Von Neumann ergodic theorem and the ergodicity of the geodesic flow

Misha Verbitsky

Estruturas geométricas em variedades,

IMPA, July 21, 2022

Hilbert spaces (reminder)

DEFINITION: Hilbert space is a complete, infinite-dimensional Hermitian space which is second countable (that is, has a countable dense set).

DEFINITION: Orthonormal basis in a Hilbert space H is a set of pairwise orthogonal vectors $\{x_\alpha\}$ which satisfy $|x_\alpha| = 1$, and such that H is the closure of the subspace generated by the set $\{x_\alpha\}$.

THEOREM: Any Hilbert space has a basis, and all such bases are countable.

Proof: A basis is found using Zorn lemma. If it's not countable, open balls with centers in x_α and radius $\varepsilon < 2^{-1/2}$ don't intersect, which means that the second countability axiom is not satisfied. ■

THEOREM: All Hilbert spaces are isometric.

Proof: Each Hilbert space has a countable orthonormal basis. ■

EXAMPLE: Let M be a space with measure, and $L^2(M)$ be the quotient of the space of measurable, square-integrable functions by the space of functions which vanish almost everywhere. Define the scalar product on $L^2(M)$ by $(f, g) := \int_M fg$. **Then $L^2(M)$ is a Hilbert space** (F. Riesz, E. S. Fischer).

Real Hilbert spaces

DEFINITION: A Euclidean space is a vector space over $\mathbb{R}$ equipped with a positive definite scalar product g. **Real Hilbert space** is a complete, infinite-dimensional Euclidean space which is second countable (that is, has a countable dense set).

DEFINITION: Orthonormal basis in a Hilbert space H is a set of pairwise orthogonal vectors $\{x_\alpha\}$ which satisfy $|x_\alpha| = 1$, and such that H is the closure of the subspace generated by the set $\{x_\alpha\}$.

THEOREM: Any real Hilbert space has a basis, and all such bases are countable.

Proof: A basis is found using Zorn lemma. If it's not countable, open balls with centers in x_α and radius $\varepsilon < 2^{-1/2}$ don't intersect, which means that the second countability axiom is not satisfied. ■

THEOREM: All real Hilbert spaces are isometric.

Proof: Each Hilbert space has a countable orthonormal basis. ■

REMARK: Further on, **all linear operators on Hilbert spaces are tacitly assumed continuous.**

Adjoint maps

EXERCISE: Let (H, g) be a Hilbert space. Show that **the map $x \longrightarrow g(x, \cdot)$ defines an isomorphism $H \longrightarrow H^*$.**

DEFINITION: Let $A: H \longrightarrow H$ be a continuous linear endomorphism of a Hilbert space (H, g). Then $\lambda \longrightarrow \lambda(A(\cdot))$ is a **an adjoint map** $A^*: H^* \longrightarrow H^*$. Identifying H and H^* as above, we interpret A^* as an endomorphism of H.

REMARK: The map A^* satisfies $g(x, A(y)) = g(A^*(x), y)$. This relation is often taken as a definition of the adjoint map.

DEFINITION: An operator $U: H \longrightarrow H$ is **orthogonal** if $g(x, y) = g(U(x), U(y))$ for all $x, y \in H$.

CLAIM: An invertible operator **U is orthogonal if and only if $U^* = U^{-1}$.**

Proof: Indeed, orthogonality is equivalent to $g(x, y) = g(U^*U(x), y)$, which is equivalent to $U^*U = \mathrm{Id}$ because the form $g(\cdot, y)$ is non-zero for non-zero y. ■

Orthogonal maps and direct sum decompositions

LEMMA: Let $U: H \longrightarrow H$ be an invertible orthogonal map. Denote by H^U the kernel of $1 - U$, that is, the space of U-invariant vectors, and let H_1 be the closure of the image of $1 - U$. **Then $H = H^U \oplus H_1$ is an orthogonal direct sum decomposition.**

Proof. Step 1: Clearly, $H^U \subset$ is a closed subspace. Let $x \in H^U$. Then

$$(U^* - 1)(x) = (U^* - 1)U(x) = (U^{-1} - 1)U(x) = (1 - U)x = 0.$$

This gives $g(x, (U - 1)y) = g((U^* - 1)x, y) = 0$, hence $x \perp H_1$. We obtain that $H^U \perp H_1$.

