AVERAGING AND INVARIANT MEASURES.

DMITRY DOLGOPYAT

Abstract. An important approach to establishing stochastic behavior of dynamical systems is based on the study of systems expanding a foliation and of measures having smooth densities along the leaves of this foliation [44, 46, 38]. We review recent results on this subject and present some extensions and open questions.

Dedicated to Yakov Sinai on the occasion of his 70th birthday.

1. Introduction.

The study of statistical properties of dynamical systems constitutes an important branch of smooth ergodic theory. A central role in such studies is played by Sinai-Ruelle-Bowen (SRB) measures. To define them let f be a smooth diffeomorphism of a smooth compact manifold M and let μ be an f-invariant measure. Define its *basin*

$$\mathbb{B}(\mu) = \{x : \forall A \in C(M) \quad \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} A(f^j x) = \mu(A)\}.$$

Call μ an *SRB measure* if the Lebesgue measure of its basin is positive. Important problems are the existence and uniqueness of SRB measures, their statistical properties such as the rate of mixing and Central Limit Theorem and their dependence on parameters. There is a good evidence [37] that if one wants some stability results then it is natural to restrict the attention to partially hyperbolic systems. In this note we review recent results about statistical properties and stability of SRB measures [18, 19] and then show how the methods of the these two papers can be combined to obtain new results in this area.

Recall that f is called *partially hyperbolic* if there is an f–invariant splitting

$$T_xM = E_u \oplus E_c \oplus E_s$$

and constants $\lambda_1 \leq \lambda_2 < \lambda_3 \leq \lambda_4 < \lambda_5 \leq \lambda_6$, $\lambda_2 < 1$, $\lambda_5 > 1$ such that for some Riemannian metric on M we have

$$\forall v \in E_s \quad \lambda_1||v|| \leq ||df(v)|| \leq \lambda_2||v||, \tag{1}$$

$$\forall v \in E_c \quad \lambda_3||v|| \leq ||df(v)|| \leq \lambda_4||v||, \tag{2}$$

$$\forall v \in E_u \quad \lambda_5||v|| \le ||df(v)|| \le \lambda_6||v||. \tag{3}$$

For partially hyperbolic systems there is an important *a priori* information about SRB measures [38]. Recall that E_u is uniquely integrable, that is, there is a foliation W^u such that $TW^u = E_u$. A measure μ is called *u-absolutely continuous* if any set which meets each leaf at a set of zero leaf measure has μ-measure zero. An invariant u-absolutely continuous measure is called *u-Gibbs state.* u-Gibbs states have good regularity properties as we explain next. Fix $\bar\delta$ and consider the cone in the tangent space

$$\mathcal{K}_u = \{v = v_u + v_c + v_s : \quad ||v_c|| \le \bar\delta||v_u|| \quad ||v_s|| \le \bar\delta||v_u||\}. \tag{4}$$

Then for small $\bar\delta$ we have $df(\mathcal{K}_u(x)) \subset \mathcal{K}_u(fx)$. We call a set D $(\mathbf{r}, C_1, C_3, \alpha_1)$–admissible if there is an embedding ϕ from the standard unit d_u–dimensional disc $\mathbb{D}$ to M such that $||\phi||_{C^2(\mathbb{D})} \le C_3$ and if $V = \phi(\mathbb{D})$ then $TV \in \mathcal{K}_u$, $D \subset V$ and in the induced Riemannian structure on V $\text{mes}(D) > \mathbf{r}$ and $\text{mes}(\partial_\varepsilon D) \le C_1\varepsilon^{\alpha_1}$. We call a pair $\ell = (D, \rho)$ $(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ standard if D is $(C_1, C_3, \mathbf{r}, \alpha_1)$ admissible ρ is a probability density on D and $||\rho||_{C^{\alpha_2}(D)} \le C_2$. Denote

$$\mathbb{E}_\ell(A) = \int_D A(x)\rho(x)dx.$$

Let $E_1(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ be the set of all such measures. We denote by $E_2(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ the convex hall of $E_1(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ and by $E(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ the closure of $E_2(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$. Usually we will drop (some of the) parameters $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$ since their precise values are not important. In the proofs it is usually convenient to assume that C_1, C_2, C_3 are so large and $\mathbf{r}, \alpha_1, \alpha_2$ are so small that the Lebesgue measure is in $E(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ and that Theorem 1(a) holds. But occasionally it is convenient to have larger C_1, C_2, C_3 and smaller $\mathbf{r}, \alpha_1, \alpha_2$. For example if $\ell \in E(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ and $B \in C^{\alpha_2}(M)$ is a function such that $\mathbb{E}_\ell(B) = 1$ then $\tilde\ell$ defined by $\mathbb{E}_{\tilde\ell}(A) = \mathbb{E}_\ell(BA)$ is in $E(C_1, (C_2 + 1)||B||_{C^{\alpha_2}}, C_3, \mathbf{r}, \alpha_1, \alpha_2)$. On the other hand the proofs of our result depend only on mixing assumptions formulated below and if those assumptions hold for some $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$ then they hold for all $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$ (see [18]). So the results of this paper are valid for all $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$.

Observe that E_2 and hence E is almost invariant. Indeed if C_1, C_2, C_3 are large enough and $\mathbf{r}, \alpha_1, \alpha_2$ are small enough then for all admissible D for all n

there are admissible D_i such that

$$f^n D = \left(\bigcup_i D_i\right) \bigcup Z, \text{ where} \tag{5}$$

$$\text{mes}(f^{-n} Z) \leq \text{Const}\theta^n. \tag{6}$$

Here and below θ is a constant which is less than 1. Its precise value can change from entry to entry. (In (6) Z consists of the points such that $\text{dist}(z, \partial(f^n D)) \leq \text{Const}$. Thus $\text{dist}(f^{-n}z, \partial D) \leq \text{Const}\theta^n$ giving (6).)

Consequently for all $\ell \in E_1$ for all n there exist $\ell_i \in E_1$ such that

$$\mathbb{E}_\ell(A \circ f^n) = \sum_i c_i \mathbb{E}_{\ell_i}(A) + \zeta(A) \text{ where } ||\zeta|| \leq \text{Const}\theta^n. \tag{7}$$

We shall call (5) and (7) *almost Markov decompositions* of $f^n D$ and $\mathbb{E}_\ell(\cdot \circ f^n)$ respectively.

Define a *Markov family* $\mathcal{P}$ as a collection of $(\mathbf{r}, C_1, C_3, \alpha_1)$-admissible sets such that for any $D \in \mathcal{P}$ for any $n > 0$ there is a *Markov decomposition*

$$f^n D = \bigcup_j D_j.$$

An argument of [44] (see also [45]) shows that for any admissible set $\mathcal{D}$ for any δ there is a Markov family $\mathcal{P}$ such that $f^n D$ has a decomposition (5) with $\text{mes}(f^{-n} Z) \leq \delta\theta^n$. (To show this one starts with a family of $(\mathbf{r}, C_1, C_3, \alpha_1)$-admissible sets and then modifies the elements of the decomposition (5) for D to get the Markov property by consecutive approximations.)

We are now ready to explain the relation between SRB measures and u-Gibbs states.

Theorem 1. [38, 18] *(a) There are constants $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$ such that any u-Gibbs state is in $E(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$.*

(b) If $f_j \to f$ in C^2 and $\nu_j \to \nu$ where ν_j are u-Gibbs for f_j then ν is u-Gibbs for f.

(c) For any $C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2$ for any sequence $\ell_n \in E(C_1, C_2, C_3, \mathbf{r}, \alpha_1, \alpha_2)$ any limit point of

$$\mu_n(A) = \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}_{\ell_n}(A \circ f^j)$$

is u-Gibbs.

(d) For $A \in C^\alpha(M)$ let $I(A) = \{\nu(A)\}_{\nu \text{ is u-Gibbs}}$. Then for each $\varepsilon > 0$ there are constants $c_\varepsilon, C_\varepsilon$ such that for each $\ell \in E$

$$\ell\left(\text{dist}\left(\frac{1}{n}\sum_{j=0}^{n-1} A(f^j x), I(A)\right) \geq \varepsilon\right) \leq C_\varepsilon e^{-c_\varepsilon n}.$$

(e) Any SRB measure is u-Gibbs.

(f) If there is unique u-Gibbs measure then it is SRB.

Call f *uuniquely ergodic* if it has unique u-Gibbs state. In [18] we obtained several limit theorems for uuniquely ergodic systems which enjoy sufficiently fast convergence to equilibrium.

Namely we suppose that there is a Banach algebra $\mathcal{B}$ of functions on M such that $||A||_{C^{\bar{\alpha}}(M)} \leq ||A||_{\mathcal{B}}$ for some $\bar{\alpha} > 0$ and that there exists a sequence $a(n)$ with $\sum_n a(n) < \infty$ such that for all $\ell \in E$ for all $A \in \mathcal{B}$

$$|\mathbb{E}_\ell(A \circ f^n) - \nu(A)| \leq a(n)||A||_{\mathcal{B}} \tag{8}$$

Let $s > 0$ be given. Let $A : M \to \mathbb{R}^s$ be a function such that each coordinate map A_β is in $\mathcal{B}$. (8) and Theorem 1 imply that there exist the limits

$$\sigma^2_{\alpha\beta}(A) = \sum_{j=-\infty}^{\infty} \left[\nu(A_\alpha(A_\beta \circ f^j)) - \nu(A_\alpha)\nu(A_\beta)\right].$$

Theorem 2. *([18]) If x is distributed according to some $\ell \in E$ then*

$$\frac{\left(\sum_{j=0}^{n-1} A(f^j x)\right) - n\nu(A)}{\sqrt{n}} \to \mathcal{N}(0, \sigma^2(A)).$$

In case $a(n) < C/n^2$ we also obtained nonlinear versions of Theorem 2 known as averaging theorems.

Consider the sequence $z_n \in \mathbb{R}^s$ given by

$$z_{n+1} - z_n = \varepsilon A(f^n x, z_n), \quad z_0 = z^* \tag{9}$$

where function $A(x, z)$ is three times differentiable with respect to z and the norms $||\frac{\partial^\alpha A(\cdot,z)}{\partial^\alpha z}||_{\mathcal{B}^s}$, are uniformly bounded for $0 \leq |\alpha| \leq 3$. Let q_n be the solution of the averaged equation

$$q_{n+1} - q_n = \varepsilon \bar{A}(q_n), \quad q_0 = z^*.$$

where

$$\bar{A}(q) = \int A(x, q) d\nu(x).$$

Let $\Delta_n = z_n - q_n$. Denote $\Delta^\varepsilon_t = \frac{\Delta_{[\frac{t}{\varepsilon}]}}{\sqrt{\varepsilon}}$. Let $q(s)$ be the solution of

$$\dot{q} = \bar{A}(q), \quad q(0) = z^*.$$

Theorem 3. *(a) Let f satisfy (8) with $a(n) = C/n^2$. If x is distributed according to some $\ell \in E$ then as $\varepsilon \to 0$ Δ_t^ε converges weakly to the solution of*

$$d\Delta(t) = D\bar{A}(q(t))\Delta dt + d\mathbf{B}$$

where $\mathbf{B}$ is a Gaussian process with independent increments, zero mean and covariance matrix

$$\mathbb{E}\left(\mathbf{B}_\alpha(t)\mathbf{B}_\beta(t)\right) = \int_0^t \sigma^2_{\alpha\beta}(A(\cdot, q(s)))ds. \tag{10}$$

(b) Let f satisfy (8) with $a(n) = o(1/n^2)$. Suppose that A in (9) has zero mean

$$\bar{A}(z) = \int A(x, z)d\nu(x) \equiv 0.$$

Let $Z_t^\varepsilon = z^\varepsilon_{[\frac{t}{\varepsilon^2}]}$, then as $\varepsilon \to 0$ Z_t^ε converges weakly to the diffusion process $Z(t)$ with drift

$$a(z) = \sum_{n=1}^{\infty} \int \frac{\partial A}{\partial z}(f^n x, z)A(x, z)d\nu(x)$$

and diffusion matrix $\sigma(A(\cdot, z))$.

Examples of systems satisfying (8) with $a(n) < o(1/n^2)$ include (generic elements of the) following classes of systems.

1. Anosov diffeomorphisms [6];
2. time one maps of Anosov flows [15, 34, 22];
3. partially hyperbolic translations on homogeneous spaces [32];
4. compact group extensions of Anosov diffeomorphisms [16];
5. partially hyperbolic toral automorphisms [28];
6. mostly contracting systems (see Section 9 for the definition) [17, 10, 11].

We refer to the above mentioned papers for the precise statements. Later we shall need stronger mixing assumptions. We say that f has *stretched exponential decay of correlations* if (8) holds with

$$a(n) = c_1 \exp(-c_2 n^\gamma)$$

and $\mathcal{B} = C^{\bar{\alpha}}(M)$ for some $c_1, c_2, \gamma, \bar{\alpha}$. We say that a family of diffeomorphisms $\{f_z\}$ has uniform stretched exponential decay of correlations if $c_1, c_2, \bar{\alpha}$ and γ can be chosen the same for all values of the parameter. We say that f is *exponentially mixing* if $\gamma = 1$. Examples (1), (3), (5) and (6) are exponentially mixing as well as contact Anosov flows in example (2).

Theorem 1(b) shows that u-Gibbs states have better continuity properties than SRB measures (which need not exist and may vary discontinuously with parameters [48]). Thus it is natural to study stability of u-Gibbs states in more details. This gives information about the SRB measures via Theorem 1(e),(f). In [19] this is achieved for Anosov elements of abelian Anosov actions. By this we mean that E_c is tangent to orbits of an $\mathbb{R}^d$-action $\mathbf{g}_t : \mathbb{R}^d \times M \to M$ and $f\mathbf{g}_t = \mathbf{g}_t f$. This setting includes classes (1)-(4) described above. However we need stronger mixing assumptions (satisfied in the above examples). To describe those assumptions we need to introduce a relevant function space.

Let $C_k^{\bar\alpha}(M)$ be the space of functions A such that for all $x \in M$, the function $t \to A(\mathbf{g}_t x) \in C^k(\mathbb{R}^d)$ and $[\partial_t^j(A(\mathbf{g}_t x))]|_{t=0} \in C^{\bar\alpha}(M)$ for $0 \leq |j| \leq k$. Here j is a multiindex $(j_1, j_2 \dots j_d)$ and $\partial_t^j = \partial_{t_d}^{j_d} \dots \partial_{t_1}^{j_1}$, $|j| = \sum_{k=1}^d j_k$. Denote

$$||A||_{C_k^{\bar\alpha}}(M) = \max_{o \leq |j| \leq k} \left\| \left[\partial_t^j (A(\mathbf{g}_t x))\right]_{t=0} \right\|_{C^\alpha(M)}.$$

We assume that for all m there exist k and C such that for all $\ell \in E$

$$\left|\mathbb{E}_\ell(A \circ f^N) - \nu(A)\right| \leq C(m)||A||_{C^{\bar\alpha}_{k(m)}(M)} N^{-m}. \tag{11}$$

We shall call a diffeomorphism satisfying (11) *rapidly mixing.* We call a family of diffeomorphisms $\{f_z\}$ uniformly rapidly mixing if each f_z is an Anosov element in an abelian Anosov action and for each m there are $C(m)$, $k(m)$ such that (11) holds for all f_z.

Observe that rapid mixing can be defined by the requirement that for any $(C_1, C_3, \mathbf{r}, \alpha_1)$ admissible D for any $\rho \in C^{\alpha_2}(D)$ for all $A \in C_k^{\bar\alpha}(M)$ we have

$$\left| \int_D \rho(x) A(f^N x) dx - \nu(A) \int_D \rho(x) dx \right| \leq C(m) ||A||_{C_k^{\bar\alpha}(M)} ||\rho||_{C^{\alpha_2}(D)} N^{-m}. \tag{12}$$

Indeed (12) implies (11) for all $\ell \in E_1$ and by convexity (11) then holds for all $\ell \in E$. Conversely if (11) holds for all $\ell \in E_1$ then splitting

$$\rho = 2||\rho||_{C^{\alpha_2}(M)} - \left[2||\rho||_{C^{\alpha_2}(M)} - \rho\right]$$

and applying (11) to each term (normalized to integrate to 1) we get (12).

Let $VC_k^{\bar\alpha}(M)$ denote the set of the vector fields on M which are C^k along the orbits of $\mathbf{g}_t$ and the derivatives are $\bar\alpha$-Holder. Define $||\cdot||_{VC_k^{\bar\alpha}(M)}$ similarly to $||\cdot||_{C_k^{\bar\alpha}(M)}$. Let $C_{1,k}^{\bar\alpha}(M)$ be the space of functions such that for any $VC^{\bar{\alpha}}_k-$ vectorfield v the function $\partial_v A \in C^{\bar{\alpha}}_k(M)$. Let

$$||A||_{C^{\bar{\alpha}}_{1,k}} = \sup_{||v||_{VC^{\bar{\alpha}}_k}\leq 1} ||\partial_v A||_{C^{\bar{\alpha}}_k} + \sup_{x\in M} |A(x)|.$$

Let $\mathbb{V}^k(M)$ be the set of C^k vectorfields on M.

