Teoría de Billares

En los últimos treinta años se ha desarrollado principalmente debido a las siguientes razones:

1. Algunas clases de billares presentan un fuerte comportamiento caótico y pueden ser considerados entre los mejores ejemplos de caos determinístico;

2. Muchos ejemplos interesantes de sistemas dinámicos de origen físico (especialmente aquellos en que la interacción entre partículas envuelve choques elásticos) pueden ser reducidos a billares. Algunos de estos ejemplos serán estudiados;

3. Importantes problemas en la teoría de caos cuántico involucran un análisis profundo de los billares clásicos;

4. El estudio de los billares sugiere muchos problemas bonitos e interesantes en geometría y probabilidad.

Gases of hard balls. Consider a more realistic model of n balls moving in space. For simplicity, let all balls have the same radius r and the same mass m. Each ball moves freely, i.e. with constant velocity, until it hits another moving ball. When two balls collide, they change their velocities according to the laws of elastic collision.

This law means the following. Let two balls collide. Denote by q_1 and q_2 their centers and by v_1 and v_2 their velocity vectors at the moment of collision. Let L be the line through the centers q_1 and q_2. We decompose

$$v_i = v_i^0 + v_i^\perp$$

for $i = 1, 2$, where v_i^0 is the component of the vector v_i parallel to L and $v_i^\perp$ is the one perpendicular to L. Then the new, outgoing, velocities of the balls are

$$v_1^{\text{new}} = v_1^\perp + v_2^0 \quad \text{and} \quad v_2^{\text{new}} = v_2^\perp + v_1^0$$

In other words, the balls exchange the velocity components parallel to the center line L and retain the orthogonal components.

We note that the laws of elastic collision imply preservation of the total kinetic energy $\sum m\|v_i\|^2/2$ and the total momentum $\sum mv_i$ of the system of n balls. We also note that a collision of two hard balls with centers q_1 and q_2 can only occur if

$$\mathrm{dist}(q_1, q_2) = 2r\,, \quad i.e.\ \ \|q_1 - q_2\|^2 = (2r)^2.$$

The system of n balls moving in the open space without walls is dynamically not very interesting. As it is intuitively clear (and proven mathematically), the total number of collisions between balls is always finite, and after the last collision the balls will fly freely forever. Furthermore, the number of collisions between n balls in the open space is uniformly bounded by a constant M that only depends on n. This last fact was proved very recently – in 1998 – by Burago, Ferleger and Kononenko.

$$M = \left(32\sqrt{\frac{m_{\max}}{m_{\min}}\frac{r_{\max}}{r_{\min}}}\,n^{\frac{3}{2}}\right)^{n^2}$$

Let us consider n balls enclosed in a bounded domain R, called a container (or reservoir).

The balls collide elastically with each other and with the walls of the container. Precisely, if a ball with center q hits a wall at a point $w \in \partial R$, then we decompose its velocity vector as $v = v^0 + v^\perp$, where v^0 is the component parallel to the line passing through q and w, and $v^\perp$ is perpendicular to that line. The new, outgoing, velocity of the ball is $v^{\rm new} = v^\perp - v^0$. Note that this rule preserves the total kinetic energy of the system, but *not* its total momentum.

Now we reduce the system of n hard balls in a container R to a billiard. We denote by $q_i = (q_i^1, q_i^2, q_i^3)$ the center of the ith ball and by $v_i = (v_i^1, v_i^2, v_i^3)$ its velocity vector, $1 \le i \le n$. Now the entire system can be described by a configuration point

$$q = (q_1^1, q_1^2, q_1^3, q_2^1, \dots, q_n^2, q_n^3) \in \mathbb{R}^{3n}$$

and its velocity vector

$$v = (v_1^1, v_1^2, v_1^3, v_2^1, \dots, v_n^2, v_n^3) \in \mathbb{R}^{3n}. \quad (1)$$

We note that $q \in R^n = R \times \cdots \times R$. It is also important to observe that not the entire region R^n is available for the configuration point q. By the rules of elastic collisions, the balls cannot overlap – the moment they bump into each other, they collide. This rule requires exclusion of configurations that satisfy

$$(q_i^1 - q_j^1)^2 + (q_i^2 - q_j^2)^2 + (q_i^3 - q_j^3)^2 < (2r)^2 \quad (2)$$

for some $1 \leq i < j \leq n$ (here r is the radius of the balls). The inequality (2) specifies a spherical cylinder in $\mathbb{R}^{3n}$, which we denote by C_{ij}. For the model of hard disks on a plane, we get a circular cylinder C_{ij} in $\mathbb{R}^{2n}$. The cylinders C_{ij}, $1 \leq i < j \leq n$, contain all forbidden configurations of the balls (disks), hence they must be removed from the available space. As a result, we get a smaller domain

$$Q = R^n \setminus \cup_{i \neq j} C_{ij}$$

This domain Q is called the *configuration space* of the system.

Now one can check by direct inspection (it is a rather tedious exercise) that the trajectory of the configuration point q in Q is governed by the billiard rules:

$$\dot{q} = v \qquad \text{and} \qquad \dot{v} = 0 \tag{3}$$

where dots indicate the derivative with respect to time. When $q \in \partial Q$, the velocity v of the particle changes discontinuously, according to the classical rule *the angle of incidence is equal to the angle of reflection.* So, the new (outgoing) vector v_+ is related to the old (incoming) vector v_- by

$$v_+ = v_- - 2\langle v_-, n(q)\rangle n(q). \tag{4}$$

Specular reflections at the surface of a cylinder C_{ij} correspond to collisions between the balls i and j. Thus, the study of the mechanical model of n balls or disks is reduced to the study of billiard dynamics in the domain Q. We note that the conservation of the total kinetic energy $\sum_i m\|v_i\|^2/2$ is equivalent to the preservation of the norm $\|v\|$ of the velocity vector (1).

The singularity set Γ^* contains all intersection of the cylindrical surfaces ∂C_{ij} with each other. Such intersections correspond to simultaneous collisions of three or more balls. The outcome of such multiple collisions is not defined. It is our general rule, though, to ignore billiard trajectories that hit Γ^*.

The gas of hard balls is a classical model in statistical physics. Its study goes back to L. Boltzmann in the XIX century. Many physical laws have been first established for gases of hard balls, and then experimentally verified for other gases. Boltzmann was first to state the celebrated *ergodic hypothesis*. He assumed that gases of hard balls are, in general, ergodic and used this assumption to justify the laws of statistical mechanics (on a "heuristic" level). Since then, it remains a major challenge for physicists and mathematicians to prove this hypothesis, as well as to make use of the ergodicity to build the mathematical foundation of statistical mechanics.

In early sixties, Ya. Sinai studied a specific version of Boltzmann's model – the gas of n hard balls (or disks) on a torus $\mathbb{T}^d$, $d \geq 2$. In that case the container R is a torus, so there are no walls (i.e., $\partial R = \emptyset$). Hence, the balls only collide with each other. Therefore, in addition to the total kinetic energy, the total momentum is conserved. Sinai conjectured that if one sets the total momentum to zero and fixes the center of mass, then the resulting reduced system would be ergodic.

Attempts to prove the Boltzmann-Sinai conjecture spanned almost 40 years, and they had a colorful and sometimes dramatic history. It appears that the problem is almost solved by now due to very recent works of N. Simanyi and D. Szasz (1999, 2000). But that solution is beyond the scope of these lectures.

Estimates for the number of reflections. Here we consider the following problem: given a piece of a billiard trajectory of length L, how many reflections at ∂Q can be there on that piece? In particular, can the number of reflections, n, be infinite? It is not difficult to prove that $n < \infty$ with probability one. But, from the geometric point of view, one would like to know if n can ever be infinite, and how large n can be. These questions also arise in the studies of ergodic properties of billiards.

