Averaging and Rates of Averaging for Uniform Families of Deterministic Fast-Slow Skew Product Systems

A. Korepanov Z. Kosloff I. Melbourne

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

21 September 2015. Updated 26 October 2015

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 $\epsilon \to 0$. 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 [30] 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 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 the sense of convergence 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

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

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 concerned with the 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 $O(\epsilon^{\frac{1}{2}-})$.[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 continuous and we are able to present the main ideas without going into the technical issues presented by

[1] $q-$ denotes $q-a$ for all $a>0$.

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 first order and second order averaging theorems. These prerequisites are presented in Appendices A and B 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 $T_\epsilon : M \to M$, $0 \le \epsilon < \epsilon_0$, 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

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

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 $a : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^d$ 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

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

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 $\hat{x}^{(\epsilon)} : [0, 1] \to \mathbb{R}^d$ 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 $g : \mathbb{R}^d \to \mathbb{R}^n$, we define $\|g\|_{\mathrm{Lip}} = \max\{|g|_\infty, \mathrm{Lip}\, g\}$ where $\mathrm{Lip}\, g = \sup_{x\neq x'} |g(x) - g(x')|/|x - x'|$ and $|x - x'| = \max_{i=1,\dots,n} |x_i - x_i'|$.

In this section, and also in Appendices A and B, we consider functions $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^n$ where there is no metric structure assumed on M. In that case, $\|g\|_{\mathrm{Lip}} = \sup_{y\in M} \sup_{\epsilon\in[0,\epsilon_0)} \|g(\cdot, y, \epsilon)\|_{\mathrm{Lip}}$. If $E \subset \mathbb{R}^d$, then $\|g|_E\|_{\mathrm{Lip}}$ 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 $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^n$, then $Dg : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^{n\times d}$ and $\|Dg\|_{\mathrm{Lip}}$ is defined accordingly. Similarly for $\|Dg|_E\|_{\mathrm{Lip}}$ when $E \subset \mathbb{R}^d$.

Below, $L_1, L_2, L_3 \geq 1$ are constants.

For first order averaging we require that a is globally Lipschitz in x:

$$\|a\|_{\mathrm{Lip}} \leq L_1. \tag{2.3}$$

Set $E = \{x \in \mathbb{R}^d : |x - x_0| \leq L_1\}$.

For second order averaging we require in addition that $Da|_E$ is Lipschitz in x:

$$\|Da|_E\|_{\mathrm{Lip}} \leq L_2. \tag{2.4}$$

In both cases, we suppose that

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

In the sequel we let $L = \max\{L_1, L_3\}$ when performing first order averaging, and $L = \max\{L_1, L_2, L_3\}$ when performing second order averaging.

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 $v_{\epsilon,x}(y) = a(x, y, \epsilon) - \bar{a}(x, \epsilon)$. We define the *first order function* $\delta_{1,\epsilon} : M \to \mathbb{R}$,

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

Next, we define the *second order function* $\delta_\epsilon : M \to \mathbb{R}$,

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

Theorem 2.2 *Let* $S_\epsilon = \sup_{x\in E} |\int_M a(x, y, 0)\, (d\nu_\epsilon - d\nu_0)(y)| + \epsilon$.

(a) *Assume conditions* (2.3) *and* (2.5). *Then* $z_\epsilon \le 6e^{2L}(\sqrt{\delta_{1,\epsilon}} + S_\epsilon)$.

(b) *Assume conditions* (2.3)—(2.5). *If* $\delta_\epsilon \le \frac{1}{2}$, *then* $z_\epsilon \le 6e^{2L}(\delta_\epsilon + S_\epsilon)$.

The proof of Theorem 2.2 is postponed to Appendices A and B.

Remark 2.3 For averaging without rates, it suffices instead of condition (2.5) that $\lim_{\epsilon\to 0} |a(x, y, \epsilon) - a(x, y, 0)| = 0$ 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 $z_\epsilon \le 2L$ and $\delta_\epsilon \le 4L$, 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.

Throughout the remainder of this paper, it is assumed that conditions (2.3)—(2.5) are satisfied, and we apply Theorem 2.2(b). By Theorem 2.2(a), the results without rates go through unchanged when condition (2.4) fails, and results with rates hold usually with weaker rates of convergence, but for brevity these rates are not stated explicitly.

According to Theorem 2.2(b), 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 $\epsilon \to 0$ ($\nu_\epsilon \to_w \nu_0$). This means that $\int_M \phi\, d\nu_\epsilon \to \int_M \phi\, d\nu_0$ for all continuous bounded functions $\phi : M \to \mathbb{R}$.

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 $\epsilon \ge 0$. Then ν_0 is *strongly statistically stable* if the densities $\rho_\epsilon = d\nu_\epsilon/dm$ satisfy $\lim_{\epsilon\to 0} R_\epsilon = 0$ where $R_\epsilon = \int_M |\rho_\epsilon - \rho_0|\, dm$. We note that $S_\epsilon \le LR_\epsilon + \epsilon$.

Proposition 2.5 *If* $\nu_\epsilon \to_w \nu_0$, *then* $\lim_{\epsilon\to 0} S_\epsilon = 0$.

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 $\nu_\epsilon \to_w \nu_0$, we have that $A_\epsilon(x) \to 0$ 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 2.6 *Let $q\ge 1$. There is a constant $C>0$ such that*

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

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

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

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

Proof Let $A_\epsilon=\{y\in M:\delta_\epsilon(y)\le\frac12\}$. Then Theorem 2.2(b) applies on A_ϵ and

$$\begin{aligned}\int_M z_\epsilon^q\,d\nu_\epsilon&=\int_{M\setminus A_\epsilon} z_\epsilon^q\,d\nu_\epsilon+\int_{A_\epsilon} z_\epsilon^q\,d\nu_\epsilon\le(2L)^q\nu_\epsilon(\delta_\epsilon>\tfrac12)+(6e^{2L})^q\int_M(\delta_\epsilon+S_\epsilon)^q\,d\nu_\epsilon\\&\le(4L)^q\int_M\delta_\epsilon^q\,d\nu_\epsilon+(6e^{2L})^q\int_M(\delta_\epsilon+S_\epsilon)^q\,d\nu_\epsilon\le(12e^{2L})^q\int_M(\delta_\epsilon+S_\epsilon)^q\,d\nu_\epsilon.\end{aligned}$$

Hence

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

yielding the first estimate.

Next,

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

Hence

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

yielding the second estimate. ∎

Corollary 2.7 *(a) Assume statistical stability and that* $\lim_{\epsilon\to 0}\int_M \delta_\epsilon\, d\nu_\epsilon = 0$*. Then* $\lim_{\epsilon\to 0}\int_M z_\epsilon\, d\nu_\epsilon = 0$*.*

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

Proof Part (a), and part (b) in the special case $\mu = \nu_0$, are immediate from Proposition 2.6. To prove the general case of part (b), suppose for contradiction that $\int_M z_{\epsilon_k}\, d\mu \to b > 0$ along some subsequence $\epsilon_k \to 0$. Since $\int_M z_{\epsilon_k}\, d\nu_0 \to 0$, by passing without loss to a further subsequence, we can suppose also that $z_{\epsilon_k} \to 0$ on a set of full measure with respect to ν_0 and hence with respect to μ. By the bounded convergence theorem, $\int_M z_{\epsilon_k}\, d\mu \to 0$ which is the desired contradiction. ∎

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

Corollary 2.8 *Assume strong statistical stability and that* $\lim_{\epsilon\to 0}\int_M \delta_\epsilon\, d\nu_0 = 0$*. Suppose further that each* T_ϵ *is nonsingular with respect to* m *and that for almost every* $y \in M$*, there exists* $N \ge 1$ *such that* $T_\epsilon^N y \in \operatorname{supp}\nu_0$ *for all* $\epsilon \in [0,\epsilon_0)$*. Then* $\lim_{\epsilon\to 0}\int_M z_\epsilon\, d\mu = 0$ *for every probability measure* μ *on* M *with* $\mu \ll m$*.*

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

$$|\delta_\epsilon \circ T_\epsilon^N - \delta_\epsilon|_\infty \le 8LN\epsilon. \tag{2.6}$$

By the arguments in the proof of Corollary 2.7, it suffices to prove that $\int_M \delta_\epsilon\, dm \to 0$. Suppose this is not the case. By Corollary 2.7(b), $\int_{\operatorname{supp}\nu_0} \delta_\epsilon\, dm \to 0$. Hence there exists a subsequence $\epsilon_k \to 0$ and a subset $A \subset \operatorname{supp}\nu_0$ with $m(\operatorname{supp}\nu_0 \setminus A) = 0$ such that (i) $\delta_{\epsilon_k} \to 0$ pointwise on A and (ii) $\int_M \delta_{\epsilon_k}\, dm \to b > 0$.

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

Remark 2.9 The hypotheses of Corollary 2.8 are particularly straightforward to apply when $\operatorname{supp}\nu_0$ has nonempty interior. This property is automatic for large classes of nonuniformly expanding maps, see [3, Lemma 5.6] (taking $G = \operatorname{supp}\nu_0 \cap H(\sigma)$, it follows that $\bar{G}$ contains a disk).

