Introduction to Ergodic Theory Mon=15:30–16:50=108 HSE [2 декабря 2019 г.]
Literature:
– M. Blank. "Stability and localization in chaotic dynamics", MCCME, Moscow, 2001.
– I.P. Kornfeld, Ya.G. Sinai, S.V. Fomin. "Ergodic theory", Nauka, Moscow, 1980.
– A. Katok, B. Hasselblatt. "Introduction to the modern theory of dynamical systems", Cambridge Univ. Press, 1995.

Lecture 1. Birkhoff ergodic theorem [02.12.19] In $\mathbf{L}^1(X, \mathcal{B}, \mu)$ we introduce an operator: $S_n\varphi := \frac{1}{n}\sum_{k=0}^{n-1}\varphi \circ T^k$, and set $S^-\varphi := \liminf_{n\to\infty} S_n\varphi$, $S^+\varphi := \limsup_{n\to\infty} S_n\varphi$. The following result shows that $S^\pm$ are well defined.

Theorem 0.1 *(Birkhoff ergodic theorem) $\forall \mu \in \mathcal{M}_T, \varphi \in \mathbf{L}^1(X, \mathcal{B}, \mu)$ $S_n\varphi \stackrel{n\to\infty}{\longrightarrow} \tilde{\varphi} \in \mathbf{L}^1(X, \mathcal{B}, \mu)$, where $\tilde{\varphi}$ is T-invariant and $\mu(\tilde{\varphi}) = \mu(\varphi)$. Moreover, if $\mu \in \mathcal{M}_T^0$, then $S_n\varphi \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)$.*

Corollary 0.1 *Choosing $\varphi := 1_A$ we get $\#\{0 \le k \le n-1 : \ T^kx \in A\}/n \stackrel{n\to\infty}{\longrightarrow} \mu(A)$ for $\mu \in \mathcal{M}_T^0$ $\forall A \in \mathcal{B}$ and μ-a.a. $x \in X$.*

In other words, if $\mu \in \mathcal{M}_T^0$ then the time average is equal to the space average μ-a.e.

Proof (formal and non-constructive) based on the notion of a conditional mathematical expectation $E_\mu(f|\mathcal{B}_T)$ wrt the σ-algebra of T, μ-invariant sets $\mathcal{B}_T := \{B \in \mathcal{B} : \ \mu(T^{-1}B\Delta B) = 0\}$. Note that $f \circ T = f$ μ-a.e. iff f is $\mathcal{B}_T$-measurable. Definition of $E_\mu(f|\mathcal{B}_T)$: $\mu(E_\mu(f|\mathcal{B}_T), \ B) = \mu(f, B)$ $\forall B \in \mathcal{B}_T$, $f \in \mathbf{L}^1(X, \mathcal{B}, \mu)$.

The idea is to show that $S^\pm\varphi = E_\mu[\varphi|\mathcal{B}_T]$. We begin as follows: for $\psi \in \mathbf{L}^1$ we define a sequence $\Psi_n := \max_{k=1}^n\{kS_k\psi\} \in \mathbf{L}^1$. Obviously, this sequence is non-decreasing and thus either diverges to ∞ or otherwise converges. Let $B_\psi := \{x \in X : \ \sup_{n\ge1}\Psi_n(x) < \infty\} \in \mathcal{B}$, then on its complement B_ψ^c this sequence diverges. From the identity $\Psi_{n+1} - \Psi_n \circ T = \Psi_1 - \min\{0, \Psi_n \circ T\}$ we deduce that $B_\psi \in \mathcal{B}_T$. We have $0 \le \mu(1_{B_\psi^c} \cdot (\Psi_{n+1} - \Psi_n)) = \mu(1_{B_\psi^c} \cdot \underbrace{(\Psi_{n+1} - \Psi_n \circ T)}_{\psi - \min\{0, \Psi_n\circ T\}}) \stackrel{n\to\infty}{\longrightarrow} \mu(1_{B_\psi^c} \cdot \psi) = \mu(1_{B_\psi^c} \cdot E_\mu[\psi|\mathcal{B}_T])$ by dominated convergence and the definition of conditional expectation, respectively. Thus, if $E_\mu[\psi|\mathcal{B}_T] < 0$ on X we must have $\mu(B_\psi^c) = 0$.

Now we ready to complete the proof. Take any $\varphi \in \mathbf{L}^1, \varepsilon > 0$ and set $\psi := \varphi - E_\mu[\varphi|\mathcal{B}_T] - \varepsilon$. Then on B_ψ the relation $S^+\psi \le 0$ holds (otherwise, $\Psi_n(x)$ should diverge), therefore,
$0 \ge S^+\psi = \limsup_{n\to\infty}(S_n\varphi - E_\mu[\varphi|\mathcal{B}_T] - \varepsilon) = S^+\varphi - E_\mu[\varphi|\mathcal{B}_T] - \varepsilon$ μ-a.e. A similar argument for $\tilde{\psi} := -\varphi + E_\mu[\varphi|\mathcal{B}_T] - \varepsilon$ shows that $S^-\varphi - E_\mu[\varphi|\mathcal{B}_T] + \varepsilon \ge 0$ μ-a.e. Thus $S^-\varphi = S^+\varphi = E_\mu[\varphi|\mathcal{B}_T]$ μ-a.e. (since $\varepsilon > 0$ is arbitrary) and hence the limit exists. If $\mu \in \mathcal{M}_T^0$ then $\tilde{\varphi} =$ Const μ-a.e.by the invariance of $\tilde{\varphi}$ from where the claim follows. □

Corollary 0.2 *(von Neumann) $S_n\varphi \stackrel{n\to\infty}{\longrightarrow} \tilde{\varphi}$ in $\mathbf{L}^1$.*

Proof. If $|\varphi| < \infty$ then the claim follows from above result. For a general φ choose a bounded function $\varphi_\varepsilon : \ ||\varphi - \varphi_\varepsilon||_{\mathbf{L}^1} \le \varepsilon$. Then $||S_n\varphi - E_\mu[\varphi|\mathcal{B}_T]||_{\mathbf{L}^1} \le ||S_n\varphi_\varepsilon - E_\mu[\varphi_\varepsilon|\mathcal{B}_T]||_{\mathbf{L}^1} + 2\varepsilon$. Now since $\varepsilon > 0$ is arbitrary we get the result. □

Theorem 0.2 *(Structure of the set of invariant measures) Let $\mu, \mu' \in \mathcal{M}_T$. Then*
(a) $\mu \in \mathcal{M}_T^0$ and $\mu' \ll \mu$ (absolutely continuous[1]) yields $\mu = \mu'$;
(b) $\mu, \mu' \in \mathcal{M}_T^0$ yields either $\mu = \mu'$ or $\mu \perp \mu'$ (mutually singular[2]).

Proof. $\lim_{n\to\infty} S_n 1_A = \mu(A)$ μ-a.e. $\forall A \in \mathcal{B}$. Since $\mu' \ll \mu$ this equality holds also μ'-a.e. Now by Lebesgue theorem on the limit transition for uniformly bounded sequences of functions we have $\lim_{n\to\infty}\mu'(S_n1_A) = \mu'(\mu(A)) = \mu(A)$. On the other hand, $\mu'(1_A \circ T) = \mu'(1_A)$ (since $\mu' \in \mathcal{M}_T$), thus $\mu'(S_n1_A) = \mu'(A)$ and, hence, $\mu(A) = \mu'(A)$. The statement (a) follows due to an arbitrary choice of $A \in \mathcal{B}$.

To prove (b) assume that $\mu \ne \mu'$, i.e. $\mu(A) \ne \mu'(A)$ for some $A \in \mathcal{B}$. Denote $B := \{x \in X : \ S_n1_A(x) \stackrel{n\to\infty}{\longrightarrow} \mu(A)\}$ and $B' := \{x \in X : \ S_n1_A(x) \stackrel{n\to\infty}{\longrightarrow} \mu'(A)\}$. By Theorem 0.1 we have $\mu(B) = \mu'(B') = 1$. On the other hand, by the construction $B \cap B' = \emptyset \Longrightarrow \mu \perp \mu'$. □

Let us extend the action of the family of operators S_n to the space of measures: $S_n^*\mu := \frac{1}{n}\sum_{k=0}^{n-1}T^{*k}\mu$. Then $\forall\varphi \in \mathbf{L}^1$ we have $S_n^*\mu(\varphi) = \frac{1}{n}\sum_{k=0}^{n-1}T^{*k}\mu(\varphi) = \frac{1}{n}\sum_{k=0}^{n-1}\mu(\varphi \circ T^k) = \mu(S_n\varphi)$ which explains the notation. Let $\mu \in \mathcal{M}_T^0$, hence, by Birkhoff Theorem $S_n^*\delta_x \stackrel{n\to\infty}{\longrightarrow} \mu$ μ-a.e., where δ_x stays for the Dirac (δ)

---

[1] i.e., $\frac{d\mu'}{d\mu} \in \mathbf{L}^1 \Longleftrightarrow \mu(A) = 0 \Longrightarrow \mu'(A) = 0 \ \forall A \in \mathcal{B}$
[2] i.e., $\exists A \in \mathcal{B} : \ \mu(A) = \mu'(X \setminus A) = 0$

measure at the point x. Indeed, $(S_n^*\delta_x)(\varphi) = S_n\varphi(x) \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)$ μ-a.e. We shall say that a point x (and the corresponding trajectory) is T,μ-*typical* if $S_n^*\delta_x \stackrel{n\to\infty}{\longrightarrow} \mu$. By Birkhoff Theorem μ-a.a. points are T,μ-typical if $\mu \in \mathcal{M}_T^0$. *Question*: are there typical points for a general invariant measure $\mu \in \mathcal{M}_T$?
Applications:
(1) For m-a.a. $x \in [0,1]$ the average number of zeros in the decimal expansion $x = 0.x_1x_2\dots$ is equal to $1/10$. Indeed, let $T_{10}x := \{10x\}$, hence the Lebesgue measure is invariant and ergodic (similar to T_2). Let $A_0 := [0,1/10]$, then $x_i = 0 \iff T^ix \in A_0$. Thus the result follows by applying the Corollary 0.1. A number is called *normal* if in its expansion in any integer base b the density of any 'pattern' of digits (from the set $\{0,1,\dots,b-1\}$) of length ℓ is equal to $b^{-\ell}$. Prove that m-a.e. real numbers are normal. The proof is almost the same: for the map $T_bx := \{bx\}$ each pattern of digits corresponds to an interval (or a union of several intervals if the pattern has 'holes'). It remains to observe that for any b the Lebesgue measure of the exceptional set is zero.
(2) A measure $\mu \in \mathcal{M}_T^0$ is *uniquely ergodic* if this is the only invariant measure. If $T \in \mathbf{C}^0$ this is equivalent to the statement that $S_n\varphi \stackrel{n\to\infty}{\longrightarrow}$ Const $\forall\varphi \in \mathbf{C}(X)$ and the convergence is either pointwise or uniform.
Proof. The family of functions $S_n\varphi$ is equicontinuous. Thus by the Arzela-Ascoli Theorem there exists its uniformly converging subsequence. On the other hand, the limit point is defined by the ergodic theorem.
(4) *Benford Law*: calculate the density of n such that the decimal number 2^n starts from the digit $b \in \{1,\dots,9\}$. Observe that this is the case iff $\exists k \in \mathbb{Z}_+$ such that $10^kb \le 2^n < 10^k(b+1)$, i.e., $\log b + k \le n\log 2 < \log(b+1) + k$, which happens iff the point $\{n\log 2\}$ belongs to the interval of length $\log(b+1)-\log b$. Considering the rotation of the unit circle by the angle $\log 2$ we see that the answer is equal to $\log(b+1)-\log b$. It is not hard to extend this result for the case of a given number (say 2) of the first digits in the decimal expansion of 2^n, but not to the k-th digit with $k>1$ in this expansion. On the other hand, a completely different (uniform) distribution holds true for the last (few) digits ($2\to4\to8\to6\to2$).

Theorem 0.3 *(Kingman subadditive ergodic theorem) Let $\{\varphi_n\}$ be a sequence of measurable functions such that $\varphi_{n+k} \le \varphi_n + \varphi_k \circ T^n$ μ-a.e. Then there exists a T-invariant function $\tilde\varphi$ such that $\lim_{n\to\infty}\frac1n\varphi_n = \tilde\varphi$ μ-a.e. and $\lim_{n\to\infty}\frac1n\mu(\varphi_n) = \inf\frac1n\mu(\varphi_n) = \mu(\tilde\varphi)$.*

An immediate collorary to this result is the famous Furstenberg-Kesten theorem about the product of random matrices. Let $\Phi : X \to \mathbb{R}^{d^2}$ where we associate each element of $\mathbb{R}^{d^2}$ with a $d\times d$ matrix. Set $\Phi_n := \prod_{k=0}^{n-1}\Phi\circ T^k$.

Theorem 0.4 *(Furstenberg-Kesten) If $\log^+||\Phi|| \in \mathbf{L}^1(X,\mathcal{B},\mu)$, then for μ-a.a. $x \in X$ there exists a T-invariant function $\lambda(x) := \lim_{n\to\infty}\frac1n\log||\Phi_n|| \in \mathbf{L}^1$.*

Moreover, the following multiplicative ergodic theorem holds.

Theorem 0.5 *(Oseledets) If $\log^+||\Phi|| \in \mathbf{L}^1(X,\mathcal{B},\mu)$, then*
(a) the limit $\Lambda(x) := \lim_{n\to\infty}\sqrt[2n]{\Phi_n^*(x)\Phi_n(x)}$ is well defined μ-a.e.;
(b) $\exists e^{\lambda_1(x)} < e^{\lambda_2(x)} < \dots < e^{\lambda_k(x)}$ – the ordered collection of eigenvalues of the matrix $\Lambda(x)$, which together with their multiplicities are measurable and T-invariant functions.

Lecture 2. Mixing [09.12.19] For an ergodic DS $(T,X,\mathcal{B},\mu)$ and any $\varphi \in \mathbf{L}^2(X,\mathcal{B},\mu)$ by Birkhoff theorem we have $S_n\varphi \to \mu(\varphi)$ μ-a.e. Multiplying both hands of this relation by an arbitrary function $\psi \in \mathbf{L}^2(X,\mathcal{B},\mu)$ and integrating we get $\frac1n\sum_{k=0}^{n-1}\mu(\psi\cdot\varphi\circ T^k) \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)\mu(\psi)$.

Lemma 0.3 *The relation $\frac1n\sum_{k=0}^{n-1}\mu(\psi\cdot\varphi\circ T^k) \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)\mu(\psi)$ $\forall\varphi,\psi \in \mathbf{L}^2(X,\mathcal{B},\mu)$ is equivalent to ergodicity.*

Proof. The direct statement is already proven, so let us assume that this relation yields that for any pair of T,μ-invariant sets $A,A' \in \mathcal{B}$ we have: $\lim\limits_{n\to\infty}\frac1n\sum_{k=0}^{n-1}\mu(1_{A'}\cdot 1_A\circ T^k) = \mu(1_{A'}\cdot 1_A) = \mu(A\cap A') = \mu(A)\mu(A')$. Hence, for $A = A'$ we get $\mu(A) = (\mu(A))^2$, i.e. $\mu(A) \in \{0,1\}$. □

Observe that $\mu(1_{A'}\cdot 1_A\circ T^k) = \mu(T^{-k}A\cap A')$ $\forall A,A' \in \mathcal{B}$. Thus $\lim_{n\to\infty}\frac1n\sum_{k=0}^{n-1}\mu(T^{-k}A\cap A') = \mu(A)\mu(A')$ if the DS is ergodic. This relation is often called independence on the average.

A DS $(T,X,\mathcal{B},\mu)$ is called *weakly mixing* if $\frac1n\sum_{k=0}^{n-1}|\mu(T^{-k}A\cap B) - \mu(A)\mu(B)| \stackrel{n\to\infty}{\longrightarrow} 0$ $\forall A,B \in \mathcal{B}$ or $\frac1n\sum_{k=0}^{n-1}\left|\mu(\varphi\circ T^k\cdot\psi) - \mu(\varphi)\mu(\psi)\right| \stackrel{n\to\infty}{\longrightarrow} 0$ $\forall\varphi,\psi \in \mathbf{L}^2(X,\mathcal{B},\mu)$.

By Lemma 0.4 below this is equivalent to $\frac1n\sum_{k=0}^{n-1}\left(\mu(\varphi\circ T^k\cdot\psi) - \mu(\varphi)\mu(\psi)\right)^2 \stackrel{n\to\infty}{\longrightarrow} 0$ $\forall\varphi,\psi \in \mathbf{L}^2(X,\mathcal{B},\mu)$.

Lemma 0.4 *For any sequence of bounded complex numbers $\{a_n\}_{n=0}^{\infty}$ the conditions $\frac{1}{n}\sum_{k=0}^{n-1}|a_k| \stackrel{n\to\infty}{\longrightarrow} 0$ and $\frac{1}{n}\sum_{k=0}^{n-1}|a_k|^2 \stackrel{n\to\infty}{\longrightarrow} 0$ are equivalent.*

Lemma 0.5 *Weak mixing yields ergodicity.*
Proof. Due to weak mixing
$0 = \lim\limits_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \left|\mu(\varphi\circ T^k\cdot\psi) - \mu(\varphi)\mu(\psi)\right| \ge \lim\limits_{n\to\infty} \left|\frac{1}{n}\sum_{k=0}^{n-1} \mu(\varphi\circ T^k\cdot\psi) - \mu(\varphi)\mu(\psi)\right| \ge 0,$
(since $|a|+|b| \ge |a+b|$) which is equivalent to ergodicity by Lemma 0.3. □

A DS is called *mixing* if $\mu(\varphi\circ T^n\cdot\psi) \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)\mu(\psi) \ \ \forall\varphi,\psi\in\mathbf{L}^2(X,\mathcal{B},\mu)$. Setting $\varphi := 1_A, \psi := 1_B$, we get that in the case of mixing $\mu(T^{-n}A\cap B) \stackrel{n\to\infty}{\longrightarrow} \mu(A)\mu(B) \ \ \forall A,B\in\mathcal{B}$. In words this is equivalent to say that under the action of the DS any set of positive measure is uniformly spreading out the phase space, and in a such way any "nonequilibrium" distribution converges in time to the "equilibrium" one.

It is of interest that it is enough to check this property only in the case of equal test-functions $\varphi = \psi$:

Theorem 0.6 *(Renyi) Let $\mu(\varphi\circ T^n\cdot\varphi) \stackrel{n\to\infty}{\longrightarrow} \mu^2(\varphi) \ \ \forall\varphi\in\mathbf{L}^2(X,\mathcal{B},\mu)$. Then the DS is mixing.*
Proof. It is enough to consider test-functions with 0-average. Fix some φ and consider a linear subspace L_0 of $\mathbf{L}^2$ spanned by the functions $1,\varphi,\varphi\circ T^n, \ n\in\mathbb{Z}_+$ and its orthogonal complement L_1. Clearly both these subspaces are T-invariant. Denoting projection operators to these subspaces by π_i we represent any function $\psi\in\mathbf{L}^2$ as $\psi = \pi_0\psi + \pi_1\psi$. Since $\mu(\varphi\circ T^n\cdot\varphi\circ T^k) \stackrel{n\to\infty}{\longrightarrow} 0$ for any $k\in\mathbb{Z}_+$ by the invariance of μ, we have $\mu(\varphi\circ T^n\cdot\pi_0\psi) \stackrel{n\to\infty}{\longrightarrow} 0$. On the other hand, $\mu(\varphi\circ T^n\cdot\pi_1\psi) \equiv 0$ by the construction. □

In the Hilbert space $\mathbf{L}^2(X,\mathcal{B},\mu)$ of measurable complex valued functions on X with the scalar product $(\varphi,\psi) := \mu(\varphi\cdot\psi^*)$, where ψ^* stays for the complex conjugation of the function ψ, we consider the so called Koopman operator U_T defined by the formula $U_T\varphi(x) := \varphi(Tx)$. Due to this relation the Koopman operator is conjugated to the map T on $\mathbf{L}^2$. Clearly, $U_T 1_X = 1_X$, hence, any constant is an eigenfunction of this operator with the eigenvalue 1.

Lemma 0.6 *$U_T : \mathbf{L}^2(X,\mathcal{B},\mu) \to \mathbf{L}^2(X,\mathcal{B},\mu)$ and this operator is isometric.*
Proof. Due to the invariance of the measure μ, we have $\mu(|\varphi\circ T|^2) = \mu(|\varphi|^2) \ \ \forall\varphi\in\mathbf{L}^2(X,\mathcal{B},\mu)$, which proves the first assertion. Further, for any $\varphi,\psi\in\mathbf{L}^2(X,\mathcal{B},\mu)$ we get $(U_T\varphi,U_T\psi) = \mu(\varphi\circ T\cdot(\psi\circ T)^*) = \mu(\varphi\cdot\psi^*) = (\varphi,\psi)$, i.e. the operator U_T preserves the scalar product in $\mathbf{L}^2(X,\mathcal{B},\mu)$, which proves the second assertion. □

Two DS $(T_i,X_i,\mathcal{B}_i,\mu_i), \ i=1,2$ are called *spectrally isomorphic* if there exists a linear isomorphism P of Hilbert spaces $\mathbf{L}^2(X_i,\mathcal{B}_i,\mu_i), \ i=1,2$ preserving the inner product, i.e. $\mu_1(\varphi\cdot\psi^*) = \mu_2(P\varphi\cdot(P\psi)^*) \ \forall\varphi,\psi\in\mathbf{L}^2(\mu_1)$ and having the property that $PU_{T_1} = U_{T_2}P$. A property is called *spectrally invariant* if it is shared by any spectrally isomorphic DSs. Examples: ergodicity, weak mixing, mixing (see Theorem 0.7).

