Proceedings of Symposia in Pure Mathematics
Volume **89**, 2015
http://dx.doi.org/10.1090/pspum/089/01490

The martingale approach after Varadhan and Dolgopyat

Jacopo De Simoi and Carlangelo Liverani

Abstract. We present, in the simplest possible form, the so called *martingale problem* strategy to establish limit theorems. The presentation is specially adapted to problems arising in partially hyperbolic dynamical systems. We will discuss a simple partially hyperbolic example with fast-slow variables and use the martingale method to prove an averaging theorem and study fluctuations from the average. The emphasis is on ideas rather than on results. Also, no effort whatsoever is done to review the vast literature of the field.

Contents

1. Introduction
2. Preliminaries and results
3. Standard Pairs
4. Conditioning
5. Averaging (the Law of Large Numbers)
6. A recap of what we have done so far
7. Fluctuations (the Central Limit Theorem)
Appendix A. Geometry
Appendix B. Shadowing
Appendix C. Martingales, operators and Itō's calculus
References

C.L. thanks Konstantin Khanin for pointing out (long time ago, during the 2011 *Thematic Program on Dynamics and Transport in Disordered Systems* at the Fields Institute) the need for this note and forcing him to start thinking about it. We thank Mikko Stenlund and Denis Volk for several comments on a preliminary version. We thank Dmitry Dolgopyat, Alexey Korepanov, Zemer Kosloff, Ian Melbourne, Mark Pollicott and the anonymous referees for providing several suggestions which improved the readability of the current version. We thank the Centre Bernoulli, Lausanne, Switzerland where part of this notes were written. Both authors have been partially supported by the ERC Grant MALADY (ERC AdG 246953). J.D.S. acknowledges partial NSERC support.

©2015 American Mathematical Society

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

1. Introduction

In this note[1] we purport to explain in the simplest possible terms a strategy to investigate the statistical properties of dynamical systems put forward by Dmitry Dolgopyat [**3**]. It should be remarked that Dolgopyat has adapted to the field of Dynamical Systems a scheme developed by Srinivasa Varadhan and collaborators first for the study of stochastic process arising from a diffusion [**13**], then for the study of limit theorems (e.g. the hydrodynamics limit), starting with the pioneering [**7**], and large deviations, e.g. [**4**].[2] The adaptation is highly non trivial as in the case of Dynamical Systems two basic tools commonly used in probability (conditioning and Itō calculus) are missing. The lesson of Dolgopyat is that such tools can be recovered nevertheless, provided one looks at the problem in the *right* way.

Rather than making an abstract exposition, we prefer a hands-on presentation. Hence, we will illustrate the method by discussing a super simple (but highly non trivial) example.

The presentation is especially aimed at readers in the field of Dynamical Systems. Thus probabilists could find the exposition at times excessively detailed and/or redundant and at other times a bit too fast.

1.1. Fast-Slow partially hyperbolic systems.

We are interested in studying fast-slow systems in which the fast variable undergoes a strongly chaotic motion. Namely, let M, S be two compact Riemannian manifolds, let $X = M \times S$ be the configuration space of our systems and let m_{Leb} be the Riemannian measure on M. For simplicity, we consider only the case in which $S = \mathbb{T}^d$ for some $d \in \mathbb{N}$. We consider a map $F_0 \in \mathcal{C}^r(X, X)$, $r \geq 3$, defined by

$$F_0(x, \theta) = (f(x, \theta), \theta)$$

where the maps $f(\cdot, \theta)$ are uniformly hyperbolic for every θ. If we consider a small perturbation of F_0 we note that the perturbation of f still yields a uniformly hyperbolic system, by structural stability. Thus such a perturbation can be subsumed in the original maps. Hence, it suffices to study families of maps of the form

$$F_\varepsilon(x, \theta) = (f(x, \theta), \theta + \varepsilon\omega(x, \theta))$$

with $\varepsilon \in (0, \varepsilon_0)$, for some ε_0 small enough, and $\omega \in \mathcal{C}^r$.

Such systems are called fast-slow, since the variable θ, the slow variable, needs a time at least ε^{-1} to change substantially.

The basic question is **what are the statistical properties of F_ε?**

The answer to such a question is at the end of a long road that starts with the attempt to understand the dynamics for times of order ε^{-1}. In this note we will concentrate on such a preliminary problem and will describe how to overcome the first obstacles along the path we would like to walk.

[1] A first, preliminary, version of this note was prepared by the second author for a mini course at the conference *Beyond Uniform Hyperbolicity* in Bedlewo, Poland, held at the end of May 2013, which, ultimately, he could not attend. The note was then extended and presented during the semester *Hyperbolic dynamics, large deviations and fluctuations* held at the Bernoulli Centre, Lausanne, January–June 2013.

[2] It should be noted that the above has no pretension of being an exact historical reconstruction, it just describes the way we learned this material. Indeed, some of the relevant ideas were previously present. See, e.g., the reference [**9**] pointed out to us by Sergei Kuksin.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

1.2. The unperturbed system: $\varepsilon = 0$.
The statistical properties of the system are well understood in the case $\varepsilon = 0$. In such a case θ is an invariant of motion, while for every θ the map $f(\cdot, \theta)$ has strong statistical properties. We will need such properties in the following discussion which will be predicated on the idea that, for times long but much shorter than ε^{-1}, on the one hand θ remains almost constant, while, on the other hand, its change depends essentially on the behavior of an ergodic sum with respect to a fixed dynamics $f(\cdot, \theta)$. It is not obvious which exact general properties are necessary to prove the type of results we are interested in. Yet, let us give an idea of the situation by listing the main properties that we will need, and use, in the following.

(1) the maps $f(\cdot, \theta)$ admit a unique SRB (Sinai–Ruelle–Bowen) measure m_θ.
(2) the measure m_θ, when seen as an element of $\mathcal{C}^1(M, \mathbb{R})'$, is differentiable in θ.
(3) there exists $C_0, \alpha > 0$ such that, for each $g, h \in \mathcal{C}^1(M, \mathbb{R})$, we have[3]

$$|m_{\mathrm{Leb}}(h \cdot g \circ f^n(\cdot, \theta)) - m_\theta(g) m_{\mathrm{Leb}}(h)| \leq C_0 e^{-\alpha n} \|h\|_{\mathcal{B}_1} \|g\|_{\mathcal{B}_2},$$
$$|m_\theta(h \cdot g \circ f^n(\cdot, \theta)) - m_\theta(g) m_\theta(h)| \leq C_0 e^{-\alpha n} \|h\|_{\mathcal{B}_1} \|g\|_{\mathcal{B}_2},$$

where $\mathcal{B}_1, \mathcal{B}_2$ are appropriate Banach spaces.[4]

The above properties hold for a wide class of uniformly hyperbolic systems, [**1**, **2**, **5**, **6**], yet here, to further simplify the exposition, we assume that $M = \mathbb{T}^1$ and

$$\partial_x f \geq \lambda > 2. \tag{1.1}$$

Then a SRB measure is just a measure absolutely continuous with respect to Lebesgue and all the above properties are well known with the choices $\mathcal{B}_1 = \mathcal{C}^1$ and $\mathcal{B}_2 = \mathcal{C}^0$ or $\mathcal{B}_1 = \mathrm{BV}$ and $\mathcal{B}_2 = L^1$ (see [**10**] for a fast and elementary exposition or [**1**] for a more complete discussion).

REMARK 1.1. *For the wondering reader: in all the following arguments the case of an higher dimensional expanding map can be treated in almost exactly the same way (a part from a slightly heavier notation).*[5] *On the contrary, the case of a hyperbolic map is a bit more complex (although the logic of the argument remains exactly the same) due to the different definition of standard pairs necessary to handle the stable direction. See* [**3**] *for details.*

REMARK 1.2. *Note that in the following we do not require or use the exact knowledge of the spectrum of the transfer operator.*[6] *Yet, a detailed understanding of the transfer operator might be necessary in order to obtain sharper results.*

[3] We remark that a slower decay of correlation could suffice, but let us keep things simple.

[4] The exact required properties for the Banach spaces vary depending on the context. In the context that we are going to consider nothing much is needed. Yet, in general, it could be helpful to have properties that allow to treat automatically multiple correlations: let $\{g_1, g_2, g_3\} \subset C^1$, then

$$m_{\mathrm{Leb}}(g_1 \cdot (g_2 \circ f^n \cdot g_3) \circ f^m) = m_{\mathrm{Leb}}(g_1) m_\theta(g_2 \circ f^n \cdot g_3) + \mathcal{O}(e^{-\alpha m} \|g_1\|_{\mathcal{B}_1} \|g_2 \circ f^n \cdot g_3\|_{\mathcal{B}_2}).$$

Thus, in order to have automatically decay of multiple correlations we need, at least, $\|g_2 \circ f^n\|_{\mathcal{B}_2} \leq C_\# \|g_2\|_{\mathcal{B}_2}$, which is false, for example, for the $\mathcal{C}^1$ norm.

[5] Simply, the support of a standard pair will be a ball rather than a segment.

[6] The transfer operator $\mathcal{L}_\theta$ is simply the adjoint of the dynamics, i.e. $\mathcal{L}_\theta \mu(g) = \mu(g \circ f(\cdot, \theta))$, when acting on an appropriate class of measures.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

It follows that the dynamical systems (X, F_0) has uncountable many SRB measures: all the measures of the form $\mu(\varphi) = \int \varphi(x,\theta) m_\theta(dx)\nu(d\theta)$ for an arbitrary measure ν. The ergodic measures are the ones in which ν is a point mass. The system is partially hyperbolic and has a central foliation. Indeed, the $f(\cdot,\theta)$ are all topologically conjugate by structural stability of expanding maps [**8**]. Let $h(\cdot,\theta)$ be the map conjugating $f(\cdot,0)$ with $f(\cdot,\theta)$, that is $h(f(x,0),\theta) = f(h(x,\theta),\theta)$. Thus the foliation $W^c_x = \{(h(x,\theta),\theta)\}_{\theta\in S}$ is invariant under F_0 and consists of points that stay, more or less, always at the same distance, hence it is a center foliation. Note however that, since in general h is only a Hölder continuous function (see [**8**]) the foliation is very irregular and, typically, not absolutely continuous.

In conclusion, the map F_0 has rather poor statistical properties and a not very intuitive description as a partially hyperbolic system. It is then not surprising that its perturbations form a very rich universe to explore and already the study of the behavior of the dynamics for times of order ε^{-1} (a time long enough so that the variable θ has a non trivial evolution, but far too short to investigate the statistical properties of F_ε) is interesting and non trivial.

2. Preliminaries and results

Let μ_0 be a probability measure on X. Let us define $(x_n,\theta_n) = F^n_\varepsilon(x,\theta)$, then (x_n,θ_n) are random variables [7] if (x_0,θ_0) are distributed according to μ_0.[8] It is natural to define the polygonalization[9]

$$\Theta_\varepsilon(t) = \theta_{\lfloor\varepsilon^{-1}t\rfloor} + (t - \varepsilon\lfloor\varepsilon^{-1}t\rfloor)(\theta_{\lfloor\varepsilon^{-1}t\rfloor+1} - \theta_{\lfloor\varepsilon^{-1}t\rfloor}), \quad t\in[0,T]. \tag{2.1}$$

Note that Θ_ε is a random variable on X with values in $\mathcal{C}^0([0,T],S)$. Also, note the time rescaling done so that one expects non trivial paths.

It is often convenient to consider random variables defined directly on the space $\mathcal{C}^0([0,T],S)$ rather than X. Let us discuss the set up from such a point of view. The space $\mathcal{C}^0([0,T],S)$ endowed with the uniform topology is a separable metric space. We can then view $\mathcal{C}^0([0,T],S)$ as a probability space equipped with the Borel σ-algebra. It turns out that such a σ-algebra is the minimal σ-algebra containing the open sets $\bigcap_{i=1}^n\{\vartheta\in\mathcal{C}^0([0,T],S)\mid \vartheta(t_i)\in U_i\}$ for each $\{t_i\}\subset[0,T]$ and open sets $U_i\subset S$, [**13**, Section 1.3]. Since Θ_ε can be viewed as a continuous map from X to $\mathcal{C}^0([0,T],S)$, the measure μ_0 induces naturally a measure $\mathbb{P}^\varepsilon$ on $\mathcal{C}^0([0,T],S)$: $\mathbb{P}^\varepsilon = (\Theta_\varepsilon)_*\mu_0$.[10] Also, for each $t\in[0,T]$ let $\Theta(t)\in\mathcal{C}^0(\mathcal{C}^0([0,T],S),S)$ be the random variable defined by $\Theta(t,\vartheta) = \vartheta(t)$, for each $\vartheta\in\mathcal{C}^0([0,T],S)$. Next, for each $\mathcal{A}\in\mathcal{C}^0(\mathcal{C}^0([0,T],S),\mathbb{R})$, we will write $\mathbb{E}^\varepsilon(\mathcal{A})$ for the expectation with respect to $\mathbb{P}^\varepsilon$. For $A\in\mathcal{C}^0(S,\mathbb{R})$ and $t\in[0,T]$, $\mathbb{E}^\varepsilon(A\circ\Theta(t)) = \mathbb{E}^\varepsilon(A(\Theta(t)))$ is the expectation of the function $\mathcal{A}(\vartheta) = A(\vartheta(t))$, $\vartheta\in\mathcal{C}^0([0,T],S)$.

To continue, a more detailed discussion concerning the initial conditions is called for. Note that not all measures are reasonable as initial conditions. Just think of the possibility to start with initial conditions given by a point mass, hence killing any trace of randomness. The best one can reasonably do is to fix the

[7] Recall that a *random variable* is a measurable function from a probability space to a measurable space.

[8] That is, the probability space is X equipped with the Borel σ-algebra, μ_0 is the probability measure and (x_n,θ_n) are functions of $(x,\theta)\in X$.

[9] Since we interpolate between close points the procedure is uniquely defined in $\mathbb{T}$.

[10] Given a measurable map $T: X\to Y$ between measurable spaces and a measure P on X, T_*P is a measure on Y defined by $T_*P(A) = P(T^{-1}(A))$ for each measurable set $A\subset Y$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

slow variable and leave the randomness only in the fast one. Thus we will consider measures μ_0 of the following type: for each $\varphi \in \mathcal{C}^0(X, \mathbb{R})$, $\mu_0(\varphi) = \int \varphi(x, \theta_0) h(x) dx$ for some $\theta_0 \in S$ and $h \in \mathcal{C}^1(M, \mathbb{R}_+)$. Our first problem is to understand $\lim_{\varepsilon \to 0} \mathbb{P}^\varepsilon$. After some necessary preliminaries, in Section 5 we will prove the following result.

