EVOLUTION OF ADIABATIC INVARIANTS IN STOCHASTIC AVERAGING.

DMITRY DOLGOPYAT.

Abstract. An averaging problem with Markov fast motion is considered. The diffusive limit is obtained for the evolution of adiabatic invariants under the assumption that the averaged motion is ergodic on almost every energy level.

1. Introduction.

Averaging theory constitutes one of the most developed branches of differential equations. The general setting is the following. Consider an equation

$$\dot{x} = \varepsilon X(x, \xi_t) \tag{1}$$

where ξ_t is a process changing with unit speed. In applications it is often important to allow a coupling between ξ and x. For example, ξ_t can be a random process whose distribution depends on x or it can satisfy a differential equation whose coefficients depend on x.

If ε is small, numerical solution of (1) is costly because x changes after time $1/\varepsilon$ whereas any numerical scheme should have steps $o(1)$ to capture the oscillations of ξ. Therefore one would like to eliminate ξ by asserting that the solution of (1) satisfies

$$x(t/\varepsilon) \sim \bar{x}(t) \tag{2}$$

where $\bar{x}(t)$ satisfies the averaging equation

$$\dot{\bar{x}} = \bar{X}(\bar{x}) \text{ with } \bar{X}(\bar{x}) = \mathbf{Average}(X(\bar{x}, \xi)) \tag{3}$$

and the meaning of the operation **Average**$(\dots)$ depends on the problem at hand. In case X does not depend on the first variable, x is given by the expression

$$x(t/\varepsilon) = \varepsilon \int_0^{t/\varepsilon} X(\xi_s)ds$$

which is covered by the ergodic theorem. So the general averaging problem can be considered as a non-linear version of the ergodic theory. Hence, the methods used to analyze ergodic properties of a certain process can be extended to get an averaging theorem on scale $1/\varepsilon$.

However, in many cases one wants to justify the averaging approach on longer time scales. This problem is quite subtle. Indeed the long time behavior of the averaged equation (3) can be quite sensitive to small perturbations. (For example, (3) could have invariant measures of full support whereas its small perturbations can have sinks (see e.g. [12] for the discussion of this phenomenon).) In this case terms neglected in (2) can dramatically alter long time dynamics. Hence one has to require nice dynamical properties of (3) in order to get the long-scale averaging so the problem is non-trivial.

One important case where long time analysis is required is the problem of adiabatic invariance. Namely we say that a vector valued function $h(x)$ is an almost adiabatic invariant if $< \nabla h, \bar{X} >= 0$. In this case (2) and (3) say that h does not change on scale $1/\varepsilon$ so it is an interesting question how long it takes before h changes significantly.

For the reason explained above previous papers dealing with this problem assumed that the averaged dynamics is very simple, namely quasiperiodic. In this paper we present a general approach to the problem of adiabatic invariance. It extends a method developed by Anosov [1] to prove short time averaging for ODEs by combining it with the moment technique used in averaging theory by Khasminskii and others. In order to present the main idea without unnecessary technicalities we discuss the simplest case where our approach work. In particular in order to simplify the computations we deal with discrete rather than continuous time.

Acknowledgment. *I thank Mark Freidlin and Yuri Kifer for useful comments. This research is supported by NSF, Sloan Foundation and IPST.*

2. The main theorem.

Consider a recurrence

$$x_{n+1} = x_n + \varepsilon F(x_n, \xi_n) + \varepsilon^2 G(x_n, \xi_n) + \varepsilon^3 H(x_n, \xi_n, \varepsilon), \quad x \in \mathbb{R}^d \tag{4}$$

where ξ_n are i.i.d. random variables of compact support. We assume that there is $K > 0$ such that

$$||F||_{C^4} \leq K, \quad ||G||_{C^3} \leq K, \quad ||H||_{C^1} \leq K. \tag{5}$$

Then averaging principle ([9]) says that with probability close to 1, $x_{[t/\varepsilon]}$ is close to $y(t)$ where y solves the the averaged equation

$$\dot{y} = \bar{F}(y), \text{ where } \bar{F}(y) = \mathbb{E}(F(y, \xi)). \tag{6}$$

Let $h_1, h_2 \dots h_m$ be first integrals of (6)

$$< \nabla h_i, \bar{F} >= 0$$

which are independent in the sense that

$$\text{rk}\left(\frac{\partial h}{\partial x}\right) = m \tag{7}$$

at every point. We also assume that

$$||\nabla h||_{C^3} \leq K.$$

(6) implies that h_i change little on times of order $1/\varepsilon$. So it is interesting to describe their evolution on longer time scales. We impose two additional conditions. Let $\Phi(t)$ denote the flow generated by $\bar{F}$. Observe that (7) implies that level sets $M_c = \{\vec{h} = c\}$ are smooth submanifolds.