2.3 Estimating the order function

By Proposition 2.6 and Corollary 2.7, 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 2.10 *Let μ be a probability measure on M. Then for all $p \geq 0$, $\epsilon \in [0, \epsilon_0)$,*

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

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

$$A = \{x \in E : |x - \tilde{x}| \leq \tfrac{1}{8L}\delta_{1,\epsilon}\}, \quad B = \{n \in [0, \epsilon^{-1}] : \epsilon|n - \tilde{n}| \leq \tfrac{1}{8L}\delta_{1,\epsilon}\}.$$

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 $\operatorname{Leb}(A) \geq \Gamma\delta_{1,\epsilon}^d$ and $\#B \geq \epsilon^{-1}\delta_{1,\epsilon}/8L$. Hence

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

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 2.11 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

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

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

3 Examples: Uniformly expanding maps

Let $T_\epsilon : M \to M$ 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 $g : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^n$ we define $\|g\|_{\mathrm{Lip}} = |g|_\infty + \sup_{\epsilon \in [0,\epsilon_0)} \mathrm{Lip}\, g(\cdot, \cdot, \epsilon)$ where $\mathrm{Lip}\, g(\cdot, \cdot, \epsilon) = \sup_{x \neq x'} \sup_{y \neq y'} |g(x, y, \epsilon) - g(x', y', \epsilon)|/(|x - x'| + d_M(y, y'))$. We continue to assume conditions (2.3)—(2.5) with this modified definition of $\|\ \|_{\mathrm{Lip}}$.

Proposition 3.1 *Suppose that there is a sequence of constants $a_n \to 0$ such that $\int_M |P_\epsilon^n v - \int_M v\, d\mu_\epsilon| \le a_n \|v\|_{\mathrm{Lip}}$ for all Lipschitz $v : M \to \mathbb{R}$ and all $n \ge 1$, $\epsilon \ge 0$. Then $\lim_{\epsilon \to 0} \int_M \delta_\epsilon\, d\nu_\epsilon = 0$.*

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

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

Hence for v Lipschitz and mean zero,

$$
\int_M \Big(\sum_{j=0}^{n-1} v \circ T_\epsilon^j\Big)^2 d\nu_\epsilon \le b_n \|v\|_{\mathrm{Lip}}^2,
$$

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

By condition (2.3), $\|v_{\epsilon,x}\|_{\mathrm{Lip}} \le 2L$ for all ϵ, x, so $\sup_{x \in E} \sup_{1 \le n \le 1/\epsilon} \int_M |v_{\epsilon,x,n}|^2\, d\nu_\epsilon \le 4L^2 b_n$. By Lemma 2.10, it follows that $\lim_{\epsilon \to 0} \int_M \delta_{1,\epsilon}^{d+3}\, d\nu_\epsilon = 0$. Similarly, by condition (2.4), $\|V_{\epsilon,x}\|_{\mathrm{Lip}} \le 2L$, so $\sup_{x \in E} \sup_{1 \le n \le 1/\epsilon} \int_M |V_{\epsilon,x,n}|^2\, d\nu_\epsilon \le 4L^2 b_n$ and hence $\lim_{\epsilon \to 0} \int_M \delta_{2,\epsilon}^{d+3}\, d\nu_\epsilon = 0$. ∎

Remark 3.2 The proof uses only that $n^{-1} \sum_{k=1}^n a_n \to 0$.

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

$\lim_{\epsilon\to 0}\int_M \delta_\epsilon^q\,d\nu_\epsilon = 0$ 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 $\operatorname{diam} M = 1$. We say that a C^2 map $T : M \to M$ is *uniformly expanding* if there exists $\lambda > 1$ such that $|(DT)_y v| \ge \lambda |v|$ for all $y \in M$, $v \in \mathbb{R}^k$. There is a unique absolutely continuous invariant probability measure, and the density is C^1 and nonvanishing.

If $T_\epsilon : M \to M$, $\epsilon \in [0,1]$, 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 $\lim_{\epsilon\to 0}\int_M \delta_\epsilon\,d\nu_\epsilon = 0$.

Moreover, it is well-known that ν_0 is uniformly equivalent to m and is strongly statistically stable. Hence by Corollary 2.7 we obtain the averaging result $\lim_{\epsilon\to 0}\int_M z_\epsilon\,d\nu_\epsilon = 0$ and $\lim_{\epsilon\to 0}\int_M z_\epsilon\,d\mu = 0$ for every absolutely continuous probability measure μ.

Suppose further that $T_\epsilon : M \to M$, $\epsilon \in [0,1]$, 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 $k \in (0,\frac{1}{2})$, 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 $k \ge \frac{1}{2}$, then the convergence rate for z_ϵ is $O(\epsilon^{\frac{1}{2}-})$.

Example 3.4 (Piecewise uniformly expanding maps) Let $M = [-1,1]$ with Lebesgue measure m. We consider continuous maps $T : M \to M$ 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 $T' \ge \lambda$ on $[-1,0)$ and $T' \le -\lambda$ 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 $\epsilon \to T_\epsilon$ 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 2.7 and 2.8 we obtain averaging in $L^1(\nu_\epsilon)$ and also in $L^1(\mu)$ for every absolutely continuous probability measure μ.

Now suppose that $\epsilon \to T_\epsilon$ is a C^1 family of such maps on $[-1,1]$ with densities $\rho_\epsilon = d\nu_\epsilon/dm$. Keller [20] showed that $\epsilon \to \rho_\epsilon$ is C^{1-} as a map into L^1 densities. Hence we obtain the convergence rate $O(\epsilon^{\frac{1}{2}-})$ 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 on the family T_ϵ and state our main result on averaging, Theorem 4.3, for such families. In Subsection 4.3 we establish some basic estimates for nonuniformly expanding maps. In Subsection 4.4 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 $T : M \to M$ be a nonsingular transformation for which m is ergodic. Let $Y \subset M$ be a subset of positive measure, and normalise m so that $m(Y) = 1$. 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 $\tau : Y \to \mathbb{Z}^+$, constant on each a with value $\tau(a) \geq 1$, and constants $\lambda > 1$, $\eta \in (0, 1]$, $C_0, C_1 \geq 1$ 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 $x, y \in a$, $0 \leq \ell < \tau(a)$.

(4) $\zeta = \frac{dm|_Y}{dm|_Y \circ F}$ 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 $T : M \to M$ is called *nonuniformly expanding*. We refer to $F = T^\tau : Y \to Y$ 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, condition (3) is vacuously satisfied, and condition (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 $T_\epsilon : M \to M$, $\epsilon \in [0, \epsilon_0)$, 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 $T_\epsilon : M \to M$ is a *uniform family* of nonuniformly expanding maps (of order p) if

(i) The constants $C_0, C_1 \ge 1$, $\lambda > 1$, $\eta \in (0,1]$ can be chosen independent of $\epsilon \in [0,\epsilon_0)$.

(ii) The return time functions $\tau_\epsilon : Y_\epsilon \to \mathbb{Z}^+$ 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 $T_\epsilon : M \to M$ is a uniform family of nonuniformly expanding maps of order p, then there is a constant $C>0$ such that for all $\epsilon \in [0,\epsilon_0)$*

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

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 $q \le p+d-1$. 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 we obtain that $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{\frac{1}{2}-})$ for all $q>0$.

If in addition ν_0 is strongly statistically stable with $R_\epsilon = \int_M |\rho_\epsilon - \rho_0|\, dm = O(\epsilon^{\frac{1}{2}-})$, then by Proposition 2.6 we obtain $|z_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{\frac{1}{2}-})$ for all $q>0$ and $|z_\epsilon|_{L^1(\nu_0)} = O(\epsilon^{\frac{1}{2}-})$.

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 it is almost the case that their definition includes our definition, so it is almost the case that verifying the assumptions of [5] is 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 assumption (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, then 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 by [2], assumption (3) holds with uniform C_0 and hence our main results hold. Certain quantities $\delta_1 > 0$ and $N_0 \ge 1$ 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 $m \le N_0$ (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 the proof of [2, 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 mention also 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 nonuniformly expanding maps

Throughout this subsection, we work with a fixed nonuniformly expanding map $T : M \to M$. Some standard constructions and estimates are described. The main novelty is that we stress the dependence of various constants on the underlying constants C_0, C_1, λ and η. For convenience, we normalise the metric d_M so that $\operatorname{diam} M = 1$.

Symbolic metric

Fix $\theta \in [\lambda^{-\eta}, 1)$ and define the *symbolic metric* $d_\theta(x,y) = \theta^{s(x,y)}$ where the *separation time* $s(x,y)$ is the least integer $n \ge 0$ such that $F^n x$ and $F^n y$ lie in distinct partition element. It is assumed that the partition α separates orbits of F, so $s(x,y)$ is finite for all $x \ne y$ guaranteeing that d_θ is a metric.

