4) holds. Since $\|A_n\| \le 1$ and $\|P_G\| \le 1$, the general case follows by triangle inequality. ☐

Amenable groups were introduced by von Neumann in relation to the Banach–Tarski paradox. Since then the notion of amenability has found an amazing number of applications. We mention only several characterizations of amenability which appear naturally in our context:

- Existence of invariant means,
- Existence of almost invariant vectors (this will lead to the notion of Kazhdan groups in Section 5),
- Bounds on the spectrum of the averaging operators (this will lead to the notion of spectral gap).

The important property of amenable groups (which, in fact, characterizes amenability) is existence of invariant measures.

Theorem 3.5 (Bogolubov, Krylov)**.** *Consider a continuous action of an amenable group G on a compact space X. Then X support a G-invariant probability measure.*

Proof. By the Riesz representation theorem, the space $\mathcal{M}(X)$ of Borel measures on X can be identified with the dual of the space $C(X)$ of continuous functions. We equip $\mathcal{M}(X)$ with the weak* topology. Namely, $\mu_n \to \mu$ if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu \quad \text{for every } f \in C(X).$$

The key ingredient of the proof is the compactness of the space $\mathcal{M}_1(X)$ of probability measures, which is a special case of the Banach–Alaoglu theorem. Consider the sequence of the probability measures μ_n defined by

$$\int_X f \, d\mu_n = \frac{1}{m(B_n)} \int_{B_n} f(g^{-1}x) \, dm(g), \quad f \in C(X),$$

where B_n is a Følner sequence. By compactness, we have convergence $\mu_{n_i} \to \mu$ along a subsequence to a probability measure μ. We claim that μ is G-invariant. As in the proof of Theorem 3.4, we have, for $g_0 \in G$ and $f \in C(X)$,

$$\begin{aligned}
&\left|\int_X f(g_0^{-1}x)\,d\mu_n(x) - \int_X f(x)\,d\mu_n(x)\right| \\
=&\frac{1}{m(B_n)}\left|\int_{B_n} f(g_0^{-1}g^{-1}x)\,dm(g) - \int_{B_n} f(g^{-1}x)\,dm(g)\right| \\
\le&\frac{1}{m(B_n)}m(B_n g_0 \triangle B_n)\|f\|_\infty \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Hence,

$$\int_X f(g_0^{-1}x)\,d\mu(x) = \int_X f(x)\,d\mu(x)$$

for every $g_0 \in G$ and $f \in C(X)$, as claimed. □

Exercise 3.6. (1) Consider the action of $\mathrm{SL}_2(\mathbb{R})$ on the space $X = \mathbb{R} \cup \{\infty\}$ defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \frac{ax+b}{cx+d}.$$

Show that there are no finite $\mathrm{SL}_2(\mathbb{R})$-invariant measure X. (Hint: the north-south pole dynamics of the diagonal subgroup.)

Note that it follows from Theorem 3.5 that $\mathrm{SL}_2(\mathbb{R})$ is not amenable.

(2) Show that a nonabelean free group, equipped with the discrete topology, is not amenable.

Let π be a unitary representation of a group G on a Hilbert space $\mathcal{H}$. Given subset $Q \subset G$ and $\varepsilon > 0$, we call a vector v a (Q, ϵ)-*invariant* if

$$\|\pi(g)v - v\| \le \varepsilon\|v\| \quad \text{for every } g \in Q.$$

We say that π has *almost invariant vectors* if for every compact subset Q and $\varepsilon > 0$, there exists a nonzero (Q, ϵ)-invariant vector.

Another characterization of amenability is in terms of almost invariant vectors for the regular representation. The *regular representation* λ of a group G is defined by

$$\lambda(g)f(x) = f(xg), \quad g \in G,\ f \in L^2(G).$$

Theorem 3.7 (Hulanicki, Reiter). *A group G is amenable if the regular representation of G has almost invariant vectors.*

Note that $L^2(G)$ contains no invariant vectors when G is not compact.

Lemma 3.8. *Let ϕ, ψ be nonnegative functions in $L^1(G)$. Then for $E_t = \{g \in G : \phi(g) \ge t\}$ and $F_t = \{g \in G : \psi(g) \ge t\}$,*

$$\|\phi - \psi\|_1 = \int_0^\infty m(E_t \triangle F_t)dt.$$

Exercise 3.9. Prove this lemma. (Hint: Observe that by the Fubini theorem, $\phi(x) = \int_0^\infty \chi_{E_t}(x)dt$.)

Proof of Theorem 3.7. Let B_n be a Følner sequence in G and $f_n = \frac{\chi_{B_n}}{m(B_n)^{1/2}}$. Then $\|f_n\|_2 = 1$ and

$$\|\lambda(g)f_n - f_n\|_2 = \frac{m(B_n g^{-1} \triangle B_n)^{1/2}}{m(B_n)^{1/2}} \to 0,$$

uniformly on g in compact sets. Hence, $L^2(G)$ contains almost invariant vectors.

Conversely, let Q be a compact generating set of G and f_n be a sequence of vectors in $L^2(G)$ such that $\|f_n\| = 1$ and

$$\|\lambda(g)f_n - f_n\|_2 \to 0$$

uniformly on $g \in Q$. Consider $\phi_n = f_n^2 \in L^1(G)$. By the Cauchy-Schwartz inequality,

$$\begin{aligned}\|\lambda(g)\phi_n - \phi_n\|_1 &\le \|\lambda(g)f_n + f_n\|_2 \|\lambda(g)f_n - f_n\|_2 \\ &\le 2\|\lambda(g)f_n - f_n\|_2 \to 0.\end{aligned}$$

Let $B_{n,t} = \{g \in G : \phi_n(g) \ge t\}$. Then by Lemma 3.8, we have

$$\|\lambda(g)\phi_n - \phi_n\|_1 = \int_0^\infty m(B_{n,t} g^{-1} \triangle B_{n,t})dt$$

and for every compact $K \subset G$,

$$\alpha_n := \int_0^\infty m(B_{n,t}) \left(\int_K \frac{m(B_{n,t} g^{-1} \triangle B_{n,t})}{m(B_{n,t})} dm(g) \right) dt \to 0.$$

Since $\int_0^\infty m(B_{n,t})dt = \|\phi_n\|_1 = 1$, the measure $d\nu_n(t) = m(B_{n,t})dt$ is a probability measure. For the set

$$\Omega_n = \left\{ t > 0 : \int_K \frac{m(B_{n,t} g^{-1} \triangle B_{n,t})}{m(B_{n,t})} dm(g) \ge \alpha_n^{1/2} \right\},$$

we have $\nu_n(\Omega_n) \le \alpha_n^{1/2} \to 0$. Hence, there exists t_n such that for $B_n = B_{n,t_n}$, we have $0 < m(B_n) < \infty$ and

$$\int_K \frac{m(B_n g^{-1} \triangle B_n)}{m(B_n)} dm(g) \to 0.$$

This already implies convergence in measure, but we need to prove the uniform convergence!

Let $\varepsilon > 0$, and

$$K_{n,\varepsilon} = \{k \in K : \frac{m(B_n k^{-1} \triangle B_n)}{m(B_n)} \le \epsilon\}.$$

Note that for $g \in G$, we have

$$m(K_{n,\varepsilon} g \cap K_{n,\varepsilon}) \to m(Kg \cap K) \quad \text{as } n \to \infty, \tag{3.5}$$

uniformly on $g \in G$, and

$$K_{n,\varepsilon}^{-1} \cdot K_{n,\varepsilon} \subset K_{n,2\varepsilon}.$$

Now we assume that Q is a compact symmetric generating set that contains identity, $m(Q) > 0$, and $K = Q^2$. Then for $g \in Q$, we have $m(Kg \cap K) \ge m(Q)$. Then it follows from (3.5) that for $n > n_0(\epsilon)$, we have $K_{n,\varepsilon} g \cap K_{n,\varepsilon} \ne \emptyset$. Hence, $g \in K_{n,2\varepsilon}$. This proves that

$$\frac{m(B_n g^{-1} \triangle B_n)}{m(B_n)} \le 2\epsilon$$

for every $g \in Q$ and $n \ge n_0(\varepsilon)$ as required. □

It would be convenient to generalise definition (3.2). Given a probability Borel measure μ on G and a unitary representation π of G on a Hilbert space $\mathcal{H}$, we define an operator $\pi(\mu) : \mathcal{H} \to \mathcal{H}$:

$$\langle \pi(\mu) v, w \rangle := \int_G \langle \pi(g) v, w \rangle \, d\mu(g), \quad v, w \in \mathcal{H}.$$

Note that we always have $\|\pi(\mu)\| \le 1$, and nontrivial upper bounds are very useful.