Step 2: It remains to show only that $H_1^\perp = H_U$. Any vector x which is orthogonal to H_1 satisfies $0 = g(x, (U - 1)y) = g((U^* - 1)x, y)$, giving

$$0 = (U^* - 1)(x) = U(U^* - 1)(x) = U(U^{-1} - 1)(x) = (1 - U)x,$$

hence $x \in H_U$. ■

Von Neumann erodic theorem

Corollary 1: Let $U: H \longrightarrow H$ be an invertible orthogonal map, and $U_n := \frac{1}{n}\sum_{i=0}^{n-1} U^i(x)$. Consider the orthogonal projection to $P: H \longrightarrow H^U$. **Then $\lim_n U_n(x) = P(x)$, for all $x \in H$.**

Proof: By the previous lemma, it suffices to show that $\lim_n U_n = 0$ on H_1. However, the vectors of form $x = (1-U)(y)$ are dense in H_1, and for such x we have $U_n(x) = U_n(1-U)(y) = \frac{1-U^n}{n}(y)$, and it converges to 0 because $\|U^n\| = 1$. ■

THEOREM: Let (M, μ) be a measured space and $T: M \longrightarrow M$ a map preserving the measure. Consider the space $L^2(M)$ of functions $f: M \longrightarrow \mathbb{R}$ with f^2 integrable, and let $T^*: L^2(M) \longrightarrow L^2(M)$ map f to T^*f. **Then the series $T_n(f) := \frac{1}{n}\sum_{i=0}^{n-1}(T^*)^i(f)$ converges in $L^2(M)$ to a T^*-invariant function.**

Proof: Corollary 1 implies that $T_n(f)$ converges to $P(f)$. ■

The Hopf Argument

DEFINITION: Let M be a metric space with a Borel measure and $F : M \longrightarrow M$ a continuous map preserving measure. The **"stable foliation"** is an equivalence relation on M, with $x \sim y$ when $\lim_i d(F^n(x), F^n(y)) = 0$. The **"leaves"** of stable foliation are the equivalence classes.

THEOREM: (Hopf Argument) Any measurable, F-invariant function **is constant on the leaves of stable foliation** outside of a measure 0 set.

Proof: Let $A(f) := \lim_n \frac{1}{n}\sum_{i=0}^{n-1}(F^i)^* f$ be the map defined above. Since $A(f) = f$ for any F-invariant f, it suffices to prove that $A(f)$ is constant on leaves of the stable foliation only for $f \in \operatorname{im} A$. The Lipschitz L^2-integrable functions are dense in $L^1(M)$ by Stone-Weierstrass. Therefore it suffices to show that $A(f)$ is constant on leaves of the stable foliation when f is C-Lipschitz for some $C > 0$ and square integrable.

For any sequence $\alpha_i \in \mathbb{R}$ converging to 0, the sequence $\frac{1}{n}\sum_{i=0}^{n-1}\alpha_i$ also converges to 0. Therefore, whenever $x \sim y$, one has

$$A(f)(x) - A(f)(y) = \lim_n \sum_{i=0}^{n-1} f(F^i(x)) - f(F^i(y)) = 0$$

because $\alpha_i = |f(F^i(x)) - f(F^i(y))| \leqslant Cd(F^i(x), F^i(y))$ converges to 0. ∎

Stable and unstable foliations

DEFINITION: Let M be a metric space with a Borel measure and $F: M \longrightarrow M$ a homeomorphism preserving measure. The **"unstable foliation"** is a stable foliation for F^{-1}.

DEFINITION: The map F is called **pseudo-Anosov** if any leaf of stable foliation intersects any leaf of unstable foliation.

COROLLARY: A pseudo-Anosov map $F: M \longrightarrow M$ is always ergodic.

Proof: F is ergodic if all F-invariant $f \in L^2(M)$ are constant. However, al such f are constant on leaves of stable foliation and leaves on unstable foliation and these leaves intersect. ■

