On statistical properties of chaotic dynamical systems

N.I. Chernov
Department of Mathematics
University of Alabama at Birmingham
Birmingham, AL 35294
chernov@vorteb.math.uab.edu

September 14, 2006

1 Introduction

This paper is devoted to the methods of investigating statistical properties of chaotic dynamical systems. By statistical properties we mean the rate of the decay of correlations, the central limit theorem and other probabilistic limit theorems. Surveys of known results in this area may be found in [3, 9, 8].

An effective method of proving these properties is based on Markov approximations to dynamical systems. This approach is an alternative to the conventional Perron-Frobenius operator techniques. It was Ya. Sinai and his school [14, 15, 4, 5, 6] who systematically developed the methods of Markov partitions, Markov symbolic dynamics and measure-theoretic Markov approximations to Anosov diffeomorphisms and billiards. Bowen [2] extended these methods to Smale's Axiom A diffeomorphisms.

A general construction of Markov chains approximating discrete-time dynamical systems was introduced by Bunimovich and Sinai in [4, 5] and later studied in [6, 8]. It was shown that the chaotic behavior of the dynamical system ensures special conditions on transition probabilities of the approximating Markov chain. In turn, under those conditions one can prove probabilistic limit theorems and establish strong bounds on correlation functions for the original dynamical system – some general results are displayed in [8].

Here we continue studying the techniques of Markov approximations to dynamical systems. We explain how approximating Markov chains can be constructed for dynamical systems with discrete and continuous time. We also introduce a brand new condition on the transition probabilities of Markov chains that implies strong bounds on the correlations. Our new condition is weaker than all those studied earlier in [4, 6, 8].

2 Markov approximations for dynamical systems

A discrete-time dynamical system is a measurable transformation $T : M \to M$ of a measurable space M preserving a probability measure μ.

Let $\mathcal{A} = \{A_i\}$ be a finite or countable measurable partition of the space M into subsets of positive measure. By Markov approximation for the map T we mean a probabilistic stationary Markov chain, whose transition probabilities are

$$\pi_{ij} = \mu(T^{-1}A_j/A_i) = \mu(T^{-1}A_j \cap A_i)/\mu(A_i) \tag{1}$$

and whose stationary distribution is

$$p_i = \mu(A_i). \tag{2}$$

This definition of Markov approximations for arbitrary measure-preserving transformations was introduced in [8]. It is one of possible implementations of the idea of 'coarse-graining' of the phase space popular among physicists (see, e.g., [13]). This definition is also very close to Ulam's construction [16] of Markov chains approximating interval maps.

The 'discrepancy' of the Markov approximation generated by the Markov chain (1)-(2), within N iterates of the map T, is measured by the following quantity:

$$\nu_N := \sup_{n\le N} \sum_{i_0,\dots,i_n} |\mu(T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}} \cap \dots \cap A_{i_0}) - \mu(T^{-1}A_{i_n}/A_{i_{n-1}})|$$

$$\times \mu(T^{-(n-1)}A_{i_{n-1}} \cap \dots \cap A_{i_0}) \tag{3}$$

Here and further on $\mu(A/B)$ means the conditional measure, $\mu(A \cap B)/\mu(B)$, and we always set it to zero whenever $\mu(B) = 0$. The quantity ν_N measures how close (better to say, how distant!) the 'long-memory' and 'short-memory' conditional distributions are within the first N iterates of T.

Recall that given two probability distributions $P = \{p_i\}$ and $Q = \{q_i\}$ on the same index set $\{i\}$, the distance in variation between P and Q is defined to be

$$\mathrm{Var}\,(P,Q) = \frac{1}{2}\sum_i |p_i - q_i|. \tag{4}$$

Now (3) estimates twice the mean distance in variation between the long- and short-memory conditional distributions on $\{A_i\}$.

By means of (3) one can estimate how the finite dimensional distributions of the Markov chain,

$$p_{i_0 i_1 \cdots i_n} = p_{i_0}\pi_{i_0 i_1}\cdots\pi_{i_{n-1}i_n} \tag{5}$$

are close to those of the dynamical system in the variational metric (4). It is shown in [8] that

$$\sum_{i_0,\dots,i_n} |\mu(T^{-n}A_{i_n} \cap \dots \cap A_{i_0}) - p_{i_0 i_1 \cdots i_n}| \le (n-1)\nu_N \tag{6}$$

for any $n \le N$.

We now explain how Markov approximations with good properties can be constructed. Let (M, T, μ) be an Anosov diffeomorphism and μ a smooth invariant measure. Let $\mathcal{B}$ be a Markov partition of M into sufficiently small rectangles [2, 14]. Fix a large integer $K > 0$ and take $\mathcal{A} = \vee_{-K}^{K} T^k \mathcal{B}$. Then $\mathcal{A}$ is another Markov partition, whose atoms are exponentially (in K) small. Precisely, there are constants $c_i, a_i > 0$ depending on the system (M, T, μ) such that for any $A \in \mathcal{A}$ we have $c_1 e^{-a_1 K} \le \operatorname{diam} A \le c_2 e^{-a_2 K}$ and $c_3 e^{-a_3 K} \le \mu(A) \le c_4 e^{-a_4 K}$. We do not go into detail, but it is a standard argument that on any $A \in \mathcal{A}$ there is a product measure μ_A^p which approximates μ to the following degree of accuracy:

$$\left| \frac{d\mu_A^p}{d\mu}(x) - 1 \right| \le c_5 e^{-a_5 K} \tag{7}$$

for every $x \in A$. Now, for any $n \ge 0$ and any atoms $A_{i_0}, \dots, A_{i_n} \in \mathcal{A}$ we have

$$\mu^p_{A_{i_{n-1}}}(T^{-n} A_{i_n} / T^{-(n-1)} A_{i_{n-1}} \cap \cdots \cap A_{i_0}) = \mu^p_{A_{i_{n-1}}}(T^{-1} A_{i_n} / A_{i_{n-1}})$$

which follows directly from the Markov property of the partition $\mathcal{A}$. It is now an immediate consequence of (7) that for the partition $\mathcal{A}$ we have $\nu_N \le 2c_5 e^{-a_5 K}$ for all $N > 0$. This approximation has the following advantage: ν_N is exponentially (in K) small, so that, according to (6), the finite-dimensional distributions of the Markov chain and those of the dynamical system stay exponentially (in K) close on very long intervals of time, $(0, N)$, at least for $N \approx e^{aK}$ with any $a < a_5$.