Theorem 0.7 *Let U_T be Koopman operator of the DS $(T,X,\mathcal{B},\mu)$. The DS is (a) ergodic iff the space of eigenfunctions corresponding to the eigenvalue 1 of the operator U_T is one dimensional; (b) weakly mixing iff each eigenfunction of U_T is a constant; (c) mixing iff $\lim_{n\to\infty}(U_T^n\varphi,\varphi) = (\mu(\varphi))^2 \ \ \forall\varphi\in\mathbf{L}^1$.*

Let us formulate also another important ergodic result, where in distinction to the Birkhoff theorem the convergence is in Hilbert norm rather than a.e. Let U be an isometric operator in a (complex) Hilbert space H, $H_U^{\rm inv} := \{\varphi\in H: \ U\varphi = \varphi\}$, and denote by P_U the orthogonal projection operator to $H_U^{\rm inv}$.
Theorem 0.8 *(von Neumann)* $\left|\left|\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi - P_U\varphi\right|\right|_H \stackrel{n\to\infty}{\longrightarrow} 0$.

Observe that if the DS (T,μ) is ergodic and $U := U_T$, then $\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi = S_n\varphi$ and $P_U\varphi = \mu(\varphi)$.
Lemma. The irrational rotation T_α of the unit circle is ergodic but not weakly mixing.
Indeed, let $A = B := [0,1/4]$. Then $\liminf\limits_{n\to\infty} m(T_\alpha^{-n}A\cap B) = 0$ while $\limsup\limits_{n\to\infty} m(T_\alpha^{-n}A\cap B) = 1/4$ which proves that the system is non mixing. To prove the absence of weak mixing, observe that the density of the set N_ε of positive integers for which $m(T_\alpha^{-n}A\cap B) \ge \varepsilon$ is equal to ε. Therefore
$\frac{1}{n}\sum_{k=0}^{n-1}|\mu(T^{-k}A\cap B) - \mu(A)\mu(B)| \ge \varepsilon\sum_{k\in N_\varepsilon}|\mu(T^{-k}A\cap B) - \mu(A)\mu(B)| \ge \varepsilon((1/4-\varepsilon)-1/16) > 0.$ □

It is rather difficult to demonstrate a weak mixing but not mixing dynamical system. Example of weak mixing – a "typical" interval exchange transformation for 3 intervals.
Lemma. The dyadic map T_2 is mixing.
First we give a formal proof. Set $z := e^{2\pi ix}$, then the dyadic map is equivalent to the map $Tz := z^2$ restricted to the unit circle on the complex plane $\{z: \ |z| = 1\}$. Consider Fourier expansions of two functions $\varphi,\psi\in\mathbf{L}_0^2(m)$, i.e. $\varphi(z) := \sum_{k\in\mathbb{Z}} a_k z^k, \ \psi(z) := \sum_{k\in\mathbb{Z}} b_k z^k$ with $a_0 = b_0 = 0$. Since $\varphi\circ T^n(z) =$

$\sum_{k\in\mathbb{Z}} a_k(z^{2^n})^k = \sum_{k\in\mathbb{Z}} a_k z^{2^n k}$, we have $m(\varphi\circ T^n\cdot\psi) = \sum_{k,k'\in\mathbb{Z}} m(a_k z^{2^n k} b_{k'} z^{k'}) = \sum_{k\neq 0} a_k b_{-2^n k} \overset{n\to\infty}{\longrightarrow} 0$. Here we have used that $m(z^k) = \int_0^1 e^{2\pi i x k}dx = 0\ \forall k\neq 0$, $m(1)=1$, and that $b_n \overset{n\to\pm\infty}{\longrightarrow} 0$ since $\psi\in\mathbf{L}^2$. $\square$

To have a more "geometrical" explanation, observe that any interval whose endpoints are binary fractions after a finite number of iterations covers the entire phase space. A similar statement about the covering for a general measurable set with positive Lebesgue measure follows since finite binary fractions are dense in $[0,1]$. This property of the dyadic map can be generalized as follows. A DS $(T,X,\mathcal{B},\mu)$ is called *exact* if its *tail* σ-algebra $\mathcal{B}_\infty := \cap_{n\geq 0} T^{-n}\mathcal{B}$ is trivial, i.e. $\mu(A)\in\{0,1\}\ \forall A\in\mathcal{B}_\infty$. The connection of this property to the covering one is described by the following result.

Lemma 0.7 *Let $TA\in\mathcal{B}\ \forall A\in\mathcal{B}$. Then $\mu(T^nA)\overset{n\to\infty}{\longrightarrow} 1\ \forall A\in\mathcal{B},\ \mu(A)>0$ iff the tail σ-algebra is trivial.*
Proof. We start with the direct statement. Assume the exactness. The collection of sets $B_n := T^{-n}\circ T^n A\in T^{-n}\mathcal{B}$ has the property: $B_n\subseteq B_{n+1}$, and $B_0 = A$. Thus $\cup_{n\geq 0}B_n\in\mathcal{B}_\infty$. Since $A\subseteq B_n\ \forall n$ we have $\mu(B_n)>0$, and if the tail σ-algebra is trivial, then $1 = \mu(\cup_{n\geq 0}B_n) = \lim_{n\to\infty}\mu(T^nA)$.

Conversely, let the covering property holds true. If $A'\in\mathcal{B}_\infty$ then $\forall n\ \exists A'_n\in\mathcal{B}:\ \ A' = T^{-n}A'_n$. Assuming that $\mu(A')>0$ we get $1 = \lim_{n\to\infty}\mu(T^nA')\leq\mu(A'_n)\leq 1\Longrightarrow\mu(A'_n)=1$. Thus $\mu(A') = \mu(A'_n) = 1$. $\square$

An exact map is necessarily ergodic since every invariant set is contained in $\mathcal{B}_\infty$. It is not so evident, but still correct, that any exact map is mixing. Observe that an invertible map cannot be exact: $\mu(T^nA) = \mu(T^{-n}\circ T^nA) = \mu(A)$ does not depend on n.

Symbolic dynamics

A finite set $\mathcal{A}$ with $\#\mathcal{A} = r<\infty$ we shall call the *alphabet*. Let $\vec{X} := \mathcal{A}^{\mathbb{Z}}$ be the space of two-sided sequences with elements from the alphabet $\mathcal{A}$, i.e. $\vec{X}\ni\vec{x} := \{x_i\}_{i=-\infty}^{\infty}$, $x_i\in\mathcal{A}$, and let $\vec{X}_+$ be the space of one-sided sequences $\{x_i\}_{i=1}^{\infty}$.

We equip the finite set $\mathcal{A}$ with the discrete topology (i.e., all its subsets are open) and consider the *product topology* on $\vec{X}$ generated by *cylinders*: $C_{i_1,\dots,i_n}^{a_1,\dots,a_n} := \{\vec{x}\in\vec{X}:\ x_{i_k} = a_k\ \forall\, 1\leq k\leq n,\ i_k\in\mathbb{Z},\ a_k\in\mathcal{A}\}$. In other words, the cylinder $C_{i_1,\dots,i_n}^{a_1,\dots,a_n}$ consists of all sequences whose $\{i_k\}$ "coordinates" are fixed (equal to the given letters $\{a_k\}$). Cylinders make a countable basis in the product topology and play the same important role as intervals in $\mathbb{R}$.

Exercise 10.2. Let $C = C_{i_1,\dots,i_n}^{a_1,\dots,a_n}$ and $C' = C_{i'_1,\dots,i'_{n'}}^{a'_1,\dots,a'_{n'}}$ be two cylinders. Describe $\vec{X}\setminus C,\ C\cap C',\ C\cup C',\ C\setminus C'$. (Finite unions of disjoint cylinders or empty sets.) Useful observation $C_{i_1,\dots,i_n}^{a_1,\dots,a_n} = \cap_{k=1}^n C_{i_k}^{j_k}$.

Corollary 10.3. Finite disjoint unions of cylinders make an algebra. This algebra generates the Borel σ-algebra on $\vec{X}$ and $\vec{X}_+$.

Remarks 10.4. (a) In the product topology, the convergence $\vec{x}^{(k)}\to\vec{x}$ is equivalent to the following: for any $n\geq 1$ there is a $k_n\geq 1$ such that $x_i^{(k)} = x_i\ \ \forall|i|\leq n$ for all $k\geq k_n$ i.e. the variable sequence $\vec{x}^{(k)}$ stabilizes as $k\to\infty$.
(b) The product topology is metrisable. The corresponding metric on $\vec{X}$ can be defined by $\mathrm{dist}(\vec{x},\vec{x}') = \sum_{n=-\infty}^{\infty}\frac{1-1_{x_n}(x'_n)}{r^{|n|}} = \sum_{n=-\infty}^{\infty}|x_n - x'_n|/r^{|n|}$. (in the space $\vec{X}_+$, we only need to sum over $n\geq 0$).
(c) In the product topology, both spaces $\vec{X}$ and $\vec{X}_+$ are compact and totally disconnected.

Definition 10.5. Let μ_0 be a probability measure on the finite set $\mathcal{A}$. It is characterized by r numbers $p_i = \mu_0(\{i\})$ such that $p_i\geq 0$ and $\sum_{i=1}^r p_i = 1$.

Then μ_0 induces the *product measure* μ on $\vec{X}$ (and on $\vec{X}_+$). For any cylinder $C_{i_1,\dots,i_n}^{j_1,\dots,j_n}$ its measure is given by $\mu(C_{i_1,\dots,i_n}^{j_1,\dots,j_n}) = \prod_{k=1}^n p_{j_k}$. The measure μ is also called the *Bernoulli measure* on $\vec{X}$ (resp., on $\vec{X}_+$).

Remarks 10.6. The measure μ is nonatomic (has no atoms), unless $p_i = 1$ for some i, in which case μ is concentrated on one sequence $\vec{x}$, for which $x_n = i\ \forall n$. The product measure μ makes the coordinates x_n, $n\in\mathbb{Z}$, independent random variables, in terms of probability theory. This explains the name *Bernoulli*.

Definition 10.7. The *left shift* map σ can be defined on the spaces $\vec{X}$ and $\vec{X}_+$. For every $\vec{x}\in\vec{X}$ we define $\vec{x}' = \sigma(\vec{x})$ by $x'_i = x_{i+1}$ for all $i\in\mathbb{Z}$. For every $\vec{x}\in\vec{X}_+$ we define $\vec{x}' = \sigma(\vec{x})$ by $x'_i = x_{i+1}$ for all $i\geq 0$.

Exercise 10.8. Show that the left shift map σ is "onto" and continuous. On the space $\vec{X}$ it is a homeomorphism, and on $\vec{X}_+$ it is an r-to-one map.

Exercise 10.9. Show that σ preserves the Bernoulli measure μ defined by 10.5. Hint: take a cylinder, describe its image, and then use the algebra made by finite disjoint unions of cylinders.

Definition 10.10. The symbolic space $\vec{X}$ (or $\vec{X}_+$) with a Bernoulli measure μ defined by 10.5 and the left shift map σ is called a *Bernoulli system* (or a *Bernoulli shift*). We denote it by $B_r(p_1,\dots,p_r)$ (resp., $B_{+,r}(p_1,\dots,p_r)$). Note: its only parameters are r and $p_1,\dots,p_r$.

PROPOSITION 10.11. The doubling map $T(x) = 2x \pmod 1$ with the Lebesgue measure is isomorphic to the (one-sided) Bernoulli shift $B_{+,2}(1/2, 1/2)$. Denote the corresponding Bernoulli measure by μ.

Proof. Let $B' \subset [0,1)$ be the set of binary rational numbers, i.e. $B' = \{k/2^n : \ n \geq 0, \ 0 \leq k < 2^n\}$. Clearly, B' is a countable set, so $m(B') = 0$. Let $r = 2$ and $\vec{X}'_+ \subset \vec{X}_+$ be the set of eventually constant sequences, i.e. $\vec{X}'_+ = \{\vec{x} : \ x_i = x_{i+1} \ \forall i > i_0\}$. Clearly, $\vec{X}'_+$ is countable set, so $\mu(\vec{X}'_+) = 0$.

Now, for each $x \in [0,1) \setminus B'$ the binary representation $x = 0.i_0 i_1 i_2 \ldots$. We define a sequence $\vec{x} = \varphi(x) \in \vec{X}_+$ by $x_n = i_n + 1$ for all $n \geq 0$. This defines a bijection between $[0,1) \setminus B'$ and $\vec{X}_+ \setminus \vec{X}'_+$. One can check by direct inspection that φ preserves the measure and dynamics, i.e. it is an isomorphism. $\square$

PROPOSITION 10.12. The baker's map with the Lebesgue measure m is isomorphic to the (two-sided) Bernoulli shift $B_2(1/2, 1/2)$.

Proof. Similar to the previous one. Define a map $G : \{0,1\}^{\mathbb{Z}} \to \mathrm{Tor}^2$ as follows: $G\vec{x} := (u := \sum\limits_{k=0}^{\infty} x_k/2^{k+1}, \ v := \sum\limits_{k=1}^{\infty} x_{-k}/2^k)$. Observe that this map is bijective except the Lebesgue measure zero set of points $(u,v) \in \mathrm{Tor}^2$ such that either u or v is a finite dyadic fraction. Our aim is to show that $G \circ T_B = T_{\text{baker}} \circ G$ μ_B-a.e. For (u,v) defined as above, we have $(G^{-1}(u,v))_i = x_i, \ (T_B \circ G^{-1}(u,v))_i = x_{i+1} \ \forall i$. Thus $G \circ T_B \circ G^{-1}(u,v) = (\sum\limits_{k=1}^{\infty} x_k/2^{k+1}, \ \sum\limits_{k=0}^{\infty} x_{-k}/2^{k+1}) = \begin{cases} (2u, v/2) & \text{if } 0 \leq u < 1/2 \\ (\{2u\}, (v+1)/2) & \text{if } 1/2 \leq u < 1 \end{cases} = T_{\text{baker}}(u,v)$, since $x_0 = 0 \iff u \in [0, 1/2)$ and $x_0 = 1 \iff u \in [1/2, 1)$.

It remains to check that G maps μ_B to the Lebesgue measure. As above we shall do this only for the simplest generating cylinders. Observe that $GC_i^j = \{(\sum_{k=0}^{\infty} x_k/2^{k+1}, \ \sum_{k=1}^{\infty} x_{-k}/2^k) : \ x_i = j\}$ is a set consisting of $2^{|i|}$ rectangles with sides $1 \times 2^{-|i|-1}$, i.e. $m(GC_i^j) = 1/2 = \mu_B(C_i^j)$. $\square$

EXERCISE 10.13. Let C and C' be two cylinders and μ a Bernoulli measure. Show that there is an $n_0 \geq 0$ such that $\mu(C \cap \sigma^{-n}(C')) = \mu(C)\,\mu(C')$ for all $n \geq n_0$. Note: this applies to both $\vec{X}$ and $\vec{X}_+$.

THEOREM 10.14. Every Bernoulli shift is mixing and hence ergodic.

Proof. Let A and B be two Borel subsets of $\vec{X}$ (or $\vec{X}_+$). By the approximation theorem 1.19, for any $\varepsilon > 0$ there are sets A_ε and B_ε, each being a finite disjoint union of some cylinders, such that $m(A \Delta A_\varepsilon) < \varepsilon$ and $m(B \Delta B_\varepsilon) < \varepsilon$. The result of Exercise 10.13 implies that there is an $n_0 \geq 0$ such that $\mu(A_\varepsilon \cap \sigma^{-n}(B_\varepsilon)) = \mu(A_\varepsilon)\,\mu(B_\varepsilon)$ for all $n \geq n_0$. Since A_ε approximates A and B_ε approximates B, it is easy to derive that $|\mu(A \cap \sigma^{-n}(B)) - \mu(A)\,\mu(B)| < 4\varepsilon$ for all $n \geq n_0$. Hence, $\mu(A \cap \sigma^{-n}(B)) \stackrel{n\to\infty}{\longrightarrow} \mu(A)\,\mu(B)$. $\square$

CLAIM. Any two different Bernoulli measures μ, ν are mutually singular.

Let us give an ergodic proof of this result. We already know that these two measures are ergodic measures for the shift map. This implies the desired result. However, it is worth to discuss the reason for this observation. Using the notation Z_μ for the set of μ-typical points and that $\mu(Z_\mu) = 1$ we are getting that $Z_\mu \cap Z_\nu = \emptyset$ and hence $\mu(Z_\nu) = 0$. To simplify the discussion assume that $r = 2$, i.e. we deal only with binary sequences. For each μ-typical point the density of indices corresponding to a given letter (say 1) is equal to the "probability" of this letter. Therefore the support of each of these measures is completely defined by the density of ones on its support.

The definition of the Bernoulli measure may be rewritten as follows $\mu(C_{1,\ldots,n,n+1}^{j_1,\ldots,j_n,j_{n+1}}) = \mu(C_{1,\ldots,n}^{j_1,\ldots,j_n}) p_{j_{n+1}}$. Our aim now is to extend this definition to the case of Markov measures. Here we consider a subset X_M of the space of sequences defined be the transition matrix M whose elements $m_{ij} \in \{0,1\}$. Then $x \in \mathbb{X}_M$ iff $x_k = i, x_{k+1} = j$ only if $m_{ij} = 1$. We say that a stochastic matrix P is compatible with M if $p_{ij}/m_{ij} > 0 \ \ \forall i,j$ (as usual we assume that 0/0=1). Denote by p the left eigenvector (with the largest eigenvalue) of the matrix P and introduce inductively the functional $\mu(C_{1,\ldots,n,n+1}^{j_1,\ldots,j_n,j_{n+1}}) = \mu(C_{1,\ldots,n}^{j_1,\ldots,j_n}) p_{j_n j_{n+1}}$. Finally setting $\mu(C_1^j) := p_j$ we obtain a probabilistic measure, called *Markov measure*. It is straightforward to check that this is indeed a measure and it is invariant with respect to the shift-map. Moreover, it is positive on each finite cylinder in X_M.

CLAIM. If the matrix P is transitive, then Markov measures are mixing.

The prove of this result is not so obvious as in the case of Bernoulli measures. The point is that again one can use the fact for a given finite cylinder C its base and the base of the cylinder $\sigma^{-n}C$ are disjoint. The problem is that the calculation of $\mu(\sigma^{-n}A \cap B)$ is not straightforward even for arbitrary large n. Here one needs to sum up contributions from all admissible sequences connected these two bases. Denote by α the first letter in the base of A, by β the last letter in the base of B, and by $N > 0$ the number of letters between them. Then $\mu(\sigma^{-n}A \cap B) = \mu(B) p_{\beta,\alpha}^N \frac{1}{p_\alpha} \mu(A)$, where $p_{\beta,\alpha}^N$ is the β,α-element of the matrix

P^N. Now transitivity of the Markov matrix P implies that for any i, j we have $p^N_{i,j} \overset{N\to\infty}{\longrightarrow} p(j)$. Therefore $p^N_{\beta,\alpha}\frac{1}{p_\alpha} \overset{N\to\infty}{\longrightarrow} 1$, which proves the claim. □

Whenever we establish an isomorphism between a given dynamical system (X, T, μ) and a symbolic system $(\vec{X}, \sigma, \nu)$ or $(\vec{X}_+, \sigma, \nu)$ with some σ-invariant measure ν, we call this a *symbolic representation* of (X, T, μ). We now outline a standard method of constructing symbolic representations.

Definition 10.15. Let $X = X_1 \cup \cdots \cup X_r$ be a finite partition of X into disjoint parts, $X_i \cap X_j = \emptyset$ for $i \neq j$. Let $T : X \to X$ be a map. For every point $x \in X$ its *itinerary* is a sequence defined by $\{x_n\}_{n=0}^\infty : \quad T^n(x) \in X_{x_n} \quad \forall n \geq 0$. If the map T is a bijection, i.e. $T^{-1} : X \to X$ is also defined, then the *full itinerary*[3] of a point $x \in X$ is a double infinite sequence defined by $\{x_n\}_{n=-\infty}^\infty : \quad T^n(x) \in X_{x_n} \quad \forall n \in \mathbb{Z}$.

Definition 10.16. A partition $X = X_1 \cup \cdots \cup X_r$ is called a *generating partition* if distinct points have distinct itineraries. Equivalently, $\forall x \neq y \quad \exists n$ such that $T^n(x) \in X_i$ and $T^n(y) \in X_j$ with some $i \neq j$.

10.17 Construction of a symbolic representation. Let $T : X \to X$ be a map and $X = X_1 \cup \cdots \cup X_r$ a generating partition. Let $\varphi : X \to \vec{X}_+$ (or $\varphi : X \to \vec{X}$, if T is an automorphism) be the map that takes every point $x \in X$ to its itinerary. This map is injective for any generating partition.

Let $\vec{X} := \varphi(X)$ be the image of X. Then $\vec{X}$ is σ-invariant, i.e. $\sigma(\vec{X}) \subset \vec{X}$. Moreover, $\varphi \circ T = \sigma \circ \varphi$. If T has an invariant measure μ on X, one can define a measure ν on $\vec{X}$ by $\nu(B) = \mu(\varphi^{-1}(B))$. Then the dynamical systems (X, T, μ) and $(\vec{X}, \sigma, \nu)$ will be isomorphic.

This is a general principle for the construction of a symbolic representation.

Remark 10.18. In the above symbolic representation of $T : X \to X$, any cylinder $C = C^{a_1,\ldots,a_n}_{m,\ldots,m+n-1} \subset \vec{X}$ corresponds to the set $\cap_{k=m}^n T^{-k} X_{\omega_k}$ that is, $\varphi^{-1}(C) = \cap_{k=m}^{m+n-1} T^{-k} X_{\omega_k}$ with $\omega \in C$.

Remarks 10.19. The symbolic representation of the doubling map corresponds to the partition $X_1 = [0, 0.5)$ and $X_2 = [0.5, 1)$. The symbolic representation of the baker's map corresponds to the partition of the square X by the line $x = 0.5$.

