YURI KIFER

Averaging in difference equations driven by dynamical systems

Astérisque, tome 287 (2003), p. 103-123

<http://www.numdam.org/item?id=AST_2003__287__103_0>

Astérisque
287, 2003, p. 103–123

AVERAGING IN DIFFERENCE EQUATIONS DRIVEN BY DYNAMICAL SYSTEMS

by

Yuri Kifer

Dedicated to Jacob Palis for his sixtieth birthday

Abstract. — The averaging setup arises in the study of perturbations of parametric families of dynamical systems when parameters start changing slowly in time. Usually, averaging methods are applied to systems of differential equations which combine slow and fast motions. This paper deals with difference equations case which leads to wider class of models and examples. The averaging principle is justified here under a general condition which is verified when unperturbed transformations either preserve smooth measures or they are hyperbolic. The convergence speed in the averaging principle is estimated for some cases, as well.

1. Introduction

In the study of evolution of many real systems we can usually observe only few parameters while other less significant ones are regarded as constant in time. A more precise investigation may reveal that these parameters change, as well, but much slower than the others. These leads to complicated double scale equations describing slow and fast motions which are difficult to solve directly. Such problems were encountered with already long ago in celestial mechanics in the study of perturbations of planetary motion. People noticed that good approximations of the slow motion on long time intervals can be obtained by averaging coefficients of its equation in fast variables. This averaging principle was applied in celestial mechanics long before it was rigourously justified in some cases in the middle of the 20th century (see [**18**] and historical remarks there).

Traditionally, averaging methods were employed in the study of two scale ordinary differential equations describing a continuous time motion. On the other hand, it is well known that the study of discrete time dynamical systems, i.e. of iterates of

---

2000 ***Mathematics Subject Classification.*** — Primary: 34C05; Secondary: 39A11, 60J05, 37D20.
Key words and phrases. — Averaging, difference equations, dynamical systems.

transformations (not necessarily invertible), enables us to deal with a wider class of models and examples and to reveal new effects. Suppose that an idealized physical system can be described by a transformation F_0 of a $(d+m)$-dimensional space and there exist functions $x_1, \ldots, x_d$ which do not change along orbits of F_0 (integrals of motion). Then, generically, F_0 can be written as a transformation of a locally trivial fiber bundle $\mathcal{M} = \{(x,y) : x \in \mathbb{R}^d, y \in M_x\}$ with base $\mathbb{R}^d$ and fibers M_x being m-dimensional manifolds acting by the formula $F_0(x,y) = (x, f_x y)$ where $f_x = f(x,\cdot) : M_x \to M_x$ is a transformation of M_x. It is natural to view a real physical system as a perturbation of the above idealized one, and so it should be described by a transformation

$$F_\varepsilon(x,y) = (x + \varepsilon\Phi(x,y,\varepsilon), f(x,y,\varepsilon)) \tag{1.1}$$

where $\Phi(\cdot,\cdot,\varepsilon) : \mathcal{M} \to \mathbb{R}^d$ and $f(x,\cdot,\varepsilon) : M_x \to M_x$. Since locally $\mathcal{M}$ has a product structure $U \times M$, where U is an open subset of $\mathbb{R}^d$ and M is an m-dimensional manifold, and iterates $F_\varepsilon^n(x,y)$ of any point (x,y) in $U \times M$ stay there for all $n \leqslant \delta/\varepsilon$ with small but fixed $\delta = \delta(x) > 0$ we conclude that it suffices to study the evolution on time intervals of order $1/\varepsilon$ only on product spaces and then glue pieces of orbits together.

In this paper we consider difference equations of the form

$$\begin{aligned} X^\varepsilon(n+1) - X^\varepsilon(n) &= \varepsilon\Phi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \quad X^\varepsilon(0) = x, \\ Y^\varepsilon(n+1) &= f(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \qquad Y^\varepsilon(0) = y \end{aligned} \tag{1.2}$$

where $X^\varepsilon(n) = X^\varepsilon_{x,y}(n) \in \mathbb{R}^d$, $Y^\varepsilon(n) = Y^\varepsilon_{x,y}(n)$ runs on a compact m-dimensional Riemannian manifold M, $\Phi = \Phi(x,y,\varepsilon)$ is a Lipschitz in x, y, ε vector function, $f_x(\cdot,\varepsilon) = f(x,\cdot,\varepsilon)$ is a family of smooth maps (usually, endomorphisms or diffeomorphisms) of M close to f_x. Thus $(X^\varepsilon_{x,y}(n), Y^\varepsilon_{x,y}(n)) = F_\varepsilon^n(x,y)$. The equations (1.2) usually cannot be solved explicitly and it is desirable to approximate its solutions for small ε. Returning back to the unperturbed $\varepsilon = 0$ case eliminates the slow motion X^ε completely and gives a rather pure approximation valid only for bounded time intervals. The averaging principle is supposed to give a prescription how to approximate the slow motion X^ε on time intervals of order $1/\varepsilon$. Recurrent relations (1.2) can be regarded as a more general than usual setup for perturbations of dynamical systems where not only the transformation itself is perturbed but also we begin to take into account evolution of some parameters whose change was disregarded before.

We note that the standard continuous time averaging setup (see [**13**]) can be always reduced by discretizing time to a model described by difference equations of type (1.2). On the other hand, an attempt to go the other way around faces substantial difficulties since the standard suspension construction should be implemented now for different transformations f_x and it is not clear how to glue everything together in an appropriate way. Observe, that (1.2) can be generalized adding some randomness in the right hand sides there so that $f_x(\cdot,\varepsilon)$ become random endomorphisms, but we will not discuss this setup here.

Assume, first, that the fast motion $Y^\varepsilon(n)$ is independent of the slow variables, i.e. $f(x,y,\varepsilon)=fy$, and so $Y^\varepsilon_{x,y}(n)=f^ny$. For an ergodic f-invariant probability measure μ the limit

$$\overline{\Phi}_\mu(x)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\Phi(x,f^ny)=\int\Phi(x,y)d\mu(y) \tag{1.3}$$

exists for μ-almost all y. For such y's uniformly in n the solution $X^\varepsilon_{x,y}$ of (1.2) is close on any time interval of order $1/\varepsilon$ to the solution $\overline{X}^\varepsilon=\overline{X}^\varepsilon_\mu=\overline{X}^\varepsilon_{x,\mu}$, taken at integer times, of the differential equation

$$\frac{d\overline{X}^\varepsilon(t)}{dt}=\varepsilon\overline{\Phi}(\overline{X}^\varepsilon(t)),\quad \overline{X}^\varepsilon(0)=x \tag{1.4}$$

where $\overline{\Phi}=\overline{\Phi}_\mu$ (see similar continuous time results in [**18**]). Already in this case the averaging principle works only for μ-almost all initial points y and for different y's averaged solutions may be different. In the particular case when f is uniquely ergodic the convergence in (1.3) is uniform in y and for all y, whence the averaged equation (1.4) and its solution are unique and the latter approximates $X^\varepsilon(n)$, $n\in[0,N/\varepsilon]$ uniformly.

The general case (1.2) when the fast and the slow motions are fully coupled is much more complicated. The averaging principle suggests here to approximate X^ε_x by $\overline{X}^\varepsilon_x$ satisfying (1.4) but with $\overline{\Phi}$ given by

$$\overline{\Phi}(x)=\overline{\Phi}_y(x)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\Phi(x,f^k_xy) \tag{1.5}$$

provided the last limit exists for "most" x and y. If μ_x is an ergodic invariant measure of f_x then the limit (1.5) exists for μ_x-almost all y's and

$$\overline{\Phi}(x)=\overline{\Phi}_{\mu_x}(x)=\int\Phi(x,y)d\mu_x(y). \tag{1.6}$$

Observe that Lipschitz continuity of $\overline{\Phi}$ cannot be guaranteed now without further assumptions even for smooth Φ, and so we do not have automatically existence and, especially, uniqueness of solutions in (1.4) in these general circumstances. On the other hand, consider the recurrent relation for $\overline{\overline{X}}^\varepsilon(n)=\overline{\overline{X}}^\varepsilon_x(n)$,

$$\overline{\overline{X}}^\varepsilon(n+1)=\overline{\overline{X}}^\varepsilon(n)+\varepsilon\overline{\Phi}(\overline{\overline{X}}^\varepsilon(n)),\quad \overline{\overline{X}}^\varepsilon(0)=x \tag{1.7}$$

which determines $\overline{\overline{X}}^\varepsilon(n)$ without any conditions on $\overline{\Phi}$ and it is easy to see that if $\overline{\Phi}$ is Lipschitz continuous and bounded then

$$\max_{0\leqslant n\leqslant T/\varepsilon}|\overline{X}^\varepsilon_x(n)-\overline{\overline{X}}^\varepsilon_x(n)|\leqslant C_T\varepsilon \tag{1.8}$$

for some $C_T > 0$ independent of ε. Thus we may discuss the approximation of $X^\varepsilon(n)$ by $\overline{\overline{X}}^\varepsilon(n)$ under more general conditions when we even do not have uniquely defined solutions of (1.4).

In general, there exists no natural family of invariant measures μ_x, $x \in \mathbb{R}^d$, since the transformations f_x may have rather different properties for different x's and the averaging principle can be justified here only under substantial restrictions. First, the averaging prescription relies here on existence of a family of probability measures μ_x such that the limit (1.5) exists μ_x-almost everywhere (a.e.) and it is given by (1.6) (at least, Lebesgue a.e. in x). Of course, in addition, we need sufficiently good dependence of Φ and f in (1.2) on ε but still, this does not seem to be enough, in general. The problem here is that the average in (1.5) is taken along orbits of the unperturbed fast motion but in the perturbed evolution (1.2) we cannot disregard now changes in the slow variable parameter of the fast motion, and so we have to study the interplay between unperturbed and perturbed dynamics. Namely, the method of this paper relies on measure estimates of sets of pairs (x, y) which arrive under the action of F_ε^k to sets of points with a specified behavior of averages for the unperturbed evolution. Then we will show that the slow motion is close to the averaged one in certain L^1-sense. Required estimates can be done assuming, for instance, that each f_x is a smooth endomorphism or a diffeomorphism of M preserving a smooth measure μ_x on M which is ergodic for Lebesgue almost all (a.a.) x. This result is a discrete time version of Anosov's theorem [**1**] which is one of few general results about fully coupled averaging. Actually, we prove our result under a general condition which is satisfied in essentially all known cases where the averaging principle holds true and it does not rely on existence of smooth invariant measures as in Anosov's approach.