Proposition 4.6 *$d_M(x,y)^\eta \le d_\theta(x,y)$ for all $x, y \in Y$.*

Proof Let $n = s(x,y)$. By condition (2),

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

Hence $d_M(x,y)^\eta \le \theta^n = d_\theta(x,y)$. ∎

Invariant measures and transfer operators

For a positive observable $\psi : Y \to (0,\infty)$, define

$$|\psi|_{\theta\ell} = \sup_{x,y\in Y,\, x\neq y} \frac{|\log\psi(x)-\log\psi(y)|}{d_\theta(x,y)}.$$

We note that

$$e^{-|\psi|_{\theta\ell}} \int_Y \psi\, dm \le \psi \le e^{|\psi|_{\theta\ell}} \int_Y \psi\, dm. \tag{4.1}$$

Also, for $\psi_k : Y \to (0,\infty)$, $k \ge 1$,

$$\Big|\sum_{k=1}^\infty \psi_k\Big|_{\theta\ell} \le \sup_{k\ge1} |\psi_k|_{\theta\ell}. \tag{4.2}$$

Let $\tilde P : L^1(Y) \to L^1(Y)$ denote the transfer operator corresponding to F and m, so $\int_Y \phi\circ F\,\psi\,dm = \int_Y \phi\,\tilde P\psi\,dm$ for all $\phi\in L^\infty$ and $\psi \in L^1$.

Proposition 4.7 *Let $\psi : Y \to (0,\infty)$. Then*

(a) $|\tilde P\psi|_{\theta\ell} \le C_1 + \theta|\psi|_{\theta\ell}$, and

(b) $e^{-(C_1+\theta|\psi|_{\theta\ell})}\int\psi\,dm \le \tilde P\psi \le e^{C_1+\theta|\psi|_{\theta\ell}}\int\psi\,dm$.

Proof Let $a\in\alpha$ and write $\psi_a = 1_a\psi$. For $y\in Y$, we have $(\tilde P\psi_a)(y) = \zeta(y_a)\psi(y_a)$ where y_a is the unique preimage of y under F lying in a.

Let $x,y\in Y$ with preimages $x_a, y_a \in a$. By condition (4) and Proposition 4.6, $|\log\zeta(x_a)-\log\zeta(y_a)| \le C_1 d_\theta(x,y)$. Hence

$$\begin{aligned}|\log(\tilde P\psi_a)(x) - \log(\tilde P\psi_a)(y)| &\le |\log\zeta(x_a)-\log\zeta(y_a)| + |\log\psi(x_a)-\log\psi(y_a)| \\ &\le C_1 d_\theta(x,y) + |\psi|_{\theta\ell} d_\theta(x_a,y_a) \le (C_1+\theta|\psi|_{\theta\ell})d_\theta(x,y),\end{aligned}$$

and so $|\tilde P\psi_a|_{\theta\ell} \le C_1+\theta|\psi|_{\theta\ell}$. Part (a) follows from (4.2).

Since $\int_Y \tilde P\psi\,dm = \int_Y\psi\,dm$, part (b) follows from (4.1) and part (a). ∎

Proposition 4.8 *There is a unique ergodic F-invariant probability measure μ on Y equivalent to $m|_Y$. Moreover, the density $h = d\mu/dm|_Y$ satisfies*

$$e^{-C_1(1-\theta)^{-1}} \le h \le e^{C_1(1-\theta)^{-1}} \quad \textit{and} \quad |h|_{\theta\ell} \le C_1(1-\theta)^{-1}.$$

Proof There is at most one ergodic invariant probability measure equivalent to $m|_Y$. To prove existence of such a measure with the desired properties, we define the sequence of positive functions $\hat h_0 = 1$, $\hat h_{n+1} = \tilde P\hat h_n$. Then $\int_Y \hat h_n\,dm = 1$ for all $n\ge 0$.

Inductively, it follows from Proposition 4.7(a) that $|\hat{h}_n|_{\theta\ell} \le C_1(1+\theta+\cdots+\theta^{n-1}) \le C_1(1-\theta)^{-1}$ for all $n \ge 1$. Also, by Proposition 4.7(b),

$$\hat{h}_n = \tilde{P}\hat{h}_{n-1} \le e^{C_1+\theta|\hat{h}_{n-1}|_{\theta\ell}} \le e^{C_1+C_1\theta(1-\theta)^{-1}} = e^{C_1(1-\theta)^{-1}},$$

and similarly $\hat{h}_n \ge e^{-C_1(1-\theta)^{-1}}$ for all $n \ge 1$.

Define $h_n = n^{-1}\sum_{j=0}^{n-1}\hat{h}_j = n^{-1}\sum_{j=0}^{n-1}\tilde{P}^j 1$. It is immediate that $\int_Y h_n\,dm = 1$ and $e^{-C_1(1-\theta)^{-1}} \le h_n \le e^{C_1(1-\theta)^{-1}}$. By (4.2), $|h_n|_{\theta\ell} = |\sum_{j=0}^{n-1}\hat{h}_j|_{\theta\ell} \le C_1(1-\theta)^{-1}$. In particular, the sequence h_n is bounded and equicontinuous. By Arzela-Ascoli, there exists a subsequential limit h. The limit inherits the properties

$$\int_Y h\,dm = 1, \quad e^{-C_1(1-\theta)^{-1}} \le h \le e^{C_1(1-\theta)^{-1}}, \quad |h|_{\theta\ell} \le C_1(1-\theta)^{-1}.$$

Moreover, $\tilde{P}h_n = h_n + n^{-1}(\tilde{P}^n 1 - 1)$ and so $\tilde{P}h = h$. Hence $d\mu = h\,dm$ defines an F-invariant probability measure. ∎

Define $g : Y \to \mathbb{R}$ by setting $g|_a = d\mu|_a/d(\mu\circ F|_a)$ for $a \in \alpha$. Let $\alpha_n = \bigvee_{j=0}^{n-1} F^{-j}\alpha$ denote the set of n-cylinders in Y, and write $g_n = (g\circ F^{n-1})\cdots(g\circ F)\cdot g$. Thus for $a \in \alpha_n$, $g_n|_a = d\mu|_a/d(\mu\circ F^n|_a)$ is the inverse of the Jacobian of $F^n|a$.

Let $P : L^1(Y) \to L^1(Y)$ 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 $\psi \in L^1$. Then $(P^n v)(y) = \sum_{a\in\alpha_n} g(y_a)v(y_a)$ where y_a is the unique preimage of y under F^n lying in a.

Proposition 4.9 *For all $x, y \in a$, $a \in \alpha_n$, $n \ge 1$,*

$$g_n(y) \le C_2\mu(a) \quad \textit{and} \quad |g_n(x) - g_n(y)| \le C_2\mu(a)d_\theta(F^n x, F^n y), \tag{4.3}$$

where $C_2 = 2C_1(1-\theta)^{-2}e^{4C_1(1-\theta)^{-2}}$.

Proof First, suppose that $n = 1$ and let $x, y \in a$, $a \in \alpha$. We have $g = \zeta h/h\circ F$, so using Proposition 4.8,

$$\begin{aligned}
&|\log g(x) - \log g(y)| \\
&\quad \le |\log\zeta(x) - \log\zeta(y)| + |\log h(x) - \log h(y)| + |\log h(Fx) - \log h(Fy)| \\
&\quad \le C_1 d_\theta(Fx, Fy) + |h|_{\theta\ell}d_\theta(x,y) + |h|_{\theta\ell}d_\theta(Fx,Fy) \\
&\quad \le C_1(1+\theta(1-\theta)^{-1} + (1-\theta)^{-1})d_\theta(Fx,Fy) = 2C_1(1-\theta)^{-1}d_\theta(Fx,Fy).
\end{aligned}$$

Next, let $x, y \in a$, $a \in \alpha_n$ for general $n \ge 1$. Then

$$\begin{aligned}
|\log g_n(x) - \log g_n(y)| &\le \sum_{j=0}^{n-1} |\log g(F^j x) - \log g(F^j y)| \\
&\le 2C_1(1-\theta)^{-1}\sum_{j=1}^{n} d_\theta(F^j x, F^j y) \le 2C_1(1-\theta)^{-1}\sum_{j=1}^{n}\theta^{n-j}d_\theta(F^n x, F^n y) \\
&\le 2C_1(1-\theta)^{-2}d_\theta(F^n x, F^n y).
\end{aligned} \tag{4.4}$$

In particular $g_n(x)/g_n(y) \le e^{2C_1(1-\theta)^{-2}}$. Hence

$$\mu(a) = \int_Y 1_a \, d\mu = \int_Y P^n 1_a \, d\mu \ge \inf P^n 1_a = \inf_{y \in a} g_n(y) \ge e^{-2C_1(1-\theta)^{-2}} \sup_{y\in a} g_n(y),$$

and so

$$g_n(y) \le e^{2C_1(1-\theta)^{-2}} \mu(a). \tag{4.5}$$

Next, we note the inequality $t - 1 \le t \log t$ which is valid for all $t \ge 1$. Suppose that $g_n(y) \le g_n(x)$. Setting $t = g_n(x)/g_n(y) \ge 1$ and using (4.4),

$$\frac{g_n(x)}{g_n(y)} - 1 \le \frac{g_n(x)}{g_n(y)} \log \frac{g_n(x)}{g_n(y)} \le e^{2C_1(1-\theta)^{-2}} 2C_1(1-\theta)^{-2} d_\theta(F^n x, F^n y).$$

Hence

$$g_n(x) - g_n(y) \le g_n(y) e^{2C_1(1-\theta)^{-2}} 2C_1(1-\theta)^{-2} d_\theta(F^n x, F^n y),$$

and the result follows from (4.5), ∎

There is a standard procedure to pass from μ to a T-invariant ergodic absolutely continuous probability measure ν on M, which we now briefly recall, since the construction is required in the proof of Lemma 4.18. Define the Young tower [33]

$$\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \le \ell < \tau(y)\}, \tag{4.6}$$

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

Notation 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.