More generally one can define 'symbolic' maps using the so called *transition matrix* P with $p_{i,j} \in \{0, 1\}$ and allow the symbol j to follow the symbol i from the alphabet $\mathcal{A}$ iff $p_{ij} = 1$. Such DS are called subshifts of finite type or topological Markov chains. Discuss connections to one-dimensional maps. **[⋆]**

A DS is called *mixing of multiplicity* $r \in \mathbb{Z}_+$ if $\mu(\varphi_0 \cdot \varphi_1 \circ T^{k_1} \cdot \ldots \cdot \varphi_r \circ T^{k_1+k_2+\ldots+k_r}) \overset{k_1,\ldots,k_r\to\infty}{\longrightarrow} \prod_{i=0}^r \mu(\varphi_i) \quad \forall \varphi_i \in \mathcal{B}$. In terms of measurable sets this property reads: $\mu(A_0 \cap T^{-k_1}A_1 \cap \ldots \cap T^{-k_1-\ldots-k_r}A_r) \overset{k_1,\ldots,k_r\to\infty}{\longrightarrow} \prod_{i=0}^r \mu(A_i)$. Clearly the usual mixing is equivalent to the mixing of multiplicity 1, while for any r the mixing of multiplicity r yields the mixing of multiplicity r for any $r' < r$.

Proposition 11.17. If T is mixing of multiplicity $r \geq 2$, then it is mixing of multiplicity $(r - 1)$.

Proof. Just set $A_r = X$ in the above definition. □

Remark 11.18. The mixing (of multiplicity r) properties are invariant under isomorphisms.

Proposition 11.19. Every Bernoulli shift is mixing of multiplicity r for all $r \geq 2$.

Proof. As in 10.14, we can approximate arbitrary sets $A_1, \ldots, A_r$ with finite unions of cylinders. Thus, it is enough to prove the mixing of multiplicity r for cylinders only. We omit details. □

A DS is called *K-mixing* if $\sup_{B\in\mathcal{B}_n^\infty(A_1,\ldots,A_r)} |\mu(A_0 \cap B) - \mu(A_0)\mu(B)| \overset{n\to\infty}{\longrightarrow} 0 \quad \forall A_i \in \mathcal{B}, r \geq 0$, where $\mathcal{B}_n^\infty(A_1, \ldots, A_r)$ is the smallest σ-algebra generated by the sets $T^{-k}A_i$ for $k \geq n, \ i = 1, 2, \ldots, r$. The letter "K" here stands for Kolmogorov. K-mixing means that the set A_0 does not depend on any event defined by a sufficiently faraway part of a 'trajectory' of pre-images of the sets A_i for $i > 0$.

Theorem 0.9 *For an automorphism the K-mixing yields mixing of multiplicity r $\forall r \in \mathbb{Z}_+$, but there exists a not K-mixing automorphism with the mixing of multiplicity r $\forall r \in \mathbb{Z}_+$. (The case $r = 1$ will be proven in Lemma 0.14)*

Let us mention two important properties related to mixing. The first of them measures the 'rate' of mixing and is called the *correlation coefficient*: $\mathrm{Cor}^n_T(\varphi, \psi) := |\mu(\varphi \circ T^n \cdot \psi) - \mu(\varphi)\mu(\psi)|$, which depends on the choice of *observables* $\varphi, \psi \in \mathbf{L}^2$. In a number of interesting examples (which we shall discuss later) the dependence on n here is exponential.

Another property describes the distribution of the Cesaro means around the limit value. We say that the DS $(T, X, \mathcal{B}, \mu)$ satisfies the Central Limit Theorem (CLT) if $\forall \varphi \in \mathbf{L}^2(X, \mathcal{B}, \mu) \quad \exists \sigma = \sigma(\varphi) > 0$

[3] itinerary = маршрут

such that $\mu\{x \in X: \sqrt{n}(S_n\varphi(x) - \mu(\varphi)) \le t\} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/(2\sigma^2)}dx \quad \forall t \in \mathbb{R}^1$, i.e. to the normal distribution.[4]

Let us define also a few purely topological notions related to the mixing. A map T is called *Lyapunov stable* if $\forall \varepsilon > 0 \ \exists \delta > 0$ such that if $\varrho(x,y) < \delta$ then $\varrho(T^n x, T^n y) < \varepsilon \quad \forall n \in \mathbb{Z}_+$. Here ϱ is a metric in X.

A map T is called *sensitive to initial conditions* if $\exists \delta > 0$ such that $\forall x \in X$ in any its open neighborhood $U \ni x$ there exists $y \in U$ and $n \in \mathbb{Z}_+$ for which $\varrho(T^n x, T^n y) > \delta$.

A map T is called *topologically transitive* if $\exists x \in X$ such that $\mathrm{Clos}(\cup_{n\ge0} T^n x) = X$, *topologically ergodic* if $\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} 1_B(T^k A) > 0$ for any two open sets A, B, and *topologically mixing* if for any two open sets A, B there exists $N \in \mathbb{Z}_+$ such that $T^{-n}A \cap B \neq \emptyset \ \forall n > N$. One often says that a DS is *chaotic* if it is sensitive to initial conditions and topologically transitive.

Examples: T_α with irrational α is topologically transitive (but neither sensitive, nor mixing); the dyadic map T_2 is topologically mixing.

Lecture 3. Ergodic constructions [16.12.19] Direct and skew products, induced and integral maps, and the natural extension.

For a pair of DS $\{(T_i, X_i, \mathcal{B}_i, \mu_i)\}_{i=1,2}$ by their *direct product* we mean a new DS $(T, X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_1, \mu_1 \times \mu_2)$ defined by the relation $T(x_1, x_2) := (T_1 x_1, T_2 x_2)$. Example: a 2-dimensional torus rotation.

Theorem 0.10 *The direct product map is ergodic iff one is the only joint eigenvalue of the Koopman operators corresponding to the original maps. (Counterexample: $T_\alpha \times T_\alpha$, $\lambda = e^{i\alpha}$, $\varphi(x) = e^{ix}$.)*

Proof. Let $U_{T_i}\varphi_i = \lambda\varphi_i$, $i = 1, 2$ with $\lambda \neq 1$. Observe that $U_{T_2}\varphi_2^*(x_2) = \varphi_2^*(T_2 x_2) = \lambda^*\varphi_2^*(x_2) = \lambda^{-1}\varphi_2^*(x_2)$. Consider $\varphi(x_1, x_2) := \varphi_1(x_1)\varphi_2^*(x_2)$. Then this function is orthogonal to the subspace of constants, and $U\varphi = U_{T_1}\varphi_1 \cdot U_{T_2}\varphi_2^* = \lambda\lambda^{-1}\varphi_1 \cdot \varphi_2^* = \varphi_1 \cdot \varphi_2^* = \varphi$, which contradicts to the ergodicity of the direct product.

Conversely, let $(\lambda_i, \varphi_i), i = 1, 2$ be two pairs of eigenvalue/eigenfunction of the Koopman operators corresponding to the original maps. Then $U(\varphi_1\varphi_2) = \lambda_1\lambda_2\varphi_1\varphi_2$. On the other hand, due to general properties of isometric operators (see details, e.g. in [KSF], p.187-188) all eigenpairs of the "product" Koopman operator can be constructed in this way. Thus U has the unit eigenvalue iff $\exists \lambda_1, \lambda_2: \ \lambda_1\lambda_2 = 1$, i.e., $\lambda_2^* = \lambda_2^{-1} = \lambda_1$ (λ_2^* is an eigenvalue of U_{T_2} since it is isometric), which contradicts the assumption that the only joint eigenvalue is 1. □

Corollary. The direct product of a weakly mixing DS and an ergodic one is ergodic, and the direct product of two (weakly) mixing DS is (weakly) mixing. Using spectral properties of isometric operators one can derive these assertions from the above proof (see details, e.g. in [KSF], p.188-189).

Note that the direct product system might have invariant measures not represented by direct product of invariant measures. Example: direct square map and the restriction to the diagonal.

A DS $(T', X', \mathcal{B}', \mu')$ is called a *factor-system* of a DS $(T, X, \mathcal{B}, \mu)$ if there exists a measurable map $\varphi: X \to X'$ having the property that $\varphi \circ T = T' \circ \varphi$. Observe that $\varphi^*\mu \in \mathcal{M}_{T'}$ whence $\mu \in \mathcal{M}_T$. Each DS has at least two trivial factor-systems: the system itself (φ is the identity) and a 'trivial' DS: X' is a point and φ maps X into this point. Each component of a direct product is its factor-system ($\varphi: X_1 \times X_2 \to X_i$).

Theorem 0.11 *The transition to a factor-system preserves ergodicity, weak mixing and mixing.*

Proof. If $A \in \mathcal{B}'$ is T', μ'-invariant then $\varphi^{-1}A$ is T, μ-invariant. Now, from the ergodicity we have $\mu'(A) = \mu(\varphi^{-1}A) \in \{0, 1\}$, which proves the preservation of ergodicity. Two other assertions can be proven similarly. □

Let $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$ be a direct product of two measurable spaces. For a given DS $(T, X, \mathcal{B}, \mu)$ we consider a family of maps T_x, $x \in X$ measurably depending on x, i.e. for any measurable function $\varphi(x, y)$ on $X \times Y$ the function $\varphi(T^n x, T_x^n y)$ is measurable on $X \times Y \times \mathbb{Z}_+$. The map $\tilde{T}(x, y) := (Tx, T_x y)$ is called a *skew product* with the *base* X. Let $\mu_Y \in \mathcal{M}_{T_x} \ \forall x \in X$, then $\tilde{\mu} := \mu_X \times \mu_Y \in \mathcal{M}_{\tilde{T}}$. Indeed, $U_{\tilde{T}} 1_{A_X \times A_Y}(x, y) = 1_{T^{-1}A_X}(x) \cdot 1_{T_x^{-1}A_Y}(y)$ and $\tilde{\mu}(U_{\tilde{T}} 1_{A_X \times A_Y}) = \mu_X(1_{T^{-1}A_X}) \cdot \mu_Y(1_{T_x^{-1}A_Y}) = \mu_X(A_X)\mu_Y(A_Y) = \tilde{\mu}(A_X \times A_Y)$.

Let $(T, X, \mathcal{B}, \mu)$ be a DS with the invariant measure μ. Consider a set $E \in \mathcal{B}$ with $\mu(E) > 0$. Setting $\mathcal{B}_E := \{A \subseteq E: \ A \in \mathcal{B}\}$ and $\mu_E := \mu(A)/\mu(E)$, we get a measurable space $(E, \mathcal{B}_E, \mu_E)$. Introduce the *return time function* $t_E: E \to \mathbb{Z}_+$ by the relation $t_E(x) := \min\{n \ge 1: \ x \in E, \ T^n x \in E\}$, and the *first return* or *Poincare* or *induced* map wrt the set E as $T_E x := T^{t_E(x)}x$ for $x \in E$. By Poincare recurrence result the function t_E is finite μ_E-a.e. In what follows we always assume that t_E is finite everywhere.

---

[4] It is of interest that the constant σ called the *asymptotic variance* may vanish, and it can be shown (M. Ratner, "The CPT for geodesic flows on n-dimensional manifolds of negative curvature", Isr. J. Math., 16(1973), 181-197) that $\sigma(\varphi) = 0$ iff the homological equation $\varphi - \mu(\varphi) = \Phi \circ \sigma(\varphi) - \Phi$ has a solution $\Phi \in \mathbf{L}^2(\mu)$.

Lemma 0.8 *$\mu(1_E \cdot t_E) = \mu(\cup_{n\ge0}T^nE)$ for each automorphism.*

Proof. Let $E_n := \{x \in E : \ t_E(x) = n\}$. Observe that $E_n \in \mathcal{B}$, indeed, $E_n \equiv (T^{-n}E \cap E) \setminus (\cup_{k<n}E_k)$. The sets T^kE_n are mutually disjoint for all $0 \le k < n < \infty$ and $\cup_{n\ge0}T^nE = \cup_{n\ge1}\cup_{k=1}^{n-1} T^kE_n$. Hence, $\mu(\cup_{n\ge0}T^nE) = \mu(\cup_{n\ge1}\cup_{k=1}^{n-1} T^kE_n) = \sum_{n=1}^{\infty}\sum_{k=0}^{n-1}\mu(T^kE_n) = \sum_{n=1}^{\infty} n\mu(E_n) = \mu(1_E \cdot t_E)$. □

Corollary. $t_E \in \mathbf{L}^1(X, \mathcal{B}_E, \mu_E)$ for each automorphism.

Observe that the proof above does not work for endomorphisms, namely, in general, $\mu(T^kE_n) \ne \mu(E_n)$.

Lemma 0.9 *$\mu_E \in \mathcal{M}_{T_E}$ for each automorphism.*

Proof. By the characterization of invariant measures (Lemma **??**) it is enough to check that $\mu_E(\varphi\circ T_E) = \mu_E(\varphi)\ \forall\varphi \in \mathbf{L}^1(E, \mathcal{B}_E, \mu_E)$. For $\varphi := 1_{E_n}$ we have $\mu_E(1_{E_n}\circ T_E) = \frac{1}{\mu(E)}\mu(1_{E_n}\circ T^n) = \frac{1}{\mu(E)}\mu(1_{E_n}) = \mu_E(1_{E_n})$ since the measure $\mu \in \mathcal{M}_T$. On the other hand, each $\varphi \in \mathbf{L}^1(E, \mathcal{B}_E, \mu_E)$ can be approximated by a linear combination of indicator functions, which yields the result. □

Now let T be a general endomorphism.

Lemma 0.10 *Let $(T, X, \mathcal{B}, \mu)$ be a measurable DS and let $\mu(E) > 0$. Then*
(a) $\mu(T_E^{-1}B) = \mu(B)\ \ \forall B \in \mathcal{B}\cap E$;
(b) if $\mu \in \mathcal{M}_T^0$ then $\mu_E \in \mathcal{M}_{T_E}^0$.

Proof. Consider collections of sets:
$E_n := \{x \in E : \ t_E(x) = n\}$ and $G_n := \{x \in E^c : Tx, \dots, T^{n-1}x \notin E, T^nx \in E\}$.
This collection may be viewed as a *tower* construction:

E_1	E_2	E_3	E_4	$\dots$	$(E,1) := E$
.	G_1	G_2	G_3	$\dots$	$(E,2) := TE\setminus E$
.	.	G_1	G_2	$\dots$	$(E,3) := T^2E\setminus(E\cup TE)$
.	.	.	.	$\dots$	$\dots$

Notice that $E = \cup_{n\ge1}E_n$ and $T^{-1}E = E_1 \cup G_1$, $T^{-1}G_n = E_{n+1}\cup G_{n+1}$, $E_n = \cap_{k=1}^{n-1}T^{-n}E^c \cap E \cap T^{-n}E$.

We have $T_E^{-1}B = \cup_{n\ge1}(E_n \cap T_E^{-1}B) = \cup_{n\ge1}(E_n\cap T^{-n}B)$. Hence $\mu(T_E^{-1}B) = \sum_{n\ge1}\mu(E_n\cap T^{-n}B)$.

On the other hand, applying repeatedly the main property of the sets E_k, G_k one gets $\forall n \ge 1$ that
$\mu(T^{-1}B) = \mu(E_1\cap T^{-1}B) + \mu(G_1\cap T^{-1}B) = \mu(E_1\cap T^{-1}B) + \mu(T^{-1}(G_1\cap T^{-1}B))$
$= \mu(E_1\cap T^{-1}B) + \mu(E_2\cap T^{-2}B) + \mu(G_2\cap T^{-2}B) = \ldots = \sum_{k=1}^{n}\mu(E_k\cap T^{-k}B) + \mu(G_n\cap T^{-n}B)$.

From $1 \ge \mu(\cup_{k\ge1}(G_k\cap T^{-k}B)) = \sum_{k\ge1}\mu(G_k\cap T^{-k}B)$ it follows that $\mu(G_k\cap T^{-k}B) \stackrel{k\to\infty}{\longrightarrow} 0$.

Thus $\mu(B) = \mu(T^{-1}B) = \sum_{k\ge1}\mu(E_k\cap T^{-k}B) = \mu(T_E^{-1}B)$, which proves (a).

To prove (b) consider a (μ, T_E)-invariant set $A \subseteq E$ (i.e. $\mu(A\Delta T_E^{-1}A) = 0$). $\forall n \ge 1$ the set $A\cap E_n$ can be pushed forward along its 'trajectory' to obtain the set $A_\infty := \cup_{n\ge1}\cup_{k=0}^{n-1}T^k(A\cap E_n)$. One shows that $\mu(A_\infty\Delta T^{-1}A_\infty) = 0$. Thus by the ergodicity of (T,μ) we get $\mu(A_\infty) \in \{0,1\}$. The equality $\mu(A_\infty) = 0$ implies $\mu(A) = 0$, so it remains to consider the case $\mu(A_\infty) = 1$. In this case $\mu(E\setminus A_\infty) = 0$ and thus $\mu(A) = \mu(E)$ since $\mu((A_\infty\cap E)\Delta A) = 0$. This finishes the proof, since it implies that each (μ, T_E)-invariant set has measure 0 or 1. □

Denote by $t_E^{(n)}(x) := \sum_{k=0}^{n-1}(t_E\circ T_E^k)(x)$ for $n \ge 0$ the n-th return time of the point $x \in E$ to E.

Lemma 0.11 *(Kac's recurrence Lemma) Let $(T, X, \mathcal{B}, \mu)$ be ergodic and let $\mu(E) > 0$. Then*
(a) $\mu_E(t_E) = 1/\mu(E)$; (b) $n^{-1}t_E^{(n)}(x) \stackrel{n\to\infty}{\longrightarrow} 1/\mu(E)\ \ \mu_E$-a.e.; (c) $\mu(T_E^{-1}B) = \mu(B)\ \ \forall B \in \mathcal{B}\cap E$.

Proof. $\forall\varphi \in \mathbf{L}^1(E,\mu)$ due to the ergodicity by Birkhoff's Ergodic Theorem $S_n\varphi \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)\ \ \mu_E$-a.e. For $x \in E, k \ge 1$ we have $T^kx \in E$ iff $k = t_E^{(n)}(x)$ for some $n \ge 0$, i.e. $T^kx = T_E^nx$. Therefore
$(1_E\cdot\varphi)(T^kx) = \begin{cases}\varphi(T_E^n(x)) & \text{if } k = t_E^{(n)}(x)\\ 0 & \text{otherwise}\end{cases}$. Setting $\varphi := 1_E\cdot t_E$ we get $S_n(1_E\cdot t_E) \stackrel{n\to\infty}{\longrightarrow} \mu(1_E\cdot t_E)\ \ \mu$-a.e.

The l.h.s. of this relation can be rewritten passing from t_E to $t_E^{(k)}$ as follows:
$\lim_{n\to\infty} S_n(1_E\cdot t_E)(x) = \lim_{k\to\infty}\frac{1}{t_E^{(k)}(x)}\sum_{j=0}^{k-1}t_E(T_E^jx) = 1$ since $t_E^{(k)} = \sum_{j=0}^{k-1}t_E\circ T_E^j$. Thus $\mu(t_E) = 1$, which using that $\mu_E(\cdot) = \mu(\cdot)/\mu(E)$ proves the first assertion.

The second assertion can be derived similarly using $\varphi := 1_E$:
$\mu(E) = \lim_{n\to\infty} S_n1_E(x) = \lim_{k\to\infty}\frac{1}{t_E^{(k)}(x)}\sum_{j=0}^{k-1}1_E(T_E^jx) = \lim_{k\to\infty}\frac{k}{t_E^{(k)}(x)}$.

To prove (c) consider $\varphi := 1_B\circ T_E$, $\ B \subseteq E$. Again by Birkhoff's Ergodic Theorem for μ-a.a. $x \in E$
$\mu(T_E^{-1}B\cap E) = \lim_{n\to\infty} S_n(1_B\circ T_E)(x) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}((1_B\circ T_E)\cdot 1_E)(T^kx) = \lim_{k\to\infty}\frac{1}{t_E^{(k)}(x)}\sum_{j=0}^{k-1}1_B(T_E^{j+1}x)$
$= \lim_{k\to\infty}\frac{t_E^{(k+1)}(x)}{t_E^{(k)}(x)}\cdot\frac{1}{t_E^{(k+1)}(x)}\sum_{j=0}^{k-1}1_B(T_E^{j+1}x) = \lim_{k\to\infty}\frac{1}{t_E^{(k+1)}(x)}\sum_{j=0}^{k-1}1_B(T_E^{j+1}x) = \lim_{k\to\infty}\frac{k}{t_E^{(k+1)}(x)}\cdot\mu(B) = \mu(B)$.

Here we used that $\frac{t_E^{(k+1)}(x)}{t_E^{(k)}(x)} \stackrel{k\to\infty}{\longrightarrow} 1$ by (b), and the last 2 steps follow from the argument similar to the case (b) but applied to $\varphi = 1_B$. Note that $T_{E'} = (T_E)_{E'}$ for $E' \subset \mathcal{B}_E$. □

One of the main applications of those constructions is the reduction of the analysis of ergodic properties of a given map T to those of a suitably chosen Poincare or integral map. This approach is especially fruitful in the former case. It turns out that by any invariant measure of a Poincare map we can explicitly construct the invariant measure of the genuine map. Moreover, we can do it even in a much more general situation. Setting $t_E(x) = 0$ for all $x \notin E$ we extend the map $T_E x := T^{t_E(x)}x$ to the map from the entire X into itself, for which any measure invariant with respect to the original Poincare map remains invariant. This simple extension shows that it is reasonable to make yet another step and to introduce a generalized Poincare map $T_t x := T^{t(x)}x$ by means of an arbitrary measurable function $t : X \to \mathbb{Z}_+ \cup \{0\}$. Denote $E_k := \{x \in X : \ t(x) = k\}$, $E_{>k} := \cup_{i>k} E_i = \{x \in X : \ t(x) > k\}$, $k \in \mathbb{Z}_+ \cup \{0\}$. The following result gives an explicit formula for a T-invariant measure in terms of a given T_t-invariant measure μ_t.

Theorem 0.12 *Let $\mu_t \in \mathcal{M}_{T_t}$. The measure μ_T defined by the relation:*

$$\mu_T(A) := \sum_{k=0}^{\infty} \mu_t(T^{-k}A \cap E_{>k}) \ \ \forall A \in \mathcal{B} \tag{0.1}$$

is T-invariant. Moreover, $\mu_T(X) < \infty$ if $\sum_{k=1}^{\infty} k\mu_t(E_k) < \infty$, and, hence, $\frac{\mu_T}{\mu_T(X)} \in \mathcal{M}_T$.