(COMP) M_h are compact and Φ restricted to M_h preserves a probability measure μ_h which is smooth and depends smoothly on h. More precisely let $\zeta(h)$ be a smooth probability density on $\mathbb{R}^m$ which is positive everywhere. Let $d\mu = \zeta(h) dh d\mu_h$. We require that

$$d\mu = \rho(x)dx \tag{8}$$

where ρ is a smooth positive density (it is easy to see that this condition is independent of the choice of ζ).

(ERG) (Φ, μ_h) is ergodic for almost every h.

Define

$$a_i(x) = \mathbb{E}\left(< \nabla h_i(x), G(x,\xi) > + \frac{1}{2} D^2 h(\nabla F(x,\xi), \nabla F(x,\xi))\right),$$

$$\sigma^2_{ij}(x) = \mathbb{E}(< \nabla h_i(x), F(x,\xi) >< \nabla h_j(x), F(x,\xi) >).$$

Let

$$\bar{a}(h) = \int a(x) d\mu_h(x), \quad \bar{\sigma}^2(h) = \int \sigma^2(x) d\mu_h(x).$$

Define $h_\varepsilon(t)$ by setting $h_\varepsilon(n\varepsilon^2) = h(x_{\varepsilon,n})$ and interpolating linearly in between. Let $p(x)$ be a probability density on $\mathbb{R}^d$. Let x_0 be chosen according to p.

Theorem 1. *As $\varepsilon \to 0$ the process h_ε converges weakly to the random process $\mathbf{h}$ satisfying the following SDE*

$$d\mathbf{h} = \bar{a}(\mathbf{h})dt + \bar{\sigma}(\mathbf{h})dw.$$

The assumptions of Theorem 1 appear quite reasonable at least in the volume preserving setting. In fact (COMP) guarantees some stationarity of **h** dynamics preventing the points of M_h from running off to infinity. Otherwise some extra assumptions on behavior of Φ at infinity would be necessary. (ERG) says that there are no additional first integrals so to describe the evolution of h we need no extra information. Finally since we only require ergodicity for **almost** all h, the assumption that $p \in L^1(\mathbb{R}^d)$ is needed to guarantee that we give the points with ergodic behavior positive weight.

3. Previous results.

Before giving the proof let us compare our result with the previous works. The readers should keep in mind however that the papers described below deal with a more general setting. Namely independence is often replaced by sufficiently fast mixing, L^∞ bounds (5) are replaced by L^p bounds for some $p < \infty$ and they work with continuous rather than discrete time. However we ignore these technical differences in the discussion below in order to compare the ideas.

The first work on this problem was done by Anosov [1]. He deals with the case when the RHS of (4) does not depend on ξ. The idea to use Corollary 3 to justify the averaging is taken from [1]. An improvement our paper is that we deal with three scale problem (ξ changes on the unit scale, x changes on the scale $1/\varepsilon$ and h changes on scale $1/\varepsilon^2$), whereas in [1] only the last two scales are present. Therefore in [1] Corollary 3 holds for all initial conditions whereas in our case it holds for most initial conditions. However this weaker version of Corollary 3 suffices to establish averaging. Let us also mention that the idea to use Anosov's approach in stochastic averaging is due to Kifer [11]. However he makes an extra assumption that $(\nabla h, F)$ depends on x only via h which makes the problem effectively two scale by eliminating the intermediate x-scale.

Another important advancement in this subject is the work of Khasminskii [9, 10] which gave the first rigorous results on stochastic averaging. While [9] deals with short time averaging, the setting of [10] is close to ours except that he has an extra assumption $F \equiv 0$. Some technical improvements in this setup are presented in [4].

[6, 5] deal with the case where the averaged motion is periodic and [7, 8] treat quasiperiodic case.