Recently, quite a few papers dealt with a class of diffeomorphisms called stably ergodic (see, for instance, [**5**]) which are volume preserving ergodic diffeomorphisms having a C^2-neighborhood of volume preserving ergodic diffeomorphisms. If each f_x from our parametric family belongs to such a neighborhood then our results yield an L^1-convergence in the averaging principle. Moreover, we need ergodicity only for almost all x's which suggests to study parametric families of volume preserving diffeomorphisms which are ergodic for almost all parameter values. When convergence in the averaging principle in a fully coupled setup (1.2) holds true for any reasonable Φ we can naturally regard this as a manifestation of compatibility of f_x's or their stability within our parametric family.

Observe that our result works in the case when all f_x's are C^2 expanding transformations of M which always possess fast mixing smooth invariant measures μ_x. On the other hand, close relatives of expanding transformations Anosov and Axiom A diffeomorphism do not possess, generically, smooth invariant measures. Still, relying on specific properties of Axiom A system in a neighborhood of an attractor we will be able to carry out necessary estimates for μ_x being either Lebesgue or corresponding

Sinai-Ruelle-Bowen (SRB) measures, and so the averaging principle will be justified in this case, as well. Moreover, using moderate deviations estimates from [**12**] for this case we will give an estimate of deviation of the slow motion from the averaged one. More delicate limit theorems (large deviations, central limit theorem etc.) for these deviations will be studied in another paper. Some relevant results in this direction were obtained recently in [**3**].

Our conditions need ergodicity of measures μ_x only for a.a. and not all x's which is important in the presence of resonances. For instance, let f_x be a parametric family of toral translations. All of them preserve the Lebesgue measure but only translations with rationally independent mod 1 frequencies are ergodic. Assuming that these frequencies depend only on the slow variable x we see that, generically, they will be rationally independent mod 1 for Lebesgue almost all and not all x's. For such translations and also for some skew translations of the torus (which are both uniquely ergodic) we will be able to estimate the speed of convergence in the averaging principle deriving a discrete time version of Neistadt's theorem (see a comprehensive exposition of Anosov's and Neistadt's theorems in [**13**]).

The author is grateful to the anonimous referee for several useful suggestions improving the exposition.

2. Preliminaries and main results

Assume that the right hand sides in (1.2) satisfy

$$|\Phi(x,y,\varepsilon)-\Phi(z,v)|+d_M(f(x,y,\varepsilon),f(z,v))\leqslant L(\varepsilon+|x-z|+d_M(y,v))\\ \text{and } |\Phi(x,y,\varepsilon)|\leqslant L \tag{2.1}$$

for some $L>0$ independent of $\varepsilon>0$, $x,z\in\mathbb{R}^d$ and $y,v\in M$, where $\Phi(z,v)=\Phi(z,v,0)$, $f(z,v)=f_zv=f(z,v,0)$, $f_x=f(x,\cdot):M\to M$ for each $x\in\mathbb{R}^d$ is a Lipschitz map and d_M is the Riemannian metric on M. In Corollaries 2.2 and 2.3 below we will assume also that

$$\|\Phi(\cdot,\cdot,\varepsilon)\|_{C^1}+\|f(\cdot,\cdot,\varepsilon)\|_{C^1}\leqslant L \quad\text{and}\quad \|D_{x,y}f(\cdot,\cdot,\varepsilon)-D_{x,y}f(\cdot,\cdot)\|\leqslant L\varepsilon \tag{2.2}$$

where $\|\cdot\|_{C^k}$ is the C^k norm of the corresponding map and $D_{x,y}f$ is the differential at (x,y) of the map $f:\mathbb{R}^d\times M\to M$. Our setup includes also a family of probability measure μ_x, $x\in\mathbb{R}^d$ on M depending measurably on x. For each $n\in\mathbb{N}$ and $\delta>0$ set

$$E(n,\delta)=\Big\{(x,y):\ \Big|\frac{1}{n}\sum_{k=0}^{n-1}\Phi(x,f_x^ky)-\overline{\Phi}(x)\Big|>\delta\Big\}$$

where $\overline{\Phi}(x)=\int_M\Phi(x,y)d\mu_x(y)$. Assume that for all $x,z\in\mathbb{R}^d$,

$$\Big|\overline{\Phi}(x)-\int_M\Phi(x,y)d\mu_z(y)\Big|\leqslant L^2|x-z| \tag{2.3}$$

and there exist $\alpha, \varepsilon_0 > 0$ such that for any $T, \delta > 0, k \in \mathbb{N}$, and a compact $K \subset \mathbb{R}^d$ we can find $d_{T,K}(k,\delta) \to 0$ as $k \to \infty$ and $\eta(\varepsilon) \to \infty$ as $\varepsilon \to 0$ such that for any $\varepsilon \leqslant \varepsilon_0$ and $k \leqslant \eta(\varepsilon)$,

$$\mu\big((K \times M) \cap F_\varepsilon^{-n} E(k,\delta)\big) \leqslant d_{T,K}(k,\delta) \quad \text{if } n \leqslant T/\varepsilon - k, \tag{2.4}$$

where $d\mu(x,y) = d\mu_x(y)d\ell(x)$ and ℓ is the Lebesgue measure on $\mathbb{R}^d$.

Theorem 2.1. — *Suppose that (2.1), (2.3) and (2.4) hold true. Then for any $T > 0$ and a compact set $K \subset \mathbb{R}^d$,*

$$\lim_{\varepsilon \to 0} \int_K \int_M \sup_{0 \leqslant n \leqslant T/\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X}^\varepsilon_x(n)| d\mu_x(y) d\ell(x) = 0 \tag{2.5}$$

where $\overline{X}^\varepsilon_x(t)$ is the solution of (1.4).

The conditions (2.1) and (2.2) are clear and rather standard, the condition (2.3) is less straightforward, in general, while the assumption (2.4) is far from being transparent. We will provide in Corollaries 2.2 and 2.3 two important classes of transformations f_x such that (2.3) and (2.4) hold true for any perturbation satisfying (2.1) and (2.2). It is instructive to verify these conditions in the simpler well known setup when the fast motion does not depend on ε and on the slow one, i.e. when $F_\varepsilon(x,y) = (x + \varepsilon\Phi(x,y,\varepsilon), fy)$ where f is a map of M. Suppose that all measures μ_x coincide with the same ergodic f-invariant probability measure μ_0 on M so that $\mu = \ell \times \mu_0$. Then (2.3) follows automatically and

$$\mu\big((K \times M) \cap F_\varepsilon^{-n} E(k,\delta)\big) = \int_K \mu_0(E_x) d\ell(x)$$

where

$$E_x = E_x(n,k,\delta) = \{y \in M : (X^\varepsilon_{x,y}(n), f^n y) \in E(k,\delta)\}.$$

Set $E_z(k,\delta) = \{y : (z,y) \in E(k,\delta)\}$. By (2.1) we have that $E_u(k,\delta) \subset E_z(k,\delta/2)$ provided $|u - z| \leqslant \delta/4L$. Hence,

$$\begin{aligned} E_{x,z} = E_{x,z}(n,k,\delta) = \{y \in M : |X^\varepsilon_{x,y}(n) - z| \leqslant \delta/4L \text{ and } f^n y \in E_{X^\varepsilon_{x,y}(n)}(k,\delta)\} \\ \subset f^{-n} E_z(k,\delta/2). \end{aligned}$$

Let K_r denotes the closed r-neighborhood of K. By (2.1), $X^\varepsilon_{x,y}(n) \in K_{LT}$ if $x \in K$ and $n \leqslant T/\varepsilon$. Thus, if $z_1, \ldots, z_l$ is a minimal $\delta/4L$-net in K_{LT} then $X^\varepsilon_{x,y}(n)$ belongs to a $\delta/4L$-ball around some z_i provided $n \leqslant T/\varepsilon$. Then, for any $x \in K$,

$$E_x(n,k,\delta) \subset \cup_{i=1}^l E_{x,z_i}(n,k,\delta) \subset \cup_{i=1}^l f^{-n} E_{z_i}(k,\delta/2).$$

Since μ_0 is f-invariant and ergodic we obtain from here that

$$\mu_0(E_x(n,k,\delta)) \leqslant \sum_{i=1}^l \mu_0(E_{z_i}(k,\delta/2)) \to 0 \quad \text{as } k \to \infty$$

and (2.4) follows.

We note that though the assumption (2.4) does not seem to be weakest possible it is rather clear that without some compatibility between measures μ_x, $x \in \mathbb{R}^d$ the averaging principle is not going to work, in general. Consider, for instance, the following simplest example where $d = 1$, M is a circle $\mathbb{T}^1$ of length 1, all f_x coincide with the identity transformation of $\mathbb{T}^1$, $\Phi(x, y, \varepsilon) = \Phi(y)$ is a C^1 function depending only on y. We define μ_x to be the unit mass at x (mod 1) regarded as a point of $\mathbb{T}^1$ which is identified with the unit interval whose end points are glued together. Of course, each μ_x is an ergodic invariant measure of the identity transformation and (2.3) holds true, as well. Extending Φ as a 1-periodic function to the whole $\mathbb{R}^1$ we can write $\overline{\Phi}(x) = \int \Phi(y) d\mu_x(y) = \Phi(x)$ for any $x \in \mathbb{R}^1$. Then $d\overline{X}^\varepsilon(t)/dt = \varepsilon\Phi(\overline{X}^\varepsilon(t))$ and $\overline{Z}(t) = \overline{X}^\varepsilon(t/\varepsilon)$ satisfies $d\overline{Z}(t)/dt = \Phi(\overline{Z}(t))$. Clearly, $X^\varepsilon_{x,y}(n) = x + \varepsilon n \Phi(y)$, and so

$$\int_0^1 \int_{\mathbb{T}^1} \sup_{0 \leqslant n \leqslant T/\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X}^\varepsilon_x(n)| d\mu_x(y) d\ell(x) = \int_0^1 \sup_{0 \leqslant t \leqslant T} |t\Phi(x) + x - \overline{Z}_x(t)| d\ell(x) + O(\varepsilon).$$

The last integral is positive, in general, (take, for instance, $\Phi(x) = \cos^2 2\pi x$ obtaining $\overline{Z}_x(t) = (2\pi)^{-1} \arctan(2\pi t + \tan 2\pi x))$ and it does not depend on ε, so we do not have (2.5) in this case. More substantial examples of nonconvergence in (2.5) can be constructed, as well, but the whole question is not yet completely understood.

