Mautner Phenomenon, Mixing Transformations and the Howe-Moore Theorem

D. Blättler, A. Furlong, A. Sandamirskaya

November 2023

1 Mixing transformations

Strong-mixing

Note: For a measure preserving system $(X, \mathscr{B}, \mu, T)$ we have that

T is ergodic $\iff \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-n}B) = \mu(A)\mu(B) \quad \forall A, B \in \mathscr{B}$.
This describes a sort of mixing property of a transformation. There are even stronger notions of mixing, which we will take a look at now:

Definition (strong-mixing):
A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is called strong-mixing if

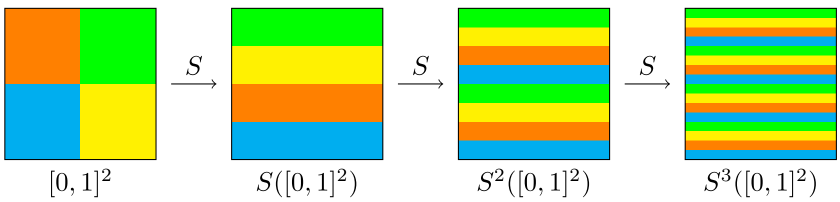
$$\lim_{n\to\infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B) \quad \forall A, B \in \mathscr{B}$$

We can see that strong-mixing implies ergodicity but ergodicity does not imply strong-mixing.

Examples:

- Baker's map
  The map $S : [0,1]^2 \to [0,1]^2$ given by $(x, y) \mapsto \big(2x - \lfloor 2x \rfloor, \frac{1}{2}(y + \lfloor 2x \rfloor)\big)$ is strong-mixing.

  This transformation can be thought of as cutting the square in half veritcally and linearly mapping the left side to the bottom half of the image and the right side to the top half of the image.



- Circle-rotation
  The map $R_\alpha : \mathbb{T} \to \mathbb{T}, x \mapsto x + \alpha \pmod 1$ is ergodic if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ but is not strong-mixing for any α.

  Proof: If $\alpha \in \mathbb{Q}$, then R_α is not ergodic and thus not strong-mixing. So we assume α is irrational. We can approximate any irrational number through a sequence of ration approximations. So we know that $\exists n_1, n_2, ... \in \mathbb{N}$ such that $\lim_{j\to\infty} \alpha n_j \pmod 1 = 0$. In fact the denominators of the rational approximations of α suffice. Now taking $A = B = \left[0, \frac{1}{2}\right]$ we get that

  $\lim_{j\to\infty} \mu\left(A \cap T_\alpha^{-n_j}\right) = \frac{1}{2} \neq \frac{1}{4} = \mu(A)\mu(B)$,
  so the circle-rotation is not strong-mixing. $\square$

There is an even stronger kind of mixing called k-fold mixing which we will introduce:

Definition (k-fold mixing):
A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is called k-fold mixing if

$$\mu(A_0 \cap T^{-n_1} A_1 \cap T^{-n_2} A_2 \cap ... \cap T^{-n_k} A_k) \to \prod_{i=0}^{k} \mu(A_i))$$

as $n_1, n_2 - n_1, n_3 - n_2, ...n_k - n_{k-1} \to \infty$
for any sets $A_0, A_1, ..., A_k \in \mathscr{B}$.

Thus strong-mixing can be thought of as being 1-fold mixing. One of the outstanding open problems in classical ergodic theory is that it is not known if strong-mixing implies k-fold mixing for every $k \geq 1$.

Weak-mixing

It is actually more interesting looking at the following, slightly weaker property.

Definition (weak-mixing):
A measure-preserving system $(X, \mathscr{B}, \mu, T)$ is called weak-mixing if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mu(A \cap T^{-n} B) - \mu(A)\mu(B)| = 0 \quad \forall A, B \in \mathscr{B}$$

If we define $a_n = \mu(A \cap T^{-n} B) - \mu(A)\mu(B)$, we can rewrite our three main mixing properties:

- **Ergodicity:** $\frac{1}{N} \sum_{n=0}^{N-1} a_n \to 0$
- **Weak-mixing:** $\frac{1}{N} \sum_{n=0}^{N-1} |a_n| \to 0$
- **Strong-mixing:** $a_n \to 0$

We can clearly see, that **strong-mixing** $\implies$ **weak-mixing** $\implies$ **ergodicity**.

Examples:

- Finding a transformation that is weak-mixing but strong-mixing is quite difficult. To imagine how such a transformation might work, imagine the mixing sequence $(a_n)_{n\in\mathbb{N}} = (1, \frac{1}{2}, 1, \frac{1}{3}, \frac{1}{3}, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, ...)$, then this does not converge to zero but it's Cesàro sum limits to zero. Such a map would be weak-mixing but not strong-mixing. To see examples, see the paper "Mixing Properties of Substitutions" [Dekking, Keane], or look at the Chacon transformation.

- The circle-rotation map R_α is not weak-mixing. Thus for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, it is an example of an ergodic, but not weak-mixing map. To prove that it is not weak-mixing we will first introduce some equivalencies of weak mixing.

Proposition: The following are equivalent properties for a system $(X, \mathscr{B}, \mu, T)$:

(1) T is weak-mixing.

(2) For any ergodic, measure-preserving system $(Y, \mathscr{B}_Y, \nu, S)$, the system $(X \times Y, \mathscr{B} \otimes \mathscr{B}_Y, \mu \times \nu, T \times S)$ is ergodic.

(3) $T \times T$ is ergodic with respect to $\mu \times \mu$.

(4) The associated operator U_T has no non-constant measurable eigenfunctions.