THEOREM 2.1. *The measures $\{\mathbb{P}^\varepsilon\}$ have a weak limit $\mathbb{P}$, moreover $\mathbb{P}$ is a measure supported on the trajectory determined by the O.D.E.*

$$\begin{aligned} \dot{\overline{\Theta}} &= \bar{\omega}(\overline{\Theta}) \\ \overline{\Theta}(0) &= \theta_0 \end{aligned} \tag{2.2}$$

where $\bar{\omega}(\theta) = \int_M \omega(x, \theta) m_\theta(dx)$.

The above theorem specifies in which sense the random variable Θ_ε converges to the average dynamics described by equation (2.2).

The next natural question is how fast the convergence takes place. To this end it is natural to consider, for each $t \in [0, T]$,

$$\zeta_\varepsilon(t) = \varepsilon^{-\frac{1}{2}} \left[\Theta_\varepsilon(t) - \overline{\Theta}(t) \right].$$

Note that ζ_ε is a random variable on X with values in $\mathcal{C}^0([0, T], \mathbb{R}^d)$ which describes the fluctuations around the average.[11] Let $\widetilde{\mathbb{P}}^\varepsilon$ be the path measure describing ζ_ε when (x_0, θ_0) are distributed according to the measure μ_0. That is, $\widetilde{\mathbb{P}}^\varepsilon = (\zeta_\varepsilon)_* \mu_0$. Our second task, and the last in this note, will be to understand the limit behavior of $\widetilde{\mathbb{P}}^\varepsilon$, hence of the fluctuation around the average. Section 7 will be devoted to proving the following result.

THEOREM 2.2. *The measures $\{\widetilde{\mathbb{P}}^\varepsilon\}$ have a weak limit $\widetilde{\mathbb{P}}$. Moreover, $\widetilde{\mathbb{P}}$ is the measure of the zero average Gaussian process defined by the Stochastic Differential Equation (SDE)*

$$\begin{aligned} d\zeta &= D\bar{\omega}(\overline{\Theta}) \zeta dt + \sigma(\overline{\Theta}) dB \\ \zeta(0) &= 0, \end{aligned} \tag{2.3}$$

where B is the $\mathbb{R}^d$ dimensional standard Brownian motion and the diffusion coefficient σ is given by [12]

$$\begin{aligned} \sigma(\theta)^2 =& m_\theta \left(\hat{\omega}(\cdot, \theta) \otimes \hat{\omega}(\cdot, \theta) \right) + \sum_{m=1}^{\infty} m_\theta \left(\hat{\omega}(f_\theta^m(\cdot), \theta) \otimes \hat{\omega}(\cdot, \theta) \right) + \\ &+ \sum_{m=1}^{\infty} m_\theta \left(\hat{\omega}(\cdot, \theta) \otimes \hat{\omega}(f_\theta^m(\cdot), \theta) \right). \end{aligned} \tag{2.4}$$

where $\hat{\omega} = \omega - \bar{\omega}$ and we have used the notation $f_\theta(x) = f(x, \theta)$. In addition, σ^2 is symmetric and non-negative, hence σ is uniquely defined as a symmetric positive definite matrix. Finally, $\sigma(\theta)$ is strictly positive, unless $\hat{\omega}(\theta, \cdot)$ is a coboundary for f_θ.

[11] Here we are using that $S = \mathbb{T}^d$ can be lifted to its universal cover $\mathbb{R}^d$.

[12] In our notation, for any measure μ and vectors v, w, $\mu(v \otimes w)$ is a matrix with entries $\mu(v_i w_j)$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

REMARK 2.3. *Note that, setting* $\psi(\lambda,t)=\mathbb{E}(e^{i\langle\lambda,\zeta(t)\rangle})$, *equation* (2.3) *implies, by Itō's formula, that*

$$\partial_t\psi=\langle\lambda,D\bar\omega\partial_\lambda\psi\rangle-\frac{1}{2}\langle\lambda,\sigma^2\lambda\rangle\psi$$
$$\psi(\lambda,0)=1$$

which implies that ψ *is a zero mean Gaussian. In turn, this implies that* ζ *is a zero mean Gaussian process, see the proof of Proposition* 7.6 *for more details.*

REMARK 2.4. *It is interesting to notice that equation* (2.3) *with* $\sigma\equiv 0$ *is just the equation for the evolution of an infinitesimal displacement of the initial condition, that is the linearised equation along an orbit of the averaged deterministic system. This is rather natural, since in the time scale we are considering, the fluctuations around the deterministic trajectory are very small.*

REMARK 2.5. *Note that the condition that insures that the diffusion coefficient* σ *is non zero can be constructively checked by finding periodic orbits with different averages.*

Having stated our goals, let us begin with a first, very simple, result.

LEMMA 2.6. *The measures* $\{\mathbb{P}^\varepsilon\}$ *are tight.*

PROOF. By (2.1) it follows that the path Θ_ε is made of segments of length ε and maximal slope $\|\omega\|_{L^\infty}$, thus for all $h>0$,[13]

$$\|\Theta_\varepsilon(t+h)-\Theta_\varepsilon(t)\|\le C_\#h+\varepsilon\sum_{k=\lceil\varepsilon^{-1}t\rceil}^{\lfloor\varepsilon^{-1}(t+h)\rfloor-1}\|\omega(x_k,\theta_k)\|\le C_\#h.$$

Thus the measures $\mathbb{P}^\varepsilon$ are all supported on a set of uniformly Lipschitz functions, that is a compact set. □

The above means that there exist converging subsequences $\{\mathbb{P}^{\varepsilon_j}\}$. Our next step is to identify the set of accumulation points.

An obstacle that we face immediately is the impossibility of using some typical probabilistic tools. In particular, conditioning with respect to the past and Itō's formula. In fact, even if the initial condition is random, the dynamics is still deterministic, hence conditioning with respect to the past seems hopeless as it might kill all the randomness at later times.

To solve the first problem it is therefore necessary to devise a systematic way to use the strong dependence on the initial condition (typical of hyperbolic systems) to show that the dynamics, in some sense, *forgets the past*. One way of doing this effectively is to use standard pairs, introduced in the next section, whereby slightly enlarging our allowed initial conditions. Exactly how this solves the conditioning problem will be explained in Section 4. The lack of Itō's formula will be overcome by taking the point of view of the Martingale problem to define the solution of a SDE. To explain what this means in the present context is the goal of the present note, but see Appendix C for a brief comment on this issue in the simple case of an SDE. We will come back to the problem of studying the accumulation points of $\{\mathbb{P}^\varepsilon\}$ after having settled the issue of conditioning.

[13]The reader should be aware that we use the notation $C_\#$ to designate a generic constant (depending only on f and ω) which numerical value can change from one occurrence to the next, even in the same line.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

3. Standard Pairs

Let us fix $\delta > 0$ small enough, and $\mathfrak{D} > 0$ large enough, to be specified later; for $c_1 > 0$ consider the set of functions

$$\Sigma_{c_1} = \{G \in \mathcal{C}^2([a,b], S) \mid a, b \in \mathbb{T}^1, b-a \in [\delta/2, \delta], \\ \|G'\|_{\mathcal{C}^0} \leq \varepsilon c_1, \|G''\|_{\mathcal{C}^0} \leq \varepsilon \mathfrak{D} c_1, \}.$$

Let us associate to each $G \in \Sigma_{c_1}$ the map $\mathbb{G} \in \mathcal{C}^2([a,b], X)$ defined by $\mathbb{G}(x) = (x, G(x))$ whose image is a curve –the graph of G– which will be called a *standard curve*. For $c_2 > 0$ large enough, let us define the set of c_2-*standard* probability densities on the standard curve as

$$D_{c_2}(G) = \left\{ \rho \in \mathcal{C}^1([a,b], \mathbb{R}_+) \;\middle|\; \int_a^b \rho(x)dx = 1, \; \left\| \frac{\rho'}{\rho} \right\|_{\mathcal{C}^0} \leq c_2 \right\}.$$

A *standard pair* ℓ is given by $\ell = (\mathbb{G}, \rho)$ where $G \in \Sigma_{c_1}$ and $\rho \in D_{c_2}(G)$. Let $\overline{\mathfrak{L}}_\varepsilon$ be the collection of all standard pairs for a given $\varepsilon > 0$. A standard pair $\ell = (\mathbb{G}, \rho)$ induces a probability measure μ_ℓ on $X = \mathbb{T}^{d+1}$ defined as follows: for any continuous function g on X let

$$\mu_\ell(g) := \int_a^b g(x, G(x))\rho(x)dx.$$

We define[14] a *standard family* $\mathfrak{L} = (\mathcal{A}, \nu, \{\ell_j\}_{j\in\mathcal{A}})$, where $\mathcal{A} \subset \mathbb{N}$ and ν is a probability measure on $\mathcal{A}$; i.e. we associate to each standard pair ℓ_j a positive weight $\nu(\{j\})$ so that $\sum_{j\in\mathcal{A}} \nu(\{j\}) = 1$. For the following we will use also the notation $\nu_{\ell_j} = \nu(\{j\})$ for each $j \in \mathcal{A}$ and we will write $\ell \in \mathfrak{L}$ if $\ell = \ell_j$ for some $j \in \mathcal{A}$. A standard family $\mathfrak{L}$ naturally induces a probability measure $\mu_{\mathfrak{L}}$ on X defined as follows: for any measurable function g on X let

$$\mu_{\mathfrak{L}}(g) := \sum_{\ell\in\mathfrak{L}} \nu_\ell \mu_\ell(g).$$

Let us denote by $\sim$ the equivalence relation induced by the above correspondence i.e. we let $\mathfrak{L} \sim \mathfrak{L}'$ if and only if $\mu_{\mathfrak{L}} = \mu_{\mathfrak{L}'}$.

Proposition 3.1 (Invariance). *There exist δ and $\mathfrak{D}$ such that, for any c_1, c_2 sufficiently large, and ε sufficiently small, for any standard family $\mathfrak{L}$, the measure $F_{\varepsilon *}\mu_{\mathfrak{L}}$ can be decomposed in standard pairs, i.e. there exists a standard family $\mathfrak{L}'$ such that $F_{\varepsilon *}\mu_{\mathfrak{L}} = \mu_{\mathfrak{L}'}$. We say that $\mathfrak{L}'$ is a* standard decomposition of $F_{\varepsilon *}\mu_{\mathfrak{L}}$.

Proof. For simplicity, let us assume that $\mathfrak{L}$ is given by a single standard pair ℓ; the general case does not require any additional ideas and it is left to the reader. By definition, for any measurable function g:

$$F_{\varepsilon *}\mu_\ell(g) = \mu_\ell(g \circ F_\varepsilon) = \\ = \int_a^b g(f(x, G(x)), G(x) + \varepsilon\omega(x, G(x))) \cdot \rho(x)dx.$$

It is then natural to introduce the map $f_{\mathbb{G}} : [a,b] \to \mathbb{T}^1$ defined by $f_{\mathbb{G}}(x) = f \circ \mathbb{G}(x)$. Note that, by assumption (1.1), $f'_{\mathbb{G}} \geq \lambda - \varepsilon c_1 \|\partial_\theta f\|_{\mathcal{C}^0} > 3/2$ provided that ε is small enough (depending on how large is c_1). Hence all $f_{\mathbb{G}}$'s are expanding maps, moreover they are invertible if δ has been chosen small enough. In addition, for any

[14] This is not the most general definition of standard family, yet it suffices for our purposes.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

sufficiently smooth function A on X, it is trivial to check that, by the definition of standard curve, if ε is small enough (once again depending on c_1)[15]

$$\|(A\circ\mathbb{G})'\|_{\mathcal{C}^0}\le\|\mathrm{d}A\|_{\mathcal{C}^0}+\varepsilon\|\mathrm{d}A\|_{\mathcal{C}^0}c_1 \tag{3.1a}$$

$$\|(A\circ\mathbb{G})''\|_{\mathcal{C}^0}\le 2\|\mathrm{d}A\|_{\mathcal{C}^1}+\varepsilon\|\mathrm{d}A\|_{\mathcal{C}^0}\mathfrak{D}c_1. \tag{3.1b}$$

Then, fix a partition (mod 0) $[f_{\mathbb{G}}(a),f_{\mathbb{G}}(b)]=\bigcup_{j=1}^m[a_j,b_j]$, with $b_j-a_j\in[\delta/2,\delta]$ and $b_j=a_{j+1}$; moreover let $\varphi_j(x)=f_{\mathbb{G}}^{-1}(x)$ for $x\in[a_j,b_j]$ and define

$$G_j(x)=G\circ\varphi_j(x)+\varepsilon\omega(\varphi_j(x),G\circ\varphi_j(x));$$
$$\tilde\rho_j(x)=\rho\circ\varphi_j(x)\varphi_j'(x).$$

By a change of variables we can thus write:

$$F_{\varepsilon*}\mu_\ell(g)=\sum_{j=1}^m\int_{a_j}^{b_j}\tilde\rho_j(x)g(x,G_j(x))dx. \tag{3.2}$$

Observe that, by immediate differentiation we obtain, for φ_j:

$$\varphi_j'=\frac{1}{f_{\mathbb{G}}'}\circ\varphi_j\qquad\qquad\varphi_j''=-\frac{f_{\mathbb{G}}''}{f_{\mathbb{G}}'^3}\circ\varphi_j. \tag{3.3}$$

Let $\omega_{\mathbb{G}}=\omega\circ\mathbb{G}$ and $\bar G=G+\varepsilon\omega_{\mathbb{G}}$. Differentiating the definitions of G_j and $\tilde\rho_j$ and using (3.3) yields

$$G_j'=\frac{\bar G'}{f_{\mathbb{G}}'}\circ\varphi_j\qquad\qquad G_j''=\frac{\bar G''}{f_{\mathbb{G}}'^2}\circ\varphi_j-G_j'\cdot\frac{f_{\mathbb{G}}''}{f_{\mathbb{G}}'^2}\circ\varphi_j \tag{3.4}$$

and similarly

$$\frac{\tilde\rho_j'}{\tilde\rho_j}=\frac{\rho'}{\rho\cdot f_{\mathbb{G}}'}\circ\varphi_j-\frac{f_{\mathbb{G}}''}{f_{\mathbb{G}}'^2}\circ\varphi_j. \tag{3.5}$$