Next, we will provide more specific conditions which ensure that (2.4) is satisfied. We assume now that (2.1) and (2.2) hold true and that for each $x \in \mathbb{R}^d$ we are given an f_x-invariant probability measure μ_x where f_x is supposed to be now a C^2-endomorphism of M, i.e. its differential $D_y f_x$ is nondegenerate at any point $y \in M$.

Corollary 2.2. — *Suppose that each measure μ_x has a Radon-Nikodim derivative $q(x, y) = q_x(y) = d\mu_x(y)/d\rho(y)$ with respect to the normalized Riemannian volume ρ on M such that*

$$\|q\|_{C^1} + \|1/q\| \leqslant L \tag{2.6}$$

where $\|\cdot\|_{C^1}$ and $\|\cdot\|$ are corresponding C^1 and supremum norms. Then there exists $C = C_{T,K}$ such that for all $n \leqslant T/\varepsilon$ and $k \in \mathbb{N}$,

$$\mu((K \times M) \cap F_\varepsilon^{-n} E(k, \delta)) \leqslant C\mu((K_{LT} \times M) \cap E(k, \delta)) \tag{2.7}$$

were, again, $d\mu(x, y) = d\mu_x(y) d\ell$. Assume, in addition, that for ℓ-a.a. x the limit (1.5) exists μ_x-a.e. and it is given by (1.6). Then $\mu((K_{LT} \times M) \cap E(k, \delta)) \to 0$ as $k \to \infty$ and (2.4) follows. Since (2.1) and (2.6) imply (2.3) then (2.5) holds true, as well. Clearly, the measures μ_x can be replaced there by the Riemannian volume ρ.

We claim that (2.4) is also satisfied in the setup of hyperbolic diffeomorphisms. Namely, we assume now that f_x, $x \in \mathbb{R}^d$ are diffeomorphisms and for each x there exists a compact f_x-invariant set $\Lambda_x \subset M$ which is a basic hyperbolic attractor for f_x (see [**11**]). Moreover, we assume that there exists an open set $W \subset M$ such that for all $x \in \mathbb{R}^d$,

$$(2.8) \qquad \Lambda_x \subset W, \quad f_x\overline{W} \subset W, \quad \text{and} \quad \cap_{n>0} f_x^n W = \Lambda_x.$$

Denote by μ_x the Sinai-Ruelle-Bowen (SRB) invariant measure of f_x on Λ_x. Recall, that μ_x can be obtained as a weak limit of $f_x^n \rho_W$ as $n \to \infty$ where ρ_W is the normalized restriction of the Riemannian volume ρ on M to W (see [**11**]). There are several other important characterizations of the SRB measure μ_x, in particular, it is the unique equilibrium state of f_x for the function

$$(2.9) \qquad \varphi_x(y) = -\log J_x^u(y)$$

where $J_x^u(y)$ is the absolute value of the Jacobian with respect to the Riemannian inner products of the linear map $D_y f_x : \Gamma_{x,y}^u \to \Gamma_{x,f_x y}^u$ where $T_{\Lambda_x} M = \Gamma_x^s \oplus \Gamma_x^u$ is the hyperbolic splitting. The measure μ_x sits on Λ_x and, in general, even when $\Lambda_x = M$ (Anosov diffeomorphism case) is singular with respect to the Riemannian volume ρ so that Corollary 2.2 is not applicable here.

Corollary 2.3. — *Suppose that (2.1) and (2.2) hold true and for each $x \in \mathbb{R}^d$ a C^2 diffeomorphism f_x of M is given which C^2 depends on x and possesses a basic hyperbolic attractor Λ_x satisfying (2.8). Then (2.3)–(2.5) hold true if each μ_x is taken to be the corresponding SRB measure. This remains true if instead of SRB measures we take in Theorem 2.1 μ_x coinciding for each x with the Riemannian volume ρ_W restricted to the set W satisfying (2.8). Moreover, for each $\gamma > 0$ and a compact set K there exists $C_{K,\gamma} > 0$ such that*

$$(2.10) \qquad \int_K \int_M \sup_{0 \leqslant n \leqslant T/\varepsilon} |X_{x,y}^\varepsilon(n) - \overline{X}_x^\varepsilon(n)| d\rho_W(y) d\ell(x) \leqslant C_{K,\gamma}(\log 1/\varepsilon)^{-\frac{1}{2}+\gamma}.$$

Note that the convergence (2.5) can be derived from the results announced in [**3**] but we consider it useful to have an independent direct proof based on Theorem 2.1. The bound (2.10) will be derived from estimates of the next section and the moderate deviations asymptotics obtained in [**12**]. The estimate (2.10) holds true also when f_x, $x \in \mathbb{R}^d$ are C^2 expanding endomorphisms of M but (2.5) follows for them already from Corollary 2.2 since they preserve smooth ergodic invariant measures (see, for instance, [**14**]). Moreover, it is possible to extend Corollary 2.3 to the continuous time case of flows with hyperbolic attractors. If f_x do not depend on x then methods from [**12**] yield easily much better estimate of order $\sqrt{\varepsilon}$ for the left hand side of (2.10) (cf. [**8**]). In the general case $Y_{x,y}^\varepsilon(n)$ and $f_x^n y$ diverge exponentially fast and the arguments of the next section yield only a logarithmic estimate of speed of convergence in (2.10). Still, a more precise study of normalized deviations $\varepsilon^{-1/2}(X_{x,y}^\varepsilon([t/\varepsilon]) - \overline{X}_x^\varepsilon([t/\varepsilon]))$ which should lead also to the central limit theorem here is likely to provide the order $\sqrt{\varepsilon}$ estimate for the left hand side of (2.10) in the general case, as well.

Next, we consider two types of specific uniquely ergodic diffeomorphisms f_x of an m-dimensional torus $\mathbb{T}^m$ falling into the framework of Corollary 2.2 for which we will be able to obtain good estimates of speed of convergence in (2.5). First type of these diffeomorphisms consists of translations of $\mathbb{T}^m$ defined by

$$f_x(y_1,\dots,y_m) = \big(\{y_1+\omega_1(x)\},\dots,\{y_m+\omega_m(x)\}\big) \tag{2.11}$$

where $\omega(x) = (\omega_1(x),\dots,\omega_m(x))$ is a vector function of frequencies and $\{a\}$ denotes the fractional part of a. All these f_x preserve the Lebesgue measure ρ on $\mathbb{T}^m$ and it is well known (see, for instance, [**7**]) that f_x is ergodic (and even uniquely ergodic) if and only if the vector $\omega(x)$ has rationally independent mod 1 components. The second class of diffeomorphisms consists of skew translations of $\mathbb{T}^m$ having the form

$$\begin{aligned} &f_x(y_1,\dots,y_m)\\ &\quad= \big(\{y_1+\alpha(x)\},\{y_2+p_{21}y_1\},\dots,\{y_m+p_{m1}y_1+\dots+p_{m,m-1}y_{m-1}\}\big)\end{aligned} \tag{2.12}$$

where p_{ij} are positive integers and α is a function on $\mathbb{R}^d$. Again, each f_x preserves the Lebesgue measure ρ and it is ergodic (and uniquely ergodic) if and only if $\alpha(x)$ is irrational.

Since M is now the torus $\mathbb{T}^m$ we can regard $\Phi(x,y)$, $y=(y_1,\dots,y_m)$ as a vector function on $\mathbb{R}^d\times\mathbb{R}^m$ 1-periodic in each y_j, $j=1,\dots,m$. Furthermore, we assume that $\Phi(x,y)$ can be extended as an analytic function $\Phi(x,y+iz)$ to a strip

$$\{y+iz:\, y\in\mathbb{R}^m,\, |z_i|<\kappa,\, i=1,2,\dots,m\}\subset\mathbb{C}^m,\qquad \kappa>0$$

with $|\Phi|\leqslant L$. The latter condition can be relaxed to finite differentiability similar to [**13**] and it will be used only to get appropriate estimates on remainders of Fourier series. Following, [**13**] we say that a map $\xi:\mathbb{R}^d\to\mathbb{R}^l$ satisfies Kolmogorov's non-degeneracy condition if its Jacobi matrix $(\partial\xi_j/\partial x_k)$ has rank l at any point x which means that $d\geqslant l$ and the maximal absolute value $\lambda_\xi(x)$ of determinants of $l\times l$ submatrices of $(\partial\xi_j/\partial x_k)$ is positive.

Theorem 2.4. — *Suppose that Φ and f satisfy (2.1) and (2.2) and, in addition, Φ satisfies the above analyticity condition and f_x, $x\in\mathbb{R}^d$ are either all translations of $\mathbb{T}^m$ defined by (2.11) or all skew translations defined by (2.12) with $\omega(x):\mathbb{R}^d\to\mathbb{R}^m$ and $\alpha(x):\mathbb{R}^d\to\mathbb{R}$ satisfying Kolmogorov's nondegeneracy condition. Then there exists $c_0>0$ such that for every $c<c_0$, each compact set $K\subset\mathbb{R}^d$ and any $T>0$ there exists $C_{T,K,c}>0$ such that*

$$\int_K\int_M \sup_{0\leqslant n\leqslant T/\varepsilon} |X^\varepsilon_{x,y}(n)-\overline{X}^\varepsilon_x(n)|\,d\rho(y)\,d\ell(x)\leqslant C_{T,K,c}\varepsilon^c. \tag{2.13}$$

Moreover, if f_x, $x\in\mathbb{R}^d$ are given by (2.11) then we can take $c_0=1/5$ and if they are given by (2.12) then $c_0=1/(3m+7)$ will do.

