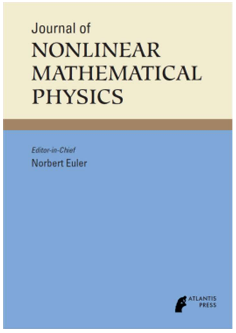

Journal of Nonlinear Mathematical Physics

ISSN (Online): 1776-0852 **ISSN (Print): 1402-9251**
Journal Home Page: https://www.atlantis-press.com/journals/jnmp

Averaging in Weakly Coupled Discrete Dynamical Systems

Niklas Brännström

To cite this article: Niklas Brännström (2009) Averaging in Weakly Coupled Discrete Dynamical Systems, Journal of Nonlinear Mathematical Physics 16:4, 465–487, DOI: https://doi.org/10.1142/S1402925109000492

To link to this article: https://doi.org/10.1142/S1402925109000492

Published online: 04 January 2021

Article

Journal of Nonlinear Mathematical Physics, Vol. 16, No. 4 (2009) 465–487

AVERAGING IN WEAKLY COUPLED DISCRETE DYNAMICAL SYSTEMS

NIKLAS BRÄNNSTRÖM

University of Helsinki, Department of Mathematics
P. O. Box 68, FIN-00014 University of Helsinki, Finland
niklas.brannstrom@helsinki.fi

Received 19 February 2009
Accepted 10 May 2009

In [Y. Kifer, Averaging in difference equations driven by dynamical systems, *Asterisque* **287** (2003) 103–123] a general averaging principle for slow-fast discrete dynamical systems was presented. In this paper we extend this method to weakly coupled slow-fast systems. For this setting we obtain sharper estimates than in the mentioned paper.

Keywords: Averaging principle; slow-fast dynamics; difference equations; full coupling; dynamical systems.

1. Introduction

Averaging methods is a collective name for the different methods that have been developed to approximate the solutions of dynamical systems with two (or more) time scales. Such systems are also referred to as slow-fast dynamical systems as some of the variables evolve slowly while the other ones evolve rapidly. In general these two subsystems are interdependent, i.e. they are fully coupled to each other. This is for example the standard case in slow-fast Hamiltonian systems. The problem with slow-fast systems is that in general it is very hard to resolve their exact dynamics. This is why averaging methods are useful. The main idea behind averaging is to approximate (often only to first order) the evolution of the slow variables by replacing the exact contribution of the fast variables by their average contribution. This average contribution is computed as space averages over the fast phase space for fixed (or frozen) slow variables. A key point in the justification of the averaging principle is to prove that the time average of the fast dynamics converges sufficiently fast to its space average. Once the average contribution of the fast variables has been computed it is substituted into the equations of motion of the slow variables to obtain the averaged equations of motion of the slow dynamics. Solving this equation, which is now independent of the fast variables, gives the averaged slow dynamics.

Averaging methods had been used for a long time before they were put on a formal footing in the 20th century [2]. The methods have typically been tailored for the particular problem at hand. The methods strongly depended on the form of the fast subsystem. For an introduction to averaging problems and results we refer to e.g. [16, 13, 9]. Anosov [1] was the first to investigate fully coupled systems

$$\begin{aligned} \frac{dX}{dt} &= \varepsilon\varphi(X, Y, \varepsilon), \\ \frac{dY}{dt} &= f(X, Y, \varepsilon), \end{aligned} \tag{1.1}$$

where Y belongs to a d-dimensional compact space. Under suitable assumptions, including the unperturbed fast dynamics ($\varepsilon = 0$) being ergodic on almost every fibre, Anosov derived an averaging method for this system. The proof involves replacing time averages of the fast dynamics with space averages. In order to justify this replacement it is crucial to understand how fast the time average of f converges to the space average of f. This is dealt with by studying the rate of "ergodization" [14] on subsets of the space of initial conditions and the amount of time the trajectories of (1.1) spend there. We make use of similar subsets in our proof and refer to it as "the set of slow ergodization".

Over the last decade Kifer has written several papers with new results for systems of the general form (1.1). First in the time discrete setting [10],

$$\begin{aligned} X^\varepsilon(n+1) - X^\varepsilon(n) &= \varepsilon\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \\ Y^\varepsilon(n+1) &= f(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \end{aligned} \tag{1.2}$$

and later in the time continuous setting (1.1), see [9,11]. In these papers Kifer gives sufficient conditions for the averaging principle to hold. For the averaging principle to work, there is a crucial condition (concerning the rate of ergodization of the unperturbed fast system) which, unfortunately, is difficult to check. Kifer proves, in [9–11] and using a different method in [12] that this condition is met if the unperturbed fast motion for each frozen slow variable, that is $f = f(X^\varepsilon(0), Y^\varepsilon(n), 0)$, is hyperbolic. Kifer also verifies that his general averaging theory yields Neishtadt's [15] optimal averaging results for multifrequency systems. Hence for the time continuous setting Kifer shows that his general theory works in the already known cases of Anosov and Neishtadt, and for the time discrete case he proves the analogous results. Recently Dolgopyat [4] proved that the averaging principle works if the unperturbed fast motion is uniformly partially hyperbolic, which Kifer had conjectured.

The system we investigate in this paper is of the same type as (1.2), but we replace the right-hand side of the first equation by $\varepsilon G(X^\varepsilon) + \varepsilon^2\Phi(X^\varepsilon, Y^\varepsilon, \varepsilon)$ to weaken the coupling. We find, using the method developed by Kifer [10], that we can make finer estimates for the difference between the true and averaged slow dynamics (Theorem 1).

As a motivating example we note that the system

$$\begin{aligned} X(n+1) - X(n) &= \varepsilon G(X(n)) + \varepsilon^2\Phi(X(n), Y(n), \varepsilon) \\ Y(n+1) &= f(Y(n)) + \varepsilon g(X(n), Y(n), \varepsilon) \end{aligned} \tag{1.3}$$

falls into the class of systems we are studying. Compared to [15] this system has more complicated fast dynamics and as opposed to [5] we allow for fast systems which are not only diagonal. System (1.3) fits within the framework of Anosov [1], compare (1.1), but we are able to give a sharper upper bound for the approximation in this case.

We have to impose a number of conditions on the system we are considering to prove the averaging principle in the weakly coupled setting. Most of these conditions are quite standard but one of the conditions, Condition 1, is not. This is the same condition Kifer uses on the set of points in the phase space for which the time average of the unperturbed fast map converges too slowly to its ergodic average. The condition on the rate of convergence is stated in terms of the coupling parameter ε in the full dynamics. It is hard to verify whether any given system satisfies Condition 1. In Sec. 3 we present three classes of examples where Condition 1 can be verified.

For one of these examples we can strengthen the averaging theorem by sharpening some of the estimates in the proof, that is, if the unperturbed fast map is cohomologous to zero then we can give an upper bound on the rate of convergence (as $\varepsilon \to 0$) between the slow component of the true solution and the averaged solution. We use a result from Kachurovskii's survey paper [8] on rates of convergence in ergodic theorems to prove our result. We note that despite having the rate of convergence in the ergodic theorem we are not able to extend the time scale for which the approximation given by the averaging theory is valid. This suggests that in general the averaging principle cannot be pushed further in this direction.

Let us also mention that, in general, if an averaging method applies to the problem it is not valid for all initial conditions, see [9] for a discussion. For an explicit example of where the exact trajectory deviates from the averaged solution at the fastest rate possible, see [7] and [3] where slow-fast Hamiltonian systems are studied under the assumption that the fast dynamics spends most of its time close to periodic orbits.

The outline of this paper is as follows. We start by considering an example in Sec. 2 which illustrates how the proof of the main theorem works. In Sec. 3 we state the problem, the conditions that need to be satisfied and the main result. This is followed by some examples where the theorem is applicable. The section is concluded with a theorem for the cohomologous case mentioned above. In Sec. 4 and 5 the main result and the cohomologous case are proved respectively.

2. A First Approach

In this section we use first principles to give an understanding of the general mechanism of averaging. Furthermore this example clearly highlights which estimates we have to improve to push the results further, which we will do in the subsequent sections of this paper.

Our main tool is the following lemma of Gronwall type.

Lemma 1. *Let $b \geq 0$ and $a_n \geq 0$ for all $n \in \mathbb{N}$. If $z_n \leq a_n + b\sum_{k=1}^{n-1} z_k$ and a_n is increasing, then $z_n \leq a_n(1+b)^{n-1}$.*

The proof is elementary and goes by induction. Similar discrete Gronwall type of inequalities can be found in e.g. [6] (Lemma 4.20 therein).

Consider the equation

$$X(n+1) = X(n) + \varepsilon G(X(n)) + \varepsilon^2 \Phi(X(n), Y(n)), \tag{2.1}$$

where $Y(n)$ is given by some dynamical system. Suppose we approximate it with the following equation

$$\overline{X}(n+1) = \overline{X}(n) + \varepsilon G(\overline{X}(n)) + \varepsilon^2 \overline{\Phi}(\overline{X}(n)), \tag{2.2}$$

where, for now, $\overline{\Phi}$ is a function $\overline{\Phi} : \mathbb{R}^n \to \mathbb{R}^n$. ($\overline{\Phi}$ will later play the role of averaged Φ.) We now estimate the size of the error. The assumptions we make on the system are that $X(n) \in \mathbb{R}^n$ and $Y \in M$ a d-dimensional compact Riemannian manifold, that $G, \Phi, \overline{\Phi}$ are bounded by L and that G, Φ, f are Lipschitz in all their arguments with Lipschitz constant L.

