

Averaging and rates of averaging for uniform families of deterministic fast-slow skew product systems

by

ALEXEY KOREPANOV (Warwick),
ZEMER KOSLOFF (Warwick and Jerusalem)
and IAN MELBOURNE (Warwick)

Abstract. We consider families of fast-slow skew product maps of the form

$$x_{n+1} = x_n + \epsilon a(x_n, y_n, \epsilon), \quad y_{n+1} = T_\epsilon y_n,$$

where T_ϵ is a family of nonuniformly expanding maps, and prove averaging and rates of averaging for the slow variables x as $\epsilon \rightarrow 0$. Similar results are obtained also for continuous time systems

$$\dot{x} = \epsilon a(x, y, \epsilon), \quad \dot{y} = g_\epsilon(y).$$

Our results include cases where the family of fast dynamical systems consists of intermittent maps, unimodal maps (along the Collet–Eckmann parameters) and Viana maps.

1. Introduction. The classical Krylov–Bogolyubov averaging method [32] deals with skew product flows of the form

$$\dot{x} = \epsilon a(x, y, \epsilon), \quad \dot{y} = g(y).$$

Let ν be an ergodic invariant probability measure for the fast flow generated by g . Under a uniform Lipschitz condition on a , it can be shown that solutions to the slow x dynamics, suitably rescaled, converge almost surely to solutions of an averaged ODE $\dot{X} = \bar{a}(X)$ where $\bar{a}(x) = \int a(x, y, 0) d\nu(y)$.

A considerably harder problem is to handle the fully coupled situation

$$\dot{x} = \epsilon a(x, y, \epsilon), \quad \dot{y} = g(x, y, \epsilon).$$

Here it is supposed that there is a distinguished family of ergodic invariant probability measures $\nu_{x,\epsilon}$ for the fast vector fields $g(x, \cdot, \epsilon)$ and the averaged

2010 *Mathematics Subject Classification*: Primary 34C29; Secondary 37D25, 34E13.

Key words and phrases: averaging, deterministic, fast-slow, nonuniformly hyperbolic, dimension reduction.

Received 13 February 2016; revised 14 January 2017.

Published online *.

vector field is given by $\bar{a}(x) = \int a(x, y, 0) d\nu_{x,0}(y)$. The first results on averaging for fully coupled systems were due to Anosov [6] who considered the case where the fast vector fields are Anosov with $\nu_{x,\epsilon}$ absolutely continuous. Convergence here is in probability with respect to Lebesgue measure.

Kifer [23, 24] extended the results of [6] to the case where the fast vector fields are Axiom A (uniformly hyperbolic) with SRB measures ν_ϵ . More generally, Kifer considers the case where $x \mapsto \nu_{x,0}$ is sufficiently regular so that \bar{a} is Lipschitz, and gives necessary and sufficient conditions for averaging to hold. However, the only situations where the conditions in [23, 24] are verified are in the Axiom A case, even though it is hoped [24] that the conditions are verifiable for nonuniformly hyperbolic examples. Analogous results for the discrete time case are obtained in [22]. See also [14, Theorem 5] for certain partially hyperbolic fast vector fields.

Here, we consider an intermediate class of examples that lies between the classical uncoupled situation and the fully coupled systems of [6, 23], namely families of skew products of the form

$$(1.1) \quad \dot{x} = \epsilon a(x, y, \epsilon), \quad \dot{y} = g(y, \epsilon),$$

with distinguished family of ergodic invariant measures ν_ϵ and averaged vector field $\bar{a}(x) = \int a(x, y, 0) d\nu_0(y)$. Notice that in this way we avoid issues of regularity of the averaged vector field \bar{a} , but we still have to deal with the ϵ -dependence of the measures ν_ϵ as well as the fast vector fields. In other words, linear response (differentiability) of the invariant measures is replaced by statistical stability (weak convergence), which is more tractable. Indeed, one aspect of the general framework in this paper is that our averaging theorems hold in a similar generality to the methods of Alves & Viana [2, 5] for proving statistical stability.

Hence, we obtain results on averaging and rates of averaging for a large class of families of skew products (1.1), going far beyond the uniformly hyperbolic setting, both in discrete and continuous time. Our examples include situations where the fast dynamics is given by intermittent maps with arbitrarily poor mixing properties, unimodal maps where linear response fails, and flows built as suspensions over such maps.

We obtain results also on rates of averaging. In the very simple situation $\dot{x} = \epsilon a(y)$, $\dot{y} = g(y)$, where g is a uniformly expanding semiflow or uniformly hyperbolic flow, it is easily seen that the optimal rate of averaging in L^1 is $O(\epsilon^{1/2})$. For systems of the form (1.1), we often obtain the essentially optimal rate $O(\epsilon^{(1/2)^-})$ ⁽¹⁾.

We have chosen to focus in this paper on the case of noninvertible dynamical systems. In this situation, the measures of interest are absolutely

⁽¹⁾ $q-$ denotes $q - a$ for all $a > 0$.

continuous and we are able to present the main ideas without going into the technical issues presented by dealing with nonabsolutely continuous measures as required in the invertible setting. The invertible case will be covered in a separate paper.

Even in the noninvertible setting, our results depend strongly on extensions and clarifications of the classical second order averaging theorem. These prerequisites are presented in the Appendix and may be of independent interest.

The remainder of the paper is organised as followed. In Section 2, we set up the averaging problem for families of fast-slow skew product systems in the discrete time case, leading to a general result, Theorem 2.2, for such systems. In Section 3, we show that Theorem 2.2 leads easily to averaging when the fast dynamics is a family of uniformly expanding maps. Section 4 is the heart of the paper and deals with the case when the fast dynamics is a family of nonuniformly expanding maps. Our main examples are presented in Section 5. In Section 6, we show how the continuous time case reduces to the discrete time case. In Section 7, we present a simple example to show that almost sure convergence fails for families of skew products.

2. General averaging theorem for families of skew products.

Let $T_\epsilon : M \rightarrow M$, $0 \leq \epsilon < \epsilon_0$, be a family of transformations defined on a measurable space M . For each $\epsilon \in [0, \epsilon_0)$, let ν_ϵ denote a T_ϵ -invariant ergodic probability measure on M .

We consider the family of fast-slow systems

$$(2.1) \quad \begin{aligned} x_{n+1}^{(\epsilon)} &= x_n^{(\epsilon)} + \epsilon a(x_n^{(\epsilon)}, y_n^{(\epsilon)}, \epsilon), & x_0^{(\epsilon)} &= x_0, \\ y_{n+1}^{(\epsilon)} &= T_\epsilon y_n^{(\epsilon)}, & y_0^{(\epsilon)} &= y_0, \end{aligned}$$

where the initial condition $x_0^{(\epsilon)} = x_0$ is fixed throughout. The initial condition $y_0 \in M$ is chosen randomly with respect to various measures that are specified in the statements of the results. Here $a : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^d$ is a family of functions satisfying certain regularity hypotheses.

Define $\bar{a}(x) = \int_M a(x, y, 0) d\nu_0(y)$ and consider the ODE

$$(2.2) \quad \dot{X} = \bar{a}(X), \quad X(0) = x_0.$$

We are interested in the convergence, and rate of convergence, of the slow variables $x_n^{(\epsilon)}$, suitably rescaled, to solutions $X(t)$ of this ODE. More precisely, define $\hat{x}^{(\epsilon)} : [0, 1] \rightarrow \mathbb{R}^d$ by setting $\hat{x}^{(\epsilon)}(t) = x_{[t/\epsilon]}^{(\epsilon)}$. We study convergence of the difference

$$z_\epsilon = \sup_{t \in [0, 1]} |\hat{x}^{(\epsilon)}(t) - X(t)|.$$

REMARK 2.1. The restriction to the time interval $[0, 1]$ entails no loss of generality: the results apply to arbitrary bounded intervals by rescaling ϵ .

Regularity assumptions. Given a function $g : \mathbb{R}^d \rightarrow \mathbb{R}^n$, we define $\|g\|_{\text{Lip}} = \max\{|g|_\infty, \text{Lip } g\}$ where $\text{Lip } g = \sup_{x \neq x'} |g(x) - g(x')|/|x - x'|$ and $|x - x'| = \max_{i=1, \dots, n} |x_i - x'_i|$.

In this section, and also in the Appendix, we consider functions $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^n$ where there is no metric structure assumed on M . In that case, $\|g\|_{\text{Lip}} = \sup_{y \in M} \sup_{\epsilon \in [0, \epsilon_0)} \|g(\cdot, y, \epsilon)\|_{\text{Lip}}$. If $E \subset \mathbb{R}^d$, then $\|g|_E\|_{\text{Lip}}$ is computed by restricting to $x, x' \in E$ (and $y \in M, \epsilon \in [0, \epsilon_0)$).

Throughout, we write $D = \frac{\partial}{\partial x}$. If $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^n$, then $Dg : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^{n \times d}$ and $\|Dg\|_{\text{Lip}}$ is defined accordingly. Similarly for $\|Dg|_E\|_{\text{Lip}}$ when $E \subset \mathbb{R}^d$.

Below, $L_1, L_2, L_3 \geq 1$ are constants. We require that a is globally Lipschitz in x :

$$(2.3) \quad \|a\|_{\text{Lip}} \leq L_1.$$

Set $E = \{x \in \mathbb{R}^d : |x - x_0| \leq L_1\}$. We assume that $a|_E$ is differentiable as a function of x with Lipschitz derivative:

$$(2.4) \quad \|Da|_E\|_{\text{Lip}} \leq L_2,$$

and that

$$(2.5) \quad \sup_{x \in E} \sup_{y \in M} |a(x, y, \epsilon) - a(x, y, 0)| \leq L_3 \epsilon.$$

We let $L = \max\{L_1, L_2, L_3\}$.

2.1. Order functions and a general averaging theorem. Define $\bar{a}(x, \epsilon) = \int_M a(x, y, \epsilon) d\nu_\epsilon(y)$ and let $v_{\epsilon, x}(y) = a(x, y, \epsilon) - \bar{a}(x, \epsilon)$. Set

$$\delta_{1, \epsilon} = \sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \epsilon |v_{\epsilon, x, n}|, \quad \text{where} \quad v_{\epsilon, x, n} = \sum_{j=0}^{n-1} v_{\epsilon, x} \circ T_\epsilon^j,$$

$$\delta_{2, \epsilon} = \sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \epsilon |\mathbf{V}_{\epsilon, x, n}|, \quad \text{where} \quad \mathbf{V}_{\epsilon, x, n} = \sum_{j=0}^{n-1} (Dv_{\epsilon, x}) \circ T_\epsilon^j.$$

Then we define the *order function* $\delta_\epsilon = \delta_{1, \epsilon} + \delta_{2, \epsilon} : M \rightarrow \mathbb{R}$.

Finally, define $S_\epsilon = \sup_{x \in E} |\int_M a(x, y, 0) (d\nu_\epsilon - d\nu_0)(y)| + \epsilon$.

THEOREM 2.2. *If $\delta_\epsilon \leq 1/2$, then $z_\epsilon \leq 6e^{2L}(\delta_\epsilon + S_\epsilon)$.*

The proof of Theorem 2.2 is postponed to the Appendix.

REMARK 2.3. For averaging without rates, it suffices instead of condition (2.5) that $\lim_{\epsilon \rightarrow 0} |a(x, y, \epsilon) - a(x, y, 0)| = 0$ for all $x \in E, y \in M$.

REMARK 2.4. As shown in Section 7, almost sure convergence in the averaging theorem is not likely to hold for fast-slow systems of type (2.1). Hence we consider convergence in L^q with respect to certain absolutely continuous probability measures on M . Since $z_\epsilon \leq 2L$ and $\delta_\epsilon \leq 4L$, convergence in L^p is equivalent to convergence in L^q for all $p, q \in (0, \infty)$. For brevity, we restrict statements to convergence in L^1 except when speaking of rates.

REMARK 2.5. If condition (2.4) fails, then all of our results without rates go through unchanged. Moreover, it is still possible to obtain results with rates but usually with weaker rates of convergence (the best rates are $O(\epsilon^{(1/4)-})$ instead of $O(\epsilon^{(1/2)-})$). These results are obtained by using $\delta_{1,\epsilon}$ (first order averaging) instead of $\delta_\epsilon = \delta_{1,\epsilon} + \delta_{2,\epsilon}$ (second order averaging) and can be found in an earlier (much longer) version of this paper [26].

According to Theorem 2.2, results on averaging reduce to estimating the scalar quantity S_ϵ and the random variable $\delta_\epsilon = \delta_\epsilon(y_0)$. These quantities are discussed below in Subsections 2.2 and 2.3 respectively.

2.2. Statistical stability. In this subsection, we suppose that M is a topological space and that the σ -algebra of measurable sets is the σ -algebra of Borel sets. Recall that the family of measures ν_ϵ is *statistically stable* at $\epsilon = 0$ if ν_0 is the weak limit of ν_ϵ as $\epsilon \rightarrow 0$ (written $\nu_\epsilon \rightarrow_w \nu_0$). This means that $\int_M \phi d\nu_\epsilon \rightarrow \int_M \phi d\nu_0$ for all continuous bounded functions $\phi : M \rightarrow \mathbb{R}$.

In the noninvertible setting, often a stronger property known as strong statistical stability holds. Let m be a reference measure on M and suppose that ν_ϵ is absolutely continuous with respect to m for all $\epsilon \geq 0$. Then ν_0 is *strongly statistically stable* if the densities $\rho_\epsilon = d\nu_\epsilon/dm$ satisfy $\lim_{\epsilon \rightarrow 0} R_\epsilon = 0$ where $R_\epsilon = \int_M |\rho_\epsilon - \rho_0| dm$. We note that $S_\epsilon \leq LR_\epsilon + \epsilon$.