Theorem 4. *([19]) Let f be a C^∞ rapidly mixing Anosov element in an abelian Anosov action. Let ν be its u-Gibbs state. Then there exists a number k_0 and a bilinear form $\omega : C^{\bar{\alpha}}_{1,k_0}(M) \times \mathbb{V}^k(M) \to \mathbb{R}$ such that the following holds. Let $\{f_\varepsilon\}$ be a C^∞ one parameter family of diffeomorphisms such that $f = f_0$. Set*

$$X(x) = \frac{d}{d\varepsilon}|_{\varepsilon=0}(f_\varepsilon(f^{-1}x)).$$

For each ε choose a u-Gibbs state ν_ε for f_ε. Then for all $B \in C^{\bar{\alpha}}_{1,k_0}(M)$ we have

$$\nu_\varepsilon(B) - \nu(B) = \varepsilon\omega(B, X) + o\left(\varepsilon||B||_{C^{\bar{\alpha}}_{1,k_0}(M)}\right). \tag{13}$$

This theorem was proven before for Anosov diffeomorphisms and flows (see [1, 14, 26, 39]) but the novelty of [19] is that we only assume that f_ε is strongly mixing for $\varepsilon = 0$ rather than for *all* small ε.

We shall write $\omega(B, X, f)$ instead of $\omega(B, X)$ if we want to emphasize the dependence on the diffeomorphism.

2. Extensions.

Here we present several extensions of the above results. Observe that in (9) the fast motion is independent of slow one. Our first result removes this restriction. Consider the recurrence

$$x_{n+1} = f(x_n, z_n, \varepsilon).$$

$$z_{n+1} = z_n + \varepsilon A(x_n, z_n, \varepsilon) \tag{14}$$

We assume throughout this section that $A \in C^\infty(M \times \mathbb{R}^s)$, that A and all its partial derivatives of any order are uniformly bounded and that the maps $f_z(x) = f(x, z, 0)$ are uniformly partially hyperbolic, that is that they satisfy (1)(3) with the same λ_1–λ_6 and the same Riemannian metric on M. Let

$$F_\varepsilon(x, z) = (f(x, z, \varepsilon), z + \varepsilon A(x, z, \varepsilon))$$

so that $F_\varepsilon(x_n, z_n) = (x_{n+1}, z_{n+1})$. Then F_ε are partially hyperbolic for small ε. Let $\mathcal{K}_u(x) \subset TM$ be the cones (4) for f_z. Then

$$\mathcal{K}^\varepsilon_u(x, z) = \{(v, u) \in TM \times \mathbb{R}^s : \quad v \in \mathcal{K}_u(x), \quad ||u|| \leq C_4\varepsilon||v||\}$$

satisfy $dF_\varepsilon(\mathcal{K}_u^\varepsilon(x,z)) \subset \mathcal{K}_u^\varepsilon(F_\varepsilon(x,z))$ provided that C_4 is large enough. We define the spaces of measures $E_1(F_\varepsilon)$, $E_2(F_\varepsilon)$ and $E(F_\varepsilon)$ as above using $\mathcal{K}_u^\varepsilon$ to define admissible sets. Let $E(z^*,\varepsilon)$ ($E_1(z^*,\varepsilon)$, $E_2(z^*,\varepsilon)$) denote the subset of $E(F_\varepsilon)$ (respectively $E_1(F_\varepsilon), E_2(F_\varepsilon)$) consisting of measures such that almost surely $|z-z^*| \le C_4\varepsilon$. (E.g. if ν is a measure in $E(f_{z^*})$ Then $\nu \times \delta_{z^*}$ belongs to $E(z^*,\varepsilon)$ for all ε.)

Theorem 5. *Suppose that for each z the map $f_z(x)$ is uuniquely ergodic with u-Gibbs state ν_z. Assume that the vectorfield*

$$\bar{A}(q) = \int A(x,q,0)d\nu_q(x)$$

is Lipschitz. Denote by $q(z_0,t)$ the solution of the equation

$$\dot{q} = \bar{A}(q), \quad q(0) = z_0.$$

Let (z_0,x_0) be distributed according to some measure from $E(F_\varepsilon)$. Then for all $T>0$

$$\lim_{\varepsilon\to 0} \left| z^\varepsilon_{T/\varepsilon} - q(z_0,T)\right| \to 0 \textit{ in probability.}$$

We say that f_z is a *family of Anosov elements* (FAE) if each f_z is an Anosov element for an abelian action $\mathbf{g}_{z,t}$ and the map $(x,z,t) \to \mathbf{g}_{z,t}(x)$ is C^∞ in all the variables.

Theorem 6. *Suppose that f_z are uniformly rapidly mixing FAE. Let*

$$\Delta_n = z_n^\varepsilon - q(z_0,\varepsilon n).$$

Denote $\Delta_t^\varepsilon = \frac{\Delta_{[t/\varepsilon]}}{\sqrt{\varepsilon}}$. Fix $z^* \in M$.

(a) Let (x_0,z_0) be chosen according to some measure in $E(z^*,\varepsilon)$. Then as $\varepsilon \to 0$ $\Delta_t^\varepsilon \to \Delta_t$, the process given by Theorem 3(a).

(b) For all $l \in E(F_\varepsilon)$ for all $1 \le R \le 1/\sqrt{\varepsilon}$

$$\ell\left(\sup_{[0,T]} |\Delta_t^\varepsilon| \ge R\right) \le c_1 e^{-c_2 R^2}.$$

Theorems 5 and 6 are not really new. In case when f_z are Anosov diffeomorphisms they are proven in [31] and [2, 3] respectively. The general case requires little modifications.

To formulate the analogue of Theorem 3(b) we need more notation. Introduce vectorfields on M

$$X_z(x) = \frac{d}{d\varepsilon}|_{\varepsilon=0}(f(x,z,\varepsilon)\circ f_z^{-1}),$$

$$Y_{z,n}(x) = \frac{\partial f}{\partial z}(f_z^{-1}x)A(f_z^{-(n+1)}x,z,0).$$

Theorem 7. *Let f_z be FAE having uniform stretched exponential decay of correlations. Suppose that $\bar{A} \equiv 0$.*

(a) Define

$$\sigma^2_{\alpha\beta}(z) = \sum_{j=-\infty}^{\infty} \int A_\alpha(x,z,0) A_\beta(f_z^j x, z, 0) d\nu_z,$$

$a(z) = a_1(z) + a_2(z) + a_3(z) + a_4(z)$ *where*

$$a_1(z) = \int \frac{\partial}{\partial \varepsilon}|_{\varepsilon=0} A(x,z,\varepsilon) d\nu_z(x),$$

$$a_2(z) = \sum_{n=1}^{\infty} \int \frac{\partial A}{\partial z}(f_z^n x, z, 0) A(x,z,0) d\nu_z,$$

$$a_3(z) = \omega\left(A(x,z,0), X_z, f_z\right)$$

$$a_4(z) = \sum_{n=0}^{\infty} \omega\left(A(x,z,0), Y_{z,n}, f_z\right)$$

where ω is the form defined in Theorem 4. Then σ^2 and a are uniformly bounded and uniformly continuous. Moreover σ^2 is C^1 with uniformly bounded derivatives.

(b) Choose (x_0, z_0) according to some measure in $E(z^*, \varepsilon)$. Then the family $\{z^\varepsilon_{[t/\varepsilon^2]}\}$ is tight.

(c) Any limit point of $\{z^\varepsilon_{[t/\varepsilon^2]}\}$ is a diffusion process with drift $a(z)$ and diffusion matrix $\sigma(z)$.

In particular if there is unique process with drift a and diffusion matrix σ then $z^\varepsilon_{[t/\varepsilon^2]}$ converges to this process. Examples of uniqueness include the following.

(1) Skew products. (See Theorem 3(b).) If $f(x,z,\varepsilon) \equiv f$ then a_3 and a_4 vanish and a_1 and a_2 are Lipshitz since correlations sums are smooth (in fact quadratic) functions of A. Then

we can use uniqueness result for diffusions with Lipschitz coefficients ([47]).

(2) Suppose that for each $z \in M$, for each $\xi \in \mathbb{R}^s - 0$ the function $< \xi, A(\cdot, z) >$ is not $L^2(\nu_z)$ coboundary with respect to f_z ($<,\cdot,\cdot>$ denotes the scalar product). In this case $\sigma^2(z)$ is not degenerate and boundedness and continuity of $a(z)$ suffice for uniqueness ([47]). Observe that in many cases L^2 coboundaries are smooth ([35, 36, 20]) and then one can show (see e.g [27]) that the set of coboundaries is a codimension infinity closed subspace so the uniqueness holds for generic system (14) with $\bar{A} \equiv 0$.

Theorems 5–7 allow extensions to the case of fiber bundles. Let N, Z be compact manifolds and $\pi : N \to Z$ be a fibering with compact fibers. Let $F : N \to N$ be a diffeomorphism such that $\pi \circ F = \pi$ and F restricted to each fiber is partially hyperbolic and uuniquely ergodic with u-Gibbs state ν_z. Call f_z the restriction of F to $\pi^{-1}z$. Let F_ε be a small perturbation of F. Consider a vectorfield $\mathcal{V} = \frac{d}{d\varepsilon}|_{\varepsilon=0}(F_\varepsilon \circ F^{-1})$ and let

$$Y(z) = \int (d\pi)(\mathcal{V})(x) d\nu_z(x).$$

We assume that $Y(z)$ is Lipshitz continuous. Define $E(F)$ and $E(z^*, \varepsilon)$ and FAE as in the product case. For $w \in N$ let $z_n^\varepsilon = \pi(F_\varepsilon^n w)$. We assume in Theorems 5*–7* that for each z the map f_z is uuniquely ergodic with u-Gibbs state ν_z.

Theorem 5* *Let $w = (x, z)$ be chosen according to some measure in $E(F_\varepsilon)$. Let $q(t, w)$ be the solution of*

$$\dot{q} = Y(q), \quad q(0) = \pi(w),$$

then as $\varepsilon \to 0$ $\mathrm{dist}(z^\varepsilon_{[t/\varepsilon]}, q(t, w)) \to 0$ in probability.

Theorem 6* *Suppose that f_z are uniformly rapidly mixing FAE Choose $w = (x, z)$ according to some measure in $E(z^*, \varepsilon)$. Let*

$$\Delta_t^\varepsilon = \frac{\exp^{-1}_{q(t)}(z^\varepsilon_{[t/\varepsilon]})}{\sqrt{\varepsilon}}$$

(a) Δ_t^ε converges to a Gaussian vector.
(b) For all $1 \leq R \leq 1/\sqrt{\varepsilon}$

$$\ell(\Delta_t^\varepsilon > R) \leq c_1 e^{-c_2 R^2}.$$

Theorem 7* *Suppose that in Theorem 6* $Y \equiv 0$ and that f_z are FAE having uniform stretched exponential decay of correlations. Let $w = (x, z)$ be chosen according to some measure in $E(z^*, \varepsilon)$. Then*
(a) The limit

$$(\mathcal{L}\phi)(z^*) = \lim_{h \to 0} \lim_{\varepsilon \to 0} \frac{\nu_z(\phi(z^\varepsilon_{h\varepsilon^{-2}}) - \phi(z))}{h}$$

is a second order differential operator.
(b) The family $\{z^\varepsilon_{[t/\varepsilon^2]}\}$ is tight.
(c) Any limit point of $\{z^\varepsilon_{[t/\varepsilon^2]}\}$ is a diffusion process with generator $\mathcal{L}$.

(In (a) $\lim_{\varepsilon \to 0}$ need not exist so $\lim\lim$ means that the set of limit points $\varepsilon \to 0$ of the above expression rescaled by h shrinks to the unique value.)

Continuing the study of small perturbation of F let ν_ε be a u-Gibbs state for F_ε. Let ν be a limit point for ν_ε. By Theorem 1(b) ν is u-Gibbs for F. Thus there is measure η on Z such that

$$\nu = \int \nu_z d\eta(z).$$

Our first goal is to reduce the set of possible limit measures. We follow [23] (see also [30]).

Theorem 8. *Under the conditions of Theorem 5* η is invariant with respect to the flow generated by Y.*

Still the set of invariant measures can be quite large, so one can ask if one can further restrict the range of possibilities. For example, suppose that Y is Morse-Smale so that the set of non-wandering points consists of finite number of periodic orbits $\gamma_1, \gamma_2, \dots \gamma_r$ with periods $T_1, T_2, \dots T_r$ (we let $T_j = 1$ if γ_j is a fixed point). Then Theorem 8 implies that $\eta = \sum_i c_i \delta_{\gamma_i}$. In fact all those measures will be present in case F is a direct product of some partially hyperbolic map f and a time ε map of a Morse-Smale flow Y.

Theorem 9. *If F satisfies the conditions of Theorem 6* then $c_j = 0$ unless γ_j is a sink or*

$$\int_0^{T_j} \sigma^2(\gamma_j(s))ds = 0 \tag{15}$$

is degenerate.

In case $Y \equiv 0$ Theorem 8 gives no information about η.

Theorem 10. *If F satisfies the conditions of Theorem 7* and there is a unique process with generator $\mathcal{L}$ then η is an invariant measure for this diffusion process.*

Theorems 5–7 are obtained by modifications of the arguments of [18, 19, 13, 3, 31] so we only sketch the proofs referring to the above papers for the complete details. Theorem 5 is proven in Section 3, Theorem 6 is proven in Section 6 and Theorem 7 is proven in Section 7. The proofs of Theorems 5^*–7^* are very similar and will be omitted. Sections 4-5 contain an extension of Theorem 4 needed in our proofs. The proofs of Theorems 8–10 are given are Section 8.

3. Averaging principle.

Here we present the proof of Theorem 5.

Proof of Theorem 5. We have

$$z_{n_2} - z_{n_1} = \varepsilon \sum_{j=n_1}^{n_2-1} \bar{A}(z_j) + \varepsilon \sum_{j=n_1}^{n_2-1} \left[A(x_j, z_j, \varepsilon) - \bar{A}(z_j)\right]. \tag{16}$$

We claim that for each δ there is ε_0 such that for $\varepsilon \leq \varepsilon_0$ for all $\ell \in E(F_\varepsilon)$

$$\ell\left(\left|\sum_{j=0}^{T/\varepsilon-1} \left[A(x_j, z_j, \varepsilon) - \bar{A}(z_j)\right]\right| \geq \frac{\delta T}{\varepsilon}\right) \tag{17}$$
$$\leq C_1(\delta, T) \exp\left(-\frac{c_2(\delta, T)}{\varepsilon}\right).$$

The proof of (17) is very similar to the proof of Theorem 1 in [18], (see also [31]). We recall briefly the argument since it plays an important role in the proofs of the other results as well. The proof consists of several steps.

Step 1. Let $\xi \in \mathbb{R}^s$ be a unit vector. It is enough to show that

$$\ell\left(\left|< \xi, \sum_{j=0}^{T/\varepsilon-1} \left[A(x_j, z_j, \varepsilon) - \bar{A}(z_j)\right] >\right| \geq \frac{\delta T}{\varepsilon}\right) \tag{18}$$
$$\leq C_1(\delta, T) \exp\left(-\frac{c_2(\delta, T)}{\varepsilon}\right)$$

($< \cdot, \cdot >$ denotes the scalar product). Indeed applying (18) to each coordinate map we get (17).

Step 2. By linearity it is enough to establish (18) for $\ell = (D, \rho)$.

Step 3. It is enough to prove (18) for the case then D belongs to a Markov family $\mathcal{P}$. Indeed let $n_1 = \frac{\delta}{100\varepsilon ||A||_{C^0}}$. For arbitrary D consider an almost Markov decomposition

$$f^{n_1} D = \bigcup_j D_j \bigcup Z$$

with $D_j \in \mathcal{P}$ and use that each D_j satisfies (18) (with δ replaced by $\delta/2$).

Step 4. We claim that for all $\delta > 0$ there is n_0 such that for ε small enough for each $\ell \in E_1(F_\varepsilon)$

$$\mathbb{E}_\ell\left(< \xi \sum_{j=0}^{n_0-1} \left[A(x_j, z_j, \varepsilon) - \bar{A}(z_j)\right] > -\frac{\delta n_0}{2}\right) \leq 0.$$

Indeed for fixed j the expression in brackets is

$$A(f_{z_0}^j x, z_0, 0) - \bar{A}(z_0) + O(\varepsilon).$$

On the other hand for large n_0 we must have

$$\mathbb{E}_\ell\left(<\xi, \sum_{j=0}^{n_0-1}\left[A(z_0, f_{z_0}^j x_0, 0)-\bar{A}(z_0)\right]> -\frac{\delta n_0}{2}\right)<0 \tag{19}$$

by (11).