We start with a simple case - a billiard trajectory moving between two lines, l_1 and l_2, which intersect at a point A at angle $\alpha > 0$. We call Q the (infinite) domain bounded by these lines. When the trajectory hits either line, say l_1, it gets reflected, but its mirror image across l_1 will continue moving straight on the other side of l_1. We will follow that mirror image, rather than the trajectory itself.

It will continue moving in the domain Q_1 that is the mirror images of Q across the line l_1. When our trajectory hits the other line l_2, its mirror image also hits the other side of the domain Q_1, etc. We will keep reflecting the domains Q_i across their sides and following the straight line made by the mirror images of our trajectory. This will look like a mirror room in an amusement park, with multiple reflections in different mirrors.

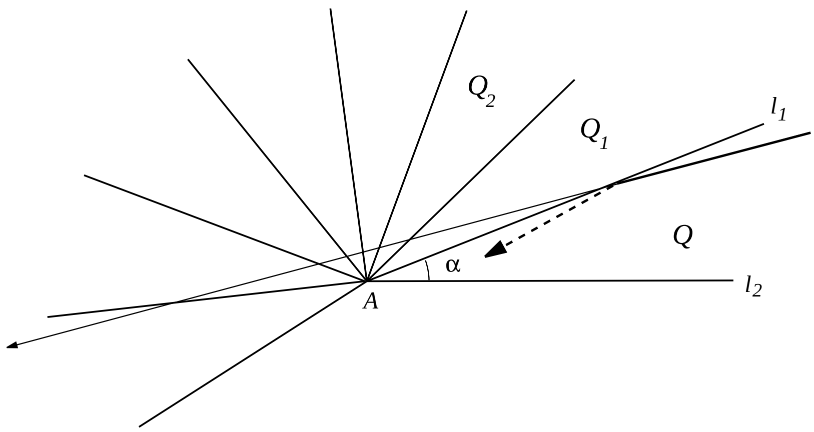

Figure 1: The unfolding of a billiard trajectory.

It is clear from Fig. 1 that the total angle made by the unfolding images $Q_1, Q_2, \ldots$, between the first reflection and the last one, cannot exceed π, hence

$$n < \frac{\pi}{\alpha} + 1$$

This simple estimate gives an upper bound on the number of reflections in Q. We note that this bound is uniform, i.e. the same for all billiard trajectories in Q.

The above estimate has a multidimensional version. Suppose that several hyperplanes in $\mathbb{R}^d$, $d \geq 2$, intersect at one point A, so that they make a "polyhedral angle" with vertex at A. Ya. Sinai proved in 1978 that the number of reflections experienced by any billiard trajectory inside such an angle is uniformly bounded, the bound only depends on the configuration of the hyperplanes.

It is more difficult to estimate the number of reflections in billiard domains with curvilinear boundary. Here we have two distinct cases. One is a billiard domain Q with a convex boundary, such as a disk on a plane or a ball in $\mathbb{R}^d$. Near a convex boundary ∂Q, a short piece of trajectory clearly can experience arbitrary many reflections. This happens when the velocity vector v at a point of reflection $q \in \partial Q$ is almost tangent to the boundary ∂Q.

Such a trajectory would simply "slide" along ∂Q experiencing many "grazing" collisions with ∂Q in rapid succession.

It is even possible to construct convex billiard tables $Q \subset \mathbb{R}^2$ where a short piece of trajectory experiences infinitely many reflections accumulating at a point of ∂Q where the curvature vanishes. Such "anomalous" examples were found by Halpern (1977).

On the contrary, when the boundary of Q is concave (i.e., convex inward), the number of reflections can be well bounded. To picture a billiard table with concave boundary, take a polygon and bow each side inward a bit. Or recall the table on a torus where a disk is removed.

It is clear that near one smooth concave piece of ∂Q any short billiard trajectory can only have one reflection. Consider now a corner point, i.e. a vertex A where two (or more) concave pieces of ∂Q meet. Estimates on the number of reflections near such corners have been extensively studied by Galperin (1981), Vasserstein (1979) and others. Here we present the most general estimate obtained by Burago, Ferleger & Kononenko in 1998.

To define a corner point with concave walls in $\mathbb{R}^d$, one can consider finitely many closed convex subsets $B_i \subset \mathbb{R}^d$, $i = 1, \ldots, n$, whose boundaries are C^1 hypersurfaces and define a billiard domain by

$$Q = \mathbb{R}^d \setminus (\cup_{i=1}^n B_i) \tag{5}$$

It is clearly enough to assume that

$$B := \bar{Q} \cap (\cap_{i=1}^n B_i) \neq \emptyset$$

and consider short billiard trajectories near B (the set B plays the role of a vertex).

For any two points $X, Y \in Q$ we denote by $T(X, Y)$ the piece of billiard trajectory starting at X and ending at Y (if one exists), and by $|T(X, Y)|$ its length. The following lemma compares $|T(x, y)|$ to the distance from X and Y to the "bottom of the corner" – the set B.

Lemma [Comparison Lemma]. *For every $X, Y \in Q$ and every $A \in B$*

$$|XA| + |AY| \geq |T(X, Y)|$$

Inequality is strict if one of the reflections occurs at a strictly concave part of the boundary of Q.

Theorem [Galperin, Vaserstein]. *For any billiard trajectory of finite length in Q the number of reflections is finite.*

Proof. Assume the opposite – a trajectory T starting at $X \in Q$ has infinitely many reflection points that accumulate at a point $A \in B$ (if $A \notin B$, we can remove some B_i's from our construction). Let $X_1, X_2, \ldots$ be the points of reflection, and $X_i \to A$ as $i \to \infty$. Clearly the length of the straight segment X_1A is smaller than the length $|T(X_1, A)|$ of the entire trajectory between X_1 and A. Therefore we can find X_k sufficiently close to A so that $|X_1A| + |AX_k| < |T(X_1, X_k)|$, which contradicts to the comparison Lemma. □

A uniform bound on the number of reflections requires some extra conditions on the billiard domain. Indeed, if a corner point A of a planar billiard table $Q \subset \mathbb{R}^2$ with concave boundary is a cusp, i.e. made by two concave curves tangent to each other at A, then a short billiard trajectory can experience arbitrary many reflections near A, see Exercise IV.2.2. Therefore, some sort of transversality of B_i's at their intersection B is necessary.

Such a condition was found by Buraro, Federer and Kononenko (1998):

Definition. A billiard domain Q given by (5) is *nondegenerate* in a subset $U \subset \mathbb{R}^d$ with constant $C > 0$ if for any $I \subset \{1, \cdots, n\}$ and for any $y \in (U \cap Q) \setminus (\cap_{i \in I} B_i)$

$$\max_{k \in I} \frac{\operatorname{dist}(y, B_k)}{\operatorname{dist}(y, \cap_{i \in I} B_i)} \geq C$$

whenever $\cap_{i \in I} B_i$ is nonempty.

Roughly speaking, this means that if a point is d-close to all the walls from I, then it is d/C-close to their intersection.

Theorem. *Let a semidispersing billiard Q be nondegenerate in an open domain $U \subset \mathbb{R}^d$. Then for any point $x \in U$ there exist a number $M_x < \infty$ and a smaller neighborhood U_x of x such that every billiard trajectory entering U_x leaves it after making no more than M_x collisions with the boundary ∂Q.*

Let $x = (r, \phi) \in M$ and $Tx = (r_1, \phi_1)$. Denote by $\tau = \tau(x)$ the return time (travel time) between x and Tx, and by $K = K(r)$ the curvature of the boundary ∂Q at r (so that the angle between normal vectors $n(r)$ and $n(r + dr)$ equals $K\,dr + o(dr)$). Similarly, we put $K_1 = K(r_1)$. Remember our conventions on the signs of curvatures.