PROPOSITION 2.6. *If $\nu_\epsilon \rightarrow_w \nu_0$, then $\lim_{\epsilon \rightarrow 0} S_\epsilon = 0$.*

Proof. Let $A_\epsilon(x) = \int_M a(x, y, 0) d\nu_\epsilon(y) - \int_M a(x, y, 0) d\nu_0(y)$. Let $\delta > 0$. Since $\nu_\epsilon \rightarrow_w \nu_0$, we have $A_\epsilon(x) \rightarrow 0$ for each x , so there exists $\epsilon_x > 0$ such that $|A_\epsilon(x)| < \delta$ for all $\epsilon \in (0, \epsilon_x)$. Moreover, $|A_\epsilon(z)| < 2\delta$ for all $\epsilon \in (0, \epsilon_x)$ and $z \in B_{\delta/(2L)}(x)$. Since E is covered by finitely many such balls $B_{\delta/(2L)}(x)$, there exists $\bar{\epsilon} > 0$ such that $\sup_{x \in E} |A_\epsilon(x)| < 2\delta$ for all $\epsilon \in (0, \bar{\epsilon})$. Hence $\int_M a(x, y, 0) (d\nu_\epsilon - d\nu_0)(y)$ converges to zero uniformly in x . ■

Hence for proving averaging theorems, statistical stability takes care of the term S_ϵ in Theorem 2.2. In specific examples, we are able to appeal to results on statistical stability with rates, yielding effective estimates.

PROPOSITION 2.7. *Let $q \geq 1$. There is a constant $C > 0$ such that*

$$|z_\epsilon|_{L^q(\nu_\epsilon)} \leq C(|\delta_\epsilon|_{L^q(\nu_\epsilon)} + S_\epsilon)$$

for all $\epsilon \in [0, \epsilon_0)$. If the measures ν_ϵ are absolutely continuous with respect to m , then there is a constant $C > 0$ such that

$$|z_\epsilon|_{L^q(\nu_0)} \leq C(|\delta_\epsilon|_{L^q(\nu_\epsilon)} + R_\epsilon^{1/q} + \epsilon)$$

for all $\epsilon \in [0, \epsilon_0)$.

Proof. Let $A_\epsilon = \{y \in M : \delta_\epsilon(y) \leq 1/2\}$. Then Theorem 2.2 applies on A_ϵ and

$$\begin{aligned} \int_M z_\epsilon^q d\nu_\epsilon &= \int_{M \setminus A_\epsilon} z_\epsilon^q d\nu_\epsilon + \int_{A_\epsilon} z_\epsilon^q d\nu_\epsilon \\ &\leq (2L)^q \nu_\epsilon(\delta_\epsilon > 1/2) + (6e^{2L})^q \int_M (\delta_\epsilon + S_\epsilon)^q d\nu_\epsilon \\ &\leq (4L)^q \int_M \delta_\epsilon^q d\nu_\epsilon + (6e^{2L})^q \int_M (\delta_\epsilon + S_\epsilon)^q d\nu_\epsilon \\ &\leq (12e^{2L})^q \int_M (\delta_\epsilon + S_\epsilon)^q d\nu_\epsilon. \end{aligned}$$

Hence

$$|z_\epsilon|_{L^q(\nu_\epsilon)} \leq 12e^{2L} |\delta_\epsilon + S_\epsilon|_{L^q(\nu_\epsilon)} \leq 12e^{2L} |\delta_\epsilon|_{L^q(\nu_\epsilon)} + 12e^{2L} S_\epsilon,$$

yielding the first estimate.

Next,

$$\begin{aligned} \int_M z_\epsilon^q d\nu_0 &= \int_M z_\epsilon^q d\nu_\epsilon + \int_M z_\epsilon^q (d\nu_0 - d\nu_\epsilon) \\ &\leq (12e^{2L})^q \int_M (\delta_\epsilon + S_\epsilon)^q d\nu_\epsilon + (2L)^q R_\epsilon \\ &\leq (12e^{2L})^q \int_M (\delta_\epsilon + LR_\epsilon + \epsilon)^q d\nu_\epsilon + (2L)^q R_\epsilon \\ &\leq (12Le^{2L})^q \int_M (\delta_\epsilon + R_\epsilon^{1/q} + \epsilon)^q d\nu_\epsilon + (2L)^q R_\epsilon \\ &\leq (24Le^{2L})^q \int_M (\delta_\epsilon + R_\epsilon^{1/q} + \epsilon)^q d\nu_\epsilon. \end{aligned}$$

Hence

$$|z_\epsilon|_{L^q(\nu_0)} \leq 24Le^{2L} |\delta_\epsilon + R_\epsilon^{1/q} + \epsilon|_{L^q(\nu_\epsilon)} \leq 24Le^{2L} (|\delta_\epsilon|_{L^q(\nu_\epsilon)} + R_\epsilon^{1/q} + \epsilon),$$

yielding the second estimate. ■

COROLLARY 2.8.

- (a) Assume that statistical stability holds and $\lim_{\epsilon \rightarrow 0} \int_M \delta_\epsilon d\nu_\epsilon = 0$. Then $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\nu_\epsilon = 0$.
- (b) Assume in addition that strong statistical stability holds and μ is a probability measure on M with $\mu \ll \nu_0$. Then $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\mu = 0$.

Proof. Part (a), and part (b) in the special case $\mu = \nu_0$, are immediate from Proposition 2.7. To prove the general case of part (b), suppose for a contradiction that $\int_M z_{\epsilon_k} d\mu \rightarrow b > 0$ along some subsequence $\epsilon_k \rightarrow 0$. Since $\int_M z_{\epsilon_k} d\nu_0 \rightarrow 0$, by passing to a further subsequence we can also suppose that $z_{\epsilon_k} \rightarrow 0$ on a set of full measure with respect to ν_0 and hence with respect to μ . By the bounded convergence theorem, $\int_M z_{\epsilon_k} d\mu \rightarrow 0$, which is the desired contradiction. ■

The next result is useful in situations where ν_0 is absolutely continuous but its support is not the whole of M .

COROLLARY 2.9. *Assume that strong statistical stability holds and*

$$\lim_{\epsilon \rightarrow 0} \int_M \delta_\epsilon d\nu_0 = 0.$$

Suppose further that each T_ϵ is nonsingular with respect to m and that for almost every $y \in M$ there exist $N \geq 1$ and $\epsilon_1 \in (0, \epsilon_0)$ such that $T_\epsilon^N y \in \text{supp } \nu_0$ for all $\epsilon \in [0, \epsilon_1]$. Then $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\mu = 0$ for every probability measure μ on M with $\mu \ll m$.

Proof. First, we note that for all $N \geq 1$, $\epsilon \geq 0$,

$$(2.6) \quad |\delta_\epsilon \circ T_\epsilon^N - \delta_\epsilon|_\infty \leq 8LN\epsilon.$$

By the arguments in the proof of Corollary 2.8, it suffices to prove that $\int_M \delta_\epsilon dm \rightarrow 0$. Suppose this is not the case. By Corollary 2.8(b), $\int_{\text{supp } \nu_0} \delta_\epsilon dm \rightarrow 0$. Hence there exists a subsequence $\epsilon_k \rightarrow 0$ and a subset $A \subset \text{supp } \nu_0$ with $m(\text{supp } \nu_0 \setminus A) = 0$ such that (i) $\delta_{\epsilon_k} \rightarrow 0$ pointwise on A and (ii) $\int_M \delta_{\epsilon_k} dm \rightarrow b > 0$.

Since each T_ϵ is nonsingular, there exists $M' \subset M$ with $m(M') = 1$ such that $M' \cap T_{\epsilon_k}^{-n}(\text{supp } \nu_0 \setminus A) = \emptyset$ for all $k, n \geq 1$. By hypothesis, there is a subset $M'' \subset M'$ with $m(M'') = 1$ such that for any $y \in M''$ there exists $N \geq 1$ such that $T_{\epsilon_k}^N y \in A$ for all k sufficiently large. Hence it follows from (i) and (2.6) that $\delta_{\epsilon_k} \rightarrow 0$ pointwise on M'' . By the bounded convergence theorem, $\int_M \delta_{\epsilon_k} dm \rightarrow 0$. Together with (ii), this yields the desired contradiction. ■

REMARK 2.10. The hypotheses of Corollary 2.9 are particularly straightforward to apply when $\text{supp } \nu_0$ has nonempty interior. This property is automatic for large classes of nonuniformly expanding maps (see [3, Lemma 5.6]; if $G = \text{supp } \nu_0 \cap H(\sigma)$, then \bar{G} contains a disk).

To be more specific, we consider the examples of the logistic family restricted to Collet–Eckmann parameters and of Viana maps, referring forwards to Examples 5.2 and 5.4 respectively. In both cases the attractors $\Lambda_\epsilon = \text{supp } \nu_\epsilon$ have nonempty interior (for the logistic family, it is standard that Λ_ϵ is a finite union of intervals; for the Viana maps, we apply [3, Lemma 5.6]).

In the case of the logistic family, defined on $M = [-1, 1]$, almost every initial condition $y \in M$ is attracted under T_ϵ to Λ_ϵ . The same is true for the Viana maps on restricting to $M = S^1 \times I$.

Choose a nonempty open set $U \subset \Lambda_0$. Almost every point in M is attracted to Λ_0 and almost every point in Λ_0 has an iterate inside U . Hence for almost every fixed $y \in M$, we can choose N so that $T_0^N y \in U$. By continuity, we can choose ϵ_1 so that $T_\epsilon^N y \in U$ for all $\epsilon \in [0, \epsilon_1]$. Hence the hypotheses of Corollary 2.9 are satisfied, and we conclude that $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\mu = 0$ for every absolutely continuous measure μ on M .

2.3. Estimating the order function. By Proposition 2.7 and Corollary 2.8, it remains to deal with the order function δ_ϵ . This is a random variable depending on the initial condition y_0 . Here we give a useful estimate.

Since $\delta_\epsilon = \delta_{1,\epsilon} + \delta_{2,\epsilon}$ and the definition of $\delta_{2,\epsilon}$ is identical to that of $\delta_{1,\epsilon}$ with a replaced by Da , it suffices to consider $\delta_{1,\epsilon}$. From now on, Γ denotes a constant that only depends on d, p, L and whose value may change from line to line.

LEMMA 2.11. *Let μ be a probability measure on M . Then for all $p \geq 0$ and $\epsilon \in [0, \epsilon_0)$,*

$$\int_M \delta_{1,\epsilon}^{p+d+1} d\mu \leq \Gamma \epsilon^p \sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \int_M |v_{\epsilon,x,n}|^p d\mu.$$

Proof. For most of the proof, we work pointwise on M suppressing the initial condition $y_0 \in M$. There exist $\tilde{x} \in E$ and $\tilde{n} \in [0, 1/\epsilon]$ such that $\delta_{1,\epsilon} = \epsilon |v_{\epsilon,\tilde{x},\tilde{n}}|$. Observe that

$$|v_{\epsilon,x,n} - v_{\epsilon,\tilde{x},\tilde{n}}| \leq |v_{\epsilon,x,n} - v_{\epsilon,\tilde{x},n}| + |v_{\epsilon,\tilde{x},n} - v_{\epsilon,\tilde{x},\tilde{n}}| \leq 2L\epsilon^{-1}|x - \tilde{x}| + 2L|n - \tilde{n}|$$

for every $x \in E$ and $n \leq 1/\epsilon$. Define

$$A = \{x \in E : |x - \tilde{x}| \leq \delta_{1,\epsilon}/(8L)\},$$

$$B = \{n \in [0, \epsilon^{-1}] : \epsilon|n - \tilde{n}| \leq \delta_{1,\epsilon}/(8L)\}.$$

Then for every $x \in A$ and $n \in B$ we have $\epsilon |v_{\epsilon,x,n}| \geq \delta_{1,\epsilon}/2$. Moreover, since $\delta_{1,\epsilon} \leq 2L$, we have $\text{Leb}(A) \geq \Gamma \delta_{1,\epsilon}^d$ and $\#B \geq \epsilon^{-1} \delta_{1,\epsilon}/(8L)$. Hence

$$\epsilon^p \sum_{n=0}^{[1/\epsilon]-1} \int_E |v_{\epsilon,x,n}|^p dx \geq (\#B) \text{Leb}(A) (\delta_{1,\epsilon}/2)^p \geq \Gamma \epsilon^{-1} \delta_{1,\epsilon}^{p+d+1}.$$

Finally,

$$\int_M \delta_{1,\epsilon}^{p+d+1} d\mu \leq \Gamma \epsilon^{p+1} \sum_{n=0}^{[1/\epsilon]-1} \int_E \int_M |v_{\epsilon,x,n}|^p d\mu dx \leq \Gamma \epsilon^p \sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \int_M |v_{\epsilon,x,n}|^p d\mu,$$

as required. ■

REMARK 2.12. Often, estimating $\int_M |v_{\epsilon,x,n}| d\mu$ leads to an essentially identical estimate for $\int_M \sup_{1 \leq n \leq 1/\epsilon} |v_{\epsilon,x,n}| d\mu$. In this case, slightly better convergence rates for $\delta_{1,\epsilon}$ can be obtained using the estimate

$$(2.7) \quad \int_M \delta_{1,\epsilon}^{p+d} d\mu \leq \Gamma \epsilon^p \sup_{x \in E} \int_M \sup_{1 \leq n \leq 1/\epsilon} |v_{\epsilon,x,n}|^p d\mu$$

for all $p \geq 0$ and $\epsilon \in [0, \epsilon_0)$.

3. Examples: Uniformly expanding maps. Let $T_\epsilon : M \rightarrow M$ be a family of maps defined on a metric space (M, d_M) with invariant ergodic Borel probability measures ν_ϵ . Let P_ϵ denote the corresponding transfer operators, so $\int_M P_\epsilon v w d\nu_\epsilon = \int_M v w \circ T_\epsilon d\nu_\epsilon$ for all $v \in L^1(\nu_\epsilon)$, $w \in L^\infty(\nu_\epsilon)$.