Using the above equations it is possible to conclude our proof: first of all, using (3.4), the definition of $\bar G$ and equations (3.1) we obtain, for small enough ε:

$$\begin{aligned}\|G_j'\|&\le\left\|\frac{G'+\varepsilon\omega_{\mathbb{G}}'}{f_{\mathbb{G}}'}\right\|\le\frac{2}{3}(1+C_\#\varepsilon)\varepsilon c_1+C_\#\varepsilon\le\\&\le\frac{3}{4}\varepsilon c_1+C_\#\varepsilon\le\varepsilon c_1,\end{aligned}$$

provided that c_1 is large enough; then:

$$\begin{aligned}\|G_j''\|&\le\left\|\frac{G''+\varepsilon\omega_{\mathbb{G}}''}{f_{\mathbb{G}}'^2}\right\|+C_\#(1+\varepsilon\mathfrak{D}c_1)\varepsilon c_1\le\\&\le\frac{3}{4}\varepsilon\mathfrak{D}c_1+\varepsilon C_\#c_1+\varepsilon C_\#\le\varepsilon\mathfrak{D}c_1\end{aligned}$$

provided c_1 and $\mathfrak{D}$ are sufficiently large. Likewise, using (3.1) together with (3.5) we obtain

$$\left\|\frac{\tilde\rho_j'}{\tilde\rho_j}\right\|\le\frac{2}{3}c_2+C_\#(1+\mathfrak{D}c_1)\le c_2,$$

provided that c_2 is large enough. This concludes our proof: it suffices to define the family $\mathfrak{L}'$ given by $(\mathcal{A},\nu,\{\ell_j\}_{j\in\mathcal{A}})$, where $A=\{1,\dots,m\}$, $\nu(\{j\})=\int_{a_j}^{b_j}\tilde\rho_j$,

[15]Given a function A by dA we mean the differential.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

$\rho_j = \nu(\{j\})^{-1}\tilde{\rho}_j$ and $\ell_j = (\mathbb{G}_j, \rho_j)$. Our previous estimates imply that $(\mathbb{G}_j, \rho_j)$ are standard pairs; note moreover that (3.2) implies $\sum_{\tilde{\ell}\in\mathfrak{L}'} \nu_{\tilde{\ell}} = 1$, thus $\mathfrak{L}'$ is a standard family. Then we can rewrite (3.2) as follows:

$$F_{\varepsilon *}\mu_\ell(g) = \sum_{\tilde{\ell}\in\mathfrak{L}'} \nu_{\tilde{\ell}}\mu_{\tilde{\ell}}(g) = \mu_{\mathfrak{L}'}(g). \qquad \square$$

REMARK 3.2. *Given a standard pair $\ell = (\mathbb{G}, \rho)$, we will interpret (x_k, θ_k) as random variables defined as $(x_k, \theta_k) = F_\varepsilon^k(x, G(x))$, where x is distributed according to ρ.*

4. Conditioning

In probability, conditioning is one of the most basic techniques and one would like to use it freely when dealing with random variables. Yet, as already mentioned, conditioning seems unnatural when dealing with deterministic systems. The use of standard pairs provides a very efficient solution to this conundrum. The basic idea is that one can apply repeatedly Proposition 3.1 to obtain at each time a family of standard pairs and then "condition" by specifying to which standard pair the random variable belongs at a given time.[16]

Note that if ℓ is a standard pair with $G' = 0$, then it belongs to $\overline{\mathfrak{L}}_\varepsilon$ for all $\varepsilon > 0$. In the following, abusing notations, we will use ℓ also to designate a family $\{\ell_\varepsilon\}$, $\ell_\varepsilon \in \overline{\mathfrak{L}}_\varepsilon$ that weakly converges to a standard pair $\ell \in \bigcap_{\varepsilon>0}\overline{\mathfrak{L}}_\varepsilon$. For every standard pair ℓ we let $\mathbb{P}_\ell^\varepsilon$ be the induced measure in path space and $\mathbb{E}_\ell^\varepsilon$ the associated expectation.

Before continuing, let us recall and state a bit of notation: for each $t \in [0, T]$ recall that the random variable $\Theta(t) \in \mathcal{C}^0(\mathcal{C}^0([0, T], S), S)$ is defined by $\Theta(t, \vartheta) = \vartheta(t)$, for all $\vartheta \in \mathcal{C}^0([0, T], S)$. Also we will need the filtration of σ-algebras $\mathcal{F}_t$ defined as the smallest σ-algebra for which all the functions $\{\Theta(s) \ : \ s \le t\}$ are measurable. Last, we consider the shift $\tau_s : \mathcal{C}^0([0, T], S) \to \mathcal{C}^0([0, T - s], S)$ defined by $\tau_s(\vartheta)(t) = \vartheta(t + s)$. Note that $\Theta(t) \circ \tau_s = \Theta(t + s)$. Also, it is helpful to keep in mind that, for all $A \in \mathcal{C}^0(S, \mathbb{R})$, we have[17]

$$\mathbb{E}_\ell^\varepsilon(A(\Theta(t + k\varepsilon))) = \mu_\ell(A(\Theta_\varepsilon(t + k\varepsilon))) = \mu_\ell(A(\Theta_\varepsilon(t) \circ F_\varepsilon^k)).$$

Our goal is to compute, in some reasonable way, expectations of a function of $\Theta(t + s)$ conditioned to $\mathcal{F}_t$, notwithstanding the above mentioned problems due to the fact that the dynamics is deterministic. Obviously, we can hope to obtain a result only in the limit $\varepsilon \to 0$. Note that we can always reduce to the case in which the conditional expectation is zero by subtracting an appropriate function, thus it suffices to analyze such a case.

The basic fact that we will use is the following.

LEMMA 4.1. *Let $t' \in [0, T]$ and $\mathcal{A}$ be a continuous bounded random variable on $\mathcal{C}^0([0, t'], S)$ with values in $\mathbb{R}$. If we have*

$$\lim_{\varepsilon\to 0} \sup_{\ell\in\overline{\mathfrak{L}}_\varepsilon} |\mathbb{E}_\ell^\varepsilon(\mathcal{A})| = 0,$$

[16] Note that the set of standard pairs does not form a σ-algebra, so to turn the above into a precise statement would be a bit cumbersome. We thus prefer to follow a slightly different strategy, although the substance is unchanged.

[17] To be really precise, maybe one should write, e.g., $\mathbb{E}_\ell^\varepsilon(A \circ \Theta(t + k\varepsilon))$, but we conform to the above more intuitive notation.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

then, for each $s \in [0, T-t']$, standard pair ℓ, uniformly bounded continuous functions $\{B_i\}_{i=1}^m$, $B_i : S \to \mathbb{R}$ and times $\{t_1 < \cdots < t_m\} \subset [0, s)$,

$$\lim_{\varepsilon\to 0} \mathbb{E}_\ell^\varepsilon \left(\prod_{i=1}^m B_i(\Theta(t_i)) \cdot \mathcal{A} \circ \tau_s \right) = 0.$$

Proof. The quantity we want to study can be written as

$$\mu_\ell \left(\prod_{i=1}^m B_i(\Theta_\varepsilon(t_i)) \cdot \mathcal{A}(\tau_s(\Theta_\varepsilon)) \right).$$

To simplify our notation, let $k_i = \lfloor t_i \varepsilon^{-1} \rfloor$ and $k_{m+1} = \lfloor s\varepsilon^{-1} \rfloor$. Also, for every standard pair $\tilde{\ell}$, let $\mathfrak{L}_{i,\tilde{\ell}}$ denote an arbitrary standard decomposition of $(F_\varepsilon^{k_{i+1}-k_i})_* \mu_{\tilde{\ell}}$ and define $\theta_\ell^* = \mu_\ell(\theta) = \int_{a_\ell}^{b_\ell} \rho_\ell(x) G_\ell(x) dx$. Then, by Proposition 3.1,

$$\mu_\ell \left(\prod_{i=1}^m B_i(\Theta_\varepsilon(t_i)) \cdot \mathcal{A}(\tau_s(\Theta_\varepsilon)) \right) = \mu_\ell \left(\prod_{i=1}^m B_i(\Theta_\varepsilon(t_i)) \cdot \mathcal{A}(\tau_{s-\varepsilon k_{m+1}}(\Theta_\varepsilon \circ F_\varepsilon^{k_{m+1}})) \right)$$

$$= \sum_{\ell_1 \in \mathfrak{L}_{1,\ell}} \cdots \sum_{\ell_{m+1} \in \mathfrak{L}_{m,\ell_m}} \left[\prod_{i=1}^m \nu_{\ell_i} B_i(\theta_{\ell_i}^*) \right] \nu_{m+1} \mu_{\ell_{m+1}}(\mathcal{A}(\Theta_\varepsilon)) + o(1)$$

$$= \sum_{\ell_1 \in \mathfrak{L}_{1,\ell}} \cdots \sum_{\ell_{m+1} \in \mathfrak{L}_{m,\ell_m}} \left[\prod_{i=1}^m \nu_{\ell_i} B_i(\theta_{\ell_i}^*) \right] \nu_{m+1} \mathbb{E}_{\ell_{m+1}}^\varepsilon(\mathcal{A}) + o(1)$$

where $\lim_{\varepsilon\to 0} o(1) = 0$. The lemma readily follows. □

Lemma 4.1 implies that, calling $\mathbb{P}$ an accumulation point of $\mathbb{P}_\ell^\varepsilon$, we have[18]

$$\mathbb{E} \left(\prod_{i=1}^m B_i(\Theta(t_i)) \cdot \mathcal{A} \circ \tau_s \right) = 0. \tag{4.1}$$

This solves the conditioning problems thanks to the following

Lemma 4.2. *Property* (4.1) *is equivalent to*

$$\mathbb{E}\left(\mathcal{A} \circ \tau_s \mid \mathcal{F}_{s_1}\right) = 0,$$

for all $s_1 \le s$.

Proof. Note that the statement of the Lemma immediately implies (4.1), we thus worry only about the other direction. If the lemma were not true then there would exist a positive measure set of the form

$$\mathcal{K} = \bigcap_{i=0}^{\infty} \{\vartheta(t_i) \in K_i\},$$

where the $\{K_i\}$ is a collection of compact sets in S, and $t_i < s$, on which the conditional expectation is strictly positive (or strictly negative, which can be treated in exactly the same way). For some arbitrary $\delta > 0$, consider open sets $U_i \supset K_i$ be

[18] By $\mathbb{E}$ we mean the expectation with respect to $\mathbb{P}$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

such that $\mathbb{P}(\{\vartheta(t_i) \in U_i \setminus K_i\}) \leq \delta 2^{-i}$. Also, let $B_{\delta,i}$ be a continuous function such that $B_{\delta,i}(\vartheta) = 1$ for $\vartheta \in K_i$ and $B_{\delta,i}(\vartheta) = 0$ for $\vartheta \notin U_i$. Then

$$0 < \mathbb{E}(\mathbb{1}_{\mathcal{K}} \mathcal{A} \circ \tau_s) = \lim_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^{n} B_{\delta,i}(\Theta(t_i)) \cdot \mathcal{A} \circ \tau_s\right) + C_{\#}\delta = C_{\#}\delta$$

which yields a contradiction by the arbitrariness of δ. □

In other words, we have recovered the possibility of conditioning with respect to the past *after* the limit $\varepsilon \to 0$.

5. Averaging (the Law of Large Numbers)

We are now ready to provide the proof of Theorem 2.1. The proof consists of several steps; we first illustrate the global strategy while momentarily postponing the proof of the single steps.

Proof of Theorem 2.1. As already mentioned we will prove the theorem for a larger class of initial conditions: any initial condition determined by a sequence of standard pairs. Note that for any fixed flat standard pair ℓ, i.e. $G_\ell(x) = \theta$, we have an initial condition assumed in the statement of the Theorem. Given a standard pair ℓ let $\{\mathbb{P}^\varepsilon_\ell\}$ be the associate measures in path space (the latter measures being determined, as explained at the beginning of Section 2, by the standard pair ℓ and (2.1)). We have already seen in Lemma 2.6 that the set $\{\mathbb{P}^\varepsilon_\ell\}$ is tight.

Next we will prove in Lemma 5.1 that, for each $A \in \mathcal{C}^2(S, \mathbb{R})$, we have

$$\lim_{\varepsilon\to 0} \sup_{\ell \in \overline{\mathfrak{L}}_\varepsilon} \left| \mathbb{E}^\varepsilon_\ell \left(A(\Theta(t)) - A(\Theta(0)) - \int_0^t \langle \overline{\omega}(\Theta(\tau)), \nabla A(\Theta(\tau))\rangle d\tau \right) \right| = 0. \tag{5.1}$$

Accordingly, it is natural to consider the random variables $\mathcal{A}(t)$ defined by

$$\mathcal{A}(t, \vartheta) = A(\vartheta(t)) - A(\vartheta(0)) - \int_0^t \langle \overline{\omega}(\vartheta(\tau)), \nabla A(\vartheta(\tau))\rangle d\tau,$$

for each $t \in [0, T]$ and $\vartheta \in \mathcal{C}^0([0, T], S)$, and the first order differential operator

$$\mathcal{L}A = \langle \overline{\omega}, \nabla A\rangle. \tag{5.2}$$

Then equation (5.1), together with Lemmata 4.1 and 4.2, means that each accumulation point $\mathbb{P}_\ell$ of $\{\mathbb{P}^\varepsilon_\ell\}$ satisfies, for all $s \in [0, T]$ and $t \in [0, T - s]$,

$$\mathbb{E}_\ell\left(\mathcal{A} \circ \tau_s \mid \mathcal{F}_s\right) = \mathbb{E}_\ell\left(A(\Theta(t+s)) - A(\Theta(s)) - \int_s^{t+s} \mathcal{L}A(\Theta(\tau)) d\tau \;\middle|\; \mathcal{F}_s \right) = 0 \tag{5.3}$$

this is the simplest possible version of the *Martingale Problem*. Indeed it implies that, for all θ, A and standard pair ℓ such that $G_\ell(x) = \theta$,

$$M(t) = A(\Theta(t)) - A(\Theta(0)) - \int_0^t \mathcal{L}A(\Theta(s)) ds$$

is a martingale with respect to the measure $\mathbb{P}_\theta$ and the filtration $\mathcal{F}_t$ (i.e., for each $0 \leq s \leq t \leq T$, $\mathbb{E}_\theta(M(t) \mid \mathcal{F}_s) = M(s)$).[19] Finally we will show in Lemma 5.2 that there is a unique measure that has such a property: the measure supported on the unique solution of equation (2.2). This concludes the proof of the theorem. □

In the rest of this section we provide the missing proofs.

[19] We use $\mathbb{P}_\theta$ to designate any measure $\mathbb{P}_\ell$ with $G_\ell(x) = \theta$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

5.1. Differentiating with respect to time.
Let us start with the proof of (5.1).

LEMMA 5.1. *For each $A \in \mathcal{C}^2(S,\mathbb{R})$ we have*[20]

$$\lim_{\varepsilon\to 0}\sup_{\ell\in\overline{\mathfrak{L}}_\varepsilon}\left|\mathbb{E}^\varepsilon_\ell\left(A(\Theta(t))-A(\Theta(0))-\int_0^t\langle\overline{\omega}(\Theta(s)),\nabla A(\Theta(s))\rangle ds\right)\right|=0,$$

where (we recall) $\overline{\omega}(\theta)=m_\theta(\omega(\cdot,\theta))$ and m_θ is the unique SRB measure of $f(\cdot,\theta)$.