A detailed (but elementary) geometric analysis gives the derivative of T:

$$D_xT = -\frac{1}{\cos\phi_1}\begin{pmatrix} \tau K + \cos\phi & \tau \\ \tau K K_1 + K\cos\phi_1 + K_1\cos\phi & \tau K_1 + \cos\phi_1 \end{pmatrix}$$

Note that, since $\cos\phi_1 \neq 0$ and $\cos\phi \neq 0$, this matrix is defined and is nonsingular. Also, since the first derivative of T involves the curvature K of ∂Q (related to the second derivative of Γ_i's), then the smoothness of T is only C^{k-1}.

We now want to estimate the *mean free path*, i.e. the asymptotic of value

$$\bar{\tau}(x) = \lim_{n\to\infty} \frac{\tau(x) + \tau(Tx) + \cdots + \tau(T^{n-1}x)}{n}$$

By the Birkhoff ergodic theorem, the value $\bar{\tau}(x)$ exists a. e. in M and its average value is

$$\bar{\tau} := \int_M \bar{\tau}(x)\, d\nu(x) = \int_M \tau(x)\, d\nu(x) \qquad (6)$$

If the billiard map T is **ergodic** then, the function $\bar{\tau}(x)$ is constant almost everywhere, and it equals $\bar{\tau}$.

$$\bar{\tau} = \frac{|Q| \cdot |S^{d-1}|}{|\partial Q| \cdot |B^{d-1}|} \qquad (7)$$

It depends on the volume of Q and the surface area of its boundary, but not on its shape.

In particular, for planar billiards $d = 2$, we have $|S^1| = 2\pi$, $|B^1| = 2$, hence the formula (7) turns very simple:

$$\bar{\tau} = \frac{\pi\, |Q|}{|\partial Q|} \qquad (8)$$

The formula (7) is well known in geometric probability and integral geometry. Its planar version (8) is often referred to as Santaló formula, since it was first given in Santaló's book (Integral geometry and geometric probability, 1977).

For example, consider again a billiard table Q on a unit torus $\mathbb{T}^2$ where a small disk D of radius r is removed. Clearly, for small r the billiard particle can move freely for a long time between collisions with the disk D, and the function $\tau(x)$ can take arbitrarily large values. The Santaló formula (8) gives its mean value:

$$\bar{\tau} = \frac{\pi(1-\pi r^2)}{2\pi r} = \frac{1-\pi r^2}{2r}$$

i.e. the mean free path is asymptotically equal $\frac{1}{2r}$ as $r \to 0$. We will see later that the map T is ergodic in this example, so that $\bar{\tau}(x) = \bar{\tau}$ almost everywhere.

As a far more complicated example, consider a system of N hard balls on a unit (d-dimensional) torus. It reduces to a billiard in multidimensional domain Q whose boundary consists of cylinders (note: $\dim Q = Nd$). The mean free time between collisions can now be estimated by (7). This requires computing the volume of Q and surface area of its boundary. This is a difficult but feasible job, which was done by Chernov (1997). Quite remarkably, the final expression coincided with the classical Boltzmann's formula for the mean intercollision time used in statistical physics for decades.

Measure Preserving Maps

$(X, \mathcal{O}, \mu)$, $(Y, \mathcal{S}, \nu)$ measure spaces
The map $T : X \to Y$ is *measure preserving* if
$B \in \mathcal{S} \Rightarrow T^{-1}(B) \in \mathcal{O}$ and $\mu(T^{-1}(B)) = \nu(B)$.

The following is a simple but useful method of checking that a map is measure preserving.

Proposition. *Let* $(X, \mathcal{O}, \mu)$ *and* $(Y, \mathcal{S}, \nu)$ *be probability spaces and* $\mathcal{S}_0 \subset \mathcal{S}$ *an algebra that generates* $\mathcal{S}$ *. If* $T^{-1}(B) \in \mathcal{O}$ *and* $\mu(T^{-1}(B)) = \nu(B)$ *for all* $B \in \mathcal{S}_0$, *then* T *is measure preserving.*

Proof. Let $\hat{\mathcal{S}}$ be the family of sets $B \in \mathcal{S}$ such that $T^{-1}(B) \in \mathcal{O}$ and $\mu(T^{-1}(B)) = \nu(B)$. It is easy to see that $\hat{\mathcal{S}}$ is a $\sigma-$algebra, and it obviously contains $\mathcal{S}_0$. Then $\hat{\mathcal{S}} \supset \mathcal{S}$. Hence $B \in \mathcal{S}$ implies $T^{-1}(B) \in \mathcal{O}$ and $\mu(T^{-1}(B)) = \nu(B)$. □

Examples:

1. Translations of $\mathbb{T}^n$. For any $k = (k_1, \ldots, k_n) \in \mathbb{T}^n$ define the translation $L_k : \mathbb{T}^n \to \mathbb{T}^n$ by

$$L_k(x_1, \ldots, x_n) = (k_1 x_1, \ldots, k_n x_n)$$

2. Linear maps of $\mathbb{T}^n$. Define a map $\pi : \mathbb{R}^n \to \mathbb{T}^n = S^1 \times \cdots \times S^1$ by

$$\pi(t_1, \ldots, t_n) = (e^{2\pi i t_1}, \ldots, e^{2\pi i t_n}).$$

Clearly $\pi(x) = \pi(y) \iff x - y \in \mathbb{Z}^n$ and $\pi(\mathbb{Z}^n) = \{(1, \ldots, 1)\}$.

Given a linear isomorphism $\hat{T} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\hat{T}(\mathbb{Z}^n) \subset \mathbb{Z}^n$, or, equivalently, whose matrix in the canonical basis has integral components, there exists $T : \mathbb{T}^n \to \mathbb{T}^n$ such that $\pi \circ \hat{T} = T \circ \pi$. The map T is defined by taking, for each $p \in \mathbb{T}^n$, a point $x \in \pi^{-1}(p)$ and setting $T(p) = \pi[\hat{T}(x)]$.
This definition is consistent because if $y \in \pi^{-1}(p)$ is another point, then $x - y \in \mathbb{Z}^n$ hence $\hat{T}(x - y) \in \mathbb{Z}^n$, and so $\pi[\hat{T}(y)] = \pi[\hat{T}(x)]$.

$\hat{T}$ is called the linear lifting of the linear map T to the covering space $\mathbb{R}^n$.

Proposition. *The Lebesgue probability on* $\mathbb{T}^n$ *is invariant under all translations. Moreover, it is the unique probability on the Borel* $\sigma-$*algebra of* $\mathbb{T}^n$ *with this property.*

The Lebesgue probability on $\mathbb{T}^n$ *is invariant under every linear map of* $\mathbb{T}^n$.

3. Bernoulli Shifts. Let X be a compact metric space. Let $B(X)$ denote the space of double-sided sequences $\theta : \mathbb{Z} \to X$ endowed with metric

$$d(\alpha, \beta) = \sum_{n=-\infty}^{\infty} \frac{1}{k^{|n|}} d_0(\alpha(n), \beta(n))$$

where $k > 1$ is a constant and d_0 the metric on X. Observe that, in this metric, a sequence of sequences $\{\theta_n\} \in B(X)$ converges to a sequence $\theta \in B(X)$ if and only if it converges componentwise, i.e.

$$\lim_{n\to\infty} \theta_n(j) = \theta(j)$$

for all $j \in \mathbb{Z}$. Hence the convergence is independent of the constant $k > 1$ used to define the metric d. The *shift* $\sigma : B(X) \to B(X)$ is defined by

$$(\sigma\theta)(n) = \theta(n+1)$$

Clearly σ is a homeomorphism. When X is a finite set, $X = \{1, \ldots, m\}$, then we denote $B(\{1, \ldots, m\})$ simply by $B(m)$.

