Markov Processes Relat. Fields **21**, 381–412 (2015)

Markov Processes and Related Fields

Liouville Ergodicity of Linear Multi-Particle Hmiltonian System with One Marked Particle Velocity Flips

A.A. Lykov and V.A. Malyshev

Faculty of Mechanics and Mathematics, Lomonosov Moscow State University. Vorobievy Gory, Main Building, 119991, Moscow Russia. E-mail: 2malyshev@mail.ru

Received December 10, 2014

Abstract. We consider multi-particle systems with linear deterministic hamiltonian dynamics. Besides Liouville measure it has continuum of invariant tori and thus continuum of invariant measures. But if one specified particle is subjected to a simple linear deterministic transformation (velocity flip) in random time moments, we prove convergence to Liouville measure for any initial state. For the proof it appeared necessary to study non-linear transformations on the energy surface.

KEYWORDS: non-equilibrium statistical mechanics, convergence to equilibrium, Liouville measure

AMS SUBJECT CLASSIFICATION: 82C05

Contents

1. Introduction

Ergodicity problem for hamiltonian multi-particle systems produced many deep results. First of all, many examples of non-ergodic systems appeared — linear, non-linear with additional integrals and close to them (KAM theory). One could expect then that for generic hamiltonians one also has non-ergodic behaviour. However, as far as we know, this is still an open difficult problem with many partial results, see [19, 20], and, after the century of existence of the ergodicity hypothesis it is reasonable to look for simpler alternative approaches to it.

Namely, one could assume that any physical system has always some contact with external world. Such contact can be of quite various extent: 1) all particles can have contact with external world and stochastic behaviour (for example, with dynamics of Glauber type), 2) only particles on the boundary, etc. But then it is quite natural to ask — what is the minimal contact which definitely provides ergodic behavior. Possible reformulation of the ergodicity hypothesis could be the following: for generic system even the minimalistic contact produces ergodic behaviour.

Here we consider an example of such minimalistic contact which consists, first of all, in that we allow some contact with external world for only one (marked) particle. In our earlier papers [9–11] this particle was subjected to some random force, that guaranteed convergence to Gibbs equilibrium. Here we assume even less randomness. Namely, the marked particle is subjected to a simple deterministic transformation (velocity flip, very popular in other problems [3–8]) but in discrete random time moments

$$0 < t_1 < \ldots < t_m < \ldots \tag{1.1}$$

Liouville measure is evidently invariant w.r.t. such dynamics on the energy surface. We prove then that ergodicity in the stronger form holds — for any initial state we have convergence to Liouville measure on the energy surface.

It is very interesting that it works even for the systems having the worst possible non-ergodic behaviour — linear systems. The only price we pay is that ergodicity holds not for any linear system but for almost any — this is purely algebraic phenomenon which was discussed in our earlier papers and cannot be avoided. It seems reasonable that the same result holds for non-linear systems as well (following the common belief that the latter have better mixing properties than linear systems). The convergence results for anharmonic chains, see [3, 21, 22] and reviews [23, 24], support this belief. However, the principal question (which we answered for linear hamiltonians) — what class of non-linear hamiltonians could be called generic and be ergodic — is largely open.

The paper is naturally subdivided in two parts. The first part uses no probability at all but only elaborates non-linear analysis to prove that the trajectory visits all invariant tori and even any point — we had to use coordinates on the energy surface where velocity flips are strongly non-linear. Second part, on the contrary, essentially uses non-trivial parts of Markov processes theory on continuous state space.

Now we come to rigorous definitions.

2. Model and main results

Hamiltonian dynamics We consider the linear space

$$L=\mathbb{R}^{2N}=\{\psi=\left(\begin{array}{c} q \\ p \end{array}\right):\ q=(q_1,\ldots,q_N)^T,\ p=(p_1,\ldots,p_N)^T,\ q_i,p_i\in\mathbb{R}\}.$$

where T denotes transposition (thus ψ is a column-vector). It can be presented as the direct sum $L=l_N^{(q)}\oplus l_N^{(p)}$ of two orthogonal (coordinate and momenta) spaces of dimension N with the standard scalar product in $\mathbb{R}^{2N}$

$$(\psi,\psi')_2=(q,q')_2+(p,p')_2=\sum_{i=1}^{N}(q_iq_i'+p_ip_i').$$

We consider quadratic hamiltonian

$$H(\psi)=\sum_{k=1}^{N}\frac{p_k^2}{2}+U(q),U(q)=\frac{1}{2}(q,Vq)_2 \tag{2.1}$$

where the matrix $V>0$ acting in $\mathbb{R}^N$ is assumed to be real and positive definite (thus the particles cannot escape to infinity). This defines hamiltonian system of linear ODE with $k=1,\ldots,N$

$$\dot{q}_k=p_k,\dot{p}_k=-\sum_{l=1}^{N}V_{kl}q_l. \tag{2.2}$$

For any $h > 0$ define the constant energy surface

$$\mathcal{M}_h = \{\psi \in L : H(\psi) = h\}.$$

Then $\mathcal{M}_h$ is a smooth manifold (ellipsoid) in L of codimension 1.

Allowed hamiltonians Define the mixing subspace

$$L_- = L_-(V) = \left\{ \begin{pmatrix} q \\ p \end{pmatrix} \in L : q, p \in l_V \right\}$$

where $l_V = l_{V,1}$ is the subspace of $\mathbb{R}^N$, generated by the vectors $V^k e_1,\ k = 0, 1, 2 \ldots$, where $e_1, \ldots, e_N$ is the standard basis in $\mathbb{R}^N$.

Let $\mathbf{V}$ be the set of all positive-definite $(N \times N)$-matrices, and let $\mathbf{V}^+ \subset \mathbf{V}$ be the subset of matrices for which

$$L_-(V) = L. \tag{2.3}$$

Denote by $\mathbf{V}_{ind}$ the set $V \in \mathbf{V}$ such that the eigenvalues of the matrices, denoted by $\omega_1^2, \ldots, \omega_N^2$, are independent over the field of rational numbers.

Lemma 2.1. *The set $\mathbf{V}^+$ is open and everywhere dense (assuming topology of $R^{N(N+1)/2}$) in $\mathbf{V}$, and the set $\mathbf{V}^+ \cap \mathbf{V}_{ind}$ is dense both in $\mathbf{V}^+$and in $\mathbf{V}$.*

See more in Section 5.3.

Piecewise deterministic process Assume that at time moments (1.1) the following deterministic transformation $I : L \to L$ occurs: all q_k, p_k are left unchanged, except for p_1, the sign of which becomes inverted,

$$p_1(t_m - 0) \to p_1(t_m) = -p_1(t_m - 0), m \geq 1.$$

For example, one can consider L as the phase space for N identical point particles in $\mathbb{R}$, with mass $m = 1$, and real numbers q_i, p_i are their coordinates and velocities (momenta). Then this transformation can be interpreted as the elastic collision of the particle 1 with a wall. Alternatively, taking dN instead of N, one can imagine N particles in R^d where only one velocity component of particle 1 is flipped. Reflections w.r.t. any hyperplane in R^d could be considered quite similarly.

In-between these moments the system evolves via hamiltonian dynamics (2.2).

With $(2N \times 2N)$-matrix

$$A = \begin{pmatrix} 0 & E \\ -V & 0 \end{pmatrix}$$

the system (2.2) can be rewritten as

$$\dot{\psi} = A\psi. \tag{2.4}$$

and the solution $\psi(t)$ of (2.4) with initial vector $\psi(0)$ will be

$$\psi(t) = e^{tA}\psi(0).$$

For given sequence (1.1) the dynamics of our process is defined for $t_m \leq t < t_{m+1}$ as

$$\psi(t) = e^{A(t-t_m)} I e^{A\tau_m} I e^{A\tau_{m-1}} I \dots I e^{A\tau_2} I e^{A\tau_1} \psi(0)$$

where

$$\tau_1 = t_1, \tau_2 = t_2 - t_1, \dots, \tau_m = t_m - t_{m-1}, \dots$$

For any $t \geqslant 0$ define linear maps $L \to L$

$$J(t)\psi = I e^{tA}\psi, \ \psi \in L.$$

It is clear that $\mathcal{M}_h$ is invariant w.r.t. $J(t)$ for any $h > 0$ and $t > 0$. For any $\psi \in L$ and any integer $m \geqslant 1$ define the set of states

$$\mathcal{J}_m(\psi) = \{J(\tau_m) \dots J(\tau_1)\psi : \ 0 \leq \tau_1, \dots, \tau_m\} \subset \mathcal{M}_h$$

which the system can visit at the m-th flip.

Theorem 2.1 (Covering theorem). *Assume that $V \in \mathbf{V}^{+} \cap \mathbf{V}_{ind}$, then there exists $m \geqslant 1$ such that for any $\psi \in L$ we have $\mathcal{J}_m(\psi) = \mathcal{M}_h$.*

We introduce randomness by the following assumption.
Assumption A0. Positive random variables

$$\tau_1 = t_1, \tau_2 = t_2 - t_1, \dots, \tau_n = t_n - t_{n-1}, \dots$$

are assumed to be independent, identically distributed with measure $P_\tau = \rho(s)ds$ where the density ρ (w.r.t. Lebesgue measure ds) is positive everywhere on R_+, and moreover the first moment $E\tau_1 < \infty$.

If, for example, τ_i are exponentially distributed with the density $\lambda \exp(-\lambda\tau)$, $\lambda > 0$, then it defines Markov process $\psi(t)$ with right continuous deterministic trajectories and random jumps. Such processes are often called piecewise deterministic Markov processes, see for example [14]. At the same time, this can be considered as an example from random perturbation theory, see [12] where the problem of invariant measures is studied.

Let π be Liouville measure on the energy surface, defined by the surface form $d\sigma$ divided by $|\nabla H|$. It is well known that π is invariant w.r.t. hamiltonian dynamics and also w.r.t. velocity flips.

Theorem 2.2 (Convergence theorem). *Assume that $V \in \mathbf{V}^+ \cap \mathbf{V}_{ind}$. Then under assumption A0) for any initial $\psi(0)$ and any bounded measurable real function f on $\mathcal{M}_h$ we have a.s.*

$$M_f(T) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T f(\psi(t))dt \to \pi(f) \stackrel{\text{def}}{=} \int_{\mathcal{M}_h} f d\pi, \quad \text{as } T \to \infty.$$

We call this property Liouville ergodicity.

3. Deterministic part — proof of covering theorem

Plan of the proofs In all assertions below we always assume that $V \in \mathbf{V}^+ \cap \mathbf{V}_{ind}$. Theorem 2.2 will follow from Theorem 2.1. Theorem 2.1 will follow from the following weaker results.

Theorem 3.1 (Closure theorem). *There exists $m \geqslant 1$ such that for all $\psi \in L$ we have $\overline{\mathcal{J}_m(\psi)} = \mathcal{M}_h$.*

This way of proof is very natural — it is important to know whether one can reach any point from any other. In our case it is absolutely not evident. We want to note that after the paper was finished, we found an analog of Theorem 3.1, but not of the Theorem 2.1, in completely different situation in [21].