Actually, employing the approach from [**15**] it is possible to prove (2.13) for f_x given by (2.11) with $c=1/2$.

3. General estimates and convergence

We begin with a general basic estimate which will be used in the proof of both Theorems 2.1 and 2.4. Set $R_k(x,y) = d_M(Y^\varepsilon_{x,y}(k), f^k_x y)$.

Proposition 3.1. — *Suppose that (2.1) and (2.3) hold true. There exists $C > 0$ such that if $1 \leqslant n(\varepsilon) \leqslant T/\varepsilon$ and $N(\varepsilon)$ is the integral part of $T(\varepsilon n(\varepsilon))^{-1}$ then for all $x \in \mathbb{R}^d$, $y \in M$, $\varepsilon \in (0,1)$ and $\delta > 0$,*

$$\tag{3.1} \sup_{0\leqslant n\leqslant T/\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X}^\varepsilon_x(n)| \leqslant Ce^{CT} \times \Big(T\delta + \varepsilon n(\varepsilon)(T+1) + \varepsilon T \textstyle\sum_{j=0}^{N(\varepsilon)-1} \big(\sum_{k=0}^{n(\varepsilon)-1} R_k(X^\varepsilon_{x,y}(jn(\varepsilon)), Y^\varepsilon_{x,y}(jn(\varepsilon))) + n(\varepsilon)\mathbb{I}_{E(n(\varepsilon),\delta)}(X^\varepsilon_{x,y}(jn(\varepsilon)), Y^\varepsilon_{x,y}(jn(\varepsilon)))\Big)$$

where $\mathbb{I}_\Gamma(v) = 1$ if $v \in \Gamma$ and $= 0$, otherwise.

Proof. — By (1.4)–(1.6) and (2.1),

$$\sup_{k\leqslant s\leqslant k+1} |\overline{X}^\varepsilon_x(s) - \overline{X}^\varepsilon_x(k)| \leqslant L\varepsilon,$$

and so by (2.3),

$$\Big| \int_k^{k+1} \overline{\Phi}(\overline{X}^\varepsilon(s))ds - \overline{\Phi}(\overline{X}^\varepsilon(k))\Big| \leqslant \varepsilon L^2(L+1).$$

Hence, by (1.2) and (1.4) for $X^\varepsilon(n) = X^\varepsilon_{x,y}(n)$ and $\overline{X}^\varepsilon(n) = \overline{X}^\varepsilon_x(n)$ we have

$$\tag{3.2} \begin{aligned} |X^\varepsilon(n) - \overline{X}^\varepsilon(n)| &\leqslant \varepsilon^2 nL^2(L+1) + \varepsilon\Big| \textstyle\sum_{k=0}^{n-1} \big(\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \overline{\Phi}(\overline{X}^\varepsilon(k))\big)\Big| \\ &\leqslant \varepsilon^2 nL^2(L+1) + \varepsilon \textstyle\sum_{k=0}^{n-1} \big|\big(\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \Phi(X^\varepsilon(k), Y^\varepsilon(k))\big| \\ &\quad + \varepsilon\Big| \textstyle\sum_{k=0}^{n-1} \big(\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))\big)\Big| + \varepsilon \textstyle\sum_{k=0}^{n-1} |\overline{\Phi}(X^\varepsilon(k)) - \overline{\Phi}(\overline{X}^\varepsilon(k))| \\ &\leqslant \varepsilon^2 nL(L^2+L+1) + \varepsilon\Big| \textstyle\sum_{k=0}^{n-1} \big(\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))\big)\Big| \\ &\quad + \varepsilon L^2(L+1) \textstyle\sum_{k=0}^{n-1} |X^\varepsilon(k) - \overline{X}^\varepsilon(k)|. \end{aligned}$$

By a version of the discrete Gronwall inequality (see, for instance, Lemma 4.20 in [**9**]) we derive from (3.2) that

$$\tag{3.3} |X^\varepsilon(n) - \overline{X}^\varepsilon(n)| \leqslant (1+\varepsilon L(L+1))^{n-1}\Big(\varepsilon^2 nL(L^2+L+1) + \varepsilon\Big| \textstyle\sum_{k=0}^{n-1} \big(\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))\big)\Big|\Big).$$

Next, setting $x^\varepsilon_j = X^\varepsilon_{x,y}(jn(\varepsilon))$ and $y^\varepsilon_j = Y^\varepsilon_{x,y}(jn(\varepsilon))$ we obtain by (2.1) that

$$\tag{3.4} \begin{aligned} &\sup_{0\leqslant n\leqslant T/\varepsilon} \Big| \textstyle\sum_{k=0}^{n-1} \big(\Phi(X^\varepsilon_{x,y}(k), Y^\varepsilon_{x,y}(k)) - \overline{\Phi}(X^\varepsilon_{x,y}(k))\big)\Big| \\ &\quad \leqslant 2Ln(\varepsilon) + \textstyle\sum_{j=0}^{N(\varepsilon)-1} \Big| \sum_{k=0}^{n(\varepsilon)-1} \big(\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k), Y^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k)) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))\big)\Big|, \end{aligned}$$

$$(3.5)\quad \Big|\sum_{k=0}^{n(\varepsilon)-1}\big(\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k),Y^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))-\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k),f^k_{x^\varepsilon_j}y^\varepsilon_j)\big)\Big| \leqslant L\sum_{k=0}^{n(\varepsilon)-1}R_k(x^\varepsilon_j,y^\varepsilon_j),$$

$$(3.6)\quad \Big|\sum_{k=0}^{n(\varepsilon)-1}\big(\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k),f^k_{x^\varepsilon_j}y^\varepsilon_j)-\Phi(x^\varepsilon_j,f^k_{x^\varepsilon_j}y^\varepsilon_j)\big)\Big|\leqslant L\sum_{k=0}^{n(\varepsilon)-1}|X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k)-x^\varepsilon_j|,$$

and

$$(3.7)\quad |X^\varepsilon_{z,v}(k)-z|\leqslant \varepsilon Lk$$

for any $z\in\mathbb{R}^d$ and $v\in M$. In addition, by (2.1) and (2.3),

$$(3.8)\quad \Big|\sum_{k=0}^{n(\varepsilon)-1}\overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))-\overline{\Phi}(x^\varepsilon_j)n(\varepsilon)\Big|\leqslant L(L+1)\sum_{k=0}^{n(\varepsilon)-1}|X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k)-x^\varepsilon_j|.$$

Observe also that

$$(3.9)\quad \Big|\sum_{k=0}^{n(\varepsilon)-1}\Phi(z,f^k_zv)-\overline{\Phi}(z)n(\varepsilon)\Big|\leqslant n(\varepsilon)(\delta+2L\mathbb{I}_{E(n(\varepsilon),\delta)}(z,v)).$$

Finally, Proposition 3.1 follows from (3.3)–(3.9). □

Now we can complete the proof of Theorem 2.1. Observe that by (2.1) and (3.7),

$$(3.10)\quad \begin{aligned} R_k(x,y) &= d_M(f(X^\varepsilon_{x,y}(k-1),Y^\varepsilon_{x,y}(k-1),\varepsilon),f^k_xy)\\ &\leqslant d_M(f(X^\varepsilon_{x,y}(k-1),Y^\varepsilon_{x,y}(k-1),\varepsilon),f(X^\varepsilon_{x,y}(k-1),Y^\varepsilon_{x,y}(k-1))\\ &\qquad + d_M(f(X^\varepsilon_{x,y}(k-1),Y^\varepsilon_{x,y}(k-1),\varepsilon),f(x,Y^\varepsilon_{x,y}(k-1))\\ &\qquad + d_M(f_x(Y^\varepsilon_{x,y}(k-1)),f_x(f^{k-1}_xy))\\ &\leqslant L(\varepsilon+L\varepsilon k+R_{k-1}(x,y))\\ &=\varepsilon L\sum_{l=0}^{k-1}L^l(1+L(k-l))\\ &\leqslant \varepsilon L(1+Lk)(L^k-1)(L-1)^{-1}. \end{aligned}$$

Integrating (3.1) against μ over $K\times M$ and taking $n(\varepsilon)=\min([\eta(\varepsilon)],[(log\frac{1}{\varepsilon})^{1-\alpha}])$ for some $\alpha\in(0,1)$ we derive from (2.4), (3.1), and (3.10) that for any $T,\gamma>0$ there exists $C(T,\gamma)>0$ such that

$$(3.11)\quad \int_K\int_M\sup_{0\leqslant n\leqslant T/\varepsilon}|X^\varepsilon_{x,y}(n)-\overline{X}^\varepsilon_x(n)|d\mu(x,y)\leqslant C(T,\gamma)(\delta+\varepsilon^{1-\gamma}+d_{T,K}(n(\varepsilon),\delta)).$$

Letting, first, $\varepsilon\to0$ and then $\delta\to0$ we obtain (2.5).