Theorem 3.10 (Kesten). *Let μ be a probability measure on G which is absolutely continuous with respect to Haar measure, and* $\mathrm{supp}(\mu)$ *generates a dense subgroup of G. Then G is amenable iff the spectrum of the operator $\lambda(\mu)$ contains* 1 *(i.e., the operator $\lambda(\mu) - I$ does not have a bounded inverse).*

Proof. We will prove that the condition on the spectrum is equivalent to the condition from Theorem 3.7. Assume that the regular representation λ contains almost invariant vectors. Let Q_n be a compact subset of G such that $\mu(Q_n) \ge 1 - 1/n$. There exists a sequence of unit vectors $v_n \in L^2(G)$ such that $\|\lambda(g) v_n - v_n\| \le 1/n$ for all $g \in Q_n$.

Then it follows that

$$\begin{aligned}\|\lambda(\mu)v_n - v_n\| &\le \int_G \|\lambda(g)v_n - v_n\| d\mu(g) \\ &\le \int_{Q_n} \|\lambda(g)v_n - v_n\| d\mu(g) + \frac{2}{n} \le \frac{3}{n} \to 0.\end{aligned}$$

This implies that $\lambda(\mu) - I$ does not have a bounded inverse.

Now we assume that $1 \in \operatorname{spec}(\pi(\mu))$. We first claim that there exists a sequence of unit vectors $v_n \in L^2(G)$ such that

$$\|\lambda(\mu)v_n - v_n\| \to 0. \tag{3.6}$$

It suffices to consider the case when $\operatorname{Ker}(\lambda(\mu) - I) = 0$. If $\operatorname{Im}(\lambda(\mu) - I)$ is dense in $L^2(G)$, then the inverse of $\lambda(\mu) - I$ should be unbounded, and this implies (3.6). Otherwise, it follows from (1.2) that there exists $v \in \operatorname{Ker}(\lambda(\mu)^* - I)$, $v \ne 0$, such that

$$\langle \lambda(\mu)v, v\rangle = \langle v, \lambda(\mu)^* v\rangle = \|v\|^2,$$

i.e., we have equality in the Cauchy-Schwartz inequality. Hence $\lambda(\mu)v = v$. This proves (3.6).

Since μ is absolutely continuous, $d\mu(g) = \phi(g)dm(g)$ for some $\phi \in L^1(G)$, so

$$\lambda(\mu)v_n(x) = \int_G v_n(xy)\phi(y)dm(y)$$

and

$$\lambda(g)\lambda(\mu)v_n(x) = \int_G v_n(xy)\phi(g^{-1}y)\Delta(g)^{-1}dm(y).$$

This implies that for every $\varepsilon > 0$, there exists a neighborhood U of identity in G such that

$$\|\lambda(g)\lambda(\mu)v_n - \lambda(\mu)v_n\| < \varepsilon \quad \text{for all } g \in U \text{ and } n \ge 0. \tag{3.7}$$

We set $w_n = \lambda(\mu)v_n / \|\lambda(\mu)v_n\|$. Since $\|\lambda(\mu)\| \le 1$, it follows from (3.6) that

$$\|\lambda(\mu)w_n - w_n\| \to 0,$$

and

$$|\langle \lambda(\mu)w_n, w_n\rangle - 1| = |\langle \lambda(\mu)w_n - w_n, w_n\rangle| \le \|\lambda(\mu)w_n - w_n\| \to 0.$$

Hence,

$$\int_G (1 - \langle \lambda(g)w_n, w_n\rangle) dm(g) \to 0,$$

i.e., $\langle \lambda(g)w_n, w_n\rangle \to 1$ in measure. Passing to a subsequence, we may assume that convergence holds almost everywhere. Since

$$\|\lambda(g)w_n - w_n\|^2 = 2 - 2\operatorname{Re}\langle \lambda(g)w_n, w_n\rangle,$$

the set

$$H = \{g \in G : \|\lambda(g)w_n - w_n\| \to 0\}$$

has full measure. Then $\mathrm{supp}(\mu) \subset \bar{H}$, and by the assumption on μ, $\bar{H} = G$.

Let $\varepsilon > 0$ and Q be a compact generating set of G. It follows from (3.8) that for a suitable neighborhood U of identity in G,

$$\|\lambda(u)w_n - w_n\| < \varepsilon \quad \text{for all } u \in U \text{ and } n \geq n_0. \tag{3.8}$$

Since H is dense, there exist $h_1, \dots, h_k \in H$ such that $Q \subset \cup_{i=1}^k h_i U$. Note that for all sufficiently large n,

$$\|\lambda(h_i)w_n - w_n\| < \varepsilon \quad \text{for all } i = 1, \dots, k.$$

Writing $g \in Q$ as $g = h_i u$ for some $i = 1, \dots, k$ and $u \in U$, we get

$$\begin{aligned}\|\lambda(g)w_n - w_n\| &\leq \|\lambda(h_i u)w_n - \lambda(h_i)w_n\| + \|\lambda(h_i)w_n - w_n\| \\ &\leq \|\lambda(u)w_n - w_n\| + \|\lambda(h_i)w_n - w_n\| < 2\varepsilon.\end{aligned}$$

This shows that $L^2(G)$ contains almost invariant vectors, and completes the proof of the theorem. □

Theorem 3.10 implies in particular that for amenable groups, we always have

$$\|\lambda(\mu)\| = 1.$$

This hints that there is no rate of convergence in the mean ergodic theorem. The situation is quite different for semisimple Kazhdan groups such as $\mathrm{SL}_d(\mathbb{R})$, $d \geq 3$, as we will see later.

We say that the unitary representation π of a group G has a *spectral gap* if $1 \notin \mathrm{spec}(\pi(\mu))$ for some absolutely continuous probability measure μ on G whose support generates G topologically. The action of G on a probability space has a *spectral gap* if the representation π_X^0 on the space $L_0^2(X)$, the subspace of functions orthogonal to the constant, has spectral gap.

Exercise 3.11. Show that the notion of spectral gap does not depend on a choice of the measure μ.

Exercise 3.12. (1) Consider the rotation $T : x \mapsto x + \alpha \mod 1$ on the circle $X = \mathbb{R}/\mathbb{Z}$ and the measures $\mu_n = \frac{1}{N}\sum_{n=0}^{N-1} \delta_{T^i}$ where δ_x denotes the Dirac measure. Show that the spectrum of $\pi_X(\mu_n)$ contains one for all n.

(2*) Let T be an invertible measure-preserving transformation of a general probability measure space (X, μ) and μ_n is defined as above. Show that $\|\pi_X(\mu_n)\| = 1$. (Hint: One can use the notion of Rohklin tower (see, for instance, Halmos' book on ergodic theory).

Exercise 3.13. Let G be a free group with generators a, b and $\mu = \frac{1}{2}\delta_a + \frac{1}{2}\delta_b$. Prove that $\|\lambda(\mu)\| = 1$, but $1 \notin \operatorname{spec}(\pi(\mu))$.

Although the regular representation of an amenable group never has a spectral gap, there are many natural examples of actions of amenable groups with spectral gap.

Exercise 3.14. Consider the action of $G = \mathbb{R}^2$ on $X = \mathbb{R}^2/\mathbb{Z}^2$ by translations. Show that this action has a spectral gap.

4. Decay of matrix coefficients

A measure-preserving action of a group G on a probability space (X, μ) is called mixing if for every $f_1, f_2 \in L^2(X)$, we have

$$\langle \pi_X(g) f_1, f_2 \rangle \to \left(\int_X f_1 \, d\mu \right) \left(\int_X f_2 \, d\mu \right)$$

as $g \to \infty$ ("$g \to \infty$" means that the sequence g has no accumulation points).

Exercise 4.1.
(1) Consider the measure-preserving transformation T of a probability space (X, μ). If $L^2(X)$ has non-constant eigen-functions, then the action is not mixing. In particular, an ergodic rotation on the torus is not mixing.
(2) Let T be a hyperbolic matrix in $\mathrm{SL}_2(\mathbb{Z})$. Then the corresponding $\mathbb{Z}$-action on the torus $\mathbb{R}^2/\mathbb{Z}^2$ is mixing. (Hint: consider what happens for characters first.)
(3) Generalise the previous example to automorphisms of d-dimensional torus.

The following theorem implies any ergodic action of the group $G = \mathrm{SL}_d(\mathbb{R})$ is mixing.