The covering and closure theorems have simpler local analogue. And moreover it shows that the dimension of $\mathcal{J}_m(\psi)$ grows as m for $m = 1, 2, \dots, 2N-1$. For exact formulation we need some definitions. For any $\tau_1, \dots, \tau_m > 0$ denote $J(\tau_1, \dots, \tau_m) = J(\tau_m) \dots J(\tau_1)$. It can be considered as the mapping from L to $\mathcal{M}_h$ for fixed $(\tau_1, \dots, \tau_m)$, but also $J(\tau_1, \dots, \tau_m)\psi$ can be considered as the map

$$J_m^\psi : \Omega_m \to \mathcal{M}_h, \quad \bar{\tau} = (\tau_1, \dots, \tau_m) \mapsto J(\tau_1, \dots, \tau_m)\psi$$

from the m-dimensional orthant $\Omega_m = \{(\tau_1, \dots, \tau_m) : \ \tau_i > 0, \ \ i = 1, \dots, m\} = \mathbb{R}_+^m$, to $\mathcal{M}_h$, for fixed $\psi \in \mathcal{M}_h$.

Theorem 3.2 (Local covering theorem). *For any point $\psi \in \mathcal{M}_h$ and any $k = 1, \dots, 2N-1$ the dimension of $J_k^\psi \Omega_k$ is equal to k. Moreover, there exists open subset $U \subset \Omega_k$ such that the mapping $J_k^\psi : U \hookrightarrow \mathcal{M}_h$ is a smooth embedding.*

3.1. Key definitions, notation and some intuition

The sequence (1.1) completely defines the trajectory, or the path. Proof of convergence and covering theorems could be obtained, in some sense, by "summation" over all possible paths. For this we have to use various coordinate systems on the energy surface. To get some intuition it is useful to see how it works for easier cases $N = 1, 2$.

Action-angle coordinates Let $v_1, \dots, v_N$ be the orthonormal eigenvectors of V in $\mathbb{R}^N$, and let $\omega_1^2, \dots, \omega_N^2$ be the corresponding eigenvalues of V. For any $k = 1, \dots, N$ define $2N$-vectors in L:

$$Q_k = (v_k, 0)^T, \quad P_k = (0, v_k)^T \tag{3.1}$$

The coordinates of $\psi \in L$ in the basis $Q_1, P_1, \dots, Q_N, P_N$ are denoted by $(\tilde{q_k}, \tilde{p_k})^T$, that is,

$$\psi = \sum_{k=1}^{N} \tilde{q}_k Q_k + \sum_{k=1}^{N} \tilde{p}_k P_k. \tag{3.2}$$

They correspond to coordinates and momenta of the independent "quasiparticles", that is, one-dimensional oscillators with energies $r_k^2/2$ where

$$r_k^2(\psi) = \tilde{p}_k^2 + \omega_k^2 \tilde{q}_k^2 = (\psi, P_k)_2^2 + \omega_k^2 (\psi, Q_k)_2^2 = (p, P_k)_2^2 + \omega_k^2 (q, Q_k)_2^2, k = 1, \dots, N$$

are the action coordinates of the point $\psi = (q, p)^T$. We agree that $r_k(\psi) = \sqrt{r_k^2(\psi)} \geqslant 0$. It is easy to see that r_k are integrals of the hamiltonian dynamics, that is, for any $t \geqslant 0$,

$$r_k^2(\psi) = r_k^2(e^{tA}\psi). \tag{3.3}$$

The angle variables then are the angles for these oscillators.

The following assertions easily follow from the known facts, see for example [1], pp. 103, 272. But the proof is very elementary and we give it for the reader's convenience in Appendix.

Lemma 3.1. *Consider hamiltonian dynamics* (2.2)*, then*

1. *For any $r_1 \geqslant 0, \dots, r_N \geqslant 0$ the set*

   $$T(r_1, \dots, r_N) = \{\psi \in L : r_k(\psi) = r_k, k = 1, \dots, N\}$$

   is invariant and diffeomorphic to torus of dimension $N - n$, where n equals the number of zeros among $r_1, \dots, r_N$.

2. *For any point $\psi \in L$ the closure of its orbit coincides with the torus defined by it, that is*

   $$\overline{\{e^{tA}\psi : t \geqslant 0\}} = T(r_1(\psi), \dots, r_N(\psi)).$$

3. *Thus the torus defines the vector $\bar{r} = r_1, r_2, \dots, r_N$. Vice-versa, any such vector with non-negative coordinates uniquely defines the torus. This torus lies on the energy surface $\mathcal{M}_h$ iff*

   $$\sum_{k=1}^{N} r_k^2 = 2h.$$

For convenience we put $h = 1/2$, and denote the set of all invariant tori on $\mathcal{M} = \mathcal{M}_{1/2}$ by

$$\mathbf{T} = \{\bar{r} = (r_1, \ldots, r_N) : r_i \geqslant 0, \sum_{k=1}^{N} r_k^2 = 1\}.$$

Case $N = 1$. Here ω_1^2 can be arbitrary, and we take $h = \omega_1^2 = 1$. Let S be the corresponding circle. Then the particle moves along S with constant angle velocity in the clock-wise direction. Covering theorem is evident in this case (with $m = 1$). However, for $N = 1$ we have much stronger statement.

For any ψ, ψ' denote by $T(\psi, \psi')$ the minimal $t > 0$ such that $\psi' = e^{At}\psi$. Then $T_0 = T(\psi, \psi)$ is the first return time, or the time of complete rotation around the circle, and for any ψ

$$T(\psi, -\psi) = \frac{1}{2} T_0.$$

Theorem 3.3. *For any ψ, ψ' and any $t \geq T_0$ there exists $0 < t_1 = t_1(\psi, \psi', t) < T_0$ such that $\exp\{A(t - t_1)\} I \exp\{At_1\}\psi = \psi'$.*

See Appendix for the proofs.

$(r, \tilde{p})$-coordinates and (r, p)-coordinates Note that I can be written as

$$I = E - 2P_1,$$

where E is the identity matrix and P_1 is the orthogonal projector on the vector $g_1 = (0, e_1)^T \in L$. The expansion

$$e_1 = \sum_{k=1}^{N} \beta_k v_k$$

of the vector $e_1 \in R^N$ defines the numbers β_k. The p_k and $\tilde{p}_k$ coordinates of the vector p in R^N are related by the formulas:

$$p = \sum_{i=1}^{N} p_i e_i = \sum_{k=1}^{N} \tilde{p}_k v_k,$$

$$\tilde{p}_k = \sum_{i=1}^{N} p_i (e_i, v_k)_2, \quad p_k = \sum_{i=1}^{N} \tilde{p}_i (v_i, e_k)_2. \tag{3.4}$$

Note also that I acts only on p_k-coordinates and $\tilde{p}_k$-coordinates, but not on q_k and $\tilde{q}_k$. Thus the following notation is justified

$$Ip = p - 2(p, e_1)_2 e_1 = \sum_{k=1}^{N} \tilde{p}'_k v_k, \quad \tilde{p}'_k = \tilde{p}_k - 2p_1\beta_k, \tag{3.5}$$

where

$$p_1 = (p, e_1)_2 = \sum_{k=1}^{N} \tilde{p}_k \beta_k.$$

By (3.4) the velocity flip changes the action variables as follows

$$\begin{aligned} r_k^2(I\psi) &= (Ip, v_k)_2^2 + \omega_k^2(q, v_k)_2^2 = (\tilde{p}_k')^2 + \omega_k^2(q, v_k)_2^2 \\ &= \tilde{p}_k^2 + 4p_1^2\beta_k^2 - 4\tilde{p}_k p_1 \beta_k + \omega_k^2(q, v_k)_2^2 \\ &= r_k^2(\psi) + 4p_1^2\beta_k^2 - 4\tilde{p}_k p_1 \beta_k = r_k^2(\psi) + 4p_1^2\beta_k^2 - 4p_1\beta_k(p, v_k)_2. \end{aligned}$$

The following formula defines the time evolution of the tori, in terms of initial action variable and momenta at time $t-0$. Namely, for any k,

$$r_k^2(Ie^{tA}\psi) = r_k^2(I\psi(t-0)) = r_k^2(\psi) + 4p_1^2(t-0)\beta_k^2 - 4p_1(t-0)\beta_k(p(t-0), v_k)_2, \tag{3.6}$$

where

$$\begin{aligned} &\psi(t-0) = e^{tA}\psi = (q(t-0), \quad p(t-0))^T \in L, \\ &p(t-0) = (p_1(t-0), \dots, p_N(t-0))^T. \end{aligned}$$

We can rewrite (3.6) as

$$\bar{r}(J(t)\psi) = \Psi(\bar{r}(\psi), p(t-0)), \tag{3.7}$$

where

$$\Psi = (\Psi_1, \dots, \Psi_N), \Psi_k^2(\bar{r}, p) = r_k^2 + 4p_1^2\beta_k^2 - 4p_1\beta_k(p, v_k)_2$$

has the domain of definition

$$D(\Psi) = \{(\bar{r}, p) : \bar{r} \in \mathbf{T}, p \in T_p(\bar{r})\} \subset \mathbf{T} \times \mathbb{R}^N$$

where $T_p(\bar{r})$ is the projection of the torus $T(\bar{r})$ onto the space of momenta, that is the set of all $p \in \mathbb{R}^N$ such that for some $q \in \mathbb{R}^N$ we have $(q,p)^T \in T(\bar{r})$. For any $\psi = \psi(0) = (q,p)^T, q = q(0), p = p(0)$ consider the following set of tori

$$\mathcal{T}_1(\psi) = \overline{\{\bar{r}(J(t)\psi) : t \geqslant 0\}} = \overline{\{\Psi(\bar{r}(\psi), p(t-0)) : t > 0, p(0) = p\}} \subset \mathbf{T}.$$

From continuity of Ψ it follows that $\mathcal{T}_1(\psi)$ depends only on the invariant torus containing ψ, that is if $\bar{r}(\psi) = \bar{r}(\psi')$ for two points $\psi, \psi' \in L$, then

$$\mathcal{T}_1(\psi) = \mathcal{T}_1(\psi'). \tag{3.8}$$

Then, the following notation is correct

$$\mathcal{T}_1(\psi) = \mathcal{T}_1(\bar{r}(\psi)). \tag{3.9}$$

It is useful to note that by continuity of Ψ for any $\bar{r} \in \mathbf{T}$ we have

$$\mathcal{T}_1(\bar{r}) = \{\Psi(\bar{r}, p) : p \in T_p(\bar{r})\}.$$

Case $N = 2$ Here we will only give intuitive arguments why for any initial state ψ we can enter any torus in finite number of jumps. From (3.6) one can get the following formula for the evolution of the action variables

$$r_1^2(J(t)\psi) - r_1^2(\psi) = 4\beta_1\beta_2(\tilde{p_1}(t-0)\beta_1 + \tilde{p_2}(t-0)\beta_2)(\tilde{p_2}(t-0)\beta_1 - \tilde{p_1}(t-0)\beta_2), \tag{3.10}$$

and also

$$r_2^2(J(t)\psi) = 1 - r_1^2(J(t)\psi).$$

Denote the right-hand side of (3.10) by $D(t)$. It appears that for some t' the set $\{D(t) : 0 \le t \le t'\}$ contains an open interval, and moreover its length has lower bound $\delta > 0$ uniformly in ψ. Thus the measure of the set of visited tori, after each application of I, enlarges by an additive constant. It follows that all tori may be visited for finite number of I transformations. To see this for general N is more difficult.