Step 5. Using the Markov decomposition we obtain by induction

$$\mathbb{E}_\ell\left(<\xi, \sum_{j=0}^{kn_0-1}\left[A(x_j, z_j, \varepsilon)-\bar{A}(z_j)\right]> -\frac{\delta k n_0}{2}\right)\leq 0.$$

Step 6. Let $n_1 = kn_0$ by sufficiently large and let $f^{n_1}D = \bigcup_m D_m$. Since the oscillation of the sum is $O(1)$ on each $f^{-n_1}D_j$ we get

$$\mathbb{E}_\ell\left(\sum_m c_m \sup_{f^{-n_1}D_m}\left(<\xi, \sum_{j=0}^{n_1-1}\left[A(x_j, z_j, \varepsilon)-\bar{A}(z_j)\right]>\right)-\frac{3\delta n_1}{4}\right)\leq 0$$

where $c_m = \ell(f^{-n_1}D_m)$. Since $e^{ct}\sim 1+ct$ for small c we conclude that if c is sufficiently small then

$$\mathbb{E}_\ell\left(\exp\left(c\left(\sum_m c_m \sup_{f^{-n_1}D_m}\left(<\xi, \sum_{j=0}^{n_1-1}\left[A(x_j, z_j, \varepsilon)-\bar{A}(z_j)\right]>\right)-\frac{3\delta n_0}{4}\right)\right)\right)\leq \theta$$

for some $\theta<1$.

Step 7. Using again Markov decomposition we obtain by induction that for all k

$$\mathbb{E}_\ell\left(\exp\left(c\left(\sum_m c_{km} \sup_{f^{-kn_1}D_{km}}\left(<\xi, \sum_{j=0}^{kn_1-1}\left[A(x_j, z_j, \varepsilon)-\bar{A}(z_j)\right]>\right)-\frac{3\delta n_1}{4}\right)\right)\right)\leq \theta^k$$

for some Markov decomposition $f^{kn_1}D=\bigcup_m D_{km}$ and $c_{km}=\ell(f^{-kn_1}D_{km})$.

Step 8. Using once more the fact that the LHS oscillates little on each $f^{-kn_1}D_{km}$ we obtain that for large N

$$\mathbb{E}_\ell\left(\exp\left(c\left(\left(<\xi, \sum_{j=0}^{N-1}\left[A(x_j, z_j, \varepsilon)-\bar{A}(z_j)\right]>\right)-\frac{3\delta N}{4}\right)\right)\right)\leq \mathrm{Const}\theta^N$$

which implies (18) via the Markov inequality.

Hence the main contribution to (16) comes from the first term. The result follows. $\square$

4. Smoothness of u-Gibbs measures. Local results.

Before giving the proofs of other results we need to extend Theorem 4 to apply to F_ε. Let us recall the idea of the proof of Theorem 4. Take a small δ. Let

$$\bar{N} = \varepsilon^{-\delta}. \tag{20}$$

We show that for any standard pair $\ell = (D, \rho)$ for $n \sim \bar{N}$, $f_\varepsilon^n D$ is close to $f^n D$. Denote $X = \frac{d}{d\varepsilon}|_{\varepsilon=0}(f_\varepsilon \circ f^{-1})$. Choose a smooth distribution E_{as} close to E_s. We show that for $0 \le n \le \bar{N}$ there is $y_n(x)$ such that

$$f_\varepsilon^n y_n = \exp_{f^n x}(Z_n(x)),$$

where $Z_n \in E_c \oplus E_{as}$. Then

$$Z_{n+1} = \pi_{ac}^n (df(Z_n) + \varepsilon X) + o(\varepsilon^{1+\hat{\alpha}}) \tag{21}$$

for some $\hat{\alpha} > 0$ where π_{ac}^n is the projection to $E_c \oplus E_{as}$ along $T(f^n D)$. Also

$$\exp_{f^n x}^{-1}(f_\varepsilon^{n+1} y_{n+1}) - \exp_{f^n x}^{-1}(f_\varepsilon^{n+1} y_n)$$

equals up to higher order terms to

$$-\pi_{T(f^n D)}^n (df(Z_n) + \varepsilon X)$$

(here $\pi_{T(f^n D)}^n = 1 - \pi_{ac}^n$.) Next, $T(f^n D)$ is exponentially close to E_u. This allows us to obtain an asymptotic expansion for $Z_{\bar{N}}$. To describe it we need some notation. Let $1 = \pi_u + \pi_c + \pi_{as}$ be the decomposition corresponding to the splitting

$$TM = E_u \oplus E_c \oplus E_{as}. \tag{22}$$

Let $\Gamma_* = \pi_* df$ and $\Gamma_*^j(x) = \Gamma_*(f^{j-1}x) \dots \Gamma_*(fx)\Gamma_*(x)$. Let $X^* = \pi_* X$. Set

$$V = \sum_{j=0}^{\infty} (\Gamma_{as}^j X^{as})(f^{-j} x),$$

Let $\{e_l\}$ is a standard basis in $\mathbb{R}^d$. Consider vectorfields $e_l(x) = \frac{\partial g_t}{\partial t}(e_l)$. Define functions $a_l(x)$ by

$$X^c + \Gamma_c(V) = \sum_l a_l e_l.$$

Finally we need the notion of canonical divergence on the unstable leaves ([38]). If S is a subset of an unstable leaf define the density ρ_S on S by the conditions

$$\frac{\rho_S(y_1)}{\rho_S(y_2)} = \prod_{j=0}^{\infty} \frac{\det(df^{-1}|E_u)(f^{-j} y_1)}{\det(df^{-1}|E_u)(f^{-j} y_2)} \text{ and } \int_S \rho_S(y) dy = 1.$$

Then the volume form $\Omega_S = \rho_S(y)dy$ is defined up to the multiplicative constant (that is $\Omega_{S_1} = c_{S_1,S_2}\Omega_{S_2}$) so its divergence $\mathrm{div}_u^{can}(Y) = \frac{L_Y \Omega_S}{\Omega_S}$ where L denotes the Lie derivative does not depend on S. (Geometrically ρ_S is a conditional density of our u-Gibbs state ν on S.)

With this notation we have

$$\frac{Z_{\bar{N}}(x)}{\varepsilon} \approx \sum_{j=0}^{\bar{N}-1} (\Gamma_{as}^j X_{as})(f^{-j}x) + \sum_l \sum_{j=0}^{\bar{N}-1} a_l(f^{-j}x)e_l. \tag{23}$$

Here $\approx$ means that this equalities have to be understood in a weak sense. That is we say that

$$\sum_{j=0}^{\bar{N}-1} b_{j,\varepsilon}(f^{-j}x) \approx \sum_{j=0}^{\bar{N}-1} b_j(f^{-j}x)$$

if given δ_0, m we can find ε_0, k such that for $\varepsilon \le \varepsilon_0$ for all $B \in C_k^{\bar{\alpha}}(M)$ such that $\nu(B) = 0$ we have

$$\left| \int_D \rho(x) \left[b_{j,\varepsilon}(f^{\bar{N}-j}x) - b_j(f^{\bar{N}-j}x) \right] B(f^{\bar{N}}x)dx \right| \le \frac{\delta_0}{j^m}. \tag{24}$$

uniformly for all ρ such that $||\rho||_{C^\alpha(M)} \le 1$. (In practise to verify (24) one shows that both

$$\int_D \rho(x) b_{j,\varepsilon}(f^{\bar{N}-j}x)B(f^{\bar{N}}x)dx \text{ and } \int_D \rho(x) b_j(f^{\bar{N}-j}x)B(f^{\bar{N}}x)dx \text{ are } O(j^{-(m+1)})$$

and then shows that (24) converges to 0 for fixed j.) Similarly (23) implies that

$$\frac{1}{\varepsilon} \sum_{j=0}^{\bar{N}-1} \left(\frac{dy_{j+1}}{dy_j} - 1 \right) \approx -\sum_j \mathrm{div}_u^{can} \left[X^u + \Gamma_u(V) \right] (f^{\bar{N}-j}x) \tag{25}$$

(observe that Γ_u is 0 on E_c).

Now take $B \in C_k^{\bar{\alpha}}(M)$ for sufficiently large k. We want to compare $\nu_\varepsilon(B)$ with $\nu(B)$. By Theorem 1(c) for this it is sufficient to control $E_l(\cdot \circ f_\varepsilon^n)$ for all sufficiently large n and (7) shows that it is enough to show that for all $\ell \in E_1$

$$\mathbb{E}_\ell(B(f_\varepsilon^N)) = \nu(B) + \varepsilon\omega(B, X) + o(\varepsilon). \tag{26}$$

Since $\nu(1) = \mathbb{E}_\ell(1)$ we can assume without loss of generality that $\nu(B) = 0$. By our choice of $\bar{N}$

$$\mathbb{E}_\ell(B(F^{\bar{N}}x)) - \nu(B) = o(\varepsilon) \tag{27}$$

if B is smooth enough. Thus the main contribution difference between $\mathbb{E}_\ell(B(f_\varepsilon^{\bar{N}}x))$ and $\nu(B)$ comes from two sources.

(1) The difference

$$\mathbb{E}_l(B(f^{\bar{N}}x)) - \mathbb{E}_\ell(B(\exp_{f^{\bar{N}}x} Z_{\bar{N}}(x))). \tag{28}$$

According to (23) we can express this contribution as follows.

Split $Z_n = Z_n^{as} + Z_n^c$. Then (28) contributes $\omega^{as} + \omega^c$, where

$$\omega^{as}(B) = \nu(\partial_V B)$$

$$\omega^c(B) = \sum_l \sum_{j=0}^{\infty} \nu(a_l(f^{-j}x)(\partial_{e_l} B)(x)).$$

(The splitting (22) appears in the above formulas since $T(f^n D)$ approaches E_u exponentially fast.)

(2) The difference

$$\mathbb{E}_\ell\left(B(f^{\bar{N}}x)\left(\frac{dy_{\bar{N}}}{dx} - 1\right)\right).$$

Writing

$$\frac{dy_{\bar{N}}}{dx} = \frac{dy_{\bar{N}}}{dy_0} = \prod_{j=0}^{\bar{N}-1} \frac{dy_{j+1}}{dy_j} \sim 1 + \sum_{j=0}^{\bar{N}-1}\left(\frac{dy_{j+1}}{dy_j} - 1\right)$$

and using (25) we get the contribution of this part as

$$\omega^u(B) = -\sum_{j=0}^{\infty} \nu\left([\mathrm{div}_u(X^u + \Gamma_u V)](f^{-j}x)B(x)\right).$$

Thus we get

$$\mathbb{E}_\ell(B \circ f_\varepsilon^n) = \varepsilon\omega(B, X) + o(\varepsilon),$$

where

$$\omega = \omega^{as} + \omega^c + \omega^u. \tag{29}$$

Combining this with (26) (recall that $\nu(B) = 0$) we get Theorem 4.

The proof of Theorem 4 shows that it is useful to control $\mathbb{E}_\ell(B \circ f_\varepsilon^n)$. Our next goal is to do the same for F_ε. Recall that by definition the standard pair $\ell = (D, \rho)$ for F_ε satisfies $\angle(TD, TM) \le C_4\varepsilon$. Fix $\ell = (D, \rho)$ as above and let (x^*, z^*) be a point in D. Take $B \in C_k^{\bar{\alpha}}(M)$ for k large enough (depending on δ in (20)).

Proposition 1. *Let f_z be FAE such that f_z are uniformly rapidly mixing.*

(a) $\mathbb{E}_\ell(B \circ F_\varepsilon^{\bar{N}}) = \bar{B}(z^*) + o(\varepsilon\bar{N}^2)$.

Suppose in addition that $\bar{B} \equiv 0$ then *(b) For all* $n > 0$

$$|\mathbb{E}_\ell(B \circ F_\varepsilon^n)| \leq \text{Const}\left(\frac{1}{n^4} + \varepsilon \bar{N}^2\right) = \text{Const}\left(\frac{1}{n^4} + \varepsilon^{1-2\delta}\right).$$

(c) If f_z *have uniform stretched exponential decay of correlations and* $\bar{A} \equiv 0$ *then*

$$\mathbb{E}_\ell(B \circ F_\varepsilon^{\bar{N}}) = \varepsilon a(z^*, B) + o(\varepsilon)$$

where

$$a(z, B) = \sum_{n=1}^{\infty} \nu_{z^*}\left(\frac{\partial B}{\partial z}(f_{z^*}^n x, z^*) A(x, z^*)\right) + \omega(B, X_{z^*}, f_{z^*}) + \sum_{n=0}^{\infty} \omega(B, Y_{z^*,n}, f_{z^*})$$

(d) For all $n > 0$

$$|\mathbb{E}_\ell(B \circ F_\varepsilon^n)| \leq \text{Const}\left(\frac{1}{n^4} + \varepsilon\right).$$

(e) for all $n \geq \bar{N}$

$$\mathbb{E}(B \circ F_\varepsilon^n) = \varepsilon \mathbb{E}_\ell(a(z_{n-\bar{N}}, B)) + o(\varepsilon).$$

(f) In particular if $\bar{N} \leq n \ll 1/\varepsilon$ *then*

$$\mathbb{E}(B \circ F_\varepsilon^n) = \varepsilon a(z^*, B) + o(\varepsilon).$$

5. Proof of Proposition 1.

Proof. We follow the proof of Theorem 4 making heavy use of the mixing properties of F described in Appendix A. Let $\ell = (D, \rho)$ be F_ε-standard pair. Let w be a point in D. By induction we shall find for $0 \leq n \leq \bar{N}$ a point $v_n \in D$ such that $F_\varepsilon^n v_n = \exp_{f^n w}(\mathcal{Z}_n(w))$, where $\mathcal{Z}_n = (Z_n, Q_n)$ with $Z_n \in E_{as} \oplus E_c(f_z)$, $Q_n \in \mathbb{R}^s$. Then (21) is replaced by

$$Z_{n+1} = \pi_{as}^n\left(df(Z_n) + \varepsilon X + \frac{\partial f}{\partial z}(Q_n)\right) + o(\varepsilon^{1+\hat{\alpha}}), \tag{30}$$

$$Q_{n+1} = Q_n + A(F^n w) + o(\varepsilon^{1+\hat{\alpha}}). \tag{31}$$

The system (30)–(31) is upper triangular so it can be solved explicitely. However, we will not present the solution in a single formula since it would be too cumbersome. Rather we divide it into several pieces and analyze each piece separetely.

(25) is replaced by

$$\exp^{-1}_{F^{n+1}w}(F^{n+1} v_{n+1}) - \exp^{-1}_{F^{n+1}w}(F^{n+1} v_n) \sim -\pi^n_{T(F^n D)}\left(\frac{\partial f}{\partial x}(Z_n) + \varepsilon X + \frac{\partial f}{\partial z}(Q_n)\right).$$

Now we estimate $\mathbb{E}_\ell(B\circ F_\varepsilon^{\bar N})-\mathbb{E}_\ell(B\circ F^{\bar N})$ by splitting it into two parts.

(1) The contribution of

$$\mathbb{E}_\ell(B(\exp_{F^{\bar N}w}(\mathcal{Z}_{\bar N}(w))))-\mathbb{E}_\ell(B(F^{\bar N}w))=O(\varepsilon\bar N^2).$$

Indeed it is easy to see that $||Q_n||_{C^0}\leq \text{Const}\varepsilon n$ and then $||Z_n||_{C^0}\leq \text{Const}\varepsilon n^2$ by induction.

(2) $\frac{dv_{n+1}}{dv_n}-1=O(\varepsilon)$ which again can be proven by induction following the line of [19], Proposition 2.4(f). (Observe that even though $||Q_n||_{C^0}$ grows linearly with n its derivatives in the directions of $T(F^nD)$ are uniformly bounded since the derivatives of $A\circ F^{-j}$ decay exponentially in j because F^{-j} restricted to $T(F^nD)$ is strongly contracting.) Since

$$\frac{dv_{\bar N}}{dw}=\prod_{n=0}^{\bar N-1}\frac{dv_{n+1}}{dv_n}$$

we have

$$\mathbb{E}\left(B(F^{\bar N}w)\left(\frac{dv_{\bar N}}{dw}-1\right)\right)=O(\varepsilon\bar N).$$

This proves part (a).

To get (b) consider two cases.

(I) $n\leq 2\bar N$. We can argue as before except that for small n we can not neglect the term

$$\mathbb{E}_\ell(B\circ F^n)=\mathbb{E}_\ell(B\circ F^n)-\bar B=O(n^{-m}||B||_{C^\alpha_k(M)})$$

(cf (27)) which accounts for $O(n^{-4})$ term in (b).

(II) $n>2\bar N$. Then we split

$$\mathbb{E}_\ell(B\circ F_\varepsilon^n)=\sum_j\mathbb{E}_{\ell_j}(B\circ F^{\bar N})+O(\theta^{n-\bar N})$$

and apply part (a) to each ℓ_j.

To prove part (c) we need to analyze $\mathcal{Z}_n$ more carefully. Again we deal with the two parts of the difference

$$\mathbb{E}_\ell(B\circ F_\varepsilon^{\bar N})-\mathbb{E}_\ell(B\circ F^{\bar N})$$

separately.