We are interested in estimating the difference between the true slow dynamics $X(n)$ and the approximated slow dynamics $\overline{X}(n)$. Let the initial condition be $(X(0), Y(0))$ and choose $X(0) = \overline{X}(0)$, then the nth iterates of this point can be written as

$$X(n) = X(0) + \varepsilon \sum_{k=0}^{n-1} G(X(k)) + \varepsilon^2 \sum_{k=0}^{n-1} \Phi(X(k), Y(k)) \tag{2.3}$$

and

$$\overline{X}(n) = \overline{X}(0) + \varepsilon \sum_{k=0}^{n-1} G(\overline{X}(k)) + \varepsilon^2 \sum_{k=0}^{n-1} \overline{\Phi}(\overline{X}(k)) \tag{2.4}$$

respectively. Using the assumptions that G is Lipschitz and Φ and $\overline{\Phi}$ are bounded the difference between the true dynamics and the approximated dynamics can be estimated by

$$|X(n) - \overline{X}(n)| \leq \varepsilon L \sum_{k=1}^{n-1} |X(k) - \overline{X}(k)| + \varepsilon^2 2Ln. \tag{2.5}$$

Next we want to apply Lemma 1, hence let

$$\begin{aligned} z_n &= |X(n) - \overline{X}(n)|, \\ b &= \varepsilon L, \\ a_n &= 2\varepsilon^2 Ln, \end{aligned}$$

from which it is clear that $b > 0$ and a_n is an increasing sequence. Hence the lemma is applicable to (2.5) and we get

$$z_n \leq (1 + \varepsilon L)^{n-1} 2\varepsilon^2 Ln. \tag{2.6}$$

That is

$$|X(n) - \overline{X}(n)| \leq ((1 + \varepsilon L)^{\frac{1}{\varepsilon L}})^{L\varepsilon(n-1)} 2\varepsilon^2 Ln \leq e^{L\varepsilon(n-1)} 2\varepsilon^2 Ln. \tag{2.7}$$

Hence, for all $0 \leq n \leq \frac{T}{\varepsilon}$ we have

$$|X(n) - \overline{X}(n)| \leq 2\varepsilon LTe^{LT}, \tag{2.8}$$

and

$$\lim_{\varepsilon \to 0} |X(n) - \overline{X}(n)| = 0. \tag{2.9}$$

This is as far as first principles take us without even specifying what the averaged system looks like (only assuming that $\overline{\Phi}$ is bounded). In the next section we strengthen this result by sharpening the estimate of the second term on the right-hand side of (2.5) making it $O(\varepsilon^3)$.

3. Set-Up of the Problem and Results

Consider a discrete dynamical system $Z(n+1) = F(Z(n))$ on a manifold $\mathcal{M}$. Suppose that there is a separation of timescales on which the variables $Z(n)$ evolve. Call the slow variables $X(n)$ and the fast variables $Y(n)$, i.e. $Z(n) = (X(n), Y(n))$. Assume that the manifold $\mathcal{M}$ can be decomposed to a direct product of a slow and fast space $\mathcal{M} = \{(x, y) : x \in \mathbb{R}^d, y \in M\}$, where M is an m-dimensional compact Riemannian manifold with Riemannian metric d_M. Let $(X^\varepsilon_{x,y}(n), Y^\varepsilon_{x,y}(n)) = F^n_\varepsilon(x, y)$ be the nth iterate of an initial point (x, y) defined by the following difference equations

$$\begin{cases} X^\varepsilon(n+1) - X^\varepsilon(n) = \varepsilon\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon) \\ Y^\varepsilon(n+1) = f(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \end{cases} \tag{3.1}$$

where

$$\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon) = G(X^\varepsilon(n)) + \varepsilon\Phi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \tag{3.2}$$

with the initial condition $X^\varepsilon(0) = x$, $Y^\varepsilon(0) = y$ and where $\varepsilon \in \mathbb{R}$ is a small positive parameter, $\varphi : \mathcal{M} \to \mathbb{R}^d$, $f : \mathcal{M} \to M$. We will refer to the map F as the full map, the dynamics on M as the fast dynamics and the dynamics on $\mathbb{R}^d$ as the slow dynamics. The form (3.2) of $\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon)$ gives us a weak coupling between the fast and the slow systems. A word on the notation in what follows: let $f(x, y) = f_x(y) = f(x, y, 0)$ and let $\varphi_0(x, y, 0) = G(x) + \varepsilon\Phi(x, y, 0) = G(x) + \varepsilon\Phi(x, y)$.

We have to make a number of assumptions on the system described by (3.1) to prove the averaging theorem. First we assume that the functions $G(x), \Phi(x, y, \varepsilon)$ and $f(x, y, \varepsilon)$ are Lipschitz in all their arguments, with Lipschitz constant L. Also assume that $G(x)$ and $\Phi(x, y, \varepsilon)$ are bounded by the constant L, where $L > 2$ is independent of ε. Let μ_x be a family of probability measures on M depending measurably on x, and define

$$\overline{\varphi}(x, \varepsilon) := \int_M \varphi_0(x, y, 0)d\mu_x(y) = G(x) + \varepsilon \int_M \Phi(x, y, 0)d\mu_x(y), \tag{3.3}$$

and

$$\overline{\Phi}(x) = \int_M \Phi(x, y, 0)d\mu_x(y). \tag{3.4}$$

From this it follows that $\overline{\varphi}(x, \varepsilon) = G(x) + \varepsilon\overline{\Phi}(x)$. Using these space averages we can define the set of points where the space average of the perturbation deviates more than δ from the time average calculated over n time steps.

Remark 1. We will assume that the fast system with a fixed slow variable has good ergodization properties. These properties depend strongly on the choice of the measure μ_x. Therefore dependence on x could be essential for Condition 1.

In the case where μ_x depends on x we have to make the following assumption on the perturbation Φ

$$\left|\overline{\Phi}(x) - \int_M \Phi(x, y)d\mu_z(y)\right| \le L^2|x - z| \quad \forall\, x, z \in \mathbb{R}^d. \tag{3.5}$$

Definition 1. (The set of slow ergodization) Let the set of slow ergodization $E(n,\delta)$ be given by

$$E(n,\delta)=\left\{(x,y):\left|\frac{1}{n}\sum_{k=0}^{n-1}\Phi(x,f_x^k y)-\overline{\Phi}(x)\right|>\delta\right\}. \tag{3.6}$$

Finally we have to impose a condition on the measure of the set of slow ergodization of system (3.1). The idea behind the condition is that in the proof of our averaging theorem we divide the time interval of length $n=\frac{T}{\varepsilon}$ into sections each of length n_0^ε, giving us $N(\varepsilon)\in\mathbb{N}$ such sections and, possibly, a remainder r of length less than n_0^ε. On each of these sections we want to replace the time average by a space average. Therefore we assume that the measure of the set of slow ergodization $E(n_0^\varepsilon,\delta)$ and its $N(\varepsilon)$ preimages tend to 0 as $\varepsilon\to 0$. More formally, let $n=n_0^\varepsilon N(\varepsilon)+r$, where $0\le r<n_0^\varepsilon$ and $n_0^\varepsilon\le(\log\frac{1}{\varepsilon})^{1-\alpha}$ for some $\alpha\in(0,1)$. Let $K\subset\mathbb{R}^d$ be compact.

Condition 1. Let n_0^ε be an integer-valued function such that $n_0^\varepsilon\le(\log\frac{1}{\varepsilon})^{1-\alpha}$ for some $\alpha\in(0,1)$. Assume that for any $T,\delta>0$

$$\max_{k=1,\dots,\left\lfloor\frac{T}{n_0^\varepsilon}\right\rfloor}\mu\{(x,y)\in K\times M: F_\varepsilon^{kn_0^\varepsilon}(x,y)\in E(n_0^\varepsilon,\delta)\}\to 0, \tag{3.7}$$

as $\varepsilon\to 0$, where $d\mu(x,y)=d\mu_x(y)dl(x)$ and l is the Lebesgue measure on $\mathbb{R}^d$.

It is not straightforward to check that this last assumption is fulfilled. We provide some examples later in this section.

Remark 2. For notational purposes we define a sequence $d_{T,K}(n_0^\varepsilon,\delta)$ which tends to 0 as $\varepsilon\to 0$. We use this sequence to bound the expression in (3.7) from above.

Now we are ready to state the main theorem.

Theorem 1. *Suppose that the Lipschitz and boundedness assumptions on the functions in* (3.1) *hold as well as* (3.5). *Assume that Condition* 1 *is satisfied. Then for any* $T>0$ *and any compact set* $K\subset\mathbb{R}^d$ *such that* (3.7) *holds*

$$\lim_{\varepsilon\to 0}\int_K\int_M\sup_{0\le n\le\frac{T}{\varepsilon}}\frac{1}{\varepsilon}|X_{x,y}^\varepsilon(n)-\overline{X_x^\varepsilon}(n)|d\mu_x(y)dl(x)=0, \tag{3.8}$$

where $\overline{X_x^\varepsilon}(n)$ *is the solution of the averaged equation*

$$\overline{X^\varepsilon}(n+1)-\overline{X^\varepsilon}(n)=\varepsilon\overline{\varphi}(\overline{X^\varepsilon}(n),\varepsilon),\quad \overline{X^\varepsilon}(0)=x, \tag{3.9}$$

and $\overline{\varphi}$ *is defined by* (3.3).

Before we consider some examples where this theorem is applicable we state some remarks on the assumptions required by the theorem.