3.2. Proof of closure theorem

3.2.1. Contraction property

Define the function ρ in $\mathbf{T}$ by

$$\rho(\bar{r}, \bar{r}') = \sum_{k=1}^{N} |r_k^2 - (r_k')^2|,$$

where $\bar{r} = (r_1, \ldots, r_N)$, $\bar{r}' = (r_1', \ldots, r_N')$. Denote $\bar{r}^* = (|\beta_1|, \ldots, |\beta_N|) \in \mathbf{T}$, thus it is the invariant torus, containing the point $g_1 = \sqrt{2}(0, e_1)^T \in L$. Note that all β_k are nonzero, that follows from the assumption $V \in \mathbf{V}^+$, see Section 5.3. The point g_1 corresponds to the configuration of particles where all particles have zero velocity and zero coordinates, except for the particle 1.

Theorem 3.4 (Contraction theorem). *For any $\bar{r} = (r_1, \ldots, r_N) \in \mathbf{T}$ we have the following contraction bound*

$$\rho(\mathcal{T}_1(\bar{r}), \bar{r}^*) \leqslant (1 - c(\bar{r}))\rho(\bar{r}, \bar{r}^*),$$

where the constant $c(\bar{r})$ is given by

$$c(\bar{r}) = \frac{1}{\max\{1, D^2(\bar{r})\}}, \quad D(\bar{r}) = \max_{k=1,\ldots,N} \frac{r_k}{|\beta_k|} - \min_{k=1,\ldots,N} \frac{r_k}{|\beta_k|}. \tag{3.11}$$

Proof. We will find a point $p' \in T_p(\bar{r})$ such that

$$\rho(\Psi(\bar{r}, p'), \bar{r}^*) = (1 - c(\bar{r}))\rho(\bar{r}, \bar{r}^*).$$

Note first that in coordinates $\tilde{p}$ the set $T_p(\bar{r})$ is the N-dimensional cube with sides $(2r_1, \dots, 2r_N)$, this follows from the oscillator representation. In other words, the point $p = (\tilde{p}_1, \dots, \tilde{p}_N)^T \in T_p(\bar{r})$ iff for any $k = 1, \dots, N$

$$|\tilde{p}_k| \leqslant r_k. \tag{3.12}$$

For $k = 1, \dots, N$ denote

$$\gamma_k = \frac{r_k}{|\beta_k|}$$

and denote the minimal of them by γ_n and maximal by γ_N. We will need the following functions

$$f^{\pm}(x) = \frac{1}{2}\left(x \pm \sqrt{x^2 + c(1 - x^2)}\right)$$

of $x \in \mathbb{R}$, where the constant c is defined in (3.11). Since $c \leqslant 1$, we have for all x,

$$x^2 + c(1 - x^2) \geq 0.$$

Define the point $p' \in \mathbb{R}^N$ with coordinates $(\tilde{p}'_1, \dots, \tilde{p}'_N)$ in the basis $v_1, \dots, v_N$ by the formula

$$\tilde{p}'_k = y\beta_k - c\frac{\beta_k^2 - r_k^2}{4y\beta_k}, \quad y = f^+(\gamma_n),$$

and find the value of $\Psi(\bar{r}, p')$. As

$$p_1 = (p', e_1)_2 = \sum_{k=1}^{N} \beta_k \tilde{p}'_k = y\sum_{k=1}^{N} \beta_k^2 - \frac{c}{4y}\sum_{k=1}^{N}(\beta_k^2 - r_k^2) = y,$$

we have for any $k = 1, \dots, N$

$$\Psi_k^2 = r_k^2 + 4p_1^2\beta_k^2 - 4p_1\beta_k\tilde{p}'_k = r_k^2 + 4y^2\beta_k^2 - 4y\beta_k\left(y\beta_k - c\frac{\beta_k^2 - r_k^2}{4y\beta_k}\right) = r_k^2 + c(\beta_k^2 - r_k^2).$$

Then we have the distance

$$\rho(\Psi(\bar{r}, p'), \bar{r}^*) = \sum_{k=1}^{N} |\Psi_k^2 - \beta_k^2| = (1 - c)\rho(\bar{r}, \bar{r}^*).$$

Lemma 3.2. *$p' \in T_p(\bar{r})$, that is the inequalities* (3.12) *hold.*

Proof consists of simple calculations and is given in Appendix.

From the contraction theorem only, one cannot prove the closure theorem because $c = c(\bar{r})$ depends on $\bar{r}$. That is why we need one more assertion. For any $\bar{r} = (r_1, \dots, r_N) \in \mathbf{T}$ put

$$A(\bar{r}) = \min_{k=1,\dots,N} \frac{r_k}{|\beta_k|}, \quad B(\bar{r}) = \max_{k=1,\dots,N} \frac{r_k}{|\beta_k|}, \quad \Delta(\bar{r}) = B^2(\bar{r}) - A^2(\bar{r}).$$

The function $\Delta(\bar{r})$ defined on $\mathbf{T}$ also looks like a distance to the point $\bar{r}^*$. Let us explain this in more detail. In particular, $\Delta(\bar{r}) = 0$ iff $r_k = |\beta_k|$ for all $k = 1, \ldots, N$. Moreover,

$$\sum_{k=1}^{N} r_k^2 = 1, \tag{3.13}$$

and hence, if $\Delta(\bar{r})$ is small, then the numbers r_k are close to $|\beta_k|$. Even more, it gives an upper bound for the distance, more exactly, for any $\bar{r} \in \mathbf{T}$,

$$\begin{aligned}\rho(\bar{r}, \bar{r}^*) &= \sum_k |r_k^2 - \beta_k^2| = \sum_k \beta_k^2 |\frac{r_k^2}{\beta_k^2} - 1| \\ &\leqslant \sum_k \beta_k^2 (B^2(\bar{r}) - A^2(\bar{r})) = \sum_k \beta_k^2 \Delta(\bar{r}) \leqslant \Delta(\bar{r}). \end{aligned} \tag{3.14}$$

The following result shows that each velocity flip at appropriate moments makes point g_1 closer by 1.

Corollary 3.1. *For any point $\bar{r} \in \mathbf{T}$ there exists $p \in T_p(\bar{r})$ such that*

$$\Delta(\Psi(\bar{r}, p)) \leqslant \max\{\Delta(\bar{r}) - 1, 0\}.$$

Proof. From the proof of contraction theorem it follows that there exists $p' \in T_p(\bar{r})$ such that

$$\Psi_k^2(\bar{r}, p') = r_k^2 + c(\bar{r})(\beta_k^2 - r_k^2).$$

Denote $\bar{r}' = (r_1', \ldots, r_N') = \Psi(\bar{r}, p')$. Without loss of generality one can choose the indices so that

$$\frac{r_1'}{|\beta_1|} = A(\bar{r}'), \quad \frac{r_N'}{|\beta_N|} = B(\bar{r}').$$

Then

$$\Delta(\bar{r}') = \frac{(r_N')^2}{\beta_N^2} - \frac{(r_1')^2}{\beta_1^2} = (1 - c(\bar{r}))\Big(\frac{r_N^2}{\beta_N^2} - \frac{r_1^2}{\beta_1^2}\Big),$$

from where we get

$$\Delta(\bar{r}') \leqslant (1 - c(\bar{r}))\Delta(\bar{r}), \tag{3.15}$$

$$c(\bar{r}) = \frac{1}{\max\{1, (B(\bar{r}) - A(\bar{r}))^2\}}.$$

Note that the following inequality holds

$$(B - A)^2 \leqslant B^2 - A^2 = \Delta$$

and then

$$c(\bar{r}) = \frac{1}{\max\{1, (B(\bar{r}) - A(\bar{r}))^2\}} \geqslant \frac{1}{\max\{1, B^2(\bar{r}) - A^2(\bar{r})\}} = \frac{1}{\max\{1, \Delta(\bar{r}\}}.$$

From this inequality and the bound (3.15) finally we have

$$\Delta(\bar{r}') \leqslant \Big(1 - \frac{1}{\max\{1, \Delta(\bar{r})\}}\Big)\Delta(\bar{r}) = \max\{\Delta(\bar{r}) - 1, 0\}.$$

□

3.2.2. Closure theorem

Before proving the theorem we have to prove a weaker assertion.

Lemma 3.3. *There exists an integer $m \geqslant 1$ such that for any $\psi \in \mathcal{M}$,*

$$g_1 \in \overline{\mathcal{J}_m(\psi)}.$$

In other words, there exists m such that for any $\psi \in \mathcal{M}$ and any $\varepsilon > 0$ one can find moments $0 < t_1 < \ldots < t_m$ so that

$$||J(\tau_m)\ldots J(\tau_1)\psi - g_1||_2 < \varepsilon.$$

By continuity of maps Ψ and Δ, and by Corollary 3.1 it follows that for any point $\psi \in L$ and any $\varepsilon > 0$ one can find a time moment $t \geqslant 0$ such that

$$\Delta(\bar{r}(J(t)\psi)) = \Delta(\Psi(\bar{r}(\psi), p(t-0))) \leqslant \max\{\Delta(\bar{r}(\psi)) - 1, 0\} + \varepsilon.$$

It follows that there exists $m \geqslant 1$ such that for any $\varepsilon > 0$ there exist time moments $t_1, \ldots, t_m \geqslant 0$ such that

$$\Delta(\bar{r}(\psi')) \leqslant \varepsilon, \quad \psi' = J(\tau_m)\ldots J(\tau_1)\psi,$$

and moreover one can take $m = [\Delta(\bar{r}(\psi))] + 1$. But as for any $\bar{r} \in \mathbf{T}$ the following inequality holds

$$\Delta(\bar{r}) \leqslant \max_{k=1,\ldots,N} \frac{1}{\beta_k^2},$$

m can be chosen uniformly in ψ. By formula (3.14) we get

$$\rho(\bar{r}(\psi'), \bar{r}^*) \leqslant \varepsilon.$$

Alternatively it is evident that there exists a constant $c > 0$ such that for any $\varepsilon' > 0$ there is $t \geqslant 0$ such that

$$||e^{tA}\psi' - g_1||_2^2 \leqslant c\rho(\bar{r}(\psi'), \bar{r}^*) + \varepsilon'.$$