In the formulas below we use summations over various indeces. l will always change from 1 to d and β will always change from 1 to s whereas other variables (j,k,m etc) will be non-negative and possibly satisfy an extra requirement described on the case-by-case basis. We shall also use the convention of Appendix A that $\bar c_2$ is a positive constant whose value can change from entry-to-entry.

(1) $\left[\mathbb{E}_\ell(B(\exp_{F^{\bar N}w}\mathcal{Z}_{\bar N}))-\mathbb{E}(B(F^{\bar N}w))\right]/\varepsilon$. We can split this into two parts $\mathcal{I}_*$ and $\mathcal{I}_{**}$ where $\mathcal{I}_*$ consists of terms not containing A and $\mathcal{I}_{**}$ consists of terms containing A. $\mathcal{I}_*$ can be analyzed as before, thus

$$\mathcal{I}_* = [\omega^{as}(B, X_{z^*}, f_{z^*}) + \omega^c(B, X_{z^*}, f_{z^*})] + o(1).$$

To handle $\mathcal{I}_{**}$ we observe that there are four types of terms containing A. We consider them separately.

(I) Contribution of $Q_N = \sum_{j=0}^{\bar N-1} A(F^{N-j}w)$. Thus we need to analyze $\sum_{\beta j}\mathcal{I}^I_{\beta j}$ where

$$\mathcal{I}^I_{\beta j} = \mathbb{E}_\ell\left(\frac{\partial B}{\partial z_\beta}(F^{\bar N}w)A_\beta(F^{\bar N-j}w)\right).$$

Since $\bar A\equiv 0$ Lemma 4(d) tells us that

$$\mathcal{I}^I_{\beta j} = O(c_1\exp(-\bar c_2\min(j,\bar N-j)^{\bar\gamma})). \tag{32}$$

Hence $\sum_j \mathcal{I}_{\beta j}$ is uniformly bounded and to get the asymptotics of this series we can restrict our attention to the cases where the minimum in (32) is small.

(a) j is small. In this case by Lemma 4(b)

$$\mathcal{I}^I_j \sim \nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}(F^j w)A_\beta(w)\right).$$

(b) $\bar N-j$ is small. In this case Lemma 3 and the continuity of ν_z imply

$$\mathbb{E}_\ell\left(A_\beta(F^{\bar N-j}w)\frac{\partial B}{\partial z_\beta}(F^{\bar N}w)\right)$$

$$\sim \mathbb{E}_\ell\left(A_\beta(F^{\bar N-j}w)\left(\frac{\partial B}{\partial z_\beta}(F^{\bar N}w)-\nu_z(\frac{\partial B}{\partial z_\beta}(F^{\bar N}w))\right)\right)+\mathbb{E}_\ell(A_\beta(F^{\bar N-j}w))\nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}\right)$$

$$\sim \nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}\right)\mathbb{E}_\ell(A_\beta(F^{\bar N-j}w)).$$

Thus the total contribution of type I terms is

$$\sum_{n=1}^{\infty}\nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}(F^n w)A_\beta(w)\right)+\nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}\right)\mathcal{A}_\beta \tag{33}$$

where

$$\mathcal{A}_\beta = \sum_n \mathbb{E}_\ell(A_\beta\circ F^n) \tag{34}$$

(Observe that both (33) and (34) converge exponentially fast because $\bar A\equiv 0$.)

We now pass to terms containing $\frac{\partial f}{\partial z}$. Let

$$\frac{\partial f}{\partial z_\beta} = R_u^\beta + R_{as}^\beta + \sum_l a_{\beta l} e_l$$

where $R_u^\beta \in E_u$, $R_{as}^\beta \in E_{as}$. There are three kinds of terms

(*II*) Call

$$\mathcal{I}^{II}_{\beta ljk} = \mathbb{E}_\ell((\partial_{e_l} B)(F^{\bar N} w) a_{\beta l}(F^{\bar N-j}(w)) A_\beta(F^{\bar N-j-k-1}w)) \quad (\text{Here } j+k \leq \bar N - 1.)$$

Observe that $\bar A \equiv 0$ and $\nu_z(\partial_{e_l} B) \equiv 0$ since e_l preserves ν_z. Hence by Lemma 5(a), (c) and (f) $\sum_{jk} \mathcal{I}^{II}_{jk}$ is uniformly bounded and the main contribution comes from the terms where j and either k or $\bar N - k$ are small.

So the total contribution of type *II* terms is

$$\sum_{\beta l} \sum_{jk} \nu_{z^*} \left(A_\beta(w) a_{\beta l}(F^{k+1} w)(\partial_{e_l} B)(F^{j+k+1} w)\right) \tag{35}$$
$$+ \sum_{\beta l} \mathcal{A}_\beta \sum_j \nu_{z^*} \left(a_{\beta l}(w)(\partial_{e_l} B)(F^j w)\right).$$

Let us remark that similarly to the proof of Theorem 4 the terms containing R_u^β and R_{as}^β provide only weak approximation to the actual Z_n/ε since we are replacing the splitting

$$TM = E_{as} \oplus E_c \oplus (T(f^n D)) \text{ by}$$

$$TM = E_{as} \oplus E_c \oplus E_u \tag{36}$$

but we ignore this isssue here and below since the argument presented above allows us to neglect the terms where n is small and the approximation (36) works poorely. On the other hand we have strong approximation for the terms containing A_β (like $\mathcal{A}_\beta$) since we always split

$$T(M\mathbb{R}^s) = TM \oplus \mathbb{R}^s.$$

In particular the foregoing analysis shows that the first series in (35) converges as

$$\sum_j j^{\gamma^*} e^{-c_2 j^\gamma}$$

(the second series converges as for type I terms).

Now we pass to the terms containing Γ_{as}.

(*III*) Let

$$\mathcal{I}^{III}_{\beta jk} = \mathbb{E}_\ell \left(A_\beta(F^{\bar N-j-k-1} w)(\partial_{\Gamma^j_{as} R^\beta_{as}} B)(F^{\bar N} w)\right) \quad (j+k \leq \bar N - 1)$$

The analysis of this term is similar to $\mathcal{I}_{\beta ljk}$ but it is simpler since for any admissible D $||(\Gamma^j_{as} R^\beta_{as}|| \leq \text{Const}\theta^j$. The contribution of type *III* terms is

$$\sum_\beta \sum_{jk} \nu_{z^*} \left(A_\beta(w)(\partial_{\Gamma^j_{as} R^\beta_{as}} B)(F^j w) \right) + \sum_\beta \mathcal{A}_\beta \sum_j \nu_{z^*} \left(\partial_{\Gamma^j_{as} R^\beta_{as}} B \right).$$

(*IV*) Let

$$\mathcal{I}^{IV}_{\beta ljkm} = \mathbb{E}_\ell \left((\partial_{e_l} B)(F^{\bar{N}} w) a_{\beta lk}(F^{\bar{N}-j} w) A_\beta(F^{j-k-m-1} w) \right) \quad (j+k+m \leq \bar{N}-2)$$

where $a_{\beta lk}$ are defined as follows

$$\Gamma_c \Gamma^k_{as} R^\beta_{as} = \sum_l a_{\beta lk} e_l.$$

Lemma 5(a) gives

$$(a) \quad \left| \mathcal{I}^{IV}_{\beta ljkm} \right| \leq \text{Const} e^{-\bar{c}_2 j^\gamma} \theta^k.$$

On the other hand if $m > (j+k)^{\gamma^*}$ then by Lemma 5(c)

$$\mathcal{I}^{IV}_{\beta ljkm} = \mathbb{E}_\ell \left(A \circ F^{\bar{N}-j-k-m-1} \right) \nu_{z^*} \left(a_{\beta lk}(w) (\partial_{e_l} B)(F^j w) \right) + O(e^{-\bar{c}_2 m^\gamma}).$$

Thus the contribution of type *IV* terms is

$$\sum_{\beta l} \sum_{jkm} \nu_{z^*} \left(A_\beta(w) a_{\beta lk}(F^{k+m+1} w)(\partial_{e_l} B)(F^{j+k+m+1} w) \right)$$
$$+ \sum_{\beta l} \mathcal{A}_\beta \sum_{jk} \nu_{z^*} \left(a_{\beta lk}(w)(\partial_{e_l} B)(F^j w) \right).$$

Now we consider terms coming from

$$\mathbb{E}_\ell \left(B(F^{\bar{N}}_\varepsilon w) \left[\frac{dv_{\bar{N}}}{dw} - 1 \right] \right) / \varepsilon.$$

Again we split it as $\mathcal{J}_* + \mathcal{J}_{**}$ where $\mathcal{J}_*$ contains A and $\mathcal{J}_{**}$ does not contain A. $\mathcal{J}_*$ can be treated as in Theorem 4. Thus

$$\mathcal{J}_* = \omega^u(B, X_{z^*}, f_{z^*}) + o(1).$$

Before handling $\mathcal{J}_{**}$ let us observe that we have an *a priori* estimate

$$\left| \mathbb{E}_\ell \left(\left[\frac{dv_{j+1}}{dv_j} \right] B(F^{\bar{N}} w) \right) \right| \leq \text{Const} e^{-\bar{c}_2 j^\gamma}$$

coming from the fact that C^1-norm of Z_{N-j} is uniformly bounded and B has zero mean.

Now we split $\mathcal{J}_{**}$ into two parts.

(I) Terms containing R^β_u. Those are

$$\mathcal{J}^I_{\beta jk} = -\mathbb{E}_\ell \left(\text{div}^{can}_u \left[A_\beta \circ F^{-(k+1)} R_u \right] (F^{\bar{N}-j} w) B(F^{\bar{N}} w) \right) \quad (j+k \leq \bar{N}-1).$$

Next
(37)
$\text{div}_u^{can}[A_\beta \circ F^{-(k+1)} R_u] = (A_\beta \circ F^{-(k+1)})\text{div}_u^{can}(R_u) + \partial_{R_u}(A \circ F^{-(k+1)})$.

The second term here is $O(\theta^k)$ since F^{-k} is strong contraction on $F^{N-j}D$. Now $\mathcal{J}^I_{\beta jk}$ can be treated similarly to $\mathcal{I}$-terms. So the total contribution of type I $\mathcal{J}$-terms is

$$-\sum_\beta \sum_{jk} \nu_{z^*}\left(\text{div}_u^{can}[(A_\beta \circ F^{-(k+1)})R_u](w)B(F^j w)\right) - \sum_\beta \mathcal{A}_\beta \sum_j \nu_{z^*}\left(\text{div}_u^{can}(R_u)(w)B(F^j w)\right).$$

(*II*) The terms containing Γ_{as}. They are

$$\mathcal{J}^{II}_{\beta jkm} = -\mathbb{E}_\ell\left(B(F^{\bar N} w)\text{div}_u^{can}\left[(A \circ F^{-(m+k+2)})\Gamma_u \Gamma^k_{as} R^\beta_{as}\right](F^{\bar N - j} w)\right) \quad (j+k+m \leq \bar N - 2).$$

To estimate $\mathcal{J}^{II}_{\beta jkm}$ we use the fact ([19], Lemma B.1) that C^1-norm of $(\Gamma^k_{as} R^\beta_u) \circ F^{-k}$ is $O(\theta^k)$. Now the analysis is similar to $\mathcal{I}^{IV}_{\beta lmjk}$ taking into account the remark after (37). The resulting contribution of type *II* $\mathcal{J}$-terms is

$$-\sum_\beta \sum_{mjk} \nu_{z^*}\left(\text{div}_u^{can}\left[(A \circ F^{-(m+k+2)})\Gamma_u \Gamma^k_{as} R^\beta_u\right](w)B(F^j w)\right)$$
$$-\sum_\beta \mathcal{A}_\beta \sum_{jk} \nu_{z^*}\left(\text{div}_u^{can}\left[\Gamma_u \Gamma^k_{as} R^\beta_u\right] B(F^j w)\right).$$

It remains to sum up the extra terms appearing here. The terms containing $\mathcal{A}_\beta$ can be combined to yield

$$\mathcal{A}_\beta\left[\nu_{z^*}\left(\frac{\partial B}{\partial z_\beta}\right) + \omega\left(B, \frac{\partial f}{\partial z}, f_{z^*}\right)\right].$$

The expression in brackets is

$$\frac{\partial}{\partial z_\beta}\left(\int B(x,z)d\nu_z(x)\right) = 0$$

since $\bar B = 0$. On the other hand the terms not containing $\mathcal{A}_\beta$ add up to $\sum_{n=0}^\infty \omega(B, Y_{z^*,n}, f_{z^*})$ (this series converges uniformly due to the estimates of terms $\mathcal{I}^I$–$\mathcal{I}^{IV}$, $\mathcal{J}^I$, $\mathcal{J}^{II}$). This proves (c).

(d) follows from (c) the same way (b) follows from (a).

To get (e) let $N_1 = \bar N/2$ and observe that the estimates of part (c) remain valid for $\bar N$ replaced by N_1 (with slightly worse constants). Now we split

$$\mathbb{E}_\ell(B \circ F^n_\varepsilon) = \sum_r c_r \mathbb{E}_{\ell_r}(B \circ F^{N_1}) + O(\theta^{n-N_1})$$

and apply part (c) to each $\mathbb{E}_{\ell_r}$. (Observe that $n - N_1 \geq N_1$ so $O(\theta^{N_1}) = O(\varepsilon^2)$.)

To get part (f) we observe that due to *a priori* bound

$$|z_n - z^*| \leq \varepsilon n ||A||_{C^0}$$

we have $z_{n-\bar{N}} + o(1)$ for $n \ll 1/\varepsilon$ so (f) follows from (e). □

6. Short time averaging.

Here we present the proof of Theorem 6. Recall that to prove weak convergence on $[0, \infty)$ it is enough to establish weak convergence on $[0, T]$ for all T. From now on we fix some $T > 0$. The proof of Theorem 6 depends on two lemmas whose proofs are given after the proof of the Theorem.

Proof of Theorem 6. (We suppress the dependence of A on ε in order to simplify the notation.) By a standard approximation argument it suffices to prove the result for $\ell \in E_1(z^*, \varepsilon)$.

Decompose

$$\Delta_{n+1} - \Delta_n = \varepsilon[A(x_n, z_n) - \bar{A}(q_n)] = \varepsilon[A(x_n, z_n) - \bar{A}(z_n)] + \varepsilon[\bar{A}(z_n) - \bar{A}(q_n)]. \tag{38}$$

Using the Hadamard Lemma we represent the second term in the form

$$\bar{A}(z_n) - \bar{A}(q_n) = D\bar{A}(q_n)\Delta_n + \zeta(q_n, \Delta_n)\Delta_n \tag{39}$$

where ζ is a bounded smooth function of its arguments satisfying $\zeta(x_n, 0) = 0$. Let $W_n = \sum_{j=0}^{n-1} \left[A(x_n, z_n) - \bar{A}(z_n)\right]$, then by Gronwell argument

$$|\Delta_n| \leq \text{Const}\varepsilon \max_{j \leq n} |W_j|. \tag{40}$$

Lemma 1. *As $\varepsilon \to 0$ $\sqrt{\varepsilon} W^{\varepsilon}_{t/\varepsilon}$ converges to $\mathbf{B}$, the process defined by (10).*

Next we estimate the second term in (39) as follows

$$\left| \sum_{j=1}^{n} \zeta(q_j, \Delta_j)\Delta_j \right| \leq \text{Const}(n\varepsilon) \max_{j \leq n} \varepsilon W_j^2 \leq T \max_{j \leq n} (\sqrt{\varepsilon}|W_j|)^2$$

As $\varepsilon \to 0$ the second term converges to $\max_{s \leq T} |\mathbf{B}(s)|^2$. Thus as $\varepsilon \to 0$ $\frac{1}{\sqrt{\varepsilon}}(\varepsilon \sum_j \zeta(q_j, \Delta_j)\Delta_j)$ converges to 0. Hence dividing (38) by $\sqrt{\varepsilon}$ and taking $\varepsilon \to 0$ we obtain

$$\Delta(t) = \int_0^t D\bar{A}(q(s))\Delta(s)ds + \mathbf{B}(t).$$

This proves (a).

To prove (b) we use (40). Arguing as in Theorem 5 we can reduce part (b) to showing that if $\xi \in \mathbb{R}^s - 0$ then for each $\ell = (D, \rho) \in E_1(z^*, \varepsilon)$ we have

$$\ell\left(< \xi, \sqrt{\varepsilon} W_{[t/\varepsilon]} > \geq R\right) \leq c_1 e^{-c_2 R^2}. \tag{41}$$

Moreover we can assume that D belongs to a Markov family $\mathcal{P}$. To establish (41) we use the following bound.

