Upgrading ergodicity to mixing

Sayantan Khan
University of Michigan

Monday 22nd April, 2019

Abstract

When studying dynamical systems, one is often interested in proving strong chaotic properties about a dynamical system like ergodicity, and mixing. Ergodicity, being the weaker property, is usually easier to prove. However, under certain conditions, we can upgrade ergodicity to mixing using a seemingly unrelated result from representation theory, the Howe-Moore theorem. In this talk, we will see how this is done, and also give a proof of the Howe-Moore theorem.

1 Basic definitions

For the purposes of this talk, a dynamical system will be a probability space (X, μ), along with an action ϕ of a Lie group G. This action must preserve the measure μ, i.e. for any measurable set $A \subset X$ and any $g \in G$, we must have the following identity.

$$\mu(\phi(g^{-1})(A)) = \mu(A)$$

Ergodicity can be thought of as a measure-theoretic generalization of the notion of transitivity. The group action is said to be ergodic if the orbit of every positive measure subset has full measure. Another way to say the same thing is that if every ϕ-invariant subset of X has measure 0 or 1.

The notion of mixing is slightly harder to formulate. To help with a formulation, let's take a real life example where we know what mixing means: imagine a pile of red M&Ms, and equally large pile of blue M&Ms to its right. If we randomly pick two M&Ms from the same spot on the pile, the probability that they're both red will be 0.5. On the other hand, if we pick one M&M, and then mix the pile, and then pick an M&M from the same spot, the probability that they are both red will drop down to 0.25, as if they'd been sampled independently. In some sense, mixing makes any two events independent. That shall be our definition of the notion of mixing for a group action. An action ϕ of the group G is said to be mixing if for any two measurable subset A and B, we have the following identity.

$$\lim_{g \to \infty} \mu(A \cap \phi(g)B) = \mu(A) \cdot \mu(B)$$

Note that $g \to \infty$ just indicates any sequence of elements in G that leaves every compact set. In particular, this definition doesn't make sense for compact groups G, and we'll only be talking about mixing for non-compact groups G.

Now suppose A is a ϕ-invariant set. That means $\phi(g)A = A$ for any $g \in G$, and if ϕ is mixing, that means $\mu(A) = \mu(A)^2$, i.e. $\mu(A) = 0$ or $\mu(A) = 1$, which means the action is ergodic. This shows that mixing is a much stronger property of group actions than ergodicity. We'll see an example of an action that's ergodic, but not mixing, which will prove that mixing is strictly stronger than ergodicity.

It's useful to formulate ergodicity and mixing in terms of L^2 functions on X, because then we bring in the machinery of representation theory into the picture. Note that since G acts via measure preserving action on (X, μ), the corresponding action on $L^2(X)$, given by $\phi(g)(f) := f \circ g^{-1}$ is a unitary action, i.e. a dynamical system is really just a unitary representation of G on the vector space $L^2(X, \mu)$. With this formalism, we say the action of G is ergodic if the only G invariant vectors in $L^2(X, \mu)$ are the constant functions. Similarly, we say the action of G is mixing if we have the following identity for any f and h in $L^2(X, \mu)$.

$$\lim_{g\to\infty} \int_X \phi(g)f \cdot h d\mu = \int_X f d\mu \int_X h d\mu$$

With all the machinery set up, we can now work with some examples. Consider the probability space (S^1, μ), where μ is the Lebesgue measure, and consider the group $\mathbb{Z}$ acting on this space via the map $\phi(1)(x) := 2x$. This is certainly a measure-preserving map, which means it gives a unitary action of $\mathbb{Z}$ on $L^2(S^1)$. The action of $1 \in \mathbb{Z}$ has an easy description on the basis of $L^2(S^1)$. It sends the basis element $\exp(2i\pi nx)$ to $\exp(4i\pi nx)$. Using this description of the $\mathbb{Z}$-action, it's easy to deduce that the action is indeed mixing, and hence also ergodic.