Remark 3. Condition 1 is simplified to

$$\mu\{(x,y)\in K\times M\cap E(n_0^\varepsilon,\delta)\}\le d_K(n_0^\varepsilon,\delta), \tag{3.10}$$

if the full map $F_\varepsilon^n(x,y)$ preserves the measure $d\mu(x,y)=d\mu_x(y)dl(x)$ above. For example, this is the case if the map is symplectic and the measure given by the symplectic form coincides with the measure $d\mu(x,y) = d\mu_x(y)dl(x)$. In general the symplectic form $\Omega = \sum \omega_{ij}(z)dz_i \wedge dz_j$ may depend on $z = (x, y)$.

Remark 4. To satisfy (3.7) we are looking for pairs of functions (φ, f) that have a rate of ergodization (using the terminology of [14]) that matches the length of the interval n_0^ε. The rate of convergence in the ergodic theorems depends on both φ and f, not only on either of these. See [8].

Remark 5. Note that (3.5) follows automatically if the measure on the fast space is independent of x, that is $\mu_x = \mu_0$ for $\forall\, x$.

Remark 6. The assumption on Lipschitz continuity in ε may be weakened in the following way. If system (3.1) satisfies this Lipschitz condition

$$|\varphi(x,y,\varepsilon) - \varphi(z,v,0)| + d_M(f(x,y,\varepsilon), f(z,v)) \le L(\varepsilon + |x-z| + d_M(y,v)) \quad (3.11)$$

$$|\Phi(x,y,\varepsilon) - \Phi(z,v,0)| + d_M(f(x,y,\varepsilon), f(z,v)) \le L(\varepsilon + |x-z| + d_M(y,v)) \quad (3.12)$$

for $\forall\, x, z \in \mathbb{R}^d$ and $\forall\, y, v \in M$, and where $L > 2$ is independent of ε then Theorem 1 is still valid.

Here follow some classes of examples for which the above assumptions can be verified.

Example. Consider a system (3.1) where the slow system has the form

$$\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon) = G(X^\varepsilon(n)) + \varepsilon\alpha(X^\varepsilon(n), \varepsilon) + \varepsilon\beta(Y^\varepsilon(n), \varepsilon), \quad (3.13)$$

and the fast system is given by

$$f(X^\varepsilon(n), Y^\varepsilon(n), 0) = \gamma(Y^\varepsilon(n)). \quad (3.14)$$

Assume that system (3.1) is measure-preserving. Assume also that $\gamma(Y^\varepsilon(n))$ is ergodic with f-invariant measure μ_0. In this case (3.5) follows automatically by Remark 5 since the unperturbed fast system is independent of x and consequently the ergodic measure of the fast space is also independent of x. Now consider the following estimate, which comes from the definition of the set of slow ergodization (3.6), and let us rewrite it using (3.13)

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}\Phi(x, f_x^k y) - \overline{\Phi}(x)\right| = \left|\frac{1}{n}\sum_{k=0}^{n-1}(\alpha(x,0) + \beta(f_x^k y, 0)) - \int_M (\alpha(x,0) + \beta((y),0))d\mu_0(y)\right|$$

$$= \left|\frac{1}{n}\sum_{k=0}^{n-1}\beta(f_x^k y) - \int_M \beta((y),0)d\mu_0(y)\right|. \quad (3.15)$$

Since F is volume-preserving, condition (3.7) is replaced by (3.10). Next, the special form of the fast system (3.14) gives us $f_x^k y = \gamma^k y$, hence the set of slow ergodization $E(n_0^\varepsilon, \delta)$ is the set of all points (x,y) satisfying

$$\left|\frac{1}{n_0^\varepsilon}\sum_{k=0}^{n_0^\varepsilon-1}\beta(f_x^k y)-\int_M \beta(y)d\mu_0(y)\right| = \left|\frac{1}{n_0^\varepsilon}\sum_{k=0}^{n_0^\varepsilon-1}\beta(\gamma^k y)-\int_M \beta(y)d\mu_0(y)\right| > \delta. \quad (3.16)$$

The Birkhoff Ergodic Theorem, e.g. [17], implies

$$\left|\frac{1}{n}\sum_{k=0}^{n}\beta(\gamma^k y)-\int_M \beta(y)d\mu_0(y)\right| \to 0 \quad \text{a.e. } y \text{ as } n\to\infty. \quad (3.17)$$

Since $n_0^\varepsilon \to \infty$ as $\varepsilon \to \infty$, there is a sequence $d(n_0^\varepsilon,\delta)$ such that

$$\mu_0(E_y(n_0^\varepsilon,\delta)) = \mu_0\left\{y\in M : \left|\frac{1}{n_0^\varepsilon}\sum_{k=0}^{n_0^\varepsilon-1}\beta(\gamma^k y)-\int_M \beta(y)d\mu_0(y)\right| > \delta\right\} < d(n_0^\varepsilon,\delta), \quad (3.18)$$

and $d(n_0^\varepsilon,\delta)\to 0$ as $\varepsilon\to 0$. Consequently

$$\mu\{(x,y)\in K\times M\cap E(n_0^\varepsilon,\delta)\} = \int_K \mu_0(E_y(n_0^\varepsilon,\delta))dl(x) < d(n_0^\varepsilon,\delta)\cdot l(K), \quad (3.19)$$

and (3.10) is satisfied as required. All other assumptions are relatively easy to verify and the averaging principle of Theorem 1 applies to the map with φ, f of the form (3.13) and (3.14).

Next follows another class of examples where we can apply our averaging mechanism, but first we introduce the concept of equidistribution.

Definition 2. A sequence $\{b_n\}$ of real numbers is said to be equidistributed on an interval (a,b) if for any subinterval $(c,d)\subset(a,b)$ we have

$$\lim_{n\to\infty}\frac{\#(\{b_1,b_2,\dots,b_n\}\cap(c,d))}{n} = \frac{d-c}{b-a},$$

where $\#(A)$ denotes the cardinality of A.

The definition of equidistribution generalizes to the d-dimensional torus as follows.

Definition 3. Let $\{x_n\}$ be a sequence of vectors where $x_n\in\mathbb{T}^d=(0,1)^d\ \forall n\in\mathbb{N}$. $\{x_n\}$ is said to be equidistributed on the unit torus $\mathbb{T}^d$ if for any subset

$$S=(a_1,b_1)\times(a_2,b_2)\times\cdots\times(a_d,b_d)\subset\mathbb{T}^d$$

we have

$$\lim_{n\to\infty}\frac{\#(\{x_1,x_2,\dots,x_n\}\cap S)}{n} = (b_1-a_1)(b_2-a_2)\cdots(b_n-a_n),$$

where $\#(A)$ denotes the cardinality of A.

The following theorem by Weyl [18] (Satz 3) gives an algebraic method to check whether a sequence is equidistributed.

Theorem 2 (Weyl's Theorem). *Let $\{b_n\}$ be a sequence of vectors, where $b_n \in \mathbb{R}^d \; \forall \, n \in \mathbb{N}$. Then $\{b_n\}$ is uniformly distributed modulo* 1 *(or equidistributed in $\mathbb{T}^d$, the $d-\dim$ torus) if and only if $\forall \, l \in \mathbb{Z}^d \backslash \{0\}$ we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} e^{2\pi i l \cdot b_j} = \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} e^{2\pi i (l_1 b_{j,1} + l_2 b_{j,2} + \cdots + l_d b_{j,d})} = 0. \tag{3.20}$$

Now, let us consider the sequence y_l defined by the unperturbed fast map, $y_l = f_x^l(y) = f^l(x, y, 0)$. We construct a class of examples for which Theorem 1 is applicable using Weyl's theorem.

Proposition 3. *Let the fast manifold $M = \mathbb{T}^d$. Consider system* (3.1). *Suppose that the function $\Phi(x, y)$ is a trigonometric polynomial in y. Also assume that f_x preserves the Lebesgue measure on $\mathbb{T}^d$ and that y_l is equidistributed on $\mathbb{T}^d$ for a.e. y. Finally assume that the full map F preserves the Lebesgue measure on $\mathbb{R}^d \times M$. Then F satisfies the assumptions of Theorem* 1.

Proof. Let us consider the set of slow ergodization $E(n_0^\varepsilon, \delta)$ defined by (3.6). Now we show that the expression between the modulus signs tends to 0 as $\varepsilon \to 0$ for almost every (x, y). The first step is to expand $\Phi(x, y)$ into Fourier series,

$$\Phi(x, y) = \sum_{k \in \mathbb{Z}^d} \Phi_k(x) e^{2\pi i k \cdot y}, \tag{3.21}$$

where the Fourier coefficients $\Phi_k(x)$ are given by

$$\Phi_k(x) = \int_{\mathbb{T}^d} \Phi(x, y) e^{-2\pi i k \cdot y} dy. \tag{3.22}$$

Note that $\overline{\Phi}(x)$ coincides with the Fourier coefficient with $k = 0$,

$$\overline{\Phi}(x) = \int_{\mathbb{T}^d} \Phi(x, y) d\rho(y) = \Phi_0(x). \tag{3.23}$$

Now using (3.21), (3.22) and (3.23) we can rewrite the expression to be estimated in the set of slow ergodization (3.6) as

$$\begin{aligned} \left| \frac{1}{n_0^\varepsilon} \sum_{l=0}^{n_0^\varepsilon - 1} \Phi(x, f_x^l y) - \overline{\Phi}(x) \right| &= \left| \frac{1}{n_0^\varepsilon} \sum_{l=0}^{n_0^\varepsilon - 1} \sum_{k \in \mathbb{Z}^d} \Phi_k(x) e^{2\pi i k \cdot f_x^l(y)} - \Phi_0(x) \right| \\ &= \left| \frac{1}{n_0^\varepsilon} \sum_{l=0}^{n_0^\varepsilon - 1} \sum_{k \in \mathbb{Z}^d \backslash \{0\}} \Phi_k(x) e^{2\pi i k \cdot f_x^l(y)} \right|. \end{aligned} \tag{3.24}$$