Induced observables

Given $\phi : Y \to \mathbb{R}^d$, we define $\|\phi\|_\theta = |\phi|_\infty + |\phi|_\theta$ where $|\phi|_\theta = \sup_{x \ne y} |\phi(x) - \phi(y)|/d_\theta(x, y)$. Then ϕ is *d_θ-Lipschitz* if $\|\phi\|_\theta < \infty$. We say that $\phi : Y \to \mathbb{R}^d$ is *locally d_θ-Lipschitz* if $\sup_{y \in a} |\phi(y)| < \infty$ and $\sup_{x,y \in a,\, x \ne y} |\phi(x) - \phi(y)|/d_\theta(x, y) < \infty$ for each $a \in \alpha$.

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

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

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

Proposition 4.10 *If $v : M \to \mathbb{R}^d$ is d_M-Lipschitz, then $V : Y \to \mathbb{R}^d$ is locally d_θ-Lipschitz and $PV : Y \to \mathbb{R}^d$ is d_θ-Lipschitz.*

Moreover,

$$|V(y)| \le \tau(a)|v|_\infty, \quad |V(x) - V(y)| \le C_0\theta^{-1}\tau(a)\mathrm{Lip}\, v\, d_\theta(x,y), \ \textit{for all } x, y \in a, a \in \alpha,$$

and

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

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

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

completing the estimates for V.

Next, $(PV)(y) = \sum_{a\in\alpha} g(y_a)V(y_a)$, so by (4.3),

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

Also,

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

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

Explicit coupling argument

Note that

$$e^{-|\psi|_{\theta\ell}} \int_Y \psi\, d\mu \le \psi \le e^{|\psi|_{\theta\ell}} \int_Y \psi\, d\mu. \tag{4.7}$$

Proposition 4.11 *Let $\psi : Y \to (0,\infty)$. Then*

(a) $|P\psi|_{\theta\ell} \le C_2 + \theta|\psi|_{\theta\ell}$, *and*

(b) $e^{-(C_2+\theta|\psi|_{\theta\ell})} \int \psi\, d\mu \le P\psi \le e^{C_2+\theta|\psi|_{\theta\ell}} \int \psi\, d\mu$.

Proof By the beginning of the proof of Proposition 4.9,

$$|\log g(x) - \log g(y)| \le 2C_1(1-\theta)^{-1}d_\theta(Fx,Fy) \le C_2 d_\theta(Fx,Fy)$$

for all x, y lying in a common partition element. The proof now proceeds exactly as in Proposition 4.7 with $\tilde{P}$, m, ζ and C_1 replaced by P, μ, g and C_2. ∎

By (4.7), $\psi - \xi \int_Y \psi \, d\mu$ is positive for all $\xi \in [0, e^{-|\psi|_{\theta l}})$.

Proposition 4.12 *Let $\psi : Y \to (0,\infty)$. For each $\xi \in [0, e^{-|\psi|_{\theta l}})$*

$$\Big|\psi - \xi \int_Y \psi \, d\mu\Big|_{\theta l} \le \frac{|\psi|_{\theta l}}{1 - \xi e^{|\psi|_{\theta l}}}.$$

Proof Let $\kappa(y) = \log \psi(y)$. Note that

$$\frac{d}{d\kappa} \log\Big(e^\kappa - \xi \int_Y \psi \, d\mu\Big) = \frac{e^\kappa}{e^\kappa - \xi \int_Y \psi \, d\mu} = \frac{1}{1 - \xi e^{-\kappa} \int_Y \psi \, d\mu}.$$

By (4.7),

$$\frac{1}{1 - \xi e^{-\kappa(y)} \int_Y \psi \, d\mu} = \frac{1}{1 - \xi \psi(y)^{-1} \int_Y \psi \, d\mu} \le \frac{1}{1 - \xi e^{|\psi|_{\theta l}}},$$

for all $y \in Y$. Hence, by the mean value theorem, for $x, y \in Y$,

$$\Big|\log\Big(e^{\kappa(x)} - \xi \int_Y \psi \, d\mu\Big) - \log\Big(e^{\kappa(y)} - \xi \int_Y \psi \, d\mu\Big)\Big| \le \frac{|\kappa(x) - \kappa(y)|}{1 - \xi e^{|\psi|_{\theta l}}} \le \frac{|\psi|_{\theta l}\, d_\theta(x,y)}{1 - \xi e^{|\psi|_{\theta l}}}.$$

This completes the proof. ∎

Choose $C_3 \ge 1$ large enough so that $C_2 + \theta C_3 < C_3$. Define $\xi = e^{-C_3}\Big(1 - \frac{C_2 + \theta C_3}{C_3}\Big)$.

Lemma 4.13 *Let $\phi : Y \to \mathbb{R}$ be an observable with $\int \phi \, d\mu = 0$, and $|\phi|_\theta \le C_3$. Then*

$$|P^n \phi|_\infty \le 2e^{C_3}(1-\xi)^n(1 + |\phi|_1) \quad \textit{and} \quad \|P^n \phi\|_\theta \le 2(C_3 + 1)e^{C_3}(1-\xi)^n(1 + |\phi|_1),$$

for all $n \ge 1$.

Proof Write $\phi = \psi_0^+ - \psi_0^-$, where $\psi_0^- = 1 - \min\{0, \phi\}$ and $\psi_0^+ = 1 + \max\{0, \phi\}$. Then $\psi_0^\pm : Y \to [1, \infty)$ and $\int_Y \psi_0^+ \, d\mu = \int_Y \psi_0^- \, d\mu \le 1 + |\phi|_1$.

Define

$$\psi_{n+1}^\pm = P\psi_n^\pm - \xi \int_Y \psi_n^\pm \, d\mu, \; n \ge 0.$$

Then $\int_Y \psi_{n+1}^\pm \, d\mu = (1 - \xi) \int_Y \psi_n^\pm \, d\mu$, so inductively we have that $\int_Y \psi_n^+ \, d\mu = \int_Y \psi_n^- \, d\mu = (1-\xi)^n \int_Y \psi_0^\pm \, d\mu$. Moreover, $P^n \phi = P^n \psi_0^+ - P^n \psi_0^- = \psi_n^+ - \psi_n^-$.

We claim that the functions $\psi_n^\pm$ are nonnegative and $|\psi_n^\pm|_{\theta l} \le C_3$ for all n. Then

$$\int \psi_n^\pm \, d\mu = (1-\xi)^n \int_Y \psi_0^\pm \, d\mu \le (1-\xi)^n (1+|\phi|_1).$$

Thus by equation (4.7), $\psi_n^\pm \le e^{C_3}(1-\xi)^n(1+|\phi|_1)$ and so

$$|P^n\phi|_\infty \le |\psi_n^+|_\infty + |\psi_n^-|_\infty \le 2e^{C_3}(1-\xi)^n(1+|\phi|_1).$$

Next, we use the inequality $|e^a - e^b| \le \max(e^a, e^b)\,|a-b|$ which holds for all $a, b \in \mathbb{R}$. By definition of $|\cdot|_{\theta l}$, for $x, y \in Y$,

$$\begin{aligned} \big|\psi_n^\pm(x) - \psi_n^\pm(y)\big| &\le \max(\psi_n^\pm(x), \psi_n^\pm(y))\,\big|\log\psi_n^\pm(x) - \log\psi_n^\pm(y)\big| \\ &\le e^{C_3}(1-\xi)^n(1+|\phi|_1)\, C_3\, d_\theta(x,y). \end{aligned}$$

Hence $|\psi_n^\pm|_\theta \le C_3\, e^{C_3}(1-\xi)^n(1+|\phi|_1)$ and so

$$|P^n\phi|_\theta \le |\psi_n^+|_\theta + |\psi_n^-|_\theta \le 2C_3 e^{C_3}(1-\xi)^n(1+|\phi|_1),$$

as required.

It remains to verify the claim. Since $\xi < e^{-C_3}$, it suffices by Proposition 4.11(b) and the choice of C_3 that $|\psi_n^\pm|_{\theta l} \le C_3$ for all n.

For $n = 0$, we suppose first that $\phi(x) \ge \phi(y) \ge 0$. Then

$$\begin{aligned} \log\psi_0^+(x) - \log\psi_0^+(y) &= \log(1+\phi(x)) - \log(1+\phi(y) = \log\Big(1 + \frac{\phi(x)-\phi(y)}{1+\phi(y)}\Big) \\ &\le \frac{\phi(x)-\phi(y)}{1+\phi(y)} \le \phi(x) - \phi(y). \end{aligned}$$

Also, if $\phi(x) \ge 0 \ge \phi(y)$,

$$\log\psi_0^+(x) - \log\psi_0^+(y) = \log(1+\phi(x)) \le \phi(x) \le \phi(x) - \phi(y).$$

It follows that $|\psi_0^+|_{\theta\ell} \le |\phi|_\theta \le C_3$. Noting that $\psi_0^- = 1 + \max\{-\phi, 0\}$, we obtain that $|\psi_0^\pm|_{\theta\ell} \le C_3$.