[6] also presents another version of the averaging theorem where our condition (8) is replaced by the requirement that the distribution of x_n has density for all $n > \delta/\varepsilon^2$. This assumption can be verified only

in special cases by quite delicate arguments [3]. Also our result has additional aesthetic appeal since all our assumptions are imposed on the averaged system whereas [6] impose an additional very strong restriction on the original system (4). On the other hand our theorem requires that x_0 has absolutely continuous distribution while the result of [6] works for arbitrary initial distribution. It is an important open problem to find viable conditions on distributions of F, G, and H implying absolute continuity assumption of [6].

4. *A priori* BOUNDS.

Given a set Ω we shall denote its compliment by Ω^c. Let $\mathcal{F}_n$ denote the sigma-algebra generated by $\xi_0, \xi_1 \dots \xi_{n-1}$.

Let vector sequence $S_{\varepsilon,n}$ satisfy

$$S_{n+1} = S_n + \varepsilon P(x_n, \xi_n) + \varepsilon^2 Q(x_n, \xi_n) + \varepsilon^3 R(x_n, \xi_n, \varepsilon). \tag{9}$$

Let $D \subset \mathbb{R}^d$ be a domain such that

$$||P||_{C^3(D)} \le \tilde{K}, \quad ||Q||_{C^2(D)} \le \tilde{K}, \quad ||R||_{C^0(D)} \le \tilde{K}$$

for some $\tilde{K}$. Assume in addition what

$$\bar{P}(x) = \mathbb{E}(P(x, \xi)) = 0. \tag{10}$$

Observe that we require one derivative less than for (5) because we shall apply the estimates below to an expression containing $\frac{dF}{dx}$. Let τ_ε be the first moment when x_n leaves D. Denote $\tilde{S}_n = S_{\min(n,\tau_\varepsilon)}$.

The next two lemmas are standard (see e.g [9], Section 2) but since the proofs in our case are much simpler than in more general settings considered elsewhere we provide the proofs in the appendix in order to make our presentation self-contained. Fix $T > 0$.

Lemma 1. *For all* $n_1 < n_2 \le T/\varepsilon^2$
(a) $\mathbb{E}(\tilde{S}_{n_2} - \tilde{S}_{n_1} | \mathcal{F}_{n_1}) = O(\varepsilon^2 (n_2 - n_1))$.
(b) $\mathbb{E}(|\tilde{S}_{n_2} - \tilde{S}_{n_1}|^2 | \mathcal{F}_{n_1}) = O(\varepsilon^2 (n_2 - n_1))$.
(c) $\mathbb{E}(|\tilde{S}_{n_2} - \tilde{S}_{n_1}|^4 | \mathcal{F}_{n_1}) = O(\varepsilon^4 (n_2 - n_1)^2)$.

Lemma 2. *Define* $S_\varepsilon(t)$ *be setting* $S_\varepsilon(n\varepsilon^2) = \tilde{S}_{\varepsilon,n}$ *and interpolating linearly in between. Then* $\{S_\varepsilon(t)\}$ *is a tight family.*

Corollary 3. *Given* $\delta > 0$ *there exists a constant* $C(\delta)$ *such that for all* $n \le T/\varepsilon^2$ *for all subsets* $\Omega \subset \mathbb{R}^d$

$$\mathbb{P}(x_n \in \Omega) \le C(\delta) \mu(\Omega) + \delta. \tag{11}$$

Proof. Recall (8). $\bar{F}$ preserves μ which means

$$\mathrm{div}_\rho \bar{F} = 0 \text{ where } \mathrm{div}_\rho \bar{F} = \mathrm{div} \bar{F} + \frac{< \bar{F}, \nabla \rho >}{\rho}.$$

Observe that (4) determines dynamics of each x_n but we can also use it to define a stochastic system of diffeomorphisms by iterating all initial conditions using the same $\{\xi_n\}$. Let $\Psi(n)$ denote the resulting family of diffeomorphisms. Let $\tilde{\mathbb{P}}$ denotes the distribution of x_n if x_0 is distributed according to μ. Denote

$$S_n = \ln \left(\frac{\rho(x_n)}{\rho(x_0)} \frac{dx_n}{dx_0} \right).$$

Observe that

$$S_{n+1} - S_n = \ln \left(\frac{\rho(x_{n+1})}{\rho(x_n)} \right) + \ln \det \left(1 + \varepsilon F(x,\xi) + \varepsilon^2 G(x,\xi) + \varepsilon^3 H(x,\xi,\varepsilon)\right).$$