Theorem 4.2 (Howe, Moore)**.** *Consider a unitary representation π of $G = \mathrm{SL}_d(\mathbb{R})$ on a Hilbert space $\mathcal{H}$ such that $\mathcal{H}$ contains no nonzero vectors fixed by G. Then for every $v, w \in \mathcal{H}$,*

$$\langle \pi(g) v, w \rangle \to 0 \quad \text{as } g \to \infty.$$

We will use the following notation:

$$\begin{aligned}
K &= \mathrm{SO}(d) = \text{ the orthogonal group,}\\
A &= \{\mathrm{diag}(a_1,\dots,a_d) : \prod_i a_i = 1, a_i > 0\},\\
A^+ &= \{\mathrm{diag}(a_1,\dots,a_d) : \prod_i a_i = 1, a_1 \geq \cdots \geq a_d > 0\},\\
N &= \text{ the group of unipotent upper triangular matrices.}
\end{aligned}$$

We also set

$$k_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad a_t = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, \quad u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

The following decompositions hold:

$$\begin{aligned}
G &= NAK \quad \text{(Iwasawa decomposition)},\\
G &= KA^+K \quad \text{(Cartan decomposition)}.
\end{aligned}$$

Exercise 4.3. Check these decompositions. (Hint: you already know them from a linear algebra course. Review the Gramm-Schmidt orthogonalization and canonical forms of quadratic forms.)

Proof. We prove the theorem by contradiction. Suppose that for some sequence $g_n \to \infty$ and $v, w \in \mathcal{H}$, we have $\langle \pi(g_n)v, w\rangle \nrightarrow 0$. We write $g_n = k_n a_n l_n$ in terms of the Cartan decomposition. Then clearly, $a_n \to \infty$, and after passing to a subsequence, we may assume that $l_n v \to v'$ and $k_n^{-1} w \to w'$ for some $v', w' \in \mathcal{H}$. It is easy to check that

$$\langle \pi(g_n)v, w\rangle - \langle \pi(a_n)v', w'\rangle \to 0.$$

Hence, $\langle \pi(a_n)v', w'\rangle \nrightarrow 0$.

Let's consider $G = \mathrm{SL}_2(\mathbb{R})$. Our intermediate aim is to show that *there exists a nonzero vector fixed by* $\{u_s\}$. From the previous paragraph, we have vectors $v, w \in \mathcal{H}$ and $t_n \to +\infty$ such that $\langle \pi(a_{t_n})v, w\rangle \nrightarrow 0$. After passing to a subsequence, we may assume that $\lim \pi(a_{t_n})v = v'$, where the limit is in the weak! topology. Then $\langle \pi(a_{t_n})v, w\rangle \to \langle v', w\rangle$, so $v' \neq 0$. We have

$$\pi(u_s)v' = \lim \pi(u_s a_{t_n})v = \lim \pi(a_{t_n} u_{s/t_n^2})v.$$

Since

$$\|\pi(a_{t_n} u_{s/t_n^2})v - \pi(a_{t_n})v\| \to 0,$$

it follows that

$$\pi(u_s)v' = \lim \pi(a_{t_n})v = v',$$

which proves the claim.

Exercise 4.4. Generalize this argument to $\mathrm{SL}_d(\mathbb{R})$. Namely, show that there exists a nonzero vector fixed by the subgroups of the form $\begin{pmatrix} id & * \\ 0 & id \end{pmatrix}$.

Now we fix a nonzero vector v which is invariant under $N = \{u_s\}$, and consider the function $\phi(g) = \langle \pi(g)v, v\rangle$, which N-biinvariant. Consider the action of G on $\mathbb{R}^2$. Since $\mathrm{Stab}_G(e_1) = N$ and the action is trivial, we have identification

$$\mathbb{R}^2 - \{0\} = G/N,$$

and ϕ can be consider as an N-invariant function on $\mathbb{R}^2 - \{0\}$. Note the N-orbits in $\mathbb{R}^2$ are the horizontal lines, except the x-axis, and single points on the x-axis. The function ϕ is constant on horizontal lines $y = c$, $c \neq 0$. Hence, by continuity, it is constant on the line $y = 0$. We have

$$\langle \pi(a_t)v, v\rangle = \phi(t, 0) = \phi(0, 0) = \|v\|^2.$$

By the equality case, of the Cauchy-Schwartz inequality, $\pi(a_t)v = v$. This implies that the function ϕ is biinvariant under AN. Since AN has a dense orbit in $\mathbb{R}^2$, ϕ is constant, i.e., $\langle \pi(g)v, v\rangle = \|v\|^2$. As above, this implies that v is G-invariant, which is a contradiction.

Exercise 4.5. Complete the proof for $\mathrm{SL}_d(\mathbb{R})$: using Exercise 4.4, deduce that there exists a nonzero vector which is invariant under copies of $\mathrm{SL}_2(\mathbb{R})$ which generate $\mathrm{SL}_d(\mathbb{R})$.

□

Now we observe that decay of matrix coefficient implies a mean ergodic theorem:

Corollary 4.6. *Consider an ergodic action of $G = \mathrm{SL}_d(\mathbb{R})$ on a probability space (X, μ). Let B_n be a sequence of Borel subsets of G such that $0 < m(B_n) < \infty$ and $m(B_n) \to \infty$. Then for every $f \in L^2(X)$,*

$$\frac{1}{m(B_n)} \int_{B_n} f(g^{-1}x)\, dm(g) \to \int_X f\, d\mu \quad \textit{as } n \to \infty$$

in L^2-norm.

Proof. It suffices to consider a function f with $\int_X f\, d\mu = 0$.

Let $\varepsilon > 0$ and Q be a compact subset of G such that

$$|\langle \pi_X(g)f, f\rangle| < \varepsilon \quad \text{for all } g \notin Q.$$

Then we have

$$\left\| \frac{1}{m(B_n)} \int_{B_n} \pi_X(g) f \, dm(g) \right\|^2 = \frac{1}{m(B_n)^2} \int_{B_n \times B_n} \langle \pi_X(g_2^{-1} g_1) f, f \rangle$$
$$\leq \frac{(m \otimes m)(\{(g_1, g_2) \in B_n \times B_n : \, g_2^{-1} g_1 \in Q\})}{m(B_n)^2} \|f\|^2 + \varepsilon.$$

Making a change of variable $(g_1, g_2) \mapsto (g_1, g_2^{-1} g_1)$, we deduce that

$$(m \otimes m)(\{(g_1, g_2) \in B_n \times B_n : \, g_2^{-1} g_1 \in Q\}) \leq m(B_n) m(Q).$$

Since $m(B_n) \to \infty$, the corollary follows. □

5. Rates of decay of matrix coefficients and Kazhdan property

In this section, we prove a quantitative version of Theorem 4.2. A vector v is called K-finite if the span of Kv is finite. Then we set $d(v) = \dim \langle Kv \rangle$.

Theorem 5.1 (Howe, Tan, Oh)**.** *Let $G = \mathrm{SL}_d(\mathbb{R})$, $d \geq 3$, and π be a unitary representation of G on a Hilbert space $\mathcal{H}$ such that $\mathcal{H}$ contains no nonzero invariant vectors. Then for every K-finite vectors $v, w \in \mathcal{H}$,*

$$|\langle \pi(a) v, w \rangle| \leq c(\varepsilon) d(v)^{1/2} d(w)^{1/2} \|v\| \|w\| \left(\max_{i<j} \frac{a_i}{a_j} \right)^{-1/4+\varepsilon}$$

for every $a \in A^+$ and $\varepsilon > 0$.

We note that the set of K-finite vectors is always dense in the ambient space (this is the Peter–Weyl theorem).

Remark 5.2. With a little more work, the rate can be improved by a factor of two, but the aim of this notes is to give a proof with minimal technology.

We note that an analogue of this theorem fails for $\mathrm{SL}_2(\mathbb{R})$, and there is no uniform general rate in this case. This is related to the fact that $\mathrm{SL}_2(\mathbb{R})$ does not have Kazhdan property (see Corollary 5.12 below). Nonetheless, uniform rates have been established for some special families of representations of $\mathrm{SL}_2(\mathbb{R})$ of number-theoretic significance. This is related to the Selberg conjecture and property τ, which we don't have time to discuss in these lectures.

The proof of Theorem 5.1 uses copies of subgroups $\mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ embedded in $\mathrm{SL}_d(\mathbb{R})$. In fact, the crucial step is the following

Proposition 5.3. *Let π be a unitary representation of the group $G = \mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ on a Hilbert space $\mathcal{H}$. Assume that $\mathcal{H}$ contains no nonzero vectors fixed by $\mathbb{R}^2$. Then for v, w in a dense subset of $\mathcal{H}$ consisting of* $\mathrm{SO}(2)$*-eigenvectors, we have*

$$|\langle \pi(a_t)v, w\rangle| \le c(v,w)\, t^{-1} \quad \text{for every } t > 1.$$

where $a_t = \mathrm{diag}(t, t^{-1})$.