If the dynamical system (M, T, μ) is a smooth hyperbolic system with singularities and the measure μ is a Sinai-Bowen-Ruelle measure [15] (not necessarily absolutely continuous), the construction of partitions A with the above properties goes through, with some technical modifications, see [6, 1].

Lastly, Markov approximations can be constructed for dynamical systems with continuous time (flows). It is common to study flows by their special representations, which are called suspension flows or Kakutani flows, as defined below.

Let (M, T, μ) be a discrete time dynamical system and $l(x)$ a positive integrable function on M. A suspension flow build under the function $l(x)$ (this is called the ceiling function) is defined on the measurable space $\mathcal{M} = \{(x, s) : x \in M,\ 0 \le s < l(x)\}$ by the rule

$$\Phi^t(x, s) = \begin{cases} (x, s + t) & \text{for } \ 0 \le t < l(x) - s \\ (Tx, s + t - l(x)) & \text{for } \ l(x) - s \le t < l(Tx) + l(x) - s \end{cases} \tag{8}$$

This flow is measurable and preserves the probability measure μ_f on $\mathcal{M}$ defined by $d\mu_f = c \cdot d\mu \times ds$, where $c^{-1} = \int_M l(x) d\mu(x)$ is the normalizing factor.

Now let $\mathcal{A}$ be a partition of M generating a Markov approximation for T. Then we can construct a Markov approximation to the flow Φ^t in the following way. First, let $\bar{l}(x)$ be the ceiling function l conditioned on the partition $\mathcal{A}$. Fix then a small $\delta > 0$ (a 'quantum of time') and set

$$\hat{l}(x) = \hat{l}_{\delta, \mathcal{A}}(x) := ([\bar{l}(x)/\delta] + 2)\delta \tag{9}$$

where $[a]$ stands for the integral part of a real number a. Now, consider another suspension flow, $\hat{\Phi}^t$, over T, build under the function $\hat{l}$. Its phase space, $\hat{\mathcal{M}} = \{(x,s) : x \in M, 0 \leq s < \hat{l}(x)\}$, is naturally partitioned into the following blocks

$$X = A \times [k\delta, (k+1)\delta), \ A \in \mathcal{A}, \ k = 0, 1, \dots \hat{l}(A)/\delta - 1$$

Denote this partition by $\hat{\mathcal{A}}$. Now consider the map $\hat{T} = \hat{\Phi}^\delta$ on $\hat{M}$. This map moves every atom of $\hat{\mathcal{A}}$ exactly onto another atom above it, and the top atoms of this partition are broken by $\hat{T}$ to pieces and transformed down to the bottom of $\hat{\mathcal{A}}$, according to the action of T on M.

The Markov chain approximating the map $\hat{T} = \hat{\Phi}^\delta$ is constructed by the same rules as before: its transition probabilities are $\hat{\pi}_{ij} = \hat{\mu}(X_j/\hat{T}X_i)$ and its stationary distribution is $\hat{p}_i = \hat{\mu}(X_i)$, where X_i, X_j are atoms of $\hat{\mathcal{A}}$ and $\hat{\mu}$ stands for the invariant measure of the suspension flow $\{\hat{\Phi}^t\}$. Since the map $\hat{T}$ acts very straightforwardly in the bulk of the partition $\hat{\mathcal{A}}$, this Markov chain provides very good approximation to the dynamical system $(\hat{\mathcal{M}}, \hat{T}, \hat{\mu})$. In fact, if the function l is bounded away from zero and infinity, $0 < l_{\min} \leq l(x) \leq l_{\max} < \infty$, then the quantity $\hat{\nu}_N$ defined by the rule (3) for the Markov chain $||\hat{\pi}_{ij}||, ||\hat{p}_i||$, satisfies

$$\hat{\nu}_N \leq \text{const} \cdot \delta \nu_{[bN\delta]}$$

where ν is the quantity (3) for the Markov chain approximating the map T and generated by $\mathcal{A}$. Here $b > 0$ is a constant depending on the original flow $\{\Phi^t\}$ alone. We do not prove the above bound here.

3 Mixing coefficients in Markov chains

In the previous section we have shown how Markov chains approximating dynamical systems can be constructed.

Consider now an abstract homogeneous Markov chain with a finite number of states, which approximates a dynamical system. We denote the sates by $1, 2, \dots, I$, the matrix of transition probabilities by $\Pi = ||\pi_{ij}||$ with $1 \leq i, j \leq I$ and the stationary distribution by $P = ||p_i||$. We denote by $\pi_{ij}^{(m)}$ the m-step transition probabilities, i.e. $||\pi_{ij}^{(m)}|| = \Pi^m$. We denote by $\mathcal{J}$ the set of indices $\{1, 2, \dots, I\}$.

In order to establish statistical properties for the underlying dynamical system, one usually has to bound the following quantities, which we call the mixing coefficients. For any $m \geq 1$ let

$$V_i^{(m)} = \frac{1}{2} \sum_{j=1}^{I} |\pi_{ij}^{(m)} - p_j|$$

and

$$V^{(m)} = \sum_{i=1}^{I} p_i V_i^{(m)} \tag{10}$$

This last quantity is the mean distance in variation between the m-step transition probabilities and the stationary distribution.

If the dynamical system is ergodic (mixing), then the approximating Markov chain is irreducible (aperiodic). For such chains, the mixing coefficient (10) monotonically decreases to zero as $m \to \infty$. The rate of the decay of this coefficient essentially represents the mixing rates of the original dynamical system. Moreover, it is possible to prove various statistical properties of the dynamical system based on available bounds on the mixing coefficients (10). Such proofs were developed in [8].

A far more difficult problem is to establish any bounds on the mixing coefficients (10) for Markov chains approximating dynamical systems. Dynamical properties of the system (e.g., hyperbolicity) seldom can provide such bounds directly. However, there are certain conditions on the transition probabilities of the Markov chain under which the coefficients (10) can be effectively bounded. It is also possible to verify those conditions by using the dynamical properties of the underlying system, such as hyperbolicity, the existence of Markov partitions, etc. Several implementations of this strategy were described in [8]. The rest of this section is devoted to two conditions on the transition probabilities used in [8].

One of them is the so called Doeblin condition:

$$d := 1 - \max_{i,j} \frac{1}{2} \sum_{k=1}^{I} |\pi_{ik}^{(s)} - \pi_{jk}^{(s)}| > 0 \tag{11}$$

for some $s \geq 1$. This condition is motivated by the classical D-condition [11] and Dobrushin's coefficient of ergodicity [10]. It was explicitly introduced by Bunimovich and Sinai [4, 5] and later used in [1]. This condition implies the bound

$$V_i^{(m)} \leq (1 - d/2)^{[m/s]}$$

for any $m \geq 1$ and all $i \in \mathcal{J}$, see proofs in [1, Lemma 15] and [8, Proposition 5.3].