Changing the order of summation gives us

$$\left| \frac{1}{n_0^\varepsilon} \sum_{l=0}^{n_0^\varepsilon - 1} \Phi(x, f_x^l y) - \overline{\Phi}(x) \right| = \left| \frac{1}{n_0^\varepsilon} \sum_{k \in \mathbb{Z}^d \backslash \{0\}} \Phi_k(x) \sum_{l=0}^{n_0^\varepsilon - 1} e^{2\pi i k \cdot f_x^l(y)} \right|. \tag{3.25}$$

Now, since $\Phi(x, y)$ is a trigonometric polynomial only a finite number of its Fourier coefficients are nonzero. Suppose all nonzero Fourier coefficients have index $|k| < N$. Using this and that $\Phi(x, y)$ is bounded gives us

$$\left|\frac{1}{n_0^\varepsilon}\sum_{l=0}^{n_0^\varepsilon-1}\Phi(x, f_x^l y) - \overline{\Phi}(x)\right| = \left|\frac{1}{n_0^\varepsilon}\sum_{\substack{k\in\mathbb{Z}^d\setminus\{0\}\\ |k|\le N}}\Phi_k(x)\sum_{l=0}^{n_0^\varepsilon-1} e^{2\pi i k\cdot f_x^l(y)}\right|$$

$$\le L \sum_{\substack{k\in\mathbb{Z}^d\setminus\{0\}\\ |k|\le N}} \frac{1}{n_0^\varepsilon}\left|\sum_{l=0}^{n_0^\varepsilon-1} e^{2\pi i k\cdot f_x^l(y)}\right|. \tag{3.26}$$

Consider the inner sum on the right-hand side. Since $f_x^l(y)$ is equidistributed on $\mathbb{T}^d$

$$\lim_{n_0^\varepsilon\to\infty}\frac{1}{n_0^\varepsilon}\left|\sum_{l=0}^{n_0^\varepsilon-1} e^{2\pi i k\cdot f_x^l(y)}\right| = 0 \quad \forall\, k\in\mathbb{Z}^d\setminus\{0\}. \tag{3.27}$$

Since we are only summing over a finite number of terms in (3.26) and each of these tends to zero as $n_0^\varepsilon \to \infty$ we get

$$\lim_{n_0^\varepsilon\to\infty}\left|\frac{1}{n_0^\varepsilon}\sum_{l=0}^{n_0^\varepsilon-1}\Phi(x, f_x^l y) - \overline{\Phi}(x)\right| = 0$$

for almost every $y \in M$. Since $n_0^\varepsilon \to \infty$ as $\varepsilon \to \infty$, there is a sequence $d(n_0^\varepsilon, \delta)$ such that

$$\mu_0(E_y(n_0^\varepsilon, \delta)) = \mu_0\left\{y\in M : \left|\frac{1}{n_0^\varepsilon}\sum_{l=0}^{n_0^\varepsilon-1}\Phi(x, f_x^l y) - \overline{\Phi}(x)\right| > \delta\right\} < d(n_0^\varepsilon, \delta), \tag{3.28}$$

and $d(n_0^\varepsilon, \delta) \to 0$ as $\varepsilon \to 0$. Consequently

$$\mu\{(x, y) \in K \times M \cap E(n_0^\varepsilon, \delta)\} = \int_K \mu_0(E_y(n_0^\varepsilon, \delta))dl(x) < d(n_0^\varepsilon, \delta)\cdot l(k), \tag{3.29}$$

and (3.10) is satisfied as required. All other assumptions are relatively easy to verify and the averaging principle of Theorem 1 applies to the map with φ, f of the form (3.13) and (3.14). □

Now we will turn to another class of examples where in addition to checking the assumptions of Theorem 1 we can estimate the rate of convergence of the limit (3.8). We achieve this by estimating how fast the set of slow ergodization is vanishing. In Chapter 1 of [8] Kachurovskii points out that in general the rate of convergence of ergodic averages depends on the pair (Φ, f). This dependence is complicated and at present not yet fully understood. There is a remarkable class of functions where the rate of convergence can be explicitly bounded. The following assumptions will all be placed on the fast variables for each slow variable x fixed. Let $f_x(y) = (x, y, 0)$ be an ergodic automorphism on every fibre x and $\Phi(x, y, 0) \in L^1(M)$ for every x. Then $\Phi \in L^1(M)$ is said to be cohomologous to zero if it can be written as $\Phi = h \circ f_x - h$ for some $h \in L^1(M)$.

We will use the following theorem for functions f_x and $\Phi(x, y, 0)$ for x fixed, i.e. on each fibre.

Theorem 4 ([8]). *Let f be an ergodic automorphism of M. For any $\Phi \in L^1(M)$ the following conditions are equivalent*

$$\begin{aligned} &(1) \quad \text{There is a constant } C \text{ such that } |A_n\Phi| \le C/n \quad \text{a.e. for all } n,\\ &(2) \quad \mu_x\{A_n\Phi = O(1/n) \text{ as } n \to \infty\} > 0,\\ &(3) \quad \Phi = h \circ f - h \text{ for some } h \in L^\infty, \end{aligned} \tag{3.30}$$

where $A_n\Phi = \frac{1}{n}\sum_{k=0}^{n-1} \Phi \circ f^k$ and μ_x is the ergodic measure.

See [8] for a proof and discussion. If $f_x(y) = f(x, y, 0)$ is an ergodic automorphism on every fibre x and $\Phi(x, y, 0) \in L^1(M)$ for every x and $\Phi(x, y, 0)$ is cohomologous to zero for every x, then Condition 1 is automatically satisfied and furthermore we obtain the rate of convergence in Theorem 1.

Theorem 5. *Suppose that the Lipschitz and boundedness assumptions on the functions in (3.1) hold as well as (3.5). If $f_x(y)$ is an ergodic automorphism on every fibre x and $\Phi(x, y, 0) \in L^1(M)$ is cohomologous to zero for every x, then for any $T > 0$ and any compact set $K \subset \mathbb{R}^d$*

$$\int_K \int_M \sup_{0 \le n \le \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X^\varepsilon_x}(n)| d\mu_x(y) dl(x) = O\left(\left(\log \frac{1}{\varepsilon}\right)^{\alpha - 1}\right), \tag{3.31}$$

as $\varepsilon \to \infty$ for some $\alpha \in (0, 1)$.

The proof of this theorem is placed in Sec. 5.

4. Proof of Theorem 1

To prove Theorem 1 we use Lemma 1 and two auxiliary lemmas.

For ease of notation define $R_k(x, y)$ to be the distance between the kth iterate of the fast component of the true orbit and the kth iterate of the unperturbed fast map, i.e.

$$R_k(x, y) = d_M(Y^\varepsilon_{x,y}(k), f^k_x y). \tag{4.1}$$

Then we can state the following lemma about the distance between the true orbit and the averaged one.

Lemma 2. *Suppose that the Lipschitz and boundedness assumptions on the functions in (3.1) hold as well as (3.5). Let $1 \le n \le \frac{T}{\varepsilon}$ and $N(\varepsilon)$ be the integer part of $\frac{n}{n_0^\varepsilon}$. Then for $\forall\, x \in \mathbb{R}^d$, $y \in M$, $\varepsilon \in (0, 1)$ and $\delta > 0$,*

$$\begin{aligned} &\sup_{0 \le n \le \frac{T}{\varepsilon}} |X^\varepsilon_{x,y}(n) - \overline{X^\varepsilon_x}(n)| \\ &\quad \le \varepsilon e^{LT} \frac{e^{\varepsilon T(L^2 + L)}}{1 + \varepsilon L + \varepsilon^2(L^2 + L)} \left(\varepsilon LT + 2\varepsilon n_0^\varepsilon L + \varepsilon L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x^\varepsilon_j, y^\varepsilon_j) + \delta T \right.\\ &\qquad \left. + \varepsilon n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon, \delta)}(x^\varepsilon_j, y^\varepsilon_j) + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \right), \end{aligned} \tag{4.2}$$

where $\mathbf{1}_{E(n_0^\varepsilon,\delta)}(x,y)$ *is the characteristic function*, i.e. $\mathbf{1}_{E(n_0^\varepsilon,\delta)}(x,y) = 1$ *if* $(x,y) \in E(n_0^\varepsilon,\delta)$ *and* $= 0$ *otherwise.*

Proof. The proof of this lemma may seem lengthy, but each step is straightforward. We start with some general estimates on the entire time domain before breaking it down into smaller intervals to get sharper estimates.