Proof. We identify the space $\hat{\mathbb{R}^2}$ of unitary characters of $\mathbb{R}^2$ with $\mathbb{R}^2$ by setting $\chi_{u,v}(x,y) = e^{i(vx-uy)}$. The group $\mathrm{SL}_2(\mathbb{R})$ acts on $\hat{\mathbb{R}^2}$ by $(g\cdot\chi)(v) = \chi(g^{-1}v)$, and under such identification, this is the standard action on $\mathbb{R}^2$ (note that $\mathrm{SL}_2(\mathbb{R})$ preserves the symplectic form $vx - uy$). Note also that these actions agree with the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$ by conjugations as a subgroup of G.

Although the group of unitary operators $\pi(\mathbb{R}^2)$ may not have a basis consisting of eigenvectors as in the finite-dimensional case, there is a natural substitute — a projection-valued measure P_B, B is a Borel subset of $\mathbb{R}^2$, such that

$$\pi(r) = \int_{\mathbb{R}^2} \chi_z(r) dP_z, \quad r \in \mathbb{R}^2.$$

This equality means that for $v, w \in \mathcal{H}$,

$$\langle \pi(r)v, w\rangle = \int_{\mathbb{R}^2} \chi_z(r)\, d\,\langle P_z v, w\rangle.$$

We have equivarience relation:

$$\pi(g)P_B\pi(g)^{-1} = P_{gB} \quad \text{for Borel } B \subset \mathbb{R}^2. \tag{5.1}$$

For $s > 1$, we consider

$$\Omega_s = \{x \in \mathbb{R}^2 :\, s^{-1} \le \|x\| \le s\}.$$

Note that it follows from (5.1) that P_{Ω_s} commutes if $\pi(k_\theta)$. Since $\mathcal{H}$ has no vectors fixed by $\mathbb{R}^2$, $P_{\{(0,0)\}} = 0$, and by continuity of the measure, $P_{\Omega_s}v \to v$ as $s \to \infty$ for every $v \in \mathbb{H}$. Hence, $\cup_{s>1}\mathrm{Im}(P_{\Omega_s})$ is dense in $\mathcal{H}$. It is a classical result that $\mathrm{SO}(2)$-eigenvectors are dense in any unitary representation. In particular, each space $\mathrm{Im}(P_{\Omega_s})$ contains a dense family of eigenvectors. Hence, it remains to prove the estimate for $v, w \in \mathrm{Im}(P_{\Omega_s})$. Setting $a = \mathrm{diag}(t, t^{-1})$, we have

$$\begin{aligned}
\langle \pi(a)v, w\rangle &= \langle \pi(a)P_{\Omega_s}v, P_{\Omega_s}w\rangle = \langle P_{a\Omega_s}\pi(a)v, P_{\Omega_s}w\rangle \\
&= \langle \pi(a)v, P_{a\Omega_s}P_{\Omega_s}w\rangle = \langle P_{\Omega_s}\pi(a)v, P_{a\Omega_s\cap\Omega_s}w\rangle \\
&= \langle \pi(a)P_{a^{-1}\Omega_s}v, P_{a\Omega_s\cap\Omega_s}w\rangle = \langle \pi(a)P_{a^{-1}\Omega_s}P_{\Omega_s}v, P_{a\Omega_s\cap\Omega_s}w\rangle \\
&= \langle \pi(a)P_{a^{-1}\Omega_s\cap\Omega_s}v, P_{a\Omega_s\cap\Omega_s}w\rangle.
\end{aligned}$$

Hence, by the Cauchy-Schwartz inequality,

$$|\langle \pi(a)v, w\rangle| \leq \|P_{a^{-1}\Omega_s\cap\Omega_s}v\|\|P_{a\Omega_s\cap\Omega_s}w\|. \tag{5.2}$$

Note that the sets $a_t^{-1}\Omega_s\cap\Omega_s$ are contained in the strip around the x-axis of size s/t, so we expect that the norm $\|P_{a^{-1}\Omega_s\cap\Omega_s}v\|$ decay as $t\to\infty$. We prove that this is the case when v is an eigenfunction of $\{k_\theta\}$. Let $\pi(k_\theta)v = e^{2\pi i n\theta}v$. We decompose $\mathbb{R}^2$ into disjoint sectors S_1, $\ldots$, S_m of equal size $\theta = 2\pi/m$. Then

$$\pi(k_\theta)P_{S_i}v = P_{k_\theta S_i}\pi(k_\theta)v = e^{2\pi i n\theta}P_{S_{i+1}}v.$$

This shows that the vectors $P_{S_1}v$, $\ldots$, $P_{S_m}v$ have the same norms. Since we have orthogonal decomposition

$$v = \sum_{i=1}^m P_{S_i}v.$$

It follows that $\|P_{S_i}v\| = \sqrt{\frac{1}{m}}\|v\| = \sqrt{\frac{\theta}{2\pi}}\|v\|$. This implies that

$$\|P_{a^{-1}\Omega_s\cap\Omega_s}v\| \leq \sqrt{\frac{2\sin^{-1}(s^2/t)}{2\pi}}\|v\|,$$

and a similar estimate holds for $\|P_{a\Omega_s\cap\Omega_s}w\|$. Now the proposition follows from (5.2). $\square$

The main draw-back of Proposition 5.3 is that the bound is not explicit in terms of v and w. This problem is rectified in Proposition 5.10 below. The proof will be carried out in several steps. With the help of Proposition 5.3, we show that the tensor square of $\pi|_{\mathrm{SL}_2(\mathbb{R})}$ embeds in a sum of the regular representations, and for the regular representation, we establish explicit estimate in terms of the Harish-Chandra function.

Exercise 5.4. Prove the following formulas for the Haar measure on $G = \mathrm{SL}_2(\mathbb{R})$:

$$\int_G f(g)\,dm(g) = \int_{\mathbb{R}\times(0,\infty)\times[0,2\pi)} f(u_s a_t k_\theta)t^{-2}d\theta\frac{dt}{t}ds,$$

$$\int_G f(g)\,dm(g) = \int_{[0,2\pi)\times\times[1,\infty)\times[0,2\pi)} f(k_{\theta_1}a_t k_{\theta_2})(t^2-1/t^2)d\theta_1\frac{dt}{t}d\theta_2.$$

Exercise 5.5. Prove the formula for the Haar measure on $G = \mathrm{SL}_d(\mathbb{R})$:

$$\int_G f(g)\,dm(g) = \int_{N\times A\times K} f(nak)\Delta(a)dndadk$$

where Δ is the modular function on NA, and dn, da, dk denote Haar measures on corresponding subgroups.

A unitary representation π of a group G is called L^p here if for vectors v, w in a dense subspace, one has $\langle \pi(g)v, w\rangle \in L^p(G)$.

Exercise 5.6. Prove that under the assumptions of Proposition 5.3, the representation $\pi|_{\mathrm{SL}_2(\mathbb{R})}$ is L^p for $p > 2$.

We will need the following characterization of the L^2-representations:

Theorem 5.7 (Godement)**.** *A representation π of a group G is L^2 iff it can be embedded as a subrepresentation of the sum $\oplus_{n=1}^{\infty}\lambda_G$ where λ_G is the regular representation.*

Proof. (sketch) Suppose that there exists a dense subset $\mathcal{H}_0$ of $\mathcal{H}$ such that $\langle v, \pi(g)w\rangle \in L^2(G)$. Without loss of generality, we may assume that $\mathcal{H}_0$ is countable. Consider the map $T : \mathcal{H}_0 \to \oplus_{\mathcal{H}_0} L^2(G)$ defined by

$$Tw = \oplus_{v\in\mathcal{H}_0} \langle v, \pi(g)w\rangle\,, \quad w \in \mathcal{H}_0.$$

It is clear that it is injective, and satisfies the equivarience relation:

$$T \circ \pi(g) = (\oplus_{\mathcal{H}_0}\lambda(g)) \circ T, \quad g \in G.$$

To finish the proof, we need to show that the map T extends to the whole space $\mathcal{H}$... $\square$

A remarkable property of semisimple groups is that the matrix coefficients of K-invariant vectors in the regular representation $L^2(G)$ can be bounded by an explicit decaying function — the *Harish-Chandra function* Ξ, which we now introduce. Recall that for $G = \mathrm{SL}_d(\mathbb{R})$, we have the Iwasawa decomposition $G = NAK$. For $g \in G$, we denote by $\mathbf{a}(g)$ its A-component. Let Δ be the modular function of the upper triangular group NA. The *Harish-Chandra function* is defined by

$$\Xi(g) = \int_K \Delta(\mathbf{a}(kg))^{-1/2}\, dk.$$

Note that Ξ is K-biinvariant.