Proof: (1) $\Longrightarrow$ (2): So assume $(X, \mathscr{B}, \mu, T)$ is weak-mixing amd let $(Y, \mathscr{B}_Y, \nu, S)$ be ergodic.

Choose $A_1, B_1 \in \mathscr{B}$ and $A_2, B_2 \in \mathscr{B}_Y$. We can then calculate

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (\mu \times \nu) \left(A_1 \times A_2 \cap (T \times S)^{-n}(B_1 \times B_2)\right)$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A_1 \cap T^{-n}B_1)\nu(A_2 \cap S^{-n}B_2)$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left(\mu(A_1 \cap T^{-n}B_1) - \mu(A_1)\mu(B_1) + \mu(A_1)\mu(B_1)\right) \nu(A_2 \cap S^{-n}B_2)$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left(\mu(A_1 \cap T^{-n}B_1) - \mu(A_1)\mu(B_1)\right) \nu(A_2 \cap S^{-n}B_2)$$
$$+ \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A_1)\mu(B_1)\nu(A_2 \cap S^{-n}B_2) = S_1 + S_2$$

Because $(Y, \mathscr{B}_Y, \nu)$ is a probability space $\nu(A_2 \cap S^{-n}B_2) \leq 1$ and thus $|S_1|$ is dominated as follows:

$$
\begin{aligned}
|S_1| &\leq \lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |(\mu(A_1 \cap T^{-n}B_1) - \mu(A_1)\mu(B_1))\, \nu(A_2 \cap S^{-n}B_2)| \\
&\leq \lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |\mu(A_1 \cap T^{-n}B_1) - \mu(A_1)\mu(B_1)| = 0
\end{aligned}
$$

Where the last limit is goes to 0 because T is weak-mixing. So we have $S_1 = 0$.

We can calculate S_2 using the property that S is ergodic.

$$
\begin{aligned}
S_2 &= \mu(A_1)\mu(B_1) \cdot \lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} \nu(A_2 \cap S^{-n}B_2) \\
&= \mu(A_1)\mu(B_1)\nu(A_2)\nu(B_2) = (\mu \times \nu)(A_1 \times A_2)(\mu \times \nu)(B_1 \times B_2)
\end{aligned}
$$

We've shown that

$$
\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} (\mu \times \nu)\left(A_1 \times A_2 \cap (T \times S)^{-n}(B_1 \times B_2)\right)
$$
$$
= (\mu \times \nu)(A_1 \times A_2)(\mu \times \nu)(B_1 \times B_2)
$$

meaning that $(X \times Y, \mathscr{B} \otimes \mathscr{B}_Y, \mu \times \nu, T \times S)$ is ergodic. $\square$

(2) $\Longrightarrow$ (3):

We'll take $(Y{=}\{y\}, \mathscr{B}_Y, \nu, S{=}\mathrm{id})$ to be the identity on the singleton. Then $(Y, \mathscr{B}_Y, \nu, S)$ is ergodic and thus we can use (2), which gives us that $T \times S$ is ergodic. But because $(T \times S)$ is isomorphic to T, this means that T itself is ergodic.

We can now use (2) again, this time with T as the ergodic system and we get that $T \times T$ is ergodic, which is what we wanted to show. $\square$

(3) $\Longrightarrow$ (4): Let f be a measurable eigenfunction of T, which we will show has to be constant.

Meaning that $f \circ T = \lambda f$ for some $\lambda \in \mathbb{S}^1$.

(Note that $\lambda \in \mathbb{S}^1$ because U_T is an isometry of L^2_μ)

We now define the measurable function g as $g(x_1, x_2) = f(x_1)\overline{f(x_2)}$.

We can see that g is $T \times T$-invariant by calculating

$$
\begin{aligned}
(g \circ T \times T)(x_1, x_2) &= g(T(x_1), T(x_2)) \\
&= f(T(x_1))\overline{f(T(x_2))} \\
&= \lambda f(x_1)\overline{\lambda f(x_2)} \\
&= \lambda\overline{\lambda} f(x_1)\overline{f(x_2)} \\
&= f(x_1)\overline{f(x_2)} = g(x_1, x_2)
\end{aligned}
$$

Because $T \times T$ is ergodic and g is $T \times T$-invariant, we know that g must be constant almost everywhere. Thus f is also constant almost everywhere. $\square$

(4) $\Longrightarrow$ (1):

The Definition of weak-mixing is equivalent to the property that

$$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, g\rangle - \langle f, 1\rangle\langle 1, g\rangle| = 0 \quad \text{for any } f, g \in L^2_\mu$$

Using the polarisation identity we get that this is equivalent to

$$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle - \langle f, 1\rangle\langle 1, f\rangle| = 0 \quad \text{for any } f \in L^2_\mu$$

By subtracting $\int_X f \mathrm{d}\mu$ from f, it is therefore enough to show that $\int_X f \mathrm{d}\mu = 0$, then

$$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle| = 0$$

which is equivalent to

$$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, f\rangle|^2 = 0$$

By the spectral theorem, it is sufficient to show, that for the non-atomic measure μ_f on $\mathbb{S}^1$, we have

$$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n=0}^{N-1} \left| \int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z) \right|^2 = 0$$

We will now calculate this limit.