First we estimate the difference between the nth iterate of the slow component of the true orbit and the averaged orbit. Using the first line of (3.1) and the averaged equation (3.9) repeatedly we obtain

$$|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| = \left| X^\varepsilon(0) + \varepsilon \sum_{k=0}^{n-1} \varphi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \overline{X^\varepsilon}(0) - \varepsilon \sum_{k=0}^{n-1} \overline{\varphi}(\overline{X^\varepsilon}(k), \varepsilon) \right|. \quad (4.3)$$

We note that $X^\varepsilon(0) = \overline{X^\varepsilon}(0)$. Then we substitute φ from (3.2) and $\overline{\varphi}$ from (3.3) into the expression and rearrange the terms

$$\begin{aligned} &|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \\ &\quad = \varepsilon \left| \sum_{k=0}^{n-1} (G(X^\varepsilon(k) - G(\overline{X^\varepsilon}(k)))) + \varepsilon \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \overline{\Phi}(\overline{X^\varepsilon}(k))) \right|. \end{aligned} \quad (4.4)$$

Since G is Lipschitz, $|G(x) - G(z)| \le L|x - z|$, the triangle inequality gives

$$\begin{aligned} &|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \\ &\quad \le \varepsilon L \sum_{k=0}^{n-1} |X^\varepsilon(k) - \overline{X^\varepsilon}(k)| + \varepsilon^2 \left| \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \overline{\Phi}(\overline{X^\varepsilon}(k))) \right|. \end{aligned} \quad (4.5)$$

Using the triangle inequality again, we bound the last term of the inequality

$$\begin{aligned} &\varepsilon^2 \left| \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \overline{\Phi}(\overline{X^\varepsilon}(k))) \right| \\ &\quad \le \varepsilon^2 \sum_{k=0}^{n-1} |\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \Phi(X^\varepsilon(k), Y^\varepsilon(k))| \\ &\qquad + \varepsilon^2 \left| \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))) \right| + \varepsilon^2 \sum_{k=0}^{n-1} |\overline{\Phi}(X^\varepsilon(k)) - \overline{\Phi}(\overline{X^\varepsilon}(k))|. \end{aligned} \quad (4.6)$$

Let us consider these three terms separately. First we note that the assumption that $\Phi(x,y,\varepsilon)$ is Lipschitz gives us

$$\varepsilon^2 \sum_{k=0}^{n-1} |\Phi(X^\varepsilon(k), Y^\varepsilon(k), \varepsilon) - \Phi(X^\varepsilon(k), Y^\varepsilon(k))| \le \varepsilon^3 n L. \quad (4.7)$$

Then using the definition of the space average (3.4) and the triangle inequality we find that

$$\begin{aligned}
&\varepsilon^2 \sum_{k=0}^{n-1} |\overline{\Phi}(X^\varepsilon(k)) - \overline{\Phi}(\overline{X^\varepsilon}(k))| \\
&\quad \leq \varepsilon^2 \sum_{k=0}^{n-1} \left(\left| \overline{\Phi}(X^\varepsilon(k)) - \int_M \Phi(X^\varepsilon(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) \right| \right. \\
&\qquad \left. + \left| \int_M \Phi(X^\varepsilon(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) - \int_M \Phi(\overline{X^\varepsilon}(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) \right| \right). \qquad (4.8)
\end{aligned}$$

Using assumption (3.5) we get

$$\left| \overline{\Phi}(X^\varepsilon(k)) - \int_M \Phi(X^\varepsilon(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) \right| \leq L^2 |X^\varepsilon(k) - \overline{X^\varepsilon}(k)|, \qquad (4.9)$$

and rewriting the following integrals as one integral and then using that $\Phi(x, y)$ is Lipschitz we obtain

$$\left| \int_M \Phi(X^\varepsilon(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) - \int_M \Phi(\overline{X^\varepsilon}(k), y) d\mu_{\overline{X^\varepsilon}(k)}(y) \right| \leq L|X^\varepsilon(k) - \overline{X^\varepsilon}(k)|. \quad (4.10)$$

Substituting these expressions into (4.8) gives

$$\varepsilon^2 \sum_{k=0}^{n-1} |\overline{\Phi}(X^\varepsilon(k)) - \overline{\Phi}(\overline{X^\varepsilon}(k))| \leq \varepsilon^2 (L^2 + L) \sum_{k=0}^{n-1} |X^\varepsilon(k) - \overline{X^\varepsilon}(k)|. \qquad (4.11)$$

In order to estimate the remaining third term on the right-hand side of (4.6), we split the time interval of length n into $N(\varepsilon)$ sections, each of length n_0^ε. Let us introduce the following notation: set $x_j^\varepsilon = X_{x,y}^\varepsilon(jn_0^\varepsilon)$ and analogously $y_j^\varepsilon = Y_{x,y}^\varepsilon(jn_0^\varepsilon)$, i.e. $(x_j^\varepsilon, y_j^\varepsilon)$ is the coordinate of the orbit at the beginning of the jth section. Note that $n = N(\varepsilon)n_0^\varepsilon + r$, where $0 \leq r < n_0^\varepsilon$ is the remainder and $N(\varepsilon)$ the integer part of $\frac{n}{n_0^\varepsilon}$. This remainder gives rise to the term $2n_0^\varepsilon L$ in the following estimate

$$\begin{aligned}
&\varepsilon^2 \left| \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))) \right| \\
&\quad \leq \varepsilon^2 \sum_{k=0}^{r} (|\Phi(X^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k), Y^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k))| + |\overline{\Phi}(X^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k))|) \\
&\qquad + \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), Y^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k)) - \overline{\Phi}(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k))) \right|. \qquad (4.12)
\end{aligned}$$

Using the boundedness of Φ and then that $r < n_0^\varepsilon$ we find that

$$\begin{aligned}
&\varepsilon^2 \sum_{k=0}^{r} (|\Phi(X^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k), Y^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k))| + |\overline{\Phi}(X^\varepsilon_{x^\varepsilon_{N(\varepsilon)}, y^\varepsilon_{N(\varepsilon)}}(k))|) \\
&\quad \leq \varepsilon^2 \sum_{k=0}^{r} (L + L) \leq 2L\varepsilon^2 n_0^\varepsilon.
\end{aligned}$$

Hence

$$\varepsilon^2 \left| \sum_{k=0}^{n-1} (\Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k))) \right|$$

$$\leq 2L\varepsilon^2 n_0^\varepsilon + \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), Y^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k)) - \overline{\Phi}(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k))) \right|. \quad (4.13)$$

So, let us continue our analysis on each such section of length n_0^ε. To estimate the last term in (4.13) we need to establish a number of auxiliary estimates, starting with equation (4.14) and ending with (4.24). First let us use the triangle inequality and that $\Phi(x, y)$ is Lipschitz to obtain

$$\left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), Y^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k)) - \Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), f^k_{x_j^\varepsilon} y_j^\varepsilon)) \right| \leq L \sum_{k=0}^{n_0^\varepsilon - 1} d_M(Y^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), f^k_{x_j^\varepsilon} y_j^\varepsilon). \quad (4.14)$$

Substituting $R_k(x, y)$ as defined in (4.1) into this expression yields

$$\left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), Y^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k)) - \Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), f^k_{x_j^\varepsilon} y_j^\varepsilon)) \right| \leq L \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon). \quad (4.15)$$

In a similar way we use the triangle inequality and the assumption that $\Phi(x, y)$ is Lipschitz to get

$$\left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k), f^k_{x_j^\varepsilon} y_j^\varepsilon) - \Phi(x_j^\varepsilon, f^k_{x_j^\varepsilon} y_j^\varepsilon)) \right| \leq L \sum_{k=0}^{n_0^\varepsilon - 1} |X^\varepsilon_{x_j^\varepsilon, y_j^\varepsilon}(k) - x_j^\varepsilon|. \quad (4.16)$$

Then, to estimate how far the slow component of the true orbit has moved from its starting point in k iterates we apply the first line of the map (3.1)

$$\begin{cases} X^\varepsilon(n+1) - X^\varepsilon(n) = \varepsilon\varphi(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon) \\ Y^\varepsilon(n+1) = f(X^\varepsilon(n), Y^\varepsilon(n), \varepsilon), \end{cases}$$

repeatedly to obtain

$$|X^\varepsilon_{z,v}(k) - z| = \left| \varepsilon \sum_{n=0}^{k-1} \varphi(X^\varepsilon_{z,v}(n), Y^\varepsilon_{z,v}(n), \varepsilon) + X^\varepsilon_{z,v}(0) - z \right|. \quad (4.17)$$

Noting that $X^\varepsilon_{z,v}(0) = z$ by definition and then using the triangle inequality and the boundedness of φ we get

$$|X^\varepsilon_{z,v}(k) - z| \leq \varepsilon \sum_{n=0}^{k-1} |\varphi(X^\varepsilon_{z,v}(n), Y^\varepsilon_{z,v}(n), \varepsilon)| \leq \varepsilon k L \quad \forall\, z \in \mathbb{R}^d, \forall\, v \in M. \quad (4.18)$$

Now, expand the following expression, rearrange the terms and apply the triangle inequality,

$$\left|\sum_{k=0}^{n_0^\varepsilon-1}\overline{\Phi}(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k))-\overline{\Phi}(x_j^\varepsilon)n_0^\varepsilon\right|\le\sum_{k=0}^{n_0^\varepsilon-1}\left|\left(\overline{\Phi}(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k))-\int_M\Phi(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k),y)d\mu_{x_j^\varepsilon}(y)\right)\right|$$
$$+\sum_{k=0}^{n_0^\varepsilon-1}\int_M|\Phi(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k),y)-\Phi(x_j^\varepsilon,y)|d\mu_{x_j^\varepsilon}(y).$$