Proof. Observe that $T_t^{-1}A = \cup_{k=1}^{\infty}(T^{-k}A \cap E_k) \ \ \forall A \in \mathcal{B}$ which proves that $T_t^*\mu(A) = \sum_{k=1}^{\infty} \mu(T^{-k}A \cap E_k)$ for any $\mu \in \mathcal{M}$. Now let $\mu_t \in \mathcal{M}_{T_t}$, i.e. $T_t^*\mu_t = \mu_t$. Our next step is to prove that the measure defined by the relation (0.1) is T-invariant. For any given set $A \in \mathcal{B}$ by the formula (0.1) we get:

$$\mu_T(T^{-1}A) = \sum_{k=0}^{\infty} \mu_t(T^{-k-1}A \cap E_{>k}) = \sum_{k=0}^{\infty} \mu_t(T^{-k-1}A \cap (E_{>k+1} \cup E_{k+1}))$$

$$= \sum_{k=1}^{\infty} \mu_t(T^{-k}A \cap E_{>k}) + \sum_{k=1}^{\infty} \mu_t(T^{-k}A \cap E_k) = \sum_{k=1}^{\infty} \mu_t(T^{-k}A \cap E_{>k}) + T_t^*\mu_t(A)$$

$= \sum_{k=0}^{\infty} \mu_t(T^{-k}A \cap E_{>k}) = \mu_T(A)$. Here we used that $\mu_t \in \mathcal{M}_{T_t}$, i.e. $T_t^*\mu_t = \mu_t$.

The last assertion follows from $\mu_T(X) = \sum_{k=0}^{\infty} \mu_t(T^{-k}X \cap E_{>k}) \le \sum_{k=0}^{\infty} \mu_t(E_{>k}) = \sum_{k=0}^{\infty} k\mu_t(E_k) < \infty$. □

Note that the construction of T_t can be considered as a measurable time change for the dynamical system $(T, X, \mathcal{B}, \mu)$. An important example of the application of above construction is the analysis of a *logistic* (quadratic) map $Tx := ax(1-x)$. It is worth note that the measure μ_T constructed by a finite T_t-invariant measure is not necessarily finite, e.g. in the case of a Poincare map of a piecewise expanding maps with a neutral singularity.

Consider now a dual construction. For a positive integer valued function $I \in \mathbf{L}^1(X, \mathcal{B}, \mu)$ construct a new measurable space $X^{(I)} := \{(x, k) : \ x \in X, k \in 1, 2, \dots, I(x)\}$. Using a standard Borel σ-algebra on this set we define a measure $\mu^{(I)}$ by the relation $\mu^{(I)}((A, k)) := \mu(A)/\mu(I) \ \forall k, A \in \mathcal{B}$. On this space we consider the so called *integral map* $T^{(I)}(x, k) := \begin{cases} (x, k+1) & \text{if } k+1 \le I(x) \\ (Tx, 1) & \text{otherwise} \end{cases}$. Then $\mu^{(I)} \in \mathcal{M}_{T^{(I)}}$ (exercise).

It is natural to think about the space $X^{(I)}$ as a *tower* whose base is the space X with $I(x)$ floors above each point $x \in X$. Under the action of $T^{(I)}$ a point (x, k) goes up vertically by 1 floor if it is possible, or goes down till the first floor to the point $(Tx, 1)$ otherwise. In this construction we identify the original space X with the set of pairs $(x, 1)$ – the first floor.

Lemma 0.12 $T = (T^{(I)})_{(X,\{1\})}$ *(derivative map), and if $\cup_{n\ge0} T^n E = X$ then $T = T_E^{(t_E)}$ (integral map).*

Proof. The first statement follows immediately from the definition. To prove the second one, observe that $T_E \in \mathbf{L}^1(E, \mathcal{B}_E, \mu_E)$ and each point $x \in T^k E_n$ for $0 \le k < n$ can be represented as $(T^{-k}x, k)$ with $T^{-k}x \in E_n$. Therefore, X can be identified with $X^{(t_E)}$, while T acts in this space as the integral map. □

Theorem 0.13 *Let $(T, X, \mathcal{B}, \mu)$ be ergodic (weak mixing, mixing) and let $T\mathcal{B} \subseteq \mathcal{B}$, then both Poincare and integral maps are ergodic (weak mixing, mixing).*

Proof. We prove only ergodicity and start with the Poincare map (which was already proven). Let $\mu(A) > 0$ and $\mu_E(T_E^{-1}A \Delta A) = 0$, i.e. it is "measurably" invariant. Since $\mu \in \mathcal{M}_T^0$ we have $\mu(\cup_{n\ge0} T^n A \Delta X) = 0$ hence $0 = \mu_E(A \Delta (E \cap (\cup_{n\ge0} T^n A))) = \mu_E(A \Delta (E \cap X))$ and thus $\mu_E(A \Delta E) = 0$.

In the case of the integral map consider a $T^{(I)}, \mu^{(I)}$-invariant set A with $\mu(A) > 0$. Then $A \cap X$ is T, μ-invariant. Since $\mu \in \mathcal{M}_T$ we have $\mu((A \cap X) \Delta X) = 0$. But then $\mu^{(I)}(A \Delta X^{(I)}) = 0$. □

Another important construction is related to the problem of the *noninvertability* of T, i.e. that $\#\{T^{-1}x\} \ne 1$ for some $x \in X$. By an invertible *extension* of a DS (T, X) one means a bijective (one-to-one) map $(\vec{T}, \vec{X})$

defined on the phase space $\vec{X} := X \times Y$ and having the property $\vec{T}^n(x,y) = (T^n x, y^{(n)})$ for all n and some $y^{(n)} = y^{(n)}(x,y)$, i.e. the projection to X of the point $\vec{T}^n(x,y)$ is equal to $T^n x \quad \forall x \in X, y \in Y, n \in \mathbb{Z}_+$. It turns out that in some cases there exist finite dimensional extensions, e.g. the baker map $\vec{T}_2(x,y) := \begin{cases}(2x, y/2) & \text{if } x \in [0,1/2] \\ (2x-1, (y+1)/2) & \text{otherwise}\end{cases}$ – is an extension for the dyadic map. The latter construction can be further extended for any map T satisfying the property that there exists a partition $\{X_i\}$ of the phase space X_i on the components on each of which the map T is a homeomorphism onto X. In this case one can define the extension as follows: $\vec{T}(x,y) := (Tx, T^{-1}_{|X_{i(x)}} y)$, where $i(x)$ stands for the index of the element of the partition containing the point x. How this this works for a non-complete tent map?

Now using a more sophisticated construction we shall construct an invertible extension for a general endomorphism $(T, X, \mathcal{B}, \mu)$. Let $\vec{X} := \{\vec{x} := \{x_i\}_{i=0}^{\infty} : \quad Tx_i = x_{i-1} \in X \ \forall i > 0\}$. Note that $\vec{x}$ is not a trajectory of the point x_0 but rather a trajectory of the point x_∞, and that not all possible sequences with elements from X are contained in $\vec{X}$. On this space we introduce the map $\vec{T} : \vec{X} \to \vec{X}$ defined by the relation $(\vec{T}\vec{x})_i := Tx_i \ \forall i \geq 0$, which is called the *natural extension* of the map T. Observe that this map is *invertible*, indeed, $(\vec{T}^{-1}\vec{x})_i = x_{i+1} \quad \forall \vec{x} \in \vec{X}, i \geq 0$. Equipping the space $\vec{X}$ with the σ-algebra $\vec{\mathcal{B}}$ generated by sets of type $B_{i,A} := \{\vec{x} \in \vec{X} : \ x_i \in A\}, \quad A \in \mathcal{B}, \ i \geq 0$, we define on this space a measure $\vec{\mu}$ by the relation $\vec{\mu}(B_{i,A}) := \mu(A)$. To show that this measure is well defined, consider $\vec{\mu}(\vec{x} \in \vec{X} : \ x_i \in A_i, \ i = 0, 1, \ldots, n) = \mu(T^{-n}A_0 \cap T^{-n+1}A_1 \cap \ldots \cap A_n)$. Thus, we obtained a collection of compatible finite dimensional probability distributions, which by Kolmogorov's theorem can be uniquely extended to a probabilistic measure on the σ-algebra $\vec{\mathcal{B}}$.

Theorem 0.14 *$\vec{\mu} \in \mathcal{M}_{\vec{T}}$, and $(\vec{T}, \vec{\mu})$ is ergodic/mixing iff (T, μ) is ergodic/mixing respectively.*

This result shows that $(\vec{T}, \vec{X}, \vec{\mathcal{B}}, \vec{\mu})$ is an automorphism, i.e. a measure preserving one-to-one system.

Proof. It is enough to check the question about the invariance on sets of type $B_{i,A}$, where it follows from the definition of $\vec{T}$. Indeed, $\vec{\mu}(\vec{T}^{-1}B_{i,A}) = \vec{\mu}(B_{i+1,A}) = \vec{\mu}(B_{i,A})$. To prove the ergodicity let $B := \{\vec{x} \in \vec{X} : \ x_0 \in A \in \mathcal{B}\}$. If A is T,μ-invariant, then B is clearly $\vec{T},\vec{\mu}$-invariant, and $\vec{\mu}(B) = \mu(A)$. Hence, if (T,μ) is nonergodic, then the same holds for $(\vec{T},\vec{\mu})$. Assume now that (T,μ) is ergodic. Then $S_n\varphi(x_0) \stackrel{n\to\infty}{\longrightarrow} \mu(\varphi)$ μ-a.e. for each $\varphi \in \mathbf{L}^1(X, \mathcal{B}, \mu)$. Thus, for any function $\Phi \in \mathbf{L}^1(\vec{X}, \vec{\mathcal{B}}, \vec{\mu})$ of type $\Phi(\vec{x}) = \varphi(x_i)$ we have $\frac{1}{n}\sum_{k=0}^{n-1}\Phi(\vec{T}^k\vec{x}) \stackrel{n\to\infty}{\longrightarrow} \vec{\mu}(\Phi)$. On the other hand, due to the condition $Tx_i = x_{i-1}$, for any function $\Psi(x_0, \ldots, x_i)$ there exists a function $\psi \in \mathbf{L}^1(X, \mathcal{B}, \mu)$ for which $\Psi(x_0, \ldots, x_i) = \psi(x_i)$. Therefore, such functions are dense in $\mathbf{L}^1(\vec{X}, \vec{\mathcal{B}}, \vec{\mu})$, which yields the ergodic theorem (and, hence, ergodicity) for $(\vec{T}, \vec{\mu})$.

To prove mixing, consider $\varphi \in \mathbf{L}^2(X, \mathcal{B}, \mu)$ such that $\mu(\varphi) = 0$ and let $\Phi(\vec{x}) := \varphi(x_i)$. Then $(U^n_{\vec{T}}\Phi, \Phi) = (U^n_T\varphi, \varphi)$ and both the right hand side and the left hand side vanish simultaneously when $n \to \infty$. The general case follows from this argument again due to the density of the functions Φ in $\mathbf{L}^2(\vec{X}, \vec{\mathcal{B}}, \vec{\mu})$. □

Now we are in a position to prove using the natural extension construction that an exact endomorphism is mixing. This will be done in two steps. First we shall show that the natural extension of an exact endomorphism is a *K-automorphism*, i.e. an invertible DS $(T, X, \mathcal{B}, \mu)$ such that $\forall r \in \mathbb{Z}_+$ and $\forall A_0, A_1, \ldots, A_r \in \mathcal{B}$ we have $\lim\limits_{n\to\infty} \sup_{B \in \mathcal{B}_n^\infty(A_1,\ldots,A_r)} |\mu(A_0 \cap B) - \mu(A_0)\mu(B)| = 0$, where the supremum is taken over sets B from the minimal σ-algebra $\mathcal{B}_n^\infty(A_1, \ldots, A_r)$ generated by the sets of type $T^k A_i, \ k \geq n, \ 1 \leq i \leq r$. On the second step we shall show that any K-automorphism is mixing, which proves the mixing of the initial exact endomorphism by Theorem 0.14.

Lemma 0.13 *If an endomorphism is exact then its natural extension is a K-automorphism. ([KSF], p.240)*

Proof. Let $(T, X, \mathcal{B}, \mu)$ be an exact DS. Recall that its natural extension – the automorphism $(\vec{T}, \vec{X}, \vec{\mathcal{B}}, \vec{\mu})$ is a $\vec{\mu}$-measure preserving invertible map. One can show that the following is equivalent to the definition of the K-automorphism: there exists a σ-subalgebra $\vec{\mathcal{B}}^0 \subset \vec{\mathcal{B}}$ such that (a) $\vec{T}\vec{\mathcal{B}}^0 \supset \vec{\mathcal{B}}^0$, (b) $\bigvee_{n\in\mathbb{Z}} \vec{T}^n\vec{\mathcal{B}}^0 = \vec{\mathcal{B}}$, (c) $\cap_{n\in\mathbb{Z}}\vec{T}^n A$ belongs to the trivial σ-algebra $\forall A \in \vec{\mathcal{B}}^0$. Here $\mathcal{B} \bigvee \mathcal{B}'$ means the minimal σ-algebra generated by sets of type $A \cap A'$ with $A \in \mathcal{B}, \ A' \in \mathcal{B}'$.

We define $\vec{\mathcal{B}}^0$ as a σ-subalgebra consisting of sets $\vec{A} := \{\vec{x} = \{x_i\}_{i=0}^{\infty} \in \vec{X} : \ x_0 \in A_0 \in \mathcal{B}\}$ for all $A_0 \in \mathcal{B}$. To check the condition (a) observe that for $\vec{A} \in \vec{\mathcal{B}}^0$ we have $\vec{T}\vec{A} = \{\vec{x} \in \vec{X} : \ \vec{T}^{-1}\vec{x} \in \vec{A}\} = \{\vec{x} \in \vec{X} : \ x_1 \in A_0 \in \mathcal{B}\}$. If $\exists A_1 \in \mathcal{B}$ such that $A_0 = T^{-1}A_1$ then $\vec{T}\vec{A} = \{\vec{x} \in \vec{X} : \ x_0 \in A_1\} \in \vec{\mathcal{B}}^0$, i.e. (a) holds.

(b) The σ-algebra $\vec{T}^n\vec{\mathcal{B}}^0$ consists of sets of type $\{\vec{x} \in \vec{X} : \ x_0 \in A_n \in \mathcal{B}\}$. On the other hand, by definition such sets generate the entire σ-algebra $\vec{\mathcal{B}}$, which yields (b).

(c) For $n \geq 0$ the sets from the σ-algebra $\vec{T}^{-n}\vec{\mathcal{B}}^0$ are in an isometrical one-to-one correspondence with the subsets of $\vec{T}^{-n}\vec{\mathcal{B}}$, which finishes the proof. □

Lemma 0.14 *Any K-automorphism $(T, X, \mathcal{B}, \mu)$ is mixing. ([KSF], p.234)*
Proof. Observe that it is enough to check the mixing for all functions $\varphi, \psi \in \mathbf{L}^2_0(X, T^r\mathcal{B}^0, \mu)$ for some $r \in \mathbb{Z}$ since such functions by the condition (b) (see the proof of the previous result) generate the entire space $\mathbf{L}^2_0(X, \mathcal{B}, \mu)$. Observe now that $\varphi \circ T^k, \psi \circ T^k \in \mathbf{L}^2_0(X, T^{r+k}\mathcal{B}^0, \mu)$. Therefore, denoting by P_k the orthogonal projecting operator to $\mathbf{L}^2_0(X, T^r\mathcal{B}^0, \mu)$ from $\mathbf{L}^2_0(X, T^{r+k}\mathcal{B}^0, \mu)$, we get:
$\mu(\varphi \circ T^n \cdot \psi) = \mu(\varphi \cdot \psi \circ T^{-n}) = (\varphi, (\psi \circ T^{-n})^*)_{\mathbf{L}^2} = (P_{r-n}\varphi, (\psi \circ T^{-n})^*)_{\mathbf{L}^2} \le ||P_{r-n}\varphi|| \cdot ||\psi|| \stackrel{n\to\infty}{\longrightarrow} 0$, since $||P_{r-n}\varphi|| \stackrel{n\to\infty}{\longrightarrow} 0$ due to the condition (c). □

Entropy. Classical approach.

Let T be a continuous map $T : X \to X$. Denote by $\xi := \{\xi_i\}_1^r$ with $\xi_i \in \mathcal{B}$ a finite measurable *partition* of $(X, \mathcal{B}, \mu)$, i.e. $\mu(\cup_i \xi_i) = 1$, $\mu(\xi_i) > 0\ \forall i$, $\mu(\xi_i \cap \xi_j) = \emptyset\ \forall i, j$. For a pair of partitions ξ, η we define their common *refinement*: $\xi \bigvee \eta := \{\xi_i \cap \eta_j :\ \mu(\xi_i \cap \eta_j) > 0\}$. Observe that making the refinement of the collection of sets $\{T^{-1}\xi_i\}$ we again get a finite measurable partition which we denote $T^{-1}\xi$. Consider the n-th refinement ξ^n of the partition ξ, which can be defined inductively $\xi^n := \xi^{n-1} \bigvee T^{-1}\xi^{n-1}$, $\xi^0 := \xi$. Then the measure μ induces a distribution $\{\mu(\xi^n_i)\}$. Setting $H(\xi) := -\sum_i \mu(\xi_i)\ln\mu(\xi_i)$ we define *the conditional entropy of a partition*: $h_\mu(T, \xi) := \inf_{n\to\infty} \frac{1}{n}H(\xi^n) = \lim_{n\to\infty} \frac{1}{n}H(\xi^n)$, and finally *the metric entropy*: $h_\mu(T) := \sup_\xi h_\mu(T, \xi)$.

Topological entropy. Now let $\xi := \{\xi_i\}_1^r$ be a covering of X by open sets. Define a transition matrix $M := \{m_{ij}\}$, where $m_{ij} = 1$ if $\mu(\xi_i \cap T^{-1}\xi_j) > 0$ and $= 0$ otherwise. Then on the Cantor set X_M – the space of sequences with the alphabet $\mathcal{A} := \{1, 2, \dots, r\}$ with the transition matrix M the left shift map σ defines a symbolic dynamical system. Denoting by A^n_ξ the set of all admissible words of length n, i.e. different pieces of length n of all trajectories of (σ, X_M) and by $\#A^n_\xi$ – their number, we define $h(T, \xi) := \inf\limits_{n\to\infty} \dfrac{1}{n}\log(\#A^n_\xi) = \lim\limits_{n\to\infty} \dfrac{1}{n}\log(\#A^n_\xi)$. Then the *topological entropy* is $h(T) := \sup_\xi h(T, \xi)$.

Variational principle. For "good enough" maps we have $h(T) = \sup_{\mu\in\mathcal{M}^0_T} h_\mu(T)$.

Entropy. Approach based on Bowen's metric.

Let (X, ρ) be a compact metric space and let T be a continuous map $T : X \to X$. For any $n \in \mathbb{Z}_+$, the *n-th Bowen metric ρ_n on X* is defined by $\rho_n(x, y) := \max\left\{\rho\left(T^k(x), T^k(y)\right) :\ k = 0, \dots, n-1\right\}$. For every $\varepsilon > 0$ we denote by $B^n_\varepsilon(x)$ the open ball of radius ε in the metric ρ_n around x.

The *local measure-theoretical entropy* of $\mu \in \mathcal{M}_T(X)$ at a point $x \in X$ is defined by
$$h_\mu(T, x) := -\lim_{\varepsilon\to0}\lim_{n\to\infty} \frac{1}{n}\log\mu(B^n_\varepsilon(x)).$$
M. Brin and A. Katok (On local entropy. Geometric dynamics (Rio de Janeiro, 1981), 30–38, Lecture Notes in Math., 1007, Springer, Berlin, 1983) proved that for an ergodic $\mu \in \mathcal{M}_T(X)$, $h_\mu(T, x)$ is well defined and does not depend on x for μ-a.e $x \in X$. Roughly speaking it measures the exponential rate of decay of the measure of points that stay ε-close to the point x under forward iterates of the map.

For a general (non-ergodic) $\mu \in \mathcal{M}_T(X)$ one defines *measure-theoretical entropy* by $h_\mu(T) := \mu(h_\mu(T, \cdot))$. Finally, the *topological entropy* is defined by $h_{\text{top}}(T) := \sup\{h_\mu(T) :\ \ \mu \in \mathcal{M}(X)\}$.
Calculate the local entropy for contracting, rotation and doubling maps.

Lecture 4. Linear Toral Automorphisms [16.12.19]

Rich classes of dynamical systems can be constructed by using linear transformations on a torus.

Definition 12.1. A d-dimensional torus Tor^d (for short, a d-torus) is a unit cube in $\mathbb{R}^d$ whose opposite faces are identified, i.e. we assume $x_i + 1 = x_i$ for all $i = 1, \dots, d$. Alternatively, Tor^d can be defined as the factor space $\mathrm{Tor}^d = \mathbb{R}^d/\mathbb{Z}^d$.

The 1-torus Tor^1 is just the unit circle. The 2-torus Tor^2 is a square with identified opposite sides. One is used to visualize Tor^2 as the surface of a doughnut.

Definition 12.2. Let $\vec{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$. A *translation* of the d-torus Tor^d is $T_{\vec{a}}(\vec{x}) = \vec{x} + \vec{a} \pmod 1$ i.e. $T_{\vec{a}}(x_1, \dots, x_d) = (x_1 + a_1, \dots, x_d + a_d)$, and each $x_i + a_i$ is taken modulo 1.

Note: if $d = 1$, the translation $T_{\vec{a}}$ is just the rotation of the unit circle Tor^1 through the angle a_1.

Remark 12.3. Every translation $T_{\vec{a}}$ is a diffeomorphism of the torus Tor^d . Moreover, $T_{\vec{a}}$ is a linear map and an isometry (i.e., it preserves distances between points and angles between vectors). Also, for each $n \in \mathbb{Z}$ the nth iteration of $T_{\vec{a}}$ is $T^n_{\vec{a}} = T_{n\vec{a}}$.

Proposition 12.4. $T_{\vec{a}}$ preserves the Lebesgue measure m on the torus Tor^d.