PROOF. We will use the notation of Appendix B. Given a standard pair ℓ let $\rho_\ell=\rho$, $\theta^*_\ell=\mu_\ell(\theta)$ and $f_*(x)=f(x,\theta^*_\ell)$. Then, by Lemmata B.1 and B.2, we can write, for $n\le C\varepsilon^{-\frac12}$,[21]

$$\begin{aligned}
\mu_\ell\left(A(\theta_n)\right)&=\int_a^b\rho(x)A\left(\theta_0+\varepsilon\sum_{k=0}^{n-1}\omega(x_k,\theta_k)\right)dx\\
&=\int_a^b\rho(x)A\left(\theta^*_\ell+\varepsilon\sum_{k=0}^{n-1}\omega(x_k,\theta^*_\ell)\right)dx+\mathcal{O}(\varepsilon^2n^2+\varepsilon)\\
&=\int_a^b\rho(x)A(\theta^*_\ell)dx+\varepsilon\sum_{k=0}^{n-1}\int_a^b\rho(x)\langle\nabla A(\theta^*_\ell),\omega(x_k,\theta^*_\ell)\rangle dx+\mathcal{O}(\varepsilon)\\
&=\int_a^b\rho(x)A(G_\ell(x))dx+\mathcal{O}(\varepsilon)\\
&\quad+\varepsilon\sum_{k=0}^{n-1}\int_a^b\rho(x)\langle\nabla A(\theta^*_\ell),\omega(f_*^k\circ Y_n(x),\theta^*_\ell)\rangle dx\\
&=\mu_\ell(A(\theta_0))+\varepsilon\sum_{k=0}^{n-1}\int_{\mathbb{T}^1}\tilde\rho_n(x)\langle\nabla A(\theta^*_\ell),\omega(f_*^k(x),\theta^*_\ell)\rangle dx+\mathcal{O}(\varepsilon)
\end{aligned}$$

where $\tilde\rho_n(x)=\left[\frac{\chi_{[a,b]}\rho}{Y_n'}\right]\circ Y_n^{-1}(x)$. Note that $\int_{\mathbb{T}^1}\tilde\rho_n=1$ but, unfortunately, $\|\tilde\rho\|_{BV}$ may be enormous. Thus, we cannot estimate the integral in the above expression by naively using decay of correlations. Yet, equation (B.3) implies $|Y_n'-1|\le C_\#\varepsilon n^2$. Moreover, $\bar\rho=(\chi_{[a,b]}\rho)\circ Y^{-1}$ has uniformly bounded variation.[22] Accordingly, by the decay of correlations and the $\mathcal{C}^1$ dependence of the invariant measure on θ (see Section 1.2) we have

$$\begin{aligned}
\int_{\mathbb{T}^1}\tilde\rho_n(x)\langle\nabla A(\theta^*_\ell),\omega(f_*^k(x),\theta^*_\ell)\rangle dx&=\int_{\mathbb{T}^1}\bar\rho_n(x)\langle\nabla A(\theta^*_\ell),\omega(f_*^k(x),\theta^*_\ell)\rangle dx+\mathcal{O}(\varepsilon n^2)\\
&=m_{\mathrm{Leb}}(\tilde\rho_n(x))m_{\theta^*_\ell}\left(\langle\nabla A(\theta^*_\ell),\omega(\cdot,\theta^*_\ell)\rangle\right)+\mathcal{O}(\varepsilon n^2+e^{-c_\#k})\\
&=\mu_\ell\left(\langle\nabla A(\theta_0),\overline{\omega}(\theta_0)\rangle\right)+\mathcal{O}(\varepsilon n^2+e^{-c_\#k}).
\end{aligned}$$

Hence,

$$\mu_\ell\left(A(\theta_n)\right)=\mu_\ell(A(\theta_0)+\varepsilon n\langle\nabla A(\theta_0),\bar\omega(\theta_0)\rangle)+\mathcal{O}(n^3\varepsilon^2+\varepsilon).\tag{5.4}$$

[20] See Lemma C.1 for the relation between the differentiability mentioned in the title of the subsection and the present integral formula.

[21] By $\mathcal{O}(\varepsilon^a n^b)$ we mean a quantity bounded by $C_\#\varepsilon^a n^b$, where $C_\#$ does not depend on ℓ.

[22] Indeed, for all $\varphi\in\mathcal{C}^1$, $|\varphi|_\infty\le 1$, $\int\bar\rho\varphi'=\int_a^b\rho\cdot\varphi'\circ Y\cdot Y'=\int_a^b\rho(\varphi\circ Y)'\le\|\rho\|_{BV}$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Finally, we choose $n = \lceil \varepsilon^{-\frac{1}{3}} \rceil$ and set $h = \varepsilon n$. We define inductively standard families such that $\mathfrak{L}_{\ell_0} = \{\ell\}$ and for each standard pair $\ell_{i+1} \in \mathfrak{L}_{\ell_i}$ the family $\mathfrak{L}_{\ell_{i+1}}$ is a standard decomposition of the measure $(F_\varepsilon^n)^* \mu_{\ell_{i+1}}$. Thus, setting $m = \lceil t\varepsilon^{-\frac{2}{3}} \rceil - 1$, recalling equation (5.4) and using repeatedly Proposition 3.1,

$$\mathbb{E}_\ell^\varepsilon(A(\Theta(t))) = \mu_\ell(A(\theta_{t\varepsilon^{-1}})) = \mu_\ell(A(\theta_0)) + \sum_{k=0}^{m-1} \mu_\ell(A(\theta_{\varepsilon^{-1}(k+1)h})) - A(\theta_{\varepsilon^{-1}kh}))$$

$$= \mu_\ell(A(\theta_0)) + \sum_{k=0}^{m-1} \sum_{\ell_1 \in \mathfrak{L}_{\ell_0}} \cdots \sum_{\ell_{k-1} \in \mathfrak{L}_{\ell_{k-2}}} \prod_{j=1}^{k-1} \nu_{\ell_j} \left[\mu_{\ell_{k-1}}(\varepsilon^{\frac{2}{3}} \langle \nabla A(\theta_0), \bar{\omega}(\theta_0) \rangle) + \mathcal{O}(\varepsilon) \right]$$

$$= \mathbb{E}_\ell^\varepsilon \left(A(\Theta(0)) + \sum_{k=0}^{m-1} \langle \nabla A(\Theta(kh)), \bar{\omega}(\Theta(kh)) \rangle h \right) + \mathcal{O}(\varepsilon^{\frac{1}{3}} t)$$

$$= \mathbb{E}_\ell^\varepsilon \left(A(\Theta(0)) + \int_0^t \langle \nabla A(\Theta(s)), \bar{\omega}(\Theta(s)) \rangle ds \right) + \mathcal{O}(\varepsilon^{\frac{1}{3}} t).$$

The lemma follows by taking the limit $\varepsilon \to 0$. □

5.2. The Martingale Problem at work.

First of all let us finally specify precisely what we mean by the *martingale problem.*

Definition 1 (Martingale Problem). *Given a Riemannian manifold S, a linear operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \subset \mathcal{C}^0(S, \mathbb{R}^d) \to \mathcal{C}^0(S, \mathbb{R}^d)$, a set of measures $\mathbb{P}_y$, $y \in S$, on $\mathcal{C}^0([0,T], S)$ and a filtration $\mathcal{F}_t$ we say that $\{\mathbb{P}_y\}$ satisfies the martingale problem if for each function $A \in \mathcal{D}(\mathcal{L})$,*

$$\mathbb{P}_y(\{z(0) = y\}) = 1$$

$$M(t,z) := A(z(t)) - A(z(0)) - \int_0^t \mathcal{L}A(z(s))ds \text{ is } \mathcal{F}_t\text{-martingale under all } \mathbb{P}_y.$$

We can now prove the last announced result.

Lemma 5.2. *If $\bar{\omega}$ is Lipschitz, then the martingale problem determined by* (5.2) *has a unique solution consisting of the measures supported on the solutions of the ODE*

$$\begin{aligned} \dot{\overline{\Theta}} &= \overline{\omega}(\overline{\Theta}) \\ \overline{\Theta}(0) &= y. \end{aligned} \tag{5.5}$$

Proof. Let $\overline{\Theta}$ be the solution of (5.5) with initial condition $y \in \mathbb{T}^d$ and $\mathbb{P}_y$ the probability measure in the martingale problem. The idea is to compute

$$\begin{aligned} \frac{d}{dt}\mathbb{E}_y(\|\Theta(t) - \overline{\Theta}(t)\|^2) =& \frac{d}{dt}\mathbb{E}_y(\langle \Theta(t), \Theta(t) \rangle) - 2\langle \bar{\omega}(\overline{\Theta}(t)), \mathbb{E}_y(\Theta(t)) \rangle \\ & - 2\langle \overline{\Theta}(t), \frac{d}{dt}\mathbb{E}_y(\Theta(t)) \rangle + 2\langle \bar{\omega}(\overline{\Theta}(t)), \overline{\Theta}(t) \rangle. \end{aligned}$$

To continue we use Lemma C.1 where, in the first term $A(\theta) = \|\theta\|^2$, in the third $A(\theta) = \theta_i$ and the generator in (5.3) is given by $\mathcal{L}A(\theta) = \langle \nabla A(\theta), \bar{\omega}(\theta) \rangle$.

$$\begin{aligned} \frac{d}{dt}\mathbb{E}_y(\|\Theta(t) - \overline{\Theta}(t)\|^2) &= 2\mathbb{E}_y(\langle \Theta(t), \bar{\omega}(\Theta(t)) \rangle) - 2\langle \bar{\omega}(\overline{\Theta}(t)), \mathbb{E}_y(\Theta(t)) \rangle \\ &\quad - 2\langle \overline{\Theta}(t), \mathbb{E}_y(\bar{\omega}(\Theta(t))) \rangle + 2\mathbb{E}_y(\langle \overline{\Theta}(t), \bar{\omega}(\overline{\Theta}(t)) \rangle) \\ &= \mathbb{E}_y(\langle \Theta(t) - \overline{\Theta}(t), \overline{\omega}(\Theta(t)) - \overline{\omega}(\overline{\Theta}(t)) \rangle). \end{aligned}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

By the Lipschitz property of $\bar{\omega}$ (let C_L be the Lipschitz constant), using the Schwartz inequality and integrating we have

$$\mathbb{E}_y(\|\Theta(t)-\overline{\Theta}(t)\|^2) \le 2C_L \int_0^t \mathbb{E}_y(\|\Theta(s)-\overline{\Theta}(s)\|^2)ds$$

which, by Gronwall's inequality, implies that

$$\mathbb{P}_y(\{\overline{\Theta}\}) = 1. \qquad \square$$

6. A recap of what we have done so far

We have just seen that the martingale method (in Dolgopyat's version) consists of four steps

(1) Identify a suitable class of measures on path space which allow one to handle the conditioning problem (in our case: the one coming from standard pairs)
(2) Prove tightness for such measures (in our case: they are supported on uniformly Lipschitz functions)
(3) Identify an equation characterizing the accumulation points (in our case: an ODE)
(4) Prove uniqueness of the limit equation in the martingale sense.

The beauty of the previous scheme is that it can be easily adapted to a variety of problems. To convince the reader of this fact we proceed further and apply it to obtain more refined information on the behavior of the system.

7. Fluctuations (the Central Limit Theorem)

It is possible to study the limit behavior of ζ_ε using the strategy summarized in Section 6, even though now the story becomes technically more involved. Let us discuss the situation a bit more in detail. Let $\widetilde{\mathbb{P}}^\varepsilon_\ell$ be the path measure describing ζ_ε when (x_0,θ_0) are distributed according to the standard pair ℓ.[23] Note that, $\widetilde{\mathbb{P}}^\varepsilon_\ell = (\zeta_\varepsilon)_*\mu_\ell$. Again, we provide a proof of the claimed results based on some facts that will be proven in later sections.

Proof of Theorem 2.2. First of all, the sequence of measures $\widetilde{\mathbb{P}}^\varepsilon_\ell$ is tight, which will be proven in Proposition 7.1. Next, by Proposition 7.4, we have that

$$\lim_{\varepsilon\to 0}\sup_{\ell\in\overline{\mathfrak{L}}_\varepsilon}\left|\widetilde{\mathbb{E}}^\varepsilon_\ell\left(A(\zeta(t))-A(\zeta(0))-\int_0^t \mathcal{L}_s A(\zeta(s))ds\right)\right| = 0, \tag{7.1}$$

where

$$(\mathcal{L}_s A)(\zeta) = \langle \nabla A(\zeta), D\bar{\omega}(\overline{\Theta}(s))\zeta\rangle + \frac{1}{2}\sum_{i,j=1}^{d}[\sigma^2(\overline{\Theta}(s))]_{i,j}\partial_{\zeta_i}\partial_{\zeta_j}A(\zeta), \tag{7.2}$$

with diffusion coefficient σ^2 given by (2.4). In the following we will often write, slightly abusing notations, $\sigma(t)$ for $\sigma(\overline{\Theta}(t))$.

[23] As already explained, here we allow ℓ to stand also for a family $\{\ell_\varepsilon\}$ which weakly converges to ℓ. In particular, this means that $\overline{\Theta}$ is also a random variable, as it depends on the initial condition θ_0.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

We can then use equation (7.1) and Lemma 4.1 followed by Lemma 4.2 to obtain that

$$A(\zeta(t)) - A(\zeta(0)) - \int_0^t \mathcal{L}_s A(\zeta(s))ds$$

is a martingale under any accumulation point $\widetilde{\mathbb{P}}$ of the measures $\widetilde{\mathbb{P}}^\varepsilon_\ell$ with respect to the filtration $\mathcal{F}_t$ with $\widetilde{\mathbb{P}}(\{\zeta(0) = 0\}) = 1$. In Proposition 7.6 we will prove that such a problem has a unique solution thereby showing that $\lim_{\varepsilon\to 0} \widetilde{\mathbb{P}}^\varepsilon_\ell = \widetilde{\mathbb{P}}$.