Exercise 5.8. Prove that the Harish-Chandra function for $\mathrm{SL}_2(\mathbb{R})$ is given by

$$\Xi(a_t) = \frac{1}{2\pi}\int_0^{2\pi} (t^{-2}\cos^2\theta + t^2\sin^2\theta)^{-1/2} d\theta,$$

and for every $\varepsilon > 0$,

$$\Xi(a_t) \le c(\varepsilon)t^{-1+\varepsilon}, \quad t > 1.$$

We note that the asymptotic behavior of the Harish-Chandra function is well-understood. It is known that $\Xi \in L^{2+\varepsilon}(G)$ for every $\varepsilon > 0$, and there are very sharp pointwise bounds on decay of Ξ.

Theorem 5.9. *Let $G = \mathrm{SL}_d(\mathbb{R})$ and $\phi, \psi \in L^2(G)$ be K-eigenfunction. Then*

$$|\,\langle \lambda(g)\phi, \psi\rangle\,| \le \|\phi\|_2 \|\psi\|_2\, \Xi(g), \quad g \in G.$$

Proof. Taking absolute values, we may assume that ϕ and ψ are K-invariant.

The proof is based on the Herz's "principe de majoration". Clearly, we only need to prove the estimate for $g = a \in A$. Using Exercise 5.5 and Fubini theorem, we obtain

$$\begin{aligned}
\langle \lambda(g)\phi, \psi\rangle &= \int_G \phi(ga)\psi(g)dm(g) \\
&= \int_{N\times A\times K} \phi(nbka)\psi(nb)\Delta(b)dndbdk \\
&\le \int_K \left(\int_{N\times A} \phi^2(nbka)\Delta(b)dndb\right)^{1/2} \left(\int_{N\times A} \psi^2(nbk)\Delta(b)dndb\right)^{1/2} dk,
\end{aligned}$$

where the last estimate is the Cauchy-Schwartz inequality in $L^2(NA)$. Since ψ is K-invariant,

$$\int_{N\times A} \psi^2(nbk)\Delta(b)dndb = \int_{N\times A\times K} \psi^2(nbk)\Delta(b)dndbdk = \|\psi\|_2^2.$$

In terms of the Iwasawa decomposition, $ka = \mathbf{n}(ka)\mathbf{a}(ka)\mathbf{k}(ka)$, and

$$nbka = nb\mathbf{n}(ka)\mathbf{a}(ka)\mathbf{k}(ka) = (nb\mathbf{n}(ka)b^{-1})(b\mathbf{a}(ka))\mathbf{k}(ka).$$

Then using invariance of the measures, we get

$$\begin{aligned}
\int_{N\times A} \phi^2(nbka)\Delta(b)dndb &= \int_{N\times A} \phi^2((nb\mathbf{n}(ka)b^{-1})(b\mathbf{a}(ka)))\Delta(b)dndb \\
&= \Delta(\mathbf{a}(ka))^{-1} \int_{N\times A} \phi^2(nb)\Delta(b)dndb \\
&= \Delta(\mathbf{a}(ka))^{-1}\|\phi\|^2.
\end{aligned}$$

This implies the theorem. □

Now we can prove a more explicit form of Proposition 5.3 for SO(2)-finite vectors:

Proposition 5.10. *Let notation be as in Proposition 6. Then for every $v, w \in \mathcal{H}$ which are* SO(2)*-finite,*

$$|\,\langle \pi(a_t)v, w\rangle\,| \le c(\varepsilon)d(v)^{1/2}d(w)^{1/2}\|v\|\|w\|\, t^{-\frac{1}{2}+\varepsilon}, \quad t > 1, \tag{5.3}$$

for every $\epsilon > 0$.

Proof. First, we reduce the proof to the case when v and w are SO(2)-eigenfunctions. Every SO(2)-finite vector v can be written as $v = \sum_{i=1}^{n} v_i$ where v_i's are orthogonal SO(2)-eigenfunctions. Then by the Cauchy–Schwartz inequality,

$$\sum_{i=1}^{n} \|v_i\| = d(v)^{1/2} \left(\sum_{i=1}^{n} \|v_i\|^2 \right)^{1/2} = d(v)^{1/2}\|v\|.$$

Hence, estimate (5.3) follows from the estimate for SO(2)-eigenfunctions by linearity.

It follows from the estimate in Proposition 5.3 and Exercise 5.4 that the matrix coefficients $\langle \pi(g)v, w\rangle$ belong to $L^4(\mathrm{SL}_2(\mathbb{R}))$ for v, w in a dense subspace $\mathcal{H}_0$ of $\mathcal{H}$.

Let $\rho = \pi|_{\mathrm{SL}_2(\mathbb{R})}$. Consider the tensor square $\rho \otimes \rho$ of the representation ρ. For $v_1, v_2, w_1, w_2 \in \mathcal{H}_0$, we have

$$\langle (\rho \otimes \rho)(g)(v_1 \otimes v_2), (w_1 \otimes w_2)\rangle = \langle \rho(g)v_1, w_1\rangle \langle \rho(g)v_2, w_2\rangle,$$

and it follows from the Cauchy-Schwartz inequality that this expression is in $L^2(G)$. Since linear combinations of vectors $v_1 \otimes v_2$ with $v_1, v_2 \in \mathcal{H}_0$ form a dense subspace of $\mathcal{H} \otimes \mathcal{H}$, we conclude that $\rho \otimes \rho$ is an L^2 representation. Hence, by Theorem 5.7, $\rho \otimes \rho$ is a subrepresentation of a direct sum of regular representations. Applying Theorem 5.9 (it is easy to check that it extends to direct sums), we get that for every SO(2)-invariant vectors $v, w \in \mathcal{H}$,

$$|\langle \rho(g)v, w\rangle|^2 \le \|v\|^2\|w\|^2\Xi(g).$$

Now the proposition follows from Exercise 5.8. $\square$

Proof of Theorem 5.1. We use various embedded copies of $\mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ sitting in $\mathrm{SL}_d(\mathbb{R})$. For simplicity, we carry out the computation for the subgroup

$$\begin{pmatrix} \mathrm{SL}_2(\mathbb{R}) & \mathbb{R}^2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Note that $\mathcal{H}$ contains no nonzero $\mathbb{R}^2$-invariant vectors because of Theorem 4.2. Hence, Proposition 5.10 applies. For $a = \mathrm{diag}(a_1, \dots, a_d) \in A^+$, we write $a = a'a''$ where

$$a' = \mathrm{diag}\left(\left(\frac{a_1}{a_2}\right)^{1/2}, \left(\frac{a_2}{a_1}\right)^{1/2}, 1, \dots, 1 \right),$$
$$a'' = \mathrm{diag}((a_1a_2)^{1/2}, (a_1a_2)^{1/2}, a_3, \dots, a_d).$$

Note that a'' commutes with $\mathrm{SL}_2(\mathbb{R})$. In particular, $\pi(a'')v$ is SO(2)-finite, and

$$\dim \langle \mathrm{SO}(2)\pi(a'')v\rangle = \dim \langle \mathrm{SO}(2)v\rangle \le d(v).$$

It is also clear that $\dim \langle \mathrm{SO}(2)w\rangle \le d(w)$. By Proposition 5.10,

$$|\langle \pi(a)v, w\rangle| = |\langle \pi(a')\pi(a'')v, w\rangle| \le c(\varepsilon) d(v)^{1/2} d(w)^{1/2} \|v\|\|w\| \left(\frac{a_1}{a_2}\right)^{-1/4+\varepsilon}$$

for every $\varepsilon > 0$. Using similar estimate for other copies of $\mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ embedded in $\mathrm{SL}_d(\mathbb{R})$, we finally deduce that

$$|\langle \pi(a)v, w\rangle| \le c(\varepsilon) d(v)^{1/2} d(w)^{1/2} \|v\|\|w\| \left(\max_{i<j} \frac{a_i}{a_j}\right)^{-1/4+\varepsilon}$$

for every $\varepsilon > 0$, as required. □

Corollary 5.11. *There exists $p(d) > 0$ such that every unitary representation of $\mathrm{SL}_d(\mathbb{R})$, $d \ge 3$, without fixed vectors is L^p for every $p > p(d)$.*

Proof. This is just a computation. The main ingredient is the estimate from Theorem 5.1. □

A group G is called *Kazhdan group* if every unitary representation of G which contains almost invariant vectors also contains invariant vectors.