Suppose inductively that $|\psi_n^\pm|_{\theta l} \le C_3$ for some n. By Proposition 4.11(a), $|P\psi_n^\pm|_{\theta l} \le C_2 + \theta C_3 < C_3$, and so $\xi < e^{-C_3} < e^{-|P\psi_n^\pm|_{\theta l}}$. By Proposition 4.12 and the definition of ξ,

$$|\psi_{n+1}^\pm|_{\theta l} = \Big|P\psi_n^\pm - \xi \int_Y P\psi_n^\pm \, d\mu\Big|_{\theta l} \le \frac{C_2 + \theta C_3}{1 - \xi e^{C_3}} = C_3,$$

completing the induction argument. ∎

Corollary 4.14 *Let $\phi : Y \to \mathbb{R}$ be d_θ-Lipschitz with $\int \phi \, d\mu = 0$. Then*

$$|P^n\phi|_\infty \le 4e^{C_3}(1-\xi)^n\|\phi\|_\theta \quad \text{and} \quad \|P^n\phi\|_\theta \le 4(C_3+1)e^{C_3}(1-\xi)^n\|\phi\|_\theta,$$

for all $n \ge 1$.

Proof Let $\mathcal{B}$ denote the Banach space of d_θ-Lipschitz observables of mean zero. If $\phi \in \mathcal{B}$ with $\|\phi\|_\theta = 1$, then certainly $|\phi|_\theta \le C_3$ so by Lemma 4.13 $|P^n\phi|_\infty \le 4e^{C_3}(1-\xi)^n$. Hence viewed as operators $P^n : \mathcal{B} \to L^\infty(Y)$,

$$\|P^n\| = \sup_{\|\phi\|_\theta = 1} |P^n\phi|_\infty \le 4e^{C_3}(1-\xi)^n,$$

yielding the first estimate. The second estimate is proved in the same way. ∎

Remark 4.15 In this paper, we make use of the L^∞ estimate in Corollary 4.14. The $\|\cdot\|_\theta$ estimate will be used elsewhere.

Explicit moment estimate

Proposition 4.16 *Let $p \ge 1$. There exists $m \in L^p(Y, \mathbb{R}^d)$ and $\chi \in L^\infty(Y, \mathbb{R}^d)$ such that $V = m + \chi \circ F - \chi$, and $m \in \ker P$. Moreover,*

$$|m|_p \le 3C_4|\tau|_p\|v\|_{\mathrm{Lip}}, \quad \text{and} \quad |\chi|_\infty \le C_4|\tau|_1\|v\|_{\mathrm{Lip}},$$

where $C_4 = 8C_0C_2e^{C_3}\theta^{-1}\xi^{-1}$.

Proof By Proposition 4.10 and Corollary 4.14 with $\phi = PV$, for $n \ge 1$,

$$|P^nV|_\infty \le 4e^{C_3}(1-\xi)^{n-1}\|PV\|_\theta \le 8C_0C_2\theta^{-1}e^{C_3}(1-\xi)^{n-1}|\tau|_1\|v\|_{\mathrm{Lip}}.$$

It follows that $\chi = \sum_{k=1}^\infty P^kV$ lies in L^∞ and $|\chi|_\infty \le C_4|\tau|_1\|v\|_{\mathrm{Lip}}$.

Write $V = m + \chi \circ F - \chi$; then $m \in L^p$ and $Pm = 0$. Finally, $|m|_p \le |V|_p + 2|\chi|_\infty \le |\tau|_p|v|_\infty + 2|\chi|_\infty \le 3C_4|\tau|_p\|v\|_{\mathrm{Lip}}$. ∎

Corollary 4.17 *Define $V_n = \sum_{j=0}^{n-1} V \circ F^j$. Let $p > 1$. There exists a constant $C_5 \ge 1$ depending only on p and C_4 such that*

$$\big|\max_{1\le j\le n} |V_j|\,\big|_p \le C_5|\tau|_p\|v\|_{\mathrm{Lip}}\; n^{\max\{1/2,1/p\}}.$$

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

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

(see for example the proof of [28, Proposition 4.3]). Hence

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

The result follows from Proposition 4.16 with $C_5 = 5C_4C(p)$. ∎

Lemma 4.18 *Let $p>1$. Let $v_n = \sum_{j=0}^{n-1} v\circ T^j$. Then*

$$\big|\max_{j\le n}|v_j|\big|_{L^{p-1}(\nu)} \le 5C_5|\tau|_p^{p/(p-1)}\|v\|_{\mathrm{Lip}}\, n^{\max\{1/2,1/p\}}.$$

Proof Let $q=p-1$. Define the tower Δ as in (4.6) with tower map $f:\Delta\to\Delta$ where

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

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 $\pi_\Delta:\Delta\to M$ 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 $N_n:\Delta\to\{0,1,\dots,n\}$ 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 $|H(y,\ell)|\le |v|_\infty\tau(y)$ 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

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

where $\hat\tau:\Delta\to\mathbb{Z}^+$ 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 $\bar\tau\ge1$,

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

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

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

where we used Proposition 4.10 with $v=1$ (and hence $V=\tau$) for the final inequality. Hence

$$\big|\max_{1\le j\le n}\hat\tau(F^jy)\big|_{L^q(\mu_\Delta)}\le\big|\max_{1\le j\le n}\hat\tau(F^jy)\big|_{L^p(\mu_\Delta)}\le C_2^{1/p}n^{1/p}|\tau|_p.$$

Combining these estimates, we obtain that

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

Next, using Hölder's inequality,

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

By Corollary 4.17,

$$\Big|\max_{j\le n}|V_{N_j}|\Big|_{L^q(\mu_\Delta)}\le C_5|\tau|_p^{p/(p-1)}\|v\|_{\rm Lip}\;n^{\max\{1/2,1/p\}}.$$

By the triangle inequality, using that $C_2^{1/p}\le C_5$,

$$\big|\,|\max_{j\le n}v_n|\big|_{L^{p-1}(\nu)}=\big|\max_{j\le n}|\hat v_n|\big|_{L^q(\mu_\Delta)}\le 5C_5|\tau|_p^{p/(p-1)}\|v\|_{\rm Lip}\;n^{\max\{1/2,1/p\}},$$

as required. ∎

4.4 Proof of Theorem 4.3

Define $v_{\epsilon,x}$ and $v_{\epsilon,x,n}$ as in Section 2. Note that $\operatorname{Lip}v_{\epsilon,x}\le 2L$ for all ϵ,x and $\int_M v_{\epsilon,x}\,d\nu_\epsilon=0$.

It follows from Lemma 4.18 that

$$\big|\max_{j\le n}|v_{\epsilon,x,j}|\big|_{L^{p-1}(\nu_\epsilon)}\le 10C_5L|\tau_\epsilon|_p^{p/(p-1)}n^{\max\{1/2,1/p\}},$$

for all $\epsilon\ge0$, $x\in\mathbb{R}$, $n\ge1$. By (2.7),

$$\begin{aligned}\int_M\delta_{1,\epsilon}^{p+d-1}\,d\nu_\epsilon&\le\Gamma\epsilon^{p-1}\sup_{x\in E}\int_M\max_{n\le1/\epsilon}|v_{\epsilon,x,n}|^{p-1}\,d\nu_\epsilon\le\Gamma C_5^{p-1}|\tau_\epsilon|_p^p\epsilon^{p-1}\epsilon^{-(p-1)\max\{1/2,1/p\}}\\&=\Gamma C_5^{p-1}|\tau_\epsilon|_p^p\epsilon^{p-1}\epsilon^{\min\{-(p-1)/2,-(p-1)/p\}}=\Gamma C_5^{p-1}|\tau_\epsilon|_p^p\epsilon^{\min\{(p-1)/2,(p-1)^2/p\}}.\end{aligned}$$

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

5 Examples: Nonuniformly expanding maps

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

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

These were studied in [26]. 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 $T_\epsilon : M \to M$, $\epsilon \in [0, \epsilon_0)$, 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 $Y = [\frac{1}{2}, 1]$ and let $\tau_\epsilon : Y \to \mathbb{Z}^+$ be the first return time $\tau_\epsilon(y) = \inf\{n \geq 1 : T_\epsilon^n y \in Y\}$. Define the first return map $F_\epsilon = T_\epsilon^{\tau_\epsilon} : Y \to Y$. 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 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 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 $\zeta_\epsilon = \frac{dm|_Y}{dm|_Y \circ F_\epsilon}$. Note that $\zeta_\epsilon(y) = 1/F_\epsilon'(y)$. For each $Y_\epsilon(n) \in \alpha_\epsilon$, define the bijection $F_{\epsilon,n} = F_\epsilon|_{Y_\epsilon(n)} : Y_\epsilon(n) \to M$. 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 $x, y \in Y_\epsilon(n)$,

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

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

Define $x_1 = \frac{1}{2}$ 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 $x_n \leq K n^{-1/\alpha_+}$ for all $n \geq 1$, $\epsilon \in [0, \epsilon_0)$. Hence

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

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 [11, 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 2.7, 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 $T : [-1,1] \to [-1,1]$ given by $T(x) = 1 - ax^2$, $a \in [-2,2]$, 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

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

In this case, we write $a \in Q_{b,c}$. The set of Collet-Eckmann parameters is given by $P_1 = \bigcup_{b,c>0} Q_{b,c}$ and is a Cantor set of positive Lebesgue measure [19, 10]. 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 $P_0 \subset [-2,2]$ for which T possesses a periodic sink attracting Lebesgue almost every trajectory in $[-1,1]$. By Lyubich [27], $P_0 \cup P_1$ 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 [31, Corollary 1] shows that on any full measure subset of $E \subset [-2,2]$ the map $a \to \nu_a$ 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 $a \to \rho_a = d\nu_a/dm$ from $Q_{b,c} \to L^1$ is continuous. (See also [15, 16] for the same result restricted to the Benedicks-Carleson parameters [10].)