Note that the time dependent operator $\mathcal{L}_s$ is a second order operator, this means that the accumulation points of ζ_ε do not satisfy a deterministic equation, but rather a stochastic one. Indeed our last task is to show that $\widetilde{\mathbb{P}}$ is equal in law to the measure determined by the stochastic differential equation

$$\begin{aligned} d\zeta &= D\bar{\omega} \circ \overline{\Theta}(t)\, \zeta dt + \sigma dB \\ \zeta(0) &= 0 \end{aligned} \tag{7.3}$$

where B is a standard $\mathbb{R}^d$ dimensional Brownian motion. Note that the above equation is well defined in consequence of Lemma 7.5 which shows that the matrix σ^2 is symmetric and non negative, hence $\sigma = \sigma^T$ is well defined and strictly positive if $\hat{\omega}$ is not a coboundary (see Lemma 7.5). To conclude it suffices to show that the probability measure describing the solution of (7.3) satisfies the martingale problem.[24] It follows from Itō's calculus, indeed if ζ is the solution of (7.3) and $A \in \mathcal{C}^r$, then Itō's formula reads

$$dA(\zeta) = \sum_i \partial_{\zeta_i} A(\zeta) d\zeta_i + \frac{1}{2} \sum_{i,j,k} \partial_{\zeta_i}\partial_{\zeta_j} A(\zeta)\sigma_{ik}\sigma_{jk} dt.$$

Integrating it from s to t and taking the conditional expectation we have

$$\mathbb{E}\left(A(\zeta(t)) - A(\zeta(s)) - \int_s^t \mathcal{L}_\tau A(\zeta(\tau))d\tau \;\middle|\; \mathcal{F}_s \right) = 0.$$

See Appendix C for more details on the relation between the Martingale problem and the theory of SDE and how this allows to dispense form Itō's formula altogether.

We have thus seen that the measure determined by (7.3) satisfies the martingale problem, hence it must agree with $\widetilde{\mathbb{P}}$ since $\widetilde{\mathbb{P}}$ is the unique solution of the martingale problem. The proof of the Theorem is concluded by noticing that (7.3) defines a zero mean Gaussian process (see the end of the proof of Proposition 7.6). □

7.1. Tightness.

PROPOSITION 7.1. *For every standard pair ℓ, the measures $\{\widetilde{\mathbb{P}}^\varepsilon_\ell\}_{\varepsilon>0}$ are tight.*

PROOF. Now the proof of tightness is less obvious since the paths have a Lipschitz constant that explodes. Luckily, there exists a convenient criterion for tightness: Kolmogorov criterion [**15**, Remark A.5].

THEOREM 7.2 (Kolmogorov). *Given a sequence of measures $\mathbb{P}^\varepsilon$ on $\mathcal{C}^0([0,T],\mathbb{R})$, if there exists $\alpha, \beta, C > 0$ such that*

$$\mathbb{E}^\varepsilon(|z(t) - z(s)|^\beta) \le C|t-s|^{1+\alpha}$$

for all $t, s \in [0,T]$ and the distribution of $z(0)$ is tight, then $\{\mathbb{P}^\varepsilon\}$ is tight.

[24] We do not prove that such a solution exists as this is a standard result in probability, [**15**].

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Note that $\zeta_\varepsilon(0) = 0$. Of course, it is easier to apply the above criteria with $\beta \in \mathbb{N}$. It is reasonable to expect that the fluctuations behave like a Brownian motion, so the variance should be finite. To verify this let us compute first the case $\beta = 2$. Note that, setting $\hat\omega(x,\theta) = \omega(x,\theta) - \bar\omega(\theta)$,

$$\begin{aligned}\zeta_\varepsilon(t) &= \sqrt{\varepsilon}\sum_{k=0}^{\lceil\varepsilon^{-1}t\rceil-1}\left[\omega(x_k,\theta_k) - \bar\omega(\overline{\Theta}(\varepsilon k))\right] + \mathcal{O}(\sqrt{\varepsilon})\\ &= \sqrt{\varepsilon}\sum_{k=0}^{\lceil\varepsilon^{-1}t\rceil-1}\left[\hat\omega(x_k,\theta_k) + \bar\omega(\theta_k) - \bar\omega(\overline{\Theta}(\varepsilon k))\right] + \mathcal{O}(\sqrt{\varepsilon})\\ &= \sqrt{\varepsilon}\sum_{k=0}^{\lceil\varepsilon^{-1}t\rceil-1}\left[\hat\omega(x_k,\theta_k) + \sqrt{\varepsilon}D\bar\omega(\overline{\Theta}(\varepsilon k))\zeta_\varepsilon(k\varepsilon)\right] + \mathcal{O}(\sqrt{\varepsilon})\\ &\quad + \sum_{k=0}^{\lceil\varepsilon^{-1}t\rceil-1}\mathcal{O}(\varepsilon^{\frac{3}{2}}\|\zeta_\varepsilon(\varepsilon k)\|^2).\end{aligned} \tag{7.4}$$

We start with a basic result.

LEMMA 7.3. *For each standard pair ℓ and $k, l \in \{0,\dots,\varepsilon^{-1}\}$, $k \geq l$, we have*

$$\mu_\ell\left(\left\|\sum_{j=l}^{k}\hat\omega(x_j,\theta_j)\right\|^2\right) \leq C_\#(k-l).$$

The proof of the above Lemma is postponed to the end of the section. Let us see how it can be profitably used. Note that, for $t = \varepsilon k, s = \varepsilon l$,

$$\widetilde{\mathbb{E}}^\varepsilon_\ell(\|\zeta(t)-\zeta(s)\|^2) \leq C_\#|t-s| + C_\#|t-s|\varepsilon\sum_{j=l}^{k}\mu_\ell(\|\zeta_\varepsilon(\varepsilon j)\|^2) + C_\#\varepsilon, \tag{7.5}$$

where we have used Lemma 7.3 and the trivial estimate $\|\zeta_\varepsilon\| \leq C_\#\varepsilon^{-\frac{1}{2}}$. If we use the above with $s = 0$ and define $M_k = \sup_{j\leq k}\mu_\ell(|\zeta_\varepsilon(\varepsilon j)|^2)$ we have

$$M_k \leq C_\#\varepsilon k + C_\#k^2\varepsilon^2 M_k.$$

Thus there exists $C > 0$ such that, if $k \leq C\varepsilon^{-1}$, we have $M_k \leq C_\#\varepsilon k$. Hence, we can substitute such an estimate in (7.5) and obtain

$$\widetilde{\mathbb{E}}^\varepsilon_\ell(\|\zeta(t)-\zeta(s)\|^2) \leq C_\#|t-s| + C_\#\varepsilon. \tag{7.6}$$

Since the estimate for $|t-s| \leq C_\#\varepsilon$ is trivial, we have the bound,

$$\widetilde{\mathbb{E}}^\varepsilon_\ell(\|\zeta(t)-\zeta(s)\|^2) \leq C_\#|t-s|.$$

This is interesting but, unfortunately, it does not suffice to apply the Kolmogorov criteria. The next step could be to compute for $\beta = 3$. This has the well known disadvantage of being an odd function of the path, and hence one has to deal with the absolute value. Due to this, it turns out to be more convenient to consider directly the case $\beta = 4$. This can be done in complete analogy with the above computation, by first generalizing the result of Lemma 7.3 to higher momenta. Doing so we obtain

$$\widetilde{\mathbb{E}}^\varepsilon_\ell(\|\zeta(t)-\zeta(s)\|^4) \leq C_\#|t-s|^2, \tag{7.7}$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

which concludes the proof of the proposition. Indeed, the proof of Lemma 7.3 explains how to treat correlations. Multiple correlations can be treated similarly and one can thus show that they do not contribute to the leading term. Thus the computation becomes similar (although much more involved) to the case of the sum independent zero mean random variables X_i (where no correlations are present), that is

$$\mathbb{E}([\sum_{i=l}^{k} X_i]^4) = \sum_{i_1,\dots,i_4=l}^{k} \mathbb{E}(X_{i_1}X_{i_2}X_{i_2}X_{i_4}) = \sum_{i,j=l}^{k} \mathbb{E}(X_i^2 X_j^2) = \mathcal{O}((k-l)^2).$$

For future use let us record that, by equation (7.7) and the Young inequality,

$$\widetilde{\mathbb{E}}_\ell^\varepsilon(\|\zeta(t)-\zeta(s)\|^3) \le C_\#|t-s|^{\frac{3}{2}}. \tag{7.8}$$ □

We still owe the reader the

Proof of Lemma 7.3. The proof starts with a direct computation:[25]

$$\begin{aligned}\mu_\ell\left(\left|\sum_{j=l}^{k}\hat\omega(x_j,\theta_j)\right|^2\right) &\le \sum_{j=l}^{k}\mu_\ell\left(\hat\omega(x_j,\theta_j)^2\right)\\ &\quad + 2\sum_{j=l}^{k}\sum_{r=l+1}^{k}\mu_\ell\left(\hat\omega(x_j,\theta_j)\hat\omega(x_r,\theta_r)\right)\\ &\le C_\#|k-l| + 2\sum_{j=l}^{k}\sum_{r=j+1}^{k}\mu_\ell\left(\hat\omega(x_j,\theta_j)\hat\omega(x_r,\theta_r)\right).\end{aligned}$$

To compute the last correlation, remember Proposition 3.1. We can thus call $\mathfrak{L}_j$ the standard family associated to $(F_\varepsilon^j)_*\mu_\ell$ and, for $r \ge j$, we write

$$\begin{aligned}\mu_\ell\left(\hat\omega(x_j,\theta_j)\hat\omega(x_r,\theta_r)\right) &= \sum_{\ell_1\in\mathfrak{L}_j}\nu_{\ell_1}\mu_{\ell_1}\left(\hat\omega(x_0,\theta_0)\hat\omega(x_{r-j},\theta_{r-j})\right)\\ &= \sum_{\ell_1\in\mathfrak{L}_j}\nu_{\ell_1}\int_{a_{\ell_1}}^{b_{\ell_1}}\rho_{\ell_1}(x)\hat\omega(x,G_{\ell_1}(x))\hat\omega(x_{r-j},\theta_{r-j}).\end{aligned}$$

We would like to argue as in the proof of Lemma 5.1 and try to reduce the problem to

$$\begin{aligned}\int_{a_{\ell_1}}^{b_{\ell_1}}\rho_{\ell_1}(x)\hat\omega(x,\theta^*_{\ell_1})\hat\omega(x_{r-j},\theta^*_{\ell_1}) &= \int_{a_{\ell_1}}^{b_{\ell_1}}\rho_{\ell_1}(x)\hat\omega(x,\theta^*_{\ell_1})\hat\omega(f_*^{r-j}(Y_{r-j}(x)),\theta^*_{\ell_1})\\ &= \int_{\mathbb{T}^1}\tilde\rho(x)\hat\omega(Y_{r-j}^{-1}(x),\theta^*_{\ell_1})\hat\omega(f_*^{r-j}(x),\theta^*_{\ell_1}),\end{aligned}$$

but then the mistake that we would make substituting $\tilde\rho$ with $\bar\rho$ is too big for our current purposes. It is thus necessary to be more subtle. The idea is to write $\rho_{\ell_1}(x)\hat\omega(x,G_{\ell_1}(x)) = \alpha_1\hat\rho_1(x)+\alpha_2\hat\rho_2(x)$, where $\hat\rho_1,\hat\rho_2$ are standard densities.[26] Note that α_1,α_2 are uniformly bounded. Next, let us fix $L>0$ to be chosen later and

[25] To simplify notation we do the computation in the case $d=1$, the general case is identical.

[26] In fact, it would be more convenient to define standard pairs with signed (actually even complex) measures, but let us keep it simple.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

assume $r-j\geq L$. Since $\ell_{1,i}=(G,\hat\rho_i)$ are standard pairs, by construction, calling $\mathfrak{L}_{\ell_{1,i}}=(F^{r-j-L})_*\mu_{\ell_{1,i}}$ we have

$$\begin{aligned}\int_{a_{\ell_1}}^{b_{\ell_1}}\hat\rho_i(x)\hat\omega(x_{r-j},\theta_{r-j})&=\sum_{\ell_2\in\mathfrak{L}_{\ell_{1,i}}}\nu_{\ell_2}\int_{a_{\ell_2}}^{b_{\ell_2}}\rho_{\ell_2}(x)\hat\omega(f_*^L(Y_L(x)),\theta^*_{\ell_2})+\mathcal{O}(\varepsilon L)\\&=\sum_{\ell_2\in\mathfrak{L}_{\ell_{1,i}}}\nu_{\ell_2}\int_{\mathbb{T}^1}\tilde\rho(x)\hat\omega(f_*^L(x),\theta^*_{\ell_2})+\mathcal{O}(\varepsilon L)\\&=\sum_{\ell_2\in\mathfrak{L}_{\ell_{1,i}}}\nu_{\ell_2}\int_{\mathbb{T}^1}\bar\rho(x)\hat\omega(f_*^L(x),\theta^*_{\ell_2})+\mathcal{O}(\varepsilon L^2)\\&=\mathcal{O}(e^{-c_\#L}+\varepsilon L^2),\end{aligned}$$

due to the decay of correlations for the map f_* and the fact that $\hat\omega(\cdot,\theta^*_{\ell_2})$ is a zero average function for the invariant measure of f_*. By the above we have

$$\mu_\ell\left(\left|\sum_{j=l}^k\hat\omega(x_j,\theta_j)\right|^2\right)\leq C_\#\sum_{j=l}^k\left\{[e^{-c_\#L}+\varepsilon L^2](k-j)+1+\varepsilon L^3\right\}$$

which yields the result by choosing $L=c\log(k-j)$ for c large enough. □

7.2. Differentiating with respect to time (poor man's Itō's formula).

PROPOSITION 7.4. *For every standard pair ℓ and $A\in\mathcal{C}^3(S,\mathbb{R})$ we have*

$$\lim_{\varepsilon\to0}\sup_{\ell\in\overline{\mathfrak{L}}_\varepsilon}\left|\widetilde{\mathbb{E}}^\varepsilon_\ell\left(A(\zeta(t))-A(\zeta(0))-\int_0^t\mathcal{L}_sA(\zeta(s))ds\right)\right|=0.$$

PROOF. As in Lemma 5.1, the idea is to fix $h\in(0,1)$ to be chosen later, and compute

$$\begin{aligned}\widetilde{\mathbb{E}}^\varepsilon_\ell(A(\zeta(t+h))-A(\zeta(t)))&=\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle\nabla A(\zeta(t)),\zeta(t+h)-\zeta(t)\rangle)\\&+\widetilde{\mathbb{E}}^\varepsilon_\ell(\frac12\langle(D^2A)(\zeta(t))(\zeta(t+h)-\zeta(t)),\zeta(t+h)-\zeta(t)\rangle)+\mathcal{O}(h^{\frac32}),\end{aligned}\tag{7.9}$$

where we have used (7.8). Unfortunately this time the computation is a bit lengthy and rather boring, yet it basically does not contain any new idea, it is just a brute force computation.