To see an example of an ergodic action that's not mixing, consider the same space and the same group, but consider a slightly different action: $\phi(1)$ sends x to $x + \alpha$, where α is an irrational angle. Again, by looking at the action on the basis of $L^2(S^1)$, it's not too hard to see that the action is ergodic. To see that this action is not mixing, consider the L^2 functions $\exp(2i\pi x)$ and $\exp(-2i\pi x)$. The action of $n \in \mathbb{Z}$ on the former function sends it to $\exp(2i\pi(x + n))$. If we write down the limit, we get the following.

$$\begin{aligned}\lim_{n\to\infty} \int_{S^1} \exp(2i\pi(x + n\alpha)) \exp(-2i\pi x) d\mu &= \int_{S^1} \exp(2i\pi n\alpha) d\mu \\ &\neq 0\end{aligned}$$

This shows the action is not mixing. We'll come back to this example later, and see what exactly goes wrong that prevents the upgrade of ergodicity to mixing.

2 The geodesic flow

Consider a finite volume hyperbolic surface M. A dynamical system of interest to geometers is the geodesic flow on M. It's a dynamical system where the group $\mathbb{R}$ acts on the space (S^1M, μ), where S^1M is the unit tangent bundle of M, and μ is the Liouville measure. The action of $\mathbb{R}$ can be thought of as a time parameter, where at time t, the point (p, v) flows along the geodesic from p in the direction of v for time t. One might ask the same questions about this dynamical system: is the action of $\mathbb{R}$ ergodic? Is it also mixing?

The answer to the first question is yes. We won't be discussing the proof of that, because that's another talk by itself. What we will do is answer the second question, i.e. whether the flow is mixing or not, using the fact that it is ergodic.

To prove that the geodesic flow is mixing, we will reinterpret the geodesic flow in more group theoretic terms. Recall that we're looking at the geodesic flow on a finite volume hyperbolic surface. Such a surface is a quotient of $\mathbb{H}^2$ by a discrete subgroup Γ of the isometry group PSL(2). Furthermore, $\mathbb{H}^2$ itself is the quotient space PSL(2)/PSO(2). We can thus think of our surface M as the following double quotient.

$$M \cong \Gamma\backslash PSL(2)/PSO(2)$$

The unit tangent bundle S^1M of M can similarly be identified as a quotient.

$$S^1M \cong \Gamma\backslash PSL(2)$$

The Liouville measure becomes the Haar measure in this translation, and the action of $t \in \mathbb{R}$ is given as multiplication by a matrix.

$$\phi(t, \Gamma g) = \Gamma g \cdot \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$$

We want to show that this $\mathbb{R}$ action is mixing. In general, doing so for $\mathbb{R}$ or $\mathbb{Z}$ actions is hard, because these groups are abelian, which doesn't provide enough structure to say anything useful. However, in this case, we have an action of a larger group staring at us: the geodesic flow is really the restriction of an $SL(2)$ action in this case. The action of any element $h \in SL(2)$ is simply given by multiplying Γg with h. Since the original $\mathbb{R}$ action was ergodic, the extended $SL(2)$ action is also ergodic, and if we prove that the $SL(2)$ action is mixing, that will also show that the $\mathbb{R}$ action is mixing, since $\mathbb{R}$ is an embedded submanifold of $SL(2)$. To show that the $SL(2)$ action is mixing, we'll need (and eventually prove) the following result about representations of $SL(2)$.

Theorem 1 (Howe-Moore vanishing theorem). *Let H be a Hilbert space, and let $SL(2)$ act on H via a continuous and unitary representation. If H contains no $SL(2)$-invariant vectors, i.e. eigenvectors with eigenvalue 1, then the following identity holds for any v and w in H.*

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

Let's see how we can apply this theorem to prove that the $SL(2)$ action is mixing. Clearly, the natural Hilbert space on which $SL(2)$ acts continuously and via unitary actions is $L^2(\Gamma\backslash PSL(2))$. However, the second hypothesis of Theorem 1 is not satisfied here, i.e. there are $SL(2)$-invariant vectors in $L^2(\Gamma\backslash PSL(2))$, e.g. the constant functions. But we know that the action is ergodic, which means the constant functions are the only $SL(2)$ invariant vectors. The obvious thing to do here is quotient out $L^2(\Gamma\backslash PSL(2))$ by the constant functions, which gives us a representation on the space $L^2_0(\Gamma\backslash PSL(2))$, the space of L^2 functions orthogonal to 1. Now we can apply Theorem 1. That means for any functions f_0 and h_0 in $L^2_0(\Gamma\backslash PSL(2))$, we have the following identity.