This should be compared with the notion of amenable groups. Note that it follows from Theorem 3.7 that if a group is both amenable and has Kazhdan property, then it is compact.

Corollary 5.12. *The group $\mathrm{SL}_d(\mathbb{R})$, $d \ge 3$, has Kazhdan property.*

Proof. Let π be a representation of G on a Hilbert space $\mathcal{H}$ which has no invariant vectors, but has a sequence v_n of almost invariant vectors, namely,

$$\|\pi(g)v_n - v_n\| \to 0$$

uniformly on compact sets. Let $w_n = \int_K \pi(k)v_n\, dk$. Then

$$\|w_n - v_n\| \le \int_K \|\pi(k)v_n - v_n\|\, dk \to 0.$$

This implies that the sequence w_n is also almost invariant and $\|w_n\| \to 1$, but w_n's are K-invariant, so we can apply the estimate from Theorem 5.1 to get a contradiction. For $a \in A^+$,

$$\|\pi(a)w_n - w_n\|^2 = 2 - 2\mathrm{Re}\,\langle \pi(a)w_n, w_n\rangle \ge 2 - 2\sigma(a)\|w_n\|_2^2$$

σ is an explicit decaying function. Since $\|\pi(a)w_n - w_n\| \to 0$ uniformly on compact sets, this is impossible. □

6. Quantitative mean ergodic theorems

In this section, we work with the group $G = \mathrm{SL}_d(\mathbb{R})$, $d \geq 3$. The goal is to prove the quantitative mean ergodic theorem for G (Theorem 6.5 below).

For a finite Borel measures μ and ν on G, we define

$$(\mu * \nu)(B) = (\mu \otimes \nu)(\{(x, y) : xy \in B\})$$

and

$$\mu^*(B) = \mu(B^{-1})$$

where B is a Borel set.

Given a unitary representation π of G on a Hilbert space $\mathcal{H}$, we define the operator $\pi(\mu)$ on $\mathcal{H}$ by

$$\langle \pi(\mu)v, w\rangle := \int_G \langle \pi(g)v, w\rangle \, d\mu(g), \quad v, w \in \mathcal{H}.$$

Exercise 6.1. Check that

(1) $\pi(\mu * \nu) = \pi(\mu) * \pi(\nu)$.
(2) $\pi(\mu^*) = \pi(\mu)^*$.

Theorem 6.2 (Nevo's transfer principle)**.** *Consider a measure-preserving action of G on a probability space (X, μ), and assume that the representation π on $L^2_0(X)$ is L^{4k} for some $k \in \mathbb{N}$, and ν a Borel probability measure on G. Then*

$$\|\pi(\nu)\| \leq 2\|\lambda(\nu)\|^{\frac{1}{2k}}.$$

Proof. We consider the tensor power representation $\pi^{\otimes 2k}$. Note that it is an L^2-representation, and hence, by Theorem 5.7, a subrepresentation of a direct sum of the regular representations λ. In particular, $\|\pi^{\otimes 2k}(\nu)\| \leq \|\lambda(\nu)\|$. For a real-valued function $f \in L^2_0(X)$, we have by the Jensen inequality,

$$\begin{aligned}
\|\pi(\nu)f\|^{4k} &= \left(\int_{G\times G} \langle \pi(g_1)f, \pi(g_2)f\rangle \, d\nu(g_1)d\nu(g_2)\right)^{2k} \\
&= \left(\int_G \langle \pi(g)f, f\rangle \, d(\nu^* * \nu)(g)\right)^{2k} \\
&\le \int_G \langle \pi(g)f, f\rangle^{2k} \, d(\nu^* * \nu)(g) \\
&= \int_G \left\langle \pi^{\otimes 2k}(g)f^{\otimes k}, f^{\otimes k}\right\rangle d(\nu^* * \nu)(g) \\
&= \int_{G\times G} \left\langle \pi^{\otimes 2k}(g_1)f^{\otimes 2k}, \pi^{\otimes 2k}(g_2)f^{\otimes 2k}\right\rangle d\nu(g_1)d\nu(g_2) \\
&= \|\pi^{\otimes 2k}(\nu)f^{\otimes 2k}\|^2 \le \|\pi^{\otimes 2k}(\nu)\|^2 \|f^{\otimes 2k}\|^2 \\
&\le \|\lambda(\nu)\|^2 \|f\|^{4k}.
\end{aligned}$$

For general functions f, we write $f = f_1 + if_2$ where f_1 and f_2 are real-valued, and it follows from the previous estimate that

$$\|\pi(\nu)f\| \le \|\lambda(\nu)\|^{\frac{1}{2k}}(\|f_1\| + \|f_2\|) \le 2\|\lambda(\nu)\|^{\frac{1}{2k}}\|f\|.$$

This implies the theorem. □

Similarly to measures, one defines convolutions of functions:

$$(f_1 * f_2)(x) = \int_G f_1(xy^{-1})f_2(y)\, dm(y).$$

Theorem 6.3 (Kunze–Stein inequality; Cowling)**.** *For $p \in [1, 2)$, $\phi \in L^2(G)$ and $f \in L^1(G) \cap L^p(G)$,*

$$\|\phi * f\|_2 \le c_p \|\phi\|_2 \|f\|_p.$$

Proof. We give a proof in the case when f is K-biinvariant, but some parts of the argument will be general.

We need to check that for every $\psi \in L^2(G)$,

$$\left|\int_G (\phi * f)\psi \, dm\right| \le c_p \|f\|_p \|\phi\|_2 \|\psi\|_2.$$

Using the Cauchy-Schwartz inequality, we obtain

$$
\begin{aligned}
&\left|\int_G \left(\int_G \phi(xy^{-1})f(y)\,dm(y)\right)\psi(x)\,dm(x)\right|\\
\le &\int_G |f(y)| \left(\int_G |\phi(xy^{-1})||\psi(x)|\,dm(x)\right)\,dm(y)\\
= &\int_G |f(y)| \left(\int_K\int_P |\phi(pky^{-1})||\psi(pk)|\,dpdk\right)\,dm(y)\\
\le &\int_G\int_K |f(y)| \left(\int_P |\phi(pky^{-1})|^2 dp\right)^{1/2}\left(\int_P |\psi(pk)|^2\,dp\right)^{1/2}\,dk\,dm(y)\\
= &\int_G |f(y)| \left(\int_K \tilde{\phi}(ky^{-1})\tilde{\psi}(k)dk\right)\,dm(y).
\end{aligned}
$$

where we use the notation:

$$
\tilde{\phi}(g) = \left(\int_P |\phi(pg)|^2\,dp\right)^{1/2} \quad \text{and} \quad \tilde{\psi}(g) = \left(\int_P |\psi(pg)|^2\,dp\right)^{1/2}.
$$

Note that for $n\in N$, $a\in A$, and $g\in G$, we have

$$
\tilde{\phi}(nag) = \Delta(a)^{-1/2}\tilde{\phi}(g) \tag{6.1}
$$

(see the computation in the Herz principe de majoration from the previous lecture).

Note that by the Jensen inequality,

$$
\int_K \tilde{\psi}(k)dk \le \left(\int_K\int_P |\psi(pk)|^2\,dpdk\right)^{1/2} = \|\psi\|_2. \tag{6.2}
$$

Now we assume that f is right K-invariant. Then using invariance of the measures and (6.2), we obtain

$$
\begin{aligned}
&\int_G |f(y)| \left(\int_K \tilde{\phi}(ky^{-1})\tilde{\psi}(k)dk\right)\,dm(y)\\
= &\int_K\int_G |f(yk_1)| \left(\int_K \tilde{\phi}(ky^{-1})\tilde{\psi}(k)dk\right)\,dm(y)dk_1\\
= &\int_G |f(y)| \left(\int_{K\times K} \tilde{\phi}(kk_1y^{-1})\tilde{\psi}(k)\,dkdk_1\right)dm(y)\\
= &\int_G |f(y)| \left(\int_K \tilde{\phi}(k_1y^{-1})dk_1\right)\left(\int_K \tilde{\psi}(k)\,dk\right)dm(y)\\
\le &\int_G |f(y)| \left(\int_K \tilde{\phi}(k_1y^{-1})dk_1\right)dm(y)\cdot\|\psi\|_2.
\end{aligned}
$$

Recall that we use the notation $g = \mathbf{n}(g)\mathbf{a}(g)\mathbf{k}(g)$ for the Iwasawa decomposition.