Let us start computing the last term of (7.9). Setting $\zeta^h(t)=\zeta(t+h)-\zeta(t)$ and $\Omega^h=\sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}}\hat\omega(x_k,\theta_k)$, by equations (7.4) and using the trivial estimate

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

$\|\zeta_\varepsilon(t)\| \le C_\# \varepsilon^{-\frac{1}{2}}$, we have

$$\begin{aligned}\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle (D^2A)(\zeta(t))\zeta^h(t), \zeta^h(t)\rangle) &= \varepsilon \sum_{k,j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle (D^2A)(\zeta_\varepsilon(t))\hat\omega(x_k,\theta_k), \hat\omega(x_j,\theta_j)\rangle)\\
&+ \mathcal{O}\left(\varepsilon^{\frac{3}{2}} \sum_{j=\varepsilon^{-1}t}^{(t+h)\varepsilon^{-1}} \mu_\ell\left(\left\|\Omega^h\right\| \, \|\zeta_\varepsilon(j\varepsilon)\|\right)\right) + \mathcal{O}\left(\varepsilon \mu_\ell(\|\Omega^h\|)\right)\\
&+ \mathcal{O}\left(\varepsilon^2 \sum_{j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\Omega^h\| \, \|\zeta_\varepsilon(j\varepsilon)\|^2) + \varepsilon^2 \sum_{k,j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\| \, \|\zeta_\varepsilon(j\varepsilon)\|)\right)\\
&+ \mathcal{O}\left(\varepsilon^{\frac{3}{2}} \sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\|) + \varepsilon^{\frac{5}{2}} \sum_{k,j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\| \, \|\zeta_\varepsilon(j\varepsilon)\|^2) + \varepsilon\right)\\
&+ \mathcal{O}\left(\varepsilon^2 \sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\|^2) + \varepsilon^3 \sum_{k,j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\|^2 \, \|\zeta_\varepsilon(j\varepsilon)\|^2)\right).\end{aligned}$$

Observe that (7.6), (7.8) and (7.7) yield

$$\mu_\ell(\|\zeta_\varepsilon(k\varepsilon)\|^m) = \mu_\ell(\|\zeta_\varepsilon(k\varepsilon) - \zeta_\varepsilon(0)\|^m) \le C_\# (\varepsilon k)^{\frac{m}{2}} \le C_\#$$

for $m \in \{1,2,3,4\}$ and $k \le C_\# \varepsilon^{-1}$. We can now use Lemma 7.3 together with Schwartz inequality to obtain

(7.10)
$$\begin{aligned}\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle (D^2A)(\zeta(t))\zeta^h(t), \zeta^h(t)\rangle) =& \varepsilon \sum_{k,j=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle (D^2A)(\zeta_\varepsilon(t))\hat\omega(x_k,\theta_k), \hat\omega(x_j,\theta_j)\rangle)\\
&+ \mathcal{O}(\sqrt{\varepsilon h} + h^2 + \varepsilon).\end{aligned}$$

Next, we must perform a similar analysis on the first term of equation (7.9).

(7.11)
$$\begin{aligned}\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle \nabla A(\zeta(t)), \zeta^h(t)\rangle) &= \sqrt{\varepsilon} \sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), \hat\omega(x_k,\theta_k)\rangle)\\
&+ \varepsilon \sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(\varepsilon k))\zeta_\varepsilon(\varepsilon k)\rangle) + \mathcal{O}(\sqrt{\varepsilon}).\end{aligned}$$

To estimate the term in the second line of (7.11) we have to use again (7.4):

$$\begin{aligned}\sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(\varepsilon k))\zeta_\varepsilon(\varepsilon k)\rangle) &= h\varepsilon^{-1} \mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(t))\zeta_\varepsilon(t)\rangle)\\
&+ \mathcal{O}(\varepsilon^{-1}h^2 + \varepsilon^{-\frac{1}{2}}h) + \sqrt{\varepsilon} \sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \sum_{j=t\varepsilon^{-1}}^{k} \mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(t))\hat\omega(x_j,\theta_j)\rangle).\end{aligned}$$

To compute the last term in the above equation let $\mathfrak{L}_\ell$ be the standard family generated by ℓ at time $\varepsilon^{-1}t$, then, setting $\alpha_\varepsilon(\theta,t) = \nabla A(\varepsilon^{-\frac{1}{2}}(\theta - \overline{\Theta}(t)))$ and $\hat j =$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

$j - t\varepsilon^{-1}$, we can write

$$\mu_\ell(\langle \nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(t))\hat\omega(x_j,\theta_j)\rangle)$$
$$= \sum_{\ell_1\in\mathfrak{L}_\ell}\sum_{r,s=1}^{d} \nu_{\ell_1}\mu_{\ell_1}(\alpha_\varepsilon(\theta_0,t)_r D\bar\omega(\overline{\Theta}(t))_{r,s}\hat\omega(x_{\hat\jmath},\theta_{\hat\jmath})_s).$$

Next, notice that for every r, the signed measure $\mu_{\ell_1,r}(\phi) = \mu_{\ell_1}(\alpha_\varepsilon(\theta_0,t)_r\phi)$ has density $\rho_{\ell_1}\alpha_\varepsilon(G_{\ell_1}(x),t)_r$ whose derivative is uniformly bounded in ε, t. We can then write $\mu_{\ell_1,r}$ as a linear combination of two standard pairs $\ell_{1,i}$. Finally, given $L\in\mathbb{N}$, if $\hat\jmath \geq L$, we can consider the standard families $\mathcal{L}_{\ell_{1,i}}$ generated by $\ell_{1,i}$ at time $\hat\jmath - L$ and write, arguing as in the proof of Lemma 7.3,

$$\begin{aligned}\mu_{\ell_{1,i}}(\hat\omega(x_{\hat\jmath},\theta_{\hat\jmath})_s) &= \sum_{\ell_2\in\mathcal{L}_{\ell_{1,i}}} \nu_{\ell_2}\mu_{\ell_2}(\hat\omega(x_L,\theta_L)_s)\\ &= \sum_{\ell_2\in\mathcal{L}_{\ell_{1,i}}} \nu_{\ell_2}\int_{a_{\ell_2}}^{b_{\ell_2}} \rho_{\ell_2}(x)\hat\omega(f^L_{\theta^*_{\ell_2}}(x),\theta^*_{\ell_2})_s + \mathcal{O}(\varepsilon L^2)\\ &= \mathcal{O}(e^{-C_\# L}+\varepsilon L^2).\end{aligned}$$

Collecting all the above estimates yields

$$\begin{aligned}\varepsilon\sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle\nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(\varepsilon k))\zeta_\varepsilon(\varepsilon k)\rangle) &= \mathcal{O}(h^2+\varepsilon^{\frac12}h)\\ + h\mu_\ell(\langle\nabla A(\zeta_\varepsilon(t)), D\bar\omega(\overline{\Theta}(t))\zeta_\varepsilon(t)\rangle) + \mathcal{O}(h^2\varepsilon^{\frac12}L^2 + h^2\varepsilon^{-\frac12}e^{-C_\# L} + \varepsilon^{\frac12}Lh).\end{aligned} \tag{7.12}$$

To deal with the second term in the first line of equation (7.11) we argue as before:

$$\begin{aligned}\sum_{k=t\varepsilon^{-1}}^{(t+h)\varepsilon^{-1}} \mu_\ell(\langle\nabla A(\zeta_\varepsilon(t)),\hat\omega(x_k,\theta_k)\rangle) &= \sum_{k=t\varepsilon^{-1}}^{t\varepsilon^{-1}+L} \mu_\ell(\langle\nabla A(\zeta_\varepsilon(t)),\hat\omega(x_k,\theta_k)\rangle)\\ &\quad + \mathcal{O}(hL^2+\varepsilon^{-1}he^{C_\# L})\\ &= \mathcal{O}(L+hL^2+\varepsilon^{-1}he^{C_\# L}).\end{aligned}$$

Collecting the above computations and remembering (7.4) we obtain

$$\begin{aligned}\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle\nabla A(\zeta(t)),\zeta^h(t)\rangle) =& h\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle\nabla A(\zeta(t)), D\bar\omega(\overline{\Theta}(t))\zeta(t)\rangle)\\ &+ \mathcal{O}(h^2 + L\sqrt{\varepsilon} + h\sqrt{\varepsilon}L^2)\end{aligned} \tag{7.13}$$

provided L is chosen in the interval $[C_*\ln\varepsilon^{-1}, \varepsilon^{-\frac14}]$ with $C_*>0$ large enough.

To conclude we must compute the term on the right hand side of the first line of equation (7.10). Consider first the case $|j-k|>L$. Suppose $k>j$, the other case being equal, then, letting $\mathfrak{L}_\ell$ be the standard family generated by ℓ at time $\varepsilon^{-1}t$, and set $\hat k = k-\varepsilon^{-1}t, \hat\jmath = j-\varepsilon^{-1}t$, $B(x,\theta,t) = (D^2A)(\varepsilon^{-\frac12}(\theta-\overline{\Theta}(t)))$

$$\mu_\ell(\langle(D^2A)(\zeta_\varepsilon(t))\hat\omega(x_k,\theta_k),\hat\omega(x_j,\theta_j)\rangle)$$
$$= \sum_{\ell_1\in\mathfrak{L}_\ell}\nu_{\ell_1}\mu_{\ell_1}(\langle B(x_0,\theta_0,t)\hat\omega(x_{\hat k},\theta_{\hat k}),\hat\omega(x_{\hat\jmath},\theta_{\hat\jmath})\rangle).$$

Note that the signed measure $\hat\mu_{\ell_1,r,s}(g) = \mu_{\ell_1}(B_{r,s}g)$ has a density with uniformly bounded derivative given by $\hat\rho_{\ell_1,r,s} = \rho_{\ell_1}B(x,G_{\ell_1}(x),t)_{r,s}$. Such a density can then be written as a linear combination of standard densities $\hat\rho_{\ell_1,r,s} = \alpha_{1,\ell_1,r,s}\rho_{1,\ell_1,r,s} +$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

$\alpha_{2,\ell_1,r,s}\rho_{2,\ell_1,r,s}$ with uniformly bounded coefficients $\alpha_{i,\ell_1,r,s}$. We can then use the same trick at time j and then at time $k-L$ and obtain that the quantity we are interested in can be written as a linear combination of quantities of the type

$$\begin{aligned}\mu_{\ell_3,r,s}(\hat{\omega}_s(x_L,\theta_L)) &= \mu_{\ell_3,r,s}(\hat{\omega}_s(x_L,\theta^*_{\ell_3}) + \mathcal{O}(L\varepsilon)\\ &= \int_a^b \tilde{\rho}_{r,s}\hat{\omega}_s(f^L_{\theta^*_{\ell_3}}(x),\theta^*_{\ell_3}) + \mathcal{O}(L^2\varepsilon)\\ &= \mathcal{O}(e^{-C_{\#}L} + L^2\varepsilon)\end{aligned}$$

where we argued as in the proof of Lemma 7.3. Thus the total contribution of all such terms is of order $L^2h^2 + \varepsilon^{-1}e^{-C_{\#}L}h^2$. Next, the terms such that $|k-j| \le L$ but $j \le \varepsilon^{-1}t + L$ give a total contribution of order $L^2\varepsilon$ while to estimate the other terms it is convenient to proceed as before but stop at the time $j-L$. Setting $\tilde{k} = k-j+L$ we obtain terms of the form

$$\mu_{\ell_2,r,s}(\hat{\omega}_s(x_{\tilde{k}},\theta_{\tilde{k}})\hat{\omega}_r(x_L,\theta_L)\rangle) = \Gamma_{k-j}(\theta^*_{\ell_2})_{r,s} + \mathcal{O}(e^{-C_{\#}L} + L^2\varepsilon)$$

where

$$\Gamma_k(\theta) = \int_S \hat{\omega}(f^k_\theta(x),\theta)\otimes\hat{\omega}(x,\theta)\, m_\theta(dx).$$

The case $j > k$ yields the same results but with Γ^*_j. Remembering the smooth dependence of the covariance on the parameter θ (see [**10**]), substituting the result of the above computation in (7.10) and then (7.10) and (7.13) in (7.9) we finally have

$$\begin{aligned}&\widetilde{\mathbb{E}}^\varepsilon_\ell(A(\zeta(t+h)) - A(\zeta(t))) = h\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle\nabla A(\zeta(t)), D\bar{\omega}(\overline{\Theta}(t))\zeta(t)\rangle)\\ &\quad + h\widetilde{\mathbb{E}}^\varepsilon_\ell(\mathrm{Tr}(\sigma^2(\overline{\Theta}(t))D^2A(\zeta(t)))) + \mathcal{O}(L\sqrt{\varepsilon} + hL^2\sqrt{\varepsilon} + h^2L^2)\\ &= \int_t^{t+h}\left[\widetilde{\mathbb{E}}^\varepsilon_\ell(\langle\nabla A(\zeta(s)), D\bar{\omega}(\overline{\Theta}(s))\zeta(s)\rangle) + \widetilde{\mathbb{E}}^\varepsilon_\ell(\mathrm{Tr}(\sigma^2(\overline{\Theta}(s))D^2A(\zeta(s)))\right]ds\\ &\quad + \mathcal{O}(L\sqrt{\varepsilon} + hL^2\sqrt{\varepsilon} + h^{\frac{3}{2}} + h^2L^2).\end{aligned}$$

The proposition follows by summing the $h^{-1}t$ terms in the interval $[0,t]$ and by choosing $L = \varepsilon^{-\frac{1}{100}}$ and $h = \varepsilon^{\frac{1}{3}}$. □

In the previous Lemma the expression σ^2 just stands for a specified matrix, we did not prove that such a matrix is positive definite and hence it has a well defined square root σ, nor we have much understanding of the properties of such a σ (provided it exists). To clarify this is our next task.