As ε and ε' are arbitrary and by $e^{tA} = J(0)J(t)$ we get the proof. □

Now we prove the closure theorem. Define another norm on L

$$||\psi||_H = \sqrt{H(\psi)}.$$

Let us fix two points $\psi, \psi' \in \mathcal{M}$ and show that $\psi' \in \overline{\mathcal{J}_m(\psi)}$ for some $m \geqslant 1$ not depending on ψ, ψ'. By Lemma 3.3 there exist $\tau_1, \dots, \tau_m > 0$ and $\tau_1', \dots, \tau_m' > 0$ such that

$$||J(\tau_1, \dots, \tau_m)\psi - J(\tau_1', \dots, \tau_m')\psi'||_H < \varepsilon.$$

It is clear that the transform $J(\tau_1', \dots, \tau_m')$ is invertible and conserves the norm $|| \ ||_H$, that is why

$$||J^{-1}(\tau_1', \dots, \tau_m')J(\tau_1, \dots, \tau_m)\psi - \psi'||_H < \varepsilon.$$

One can find $m' \leqslant 2m$ and $\tau_1'', \dots, \tau_{m'}'' > 0$ such that

$$||J(\tau_1'', \dots, \tau_{m'}'')J(\tau_1, \dots, \tau_m)\psi - \psi'||_H < 2\varepsilon. \tag{3.16}$$

In fact

$$J^{-1}(\tau_1, \dots, \tau_m) = e^{-\tau_1 A} I e^{-\tau_2 A} I \dots e^{-\tau_m A} I. \tag{3.17}$$

As e^{tA} conserves the norm $||.||_H$ and $I = J(0)$, it is sufficient to show that for any point $\psi \in \mathcal{M}$, any $\varepsilon > 0$ and $t \geqslant 0$ there exists $s = s(\psi, t, \varepsilon)$ such that

$$||e^{-tA}\psi - e^{sA}\psi||_H < \varepsilon.$$

But as the closure of the orbit of any point is the torus, it is Poincare recurrence theorem.

From (3.16) and equivalence of the norms $|| \ ||_2$ and $|| \ ||_H$ the closure theorem follows.

3.3. Proof of local covering theorem

Fix ψ. It is sufficient to show that for some point $\bar{\tau} \in \Omega_k$ the rank of $dJ_k^{\psi}(\bar{\tau})$ equals k. The columns of the Jacobian matrix J_k^{ψ} are the vectors

$$\theta_k(\tau) = \frac{d}{d\tau_k} J(\tau_1, \dots, \tau_k)\psi = I e^{\tau_k A} A J(\tau_1, \dots, \tau_{k-1})\psi,$$
$$\theta_i(\tau) = \frac{d}{d\tau_i} J(\tau_1, \dots, \tau_k)\psi = I e^{\tau_k A} \frac{d}{d\tau_i} J(\tau_1, \dots, \tau_{k-1})\psi, \quad i = 1, \dots, k-1.$$

Denote by $\dim\langle w_1, w_2, \dots\rangle$ the dimension of the linear span of vectors $w_1, w_2, \dots$ As the mapping $I \exp\{\tau_k A\}$ is non-degenerate, one can derive for any $k = 1, 2, \dots$ that there is $\bar{\tau} = (\tau_1, \dots, \tau_k) \in \Omega_k$ such that

$$\dim\langle\theta_1(\tau), \dots, \theta_k(\tau)\rangle = \dim\Big\langle AJ(\bar{\tau}_c)\psi, \frac{d}{d\tau_1} J(\bar{\tau}_c)\psi, \dots, \frac{d}{d\tau_{k-1}} J(\bar{\tau}_c)\psi\Big\rangle = k,$$

where $\bar{\tau}_c = (\tau_1, \ldots, \tau_{k-1}) \in \Omega_{k-1}$. We shall do it by induction in k. Remind that for any $\psi \in L$,

$$I\psi = \psi - 2(\psi, g_1)_2 g_1.$$

In fact, for $k = 1$ this is trivial.

Induction step. Assuming that we have proved the assertion for $k \leq 2N-2$, we shall prove it for $k+1$.

For $\bar{\tau} = (\tau_1, \ldots, \tau_k) \in \Omega_k$ denote $\bar{\tau}_c = (\tau_1, \ldots, \tau_{k-1}) \in \Omega_{k-1}$. Then

$$\begin{aligned} AJ(\bar{\tau})\psi &= AI\exp\{\tau_k A\}J(\tau_1, \ldots, \tau_{k-1})\psi \\ &= A\exp\{\tau_k A\}J(\tau_1, \ldots, \tau_{k-1})\psi + a_0(\bar{\tau})Ag_1 \\ &= \exp\{\tau_k A\}\left(AJ(\tau_1, \ldots, \tau_{k-1})\psi + a_0(\bar{\tau})Ae^{-\tau_k A}g_1\right) \\ &= \exp\{\tau_k A\}\left(AJ(\bar{\tau}_c)\psi + a_0(\bar{\tau})A\exp\{-\tau_k A\}g_1\right), \end{aligned}$$

where $a_0(\bar{\tau}) = -2(\exp\{\tau_k A\}J(\bar{\tau}_c)\psi, g_1)_2$. Similarly we get equalities for the derivatives for $i = 1, \ldots, k-1$:

$$\begin{aligned} \frac{d}{d\tau_i}J(\bar{\tau})\psi &= I\exp\{\tau_k A\}\frac{d}{d\tau_i}J(\tau_1, \ldots, \tau_{k-1})\psi \\ &= \exp\{\tau_k A\}\frac{d}{d\tau_i}J(\tau_1, \ldots, \tau_{k-1})\psi + a_i(\bar{\tau})g_1 \\ &= \exp\{\tau_k A\}\Big(\frac{d}{d\tau_i}J(\bar{\tau}_c)\psi + a_i(\bar{\tau})\exp\{-\tau_k A\}g_1\Big), \\ a_i(\bar{\tau}) &= -2(\exp\{\tau_k A\}\frac{d}{d\tau_i}J(\bar{\tau}_c)\psi, g_1)_2. \end{aligned}$$

Derivative in τ_k has the following expansion:

$$\begin{aligned} \frac{d}{d\tau_k}J(\bar{\tau})\psi &= IA\exp\{\tau_k A\}J(\tau_1, \ldots, \tau_{k-1})\psi \\ &= A\exp\{\tau_k A\}J(\tau_1, \ldots, \tau_{k-1})\psi + a_k(\bar{\tau})g_1 \\ &= \exp\{\tau_k A\}\left(AJ(\bar{\tau}_c)\psi + a_k(\bar{\tau})\exp\{-\tau_k A\}g_1\right), \\ a_k(\bar{\tau}) &= -2(A\exp\{\tau_k A\}J(\bar{\tau}_c)\psi, g_1)_2. \end{aligned}$$

Then by nondegeneracy of the mapping $\exp\{\tau_k\}A$, we get that the dimension of the subspace

$$\langle AJ(\bar{\tau})\psi, \frac{d}{d\tau_1}J(\bar{\tau})\psi, \ldots, \frac{d}{d\tau_k}J(\bar{\tau})\psi\rangle$$

equals the dimension $\aleph$ of the linear span of the following $(k+1)$ vectors:

$$\begin{aligned} &AJ(\bar{\tau}_c)\psi + a_0(\bar{\tau})A\exp\{-\tau_k A\}g_1, AJ(\bar{\tau}_c)\psi + a_k(\bar{\tau})\exp\{-\tau_k A\}g_1, \\ &\frac{d}{d\tau_1}J(\bar{\tau}_c)\psi + a_1(\bar{\tau})\exp\{-\tau_k A\}g_1, \ldots, \frac{d}{d\tau_{k-1}}J(\bar{\tau}_c)\psi + a_{k-1}(\bar{\tau})\exp\{-\tau_k A\}g_1. \end{aligned}$$

By inductive assumption there exists a point $\bar{\tau}_c$ such that the vectors

$$AJ(\bar{\tau}_c)\psi, \frac{d}{d\tau_1}J(\bar{\tau}_c)\psi, \dots, \frac{d}{d\tau_{k-1}}J(\bar{\tau}_c)\psi$$

are linearly independent. By Lemmas 3.4 and 3.5 there exists $\tau_k \geqslant 0$ such that the vectors

$$AJ(\bar{\tau}_c)\psi, \frac{d}{d\tau_1}J(\bar{\tau}_c)\psi, \dots, \frac{d}{d\tau_{k-1}}J(\bar{\tau}_c)\psi, A\exp\{-\tau_k A\}g_1, \exp\{-\tau_k A\}g_1$$

are linearly independent and $a_0(\bar{\tau}) = -2(\exp\{\tau_k A\}J(\bar{\tau}_c)\psi, g_1)_2 \neq 0$. Thus $\aleph = k+1$ and the induction step is proved. □

Lemma 3.4. *For any vector $\psi \in L \setminus \{0\}$ the set*

$$\{t > 0 : \ (e^{tA}\psi, g_1)_2 \neq 0\} \subset \mathbb{R}_+$$

is open end everywhere dense.

Proof. To prove this, note the following equalities

$$(e^{tA}\psi, g_1)_2 = (e^{tA}\psi, g_1)_H = (\psi, e^{-tA}g_1)_H.$$

The second one holds since $A^\star = -A$, where $A^\star$ is the adjoint operator to A, for the scalar product $(,)_H$, which follows from the equality

$$(Au, w)_H = (Vp, q')_2 - (Vq, p')_2 = -((Vq, p')_2 - (p, Vq')_2) = -(u, Aw)_H$$

for $u = (q,p)^T, \ w = (q', p')^T \in L$. Then we use the expansion

$$(\psi, e^{-tA}g_1)_H = \sum_{k=0}^{\infty} (-1)^k \frac{t^k}{k!}(\psi, A^k g_1)_H$$

as, by Lemma 5.1, the linear span of vectors $A^k g_1, k = 0, 1, 2, \dots$ coincides with L. Then by analyticity of the left-hand part of the latter formula we get the proof. □

Lemma 3.5. *For any linearly independent vectors $w_1, \dots, w_k \in L, \ k < 2N-1$, the set*

$$T(w_1, \dots, w_k) = \{t > 0 : \dim\langle e^{-tA}g_1, Ae^{-tA}g_1, w_1, \dots, w_k\rangle = k+2\} \subset \mathbb{R}_+$$

contains an open and everywhere dense subset.

Proof. If $k < 2N-2$, then choose the vectors $w_{k+1}, \dots, w_{2N-2}$ so that

$$\dim\langle w_1, \dots, w_{2N-2}\rangle = 2N-2.$$

Note that $T(w_1, \dots, w_{2N-2}) \subset T(w_1, \dots, w_k)$. That is why it is sufficient to prove the assertion of the lemma for the case $k = 2N-2$. Thus further on we assume that $k = 2N-2$.