Given Borel sets $A_0, \ldots, A_m$ in X and $j \in \mathbb{Z}$

we define a cylinder $C(j, A_0, \ldots, A_m)$ by

$$\{\theta \in B(X) |\ \theta(j+i) \in A_i,\ 0 \leq i \leq m\}.$$

Finite disjoint unions of cylinders make an algebra that generates the Borel $\sigma-$algebra of $B(X)$. Moreover, given a probability μ_0 on the Borel $\sigma-$algebra of X, there exists a unique probability μ on the Borel $\sigma-$algebra of $B(X)$ (called the product measure associated with μ_0) such that for every cylinder:

$$\mu(C(j, A_0, \ldots, A_m)) = \prod_{i=0}^{m} \mu_0(A_i) \qquad (9)$$

The existence and uniqueness of μ can be deduced following a construction similar to that of Lebesgue measure on $\mathbb{R}^n$ or $\mathbb{T}^n$. Moreover, μ is invariant under σ. This follows from the fact $\mu(\sigma^{-1}(C)) = \mu(C)$ for every cylinder C, as it can be checked by using the above formula, and from the fact that finite disjoint unions of cylinders make an algebra that generates the Borel $\sigma-$algebra of $B(X)$. Denote by $B_\mu(X)$ the space $B(X)$ endowed with the probability μ. The shift $\sigma : B_\mu(X) \to B_\mu(X)$ is called a *Bernoulli shift.*

When X is a finite set, $X = \{1, \dots, m\}$, the probability μ_0 is determined by the numbers $p_i = \mu_0(\{i\})$ and in this case $B_\mu(\{1, \dots, m\})$ is simply denoted by $B(p_1, \dots, p_m)$.vspace.2cm

In a similar way we define $B^+(X)$, the space of one-sided sequences $\theta : \mathbb{Z}^+ \to X$ endowed with the metric

$$d(\alpha, \beta) = \sum_{n=0}^{\infty} \frac{1}{k^n} d_0(\alpha(n), \beta(n))$$

where $k > 1$. The *one sided shift* $\sigma : B^+(X) \to B^+(X)$ is defined also by the formula $(\sigma\theta)(n) = \theta(n+1)$, but now $n \geq 0$. The map σ on $B^+(X)$ is only a continuous surjective map, and not a homeomorphism. If μ_0 is a probability on the Borel $\sigma-$algebra of X, a product measure μ on the Borel $\sigma-$algebra of $B^+(X)$ is defined by the same formula for cylinders $C(j, A_0, \cdots, A_m)$ with $j \geq 0$. Again μ turns out to be $\sigma-$invariant. Then $B^+_\mu(X)$ and $B^+(p_1, \dots, p_m)$ are defined as in the previous case.

4. Volume Preserving Diffeomorphisms. Let U and V be open sets of $\mathbb{R}^n$. We say that a diffeomorphism $f : U \to V$ is *volume preserving* if $|\det f'(x)| = 1$ for all $x \in U$. Then

$$\lambda(f^{-1}(A)) = \lambda(A)$$

for every Borel subset $A \subset V$. For every open set $A \subset \mathbb{R}^n$ with $\lambda(\partial A) = 0$ (which implies $\lambda(\partial f^{-1}(A)) = 0$) we have

$$\lambda(f^{-1}(A)) = \int \chi_{f^{-1}(A)}(x)\, dx = \int (\chi_A \circ f)\, dx =$$

$$\int (\chi_A \circ f) |\det f'|\, dx = \int \chi_A(x)\, dx = \lambda(A)$$

(fourth identity is obtained by change of variables). So the desired formula holds, in particular, when A is a cube. Then we take a covering of V by disjoint cubes $Q_1, Q_2, \cdots$ and define the $\sigma-$algebra $\hat{\mathcal{O}}$ of all the Borel sets $A \subset V$ such that

$$\lambda(f^{-1}(A \cap Q_i)) = \lambda(A \cap Q_i) \text{ for all } i.$$

This is a $\sigma-$algebra that contains the subalgebra of disjoint unions of cubes. Hence $\hat{\mathcal{O}}$ is the Borel $\sigma-$algebra and the formula is proved.

Poincaré Recurrence Theorem

We deal first with a probabilistic version, which makes no reference to topology.

Theorem. *Let* T *be a measure-preserving map of a probability space* $(X, \mathcal{O}, \mu)$. *Given* $A \in \mathcal{O}$, *let* A_0 *be the set of points* $x \in A$ *such that* $T^n(x) \in A$ *for infinitely many* $n \geq 0$. *Then* A_0 *belongs to* $\mathcal{O}$, *and* $\mu(A_0) = \mu(A)$.

Proof. Let $C_n := \{x \in A |\ T^j(x) \notin A$ for all $j \geq n\}$. It is clear that

$$A_0 = A \setminus \cup_{n=1}^{\infty} C_n$$

Thus, the theorem will be proved if we show that $C_n \in \mathcal{O}$ and $\mu(C_n) = 0$ for every $n \geq 1$. Observe that

$$C_n = A \setminus \cup_{j \geq n} T^{-j}(A) \implies C_n \in \mathcal{O}, \text{ and}$$

$$C_n \subset \cup_{j \geq 0} T^{-j}(A) \setminus \cup_{j \geq n} T^{-j}(A). \text{ Since}$$

$$\cup_{j \geq n} T^{-j}(A) = T^{-n}(\cup_{j \geq 0} T^{-j}(A))$$

we obtain that

$$\mu(\cup_{j \geq n} T^{-j}(A)) = \mu(\cup_{j \geq 0} T^{-j}(A))$$

This implies $\mu(C_n) = 0$. □

In order to state the topological version of the recurrence theorem, we need the notion of $\omega-$limit set of a point under a map. Let X be a topological space and $T : X \to X$ a map. We define the $\omega-$limit set of a point $x \in X$ as the set of points $y \in X$ such that for every neighborhood U of y the relation $T^n(x) \in U$ holds for infinitely many positive values of n. If X is a metric space, this is equivalent to saying that

$$\liminf_{n\to\infty} \operatorname{dist}(T^n(x), y) = 0$$

Theorem. *Let X be a separable metric space and $T : X \to X$ a Borel-measurable map. Let μ be a T-invariant probability measure on the Borel $\sigma-$algebra of X. Then $\mu(\{x : x \notin \omega(x)\}) = 0$. In other words, almost every point is recurrent.*

Proof. Let $\{U_n\}_{n=0}^{\infty}$ be a basis of open sets such that

$$\lim_{n\to\infty} \operatorname{diam} U_n = 0, \quad \cup_{n\geq m} U_n = X$$

for every $m \geq 0$.