4. Proof of Corollaries

4.1. We deal first with Corollary 2.2. Denote by Jac_yf_x the Jacobian of the linear map $D_yf_x: T_yM\to T_{f_xy}M$ with respect to the Riemannian norms. Since f_x is an endomorphism Jac_yf_x is bounded away from zero uniformly in $y\in M$ and in x belonging to a compact set. The density q_x of the f_x-invariant measure μ_x satisfies

$$(4.1)\quad q_x(y)=\sum_{v\in f_x^{-1}y}\frac{q_x(v)}{|\mathrm{Jac}_vf_x|}.$$

By perturbation arguments, for any compact set $V \subset \mathbb{R}^d$ there exists $\varepsilon(V) > 0$ such that if $\varepsilon \leqslant \varepsilon(V)$ then the differential $D_{z,v}F_\varepsilon : \mathbb{R}^d \times M \to \mathbb{R}^d \times M$ is nondegenerate on $V \times M$ and, moreover, its Jacobian $\mathrm{Jac}_{z,v}F_\varepsilon$ is uniformly bounded away from zero there. Then for any bounded Borel function g on $\mathbb{R}^d \times M$,

$$\begin{aligned}\int_{V\times M} g\circ F_\varepsilon(x,y)d\mu_x(y)d\ell(x) &= \int_{V\times M} g\circ F_\varepsilon(x,y)q_x(y)\times d\rho(y)d\ell(x)\\ &= \int_{F_\varepsilon(V\times M)} g(z,v)\sum_{(x,y)\in F_\varepsilon^{-1}(z,v)}\frac{q_x(y)}{|\mathrm{Jac}_{x,y}F_\varepsilon|}d\rho(v)d\ell(z).\end{aligned}\tag{4.2}$$

It follows from (2.1), (2.2), and from the implicit function theorem arguments that there exists $C_1 > 0$ depending only on V such that if $x \in V$, $F_\varepsilon(x,y) = (z,v)$, and $f_z^{-1}v = (w_1,\dots,w_k)$ then $F_\varepsilon^{-1}(z,v) = ((x_1,y_1),\dots,(x_k,y_k))$ with $|x_i - z| \leqslant C_1\varepsilon$ and $|w_i - y_i| \leqslant C_1\varepsilon$, $i = 1,\dots,k$. Thus by (2.1), (2.2), (2.6), and (4.1),

$$\sum_{(x,y)\in F_\varepsilon^{-1}(z,v)}\frac{q_x(y)}{|\mathrm{Jac}_{x,y}F_\varepsilon|} \leqslant \sum_{y\in f_z^{-1}v}\frac{q_z(y)}{|\mathrm{Jac}_y f_z|} + C_2\varepsilon = q_z(v) + C_2\varepsilon \tag{4.3}$$

for some $C_2 > 0$ depending only on V. Suppose that $g \geqslant 0$ then substituting (4.3) into (4.2) and taking into account (2.6) we obtain

$$\begin{aligned}\int_{V\times M} g\circ F_\varepsilon(x,y)d\mu_x(y)d\ell(x) &\leqslant \int_{F_\varepsilon(V\times M)} g(z,v)(q_z(v)+C_2\varepsilon)d\rho(v)d\ell(z)\\ &\leqslant (1+C_2L\varepsilon)\int_{F_\varepsilon(V\times M)} g(z,v)d\mu_x(v)d\ell(z).\end{aligned}\tag{4.4}$$

If $(x,y) \in K \times M$ then by (2.1) we see that $F_\varepsilon^n(x,y) \in K_{LT} \times M$ for all $n \in [0, T/\varepsilon]$ where, recall, K_r is the closed r-neighborhood of K. It follows that there exists $C_{T,K} > 0$ such that

$$\int_{K\times M} g\circ F_\varepsilon^n(x,y)d\mu_x(y)d\ell(x) \leqslant C_{T,K}\int_{K_{LT}\times M} g(z,v)d\mu_x(v)d\ell(z) \tag{4.5}$$

for all $n \in [0, T/\varepsilon]$ and taking $g = \mathbb{I}_{E(k,\delta)}$ we derive (2.7).

4.2. Next, we consider the setup of Corollary 2.3. First, observe that (2.3) follows from § 14 in [**2**] (see also [**17**]). If x belongs to a compact set K then by (2.1), $X_{x,y}^\varepsilon(n) \in \mathcal{X} = K_{LT}$ for all $n \leqslant T/\varepsilon$ so we will have to consider x-coordinates in $\mathcal{X}$ only. Any vector $\xi \in T(\mathbb{R}^d \times M) = \mathbb{R}^d \oplus TM$ can be uniquely written as $\xi = \xi^{\mathcal{X}} + \xi^W$ where $\xi^{\mathcal{X}} \in T\mathbb{R}^d$ and $\xi^W \in TM$ and it has the Riemannian norm $|||\xi||| = |\xi^{\mathcal{X}}| + \|\xi^W\|$ where $|\cdot|$ is the usual Euclidean norm on $\mathbb{R}^d$ and $\|\cdot\|$ is the Riemannian norm on M. The corresponding metrics on M and on $\mathbb{R}^d \times M$ will be denoted by d_M and d, respectively, so that if $z_1 = (x_1, w_1)$, $z_2 = (x_2, w_2) \in \mathbb{R}^d \times M$ then $d(z_1, z_2) = |x_1 - x_2| + d_M(w_1, w_2)$. It is known (see [**10**] and [**16**]) that the hyperbolic splitting $T_{\Lambda_x}M = \Gamma_x^s \oplus \Gamma_x^u$ over Λ_x can be continuously extended to the splitting $T_W M = \Gamma^s_x \oplus \Gamma^u_x$ over W which is forward invariant with respect to Df_x and uniformly in $x \in \mathcal{X}$ satisfies exponential estimates

$$\|Df^n_x \xi\| \leqslant e^{-\kappa n}\|\xi\| \quad \text{and} \quad \|Df_x^{-n}\eta\| \leqslant e^{-\kappa n}\|\eta\| \tag{4.6}$$

for all $\xi \in \Gamma^s_x$, $\eta \in \Gamma^u_x$, and $n \geqslant n_0$. Moreover, by [**6**] (see also [**17**]) we can choose these extensions so that Γ^s_x and Γ^u_x will be C^1 in x in the corresponding Grassmann manifold. Each vector $\xi \in T_{x,w}(\mathcal{X} \times W) = T_x\mathcal{X} \oplus T_w W$ can be represented uniquely in the form $\xi = \xi^{\mathcal{X}} + \xi^u + \xi^s$ with $\xi^{\mathcal{X}} \in T_x\mathcal{X}$, $\xi^u \in \Gamma^u_{x,w}$ and $\xi^s \in \Gamma^s_{x,w}$. For any small $\varepsilon, \beta > 0$ set $\mathcal{C}^u(\varepsilon,\beta) = \{\xi \in T(\mathcal{X}\times W) : \|\xi^s\| \leqslant \varepsilon\beta^{-2}\|\xi^u\| \text{ and } |\xi^{\mathcal{X}}| \leqslant \varepsilon\beta^{-1}\|\xi^u\|\}$ and $\mathcal{C}^u_{x,w}(\varepsilon,\beta) = \mathcal{C}^u(\varepsilon,\beta) \cap T_{x,w}(\mathcal{X} \times W)$ which are cones around Γ^u and $\Gamma^u_{x,w}$, respectively. Similarly, we define $\mathcal{C}^s(\varepsilon,\beta) = \{\xi \in T(\mathcal{X} \times W) : \|\xi^u\| \leqslant \varepsilon\beta^{-2}\|\xi^s\| \text{ and } |\xi^{\mathcal{X}}| \leqslant \varepsilon\beta^{-1}\|\xi^s\|\}$ and $\mathcal{C}^s_{x,w}(\varepsilon,\beta) = \mathcal{C}^s(\varepsilon,\beta) \cap T_{x,w}(\mathcal{X} \times W)$. We claim that there exist $n_1, \beta_0, \varepsilon(\beta) > 0$ such that if $F^k_\varepsilon z \in \mathcal{X} \times W \,\forall\, k = 0, 1, \ldots, n$, $n \geqslant n_1$, $\beta \leqslant \beta_0$, $\varepsilon \leqslant \varepsilon(\beta)$ then

$$D_z F^n_\varepsilon \mathcal{C}^u_z(\varepsilon,\beta) \subset \mathcal{C}^u_{F^n_\varepsilon z}(\varepsilon,\beta), \quad \mathcal{C}^s_z(\varepsilon,\beta) \supset D_z F_\varepsilon^{-n} \mathcal{C}^s_{F^n_\varepsilon z}(\varepsilon,\beta) \tag{4.7}$$

and for any $\xi \in \mathcal{C}^u_z(\varepsilon,\beta)$, $\eta \in \mathcal{C}^s_{F^n_\varepsilon z}(\varepsilon,\beta)$,

$$|||D_z F^n_\varepsilon \xi||| \geqslant e^{\frac{1}{2}\kappa n}|||\xi|||, \quad |||D_z F_\varepsilon^{-n}\eta||| \geqslant e^{\frac{1}{2}\kappa n}|||\eta|||. \tag{4.8}$$

Before deriving (4.7) and (4.8) we explain how to use them in order to obtain (2.4). For any linear subspace Ξ of $T_z(\mathcal{X} \times W)$ denote by $J^\Xi_\varepsilon(z)$ absolute value of the Jacobian of the linear map $D_z F_\varepsilon : \Xi \to D_z F_\varepsilon \Xi$ with respect to inner products induced by the Riemannian metric and set $J^\Xi_\varepsilon(n,z) = \prod_{k=0}^{n-1} J_\varepsilon^{D_z F^k_\varepsilon \Xi}(F^k_\varepsilon z)$. Denote also by $J^u_x(y)$ absolute value of the Jacobian of the linear map $D_y f_x : \Gamma^u_{x,y} \to \Gamma^u_{x,f_x y}$ and set $J^u_x(\varepsilon,n,y) = \prod_{k=0}^{n-1} J^u_{X^\varepsilon_{x,y}(k)}(Y^\varepsilon_{x,y}(k))$. Let n^u and n^s be the dimensions of $\Gamma^u_{x,y}$ and $\Gamma^s_{x,y}$, respectively, which do not depend on x, y by continuity considerations. If Ξ is an n^u-dimensional subspace of $T_z(\mathcal{X} \times W)$, $z = (x,y)$, and $\Xi \subset \mathcal{C}^u_{x,y}(\varepsilon,\beta)$ then it follows easily from (2.1), (2.2), and (4.7) that there exists a constant $C_1 > 0$ independent of $x \in \mathcal{X}$ and $y \in M$, such that for any small $\varepsilon > 0$ and $n \in \mathbb{N}$,

$$(1 - C_1\varepsilon)^n \leqslant J^\Xi_\varepsilon(n,z)(J^u_x(\varepsilon,n,y))^{-1} \leqslant (1 + C_1\varepsilon)^n. \tag{4.9}$$