From now on, we require Lipschitz regularity in the M variables in addition to the \mathbb{R}^d variables as was required in assumptions (2.3) and (2.4). So for $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^n$ we let $\|g\|_{\text{Lip}} = |g|_\infty + \sup_{\epsilon \in [0, \epsilon_0)} \text{Lip } g(\cdot, \cdot, \epsilon)$ where $\text{Lip } g(\cdot, \cdot, \epsilon) = \sup_{x \neq x'} \sup_{y \neq y'} |g(x, y, \epsilon) - g(x', y', \epsilon)| / (|x - x'| + d_M(y, y'))$. We continue to assume conditions (2.3)–(2.5) with this modified definition of $\|\cdot\|_{\text{Lip}}$.

PROPOSITION 3.1. *Suppose that there is a sequence of constants $a_n \rightarrow 0$ such that $\int_M |P_\epsilon^n v - \int_M v d\nu_\epsilon| d\nu_\epsilon \leq a_n \|v\|_{\text{Lip}}$ for all Lipschitz $v : M \rightarrow \mathbb{R}$ and all $n \geq 1$ and $\epsilon \geq 0$. Then $\lim_{\epsilon \rightarrow 0} \int_M \delta_\epsilon d\nu_\epsilon = 0$.*

Proof. We prove the result for $\delta_{1,\epsilon}$ and $\delta_{2,\epsilon}$ separately. By Remark 2.4, we can work in L^q for any choice of q and we take $q = d + 3$.

For every Lipschitz v , we have

$$\begin{aligned} \int_M \left(\sum_{j=0}^{n-1} v \circ T_\epsilon^j \right)^2 d\nu_\epsilon &= \sum_{j=0}^n \int_M v^2 \circ T_\epsilon^j d\nu_\epsilon + 2 \sum_{0 \leq i < j \leq n-1} \int_M v \circ T_\epsilon^i v \circ T_\epsilon^j d\nu_\epsilon \\ &= n \int_M v^2 d\nu_\epsilon + 2 \sum_{1 \leq k \leq n-1} (n-k) \int_M v v \circ T_\epsilon^k d\nu_\epsilon \\ &= n \int_M v^2 d\nu_\epsilon + 2 \sum_{1 \leq k \leq n-1} (n-k) \int_M P_\epsilon^k v v d\nu_\epsilon. \end{aligned}$$

Hence for v Lipschitz and mean zero,

$$\int_M \left(\sum_{j=0}^{n-1} v \circ T_\epsilon^j \right)^2 d\nu_\epsilon \leq b_n \|v\|_{\text{Lip}}^2,$$

where $b_n = n + 2n \sum_{1 \leq k \leq n-1} a_k = o(n^2)$ by the assumption on a_n .

By (2.3), $\|v_{\epsilon,x}\|_{\text{Lip}} \leq 2L$ for all ϵ, x , so $\sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \int_M |v_{\epsilon,x,n}|^2 d\nu_\epsilon \leq 4L^2 b_n$. By Lemma 2.11, it follows that $\lim_{\epsilon \rightarrow 0} \int_M \delta_{1,\epsilon}^{d+3} d\nu_\epsilon = 0$. Similarly, by (2.4), $\|\mathbf{V}_{\epsilon,x}\|_{\text{Lip}} \leq 2L$, so $\sup_{x \in E} \sup_{1 \leq n \leq 1/\epsilon} \int_M |\mathbf{V}_{\epsilon,x,n}|^2 d\nu_\epsilon \leq 4L^2 b_n$ and hence $\lim_{\epsilon \rightarrow 0} \int_M \delta_{2,\epsilon}^{d+3} d\nu_\epsilon = 0$. ■

REMARK 3.2. The proof uses only the fact that $n^{-1} \sum_{k=1}^n a_n \rightarrow 0$.

Proposition 3.1 is useful in situations where T_ϵ is a family of (piecewise) uniformly expanding maps. A general result of Keller & Liverani [21] guarantees uniform spectral properties of the transfer operators P_ϵ under mild conditions, and consequently $\lim_{\epsilon \rightarrow 0} \int_M \delta_\epsilon^q d\nu_\epsilon = 0$ for all q . We mention two situations where this idea can be applied. Again, for brevity we work in L^1 except when discussing convergence rates (see Remark 2.4).

EXAMPLE 3.3 (Uniformly expanding maps). Suppose that $M = \mathbb{T}^k \cong \mathbb{R}/\mathbb{Z}^k$ is a torus with Haar measure m and distance d_M inherited from Euclidean distance on \mathbb{R}^k and normalised so that $\text{diam } M = 1$. We say that a C^2 map $T : M \rightarrow M$ is *uniformly expanding* if there exists $\lambda > 1$ such that $|(DT)_y v| \geq \lambda|v|$ for all $y \in M$ and $v \in \mathbb{R}^k$. There is a unique absolutely continuous invariant probability measure, and the density is C^1 and nonvanishing.

If $T_\epsilon : M \rightarrow M$, $\epsilon \in [0, 1]$, is a continuous family of C^2 maps, each of which is uniformly expanding, with corresponding probability measures ν_ϵ , then it follows from [21] that we are in the situation of Proposition 3.1, and so $\lim_{\epsilon \rightarrow 0} \int_M \delta_\epsilon d\nu_\epsilon = 0$.

Moreover, it is well-known that ν_0 is uniformly equivalent to m and is strongly statistically stable. Hence by Corollary 2.8 we obtain the averaging result $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\nu_\epsilon = 0$ and $\lim_{\epsilon \rightarrow 0} \int_M z_\epsilon d\mu = 0$ for every absolutely continuous probability measure μ .

Suppose further that $T_\epsilon : M \rightarrow M$, $\epsilon \in [0, 1]$, is a C^k family of C^2 maps, for some $k \in (0, 1]$. By standard results (for instance [21] with Banach spaces C^0 and C^1), $R_\epsilon = \int_M |\rho_\epsilon - \rho_0| dm = O(\epsilon^k)$. If $k \in (0, 1/2)$, then by Remark 4.4 below we obtain the convergence rate $O(\epsilon^k)$ for z_ϵ in $L^q(\nu_\epsilon)$ and $L^q(m)$ for all $q > 0$. If $k \geq 1/2$, then the convergence rate for z_ϵ is $O(\epsilon^{(1/2)^-})$.

EXAMPLE 3.4 (Piecewise uniformly expanding maps). Let $M = [-1, 1]$ with Lebesgue measure m . We consider continuous maps $T : M \rightarrow M$ with $T(-1) = T(1) = -1$ such that T is C^2 on $[-1, 0]$ and $[0, 1]$. We require that there exists $\lambda > 1$ such that $T' \geq \lambda$ on $[-1, 0)$ and $T' \leq -\lambda$ on $(0, 1]$. There exists a unique absolutely continuous invariant probability measure with density of bounded variation.

The results are analogous to those in Example 3.3. Let $\epsilon \mapsto T_\epsilon$ be a continuous family of such maps on $[-1, 1]$ with associated measures ν_ϵ . We assume that T_0 is topologically mixing on the interval $[T_0^2(0), T_0(0)]$ and that 0 is not periodic (which guarantees that T_ϵ is mixing for ϵ small).

Then ν_0 is strongly statistically stable, so by Corollaries 2.8 and 2.9 we obtain averaging in $L^1(\nu_\epsilon)$ and also in $L^1(\mu)$ for every absolutely continuous probability measure μ .

Now suppose that $\epsilon \mapsto T_\epsilon$ is a C^1 family of such maps on $[-1, 1]$ with densities $\rho_\epsilon = d\nu_\epsilon/dm$. Keller [20] showed that $\epsilon \mapsto \rho_\epsilon$ is C^{1-} as a map into L^1 densities. Hence we obtain the convergence rate $O(\epsilon^{(1/2)-})$ for z_ϵ in $L^q(\nu_\epsilon)$ for all $q > 0$ and in $L^1(\nu_0)$.

More precisely, [20] shows that $\int_M |\rho_\epsilon - \rho_0| dm = O(\epsilon \log \epsilon^{-1})$. By [8], this estimate is optimal, so this is a situation where linear response fails, in contrast to Example 3.3.

4. Families of nonuniformly expanding maps. In this section, we consider the situation where the fast dynamics is generated by nonuniformly expanding maps T_ϵ , such that the nonuniform expansion is uniform in the parameter ϵ .

In Subsection 4.1, we recall the notion of nonuniformly expanding map. In Subsection 4.2, we describe the uniformity criteria for T_ϵ and state our main result on averaging, Theorem 4.3, for such families. In Subsections 4.3 and 4.4 we establish some basic estimates for uniformly and nonuniformly expanding maps. In Subsection 4.5 we prove Theorem 4.3.

4.1. Nonuniformly expanding maps. Let (M, d_M) be a locally compact separable bounded metric space with finite Borel measure m and let $T : M \rightarrow M$ be a nonsingular transformation for which m is ergodic. Let $Y \subset M$ be a subset of positive measure, and let α be an at most countable measurable partition of Y with $m(a) > 0$ for all $a \in \alpha$. We suppose that there is an L^1 return time function $\tau : Y \rightarrow \mathbb{Z}^+$, constant on each a with value $\tau(a) \geq 1$, and constants $\lambda > 1$, $\eta \in (0, 1]$ and $C_0, C_1 \geq 1$ such that for each $a \in \alpha$:

- (1) $F = T^\tau$ restricts to a (measure-theoretic) bijection from a onto Y .
- (2) $d_M(Fx, Fy) \geq \lambda d_M(x, y)$ for all $x, y \in a$.
- (3) $d_M(T^\ell x, T^\ell y) \leq C_0 d_M(Fx, Fy)$ for all $x, y \in a$ and $0 \leq \ell < \tau(a)$.
- (4) $\zeta = dm|_Y/d(m|_Y \circ F)$ satisfies $|\log \zeta(x) - \log \zeta(y)| \leq C_1 d_M(Fx, Fy)^\eta$ for all $x, y \in a$.

Such a dynamical system $T : M \rightarrow M$ is called *nonuniformly expanding*. We refer to $F = T^\tau : Y \rightarrow Y$ as the *induced map*. (It is not required that τ is the first return time to Y .) It follows from standard results (recalled later) that there is a unique absolutely continuous ergodic T -invariant probability measure ν on M .

REMARK 4.1. The uniformly expanding maps in Example 3.3 are clearly nonuniformly expanding: take $Y = M$, $\eta = 1$, $\tau = 1$. Then conditions (1) and (2) are immediate, (3) is vacuously satisfied, and (4) holds with $C_1 = \sup_{x, y \in M, x \neq y} |(DT_\epsilon)_x - (DT_\epsilon)_y|/d_M(x, y)$.

4.2. Uniformity assumptions. Now suppose that $T_\epsilon : M \rightarrow M$, $\epsilon \in [0, \epsilon_0)$, is a family of nonuniformly expanding maps as defined in Subsection 4.1, with corresponding absolutely continuous ergodic invariant probability measures ν_ϵ .

DEFINITION 4.2. Let $p > 1$. We say that $T_\epsilon : M \rightarrow M$ is a *uniform family* of nonuniformly expanding maps (of order p) if:

- (i) The constants $C_0, C_1 \geq 1$, $\lambda > 1$, $\eta \in (0, 1]$ can be chosen independent of $\epsilon \in [0, \epsilon_0)$.
- (ii) The return time functions $\tau_\epsilon : Y_\epsilon \rightarrow \mathbb{Z}^+$ lie in L^p for all $\epsilon \in [0, \epsilon_0)$, and moreover $\sup_{\epsilon \in [0, \epsilon_0)} \int_{Y_\epsilon} |\tau_\epsilon|^p dm < \infty$.

We can now state our main result for this section. Recall the set up in Section 2.

THEOREM 4.3. *If $T_\epsilon : M \rightarrow M$ is a uniform family of nonuniformly expanding maps of order p , then there is a constant $C > 0$ such that for all $\epsilon \in [0, \epsilon_0)$,*

$$\int_M \delta_\epsilon^{p+d-1} d\nu_\epsilon \leq \begin{cases} C\epsilon^{(p-1)/2}, & p > 2, \\ C\epsilon^{(p-1)^2/p}, & p \in (1, 2]. \end{cases}$$

REMARK 4.4. In the case $p > 2$, it follows from Theorem 4.3 that

$$|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(p-1)/(2(p+d-1))})$$

for all $q \leq p+d-1$. Since δ_ϵ is uniformly bounded, $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(p-1)/(2q)})$ for all $q > p+d-1$. Similar comments apply for $p \in (1, 2]$.

In particular, if p can be taken arbitrarily large in Definition 4.2, then $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(1/2)^-})$ for all $q > 0$.

If in addition ν_0 is strongly statistically stable and $R_\epsilon = \int_M |\rho_\epsilon - \rho_0| dm = O(\epsilon^{(1/2)^-})$, then by Proposition 2.7 we obtain $|z_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(1/2)^-})$ for all $q > 0$ and $|z_\epsilon|_{L^1(\nu_0)} = O(\epsilon^{(1/2)^-})$.

REMARK 4.5. Alves & Viana [5] prove strong statistical stability for a large class of noninvertible dynamical systems. These maps are uniform families of nonuniformly expanding maps in a sense that is very similar to our definition. In fact, their definition almost includes ours, so verifying the assumptions of [5] is almost sufficient to obtain averaging and rates of averaging via Theorem 4.3.

To be more precise, let us momentarily ignore assumption (3) in Subsection 4.1. Then Definition 4.2(i) with $\eta = 1$ is immediate from [5, (U1)], and Definition 4.2(ii) is immediate from [5, (U2')] which follows from their conditions (U1) and (U2).