Lemma 2. *If $\ell = (D, \rho)$ is as above then there are constants δ and K such that if $K < |p| \leq \delta/\sqrt{\varepsilon}$ then*

$$\mathbb{E}_\ell(e^{p < \xi, \sqrt{\varepsilon} W_{[t/\varepsilon]} >}) \leq c_1 e^{c_2 p^2}.$$

Lemma 2 and Markov inequality imply that

$$\ell(< \xi, \sqrt{\varepsilon} W_{[t/\varepsilon]} > \geq R) \leq e^{c_2 p^2 - Rp}.$$

Taking $p = R/(2c_2)$ we obtain (41) for R such that $K \leq R/(2c_2) \leq \delta/\sqrt{\varepsilon}$, that is $\hat{K} \leq R \leq \frac{\hat{\delta}}{\sqrt{\varepsilon}}$. Next the inequality

$$\ell\left(< \xi, \sqrt{\varepsilon} W_{[t/\varepsilon]} > \geq \frac{1}{\sqrt{\varepsilon}}\right) \leq \ell\left(< \xi, \sqrt{\varepsilon} W_{[t/\varepsilon]} > \geq \frac{\hat{\delta}}{\sqrt{\varepsilon}}\right)$$

extends (41) to $R \leq 1/(\sqrt{\varepsilon})$ at the expense of increasing c_1 and c_2. Finally increasing c_1 once more we can assume that $c_1 e^{-c_2 \hat{K}^2} \geq 1$ which makes (41) trivial for $R \leq \hat{K}$. □

Proof of Lemma 1. The proof consists of several steps.

Step 1. We shall use Proposition 1 with $\delta = 0.01$. Call $\bar{n} = \varepsilon^{-(1-2\delta)}$, $B = A - \bar{A}$ and consider

$$P = \sum_{j=0}^{\bar{n}-1} B(F_\varepsilon^j(x, z)).$$

Then by Proposition 1

$$|\mathbb{E}_\ell(P)| \leq \text{Const.} \tag{42}$$

We also claim that

$$\mathbb{E}_\ell(P_\alpha P_\beta) = \bar{n}(\sigma^2_{\alpha\beta}(A(\cdot, z^*)) + o(1)) \text{ and} \tag{43}$$

$$\mathbb{E}_\ell(|P|^4) = O(\bar{n}^2). \tag{44}$$

To prove (43) decompose

$$\mathbb{E}_\ell(P_\alpha P_\beta) = \sum_{jk} \mathbb{E}_\ell(B_\alpha(F_\varepsilon^j(x, z)) B_\beta(F_\varepsilon^k(x, z))).$$

We claim that

$$\mathbb{E}_\ell(B_\alpha(F_\varepsilon^j(x,z))B_\beta(F_\varepsilon^k(x,z))) \leq \text{Const}\left[\frac{1}{(k-j)^4} + \varepsilon^{1-2\delta}\right]. \tag{45}$$

Indeed, assume, e.g. that $k > j$ then letting $n = j + k$, $m = k - j$ we can split

$$\mathbb{E}_\ell(B_\alpha(F_\varepsilon^j(x,z))B_\beta(F_\varepsilon^k(x,z))) = \sum_r c_r \mathbb{E}_{\ell_r}(B_\alpha(F_\varepsilon^{-m/2}(x,z))B_\beta(F_\varepsilon^{m/2}(x,z))) + O(\theta^n).$$

Let $\ell_r = (D_r, \rho_r)$. On each D_r we can approximate $(B_\alpha(F_\varepsilon^{-m/2}(x,z))$ by a constant σ_r with $O(\theta^m)$-error. Thus

$$\mathbb{E}_\ell(B_\alpha(F_\varepsilon^j(x,z))B_\beta(F_\varepsilon^k(x,z))) = \sum_r c_r\sigma_r \mathbb{E}_{\ell_r}(B_\beta \circ F^{m/2}) + O(\theta^m).$$

By Proposition 1(b)

$$\mathbb{E}_{\ell_r}(B_\beta \circ F^{m/2}) \leq \text{Const}\left[\frac{1}{m^4} + \varepsilon^{1-2\delta}\right]$$

proving (45).

(45) shows that the main contribution to $\mathbb{E}_\ell(P_\alpha P_\beta)$ comes from the terms where $k - j$ is small.

Let m be fixed. For $j \geq \bar{N}$ split

$$\mathbb{E}_\ell\left(B_\alpha(F_\varepsilon^j(x,z))B_\alpha(F_\varepsilon^{j+m}(x,z))\right) = \sum_r E_{\ell_r}\left(B(x,z)B(F_\varepsilon^m(x,z))\right) + O(\theta^{j-m}).$$

Applying Proposition 1(a) to each ℓ_r with function $C(x,z) = B_\alpha(x,z)B_\beta(F^m(x,z))$ (observe that for m fixed

$$B_\alpha(x,z)B_\beta(F_\varepsilon^m(x,z)) = B_\alpha(x,z)B_\beta(F^m(x,z)) + O(\varepsilon))$$

we get

$$\mathbb{E}_\ell\left(B_\alpha(F_\varepsilon^j(x,z))B_\beta(F_\varepsilon^{j+m}(x,z))\right) = \nu_{z^*}\left(B_\alpha(x,z)B_\beta(F^m(x,z))\right) + O(\varepsilon) + O(\theta^{j-m}).$$

Summation over m proves (43). (44) follows from (43) similarly to [18], Lemma 1(d).

Step 2. Fix $\xi \in \mathbb{R}^s$. Let $\phi(\xi) = \mathbb{E}_\ell(e^{i<\xi,\sqrt{\varepsilon}P>})$. From the Taylor series

$$e^{1<\xi,\sqrt{\varepsilon}P>} = 1 + i\sqrt{\varepsilon} < \xi, P > -\varepsilon\frac{<\xi, P>^2}{2} + O(\varepsilon^{3/2}|P|^{3/2})$$

and the fact that (44) and Holder inequality imply that $\mathbb{E}_\ell(|P|^3) \leq \text{Const}\bar{n}^{3/2}$ we get

$$\phi(\xi) = 1 - \varepsilon\bar{n} < \sigma^2(\bar{A}(\cdot, z^*))\xi, \xi > +o(\varepsilon\bar{n}).$$

Step 3. We now use a big block–small block approach common in the theory weakly dependent random variables (see [25]). Let

$$\bar{P}_{(k)} = \sum_{j=(k-1)(\bar{n}+\bar{N})+1}^{(k-1)(\bar{n}+\bar{N})+\bar{n}} B(\circ F_\varepsilon^j), \quad \bar{\bar{P}}_{(k)} = \sum_{j=(k-1)(\bar{n}+\bar{N})+\bar{n}+1}^{k(\bar{n}+\bar{N})} B(\circ F_\varepsilon^j).$$

Then

$$\sum_{k=1}^{t/(\varepsilon\bar{n})} \bar{\bar{P}}_{(k)} \leq \text{Const}\frac{t}{\varepsilon\bar{n}}\bar{N} \leq \text{Const}\varepsilon^{-4\delta}.$$

Hence the main contribution to W comes from big blocks $\bar{P}_{(k)}$.

Step 4. Let

$$\psi_k(\xi) = \mathbb{E}_\ell\left(\exp\left(i < \xi, \sqrt{\varepsilon}\sum_{j=1}^k \bar{P}_{(j)} >\right)\right).$$

We prove by induction that

$$\ln\psi_k(\xi) = -\left[\varepsilon\bar{n}\sum_{j=1}^k < \sigma^2(q_{(k-1)(\bar{n}+N)})\xi, \xi >\right] + o(\varepsilon\bar{n}k) \tag{46}$$

Indeed suppose that (46) holds for $k-1$. Denote $m = (k-1)(\bar{n}+\bar{N})+\bar{N}$. Split

$$\psi_k(\xi) = \sum_r c_r \mathbb{E}_{\ell_r}(\exp(i < \xi, \sqrt{\varepsilon}\left[\sum_{j=1}^{k-1}\bar{P}_{(j)}\right]\circ F_\varepsilon^{-m} >)\exp(i < \xi, \sqrt{\varepsilon}P >)) + O(\theta^m). \tag{47}$$

Let $\ell_r = (D_r, \rho_r)$. On each D_r we can approximate the first factor in (47) by a constant σ_r with $O(\theta^{\bar{N}})$-error. Applying the result of step 2 to each ℓ_r we get

$$\psi_k(\xi) = \left[\sum_r c_r\sigma_r\left(1 - \frac{\varepsilon\bar{n}}{2} < \xi, \sigma^2(A(\cdot, z^*_{(r)}))\xi >\right)\right] + o(\varepsilon\bar{n})$$

where $(x^*_{(r)}, z^*_{(r)})$ is a point in D_r. Recall that by Theorem 5 for most r's we have $z^*_{(r)} = q_{(k-1)(\bar{n}+N)} + o(1)$. Hence

$$\psi_k(\xi) = \left[\sum_r c_r\sigma_r\right]\left(1 - \frac{\varepsilon\bar{n}}{2} < \xi, \sigma^2(A(\cdot, q_{(k-1)(\bar{n}+N)}))\xi >\right) + o(\varepsilon\bar{n}).$$

By inductive assumption

$$\sum_r c_r\sigma_r = \left[\exp(-\frac{\bar{n}\varepsilon}{2}\sum_{j=1}^{(k-1)} < \sigma^2(A(\cdot, q_{j(\bar{n}+N)}))\xi, \xi >)\right] + o(\varepsilon\bar{n}(k-1)) + O(\theta^{\bar{N}}).$$

Hence (46) holds for k.

Step 5. Applying (46) with $k = t/(\varepsilon\bar{n})$ we see that $\sqrt{\varepsilon}W_{[\varepsilon t]}$ is asymptotically Gaussian with zero mean and variance

$$\int_0^t \sigma^2(A(\cdot, q(s)))ds.$$

Step 6. The fact that for each t_1, t_2, t_m the vector

$$\sqrt{\varepsilon}(W_{[\varepsilon t_1]}, W_{[\varepsilon t_2]}, \dots W_{[\varepsilon t_m]})$$

is asymptotically Gaussian is proven by induction on m using the argument of step 4. We leave this to the reader. □

Proof of Lemma 2. The proof is similar to the second part of the proof of Theorem 5 so it will only be sketched here. Fix a large τ and let $n_0 = \tau/(\varepsilon p)$. Observe that $\delta r \leq n_0 \leq \tau K/\sqrt{\varepsilon}$. Now similarly to (42), (43) one can show that

$$\left|\mathbb{E}_\ell(p < \xi, \sqrt{\varepsilon}W_{n_0} >)\right| \leq \text{Const}\sqrt{\varepsilon}p,$$

$$\mathbb{E}_\ell(p^2 < \xi, \sqrt{\varepsilon}W_{n_0} >^2) \leq \text{Const}n_0\varepsilon p^2 = \text{Const}\sqrt{\varepsilon}p.$$

Using Taylor series $e^{ct} = 1 + ct + O((ct)^2)$ (observe that $\sqrt{\varepsilon}W_{n_0} \leq \tau K$) we obtain

$$\mathbb{E}_\ell(e^{p<\xi,\sqrt{\varepsilon}W_{n_0}>}) \leq (1 + c\sqrt{\varepsilon}p).$$

Let $f^{n_0}D = \bigcup_r D_r$. Arguing as in the proof of Theorem 5 we obtain

$$\sum_r c_r \sup_{(x_{(r)}, z_{(r)}) \in f^{-n_0}D_r} e^{p<\xi,\sqrt{\varepsilon}W_{n_0}>(x_{(r)},z_{(r)})} \leq 1 + \bar{c}\sqrt{\varepsilon}p$$

where $c_r = \ell(f^{-n_0}D_r)$. Arguing as in the proof of Theorem 5 we can deduce from this that for all k

$$\mathbb{E}_\ell(e^{p<\xi,\sqrt{\varepsilon}W_{n_0}>}) \leq (1 + c\sqrt{\varepsilon}p)^k.$$

Taking $k = \frac{t}{\varepsilon}/n_0 = t\frac{p}{\tau\sqrt{\varepsilon}}$ we obtain Lemma 2. □

7. Long time averaging.

Proof of Theorem 7. Continuity statements of part (a) follow from the uniform convergence of the series for $a(z)$ and $\sigma^2(s)$ proven in Section 5. Differentiability of σ^2 is proven in Appendix B. We do not use this differentiability result here but it can be useful for the question of uniqueness of the corresponding diffusion process which we plan to discuss elsewhere. (Estimates of Appendix B are used in the Appendix A but just continuity of σ^2 would suffice for our purposes.)

The proof of parts (b) and (c) proceeds as in ([18], Section 15) replacing the estimates of ([18], Section 13) by Proposition 1. For completeness we sketch the proof here. We divide the proof into several steps.

Step 1. Again we can assume $\ell \in E_1(z^*, \varepsilon)$. We establish *a priori* bounds on the growth of z. Let

$$S_N = \sum_{j=0}^{N-1} A(F_\varepsilon(x,z)).$$

We claim that for all $\ell \in E(\varepsilon, z^*)$ for all $\frac{1}{\varepsilon} \leq N \leq \frac{T}{\varepsilon^2}$ we have

$$\mathbb{E}_\ell(S_N) = O(\varepsilon N) \tag{48}$$

$$\mathbb{E}_\ell(S_N^2) = O(N) \tag{49}$$

$$\mathbb{E}_\ell(S_N^4) = O(N^2). \tag{50}$$

(48) is immediate from Proposition 1(d). To prove (49) we want to show as we did in Section 6 that the main contribution to $\mathbb{E}_\ell(S_N^2)$ comes from near diagonal terms. However a naive application of (45) gives only the following insufficient estimate for the offdiagonal terms

$$o(N) + O(N^2\varepsilon). \tag{51}$$

(Observe here that we can replace the RHS of (45) by $O(\theta^{j+k} + \frac{1}{(k-j)^4} + \varepsilon)$ since we can use Proposition 1 (d) instead of Proposition 1(b) used to derive (45).) Therefore we shall use multi-scale analysis to improve (51). The point is that in (45) we bound $\mathbb{E}_\ell((A_\alpha \circ F_\varepsilon^j)(A_\beta \circ F_\varepsilon^k))$ by replacing $A_\alpha \circ F_\varepsilon^j$ by $|A_\alpha \circ F_\varepsilon^j|$ but now we shall explore the cancellations between different j's as well.

Let τ be a small number. The argument of Section 6 shows that for $N_0 := \tau/\varepsilon$ for all $\xi \in \mathbb{R}^s$ we have

$$\mathbb{E}_\ell(< S_N, \xi >^2) = N_0 < \sigma^2(z^*)\xi, \xi > + O(N_0^2).$$

In particular we have the required *a priori* estimate (49) for N_0. Next let $N_k = 2^k N_0$. We show by induction on k that for $N \leq N_k$

$$\left| \mathbb{E}_\ell(< \sum_{j=0}^{N_k-1} A(x_j, z_j, \varepsilon), \xi >^2) \right| \leq \mathbf{c}_k N_k$$

where $\mathbf{c}_k \geq 1$ will be uniformly bounded if $2^k \leq \frac{T}{\varepsilon\tau}$. Let us explain the induction step. Suppose that the result holds for $N \leq N_k$. Take $N_k < N \leq N_{k+1}$. Denote $S' = S_{N/2}$, $S'' = S_{N/2} \circ F_\varepsilon^{N/2}$. Then

$$\mathbb{E}_\ell(< S' + S'', \xi >^2) = \mathbb{E}_\ell(< S', \xi >^2) + \mathbb{E}_\ell(< S', \xi >^2) + 2\mathbb{E}_\ell(< S'\xi >< S'', \xi >)$$

$$= I + II + III.$$

By induction $|I| \leq \mathbf{c}_k N/2$. Using the splitting (7) we get $|II| \leq \mathbf{c}_k N/2 + O(\theta^N)$. To estimate the cross term we split

$$\mathbb{E}_\ell(< S'\xi >< S'', \xi >) = \sum_r c_r \mathbb{E}_{\ell_r}\left(< S_{N/2} \circ F^{-N/2}, \xi >< S_{N/2}, \xi >\right) + O(\theta^N).$$

Let $\ell_r = (D_r, \rho_r)$. On each D_r we can approximate $< S_{N/2} \circ F^{-N/2}, \xi >$ by a constant σ_r with $O(1)$-error. Hence

$$III = \sum_r c_r \sigma_r \mathbb{E}_\ell(S_{N/2}) + O(\sum_r c_r \mathbb{E}_{\ell_r}(|S_{N/2}|) = III_a + III_b$$

where

$$|III_a| \leq \sum_r c_r \sigma_r \mathrm{Const}(\varepsilon N + 1)$$

$$\leq \mathrm{Const}\left[\mathbb{E}_\ell(|S_{N/2}|) + 1\right](\varepsilon N + 1) \leq \mathrm{Const}\left[\sqrt{\mathbf{c}_k N/2} + 1\right](\varepsilon N + 1),$$

$$|III_b| \leq \mathrm{Const} \sum_r c_r \mathbb{E}_{\ell_r}(|S_{N/2}|) \leq \mathrm{Const}\sqrt{\mathbf{c}_k N/2}.$$

Next $\sqrt{\mathbf{c}_k} \leq \mathbf{c}_k$ since $\mathbf{c}_k \geq 1$. Thus

$$\mathbf{c}_{k+1} \leq \mathbf{c}_k \left(1 + \frac{D(\varepsilon N + 1)}{\sqrt{N}}\right) \tag{52}$$

for some constant D. To analyze (52) we consider two cases. Let k_0, k_1 be such that $N_{k_0} \leq 1/\varepsilon < N_{k_0+1}$, $N_{k_1} \leq T/\varepsilon^2 < N_{k_1+1}$. Then for a constant $\bar{D}$ we have

$$\mathbf{c}_{k_0} \leq \mathbf{c}_0(1 + \bar{D}\sqrt{\tau}\varepsilon)^{\log_2(1/\tau)+1} \leq 2\mathbf{c}_0.$$

Next for $k_0 \leq k \leq k_1 + 1$ we have

$$\mathbf{c}_k \leq \mathbf{c}_{k_0} \prod_{j=k_0}^{k} (1+\bar{D}\varepsilon\sqrt{N_j}) = \mathbf{c}_{k_0} \prod_{j=k_0}^{k} (1+\bar{D}2^{(j-k_1)/2}) \leq \mathbf{c}_{k_0} \prod_{m=k_1-k_0}^{\infty} (1+\bar{D}2^{-m/2})$$

This proves (49). The (50) follows from (49) by a similar inductive argument (cf. [18], Lemma 13).