For each $y \in \Lambda_x$ and $\gamma > 0$ small enough set

$$W^s_x(y,\gamma) = \{v \in W : d_M(f^k_x y, f^k_x v) \leqslant \gamma \;\forall\, k \geqslant 0\}$$

and

$$W^u_x(y,\gamma) = \{v \in W : d_M(f^k_x y, f^k_x v) \leqslant \gamma \;\forall\, k \leqslant 0\}$$

which are local stable and unstable manifolds for f_x at y. According to [**10**] these families can be included into continuous families of n^s and n^u-dimensional stable and unstable discs $W^s_x(y,\gamma)$ and $W^u_x(y,\gamma)$, respectively, defined for all $y \in W$ and such that $W^s_x(y,\gamma)$ is tangent to Γ^s_x, $W^u_x(y,\gamma)$ is tangent to Γ^u_x, $f_x W^s_x(y,\gamma) \subset W^s_x(f_x y,\gamma)$ and $W^u_x(y,\gamma) \supset f_x^{-1} W^u_x(f_x y,\gamma)$. For any $z, \widetilde{z} \in \mathbb{R}^d \times M$ set

$$d^\varepsilon_n(z,\widetilde{z}) = \max\{d(F^k_\varepsilon z, F^k_\varepsilon \widetilde{z}) : 0 \leqslant k < n\}$$

and

$$B^{\varepsilon}_{x}(y,\gamma,n)=\{v\in W:\, d^{\varepsilon}_{n}((x,y),(x,v))\leqslant\gamma\}.$$

Let $v\in W^u_x(y,\gamma)$ and assume that $B^{\varepsilon}_{x}(v,\gamma,n)$ does not intersect the boundary $\partial W^u_x(y,\gamma)$ of $W^u_x(y,\gamma)$. Set

$$V^u_x(v)=V^u_x(v,\gamma)=W^u_x(y,\gamma)\cap B^{\varepsilon}_{x}(v,\gamma,n)$$

and

$$V^u_x(v,k)=V^u_x(v,\gamma,k)=F^k_{\varepsilon}(\{x\}\times V^u_x(v)),\quad k=1,2,\dots,n.$$

By (4.7), $TV^u_x(v,k)\subset\mathcal{C}^u(\varepsilon,\beta)$, $k=1,2,\dots,n$ (where TV is the tangent bundle of V) and we conclude by (4.8), (4.9) and the volume lemma type arguments (see Appendix in [**4**]) that

$$C^{-1}\leqslant\rho^u(V^u_x(v))J^u_x(\varepsilon,n,v)\leqslant C \tag{4.10}$$

where ρ^u is the induced Riemannian volume on $W^u_x(y,\gamma)$ and we denote by C here and below different positive constants depending on γ but not on $x\in\mathcal{X}$, y,ε, and $n\leqslant T/\varepsilon$. By the volume lemma arguments we derive also that

$$\rho^u(V^u_x(v)\cap F^{-n}_{\varepsilon}E(k,\delta))\leqslant C\big(J^u_x(\varepsilon,n,v)\big)^{-1}\rho_V\big(V^u_x(v,n)\cap E(k,\delta)\big) \tag{4.11}$$

where ρ_V is the induced n^u-dimensional Riemannian volume on $V^u_x(v,n)$. By (4.7) and (4.8) it is easy to see that if $w\in V^u_x(v)$ then

$$d(F^k_{\varepsilon}(x,v),F^k_{\varepsilon}(x,w))\leqslant C\exp(-C^{-1}(n-k)) \tag{4.12}$$

for all $k=0,1,\dots,n$ which together with (1.2) yield that

$$|X^{\varepsilon}_{x,v}(n)-X^{\varepsilon}_{x,w}(n)|\leqslant C\varepsilon, \tag{4.13}$$

i.e. the x-coordinate of points in $V^u_x(v,n)$ may differ at most by $2C\varepsilon$. It follows by (2.1) that if $(z,u)\in V^u_x(v,n)\cap E(k,\delta)$ then $(X^{\varepsilon}_{x,v}(n),u)\in E(k,\delta/2)$ provided $k\leqslant(\log 1/\varepsilon)^{1-\alpha}$, $\alpha>0$ and ε is small enough. Then, we will have also that for such k,

$$\{X^{\varepsilon}_{x,v}(n)\}\times W^s_{X^{\varepsilon}_{x,v}(n)}(u,\gamma)\subset E(k,\delta/3)$$

provided γ is small enough. This together with (4.7) yield that

$$\begin{aligned}\rho_V\big(V^u_x(v,n)\cap E(k,\delta)\big)&\leqslant C\rho^u\big(W^u_{X^{\varepsilon}_{x,v}(n)}(Y^{\varepsilon}_{x,v}(n),\gamma))\cap E_{X^{\varepsilon}_{x,v}(n)}(k,\delta/2)\big)\\&\leqslant C^2\rho(E_{X^{\varepsilon}_{x,v}(n)}(k,\delta/3))\end{aligned} \tag{4.14}$$

where $E_z(k,r)=\{v:\,(z,v)\in E(k,r)\}$ and $k\leqslant(\log 1/\varepsilon)^{1-\alpha}$. In the right hand side of (4.14) we can write also $\mu_{X^{\varepsilon}_{x,v}(n)}$ in place of ρ. It follows from the upper moderate deviations bound in [**12**] that

$$r_k=\sup_{z\in\mathcal{X}}\rho(E_z(k,k^{a-1}))\leqslant\exp(-c_a k^{2a-1}) \tag{4.15}$$

for each $a\in(\frac{1}{2},1)$ and some $c_a>0$. Observe that (4.15) remains true with μ_z in place of ρ.

Now, choose a maximal set of points $v_i \in W^u_x(y,\gamma)$ such that $V^u_x(v_i,\gamma)$ does not intersect the boundary of $W^u_x(y,2\gamma)$ and $d^\varepsilon_n(v_i,v_j) \geqslant \gamma$ if $i \neq j$. Then $\cup_i V^u_x(v_i,\gamma) \supset W^u_x(y,\gamma)$ and $V^u_x(v_i,\gamma/2)$ are disjoint for different i's. Applying the volume lemma style arguments as above to $V^u_x(v_i,\gamma/2)$ we conclude from (4.10) that

$$\sum_i \big(J^u_x(\varepsilon,n,v_i)\big)^{-1} \leqslant C\rho^u(W^u_x(y,\gamma)). \tag{4.16}$$

It follows from (4.11) and (4.14)–(4.16) that

$$\rho^u(W^u_x(y,\gamma)\cap F^{-n}_\varepsilon E(k,\delta)) \leqslant \sum_i \rho^u(V^u_x(v_i,\gamma)\cap F^{-n}_\varepsilon E(k,\delta)) \leqslant Cr_k \tag{4.17}$$

provided $\delta \geqslant k^{a-1}$ and $k \leqslant (\log\frac{1}{\varepsilon})^{1-\alpha}$. Let $B((x,y),c\gamma)$ be a ball in $\mathbb{R}^d \times M$ of radius $c\gamma$ centered at (x,y) for some small constant c. The family $W^u_z(v,\gamma)\cap B((x,y),c\gamma)$, $(z,v)\in B((x,y),c\gamma)$ forms a measurable partition of $B((x,y),c\gamma)$ (even a foliation) and conditional measures of $\ell\times\rho$ relative to this partition are equivalent to the corresponding measures ρ^u. Hence, (4.17) implies

$$\ell\times\rho(B((x,y),c\gamma)\cap F^{-n}_\varepsilon E(k,\delta)) \leqslant Cr_k \tag{4.18}$$

provided γ is small enough. Choose a minimal finite cover of $\mathcal{X}\times M$ by balls of radius $c\gamma$ centered at some points (x_j,y_j), making the above construction for each point (x_j,y_j) and applying (4.18) we arrive at (2.4) and (2.10) follows from (3.1), (3.10), (4.15) and (4.18). Since essential estimates concern only volumes on unstable and close to unstable manifolds where the SRB measures are equivalent to the volume there we see that (4.19) remains true if we replace $\ell\times\rho$ by μ such that $d\mu(x,y) = d\mu_x(y)d\ell(x)$, and so (2.4) and (2.10) remain true with such μ, as well.

Finally, we prove (4.7) and (4.8). Let $\xi = \xi^{\mathcal{X}} + \xi^u + \xi^s \in T_z(\mathcal{X}\times M)$, $D_zF^n\xi^{\mathcal{X}} = \zeta = \zeta^{\mathcal{X}} + \zeta^u + \zeta^s \in T_{F^nz}(\mathcal{X}\times M)$, $z=(x,w)$, $D_wf^n_x\xi^u = \eta^u$, and $D_wf^n_x\xi^s = \eta^s$. Then $D_zf^n_x\xi = \zeta^{\mathcal{X}} + (\zeta^u+\eta^u) + (\zeta^s+\eta^s)$ and $|\xi^{\mathcal{X}}| = |\zeta^{\mathcal{X}}|$, $\|\zeta^u\| \leqslant Ce^{Cn}|\xi^{\mathcal{X}}|$, $\|\zeta^s\| \leqslant Ce^{Cn}|\xi^{\mathcal{X}}|$, $\|\eta^s\| \leqslant C\|\xi^s\|$ for some $C>0$ independent of ξ and $\|\eta^u\| \geqslant e^{\kappa n}\|\xi^u\|$ if $n \geqslant n_0$. Hence, for $n \geqslant n_0$,

$$\|\xi^u+\eta^u\| \geqslant \|\eta^u\| - \|\zeta^u\| \geqslant e^{\kappa n}\|\xi^u\| - Ce^{Cn}|\xi^{\mathcal{X}}| \tag{4.19}$$

and

$$\|\zeta^s+\eta^s\| \leqslant \|\zeta^s\| + \|\eta^s\| \leqslant Ce^{Cn}|\xi^{\mathcal{X}}| + C\|\xi^s\|. \tag{4.20}$$

Suppose that $\|\xi^u\| \geqslant \beta\varepsilon^{-1}|\xi^{\mathcal{X}}|$ and $\|\xi^u\| \geqslant \beta^2\varepsilon^{-1}\|\xi^s\|$. Then by (4.19),

$$\|\zeta^u+\eta^u\| \geqslant (\tfrac12 e^{\kappa n}\beta\varepsilon^{-1} - Ce^{Cn})|\xi^{\mathcal{X}}| + \tfrac12 e^{\kappa n}\beta^2\varepsilon^{-1}\|\xi^s\|. \tag{4.21}$$

Put $n_1 = [\kappa^{-1}\ln(8C+5)]+1$, choose $\beta_0>0$ so that $e^{\kappa n} \geqslant 4\beta Ce^{Cn}$ for any $\beta \leqslant \beta_0$ and all $n \in [n_1, 2n_1]$, and set $\varepsilon(\beta) = \frac{\beta}{4}\min(\beta, C^{-1}e^{-2Cn_1})$. Then by (4.21) we see that $D_zF^n_0\xi \in \mathcal{C}^u_{F^n_0z}(\varepsilon,2\beta)$ and since we have to check (4.7) only for $n\in[n_1,2n_1]$ we obtain easily from (2.1) and (2.2) that the perturbation F_ε of F_0 satisfies the first part of (4.7) provided β_0 and $\varepsilon(\beta)$ are chosen sufficiently small. The second part of (4.7) follows in the same way.