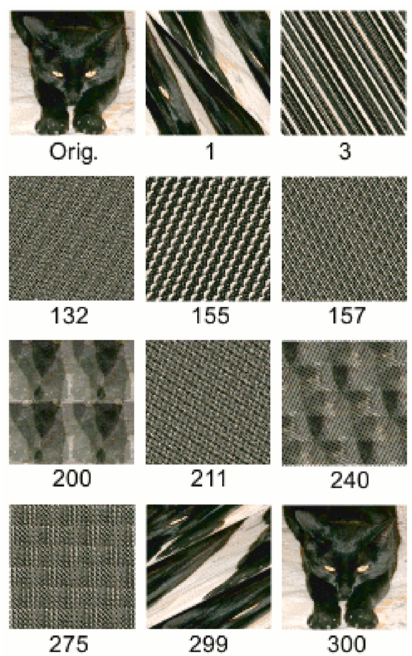
EXAMPLE: (Anosov diffeomorphism)
Let $A: T^2 \longrightarrow T^2$ be a linear map of a torus defined by $A \in SL(2, \mathbb{Z})$, with real eigenvalues $\alpha > 1$ and $\beta \in]0, 1[$, The eigenspace corresponding to β gives a stable foliation, the eigenspace corresponding to α the unstable foliation, hence **A is ergodic.**

Arnold's cat map

DEFINITION: The Arnold's cat map is $A: T^2 \longrightarrow T^2$ defined by $A \in SL(2,\mathbb{Z})$,

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

The eigenvalues of A are roots of $\det(t\operatorname{Id} - A) = (t-2)(t-1) - 1 = t^2 - 3t - 1$. This is a quadratic equation with roots $\alpha_\pm = \frac{3\pm\sqrt{5}}{2}$. On the vectors tangent to the eigenspace of α_-, the map A^n acts as $(\alpha_-)^n$, hence the stable foliation is tangent to these vectors. Similarly, unstable foliation is tangent to the eigenspace of α_+.



Boolean algebras

DEFINITION: The set of subsets of X is denoted by 2^X. **Boolean algebra of subsets if X** is a subset of 2^X closed under **boolean operations** of union, intersection and complement.

DEFINITION: Let M be a set **A σ-algebra** of subsets of X is a Boolean algebra $\mathfrak{A} \subset 2^X$ such that for any countable family $A_1, ..., A_n, ... \in \mathfrak{A}$ the union $\bigcup_{i=1}^\infty A_i$ is also an element of $\mathfrak{A}$.

DEFINITION: A function $\mu: \mathfrak{A} \longrightarrow \mathbb{R} \cup \{\infty\}$ is called **finitely additive** if for all non-intersecting $A, B \in \mathfrak{U}$, $\mu(A \coprod B) = \mu(A) + \mu(B)$. The sign $\coprod$ denotes union of non-intersecting sets. μ is called **σ-additive** if $\mu(\coprod_{i=1}^\infty A_i) = \sum \mu(A_i)$ for any pairwise disjoint countable family of subsets $A_i \in \mathfrak{A}$.

DEFINITION: **A measure** in a σ-algebra $\mathfrak{A} \subset 2^X$ is a σ-additive function $\mu: \mathfrak{A} \longrightarrow \mathbb{R} \cup \{\infty\}$.

EXAMPLE: Let X be a topological space. The **Borel σ-algebra** is a smallest σ-algebra $\mathfrak{A} \subset 2^X$ containing all open subsets. **Borel measure** is a measure on Borel σ-algebra.

Ergodic measures

REMARK: Let M, μ be a space with measure. We say that **"property P holds for almost all $x \in M$"** when property P holds for all $x \in M$ outside of a measure 0 subset.

DEFINITION: Let Γ be a group acting on a measured space (M, μ) and preserving its σ-algebra. We say that the Γ-action is **ergodic** if for each Γ-invariant, measurable set $U \subset M$, either $\mu(U) = 0$ or $\mu(M \backslash U) = 0$. In this case μ is called **an ergodic measure**.

THEOREM: Let M be a second countable topological space, and μ a Borel measure on M. Let Γ be a group acting on M by homeomorphisms. Suppose that any non-empty open subset of M has positive measure, and action of Γ is ergodic. Then **for almost all $x \in M$, the orbit $\Gamma \cdot x$ is dense in M.**

Proof. Step 1: Let U_i be a countable base of topology on M. The orbit $\Gamma \cdot x$ is dense in M if $(\Gamma \cdot x) \cap U_i \neq 0$ for all i. This is equivalent to $x \in \Gamma \cdot U_i$. Therefore, **the set of all x with dense orbits is $\bigcap_i (\Gamma \cdot U_i)$.**

Step 2: Since $\Gamma \cdot U_i$ is Γ-invariant and has positive measure, it has full measure because of ergodicity. **Then $\bigcap_i (\Gamma \cdot U_i)$ is an intersection of sets which have full measure.** ■

Ergodic measures and integrable functions

Rule of a thumb: If your group action preserves measure and almost all its orbits are dense, it is most likely ergodic. **Not always!**

THEOREM: Let (M, μ) be a space with finite measure, and Γ a group acting on M and preserving the measure. Then the following are equivalent.