Consider the following real function on L:

$$W(u) = \det[u, Au, w_1, \dots, w_{2N-2}].$$

Denote by $X = \{u \in L : \ W(u) = 0\}$ the set of zeros of the function W. It is clear that

$$T(w_1, \dots, w_{2N-2}) = \{t > 0 : W(e^{-tA}g_1) \neq 0\} = \{t > 0 : \ e^{-tA}g_1 \notin X\}.$$

It is clear that W is a quadratic function. Thus X is a quadric. Further we define the canonical type of the quadric X. Denote

$$L_1 = \langle w_1, \dots, w_{2N-2}\rangle.$$

Consider also the orthogonal complement $L_1^{\perp}$ to L_1 (in the scalar product $(,)_H$). By definition of W, we have $L_1 \subset X$. Thus in some coordinates $(u_1, u_2, \dots, u_{2N})$ on L the function W will look like

$$W(u) = a_1 u_1^2 + a_2 u_2^2,$$

for some $a_1, a_2 \in \mathbb{R}$. (As the basis for such coordinates, one can choose the vectors $w_1, \dots, w_{2N-2}, u_1, u_2$, where u_1, u_2 are the appropriate coordinates of the orthogonal compliment $L_1^{\perp}$. Since $L_1 \subset X$, the form will not depend on the first $2N-2$ coordinates, and one can choose two coordinates in $L_1^{\perp}$ by the method of Lagrange).

If we show that there exists a vector u such that $W(u) \neq 0$, then it will follow that a_1 and a_2 cannot be simultaneously zero. And consequently the canonical type of X should be one of three types: a $2N-2$-dimensional hyperplane, $2N-1$-dimensional hyperplane, the union of two $2N-1$ hyperplanes. Then applying the same argument as in Lemma 3.4 we get the proof of Lemma 3.5.

Let us show that there exists a vector u such that $W(u) \neq 0$.

For a vector $u \in L$ denote by $u^{\perp}$ and $(Au)^{\perp}$ the orthogonal projections on $L_1^{\perp}$ of the vectors u and Au correspondingly. As the determinant is polilinear we have

$$W(u) = \det[u^{\perp}, (Au)^{\perp}, w_1, \dots, w_{2N-2}].$$

That is why X is the set of all $u \in L$ such that $u^{\perp}$and $(Au)^{\perp}$ are linearly dependent. Note that for any $u = (q,p)^T \in L$

$$(u, Au)_H = (Vq, p)_2 - (p, Vq)_2 = 0.$$

Consider two cases:

1. Subspace L_1 is invariant with respect to A. Let us show that then $L_1^\perp$ is also invariant with respect to A. In fact, for $u \in L_1^\perp,\ v \in L_1$ we have

$$(Au, v)_H = (u, A^\star v)_H = -(u, Av)_H = 0,$$

as $Av \in L_1$ by invariance of L_1. Then $L_1^\perp$ is invariant with respect to A and $X \cap L_1^\perp$ is the set of all $u \in L_1^\perp$ such that u and Au are linearly dependent. As the spectrum of A is pure imaginary and does not contain zero, the vectors u and Au are linearly independent for all $u \neq 0 \in L_1^\perp$. Then $X \cap L_1^\perp = \{0\}$ and the proof is finished in this case.

2. Subspace L_1 is not invariant with respect to A. In this case there exists vector $v \in L_1$ such that $u = (Av)^\perp \neq 0$. Consider a vector $\tilde{u} \in L_1^\perp$ such that $\tilde{u}, u$ are linearly independent and $\tilde{u}', u$ are linearly dependent, where $\tilde{u}' \in L_1^\perp$ is orthogonal to $\tilde{u}$ in $L_1^\perp$, again with respect to the scalar product $(,)_H$. As $(\tilde{u}, A\tilde{u})_H = 0$, for the projection $(A\tilde{u})^\perp$ of the vector $A\tilde{u}$ on the subspace $L_1^\perp$ we have $(A\tilde{u})^\perp = c\tilde{u}'$, for some $c \in \mathbb{R}$. If $c \neq 0$, then it is clear that $W(\tilde{u}) \neq 0$. Assume that $c = 0$ and consider the vector $u^* = v + \tilde{u}$. Then

$$(u^*)^\perp = \tilde{u}, \quad (Au^*)^\perp = (Av + A\tilde{u})^\perp = u + (A\tilde{u})^\perp = u.$$

By definition of $\tilde{u}$ the vectors $\tilde{u}$ and u are linearly independent. Then $(u^*)^\perp$ and $(Au^*)^\perp$ are also linearly independent, and then $W(u^*) \neq 0$. The proof in this case is also finished □

3.4. Proof of Theorem 2.1

Fix two points $\psi_1, \psi_2 \in \mathcal{M}$. We want to show that for some m^*, not depending on ψ_1, ψ_2, there exist $\tau_1, \ldots, \tau_{m^*} \geqslant 0$ such that

$$J(\tau_1, \ldots, \tau_{m^*})\psi_1 = \psi_2.$$

By local covering theorem there exists a point $\bar{\tau} = (\tau_1, \ldots, \tau_m) \in \Omega_m$ such that there exist two neighborhoods: $O(\bar{\tau}) \subset \Omega_m$ of $\bar{\tau}$ and $O(\psi_1^*) \subset \mathcal{M}$ of $\psi_1^* = J^{\psi_1}(\bar{\tau}) = J(\tau_1, \ldots, \tau_m)\psi_1$ such that

$$J^{\psi_1}(O(\tau)) = O(\psi_1^*). \tag{3.18}$$

Consider ε-neighborhood $O_\varepsilon(\psi_1^*)$ of ψ_1^* in the norm, corresponding to the scalar product $(,)_H$, such that

$$O_\varepsilon(\psi_1^*) = \{\psi \in \mathcal{M} : \ ||\psi - \psi_1^*||_H < \varepsilon\} \subset O(\psi_1^*).$$

By closure theorem there exist $\tau_1', \ldots, \tau_m' \geqslant 0$ such that

$$||J(\tau_1', \ldots, \tau_m')\psi_1^* - \psi_2||_H < \varepsilon. \tag{3.19}$$

Note that for fixed $\tau'_1, \dots, \tau'_m$ the map $J(\tau'_1, \dots, \tau'_m) : L \to L$ is non-degenerate and conserves the norm $|| \, ||_H$. That is why from (3.19) the following inequality follows

$$||\psi_1^* - J^{-1}(\tau'_1, \dots, \tau'_m)\psi_2||_H < \varepsilon.$$

Thus, $J^{-1}(\tau'_1, \dots, \tau'_m)\psi_2 \in O_\varepsilon(\psi_1^*)$ and by (3.18) there is $\bar{\tau}^* = (\tau_1^*, \dots, \tau_m^*) \in O(\bar{\tau})$ such that

$$J(\tau_1^*, \dots, \tau_m^*)\psi_1 = J^{-1}(\tau'_1, \dots, \tau'_m)\psi_2.$$

This can be rewritten as

$$J(\tau_1^*, \dots, \tau_m^*, \tau'_1, \dots, \tau'_m)\psi_1 = J(\tau'_1, \dots, \tau'_m)J(\tau_1^*, \dots, \tau_m^*)\psi_1 = \psi_2.$$

□

4. Stochastic part — proof of convergence theorem

4.1. Embedded process

Here we consider the sequence $\psi_k = \psi(t_k)$, which is a discrete time Markov chain (embedded chain) with state space $\mathcal{M}$. We prove that it is (Markov) ergodic as it is defined in the following theorem concerning more general class of discrete time Markov chains. Namely, we consider Markov chains ξ_n on compact state space X with Borel σ-algebra $\mathcal{B}(X)$ (with countable basis) and transition probability kernels $P(x, A)$, which are probability measures on $\mathcal{B}(X)$ for any $x \in X$ and measurable functions on X for any $A \subset X$. We will consider the class of such chains satisfying the following assumptions:

(A1) for some integer $m \geq 1$ and any $x \in X$ the m-step transition probability $P^m(x, \cdot)$ is equivalent to some finite non-negative measure μ such that $\mu(O) > 0$ for any open set $O \subset X$. Moreover, for any x there exists m-step transition density $p^m(x, y)$ (with respect to μ), which is measurable on $\mathcal{M} \times \mathcal{M}$;

(A2) for any open $O \subset X$ the function $P(x, O)$ is lower semi-continuous.

Theorem 4.1. *Under assumptions* (A1), (A2) *the Markov chain ξ_n is ergodic, that is there exists probability measure π on X such that*

$$\sup_{A \in \mathcal{B}(X)} |P^n(x, A) - \pi(A)| \to 0 \tag{4.1}$$

as $n \to \infty$, uniformly in x.

This theorem follows probably from the existing deep theory of such Markov chains with continuous state space, see for example [15–17], but we did not find the statement we need, and for the reader's convenience we give a short proof below, using the ideas from [15, 17].

Corollary 4.1. *The embedded chain ψ_k has the unique invariant measure π, and moreover it is ergodic as in Theorem* 4.1.

We have only to prove the properties (A1) and (A2) for our embedded chain. For the rest of this section the integer m is the same as in the covering theorem.

Proof of (A1)

We will prove that the measures π and $P^m(\psi, A)$ are equivalent for any ψ.

Lemma 4.1. *For any measurable $B \subset \mathcal{M}$ its Liouville measure $\pi(B) = 0$ iff the Lebesgue measure λ of the set $(J_m^{\psi})^{-1}(B)$ in Ω_m is zero.*

Proof. 1) Assume that for some $B \subset \mathcal{M}$ we have $\pi(B) = 0$. Let us show that $\lambda((J_m^{\psi})^{-1}(B)) = 0$. Let A_{cr} be the set of critical points of the map J_m^{ψ} (that is points $\overline{\tau} = (\tau_1, \dots, \tau_m)$ where the rank of the Jacobian is not maximal) and let $E = J_m^{\psi}(A_{cr}) \subset \mathcal{M}$ be the set of critical values of J_m^{ψ}. By Sard's theorem $\pi(E) = 0$. But since $J_m^{\psi}(\Omega^{\psi}) = \mathcal{M}$, there exists non-critical point $\overline{\tau} = (\tau_1, \dots, \tau_m) \in \Omega_m$ such that the rank of dJ_m^{ψ} at this point equals $2N - 1$. As the map J_m^{ψ} is analytic in the variables $\tau_1, \dots, \tau_m$, the set of points A_{cr} where the rank is less than $2N - 1$, has Lebesgue measure zero. Then the equality $\lambda((J_m^{\psi})^{-1}(B)) = 0$ follows from Theorem 1 of [13].