Hence it remains to discuss (3). This assumption is not explicitly mentioned in [5] since it is not required for the statement of their main results. However, in specific applications, the hypotheses in [5] are often verified via

the method of *hyperbolic times* [1]. When the return time function τ_ϵ is a hyperbolic time, it is automatic that $C_0 = 1$ (see for example [2, Proposition 3.3(3)]).

Alves et al. [4] introduced a general method for constructing inducing schemes, where τ_ϵ is not necessarily a hyperbolic time but is close enough that C_0 can still be chosen uniformly. Alves [2] combined the methods of [4] and [5] to prove statistical stability for large classes of examples. We show now that in the situation discussed in [2], assumption (3) holds with uniform C_0 , and hence our main results hold. Certain quantities $\delta_1 > 0$ and $N_0 \geq 1$ are introduced in [2, Lemma 3.2] and [2, eq. (16)] respectively, and are explicitly uniform in ϵ . Moreover $\tau_\epsilon = n + m$ where n is a hyperbolic time and $m \leq N_0$ (see [2, Section 4.3]), so C_0 depends only on at most N_0 iterates of T_ϵ . The construction in [2] (see in particular [2, proof of Lemma 4.2]) ensures that the derivative of T_ϵ is bounded along these iterates, so assumption (3) holds and C_0 is uniform in ϵ .

We also mention the extension of [4] due to Gouëzel [18] where $C_0 = 1$ (see [18, Theorem 3.1(4)].)

Finally, we note that when [4] is used to obtain polynomial decay of correlations with rate $O(1/n^\beta)$, $\beta > 0$, the resulting uniform family is of order $p = \beta + 1 -$. (Uniformity in ϵ in Definition 4.2(ii) follows from [2, Lemma 5.1].)

4.3. Explicit estimates for uniformly expanding maps. Throughout this subsection, we work with a fixed uniformly expanding map $F : Y \rightarrow Y$ satisfying conditions (1), (2) and (4). Some standard constructions and estimates are described. The main novelty is that we stress the dependence of various constants on the underlying constants C_1 , λ and η . For convenience, we normalise the metric d_M so that $\text{diam } M = 1$.

For $\theta \in (0, 1)$, we define the *symbolic metric* $d_\theta(x, y) = \theta^{s(x, y)}$ where the *separation time* $s(x, y)$ is the least integer $n \geq 0$ such that $F^n x$ and $F^n y$ lie in distinct partition elements. It is assumed that the partition α separates the orbits of F , so $s(x, y)$ is finite for all $x \neq y$, guaranteeing that d_θ is a metric.

Given $\phi : Y \rightarrow \mathbb{R}^d$, we define $\|\phi\|_\theta = |\phi|_\infty + |\phi|_\theta$ where

$$|\phi|_\theta = \sup_{x \neq y} |\phi(x) - \phi(y)| / d_\theta(x, y).$$

Then ϕ is d_θ -Lipschitz if $\|\phi\|_\theta < \infty$.

The assumptions on F guarantee that there exists a unique absolutely continuous F -invariant probability measure μ on Y . Let $P : L^1(Y) \rightarrow L^1(Y)$ denote the (normalised) transfer operator corresponding to F and μ , so $\int_Y \phi \circ F \psi d\mu = \int_Y \phi P\psi d\mu$ for all $\phi \in L^\infty$ and $\psi \in L^1$. Define $g : Y \rightarrow \mathbb{R}$ by setting $g|_a = d\mu|_a / d(\mu \circ F|_a)$ for $a \in \alpha$. Then $(P\phi)(y) = \sum_{a \in \alpha} g(y_a) \phi(y_a)$ where y_a is the unique preimage of y under F lying in a .

LEMMA 4.6. *There exist constants $\theta, \gamma \in (0, 1)$ and $C_2 \geq 1$ depending continuously on λ, η and C_1 such that:*

(a) $d_M(x, y)^\eta \leq d_\theta(x, y)$ for all $x, y \in Y$.

(b) For all $x, y \in a$ and $a \in \alpha$,

$$(4.1) \quad g(y) \leq C_2 \mu(a) \quad \text{and} \quad |g(x) - g(y)| \leq C_2 \mu(a) d_\theta(x, y).$$

(c) Let $\phi : Y \rightarrow \mathbb{R}$ be d_θ -Lipschitz with $\int \phi d\mu = 0$. Then

$$|P^n \phi|_\infty \leq C_2 \gamma^n \|\phi\|_\theta \quad \text{for all } n \geq 1.$$

Proof. Choose $\theta = \lambda^{-\eta}$. Let $n = s(x, y)$. By condition (2),

$$1 \geq \text{diam } Y \geq d_M(F^n x, F^n y) \geq \lambda^n d_M(x, y) = (\theta^{1/\eta})^{-n} d_M(x, y).$$

Hence $d_M(x, y)^\eta \leq \theta^n = d_\theta(x, y)$, proving (a).

By (4) (see Sect. 4.1) and [27, Proposition 2.3], there is a constant K depending continuously on λ, η and C_1 such that $|\log g|_\theta \leq K$. Hence for $y \in a, a \in \alpha$,

$$\mu(a) = \int_a g d\mu \circ F \geq \inf_a g|_a \mu(Fa) = \inf_a g|_a \geq e^{-K} g(y),$$

so $g|_a \leq e^K \mu(a)$. Next, we note the inequality $t - 1 \leq t \log t$, which is valid for all $t \geq 1$. Let $x, y \in a$ and suppose without loss of generality that $g(y) \leq g(x)$. Setting $t = g(x)/g(y) \geq 1$, we get

$$\frac{g(x)}{g(y)} - 1 \leq \frac{g(x)}{g(y)} \log \frac{g(x)}{g(y)} \leq e^K K d_\theta(x, y).$$

Hence $g(x) - g(y) \leq g(y) e^K K d_\theta(x, y) \leq e^{2K} K \mu(a) d_\theta(x, y)$. Hence part (b) holds with $C_2 = e^{2K} K$.

Finally, part (c) follows for example from [27, Proposition 2.5]. ■

4.4. Explicit estimates for nonuniformly expanding maps. In this subsection, we work with a fixed nonuniformly expanding map T satisfying assumptions (1)–(4) and such that $\tau \in L^p$ for some $p > 1$.

There is a standard procedure to pass from an F -invariant ergodic absolutely continuous probability measure μ on Y to a T -invariant ergodic absolutely continuous probability measure ν on M . We briefly describe this procedure, since the construction is required in the proof of Lemma 4.10. Define the Young tower [35]

$$(4.2) \quad \Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \leq \ell < \tau(y)\},$$

with probability measure $\mu_\Delta = \mu \times \{\text{counting}\} / \int_Y \tau d\mu$. Define $\pi_\Delta : \Delta \rightarrow M$ by $\pi_\Delta(y, \ell) = T^\ell y$. Then $\nu = (\pi_\Delta)_* \mu_\Delta$ is the desired probability measure on M .

In the remainder of this subsection, L^q norms of functions defined on Y are computed using μ . For functions on other spaces, the measures are indicated explicitly in the notation.

Given an observable $v : M \rightarrow \mathbb{R}^d$, we define the *induced observable* $V : Y \rightarrow \mathbb{R}$ by

$$V(y) = \sum_{\ell=0}^{\tau(y)-1} v(T^\ell y).$$

If $v : M \rightarrow \mathbb{R}^d$ satisfies $\int_M v d\nu = 0$, then $\int_Y V d\mu = 0$.

PROPOSITION 4.7. *If $v : M \rightarrow \mathbb{R}^d$ is d_M -Lipschitz, then $PV : Y \rightarrow \mathbb{R}^d$ is d_θ -Lipschitz. Moreover, for all $x, y \in a$, $a \in \alpha$,*

$$|V(y)| \leq \tau(a)|v|_\infty, \quad |V(x) - V(y)| \leq C_0\theta^{-1}\tau(a)(\text{Lip } v) d_\theta(x, y),$$

and

$$|PV|_\infty \leq C_2|\tau|_1|v|_\infty, \quad |PV|_\theta \leq C_0C_2\theta^{-1}|\tau|_1\|v\|_{\text{Lip}}.$$

Proof. The estimate for $V(y)$ is immediate. By condition (3) and Lemma 4.6(a),

$$\begin{aligned} |V(x) - V(y)| &\leq (\text{Lip } v) \sum_{\ell=0}^{\tau(a)-1} d_M(T^\ell x, T^\ell y) \leq C_0(\text{Lip } v) \sum_{\ell=0}^{\tau(a)-1} d_M(Fx, Fy) \\ &= C_0\tau(a)(\text{Lip } v)d_M(Fx, Fy) \leq C_0\tau(a)(\text{Lip } v)d_\theta(Fx, Fy)^{1/\eta} \\ &\leq C_0\tau(a)(\text{Lip } v)d_\theta(Fx, Fy) = C_0\theta^{-1}\tau(a)(\text{Lip } v)d_\theta(x, y), \end{aligned}$$

completing the estimates for V .

Next, $(PV)(y) = \sum_{a \in \alpha} g(y_a)V(y_a)$, so by (4.1),

$$|PV|_\infty \leq C_2 \sum_{a \in \alpha} \mu(a)\tau(a)|v|_\infty = C_2|\tau|_1|v|_\infty.$$

Also,

$$\begin{aligned} |(PV)(x) - (PV)(y)| &\leq \sum_{a \in \alpha} |g(x_a) - g(y_a)| |V(x_a)| + \sum_{a \in \alpha} g(y_a) |V(x_a) - V(y_a)| \\ &\leq C_2 \sum_{a \in \alpha} \mu(a)d_\theta(x, y)\tau(a)|v|_\infty + C_2 \sum_{a \in \alpha} \mu(a)C_0\theta^{-1}\tau(a)(\text{Lip } v)d_\theta(x, y), \end{aligned}$$

yielding the estimate for $|PV|_\theta$. ■

PROPOSITION 4.8. *Let $p \geq 1$. There exist $m \in L^p(Y, \mathbb{R}^d)$ and $\chi \in L^\infty(Y, \mathbb{R}^d)$ such that $V = m + \chi \circ F - \chi$ and $m \in \ker P$. Moreover,*

$$\|m\|_p \leq 3C_3|\tau|_p\|v\|_{\text{Lip}} \quad \text{and} \quad \|\chi\|_\infty \leq C_3|\tau|_1\|v\|_{\text{Lip}},$$

where $C_3 = 2C_0C_2^2\theta^{-1}(1 - \gamma)^{-1}$.

Proof. By Proposition 4.7 and Lemma 4.6(c) with $\phi = PV$, for $n \geq 1$,

$$|P^n V|_\infty \leq C_2 \gamma^{n-1} \|PV\|_\theta \leq 2C_0 C_2^2 \theta^{-1} \gamma^{n-1} |\tau|_1 \|v\|_{\text{Lip}}.$$

It follows that $\chi = \sum_{k=1}^\infty P^k V$ lies in L^∞ and $|\chi|_\infty \leq C_3 |\tau|_1 \|v\|_{\text{Lip}}$.

Write $V = m + \chi \circ F - \chi$; then $m \in L^p$ and $Pm = 0$. Finally, $|m|_p \leq |V|_p + 2|\chi|_\infty \leq |\tau|_p \|v\|_\infty + 2|\chi|_\infty \leq 3C_3 |\tau|_p \|v\|_{\text{Lip}}$. ■

COROLLARY 4.9. *Define $V_n = \sum_{j=0}^{n-1} V \circ F^j$. Let $p > 1$. There exists a constant $C_4 \geq 1$ depending only on p and C_3 such that*

$$\left| \max_{1 \leq j \leq n} |V_j| \right|_p \leq C_4 |\tau|_p \|v\|_{\text{Lip}} n^{\max\{1/2, 1/p\}}.$$

Proof. First note that $V_n = m_n + \chi \circ F^n - \chi$ where $m_n = \sum_{j=0}^{n-1} m \circ F^j$. Since $m \in \ker P$, an application of Burkholder's inequality [12] shows that

$$\left| \max_{1 \leq j \leq n} |m_j| \right|_p \leq C(p) |m|_p n^{\max\{1/2, 1/p\}}$$

(see for example [30, proof of Proposition 4.3]). Hence

$$\left| \max_{1 \leq j \leq n} |V_j| \right|_p \leq C(p) |m|_p n^{\max\{1/2, 1/p\}} + 2|\chi|_\infty.$$

The result follows from Proposition 4.8 with $C_4 = 5C_3 C(p)$. ■

LEMMA 4.10. *Let $p > 1$. Let $v_n = \sum_{j=0}^{n-1} v \circ T^j$. Then*

$$\left| \max_{j \leq n} |v_j| \right|_{L^{p-1}(\nu)} \leq 5C_4 |\tau|_p^{p/(p-1)} \|v\|_{\text{Lip}} n^{\max\{1/2, 1/p\}}.$$

Proof. Let $q = p - 1$. Define the tower Δ as in (4.2) with tower map $f : \Delta \rightarrow \Delta$ where

$$f(y, \ell) = \begin{cases} (y, \ell + 1), & \ell \leq \tau(y) - 2, \\ (Fy, 0), & \ell = \tau(y) - 1. \end{cases}$$

Recall that $\mu_\Delta = \mu \times \{\text{counting}\} / \bar{\tau}$ on Δ where $\bar{\tau} = \int_Y \tau d\mu$. Also, $\nu = (\pi_\Delta)_* \mu_\Delta$ where $\pi_\Delta : \Delta \rightarrow M$ is the projection $\pi_\Delta(y, \ell) = T^\ell y$.

Let $\hat{v} = v \circ \pi_\Delta$ and define $\hat{v}_n = \sum_{j=0}^{n-1} \hat{v} \circ f^j$. Then $\int_M |v_n|^q d\nu = \int_\Delta |\hat{v}_n|^q d\mu_\Delta$.