We consider families $\epsilon \to T_\epsilon$ 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

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

Moreover, using Corollary 2.8 we obtain convergence in $L^1(\nu)$ for every absolutely continuous probability measure ν. 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 $\frac{|T_\epsilon^j x - T_\epsilon^j y|}{\operatorname{diam} T_\epsilon^j a} \le \tilde{C}_1 \frac{|F_\epsilon x - F_\epsilon y|}{\operatorname{diam} Y_\epsilon}$ for all partition elements a, all $x, y \in a$ and all $j \le \tau_\epsilon(a)$. In particular, $|T_\epsilon^j x - T_\epsilon^j y| \le (2\tilde{C}_1 / \operatorname{diam} Y_\epsilon)|F_\epsilon x - F_\epsilon y|$, 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 $|\delta_\epsilon|_{L^q(\nu_\epsilon)} = O(\epsilon^{\frac{1}{2}-})$. If $\epsilon \mapsto a_\epsilon$ is C^1, then it follows from Baladi *et al.* [9] that $R\epsilon = \int |\rho_\epsilon - \rho_0| \, dm = O(\epsilon^{\frac{1}{2}-})$. By Remark 4.4, we obtain averaging with rate $O(\epsilon^{\frac{1}{2}-})$ 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 [32] introduced a C^3 open class of multidimensional nonuniformly expanding maps $T_\epsilon : M \to M$. For definiteness, we restrict attention to the case $M = S^1 \times \mathbb{R}$. Let $S : M \to M$ 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, $0 \le \epsilon < \epsilon_0$ 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 $\Lambda_\epsilon = \operatorname{supp} \nu_\epsilon$ 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 2.8 is applicable by Remark 2.9. Hence we obtain averaging in $L^1(\nu_\epsilon)$ and in $L^1(\mu)$ for all absolutely continuous μ.

6 Averaging for continuous time fast-slow systems

Let $\phi_t^\epsilon : M \to M$, $0 \le \epsilon < \epsilon_0$, be a family of semiflows defined on the metric space (M, d_M). For each $\epsilon \ge 0$, let ν_ϵ denote a ϕ_t^ϵ-invariant ergodic Borel probability measure. Let $a : \mathbb{R}^d \times M \times [0, \epsilon_0) \to \mathbb{R}^d$ be a family of vector fields on $\mathbb{R}^d$ satisfying conditions (2.3)—(2.5).

We consider the family of fast-slow systems

$$\dot{x}^{(\epsilon)} = \epsilon a(x^{(\epsilon)}, y^{(\epsilon)}, \epsilon), \quad x^{(\epsilon)}(0) = x_0,$$
$$y^{(\epsilon)}(t) = \phi_t^\epsilon y_0,$$

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 $\hat{x}^{(\epsilon)} : [0,1] \to \mathbb{R}^d$ by setting $\hat{x}^{(\epsilon)}(t) = x^{(\epsilon)}(t/\epsilon)$. Let $X : [0,1] \to \mathbb{R}^d$ 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 $E = \{x \in \mathbb{R}^d : |x - x_0| \le L_1\}$. 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* $\delta_\epsilon = \delta_{1,\epsilon} + \delta_{2,\epsilon} : M \to \mathbb{R}$ where

$$\delta_{1,\epsilon} = \sup_{x\in E} \sup_{0\le t\le 1/\epsilon} \epsilon |v_{\epsilon,x,t}| \qquad \text{where} \qquad v_{\epsilon,x,t} = \int_0^t v_{\epsilon,x} \circ \phi_s^\epsilon \, ds,$$

$$\delta_{2,\epsilon} = \sup_{x\in E} \sup_{0\le t\le 1/\epsilon} \epsilon |V_{\epsilon,x,t}| \qquad \text{where} \qquad V_{\epsilon,x,t} = \int_0^t (Dv_{\epsilon,x}) \circ \phi_s^\epsilon \, ds.$$

The next results 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). *If* $\delta_\epsilon \le \frac{1}{2}$, *then* $z_\epsilon \le 6e^{2L}(\delta_\epsilon + S_\epsilon)$.

As in the discrete time setting, we say that ν_0 is statistically stable if $\nu_\epsilon \to_w \nu_0$. Proposition 2.5 goes through unchanged and statistical stability implies that $S_\epsilon \to 0$.

If the measures ν_ϵ are absolutely continuous with respect to a reference measure m on M, we define the densities $\rho_\epsilon = d\nu_\epsilon/dm$ and set $R_\epsilon = \int_M |\rho_\epsilon - \rho_0|\, dm$. Then ν_0 is strongly statistically stable if $R_\epsilon \to 0$. Proposition 2.6 and Corollaries 2.7 and 2.8 go through unchanged from the discrete time setting.

Fix a Borel subset $M' \subset M$ and a reference Borel measure m' on M'. Let $h_\epsilon : M' \to \mathbb{R}^+$ be a family of Lipschitz functions such that $\phi^\epsilon_{h_\epsilon(y)}(y) \in M'$ for almost all $y \in M'$. The map $T_\epsilon : M' \to M'$, $T_\epsilon(y) = \phi^\epsilon_{h_\epsilon(y)}(y)$, 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 $M_{h_\epsilon} = \{(y,u) \in M' \times \mathbb{R} : 0 \le u \le h_\epsilon(y)\}/\sim$ where $(y, h_\epsilon(y)) \sim (T_\epsilon y, 0)$. The suspension semiflow $f_t^\epsilon : M_{h_\epsilon} \to M_{h_\epsilon}$ is given by $f_t^\epsilon(y,u) = (y, u+t)$ computed 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 $\pi_\epsilon : M_{h_\epsilon} \to M$ 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 $K_2 \ge K_1 \ge 1$ such that for all $x, y \in M'$, $\epsilon \in [0, \epsilon_0)$,

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

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

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

$$\int_M v(d\nu_\epsilon - d\nu_0) \le 3K_2^4\|v\|_{\mathrm{Lip}}\epsilon + K_1^3|v|_\infty\Big|\int_{M'} h_0(d\nu_\epsilon' - d\nu_0')\Big| + K_1\Big|\int_{M'} \tilde v(d\nu_\epsilon' - d\nu_0')\Big|.$$

Proof We have

$$\begin{aligned}\int_M v\,(d\nu_\epsilon - d\nu_0) &= \int_{M_{h_\epsilon}} v\circ\pi_\epsilon\, d\nu_\epsilon'' - \int_{M_{h_0}} v\circ\pi_0\, d\nu_0'' \\ &= (1/\bar h_\epsilon)\int_{M'}\int_0^{h_\epsilon} v\circ\pi_\epsilon\, du\, d\nu_\epsilon' - (1/\bar h_0)\int_{M'}\int_0^{h_0} v\circ\pi_0\, du\, d\nu_0' \\ &= I_1 + I_2 + I_3 + I_4\end{aligned}$$

where

$$I_1 = (1/\bar h_\epsilon - 1/\bar h_0)\int_{M'}\int_0^{h_\epsilon} v\circ\pi_\epsilon\, du\, d\nu_\epsilon',\ I_2 = (1/\bar h_0)\int_{M'}\int_0^{h_\epsilon}(v\circ\pi_\epsilon - v\circ\pi_0)\, du\, d\nu_\epsilon',$$

$$I_3 = (1/\bar h_0)\Big(\int_{M'}\int_0^{h_\epsilon} v\circ\pi_0\, du\, d\nu_\epsilon' - \int_{M'}\int_0^{h_0} v\circ\pi_0\, du\, d\nu_\epsilon'\Big),$$

$$I_4 = (1/\bar h_0)\Big(\int_{M'}\int_0^{h_0} v\circ\pi_0\, du\, d\nu_\epsilon' - \int_{M'}\int_0^{h_0} v\circ\pi_0\, du\, d\nu_0'\Big).$$

Now

$$\begin{aligned}|I_1| &\le K_1^2|\bar h_\epsilon - \bar h_0|K_1|v|_\infty \le K_1^3|v|_\infty\Big(K_1\epsilon + \Big|\int_{M'} h_0\,(d\nu_\epsilon' - d\nu_0')\Big|\Big), \\ |I_2| &\le K_1^2 \sup_{y\in M'}\sup_{0\le u\le K_1} \operatorname{Lip} v\, d_M(\phi_u^\epsilon y, \phi_u^0 y) \le K_1^2 K_2 \mathrm{Lip}\, v\,\epsilon, \\ |I_3| &\le K_1|v|_\infty|h_\epsilon - h_0|_\infty \le K_1^2|v|_\infty\epsilon, \qquad |I_4| \le K_1\Big|\int_{M'}\tilde v\,(d\nu_\epsilon' - d\nu_0')\Big|.\end{aligned}$$

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 $S_\epsilon \le 3K_2^4 L\Big(\epsilon + \int_{M'} |\rho_\epsilon' - \rho_0'|\, dm'\Big)$.*

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

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

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

Note that $\int_{M'} w_{\epsilon,x}\, d\nu_\epsilon' = 0$ and $w_{\epsilon,x}$ is d_M-Lipschitz with $\|w_\epsilon\|_{\mathrm{Lip}} \le 2K_1K_2L$. Let

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

We can now state our main result for this section.