Let $\tilde{U}_n := \{x \in U_n |\ T^j(x) \in U_n$ for infinitely many positive values of $j\}$. From the preceding theorem

$$\mu(U_n \setminus \tilde{U}_n) = 0.$$

Put

$$\tilde{X} := \cap_{m=0}^{\infty} \cup_{n \geq m} \tilde{U}_n$$

It follows that

$$\begin{aligned} \mu(X \setminus \tilde{X}) &= \mu(\cup_{m=0}^{\infty}(X \setminus \cup_{n \geq m} \tilde{U}_n)) \\ &= \mu(\cup_{m=0}^{\infty}(\cup_{n \geq m} U_n \setminus \cup_{n \geq m} \tilde{U}_n)) \\ &\leq \mu(\cup_{m=0}^{\infty} \cup_{n \geq m} (U_n \setminus \tilde{U}_n)) \\ &= 0 \end{aligned}$$

Thus we only have to show that $x \in \tilde{X}$ implies $x \in \omega(x)$. Let $r > 0$. Choose m such that $\operatorname{diam} U_n \leq r/3$ if $n \geq m$. Since $x \in \tilde{X}$, it follows that $x \in \cup_{n \geq m} \tilde{U}_n$. Thus there exists $n \geq m$ such that $x \in \tilde{U}_n$. Since $\operatorname{diam} U_n \leq r/3$, it follows that $U_n \subset B_r(x)$, which implies that $T^j(x) \in B_r(x)$ if $T^j(x) \in U_n$. But since $x \in \tilde{U}_n$, $T^j(x) \in U_n$ for infinitely many values of j, showing that $x \in \omega(x)$. □

Existence of Invariant Measures

Let X be a compact metric space and $\mathcal{O}$ its Borel $\sigma-$algebra. The set $m(X)$ of probabilities $\mu : \mathcal{O} \to [0,1]$ admits a topology that is metrizable. With this topology $m(X)$ becomes a compact space. Then every continuous map $T : X \to X$ has at least one T-invariant probability, i.e. a probability $\mu \in m(X)$ such that $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{O}$. If T is not continuous it may happen than there is no T-invariant probability measure, see Exercise I.4.5. If T is continuous the most frequent situation is when T has infinitely many invariant probabilities. When T has only finitely many invariant probabilities it is clear that is has only one (because if it has two, μ_1 and μ_2, then all the linear combinations $\lambda\mu_1 + (1-\lambda)\mu_2$, $0 \le \lambda \le 1$, would be invariant probabilities).

Given a continuous map T, denote by $m_T(X)$ the set of T-invariant probabilities.

Theorem [Bogolyubov-Krylov] $m_T(X)$ *is non-empty.*

The Equivalence Problem

The natural notion of equivalence between two measure preserving maps is given by the following definition.

Definition. We say that two measure preserving maps T_i, $i = 1, 2$, of two measure spaces $(X_i, \mathcal{O}_i, \mu_i)$, $i = 1, 2$, respectively, are *equivalent* if there exists a measure preserving map F taking $(X_1, \mathcal{O}_1(\text{mod } 0), \mu_1)$ into $(X_2, \mathcal{O}_2(\text{mod } 0), \mu_2)$ satisfying

(a) F is invertible, i.e. $\exists\, G : X_2 \to X_1$, measurable, such that $GF(x) = x$ for a.e. $x \in X_1$ and $FG(y) = y$ for a.e. $y \in X_2$.

(b) F preserves measure, i.e. $\mu_1(F^{-1}(A)) = \mu_2(A)$ (mod 0) for every Borel $A \subset X_2$.

(c) $T_2F = FT_1$ for a.e. $x \in X_1$.

Observe that (a)-(b) imply that G is a measure preserving map of $(X_2, \mathcal{O}_2, \mu_2)$ into $(X_1, \mathcal{O}_1, \mu_1)$, and, by (c), $GT_2 = T_1G$ almost everywhere.

Hence, the equivalence is symmetric. Clearly it is transitive and reflexive, so it is a true equivalence relation.

One of the aims of ergodic theory is to classify measure preserving maps modulo this equivalence relation. One of the methods for this analysis consists in associating with a measure preserving map $T : (X, \mathcal{O}, \mu) \to (Y, \mathcal{S}, \nu)$ a linear operator $U_T : \mathcal{L}^2(Y) \to \mathcal{L}^2(X)$ defined by

$$U_T f = f \circ T.$$

The fact that T preserves measure implies that U_T is a unitary operator, i.e., denoting by $\langle \cdot, \cdot \rangle$ the inner product in $\mathcal{L}^2$ we have

$$\langle U_T f, U_T g \rangle = \langle f, g \rangle$$

for every $f, g \in \mathcal{L}^2(Y)$.

Ergodic Theorem of Birkhoff-Khinchin

Birkhoff-Khinchin theorem deals with the distribution of the orbits of a measure preserving map T of a probability space $(X, \mathcal{O}, \mu)$. In order to study how an orbit $\{x, T(x), T^2(x), \ldots\}$ is asymptotically distributed in X we introduce the *sojourn time* of x in a set $A \in \mathcal{O}$ by

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \#\{0 \le m < n \,|\, T^m(x) \in A\}$$

Birkhoff's Theorem states that this limit exists for a.e. x and that $\tau(x, A)$ is an integrable function of x whose integral is given by

$$\int_X \tau(x, A)\, d\mu(x) = \mu(A)$$

Moreover, as a function of x, τ is T-invariant, i.e.

$$\tau(x, A) = \tau(T(x), A) \qquad \text{a.e.}$$

This motivates the following definition: T is ergodic if all the T-invariant functions are constant a.e. Then, if T is ergodic , for a.e. x we have:

$$\tau(x, A) = \mu(A)$$

This remarkable conclusion poses the problem: develop methods to decide when a map T is ergodic.

Moreover, when X is a compact metric space and $T : X \to X$ is a continuous map, there always exist ergodic T-invariant probabilities on the Borel $\sigma-$algebra of X. They are important in the analysis of the dynamics of T.

An interesting example of an ergodic measure preserving map is the map $T : [0,1] \to [0,1]$ given by $T(x) = 10x - [10x]$. This map preserves the Lebesgue probability λ on [0,1] and is ergodic. A direct consequence of its ergodicity is the following important fact in number theory. Write $x \in [0,1]$ in decimal representation $x = 0.a_0 a_1 a_2 \ldots$ and let $N_n(x,j)$ be the number of times that the digit $0 \leq j \leq 9$ appears in the string $[a_0 \ldots a_{n-1}]$. Then, for a.e. $x \in [0,1]$

$$\lim_{n\to\infty} \frac{1}{n} N_n(x,j) = \frac{1}{10}$$

This is a consequence of the ergodicity of T because $a_m = j$ if and only if $T^m(x) \in [j/10,(j+1)/10)$. Then, for a.e. x :

$$\lim_{n\to\infty} \frac{1}{n} N_n(x,j) = \tau(x,[j/10,(j+1)/10))$$
$$= \lambda([j/10,(j+1)/10)) = 1/10.$$

More subtle is a similar property for continued fractions. Every irrational number $x \in (0,1)$ can be written in a unique way as a continued fraction

$$x = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}}$$

where $a_0, a_1, \ldots$ are positive integers. Let $P_n(x,k)$ be the number of times that k appears among $a_0, \ldots, a_{n-1}$. Then, for a.e. $x \in (0,1)$

$$\lim_{n\to\infty} \frac{1}{n} P_n(x,k) = \frac{1}{\log 2} \log(1 + \frac{1}{k(k+2)})$$

The proof of this property requires first transforming $P_n(x,k)$ into a sojourn time. This is done with the help of the Gauss map $T : [0,1] \to [0,1]$ defined by

$$T(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x}\right] & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Theorem [Birkhoff-Khinchin] *Let* $(X, \mathcal{O}, \mu)$ *probability space;* $T : X \to X$ *preserves* μ*;* $f : X \to \mathbb{R}$ *integrable function, then*

$$\textit{the limit} \qquad \tilde{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \quad (10)$$

exists for a.e. $x \in X$, $\tilde{f}$ *(the* time average *of* f. *) is* T*-invariant, integrable,* $\|\tilde{f}\|_1 \leq \|f\|_1$,

$$\int_X \tilde{f}\, d\mu = \int_X f\, d\mu, \quad \text{and}$$

$$\lim_{n\to\infty} \int_X \left| \tilde{f} - \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu = 0$$

Sojourn times are time averages because

$$\#\{0 \leq j < n : T^j(x) \in A\} = \sum_{j=0}^{n-1} \chi_A(T^j(x))$$

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j(x)) = \tilde{\chi}_A(x)$$

Hence we have

$$\int \tau(x, A)\, d\mu(x) = \int \tilde{\chi}_A\, d\mu = \int \chi_A\, d\mu = \mu(A).$$

Ergodic Hierarchy

It can be proved that a measure-preserving map T of a probability space $(X, \mathcal{O}, \mu)$ is ergodic if and only if, for every $A, B \in \mathcal{O}$

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=0}^{m-1} \mu(T^{-j}(A) \cap B) = \mu(A)\mu(B)$$

In this case, if the limit of $\mu(T^{-j}(A) \cap B)$ as $j \to \infty$ exists, its value must be equal to $\mu(A)\mu(B)$.