The second condition is

$$r := \min_{i,j} \frac{\pi_{ij}^{(s)}}{p_j} > 0 \tag{12}$$

for some $s \geq 1$. It was motivated by Ibragimov's regularity of stationary random processes [12] and first explicitly introduced in [6]. It was later used in [7]. This condition implies the bound

$$V_i^{(m)} \leq (1 - r)^{[m/s]} \tag{13}$$

for any $m \geq 1$ and all $i \in \mathcal{J}$, see proofs in [6, Lemma 4.3] and in Section 4 here.

The meaning of conditions (11) and (12) for the dynamics of the underlying system is the following. According to (12), the s-th image of any atom $A_i \in \mathcal{A}$ intersects all the atoms of $\mathcal{A}$ and the conditional distribution on atoms $A \in \mathcal{A}$ (conditioned on $T^s A_i$) recovers a certain fraction of the invariant measure. This is a very stringent condition. It is, however, possible to verify it for some uniformly hyperbolic maps with smooth invariant measure [6, 8]. The condition (11) means that the s-th images of any two atoms $A_i, A_j \in \mathcal{A}$ are so close to each other that they intersect the same other atoms of $\mathcal{A}$ and the conditional distributions (conditioned on $T^s A_i$ and $T^s A_j$) are close to each other in the variational metric. This is a weaker condition than (12) (it does not require that $T^s A_i$ or $T^s A_j$ intersect all the atoms of $\mathcal{A}$), but it is still pretty stringent. It is verifiable for some hyperbolic attractors with Bowen-Ruelle-Sinai invariant measures [1].

4 A new condition on transition probabilities

In this section we introduce a new condition on the transition probabilities of the Markov chain, which is weaker than the two discussed above and provides good bounds on the mixing coefficients (10). Our new condition is designed to be verifiable in the case of Markov approximations to hyperbolic flows, in particular, geodesic flows on surfaces of negative curvature. A verification of this condition is, however, beyond the scopes of this paper.

We use the notations of the previous section. Denote $p_{\min} = \min_i p_i$. For every $i, j \in \mathcal{J}$ let

$$b_{i,j} = \sum_{k=1}^{I} \frac{\pi_{ik}\pi_{jk}}{p_k}$$

Theorem 1 *Suppose that the Markov chain satisfies the following condition:*

$$b = \min_{i,j} b_{i,j} > 0 \tag{14}$$

Then for any $m \geq 1$ and all $i \in \mathcal{J}$ we have

$$V_i^{(m)} \leq 50 b^{-1/2} p_{\min}^{-1} \cdot (1 - b/2)^{m/3} \tag{15}$$

and thus

$$V^{(m)} \leq 50 b^{-1/2} p_{\min}^{-1} \cdot (1 - b/2)^{m/3}$$

Remark. If the condition (14) is satisfied for the s-step transition probabilities $\pi_{ij}^{(s)}$ instead of π_{ij}, then Theorem 1 remains true with the exponent $m/3$ replaced by $[m/s]/3$ in the above bounds.

We will first compare our condition (14) to the two conditions described in the previous section.

Lemma 2 *The regularity (12) implies the Doeblin condition (11) with $d \geq r$, and the latter (with $s = 1$) implies our condition (14) with $b \geq d^2$.*

Proof. Without loss of generality we set $s = 1$. The Doeblin condition (11) is equivalent to

$$d = \min_{i,j} \left\{ \sum_{k=1}^{I} \min\{\pi_{ik}, \pi_{jk}\} \right\} > 0 \tag{16}$$

Clearly, (12) implies (16) with $d \geq r$. We now show that (16) implies (14), with the help of Schwarz' inequality:

$$\begin{aligned} b_{i,j} &= \sum_{k=1}^{I} \frac{\pi_{ik}\pi_{jk}}{p_k} \geq \sum_{k=1}^{I} \left(\frac{\min\{\pi_{ik}, \pi_{jk}\}}{p_k} \right)^2 \cdot p_k \\ &\geq \left(\sum_{k=1}^{I} \frac{\min\{\pi_{ik}, \pi_{jk}\}}{p_k} \cdot p_k \right)^2 \geq d^2 \end{aligned}$$

The lemma is proved.

As Lemma 2 shows, our condition (14) is the weakest one among these three conditions.

We now prove Theorem 1. First, we make an additional, simplifying assumption that the stationary distribution P is uniform, i.e., $p_i = 1/I$ for all $1 \leq i \leq I$. In this case the matrix Π is a doubly stochastic one, i.e. its transpose Π^T is a stochastic matrix also. Consider the matrix $\tilde{\Pi} = \Pi\Pi^T$. It is a symmetric and doubly stochastic matrix with the same uniform stationary distribution P. We denote its components by $||\tilde{\pi}_{ij}||$. Now we can rewrite the condition (14) as follows:

$$b = \min_{i,j} \frac{\tilde{\pi}_{ij}}{p_j} > 0 \tag{17}$$

for any $i, j \in I$. Note that this is exactly the regularity condition (12) applied to the stochastic matrix $\tilde{\Pi}$, with r replaced by b.

Due to (17) the operator $\tilde{\Pi}$ is a contraction on the simplex of the probability distributions equipped with the distance in variation (4), whose only 'fixed point' is the distribution P, i.e.

$$\mathrm{Var}\,(P'\tilde{\Pi}, P) \leq (1 - b)\mathrm{Var}\,(P', P). \tag{18}$$

for any distribution $P' = ||p'_j||$. To show this, we denote by Σ_j^+ the summation over such j that $(P'\tilde{\Pi})_j > p_j$. Then

$$\mathrm{Var}\,(P'\tilde{\Pi}, P) = \Sigma_j^+ (\Sigma_i p'_i \tilde{\pi}_{ij} - \Sigma_i p_i \tilde{\pi}_{ij})$$

$$= \Sigma_i (p'_i - p_i) \Sigma_j^+ \tilde{\pi}_{ij} = \Sigma_i (p'_i - p_i) \Sigma_j^+ (\tilde{\pi}_{ij} - b/I).$$

Now denote by Σ_i^+ the summation over such i that $p'_i > p_i$. Since $\tilde{\pi}_{ij} - b/I \geq 0$ for all $i, j \in \mathcal{J}$ (this follows from (17)) the RHS of the last equation is bounded above by

$$\Sigma_i^+ (p'_i - p_i) \Sigma_j^+ (\tilde{\pi}_{ij} - b/I) \leq \Sigma_i^+ (p'_i - p_i) \Sigma_j (\tilde{\pi}_{ij} - b/I) \leq (1 - b) \cdot \mathrm{Var}\,(P', P).$$

The bound (18) is proved.