Proof. The derivative $DT_{\vec{a}}$ is the identity matrix, hence det $DT_{\vec{a}} = 1$. Now the invariance of the Lebesgue measure follows as in Exercise 5.6(a). □

THEOREM 12.5. The following are equivalent:
(a) $T_{\vec{a}}$ is ergodic with respect to the Lebesgue measure;
(b) $\forall \vec{x} \in \mathrm{Tor}^d$ the trajectory $\{T^n_{\vec{a}}(x)\}_{n=0}^\infty$ is dense in Tor^d;
(c) $\vec{a}$ is non-resonant, namely the numbers $a_1, \dots, a_d$ are rationally independent of 1, i.e. $k_0 + \sum_{i=1}^d k_i a_i \neq 0$ for any integers $k_i \in \mathbb{Z}$ unless $k_0 = k_1 = \cdots = k_d = 0$.

Proof. We start with the implication (a)⇒(b). By 7.25, if $T_{\vec{a}}$ is ergodic, then the trajectory of almost every point $\vec{x} \in \mathrm{Tor}^d$ is dense. On the other hand, if the trajectory of at least one point $\vec{x}^0 \in \mathrm{Tor}^d$ is dense, then the trajectory of every point $\vec{x} \in \mathrm{Tor}^d$ is dense. Indeed, $T^n_{\vec{a}}(\vec{x}) = \vec{x} + n\vec{a} = \vec{x}_0 + n\vec{a} + (\vec{x} - \vec{x}^0) = T^n_{\vec{a}}(\vec{x}^0) + (\vec{x} - \vec{x}^0)$ (all calculations here are done modulo 1). So, to find a subsequence $T^{n_k}_{\vec{a}}(\vec{x})$ converging to a given point $\vec{c} \in \mathrm{Tor}^d$, it is enough to find a subsequence $T^{n_k}_{\vec{a}}(\vec{x}^0)$ converging to the point $\vec{c} - \vec{x} + \vec{x}^0$ (mod 1).

Next we prove the implication (b)⇒(c). Suppose (c) fails. Consider the trajectory of the point $\vec{x}_0 = \vec{0}$, which consists of the points $\vec{x}^n = T^n_{\vec{a}}(\vec{0}) = n\vec{a}$ (mod 1). Their coordinates must satisfy the relation $\sum_{i=1}^d k_i x_i^n = n \sum_{i=1}^d k_i a_i = -nk_0 = 0 \pmod 1$. We now show that the sequence $\{\vec{x}_n\}$ cannot be dense in Tor^d. Let $M = \max\{|k_i|\}$ and consider a small cube $K' \subset \mathrm{Tor}^d$ defined by $K' = \{\vec{x} :\ 0 \le x_i \le 1/(2dM)$ for $i = 1, \dots, d\}$. For every point $\vec{x} \in K'$ the relation $\sum_i k_i x_i = 0$ (mod 1) is equivalent to $\sum_i k_i x_i = 0$ (without being taken modulo 1). This is just a hyperplane in $\mathbb{R}^d$, which cannot be dense in any open set, in particular in K'.

Lastly, we prove the implication (c)⇒(a). The proof uses Fourier analysis. Let $T_{\vec{a}}$ be not ergodic and $B \subset \mathrm{Tor}^d$ be an invariant set with measure $0 < m(B) < 1$. The function $f = 1_B - m(B)$ is $T_{\vec{a}}$-invariant and bounded, so it belongs in $L^2_m(\mathrm{Tor}^d)$. Consider functions $\varphi_{\vec{m}}(\vec{x}) = e^{2\pi i \langle \vec{k}, \vec{x}\rangle} = \cos 2\pi\langle \vec{k}, \vec{x}\rangle + i \sin 2\pi \langle \vec{k}, \vec{x}\rangle$ where $\vec{k} \in \mathbb{Z}^d$ and $\langle \vec{k}, \vec{x}\rangle = \sum_i k_i x_i$ is the scalar product in $\mathbb{R}^d$. These functions are periodic with period 1 in each coordinate, so that they are well defined on Tor^d. They make an orthonormal basis (the Fourier basis) in the space $L^2_m(\mathrm{Tor}^d)$ of complex-valued functions on Tor^d. This means that if functions f and $\varphi_{\vec{k}}$ are orthogonal $\forall \vec{k} \in \mathbb{Z}^d$ (i.e. $m(f \cdot \varphi^*_{\vec{k}}) = 0$) then $f(\vec{x}) = 0$ a.e. By using change of variable, for any $T_{\vec{a}}$-invariant function $f(\vec{x})$ we get $m(f \cdot \varphi^*_{\vec{k}}) = m(f(\cdot + \vec{a}) \cdot \varphi^*_{\vec{k}}) = m(f \cdot \varphi^*_{\vec{k}}(\cdot - \vec{a})) = e^{2\pi i \langle \vec{k}, \vec{a}\rangle} m(f \cdot \varphi^*_{\vec{k}})$. By the assumption (c), we have $e^{2\pi i \langle \vec{k}, \vec{a}\rangle} \neq 1$ for all $\vec{k} \in \mathbb{Z}^d$ except $\vec{k} = \vec{0}$. Therefore, we get the orthogonality for all $\vec{k} \neq \vec{0}$. For $\vec{k} = \vec{0}$ we have $\varphi_{\vec{k}}(\vec{x}) \equiv 1$, hence the orthogonality holds as well for our function $f(\vec{x}) = 1_B - m(B)$. Thus, $f(\vec{x}) = 0$ almost everywhere, a contradiction. □

Note: the translation $T_{\vec{a}}$ is not mixing, not even weakly mixing (just like circle rotations).

An **alternative proof** is related to the observation that the map $T_{\vec{a}}$ is a direct product of one-dimensional rotations. Therefore it is ergodic iff the corresponding one-dimensional Koopman operators have no common eigenvalues. Remark, that the eigenvalue $e^{i2\pi\alpha}$ implies also the eigenvalues $e^{-i2\pi\alpha}$ and $e^{i2\pi(1-\alpha)}$, which explains the nontrivial form of the resonance condition.

Next, we restrict ourselves to the 2-torus $\mathrm{Tor}^2 = \{(x, y) :\ 0 \le x, y < 1\}$ (just for the sake of simplicity). Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix with integral entries $a, b, c, d \in \mathbb{Z}_+$. Assume that $\det A = ad - bc = \pm 1$.

DEFINITION 12.6. Any matrix A with the above properties defines a *linear toral automorphism* $T_A : \mathrm{Tor}^2 \to \mathrm{Tor}^2$ by $T_A(\vec{x}) = A\vec{x} \pmod 1$ i.e. $T_A(x, y) = (ax + by \pmod 1,\ cx + dy \pmod 1)$.

Note: T_A is well defined on Tor^2 whenever the entries of A are integers.

EXERCISE 12.7. Show that T_A is one-to-one. Hint: if $T_A(\vec{x}) = T_A(\vec{y})$, then $T_A(\vec{z}) = \vec{0} \pmod 1$ for $\vec{z} = \vec{x} - \vec{y} \pmod 1$. Next consider the system $az_1 + bz_2 = m_1$ and $cz_1 + dz_2 = m_2$ for some $m_1, m_2 \in \mathbb{Z}$ and show that its only solution (z_1, z_2) is a pair of integers z_1, z_2.

REMARK 12.8. The map T_A is a linear diffeomorphism of the torus Tor^2. Moreover, $\forall n \in \mathbb{Z}$ its nth iterate T^n_A is $T^n_A = T_{A^n}$. In particular, $T_A^{-1} = T_{A^{-1}}$. The point $\vec{0} = (0, 0)$ is always a fixed point: $T_A(\vec{0}) = \vec{0}$.

EXERCISE 12.9. Show that every point $(x, y) \in \mathrm{Tor}^2$ with rational coordinates $x, y \in \mathbf{Q}$ is periodic.

Let us prove even more general statement for an arbitrary dimension d. First, we prove that if a point x is periodic, then it has rational coordinates. Indeed, if it is periodic, then $\exists m \in \mathbb{Z}\ A^m x - x = (A^m - I)x \in \mathbb{Z}^d$. Since $\det(A^m - I) \neq 0$ we have $(A^m - I)^{-1}$ has rational entries. Hence $x = (A^m - I)^{-1}(A^m x - x) \in (A^m - I)^{-1}\mathbb{Z}^d \subset Q^d$ – has rational entries. It remains to prove that a rational vector x is a periodic point. Let r be the smallest common denominator of the entries of x. Denote by π_r the projector from the matrix A to $A_r := A \pmod r$. Then the matrix A_r acts on the finite ring $\mathbb{Z}_r$. Therefore $\exists m \in \mathbb{Z}_+$ such that

$\pi_r^m A = \pi_r(A^m) = I$. Hence $\pi_r(A^m - I) = 0$ or $A^m = I \pmod r$. Hence all entries of the matrix $(A^m - I)$ are multiple by r and thus $A^m x - x \in \mathbb{Z}_+$, i.e. x and $A^m x$ correspond to the same torus point.

PROPOSITION 12.10. T_A preserves the Lebesgue measure m on the torus Tor^2.

Proof. The derivative DT_A is the matrix A itself, hence $\det DT_A = \pm 1$. Now the invariance of the Lebesgue measure follows as in Exercise 5.6(a).

EXAMPLE 12.11. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The line $x = y$ consists entirely of fixed points, while all the other points are periodic with period 2, because $T_A^2 = \mathrm{id}$. The map T_A "flips" the torus across its main diagonal $x = y$. There are no dense orbits, hence the map is not ergodic. Note that the eigenvalues of A are $\lambda = \pm 1$.

EXAMPLE 12.12. Let $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ with any $m \in \mathbb{Z}$. The line $y = 0$ consists entirely of fixed points. Every line $y = \mathrm{const}$ is invariant under T_A. No dense orbits exist, so the map T_A is not ergodic. Note that A has one eigenvalue $\lambda = 1$ of multiplicity 2.

Note: if we picture Tor^2 as a cylinder with the base circle $\{(x, y) : \ 0 \le x < 1, \ y = 0\}$ and the vertical coordinate y, then T_A rotates every horizontal section $y = \mathrm{const}$ by the angle my, since it acts by $x \mapsto x + my \pmod 1$. The higher the section, the larger the angle of rotation, so that the whole cylinder is twisted upward (unscrewed). Such maps are called *twist maps.*

REMARK 12.13. Examples 12.11 and 12.12 illustrate what ergodic components of a map may look like. In 12.11, each ergodic component of T_A is either a periodic orbit of period 2 or a fixed point (the latter can be ignored since they make a null set). In 12.12, most of the ergodic components are sections $y = c$ (precisely, such are all sections with irrational c). The sections $y = c$ for rational c can be further decomposed into periodic orbits or ignored altogether, since their total measure is zero.

EXAMPLE 12.14. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. In this case, the only fixed point is $\vec{0}$, and the description of the map T_A is not so simple. Note that the eigenvalues of A are $\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}$ and the corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$ are orthogonal. Let L_1 and L_2 be the perpendicular lines on Tor^2 spanned by the vectors v_1 and v_2, respectively. Both lines are invariant under T_A. Since $\lambda_1 > 1$, the line L_1 is expanded ("stretched out") by a factor of λ_1 under T_A. On the other hand, $|\lambda_2| < 1$, so the other line L_2 is compressed (contracted) by a factor of $|\lambda_2|$ under T_A (and it is flipped over, because $\lambda_2 < 0$). Locally, near the fixed point $\vec{0}$, the action of T_A is shown on Fig. 12, it looks like a "saddle". The orbit of any point near $\vec{0}$ lies on a hyperbola (or a pair of hyperbolas). In differential equations such fixed points are referred to as hyperbolic.

DEFINITION 12.15. A linear total automorphism T_A is *hyperbolic* if the eigenvalues of A are real numbers different from ± 1.

EXERCISE 12.16. For any hyperbolic toral automorphism eigenvalues are irrational. Hint: $(a+d)^2 \pm 4 \neq n^2 \ \forall a, d, n \in \mathbb{Z}_+$.

Let $\lambda = \min\{|\lambda_1|, |\lambda_2|\}$. Note that $\lambda^{-1} = \max\{|\lambda_1|, |\lambda_2|\}$. Note also that the inverse matrix A^{-1} has eigenvalues λ_1^{-1} and λ_2^{-1} and the same eigenvectors as A does. So, A^{-1} contracts L_1 by a factor of λ and expands L_2 by a factor of λ^{-1}.

DEFINITION 12.17. The line L_1 spanned by the eigenvector v_1 corresponding to the larger (in absolute value) eigenvalue of A is called the *unstable manifold.* It is expanded ("stretched out") under T_A. The line L_2 spanned by the eigenvector v_2 corresponding to the other, smaller eigenvalue of A is called the *stable manifold.* Note that both lines extend infinitely long, they wrap around the torus infinitely many times.

EXERCISE 12.18. Show that for any hyperbolic toral automorphism the lines L_1 and L_2 are dense on the torus Tor^2. Hint: verify that the equation of the line L_i for $i = 1, 2$ is $y = \gamma_i x$ where $\gamma_i = (\lambda_i - a)/b$ is an irrational number by 12.16 (assume that $b \neq 0$ for simplicity). For any real number α the points $(n\alpha, n\alpha\gamma_i) \pmod 1$ for $n = 0, 1, 2, \dots$ belong in L_i. Now use Theorem 12.5 to show that for some $\alpha \neq 0$ these points make a dense set. (Note: your α should be chosen carefully, so that (c) will be satisfied!)

Note: with a little extra effort one can show that for any $\varepsilon > 0$ there is a $d > 0$ such that every segment of length d on the line L_1 intersects every disk of radius $\varepsilon > 0$ on the torus Tor^2.

RECTANGULAR PARTITIONS 12.19. Further analysis of the map T_A involves symbolic dynamics. According to 10.17, one needs to start with a generating partition. Here we partition the torus Tor^2 into

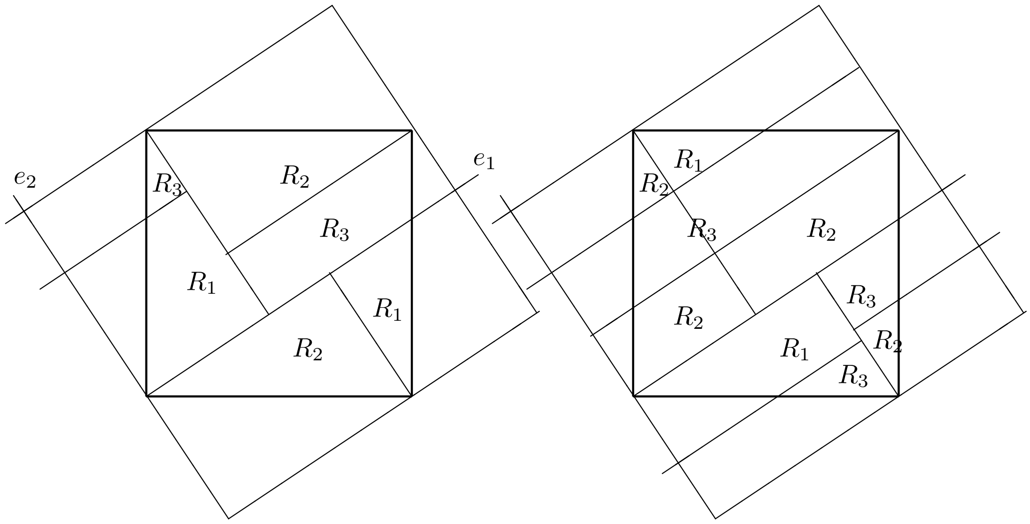

Рис. 1: Markov partition into 3 elements R_1, R_2, R_3(left), its image (right).

rectangles with sides parallel to the stable and unstable lines. Figure 1 (left) [13(a)] shows the partition of the torus into three rectangles for Example 12.14. The sides of the rectangles are made by pieces of the lines L_1 and L_2. Fig. 13(b) shows the images of those three rectangles under T_A, respectively. Note that each rectangle is stretched by T_A in the direction of L_1 (the unstable direction) and compressed in the direction of L_2 (the stable direction), but it retains its rectangular shape.

PROPER INTERSECTION 12.20. Denote the rectangles by R_1, R_2, R_3 and let us examine the intersections $T_A(R_i) \cap R_j$ for each pair i, j. If it is not empty, then it is a subrectangle in R_j, which stretches completely across R_j in the unstable direction. Also, it is a subrectangle in $T_A(R_i)$, which stretches completely across $T_A(R_i)$ in the stable direction. In other words, $T_A(R_i)$ intersects R_j *properly* (transversely).

DEFINITION 12.21. A partition into rectangles $\{R_i\}_{i=1}^r$ with sides parallel to L_1 and L_2 is called a *Markov partition* if all intersections $T_A(R_i) \cap R_j$, $1 \le i, j \le r$, with nonempty interior are connected and proper.

LEMMA 12.22. Any Markov partition is generating.

Proof. If not, then some distinct points $x \neq y$ have the same itinerary, i.e. $x, y \in \cap_{k=-\infty}^{\infty} T_A^{-k} R_{i_k}$ for some sequence $\{i_k\}_{k=-\infty}^{\infty}$. However, the diameter of the set $\cap_{k=-n}^{n} T_A^{-k} R_{i_k}$ is $O(\lambda^n)$, which converges to zero as $n \to \infty$, a contradiction. □

Note: it is essential for this proof that the intersections $T(R_i) \cap R_j$ are connected, without this assumption Lemma 12.22 may fail.

Recall that a generating partition $\mathrm{Tor}^2 = R_1 \cup \cdots \cup R_r$ into r disjoint subsets gives rise to a symbolic representation of an automorphism $T_A : \mathrm{Tor}^2 \to \mathrm{Tor}^2$ by a shift $\sigma : \vec{X} \to \vec{X}$ on a symbolic space with r symbols as defined by 10.15–10.17. By 10.18, every cylinder $C_{m,\dots,n}^{i_m,\dots,i_n}$ corresponds to the intersection $R_{m,\dots,n}^{i_m,\dots,i_n} = \cap_{k=m}^{n} T_A^{-k}(R_{i_k})$ that is, $\varphi^{-1} C_{m,\dots,n}^{i_m,\dots,i_n} = R_{m,\dots,n}^{i_m,\dots,i_n}$.

We now study the Lebesgue measure m and the induced measure μ on $\vec{X}$.

For each rectangle R_i, denote by s_i and u_i its sides parallel to the stable direction (L_2) and the unstable direction (L_1), respectively. Then $m(R_i) = s_i u_i$. Due to the properness of intersections, if $T_A R_i \cap R_j$ has nonempty interior, then $m(T_A R_i \cap R_j) = \lambda s_i u_j$, where $\lambda = \min\{|\lambda_1|, |\lambda_2|\}$.

LEMMA 12.23. (a) For any integers $m < n$ and any $i_m, i_{m+1}, \dots, i_n$ the set $R_{m,\dots,n}^{i_m,\dots,i_n}$ either has empty interior (and then zero measure) or has measure $m(R_{m,\dots,n}^{i_m,\dots,i_n}) = \lambda^{n-m} s_{i_m} u_{i_n}$

(b) The set $R_{m,n}(i_m, \dots, i_n)$ has nonempty interior if and only if $T_A(R_{i_k}) \cap R_{i_{k+1}}$ has nonempty interior for every $k = m, \dots, n-1$. *Proof* is a simple geometric inspection. □

DEFINITION 12.24. The *matrix of transition probabilities* is an $r \times r$ matrix $\Pi = (\pi_{ij})$ with entries $\pi_{ij} = \frac{m(T_A R_i \cap R_j)}{m(R_i)} = \begin{cases} \lambda u_j / u_i & \text{if } \operatorname{int}(T_A R_i \cap R_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$. The *stationary probability vector* is the row vector $\vec{p} = (p_1, \dots, p_r)$ with components $p_i = m(R_i) = s_i u_i$.

Notes: (a) $\vec{p}\Pi = \vec{p}$, i.e. $\vec{p}$ is a left eigenvector for the matrix Π with eigenvalue 1. In other words, it remains invariant (stationary) under the right multiplication by Π.

(b) For each i we have $\sum_j \pi_{ij} = 1$. Matrices with nonnegative entries whose rows sum up to one are called *stochastic matrices*.

(c) π_{ij} is the fraction of R_i that is mapped into R_j by T_A. We interpret π_{ij} as the probability for a point starting in R_i to move into R_j.

LEMMA 12.25. For any integers $m < n$ and any $i_m, \dots, i_n$ the intersection $R^{i_m,\dots,i_n}_{m,\dots,n}$ has measure $m(^{i_m,\dots,i_n}_{m,\dots,n}) = p_{i_m}\pi_{i_m i_{m+1}} \cdots \pi_{i_{n-1}i_n}$.

DEFINITION 12.26. A measure μ on $\vec{X}$ is called a *Markov measure* with a transition probability matrix $\Pi = (\pi_{ij})$ of size $r \times r$ and a stationary probability vector $\vec{p} = (p_1, \dots, p_r)$ if for every cylinder $C^{i_m,\dots,i_n}_{m,\dots,n}$ its measure is $\mu(C^{i_m,\dots,i_n}_{m,\dots,n}) = p_{i_m}\pi_{i_m i_{m+1}} \cdots \pi_{i_{n-1}i_n}$. This model is known in probability theory as a *Markov chain*. This analogy explains the term *Markov partition*.

COROLLARY 12.27. The Lebesgue measure m on Tor^2 corresponds to the Markov measure μ on Ω_r. The dynamical systems (Tor^2, T_A, m) and $(\vec{X}, \sigma, \mu)$ are isomorphic.

Now the study of the hyperbolic toral automorphism T_A reduces to the study of the shift σ on $\vec{X}$ with the Markov measure μ. Surprisingly, this reduction makes things a lot simpler.

NOTATION 12.28. For $k \geq 1$, denote by $\pi^{(k)}_{ij}$ the elements of the matrix Π^k. Note that $\vec{p}\Pi^k = \vec{p}\ \ \forall k \geq 1$.

EXERCISE 12.29. Show that $\pi^{(k)}_{ij} = \frac{m(T^k_A R_i \cap R_j)}{m(R_i)}$.