$$\begin{aligned}
\lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1}\left|\int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z)\right|^2 &= \lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1}\left(\int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z)\right)^2 \\
&= \lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1}\left(\int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z) \cdot \int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z)\right) \\
&= \lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1}\left(\int_{\mathbb{S}^1} z^n \mathrm{d}\mu_f(z) \cdot \int_{\mathbb{S}^1} w^{-n} \mathrm{d}\mu_f(w)\right) \\
&= \lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1}\left(\int_{\mathbb{S}^1\times\mathbb{S}^1} \left(\frac{z}{w}\right)^n \mathrm{d}\mu_f^2(z,w)\right) \quad \text{(by Fubini)} \\
&= \lim_{N\to\infty} \left(\int_{\mathbb{S}^1\times\mathbb{S}^1} \tfrac{1}{N}\sum_{n=0}^{N-1} \left(\frac{z}{w}\right)^n\right) \mathrm{d}\mu_f^2(z,w)
\end{aligned}$$

Because the measure μ_f is non-atomic, the diagonal set $\{(z,z)|z \in \mathbb{S}^1\} \subseteq \mathbb{S}^1 \times \mathbb{S}^1$ has zero μ_f^2-measure. So we just have to take a look at when $z \neq w$ for which we have

$$\frac{1}{N}\sum_{n=0}^{N-1}\left(\frac{z}{w}\right)^n = \frac{1}{N}\cdot\frac{1-(\frac{z}{w})^N}{1-\frac{z}{w}} \to 0 \qquad \text{as } N \to \infty$$

So by the dominated convergence theorem, we have that

$$\lim_{N\to\infty} \left(\int_{\mathbb{S}^1\times\mathbb{S}^1} \tfrac{1}{N}\sum_{n=0}^{N-1} \left(\frac{z}{w}\right)^n\right) \mathrm{d}\mu_f^2(z,w) = 0$$

meaning that indeed T is weak-mixing. $\square$

We can now prove that the circle-rotation R_α is not weak-mixing.

Proof: Define the function $f : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, $(x, y) \mapsto \mathrm{e}^{2\pi i(x-y)}$.
We can then calculate that

$$\begin{aligned}
(f \circ R_\alpha \times R_\alpha)(x,y) = f(R_\alpha(x), R_\alpha(y)) &= f(x+\alpha \pmod 1, y+\alpha \pmod 1) \\
&= \mathrm{e}^{2\pi i((x+\alpha \pmod 1)-(y+\alpha \pmod 1))} = \mathrm{e}^{2\pi i(x+\alpha-\lfloor x+\alpha\rfloor-y+\alpha+\lfloor y+\alpha\rfloor)} \\
&= \mathrm{e}^{2\pi i(x-y)} \cdot \mathrm{e}^{2\pi i\lfloor x+\alpha\rfloor} \cdot \mathrm{e}^{2\pi i\lfloor y+\alpha\rfloor} = f(x,y)\cdot 1\cdot 1 = f(x,y)
\end{aligned}$$

We see that f is $R_\alpha \times R_\alpha$-invariant. If $R_\alpha \times R_\alpha$ were ergodic, f would need to be constant almost everywhere, but it is not. Thus $R_\alpha \times R_\alpha$ mustn't be ergodic.

Thus by equivalence (3), the map R_α cannot be weak-mixing. $\square$

2 The action of $SL_2(\mathbb{R})$ is mixing

We have already seen that the action of $SL_2(\mathbb{R})$ onto a quotient space is ergodic. In this section we want to prove that it is furthermore mixing.

Theorem: Let $\Gamma \leqslant SL_2(\mathbb{R})$ be a lattice. Then the action of $SL_2(\mathbb{R})$ onto the quotient space $X = \Gamma \backslash SL_2(\mathbb{R})$ is mixing with respect to the Haar measure.

Two prove this theorem we will need a proposition and a lemma, but first we define the following things:

Definition: A matrix $M \in GL_k(\mathbb{R})$ is called **unipotent**, if $\exists k \in \mathbb{N}$ such that $(M - \mathbb{I}_k)^k = 0$. This is equivalent to all eigenvalues of M being 1.
We say a sequence of elements v_n in a Hilbert space $\mathcal{H}$ **converges weak*** to v, if $\forall w \in \mathcal{H}$ we have $\langle v_n, w\rangle \to \langle v, w\rangle$ as $n \to \infty$. Note that this correspond to pointwise convergence in the dual space $\mathcal{H}^*$.
For a group G and a sequence $\alpha = (a_n)_{n\in\mathbb{N}}$ of elements in G, define the set $S(\alpha) = \{g \in G : e \in \overline{\{a_n^{-1} g a_n : n \in \mathbb{N}\}}\}$, where e is the identity element in G. One can think of it in terms of limit points. Id est, as the $g \in G$, for which there exists a subsequence $(a_{n_k})_{k\in\mathbb{N}}$ such that $a_{n_k}^{-1} g a_{n_k} \to e$.
We also say α **converges to** ∞ if it "leaves compact subsets", that is $\forall K \subseteq G$ compact, only finitely for many $n \in \mathbb{N}$ we have $a_n \in K$.

Proposition: Let G be a locally compact group, $\alpha = (a_n)_{n\in\mathbb{N}}$ a sequence of elements in G. Let $\mathcal{H}$ be a Hilbert space carrying a unitary representation $\rho : G \to GL(\mathcal{H}$. Suppose $\exists v \in \mathcal{H}$ such that the sequence $\rho(a_n)v \in \mathcal{H}$ converges weak* to some $v_0 \in \mathcal{H}$. Then $\forall g \in \overline{S(\alpha)} : gv_0 = v_0$.