LEMMA 7.5. *The matrices $\sigma^2(s)$, $s\in[0,T]$, are symmetric and non negative, hence they have a unique real symmetric square root $\sigma(s)$. In addition, if, for each $v\in\mathbb{R}^d$, $\langle v,\bar{\omega}\rangle$ is not a smooth coboundary, then there exists $c>0$ such that $\sigma(s)\ge c\mathbb{1}$.*

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

PROOF. For each $v \in \mathbb{R}^d$ a direct computation shows that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} m_\theta \left(\left[\sum_{k=0}^{n-1} \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \right]^2 \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k,j=0}^{n-1} m_\theta \left(\langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \langle v, \omega(f_\theta^j(\cdot), \theta) \rangle \right) \\
&= m_\theta(\langle v, \omega(\cdot, \theta) \rangle^2) + \lim_{n\to\infty} \frac{2}{n} \sum_{k=1}^{n-1} (n-k) m_\theta(\langle \omega(\cdot, \theta), v \rangle \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle) \\
&= m_\theta(\langle v, \omega(\cdot, \theta) \rangle^2) + 2 \sum_{k=1}^{\infty} m_\theta \left(\langle \omega(\cdot, \theta), v \rangle \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \right) = \langle v, \sigma(\theta)^2 v \rangle.
\end{aligned}$$

This implies that $\sigma(\theta)^2 \geq 0$ and since it is symmetric, there exists, unique, $\sigma(\theta)$ symmetric and non-negative. On the other hand if $\langle v, \sigma^2(\theta) v \rangle = 0$, then, by the decay of correlations and the above equation, we have

$$\begin{aligned}
m_\theta \left(\left[\sum_{k=1}^{n-1} \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \right]^2 \right) &= n\, m_\theta(\langle v, \omega(\cdot, \theta) \rangle^2) \\
&\quad + 2n \sum_{k=0}^{n-1} m_\theta \left(\langle v, \omega(\cdot, \theta) \rangle \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \right) + \mathcal{O}(1) \\
&= 2n \sum_{k=n}^{\infty} m_\theta \left(\langle v, \omega(\cdot, \theta) \rangle \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle \right) + \mathcal{O}(1) = \mathcal{O}(1).
\end{aligned}$$

Thus the L^2 norm of $\phi_n = \sum_{k=1}^{n-1} \langle v, \omega(f_\theta^k(\cdot), \theta) \rangle$ is uniformly bounded. Hence there exist a weakly convergent subsequence. Let $\phi \in L^2$ be an accumulation point, then for each $\varphi \in \mathcal{C}^1$ we have

$$m_\theta(\phi \circ f_\theta \varphi) = \lim_{k\to\infty} m_\theta(\phi_{n_k} \circ f_\theta \varphi) = m_\theta(\phi\varphi) - m_\theta(\langle v, \omega(\cdot, \theta) \rangle \varphi)$$

That is $\langle v, \omega(x, \theta) \rangle = \phi(x) - \phi \circ f_\theta(x)$. In other words $\langle v, \omega(x, \theta) \rangle$ is an L^2 coboundary. Since the Livsic Theorem [**11**] states that the solution of the cohomological equation must be smooth, we have $\phi \in \mathcal{C}^1$. □

7.3. Uniqueness of the Martingale Problem.

We are left with the task of proving the uniqueness of the martingale problem. Note that in the present case the operator depends explicitly on time. Thus if we want to set the initial condition at a time $s \neq 0$ we need to slightly generalise the definition of martingale problem. To avoid this, for simplicity, here we consider only initial conditions at time zero, which suffice for our purposes. In fact, we will consider only the initial condition $\zeta(0) = 0$, since it is the only one we are interested in. We have then the same definition of the martingale problem as in Definition 1, apart form the fact that $\mathcal{L}$ is replaced by $\mathcal{L}_s$ and $y = 0$.

Since the operators $\mathcal{L}_s$ are second order operators, we could use well known results. Indeed, there exists a deep theory due to Stroock and Varadhan that establishes the uniqueness of the martingale problem for a wide class of second order operators, [**13**]. Yet, our case is especially simple because the coefficients of the higher order part of the differential operator depend only on time and not on

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

ζ. In this case it is possible to modify a simple proof of the uniqueness that works when all the coefficients depend only on time, [**13**, Lemma 6.1.4]. We provide here the argument for the reader's convenience.

PROPOSITION 7.6. *The martingale problem associated to the operators $\mathcal{L}_s$ in Proposition* 7.4 *has a unique solution.*

PROOF. As already noticed, $\mathcal{L}_t$, defined in (7.2), depends on ζ only via the coefficient of the first order part. It is then natural to try to change measure so that such a dependence is eliminated and we obtain a martingale problem with respect to an operator with all coefficients depending only on time, then one can conclude arguing as in [**13**, Lemma 6.1.4]. Such a reduction is routinely done in probability via the Cameron-Martin-Girsanov formula. Yet, given the simple situation at hand one can proceed in a much more naive manner. Let $S(t) : [0,T] \to M_d$, M_d being the space of $d \times d$ matrices, be the generated by the differential equation

$$\begin{aligned} \dot{S}(t) &= -D\overline{\omega}(\overline{\Theta}(t))S(t) \\ S(0) &= \mathbb{1}. \end{aligned}$$

Note that, setting $\varsigma(t) = \det S(t)$ and $B(t) = D\overline{\omega}(\overline{\Theta}(t))$, we have

$$\begin{aligned} \dot{\varsigma}(t) &= -\operatorname{tr}(B(t))\varsigma(t) \\ \varsigma(0) &= 1. \end{aligned}$$

The above implies that $S(t)$ is invertible.

Define the map $\mathcal{S} \in \mathcal{C}^0(\mathcal{C}^0([0,T],\mathbb{R}^d), \mathcal{C}^0([0,T],\mathbb{R}^d))$ by $[\mathcal{S}\zeta](t) = S(t)\zeta(t)$ and set $\overline{\mathbb{P}} = \mathcal{S}_*\widetilde{\mathbb{P}}$. Note that the map $\mathcal{S}$ is invertible. Finally, we define the operator

$$\widehat{\mathcal{L}}_t = \frac{1}{2}\sum_{i,j}[\widehat{\Sigma}(t)^2]_{i,j}\partial_{\zeta_i}\partial_{\zeta_j},$$

where $\widehat{\Sigma}(t)^2 = S(t)\sigma(t)^2S(t)^*$, $\sigma(t) = \sigma(\overline{\Theta}(t))$ as mentioned after (7.2). Let us verify that $\overline{\mathbb{P}}$ satisfies the martingale problem with respect to the operators $\widehat{\mathcal{L}}_t$. By Lemma C.1 we have

$$\begin{aligned} \frac{d}{dt}\overline{\mathbb{E}}(A(\zeta(t)) \mid \mathcal{F}_s) &= \frac{d}{dt}\widetilde{\mathbb{E}}(A(S(t)\zeta(t)) \mid \mathcal{F}_s) \\ &= \widetilde{\mathbb{E}}(\dot{S}(t)\nabla A(S(t)\zeta(t)) + \mathcal{L}_t A(S(t)\zeta(t)) \mid \mathcal{F}_s) \\ &= \frac{1}{2}\widetilde{\mathbb{E}}\left(\sum_{i,j,k,l}\sigma^2(t)_{i,j}\partial_{\zeta_k}\partial_{\zeta_l}A(S(t)\zeta(t))S(t)_{k,i}S(t)_{l,j} \,\middle|\, \mathcal{F}_s\right) \\ &= \overline{\mathbb{E}}(\widehat{\mathcal{L}}_t A(\zeta(t)) \mid \mathcal{F}_s). \end{aligned}$$

Thus the claim follows by Lemma C.1 again.

Accordingly, if we prove that the above martingale problem has a unique solution, then $\overline{\mathbb{P}}$ is uniquely determined, which, in turn, determines uniquely $\widetilde{\mathbb{P}}$, concluding the proof.

Let us define the function $B \in \mathcal{C}^1(\mathbb{R}^{2d+1}, \mathbb{R})$ by

$$B(t,\zeta,\lambda) = e^{\langle\lambda,\zeta\rangle - \frac{1}{2}\int_s^t \langle\lambda,\widehat{\Sigma}(\tau)^2\lambda\rangle d\tau}$$

then Lemma C.1 implies

$$\frac{d}{dt}\overline{\mathbb{E}}(B(t,\zeta(t),\lambda) \mid \mathcal{F}_s) = \overline{\mathbb{E}}(-\frac{1}{2}\langle\lambda,\widehat{\Sigma}(t)^2\lambda\rangle B(t,\zeta(t),\lambda) + \widehat{\mathcal{L}}_t B(t,\zeta(t),\lambda) \mid \mathcal{F}_s) = 0.$$

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Hence

$$\overline{\mathbb{E}}(e^{\langle\lambda,\zeta(t)\rangle} \mid \mathcal{F}_s) = e^{\langle\lambda,\zeta(s)\rangle+\frac{1}{2}\int_s^t\langle\lambda,\widehat{\Sigma}(\tau)^2\lambda\rangle d\tau}.$$

From this follows that the finite dimensional distributions are uniquely determined. Indeed, for each $n\in\mathbb{N}$, $\{\lambda_i\}_{i=1}^n$ and $0\le t_1<\cdots<t_n$ we have

$$\begin{aligned}\overline{\mathbb{E}}\left(e^{\sum_{i=1}^n\langle\lambda_i,\zeta(t_i)\rangle}\right) &= \overline{\mathbb{E}}\left(e^{\sum_{i=1}^{n-1}\langle\lambda_i,\zeta(t_i)\rangle}\overline{\mathbb{E}}\left(e^{\langle\lambda_n,\zeta(t_n)\rangle} \mid \mathcal{F}_{t_{n-1}}\right)\right)\\ &= \overline{\mathbb{E}}\left(e^{\sum_{i=1}^{n-2}\langle\lambda_i,\zeta(t_i)\rangle+\langle\lambda_{n-1}+\lambda_n,\zeta(t_{n-1})\rangle}\right)e^{\frac{1}{2}\int_{t_{n-1}}^{t_n}\langle\lambda_n,\widehat{\Sigma}(\tau)^2\lambda_n\rangle d\tau}\\ &= e^{\frac{1}{2}\int_0^{t_n}\langle\sum_{i=n(\tau)}^n\lambda_i,\widehat{\Sigma}(\tau)^2\sum_{i=n(\tau)}^n\lambda_i\rangle d\tau}\end{aligned}$$

where $n(\tau)=\inf\{m\mid t_m\ge\tau\}$. This concludes the Lemma since it implies that the measure is uniquely determined on the sets that generate the σ-algebra.[27] Note that we have also proven that the process is a zero mean Gaussian process; this, after translating back to the original measure, generalises Remark 2.3. □

Appendix A. Geometry

For $c>0$, consider the cones $\mathcal{C}_c=\{(\xi,\eta)\in\mathbb{R}^2 \,:\, |\eta|\le\varepsilon c|\xi|\}$. Note that

$$\mathrm{d}F_\varepsilon=\begin{pmatrix}\partial_x f & \partial_\theta f\\ \varepsilon\partial_x\omega & 1+\varepsilon\partial_\theta\omega\end{pmatrix}.$$

Thus if $(1,\varepsilon u)\in\mathcal{C}_c$,

$$\begin{aligned}\mathrm{d}_pF_\varepsilon(1,\varepsilon u) &= (\partial_x f(p)+\varepsilon u\partial_\theta f(p),\varepsilon\partial_x\omega(p)+\varepsilon u+\varepsilon^2 u\partial_\theta\omega(p))\\ &= \partial_x f(p)\left(1+\varepsilon\frac{\partial_\theta f(p)}{\partial_x f(p)}u\right)\cdot(1,\varepsilon\Xi_p(u))\end{aligned}$$

where

$$\Xi_p(u)=\frac{\partial_x\omega(p)+(1+\varepsilon\partial_\theta\omega(p))u}{\partial_x f(p)+\varepsilon\partial_\theta f(p)u}. \tag{A.1}$$

Thus the vector $(1,\varepsilon u)$ is mapped to the vector $(1,\varepsilon\Xi_p(u))$. Thus letting $K=\max\{\|\partial_x\omega\|_\infty,\|\partial_\theta\omega\|_\infty,\|\partial_\theta f\|_\infty\}$ we have, for $|u|\le c$ and assuming $K\varepsilon c\le 1$,

$$|\Xi_p(u)|\le\frac{K+(1+\varepsilon K)c}{\lambda-\varepsilon Kc}\le\frac{K+1+c}{\lambda-1}.$$

Thus, if we choose $c\in[\frac{K+1}{\lambda-2},(\varepsilon K)^{-1}]$ we have that $\mathrm{d}_pF_\varepsilon(\mathcal{C}_c)\subset\mathcal{C}_c$. Since this implies that $\mathrm{d}_pF_\varepsilon^{-1}\complement\mathcal{C}_c\subset\complement\mathcal{C}_c$ we have that the complementary cone $\complement\mathcal{C}_{K\varepsilon^{-1}}$ is invariant under $\mathrm{d}F_\varepsilon^{-1}$. From now on we fix $c=\frac{K+1}{\lambda-2}$.

Hence, for any $p\in\mathbb{T}^{1+d}$ and $n\in\mathbb{N}$, we can define the quantities v_n,u_n,s_n,r_n as follows:

$$\mathrm{d}_pF_\varepsilon^n(1,0)=v_n(1,\varepsilon u_n)\qquad\qquad \mathrm{d}_pF_\varepsilon^n(s_n,1)=r_n(0,1) \tag{A.2}$$

with $|u_n|\le c$ and $|s_n|\le K$. For each n the slope field s_n is smooth, therefore integrable; given any small $\Delta>0$ and $p=(x,\theta)\in\mathbb{T}^{1+d}$, define $\mathcal{W}_n^{\mathrm{c}}(p,\Delta)$ the *local n-step central manifold of size* Δ as the connected component containing p of the intersection with the strip $\{|\theta'-\theta|<\Delta\}$ of the integral curve of $(s_n,1)$ passing through p.

[27] See the discussion at the beginning of Section 2.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Notice that, by definition, $\mathrm{d}_p F_\varepsilon(s_n(p),1) = r_n/r_{n-1}(s_{n-1}(F_\varepsilon p),1)$; thus, by definition, there exists a constant b such that:

$$\exp(-b\varepsilon) \le \frac{r_n}{r_{n-1}} \le \exp(b\varepsilon). \tag{A.3}$$

Furthermore, define $\Gamma_n = \prod_{k=0}^{n-1} \partial_x f \circ F_\varepsilon^k$, and let

$$a = c \left\| \frac{\partial_\theta f}{\partial_x f} \right\|_\infty . \tag{A.4}$$

Clearly,

$$\Gamma_n \exp(-a\varepsilon n) \le v_n \le \Gamma_n \exp(a\varepsilon n). \tag{A.5}$$

Appendix B. Shadowing

In this section we provide a simple quantitative version of shadowing that is needed in the argument. Let $(x_k,\theta_k) = F_\varepsilon^k(x,\theta)$ with $k \in \{0,\dots,n\}$. We assume that θ belongs to the range of a standard pair ℓ (i.e., $\theta = G(x)$ for some $x \in [a,b]$).