Next, for $n \geqslant n_0$,

$$
\begin{aligned}
|||D_z F_0^n \xi||| &\geqslant \|\eta^u\| - |\zeta^{\mathcal{X}}| - \|\zeta^u\| - \|\zeta^s\| - \|\eta^s\| \\
&\geqslant e^{\kappa n}\|\xi^u\| - (1+2Ce^{Cn})|\xi^{\mathcal{X}}| - C\|\xi^s\| \\
&\geqslant (e^{\kappa n} - \beta^{-1}\varepsilon(1+2Ce^{Cn}) - \beta^{-2}\varepsilon C)\|\xi^u\| \\
&\geqslant (e^{\kappa n} - \beta^{-1}\varepsilon(1+2Ce^{Cn}) - \beta^{-2}\varepsilon C)(1+\varepsilon\beta^{-1}+\varepsilon\beta^{-2})^{-1}|||\xi|||.
\end{aligned}
\tag{4.22}
$$

Choose $\varepsilon(\beta)$ so small (for instance, $\varepsilon(\beta)=\beta^3$) that for all $\varepsilon \leqslant \varepsilon(\beta)$ and $\beta \leqslant \beta_0$,

$$
e^{\kappa n} - \varepsilon\beta^{-1}(1+2Ce^{Cn}) - \varepsilon\beta^{-2}C \geqslant (1+\varepsilon\beta^{-1}+\varepsilon\beta^{-2})e^{\frac{2}{3}\kappa n}
$$

for any $n \in [n_1, 2n_1]$. Then, $|||D_z F_0^n \xi||| \geqslant e^{\frac{2}{3}\kappa n}|||\xi|||$ for all such n, and so if ε small enough we have also $|||D_z F_\varepsilon^n \xi||| \geqslant e^{\frac{1}{2}\kappa n}|||\xi|||$. Using (4.7) and repeating this argument for $D_z F_\varepsilon^{in_1}\xi$, $i=1,2,\ldots.$ in place of ξ we derive the first assertion in (4.8) for all $n \geqslant n_1$ and the second one follows in the same way.

5. Toral translations and skew translations

First, we write the Fourier series for Φ,

$$
\Phi(x,y) = \sum_{k\in\mathbb{Z}^m} \Phi_k(x)\exp(2\pi i(k,y))
\tag{5.1}
$$

where $(k,y)=\sum_{j=1}^m k_j y_j$, the vector coefficients Φ_k are given by

$$
\Phi_k(x) = \int \Phi(x,y)\exp(-2\pi i(k,y))d\rho(y)
\tag{5.2}
$$

where ρ is the Lebesgue measure on $\mathbb{T}^m$. Set

$$
r_N(x,y) = \sum_{k:\,|k|>N} \Phi_k(x)\exp(2\pi i(k,y))
$$

then by the well known estimates of tails of Fourier expansions of analytic functions

$$
r_N(x,y) \leqslant c_\Phi^{-1} e^{-c_\Phi N}
\tag{5.3}
$$

for some $c_\Phi > 0$ which can be written explicitly via the supremum norm of the analytic extension of Φ (see Appendix 1 in [**13**]). By (5.2), $\overline{\Phi}(x)=\Phi_0(x)$, and so by (2.1) and (5.1)–(5.3),

$$
\Big|\frac{1}{n}\sum_{l=0}^{n-1}\Phi(x,f_x^k y) - \overline{\Phi}(x)\Big| \leqslant c_\Phi^{-1}e^{-c_\Phi N} + \frac{L}{n}\sum_{k:\,|k|\leqslant N}\Big|\sum_{l=0}^{n-1}\exp(2\pi i(k,f_x^l y))\Big|.
\tag{5.4}
$$

Next, we will deal separately with translations and skew translations.

5.1. Consider, first, toral translations f_x defined by (2.11). Then we have to estimate

$$\Big|\sum_{l=0}^{n-1}\exp(2\pi i(k,y+l\omega(x)))\Big| \leqslant 2|\exp(2\pi i(k,\omega(x)))-1|^{-1}. \tag{5.5}$$

Let $V\subset\mathbb{R}^d$ be a compact set and

$U_V(\eta,N)=\{x\in V:\,|\exp(2\pi i(k,\omega(x)))-1|\leqslant\eta|k|^{-m}\text{for some } k \text{ with } 0<|k|<N\}.$

It follows by Diophantine approximations type arguments (cf. Appendix 4 in [**13**]) that for some constant $C>0$,

$$\ell(U_V(\eta,N))\leqslant C\eta(\inf_{x\in V}\lambda_\omega(x))^{-1}\ell(V) \tag{5.6}$$

with λ_ω defined before the statement of Theorem 2.4. Set $N=N(\varepsilon)=[-c_1\log\varepsilon]$, $n=n(\varepsilon)=\varepsilon^{-c_2}$, $\delta=\delta(\varepsilon)=\varepsilon^{c_3}$, and $\eta=\eta(\varepsilon)=\varepsilon^{c_4}$ where $c_1,c_2,c_3,c_4>0$ will be picked up later. Then by (5.4) and (5.5) for any $x\in V\setminus U_V(\eta,N)$,

$$\Big|\frac{1}{n}\sum_{l=0}^{n-1}\Phi(x,f_x^ky)-\overline{\Phi}(x)\Big|\leqslant C(\varepsilon^{c_\Phi c_1}+(\log\varepsilon)^{2m}\varepsilon^{c_2-c_4}) \tag{5.7}$$

for some $C>0$ independent of ε. Hence, if

$$\min(c_\Phi c_1,c_2-c_4)>c_3 \tag{5.8}$$

then

$$V\cap E(\delta(\varepsilon),n(\varepsilon))\subset U_V(\eta(\varepsilon),N(\varepsilon)) \tag{5.9}$$

provided ε is small enough.

Next, we improve the estimate (3.10) in our particular case. Since f_x is an isometry then $d_M(f_x(Y^\varepsilon_{x,y}(l-1)),f_x(f_x^{l-1}y))=R_{l-1}(x,y)$ and (3.10) becomes

$$R_l(x,y)\leqslant\varepsilon L(1+Ll)+R_{l-1}(x,y)\leqslant\varepsilon Ll(1+Ll). \tag{5.10}$$

It follows from (2.7), (3.1), (5.6), (5.7), and (5.10) that (2.13) holds true with $c=\min(c_3,1-2c_2,c_4)$ provided (5.8) is satisfied. Since c_1 can be chosen arbitrarily large we have to choose only $c_2,c_3,c_4>0$ so that $\min(c_3,1-2c_2,c_4)$ is maximal possible assuming that $c_2-c_4>c_3$. Set $c_5=\min(c_2-c_4,1-2c_2,c_4)$ then $c_5\geqslant c$. Since we can increase one term in the last minimum only by decreasing another term there, it is clear that c_5 will be maximized when all three terms are equal. Solving emerging then two equations we obtain $c_5=c_4=1/5$ and $c_2=2/5$. Taking c_3 arbitrary close but less than $1/5$ we conclude that (2.13) holds true with any $c<c_0=1/5$.

5.2. Next, we consider skew translations f_x defined by (2.12). Identifying $y\in\mathbb{T}^m$ with a vector of $\mathbb{R}^m$ having coordinates $y_i\in[0,1)$, $i=1,\dots,m$ we define recursively vectors $\widetilde{f}^k_xy$, $\widetilde{Y}^\varepsilon_{x,y}(k)\in\mathbb{R}^m$ by $\widetilde{f}^0_xy=y$, $\widetilde{Y}^\varepsilon_{x,y}(0)=y$ and $\widetilde{f}^n_xy=(I+P)\widetilde{f}^{n-1}_xy+a(x)$, $\widetilde{Y}^\varepsilon_{x,y}(n)=(I+P)\widetilde{Y}^\varepsilon_{x,y}(n-1)+a(X^\varepsilon_{x,y}(n-1))$ where I is the $m\times m$ identity matrix, $P=(p_{jl})$ is the matrix whose elements for $l<j$ appear in the definition (2.12) and

for $l > j$ are equal to zero, and $a(x)$ is the m-vector whose first coordinate is $\alpha(x)$ and all other coordinates are zero. Then

$$\widetilde{f}_x^n y = (I+P)^n y + \sum_{l=1}^{n} (I+P)^{n-l} a(x) \tag{5.11}$$

and

$$\widetilde{Y}_{x,y}^{\varepsilon}(n) = (I+P)^n y + \sum_{l=1}^{n} (I+P)^{n-l} a(X_{x,y}^{\varepsilon}(l-1)). \tag{5.12}$$

The crucial fact in using the formulas (5.11) and (5.12) is that P is nilpotent, and so

$$P^j = 0 \quad \text{for all } j \geqslant m. \tag{5.13}$$

For any vector $v = (v_1, \dots, v_m) \in \mathbb{R}^m$ denote by $\{v\}$ the vector in $\mathbb{T}^m$ whose coordinates are fractional parts $\{v_1\}, \dots, \{v_m\}$ of $v_1, \dots, v_m$, i.e. $\{\cdot\} : \mathbb{R}^m \to \mathbb{T}^m$ is the natural projection. It is clear that

$$f_x^n y = \{\widetilde{f}_x^n y\} \quad \text{and} \quad Y_{x,y}^{\varepsilon}(n) = \{\widetilde{Y}_{x,y}^{\varepsilon}(n)\}. \tag{5.14}$$

Expanding binomials in (5.11) the coordinates of $f_x^n y$ can be written more explicitly. Namely, (see [**7**], §2 in Ch. 7) $f_x^n(y_1, \dots, y_m) = (y_1^{(n)}(\alpha(x)), \dots, y_m^{(n)}(\alpha(x)))$ where

$$\begin{aligned} &y_1^{(n)}(\beta) = y_1 + n\beta, \\ &y_l^{(n)} = y_l + \sum_{j=1}^{l-1} y_j \sum_{q=1}^{l-j} \binom{n}{q} p_{lj}^{(q)} + \beta \sum_{q=1}^{l-1} \binom{n}{q+1} p_{l1}^{(q)}, \quad 1 \leqslant l \leqslant m, \end{aligned} \tag{5.15}$$

where $p_{lj}^{(q)}$ are elements of the qth power of the matrix P.