Proof. It suffices to show that $\forall g \in S(\alpha) : gv_0 = v_0$. Let $g \in G$ and $(a_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} a_{n_k}^{-1} g a_{n_k} = e$. Then by the definition of weak* convergence we have $\forall w \in \mathcal{H}$:

$$\langle gv_0, w\rangle = \langle v_0, g^{-1}w\rangle = \lim_{k\to\infty} \langle a_{n_k} v, g^{-1} w\rangle = \lim_{k\to\infty} \langle g a_{n_k} v, w\rangle$$

In particular also for g = e. Computing the absolute value yields: $\forall w \in \mathcal{H}$

$$|\langle gv_0, w\rangle - \langle v_0, w\rangle| = \lim_{k\to\infty} |\langle g a_{n_k} v, w\rangle - \langle v, w\rangle| = \lim_{k\to\infty} |\langle a_{n_k}^{-1} g a_{n_k} v, w\rangle - \langle a_{n_k}^{-1} v, w\rangle|$$

$$\leq \lim_{k\to\infty} \|(a_{n_k}^{-1} g a_{n_k})v - v\| \|w\| = \|v - v\| \|w\| = 0$$

Since this hold for all $w \in \mathcal{H}$, it hence follows that $gv_0 = v_0$. □

To be able to apply this, we need to find the non-trivial element of $S(\alpha)$. The following lemma gives them for $G = SL_2(\mathbb{R})$:

Lemma: Let $\alpha = (g_n)_{n\in\mathbb{N}}$ be a sequence of elements in $SL_2(\mathbb{R})$ converging to ∞. Then the set $S(\alpha)$ contains a non-trivial unipotent element.

Proof. Consider the map $\Phi : SL_2(\mathbb{R}) \to GL(Mat_{2\times2}(\mathbb{R}))$ given by $\Phi(g)v = gvg^{-1}$, $g \in SL_2(\mathbb{R}), v \in Mat_{2\times2}(\mathbb{R})$. This is a proper linear map, thus if $g_n \to \infty$ then $\|\Phi(g)\| \to \infty$. Observe that we have two invariant subspaces under this map, whose direct sum is the whole space:

$$\mathbb{R}\mathbb{I}_2, \quad B = \{X \in Mat_{2\times2}(\mathbb{R}) : \text{trace}(X) = 0\}$$

Note that actually $B = sl_2(\mathbb{R})$, but this is not important here. The group action is trivial under the first one. Thus, there exists a sequence $(v_n)_{n\in\mathbb{N}}$ in B such that $\|v_n\| \to 0$ but $\|\Phi(g_n)v_n\| = c > 0$, $\forall n \in \mathbb{N}$, where c is small enough such that the exponential map is injective on a ball of radius 2c. We apply exp onto the sequence to get a new sequence $h_n = \exp(v_n) \to \exp(0) = \mathbb{I}$ and $g_n h_n g_n^{-1} \to u \neq \mathbb{I}$ (because $\|\Phi(g_n)h_n\| = \exp(c) \neq 1$). In particular, all the eigenvalues of h_n go to 1. Since eigenvalues are invariant under conjugation, all eigenvalues of $g_n h_n g_n^{-1}$ also all go to 1, i.e. the eigenvalues of u are all 1. Therefore, u is unipotent and it lies in $S(\alpha)$. $\square$

Now we have all the ingredients we need to prove the theorem.

Proof. Let $\alpha = (a_n)_{n\in\mathbb{N}}$ be a sequence of elements in $SL_2(\mathbb{R})$ converging to ∞. Recall that $\rho : SL_2(\mathbb{R})\times X \to X$ induces an action on $L^2(X)$ via $f \mapsto f\circ\rho(g)$. Let $f \in L^2(X)$.
Claim: $f \circ \rho(a_n)$ converges weak* to $\int f \mathrm{d}m_X$
If the claim holds, we have:

$$\langle f\circ\rho(a_n), g\rangle \to \left\langle \int f\mathrm{d}m_X, g\right\rangle = \int\int f\mathrm{d}m_X\overline{g}\mathrm{d}m_X = \int f\mathrm{d}m_X \int \overline{g}\mathrm{d}m_X = \langle f, 1\rangle\langle 1, \overline{g}\rangle$$

Where we used that $\int f\mathrm{d}m_x$ is constant with respect to the other integral. The statement $\langle f \circ \rho(a_n), g\rangle \to= \langle f, 1\rangle\langle 1, \overline{g}\rangle$ is exactly equivalent to mixing. Now all that is left to do is to prove the claim.
Observe that $\|f \circ \rho(a_n)\| = \|f\|$ since ρ is unitary. The Banach-Alaoglu theorem implies that $f\circ\rho(a_n)$ has a weak* convergent subsequence, say $f\circ\rho(a_{n_k}) \to f_0 \in L^2(\mathbb{R})$ and call αa_{n_k}. The above lemma implies that there exists a non-trivial unipotent element $u \in S(\alpha)$; and from the proposition it follows $\forall a \in S(\alpha) : f_0 \circ \rho(a) = f_0$. In particular we have that $f_0 \circ u = f_0$. Note that u is conjugate to an element of U in $SL_2(\mathbb{R})$ and recall that the action of such elements is ergodic. Furthermore, recall the implication "invariant under ergodic $\Rightarrow$ constant. Hence, f_0 is constant, and using the pointwise ergodic theorem we can write $f_0 = \int f dm_X$. This proves the claim and hence the theorem. $\square$

In the next section we will prove the following interesting result:

Theorem: (Vanishing of matrix coefficients) Let $\mathcal{H}$ be a Hilbert space equipped with a unitary representation of $SL_2(\mathbb{R})$, without any non-trivial invariant vectors. Then $\forall v, w \in \mathcal{H}$ we have that the "matrix coefficients" $\langle gv, w\rangle$ where $g \in SL_2(\mathbb{R})$ vanish at ∞. Id est if $g_n \to \infty$, then $\langle g_n v, w\rangle \to 0$. The reason for the name "matrix coefficients" is that in the finite dimensional case these $\langle gv, w\rangle$ terms for basis vectors, exactly correspond to the coefficients in the Gram-Matrix.