Step 2. Using (50) and (7) we see that if $N_1, N_2 \leq \frac{\delta}{\varepsilon}$ and $|N_2 - N_1| \geq 1/\varepsilon$ then

$$\mathbb{E}_\ell(|S_{N_1} - S_{N_2}|^4) \leq \mathrm{Const}(N_2 - N_1)^2.$$

This implies that $\{z^\varepsilon_{[t/\varepsilon]}\}$ is a tight family (see e.g. [5]).

Step 3. We claim that for $N \leq \delta/\varepsilon^2$ we have

$$\mathbb{E}_\ell(< S_N, \xi >^2) = \sum_{j=0}^{N} \mathbb{E}_\ell(< \sigma^2(z_j)\xi, \xi >) + o_{\varepsilon\to 0, \delta\to 0}(N) \tag{53}$$

Indeed let $\bar{C}$ be such that $\mathbb{E}_\ell(S_N^2) \leq \bar{C}N$. We rerun the induction procedure using the inductive assumption

$$\mathbb{E}_\ell(< S_{N_k}, \xi >^2) = \sum_{j=0}^{N_k} \mathbb{E}_\ell(< \sigma^2(z_j)\xi, \xi >) + \Gamma_k,$$

where $|\Gamma_k| \leq \gamma_k N_k$. Then

$$\gamma_{k+1} \leq \gamma_k + \mathrm{Const}\bar{C}\frac{N_k\varepsilon + 1}{\sqrt{N_k}}$$

proving (53).

Step 4. For $N = \delta/\varepsilon^2$ we have

$$\varepsilon\mathbb{E}_\ell(S_N) = \delta a(z^*) + o(\delta), \tag{54}$$

$$\varepsilon(< S_N\xi, \xi >) = \delta < \sigma^2(z^*)\xi, \xi > + o(\delta). \tag{55}$$

Indeed, (54) follows from Proposition 1(c) and the tightness of $\{z_{[t/\varepsilon]}\}$ proven on step 2, (55) follows from step 3 and tightness. Also (50) and the Holder inequality imply

$$\mathbb{E}_\ell(\varepsilon^3|S_N|^3) = O(\delta^{3/2}). \tag{56}$$

Step 5. Let $\phi(z)$ be a function bounded together with its first three derivatives. Divide $[0, T]$ into segments of length δ/ε^2. Let $t_j = \frac{\delta j}{\varepsilon^2}$. Denote

$$\mathcal{L}(\phi)(z) = \sum_\alpha a_\alpha(z)\partial_{z_\alpha}\phi + \frac{1}{2}\sum_{\alpha\beta}\sigma^2_{\alpha\beta}(z)\partial^2_{z_\alpha z_\beta}\phi.$$

Then

$$\phi(z_{T/\varepsilon^2}) - \phi(z^*) = \sum_j \left[\phi(z_{t_{j+1}}) - \phi(z_{t_{j+1}})\right].$$

Decomposing

$$\mathbb{E}_\ell(B \circ F_\varepsilon^{t_j}) = \sum_r c_{rj}\mathbb{E}_{\ell_{rj}}(B)$$

using second order Taylor series for $\phi(z_{t_{j+1}}) - \phi(z_{t_{j+1}})$ and applying step 4 to each ℓ_{rj} we get

$$\mathbb{E}_\ell(\phi(z^\varepsilon_{T/\varepsilon^2} - \phi(z^*)) = \mathbb{E}_\ell\left(\sum_j \delta\mathcal{L}(\phi(z_{t_j}))\right) + O(\sqrt{\delta}).$$

Taking the limits $\varepsilon \to 0$, $\delta \to 0$ we obtain

$$\mathbb{E}(\phi(z_T) - \phi(z_0)) = \mathbb{E}\left(\int_0^t (\mathcal{L}\phi)(z_s)ds\right).$$

Step 6. The argument used in Lemma 1 shows that for all $s_1 \leq s_2 \leq \cdots \leq s_m \leq t_1 \leq t_2$ for all bounded Holder functions $\phi_j : \mathbb{R}^s \to \mathbb{R}$ we have

$$\mathbb{E}\left(\left[\phi(z_{t_2}) - \phi(z_{t_1}) - \int_{t_1}^{t_2} (\mathcal{L}\phi)(z_s)ds\right] \prod_j \phi_j(z_{s_j})\right) = 0.$$

This completes the proof of Theorem 7. □

8. Invariant measures.

Here we prove Theorems 8–10.

Proof of Theorem 10. Assume to the contrary that η is not invariant. Then by Hahn-Banach theorem there exists a function $B(z)$ such that $\eta(B) = 1$, while $\kappa(B) < 1/2$ for any invariant measure κ. Since $C^\infty(Z)$ is dense in $C(Z)$ we can assume that $B \in C^\infty(Z)$. On the other hand there exists $T > 0$ such that for each $z \in Z$

$$\mathbb{E}_z\left(\frac{1}{T}\int_0^T B(z(s))ds\right) < 2/3 \tag{57}$$

(since if for each T there were z_T failing (57) then a limit point of

$$\kappa_T(H) = \mathbb{E}_{z_T}\left(\frac{1}{T}\int_0^T B(z(s))ds\right)$$

would have $\kappa(B) \geq 2/3$.) Let $N_T = T\varepsilon^{-2}$. Theorem 7* implies that for any $\mu \in E(F_\varepsilon)$ we have

$$\mu\left(\frac{1}{N_T}\sum_{j=0}^{N_T-1} B(F_\varepsilon^j(x,z))\right) \leq 4/5$$

Since $\nu_\varepsilon(B) \in E(F_\varepsilon)$ by Theorem 1(a) we have

$$\nu_\varepsilon(B) = \nu_\varepsilon\left(\frac{1}{N_T}\sum_{j=0}^{N_T-1} B(F_\varepsilon^j(x,z))\right) \leq 4/5$$

if ε is small enough (at the last step we have used Theorem 1(a)). Thus $\eta(B) \leq 4/5$ a contradiction. □

The proof of Theorem 8 is similar to the proof of Theorem 10 and can be left to the reader.

Proof of Theorem 9. We split the proof into several steps.

Step 1. Let $\bar{z}$ be a periodic point which is not a sink and the matrix (15) is non-degenerate. Let U be a small neighborhood of the orbit of $\bar{z}$. We shall prove that there exist constants C, σ such that for each

$\delta > 0$ the following holds for sufficiently small ε. For any standard pair $\ell = (D, \rho)$ we have

$$\ell\left(z\left(\frac{C|\ln\varepsilon|}{\varepsilon}\right) \in U \text{ or } \mathrm{dist}\left(z^*, W^{cs}(\bar{z})\right) < \sigma\right) \leq \delta \tag{58}$$

where z^* is the value of z_j at the moment it exists U.

Step 2. We introduce the ordering $\gamma_i > \gamma_j$ if $W^u(\gamma_i) \bigcap W^s(\gamma_j) \neq \emptyset$. Theorem 5* implies that there exists T such that any point on the boundary of U which is σ away from $W^{cs}(\bar{z})$ enters a small neighborhood of a smaller periodic point after a time T/ε with probability $1 - O(\exp(-C_2\varepsilon^{-1}))$.

Step 3. Iterating steps 1 and 2 using decomposition (7) we conclude that there is a constant $\bar{C}$ such that for any standard pair ℓ if $(z(0), x(0))$ is distributed according to ℓ then $z\left(\frac{\bar{C}|\ln\varepsilon|}{\varepsilon}\right)$ enters a small neighborhood of a sink with probability larger than $1 - \delta$. Applying again Theorem 5 we see that upon entering a small neighborhood of a sink it can leave in $1/\varepsilon$ steps with probability $O(\exp(-C_2\varepsilon^{-1}))$ so with probability $1 - \delta$ it stays near the sink for exponentially long time.

Step 4. Let $\phi(z)$ be a smooth positive function equal to 1 near the orbit of $\bar{z}$ and 0 outside U. Steps 1-3 show that for each $\delta > 0$ there exists $\hat{C}$ such that for each standard pair $\ell \in E_1(F_\varepsilon)$

$$\mathbb{E}_\ell\left(\frac{1}{N_\varepsilon}\sum_{j=0}^{N_\varepsilon-1}\phi(z_j)\right) \leq \delta$$

where $N_\varepsilon = |\hat{C}\ln\varepsilon|/\varepsilon$. Thus

$$\mu\left(\frac{1}{N_\varepsilon}\sum_{j=0}^{N_\varepsilon-1}\phi(z_j)\right) \leq \delta$$

for any $\mu \in E(F_\varepsilon)$ and by Theorem 1(a)

$$\nu_\varepsilon(\phi) = \nu_\varepsilon\left(\frac{1}{N_\varepsilon}\sum_{j=0}^{N_\varepsilon-1}\phi(z_j)\right) \leq \delta$$

for any u-Gibbs state ν_ε. Since δ is arbitrary we have $\eta(\phi) = 0$.

Steps 2–4 are straightforward. Let us describe the proof of (58) in more detail. To fix our ideas we consider the case when $\bar{z}$ is periodic. The case when it is fixed is similar. In a small neighborhood of the orbit of $\bar{z}$ we can choose coordinates (p, q, t) such that t is the periodic coordinate, the orbit of $\bar{z}$ is $\{p = 0, q = 0\}$, the center stable manifold of $\bar{z}$ is $\{p = 0\}$ and the unstable direction along the orbit of z is

$\{q = 0, t = 0\}$. Take some $(\tilde{z}, \tilde{x}) \in D$. Choose a large constant R. There are two cases.

(I) $|p(\tilde{z})| < \text{Const} R\sqrt{\varepsilon}$. Then let $\hat{z}$ be the point such that $p(\hat{z}) = 0$, $q(\hat{z}) = q(\tilde{z})$, $t(\hat{z}) = t(\tilde{z})$. Theorem 6 implies that if we call

$$\tilde{\Delta}_t^\varepsilon = \frac{\tilde{z}^\varepsilon_{[t\varepsilon]} - q(\hat{z}, t)}{\sqrt{\varepsilon}}$$

then $\tilde{\Delta}_t^\varepsilon$ is close to the solution of

$$d\tilde{\Delta} = D\bar{A}(q(\hat{z}, t))\tilde{\Delta} + \tilde{\mathbf{B}}, \quad \tilde{\Delta}(0) = \frac{\tilde{z} - \hat{z}}{\sqrt{\varepsilon}}$$

where $\tilde{\mathbf{B}}$ is the process deffined by (10) with $\hat{z}$ as a guiding orbit. Therefore

$$\Delta(t) = \Gamma(t) \int_0^t \Gamma^{-1}(s) d\mathbf{B}(s)$$

where Γ is the solution of the homogeneous equation

$$\dot{\Gamma} = D\bar{A}(q(\hat{z}, t))\Gamma.$$

Letting π_p, π_q, π_t to denote the projections we get

$$\pi_p[\tilde{\Delta}(t) - \tilde{\Delta}(0)] = \pi_p \Gamma(t) \int_0^t \pi_p \Gamma^{-1}(s) d\pi_p \tilde{\mathbf{B}}(s).$$

Observe that then $t \to \infty$ along the periods of $\bar{z}$ $(\pi_p\Gamma)(t)$ is growing exponentially whereas the second factor approaches in distribution

$$\int_{-\infty}^0 (\pi_p\Gamma)^{-1}(s) d(\pi_p \tilde{\mathbf{B}})(s)$$

which is non-degenerate as can be seen by computing its variance. Thus there exists T_0 such that

$$\text{Prob}(|\pi_p(\Delta)|(T_0) > 2R) \geq 1 - \frac{\delta}{100}.$$

This implies that for small ε we have

$$\ell\left(|p\left(z\left(T_0/\varepsilon\right)\right)| \geq R\sqrt{\varepsilon}\right) \geq 1 - \frac{\delta}{100}.$$

Thus we found ourselves in case (II) at the expense of loosing probability $\delta/100$.

(II) $|p(\tilde{z})| \geq \text{Const} R\sqrt{\varepsilon}$. We shall prove that there exists $\lambda > 0$ such that with large probability

$$\kappa_n := \frac{p(z_{\frac{n+1}{\varepsilon}})}{p(z_{\frac{n}{\varepsilon}})} \geq 1 + \lambda.$$

Indeed there exists $\lambda > 0$ such that if $p(z_0) > 0$ then the solution of the averaged equation satisfies $p(z_1)/p(z_0) > 1+2\lambda$. Thus if $\kappa_j \geq 1+\lambda$ for $j = 1, 2 \dots n-1$ then

$$p(z_{n/\varepsilon}) \geq (1+\lambda)^{n-1} R\sqrt{\varepsilon}$$

and if $\kappa_n < 1+\lambda$ then the difference between the actual and averaged evolution is at least $cR(1+\lambda)^n\sqrt{\varepsilon}$. According to Theorem 6(b) the probability of this is less than

$$\max(C_1 \exp(-C_2 R^2 (1+\lambda)^n), C_1 \exp(-C_2/\varepsilon))$$

so if R is large enough then the probability that $\kappa_j < 1+\lambda$ for some $j < C|\ln\varepsilon|$ can be made as small as we wish. This completes the proof of (58). Theorem 9 is established. □

9. Lyapunov exponents.

Here we apply the previous results to the dynamics of skew products. Consider $F_\varepsilon : M \times Z \to M \times Z$ given by $F_\varepsilon(z, x) = (fx, g_{x,\varepsilon}(z))$ where $f : M \to M$ is Anosov and $g_{x,\varepsilon}$ are close to id. We compute the asymptotics of the Lyapunov exponents of those products. Similar computations in slightly different setting can be found in [43, 41, 19, 21].

Recall that a partially hyperbolic diffeomorphism is called *mostly contracting* if its central exponents are negative for any u-Gibbs state. Mostly contracting diffeomorphisms has been studied in [7, 9, 10, 11, 17, 42] and their properties are well understood.

Theorem 11. *If for the averaged system the non-wandering set consists of finite number of hyperbolic* FIXED *points then F_ε is mostly contracting for small ε.*

Proof. Let $G(Z)$ be the Grassmann bundle of Z. F_ε induces a diffeomorphism $\tilde{F}_\varepsilon : G(Z) \to G(Z)$. Since the integration (averaging) commutes with differentiation the averaged equation on $G(Z)$ is induced from the averaged equation on Z. In the proof of Theorem 9 we saw that most orbits spend most of the time near the sinks. It follows that for large T we have for all $\ell \in E$

$$\mathbb{E}_\ell \left(\ln ||(dF_\varepsilon^{T/\varepsilon}|E_c)||\right) \leq -cT$$

This implies ([9]) that F_ε is mostly contracting. □

Next assume that F_ε satisfies the conditions of Theorem 10*. Then by the above argument the induced system on TZ and hence on $G(Z)$ also vanishes. Thus we can apply Theorem 10 to the induced action of F_ε on $G(Z)$. Let $\tilde{Z}_t$ denote the induced process. Assume that $\tilde{Z}$ has unique invariant measure η and let $\lambda_1(\eta) \geq \lambda_2(\eta) \geq \cdots \geq \lambda_d(\eta)$ be its

Lyapunov exponents. The argument similar to the proof of Theorem 11 give the following result.

Theorem 12. *Let ν_ε be u-Gibbs state for F_ε and let*

$$\lambda_1(\varepsilon) \geq \lambda_2(\varepsilon) \geq \cdots \geq \lambda_d(\varepsilon)$$

be its Lyapunov exponents. Then $\lambda_m(\varepsilon)/\varepsilon^2 \to \lambda_m(\eta)$.