Lemma 6.5 *Let $q \ge 1$. Then $|\delta_{1,\epsilon}|_{L^q(\nu_\epsilon)} \le (|\Delta_{1,\epsilon}|_{L^q(\nu_\epsilon')} + 4K_1L\epsilon)$.*

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 $N_{\epsilon,t} : M_{h_\epsilon} \to \{0, 1, \dots, 1 + [K_1t]\}$ 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 $|H_{\epsilon,x}|_\infty \le 2K_1L$. Hence

$$\begin{aligned}\sup_{s\le t} |v_{\epsilon,x,s}| \circ \pi_\epsilon(y, u) = \sup_{s\le t} |\hat{v}_{\epsilon,x,s}(y, u)| &\le \sup_{s\le t} |w_{\epsilon,x,N_s(y,u)}(y)| + 4K_1L \\ &\le \max_{j\le 1+K_1t} |w_{\epsilon,x,j}(y)| + 4K_1L.\end{aligned}$$

It follows that

$$\epsilon \sup_{s\le 1/\epsilon} |v_{\epsilon,x,s}| \circ \pi_\epsilon(y, u) \le \Delta_{1,\epsilon}(y) + 4K_1L\epsilon,$$

and so $\delta_{1,\epsilon} \circ \pi_\epsilon(y, u) \le \Delta_{1,\epsilon}(y) + 4K_1L\epsilon$. The result follows. ∎

As a consequence of Proposition 6.4 and Lemma 6.5, our results for maps go through immediately for semiflows. For example, suppose that the maps $T_\epsilon(y) = \phi^\epsilon_{h_\epsilon(y)}(y)$ 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 $O(\epsilon^{\frac{1}{2}-})$. If moreover, ν_0 is strongly statistically stable, then we obtain averaging in $L^1(\nu_0)$ with rate $O(\epsilon^{\frac{1}{2}-})$, 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.

Fix $\beta > 0$. We consider the family of maps $T_\epsilon : [0,1] \to [0,1]$ given by $T_\epsilon y = 2y + \epsilon^\beta \bmod 1$ 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 $y_0 \in [0,1]$, $\limsup_{\epsilon\to 0} \hat{x}^{(\epsilon)}(1) = 1$.*

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

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

Choose ϵ such that

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

Then

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

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

Now

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

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

A Proof of first order averaging

In this appendix, we prove Theorem 2.2(a). Our proof is analogous to that of [30, Theorem 4.3.6], but our setup is different, and we work with discrete time rather than continuous time.

First, we consider the case where T_ϵ is independent of ϵ. Let $T : M \to M$ be a transformation, and $a : \mathbb{R}^d \times M \to \mathbb{R}^d$ and $\bar{a} : \mathbb{R}^d \to \mathbb{R}^d$ be functions satisfying $\|a\|_{\mathrm{Lip}} = \sup_{y \in M} \|a(\cdot, y)\|_{\mathrm{Lip}} \le L$ and $\|\bar{a}\|_{\mathrm{Lip}} \le L$, where $L \ge 1$.

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

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

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

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

$$\delta_{1,\epsilon} = \sup_x \sup_{1 \le n \le 1/\epsilon} \epsilon \Big| \sum_{j=0}^{n-1} (a(x, y_j) - \bar{a}(x)) \Big|.$$

Theorem A.1 *For all $\epsilon > 0$, $x_0 \in \mathbb{R}^d$, $y_0 \in M$ and $t \in [0,1]$,*

$$|\hat{x}_\epsilon(t) - X(t)| \le 5e^{2L} \Big(\sqrt{\delta_{1,\epsilon}(y_0)} + \epsilon \Big).$$

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

Proposition A.2 *Suppose that $b_n \ge 0$ and that there exist constants $C, D \ge 0$ such that $b_n \le C + D \sum_{m=0}^{n-1} b_m$ for all $n \ge 0$. Then $b_n \le C(D+1)^n$.*

Proof This follows by induction. ∎

We are interested in the sequences $x_n \in \mathbb{R}^d$, $y_n \in M$ for $n \le \epsilon^{-1}$, and for convenience and without loss of generality we assume that $a(x, y_n) = \bar{a}(x)$ for all $n > \epsilon^{-1}$ and all $x \in \mathbb{R}^d$. This gives us

$$\Big| \sum_{m=0}^{n-1} (a(x, y_m) - \bar{a}(x)) \Big| \le \frac{\delta_{1,\epsilon}(y_0)}{\epsilon} \tag{A.1}$$

for every n, from the definition of $\delta_{1,\epsilon}$.

Define the local averages of x and a,

$$x_{N,n} = \frac{1}{N} \sum_{m=0}^{N-1} x_{n+m}, \qquad a_{N,n}(x) = \frac{1}{N} \sum_{m=0}^{N-1} a(x, y_{n+m}).$$

The proof consists mainly of five short lemmas, all of them uniform in $N, n \in \mathbb{Z}$, $x_0 \in \mathbb{R}^d$, $y_0 \in M$, $\epsilon > 0$, with $0 \le n \le \epsilon^{-1}$.

Lemma A.3 *$|x_n - x_{N,n}| \le L\epsilon N$.*

Proof We have

$$x_{N,n} = \frac{1}{N}\sum_{m=0}^{N-1}\Big[x_n + \epsilon\sum_{k=n}^{n+m-1} a(x_k,y_k)\Big] = x_n + \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=n}^{n+m-1} a(x_k,y_k),$$

and so $|x_n - X_{N,n}| \le \epsilon N|a|_\infty \le L\epsilon N$. ∎

Lemma A.4 $\Big|x_{N,n} - x_0 - \epsilon\sum_{m=0}^{n-1} a_{N,m}(x_m)\Big| \le 2L^2\epsilon N.$

Proof First compute that

$$\begin{aligned}
x_{N,n} - x_0 &= x_n - x_0 + \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=n}^{n+m-1} a(x_k,y_k) \\
&= \epsilon\sum_{k=0}^{n-1} a(x_k,y_k) + \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=n}^{n+m-1} a(x_k,y_k) = \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=0}^{n+m-1} a(x_k,y_k).
\end{aligned}$$

Next,

$$\sum_{m=0}^{n-1} a_{N,m}(x_m) = \frac{1}{N}\sum_{m=0}^{n-1}\sum_{k=0}^{N-1} a(x_m, y_{m+k}) = A(n,N) + B(n,N),$$

where

$$\begin{aligned}
A(n,N) &= \frac{1}{N}\sum_{m=0}^{n-1}\sum_{k=0}^{N-1}[a(x_m,y_{m+k}) - a(x_{m+k},y_{m+k})], \\
B(n,N) &= \frac{1}{N}\sum_{m=0}^{n-1}\sum_{k=0}^{N-1} a(x_{m+k},y_{m+k}).
\end{aligned}$$

Note that $|A(n,N)| \le \frac{L}{N}\sum_{m=0}^{n-1}\sum_{k=0}^{N-1}|x_m - x_{m+k}|$ and $|x_m - x_{m+k}| \le L\epsilon k \le L\epsilon N$ so that $|A(n,N)| \le L^2 n\epsilon N \le L^2 N$. Also

$$B(n,N) = \frac{1}{N}\sum_{m=0}^{N-1}\sum_{k=0}^{n-1} a(x_{m+k},y_{m+k}) = \frac{1}{N}\sum_{m=0}^{N-1}\sum_{k=m}^{n+m-1} a(x_k,y_k).$$

Altogether,

$$\begin{aligned}
&x_{N,n} - x_0 - \epsilon\sum_{m=0}^{n-1} a_{N,m}(x_m) \\
&\quad = \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=0}^{n+m-1} a(x_k,y_k) - \epsilon A(n,N) - \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=m}^{n+m-1} a(x_k,y_k) \\
&\quad = \frac{\epsilon}{N}\sum_{m=0}^{N-1}\sum_{k=0}^{m-1} a(x_k,y_k) - \epsilon A(n,N).
\end{aligned}$$

Hence

$$|x_{N,n} - x_0 - \epsilon \sum_{m=0}^{n-1} a_{N,m}(x_m)| \le L\epsilon N + \epsilon|A(n,N)| \le 2L^2\epsilon N,$$

as required. ∎

Define a sequence $u_n \in \mathbb{R}^d$ by setting $u_0 = x_0$ and

$$u_{n+1} = u_n + \epsilon a_{N,n}(u_n).$$

Lemma A.5 $|x_n - u_n| \le 3L^2 e^L \epsilon N$.

Proof Note that $\mathrm{Lip}(a_{N,n}) \le L$ for all N, n. Hence it follows from Lemma A.4 that

$$|x_{N,n} - u_n| \le \epsilon \sum_{m=0}^{n-1} |a_{N,m}(x_m) - a_{N,m}(u_m)| + 2L^2\epsilon N \le L\epsilon \sum_{m=0}^{n-1} |x_m - u_m| + 2L^2\epsilon N.$$

By Lemma A.3,

$$|x_n - u_n| \le L\epsilon \sum_{m=0}^{n-1} |x_m - u_m| + 3L^2\epsilon N.$$

By Proposition A.2, $|x_n - u_n| \le 3L^2\epsilon N(1+L\epsilon)^n \le 3L^2\epsilon N(1+L\epsilon)^{1/\epsilon} \le 3L^2 e^L \epsilon N$. ∎

Define a sequence $z_n \in \mathbb{R}^d$ with

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

Lemma A.6 $|u_n - z_n| \le \dfrac{2e^L \delta_{1,\epsilon}}{\epsilon N}$.