Definition. A measure preserving endomorphism T of a probability space $(X, \mathcal{O}, \mu)$ is said to be *mixing* if for any pair $A, B \in \mathcal{O}$

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

Caipirinha. A pictorial example in the classical book by Arnold and Avez explains what a mixing map does. Suppose a cocktail shaker M, $\mu(M) = 1$ is filled by 85% of cachaa and 15% of lemon juice. Let A be the part of the cocktail shaker originally occupied by the juice and B any part of the shaker. Let $T\colon M \to M$ be the transformation of the content of the shaker made during one move by the bartender (who is shaking the cocktail repeatedly). Then after n moves the fraction of juice in the part B will be $\mu(T^n(A) \cap B)/\mu(B)$. As the bartender keeps shaking the cocktail ($n \to \infty$), the fraction of juice in any part B approaches $\mu(A) = 15\%$, i.e. the lemon juice will spread uniformly in the mixture.

We note that the definition we gave for a mixing map is good for both invertible and noninvertible maps (endomorphisms).

Proposition. *Any mixing map is ergodic.*

Proof. Let A be any T-invariant measurable set, then $T^{-n}(A) = A$ and $\mu(A \cap B) = \lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$. In particular, for $A = B$ we have $\mu(A) = \mu^2(A)$. This means $\mu(A) = 0$ or 1, hence T is ergodic. □

We note that not all ergodic maps are mixing, see Exercise II.4.1. Therefore, mixing is a stronger property than ergodicity.

Definition. If X is a topological space, a transformation $T : X \to X$ is *topologically mixing* if for any pair of open sets $U, V \subset X$ there exist $N \in \mathbb{N}$ such that $T^{-n}(U) \cap V \neq \emptyset$ for all $n \geq N$.

Proposition. *If* X *is a topological space,* $\mathcal{O}$ *is the Borel* σ*-algebra and* μ *a probability measure positive on open sets, then whenever* $T : X \to X$ *is mixing, it is also topologically mixing.*

Proof. Since, for any open sets U, V we have $\lim_{n\to\infty} \mu(T^{-n}(U) \cap V) = \mu(U)\mu(V) \neq 0$ it results that, for all $n \geq N$, $\mu(T^{-n}(U) \cap V) \neq 0$, hence $T^{-n}(U) \cap V \neq \emptyset$. □

Definition. A measure preserving automorphism T of a probability space $(X, \mathcal{O}, \mu)$ is Bernoulli if it is equivalent to a Bernoulli shift (the shifts can be defined on a probability space).

Ergodic hierarchy (discussed in Mañé' book):

$$\text{Bernoulli} \Rightarrow \text{K} - \text{mixing} \Rightarrow \text{Mixing} \Rightarrow \text{Ergodic} \tag{11}$$

Each word in this row represents the set of measure preserving maps of a probability space $(X, \mathcal{O}, \mu)$ that satisfy the corresponding definition.

All the implications in (11) are one-way only, none can be reversed. This means that ergodicity does not imply mixing, mixing does not imply K-mixing, etc.

Statistical Properties of Dynamical Systems

For those familiar with probability theory: In the language of probability theory, a Bernoulli shift represents a sequence of independent random variables. This immediately follows from the formula that defines the product measure. For this reason the Bernoulli property is regarded as a *statistical* property of a dynamical system. It establishes an equivalence between a dynamical system and a purely random sequence of independent trials - a canonical model in probability theory.

This a a very interesting and important observation. A dynamical system $T : X \to X$ is, by nature, completely deterministic. This means that if you have a point $x \in X$, its entire future $\{T^n x\}$, $n \geq 1$, is uniquely determined and can be computed precisely. When the map T is invertible, the past $\{T^n x\}$, $n \leq -1$, is uniquely determined and computable, too. One can look at it this way: knowing the *present state* of a dynamical system (given by $x \in X$), one can determine its future and, often, its past. This is the precise meaning we give to the word “deterministic”.

On the other hand, in a sequence of independent trials the outcome of any trial gives no clue of what the outcomes of other trials would be (or have been).

While the Bernoulli property is a manifestation of randomness or chaoticity, strangely, it has little relevance to direct physical applications. Why? Because the equivalence between a dynamical system and a Bernoulli shift is, usually, given by just a measurable map with a very complicated structure, not at all smooth or even continuous. In physics, on the other hand, the laws of motion are usually specified by differential equations (like Newton or Hamiltonian equations), and all interesting functions (such as temperature, energy, pressure) are smooth as well. Hence, only the properties of dynamical systems expressed by smooth maps and smooth functions are relevant in physics.

For these reasons, assuming that X is a manifold, $T : X \to X$ a smooth map preserving a probability μ and $f : X \to \mathbb{R}$ a smooth function, one can characterize the system in a physically meaningful way as follows.

Consider

$$S_n = f + f \circ T + f \circ T^2 + \cdots + f \circ T^{n-1} \quad (12)$$

The quotient S_n/n is called the *time average* of the function f. Adopting physical notation, we denote by $\langle\cdot\rangle$ the expected value of a function with respect to μ, e.g. $\langle f\rangle = \int_X f\, d\mu$. The integral $\langle f\rangle$ is also called the *space average* of f.

Now the Birkhoff Ergodic Theorem asserts that if T is ergodic, then S_n/n converges almost everywhere to $\langle f\rangle$ as $n \to \infty$. In physical language, it means that **time averages converge to space averages**. In probability theory, this fact is also called the *strong law of large numbers*.

An important characteristic of a dynamical system is the *time correlation function*

$$C_f(n) = \langle f \cdot (f \circ T^n)\rangle - \langle f\rangle^2. \quad (13)$$

If the map T is mixing, one can show that $C_f(n) \to 0$ as $n \to \infty$, i.e. the correlations *decay*, as physicists call it. The asymptotic speed of convergence $C_f(n) \to 0$ characterizes the "speed of mixing" in the system.

Next, we say that f satisfies the *central limit theorem* if

$$\lim_{n\to\infty} \mu\left\{x\colon \frac{S_n(x) - n\langle f\rangle}{\sqrt{n}} < z\right\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{z} e^{-\frac{s^2}{2\sigma^2}}\,ds \text{ for all } -\infty < z < \infty. \tag{14}$$

$$\sigma = \sigma_f \geq 0, \qquad \sigma_f^2 = C_f(0) + 2\sum_{n=1}^{\infty} C_f(n) \tag{15}$$

Equation (14) is equivalent to the convergence of $(S_n - n\langle f\rangle)/\sqrt{n}$ in distribution to the normal random variable $N(0, \sigma_f^2)$. The central limit theorem is considerably more refined than the Birkhoff Ergodic Theorem: the distribution of the deviations of the time average S_n/n from its limit value $\langle f\rangle$, when scaled by $1/\sqrt{n}$, is asymptotically Gaussian.