Proof. Since η is unique it follows that for all k and δ there exists T_0 such that for any k-plane Π we have

$$\left| \mathbb{E}\left(\frac{1}{T_0} \ln \Lambda_{T_0}(\Pi)\right) - \sum_{m=1}^{k} \lambda_m(\eta) \right| \leq \frac{\delta}{2}$$

where $\Lambda_T(\Pi)$ is the expansion of the plane Π at time T. Then for small ε we have the following. Let $n_\varepsilon = T_0\varepsilon^{-2}$, then for any $\ell \in E(F_\varepsilon)$ for any plane field Π which is sufficiently close to constant we get

$$\left| \mathbb{E}_\ell\left(\frac{1}{T_0} \ln \det(df^{n_\varepsilon}|\Pi)(x)\right) - \sum_{m=1}^{k} \lambda_m(\eta) \right| \leq \delta.$$

Now Theorem 1 (applied to $\tilde{F}_\varepsilon$) implies that all limit points of $\frac{1}{n} \ln \det(df^n|\Pi)(x)$ is within $\varepsilon^2\delta$ from $\varepsilon^2 \sum_{m=1}^{k} \lambda_m(\eta)$. Since δ is arbitrary we have

$$\lim_{\varepsilon \to 0} \frac{\sum_{m=1}^{k} \lambda_m(\varepsilon)}{\varepsilon^2} \to \sum_{m=1}^{k} \lambda_m(\eta).$$

□

Remarks. (1) The assumption that $\tilde{Z}_t$ has unique invariant measure could be relaxed considerably (see [4, 8, 24]) but we do not pursue we subject here since the assumption of Theorem 12 holds for generic diffusion.

(2) In the case the perturbations destroy the skew-product structure one still can obtain the asymptotics of Lyapunov exponents (cf. [41]) by considering the induced action on $k + \dim(E_u)$ dimensional planes $(1 \leq k \leq \dim(Z))$ but the
resulting exponents have less clear probabilistic meaning.

For our final example let $Z = \mathbb{T}^d$ and let

$$F_\varepsilon(x, z) = (f(x), z + \alpha(x) + \varepsilon\beta(x, z) + \varepsilon^2\gamma(x, z, \varepsilon))$$

Suppose that F_0 satisfies (11). Let v_0 be a vector in $\mathbb{R}^d = TZ$ and $v_n = dF_\varepsilon(v_0)$, then

$$v_{n+1} = \left(1 + \varepsilon\frac{\partial\beta}{\partial z} + \varepsilon^2\frac{\partial\gamma}{\partial z}\right) v_n.$$

We have $d\nu(F) = d\mu_{SRB}(x)dz$ where μ_{SRB} is the SRB measure for f. Since

$$\int \frac{\partial \beta}{\partial z} d\nu = \int \left(\int \frac{\partial \beta}{\partial z} dz \right) d\mu_{SRB}(x) = 0$$

the averaged system vanishes. Thus we can apply Theorem 7. (Formally Theorem 7 does not apply since it is unknown if F_0 has stretched exponential decay of correlations. However the proof works in our situation because $\frac{\partial F}{\partial v} = 0$ and so there are no $Y_{z,n}$ terms involving triple correlation functions. Hence rapid mixing ([16]) suffices here.) The drift can be computed as follows. $a_4(z) = 0$ since $\frac{\partial F}{\partial v} = 0$. Next,

$$a_1(v) = \int \frac{\partial \gamma}{\partial z} d\nu = 0,$$

$$a_2(v) = \sum_{n=1}^{\infty} \frac{\partial \beta}{\partial z}(F^n(z,x)) \frac{\partial \beta}{\partial z}(z,x) v dz d\mu_{SRB}(x)$$

whereas the expression for $a_3(v)$ is given in ([19], Section 2.10)

$$a_3(v) = \sum_{n=1}^{\infty} \frac{\partial^2 \beta}{\partial z^2}(F^n(z,x))\left(\beta(z,x), v\right) dz d\mu_{SRB}(x).$$

Integration by parts shows that $a_3 = -a_2$ so $a(v) = 0$. Thus the effective equation in this case can be obtained as follows. Consider the recurrence $M_{n+1} - M_n = \varepsilon \frac{\partial \beta}{\partial z}(F_0^n(z,x))$ then as $\varepsilon \to 0$ $M_{t\varepsilon^{-2}} \to W(t)$ and the diffusion process of Theorem 7 takes form

$$dV = (dW)V. \tag{59}$$

The arguments of Theorem 12 give the following.

Theorem 13. *Suppose that the linear system (59) has unique invariant measure on each Grassmannian. Let $\lambda_m(\varepsilon)$ be defined as in Theorem 12. Then as $\varepsilon \to 0$ $\lambda_m(\varepsilon)/\varepsilon^2$ converge to the Lyapunov exponents of (59).*

10. Open problems.

We conclude with listing some open problems.

(1) Find further restrictions on the class of limiting u-Gibbs states.

Theorems 10 and 9 deal with the simple averaged motions: identity and Morse-Smale respectively. It is of interest to extend this analysis to more complicated averaged dynamics. Theorem 6 suggest that the better description of the actual motion is the averaged equation *with small Gaussian corrections.* This leads to the following refinement of question (1).

(1a) Suppose that there is unique measure which is the limit of the small noise perturbations of the averaged equation. Is it true that η is this measure?

As a special case we have the following.

(1b) Suppose that the averaged motion is partially hyperbolic. Is it true that η is a u-Gibbs state?

This question is open even in the uniformly hyperbolic case. We remark that (1a) would imply (1b) due to the results of [29, 12].

(2) In the Morse-Smale case (Theorem 9) can the coefficients c_j be specified completely (at least under some non-degeneracy assumptions)?

In the case then the fast motion is Anosov and is uncoupled from the slow one the answer is given by large deviation techniques ([30]) but little is known about large deviations for general partially hyperbolic systems.

(3) Can one rule out non-sinks as limits of SRB measures in the setting more general than that of Theorem 9?

As it was mentioned Theorem 9 is false for direct products since in that case one can have u-Gibbs measures concentrated on arbitrary periodic points. However those measures are not SRB.

(4) Extend the results of this paper to the case of flows.

For Anosov flows the time one maps are *generically* rapidly mixing ([15, 34, 22]), so Theorem 6 holds for generic flow. However in the uncoupled case there is no need to exclude non-mixing flows. So it is interesting if the same is true in the coupled setting. On the other hand Theorem 7 does not apply since it is unknown if Anosov flows have stretched exponential decay of correlation (except for contact flows ([34]). It seems that one can establish the convergence of the triple sum using an approach of [33], but we do not pursue this question here.

Appendix A. Mixing properties of F.

Here we discuss estimates of multiple correlation functions for $F : M \times \mathbb{R}^s \to M \times \mathbb{R}^s$ given by $(x, z) \to (f_z x, z)$ where f_z is FAE and there are constants c_1, c_2 such that for all z for all $\ell \in E(f_z)$ we have

$$|\mathbb{E}_\ell(B \circ f_z^n) - \nu_z(A)| \leq c_1 e^{-c_2 n^\gamma} ||B||_{C^{\bar{\alpha}}(M)} \tag{60}$$

In the estimates below we let $\bar{c}_1, \bar{c}_2$ be constants whose precise value can change from entry-to-entry.

(60) implies in particular that for all $B_1, B_2 \in C^{\bar{\alpha}}(M)$ we have

$$\left| \int B_1(x) B_2(f_z^n x) dx - \int B_1(x) dx \times \nu_z(B_2) \right| \leq \bar{c}_1 e^{-\bar{c}_2 n^\gamma} ||B_1||_{C^{\bar{\alpha}}(M)} ||B_2||_{C^{\bar{\alpha}}(M)}. \tag{61}$$

This inequality allows us to to control correlation functions for F. Namely if B_1, B_2 are Holder functions on $M \times \mathbb{R}^s$ then

$$\left| \int B_1(x,z) B_2(F^n(x,z)) dx dz - \int B_1(x,z) \bar{B}_2(z) dx dz \right|$$
$$\leq \bar{c}_1 e^{-\bar{c}_2 n^\gamma} ||B_1||_{C^{\bar{\alpha}}} ||B_2||_{C^{\bar{\alpha}}}.$$

Define F-admissible sets using cones

$$\mathcal{K}_u(x,z) = \{(u,v): \quad u \in K_u(x, f_z), \quad ||v|| \leq \bar{\delta} ||u||\}.$$

Approximating integrals over admissible sets by integrals over their small tubular neighbourhoods we get

$$\left| \mathbb{E}_\ell(B_1(B_2 \circ F^n)) - \mathbb{E}_\ell(B_1 \bar{B}_2) \right| \leq \bar{c}_1 e^{-\bar{c}_2 n^\gamma} ||B_1||_{C^{\bar{\alpha}}} ||B_2||_{C^{\bar{\alpha}}} \tag{62}$$

We now state several bounds on multiple correlation functions. Given Holder functions B_1, B_2, B_3 let

$$\rho_{jk}(\ell) = \mathbb{E}_\ell(B_1(B_2 \circ F^k)(B_3 \circ F^{k+j}).$$

Denote

$$\mathcal{N} = ||B_1||_{C^{\bar{\alpha}}} ||B_2||_{C^{\bar{\alpha}}} ||B_3||_{C^{\bar{\alpha}}}$$

Lemma 3. *If $\bar{B}_3 \equiv 0$ then*

$$|\rho_{jk}(\ell)| \leq \bar{c}_1 e^{-\bar{c}_2 j^\gamma} \mathcal{N} \tag{63}$$

Proof. It suffices to prove this for $\ell \in E_1(F)$. We consider two cases.
(I) $k \geq j$. Split

$$\rho_{jk}(\ell) = \sum_r c_r \mathbb{E}_{\ell_r}((B_1 \circ F^{-k}) B_2 (B_3 \circ F^j)) + O(\theta^k).$$

Applying (62) to each ℓ_r with function $B_4 = (B_1 \circ F^{-k}) B_2$ we get (63).
(II) $k < j$. Split

$$\rho_{jk}(\ell) = \sum_r c_r \mathbb{E}_{\ell_r}((B_1 \circ F^{-(k+j/2)})(B_2 \circ F^{-j/2})(B_3 \circ F^{j/2})) + O(\theta^j)$$

and argue as in case (I). □

Let $\gamma^* = \frac{1}{\gamma} + 1$.

Lemma 4. *In parts (a)–(c) we suppose that $k \geq j^{\gamma^*}$ Then*

$$(a) \quad \rho_{jk}(\ell) = \mathbb{E}_\ell(B_1 \nu_z(B_2(B_3 \circ F^j))) + O\left(e^{-\bar{c}_2 k^\gamma} \mathcal{N}\right).$$

(b) Suppose that $\ell(|z - z^*| \leq \varepsilon) = 1$ for some z^*, ε. Then

$$\rho_{jk}(\ell) = \mathbb{E}_\ell(B_1) \nu_{z^*}(B_2(B_3 \circ F^j)) + O(\varepsilon) + O\left(e^{-\bar{c}_2 k^\gamma} \mathcal{N}\right).$$

(c) If $\bar{B}_2\bar{B}_3 \equiv 0$ *then*

$$\rho_{jk}(\ell) = O\left(e^{-\bar{c}_2 j^\gamma}\mathcal{N}\right).$$

(d) If $\bar{B}_2\bar{B}_3 \equiv 0$ *then (without assuming* $k \geq j^{\gamma^*}$*) we have*

$$\rho_{jk}(\ell) = O\left(\exp\left(-\bar{c}_2 \min(j,k)^{\bar{\gamma}}\right)\mathcal{N}\right)$$

for some $\bar{\gamma} > 0$.

Proof. (a) Rewrite $\rho_{jk}(\ell) = \mathbb{E}_\ell(B_1(B_4 \circ F^k))$ where $B_4 = B_2(B_3 \circ F^j)$. Applying (62) and observing that

$$||B_4||_{C^{\bar{\alpha}}} \leq ||B_2||_{C^{\bar{\alpha}}}||B_3||_{C^{\bar{\alpha}}}K^j$$

for some constant K we obtain (a).

Now (b) follows from the fact that if $|z - z^*| \leq \varepsilon$ then

$$\nu_z(B_2(B_3 \circ f_z^j)) = \nu_{z^*}(B_2(B_3 \circ f_{z^*}^j)) + O(\varepsilon)$$

(see Appendix B.)

To get (c) we apply (a) and estimate $\nu_z(B_2(B_3 \circ F^j))$ using (62).

(d) follows from (c) and Lemma 3. □

Next we shall need the estimates on the triple correlation functions of the type

$$\rho_{jkm}(\ell) = \mathbb{E}_\ell((B_1 \circ F^m)(B_2 \circ F^{m+k})(B_3 \circ F^{m+k+j})).$$

The analyzis of ρ_{jkm} is similar to ρ_{jk}.

Lemma 5. *(a) If* $\bar{B}_3 \equiv 0$ *then*

$$\rho_{jkm} = O(e^{-\bar{c}_2 j^\gamma}\mathcal{N}).$$

(b) If $k \geq j^{\gamma^*}$ *then*

$$\rho_{jkm} = \mathbb{E}_\ell((B_1 \circ F^m)\nu_z(B_2(B_3 \circ F^j))) + O(e^{-\bar{c}_2 k^\gamma}\mathcal{N}).$$

(c) If $k \geq j^{\gamma^*}$ *and* $\ell(|z - z^*| < \varepsilon) = 1$ *then*

$$\rho_{jkm} = \mathbb{E}_\ell((B_1 \circ F^m))\nu_{z^*}(B_2(B_3 \circ F^j))) + O(\left[\varepsilon + e^{-\bar{c}_2 k^\gamma}\right]\mathcal{N}).$$

(d) If $m > (j+k)^{\gamma^*}$ *then*

$$\rho_{jkm} = \mathbb{E}_\ell(\nu_z(B_1(B_2 \circ F^k)(B_3 \circ F^{j+k}))) + O(e^{-\bar{c}_2 m^\gamma}\mathcal{N}).$$

There exists $\bar{\gamma} > 0$ *such that*

(e) If $m > (j+k)^{\gamma^*}$ *and* $\bar{B}_1 = \bar{B}_3 = 0$ *then*

$$\rho_{jkm}(\ell) = O\left(\exp\left(-\bar{c}_2 \max(j,k)^{\bar{\gamma}}\right)\mathcal{N}\right)$$

(f) If $\bar{B}_1 = \bar{B}_3 = 0$ *then*

$$\rho_{jkm}(\ell) = O(e^{-\bar{c}_2 n^{\bar{\gamma}}}\mathcal{N})$$

where n *is the second largest among* j, k, m.

Proof. The proofs are similar to the proofs of Lemmas 3 and 4. (To get (e) we use (d) and Lemma 4(d).) $\square$

Appendix B. Smoothness of the diffusion coefficient.

Here we prove that under the conditions of Theorem 7 σ^2 is uniformly C^1. We shall use the mixing estimates of Appendix A which clearly hold for individual f. (In other words we have $s=0$ in the setting of Appendix A.)