Exploiting the estimate (18) m times in a row (in an obvious way) yields

$$\mathrm{Var}\,(P'\tilde{\Pi}^m, P) \le (1-b)^m \mathrm{Var}\,(P', P) \tag{19}$$

In particular, the bound (13) follows from (19) and the inequality $2^{-1}\sum_j |\tilde{\pi}_{ij} - p_j| \le 1-b$, which is a consequence of (17).

We now examine the spectrum of the matrix $\tilde{\Pi}$. Since it is a symmetric matrix, all its eigenvalues are real and its eigenvectors are mutually orthogonal. P is an eigenvector with eigenvalue one. Let V be an eigenvector of $\tilde{\Pi}$ different from P, with an eigenvalue λ'. If $||V||$ is small enough, then $P' := P + V$ is a probability distribution. Then $P'\tilde{\Pi}^m = P + V\tilde{\Pi}^m = P + (\lambda')^m V$. Applying the estimate (19) gives

$$|\lambda'|^m \cdot \sum |v_i| \le 2\mathrm{Var}\,(P', P)(1-b)^m$$

and so $|\lambda| \le 1-b$. Therefore, all the eigenvalues of the matrix $\tilde{\Pi}$ but one lie in the interval $[-1+b, 1-b]$.

In all that follows we denote by U the uniform distribution, i.e. $U = (1/I, \dots, 1/I)$. (We assumed that $P = U$, but later we remove this assumption.) Denote the hyperplane in $\mathbb{R}^I$ perpendicular to U by L_0. It is parallel to the simplex made by probability distributions, in which U is a center.

We now turn to the matrix Π. Fix an $i \in \mathcal{J}$. The ith row of the matrix Π^m is the vector $||\pi_{ij}^{(m)}||$, $1 \le j \le I$. It equals $E_i\Pi^m$, where $E_i = (0, \dots, 0, 1, 0, \dots, 0)$ is a unit row vector with its ith component equal to 1. The vector $E_i - P$ belongs to the subspace L_0, and this subspace is left invariant under both $\tilde{\Pi}$ and Π. As shown above, the restriction of $\tilde{\Pi}$ on L_0 is a contraction, with all the eigenvalues lying in $[-1+b, 1-b]$. Hence, for any vector $V \in L_0$ we have $||V\tilde{\Pi}|| \le (1-b)||V||$, where $||\cdot||$ is the Euclidean norm. Hence

$$||V\Pi||^2 = (V\Pi)\cdot(V\Pi) = V\Pi\Pi^T V^T = (V\tilde{\Pi})\cdot V \le (1-b)||V||^2,$$

where the dot $(\cdot)$ stands for the scalar product of two row-vectors in $\mathbb{R}^I$. Therefore, $||(E_i - P)\Pi^m||^2 \le (1-b)^m ||E_i - P|| < (1-b)^m$. As a result,

$$V_i^{(m)} = \frac{1}{2}\sum_{j=1}^{I} |\pi_{ij}^{(m)} - p_j| \le \frac{1}{2}\sqrt{I}\cdot||(E_i - P)\Pi^m|| \le 2^{-1}p_{\min}^{-1/2}\cdot(1-b)^{m/2} \tag{20}$$

Hence, we get Theorem 1 under the additional assumption that P is uniform.

Remark. In our calculations we have never used the fact that the components of the matrix Π were positive. So, the bound (20) is still true if the matrix Π in Theorem 1 is a so-called quasi-stochastic one, i.e., the one whose components in each row sum to one but are not necessarily nonnegative. (In other words, a matrix Π' is quasistochastic iff $\Pi' U^T = U^T$.) Of course, we have to assume, additionally, that the stationary distribution P is uniform: $P = U$.

We now extend the above simplified version of Theorem 1 to its full version. The idea of the extension is to make the stationary distribution P uniform by an appropriate refinement and approximation.

By a refinement of a Markov chain we mean the splitting of each state $i \in \mathcal{J}$ into a number, $I_i \geq 1$, of 'equal' fragments. Precisely, we replace each state $i \in \mathcal{J}$ by a collection of I_i states labeled by i_r, $1 \leq r \leq I_i$. We then define a new Markov chain with the states i_r, $1 \leq r \leq I_i$ and $1 \leq i \leq I$. The total number of states is now $I' = I_1 + \cdots + I_I$. We denote the collection of new staes by $\mathcal{J}'$. The transition probabilities are defined by $\pi'_{i_r j_s} = \pi_{ij}/I_j$ for every i_r and j_s. These form a matrix $\Pi' = ||\pi'_{i_r j_s}||$ of size $I' \times I'$. The stationary distribution is $P' = ||p'_{i_r}||$ with $p'_{i_r} = p_i/I_i$.

Clearly, for any i_r and j_s we have

$$b'_{i_r,j_s} := \sum_{k_t \in \mathcal{J}'} \frac{\pi'_{i_r k_t} \pi'_{j_s k_t}}{p'_{k_t}} = b_{i,j} \tag{21}$$

so that $b' := \min_{i_r,j_s} b'_{i_r,j_s} = b$. By a direct inspection one can verify that for any $m \geq 1$ the components of the matrix $(\Pi')^m = ||\pi'^{(m)}_{i_r j_s}||$ satisfy the equation $\pi'^{(m)}_{i_r j_s} = \pi^{(m)}_{ij}/I_j$ and so

$$V'^{(m)}_{i_r} := \frac{1}{2} \sum_{j_s \in \mathcal{J}'} |\pi'^{(m)}_{i_r j_s} - p'_{j_s}| = V^{(m)}_i \tag{22}$$

for any $i_r \in \mathcal{J}'$.