Assumption (3.5) gives us the following bound for the first term

$$\sum_{k=0}^{n_0^\varepsilon-1}\left|\left(\overline{\Phi}(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k))-\int_M\Phi(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k),y)d\mu_{x_j^\varepsilon}(y)\right)\right|\le\sum_{k=0}^{n_0^\varepsilon-1}L^2|X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k)-x_j^\varepsilon|, \tag{4.19}$$

and the second term is bounded since $\Phi(x,y)$ is Lipschitz

$$\sum_{k=0}^{n_0^\varepsilon-1}\int_M|\Phi(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k),y)-\Phi(x_j^\varepsilon,y)|d\mu_{x_j^\varepsilon}(y)\le\sum_{k=0}^{n_0^\varepsilon-1}L|X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k)-x_j^\varepsilon|, \tag{4.20}$$

hence

$$\left|\sum_{k=0}^{n_0^\varepsilon-1}\overline{\Phi}(X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k))-\overline{\Phi}(x_j^\varepsilon)n_0^\varepsilon\right|\le(L^2+L)\sum_{k=0}^{n_0^\varepsilon-1}|X^\varepsilon_{x_j^\varepsilon,y_j^\varepsilon}(k)-x_j^\varepsilon|. \tag{4.21}$$

Before collecting the results we need the following estimate. Let $\mathbf{1}_{E(n_0^\varepsilon,\delta)}(z,v)=1$ if $(z,v)\in E(n_0^\varepsilon,\delta)$. Suppose that $(z,v)\notin E(n_0^\varepsilon,\delta)$, then by the definition of the set of slow ergodization (3.6)

$$\left|\sum_{k=0}^{n_0^\varepsilon-1}\Phi(z,f_z^kv)-n_0^\varepsilon\overline{\Phi}(z)\right|\le n_0^\varepsilon\delta. \tag{4.22}$$

Now suppose that $(z,v)\in E(n_0^\varepsilon,\delta)$. In this case we get a larger upper bound. Boundedness of $\Phi(x,y)$ implies that $|\overline{\Phi}|\le L$, hence

$$\left|\sum_{k=0}^{n_0^\varepsilon-1}\Phi(z,f_z^kv)-\overline{\Phi}(z)n_0^\varepsilon\right|\le\sum_{k=0}^{n_0^\varepsilon-1}|\Phi(z,f_z^kv)|+n_0^\varepsilon|\overline{\Phi}(z)|\le 2n_0^\varepsilon L. \tag{4.23}$$

Combining (4.22) and (4.23) we get

$$\left|\sum_{k=0}^{n_0^\varepsilon-1}\Phi(z,f_z^kv)-n_0^\varepsilon\overline{\Phi}(z)\right|\le n_0^\varepsilon(\delta+2L\mathbf{1}_{E(n_0^\varepsilon,\delta)}(z,v)) \tag{4.24}$$

since (z,v) either belongs to $E(n_0^\varepsilon,\delta)$ or it does not.

Now we are ready to continue with our estimate of the third term of (4.6) so let us return to Eq. (4.13) and consider the second term in that expression. Expansion, the triangle inequality (with middle point $\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k), f^k_{x^\varepsilon_j} y^\varepsilon_j))$ and (4.15) yield

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k), Y^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k)) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))) \right|$$
$$\leq \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} R_k(x^\varepsilon_j, y^\varepsilon_j) + \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k), f^k_{x^\varepsilon_j} y^\varepsilon_j) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))) \right|. \tag{4.25}$$

Using the triangle inequality (with middle point $\Phi(x^\varepsilon_j, f^k_{x^\varepsilon_j} y^\varepsilon_j))$ and (4.16), the second term in (4.25) is estimated by the following two terms

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k), f^k_{x^\varepsilon_j} y^\varepsilon_j) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))) \right|$$
$$\leq \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} |X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k) - x^\varepsilon_j| + \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(x^\varepsilon_j, f^k_{x^\varepsilon_j} y^\varepsilon_j) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))) \right|. \tag{4.26}$$

Let us estimate these two terms separately. Equation (4.18) gives us the following bound for the first term

$$\varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} |X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k) - x^\varepsilon_j| \leq \varepsilon^3 L^2 \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} k. \tag{4.27}$$

To estimate the second term we expand the expression, apply the triangle inequality (with middle point $\overline{\Phi}(x^\varepsilon_j))$ and rearrange the terms to get

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(x^\varepsilon_j, f^k_{x^\varepsilon_j} y^\varepsilon_j) - \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k))) \right|$$
$$\leq \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(x^\varepsilon_j, f^k_{x^\varepsilon_j} y^\varepsilon_j) - \overline{\Phi}(x^\varepsilon_j)) \right| + \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} \overline{\Phi}(X^\varepsilon_{x^\varepsilon_j,y^\varepsilon_j}(k)) - n_0^\varepsilon \overline{\Phi}(x^\varepsilon_j) \right|. \tag{4.28}$$

Now, substituting (4.25), (4.26), (4.27) and (4.28) into (4.13) we have

$$\varepsilon^2 \left| \sum_{k=0}^{n-1} \Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k)) \right|$$
$$\leq \varepsilon^2 2 n_0^\varepsilon L + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} R_k(x^\varepsilon_j, y^\varepsilon_j) + \varepsilon^3 L^2 \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} k$$

$$+\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(x_j^\varepsilon, f_{x_j^\varepsilon}^k y_j^\varepsilon) - \overline{\Phi}(x_j^\varepsilon)) \right|$$

$$+\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} \overline{\Phi}(X_{x_j^\varepsilon,y_j^\varepsilon}^\varepsilon(k)) - n_0^\varepsilon \overline{\Phi}(x_j^\varepsilon) \right|. \tag{4.29}$$

Equation (4.24) gives the following estimate for the second last term

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} (\Phi(x_j^\varepsilon, f_{x_j^\varepsilon}^k y_j^\varepsilon) - \overline{\Phi}(x_j^\varepsilon)) \right| \leq \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} (n_0^\varepsilon \delta + n_0^\varepsilon 2L \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon)). \tag{4.30}$$

The last term of (4.29) is bounded in the following way. First apply (4.21)

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} \overline{\Phi}(X_{x_j^\varepsilon,y_j^\varepsilon}^\varepsilon(k)) - n_0^\varepsilon \overline{\Phi}(x_j^\varepsilon) \right| \leq \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} (L^2+L) |X_{x_j^\varepsilon,y_j^\varepsilon}^\varepsilon(k) - x_j^\varepsilon|, \tag{4.31}$$

and then (4.18) to get

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon-1} \overline{\Phi}(X_{x_j^\varepsilon,y_j^\varepsilon}^\varepsilon(k)) - n_0^\varepsilon \overline{\Phi}(x_j^\varepsilon) \right| \leq \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} (L^2+L) \varepsilon k L. \tag{4.32}$$

Substituting (4.30) and (4.32) into (4.29) gives

$$\begin{aligned} &\varepsilon^2 \left| \sum_{k=0}^{n-1} \Phi(X^\varepsilon(k), Y^\varepsilon(k)) - \overline{\Phi}(X^\varepsilon(k)) \right| \\ &\quad \leq \varepsilon^2 2 n_0^\varepsilon L + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} R_k(x_j^\varepsilon, y_j^\varepsilon) + \varepsilon^2 n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) \\ &\qquad + \varepsilon^2 N(\varepsilon) n_0^\varepsilon \left(\delta + \varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \right). \end{aligned} \tag{4.33}$$

We now substitute (4.7), (4.11) and (4.33) into (4.6). Then (4.5) implies

$$\begin{aligned} &|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \\ &\quad \leq \varepsilon L \sum_{k=0}^{n-1} |X^\varepsilon(k) - \overline{X^\varepsilon}(k)| + \varepsilon^3 n L + \varepsilon^2 2 n_0^\varepsilon L + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon-1} R_k(x_j^\varepsilon, y_j^\varepsilon) \\ &\qquad + \varepsilon^2 n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) + \varepsilon^2 N(\varepsilon) n_0^\varepsilon \left(\delta + \varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \right) \\ &\qquad + \varepsilon^2 (L^2 + L) \sum_{k=0}^{n-1} |X^\varepsilon(k) - \overline{X^\varepsilon}(k)|. \end{aligned} \tag{4.34}$$

The last step in this proof is to apply Lemma 1 to (4.34). Let z_k be the difference between the kth iterate of the slow component of the true orbit and the averaged one

$$\begin{aligned} z_k &= |X^\varepsilon(k) - \overline{X^\varepsilon}(k)|, \\ b &= \varepsilon L + \varepsilon^2(L^2 + L), \end{aligned} \tag{4.35}$$

and let a_n be the sequence

$$\begin{aligned} a_n = \varepsilon^3 nL + \varepsilon^2 2n_0^\varepsilon L + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) \\ + \varepsilon^2 n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) + \varepsilon^2 N(\varepsilon) n_0^\varepsilon \left(\delta + \varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2)\right). \end{aligned} \tag{4.36}$$

Obviously $b > 0$. We now show that a_n is increasing with n. Note that $N(\varepsilon) := \lfloor \frac{n}{n_0^\varepsilon} \rfloor$ is a non-decreasing function of n and that all the values involved are non-negative. Consequently all terms of (4.36) are non-decreasing functions of n. Therefore $a_{n+1} - a_n \geq \varepsilon^3 L > 0$. Lemma 1 applied to (4.34) gives $z_n \leq (1+b)^{n-1} a_n$, i.e.

$$\begin{aligned} &|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \\ &\quad \leq (1 + \varepsilon^2(L^2 + L) + \varepsilon L)^{n-1} \left(\varepsilon^3 nL + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) \right. \\ &\qquad \left. + \varepsilon^2 2n_0^\varepsilon L + \varepsilon^2 n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) + \varepsilon^2 N(\varepsilon) n_0^\varepsilon \left(\delta + \varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2)\right) \right). \end{aligned}$$

Using that $0 \leq n\varepsilon \leq T$ and therefore $\varepsilon n_0^\varepsilon N(\varepsilon) \leq T$ we get

$$\begin{aligned} |X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \leq (1 + \varepsilon^2(L^2 + L) + \varepsilon L)^{n-1} \left(\varepsilon^2 LT + 2\varepsilon^2 n_0^\varepsilon L \right. \\ + \varepsilon^2 L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) + \varepsilon^2 n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) \\ \left. + \varepsilon\delta T + T\varepsilon^2 n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \right). \end{aligned} \tag{4.37}$$

So, finally, we have

$$\begin{aligned} |X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \leq \varepsilon e^{LT} \frac{e^{\varepsilon T(L^2+L)}}{1 + \varepsilon L + \varepsilon^2(L^2 + L)} \left(\varepsilon LT + \varepsilon L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) + 2\varepsilon n_0^\varepsilon L \right. \\ \left. + \varepsilon n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \mathbf{1}_{E(n_0^\varepsilon,\delta)}(x_j^\varepsilon, y_j^\varepsilon) + \delta T + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \right), \end{aligned} \tag{4.38}$$

which concludes the proof of Lemma 2. □

Lemma 3. *For $L > 1$ constant, we have*

$$L^{(\log \frac{1}{\varepsilon})^{1-\alpha}} = O(\varepsilon^{-\beta}) \tag{4.39}$$

when $\varepsilon \to 0$ for any $\alpha \in (0,1)$ and $\forall\, \beta > 0$.