Next, let $N_n : \Delta \rightarrow \{0, 1, \dots, n\}$ be the number of laps by time n ,

$$N_n(y, \ell) = \#\{j \in \{1, \dots, n\} : f^j(y, \ell) \in Y \times \{0\}\}.$$

Then

$$\hat{v}_n(y, \ell) = V_{N_n(y, \ell)}(y) + H \circ f^n(y, \ell) - H(y, \ell),$$

where $H(y, \ell) = \hat{v}_\ell(y, 0)$. Note that $|H(y, \ell)| \leq |v|_\infty \tau(y)$ for all $(y, \ell) \in \Delta$.

Now $f^n(y, \ell) = (F^{N_n(y, \ell)}y, \ell + n - \tau_{N_n(y, \ell)}(y))$, so

$$\begin{aligned} \max_{j \leq n} |H \circ f^j(y, \ell)| &\leq |v|_\infty \max_{j \leq n} \tau(F^{N_j(y, \ell)}y) \leq |v|_\infty \max_{j \leq n} \tau(F^j y) \\ &\leq |v|_\infty \tau(y) + |v|_\infty \max_{1 \leq j \leq n} \tau(F^j y) \\ &= |v|_\infty \hat{\tau}(y, \ell) + |v|_\infty \max_{1 \leq j \leq n} \hat{\tau}(F^j y, \ell), \end{aligned}$$

where $\hat{\tau} : \Delta \rightarrow \mathbb{Z}^+$ is given by $\hat{\tau}(y, \ell) = \tau(y)$.

We estimate the first term in $L^q(\mu_\Delta)$ and the second term in $L^p(\mu_\Delta)$. Using the definition of μ_Δ and the fact that $\bar{\tau} \geq 1$, we obtain

$$\int_{\Delta} \hat{\tau}^q d\mu_\Delta = (1/\bar{\tau}) \int_Y \tau^{q+1} d\mu \leq \int_Y \tau^p d\mu,$$

so $|\hat{\tau}|_{L^q(\mu_\Delta)} \leq |\tau|_p^{p/(p-1)}$. Also,

$$\begin{aligned} \int_{\Delta} \max_{1 \leq j \leq n} \hat{\tau}(F^j y)^p d\mu_\Delta &= (1/\bar{\tau}) \int_Y \tau \max_{1 \leq j \leq n} \tau^p \circ F^j d\mu \\ &\leq (1/\bar{\tau}) \sum_{j=1}^n \int_Y \tau \tau^p \circ F^j d\mu = (1/\bar{\tau}) \sum_{j=1}^n \int_Y P\tau \tau^p \circ F^{j-1} d\mu \\ &\leq (1/\bar{\tau}) \sum_{j=1}^n |P\tau|_\infty \int_Y \tau^p \circ F^{j-1} d\mu = (1/\bar{\tau}) n |P\tau|_\infty \int_Y \tau^p d\mu \\ &\leq C_2 n \int_Y \tau^p d\mu, \end{aligned}$$

where we have used Proposition 4.7 with $v = 1$ (and hence $V = \tau$) for the final inequality. Hence

$$\left| \max_{1 \leq j \leq n} \hat{\tau}(F^j y) \right|_{L^q(\mu_\Delta)} \leq \left| \max_{1 \leq j \leq n} \hat{\tau}(F^j y) \right|_{L^p(\mu_\Delta)} \leq C_2^{1/p} n^{1/p} |\tau|_p.$$

Combining these estimates, we obtain

$$|H|_{L^q(\mu_\Delta)} \leq \left| \max_{j \leq n} |H \circ f^j| \right|_{L^q(\mu_\Delta)} \leq 2C_2^{1/p} |v|_\infty |\tau|_p^{p/(p-1)} n^{1/p}.$$

Next, using Hölder's inequality, we get

$$\begin{aligned} \int_{\Delta} \max_{j \leq n} |V_{N_j(y, \ell)}(y)|^q d\mu_\Delta &\leq \int_{\Delta} \max_{j \leq n} |V_j(y)|^q d\mu_\Delta = (1/\bar{\tau}) \int_Y \tau \max_{j \leq n} |V_j|^q d\mu \\ &\leq |\tau|_p \left| \max_{j \leq n} |V_j|^q \right|_{p/q} = |\tau|_p \left| \max_{j \leq n} |V_j| \right|_p^q. \end{aligned}$$

By Corollary 4.9,

$$\left| \max_{j \leq n} |V_{N_j}| \right|_{L^q(\mu_\Delta)} \leq C_4 |\tau|_p^{p/(p-1)} \|v\|_{\text{Lip}} n^{\max\{1/2, 1/p\}}.$$

By the triangle inequality, using the fact that $C_2^{1/p} \leq C_4$, we obtain

$$\left| \max_{j \leq n} |v_j| \right|_{L^{p-1}(\nu)} = \left| \max_{j \leq n} |\hat{v}_j| \right|_{L^q(\mu_\Delta)} \leq 5C_4 |\tau|_p^{p/(p-1)} \|v\|_{\text{Lip}} n^{\max\{1/2, 1/p\}},$$

as required. ■

4.5. Proof of Theorem 4.3. Define $v_{\epsilon, x}$ and $v_{\epsilon, x, n}$ as in Section 2. Note that $\text{Lip } v_{\epsilon, x} \leq 2L$ for all ϵ, x and $\int_M v_{\epsilon, x} d\nu_\epsilon = 0$.

It follows from Lemma 4.10 that

$$\left| \max_{j \leq n} |v_{\epsilon, x, j}| \right|_{L^{p-1}(\nu_\epsilon)} \leq 10C_4 L |\tau_\epsilon|_p^{p/(p-1)} n^{\max\{1/2, 1/p\}}$$

for all $\epsilon \geq 0$, $x \in \mathbb{R}$ and $n \geq 1$. By (2.7),

$$\begin{aligned} \int_M \delta_{1, \epsilon}^{p+d-1} d\nu_\epsilon &\leq \Gamma \epsilon^{p-1} \sup_{x \in E} \int_M \max_{n \leq 1/\epsilon} |v_{\epsilon, x, n}|^{p-1} d\nu_\epsilon \\ &\leq \Gamma C_4^{p-1} |\tau_\epsilon|_p^p \epsilon^{p-1} \epsilon^{-(p-1) \max\{1/2, 1/p\}} \\ &= \Gamma C_4^{p-1} |\tau_\epsilon|_p^p \epsilon^{p-1} \epsilon^{\min\{-(p-1)/2, -(p-1)/p\}} \\ &= \Gamma C_4^{p-1} |\tau_\epsilon|_p^p \epsilon^{\min\{(p-1)/2, (p-1)^2/p\}}. \end{aligned}$$

We obtain the same estimate for $\delta_{2, \epsilon}$ on replacing $v_{\epsilon, x}$, $v_{\epsilon, x, n}$ by $\mathbf{V}_{\epsilon, x}$, $\mathbf{V}_{\epsilon, x, n}$.

5. Examples: Nonuniformly expanding maps

EXAMPLE 5.1 (Intermittent maps). Let $M = [0, 1]$ with Lebesgue measure m and consider the intermittent maps $T : M \rightarrow M$ given by

$$(5.1) \quad Tx = \begin{cases} x(1 + 2^a x^a), & x \in [0, 1/2], \\ 2x - 1, & x \in (1/2, 1]. \end{cases}$$

These were studied in [28]. Here $a > 0$ is a parameter. For $a \in (0, 1)$ there is a unique absolutely continuous invariant probability measure with C^∞ nonvanishing density.

We consider a family $T_\epsilon : M \rightarrow M$, $\epsilon \in [0, \epsilon_0)$, of such intermittent maps with parameter $a_\epsilon \in (0, 1)$ depending continuously on ϵ . Let ν_ϵ denote the corresponding family of absolutely continuous invariant probability measures.

For each ϵ , we take $Y = [1/2, 1]$ and let $\tau_\epsilon : Y \rightarrow \mathbb{Z}^+$ be the first return time, $\tau_\epsilon(y) = \inf\{n \geq 1 : T_\epsilon^n y \in Y\}$. Define the first return map $F_\epsilon = T_\epsilon^{\tau_\epsilon} : Y \rightarrow Y$. Let $\alpha_\epsilon = \{Y_\epsilon(n), n \geq 1\}$ where $Y_\epsilon(n) = \{y \in Y : \tau_\epsilon(y) = n\}$.

It is standard that each T_ϵ is a nonuniformly expanding map in the sense of Section 4.1 with $\tau_\epsilon \in L^p$ for all $p < 1/a_\epsilon$. Fix $p \in (1, 1/a_0)$ and choose $0 < a_- < a_0 < a_+ < 1$ such that $p < 1/a_+$. Without loss of generality we can shrink ϵ_0 so that $a_\epsilon \in [a_-, a_+]$ for all $\epsilon \in [0, \epsilon_0)$. We show that T_ϵ , $\epsilon \in [0, \epsilon_0)$, satisfies the conditions of Definition 4.2 for this choice of p .

Since $T'_\epsilon \geq 1$ on M and $T'_\epsilon = 2$ on Y , it is immediate that conditions (1)–(3) in Section 4.1 are satisfied with $\lambda = 2$ and $C_0 = 1$.

Recall that $\zeta_\epsilon = dm|_Y/d(m|_Y \circ F_\epsilon)$. Note that $\zeta_\epsilon(y) = 1/F'_\epsilon(y)$. For each $Y_\epsilon(n) \in \alpha_\epsilon$, define the bijection $F_{\epsilon,n} = F_\epsilon|_{Y_\epsilon(n)} : Y_\epsilon(n) \rightarrow M$. Let $G_{\epsilon,n} = (F_{\epsilon,n}^{-1})' = \zeta_\epsilon \circ F_{\epsilon,n}^{-1}$. By [25, Assumption A2 and Theorem 3.1], there is a constant K , depending only on a_- and a_+ , such that $|(\log G_{\epsilon,n})'| \leq K$ for all $\epsilon \in [0, \epsilon_0)$. Hence $|(\log \zeta_\epsilon \circ F_{\epsilon,n}^{-1})'| = |(\log G_{\epsilon,n})'| \leq K$. By the mean value theorem, for $x, y \in Y_\epsilon(n)$,

$$\begin{aligned} |\log \zeta_\epsilon(x) - \log \zeta_\epsilon(y)| &= |(\log \zeta_\epsilon \circ F_{\epsilon,n}^{-1})(F_\epsilon x) - (\log \zeta_\epsilon \circ F_{\epsilon,n}^{-1})(F_\epsilon y)| \\ &\leq K|F_\epsilon x - F_\epsilon y|. \end{aligned}$$

This proves condition (4) in Section 4.1, and so condition (i) in Definition 4.2 is satisfied.

Define $x_1 = 1/2$ and inductively $x_{n+1} < x_n$ (depending on ϵ) with $T_\epsilon x_{n+1} = x_n$. Then $T_\epsilon(Y_\epsilon(n)) = [x_n, x_{n-1}]$ for $n \geq 2$ and it is standard that $x_n = O(1/n^\alpha)$ as a function of n . By [25, Lemma 5.2], there is a constant K , depending only on a_- and a_+ , such that $x_n \leq Kn^{-1/\alpha_+}$ for all $n \geq 1$, $\epsilon \in [0, \epsilon_0)$. Hence

$$m(\tau_\epsilon > n) = m([1/2, (x_n + 1)/2]) = x_n/2 \leq Kn^{-1/\alpha_+}.$$

Since $p < 1/\alpha_+$ it follows that $\sup_{\epsilon \in [0, \epsilon_0)} \int_Y |\tau_\epsilon|^p dm < \infty$ so condition (ii) in Definition 4.2 is satisfied.

By [10, 25], ν_ϵ is strongly statistically stable and the densities ρ_ϵ satisfy $R_\epsilon = \int |\rho_\epsilon - \rho_0| dm = O(a_\epsilon - a_0)$. Hence, by Corollary 2.8, we obtain averaging in L^1 with respect to ν_ϵ , and also with respect to any absolutely continuous probability measure.

Finally, if $\epsilon \mapsto a_\epsilon$ is Lipschitz say, so that $R_\epsilon = O(\epsilon)$, then we obtain the rates described in Remark 4.4 with $p = (1/a_0)_-$.

EXAMPLE 5.2 (Logistic family). We consider the family of quadratic maps $T : [-1, 1] \rightarrow [-1, 1]$ given by $T(x) = 1 - ax^2$, $a \in [0, 2]$, with m taken to be Lebesgue measure.

Let $b, c > 0$. The map T satisfies the *Collet–Eckmann condition* [13] with constants b, c if

$$(5.2) \quad |(T^n)'(1)| \geq ce^{bn} \quad \text{for all } n \geq 0.$$

In this case, we write $a \in Q_{b,c}$. The set of *Collet–Eckmann parameters* is $P_1 = \bigcup_{b,c>0} Q_{b,c}$ and is a Cantor set of positive Lebesgue measure [19, 11]. When $a \in P_1$, the map T has an invariant set Λ consisting of a finite union of intervals with an ergodic absolutely continuous invariant probability measure ν_a . The density for ν_a is bounded below on Λ and lies in L^{2^-} . The invariant set attracts Lebesgue almost every trajectory in $[-1, 1]$.

There is also an open dense set of parameters $P_0 \subset [0, 2]$ for which T has a periodic sink attracting Lebesgue almost every trajectory in $[-1, 1]$. By Lyubich [29], $P_0 \cup P_1$ has full measure. For $a \in P_0$, we let ν_a denote the invariant probability measure supported on the periodic attractor, so we have a map $a \mapsto \nu_a$ defined on $P_0 \cup P_1$.