3 Howe-Moore and Applications

Firstly, two key facts about Hilbert spaces:
- For every linear functional $f \in \mathcal{H}^* \exists! v = v_f$ such that $\forall v \in \mathcal{H} f(u) =< v, u >$.
- Weak-* topology on H^* : $f_n, f \in \mathcal{H}^*$ say that $f_n \to f$ weak-* if $\forall v \in \mathcal{H}, f_n(v) \to f(v)$. Banach-Alaoglu: The unit ball in $\mathcal{H}^*$ is compact with respect to the weak-* topology.

Question If $g \in G$, does the operator $T_g : X \to X$ given by $T_g(x) = xg^{-1}$ act ergodically on X ?

Astonishingly, as long as the closed subgroup generated by g is not compact, the answer to the Question is yes. This ergodicity will play a key role in our later results. The solution uses the Howe-Moore theorem, which we prove in this section.
Definition A sequence of elements $(g_n)_{n\in\mathbb{N}}$ in G is said to tend to infinity if for every compact set $K \subseteq G$, there's some N so that if $n > N$ then $g_n \notin K$.
With Howe-Moore, we can answer the Question as follows. Take $\mathcal{H} = L^2(X)/V$, for V the space of ae constant functions in $L^2(X)$, and for $g \in G, f \in L^2(X)$, set $\pi(g)(f) : X \to \mathbb{R}$ to be the function given by $[\pi(g)(f)](x) = f(gx)$. Since T_g preserves the measure m_X, π is a unitary representation of G. Note π is continuous in the strong operator topology, which can be checked by density of $C_c(X)$ in $L^2(X)$, for if $g_n \to g$ and $\phi \in C_c(X)$, then using the uniform continuity of ϕ we can see $\pi(g_n)\phi \to \pi(g)\phi$. Take a simple factor G_i of G. Since G has no compact factors and Γ is irreducible, we find that ΓG_i is a dense subgroup.

Mixing can also be translated to a spectral property of the Koopman-von Neumann representation. If 1 is a generator of $\mathbb{Z}$, then the $\mathbb{Z}$-action is mixing iff $\lim_{n\to+\infty}\langle\pi(n)f, g\rangle = 0$ for all $f, g \in L^2_0(X, \mu_X)$. This former expression defines a matrix coefficient of π.

Definition: Let $\pi : G \to \mathrm{GL}(\mathcal{H})$ be a representation. Given $v, w \in \mathcal{H}$, the function $g \in G \mapsto f_{v,w}(g) = \langle\pi(g)v, w\rangle$ is called a matrix coefficient of π.
Using the analogy of actions and their Koopman-von Neumann representations, we define notions of ergodicity and mixing for a representation. Given a representation $\pi : G \to \mathrm{GL}(\mathcal{H})$, we say that $v \in \mathcal{H}$ is invariant if $\pi(g)v = v$ for all $g \in G$.
Definition π is called ergodic if 0 is the only invariant vector. π is called mixing if its matrix coefficients vanish at infinity: for every $v, w \in \mathcal{H}$ we have $\lim_{g\to\infty} f_{v,w}(g) = 0$. As above, by $\lim_{g\to\infty} f_{v,w}(g) = 0$ we mean the following: for every $\varepsilon > 0$ there is a compact $K \subset G$ such that $|f_{v,w}(g)| < \varepsilon$ for all $g \notin K$.

Fix a unitary representation $\pi : G \to \mathcal{U}(V)$.
Lemma (Identifying Fixed Vectors). Let $v \in V$ be a vector. Then its stabilizer $\operatorname{Stab}_v G$ can be identified with

$$\operatorname{Stab}_v G = \left\{g : f_{v,v} = \|v\|^2\right\}$$

Moreover the matrix coefficient satisfies $|f_{v,v}| \leq \|v\|^2$ pointwise.
Proof. That pointwise inequality $|f_{v,v}| \leq \|v\|^2$ follows by Cauchy-Schwarz. It also implies that if equality holds, then $gv = cv$ for some $c \in \mathbb{C}$ with $|c| = 1$. Moreover

$$\|gv - v\|^2 = 2 \cdot \left(\|v\|^2 - \operatorname{Re}\langle gv, v\rangle\right) = 2\left(\|v\|^2 - \operatorname{Re} f_{v,v}(g)\right)$$

and the equality case of Cauchy-Schwarz implies the claim.
Recall Mautner's lemma from last week:
Lemma (Mautner). Suppose that $\pi : G \to \mathcal{U}(V)$ is a unitary representation of a topological group. Suppose that $v \in V$ and $s_n, s_n' \in G$ are stabilizing v, and moreover there exists a sequence $g_n \to g$ such that

$$\lim s_n g_n s_n' = \mathbf{1} \quad \text{ in } G$$

Then g also stabilizes v.
Proof. By the Lemma, the stabilizer of v can be identified with the set of $h \in G$ such that $f_{v,v}(h) = \|v\|^2$. One can then take the limit in

$$f_{v,v}(g) = \lim_n f_{v,v}(g_n) = \lim_n f_{v,v}(s_n g_n s_n') = \|v\|^2$$

since $f_{v,v}$ is a continuous function.