(a) The action of Γ is ergodic.

(b) For each integrable, Γ-invariant function $f : M \longrightarrow \mathbb{R}$, f is constant almost everywhere.

Proof: To obtain (a) from (b), take the characteristic function χ_U of a Γ-invariant set $U \subset M$. Then it is constant almost everywhere, hence U is of full measure (in this case $\chi_U = 1$ almost everywhere) or measure zero, in later case $\chi_U = 0$ almost everywhere.

To obtain (b) from (a), let c be the average value of f on M, $M_\varepsilon^+ := f^{-1}([c+\varepsilon, \infty[)$, and $M_\varepsilon^- := f^{-1}(]-\infty, c+\varepsilon])$. Both sets are Γ-invariant and not of full measure, hence they have measure zero. This means that for all $\varepsilon > 0$, $c - \varepsilon < f(x) < c + \varepsilon$ for almost all x. ■

Geodesic flow

DEFINITION: Let M be a manifold. **Spherical tangent bundle** $SM \subset TM$ is the space of all tangent vectors of length 1.

DEFINITION: Consider the map

$$\Psi_t(v,x) = (\exp(tv), d\exp(tv)(v))$$

mapping $v \in T_xM, t \in \mathbb{R}$ to $d\exp(tv)(v)) \in T_{\exp(tv)}M$; here

$$d\exp(tv) : T_xM \longrightarrow T_{\exp(tv)}M$$

is the differential of the exponent map $\exp : T_xM \longrightarrow M$. This defines an action of $\mathbb{R}$ on SM, $t \longrightarrow \Psi_t \in \operatorname{Diff}(SM)$. This action is called **the geodesic flow**.

REMARK: Geodesic flow **takes a unit tangent vector, takes a naturally parametrized geodesic tangent to this vector, and moves this vector along this geodesic.**

THEOREM: Geodesic flow **preserves the Riemannian volume form on SM.**

Proof: Left as an exercise. ■

Space forms (reminder)

DEFINITION: Simply connected space form is a homogeneous manifold of one of the following types:

positive curvature: S^n (an n-dimensional sphere), equipped with an action of the group $SO(n+1)$ of rotations

zero curvature: $\mathbb{R}^n$ (an n-dimensional Euclidean space), equipped with an action of isometries

negative curvature: $SO(1,n)/O(n)$, equipped with the natural $SO(1,n)$-action. This space is also called **hyperbolic space**, and in dimension 2 **hyperbolic plane** or **Poincaré plane** or **Bolyai-Lobachevsky plane**

Riemannian metric on space forms

LEMMA: Let $G = SO(n)$ act on $\mathbb{R}^n$ in a natural way. **Then there exists a unique G-invariant symmetric 2-form:** the standard Euclidean metric.

Proof: Let g, g' be two G-invariant symmetric 2-forms. Since S^{n-1} is an orbit of G, we have $g(x,x) = g(y,y)$ for any $x, y \in S^{n-1}$. Multiplying g' by a constant, we may assume that $g(x,x) = g'(x,x)$ for any $x \in S^{n-1}$. **Then $g(\lambda x, \lambda x) = g'(\lambda x, \lambda x)$ for any $x \in S^{n-1}$, $\lambda \in \mathbb{R}$;** however, all vectors can be written as λx. ■

COROLLARY: Let $M = G/H$ be a simply connected space form. **Then M admits a unique, up to a constant multiplier, G-invariant Riemannian form.**

Proof: The isotropy group is $SO(n-1)$ in all three cases, and the previous lemma can be applied. ■

REMARK: **From now on, the space forms are assumed to be homogeneous Riemannian manifolds**.