Next, we assume that f is right K-invariant. Then using invariance of the measures, (6.1) and (6.2),

$$\begin{aligned}
&\int_G |f(y)| \left(\int_K \tilde{\phi}(k_1 y^{-1})dk_1\right) dm(y) \\
&= \int_K \int_G |f(k_2 y)| \left(\int_K \tilde{\phi}(k_1 y^{-1})dk_1\right) dm(y)dk_2 \\
&= \int_G |f(y)| \left(\int_{K\times K} \tilde{\phi}(k_1 y^{-1} k_2)dk_1 dk_2\right) dm(y) \\
&= \int_G |f(y)| \left(\int_{K\times K} \Delta(\mathbf{a}(k_1 y^{-1}))^{-1/2} \tilde{\phi}(\mathbf{k}(k_1 y^{-1})k_2)dk_1 dk_2\right) dm(y) \\
&= \int_G |f(y)| \left(\int_K \Delta(\mathbf{a}(k_1 y^{-1}))^{-1/2} dk_1\right) \left(\int_K \tilde{\phi}(k_2)dk_2\right) dm(y) \\
&\le \int_G |f(y)| \Xi(y^{-1}) dm(y) \cdot \|\phi\|_2.
\end{aligned}$$

To complete the proof, we use that the Harish-Chandra function is in $L^q(G)$ for $q > 2$. Take $p \in [1, 2)$ and let q be the reciprocal of p. Then by the Hölder inequality,

$$\int_G |f(y)| \Xi(y^{-1})\, dm(y) \le \|f\|_p \|\Xi\|_q.$$

This completes the proof.

If the reader does not want to use the fact that $\Xi \in L^q(G)$ for $q > 2$, we also indicate an alternative approach which shows that $\Xi \in L^q(G)$ for sufficiently large q. Then it follows that the Kunze–Stein inequality holds for some $p > 1$, which is sufficient for our purposes.

In fact, one can show that Ξ is a matrix coefficient of so-called quasi-regular representation. This representation is defined on the space of functions F on G such that

$$F(nax) = \Delta(a)^{-1/2} F(x), \quad n \in N,\ a \in A,\ x \in G,$$

where the action is given by

$$(\rho(g)F)(x) = F(xg), \quad g, x \in G,$$

and the scalar product is

$$\langle F_1, F_2 \rangle = \int_K F_1(k) \overline{F_2(k)} dk.$$

For $F(g) = \Delta(\mathbf{a}(g))^{-1/2}$, we have

$$\Xi(g) = \langle \rho(g)F, F \rangle .$$

Hence, Ξ is a matrix coefficient and we can use Corollary 5.11 to conclude that $\Xi \in L^q(G)$ for sufficiently large q. $\square$

Exercise 6.4. (1) Check that for every group, one has

$$\|\phi * f\|_2 \le \|\phi\|_2 \|f\|_1.$$

(2) Prove that if $G = \mathbb{R}^d$, then the Kunze–Stein inequality fails for every $p > 1$.

Let B_t, $t > 0$, be a family of measurable subsets of G such that $0 < m(B_t) < \infty$. We say that the family is *coarsely admissible* if the following conditions hold:

(1) For every compact subset $Q \subset G$, there exists $c > 0$ such that $QB_tQ \subset B_{t+c}$ for every $t > t_0$.
(2) There exists $\alpha > 0$ such that $m(B_{t+1}) \le \alpha\, m(B_t)$ for every $t > t_0$.

For instance, given any norm on the space of matrices $\mathrm{Mat}_d(\mathbb{R})$, one can show that the family of sets

$$B_t = \{g \in G : \log \|g\| < t\}$$

is coarsely admissible.

Now we prove a mean ergodic theorem with a rate, which would be unthinkable in the amenable world.

Theorem 6.5. *Let B_t be a coarsely admissible family of subsets of $G = \mathrm{SL}_d(\mathbb{R})$, $d \ge 3$. Then for every ergodic action of G on a probability space (X, μ) and $f \in L^2(X)$,*

$$\left\| \frac{1}{m(B_t)} \int_{B_t} f(g^{-1}x) - \int_X f\, d\mu \right\|_2 \le c\, m(B_t)^{-\delta} \|f\|_2$$

where $\delta > 0$ depends only on G.

Proof. Consider the probability measure

$$d\nu_t(g) = \frac{1}{m(B_t)} \chi_{B_t}(g) dm(g).$$

We need to show that for the representation π of G on $L^2_0(X)$, we have

$$\|\pi(\nu_t)\| \le c\, m(B_t)^{-\delta}.$$

Since the action of G on (X, μ) is ergodic, the representation G on $L^2_0(X)$ has no fixed vectors. Hence, by Corollary 5.11, it is L^{4k} for some sufficiently large $k \in \mathbb{N}$. Hence, by Theorem 6.2,

$$\|\pi(\nu_t)\| \le 2\|\lambda(\nu_t)\|^{\frac{1}{2k}}.$$

Let's set $\tilde{B}_t = KB_tK$ and consider the probability measure

$$d\tilde{\nu}_t(g) = \frac{1}{m(\tilde{B}_t)}\chi_{\tilde{B}_t}(g)dm(g).$$

Using that the sets are coarsely well-rounded, we deduce that

$$\|\lambda(\nu_t)\| \le \text{const}\, \|\lambda(\tilde{\nu}_t)\|.$$

Hence, we may assume that the sets B_t are K-biinvariant.

Finally, by Theorem 6.3, for $\phi \in L^2(G)$,

$$\begin{aligned}\|\lambda(\nu_t)\phi\|_2 &= \frac{1}{m(B_t)}\|\phi * \chi_{B_t^{-1}}\|_2 \le \frac{1}{m(B_t)}\, c_p\|\phi\|_2\|\chi_{B_t^{-1}}\|_p \\ &= c_p m(B_t)^{-(1-\frac{1}{p})}\|\phi\|_2.\end{aligned}$$

This proves the theorem. □

We note that Theorem 6.5 also hold without the assumption that the family B_t is coarsely admissible, but this requires the general version of the Kunze–Stein inequality, which we haven't proven in these lectures.

7. Application I: counting lattice points

The classical Gauss circle problem asks about the asymptotic of

$$|\{(x, y) \in \mathbb{Z}^2 :\ \sqrt{x^2 + y^2} < t\}|$$

as $t \to \infty$. It is easy to observe in this case by "counting squares" that this number is asymptotic to

$$\text{Area}(\{(x, y) \in \mathbb{Z}^2 :\ \sqrt{x^2 + y^2} < t\}).$$

More generally, we consider a locally compact group G and a discrete subgroup Γ of G such that $m(G/\Gamma) < \infty$. For a family of domains B_t in G, we are interested in the asymptotic of $|\Gamma \cap B_t|$ as $t \to \infty$. It is natural to conjecture that under some regularity conditions on the domains, we should have

$$|\Gamma \cap B_t| \sim m(B_t) \quad \text{as } t \to \infty.$$

We show that this is indeed the case provided that one can establish the mean ergodic theorem for averages along B_t.

In particular, we are interested in the asymptotics of

$$|\{\gamma \in \text{SL}_d(\mathbb{Z}) :\ \|\gamma\| < T\}|$$

for a given norm $\|\cdot\|$ on $\mathrm{Mat}_d(\mathbb{R})$. Note that $\mathrm{SL}_d(\mathbb{Z})$ has finite covolume in $\mathrm{SL}_d(\mathbb{R})$. This is a classical fact proved by Minkowski. It follows from Theorem 7.1 below that this number is asymptotic to the volume

$$\frac{1}{m(\mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z}))} m(\{g \in \mathrm{SL}_d(\mathbb{R}) : \|g\| < T\}).$$

(The asymptotic of the volume can be computed as well: it is $v_0\, T^{d^2-d}$ for some $v_0 > 0$.) We note that all these estimates can be made quantitative.

In fact, the mean ergodic theorem give a solution for the lattice point counting problem in a general group G. We fix a basis of symmetric neighborhoods of identity $\mathcal{O}_\varepsilon$, $\varepsilon > 0$, of G. An increasing family of Borel subsets B_t, $t > 0$, of G with positive measure is *admissible* if there exists $c > 0$ such that

$$\begin{aligned} \mathcal{O}_\varepsilon \cdot B_t \cdot \mathcal{O}_\varepsilon &\subset B_{t+c\varepsilon}, \\ m(B_{t+\varepsilon}) &\leq (1 + c\varepsilon) \cdot m(B_t) \end{aligned}$$

for all $t \geq t_0$ and $0 < \varepsilon < \varepsilon_0$.