2) Assume that for some $B \subset \mathcal{M}$ we have $\pi(B) > 0$, and let us show that $\lambda((J_m^{\psi})^{-1}(B)) > 0$. By Lebesgue differentiation theorem there exists a point $\psi' \in \mathcal{M} \setminus E$ and its neighbourhood $O(\psi')$ such that $\pi(O(\psi') \cap B) > 0$. Then there are a point $\overline{\tau} = \overline{\tau}(\psi') \in (J_m^{\psi})^{-1}(\psi')$ and its neighborhood $O(\overline{\tau}) \subset \Omega_m$, so that the restriction of J_m^{ψ} on $O(\overline{\tau})$ is a submersion. Then $\pi(O(\psi') \cap B) > 0$ implies $\lambda((J_m^{\psi})^{-1}(B) \cap O(\overline{\tau})) > 0$. □

Denote by $\rho^{(m)}$ the product of m densities ρ. Then, as for any $B \subset \mathcal{M}$

$$P^m(\psi, B) = \int\limits_{(J_m^{\psi})^{-1}(B)} \rho^{(m)}(\overline{\tau}) d\overline{\tau},$$

by Lemma 4.1 we get that P^m and π are equivalent measures.

The proof of measurability of the transition density follows from Theorem 1, p. 180 of [16], and Proposition 1.1, p. 5, of [17]. □

Proof of (A2)

Lemma 4.2. *Our Markov chain ψ_k is a weak Feller chain, that is for any open $O \subset \mathcal{M}$ the transition probability $P(\psi, O)$ is lower semicontinuous in ψ.*

Proof. For any ψ denote by $\mathbf{1}_{\psi}(\tau)$ the indicator function on R_+, that is, $\mathbf{1}_{\psi}(\tau) = 1$ if $Ie^{\tau A}\psi \in O$ and zero otherwise. Then we have

$$P(\psi, O) = \int\limits_{\mathbb{R}_+} \mathbf{1}_{\psi}(\tau)\rho(\tau)d\tau.$$

Let $\psi_n \to \psi$ as $n \to \infty$, $\psi_n \in \mathcal{M}$. Fix $\tau \geqslant 0$ and consider two cases:

1. $Ie^{\tau A}\psi \in O$, then starting from some n the inclusion $Ie^{\tau A}\psi_n \in O$ holds, as O is open. That is why $\lim_{n\to\infty} \mathbf{1}_{\psi_n}(\tau) = \mathbf{1}_\psi(\tau) = 1$.

2. $Ie^{\tau A}\psi \notin O$. Then $\liminf_n \mathbf{1}_{\psi_n}(\tau) \geqslant \mathbf{1}_\psi(\tau) = 0$. Thus for any τ, $\liminf_n \mathbf{1}_{\psi_n}(\tau) \geqslant \mathbf{1}_\psi(\tau)$. Then by the Fatou lemma $\liminf_n P(\psi_n, O) \geqslant P(\psi, O)$. □

4.2. Proof of Theorem 4.1

Small sets We will need the following important definition.

Definition 4.1. Below ν will be any non-zero non-negative measure not necessarily probabilistic. The Borel subset $C \subset X$ is called (ν, n)-small (or simply small, if it is (ν, n)-small for some integer $n > 0$ and some ν) if for any $x \in C$ and any Borel set B

$$P^n(x, B) \geq \nu(B).$$

Lemma 4.3. *Assume* (A1). *Then for some* $n_1 \geqslant 1$ *and some* ν *there exists* (ν, n_1)*-small subset* $C \in \mathcal{B}(X)$ *such that* $\nu(C) > 0$.

Proof. It follows from the assumption (A1) that for any $n \geqslant m$ there is a measurable function $p_n(x, y)$ such that

$$P^n(x, B) = \int_B p_n(x, y)\mu(dy),$$

for any $x \in X,\ B \in \mathcal{B}(X)$. Moreover, $p_n(x, y) > 0$ for almost any $(x, y) \in X \times X$ (with respect to $\mu \times \mu$).

We will prove more, namely, that for some $n \geqslant m$ there exist sets $B_1, B_2 \in \mathcal{B}(X)$ of positive measure μ and a constant $\delta > 0$ such that for all $x \in B_1,\ y \in B_2$ we have

$$p_n(x, y) > \delta. \tag{4.2}$$

As the density $p_m(x, y)$ is measurable and almost everywhere positive, one can find a number $c > 0$ so that the sets

$$A = \{(x, y) \in X \times X : p_m(x, y) > c\},$$
$$\{(x, y, z) \in X \times X \times X : \ (x, y) \in A \text{ and } (y, z) \in A\} = (A \times X) \cap (X \times A)$$

have positive measures $\mu^2 = \mu \times \mu$ on $X \times X$ and $\mu^3 = \mu \times \mu \times \mu$ on $X \times X \times X$ correspondingly. Denote by $O_r(x) \subset X$ the open neighborhood of $x \in X$ of radius r, and put $O_r(x, y) = O_r(x) \times O_r(y) \subset X \times X$. By Lebesgue differentiation theorem there exists a set A_0 of zero measure $\mu \times \mu$ such that for any $(x, y) \in A \setminus A_0$

$$\lim_{r\to 0+} \frac{\mu^2(A \cap O_r(x, y))}{\mu^2(O_r(x, y))} = 1.$$

It follows that the set $((A \setminus A_0) \times X) \cap (X \times (A \setminus A_0))$ also has positive measure. Consider a point (x^*, y^*, z^*) in this set. Choose r so that the following inequalities hold:

$$\mu^2(A \cap O_r(x^*, y^*)) > \frac{3}{4}\mu^2(O_r(x^*, y^*)), \quad \mu^2(A \cap O_r(y^*, z^*)) > \frac{3}{4}\mu^2(O_r(y^*, z^*)).$$

For any $x \in X$ put

$$A_L(x) = \{y \in X : (x, y) \in A\}, \quad A_R(z) = \{y \in X : (y, z) \in A\}.$$

Define B_1 by

$$B_1 = \Big\{x \in O_r(x^*) : \mu(A_L(x) \cap O_r(y^*)) > \frac{3}{4}\mu(O_r(y^*))\Big\}.$$

Otherwise speaking, the set B_1 consists of the points $x \in O_r(x^*)$, for which the set $A_L(x)$ is sufficiently large inside $O_r(y^*)$. Let us show that B_1 has positive measure. Assume the contrary — that for almost all points of $O_r(x^*)$ the inequality $\mu(A_L(x) \cap O_r(y^*)) \leqslant (3/4)\mu(O_r(y^*))$ holds. Then by the Fubini theorem

$$\begin{aligned}\mu^2(A \cap O_r(x^*, y^*)) &= \int\limits_{O_r(x^*)} \mu(A_L(x) \cap O_r(y^*))\mu(dx) \\ &\leqslant \frac{3}{4}\mu(O_r(y^*))\mu(O_r(x^*)) = \frac{3}{4}(\mu \times \mu)(O_r(x^*, y^*)).\end{aligned}$$

That contradicts the choice of the points x^*, y^*. Thus, $\mu(B_1) > 0$. Similarly, one can show that the set

$$B_2 = \Big\{z \in O_r(z^*) : \mu(A_R(z) \cap O_r(y^*)) > \frac{3}{4}\mu(O_r(y^*))\Big\}$$

has positive measure μ. But from the definition of the sets $B_1,\ B_2$ it follows that for any points $x \in B_1,\ z \in B_2$ the following inequality holds

$$\mu(A_L(x) \cap A_R(z)) \geqslant \frac{1}{2}\mu(O_r(y^*)).$$

Also we have the estimates for the density p_{2m} and any $x \in B_1,\ z \in B_2$:

$$\begin{aligned}p_{2m}(x, z) &= \int\limits_X p_m(x, y)p_m(y, z)\mu(dy) \geqslant \int\limits_{A_L(x) \cap A_R(z)} p_m(x, y)p_m(y, z)\mu(dy) \\ &> c^2\mu(A_L(x) \cap A_R(z)) \geq \frac{c^2}{2}\mu(O_r(y^*)).\end{aligned}$$

Thus we have proved (4.2). Now we finish the proof of Lemma 4.3. Note that for any $x \in B_1$ and any $B \in \mathcal{B}(X)$ we have

$$P^n(x,B) = \int_B p_n(x,y)\mu(dy) \geqslant \int_{B\cap B_2} p_n(x,y)\mu(dy) \geqslant \delta\mu(B\cap B_2).$$

Moreover, as the sets B_1, B_2 have positive measure, there exist a subset $C \subset B_2$ of positive measure μ and a constant $\delta' > 0$ such that $P^m(x,B_1) > \delta'$ for all $x \in C$. It follows that for all $x \in C,\ B \in \mathcal{B}(X)$ the following inequalities hold:

$$P^{m+n}(x,B) = \int_X P^m(x,dy)P^n(y,B) \geqslant \int_{B_1} P^m(x,dy)P^n(y,B) \geqslant \delta'\delta\mu(B\cap B_2).$$

Thus, C is $(\nu, n{+}m)$-small, where $\nu(B) = \delta\delta'\mu(B\cap B_2)$, and moreover $\nu(C) > 0$. □

Lemma 4.4. *Let $C \in \mathcal{B}(X)$ be (ν, n_1)-small and assume that for any x from some set $D \in \mathcal{B}(X)$, $P^n(x,C) > \delta$ for some $\delta > 0$ and $n \geqslant 1$. Then the set D is $(\delta\nu, n+n_1)$-small.*

Proof. For any $x \in D$ and any $B \in \mathcal{B}(X)$, by semigroup property

$$\begin{aligned} P^{n+n_1}(x,B) &= \int_X P^n(x,dy)P^{n_1}(y,B) \geqslant \int_C P^n(x,dy)P^{n_1}(y,B) \\ &\geqslant P^n(x,C)\nu(B) \geqslant \delta\nu(B). \end{aligned}$$

□

Lemma 4.5. *Under assumptions (A1) and (A2) the set X itself is small.*

Proof. Consider a (ν, n_1)-small subset $C \in \mathcal{B}(X)$ of Lemma 4.3. For $k = 1, 2, \ldots$ introduce the subsets

$$A_k = \Big\{x \in X :\ P^m(x,C) > \frac{1}{k}\Big\},$$

where m is defined in (A1). From Lemma 4.4 we have that A_k is a $((1/k)\nu, m+n_1)$-small set. Let us prove that its closure is also a small set. Note that the measure ν is regular, that is for any Borel set B, $\nu(B) = \sup\{\nu(K)\}$, where the supremum is over all compact sets $K \subset B$. As $\nu(C) > 0$, there exists a compact set $K \subset C$ such that for all $x \in A_k$,

$$P^{m+n_1}(x,K) \geqslant \frac{1}{k}\nu(K) = \delta > 0.$$

Let $\{x_n\}_{n=1,2\dots} \in A_k$ and $x_n \to x$, then by semi-continuity of the transition probability we get:

$$P^{m+n_1}(x,K) = 1 - P^{m+n_1}(x, X \setminus K) \geqslant 1 - \liminf P^{m+n_1}(x_n, X \setminus K)$$
$$= \limsup P^{m+n_1}(x_n, K) \geqslant \delta.$$

Thus, for any $x \in \bar{A}_k$ we have $P^{m+n_1}(x,K) > \delta/2$, and applying Lemma 4.4 yields that $\bar{A}_k$ is a $((\delta/2)\nu, n_2)$-small set for some n_2.