It is clear that statistical stability holds on P_0 , and that strong statistical stability fails everywhere in $P_0 \cup P_1$. Moreover, Thunberg [33, Corollary 1] showed that on any full measure subset of $E \subset [0, 2]$ the map $a \mapsto \nu_a$ is not statistically stable at any point of $P_1 \cap E$. On the other hand, Freitas & Todd [17] proved that strong statistical stability holds on $Q_{b,c}$ for all constants $b, c > 0$. That is, the map $a \rightarrow \rho_a = d\nu_a/dm$ from $Q_{b,c}$ to L^1 is continuous. (See also [15, 16] for the same result restricted to the Benedicks–Carleson parameters [11].)

We consider families $\epsilon \mapsto T_\epsilon$ where each T_ϵ is a quadratic map with parameter $a = a_\epsilon$ depending continuously on ϵ . Fix $b, c > 0$ such that $a_0 \in Q_{b,c}$. We claim that

$$\lim_{\substack{\epsilon \rightarrow 0 \\ a_\epsilon \in Q_{b,c}}} \int z_\epsilon d\nu_{a_\epsilon} = 0.$$

Moreover, using Corollary 2.9 we obtain convergence in $L^1(\mu)$ for every absolutely continuous probability measure μ . Given the above results on strong statistical stability, it suffices to verify that T_ϵ is a uniform family of nonuniformly expanding maps.

For the Benedicks–Carleson parameters, the method in [15, 16] is the approach of [2] and we can apply Remark 4.5. In the general case, a different method exploiting negative Schwarzian derivative and Koebe spaces [17, proof of Theorem B in Section 6] shows that the conditions in [5] are satisfied. By Remark 4.5, this completes the proof of averaging with the possible exception of condition (3). However, a standard consequence of negative Schwarzian derivative and the Koebe distortion property (as discussed in [17, Lemma 4.1] and used in [17, Remark 3.2]) is that bounded distortion holds at intermediate steps and not just at the inducing time as in condition (4). Hence there is a uniform constant \tilde{C}_1 such that

$$\frac{|T_\epsilon^j x - T_\epsilon^j y|}{\text{diam } T_\epsilon^j a} \leq \tilde{C}_1 \frac{|F_\epsilon x - F_\epsilon y|}{\text{diam } Y_\epsilon}$$

for all partition elements a , all $x, y \in a$ and all $j \leq \tau_\epsilon(a)$. In particular, $|T_\epsilon^j x - T_\epsilon^j y| \leq (2\tilde{C}_1/\text{diam } Y_\epsilon)|F_\epsilon x - F_\epsilon y|$, yielding condition (3) uniformly in ϵ .

Next, we discuss rates of convergence. By [17, Lemma 4.1], condition (ii) in Definition 4.2 is satisfied for any $p > 1$. Hence $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(1/2)^-})$. If $\epsilon \mapsto a_\epsilon$ is C^1 , then it follows from Baladi et al. [9] that $R_\epsilon = \int |\rho_\epsilon - \rho_0| dm = O(\epsilon^{(1/2)^-})$. By Remark 4.4, we obtain averaging with rate $O(\epsilon^{(1/2)^-})$ in $L^q(\nu_\epsilon)$ for all $q > 0$ and in $L^1(\nu_0)$.

EXAMPLE 5.3 (Multimodal maps). Freitas & Todd [17] also consider families of multimodal maps where each critical point c satisfies a Collet–Eckmann condition along the orbit of Tc with constants uniform in ϵ . Hence the averaging result for the quadratic family in Example 5.2 extends immediately to multimodal maps.

EXAMPLE 5.4 (Viana maps). Viana [34] introduced a C^3 open class of multi-dimensional nonuniformly expanding maps $T_\epsilon : M \rightarrow M$. For definiteness, we restrict attention to the case $M = S^1 \times \mathbb{R}$. Let $S : M \rightarrow M$ be the map $S(\theta, y) = (16\theta \bmod 1, a_0 + a \sin 2\pi\theta - y^2)$. Here a_0 is chosen so that 0 is a preperiodic point for the quadratic map $y \mapsto a_0 - y^2$ and a is fixed sufficiently small. Let T_ϵ , $0 \leq \epsilon < \epsilon_0$, be a continuous family of C^3 maps sufficiently close to S . It follows from [1, 5] that there is an interval $I \subset (-2, 2)$ such that, for each $\epsilon \in [0, \epsilon_0)$, there is a unique absolutely continuous T_ϵ -invariant ergodic probability measure ν_ϵ supported in the interior of $S^1 \times I$. Moreover the invariant set $\Lambda_\epsilon = \text{supp } \nu_\epsilon$ attracts almost every initial condition in $S^1 \times I$.

By Alves & Viana [5], ν_0 is strongly statistically stable. Moreover, the inducing method of [4] and the arguments in [2] apply to this example, so T_ϵ is a uniform family of nonuniformly expanding maps by Remark 4.5. Also, Corollary 2.9 is applicable by Remark 2.10. Hence we obtain averaging in $L^1(\nu_\epsilon)$ and in $L^1(\mu)$ for all absolutely continuous μ .

Finally, we discuss rates. Although not stated explicitly in [5], it follows from their estimates that T_ϵ is a uniform family of order p for all p . Hence $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{(1/2)-})$ for all $q > 0$ by Remark 4.4. To verify that p is arbitrary, we mention the following steps in [5] (their q is our p). Note by Remark 4.5 that it remains to verify their condition (U2) for all p . On page 25 (calculation for (U2)), they give an estimate that works for all p depending on two constants C_0 and γ_0 which are uniform in ϵ by [5, Remark 4.6].

6. Averaging for continuous time fast-slow systems. Let $\phi_t^\epsilon : M \rightarrow M$, $0 \leq \epsilon < \epsilon_0$, be a family of semiflows defined on the metric space (M, d_M) . For each $\epsilon \geq 0$, let ν_ϵ denote a ϕ_t^ϵ -invariant ergodic Borel probability measure. Let $a : \mathbb{R}^d \times M \times [0, \epsilon_0) \rightarrow \mathbb{R}^d$ be a family of vector fields on \mathbb{R}^d satisfying conditions (2.3)–(2.5).

We consider the family of fast-slow systems

$$\begin{aligned} \dot{x}^{(\epsilon)} &= \epsilon a(x^{(\epsilon)}, y^{(\epsilon)}, \epsilon), & x^{(\epsilon)}(0) &= x_0, \\ \dot{y}^{(\epsilon)}(t) &= \phi_t^\epsilon y_0, \end{aligned}$$

where the initial condition $x^{(\epsilon)}(0) = x_0$ is fixed throughout. The initial condition $y_0 \in M$ is again chosen randomly with respect to various measures that are specified in the statements of the results.

Define $\hat{x}^{(\epsilon)} : [0, 1] \rightarrow \mathbb{R}^d$ by setting $\hat{x}^{(\epsilon)}(t) = x^{(\epsilon)}(t/\epsilon)$. Let $X : [0, 1] \rightarrow \mathbb{R}^d$ be the solution to the ODE (2.2) and define

$$z_\epsilon = \sup_{t \in [0, 1]} |\hat{x}^{(\epsilon)}(t) - X(t)|.$$

Recall that $E = \{x \in \mathbb{R}^d : |x - x_0| \leq L_1\}$. As in Section 2.1, define $\bar{a}(x, \epsilon) = \int_M a(x, y, \epsilon) d\nu_\epsilon(y)$ and let $v_{\epsilon, x}(y) = a(x, y, \epsilon) - \bar{a}(x, \epsilon)$. We define the *order function* $\delta_\epsilon = \delta_{1, \epsilon} + \delta_{2, \epsilon} : M \rightarrow \mathbb{R}$ by

$$\begin{aligned} \delta_{1, \epsilon} &= \sup_{x \in E} \sup_{0 \leq t \leq 1/\epsilon} \epsilon |v_{\epsilon, x, t}| \quad \text{where} \quad v_{\epsilon, x, t} = \int_0^t v_{\epsilon, x} \circ \phi_s^\epsilon ds, \\ \delta_{2, \epsilon} &= \sup_{x \in E} \sup_{0 \leq t \leq 1/\epsilon} \epsilon |\mathbf{V}_{\epsilon, x, t}| \quad \text{where} \quad \mathbf{V}_{\epsilon, x, t} = \int_0^t (Dv_{\epsilon, x}) \circ \phi_s^\epsilon ds. \end{aligned}$$

The next result is the continuous time analogue of Theorem 2.2. The proof is entirely analogous, and hence is omitted.

THEOREM 6.1. *Let $S_\epsilon = \sup_{x \in E} |\int_M a(x, y, 0) (d\nu_\epsilon - d\nu_0)(y)| + \epsilon$. Assume conditions (2.3)–(2.5) hold. If $\delta_\epsilon \leq 1/2$, then $z_\epsilon \leq 6e^{2L}(\delta_\epsilon + S_\epsilon)$.*

As in the discrete time setting, we say that ν_0 is *statistically stable* if $\nu_\epsilon \rightarrow_w \nu_0$. Proposition 2.6 goes through unchanged and statistical stability implies that $S_\epsilon \rightarrow 0$.

If the measures ν_ϵ are absolutely continuous with respect to a reference measure m on M , we define $\rho_\epsilon = d\nu_\epsilon/dm$ and set $R_\epsilon = \int_M |\rho_\epsilon - \rho_0| dm$. Then ν_0 is strongly statistically stable if $R_\epsilon \rightarrow 0$. Proposition 2.7 and Corollaries 2.8 and 2.9 go through unchanged.

Fix a Borel subset $M' \subset M$ and a reference Borel measure m' on M' . Let $h_\epsilon : M' \rightarrow \mathbb{R}^+$ be a family of Lipschitz functions such that $\phi_{h_\epsilon(y)}^\epsilon(y) \in M'$ for almost all $y \in M'$. The map $T_\epsilon : M' \rightarrow M'$, $T_\epsilon(y) = \phi_{h_\epsilon(y)}^\epsilon(y)$, is then defined almost everywhere.

As usual, we suppose that there is a family ν'_ϵ of ergodic T_ϵ -invariant probability measures on M' . Define the suspension $M_{h_\epsilon} = \{(y, u) \in M' \times \mathbb{R} : 0 \leq u \leq h_\epsilon(y)\} / \sim$ where $(y, h_\epsilon(y)) \sim (T_\epsilon y, 0)$. The suspension semiflow $f_t^\epsilon : M_{h_\epsilon} \rightarrow M_{h_\epsilon}$ is given by $f_t^\epsilon(y, u) = (y, u + t)$ modulo identifications. Let $\bar{h}_\epsilon = \int_{M'} h_\epsilon d\nu'_\epsilon$. Then $\nu''_\epsilon = (\nu'_\epsilon \times \{\text{Lebesgue}\}) / \bar{h}_\epsilon$ is an ergodic absolutely continuous f_t^ϵ -invariant probability measure on M_{h_ϵ} . The projection $\pi_\epsilon : M_{h_\epsilon} \rightarrow M$ given by $\pi_\epsilon(y, u) = \phi_u^\epsilon y$ is a semiconjugacy between f_t^ϵ and ϕ_t^ϵ . Hence $\nu_\epsilon = \pi_{\epsilon*} \nu''_\epsilon$ is an ergodic ϕ_t^ϵ -invariant probability measure on M .

We suppose from now on that there are constants $K_2 \geq K_1 \geq 1$ such that for all $x, y \in M'$ and $\epsilon \in [0, \epsilon_0)$:

- $K_1^{-1} \leq h_\epsilon \leq K_1$, $\text{Lip } h_\epsilon \leq K_1$ and $|h_\epsilon - h_0|_\infty \leq K_1 \epsilon$.
- $d_M(\phi_t^\epsilon x, \phi_t^\epsilon y) \leq K_2 d_M(x, y)$ and $d_M(\phi_t^\epsilon y, \phi_t^0 y) \leq K_2 \epsilon$ for all $t \leq K_1$.

(These assumptions are easily weakened; in particular changing the ϵ estimates to $\epsilon^{1/2}$ will not affect anything.)

PROPOSITION 6.2. *Let $v : M \rightarrow \mathbb{R}^d$ be Lipschitz. Define $\tilde{v} : M' \rightarrow \mathbb{R}^d$ by $\tilde{v}(y) = \int_0^{h_0(y)} v(\phi_u^0 y) du$. Then*

$$\begin{aligned} \int_M v(d\nu_\epsilon - d\nu_0) &\leq 3K_2^4 \|v\|_{\text{Lip}} \epsilon + K_1^3 |v|_\infty \left| \int_{M'} h_0(d\nu'_\epsilon - d\nu'_0) \right| \\ &\quad + K_1 \left| \int_{M'} \tilde{v}(d\nu'_\epsilon - d\nu'_0) \right|. \end{aligned}$$

Proof. We have

$$\begin{aligned} \int_M v(d\nu_\epsilon - d\nu_0) &= \int_{M_{h_\epsilon}} v \circ \pi_\epsilon d\nu''_\epsilon - \int_{M_{h_0}} v \circ \pi_0 d\nu''_0 \\ &= (1/\bar{h}_\epsilon) \int_{M' 0}^{h_\epsilon} v \circ \pi_\epsilon du d\nu'_\epsilon - (1/\bar{h}_0) \int_{M' 0}^{h_0} v \circ \pi_0 du d\nu'_0 \\ &= I_1 + I_2 + I_3 + I_4 \end{aligned}$$

where

$$\begin{aligned} I_1 &= (1/\bar{h}_\epsilon - 1/\bar{h}_0) \int_{M' 0}^{h_\epsilon} v \circ \pi_\epsilon du d\nu'_\epsilon, \\ I_2 &= (1/\bar{h}_0) \int_{M' 0}^{h_\epsilon} (v \circ \pi_\epsilon - v \circ \pi_0) du d\nu'_\epsilon, \\ I_3 &= (1/\bar{h}_0) \left(\int_{M' 0}^{h_\epsilon} v \circ \pi_0 du d\nu'_\epsilon - \int_{M' 0}^{h_0} v \circ \pi_0 du d\nu'_0 \right), \\ I_4 &= (1/\bar{h}_0) \left(\int_{M' 0}^{h_0} v \circ \pi_0 du d\nu'_\epsilon - \int_{M' 0}^{h_0} v \circ \pi_0 du d\nu'_0 \right). \end{aligned}$$

Now

$$\begin{aligned} |I_1| &\leq K_1^2 |\bar{h}_\epsilon - \bar{h}_0| K_1 |v|_\infty \leq K_1^3 |v|_\infty \left(K_1 \epsilon + \left| \int_{M'} h_0(d\nu'_\epsilon - d\nu'_0) \right| \right), \\ |I_2| &\leq K_1^2 \sup_{y \in M' 0 \leq u \leq K_1} (\text{Lip } v) d_M(\phi_u^\epsilon y, \phi_u^0 y) \leq K_1^2 K_2 (\text{Lip } v) \epsilon, \\ |I_3| &\leq K_1 |v|_\infty |h_\epsilon - h_0| \leq K_1^2 |v|_\infty \epsilon, \quad |I_4| \leq K_1 \left| \int_{M'} \tilde{v}(d\nu'_\epsilon - d\nu'_0) \right|. \end{aligned}$$

The result follows from the combination of these estimates. ■

COROLLARY 6.3. *Statistical stability of ν'_0 implies statistical stability of ν_0 .*

Proof. This follows from Proposition 6.2. ■

COROLLARY 6.4. *Suppose that the measures ν'_ϵ are absolutely continuous with respect to m' , with densities $\rho'_\epsilon = d\nu'_\epsilon/dm'$. Then*

$$S_\epsilon \leq 3K_2^4 L \left(\epsilon + \int_{M'} |\rho'_\epsilon - \rho'_0| dm' \right).$$

Proof. This follows from Proposition 6.2 with $v(y) = a(x, y, 0)$ for each fixed x . ■

Next we show how the order function for the flows $\phi_t^\epsilon : M \rightarrow M$ is related to the order function for the maps $T_\epsilon : M' \rightarrow M'$. We restrict attention to $\delta_{1,\epsilon}$ since the corresponding statement for $\delta_{2,\epsilon}$ is identical.