Observe, that the projection $\{\cdot\}$ does not increase distances, and so by (3.7) and (5.11)–(5.14),

$$\begin{aligned} R_n(x,y) &= d_M(Y_{x,y}^{\varepsilon}(n), f_x^n y) \leqslant |\widetilde{Y}_{x,y}^{\varepsilon} - \widetilde{f}_x^n y| \\ &\leqslant CLn^m \sum_{l=1}^{n} |X_{x,y}^{\varepsilon}(l-1) - x| \leqslant \varepsilon CL^2 n^{m+2} \end{aligned} \tag{5.16}$$

for some $C > 0$ depending only on P and m.

Let $k = (k_1, \dots, k_m) \in \mathbb{Z}^m$, $|k| \neq 0$ and set $l_0 = l_0(k) = \max\{l : k_l \neq 0\}$. Define the map $\Psi_k : \mathbb{R}^{l_0} \to \mathbb{R}^{l_0}$ acting by $\Psi_k(y_1, \dots, y_{l_0 - 1}, \beta) = (\gamma_1, \dots, \gamma_{l_0})$ so that

$$\sum_{l=1}^{l_0} k_l y_l^{(n)}(\beta) = G_\gamma(n) = (k, y) + \gamma_1 n + \gamma_2 n^2 + \cdots + \gamma_{l_0} n^{l_0}$$

for any $n \in \mathbb{N}$ where $y_l^{(n)}(\beta)$, $l = 1, \dots, l_0$ are given by (5.15). Observe, that Ψ_k is one-to-one and the coordinates of $\Psi_k^{-1}(\gamma_1, \dots, \gamma_{l_0})$ can be obtained recursively from the formulas

$$
\begin{aligned}
\beta k_{l_0} p_{l_0 1}^{(l_0-1)} &= l_0! \gamma_{l_0}, \quad y_1 k_{l_0} p_{l_0 1}^{(l_0-1)} = (l_0-1)!(\gamma_{l_0-1} + c_{11}\beta),\\
y_2 k_{l_0} p_{l_0 2}^{(l_0-2)} &= (l_0-2)!(\gamma_{l_0-2} + c_{21}\beta + c_{22} y_1), \dots,\\
y_{l_0-1} k_{l_0} p_{l_0, l_0-1}^{(1)} &= (\gamma_1 + c_{l_0-1,1}\beta + \textstyle\sum_{j=1}^{l_0-2} c_{l_0-1,j+1} y_j)
\end{aligned}
\tag{5.17}
$$

where the coefficients c_{jl}, $j \geqslant l$ satisfy

$$
|c_{jl}| \leqslant c(l_0)|k| \tag{5.18}
$$

for some $c(m) > 0$ depending only on m. Next, we will employ an elementary estimate of Weyl's sums which can be found on p.p. 215–216 in [**19**]. Namely, let

$$
W_\gamma(n) = \sum_{l=1}^{n} \exp(2\pi i G_\gamma(l))
$$

then it follows easily that for any compact $V \subset \mathbb{R}^{l_0}$,

$$
\int_V |W_\gamma(n)|^2 d\ell(\gamma) \leqslant n\ell(V_{\sqrt{l_0}})
$$

where, recall, V_r is an r-neighborhood of V. By Chebyshev's inequality we conclude that for any $n \in \mathbb{N}$ and $\widetilde{c} \in (0, 1/2)$ there exists a Borel set $U = U_{V,n,\widetilde{c}} \subset V$ such that

$$
\ell(U) \leqslant \ell(V_{\sqrt{l_0}}) n^{2\widetilde{c}-1} \tag{5.19}
$$

and

$$
|W_\gamma(n)| \leqslant n^{1-\widetilde{c}} \quad \text{provided } \gamma \notin U. \tag{5.20}
$$

It follows from (5.17)–(5.20) and the nondegeneracy condition on $\alpha(x)$ that for any $T > 0$ and a compact $K \subset \mathbb{R}^d$ there exists a constant $C(K, T) > 0$ such that

$$
\Big| \sum_{l=0}^{n-1} \exp(2\pi i (k, f_x^l y)) \Big| \leqslant n^{1-\widetilde{c}} \tag{5.21}
$$

provided $x \in K_{LT}$, $y \in \mathbb{T}^m$ and $(x, y) \notin U_k \subset \mathbb{R}^d \times \mathbb{T}^m$ with

$$
\ell \times \rho(U_k) \leqslant C(K, T)|k|^{l_0} n^{2\widetilde{c}-1} \tag{5.22}
$$

where U_k depends on n, k and $\widetilde{c}$. Set now $N = N(\varepsilon) = [-c_1 \log \varepsilon]$, $n = n(\varepsilon) = \varepsilon^{-c_2}$, and $\delta = \delta(\varepsilon) = \varepsilon_3^c$. Then by (5.4) for any $(x, y) \notin U = \cup_{k:|k| \leqslant N} U_k$,

$$
\Big| \frac{1}{n} \sum_{l=0}^{n-1} \Phi(x, f_x^k y) - \overline{\Phi}(x) \Big| \leqslant C(\varepsilon^{c_\Phi c_1} + (\log \varepsilon)^{3m} \varepsilon^{\widetilde{c} c_2}) \tag{5.23}
$$

for some $C > 0$ independent of ε. Hence, if

$$
\min(c_\Phi c_1, \widetilde{c} c_2) > c_3 \tag{5.24}
$$

then

$$
K_{LT} \cap E(\delta(\varepsilon), n(\varepsilon)) \subset U. \tag{5.25}
$$

It follows from (2.7), (3.1), (5.16), (5.22), and (5.25) that (2.13) holds true with

$$c = \min(c_3, 1-(m+2)c_2, c_2(1-2\widetilde{c})) \tag{5.26}$$

provided $\widetilde{c} < 1/2$ and (5.24) is satisfied. Since c_1 can be chosen arbitrarily large we have to choose only $\widetilde{c}$, c_2, $c_3 > 0$ so that the right hand side of (5.26) is maximal possible assuming that $\widetilde{c}c_2 > c_3$. Set $c_4 = \min(\widetilde{c}c_2, 1-(m+2)c_2, c_2(1-2\widetilde{c}))$, then $c_4 \geqslant c$. Again, c_4 will be maximized when all three term in the minimum there will be equal which gives $\widetilde{c} = 1/3$, $c_2 = 3/(3m+7)$, and $c_4 = 1/(3m+7)$. It follows that (2.13) holds true with any $c < c_0 = 1/(3m+7)$.

References

[1] D.B. Anosov, *Averaging in systems of ordinary differential equations with fast oscillating solutions,* Izv. Acad. Nauk SSSR Ser. Mat., **24** (1960), 731–742 (in Russian).

[2] V.I. Bakhtin, *A direct method of constructing an invariant measure on a hyperbolic attractor,* Russian Acad. Sci. Izv. Math., **41** (1993), 207–227.

[3] V.I. Bakhtin, *On the averaging method in a system with fast hyperbolic motions,* Proc. Math. Inst. of Belarus Acad. Sci., **6** (2000), 23–26 (in Russian).

[4] R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows,* Invent. Math., **29** (1975), 181–202.

[5] K. Burns, C. Pugh, and A. Wilkinson, *Stable ergodicity and Anosov flows,* Topology, **39** (2000), 149–159.

[6] G. Contreras, *Regularity of topological and metric entropy of hyperbolic flows,* Math. Z., **210** (1992), 97–111.

[7] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory,* (1982), Springer-Verlag, Berlin.

[8] H.S. Dumas and F. Golse, *The averaging method for perturbations of mixing flows,* Ergod. Th. & Dynam. Sys. **17** (1997) 1339–1358.

[9] S.N. Elaydi, *An Introduction to Difference Equations,* (1996), Springer–Verlag, New York.

[10] M. Hirsch, J. Palis, G. Pugh and M. Shub, *Neighborhoods of hyperbolic sets,* Invent. Math., **9** (1970), 121–134.

[11] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems,* (1995), Cambridge Univ. Press, Cambridge.

[12] Yu. Kifer, *Limit theorems in averaging for dynamical systems,* Ergod. Th.& Dynam. Sys., **15** (1995), 1143–1172.

[13] P. Lochak and C. Meunier, *Multiple Averaging for Classical Systems,* (1988), Springer-Verlag, New York.

[14] R. Mañé, *Ergodic Theory and Differentiable Dynamics,* (1987), Springer-Verlag, Berlin.

[15] A.I. Neishtadt, *Averaging in multifrequency systems. II,* Sov. Phys. Dokl., **21**, No. 2 (1976), 80–82.

[16] C. Robinson, *Structural stability of C^1 diffeomorphisms,* J. Diff. Equat., **22** (1976), 28–73.

[17] D. Ruelle, *Differentiation of SRB states,* Comm. Math. Phys., **187** (1997), 227–241.

[18] J.A. Sanders and F.Verhurst, *Averaging Methods in Nonlinear Dynamical Systems,* (1985), Springer-Verlag, Berlin.

[19] I.M. Vinogradov, *Selected Works,* (1985) Springer–Verlag, Berlin.

Y. KIFER, Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem 91904, Israel
E-mail : kifer@math.huji.ac.il