Proof. In the expression for σ^2 both A and f depend on z. For f fixed the map $A \to \sigma^2_{\alpha\beta}(A)$ is smooth (quadratic). So it remains to be proven that σ^2 is a differentiable function of f. Let $B: M \to \mathbb{R}^s$ be fixed and f_ε be FAE having uniform stretched exponential decay of correlations. Let ν_ε be the SRB measure for f_ε. We need to show that

$$\varepsilon \to \sigma^2_{\alpha\beta}(\varepsilon) := \sum_{n=-\infty}^{\infty} [\nu_\varepsilon(B_\alpha(B_\beta \circ f_\varepsilon^n)) - \nu_\varepsilon(B_\alpha)\nu_\varepsilon(B_\beta)]$$

is uniformly C^1. We have

$$\sigma^2_{\alpha\beta}(\varepsilon) := \nu_\varepsilon(B_\alpha B_\beta) - \nu_\varepsilon(B_\alpha)\nu_\varepsilon(B_\beta) + \sigma^2_{\alpha\beta+}(\varepsilon) + \sigma^2_{\beta\alpha+}(\varepsilon)$$

where

$$\sigma^2_{\alpha\beta+}(\varepsilon) = \sum_{n=1}^{\infty} [\nu_\varepsilon(B_\alpha(B_\beta \circ f_\varepsilon^n)) - \nu_\varepsilon(B_\alpha)\nu_\varepsilon(B_\beta)]$$

Since the first two terms are C^1 functions of ε by Theorem 4 it remains to show that $\varepsilon \to \sigma^2_{\alpha\beta+}(\varepsilon)$ is C^1. Let us show that this map is differentiable at $\varepsilon = 0$. We follows [13]. We have

$$\sigma^2_{\alpha\beta+}(\varepsilon) = \lim_{N\to\infty} \lim_{M\to\infty} \sum_{n=1}^{N} \nu_0((B \circ f_\varepsilon^M)(B \circ f_\varepsilon^{n+M}))$$

so it is enough to show that the derivative of the above expressions converge as $M, N \to \infty$. To simplify the formulas we assume that $\nu_0(B) = 0$. Let $X = \frac{d}{d\varepsilon}|_{\varepsilon=0}(f_\varepsilon \circ f_0^{-1})$. Differentiating we get

$$\sum_n \sum_{k=1}^{M} \nu_0(\partial_X \left[(B_\alpha \circ f_0^k)(B_\beta \circ f_0^{n+k})\right] \circ f_0^{M-k}) + \sum_n \sum_{k=M+1}^{M+n} \nu_0((B_\alpha \circ f_0^M) \left(\partial_X \left[B_\beta f_0^{M+n-k}\right] \circ f_0^k\right))$$

$= P + Q$. To estimate P we decompose

$$df_0^k(X) = \Gamma^k_{as} X_{as} + \sum_l (\bar{a}_l \circ f_0^{-k}) e_l + \sum_{lm} (\bar{a}_{lm} \circ f^{k-m}) e_l + df_0^k X_u + \sum_m df_0^{k-m} \Gamma_u \Gamma^m_{as} X_{as}$$

where $X_c = \sum_l \bar{a}_l e_l$, $\Gamma_c \Gamma^m_{as} X_{as} = \sum_l \bar{a}_{lm} e_l$. Accordingly P is a sum of five terms.

$$(I) \quad P^{I}_{nk} = \nu_0 \left(\partial_{\Gamma^k_{as} X_{as}}[B_\alpha(B_\beta \circ f^n)]\right).$$

Split $P^I_{nk} = P^{Ia}_{nk} + P^{Ib}_{nk}$ corresponding to

$$\partial_{\Gamma^k_{as} X_{as}}[B_\alpha(B_\beta \circ f^n)] = (B_\beta \circ f^n)\partial_{\Gamma^k_{as} X_{as}}(B_\alpha) + B_\alpha \partial_{\Gamma^k_{as} X_{as}}[(B_\beta \circ f^n)].$$

Now

$P^{Ia}_{nk} = O(\theta^k e^{-\bar{c}_2 n^\gamma})$ by Lemma 3 (observe that $||\Gamma^k_{as}|| \leq \theta^k$). Hence $\sum_{nk} P^{Ia}_{nk}$ converges uniformly. On the other hand $P^{Ib}_{nk} = O(\theta^{n+k})$ since $B \circ f^n_0$ is almost constant on stable leaves. Hence $\sum_{nk} P^{Ib}_{nk}$ converges uniformly.

$$(\mathit{II}) \quad P^{\mathit{II}}_{lnk} = \nu_0(\bar{a}_l \, [\partial_{e_l}(B_\alpha(B_\beta \circ f^n_0))] \circ f^k_0).$$

Split it as $P^{\mathit{II}a}_{lnk} + P^{\mathit{II}b}_{lnk}$ as in part (I). Since $\nu_0(B_\alpha) = \nu_0(B_\beta) = \nu_0(\partial_{e_l} B_\alpha) = \nu_0(\partial_{e_l} B_\beta)$ Lemma 3 shows that

$$P^{\mathit{II}\sigma}_{lnk} \leq \bar{c}_1 e^{-\bar{c}_2 n^\gamma} \quad \sigma = 1, 2$$

and if $k > n^{\gamma^*}$ then by Lemma 4(a)

$$P^{\mathit{II}}_{lnk} = \nu_0(\bar{a}_l)\nu_0(\partial_{e_l}(B_\alpha(B_\beta \circ f^n_0)) + O(e^{-\bar{c}_2 n^\gamma}) = O(e^{-\bar{c}_2 n^\gamma})$$

since e_l preserves ν_0. Thus $\sum_{nk} P^{\mathit{II}}_{lnk}$ converges uniformly.

$$(\mathit{III}) \quad P^{\mathit{III}}_{lknm} = \nu_0(\bar{a}_{lm} \, [\partial_{e_l}(B_\alpha(B_\beta \circ f^n_0))] \circ f^k_0).$$

Since $||\Gamma^m_{as}|| \leq \theta^m$ the analysis of these terms is similar to the analysis of type *II* terms. Let us state the relevant bounds.

$$(a) \quad P^{\mathit{III}}_{lnkm} = O(\theta^m e^{-\bar{c}_2 n^\gamma}).$$

(b) If $k > n^{\gamma^*}$ then $P^{\mathit{III}}_{lnkm} = O(\theta^m e^{-\bar{c}_2 k^\gamma})$. Hence $\sum_{nkm} P^{\mathit{III}}_{lnkm}$ converge uniformly.

For the terms containing E_u-components we integrate by parts inside unstable leaves. Thus we get

$$(\mathit{IV}) \quad P^{\mathit{IV}}_{nk} = -\nu_0(\mathrm{div}^{can}_u(X_u)(B_\alpha \circ f^k_0)(B_\beta \circ f^{n+k}_0)).$$

For this term we have two bounds. By Lemma 3

$$P^{\mathit{IV}}_{nk} = O(e^{-\bar{c}_2 n^\gamma}).$$

On the other hand if $k > n^{\gamma^*}$ then by Lemma 4(a)

$$P^{\mathit{IV}}_{nk} = \nu_0(\mathrm{div}^{can}_u(X_u))\nu_0(B_\alpha(B_\beta \circ f^n_0)) + O(e^{-\bar{c}_2 k^\gamma}) = O(e^{-\bar{c}_2 k^\gamma})$$

since

$$\nu_0(\mathrm{div}^{can}_u(X_u)) = \nu_0(\partial_{X_u} 1) = 0.$$

Hence $\sum_{nk} P^{\mathit{IV}}_{nk}$ converges uniformly.

$$(V) \quad P^V_{lnkm} = -\nu_0(\mathrm{div}^{can}_u(\Gamma_c\Gamma^m_{as}X_{as})(B_\alpha \circ f^k_0)(B_\beta \circ f^{n+k}_0)).$$

The estimates of these terms are similar to the estimates of part *IV* terms. Namely by Lemma 3

$$P^V_{lnkm} = O(\theta^m e^{-\bar{c}_2 n^\gamma})$$

and if $k > n^{\gamma^*}$ then by Lemma 4(a)

$$P^V_{lnkm} = O(\theta^m e^{-\bar{c}_2 k^\gamma}).$$

Thus $\sum_{nkm} P^V_{lnkm}$ converges uniformly.

Now we pass to Q-terms. They can be split into five parts as P-terms. Recall that $\nu_0(B_\alpha) = 0$.

$$(I) \quad Q^I_{nk} = \nu_0(B_\alpha(\partial_{\Gamma^k_{as}X_{as}}B_\beta) \circ f^n_0).$$

This term can be estimated as follows.

$Q^I_{nk} = O(\theta^k)$ since $||\Gamma^k_{as}|| \leq \theta^k$ and by (62)

$$Q^I_{nk} = \nu_0(B_\alpha)\nu_0(\Gamma^k_{as}B_\beta) + O(e^{-\bar{c}_2 n^\gamma}) = O(e^{-\bar{c}_2 n^\gamma}).$$

Thus $\sum_{nk} Q^I_{nk}$ converges uniformly.

$$(II) \quad Q^{II}_{lnk} = \nu_0(B_\alpha(\bar{a} \circ f^{n-k}_0)((\partial_{e_l}B_\beta) \circ f^n_0)).$$

Again we have two estimates. $Q^{II}_{lnk} = O(e^{-\bar{c}_2 k^\gamma})$ by Lemma 3 and if $n-k \geq k^{\gamma^*}$ then $Q^{II}_{lnk} = O(e^{-\bar{c}_2 (n-k)^\gamma})$ by Lemma 4(a). Hence $\sum_{nk} Q^{II}_{lnk}$ converges uniformly.

$$(III) \quad Q^{III}_{lnkm} = \nu_0(B_\alpha(\bar{a}_{lm} \circ f^{n-k}_0)(\partial_{e_l}B_\beta) \circ f^n_0).$$

The bounds for these terms are

$$Q^{III}_{lnkm} = O(\theta^m e^{-\bar{c}_2 k^\gamma}) \quad \text{(by Lemma 3) and}$$

$$Q^{III}_{lnkm} = O(\theta^m e^{-\bar{c}_2 (n-k)^\gamma}) \text{ if } n-m-k > (m+k)^{\gamma^*} \quad \text{(by Lemma 4(a))}.$$

Hence $\sum_{nkm} Q^{IV}_{lnkm}$ converges uniformly. For the other terms we integrate by parts inside unstable leaves. Thus we get the following expressions.

$$(IV) \quad Q^{IV}_{nk} = -\nu_0(\mathrm{div}^{can}_u \left[(B_\alpha \circ F^{-(n-k)}_0)X_u\right](B_\beta \circ f^k_0)).$$

Since $\mathrm{div}^{can}_u \left[(B_\alpha \circ F^{-(n-k)}_0)X_u\right]$ is uniformly Holder Lemma 3 gives

$$Q^{IV}_{nk} = O(e^{-\bar{c}_2 k^\gamma}).$$

On the other hand

$$\mathrm{div}_u^{can}\left[(B_\alpha\circ F_0^{-(n-k)})X_u\right]=(B_\alpha\circ f_0^{-(n-k)})\mathrm{div}_u^{can}(X_u)+O(\theta^{n-k})$$

so if $n-k\geq k^{\gamma^*}$ then by Lemma 4(a)

$$Q_{nk}^{IV}=O(e^{-\bar{c}_2(n-k)^\gamma})$$

proving the convergence of $\sum_{nk}Q_{nk}^{IV}$.

$$(V)\quad Q_{nkm}^{V}=-\nu_0(\mathrm{div}_u^{can}\left[(B_\alpha\circ f_0^{-(n-k)})\Gamma_c\Gamma_{as}^m X_{as}\right](B_\beta\circ f_0^k)).$$

By Lemma 3

$$Q_{nkm}^{V}=O(\theta^m e^{-\bar{c}_2 k^\gamma})$$

and if $n-m-k\geq(m+k)^{\gamma^*}$ then by Lemma 5(b)

$$Q_{nkm}^{V}=O(\theta^m e^{-\bar{c}_2(n-m-k)^\gamma}).$$

Thus $\sum_{nkm}Q_{nkm}^{V}$ converges. This completes the proof of the differentiability of σ^2. □

References

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

[2] Bakhtin V. I. *Cramer theorem with two discrepancies,* (Russian) Proc. Belarussian Math. Inst. **6** (2000) 23-26.

[3] Bakhtin V. I. *Cramer asymptotics in the averaging method for systems with fast hyperbolic motions,* Tr. Mat. Inst. Steklova **244** (2004) 65–86.

[4] Baxendale P. H. *Lyapunov exponents and relative entropy for a stochastic flow of diffeomorphisms,* Prob. Th. & Rel. Fields **81** (1989) 521–554.

[5] Billingsley P. *Convergence of Probability Measures,* 1979, Wiley and Sons, New York.

[6] Bowen R. *Equilibrium states and ergodic theory of Anosov diffeomorphisms,* Lect. Notes in Math. **470** (1975) Springer New York.

[7] Bonnatti C.& Viana M. *SRB measures for partially hyperbolic systems: mostly contracting case*, Israel J. Math **115** (2000) 157–193.

[8] Brin M. & Kifer Yu. *Dynamics of Markov chains and stable manifolds for random diffeomorphisms,* Erg. Th. & Dyn. Sys. **7** (1987) 351–374.

[9] Burns K., Dolgopyat D. & Pesin Ya. B. *Partial hyperbolicity, Lyapunov exponents and stable ergodicity,* J. Stat. Phys. **108** (2002) 927–942.

[10] Castro A. A. *Backward inducing and exponential decay of correlations for partially hyperbolic attractors,* Israel J. Math. **130** (2002) 29–75.

[11] Castro A. A. *Fast mixing for attractors with a mostly contracting central direction,* Erg. Th. & Dyn. Sys. **24** (2004) 17–44.

[12] Cowieson W. & Young L.–S. *SRB measures as zero-noise limits,* preprint.

[13] Chernov N. & Dolgopyat D. *Brownian Brownian Motion-1,* preprint.

[14] Chernov N. I., Eyink G. L., Lebowitz J. L. & Sinai Ya. G. *Steady-state electrical conduction in the periodic Lorentz gas,* Comm. Math. Phys. **154** (1993), 569–601.

[15] Dolgopyat D. *On decay of correlations in Anosov flows,* Ann. Math. **147** (1998) 357-390.

[16] Dolgopyat D. *On mixing properties of compact group extensions of hyperbolic systems*, Israel J. Math. **130** (2002) 157–205.

[17] Dolgopyat D. *On dynamics of mostly contracting diffeomorphisms,* Comm. Math. Phys. **213** (2000) 181–201.

[18] Dolgopyat D. *Limit theorems for partially hyperbolic systems,* Trans. AMS **356** (2004) 1637–1689.

[19] Dolgopyat D. *On differentiability of SRB states for partially hyperbolic systems,* Invent. Math. **155** (2004) 389–449.

[20] Dolgopyat D. *Livsic theory for compact group extensions of hyperbolic systems,* Moscow Math. J. **5** (2005) 55-66.

[21] Dolgopyat D. & Pesin Ya. B. *Every manifold carries a completely hyperbolic diffeomorphism,* Erg. Th. & Dyn. Sys. **22** (2002) 409–437.

[22] Field M., Melbourne I. & Torok A. *Stability of mixing and rapid mixing for hyperbolic flows,* preprint.

[23] Freidlin M. I. & Wentzell A. D. *Random perturbations of dynamical systems,* 2d ed. Grundlehren der Mathematischen Wissenschaften **260** (1998) Springer, New York.

[24] Goldsheid I. Ya. & Margulis G. A. *Lyapunov exponents of a product of random matrices,* Russ. Math. Surv. **44** (1989) no. 5, 11–71.

[25] Ibragimov I. A. & Linnik Yu. V. *Independent and stationary sequences of random variables,* Translation from the Russian ed. by J. F. C. Kingman. Wolters-Noordhoff Publishing, Groningen, 1971.

[26] Katok A., Knieper G., Pollicott M. & Weiss H. *Differentiability and analyticity of topological entropy for Anosov and geodesic flows,* Invent. Math. **98** (1989) 581–597.

[27] Katok A. & Kononenko A. *Cocycles' stability for partially hyperbolic systems,* Math. Res. Lett. **3** (1996) 191–210.

[28] Katznelson Y. Ergodic automorphisms of $\mathbb{T}^n$ are Bernoulli, Israel J. Math. **10** (1971) 186-195.

[29] Kifer Yu. *Random perturbations of dynamical systems,* Progress in Prob. & Stat. **16** (1988) Birkhauser, Boston.

[30] Kifer Yu. *Averaging in dynamical systems and large deviations,* Invent. Math. **110** (1992) 337–370.

[31] Kifer Yu. *Averaging principle for fully coupled dynamical systems and large deviations,* Erg. Th. & Dyn. Sys. **24** (2004) 847–871.

[32] Kleinbock D. Y. & Margulis G. A. *Bounded orbits of nonquasiunipotent flows on homogeneous spaces,* Amer. Math. Soc. Transl. **171** (1996) 141-172.

[33] Kotani M. *Higher correlations for an Anosov flow,* preprint.

[34] Liverani C. *On contact Anosov flows,* Ann. of Math. **159** (2004), 1275–1312.

[35] Livsic A. N. *Cohomology of dynamical systems,* (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 1296–1320.

[36] Nicol M. & Pollicott M. *Livsic's theorem for semisimple Lie groups,* Erg. Th. & Dyn. Sys. **21** (2001) 1501–1509.

[37] Palis J. *A global view of dynamics and a conjecture on the denseness of finitude of attractors,* Geometrie complexe et systemes dynamiques (Orsay, 1995). Asterisque **261** (2000) 335–347.

[38] Pesin Ya. B. & Sinai Ya. G. *Gibbs measures for partially hyperbolic attractors* Erg. Th. & Dyn. Sys. **2** (1982) 417–438.

[39] Ruelle D. *Differentiation of SRB states,* Comm. Math. Phys. **187** (1997) 227–241 and **234** (2003) 185–190.

[40] Ruelle D. *Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics* J. Stat. Phys. **95** (1999) 393–468.

[41] Ruelle D. *Perturbation theory for Lyapunov exponents of a toral map: extension of a result of Shub and Wilkinson,* Israel J. Math. **134** (2003), 345–361

[42] Ruelle D. & Wilkinson A. *Absolutely singular dynamical foliations,* Comm. Math. Phys. **219** (2001) 481–487.

[43] Shub M. & Wilkinson A. *Pathological foliations and removable zero exponents,* Invent. Math. **139** (2000) 495–508.

[44] Sinai Ya. G. *Classical dynamic systems with countably-multiple Lebesgue spectrum. II,* Izv. Akad. Nauk SSSR Ser. Mat. **30** (1966) 15-68.

[45] Sinai Ya. G. *Markov partitions and U-diffeomorphisms,* Funkts. Anal. i Prilozhen. **2** (1968) no. 1 64–89; English transl. in Funct. Anal. Appl. **2** (1968).

[46] Sinai Ya. G. *Gibbs measures in ergodic theory,* Uspehi Mat. Nauk **27** (1972) 21-64.

[47] Stroock D. W. *Lectures on stochastic analysis: diffusion theory,* London Math. Soc. Student Texts, **6** (1987) Cambridge University Press, Cambridge.

[48] Tsujii M. *On continuity of Bowen-Ruelle-Sinai measures in families of one-dimensional maps,* Comm. Math. Phys. **177** (1996), 1–11.