$$\lim_{g\to\infty}\int (f_0\circ g^{-1})\cdot h_0\, d\mu = 0 \tag{1}$$

Recall that an action was said to be mixing if the following equation was true for all f and h in $L^2(\Gamma\backslash PSL(2))$.

$$\lim_{g\to\infty}\int (f\circ g^{-1})\cdot h\, d\mu = \int f\, d\mu \int g\, d\mu$$

Write f as $f_0+\bar{f}$, where $\bar{f}$ is the constant function $\int f\, d\mu$, and decompose g in a similar manner. The left hand side of the above equation then becomes the following.

$$\begin{aligned}\lim_{g\to\infty}\int (f\circ g^{-1})\cdot h\, d\mu &= \lim_{g\to\infty}\int ((f_0+\bar{f})\circ g^{-1})\cdot (h_0+\bar{h})\, d\mu \\ &= \lim_{g\to\infty}\left(\int (f_0\circ g^{-1})\cdot h_0\, d\mu + \int (\bar{f}\circ g^{-1})\cdot h_0\, d\mu + \int (f_0\circ g^{-1})\cdot \bar{h}\, d\mu + \int (\bar{h}\circ g^{-1})\cdot \bar{h}\, d\mu\right)\end{aligned}$$

The second and third term of the above expression are 0, since f_0 and g_0 integrate to 0. The first term is precisely the expression in equation 1, which means it goes to 0. The last term is just $\overline{fh}$, which gives us exactly the equation we need for mixing.

The only thing we need to do now is prove theorem 1, the Howe-Moore vanishing theorem. To do so, we'll need to take a small digression into the Cartan and Iwasawa decomposition of the group $SL(2)$.

Cartan and Iwasawa Decomposition

To understand the Cartan decomposition of $SL(2)$ we need to consider the Lie algebra $\mathfrak{sl}(2)$. The Lie algebra consists of 2×2 trace 0 matrices, and has an involution given by $A\mapsto -A^T$. The involution decomposes $\mathfrak{sl}(2)$ into two eigenspaces, $\mathfrak{k}$ corresponding to eigenvalue 1, and $\mathfrak{p}$ corresponding to eigenvalue -1. The eigenspace $\mathfrak{k}$ is a 1-dimensional Lie subalgebra, consisting of skew symmetric matrices, and the subgroup of $SL(2)$ corresponding to $\mathfrak{k}$ is $SO(2)$. We shall denote this subgroup by K. The eigenspace $\mathfrak{p}$ is not a Lie subalgebra, but it has a 1-dimensional maximal abelian subalgebra $\mathfrak{a}$, spanned by the matrix $\begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{pmatrix}$. The subgroup of $SL(2)$ corresponding to $\mathfrak{a}$ is denoted by A. It turns out that every element of $SL(2)$ can be written (not necessarily uniquely) as $k_1 a k_2$, where k_1 and k_2 lie in K, and a lies in A.

The Iwasawa decomposition is a refinement of the Cartan decomposition, because it gives a unique decomposition. This is useful, because the unique decomposition lets us understand the geometry of $SL(2)$. We first describe the decomposition on the Lie algebra level. Recall that $\mathfrak{a}$ is the maximal abelian subalgebra of $\mathfrak{p}$. That means we can write down the root space decomposition of $\mathfrak{p}$, and then look at the subspace consisting of positive roots. Let's denote the space of positive roots by $\mathfrak{n}$. This forms a subalgebra, and in our example, it's 1-dimensional, and spanned by the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. The Lie subgroup corresponding to this Lie algebra is denoted N. It turns out that any element $g \in SL(2)$ can be uniquely written as kan, where $k \in K$, $a \in A$, and $n \in N$. Elements of K are matrices of the form $\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$, which is diffeomorphic to S^1. Elements of A are of the form $\begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}$, for $a > 0$, which means A is diffeomorphic to $\mathbb{R}^+$. Finally, elements of N are matrices of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ for all $x \in \mathbb{R}$. Geometrically, this means $SL(2)$ looks like $(\mathbb{R}^2 \setminus 0) \times \mathbb{R}$, where (r, θ) are polar coordinates on $\mathbb{R}^2 \setminus 0$, and x is the coordinate on the second $\mathbb{R}$ coordinate. The homeomorphism is given by the following map.