Theorem 7.1 (Gorodnik, Nevo)**.** *Let Γ be a discrete subgroup of G such that $m(G/\Gamma) < \infty$. Let B_t be an admissible family of subsets of G, and assume that the mean ergodic theorem holds for the action of G on G/Γ and the averages along B_t's. Then*

$$\lim_{t\to\infty} \frac{|\Gamma \cap B_t|}{m(B_t)} = \frac{1}{m(G/\Gamma)}.$$

Moreover, if the mean ergodic theorem holds with a rate, then

$$\left| |\Gamma \cap B_t| - \frac{m(B_t)}{m(G/\Gamma)} \right| \leq c\, m(B_t)^\rho$$

for some $c > 0$ and $\rho \in (0, 1)$.

Proof. We only prove the first part of the theorem. The proof of the second part is more elaborate, but follows the same idea.

We normalize the Haar measure so that $m(G/\Gamma) = 1$, and consider the functions

$$\chi_\varepsilon = \frac{\chi_{\mathcal{O}_\varepsilon}}{m(\mathcal{O}_\varepsilon)} \quad \text{and} \quad \bar{\chi}_\varepsilon(g\Gamma) = \sum_{\gamma\in\Gamma} \chi_\varepsilon(g\gamma).$$

Note that $\bar{\chi}_\varepsilon$ is a measurable bounded function on G/Γ with compact support, and

$$\int_G \chi_\varepsilon \, dm = \int_{G/\Gamma} \bar{\chi}_\varepsilon \, dm = 1.$$

The main idea of the proof is the following observation: for $x \in \mathcal{O}_\varepsilon$,

$$\int_{B_{t-c\varepsilon}} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g) \le |B_t \cap \Gamma| \le \int_{B_{t+c\varepsilon}} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g). \tag{7.1}$$

To prove this observation, we note that

$$\int_{B_t} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g) \le \sum_{\gamma\in\Gamma} \int_{B_t} \chi_\varepsilon(g^{-1}x\gamma)\, dm(g) = \sum_{\gamma\in\Gamma} \frac{m(x\gamma\mathcal{O}_\varepsilon \cap B_t)}{m(\mathcal{O}_\varepsilon)}.$$

If $\gamma \in B_{t-c\varepsilon}$, then $x\gamma\mathcal{O}_\varepsilon \subset B_t$. Hence,

$$|\Gamma \cap B_{t-c\varepsilon}| = \sum_{\gamma\in\Gamma\cap B_{t-c\varepsilon}} \frac{m(x\gamma\mathcal{O}_\varepsilon \cap B_t)}{m(\mathcal{O}_\varepsilon)} \le \sum_{\gamma\in\Gamma} \frac{m(x\gamma\mathcal{O}_\varepsilon \cap B_t)}{m(\mathcal{O}_\varepsilon)}.$$

On the other hand, if $x\gamma\mathcal{O}_\varepsilon \cap B_t \neq \emptyset$, then $\gamma \in x^{-1}B_t\mathcal{O}_\varepsilon^{-1} \subset B_{t+c\varepsilon}$. Therefore,

$$\sum_{\gamma\in\Gamma} \frac{m(x\gamma\mathcal{O}_\varepsilon \cap B_t)}{m(\mathcal{O}_\varepsilon)} = \sum_{\gamma\in\Gamma\cap B_{t+c\varepsilon}} \frac{m(x\gamma\mathcal{O}_\varepsilon \cap B_t)}{m(\mathcal{O}_\varepsilon)} \le |\Gamma \cap B_{t+c\varepsilon}|.$$

This completes the proof of (7.1).

We use small parameters $\varepsilon, \delta > 0$. By the mean ergodic theorem,

$$\left\| \frac{1}{m(B_t)} \int_{B_t} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g) - 1 \right\|_2 \to 0$$

as $t \to \infty$, and hence,

$$m\left(\left\{ x\Gamma \,:\, \left| \frac{1}{m(B_t)} \int_{B_t} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g) - 1 \right| > \delta \right\} \right) \to 0$$

as $t \to \infty$. In particular, it will be smaller than $m(\mathcal{O}_\varepsilon\Gamma)$ for large t. Hence, there exists $x \in \mathcal{O}_\varepsilon$ such that

$$\left| \frac{1}{m(B_t)} \int_{B_t} \bar{\chi}_\varepsilon(g^{-1}x\Gamma)\, dm(g) - 1 \right| \le \delta.$$

Combining this estimate with (7.1), we obtain that the estimate

$$|\Gamma \cap B_t| \le (1+\delta)m(B_{t+c\varepsilon}) \le (1+\delta)(1+c^2\varepsilon)m(B_t)$$

which holds for all $\varepsilon, \delta > 0$ and $t > t_0(\varepsilon, \delta)$. The lower estimate is proved similarly. $\square$

8. Application II: an equidistribution example

We consider the subset $\Lambda = \mathrm{SO}_d(\mathbb{Z}[1/p])$ of the set $X = \mathrm{SO}_d(\mathbb{R})$ of orthogonal matrices. We assume that the equation

$$x_1^2 + \cdots + x_d^2 = 0$$

has a nontrivial solution modulo p. Then it is a result from number theory that Λ is dense X. We would like to prove a quantitative estimate on the distribution of this dense set.

We denote by Λ_n the subset of matrices with denominators at most p^n and by m the probability invariant measure on $X = \mathrm{SO}_d(\mathbb{R})$.

Theorem 8.1 (Gorodnik, Nevo)**.** *For every $f \in C^1(X)$,*

$$\left| \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} f(\lambda) - \int_X f\, dm \right| \le c\, |\Lambda_n|^{-\kappa}\, \|f\|_{C^1}$$

for some $c, \kappa > 0$.

Proof. (sketch) We consider the action of Λ on X by right multiplication. Ergodicity of this action follows from density of Λ in X. The proof consists of two step:

(1) prove the mean ergodic theorem with rate for the action of Λ on X,
(2) Deduce from convergence in norm pointwise convergence.

More precisely, in step (1), we show that for every $f \in L^2(X)$,

$$\left\| \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} f(x\lambda) - \int_X f\, dm \right\|_2 \le c' \|f\|_2\, |\Lambda_n|^{-\kappa'} \tag{8.1}$$

for some $c', \kappa' > 0$. To prove this, we observe that Λ is a discrete subgroup of finite covolume in $G = \mathrm{SO}_d(\mathbb{Q}_p)$, $\mathbb{Q}_p$ denote the p-adic numbers. Instead of the action of Λ on X, one can consider the action of G on $(G \times X)/\Lambda$ where Λ is embedded diagonally. We show that the mean ergodic theorem for the Λ-action follows from the mean ergodic theorem of G-action (see Gorodnik–Nevo paper for precise statement and proof). The ergodic theorem for G is proved using the method similar to the proof of Theorem 6.5. However, an important point here is that G might not have the Kazhdan property, so the mean ergodic theorem with a rate is not valid for general G-actions. The crucial input in the proof is the deep result of Clozel on property (τ), which in particular says that the action of G on $(G \times X)/\Lambda$ has a spectral gap. This result allows us to prove a mean ergodic theorem with a rate.

Now we explain step (2). We fix invariant metric d on X. Let $\varepsilon, \delta > 0$ be small parameters, which will be chosen later. It follows from (8.1) and Markov inequality that

(8.2)

$$m\left(\left\{x \in X : \left|\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} f(x\lambda) - \int_X f\,dm\right| > \delta\right\}\right)^{1/2} \le \delta^{-1}\cdot c'\|f\|_2\,|\Lambda_n|^{-\kappa'}.$$

Note that

$$m(B_\varepsilon(e)) \ge v_0\varepsilon^D, \quad \varepsilon \in (0,1),$$

where $v_0 > 0$ and $D = \dim X$. We choose $\delta > 0$ so that

$$v_0\varepsilon^D = 2\delta^{-1}\cdot c'\|f\|_2\,|\Lambda_n|^{-\kappa'}.$$

Then by (8.2), there exists $x \in B_\varepsilon(e)$ such that

$$\left|\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} f(x\lambda) - \int_X f\,dm\right| \le \delta.$$

Since the action of Λ is isometric,

$$\left|\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} f(x\lambda) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} f(\lambda)\right| \le \varepsilon\|f\|_{C^1}.$$

Hence, we obtain the estimate

$$\left|\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} f(x\lambda) - \int_X f\,dm\right| \le \frac{2c'\|f\|_2\,|\Lambda_n|^{-\kappa'}}{v_0\varepsilon^D} + \varepsilon\|f\|_{C^1}.$$

Finally, to finish the proof, we take $\varepsilon = |\Lambda_n|^{-\rho}$ for sufficiently small $\rho > 0$. $\square$