Let the stationary probabilities p_i be rational numbers whose common denominator is D. In this case we can make a refinement as described above so that the stationary distribution will be uniform with probabilities $= 1/D$. Due to (21) we can utilize the version of Theorem 1 just proved above. It gives, in virtue of (20) and (22), the following bound:

$$V^{(m)}_i \leq 2^{-1} D^{1/2} \cdot (1-b)^{m/2} \tag{23}$$

We now turn to the proof of Theorem 1. The idea is to change the matrix Π and the vector P slightly, so that the new stationary probabilities will be rational numbers with sufficiently small common denominator. Let $p > 0$ be a small parameter, $p \ll p_{\min}$. We set $D = [p^{-1}] + 1$. There is a probability distribution $\bar{P} = (\bar{p}_1, \ldots, \bar{p}_I)$ with rational components whose common denominator is D such that

$$|\bar{p}_i - p_i| \leq 1/D < p \tag{24}$$

for all $i \in \mathcal{J}$. Note that

$$2\mathrm{Var}\,(P, \bar{P}) \leq I/D < p/p_{\min} \tag{25}$$

Let Γ be an $I \times I$ matrix defined by two conditions: (i) it is left identical on the subspace L_0 defined above, i.e., $V\Gamma = V$ for every $V \in L_0$, and (ii) it moves $\bar{P}$ to P, i.e. $\bar{P}\Gamma = P$. Since both P and $\bar{P}$ are transversal to L_0, the matrix Γ is well defined and invertible. Moreover, both Γ and Γ^{-1} are quasistochastic (see Remark above). Since Γ and Γ^{-1} are left identical on L_0, for any probability distribution Q we have

$$Q\Gamma = (Q - \bar{P} + \bar{P})\Gamma = (Q - \bar{P})\Gamma + P = Q - \bar{P} + P. \tag{26}$$

Similarly,

$$Q\Gamma^{-1} = (Q - P + P)\Gamma^{-1} = Q - P + \bar{P}. \tag{27}$$

We now consider the matrix $\bar{\Pi} = \Gamma\Pi\Gamma^{-1}$. It is quasi-stochastic, although it need not be a stochastic one, and its stationary vector is $\bar{P}$, because $\bar{P}\bar{\Pi} = \bar{P}\Gamma\Pi\Gamma^{-1} = \bar{P}$.

When D is a very large number, the matrix Γ is very close to the identity matrix, and so the matrix $\bar{\Pi}$ is very close to Π. To make this precise, we employ the vectors E_i defined above and, based on (26) and (27), obtain

$$E_i\bar{\Pi} = E_i\Gamma\Pi\Gamma^{-1} = E_i\Pi - \bar{P}\Pi + \bar{P}.$$

Then, utilizing (25) gives

$$\mathrm{Var}\,(E_i\bar{\Pi}, E_i\Pi) = \mathrm{Var}\,(\bar{P}, \bar{P}\Pi) \le \mathrm{Var}\,(\bar{P}, P) + \mathrm{Var}\,(P\Pi, \bar{P}\Pi) \le p/p_{\min} \tag{28}$$

(at the last step we used a classical estimate, $\mathrm{Var}\,(P\Pi, Q\Pi) \le \mathrm{Var}\,(P, Q)$, valid for any stochastic matrix Π and any probability distributions P, Q). We will need only the following consequence of (28):

$$\delta_* := \max_{i,j} |\bar{\pi}_{ij} - \pi_{ij}| \le p/p_{\min} \tag{29}$$

where $||\bar{\pi}_{ij}||$ are the components of $\bar{\Pi}$. It is now a simple calculation based on (14), (24) and (29) that for any $i, j \in \mathcal{J}$ we have

$$\sum_k \frac{\bar{\pi}_{ik}\bar{\pi}_{jk}}{\bar{p}_k} \ge \sum_k \frac{\pi_{ik}\pi_{jk} - \delta_*\pi_{ik} - \delta_*\pi_{jk} - \delta_*^2}{p_k(1 + p/p_{\min})}$$

$$\ge \Big(b - 2p/p_{\min}^2 - p^2/p_{\min}^3\Big) \cdot \Big(1 + p/p_{\min}\Big)^{-1}.$$

Now assume that $p/p_{\min}^2 < b/10$. Then

$$\sum_k \frac{\bar{\pi}_{ik}\bar{\pi}_{jk}}{\bar{p}_k} \ge b/2$$

Therefore, the matrix $\bar{\Pi}$ is quasi-stochastic and satisfies the condition (14) with b replaced by $b/2$. Its stationary vector has rational components with the common denominator D. The bound (23) then applies and and yields the following:

$$\bar{V}_i^{(m)} := \frac{1}{2}\sum_j |\bar{\pi}_{ij}^{(m)} - \bar{p}_j| \le 2^{-1}D^{1/2} \cdot (1 - b/2)^{m/2} \le p^{-1/2} \cdot (1 - b/2)^{m/2} \tag{30}$$

The bound (28) has the following generalization:

$$\begin{aligned}\mathrm{Var}\,(E_i\bar{\Pi}^m, E_i\Pi^m) &\leq \mathrm{Var}\,(\bar{P}, \bar{P}\Pi^m) \\ &\leq \mathrm{Var}\,(\bar{P}, P) + \mathrm{Var}\,(P\Pi^m, \bar{P}\Pi^m) \\ &\leq p/p_{\min} \end{aligned} \tag{31}$$

for any $m \geq 1$, which can be obtained by the same arguments as (28), now applied to the matrix $\bar{\Pi}^m = \Gamma\Pi^m\Gamma^{-1}$. Based on (30), (31) and (25) we have

$$\begin{aligned} V_i^{(m)} &= \frac{1}{2}\sum_j |\pi_{ij}^{(m)} - p_j| = \mathrm{Var}\,(E_i\Pi^m, P) \\ &\leq \mathrm{Var}\,(E_i\Pi^m, E_i\bar{\Pi}^m) + \mathrm{Var}\,(E_i\bar{\Pi}^m, \bar{P}) + \mathrm{Var}\,(\bar{P}, P) \\ &\leq p^{-1/2}\cdot(1-b/2)^{m/2} + 1.5p/p_{\min} \end{aligned} \tag{32}$$

We now pick

$$p = 10^{-1}bp_{\min}^2\cdot(1-b/2)^{m/3}$$

and obtain Theorem 1.