It is clear from (15) that the central limit theorem only holds if $\sum_n |C_f(n)| < \infty$. Actually, the proof of this theorem requires even a more rapid convergence of $C_f(n)$ to zero. For these (and other) reasons the speed of that convergence is regarded as an important statistical characteristic of the system.

Two main types of convergence are *exponential*, when $|C_f(n)| <$ const$\cdot e^{-an}$, $a > 0$ ("most chaotic", possess many features necessary for applications in statistical physics), and
polynomial: $|C_f(n)| <$ const $\cdot\, n^{-b}$, $b > 0$ (are regarded as being intermediate ("intermittent") between "regular" and "chaotic", and their behavior is very sensitive to the exact value of the power $b > 0$ and other factors).

If one relaxes the requirement that the function f in (13) be smooth, then one totally loses control over the decay of correlations. In all known mixing dynamical systems, the convergence $C_f(n) \to 0$ is indeed arbitrarily slow for generic integrable functions, even for generic continuous functions. So, the smoothness of f is essential.

Quite surprisingly, for many interesting dyn. sys. one can actually prove the above central limit theorem, and obtain good estimates on the decay of correlations for smooth functions f. This opens the door to close interaction between the theory of dyn. sy.s and prob. th- and statistical mechanics, which is currently a very active area of research.

Definition. Lyapunov exponents. f^n is differentiable at a point $p \in M$ for all $n \in \mathbb{Z}$. Assume $\mathcal{T}_p M = E_1 \oplus \cdots \oplus E_{m(p)}$; if $\vec{0} \neq v_i \in E_i$,

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \|(f^n)'_p v_i\| = \lambda_i(p) \qquad (16)$$

Then the values $\lambda_i(p)$ are called *Lyapunov exponents* at the point p, whose multiplicities are $\dim E_i$.

The existence of the limit (16) is not guaranteed for any point $p \in A$, we will return to this issue in the next section. For now, we will say that p has all Lyapunov exponents if the above limits exist.

If a point p has all Lyapunov exponents and none of them is zero, we call p a *hyperbolic point.* For a hyperbolic point $p \in M$, we have $\mathcal{T}_p M = E_p^s \oplus E_p^u$,

$$E_p^s = \oplus_{\lambda_i(p)<0} E_i \quad \text{and} \quad E_p^u = \oplus_{\lambda_i(p)>0} E_i \quad (17)$$

Existence of submanifolds $W^s(p)$, $W^u(p)$

$$\operatorname{dist}(f^n(y), f^n(p)) \leq Ce^{-\lambda n} \qquad y \in W^s(p)$$

and

$$\operatorname{dist}(f^{-n}(y), f^{-n}(p)) \leq Ce^{-\lambda n} \qquad y \in W^u(p)$$

for all $n \geq 0$ and some constant $C > 0$.

Forward orbits of the points of $W^s(p)$ are getting close to each other (converge) exponentially fast: $W^s(p)$ is called the *stable manifolds* (the term comes from differential equations, where the convergence of solutions is interpreted as stability). Forward orbits of the points of $W^u(p)$ get separated (diverge) exponentially fast: $W^u(p)$ is called the *unstable manifold*. Note, though, that the backward orbits of $W^u(p)$ converge, and the backward orbits of $W^s(p)$ diverge.

Orbits of all the points near a hyperbolic point p are very unstable: diverge (get separated) exponentially fast either in the future or in the past, or both. If $\dim W^s(p) \neq 0$ and $\dim W^u(p) \neq 0$; y close to p not exactly lying on $W^u(p)$ or $W^s(p)$, the trajectory of y separates from that of p both in future and in past! Exponential separation of trajectories is the main source of instability, turbulence, mixing – all that we call *chaos*.

J. Hadamard (around 1900): hyperbolicity for geodesic flows on manifolds of constant negative curvature. J. G. Hedlund and E. Hopf (1930's): ergodic properties of these flows. Based on the studies of geodesic flows, in the 1960s D. Anosov (S. Smale, in a different form) introduced general classes of diffs with hyp. points.

Theorem[Oseledec] *Assume that the map* $f : N \to V$ *preserves a Borel probability measure* μ *on* M *and* $\mu(H) = 1$. *If*

$$\int_M \log^+ \|(f^{\pm 1})'_p\| \, d\mu(p) < \infty$$

where $\log^+ s = \max\{\log s, 0\}$, *then there exists an* f*-invariant set* $E \subset H$, $\mu(E) = 1$, *such that for every point* $p \in E$ *all the Lyapunov exponents exist.*

For $p \in \Gamma$ we denote by $\lambda_1(p) < \lambda_2(p) < \cdots < \lambda_{m(p)}(p)$ all distinct Lyapunov exponents and by $E_1(p), \ldots, E_{m(p)}(p)$ the corresponding subspaces in the tangent space $\mathcal{T}_p M$. For any real number $\kappa \in \mathbb{R}$ and $p \in \Gamma$ denote

$$E^-_{p,\kappa} = \oplus_{\lambda_i(p) \le \kappa} E_i(p) \qquad E^+_{p,\kappa} = \oplus_{\lambda_i(p) > \kappa} E_i(p)$$

Remark. Oseledec's theorem also includes the following fact, which we state separately. Let $\gamma_\kappa(p)$ denote the angle between the spaces $E^-_{p,\kappa}$ and $E^+_{p,\kappa}$. Then

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \gamma_\kappa(f^n(p)) = 0 \qquad (18)$$

i.e. the angle $\gamma_\kappa(f^n(p))$ slowly changes with n (more slowly than any exponential function).

We now make two additional assumptions made by Katok and Strelcyn:

(KS1) There are constants $C_1 > 0$ and $a > 0$ such that for all $\varepsilon > 0$ the μ-measure of the ε-neighborhood of S satisfies

$$\mu(U_\varepsilon(S)) \leq C_1\,\varepsilon^a$$

i.e. the measure μ does not build up too much near the singularity set S.

(KS2) There are constants $C_2 > 0$ and $b > 0$ such that for every $x \in N$ and $v \in \mathcal{T}_x N$, $\|v\| \leq r(x, N)$ we have

$$\|f''_{ox}(v)\| \leq C_2\, d(\exp_x(v), S)^{-b}$$

i.e. the second derivative f''_{ox} does not grow too fast near the singularity set S.

The set of all hyperbolic points in N is often called the *Pesin region* of f: $\Sigma(f) =$

$$\{x \in H : \lambda_i(x) \neq 0, \text{ for every } i = 1, \ldots, m(x)\}$$

Pesin region $\Sigma(f)$ is invariant under f. At $x \in \Sigma(f)$ we have the usual subspaces E^s_x and E^u_x; let

$$\lambda^+(x) = \min\{\lambda_i(x) > 0\} \quad \text{and}$$

$$\lambda^-(x) = \max\{\lambda_i(x) < 0\}$$

Theorem. *For μ-almost every $x \in \Sigma(f)$ there is a (small) neighborhood $N_{\delta(x)}$ such that for any small $\varepsilon > 0$ the set*

$$W^s(x) = \{y \in N_{\delta(x)}:$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{dist}(f^n(x), f^n(y)) \leq \lambda^-(x) + \varepsilon\}$$

is a C^r differentiable (stable) manifold.

$$\text{Similarly,} \quad W^u(x) = \{y \in N_{\delta(x)}(0):$$

$$\liminf_{n\to-\infty} \frac{1}{n} \log \operatorname{dist}(f^n(x), f^n(y)) \geq \lambda^+(x) - \varepsilon\}$$

is a C^r differentiable (unstable) manifold. We also have

$$\mathcal{T}_x W^s(x) = E^s_x \qquad \text{and} \qquad \mathcal{T}_x W^u(x) = E^u_x$$

and the angle $\gamma(f^m x)$ between $E^s_{f^m x}$ and $E^u_{f^m x}$ satisfies

$$\gamma(f^m x) \geq C^{-1}(x, \varepsilon) e^{-\varepsilon|m|}$$

Stable and unstable manifolds can be efficiently used in the study of ergodic properties of the map f. Let $B \subset H$ be an f-invariant set such that

$\mu(B) > 0$. Then the map $f_B := f|_B$ (the restriction of f to B) preserves the probability μ_B, which is obtained by conditioning the measure μ on B. We recall that the map $f_B: B \to B$ is ergodic iff every function $g \in \mathcal{L}^1(B)$ that is f_B-invariant is constant almost everywhere on B.