Proof. First we manipulate $L^{\log(\frac{1}{\varepsilon})^{1-\alpha}}$ as follows

$$L^{\log(\frac{1}{\varepsilon})^{1-\alpha}} = e^{(\log L)(\log \frac{1}{\varepsilon})^{1-\alpha}} = \varepsilon^{-\log L (\log \frac{1}{\varepsilon})^{-\alpha}}. \tag{4.40}$$

Then for any $\alpha \in (0,1)$ and any $\beta > 0\ \exists\, \varepsilon_0$ such that $(\log L)(\log \frac{1}{\varepsilon})^{-\alpha} < \beta\ \forall\, \varepsilon < \varepsilon_0$, hence

$$L^{(\log \frac{1}{\varepsilon})^{1-\alpha}} = \varepsilon^{-\log L(\log \frac{1}{\varepsilon})^{-\alpha}} = O(\varepsilon^{-\beta}) \quad \varepsilon < \varepsilon_0. \tag{4.41}$$

□

Now let us turn to the proof of Theorem 1. The main components in this proof are an estimate on $R_k(x,y)$, Lemmas 2 and 3.

Proof of Theorem 1. We begin by noting that the definition of $R_k(x,y)$, Eq. (4.1), and the triangle inequality allow us to rewrite $R_k(x,y)$ as follows

$$\begin{aligned} R_k(x,y) &= d_M(f(X^\varepsilon_{x,y}(k-1), Y^\varepsilon_{x,y}(k-1), \varepsilon), f^k_x y) \\ &\le d_M(f(X^\varepsilon_{x,y}(k-1), Y^\varepsilon_{x,y}(k-1), \varepsilon), f(X^\varepsilon_{x,y}(k-1), Y^\varepsilon_{x,y}(k-1))) \\ &\quad + d_M(f(X^\varepsilon_{x,y}(k-1), Y^\varepsilon_{x,y}(k-1)), f(x, Y^\varepsilon_{x,y}(k-1))) \\ &\quad + d_M(f(x, Y^\varepsilon_{x,y}(k-1)), f_x(f^{k-1}_x y)). \end{aligned} \tag{4.42}$$

Each of these three terms is now estimated using the Lipschitz properties of $f(x,y,\varepsilon)$ which yields

$$R_k(x,y) \le L\varepsilon + L|X^\varepsilon_{x,y}(k-1) - x| + L d_M(Y^\varepsilon_{x,y}(k-1), f^{k-1}_x y), \tag{4.43}$$

which, using (4.18) and the definition of $R_k(x,y)$ again, is rewritten as

$$R_k(x,y) \le L\varepsilon + \varepsilon L^2(k-1) + L R_{k-1}(x,y). \tag{4.44}$$

Applying the same method another $k-1$ times gives us

$$\begin{aligned} R_k(x,y) &\le L(\varepsilon + \varepsilon L(k-1) + R_{k-1}(x,y)) \\ &\le L(\varepsilon + \varepsilon L(k-1) + L(\varepsilon + \varepsilon L(k-2) + L(\varepsilon + \varepsilon L(k-3) + \dots \\ &\quad + L(\varepsilon + \varepsilon L(k-k) + R_0(x,y)))\dots) \\ &= \varepsilon L \sum_{l=0}^{k-1} L^l (1 + L(k-l-1)). \end{aligned} \tag{4.45}$$

We can then establish the following upper bound for $R_k(x,y)$

$$R_k(x,y) \le \varepsilon L \sum_{l=0}^{k-1} L^l (1 + Lk) = \varepsilon L(1+Lk)\frac{L^k - 1}{L-1}. \tag{4.46}$$

Now, let us consider the averaged difference between the slow component of the true orbit and the averaged one. First using Lemma 2 we get

$$\int_K\int_M \sup_{0\le n\le \frac{T}{\varepsilon}} \frac{1}{\varepsilon}|X^{\varepsilon}_{x,y}(n) - \overline{X^{\varepsilon}_x}(n)|d\mu(x,y)$$

$$\le \int_K\int_M e^{LT}\frac{e^{\varepsilon T(L^2+L)}}{1+\varepsilon L+\varepsilon^2(L^2+L)}\left(\varepsilon LT + 2\varepsilon n_0^{\varepsilon}L + \varepsilon L\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1} R_k(x_j^{\varepsilon}, y_j^{\varepsilon})\right.$$

$$\left. + \varepsilon n_0^{\varepsilon}2L\sum_{j=0}^{N(\varepsilon)-1}\mathbf{1}_{E(n_0^{\varepsilon},\delta)}(x_j^{\varepsilon}, y_j^{\varepsilon}) + \delta T + T\varepsilon n_0^{\varepsilon}\frac{1}{2}(L^3+2L^2)\right) d\mu(x,y). \tag{4.47}$$

Consider the sum containing $R_k(x,y)$. Using (4.46) gives

$$\begin{aligned}\varepsilon\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1} R_k(x_j^{\varepsilon}, y_j^{\varepsilon}) &\le \varepsilon\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1}\varepsilon L(1+Lk)\frac{L^k-1}{L-1}\\ &\le \varepsilon\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1}\varepsilon L(1+Lk)L^k\\ &= \varepsilon^2 LN(\varepsilon)\sum_{k=0}^{n_0^{\varepsilon}-1}(L^k + LkL^k).\end{aligned} \tag{4.48}$$

Calculating the values of these sums gives

$$\sum_{k=0}^{n_0^{\varepsilon}-1} L^k = \frac{L^{n_0^{\varepsilon}}-1}{L-1} \le L^{n_0^{\varepsilon}}, \tag{4.49}$$

and

$$L\sum_{k=0}^{n_0^{\varepsilon}-1} kL^k = L\left(n_0^{\varepsilon}\frac{L^{n_0^{\varepsilon}}}{L-1} - \frac{L^{n_0^{\varepsilon}}-1}{(L-1)^2}L\right) \le L^2 n_0^{\varepsilon}L^{n_0^{\varepsilon}}. \tag{4.50}$$

Substituting (4.49) and (4.50) into (4.48) gives

$$\varepsilon\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1} R_k(x_j^{\varepsilon}, y_j^{\varepsilon}) \le \varepsilon^2 LN(\varepsilon)(L^{n_0^{\varepsilon}} + L^2 n_0^{\varepsilon}L^{n_0^{\varepsilon}}).$$

Next we apply Lemma 3, which gives

$$\begin{aligned}\varepsilon\sum_{j=0}^{N(\varepsilon)-1}\sum_{k=0}^{n_0^{\varepsilon}-1} R_k(x_j^{\varepsilon}, y_j^{\varepsilon}) &\le \varepsilon^2 LN(\varepsilon)\widetilde{K}(\varepsilon^{-\beta} + L^2 n_0^{\varepsilon}\varepsilon^{-\beta})\\ &\le \varepsilon^2 N(\varepsilon)\widetilde{K}_1(\varepsilon^{-\beta} + n_0^{\varepsilon}\varepsilon^{-\beta}),\end{aligned} \tag{4.51}$$

for some $\beta \in (0,1)$ and $\widetilde{K}, \widetilde{K}_1 > 0$ independent of ε. Using that $N(\varepsilon) \leq \frac{T}{\varepsilon n_0^\varepsilon}$ we obtain

$$\varepsilon \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) \leq T\widetilde{K}_1 \left(\frac{\varepsilon^{1-\beta}}{n_0^\varepsilon} + \varepsilon^{1-\beta} \right). \tag{4.52}$$

Substituting (4.52) into (4.47) and performing the integration gives

$$\begin{aligned} &\int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon^2}} \frac{1}{\varepsilon} \left| X_{x,y}^\varepsilon(n) - \overline{X_x^\varepsilon}(n) \right| d\mu(x,y) \\ &\quad \leq e^{LT} \frac{e^{\varepsilon T(L^2+L)}}{1 + \varepsilon L + \varepsilon^2(L^2+L)} \left(\varepsilon LT + 2\varepsilon n_0^\varepsilon L + LT\widetilde{K}_1 \left(\frac{\varepsilon^{1-\beta}}{n_0^\varepsilon} + \varepsilon^{1-\beta} \right) + \delta T \right. \\ &\qquad \left. + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) + \varepsilon n_0^\varepsilon 2L \sum_{j=0}^{N(\varepsilon)-1} \int_K \int_M \mathbf{1}_{E(n_0^\varepsilon, \delta)}(X_{x,y}^\varepsilon(jn_0^\varepsilon), Y_{x,y}^\varepsilon(jn_0^\varepsilon)) d\mu(x,y) \right). \end{aligned} \tag{4.53}$$

Finally, using assumption (3.7) we get

$$\begin{aligned} &\int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X_{x,y}^\varepsilon(n) - \overline{X_x^\varepsilon}(n)| d\mu(x,y) \\ &\quad \leq e^{LT} \frac{e^{\varepsilon T(L^2+L)}}{1 + \varepsilon L + \varepsilon^2(L^2+L)} \left(\varepsilon LT + 2\varepsilon n_0^\varepsilon L + LT\widetilde{K}_1 \left(\frac{\varepsilon^{1-\beta}}{n_0^\varepsilon} + \varepsilon^{1-\beta} \right) \right. \\ &\qquad \left. + \delta T + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) + 2LTd_{T,K}(n_0^\varepsilon, \delta) \right), \end{aligned} \tag{4.54}$$

where $\beta \in (0,1)$ and $\widetilde{K}_1 > 0$ independent of ε. As we let $\varepsilon \to 0$ and then $\delta \to 0$ we obtain

$$\lim_{\varepsilon \to 0} \int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X_{x,y}^\varepsilon(n) - \overline{X_x^\varepsilon}(n)| d\mu(x,y) = 0. \tag{4.55}$$

□

Let us here point out some limitations of our method.