Remark. Notice that picking

$$p = 2^{-1}p_{\min}^{2/3}\cdot(1-b/2)^{m/3}$$

would yield a slightly better bound:

$$V_i^{(m)} \leq 2p_{\min}^{-1/3}(1-b/2)^{m/3}$$

But this is only valid when our assumption $p/p_{\min}^2 < b/10$ is satisfied, i.e. for

$$m \geq |\log p_{\min}^{4/3} + \log b - \log 10|\cdot|\log(1-b/2)|^{-1}$$

5 Relaxed condition on transition probabilities

In this section we relax the condition (14). Our point is that in the case of dynamical systems with singularities some atoms of the partition $\mathcal{A}$ may be very 'ugly' and their evolution may be totally out of control. In that case, it is enough to ensure a positive lower bound on $b_{i,j}$ for an 'overwhelming majority' of pairs (i,j) rather than for every single pair (i,j). Based on this we still can estimate $V^{(m)}$, as the following theorem states.

Theorem 3 *Suppose that there is a subset of pairs of indices, $\mathcal{R} \in \mathcal{J}\times\mathcal{J}$, such that for every pair $(i,j)\in\mathcal{R}$ we have $b_{i,j}\geq b>0$ and*

$$Q := \sum_{(i,j)\notin\mathcal{R}} p_ip_j < 1$$

Then for any $m \geq 1$ we have

$$V^{(m)} \leq \text{const} \cdot \left[b^{-1/2} p_{\min}^{-2} (1 - b/40)^{m/3} + m(p_{\min} + Q) \right] \tag{33}$$

where const *is an absolute constant (one can set* const $= 50$*).*

Note that Theorem 3 does not guarantee any convergence to equilibrium. It is useful only when $m(p_{\min} + Q) \ll 1$, i.e. for relatively small values of m.

The last theorem in this section shows that the hypotheses of Theorem 3 are stable under certain perturbations. Consider two Markov chains with matrices of transition probabilities $\Pi = ||\pi_{ij}||$ and $\Pi' = ||\pi'_{ij}||$ and with a common stationary distribution $P = ||p_i||$. Denote

$$b'_{i,j} = \sum_{k=1}^{I} \frac{\pi'_{ik} \pi'_{jk}}{p_k}.$$

Theorem 4 *Let the Markov chain* (Π, P) *satisfy the hypotheses of Theorem 3, and let*

$$\nu' := \frac{1}{2} \sum_{i,j=1}^{I} p_i |\pi_{ij} - \pi'_{ij}| < 1.$$

Then there is a subset $\mathcal{R}' \subset \mathcal{J} \times \mathcal{J}$ such that for every pair $(i,j) \in \mathcal{R}'$ we have $b'_{i,j} \geq b' = b/2$ and

$$Q' := \sum_{(i,j) \notin \mathcal{R}'} p_i p_j < Q + 50 b^{-1} \nu'$$

We now prove Theorem 3. The key idea of the proof is to add some new states to the Markov chain involved in this theorem so that the new, larger chain will meet the assumptions of Theorem 1 and, in some sense, will be still close enough to the original chain.

First, we notice that $b_{i,j} = b_{j,i}$, and so $(i,j) \in \mathcal{R}$ whenever $(j,i) \in \mathcal{R}$. Besides, $(i,i) \in \mathcal{R}$ for each $i \in \mathcal{J}$, since $b_{i,i} \geq 1$ for every i.

We need some preparatory work to implement our plan. That work consists in 'uniformization' of both the stationary distribution P and the set of 'bad' pairs, $(i,j) \notin \mathcal{R}$. First we make the stationary vector P 'fairly uniform', by which we mean that the ratio $\max_i p_i / p_{\min}$ will not exceed 2. To this end, we employ the refinement techniques from the proof of Theorem 1 and simply break every state in half as long as its probability is larger than $2p_{\min}$. After such a refinement, we define $\mathcal{R}'$ to be the union of all the pairs (i_r, j_s) (in the notations of Section 4) for which the pair of 'predecessors', (i,j), was in $\mathcal{R}$ in the original chain. As it follows from (21), we have $b_{i_r,j_s} \geq b$ for every pair $(i_r, j_s) \in \mathcal{R}'$. It is also clear that

$$\sum_{(i_r,j_s) \notin \mathcal{R}'} p'_{i_r} p'_{j_s} = \sum_{(i,j) \notin \mathcal{R}} p_i p_j = Q$$

The value of $V^{(m)}$ remains unchanged by virtue of (22), and, obviously, $p_{\min}$ will not be altered by the above refinement.

Therefore, it is enough to prove Theorem 3 for the new, 'refined' chain. In other words, we will simply assume, in addition to the hypotheses of Theorem 3 that

$$\max_i p_i \le 2p_{\min} \tag{34}$$

Next, we make the set of 'bad' pairs, $\mathcal{J}^2 \setminus \mathcal{R}$, 'fairly uniform', as follows. Denote $Q' = Q + p_{\min}$. For every $i \in \mathcal{J}$ let

$$q(i) := \sum_{j:(i,j)\notin\mathcal{R}} p_j$$

Then, for every $i \in \mathcal{J}$ such that $q(i) < 2Q'$ we remove from $\mathcal{R}$ one or more pairs (i, j) with some arbitrary $j \neq i$, so that the value of $q(i)$ will increase and satisfy

$$2Q' \le q(i) < 4Q'$$

This is possible, since $\max_i p_i \le 2p_{\min} \le 2Q'$. After that, for every pair (i, j) that we have removed from $\mathcal{R}$, we also remove its 'transpose' (j, i). After all that, the set $\mathcal{R}$ will be still symmetric [i.e., $(i, j) \in \mathcal{R} \Leftrightarrow (j, i) \in \mathcal{R}$] and contain the diagonal $\{(i, i) : i \in \mathcal{J}\}$. It is then an easy calculation that the new value of Q, i.e.,

$$Q_* := \sum_{(i,j)\notin\mathcal{R}} p_i p_j = \sum_{i=1}^{I} p_i q(i)$$

will satisfy the bound $Q_* \le Q + 8Q'$. Hence, $Q'_* = Q_* + p_{\min} \le 9Q'$. In addition, for any $i \in \mathcal{J}$ we now have

$$q(i) = \sum_{j:(i,j)\notin\mathcal{R}} p_j \ge 2Q' > \frac{1}{5}Q'_* \ge \frac{1}{5}Q_*. \tag{35}$$

We are now in a position to implement the plan mentioned in the beginning of the proof. For each unordered pair $(i, j) \notin \mathcal{R}$ we add a new state to our Markov chain, and we label it by (ij) [notice that (ij)=(ji)]. Next we specify a new Markov chain with the states $\{i : i \in \mathcal{J}\} \cup \{(ij) : (i, j) \notin \mathcal{R}\}$, which we call 'old' and 'new' states, respectively. The matrix of transition probabilities is defined by

$$\tilde{\pi}_{ij} = \frac{\pi_{ij}}{1 + q(i)}$$

for any pair $i, j \in \mathcal{J}$ of old states,

$$\tilde{\pi}_{i(ij)} = \frac{p_j}{1 + q(i)} \quad \text{and} \quad \tilde{\pi}_{(ij)k} = \frac{p_k q(k)}{Q_*}$$

for transitions between the old and new states and $\tilde{\pi}_{(ij)(kl)} = 0$ for any pair of new states (in particular, $\tilde{\pi}_{(ij)(ij)} = 0$). We also set $\tilde{\pi}_{k(ij)} = 0$ if k is different from i and j.