A classical method to construct a set B on which the map f_B is ergodic (with respect to the measure μ_B) uses stable and unstable manifolds and goes back to E. Hopf. Take any point $x \in \Sigma(f)$ and its stable and unstable manifolds $W^s(x)$, $W^u(x)$. Let g be an f-invariant function, and assume, for simplicity, that it is continuous on M (integrable functions then can be approximated by continuous ones, but this step is purely technical, and we omit it). Then one can show that g is constant on $W^u(x) \cup W^s(x)$, see Exercise III.3.3. Since this fact applies to any other point of the set $\Sigma(f)$ as well, we can proceed as follows. Start by fixing $x \in \Sigma(f)$ and construct a set B_1 (a first approximation to B) as the union of all unstable manifolds $W^u(y)$, $y \in W^s(x)$, and all stable manifolds $W^s(z)$, $z \in W^u(x)$. The function g then must be constant on B_1.

We can defining sets B_n, $n \geq 2$, recursively by

$$B_n = \cup\{W^u(y) \cup W^s(y) : \ y \in B_{n-1}\}$$

One can easily show, by induction on n, that the function g is constant on B_n for each n. We now put $B_\infty = \cup_n B_n$. Clearly, the function g is constant on the entire set B_∞. Then it is enough to put

$$B = \cup_{n=-\infty}^{\infty} f^n(B_\infty), \tag{19}$$

the map $f_B : B \to B$ will be ergodic and B, is called an *ergodic component* of the map f.

One can verify that B_∞ consists of all the points $y \in \Sigma(f)$ for which there exists a finite sequence $x = z_0, z_1, \ldots, z_{k-1}, z_k = y$ with the property that for all $0 \leq i \leq k-1$ either $W^s(z_i) \cap W^u(z_{i+1}) \neq \emptyset$ or $W^u(z_i) \cap W^s(z_{i+1}) \neq \emptyset$ (clearly, all $z_i \in B_\infty$). In other words, for any $y \in B_\infty$ there is a chain of stable and unstable manifolds that joins the point y with the original point x. Such a chain is usually called *Hopf chain* or a *zig-zag*. Now one can say that the set B_∞ is the union of all Hopf chains (or zig-zags) starting at x.

Now let us look how big the set B_∞ is. The stable and unstable manifolds, $W^s(y)$ and $W^u(y)$, are transversal to each other at any $y \in \Sigma(f)$, i.e. the angle between them is positive. Note also that $\dim W^s(y) + \dim W^u(y) = \dim M$ by hyperbolicity. Assume for simplicity that all the stable and unstable manifolds $W^s(y)$, $W^u(y)$ in a neighb. of x are large enough, say,

$$\mathrm{dist}(y, \partial W^s(y)) \geq c \quad \text{and} \quad \mathrm{dist}(y, \partial W^u(y)) \geq c$$

for some constant $c > 0$ and all y close to x. Then one can prove that the set B_∞ contains an open neighborhood of x in the Pesin region $\Sigma(f)$. This fact is called *local ergodicity* (or sometimes local ergodic theorem).

In systems with singularities, though, stable and unstable manifolds can be arbitrarily short. This happens for the same reason as why they sometimes fail to exist at all. Hence, B_∞ may not cover any open neighborhood of x in $\Sigma(f)$, i.e. there might be some tiny islands left out arbitrarily close to x. But it is still possible to show that the set B_∞ has a positive μ measure. The following theorem given without proof summarizes our discussion:

Theorem [Pesin, 1977]. *Let* $\mu(\Sigma(f)) > 0$. *There exist sets (ergodic components)* $\Sigma_i \subset \Sigma(f)$, $i = 0, 1, 2, \ldots, J$ $(J \leq +\infty)$ *such that*

(i) $\Sigma_i \cap \Sigma_j = \emptyset$ *for* $i \neq j$ *and* $\cup_i \Sigma_i = \Sigma(f)$;

(ii) $\mu(\Sigma_0) = 0$ *and* $\mu(\Sigma_i) > 0$ *for* $i > 0$;

(iii) $f(\Sigma_i) = \Sigma_i$ *for* $i \geq 0$;

(iv) $f|_{\Sigma_i}$ *is erg. with respect to* μ_{Σ_i} *for* $i > 0$.

Furthermore, for every $i > 0$ *we have* $\Sigma_i = \Sigma_{i,1} \cup \cdots \cup \Sigma_{i,J(i)}$ *with some* $1 \leq J(i) < \infty$ *such that*

(v) $\Sigma_{i,j} \cap \Sigma_{i,k} = \emptyset$ *for* $j \neq k$;

(vi) $f(\Sigma_{i,j}) = \Sigma_{i,j+1}$ *for* $1 \leq j < J(i)$ *and* $f(\Sigma_{i,J(i)}) = \Sigma_{i,1}$;

(vii) *the map* $f^{J(i)}$ *restricted to* $\Sigma_{i,j}$ *is* K *– mixing for every* $1 \leq j \leq J(i)$.

According to a tradition, the partition of $\Sigma(f)$ into the sets $\Sigma_{i,j}$ is called *spectral decomposition.*

Definition. If $\mu(\Sigma(f)) = 1$, that is, if the Pesin region has full measure in N, we will say the map f is *nonuniformly hyperbolic*, or has *chaotic behavior.*

Piecewise smooth maps. M contained in a Riemannian manifold of dimension $d \geq 2$,

$$M = B_1^+ \cup \cdots \cup B_r^+ = B_1^- \cup \cdots \cup B_r^-$$

$B_i^\pm$ are compact, int $B_i^\pm$ are connected and dense in $B_i^\pm$, $\partial B_i^\pm$ are made by finitely many $d-1$ dimensional submanifolds which overlap at most on their boundaries. The map Φ is defined separately on each domain B_i^+, $1 \leq i \leq r$, so that Φ is a C^k ($k \geq 2$) diffeomorphism of the interior of B_i^+ onto the interior of B_i^- and a homeomorphisms of B_i^+ onto B_i^-. Then we have $N = \cup_i \operatorname{int} B_i^+$, $\Phi(N) = \cup_i \operatorname{int} B_i^-$

μ is a Φ-inv. prob. measure abs. cont. w.r. to the Riemannian meas. on M, with bounded dens.

$S^+ = \partial N$ the singularity set for the map Φ.
$S^- = \partial \Phi(N)$ the singularity set for the map Φ^{-1}.

$$S_n^+ = S^+ \cup \Phi^{-1}(S^+) \cup \cdots \cup \phi^{-n+1}(S^+), \ n \geq 1$$

is the singularity set for Φ^n, and similarly

$$S_n^- = S^- \cup \Phi(S^-) \cup \cdots \cup \Phi^{n-1}(S^-)$$

is the singularity set for ϕ^{-n}.