Remark 7. It is Lemma 3 that determines how n_0^ε should be chosen, not condition (3.7). In fact, allowing for longer intervals n_0^ε, i.e. more time steps in each section, would make it easier to satisfy the condition, that is, a lower rate of ergodization of Φ would be permissible. The choice of $n_0^\varepsilon = (\log \frac{1}{\varepsilon})^{1-\alpha}$ with $\alpha \in (0,1)$ is the optimal choice. It is the fastest growing n_0^ε that satisfies $\varepsilon L^{n_0^\varepsilon} \to 0$ and $\varepsilon n_0^\varepsilon \to 0$ as $\varepsilon \to 0$. It can be mentioned that further weakening of the coupling, say to the kth power of ε (in Eq. (3.2)), will not change this fact as we will (depending on the way the theorem is stated) end up with terms such as $\varepsilon^k L^{n_0^\varepsilon} \to 0$ and $\varepsilon^k n_0^\varepsilon \to 0$ as $\varepsilon \to 0$. The permissible growth rate of $L^{n_0^\varepsilon}$ does not change as we multiply by a factor ε^k.

5. Proof of Theorem 5

In this section we prove Theorem 5. The proof is very similar to the proof of Theorem 1; with the exception that we sharpen the estimates concerning the set of slow ergodization.

Proof of Theorem 5. We first sharpen estimate (4.38) to

$$|X^\varepsilon(n) - \overline{X^\varepsilon}(n)| \leq \varepsilon e^{LT} \frac{e^{\varepsilon T(L^2+L)}}{1+\varepsilon L + \varepsilon^2(L^2+L)} \Bigg(\varepsilon LT + 2\varepsilon n_0^\varepsilon L + \varepsilon L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) + C_1 \frac{T}{n_0^\varepsilon} + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \Bigg), \tag{5.1}$$

for some $C_1 > 0$. Theorem 4 part (3) implies that $\overline{\Phi}(x) = 0$. Then using Theorem 4 part (1) we conclude that

$$\varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} \left| \sum_{k=0}^{n_0^\varepsilon - 1} (\Phi(x_j^\varepsilon, f^k_{x_j^\varepsilon} y_j^\varepsilon) - \overline{\Phi}(x_j^\varepsilon)) \right| \leq \varepsilon^2 \sum_{j=0}^{N(\varepsilon)-1} C_1 \leq \varepsilon^2 N(\varepsilon) C_1 \leq \frac{\varepsilon C_1 T}{n_0^\varepsilon}, \tag{5.2}$$

for some constant $C_1 > 0$ and *a.e.* y. Replacing the estimates (4.22), (4.23), (4.24) and (4.30) in the proof of Theorem 1 by (5.2) and following those calculations through gives (5.1). Therefore

$$\begin{aligned} &\int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X^\varepsilon_x}(n)| d\mu(x,y) \\ &\leq \int_K \int_M e^{LT} \frac{e^{\varepsilon T(L^2+L)}}{1+\varepsilon L + \varepsilon^2(L^2+L)} \Bigg(\varepsilon LT + 2\varepsilon n_0^\varepsilon L \\ &\quad + \varepsilon L \sum_{j=0}^{N(\varepsilon)-1} \sum_{k=0}^{n_0^\varepsilon - 1} R_k(x_j^\varepsilon, y_j^\varepsilon) + C_1 \frac{T}{n_0^\varepsilon} + T\varepsilon n_0^\varepsilon \frac{1}{2}(L^3 + 2L^2) \Bigg) d\mu(x,y). \end{aligned} \tag{5.3}$$

Using (4.52), performing the integration on the right-hand side and factorizing out $1/n_0^\varepsilon$ yields

$$\begin{aligned} &\int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X^\varepsilon_{x,y}(n) - \overline{X^\varepsilon_x}(n)| d\mu(x,y) \\ &\leq e^{LT} \frac{1}{n_0^\varepsilon} \frac{e^{\varepsilon T(L^2+L)}}{1+\varepsilon L + \varepsilon^2(L^2+L)} \Bigg(\varepsilon LT n_0^\varepsilon + 2\varepsilon (n_0^\varepsilon)^2 L \\ &\quad + LT\widetilde{K}_1(\varepsilon^{1-\beta} + \varepsilon^{1-\beta} n_0^\varepsilon) + T\varepsilon (n_0^\varepsilon)^2 \frac{1}{2}(L^3 + 2L^2) + C_1 T \Bigg) l(K), \end{aligned} \tag{5.4}$$

where $l(K)$ is the Lebesgue measure of K. So, as $\varepsilon \to 0$

$$\int_K \int_M \sup_{0 \leq n \leq \frac{T}{\varepsilon}} \frac{1}{\varepsilon} |X_{x,y}^{\varepsilon}(n) - \overline{X_x^{\varepsilon}}(n)| d\mu(x,y) = O\left(\frac{1}{n_0^{\varepsilon}}\right) = O\left(\left(\log \frac{1}{\varepsilon}\right)^{\alpha - 1}\right), \qquad (5.5)$$

for any $\alpha \in (0, 1)$. □

Acknowledgments

The author thanks Dr. Vassili Gelfreich for many useful discussions and helpful suggestions. The author was partially supported by the EU research training network CODY.

References

[1] D. V. Anosov, Averaging in systems of ODEs with rapidly oscillating solutions, *Izv. Akad. Nauk. SSSR* **24** (1960) 721–742 (in Russian).
[2] N. N. Bogoliubov and Y. A. Mitropolsky, Asymptotic methods in the theory of non-linear oscillations, 2nd edn (Hindustan Publishing Corpn., Delhi, 1961).
[3] N. Brännström and V. Gelfreich, Drift of slow variables in slow-fast Hamiltonian systems, *Physica D* **237** (2008) 2913–2912.
[4] D Dolgopyat, Averaging and invariant measures, *Mosc. Math. J.* **5** (2005) 537–576.
[5] H. S. Dumas and F. Golse, The averaging method for perturbations of mixing flows, *Ergod. Th. Dynam. Sys.* **17** (1997) 1339–1358.
[6] S. N. Elaydi, *An Introduction to Difference Equations* (Springer, New York, 1996).
[7] V. Gelfreich and D. Turaev, Unbounded energy growth in Hamiltonian systems with a slowly varying parameter, *Comm. Math. Phys.* **283** (2008) 769–794.
[8] A. G. Kachurovskii, The rate of convergence in ergodic theorems, *Russian Math. Surveys* **51**(4) (1996) 653–703.
[9] Y. Kifer, Some recent advances in averaging, in *Modern Dynamical Systems and Applications*, Editors: M. Brin, B. Hasselblatt and Y. Pesin (Cambridge Univ. Press, Cambridge, 2004), pp. 385–403.
[10] Y. Kifer, Averaging in difference equations driven by dynamical systems, *Astérisque* **287** (2003) 103–123.
[11] Y. Kifer, Averaging principle for fully coupled dynamical systems and large deviations, *Ergod. Th. Dynam. Sys.* **24** (2004) 847–871.
[12] Y. Kifer, Another proof of the averaging principle for fully coupled dynamical systems with hyperbolic fast motions, *Disc. Cont. Dynam. Sys.* **13**(5) (2005) 1187–1201.
[13] Y. Kifer, Convergence, nonconvergence and adiabatic transitions in fully coupled averaging, *Nonlinearity* **21** (2008) T27–T32.
[14] P. Lochak and C. Meunier, *Multiphase Averaging for Classical Systems* (Springer Verlag, New York, 1988).
[15] A. Neishtadt, Averaging in multi-frequency systems II, *Sov. Phys. Dokl.* **21** (1976) 80–82.
[16] J. A. Sanders, F. Verhulst and J. Murdoch, *Averaging Methods in Nonlinear Dynamical Systems*, 2nd edn (Springer, New York, 2007).
[17] P. Walters, *An Introduction to Ergodic Theory* (Springer Verlag, New York, 1982).
[18] H. Weyl, in Selecta Hermann Weyl, Birkhäuser Verlag, Basel and Stuttgart, 1956; reprint of Über die Gleichverteilung von Zahlen mod. Eins, *Mathematische Annalen* **77** (1916) 313–352 (in German).