EXERCISE 12.30. Let $C = C^{i_m,\dots,i_n}_{m,\dots,n}$ and $C' = C^{i_{m'},\dots,i_{n'}}_{m',\dots,n'})$ be two cylinders such that $n < m'$. Let $t = m' - n$. Show that $\mu(C \cap C') = \mu(C)\,\mu(C')\,\pi^{(t)}_{i_n i_{m'}}/p_{i_{m'}}$. Hint: do that first for $t = 1$ and then use induction on t. A helpful formula is $C = \cup^r_{i=1} C^{i_m,\dots,i_n,i}_{m,\dots,n+1}$.

LEMMA 12.31. There is an $s \geq 1$ such that the matrix Π^s has all positive entries.

Proof. For large $s > 1$, the set $T^s_A(R_i)$ is a very long narrow rectangle, one long side of which lies on the line L_1. Then by Exercise 12.18 (and the remark after it), $T^s_A(R_i)$ intersects every rectangle R_j of the Markov partition. Now 12.29 completes the proof. □

THEOREM 12.32 (LIMIT THEOREM FOR MARKOV CHAINS). If Π^s has all positive entries for some $s > 0$, then for all $1 \leq i, j \leq r$ we have $\pi^{(t)}_{ij} \to p_j$ as $t \to \infty$.

Proof. Fix $1 \leq i, j \leq r$ and let $n, t \geq 1$. Since $\Pi^{n+t} = \Pi^n\Pi^t$, we have $\pi^{(n+t)}_{ij} = \sum^r_{k=1} \pi^{(n)}_{ik}\pi^{(t)}_{kj}$. Let $\delta_n = \min_k \pi^{(n)}_{ik} \geq 0$. Note that $\max_k \pi^{(n)}_{ik} \leq 1 - \delta_n$. Now let $m_t = \min_k \pi^{(t)}_{kj}$ and $M_t = \max_k \pi^{(t)}_{kj}$. The following estimate is rather elementary: $(1-\delta_n)m_t + \delta_n M_t \leq \pi^{(n+t)}_{ij} \leq \delta_n m_t + (1-\delta_n)M_t$. Hence, $M_{t+n} \leq M_t$ and $m_{t+n} \geq m_t$. Next we show that $M_t - m_t \to 0$ as $t \to \infty$. Let $t = ms + n$ with $0 \leq n < s$. Then $M_t - m_t \leq M_{ms} - m_{ms}$. Now, put $n = s$ and $t = (m-1)s$ in the estimate for $\pi^{(n+t)}_{ij}$ above and get $M_{ms} - m_{ms} \leq (1-2\delta_s)\,(M_{(m-1)s} - m_{(m-1)s})$. Since $\delta_s > 0$, we have $M_{ms} - m_{ms} < C\lambda^{m-1}$ where $C = M_s - m_s$ and $\lambda = 1 - 2\delta_s < 1$.

Therefore, $\pi^{(t)}_{ij} \to q_j$ as $t \to \infty$ with some $q_j \geq 0$ (independent of i). This implies $\vec{e}_i\Pi^t \to \vec{q}$ as $t \to \infty$ for each $i = 1, \dots, r$, where $\vec{e}_i$ is the ith canonical basis row-vector and $\vec{q} = (q_1, \dots, q_r)$. By linearity, $\vec{p}\Pi^t \to \vec{q}$ as $t \to \infty$, hence $\vec{q} = \vec{p}$. □

Two important facts follow from the above proof. First, the stationary vector $\vec{p}$ is unique. Second, the convergence $\pi^{(t)}_{ij} \to p_j$ is exponentially fast (in t).

COROLLARY 12.33. For any two cylinders $C, C' \subset \Omega_r$ we have $\lim_{n\to\infty} \mu(C \cap \sigma^{-n}(C')) = \mu(C)\,\mu(C')$.

Proof. This follows from 12.30, 12.31 and 12.32.

PROPOSITION 12.34. The shift $\sigma : \vec{X} \to \vec{X}$ with the Markov measure μ is mixing, and so is the original hyperbolic toral automorphism T_A.

Proof. This follows from 12.33 and standard approximation arguments used in the proof of 10.14.

REMARK 12.35. The hyperbolic toral automorphism T_A is actually Bernoulli, but the proof of this fact is rather sophisticated, we omit it.

Lecture 5. Lyapunov Exponents [10.12.09]

In order to get a *quantitative* description of chaoticity people introduced some numerical characteristics. One of them is the spectrum of Lyapunov exponents. We start with an abstract setting.

Let $(T, X, \mathcal{B}, \mu)$ be a measurable DS. For (μ-a.e.) $x \in X$ we associate a $d \times d$ matrix $A(x)$ and consider their products along the trajectories of the map T, i.e. $A^{(n)}(x) := A(T^n x) \times A(T^{n-1}x) \times \ldots \times A(x),\ \ x \in X$. Those products generate a skew product dynamical system with our fixed DS in the base and are called *linear co-cycles*. The phase space in the layers is $\mathbb{R}^d$ with the Lebesgue measure m^d. For a given $x \in X$ for each $n \in \mathbb{Z}_+$ one calculates the spectrum of the matrix $A^{(n)}(x)$ and ordering logarithms of modules of the eigenvalues of $A^{(n)}(x)$ normalized by $1/n$ obtain a collection of d ordered numbers: $\vec{\lambda}^{(n)}(x) := \{\lambda^{(n)}_1 \geq \lambda^{(n)}_2 \geq \ldots \geq \lambda^{(n)}_d\}$. In particular $\lambda^{(n)}_1 = \lambda^{(n)}_1(x)$ is equal to the largest among the logarithms of the modules of eigenvalues multiplied by $1/n$. Now we can study asymptotic properties of the vectors $\vec{\lambda}^{(n)}(x)$ as $n \to \infty$.

Assuming that $X \subset \mathbb{R}^d$ and that $A(x) = D_xT := (\partial(Tx)_i/\partial x_j)$ is the matrix of partial derivatives of the map T we are coming to the definition of the Lyapunov exponents as the limits of $\vec{\lambda}^{(n)}(x)$ as $n \to \infty$. Note: $D_xT^n = D_{T^{n-1}x}T \cdots D_{Tx}T \cdot D_xT$ by the chain rule and hence $A^{(n)}(x) := D_xT^n(x)$.

Exercise. Consider examples $Tx := Ax$ with different A and discuss properties of solutions.

Let us study in detail Lyapunov exponents in the 1D case, i.e. $d = 1$. In this case we can write explicitly $\lambda(x) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \ln(|T'(T^kx)|) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \ln(|(T^kx)'|)$.

Example. Let $T : X \to X$ be a diffeomorphism of the unit circle $X = S^1$ given by $T(x) = x + \frac{1}{3\pi}\sin 2\pi x$, where $0 \le x < 1$ is the cyclic coordinate on X. We have two fixed points here, $x_0 = 0$ and $x_1 = 1/2$. Lyapunov exponents exist at both fixed points: $\lambda(x_0) = \ln|T'(x_0)| = \ln\frac{5}{3} > 0$ and $\lambda(x_1) = \ln|T'(x_1)| = \ln\frac{1}{3} < 0$. Since $\lambda(x_0) > 0$, the point x_0 is unstable (a *repeller*). Likewise, x_1 is a stable point (an *attractor*). For any point $x \ne x_0$ we have $T^n(x) \to x_1$ while $T^{-n}(x) \to x_0$ for $x \ne x_1$ as $n \to \infty$. Hence, by the chain rule, $\lambda(x) := \lim_{n\to-\infty} \frac{1}{n}\ln|(T^nx)'| = \ln\frac{5}{3}$ m-a.e. , while $\lim_{n\to\infty}\frac{1}{n}\ln|(T^n(x_0))'| = -\ln 3$. This shows that Lyapunov exponents may depend on x even if they are well defined.

Theorem 0.15 *Let $(T, X, \mathcal{B}, \mu)$ be an ergodic DS and let $X \subset \mathbb{R}^d$, $d = 1$ and T be piecewise smooth. Then the Lyapunov exponent is well defined and does not depend on $x \in X$ on the set of μ-full measure. If we assume only that $\mu \in \mathcal{M}_T$ (but not ergodic) then we can only claim the existence of $\lambda(x)$.*

Proof. Define $\varphi(x) := \ln|D_xT(x)|$. Then $\lambda(x) := \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\varphi\circ T(x)$, which converges by Birkhoff ergodic theorem as $n \to \infty$ to $\mu(\varphi) = \mu(\ln|D_xT|)$ for μ-a.a. $x \in X$. If the measure μ is invariant but not ergodic then Birkhoff ergodic theorem implies only the convergence to a certain function of x. $\square$

Theorem 0.16 *(Oseledets) Assume that X is a compact manifold and $T : X \to X$ is a C^1 diffeomorphism preserving a Borel probability measure μ. Then there exists a T-invariant set $X' \subset X$, $\mu(X') = 1$, such that for every point $x \in X$ all Lyapunov exponents exist.*

Theorem 0.17 *(Upper Lyapunov Exponent). Under the above assumptions, there is a $\lambda_+ \in \mathbb{R}$ such that $\lim_{n\to+\infty}\frac{1}{n}\ln\|D_xT^n\| = \lambda_+$ for μ-a.e. $x \in X$. Here $\|A\| = \sup_{\|\vec{u}\|=1}\|A\vec{u}\|$ is the norm of the matrix A.*

Proof. By the chain rule, $\|D_xT^{n+m}\| \le \|D_xT^n\|\,\|D_{T^nx}T^m\|$. Defining $\varphi_n(x) = \ln\|D_xT^n\|$, for μ-a.e. $x \in X$ we get $\varphi_{n+m}(x) \le \varphi_n(x) + \varphi_m(T^nx)$. This condition is referred to as the *subadditivity* of the sequence of functions $\{\varphi_n\}$. Now the required result follows from the next general statement:

Theorem 0.18 *(Subadditive Ergodic Theorem) Let $T : X \to X$ be a transformation preserving an ergodic measure μ, and $\{\varphi_n\} \subset \mathbf{L}^1_\mu(X)$, $n \ge 1$ be a subadditive sequence of integrable functions on X for μ-a.e. $x \in X$. Then there is a $\lambda \in \mathbb{R}\cup\{-\infty\}$ such that $\lim_{n\to\infty}\frac{1}{n}\varphi_n(x) = \lambda$ for μ-a.e. $x \in X$.*

Remarks. In Theorem 0.17, $\varphi_1(x) = \ln\|D_xT\|$ is a continuous function on X, because T is C^1. Hence there is an upper bound $\varphi_{\max} := \sup_{x\in X}\varphi_1(x) < \infty$. By iterating the subadditivity condition we obtain for all $x \in X$ that $\frac{1}{n}\varphi_n(x) \le \frac{1}{n}\sum_{i=0}^{n-1}\varphi_1(T^ix) \le \varphi_{\max}$. Also, by the chain rule $(D_{T^nx}T^{-n})(D_xT^n) = I$ (the identity matrix), hence $1 = \|I\| \le \|D_xT^n\|\,\|D_{T^nx}T^{-n}\|$ therefore
$\frac{1}{n}\varphi_n(x) = \frac{1}{n}\ln\|D_xT^n\| \ge -\frac{1}{n}\ln\|D_{T^nx}T^{-n}\| \ge -\ln\max_{x\in X}\|D_xT^{-1}\| =: \varphi_{\min} > -\infty$. As a result, λ_+ in Theorem 0.17 is finite and $\lambda_+ \in [F_{\min}, F_{\max}]$.

Theorem 0.19 *(Lyapunov Exponents versus Volume) Let $J_n(x) = |\det D_xT^n|$ be the Jacobian of the map T^n at x (this is the factor by which T^n changes volume in an infinitesimal neighborhood of $x \in X$). Then $\lim_{n\to\infty}\frac{1}{n}\ln J_n(x) = \sum_{j=1}^d \lambda_j$ for almost every x i.e. the asymptotic rate of change of volume equals the sum of all Lyapunov exponents (counting multiplicity).*

Corollary 13.36. Assume that $\ln J_1(x) \in L^1_\mu(X)$. Then $\mu(\ln J_1) = \sum_{j=1}^d \lambda_j$ i.e. the average one-step rate of change of volume equals the sum of all Lyapunov exponents (counting multiplicity).

Proof. By the chain rule, $J_n(x) = J_1(x)\cdots J_1(T^ix)$. Taking the logarithm, dividing by n and letting $n \to \infty$ proves the theorem in view of the results above.

Corollary 13.37. Let the invariant measure μ be absolutely continuous with density $f(x)$ with respect to the Lebesgue measure (volume). Assume that $\ln f(x) \in L^1_\mu(X)$. Then $\sum_{j=1}^d \lambda_j = 0$ i.e. the sum of all Lyapunov exponents vanishes.

Proof. $J_1(x) = f(x)/f(Tx)$, hence $\mu(\ln J_1) = \mu(\ln f) - \mu(\ln f(Tx)) = 0$ by the invariance of μ. $\square$

Let us discuss the multidimensional case in some more detail.

Facts 13.1 (from Linear Algebra). Let A be a $d\times d$ matrix (with real entries), and assume that $\det A \ne 0$. For a nonzero vector $\vec{u} \in \mathbb{R}^d$, consider the sequence $\{A^n\vec{u}\}$, $n \in \mathbb{Z}$. We would like to see if A^n expands or contracts the vector $\vec{u}$ as $n \to \infty$ and $n \to -\infty$.

Let $\{\nu_j\}$, $1 \le j \le q$ $(q \le n)$, be all distinct roots (real and complex) of the characteristic polynomial $P_A(\nu) = \det(\nu I - A)$. We arrange them so that $\nu_1, \dots, \nu_r$ are all the distinct real roots and $\nu_{r+1}, \bar{\nu}_{r+1}, \dots, \nu_s, \bar{\nu}_s$

are all the distinct conjugate pairs of complex roots. Denote by m_j, $1 \le j \le s$, the respective multiplicities of the roots. Jordan theorem (the real canonical form) says that the roots ν_j (the eigenvalues of A) are associated to A-invariant generalized eigenspaces E_j, $1 \le j \le s$, whose respective dimensions equal $\dim E_j = m_j$ for $1 \le j \le r$ and $\dim E_j = 2m_j$ for $r < j \le s$ (note: in the latter case the space E_j is associated to the pair $\nu_j, \bar{\nu}_j$, rather than to a single root ν_j). Moreover, $\mathbb{R}^d = \oplus_{j=1}^s E_j$. We recall that $\det A \neq 0$, hence $\nu_j \neq 0$ for all j.

Now, if $\vec{u} \in E_j$ is an eigenvector (such a vector only exists for $j \le r$), then $A^n\vec{u} = \nu_j^n\vec{u}$, hence $\ln\|A^n\vec{u}\| = n\ln|\nu_j| + \ln\|\vec{u}\|$ for all $n \in \mathbb{Z}$. While this is not true for *any* vector $\vec{u} \in E_j$, it is still true that $\lim_{n\to\infty} \frac{1}{n}\ln\|A^n\vec{u}\| = \ln|\nu_j| \quad \forall \vec{u} \in E_j, \quad \vec{u} \neq \vec{0}$ This shows that when $|\nu_j| > 1$, the vector $A^n\vec{u}$ grows exponentially fast as $n \to \infty$ and shrinks exponentially fast as $n \to -\infty$. If $|\nu_j| < 1$, then it is vice versa. If $|\nu_j| = 1$, then there is no exponential growth or contraction, but there might be a slow (subexponential) growth or contraction of the vector $A^n\vec{u}$.

Exercise 13.2. A complete proof of the equation for $\ln|\nu_j|$ above is quite lengthy and tedious. But we can easily verify it in the simple case $\dim E_j = 2$. There are two principal subcases here. If ν_j is a real root of multiplicity 2, then A restricted to E_j is given by a Jordan matrix $J = \begin{pmatrix} \nu_j & 1 \\ 0 & \nu_j \end{pmatrix}$ in some basis. Verify that $J^n = \begin{pmatrix} \nu_j^n & n\nu_j^{n-1} \\ 0 & \nu_j^n \end{pmatrix}$ for all $n \in \mathbb{Z}$ and then derive the result for $\ln|\nu_j|$. If $\nu_j = a + bi$ is a complex root and $b \neq 0$, then the corresponding Jordan canonical block is $J = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Verify that $J^n = |\nu_j|^n \begin{pmatrix} \cos n\varphi & \sin n\varphi \\ -\sin n\varphi & \cos n\varphi \end{pmatrix}$ for some $\varphi \in [0, 2\pi)$ and all $n \in \mathbb{Z}$, and then derive the result for $\ln|\nu_j|$.

Definition 13.3. The numbers $\lambda_j = \ln|\nu_j|$ are called the *characteristic exponents* or the *Lyapunov exponents* of the matrix A, while ν_j are called Lyapunov *multiplicators*.

Note: some distinct eigenvalues $\nu_i \neq \nu_j$ may correspond to the same Lyapunov exponent, this happens whenever $|\nu_i| = |\nu_j|$. In this case each $0 \neq \vec{u} \in E_i \oplus E_j$ satisfies $\lim_{n\to\infty} \frac{1}{n}\ln\|A^n\vec{u}\| = \ln|\nu_j|$.

Proposition 13.4 (Lyapunov decomposition). Every nonsingular real matrix A has distinct Lyapunov exponents $\lambda_1 > \lambda_2 > \ldots > \lambda_m$ (with $m \le s$) and there is a decomposition $\mathbb{R}^d = \mathcal{E}_1 \oplus \cdots \oplus \mathcal{E}_m$ such that $A(\mathcal{E}_j) = \mathcal{E}_j$ and $\lim_{n\to\infty} \frac{1}{n}\ln\|A^n\vec{u}\| = \lambda_j \quad \forall \vec{u} \in \mathcal{E}_j, \quad \vec{u} \neq \vec{0}$. The number $\dim \mathcal{E}_j$ is called the *multiplicity* of the Lyapunov exponent λ_j. We also call $\mathcal{E}_j$ the *characteristic spaces* for A.

13.5 Remark. We have $\sum_j \lambda_j \cdot \dim \mathcal{E}_j = \ln|\det A|$ because $\det(A)$ equals the product of all the eigenvalues of A (counting multiplicity).

Examples 13.6. For the matrices $A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $A_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $A_3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ all Lyapunov exponents are zero. In the latter case eigenvalues are $\pm i$. Note that A_2^n does expand and contract vectors, but very slowly (at most linearly in n). On the other hand, for any matrix A defining a hyperbolic toral automorphism T_A (see 12.15) one Lyapunov exponent is positive, $\lambda_1 = \ln\nu^{-1} > 0$, and the other is negative, $\lambda_2 = \ln\nu < 0$. Note that $\lambda_1 + \lambda_2 = 0$, because $\det A = \pm 1$.

Definition 13.7. A matrix A and the corresponding linear map $A : \mathbb{R}^d \to \mathbb{R}^d$ are called *hyperbolic* if none of the eigenvalues of A (real or complex) lie on the unit circle $|z| = 1$.

Equivalently A is hyperbolic iff all the Lyapunov exponents of A are different from zero.

Definition 13.8. The A-invariant subspaces $\mathcal{E}^s = \oplus_{\lambda_j<0}\mathcal{E}_j$, $\mathcal{E}^u = \oplus_{\lambda_j>0}\mathcal{E}_j$ and $\mathcal{E}^c = \mathcal{E}_j|_{\lambda_j=0}$ are called *stable*, *unstable*, and *neutral* (or *central*) subspaces, respectively. Note that $\mathbb{R}^d = \mathcal{E}^s \oplus \mathcal{E}^u \oplus \mathcal{E}^c$. If the matrix A is hyperbolic, then $\mathcal{E}^c = \{\vec{0}\}$, hence $\mathcal{E}^c$ can be omitted from the above decomposition.

By definition we get $\limsup_{n\to\infty} \frac{1}{n}\ln\|A^n\vec{u}\| < 0 \quad \forall \vec{0} \neq \vec{u} \in \mathcal{E}^s$ and $\liminf_{n\to\infty} \frac{1}{n}\ln\|A^n\vec{u}\| > 0 \quad \forall \vec{0} \neq \vec{u} \in \mathcal{E}^u$.

Let the matrix A have at least one nonzero Lyapunov exponent $\lambda_i \neq 0$. Denote $\lambda = \min\{|\lambda_i| : \lambda_i \neq 0\}$ and $\nu = e^{-\lambda}$, Note that $\lambda > 0$ and $\nu < 1$.

Proposition 13.9. For any $\varepsilon > 0$ there is a $K > 0$ such that for all $n \ge 0$
$\|A^n\vec{u}\| \le K(\nu+\varepsilon)^n\|\vec{u}\|$ and $\|A^{-n}\vec{u}\| \ge K^{-1}(\nu+\varepsilon)^{-n}\|\vec{u}\| \quad \forall \vec{u} \in \mathcal{E}^s$ and
$\|A^n\vec{u}\| \ge K^{-1}(\nu+\varepsilon)^{-n}\|\vec{u}\|$ and $\|A^{-n}\vec{u}\| \le K(\nu+\varepsilon)^n\|\vec{u}\| \quad \forall \vec{u} \in \mathcal{E}^u$.

Proof. It is enough to prove the above bounds for unit vectors only. For every unit vector $\vec{u}$ $\exists K = K(\varepsilon, \vec{u})$ such that all these bounds hold, but K may depend on $\vec{u}$. Then we pick an orthonormal basis $e_1, \ldots, e_k$ in $\mathcal{E}^s$ (resp., $\mathcal{E}^u$), ensure the above bounds with the same constant $K(\varepsilon)$ for all the vectors $e_1, \ldots, e_k$. Then we use the triangle inequality to derive the proposition for all unit vectors $\vec{u}$ in $\mathcal{E}^s$ and $\mathcal{E}^u$. □

Thus, vectors $\vec{u} \in \mathcal{E}^u$ grow exponentially fast under A^n as $n \to \infty$ and shrink exponentially fast as $n \to -\infty$. For vectors $\vec{u} \in \mathcal{E}^s$, it is exactly the opposite. Now what happens to other vectors in $\vec{u} \in \mathbb{R}^d$?