$$(r, \theta, x) \mapsto \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & \frac{1}{r} \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

Proof of the Howe-Moore theorem

We are now ready to prove Theorem 1. Recall that we want to prove that if the Hilbert space H has no $SL(2)$-invariant vectors, then for any sequence $\{g_p\}$ that goes to infinity, and any $v \in H$, the sequence $\{g_p v\}$ converges weakly to 0. Since $SL(2)$ is acting via unitary transformations, all the $g_p v$ lie in the unit ball, and since balls are weakly compact, the sequence will have a weakly convergent subsequence. The only thing we really need to show is that the sequence weakly converges to 0. We have reduced our problem to proving the following claim.

Claim 2. *If $\{g_p v\}$ converges weakly to some u, then $u = 0$.*

This is where Cartan decomposition comes in. Write g_p as $k_p a_p l_p$, for k_p and l_p in K, and a_p in A. Since K is compact, pass to a subsequence where k_p converges to some k, and l_p converges to some l. We now have that $(k a_p l)(v)$ converges weakly to 0. This is not too hard to prove, and follows essentially from the triangle inequality.

Now let $v' = lv$, and $u' = k^{-1}v$. We will now show that u' is $SL(2)$ invariant. This will suffice, since by assumption, that will force $u' = 0$, and hence $u = 0$. To show u' is $SL(2)$-invariant, we will use the Iwasawa decomposition, and begin by showing that u' is just N-invariant. First of all, note that we have the following identity.

$$\lim_{p \to \infty} a_p^{-1} n a_p = 0$$

This follows because the element $a_p = \exp(t_p H)$, where t_p is a sequence of positive real numbers going to ∞, and H is the generator of $\mathfrak{a}$. Similarly, $n = \exp(J)$ for some $J \in \mathfrak{n}$. Because $\mathfrak{n}$ is the positive root space of $\mathfrak{a}$, the above expression just becomes the following.

$$\begin{aligned} \lim_{p \to \infty} a_p^{-1} n a_p &= \lim_{t_p \to \infty} \exp(e^{-\lambda t_p}(J)) \\ &= e \end{aligned}$$

Since we want to show $nu' = u'$, it is useful to look at $n a_p v' - v'$. For any $w \in H$, we have the following.

$$\begin{aligned} \langle n a_p v' - v', w \rangle &= \langle a_p^{-1} n a_p v' - a_p^{-1} v', a_p^{-1} w \rangle \\ &\leqslant \| a_p^{-1} n a_p v' - a_p^{-1} v' \| \, \| w \| \end{aligned}$$

If we now take a limit, the right hand side goes to 0, which means na_pv' converges weakly to a_pv', which means $nu' = u'$.

The next step is to show that u' is AN-invariant. To see this consider the function $\phi(g) := \langle gu', u' \rangle$ for $g \in G/N$. This is a continuous function on G/N, which can be thought of as $\mathbb{R}^2 \setminus 0$. Furthermore, this function is constant along the orbits of the left action of N on G/N. For points of the form (x, y), for $y \neq 0$, the orbit under the N-action is $\mathbb{R} \times \{y\}$, which means the function ϕ is constant along horizontal lines.

Note that the function ϕ takes value $\|u'\|^2$ at $(1, 0)$. Consider the action of A on this point. The orbit is just the x-axis, which means $\langle au', u' \rangle = \|u'\|^2$ for all $a \in A$. By the equality case in Cauchy-Schwarz, that means $au' = u'$. This shows u' is invariant under the AN action. Now think of ϕ as a function on G/AN, which is just S^1. Again, the function ϕ must be constant along A orbits, and one A orbit is all of S^1 minus $(1, 0)$. This means ϕ is constant everywhere, and hence $gu' = u'$ for all $g \in SL(2)$. This proves the result.