Upper half-plane as a space form

DEFINITION: Poincaré half-plane is the upper half-plane equipped with an $PSL(2,\mathbb{R})$-invariant metric. By constructtion, **t is isometric to the Poincare disk and to the hyperbolic space form.**

THEOREM: Let (x,y) be the usual coordinates on the upper half-plane $\mathbb{H}$. **Then the Riemannian structure s on $\mathbb{H}$ is written as $s = const\frac{dx^2+dy^2}{y^2}$.**

Proof: Since the complex structure on $\mathbb{H}$ is the standard one and all Hermitian structures are proportional, we obtain that $s = \mu(dx^2+dy^2)$, where $\mu \in C^\infty(\mathbb{H})$. **It remains to find μ, using the fact that s is $PSL(2,\mathbb{R})$-invariant.**

For each $a \in \mathbb{R}$, the parallel transport $x \longrightarrow x+a$ fixes s, hence μ is a function of y. For any $\lambda \in \mathbb{R}^{>0}$, the map $H_\lambda(x) = \lambda x$, being holomorphic, also fixes s; since $\mathbb{H}_\lambda(dx^2+dy^2) = \lambda^2 dx^2 + dy^2$, we have $\mu(\lambda x) = \lambda^{-2}\mu(x)$. ∎

Absolute

Let $V = \mathbb{R}^d$ be a vector space with bilinear form of signature (1,d-1). Denote by V^+ the positive cone of V, that is, one of two connected components of $\{v \in V \ \mid \ (v,v) > 0\}$. Consider the hyperbolic space $\mathbb{H} = SO^+(1,d-1)/SO(d-1)$ as projectivization of V^+, $\mathbb{H} = \mathbb{P}V^+ = V^+/\mathbb{R}^{>0}$. Let $\overline{\mathbb{H}}$ be the closure of $\mathbb{P}V^+ \subset \mathbb{P}V = \mathbb{R}P^2$.

DEFINITION: The infinite circle $\partial\Delta$ considered as a boundary of the disk $\mathbb{P}V^+ = \mathbb{H}$ is called **the absolute** of the projective plane.

REMARK: Any isometry of the disk is naturally extended to the absolute. Indeed, $SO^+(1,d-1)$ acts on the real projective space $\mathbb{R}P^{d-1}$, and absolute is the boundary of $\mathbb{P}V^+$ in $\mathbb{R}P^{d-1}$.

Convergence of geodesics

REMARK: From now on, **all geodesics are considered with their natural parametrization.**

REMARK: From the description of geodesics in Poincare disc, it is clear that for any geodesic $\gamma :]\infty, \infty[\longrightarrow \Delta$ the limit points $\gamma_+ := \lim_{t \mapsto \infty} \gamma(t)$ and $\gamma_- := \lim_{t \mapsto -\infty} \gamma(t)$ are well defined in the absolute $\partial\Delta$, and, moreover, the points $\gamma_+, \gamma_- \in \partial\Delta$ determine the geodesic uniquely.

REMARK: The Poincaré metric on $\mathbb{H}$ is $d_P = \frac{dx^2 + dy^2}{y^2}$. Therefore,

$$\lim_{u \longrightarrow \infty} d_P((t_1, u_1 + u), (t_2, u_2 + u)) = 0.$$

This gives the following

COROLLARY: Let γ, δ be geodesics such that their $+\infty$-limits $\gamma_+, \delta_+ \in \partial\Delta$ are equal, and $t_1 \in \mathbb{R}$ any number. **Then there exists $t_2 \in \mathbb{R}$ such that the tangent vectors $\dot\gamma(t_1), \dot\delta(t_2) \in S\Delta$ belong to the same leaf of the stable foliation.** ■

Geodesic flow on a Riemann surface

THEOREM: (E. Hopf) Let M be a complete Riemannian manifold of finite volume and constant negative curvature. **Then the geodesic flow is ergodic.**

Proof: We need to check that the geodesic flow is pseudo-Anosov, and apply Hopf's theorem.

Proof. Step 1: Any such M **is obtained as a quotient** of the hyperbolic plane Δ/Γ, where Γ is a discrete group acting on Δ by isometries.

Step 2: To prove Hopf Theorem it would suffice to show that a function which is (*) constant on orbits of the geodesic flow and on almost all leaves of stable foliation on $S\Delta$ and (**) on orbits of the geodesic flow and on almost all leaves of unstable foliation is necessarily constant. This follows from the Hopf argument.

Step 3: For (*) f should be constant on all S_α, where S_α is all $v \in T_x\Delta$ such that the geodesic tangent to v end up in a point $\alpha \in \partial\Delta$. For (**), f should be constant on all U_β, where U_β is all vectors $v \in T_x\Delta$ such that the geodesic tangent to v begins in $\beta \in \partial\Delta$. **The sets S_α, U_β intersect in a geodesic connecting α to β** which exists whenever $\alpha \neq \beta$. ■