COROLLARY 13.10. For any vector $\vec{u} \notin \mathcal{E}^u \cup \mathcal{E}^s$ and any $\varepsilon > 0$ there is a $K > 0$ such that $\|A^n \vec{u}\| \geq K(\nu + \varepsilon)^{-|n|} \|\vec{u}\|$ for all $n \in \mathbb{Z}$. That is, the vector $\vec{u}$ grows under A^n exponentially fast in both time directions: as $n \to +\infty$ and as $n \to -\infty$.

Next, we extend the above results to nonlinear maps.

DEFINITION 13.11. Let $U \subset \mathbb{R}^d$ be an open set and $T : U \to \mathbb{R}^d$ a smooth one-to-one map with a fixed point x, i.e. $T(x) = x$. Then the matrix $A = D_xT$ acts on tangent vectors $\vec{u} \in \mathcal{T}_x\mathbb{R}^d$, and the tangent space $\mathcal{T}_x\mathbb{R}^d$ can be naturally identified with $\mathbb{R}^d$. Note that $D_xT^n = (D_xT)^n = A^n$ by the chain rule.

Assume that $\det A \neq 0$. The Lyapunov exponents of the matrix A are called the *Lyapunov exponents* of the map T at the point x. The corresponding subspaces $\mathcal{E}^s, \mathcal{E}^u, \mathcal{E}^c \subset \mathcal{T}_x\mathbb{R}^d$ are called the *stable*, *unstable*, and *neutral* (or *central*) subspaces, respectively, for the map T at the point x.

Note: the subspaces $\mathcal{E}^s, \mathcal{E}^u, \mathcal{E}^c$ are invariant under D_xT but not necessarily under the map T itself. On the other hand, D_xT is a linear approximation to the map T at the point x. This allows us to obtain the following theorem, whose proof we omit.

THEOREM 13.12 (HADAMARD-PERRON). $\exists$ two submanifolds $W^s \subset U$ and $W^u \subset U$ such that
(a) $W^s \cap W^u = \{x\}$;
(b) the spaces $\mathcal{E}^s$ and $\mathcal{E}^u$ are tangent to W^s and W^u, respectively, at the point x;
(c) $T(W^s) \subset W^s$ and $T^{-1}(W^u) \subset W^u$;
(d) $T^n(y) \to x$ for every $y \in W^s$ and $T^{-n}(y) \to x$ for every $y \in W^u$, as $n \to \infty$.

We omit the proof, but remark that the manifold W^u is constructed as a limit of $(T^n\mathcal{E}^u) \cap V(x)$, as $n \to \infty$, where $V(x)$ is a sufficiently small neighborhood of x. The existence of this limit is proved by the contraction mapping principle. Similarly, W^s is constructed as a limit of $(T^{-n}\mathcal{E}^s) \cap V(x)$, as $n \to \infty$.

Since $A = D_xT$ is a linear part of the map T at x, it is easy to obtain the following corollary.

COROLLARY 13.13. For any $\varepsilon > 0$ there is a neighborhood $V(x)$ of the point x and a $K > 0$ such that $\mathrm{dist}(T^n y, x) \leq K(\nu+\varepsilon)^n \cdot \mathrm{dist}(y, x)$ for all $n \geq 0$, $y \in W^s \cap V(x)$ and $\mathrm{dist}(T^{-n} y, x) \leq K(\nu+\varepsilon)^n \cdot \mathrm{dist}(y, x)$ for all $y \in W^u \cap V(x)$.

DEFINITION 13.14. W^s and W^u are called the *stable* and *unstable manifolds*, respectively, for the map T at the point x. The map T is called *hyperbolic* at a fixed point x (and then x is called a *hyperbolic fixed point* for T) if $\dim \mathcal{E}^c = 0$. In this case $\dim W^s + \dim W^u = d$.

DEFINITION 13.15. A hyperbolic point x is called a *source* (a *repeller*) if $\dim \mathcal{E}^s = 0$ (hence $\mathcal{E}^u$ coincides with $\mathcal{T}_x\mathbb{R}^d$). It is called a *sink* (an *attractor*) if $\dim \mathcal{E}^u = 0$ (hence $\mathcal{E}^s$ coincides with $\mathcal{T}_x\mathbb{R}^d$). It is called a *saddle* (a truly hyperbolic point) if both $\mathcal{E}^s$ and $\mathcal{E}^u$ are not trivial.

REMARK 13.16. Let x be a saddle point and $y \neq x$ another point very close to x. If $y \in W^u$, then the trajectory $T^n y$ moves away from x exponentially fast for $n > 0$, at least until $T^n y$ leaves a certain neighborhood of x. If $y \in W^s$, then the trajectory $T^n y$ moves away from x exponentially fast for $n < 0$. Now, if $y \notin W^u \cup W^s$, then the trajectory $T^n y$ moves away from x exponentially fast for both $n > 0$ and $n < 0$. This fact is known as the *separation principle*: nearby trajectories tend to separate exponentially fast, either in the future or in the past or (in most cases) both.

All the above definitions and results extend to any diffeomorphism $T : U \to T(U) \subset M$ on an open subset $U \subset M$ of a Riemannian manifold M, rather than $U \subset \mathbb{R}^d$. A Riemannian structure in M is necessary for the norm $\|\cdot\|$ to be well defined on M. Henceforth we assume that T is defined on an open subset of a Riemannian manifold M. All the above definitions and results easily apply to a periodic point $x \in U$ rather than a fixed point. If $T^p(x) = x$, we can just consider T^p instead of T.

Next, we turn to nonperiodic points. This is the most interesting and important part of the story.

DEFINITION 13.18. Let the map T^n be differentiable at a point $x \in M$ for all $n \in \mathbb{Z}$. Assume that there are numbers $\lambda_1 > \ldots > \lambda_m$ and the tangent space $\mathcal{T}_xM$ is a direct sum of subspaces $\mathcal{E}_1 \oplus \cdots \oplus \mathcal{E}_m$ such that if $\vec{0} \neq \vec{u} \in \mathcal{E}_i$, then $\lim\limits_{n\to\infty} \frac{1}{n} \ln \|(D_xT^n)\, \vec{u}\| = \lambda_i$ The values λ_i are called the *Lyapunov exponents* of the map T at the point x. The number $\dim \mathcal{E}_i$ is called the *multiplicity* of the Lyapunov exponent λ_i. The spaces $\mathcal{E}_j$ are called *characteristic subspaces* at x.

The subspaces $\mathcal{E}^s = \oplus_{\lambda_i<0}\mathcal{E}_i$, $\mathcal{E}^u = \oplus_{\lambda_i>0}\mathcal{E}_i$, and $\mathcal{E}^c = \mathcal{E}_j|_{\lambda_j=0}$ are called *stable*, *unstable*, and *neutral* (or *central*) subspaces of $\mathcal{T}_xM$, respectively.

The existence of the Lyapunov exponents λ_i and the subspaces $\mathcal{E}_i$ is not guaranteed for any point $x \in U$, as an example will show soon. We say that a point x *has all Lyapunov exponents* if λ_i and $\mathcal{E}_i$ exist.

REMARK 13.19. If a point $x \in U$ has all Lyapunov exponents, then so do points $T^n(x)$ for all $n \in \mathbb{Z}$. Moreover, the points $T^n(x)$ have the same Lyapunov exponents (with the same multiplicity) as x does, and the characteristic subspaces are invariant along the trajectory of x: $(D_xT^n)(\mathcal{E}_i(x)) = \mathcal{E}_i(T^nx)$ for all $n \in \mathbb{Z}$ and each i. Observe that the Lyapunov exponents and their multiplicities are T-invariant functions.

EXAMPLE 13.20. Let $T_A : \mathrm{Tor}^2 \to \mathrm{Tor}^2$ be a hyperbolic toral automorphism. Then all Lyapunov exponents exist everywhere on Tor^2, and they are $\lambda_1 = \ln \nu^{-1} > 0$ and $\lambda_2 = \ln \nu < 0$. The corresponding subspaces $\mathcal{E}_1$ and $\mathcal{E}_2$ are parallel to the lines L_1 and L_2, respectively.

EXAMPLE 13.21. Let $T : X \to X$ be the baker's transformation of the unit square X. Let $X' \subset X$ be the set of points where T^n is differentiable for all $n \in \mathbb{Z}$. Then for every $x \in X'$ all Lyapunov exponents exist, and they are $\lambda_1 = \ln 2 > 0$ and $\lambda_2 = -\ln 2 < 0$. The corresponding subspaces $\mathcal{E}_1$ and $\mathcal{E}_2$ are parallel to the x axis and y axis, respectively. Note that $T^{\pm n}$ fails to be differentiable on the lines $x = k/2^n$ and $y = m/2^n$, with $k, m = 0, 1, \ldots, 2^n$. Hence $m(X') = 1$, i.e. all Lyapunov exponents exist almost everywhere.

Theorem 0.20 *(Shannon, McMillan, Breiman) Let $(T, X, \mathcal{B}, \mu)$ be ergodic and let ξ be an arbitrary finite measurable partition. Then $h_\mu(T, \xi) = -\lim\limits_{n\to\infty} \frac{1}{n} \ln \mu(\xi^{(n)}(x))$.*

Thus $h_\mu(T, \xi)$ describes the rate of decay of measures of elements of the preimages of ξ. It is very natural to make the next step in the construction of the entropy of a dynamical system, i.e. to maximize it over all invariant measures. The result $h(T) := \sup_{\mu\in\mathcal{M}_T} h_\mu(T)$ is called the *topological entropy.* The measure on which the supremum above is achieved (if it exists) is called the measure *of maximal entropy.*

In the case of a Markov shift with the transition matrix π (consisting of zeros and ones) the topological entropy is equal to the logarithm of the largest eigenvalue of the matrix π. The proof is based on the simple observation that $k(\xi^n) = (\vec{1}, \pi^n\vec{1})$ for the partition ξ consisting of the simplest cylinders.

Lecture 6. Phase space discretization in dynamical systems [16.12.19]

Definition 0.1 By an *ε-discretization*, $\varepsilon > 0$, of a compact set $X \subset \mathbb{R}^d$ we mean a choice of an ordered finite lattice (a collection of points) X_ε in the set X such that the distances between its neighboring elements do not exceed ε. By an *operator of ε-discretization* we mean a map $D_\varepsilon : X \to X_\varepsilon$ that associates to each point $x \in X$ its nearest point on the lattice $x_\varepsilon = D_\varepsilon(x) \in X_\varepsilon$ (if there are several such points we choose the point with the minimal index – the lattice X_ε is ordered). The value of the parameter $\varepsilon > 0$ is called the *diameter* of the ε-discretization or the magnitude of the corresponding perturbation.

Definition 0.2 By an *ε-discretized* (perturbed) system for the map T with the compact phase space X we mean a pair $(T_\varepsilon,\ X_\varepsilon)$, where $T_\varepsilon = D_\varepsilon \circ T$.

Note that contrary to usual perturbation schemes, the discretized (perturbed) system here is given not on the original phase space X but on the lattice X_ε. In the case of round-off errors in computer modeling we deal with a uniform ε-discretization. When the phase space X of the system is the unit d-dimensional cube (or torus) and $\varepsilon = 1/N$, $N \in \mathbb{Z}_+$, the lattice X_ε consists of all points $x \in X$ with rational coordinates with the denominator N. Actually only *binary dicsretizations* with $\varepsilon = 2^{-n}$ (i.e., $N = 2^n$) may occur in the computer arithmetic. Another interesting example is a random discretization, when all the points of the lattice are chosen randomly according to some distribution law (for example, to the Poisson distribution).

We start with properties of periodic trajectories under discretization and consider the simplest case in numerical simulations, which is to find the globally stable periodic orbit of the system. It seems that for sufficiently high precision the only visible effect of perturbations is deformation and shift of the periodic orbit. However, it turns out that here "period multiplication" may take place. This phenomenon consists in the emergence in a neighborhood of the original cycle of a new cycle, the period of which is a multiple of that of the parent cycle.

Theorem 0.21 *Suppose that the map T has a cycle of period n such that T is a local homeomorphism in some neighborhood of the cycle formed of disjoint neighborhoods of points of the cycle, and such that there is a cycle of period k of the ε-discretized system. Then the fraction k/n may take values 1 or 2 in the one-dimensional case and may be an arbitrary integer in the general multidimensional case.*

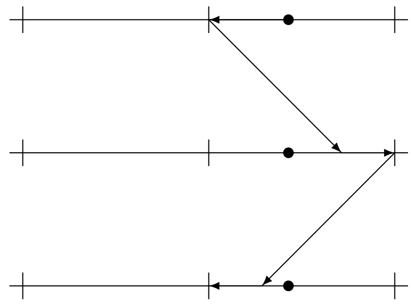

Рис. 2: Period doubling. By black dots we denote the points of the parent cycle (fixed point), and by vectors the cycle of the discretized system.

The general scheme of the period doubling of a stable fixed point is shown in Figure 2.

Proof. Let $\bar{x} = (x_1, \dots, x_n)$ be a cycle of the map T and let $U = \cup U_i$ be its neighborhood, consisting of disjoint neighborhoods of the points x_i. We assume also that in U there is a cycle $\bar{y} = (y_1, \dots, y_k)$ of the ε-discretized map (i.e., $T_\varepsilon^k y_j = y_j$, $i = 1, 2, \dots, k$).

We start with the one dimensional case ($d = 1$). The map T is a local homeomorphism of the collection of intervals U into itself. Thus, the function $T(x)$ is monotone in the set U, and, hence, the function $T_\varepsilon(x)$, restricted to the set $U \cap X_\varepsilon$, is also monotone as a superposition of two monotone functions T and T_ε. Denote by $\tilde{T}_\varepsilon$ the first return map, constructed for the dynamical system $(T_\varepsilon, X_\varepsilon)$ wrt to the set $U_1 \cap X_\varepsilon$.

Let $z_1 < z_2 < \dots < z_{m_1}$ be the points of the cycle $\bar{y}$ lying in the interval U_1. Since the map f is a local homeomorphism, for any i the number m_i of points of the cycle $\bar{y}$ belonging to U_i does not depend on i and is equal to an integer m (the period multiplier). There are three possibilities:

(a) $\tilde{T}_\varepsilon z_1 = z_1$. This means that $m = 1$ and $k = n$, i.e., there is no period multiplication.

(b) $\tilde{T}_\varepsilon z_1 \neq z_1$ and $(\tilde{T}_\varepsilon)^2 z_1 = z_1$. Then $m = 2$ and $k = 2n$, i.e., the discretized system has a cycle of double period. Notice that the necessary condition for the period doubling is monotone decreasing of $T^n(x)$ in the neighborhood of the cycle.

(c) $\tilde{T}_\varepsilon z_1 \neq z_1$ and $(\tilde{T}_\varepsilon)^2 z_1 \neq z_1$.

Let us show that the last situation cannot take place. At first suppose that the value m is even, i.e., $m = 2l$. Since $\tilde{T}_\varepsilon$ is monotone, then $(\tilde{T}_\varepsilon)^2$ is monotonically nondecreasing. Hence we have:

$$z_1 < (\tilde{T}_\varepsilon)^2 z_1 \leq (\tilde{T}_\varepsilon)^4 z_1 \leq \dots \leq (\tilde{T}_\varepsilon)^{2l} = z_1.$$

This is a contradiction, because the first inequality is strict. It remains to prove that the multiplier m cannot be odd, i.e., it cannot be represented in the form $m = 2l + 1$, $l \in \mathbb{Z}_+$. Here there are also two possibilities: the function $\tilde{T}_\varepsilon(x)$ is monotonically nondecreasing or monotonically nonincreasing. In the former case the orientations of the pairs (z_i, z_{i+1}) are the same for all $i = 1, 2, \dots, m-1$. Therefore such a sequence of points cannot form a cycle. In the latter case the orientations of the subsequent pairs are reversed, because $\tilde{T}_\varepsilon z_i = z_{i+1}$ by the construction of the first return map. Hence the pairs $(\tilde{T}_\varepsilon z_{2l+1}, \tilde{T}_\varepsilon z_{2l+2})$ and (z_1, z_2) must have a reverse orientation. We come to a contradiction again, because $\tilde{T}_\varepsilon z_{2l+1} = z_1$ and $z_1 < z_i$ for any $1 < i \leq m$.

Consider the multidimensional case. In this case the discretized system may have a cycle of arbitrary period integral multiple to the period of the parent cycle. Geometrically this result means that the orbit of the ε-discretized system curls around the original orbit. In the one-dimensional case only one sort of the rotation around the original orbit may take place – the rotation by the angle π, which we discussed above. We now show that in the two-dimensional case this phenomenon may occur with an *arbitrary* large multiplier. Consider a map of the two-dimensional plane $\mathbb{R}^2$, written in the polar coordinates (φ, r) as $(\varphi, r) \to (\varphi + \alpha, \lambda r)$. Here $\lambda \in (0, 1)$, and the irrational number $\alpha \in [0, 2\pi)$ are parameters. Each trajectory of this map has a spiral shape and tends to the zero point as time tends to infinity. Observe that if $(1 - \lambda) \ll 1$, then for a sufficiently fine uniform ε-discretization any trajectory of the discretized map, beginning at the point (φ, r), with $r \gg 1$ will end up into a nontrivial periodic trajectory around the origin rather than to hit the origin itself. The shape of this periodic trajectory is close to a circle, whose a radius is defined by the following inequality: $\varepsilon[\lambda r/\varepsilon] \geq r$, where by $[x]$ we denote the nearest integer point to x. Therefore, such a periodic trajectory corresponds to the rotation by an angle close to α around the origin and we may obtain arbitrary large periods by a suitable choice of the parameters A and α. □

Practically, there are only a few results showing that orbits of the discretized chaotic systems may be related to the original orbits. Among these results the main one is the so called shadowing property, which shows the connection of the orbit of a weakly perturbed system (with arbitrary type of perturbation) with the original orbit. Introduction of this property (in its original form) is due to Anosov , who showed that for a smooth hyperbolic system for each orbit of weakly perturbed system there exists a uniformly closed orbit of the original one. However, this shadowing orbit may be non-typical in the sense that the distribution of its points (which is the most important feature of a chaotic system) differs from the typical one, defined by the corresponding SBR measure. The following statement shows that these situations may often happen in the case of phase space discretizations.

Theorem 0.22 *Assume that the set of all preimages of a certain cycle is dense in the phase space. Then there is a sequence of discretizations with vanishing diameters such that each discretized system has only one cycle that coincides with the original one.*

Proof. Consider a sequence of discrete lattices $\{X^{(n)}\}$ such that the points of the n-th lattice coincide with the n-th preimage of the points of the cycle under consideration. Now the fact that the set of preimages is dense in the entire phase space yields the vanishing of the diameters of the lattices $\{X^{(n)}\}$. On the other hand, after n iterations of the discretized map each point of the lattice $\{X^{(n)}\}$ ends up in one of the points of our cycle. □

We can consider also a more general approach to this situation. For a given map T let us fix a family of lattices $\{X_\varepsilon\}$. Then for each cycle $\bar{x}_\varepsilon$ of the discretized map T_ε we associate the number $p(\bar{x}_\varepsilon)$, defined as the number of points in all preimages of this cycle (including the cycle itself) normalized by the total number of points in the lattice $\{X_\varepsilon\}$. This number shows how often one can observe the given cycle when initial points are chosen at random. To each discretization $\{X_\varepsilon\}$ we assign the cycle $\bar{x}_\varepsilon$ with the largest value of $p(\bar{x}_\varepsilon)$ and the probability measure μ_ε uniformly distributed along this cycle. We ask the question about the limit points (in the weak topology of measures) of these measures for the typical family of discretizations. It may be shown that for a sufficiently "good" map these limit points are invariant measures of the map. However, the family of invariant measures may be very rich (for example uniform measures on cycles and their closure). We believe that the most likely answer is the following: the limit point coincides with one of the measures: Sinai - Bowen - Ruelle measure or the measure of maximal entropy. Neither of these possibilities can be ruled out at the moment.

It may seem that the situation described in Theorem 0.22 is unrealistic. Actually, the condition about the density of preimages is typical for chaotic systems, and so the question is only about how often such particular discretizations occur. It is worth noting that for any binary discretization (i.e., computer discretization) of the dyadic map $x \to 2x \pmod 1$ the discretized system has only one cycle with period equal to 1. The same is true also for the map $x \to kx \pmod 1$ with integer k and the corresponding uniform k^{-n} discretizations. In the general case however, discretizations satisfying the assumptions of Theorem 0.22 are not uniform and this gives the hope to think that a resonance between the map and the discretization in numerical simulations may appear only by chance.

From another point of view Theorem 0.22 shows that some sort of a "localization" phenomenon takes place: trajectories that should normally be dense remain confined to a small number of points. Note that one of the most frequently used numerical methods to calculate the density of an invariant measure of a chaotic dynamical system is to compute the histogram of a sufficiently long numerical orbit of the system. Theorem 0.22 tells us that the error here can be very large even in the $\mathbf{L}^1$-norm. Indeed, in the numerical simulation (with the precision 2^{-N}) of the simplest chaotic map $x \to 2x \pmod 1$ (dyadic map) every point of the binary lattice will end up into the zero unstable fixed point of the original map. As a result this unstable fixed point becomes a globally stable fixed point of the discretized map. For example, on a VAX computer it could be verified that the orbit starting at the point 1/3 (which is a periodic point of period 2 in the true dyadic map) falls to 0 in 57 iterations, even with the double precision. One could think that, for computational purposes, there is no need to study asymptotic properties but it is enough to stop the count when the histogram is already stabilized and the accumulated error is not yet large. However, the analysis shows also that this method may not lead to the required result. In the case of the uniform ε-discretization the number of steps at which the count should be stopped is of order $-\ln(\varepsilon)$, while the guaranteed precision is of order $(-\ln(\varepsilon))^{-1/2}$.

Statistical probability for discretized systems. Now let us try to answer the question about how "typical" is the behavior of the discretized systems. First let us consider the existence of a cycle of an ε-discretized system for sufficiently small $\varepsilon > 0$ in a small neighborhood of an unstable cycle of the unperturbed system. For various values of ε a cycle of an ε-discretized system can appear and disappear with arbitrary small changes in ε and a given choice of X_ε. Therefore, to answer the above question it is necessary to make some additional assumptions about the structure of discretizations. We restrict ourselves to the most interesting case for applications, the case of uniform discretizations of the d-dimensional unit cube X. In this case, ε takes values of the form $\varepsilon = 1/n$, $n \in \mathbb{Z}_+$. Even in this case, the presence or absence of a cycle for a uniform $1/n$-discretized system depends on n in a very irregular way. The statistical approach seems to be the most natural one for the description of such a situation.