Another important and natural question that one might ask is the following: given an ergodic G-space X and a subgroup H, when is H ergodic on X ? For semisimple Lie groups, Moore's Ergodicity Theorem gives a complete answer to the above question.
Theorem (Howe-Moore 1: Lie Group version). Let G be a connected real semisimple Lie group, with finite center (Note: the center of $SL_2(\mathbb{R}) = \pm I$). Suppose that V is a unitary representation of G whose restriction to any non-compact simple factor has no nontrivial invariant vector. Then all matrix coefficients vanish at infinity, i.e. $\forall v, w \in V$:

$$\lim_{g \to \infty} \langle gv, w \rangle \to 0$$

where $g \to \infty$ means that g leaves every compact set in G.
The property $\langle gv, w \rangle \to 0$ is also called mixing, or decay of correlations.
Corollary (Ergodicity of actions). With the same assumptions as above, suppose that G acts on X preserving a probability measure μ, and the action is ergodic. Then any non-compact subgroup also acts ergodically on (X, μ).
One example is G acting on G/Γ, where ergodicity of the full G action is immediate and thus implies ergodicity (in fact mixing) of any 1-parameter subgroup.

We will first prove Theorem Howe-Moore for the case of $\mathbf{SL}_2\mathbb{R}$. For some notation, consider the subgroups

$$A := \left\{ a_t := \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix} \right\} U^+ = \left\{ u_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \right\} U^- = \left\{ u_s^- = \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \right\}$$

Proposition (Invariant vectors for $\mathbf{SL}_2$). Suppose that V is a unitary representation of $\mathbf{SL}_2\mathbb{R}$.
If v is a vector that is A, or U^+, or U^--invariant, then it is $\mathbf{SL}_2\mathbb{R}-$ invariant.
Proof. The proof is based on the Mautner Lemma.
For the case U^+-invariant implies A-invariant, take

$$s_1 = \frac{1-t}{\varepsilon}, s_2 = \frac{1-t^{-1}}{\varepsilon}$$

and compute

$$\begin{bmatrix} 1 & s_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} t & 0 \\ \varepsilon & t^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & s_2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 \end{bmatrix}$$

so as $\varepsilon \to 0$ we conclude that any U^+-invariant vector is also A-invariant, again by the Mautner lemma.

Theorem (Howe-Moore for $\mathbf{SL}_2$). Suppose that V is a unitary representation of $\mathbf{SL}_2\mathbb{R}$.

Then either V has an invariant vector, or it is mixing, i.e. $\forall v, w \in V$

$$\lim_{g \to \infty} \langle gv, w \rangle \to 0$$

Proof. Suppose by contradiction that there exists v, w and a sequence $g_i \to \infty$ such that the matrix coefficient does not go to 0, e.g. $\operatorname{Re}\langle g_i v, w\rangle \geq \varepsilon > 0$. Let $g_i = k_{i,1} a_i k_{i,2}$ be the KAK decomposition of g_i (this is what is generally called Cartan decomposition, in this case it is just the SVD decomposition, which is essentially polar coordinates "rotate stretch rotate"). By passing to a subsequence, assume that $k_{i,1} \to k_1$ and $k_{i,2} \to k_2$, since K is compact. It follows that

$$\varepsilon \leq \liminf \langle k_1 \cdot a_i \cdot k_2 \cdot v, w \rangle = \liminf \left\langle a_i \cdot (k_2 v), k_1^{-1} w \right\rangle$$

so up to replacing v, w, we can assume $\operatorname{Re}\langle a_i v, w\rangle \geq \varepsilon > 0$. Now $a_i v$ is a bounded sequence of vectors, so let v_0 be some weak limit, which is given by Banach-Alaoglu (closed unit ball is weak* compact). Then $v_0 \neq 0$ since $\operatorname{Re}\langle v_0, w\rangle \geq 0$ by construction. It suffices to check that v_0 is U^+-invariant, since by the Proposition it will be $\mathbf{SL}_2\mathbb{R}$-invariant. For $u \in U^+$, we have

$$\|uv_0 - v_0\| \leq \limsup_i \left\|a_i^{-1}(ua_i v - a_i v)\right\| = \|v - v\| = 0$$

where we used that $a_i^{-1} u a_i \to \mathbf{1}$ as $i \to \infty$.
So we are in the invariant vector case. The equivalence between vanishing coefficients and mixing is simple by taking the functional form of mixing using L^2 functions.

A similar proof can be used for general semisimple Lie group, which also have their own Cartan decompositions.

Q: Why does this correspond to mixing?

The probability measure space $(X, \mu) = \left(G/\Gamma, m_{G/\Gamma}\right)$ has a transitive measure-preserving action of the simple Lie group G by translations: $g : h\Gamma \mapsto gh\Gamma$. Consider the orthogonal complement in $L^2(X, \mu)$ to constant functions:

$$\mathcal{H} = L_0^2(X, \mu) = \left\{ f \in L^2(X, \mu) \mid \int f \mathrm{d}\mu = 0 \right\}$$

Then $\pi : G \to U(\mathcal{H})$, given by $(\pi(g)f)(h\Gamma) = f\left(g^{-1}h\Gamma\right)$, is a unitary representation without non-trivial invariant vectors (exercise!). By Howe-Moore, for any $f_1, f_2 \in \mathcal{H}$

$$\lim_{n \to \infty} \langle \pi(\alpha^n) f_1, f_2 \rangle = 0$$

We claim that this corresponds to the property of mixing for T on (X, μ) that independence of $T^{-n}B$ from A :

$$\lim_{n \to \infty} \mu\left(A \cap T^{-n}B\right) = \mu(A) \cdot \mu(B)$$

Indeed, given a measurable subset $E \subset X$ the projection f_E of the characteristic function 1_E to $\mathcal{H} = L^2_0(X, \mu)$ is

$$f_E(x) = 1_E(x) - \mu(E) = (1 - \mu(E)) \cdot 1_E(x) + (-\mu(E)) \cdot 1_{X \setminus E}(x).$$