Define the family of induced observables $w_{x,\epsilon} : M' \rightarrow \mathbb{R}$ by

$$w_{\epsilon,x}(y) = \int_0^{h_\epsilon(y)} v_{\epsilon,x}(\phi_u^\epsilon y) du.$$

Note that $\int_{M'} w_{\epsilon,x} d\nu'_\epsilon = 0$ and $w_{\epsilon,x}$ is d_M -Lipschitz with $\|w_\epsilon\|_{\text{Lip}} \leq 2K_1 K_2 L$. Let

$$\Delta_{1,\epsilon} = \sup_{x \in E} \sup_{1 \leq n \leq 1+K_1/\epsilon} \epsilon |w_{\epsilon,x,n}| \quad \text{where} \quad w_{\epsilon,x,n} = \sum_{j=0}^{n-1} w_{\epsilon,x} \circ T_\epsilon^j.$$

We can now state our main result for this section.

LEMMA 6.5. *Let $q \geq 1$. Then $|\delta_{1,\epsilon}|_{L^q(\nu_\epsilon)} \leq |\Delta_{1,\epsilon}|_{L^q(\nu_\epsilon)} + 4K_1 L \epsilon$.*

Proof. Let $\hat{v}_{\epsilon,x} = v_{\epsilon,x} \circ \pi_\epsilon$ and define $\hat{v}_{\epsilon,x,t} = \int_0^t \hat{v}_{\epsilon,x} \circ f_u^\epsilon du$. Let $N_{\epsilon,t} : M_{h_\epsilon} \rightarrow \{0, 1, \dots, 1 + [K_1 t]\}$ be the number of laps by time t ,

$$N_{\epsilon,t}(y, u) = \#\{s \in (0, t] : f_s^\epsilon(y, u) \in M' \times \{0\}\}.$$

Then

$$\hat{v}_{\epsilon,x,t}(y, u) = w_{\epsilon,x,N_{\epsilon,t}(y,u)}(y) + H_{\epsilon,x} \circ f_t^\epsilon(y, u) - H_{\epsilon,x}(y, u)$$

where $H_{\epsilon,x}(y, u) = \int_0^u \hat{v}_{\epsilon,x}(y, u') du' = \hat{v}_{\epsilon,x,u}(y, 0)$. Note that $|H_{\epsilon,x}|_\infty \leq 2K_1 L$. Hence

$$\begin{aligned} \sup_{s \leq t} |v_{\epsilon,x,s}| \circ \pi_\epsilon(y, u) &= \sup_{s \leq t} |\hat{v}_{\epsilon,x,s}(y, u)| \leq \sup_{s \leq t} |w_{\epsilon,x,N_s(y,u)}(y)| + 4K_1 L \\ &\leq \max_{j \leq 1+K_1 t} |w_{\epsilon,x,j}(y)| + 4K_1 L. \end{aligned}$$

It follows that

$$\epsilon \sup_{s \leq 1/\epsilon} |v_{\epsilon,x,s}| \circ \pi_\epsilon(y, u) \leq \Delta_{1,\epsilon}(y) + 4K_1 L \epsilon,$$

and so $\delta_{1,\epsilon} \circ \pi_\epsilon(y, u) \leq \Delta_{1,\epsilon}(y) + 4K_1 L \epsilon$. The result follows. ■

As a consequence of Corollary 6.4 and Lemma 6.5, our results for maps go through immediately for semiflows. For example, suppose that the maps $T_\epsilon(y) = \phi_{h_\epsilon}^\epsilon(y)$ are a family of quadratic maps as in Example 5.2. Then for any $q > 0$, we obtain averaging in $L^q(\nu_\epsilon)$ with rate $O(\epsilon^{(1/2)^-})$. If moreover ν_0

is strongly statistically stable, then we obtain averaging in $L^1(\nu_0)$ with rate $O(\epsilon^{(1/2)^-})$, and averaging in $L^1(\mu)$ for any absolutely continuous probability measure μ on M .

7. Counterexample for almost sure convergence. It is known [7] that almost sure convergence fails for fully coupled fast-slow systems. Here we give an example to show that almost sure convergence fails also in the simpler context of families of skew products as considered in this paper.

We consider the family of maps $T_\epsilon : [0, 1] \rightarrow [0, 1]$ given by $T_\epsilon y = 2y + \epsilon \pmod 1$ with invariant measure ν_ϵ taken to be Lebesgue for all $\epsilon \geq 0$. Let $a(x, y, \epsilon) = \cos 2\pi y$. Since a has mean zero, the averaged ODE is given by $\dot{X} = 0$. We take $x_0 = 0$ so that $X(t) \equiv 0$. Nevertheless, we prove:

PROPOSITION 7.1. *For every $y_0 \in [0, 1]$, $\limsup_{\epsilon \rightarrow 0} \hat{x}^{(\epsilon)}(1) = 1$.*

Proof. The idea of the proof is to show that every $y_0 \in [0, 1]$ is eventually fixed by T_ϵ for infinitely many arbitrarily small values of ϵ . Namely, $y_n^{(\epsilon)} = T_\epsilon^n y_0 = -\epsilon \pmod 1$ for all large enough n . Moreover, $y_n^{(\epsilon)}$ gets fixed sufficiently early to interfere with the averaging.

Let $y_0 \in [0, 1]$ and $\delta > 0$. Let $N = \lceil \delta^{-1/2} \rceil$ and choose an integer $1 \leq k \leq 2^N$ such that

$$y_0 \in [-\delta + (k-1)2^{-N}, -\delta + k2^{-N}].$$

Choose ϵ such that

$$y_0 = -\epsilon + (k-1)2^{-N}.$$

Then

$$\delta - 2^{-N} \leq \epsilon \leq \delta.$$

If δ is small enough, then $\delta - 2^{-N} = \delta - 2^{-\lceil \delta^{-1/2} \rceil} > 0$, and so $0 < \epsilon \leq \delta$.

Now

$$y_n^{(\epsilon)} = 2^n y_0 + (2^n - 1)\epsilon \pmod 1 = -\epsilon + (k-1)2^{n-N} \pmod 1$$

for all $n \geq 0$. In particular, for $n \geq N$ we have $y_n^{(\epsilon)} = -\epsilon \pmod 1$, and $\cos 2\pi y_n^{(\epsilon)} \geq 1 - \pi\epsilon$. Note that $N \leq \epsilon^{-1/2}$. Hence $\hat{x}^{(\epsilon)}(1) = \epsilon \sum_{n=0}^{\lceil \epsilon^{-1} \rceil - 1} \cos 2\pi y_n^{(\epsilon)} = 1 + O(\epsilon^{1/2}) + O(\epsilon)$. Since $\epsilon \in (0, \delta]$ is arbitrarily small, the result follows. ■

REMARK 7.2. A similar argument works for the family of maps $T_\epsilon y = 2y + \epsilon^\beta \pmod 1$ for any choice of $\beta > 0$.

Appendix: Proof of second order averaging. In this appendix, we prove Theorem 2.2. This is a quantitative version of a result due to [31] with a somewhat simplified proof. We work with discrete time rather than continuous time.

First, we consider the case where $T : M \rightarrow M$ is independent of ϵ . Suppose that $a : \mathbb{R}^d \times M \rightarrow \mathbb{R}^d$ and $\bar{a} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are functions. Assume that $\|a\|_{\text{Lip}} \leq L$ and $\|Da\|_{\text{Lip}} \leq L$ where $D = \frac{d}{dx}$ and $L \geq 1$.

For $\epsilon > 0$, consider the discrete fast-slow system

$$x_{n+1} = x_n + \epsilon a(x_n, y_n), \quad y_{n+1} = Ty_n$$

with $x_0 \in \mathbb{R}^d$ and $y_0 \in M$ given.

Define $\hat{x}_\epsilon : [0, 1] \rightarrow \mathbb{R}^d$, $\hat{x}_\epsilon(t) = x_{\lfloor t/\epsilon \rfloor}$, and let $X : [0, 1] \rightarrow \mathbb{R}^d$ be the solution of the ODE $\dot{X} = \bar{a}(X)$ with initial condition $X(0) = x_0$. As in Section 2, we define $\delta_\epsilon = \delta_{1,\epsilon} + \delta_{2,\epsilon}$ where

$$\delta_{1,\epsilon}(y_0) = \sup_x \sup_{1 \leq n \leq 1/\epsilon} \epsilon \left| \sum_{j=0}^{n-1} (a(x, y_j) - \bar{a}(x)) \right|,$$

$$\delta_{2,\epsilon}(y_0) = \sup_x \sup_{1 \leq n \leq 1/\epsilon} \epsilon \left| \sum_{j=0}^{n-1} (Da(x, y_j) - D\bar{a}(x)) \right|.$$

THEOREM A.1. *Let $\epsilon > 0$ and $x_0 \in \mathbb{R}^d$. For all $y_0 \in M$ with $\delta_\epsilon(y_0) \leq 1/2$ and $t \in [0, 1]$,*

$$|\hat{x}_\epsilon(t) - X(t)| \leq 5e^{2L}(\delta_\epsilon(y_0) + \epsilon).$$

First we recall a discrete version of Gronwall's lemma.

PROPOSITION A.2. *Suppose that $b_n \geq 0$ and that there exist constants $C, D \geq 0$ such that*

$$b_n \leq C + D \sum_{m=0}^{n-1} b_m \quad \text{for all } n \geq 0.$$

Then $b_n \leq C(D+1)^n$.

Proof. This follows by induction. ■

Define a function $u : \mathbb{R}^d \times \{0, 1, \dots, [1/\epsilon]\}$ by setting $u(x, 0) \equiv 0$ and

$$u(x, n) = \frac{\epsilon}{\delta_\epsilon} \sum_{j=0}^{n-1} (a(x, y_j) - \bar{a}(x)), \quad n \geq 1.$$

PROPOSITION A.3. *For any $n \leq 1/\epsilon$, we have $|u(\cdot, n)|_\infty \leq 1$ and $\text{Lip } u(\cdot, n) \leq 1$.*

Proof. For all x ,

$$|Du(x, n)| = \frac{\epsilon}{\delta_\epsilon} \left| \sum_{j=0}^{n-1} (Da(x, y_j) - D\bar{a}(x)) \right| \leq \frac{\delta_{2,\epsilon}}{\delta_\epsilon} \leq 1.$$

Hence the second estimate follows from the mean value theorem, and the first estimate is easier. ■

Define a new sequence w_n by setting $w_0 = x_0$ and

$$w_n = x_n - \delta_\epsilon u(w_{n-1}, n), \quad n \geq 1.$$

LEMMA A.4. For all $0 \leq n \leq 1/\epsilon$ and all $y \in M$ with $\delta_\epsilon(y) \leq 1/2$,

$$\left| w_n - w_0 - \epsilon \sum_{k=0}^{n-1} \bar{a}(w_k) \right| \leq 4L\delta_\epsilon.$$

Proof. By definition, for all $0 \leq k \leq 1/\epsilon$,

$$\delta_\epsilon[u(w_k, k+1) - u(w_k, k)] = \epsilon[a(w_k, y_k) - \bar{a}(w_k)].$$

Thus

$$\begin{aligned} \delta_\epsilon u(w_{n-1}, n) &= \delta_\epsilon \sum_{k=0}^{n-1} \{u(w_k, k+1) - u(w_{k-1}, k)\} \\ &= \delta_\epsilon \left(\sum_{k=0}^{n-1} \{u(w_k, k+1) - u(w_k, k)\} + \sum_{k=0}^{n-1} \{u(w_k, k) - u(w_{k-1}, k)\} \right) \\ &= \epsilon \sum_{k=0}^{n-1} \{a(w_k, y_k) - \bar{a}(w_k)\} + \mathbf{I}_n, \end{aligned}$$

where

$$\mathbf{I}_n = \delta_\epsilon \sum_{k=1}^{n-1} \{u(w_k, k) - u(w_{k-1}, k)\}.$$

This together with the definition of x_n yields

$$\begin{aligned} w_n &= x_0 + \epsilon \sum_{k=0}^{n-1} a(x_k, y_k) - \delta_\epsilon u(w_{n-1}, y, n) \\ &= w_0 + \epsilon \sum_{k=0}^{n-1} a(x_k, y_k) - \epsilon \sum_{k=0}^{n-1} \{a(w_k, y_k) - \bar{a}(w_k)\} - \mathbf{I}_n \\ &= w_0 + \epsilon \sum_{k=0}^{n-1} \bar{a}(w_k) - \mathbf{I}_n + \mathbf{II}_n, \end{aligned}$$

where

$$\mathbf{II}_n = \epsilon \sum_{k=0}^{n-1} \{a(x_k, y_k) - a(w_k, y_k)\}.$$

We claim that for all $0 \leq n \leq 1/\epsilon$,

$$|w_{n+1} - w_n - \epsilon \bar{a}(w_n)| \leq 4L\epsilon\delta_\epsilon.$$

The result follows by summing over n .