Proof By equation (A.1),

$$|a_{N,n}(x) - \bar{a}(x)| = \frac{1}{N}\Big| \sum_{m=0}^{n+N-1} [a(x, y_m) - \bar{a}(x)] - \sum_{m=0}^{n-1} [a(x, y_m) - \bar{a}(x)] \Big| \le \frac{2\delta_{1,\epsilon}}{\epsilon N},$$

for all x. Hence,

$$\begin{aligned}|u_n - z_n| &\le \epsilon \sum_{m=0}^{n-1} |a_{N,m}(u_m) - \bar{a}(z_m)| \\ &\le \epsilon \sum_{m=0}^{n-1} |a_{N,m}(u_m) - \bar{a}(u_m)| + \epsilon \sum_{m=0}^{n-1} |\bar{a}(u_m) - \bar{a}(z_m)| \le \frac{2\delta_{1,\epsilon}}{\epsilon N} + L\epsilon \sum_{m=0}^{n-1} |u_m - z_m|.\end{aligned}$$

By Proposition A.2, $|u_n - z_n| \le \frac{2\delta_{1,\epsilon}}{\epsilon N}(1+L\epsilon)^n \le \frac{2e^L\delta_{1,\epsilon}}{\epsilon N}$. ∎

Lemma A.7 $|X(n\epsilon) - z_n| \le L^2 e^L \epsilon$.

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 $|\bar{a}(X(t_1)) - \bar{a}(X(t_2))| \le L|X(t_1) - X(t_2)| \le L^2|t_1 - t_2|$ for all t_1, t_2, we obtain that

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

Hence

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

The result follows from Proposition A.2. ∎

Proof of Theorem A.1 For convenience, we write $L^2 e^L \le e^{2L}$. Also, $2e^L \le e^{2L}$. By Lemmas A.5 and A.6,

$$|x_n - z_n| \le \frac{e^{2L}\delta_{1,\epsilon}}{\epsilon N} + 3e^{2L}\epsilon N. \tag{A.2}$$

Choosing $N = [\delta_{1,\epsilon}^{1/2}/\epsilon] + 1$, we obtain $|x_n - z_n| \le 4e^{2L}\delta_{1,\epsilon}^{1/2} + 3e^{2L}\epsilon$.

Overall, using Lemma A.7,

$$\begin{aligned}|x_{[t/\epsilon]} - X(t)| &\le |X(t) - X([t/\epsilon]\epsilon)| + |X([t/\epsilon]\epsilon) - z_{[t/\epsilon]}| + |z_{[t/\epsilon]} - x_{[t/\epsilon]}| \\ &\le \epsilon L + L^2 e^L \epsilon + 4e^{2L}\sqrt{\delta_{1,\epsilon}} + 3e^{2L}\epsilon \le 5e^{2L}(\sqrt{\delta_{1,\epsilon}} + \epsilon).\end{aligned}$$

This completes the proof. ∎

Proof of Theorem 2.2(a) 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 that

$$|\hat{x}_\epsilon(t) - X_\epsilon(t)| \le 5e^{2L}\big(\sqrt{\delta_{1,\epsilon}} + \epsilon\big)$$

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

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

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

Next, $A_\epsilon \le L\epsilon + \sup_{x\in E}|\int_M a(x,y,0)\,(d\nu_\epsilon - d\nu_0)(y)|$. Combining these estimates we obtain that

$$|\hat{x}_\epsilon(t)-X(t)| \le 5e^{2L}\sqrt{\delta_{1,\epsilon}} + 6e^{2L}\epsilon + e^L \sup_{x\in E}\Big|\int_M a(x,y,0)\,(d\nu_\epsilon - d\nu_0)(y)\Big|,$$

yielding the result. ∎

B Proof of second order averaging

In this appendix, we prove Theorem 2.2(b). This is a quantitative version of a result due to [29] with a somewhat simplified proof. Again we work with discrete time rather than continuous time.

As in Appendix A, we consider first the case where $T : M \to M$ is independent of ϵ. Suppose that $a : \mathbb{R}^d \times M \to \mathbb{R}^d$ and $\bar{a} : \mathbb{R}^d \to \mathbb{R}^d$ are functions. Assume that $\|a\|_{\mathrm{Lip}} \le L$ and $\|Da\|_{\mathrm{Lip}} \le L$ where $D = \frac{d}{dx}$ and $L \ge 1$.

Define $\delta_\epsilon = \delta_{1,\epsilon} + \delta_{2,\epsilon}$ where

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

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

Theorem B.1 *Let $\epsilon > 0$, $x_0 \in \mathbb{R}^d$. For all $y_0 \in M$ with $\delta_\epsilon(y_0) \le \frac{1}{2}$ and $t \in [0,1]$,*

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

Define a function $u : \mathbb{R}^d \times \{0, 2, \dots, 1/\epsilon\}$ by setting $u(x,0) \equiv 0$ and

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

Proposition B.2 *For any $n \le 1/\epsilon$, we have $|u(\cdot,n)|_\infty \le 1$ and* $\mathrm{Lip}\,u(\cdot,n) \le 1$.

Proof For all x,

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

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

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

Lemma B.3 *For all $0 \le n \le 1/\epsilon$ and for all $y \in M$ with $\delta_\epsilon(y) \le \frac{1}{2}$,*

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

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

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

Thus

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

where

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

This together with the definition of x_n yields

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

where

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

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

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

The result follows by summing over n.

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

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

Also,

$$|\mathrm{I}_{n+1} - \mathrm{I}_n| \le \delta_\epsilon \mathrm{Lip} u |w_n - w_{n-1}| \le \delta_\epsilon(\epsilon|\bar{a}(w_{n-1})| + 4L\epsilon\delta_\epsilon) \le 3L\epsilon\delta_\epsilon,$$

where the second inequality follows by the induction hypothesis and the third inequality uses $\delta_\epsilon \le \frac{1}{2}$. Therefore

$$|w_{n+1} - w_n - \epsilon\bar{a}(w_n)| \le |\mathrm{I}_{n+1} - \mathrm{I}_n| + |\mathrm{II}_{n+1} - \mathrm{II}_n| \le 4L\epsilon\delta_\epsilon,$$

proving the claim. ∎

Define the sequence

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

Corollary B.4 $|x_n - z_n| \le 5\delta_\epsilon e^{2L}$ *for all* $1 \le n \le 1/\epsilon$*, and for all* $y \in M$ *with* $\delta_\epsilon(y) \le \frac{1}{2}$,

Proof Write

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

By Lemma B.3,

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

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

Proof of Theorem B.1 This is identical to the proof of Theorem A.1, using Corollary B.4 in place of (A.2) together with Lemma A.7. ∎

Proof of Theorem 2.2(b) This follows from Theorem B.1 in exactly the same way that Theorem 2.2(a) followed from Theorem A.1. ∎

Acknowledgements 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.

[7] V. Bakhtin and Y. Kifer. Nonconvergence examples in averaging. *Geometric and probabilistic structures in dynamics*. Contemp. Math. **469**, Amer. Math. Soc., Providence, RI, 2008, pp. 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.* To appear.

[10] M. Benedicks and L. Carleson. On iterations of $1-ax^2$ on $(-1,1)$. *Ann. of Math.* **122** (1985) 1–25.

[11] V. Baladi and M. Todd. Linear response for intermittent maps. Preprint, 2015.

[12] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability* **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. *Mosc. 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. *Port. 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. Gouzel. 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. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze* **XXVIII** (1999) 141–152.

[22] Y. Kifer. Averaging in difference equations driven by dynamical systems. *Astérisque* (2003), no. 287, xviii, 103–123, Geometric methods in dynamics. II.

[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. Dyn. Syst.* **13** (2005) 1187–1201.

[25] A. Korepanov. Linear response for intermittent maps with summable and non-summable decay of correlations. Preprint, 2015.

[26] C. Liverani, B. Saussol and S. Vaienti. A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems* **19** (1999) 671–685.

[27] M. Lyubich. Almost every real quadratic map is either regular or stochastic. *Ann. of Math.* **156** (2002) 1–78.

[28] I. Melbourne and R. Zweimüller. Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems. *Ann Inst. H. Poincaré (B) Probab. Statist.* **51** (2015) 545–556.

[29] J. A. Sanders. On the fundamental theorem of averaging. *SIAM J. Math. Anal.* **14** (1983) 1–10.

[30] J. A. Sanders, F. Verhulst and J. Murdock. *Averaging methods in nonlinear dynamical systems*, second ed., Applied Mathematical Sciences **59**, Springer, New York, 2007.

[31] H. Thunberg. Unfolding of chaotic unimodal maps and the parameter dependence of natural measures. *Nonlinearity* **14** (2001) 323–337.

[32] M. Viana. Multidimensional nonhyperbolic attractors. *Inst. Hautes Études Sci. Publ. Math.* (1997) 63–96.

[33] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.* **110** (1999) 153–188.