The new Markov chain has a stationary distribution with probabilities

$$\tilde{p}_i = p_i \frac{1 + q(i)}{1 + 2Q_*} \quad \text{and} \quad \tilde{p}_{(ij)} = \frac{2p_i p_j}{1 + 2Q_*}, \tag{36}$$

for the old and new states, respectively.

We now show that the new Markov chain meets the assumptions of Theorem 1. For any 'good' pair, $(i, j) \in \mathcal{R}$, of the old states we have $b_{i,j} \geq b$, and so

$$\sum_{k \in \mathcal{J}} \frac{\tilde{\pi}_{ik} \tilde{\pi}_{jk}}{\tilde{p}_k} \geq \frac{1}{8} b$$

For any 'bad' pair of the old states, $(i, j) \notin \mathcal{R}$, we have

$$\frac{\tilde{\pi}_{i(ij)} \tilde{\pi}_{j(ij)}}{\tilde{p}_{(ij)}} \geq \frac{1}{8}.$$

For any pair of new states, (ij) and (lr), we have

$$\sum_{k \in \mathcal{J}} \frac{\tilde{\pi}_{(ij)k} \tilde{\pi}_{(lr)k}}{\tilde{p}_k} \geq 1,$$

which follows from the fact that $\tilde{\pi}_{(ij)k} = \tilde{\pi}_{(lr)k}$ for all $k = 1, \ldots, I$. Lastly, for any old state, i, and any new state, (lj), we have

$$\sum_{k \in \mathcal{J}} \frac{\tilde{\pi}_{ik} \tilde{\pi}_{(lj)k}}{\tilde{p}_k} \geq \frac{1}{4} \sum_{k \in \mathcal{J}} \frac{\pi_{ik} q(k)}{Q_*} \geq \frac{1}{20},$$

which follows from (35).

Therefore, the new Markov chain satisfies the hypotheses of Theorem 1 with b replaced by $b/20$. (Note that since $\sum_{i,j} p_i p_j b_{i,j} = 1$, we always have $b \leq 1$.) According to (36), the minimum of the stationary probabilities in the new Markov chain satisfies $p_{\min}^{\text{new}} \geq p_{\min}^2/2$.

By virtue of Theorem 1, for any $m \geq 1$ and every old state $i \in \mathcal{J}$ we have

$$\frac{1}{2} \sum_{j=1}^{I} |\tilde{\pi}_{ij}^{(m)} - \tilde{p}_j| \leq 50 b^{-1/2} p_{\min}^{-2} (1 - b/40)^{m/3} \tag{37}$$

We now bound the LHS of (33) as follows:

$$\sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - p_j| p_i \leq \sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - \tilde{\pi}_{ij}^{(m)}| p_i$$

$$+ \sum_{i,j=1}^{I} |\tilde{\pi}_{ij}^{(m)} - \tilde{p}_j| p_i + \sum_{i,j=1}^{I} |\tilde{p}_j - p_j| p_i \tag{38}$$

The middle term in the RHS is readily bounded by (37). The last term in the RHS of (38) is, in view of (36), bounded by

$$\sum_{i,j=1}^{I} |\tilde{p}_j - p_j| p_i \le (1+2Q_*)^{-1} \sum_{i,j=1}^{I} (p_j q(j) + 2Q_* p_j) p_i \le 3Q_* \le 27Q'.$$

In order to bound the first term in the RHS of (38), we expand the transition probability $\tilde{\pi}_{ij}^{(m)}$ as follows:

$$\tilde{\pi}_{ij}^{(m)} = \sum_{k_1,\dots,k_{m-1}} \tilde{\pi}_{ik_1} \tilde{\pi}_{k_1k_2} \cdots \tilde{\pi}_{k_{m-1}j}.$$

where the variables $k_1, \dots, k_{m-1}$ run over all the old and new states. We break this sum into two subsums, $\tilde{\Sigma}_{ij}^{\text{old}}(m)$ and $\tilde{\Sigma}_{ij}^{\text{new}}(m)$, so that the former will be taken over the old states only (i.e., when $k_1, \dots, k_{m-1} \in \mathcal{J}$) and the latter will be taken over strings $k_1, \dots, k_{m-1}$ which include at least one new state. Respectively,

$$\tilde{\pi}_{ij}^{(m)} = \tilde{\Sigma}_{ij}^{\text{old}}(m) + \tilde{\Sigma}_{ij}^{\text{new}}(m)$$

and

$$\sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - \tilde{\pi}_{ij}^{(m)}| p_i \le \sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - \tilde{\Sigma}_{ij}^{\text{old}}(m)| p_i + \sum_{i,j=1}^{I} p_i \tilde{\Sigma}_{ij}^{\text{new}}(m) \qquad (39)$$

By using the stationarity of the distribution (36) it is a rather straightforward calculation to bound the second term in the RHS of (39) as follows:

$$\sum_{i,j=1}^{I} p_i \tilde{\Sigma}_{ij}^{\text{new}}(m) \le (1+2Q_*) \sum_{i,j=1}^{I} \tilde{p}_i \tilde{\Sigma}_{ij}^{\text{new}}(m) \le (1+2Q_*) m \sum_{(l,r)\notin\mathcal{R}} \tilde{p}_{(lr)} \le mQ_* \le 9mQ'$$

We then rewrite the first term in the RHS of (39) as follows

$$\sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - \tilde{\Sigma}_{ij}^{\text{old}}(m)| p_i$$

$$= \sum_{k_0,\dots,k_m=1}^{I} |p_{k_0} \pi_{k_0k_1} \pi_{k_1k_2} \cdots \pi_{k_{m-1}k_m} - p_{k_0} \tilde{\pi}_{k_0k_1} \tilde{\pi}_{k_1k_2} \cdots \tilde{\pi}_{k_{m-1}k_m}|$$

$$= \sum_{k_0,\dots,k_m=1}^{I} |\sum_{t=0}^{m-1} p_{k_0} \pi_{k_0k_1} \cdots \pi_{k_{t-1}k_t} (\pi_{k_tk_{t+1}} - \tilde{\pi}_{k_tk_{t+1}}) \tilde{\pi}_{k_{t+1}k_{t+2}} \cdots \tilde{\pi}_{k_{m-1}k_m}|$$