It remains to prove the claim. It is easy to check $w_1 - w_0 - \epsilon \bar{a}(w_0) = 0$. Inductively, suppose that $|w_n - w_{n-1} - \epsilon \bar{a}(w_{n-1})| \leq 4L\epsilon\delta_\epsilon$. Notice that

$$|\mathbf{II}_{n+1} - \mathbf{II}_n| \leq \epsilon(\text{Lip } a)|x_n - w_n| \leq (\text{Lip } a)\epsilon\delta_\epsilon|u(w_{n-1}, n)| \leq L\epsilon\delta_\epsilon.$$

Also,

$$|I_{n+1} - I_n| \leq \delta_\epsilon (\text{Lip } u) |w_n - w_{n-1}| \leq \delta_\epsilon (\epsilon |\bar{a}(w_{n-1})| + 4L\epsilon\delta_\epsilon) \leq 3L\epsilon\delta_\epsilon,$$

where the second inequality follows by the induction hypothesis and the third inequality uses $\delta_\epsilon \leq 1/2$. Therefore

$$|w_{n+1} - w_n - \epsilon \bar{a}(w_n)| \leq |I_{n+1} - I_n| + |\text{II}_{n+1} - \text{II}_n| \leq 4L\epsilon\delta_\epsilon,$$

proving the claim. ■

Define the sequence

$$z_{n+1} = z_n + \epsilon \bar{a}(z_n), \quad z_0 = x_0.$$

COROLLARY A.5. $|x_n - z_n| \leq 5\delta_\epsilon e^{2L}$ for all $1 \leq n \leq 1/\epsilon$ and all $y \in M$ with $\delta_\epsilon(y) \leq 1/2$.

Proof. Write

$$\begin{aligned} w_n - z_n &= w_n - x_0 - \epsilon \sum_{k=0}^{n-1} \bar{a}(z_k) \\ &= w_n - w_0 - \epsilon \sum_{k=0}^{n-1} \bar{a}(w_k) + \epsilon \sum_{k=0}^{n-1} \{\bar{a}(w_k) - \bar{a}(z_k)\}. \end{aligned}$$

By Lemma A.4 ,

$$|w_n - z_n| \leq 4L\delta_\epsilon + \epsilon \sum_{k=0}^{n-1} |\bar{a}(w_k) - \bar{a}(z_k)| \leq 4L\delta_\epsilon + L\epsilon \sum_{k=0}^{n-1} |w_k - z_k|.$$

By Proposition A.2, $|w_n - z_n| \leq 4L\delta_\epsilon e^L \leq 4\delta_\epsilon e^{2L}$. Moreover, $|x_n - w_n| \leq \delta_\epsilon |u|_\infty \leq \delta_\epsilon$ and the result follows. ■

LEMMA A.6. $|X(n\epsilon) - z_n| \leq L^2 e^L \epsilon$.

Proof. Write

$$X(n\epsilon) = x_0 + \sum_{m=0}^{n-1} \int_{m\epsilon}^{(m+1)\epsilon} [\bar{a}(X(s)) - \bar{a}(X(m\epsilon))] ds + \epsilon \sum_{m=0}^{n-1} \bar{a}(X(m\epsilon)).$$

Since

$$|\bar{a}(X(t_1)) - \bar{a}(X(t_2))| \leq L|X(t_1) - X(t_2)| \leq L^2|t_1 - t_2| \quad \text{for all } t_1, t_2,$$

we obtain

$$\left| X(n\epsilon) - x_0 - \epsilon \sum_{m=0}^{n-1} \bar{a}(X(m\epsilon)) \right| \leq L^2 n\epsilon^2 \leq L^2 \epsilon.$$

Hence

$$\begin{aligned} |X(n\epsilon) - z_n| &\leq \epsilon \sum_{m=0}^{n-1} |\bar{a}(X(m\epsilon)) - \bar{a}(z_m)| + L^2\epsilon \\ &\leq L\epsilon \sum_{m=0}^{n-1} |X(m\epsilon) - z_m| + L^2\epsilon. \end{aligned}$$

The result follows from Proposition A.2. ■

Proof of Theorem A.1. By Corollary A.5 and Lemma A.6,

$$\begin{aligned} |x_{[t/\epsilon]} - X(t)| &\leq |X(t) - X([t/\epsilon]\epsilon)| + |X([t/\epsilon]\epsilon) - z_{[t/\epsilon]}| + |z_{[t/\epsilon]} - x_{[t/\epsilon]}| \\ &\leq \epsilon L + L^2 e^L \epsilon + 5e^{2L} \delta_\epsilon \leq 5e^{2L} (\delta_\epsilon + \epsilon). \quad \blacksquare \end{aligned}$$

Proof of Theorem 2.2. Replacing T , $a(x, y)$ and $\bar{a}(x)$ in Theorem A.1 by T_ϵ , $a(x, y, \epsilon)$ and $\bar{a}(x, \epsilon)$, we obtain

$$|\hat{x}_\epsilon(t) - X_\epsilon(t)| \leq 5e^{2L} (\delta_\epsilon + \epsilon)$$

where

$$\dot{X}_\epsilon = \bar{a}(X, \epsilon), \quad X_\epsilon(0) = x_0.$$

Let $A_\epsilon = \sup_{x \in E} |\bar{a}(x, \epsilon) - \bar{a}(x, 0)|$. Then

$$\begin{aligned} |X_\epsilon(t) - X(t)| &\leq \int_0^t |\bar{a}(X_\epsilon(s), \epsilon) - \bar{a}(X(s), 0)| ds \\ &\leq tA_\epsilon + \int_0^t |\bar{a}(X_\epsilon(s), 0) - \bar{a}(X(s), 0)| ds \\ &\leq A_\epsilon + L \int_0^t |X_\epsilon(s) - X(s)| ds. \end{aligned}$$

By Gronwall's lemma, $|X_\epsilon(t) - X(t)| \leq e^L A_\epsilon$ for all $t \leq 1$.

Next, $A_\epsilon \leq L\epsilon + \sup_{x \in E} \left| \int_M a(x, y, 0) (d\nu_\epsilon - d\nu_0)(y) \right|$. Combining these estimates leads to

$$\begin{aligned} |\hat{x}_\epsilon(t) - X(t)| &\leq 5e^{2L} \delta_\epsilon + 6e^{2L} \epsilon \\ &\quad + e^L \sup_{x \in E} \left| \int_M a(x, y, 0) (d\nu_\epsilon - d\nu_0)(y) \right|, \end{aligned}$$

yielding the result. ■

Acknowledgments. This research was supported in part by a European Advanced Grant *StochExtHomog* (ERC AdG 320977). We are grateful to Vitor Araújo, Jorge Freitas, Vilton Pinheiro, Mike Todd and Paulo Varandas for helpful discussions.

References

- [1] J. F. Alves, *SRB measures for non-hyperbolic systems with multidimensional expansion*, Ann. Sci. École Norm. Sup. (4) 33 (2000), 1–32.
- [2] J. F. Alves, *Strong statistical stability of non-uniformly expanding maps*, Nonlinearity 17 (2004), 1193–1215.
- [3] J. F. Alves, C. Bonatti and M. Viana, *SRB measures for partially hyperbolic systems whose central direction is mostly expanding*, Invent. Math. 140 (2000), 351–398.
- [4] J. F. Alves, S. Luzzatto and V. Pinheiro, *Markov structures and decay of correlations for non-uniformly expanding dynamical systems*, Ann. Inst. H. Poincaré Anal. Non Linéaire 22 (2005), 817–839.
- [5] J. F. Alves and M. Viana, *Statistical stability for robust classes of maps with non-uniform expansion*, Ergodic Theory Dynam. Systems 22 (2002), 1–32.
- [6] D. V. Anosov, *Averaging in systems of ordinary differential equations with rapidly oscillating solutions*, Izv. Akad. Nauk SSSR Ser. Mat. 24 (1960), 721–742 (in Russian).
- [7] V. Bakhtin and Y. Kifer, *Nonconvergence examples in averaging*, in: Geometric and Probabilistic Structures in Dynamics, Contemp. Math. 469, Amer. Math. Soc., Providence, RI, 2008, 1–17.
- [8] V. Baladi, *On the susceptibility function of piecewise expanding interval maps*, Comm. Math. Phys. 275 (2007), 839–859.
- [9] V. Baladi, M. Benedicks and D. Schnellmann, *Whitney–Hölder continuity of the SRB measure for transversal families of smooth unimodal maps*, Invent. Math. 201 (2015), 773–844.
- [10] V. Baladi and M. Todd, *Linear response for intermittent maps*, Comm. Math. Phys. 347 (2016), 857–874.
- [11] M. Benedicks and L. Carleson, *On iterations of $1 - ax^2$ on $(-1, 1)$* , Ann. of Math. 122 (1985), 1–25.
- [12] D. L. Burkholder, *Distribution function inequalities for martingales*, Ann. Probab. 1 (1973), 19–42.
- [13] P. Collet and J.-P. Eckmann, *Positive Liapunov exponents and absolute continuity for maps of the interval*, Ergodic Theory Dynam. Systems 3 (1983), 13–46.
- [14] D. Dolgopyat, *Averaging and invariant measures*, Moscow Math. J. 5 (2005), 537–576, 742.
- [15] J. M. Freitas, *Continuity of SRB measure and entropy for Benedicks–Carleson quadratic maps*, Nonlinearity 18 (2005), 831–854.
- [16] J. M. Freitas, *Exponential decay of hyperbolic times for Benedicks–Carleson quadratic maps*, Portugaliae Math. 67 (2010), 525–540.
- [17] J. M. Freitas and M. Todd, *The statistical stability of equilibrium states for interval maps*, Nonlinearity 22 (2009), 259–281.
- [18] S. Gouëzel, *Decay of correlations for nonuniformly expanding systems*, Bull. Soc. Math. France 134 (2006), 1–31.
- [19] M. V. Jakobson, *Absolutely continuous invariant measures for one-parameter families of one-dimensional maps*, Comm. Math. Phys. 81 (1981), 39–88.
- [20] G. Keller, *Stochastic stability in some chaotic dynamical systems*, Monatsh. Math. 94 (1982), 313–333.
- [21] G. Keller and C. Liverani, *Stability of the spectrum for transfer operators*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 28 (1999), 141–152.
- [22] Y. Kifer, *Averaging in difference equations driven by dynamical systems*, Astérisque 287 (2003), 103–123.

- [23] Y. Kifer, *Averaging principle for fully coupled dynamical systems and large deviations*, Ergodic Theory Dynam. Systems 24 (2004), 847–871.
- [24] Y. Kifer, *Another proof of the averaging principle for fully coupled dynamical systems with hyperbolic fast motions*, Discrete Contin. Dynam. Systems 13 (2005), 1187–1201.
- [25] A. Korepanov, *Linear response for intermittent maps with summable and nonsummable decay of correlations*, Nonlinearity 29 (2016), 1735–1754.
- [26] A. Korepanov, Z. Kosloff and I. Melbourne, earlier (much longer) version of this paper, arXiv:1509.06772v2.
- [27] A. Korepanov, Z. Kosloff and I. Melbourne. *Explicit coupling argument for nonuniformly hyperbolic transformations*, Proc. Roy. Soc. Edinburgh Sect. A, to appear.
- [28] C. Liverani, B. Saussol and S. Vaienti, *A probabilistic approach to intermittency*, Ergodic Theory Dynam. Systems 19 (1999), 671–685.
- [29] M. Lyubich, *Almost every real quadratic map is either regular or stochastic*, Ann. of Math. 156 (2002), 1–78.
- [30] I. Melbourne and R. Zweimüller, *Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems*, Ann Inst. H. Poincaré Probab. Statist. 51 (2015), 545–556.
- [31] J. A. Sanders, *On the fundamental theorem of averaging*, SIAM J. Math. Anal. 14 (1983), 1–10.
- [32] J. A. Sanders, F. Verhulst and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, 2nd ed., Appl. Math. Sci. 59, Springer, New York, 2007.
- [33] H. Thunberg, *Unfolding of chaotic unimodal maps and the parameter dependence of natural measures*, Nonlinearity 14 (2001), 323–337.
- [34] M. Viana, *Multidimensional nonhyperbolic attractors*, Inst. Hautes Études Sci. Publ. Math. 85 (1997), 63–96.
- [35] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. 110 (1999), 153–188.

Alexey Korepanov, Ian Melbourne
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK
E-mail: a.korepanov@warwick.ac.uk
i.melbourne@warwick.ac.uk

Zemer Kosloff
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK
Permanent address:
Einstein Institute of Mathematics
The Hebrew University
Edmond J. Safra Campus (Givat Ram)
Jerusalem 91904, Israel
E-mail: zemer.kosloff@mail.huji.ac.il