Let $\theta^* \in S$ such that $\|\theta^* - \theta\| \le \varepsilon$ and set $f_*(x) = f(x,\theta^*)$. Let us denote with $\pi_x : X \to S$ the canonical projection on the x coordinate; then, for any $s \in [0,1]$, let

$$H_n(x,z,s) = \pi_x F_{s\varepsilon}^n(x,\theta^* + s(G_\ell(x) - \theta^*)) - f_*^n(z)$$

Note that, $H_n(x,x,0) = 0$, in addition, for any x,z and $s \in [0,1]$

$$\partial_z H_n(x,z,s) = -(f_*^n)'(z).$$

Accordingly, by the Implicit Function Theorem any $n \in \mathbb{N}$ and $s \in [0,1]$, there exists $Y_n(x,s)$ such that[28] $H_n(x,Y_n(x,s),s) = 0$; from now on $Y_n(x)$ stands for $Y_n(x,1)$. Note that setting $x_k^* = f_*^k(Y_n(x))$, by construction, $x_n^* = x_n$. Observe moreover that

$$\partial_x Y_n = (f_*^n)'(z)^{-1} \mathrm{d}(\pi_x F_\varepsilon^n) = \frac{(1 - G_\ell' s_n) v_n}{(f_*^n)' \circ Y_n}, \tag{B.1}$$

where we have used the notations introduced in equation (A.2). Recalling (A.5) and by the cone condition we have

$$e^{-c_\# \varepsilon n} \prod_{k=0}^{n-1} \frac{\partial_x f(x_k,\theta_k)}{f_*'(x_k^*)} \le \left| \frac{(1 - G_\ell' s_n) v_n}{(f_*^n)'} \right| \le e^{c_\# \varepsilon n} \prod_{k=0}^{n-1} \frac{\partial_x f(x_k,\theta_k)}{f_*'(x_k^*)}. \tag{B.2}$$

Next, we want to estimate to which degree x_k^* shadows the true trajectory.

Lemma B.1. *There exists $C > 0$ such that, for each $k \le n < C\varepsilon^{-\frac{1}{2}}$ we have*

$$\|\theta_k - \theta^*\| \le C_\# \varepsilon k$$
$$|x_k - x_k^*| \le C_\# \varepsilon k.$$

[28] The Implicit Function Theorem allows to define $Y_n(x,s)$ in a neighborhood of $s = 0$; in fact we claim that this neighborhood necessarily contains $[0,1]$. Otherwise, there would exist $\bar{s} \in (0,1)$ and $\bar{x}$ so that Y_n is defined at $(\bar{x},\bar{s})$ but not at $(\bar{x},s)$ with $s > \bar{s}$. We then could apply the Implicit Function Theorem at the point $(\bar{x},Y_n(\bar{x},\bar{s}),\bar{s})$ and obtain, by uniqueness, an extension of the previous function Y_n to a larger neighborhood of $s = 0$, which contradicts our assumption.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

PROOF. Observe that

$$\theta_k = \varepsilon \sum_{j=0}^{k-1} \omega(x_j, \theta_j) + \theta_0$$

thus $\|\theta_k - \theta^*\| \le C_\# \varepsilon k$. Accordingly, let us set[29] $\xi_k = x_k^* - x_k$; then, by the mean value theorem,

$$\begin{aligned} |\xi_{k+1}| &= |\partial_x f \cdot \xi_k + \partial_\theta f \cdot (\theta_k - \theta^*)| \\ &\ge \lambda |\xi_k| - C_\# \varepsilon k. \end{aligned}$$

Since, by definition, $\xi_n = 0$, we can proceed by backward induction, which yields

$$|\xi_k| \le \sum_{j=k}^{n-1} \lambda^{-j+k} C_\# \varepsilon j \le C_\# \varepsilon \sum_{j=0}^{\infty} \lambda^{-j}(j+k) \le C_\# \varepsilon k. \qquad \square$$

LEMMA B.2. *There exists $C > 0$ such that, for each $n \le C\varepsilon^{-\frac{1}{2}}$,*

$$e^{-c_\# \varepsilon n^2} \le |Y_n'| \le e^{c_\# \varepsilon n^2}. \tag{B.3}$$

In particular, Y_n is invertible with uniformly bounded derivative.

PROOF. Let us prove the upper bound, the lower bound being similar. By equations (B.1), (B.2) and Lemma B.1 we have

$$|Y_n'| \le e^{c_\# \varepsilon n} e^{\sum_{k=0}^{n-1} \ln \partial_x f(x_k, \theta_k) - \ln f_*'(x_k^*)} \le e^{c_\# \varepsilon n} e^{c_\# \sum_{k=0}^{n-1} \varepsilon k}. \qquad \square$$

Appendix C. Martingales, operators and Itō's calculus

Suppose that $\mathcal{L}_t \in L(\mathcal{C}^r(\mathbb{R}^d, \mathbb{R}), \mathcal{C}^0(\mathbb{R}^d, \mathbb{R}))$, $t \in \mathbb{R}$, is a one parameter family of bounded linear operators that depends continuously on t.[30] Also suppose that $\mathbb{P}$ is a measure on $\mathcal{C}^0([0,T], \mathbb{R}^d)$ and let $\mathcal{F}_t$ be the σ-algebra generated by the variables $\{z(s)\}_{s \le t}$.[31]

LEMMA C.1. *The two properties below are equivalent:*

1. *For all $A \in \mathcal{C}^1(\mathbb{R}^{d+1}, \mathbb{R})$, such that, for all $t \in \mathbb{R}$, $A(t, \cdot) \in \mathcal{C}^r(\mathbb{R}^d, \mathbb{R})$, and for all times $s, t \in [0,T], s < t$, the function $g(t) = \mathbb{E}(A(t, z(t)) \mid \mathcal{F}_s)$ is differentiable and $g'(t) = \mathbb{E}(\partial_t A(t, z(t)) + \mathcal{L}_t A(t, z(t)) \mid \mathcal{F}_s)$.*
2. *For all $A \in \mathcal{C}^r(\mathbb{R}^d, \mathbb{R})$, $M(t) = A(z(t)) - A(z(0)) - \int_0^t \mathcal{L}_s A(z(s)) ds$ is a martingale with respect to $\mathcal{F}_t$.*

PROOF. Let us start with (1) $\Rightarrow$ (2). Let us fix $t \in [0,T]$, then for each $s \in [0,t]$ let us define the random variables $B(s)$ by

$$B(s, z) = A(z(t)) - A(z(s)) - \int_s^t \mathcal{L}_\tau A(z(\tau)) d\tau.$$

[29] Here, as we already done before, we are using the fact that we can lift $\mathbb{T}^1$ to the universal covering $\mathbb{R}$.

[30] Here $\mathcal{C}^r$ are thought as Banach spaces, hence consist of bounded functions. A more general setting can be discussed by introducing the concept of a *local martingale*.

[31] At this point the reader is supposed to be familiar with the intended meaning: for all $\vartheta \in \mathcal{C}^0([0,T], \mathbb{R}^d)$, $[z(s)](\vartheta) = z(\vartheta, s) = \vartheta(s)$.

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

Clearly, for each $z \in \mathcal{C}^0$, $B(s,z)$ is continuous in s, and $B(t,z) = 0$. Hence, for all $\tau \in (s,t]$, by Fubini we have[32]

$$\begin{aligned}\frac{d}{d\tau}\mathbb{E}(B(\tau) \mid \mathcal{F}_s) &= -\frac{d}{d\tau}\mathbb{E}(A(z(\tau)) \mid \mathcal{F}_s) - \frac{d}{d\tau}\int_\tau^t \mathbb{E}(\mathcal{L}_r A(z(r)) \mid \mathcal{F}_s)dr \\ &= \mathbb{E}(-\mathcal{L}_\tau A(z(\tau)) + \mathcal{L}_\tau A(z(\tau)) \mid \mathcal{F}_s) = 0.\end{aligned}$$

Thus, since B is bounded, by Lebesgue dominated convergence theorem, we have

$$0 = \mathbb{E}(B(t) \mid \mathcal{F}_s) = \lim_{\tau \to 0} \mathbb{E}(B(\tau) \mid \mathcal{F}_s) = \mathbb{E}(B(s) \mid \mathcal{F}_s).$$

This implies

$$\mathbb{E}(M(t) \mid \mathcal{F}_s) = \mathbb{E}(B(s) \mid \mathcal{F}_s) + M(s) = M(s)$$

as required.

Next, let us check (2) $\Rightarrow$ (1). For each $h > 0$ we have

$$\begin{aligned}\mathbb{E}(A(t+h, z(t+h)) - A(t, z(t)) \mid \mathcal{F}_s) &= \mathbb{E}\left((\partial_t A)(t, z(t+h)) \mid \mathcal{F}_s\right) h + o(h) \\ &+ \mathbb{E}\left(M(t+h) - M(t) + \int_t^{t+h} \mathcal{L}_\tau A(t, z(\tau))d\tau \mid \mathcal{F}_s\right).\end{aligned}$$

Since M is a martingale $\mathbb{E}\left(M(t+h) - M(t) \mid \mathcal{F}_s\right) = 0$. The lemma follows by Lebesgue dominated convergence theorem. □

The above is rather general, to say more it is necessary to specify other properties of the family of operators $\mathcal{L}_s$. A case of particular interest arises for second order differential operators like (7.2). Namely, suppose that

$$(\mathcal{L}_s A)(z) = \sum_i a(z,s)_i \partial_{z_i} A(z) + \frac{1}{2}\sum_{i,j=1}^d [\sigma^2(z,s))]_{i,j} \partial_{z_i}\partial_{z_j} A(z),$$

where, for simplicity, we assume a, σ to be smooth and bounded and $\sigma_{ij} = \sigma_{ji}$. Clearly, (7.2) is a special case of the above. In such a case it turns out that it can be established a strict connection between $\mathcal{L}_s$ and the Stochastic Differential Equation

$$dz = adt + \sigma dB \tag{C.1}$$

where B is the standard Brownian motion. The solution of (C.1) can be defined in various way. One possibility is to define it as the solution of the Martingale problem [**13**], another is to use stochastic integrals [**15**, Theorem 6.1]. The latter, more traditional, approach leads to Itō's formula that reads, for each bounded continuous function A of t and z, [**15**, page 91],

$$\begin{aligned}A(z(t),t) - A(z(0),0) &= \int_0^t \partial_s A(z(s),s)ds + \int_0^t \sum_i a(z(s),s)_i \partial_{z_i} A(z(s),s)ds \\ &+ \frac{1}{2}\int_0^t \sum_{i,j} \sigma^2(z(s),s)\partial_{z_i}\partial_{z_j} A(z(s),s)ds \\ &+ \sum_{i,j}\int_0^t \sigma_{ij}(z(s),s)\partial_{z_j} A(z(s),s)_j dB_i(s)\end{aligned}$$

[32] If uncomfortable about applying Fubini to conditional expectations, then have a look at [**15**, Theorem 4.7].

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

where the last is a stochastic integral [**15**, Theorem 5.3]. This formula is often written in the more impressionistic form

$$dA = \partial_t A dt + a\partial_z A dt + \sigma\partial_z A dB + \frac{1}{2}\sigma^2\partial_z^2 A dt = \partial_t A dt + \sigma\partial_z A dB + \mathcal{L}_t A dt.$$

Taking the expectation with respect to $\mathbb{E}(\cdot \mid \mathcal{F}_s)$ we obtain exactly condition (1) of Lemma C.1, hence we have that the solution satisfies the Martingale problem.

REMARK C.2. *Note that, if one defines the solution of* (C.1) *as the solution of the associated Martingale problem, then one can dispense from Itō's calculus altogether. This is an important observation in our present context in which the fluctuations come form a deterministic problem rather than from a Brownian motion and hence a direct application of Itō's formula is not possible.*

References

[1] Viviane Baladi, *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics, vol. 16, World Scientific Publishing Co., Inc., River Edge, NJ, 2000. MR1793194 (2001k:37035)

[2] Viviane Baladi and Masato Tsujii, *Anisotropic Hölder and Sobolev spaces for hyperbolic diffeomorphisms* (English, with English and French summaries), Ann. Inst. Fourier (Grenoble) **57** (2007), no. 1, 127–154. MR2313087 (2008d:37034)

[3] Dmitry Dolgopyat, *Averaging and invariant measures* (English, with English and Russian summaries), Mosc. Math. J. **5** (2005), no. 3, 537–576, 742. MR2241812 (2007i:37063)

[4] M. D. Donsker and S. R. S. Varadhan, *Large deviations for noninteracting infinite-particle systems*, J. Statist. Phys. **46** (1987), no. 5-6, 1195–1232, DOI 10.1007/BF01011162. MR893138 (88i:60047)

[5] Sébastien Gouëzel and Carlangelo Liverani, *Banach spaces adapted to Anosov systems*, Ergodic Theory Dynam. Systems **26** (2006), no. 1, 189–217, DOI 10.1017/S0143385705000374. MR2201945 (2007h:37037)

[6] Sébastien Gouëzel and Carlangelo Liverani, *Compact locally maximal hyperbolic sets for smooth maps: fine statistical properties*, J. Differential Geom. **79** (2008), no. 3, 433–477. MR2433929 (2010a:37050)

[7] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan, *Nonlinear diffusion limit for a system with nearest neighbor interactions*, Comm. Math. Phys. **118** (1988), no. 1, 31–59. MR954674 (89m:60255)

[8] Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza. MR1326374 (96c:58055)

[9] R. Z. Has′minskiĭ, *On the principle of averaging the Itô's stochastic differential equations* (Russian, with Czech summary), Kybernetika (Prague) **4** (1968), 260–279. MR0260052 (41 #4681)

[10] Carlangelo Liverani, *Invariant measures and their properties. A functional analytic point of view*, Dynamical systems. Part II, Pubbl. Cent. Ric. Mat. Ennio Giorgi, Scuola Norm. Sup., Pisa, 2003, pp. 185–237. MR2071241 (2005d:37045)

[11] A. N. Livšic, *Cohomology of dynamical systems* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 1296–1320. MR0334287 (48 #12606)

[12] Charles Pugh and Michael Shub, *Stably ergodic dynamical systems and partial hyperbolicity*, J. Complexity **13** (1997), no. 1, 125–179, DOI 10.1006/jcom.1997.0437. MR1449765 (98e:58110)

[13] Daniel W. Stroock and S. R. Srinivasa Varadhan, *Multidimensional diffusion processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 233, Springer-Verlag, Berlin-New York, 1979. MR532498 (81f:60108)

[14] S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes in Mathematics, vol. 7, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001. MR1852999 (2003a:60001)

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.

[15] S. R. S. Varadhan, *Stochastic processes*, Courant Lecture Notes in Mathematics, vol. 16, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2007. MR2354349 (2009c:60207)

Department of Mathematics, University of Toronto, 40 St George St. Toronto, Ontario M5S 2E4

E-mail address: jacopods@math.utoronto.ca

URL: http://www.math.utoronto.ca/jacopods

Dipartimento di Matematica, II Università di Roma (Tor Vergata), Via della Ricerca Scientifica, 00133 Roma, Italy

E-mail address: liverani@mat.uniroma2.it

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.