One calculates

$$\langle f_A, f_B \rangle = \mu(A \cap B) - \mu(A)\mu(B)$$

and

$$\langle \pi(\alpha^n) f_A, f_B \rangle = \langle f_A, f_{T^{-n}B} \rangle = \mu\left(A \cap T^{-n}B\right) - \mu(A)\mu(B)$$

Finally, mixing implies ergodicity, because any set E with $\mu\left(E \triangle T^{-1}E\right) = 0$ would have

$$\mu(E) = \mu\left(E \cap T^{-n}E\right) \to \mu(E)^2$$

which is possible only if $\mu(E) = 0$ or $\mu(E) = 1$.

Theorem(Howe-Moore 0). Suppose G is a simple connected Lie group with finite center. Γ is a lattice in G . Then $H \curvearrowright G/\Gamma$ is mixing (and ergodic) if $\bar{H}$ is not compact.

Mixing for G-action $\Rightarrow$ Mixing for any subgroup $H \subseteq G$ $\Rightarrow$ Ergodicity of any unbounded subgroup H. So mixing has a property that it passes to subgroups.
This is certainly not true for ergodicity. Take any $\mathbb{R} \curvearrowright X$ and any homomorphism $\mathbb{R}^2 \to \mathbb{R}$.
Theorem(Howe-Moore 1). Suppose G is a simple connected Lie group with finite center. Then any ergodic action of G on a propabiblity space is mixing.

Vanishing of matrix coefficients implies Howe-Moore 1 . Let $L^2(X\mu)$ and take $\mathcal{H} = 1^\perp = L^2_0(X, \mu)$ to be the zero mean functions, where 1 is the constant function $f \equiv 1$. The orthogonal projection is $f \to f - \int_X f \mathrm{d}\mu$. By ergodicity $\mathcal{H}$ has no nonzero G-fixed vectors. Now we get $\forall f_1, f_2 \in \mathcal{H} < gf_1, f_2 > \to_{g \to \infty} 0$. If $f_1, f_2 \in L^2(X, \mu)$, $\bar{f}_1, \bar{f}_2$ are the orthogonal projections on $\mathcal{H}$ then

$$\begin{aligned}
< gf_1, f_2 > &=< g\left(\bar{f}_1 + \int f_1 \mathrm{d}\mu\right), \bar{f}_2 + \int f_2 \mathrm{d}\mu > \\
&=< g\bar{f}_1, \bar{f}_2 > + < g \int f_1 \mathrm{d}\mu, \bar{f}_2 > + < g\bar{f}_1, \int f_2 \mathrm{d}\mu > + < g \int f_1 \mathrm{d}\mu, \int f_2 \mathrm{d}\mu > \\
&= \int f_1 \mathrm{d}\mu \int f_2 \mathrm{d}\mu < g1, 1 >= \int f_1 \mathrm{d}\mu \int f_2 \mathrm{d}\mu
\end{aligned}$$

Why can we rule out the case of a dense compact subgroup?

Proposition If $D \leq G$ is a dense subgroup, then the action of D on X is ergodic.

Proof. For $f \in L^2(X)$ define its stabilizer as

$$\mathrm{Stab}(f) = \{g \in G \mid \pi(g)(f) = f\}$$

If f is D-invariant, then $D \subseteq \operatorname{Stab}(f)$.
If $g_n \in \operatorname{Stab}(f)$ and $g_n \to g$, then $\pi(g)f = \lim_{n\to\infty} \pi(g_n) f = f$ so that $g \in \operatorname{Stab}(f)$.
By density of D, we find $\operatorname{Stab}(f) = G$ and so f is G-invariant, meaning it is essentially constant.
By the proposition applied to $D = \Gamma G_i$, $L^2(X)$ has no non-constant functions invariant under the action of G_i. Hence, $\pi : G \to U\left(L^2(X)\right)$ descends to a representation $\pi : G \to \mathcal{H}$ for which the hypothesis of Howe-Moore applies.
So, take $\alpha, \beta \in L^2(X)$. Let $v = [\alpha], w = [\beta]$ be the projections of α, β to $\mathcal{H} = L^2(X)/V$. The inner product on $\mathcal{H}$ is given by, for $\bar{\alpha} = \frac{1}{m_X(X)} \int_X \alpha \mathrm{d} m_X$,

$$\langle v, w \rangle = \int_X (\alpha - \bar{\alpha})(\beta - \bar{\beta}) \mathrm{d} m_X$$

If g does not generate a compact subgroup, then either $g^n \to \infty$ or $g^{-n} \to \infty$. Since any g-invariant function is also g^{-1}-invariant, we may assume without loss of generality that $g^n \to \infty$. Then Howe-Moore gives $\langle \pi(g^n) v, w \rangle \to 0$ which we can rewrite as

$$\int_X \alpha(g^n x) \beta(x) \mathrm{d} m_X \to \frac{\int_X \alpha \int_X \beta}{m_X(X)}$$

Definition: If the equation above holds for every $\alpha, \beta \in C_c(X)$, then we say that the action of T_g on X is mixing.
Howe-Moore thus implies that the action of T_g on X is mixing.

4 More on Moore ergodicity

Theorem(Howe-Moore ergodicity theorem). Let G be a semisimple Lie group with finite center and no compact simple factors, and X an irreducible G space with finite G-invariant measure. If H is a closed noncompact subgroup of G, then H also acts ergodically on X.