But also by assumption (A1) $X = \cup_{k=1}^{\infty} A_k$. As X is not a countable union of sets which are nowhere dense, for some k there is an open subset $O \subset \bar{A}_k$. Thus for any $x \in X$ and any $B \in \mathcal{B}(X)$, we have

$$P^{m+n_2}(x,B) = \int\limits_X P^m(x,dy)P^{n_2}(y,A)$$
$$\geqslant \int\limits_O P^m(x,dy)P^{n_2}(y,B) \geqslant \frac{\delta}{2}\nu(B)P^m(x,O). \tag{4.3}$$

As the measures $P(x,\cdot)$ and μ are equivalent for any $x \in X$, we have $P^m(x,O) > 0$. But $P^m(x,O)$ is lower semi-continuous, and thus attains minimum on the compact X. It follows that $P^m(x,O) > \delta'$ for some $\delta' > 0$ and any $x \in X$. Using inequality (4.3), we get the proof. □

Proof of Theorem 4.1

For$A \in \mathcal{B}(X)$ and $n \geqslant 1$ denote

$$I_n(A) = \inf_{x \in X} P^n(x,A), \quad S_n(A) = \sup_{x \in X} P^n(x,A).$$

Note that

$$I_{n+1}(A) = \inf_{x \in X} \int\limits_X P(x,dy)P^n(y,A) \geqslant \inf_{x \in X} \int\limits_X P(x,dy)I_n(A) = I_n(A).$$

Thus, for fixed $A \in \mathcal{B}(X)$ the sequence $I_n(A)$ is non-decreasing. Similarly the sequence $S_n(A)$ is non-increasing. We shall prove that $S_n(A) - I_n(A)$ tends to zero as $n \to \infty$.

Take the number N and the measure ν as in Lemma 4.5. Then for any $n \geqslant 1$

$$P^{n+N}(x,A) = \int\limits_X P^N(x,dy)P^n(y,A)$$
$$= \int\limits_X (P^N(x,dy) - \nu(dy))P^n(y,A) + \int\limits_X \nu(dy)P^n(y,A).$$

As the measure $P^N(x,\cdot) - \nu(\cdot)$ is non-negative, we get from this equality that

$$\begin{aligned} P^{n+N}(x,A) &\geqslant I_n(A) \int\limits_X (P^N(x,dy) - \nu(dy)) + \int\limits_X \nu(dy) P^n(y,A) \\ &= (1-\nu(X)) I_n(A) + \int\limits_X \nu(dy) P^n(y,A) \end{aligned}$$

and then

$$I_{n+N}(A) \geqslant (1-\delta) I_n(A) + c_1, \quad \delta = \nu(X), \quad c_1 = \int\limits_X \nu(dy) P^n(y,A).$$

Similarly we get the upper bound for $S_{n+N}(A)$:

$$P^{n+N}(x,A) \leq (1-\nu(X)) S_n(A) + \int \nu(dy) P^n(y,A) \leq (1-\delta) S_n(A) + c_1$$

and

$$S_{n+N}(A) \leqslant (1-\delta) S_n(A) + c_1.$$

Then the difference is estimated as follows

$$S_{n+N}(A) - I_{n+N}(A) \leqslant (1-\delta)(S_n(A) - I_n(A)).$$

From the last inequality and monotonicity of the corresponding sequences it follows that for any $x \in X$ there exist the following limits and that they are equal

$$\lim_{n\to\infty} I_n(A) = \lim_{n\to\infty} S_n(A) = \lim_{n\to\infty} P_n(x,A) = \pi(A),$$

and moreover the convergence is uniform in $A \in \mathcal{B}(X)$ and in x. □

4.3. Proof of Theorem 2.2

We will use the following theorem (strong law of large numbers) for discrete time Markov chains on arbitrary state space X equipped with σ-algebra $\mathcal{A}$. Let $P^n(x,B)$ be n-step transition probability assumed to be measurable on X for any $B \in \mathcal{A}$ and a probability measure on $(X,\mathcal{A})$ for any x. Let us assume that there exists invariant measure π on $(X,\mathcal{A})$ such that uniformly in x

$$\sup_{A\in\mathcal{A}} |P^n(x,A) - \pi(A)| \to 0, \quad n\to\infty.$$

Denote by P_x the measure on trajectories $(x_0 = x, x_1, x_2, \ldots)$ with initial point x. Under these conditions the following assertion holds.

Theorem 4.2. *For any $f \in L^1(X,\pi)$ and any $x \in X$ we have P_x-a.s.*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f(x_k) = \int_X f(x)\pi(dx).$$

Proof. See [18], p. 140, and [16], p. 209.

To prove Theorem 2.2 we need the following lemma.

Lemma 4.6. *For any measurable bounded function f on $\mathcal{M}$ and any initial state $\psi(0) = \psi$ the following limit holds a.s.*

$$\lim_{N\to\infty} M_f(t_N) = \pi(f).$$

Proof. Denote by $X_k = (\psi_k, \tau_{k+1}),\ k = 0, 1, \dots$ the Markov chain with values in $X = \mathcal{M} \times \mathbb{R}_+$. Then

$$\int_{t_k}^{t_{k+1}} f(\psi(s))ds = \int_{t_k}^{t_{k+1}} f(e^{(s-t_k)A}\psi_k)ds = \int_0^{\tau_{k+1}} f(e^{sA}\psi_k)ds = F(X_k) \tag{4.4}$$

where

$$F(\psi, t) = \int_0^t f(e^{sA}\psi)ds,\ (\psi, t) \in X.$$

Then

$$M_f(t_N) = \frac{1}{t_N} \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} f(\psi(s))ds = \frac{1}{t_N} \sum_{k=0}^{N-1} F(X_k). \tag{4.5}$$

It is easy to show that X_k has invariant measure $\mu = \pi \times P_\tau, P_\tau = \rho ds$, which satisfies the relation stated in Theorem 4.2, since ψ_k satisfies it. Then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(X_k) = \mu(F) = \int_X F(\psi, s)d\mu$$

where

$$\mu(F) = \int_{\mathbb{R}_+} P_\tau(dt) \int_{\mathcal{M}} \pi(d\psi) \int_0^t ds f(e^{sA}\psi) = \int_{\mathbb{R}_+} P_\tau(dt) \int_0^t ds \int_{\mathcal{M}} \pi(d\psi) f(e^{sA}\psi)$$

$$= \pi(f) \int_{\mathbb{R}_+} P_\tau(dt) \int_0^t ds = \pi(f) E\tau_1.$$

Moreover, by strong law of large numbers for independent random variables τ_k we have

$$\lim_{N\to\infty} \frac{t_N}{N} = E\tau_1.$$

Then by (4.5) we get the proof of the lemma. □

To prove Theorem 2.2 we have to estimate the difference between $M_f(t)$ and $M_f(t_N)$. Using the boundedness $|f(\psi)| \leqslant c$ we have

$$|M_f(t)-M_f(t_N)| \leqslant \left|\frac{1}{t}\int_{t_N}^{t} f(\psi(s))ds|+\frac{|t-t_N|}{t}|M_f(t_N)\right| \leqslant \frac{|t-t_N|}{t}(c+|M_f(t_N)|). \tag{4.6}$$

For any $t > 0$ define the random index $N(t)$ so that

$$t_{N(t)} \leqslant t < t_{N(t)+1}$$

and note that

$$\frac{|t-t_{N(t)}|}{t} \leqslant \frac{\tau_{N(t)+1}}{t_{N(t)}} = \frac{\tau_{N(t)+1}}{\sum_{k=1}^{N(t)} \tau_k}.$$

As $E\tau_1 < \infty$, the law of large numbers, as $N \to \infty$, gives a.s.

$$\frac{\tau_{N+1}}{\sum_{k=1}^{N} \tau_k} \to 0.$$

But $N(t) \to \infty$ as $t \to \infty$. Then the right-hand side of (4.6) tends to 0 a.s. if $N = N(t)$ and $t \to \infty$. □

5. Appendix

5.1. Proof of Lemma 3.1

For any $t \geqslant 0$ one can show that

$$e^{tA} = \begin{pmatrix} \cos(t\sqrt{V}) & (\sqrt{V})^{-1}\sin(t\sqrt{V}) \\ -\sqrt{V}\sin(t\sqrt{V}) & \cos(t\sqrt{V}) \end{pmatrix},$$

where $\sqrt{V}$ is the positive square root of the matrix V. Then for any $k = 1, \dots, N$ and $t \geqslant 0$:

$$e^{tA}Q_k = \cos(\omega_k t)Q_k - \omega_k \sin(\omega_k t)P_k, \tag{5.1}$$

$$e^{tA}P_k = \frac{\sin(\omega_k t)}{\omega_k}Q_k + \cos(\omega_k t)P_k,. \tag{5.2}$$

Using (5.1)–(5.2) we have for any ψ

$$\psi(t) = e^{tA}\psi = \sum_{k=1}^{N}\Big(\cos(\omega_k t)\tilde{q}_k + \frac{\sin(\omega_k t)}{\omega_k}\tilde{p}_k\Big)Q_k + \sum_{k=1}^{N}\left(-\omega_k \sin(\omega_k t)\tilde{q}_k + \cos(\omega_k t)\tilde{p}_k\right)P_k$$

and then

$$\tilde{q}_k(t) = \cos(\omega_k t)\tilde{q}_k + \frac{\sin(\omega_k t)}{\omega_k}\tilde{p}_k, \quad \tilde{p}_k(t) = -\omega_k \sin(\omega_k t)\tilde{q}_k + \cos(\omega_k t)\tilde{p}_k.$$

From these two formulas we see that the pair of functions $(\tilde{q}_k(t), \tilde{p}_k(t)), k = 1, \dots, N$, corresponds to the dynamics of one-dimensional oscillator of unit mass and frequency ω_k, where $\tilde{q}_k(t)$ is the oscillator coordinate and $\tilde{p}_k(t)$ is its momentum. Thus the dynamics $e^{tA}\psi$ is isomorphic to the uniform movement on the torus with velocity $(\omega_1^2, \dots, \omega_N^2)$. This gives the second assertion.

5.2. Case $N = 1$

Note the identity for any ψ

$$Ie^{At_1}\psi = e^{-At_1}I\psi. \tag{5.3}$$

Let us first prove the proposition for $t = T_0$, $\psi = \psi_0 = (1, 0)$ and arbitrary ψ'. Define the one-to-one mapping $W : (0, (1/2)T_0) \to S$ as follows: for any $t_1 \in (0, (1/2)T_0)$ set

$$W(t') = \psi' = \exp\{A(T_0 - t_1)\}I\exp\{At_1\}\psi_0 = \exp\{A(T_0 - 2t_1)\}\psi_0.$$

Then it is sufficient to take $t' = (T_0 - t'')/2$ and choose minimal $t'' > 0$ so that

$$\psi' = e^{At''}\psi_0.$$

For $t > T_0$ the proof is quite similar but we will not need this case to prove convergence.