Since the distribution $\{p_k\}$ is stationary for the matrix $||\pi_{kl}||$, the last sum is bounded by

$$\sum_{t=0}^{m-1} \sum_{k_t,k_{t+1}=1}^{I} p_{k_t} |\pi_{k_tk_{t+1}} - \tilde{\pi}_{k_tk_{t+1}}| = m \sum_{i,j=1}^{I} p_i |\pi_{ij} - \tilde{\pi}_{ij}|.$$

Lastly, since $|\pi_{ij} - \tilde{\pi}_{ij}| \le \pi_{ij} q(j)$ for any pair of old states $(i,j) \in \mathcal{J}^2$, the above sum is bounded by

$$m \sum_{i,j=1}^{I} p_i \pi_{ij} q(j) = m Q_* \le 9mQ'$$

Combining the previous estimates of the three terms in the RHS of (38) gives

$$\frac{1}{2} \sum_{i,j=1}^{I} |\pi_{ij}^{(m)} - p_j| p_i \le 50 b^{-1/2} p_{\min}^{-2} (1 - b/40)^{m/3} + (9m + 14)Q'$$

Theorem 3 is proven.

We now prove Theorem 4. For any $i, j \in \mathcal{J}$ let $d_{ij} = \pi'_{ij} - \pi_{ij}$. Then,

$$\begin{aligned} b'_{i,j} &= \sum_k \frac{(\pi_{ik} + d_{ik})(\pi_{jk} + d_{jk})}{p_k} \\ &= \sum_k \frac{\pi_{ik}\pi_{jk}}{p_k} + \sum_k \frac{d_{ik}\pi_{jk}}{p_k} + \sum_k \frac{\pi_{ik} d_{jk}}{p_k} + \sum_k \frac{d_{ik} d_{jk}}{p_k} \end{aligned}$$

Denote the last three sums by $D_{ij}^{(1)}, D_{ij}^{(2)}$ and $D_{ij}^{(3)}$, respectively. Now, for each $s = 1, 2, 3$ let $\mathcal{B}^{(s)} = \{(i,j) : |D_{ij}^{(s)}| > b/6\}$. Then, clearly for any pair (i,j) in the set

$$\mathcal{R}' := \mathcal{R} \setminus \cup_{s=1}^{3} \mathcal{B}^{(s)}$$

we have $b'_{i,j} \ge b/2$.

It remains to bound the quantity Q'. First,

$$\sum_{(i,j) \in \mathcal{B}^{(1)}} p_i p_j < \frac{6}{b} \sum_{i,j} p_i p_j |D_{ij}^{(1)}| < \frac{6}{b} \sum_{i,j,k} p_k^{-1} p_i p_j |d_{ik}| \pi_{jk} = \frac{6}{b} \sum_{i,k} p_i |d_{ik}| = 12 b^{-1} \nu'$$

A similar estimate holds for $\mathcal{B}^{(2)}$. Lastly,

$$\sum_{(i,j) \in \mathcal{B}^{(3)}} p_i p_j < \frac{6}{b} \sum_{i,j} p_i p_j |D_{ij}^{(3)}| < \frac{6}{b} \sum_{i,j,k} p_k^{-1} p_i p_j |d_{ik}| \cdot |d_{jk}| \le \frac{12}{b} \sum_{i,k} p_i |d_{ik}| = 24 b^{-1} \nu'$$

where we use the following bound:

$$\sum_j p_j |d_{jk}| \le \sum_j p_j \pi'_{jk} + \sum_j p_j \pi_{jk} = 2p_k$$

Theorem 4 is proved.

Acknowledgements. This work was essentially done during my visit at the Georgia Institute of Technology (Atlanta), and I am grateful to L.A. Bunimovich, S.-N. Chow and J.K. Hale for their kind hospitality.

References

[1] V.S. Afraimovich, N.I. Chernov and E.A. Sataev, Statistical properties of 2-D generalized hyperbolic attractors, Chaos/Xaoc, **4** (1994).

[2] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lect. Notes Math., **470**, Springer-Verlag, Berlin, 1975.

[3] L.A. Bunimocvich, Decay of correlations in dynamical systems with chaotic behavior, Soviet Physics JETP, **62** (1985), 842–852.

[4] L.A. Bunimovich and Ya.G. Sinai, Markov partitions for dispersed billiards, Commun. Math. Phys., **78** (1980), 247–280.

[5] L.A. Bunimovich and Ya.G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers, Commun. Math. Phys., **78** (1981), 479–497.

[6] L.A. Bunimovich, Ya.G. Sinai and N.I. Chernov, Statistical properties of two-dimensional hyperbolic billiards, Russ. Math. Surv., **46** (1991), 47–106.

[7] N.I. Chernov, Statistical properties of the periodic Lorentz gas. Multidimensional case, J. Stat. Phys., **74** (1994), 11–53.

[8] N.I. Chernov, Limit theorems and Markov approximations for chaotic dynamical systems, to appear in Prob. Th. Rel. Fields, (1995).

[9] M. Denker, The central limit theorem for dynamical systems, Dyn. Syst. Ergod. Th. Banach Center Publ., **23**, PWN–Polish Sci. Publ., Warsaw, 1989.

[10] R.L. Dobrushin, Central limit theorem for nonstationary Markov chains I, Theory Probab. Its Appl., **1** (1956), 65–79.

[11] J.L. Doob, Stochastic Processes, Wiley, New York, 1990.

[12] I.A. Ibragimov, Y.V. Linnik, Independent and stationary sequences of random variables, Wolters-Noordhoff, Groningen, 1971.

[13] G. Nicolis, S. Martinez and E. Tirapegui, Finite coarse-graining and Chapman-Kolmogorov equation in conservative dynamical systems, Chaos, Solutions and Fractals, **1** (1991), 25–37.

[14] Ya.G. Sinai, Markov partitions and C-diffeomorphisms, Funct. Anal. Its Appl., **2** (1968), 61–82.

[15] Ya.G. Sinai, Gibbs measures in ergodic theory, Russ. Math. Surveys, **27** (1972), 21–69.

[16] S. Ulam, Problems in Modern Mathematics, Independence Publishers, New York, 1960.