Definition 0.3 By the *statistical probability* of some event (with respect to a given family of space discretizations) we mean the fraction of the discretizations for which this event takes place.

Thus, the statistical probability $p(x_1, x_2, \cdots, x_n)$ of the *stabilization* of an unstable cycle $(x_1, x_2, \cdots, x_n)$ of a map f is the density of the set of those natural numbers N for which $(D_{1/N}x_1, D_{1/N}x_2, \cdots, D_{1/N}x_n)$ is a cycle of a uniformly $1/N$-discretized system.

The meaning of this definition is that the existence, in the case of uniform discretizations, of a cycle of a discretized system in a small neighborhood of the original cycle results in the possibility to observe an unstable trajectory in the computer simulation. It is therefore natural to call this situation the "stabilization of an unstable cycle". Numerical experiments show that in the modelling of chaotic dynamical systems, with all cycles unstable, some of the cycles are observed much more frequently than others. The following statement provides an explanation for this phenomenon.

Theorem 0.23 *Let $\bar{x} = (x_1, x_2, \ldots, x_n)$ be an unstable cycle of the map T, that is continuously differentiable in a neighborhood of the cycle, and let the coordinates of its points $x_1, x_2, \ldots, x_n$ be rationally independent. Then the statistical probability of the stabilization $p(\bar{x})$ is well defined and depends only on the derivatives of the map f at the points of the cycle: it is equal to the volume of the dn-dimensional polyhedron $\Omega = \Omega(f, \bar{x})$ defined by the following system of semi-linear inequalities:*

$$\begin{aligned} |T'(x_1)z_1 - z_2| &\le 1/2 \\ |T'(x_2)z_2 - z_3| &\le 1/2 \\ &\cdots \\ |T'(x_n)z_n - z_1| &\le 1/2 \\ |z_i| &\le 1/2, \quad i = 1, 2, \ldots, n, \end{aligned}$$

where $T'(x)$ is the matrix of partial derivatives of the map f at the point x, and $|z| = \max_i\{|z^{(i)}|\}$ for a vector $z \in \mathbb{R}^d$.

Proof. Clearly the density of the set of natural numbers depends only on the distribution of large elements of the set, corresponding in our case to sufficiently fine discretizations. Therefore, working only in small neighborhoods of the points of the cycle, we may assume that the map T is linear there. Let us fix a natural number $k \gg 1$. The collection of points $\bar{x}_{1/k} = (D_{1/k}x_1, D_{1/k}x_2, \ldots, D_{1/k}x_n)$ is a cycle of the discretized map $T_{1/k}$ if and only if the following system of inequalities holds:

$$\begin{aligned} |x_2 + T'(x_1)(D_{1/k}x_1 - x_1) - D_{1/k}x_2| &< \frac{1}{2k} \\ |x_3 + T'(x_2)(D_{1/k}x_2 - x_2) - D_{1/k}x_3| &< \frac{1}{2k} \\ &\cdots \\ |x_1 + T'(x_n)(D_{1/k}x_n - x_n) - D_{1/k}x_1| &< \frac{1}{2k}. \end{aligned}$$

Indeed, if these inequalities hold, then $T_{1/k}(D_{1/k}x_i) = D_{1/k}x_{i+1}$ for all i. In addition, if one of the inequalities fails, say the i-th inequality, then $T_{1/k}(D_{1/k}x_i) \neq D_{1/k}x_i$. Recall now that $D_{1/k}x_i$ is the best rational approximation of the vector x_i with rational coordinates with the common denominator k. Therefore, multiplying both sides of these inequalities by k and denoting by $\{v\}$ the difference between a vector $v \in R^d$ and the nearest vector with integer coordinates, we have:

$$\begin{aligned}
|T'(x_1)\{kx_1\} - \{kx_2\}| &< 1/2 \\
|T'(x_2)\{kx_2\} - \{kx_3\}| &< 1/2 \\
&\cdots \\
|T'(x_n)\{kx_n\} - \{kx_1\}| &< 1/2.
\end{aligned}$$

Now, let us consider the rotation T_α on the dn-dimensional unit torus $\mathrm{Tor} = [-0.5, 0.5]^{nd}$ by a (nd)-dimensional angle $\alpha = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{nd}$. Then the trajectory of this map starting at the point $\alpha \in \mathrm{Tor}$ ends up into the domain Ω, defined in the theorem if and only if there is a stabilization effect. Thus, $p(\bar{x})$ coincides with the fraction of the time this trajectory spends in Ω and consequently coincides with the F_α-invariant measure of this domain. However, Lebesgue measure on the torus Tor is the only one invariant measure of the irrational rotation, which yields the desired statement. □

To clarify this construction, consider a related but significantly simpler problem. Let x be an arbitrary irrational number in the unit interval and let us compute how often the precision of a rational approximation of this number with a denominator n is not worse than some fixed constant $\varepsilon \ll 1$ divided by n. This amounts to calculating the density $p(x)$ of those natural numbers n for which the inequality $\left|x - \frac{q}{n}\right| \le \frac{\varepsilon}{n}$ holds for some natural $q = q(n)$. Multiplying both sides of this inequality by n, we see that it is equivalent to $|nx \,(\mathrm{mod}\ 1)\,| \le \varepsilon$. Now, considering the two intervals on the unit circle defined by this inequality and using the one-dimensional rotation by the angle x to calculate $p(x)$, we come to the same situation as in the proof of Theorem 0.23. In this case we can calculate this density in the explicit form: $p(x) = 2\varepsilon$. Note that the considered situation coincides with the stabilization of a fixed point x of a one-dimensional map, while the derivative at this point is equal to $1/(2\varepsilon)$.

Let us discuss briefly a question about the practical applicability of this statement: which magnitude of k is needed to reach the value of $p(\alpha)$ (defined by Theorem 0.23 with a sufficiently good precision? From the proof above it follows that the answer to the question depends on the rate of convergence of the fraction of the points of a typical trajectory of the torus rotation which happen to be in the specified region. In the "good" case, this convergence is of order $\ln(k)/k$, which demonstrates that it is quite fast.

Note that the statistical probability of the stabilization obtained in Theorem 0.23 does not depend on the coordinates of the points of the cycle. It should be pointed out that such a uniform estimate holds only for "typical" cycles. If the condition of rational independence of the coordinates (or irrationality of the number x above) does not hold, there are "exceptional" cycles (and even maps for which all the cycles are "exceptional") whose stabilization probability depends on the coordinates of its points. Among the "exceptional" cycles there are cycles with abnormally high probability of stabilization (for example, equal to 1) and also cycles for which this probability is abnormally small (in comparison with the "typical" situation described in the theorem).

The method used in the proof of Theorem 0.23 enables us to study even finer properties of the discretized systems. For example, one can calculate by this method the statistical probabilities of events that can appear in the "period multiplication" phenomenon.

A closely related construction can be carried out also if we want to consider not all uniform discretizations but only binary ones corresponding to the case $\varepsilon = 2^{-n}$. In this case using the idea similar to the one in the proof of Theorem 0.23 we obtain the multidimensional dyadic map T (the direct product of (dn) dyadic maps) rather than the rotation of the dn-dimensional unit torus by the angle x. In distinction to the irrational rotation the map T has a rich family of invariant measures and therefore the answer depends on the invariant measure represented by the trajectory of this system, starting at the point $(x_1, x_2, \ldots, x_n)$. However, for almost all initial conditions this measure is still equal to the Lebesgue measure on the dn-dimensional unit torus.

Theorem 0.24 *Suppose that the map T is continuously differentiable in a neighborhood of its cycle $\bar{x} = (x_1, x_2, \ldots, x_n)$. Then there is a map $\hat{T}$ arbitrary C^1-close to f with a cycle $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$ close* *to initial one such that the statistical probability of the stabilization of the cycle $\hat{x}$ in the case of binary discretizations of the map $\hat{T}$ is equal to* 1.

Proof. Notice that the preimages of the zero fixed point of the dyadic map (which is the replacement of the rotation map for binary discretizations) are dense on the interval $[0, 1]$. Therefore, we can choose a map $\hat{f}$ such that the cycle $\hat{x}$ on one hand, lies in this set of preimages, and on the other hand, is arbitrary close to the initial one. However the trajectory of the dyadic map, starting at the point $\hat{x}$ is far from being typical, because it runs into the unstable zero fixed point and stays there. Therefore the statistical probability of the stabilization of the cycle $\hat{x}$ is equal to 1. □

For systems on the unit torus one can consider also another approach to study how typical the properties of discretized systems are. For a torus (in contrast with a cube) it is natural to fix a discretization up to a shift of the origin and a rotation around it. Therefore, in this case the parameter is not discrete and by the statistical probability of the stabilization of the cycle $(x_1, x_2, \cdots, x_n)$ we mean the value $p(x_1, x_2, \cdots, x_n; \varepsilon)$, which is equal to the Lebesgue measure of the set of shifts (≤ 1) of the origin and rotations (by angles from 0 to 2π) around it, divided by 2π for normalization, for which the stabilization effect takes place. Here ε is the diameter of the discretization. This definition gives a possibility obtain the dependence of the probability of the stabilization on the fixed discretization diameter ε. Let us consider the map $\varphi : \mathbb{R}^1_+ \to \mathbb{R}^1_+$, defined by the following relation: $\varphi(x) = x/(1 + x)$. The following statement demonstrates a variant of the universality of the dependence of this probability on the parameter ε.

Theorem 0.25 *Suppose that the map T is continuously differentiable in a neighborhood of its cycle $(x_1, x_2, \cdots, x_n)$. Then for any sufficiently small $\varepsilon > 0$ the sequence of quantities $\{p(x_1, x_2, \cdots, x_n; \varphi^k(\varepsilon))\}_{k=1}^{\infty}$ is almost periodic. Here $\varphi^k(x)$ is the value of the function φ applied k times recursively at the point x.*

Case of neutral periodic trajectories. Now let us study the influence of the uniform discretizations on systems whose trajectories are neither stable nor unstable, or such that they have a local invariant component of this type. The main example here is the rotation $T^{(\alpha)}$ of the d-dimensional unit torus $X :=$Tor by an angle $\alpha \in \mathbb{R}^d$. In this case the analysis of individual trajectories does not make sense, and we study the influence of discretizations to certain global properties of the system, such as the ergodicity property. In the case of a discretized system (i.e., a system with a finite phase space), ergodicity means that the invariant set consists of only one cycle. Therefore any trajectory of this map ends up into this cycle after a finite number of iterations. Since the discretized rotation preserves distances between the points of the uniform lattice, in the ergodic case there is only one periodic trajectory filling in the entire phase space (i.e., its period is equal to the number of points in X_ε). The ergodicity property of the discretized rotation does not depend on a shift of the origin, or a rotation of our lattice around it. Thus we come to the following definition.

Definition 0.4 By the *statistical probability of ergodicity* $p(\alpha)$ we mean the density of the set of those natural numbers N for which uniformly $1/N$-discretized systems $(T^{(\alpha)}_{1/N}, X_{1/N})$ are ergodic.

Theorem 0.26 *Let the coordinates of the vector $\alpha \in \mathbb{R}^d$ be rationally independent. Then $p(\alpha)$ is well defined and does not depend on α. In the one-dimensional case it satisfies the inequalities* $0.3889 < p(\alpha) < 0.5678$*, while in the multidimensional case $p(\alpha)$ exponentially goes to zero with the dimension.*

The proof of this result is rather nontrivial and is based on the analysis of the distribution of prime numbers. Let us discuss briefly the multidimensional case. The dynamical system $(T^{(\alpha)}_{1/N}, X_{1/N})$ is ergodic if and only if $\forall i \neq j$ the pairs of numbers N and $D_1(N\alpha_i)$ and $D_1(N\alpha_i)$ and $D_1(N\alpha_j)$ are coprime.
By the first part of the proof, the density of all integers N satisfying the first condition is well defined and is independent of α_i. It remains to prove that the density of all integers N satisfying the second condition is well defined. This can be done in the same manner as in the first part of the proof. Therefore in the d-dimensional case $p(\alpha) < 0.5678^d$. □

This result shows the qualitative distinction between the influence of perturbations arising from space discretizations and those due to smooth perturbations, since in the latter case the well known KAM theory says that it is ergodicity which is "typical". However, it would be of interest to further specify the link between the chaotic appearance and disappearance of periodic trajectories in the $1/N$-discretized systems, for $N \to \infty$, and the so called "subfurcation" phenomenon arising in the case of smooth perturbations.

The one-dimensional estimates for the upper and lower bounds for $p(\alpha)$ in Theorem 0.26 are very raw and we actually use them only to prove the multidimensional case. D. Nucinkis proposed a heuristic argument to calculate $p(\alpha)$, based on the well known fact that the density of the set of integer lattice points with coprime coordinates is equal to $6/\pi^2$. The point, we are looking for the set of integers $\{n\}$ such that n and $D_1(n\alpha)$ are coprime. Thus if n and $D_1(n\alpha)$ would be random this would give the desired answer. However, $6/\pi^2 \approx 0.6085 > 0.5678$, which shows that a more delicate approach is needed.

We note that if the coordinates of the angle α are rationally dependent (i.e., the original system is nonergodic), then the statistical probability of ergodicity is also well defined but can take more or less arbitrary values (for instance, $p(0) = 0$, $p(1/3) = 2/3$).

It is surprising that in the case of the binary discretizations the result differs significantly.

Theorem 0.27 *For almost all vectors $\alpha \in \mathbb{R}^d$ the statistical probability of ergodicity with respect to the binary discretizations $p_2(\alpha)$ is well defined and identically equal to $(1/2)^d$.*

Proof. Let us start from the one-dimensional version of this statement. Fix a natural number n. Then the 2^{-n}-discretized system $(T^{(\alpha)}_{2^{-n}}, X_{2^{-n}})$ is ergodic if and only if $|2^n\alpha - (2k+1)| < 1/2$, which is equivalent to $1/4 < 2^{n-1}\alpha \pmod 1 < 3/4$. Just as we did in the proof of Theorem 0.23, we consider now the dyadic map $x \to 2x \pmod 1$ and then, for a typical initial point $x_0 = \alpha$ we obtain that $p_2(\alpha)$ is equal to the fraction of time that the trajectory beginning at the point x_0 spends in the segment $(1/4, 3/4)$. The latter is equal to $1/2$, because the Lebesgue measure is invariant for the map considered and the initial point is assumed to be typical.

To prove the multidimensional case ($d > 1$) it is enough to notice that the only difference here is that one needs to use systems of inequalities and to consider the direct product of dyadic maps. □

It is of interest that for non-typical angles α, the statistical probability of ergodicity can be far from the value in Theorem 0.27 (for instance, $p(1/2^k) = 0$ for any $k \in \mathbb{Z}_+$).

Another problem arises in the analysis of a rotation $T^{(\alpha)}$ of the two-dimensional plane $\mathbb{R}^2$ by a fixed angle α around the origin. It seems that the situation here is more or less the same as in the previous example, but there is a significant difference between them. Indeed, although the map $T^{(\alpha)}$ preserves distances, the discretized map does not preserve them (in contrast with the circle rotation). Moreover, if we consider a piecewise linear curve, consisting of consecutive points of a numerical trajectory connected by straight lines, then two such different curves may intersect each other. Without discretization, the trajectories of the map $T^{(\alpha)}$ lie on concentric circles. With round-off, it is not known when the trajectories go to the origin and when they go to infinity (neither can be ruled out at the moment). Numerical simulations show the evidence of a picture similar to the situation in the KAM theory, that is, we can see some quasi-circles (invariant tori) with gaps between them. Notice that the distribution of points on these tori may not be ergodic. In spite of this numerical evidence there is no analytical proof even of the eventual periodicity of numerical trajectories in this case.

To show that similar neutral systems may behave under the discretization as dissipative ones, let us consider a generalization of the previous model.

Definition 0.5 A map $T : \mathbb{R}^2 \to \mathbb{R}^2$ is called the *generalized rotation* around the fixed point $x_0 \in \mathbb{R}^2$ if $T(x_0) = x_0$, and for almost all $x \neq x_0$ the trajectory $\{T^n(x)\}_{n \geq 0}$ starting at this point fills in densely a closed curve, homeomorphic to a circle.

An example of such a map is a usual rotation around the origin by an irrational angle. Another example gives any twist map, i.e., a map represented in the polar coordinates (φ, r) as $(\varphi, r) \to (\varphi + \Phi(r), r)$ with $|d\Phi(r)/dr| > 0$. The latter map is Hamiltonian and, as it follows from Kolmogorov - Arnold - Moser (KAM) theory , for sufficiently "good" rotation angles and "good" sufficiently small smooth perturbations, the perturbed map has invariant curves (homeomorphic to a circle) around the origin and each trajectory of the perturbed system is bounded by these invariant curves. We have noted before, that in the case of perturbations, arising in the computer simulations the behavior of the perturbed systems looks similar. However in the case of the generalized rotation this is no longer true.

Let us fix in the two-dimensional plane $\mathbb{R}^2$ a system of orthogonal coordinates (x, y) and a family of "triangular" curves $\Gamma_t, t > 0$, the i-th side of which satisfies the equality $a_i x + b_i t = y, \quad i = 1, 2, 3$. We now define a map $T_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ as a generalized rotation around the origin along the curves Γ_t with the unit rate, i.e., as a shift of the current point by means of one of the vectors: $\alpha^{(i)} = \left(\frac{1}{\sqrt{1+a_i^2}}, \frac{a_i}{\sqrt{1+a_i^2}} \right), \quad i = 1, 2, 3,$ where i corresponds to the side of the "triangular" curve Γ_t on which the current point lies. If during the shift we are going through the corner of the "triangle", then we continue the shift along the next side. The collection of the vectors $\alpha^{(i)}$ we denote by α.

We want to study the event that any point of the $1/n$-discretized plane, outside the ball of radius $10/n$ centered in the origin, goes to infinity under the action of the $1/n$-discretized generalized rotation T_α. To do it consider a coarser situation, when after each iteration of the discretized map, we go to the "triangle" with the larger value of the parameter t. We call the latter event the *coarse destabilization.*

Theorem 0.28 *Let the vectors $\alpha^{(i)}$ and the unit vector be jointly rationally independent. Then the statistical probability of the coarse destabilization is equal to* $1/8$.

Proof. For a fixed vector $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$ with rationally independent coordinates, let us define a map $h_\beta(x) = x + \beta$, $x \in \mathbb{R}^2$ and a function $H_\beta(x)$, which is equal to the distance from the point x to the line $y = \frac{\beta_2}{\beta_1} x$. Assume first that $\beta_1 \beta_2 > 0$. Let us consider the unit square with sides parallel to the coordinate axes and the left lower corner at the origin. We divide this square into four equal squares and consider a set A_+, consisting of three parts: the left upper square, the part of the left lower square lying below the line $y = \frac{\beta_2}{\beta_1} x$, and the part of the right upper square lying below the line $y = 1 + \frac{\beta_2}{\beta_1}(x - 1)$. Then the density $p(\beta)$ of natural numbers n such that the $1/n$-discretized map $h_{\beta,1/n}$ monotonically increases the distance function H_β (i.e., $H_\beta(x) \le H_\beta(h_{\beta,1/n}(x))$ for any x), is equal to the Lebesgue measure of the set A_+. The explanation is that the $1/n$-discretized map h_β increases the distance function if and only if the n-th iterate of the map $x \to x + \beta \pmod 1$, applied to the point $\beta \in \mathbb{R}^2$, lies in the set A_+. Therefore, $p(\beta)$ is equal to the occurrence time of this set for the above rotation map, which is equal to the Lebesgue measure of this set, i.e., to $1/2$.

The situation when $\beta_1 \beta_2 < 0$ is considered in the same way. Now, applying this result coupled with the rational independency of $\alpha^{(i)}$, we obtain the statement of the theorem, because the statistical probability of the considered event is equal to the product of probabilities of events, relating to each $\alpha^{(i)}$. $\square$

Although this example shows that in the general case, trajectories of generalized rotations in the computer modeling are not bounded, all numerical simulations with Hamiltonian generalized rotations (for example, with usual rotations by irrational angles) give another result. Therefore we can formulate a conjecture.

Proposition 0.15 *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a Hamiltonian generalized rotation (and thus preserving the Lebesgue measure). Then for any n each trajectory of the $1/n$-discretized map is bounded.*

If this statement holds, then since each computer trajectory is bounded and lies on the discrete lattice, it ends up eventually into a finite cycle. As we mentioned above, the invariant cycles look like invariant tori in the KAM theory (apart from the distribution of points on such a torus).

A KAM type theorem for systems with round-off errors. Now we construct a generalization of the KAM theory for twist maps with round-off errors and discuss obstacles to apply this idea for the rotational map.

Definition 0.6 A map T of the annulus $\mathcal{A}(r_-, r_+) := \{(\varphi, r) \in \mathbb{R}^2 : \ 0 \le r_- \le r \le r_+ < \infty\}$ into $\mathbb{R}^2$ is said to be *twist* if in polar coordinates $(0 \le \varphi < 1, \ 0 \le r < \infty)$ it can be written as $(\varphi, r) \to (\varphi + \Phi(r) \pmod 1, \ r)$, where the function $\Phi(r)$ is infinitely differentiable and $|d\Phi(r)/dr| > 0$.

Theorem 0.29 *For each $\varepsilon > 0$ any trajectory of the ε-discretized twist map is eventually periodic. Moreover, for any constants $0 < \hat{r}_- < r_- < r_+ < \hat{r}_+ < \infty$ there exists $\varepsilon_0 > 0$ such that for each $0 < \varepsilon < \varepsilon_0$ if $r_- < r < r_+$ then $\hat{r}_- < (D_\varepsilon \circ T)^n(\varphi, r) < \hat{r}_+$ for any $n \in \mathbb{Z}^+$ and any angles $\varphi \in [0, 1)$.*