5.3. Mixing subspace

Here we give some properties of the mixing space. The following two lemmas show how the dimension of L_-can be explicitly characterized.

Lemma 5.1. *The space L_- is invariant with respect to A. Moreover*

$$L_- = \langle\{A^k g_1 : \ k = 0, 1, \dots\}\rangle$$

where $\langle\ \rangle$ is the linear span of the set of vectors.

Consider also the orthogonal complement to L_- in the scalar product $(,)_2$,

$$L_0 = L_-^{\perp}.$$

Then it is also invariant with respect to A. Moreover, the vector $\psi \in L_0$ iff for the hamiltonian dynamics with initial condition $\psi = \psi(0)$ the momentum $p_1(t) = 0$ for any t.

Lemma 5.2. *Assume that the spectrum of V is simple, and let $\{v_1, \dots, v_N\}$ be the eigenvectors of V, they form a basis in R^N. Then the dimension of L_0 is twice the number of v_k having coordinates $v_{k,1} = (e_1, v_k) = 0$.*

On what occurs if condition (2.3) is not fulfilled, for exact formulas for the dimension of L_- for the chain of harmonic oscillators and for other cases see [9–11].

5.4. Proof of Lemma 3.2

We have

$$|\tilde{p}_k^*| = |\beta_k| \left| y - c\frac{1-\gamma_k^2}{4y} \right| = |\beta_k| \left| \frac{2\gamma_n^2 + 2\gamma_n\sqrt{\gamma_n^2 + c(1-\gamma_n^2)} + c(\gamma_k^2 - \gamma_n^2)}{4y} \right|.$$

As for any $k = 1, \dots, N$ we have $\gamma_k \geqslant \gamma_n \geqslant 0$, the expression under module in the last formula is non-negative, and we have

$$|\tilde{p}_k'| = |\beta_k| \Big(y - c\frac{1-\gamma_k^2}{4y} \Big).$$

Consider two cases:

1. $\gamma_k \leqslant 1$. Show that $f^+(x)$ is monotone increasing, that is, its derivative

$$(f^+(x))' = \frac{1}{2}\left(1 + \frac{x(1-c)}{\sqrt{x^2 + c(1-x^2)}}\right) > 0.$$

Thus $y = f^+(\gamma_n) \leqslant f^+(\gamma_k)$. Taking into account $\gamma_k \leqslant 1$, we have the inequalities

$$\begin{aligned} |\tilde{p}_k'| &= |\beta_k| \Big(y - c\frac{1-\gamma_k^2}{4y} \Big) \leqslant |\beta_k| \Big(f^+(\gamma_k) - \frac{c(1-\gamma_k^2)}{4f^+(\gamma_k)} \Big) \\ &= |\beta_k| \frac{2\gamma_k^2 + 2\gamma_k\sqrt{\gamma_k^2 + c(1-\gamma_k^2)}}{4f^+(\gamma_k)} = |\beta_k|\gamma_k = r_k. \end{aligned}$$

2. $\gamma_k > 1$. Then we will show that $f^-(\gamma_k) \leqslant y$. It is easy to check that $f^-(x)$ is increasing and hence $0 < f^-(\gamma_k) \leqslant f^-(\gamma_N)$. Note that the left inequality holds because $f^-(1) = 0$.

Let us prove that for any $x \in \mathbb{R}$ we have $f^{-}(x) < (c/2)x$. We have

$$f^{-}(x) - \frac{c}{2}x = \frac{1}{2}\left((1-c)x - \sqrt{x^2(1-c)+c}\right)$$
$$= \frac{1}{2}\left(\frac{-c - x^2(1-c)c}{(1-c)x + \sqrt{x^2(1-c)+c}}\right) < 0.$$

Thus we have proved that

$$f^{-}(\gamma_k) \leqslant f^{-}(\gamma_N) < \frac{c}{2}\gamma_N. \tag{5.4}$$

By $f^{+}(0) = \sqrt{c}/2$ and the evident inequality $(f^{+}(x))' \geqslant c/2$, which holds as $c \leqslant 1$, we have

$$y = f^{+}(\gamma_n) \geqslant \frac{1}{2}(c\gamma_n + \sqrt{c}). \tag{5.5}$$

Then by (5.4), (5.5) and $c \leqslant (\gamma_N - \gamma_n)^{-2}$ we get

$$y - f^{-}(\gamma_k) \geqslant \frac{1}{2}(-c(\gamma_N - \gamma_n) + \sqrt{c}) = \frac{\sqrt{c}}{2}\left(-\sqrt{c}(\gamma_N - \gamma_n) + 1\right) \geqslant 0.$$

Remind that we have proved earlier that the function $f^{+}(x)$ is increasing, then taking into account the latter inequality we have the following bounds

$$|\tilde{p}_k^*| = |\beta_k|\Big(y - c\frac{1-\gamma_k^2}{4y}\Big) = |\beta_k|\Big(y + c\frac{\gamma_k^2 - 1}{4y}\Big) \leqslant |\beta_k|\Big(f^{+}(\gamma_k) + c\frac{\gamma_k^2 - 1}{4f^{-}(\gamma_k)}\Big)$$
$$= |\beta_k|\Big(\frac{\gamma_k^2 - \gamma_k^2 - c(1-\gamma_k^2) + c(\gamma_k^2 - 1)}{4f^{-}(\gamma_k)}\Big) = |\beta_k|\, c\frac{\gamma_k^2 - 1}{2f^{-}(\gamma_k)}$$
$$= |\beta_k| c\frac{2(\gamma_k^2 - 1)f^{+}(\gamma_k)}{2c(\gamma_k^2 - 1)} = |\beta_k| f^{+}(\gamma_k) \leqslant |\beta_k|\gamma_k = r_k.$$

In the last inequality we used the fact that $f^{+}(x) < x$ for $x > 1$, which holds due to $f^{+} + f^{-} = x$. □

5.5. Advertisement concerning non-linear situation

Here we summarize our intuition of general problems concerning ergodicity. We think that the difference between linear and non-linear hamiltonians is overexaggerated.

1. We think that without external influence "generic" N-particle hamiltonians will give non-ergodic system (Problem 1). More interesting question is what means "generic" and how the foliation on invariant subsets looks like in the "generic" case; we do not know the answer.

2. Possibly a bit easier problem — ergodicity in the "generic" case with most minimal random influence (Problem 2). One can think that Boltzmann did

not speak on rigorous mathematical language and did not make any difference between closed system and systems with most minimal random influence. It is funny that we did not see earlier attempts to formulate it like this.

3. One could give many examples or even classes of non-linear hamiltonians for which Problem 2 could be solved. But it is not clear whether such very time consumable activity deserves big efforts. The result can be positive — that there will be generic ergodicity for some class of non-linear hamiltonians — but this will not give any certitude concerning Problem 1. Or the result could be negative — one could imagine in this case that this class is far from the "generic" situation.

4. The authors think that after more than 100 years of hard work the strategy should be changed. Namely, some active organizer should better consolidate a goal-oriented team for the "generic" problem.

References

[1] V. Arnold (1989) *Mathematical Methods of Classical Mechanics.* Springer.

[2] D. Anosov and E. Zhuzhoma (2012) Closing lemmas. *Differ. Equ.* **48** (13), 1653–1699.

[3] J. Fritz, T. Funaki and J. Lebowitz (1994) Stationary states of random Hamiltonian systems. *Probab. Theory and Relat. Fields* **99**, 211–236.

[4] C. Bernardin, V. Kannan, J. Lebowitz and J. Lukkarinen (2012) Harmonic systems with bulk noises. *J. Stat. Phys.* **146**, 800–831.

[5] C. Bernardin, V. Kannan, J. Lebowitz and J. Lukkarinen (2011) Nonequilibrium stationary states of harmonic chains with bulk noises. *Eur. J. Phys. B* **84** (4), 685–689.

[6] C. Bernardin and S. Olla (2011) Transport properties of a chain of anharmonic oscillators with random flip of velocities. *J. Stat. Phys.* **145**, 1224–1255.

[7] J. Lukkarinen (2014) Thermalization in harmonic particle chains with velocity flips. *J. Stat. Phys.* **155**, 1143–1177.

[8] M. Simon (2013) Hydrodynamic limit for the velocity-flip model. *Stoch. Process. Appl.* **123**, 3623–3662.

[9] A. Lykov and V. Malyshev (2013) Convergence to Gibbs equilibrium — unveiling the mystery. *Markov Processes Relat. Fields* **19** (4), 634–666.

[10] A. Lykov, V. Malyshev and S. Muzychka (2013) Linear hamiltonian systems under microscopic random influence. *Theory Probab. and Appl.* **57** (4), 684–688.

[11] A. Lykov and V. Malyshev (2012) Harmonic chain with weak dissipation. *Markov Processes Relat. Fields* **18** (4), 721–729.

[12] Yu. Kifer (1988) *Random Perturbations of Dynamical Systems.* Birkhauser, Boston-Basel.

[13] S. Ponomarev (1987) Submersions and pre-images of sets of zero measure. *Sib. Math. J.* **28** (1), 199–210.

[14] R. Azais, J. Bardet, A. Genadot, N. Krell and P. Zitt (2014) Piecewise deterministic Markov process — recent results. *ESAIM: Proceedings* **44**, 276–290.

[15] S. Meyn and R. Tweedie (2009) *Markov Chains and Stochastic Stability.* Cambridge.

[16] N. Portenko, A. Skorohod and V. Shurenkov (1989) *Markov Processes.* Itogi nauki i tehniki, VINITI, Moscow.

[17] S. Orey (1971) *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand, London.

[18] D. Revuz (1984) *Markov Chains.* NorthHolland.

[19] C. Villani (2011) *Particle Systems and Nonlinear Landau Damping.* Lecture Notes, `http://cedricvillani.org/`.

[20] N. Simanyi (2009) Conditional proof of the Boltzmann – Sinai ergodic hypothesis. *Invent. Math.* **177**, 381–413.

[21] J.-P. Eckmann, C.-A. Pillet and L. Rey-Bellet (1999) Entropy production in non-linear, thermally driven hamiltonian systems. *J. Stat. Phys.* **95** (1), 305–331.

[22] L. Rey-Bellet and L. Thomas (2002) Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Commun. Math. Phys.* **225**, 305–329.

[23] L. Rey-Bellet (2003) Statistical mechanics of anharmonic lattices. In: *Advances in Differential Equations and Mathematical Physics*, AMS Contemporary Mathematics Series **327**, 283–298.

[24] J.-P. Eckmann (2002) Non-equilibrium steady states. In: *Proceedings of the International Congress of Mathematicians*, Beijing, Higher Education Press **3**, 409–418.