The previous theorem provides a powerful criterion of ergodicity for homogeneous actions, as the next corollary illustrates.
Corollary Let G be a simple noncompact Lie group with finite center and let Γ be a lattice in G. Then any closed noncompact subgroup L of G acts ergodically on G/Γ by left-translations.

Proof. G clearly acts ergodically on G/Γ, since the action is transitive. By the Howe-Moore theorem, the L-action must also be ergodic.

The Howe-Moore ergodicity theorem is in fact a spectral result. In view of the characterization of ergodicity in terms of the unitary representation of G on $L^2(X, \mu)$, the theorem results from the following fact.
For any connected noncompact simple Lie group G with finite center, and unitary representation π of G with no nonzero invariant vectors, a closed subgroup L of G such that $\pi|_L$ has nonzero invariant vectors must be compact. (Mixing implies that if elements act trivially, they must be jailed in a compact set).

Observe that given a nontrivial L-invariant vector $v \in H$, the function $f(g) := \langle \pi(g)v, w \rangle$ is constant on L. Therefore, it is sufficient to prove that for all $v, w \in H$, $\langle \pi(g)v, w \rangle$ approaches 0 as $g \to \infty$ in G, which is exactly the content of Howe-Moore.

Proposition:
i) Let $H \subset G$ be a closed subgroup of a locally compact group G and X a nonsingular G-space. Then H is ergodic on X if and only if G acting diagonally on $X \times G/H$ is ergodic with respect to the product measure class.
ii) For two closed subgroups H_1 and H_2 of G, H_1 is ergodic on G/H_2 if and only if H_2 is ergodic on G/H_1.
Proof. Let $A \subset X \times G/H$ be G-invariant and neither null nor conull. For $y \in G/H$, define the section $A_y = \{x \in X : (x, y) \in A\}$. Invariance of A implies that for any $g \in G, gA_y = A_{gy}$. Fubini then implies that A_e is neither null nor conull. But A_e is H-invariant.
Conversely, let $B \subset X$ be H-invariant subset, and choose a Borel section $\sigma : G/H \to G$ of the projection $G \to G/H$. Define

$$A = \{(x, y) \in X \times G/H : x \in \sigma(y)B\}$$

Since B is H-invariant and for $g \in G$, there exists an $h \in H$ such that $\sigma(gy) = g\sigma(y)h, A$ is a G-invariant Borel set. If B is neither null nor conull, then the same holds for A.
For the second part, by the first part, H_1 is ergodic on G/H_2 if and only if G is ergodic on $G/H_1 \times G/H_2$, which is symmetric in H_1 and H_2.

Applications: A new ergodic theorem, Horocycle flow is mixing of all orders and Furstenberg's theorem

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

Corollary 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 $f \in L^2(X)$,

$$\frac{1}{m(B_n)} \int_{B_n} f\left(g^{-1}x\right) \mathrm{d}m(g) \to \int_X f \mathrm{d}\mu \quad \text{as } n \to \infty$$

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

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

With a change of variables $(g_1, g_2) \mapsto \left(g_1, g_2^{-1} g_1\right)$, we deduce that

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

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

Even more is known than mixing of the geodesic and horocycle flows: Marcus proved that the horocycle flow is mixing of all orders. This is proved using Van der Corput's lemma, which says that if the gaps in a sequence are equidistributed, then the sequence is equidistributed:

Let u_n be a bounded sequence in a hilbert space. If

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} |\frac{1}{N} \sum_{n=1}^{N} \langle u_{n+h}, u_n \rangle| = 0$$

then we have that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u_n = 0$.

A proof of this lemma and the full theorem can be found on Joel Moreira's blog.

Furstenberg's theorem says that horocycle flows on compact homogeneous spaces are measure-theoretically rigid in the following sense:

Theorem (Furstenberg). If $X = \Gamma\backslash\mathrm{SL}(2, \mathbb{R})$ is a compact homogeneous space, then the horocycle flow is uniquely ergodic, i.e. there exists a unique U-invariant probability measure.
To prove that the horocycle flow is uniquely ergodic, Furstenberg used that μ_X is mixing for the geodesic flow, i.e. that the Koopman-von Neumann representation of the geodesic flow is mixing.

Equidistribution: The Wavefront Lemma

We write Howe-Moore under this form to talk about the final lemma:

For $\alpha, \beta \in L^2(X)$,

$$\int_X \alpha(xg)\beta(x)\mathrm{d}x \to \frac{\int_X \alpha \int_X \beta}{m_X(X)}$$

as $g \to \infty$.

Using this, we can establish the following equidistribution result, a version of which will be seen next week.
Theorem. Let $Y = (\Gamma \cap H)\backslash H$. Then, as $g \to \infty$ in $H\backslash G$, the sets Yg become equidistributed, meaning for all $f \in C_c(X)$,

$$\frac{1}{m_Y(Y)} \int_Y \tilde{f}(yg)\mathrm{d}m_Y \to \frac{1}{m_X(X)} \int_X f\mathrm{d}m_X$$

where $\tilde{f} : Y \to \mathbb{R}$ is given by $\tilde{f}((\Gamma \cap H)h) = f(\Gamma h)$.
Remark: The idea of the proof is to realize Y as the subset ΓH of X. Then, we apply Howe-Moore with $\alpha = \chi_Y$, viewing $Y = \Gamma H \subseteq X$, and $\beta = f$. The only problem is that, when $m_X(\Gamma H) = 0$, the mixing identity tells us nothing as $\alpha = \chi_Y = 0$ a.e. The solution will be to 'fatten' Y, so that it has positive measure in X.