Given r consider $D(r) = \{x : |h(x)| < r\}$. Then S_n satisfies (9) in $D(r)$ with

$$P(x,\xi) = \mathrm{div}_\rho F(x,\xi).$$

Applying Lemma 2 we see that given δ, r there exists L such that

$$\tilde{\mathbb{P}}(\max_{n \leq T/\varepsilon^2} |S_{\min(n, \tau_{\varepsilon,r})}| > \ln L) \leq \frac{\delta}{2}$$

where $\tau_{\varepsilon,r} = \min(n : \quad |h(x_n)| \geq r)$. By Lemma 2 we can choose r so that

$$\tilde{\mathbb{P}}(\max_{n \leq T/\varepsilon^2} |h(x_n)| \geq r) \leq \frac{\delta}{2}$$

Then

$$\tilde{\mathbb{P}}(\mathcal{M}^c) \leq \delta \text{ where } \mathcal{M} = \left\{ \frac{1}{L} \leq \left| \frac{\rho(x_n)}{\rho(x_0)} \frac{dx_n}{dx_0} \right| \leq L \text{ for all } n \leq T/\varepsilon \right\}.$$

We have

$$\tilde{\mathbb{P}}(x_n \in \Omega) = \iint 1_\Omega(\Psi(n)x_0) d\mu(x_0) d\xi \leq \iint 1_{\mathcal{M}}(x_0,\xi) 1_\Omega(\Psi(n)x_0) d\mu(x_0) d\xi + \tilde{\mathbb{P}}(\mathcal{M}^c).$$

The second term is less than δ. Changing variables $z = \Psi(n)x_0$ we rewrite the first term as

$$\iint 1_{\mathcal{M}}(x_0,\xi) 1_\Omega(z) \frac{d\mu(x_0)}{d\mu(z)} d\mu(z) d\xi.$$

Since $1_{\mathcal{M}}(x_0,\xi) \frac{d\mu(x_0)}{d\mu(z)} \leq L$ by the definition of $\mathcal{M}$ we can bound the last expression by

$$L \iint 1_\Omega(z) d\mu(z) d\xi = L\mu(\Omega).$$

This proves (11) with $\tilde{\mathbb{P}}$ instead of $\mathbb{P}$. To get the result for $\mathbb{P}$ observe that given δ there exists a constant $\tilde{C}(\delta)$ such that for any set Σ we have

$$\mathbb{P}(\Sigma) \leq \tilde{C}(\delta)\tilde{\mathbb{P}}(\Sigma) + \delta.$$

□

Let $\phi(h)$ be a smooth function of compact support. Denote

$$\mathcal{L}\phi = \sum_i a_i(x)\frac{d\phi}{dh_i} + \frac{1}{2}\sum_{ij}\sigma^2_{ij}(x)\frac{d^2\phi}{dh_i dh_j}, \quad \bar{\mathcal{L}}\phi = \int \mathcal{L}\phi d\mu_h.$$

Lemma 4. *For* $0 \leq n_1 \leq n_2 \leq n$

$$\mathbb{E}\left(\left[\phi(h(x_{n_2})) - \phi(h(x_{n_1})) - \varepsilon^2 \sum_{j=n_1}^{n_2} (\mathcal{L}\phi)(x_j)\right] |\mathcal{F}_{n_1}\right) = O(\varepsilon^3 n).$$

Proof. We have

$$\phi(h(x_{j+1})) - \phi(h(x_j)) = \varepsilon \sum_i \frac{d\phi}{dh_i} < \nabla h_i, F(x_j, \xi_j) >$$

$$+\varepsilon^2 \left[\frac{1}{2}\sum_{i,k}\frac{d^2\phi}{dh_i dh_k} < \nabla h_i(x_j), F(x_j,\xi_j) >< \nabla h_k(x_j), F(x_j,\xi_j) >\right.$$

$$\left. + \sum_i \frac{d\phi}{dh_i}(x_j)\left(< \nabla h_i(x_j), G(x_j,\xi_j) > + \frac{1}{2}D^2 h(x_j)(F(x_j,\xi_j), F(x_j,\xi_j))\right)\right] + O(\varepsilon^3).$$

Hence

$$\mathbb{E}(\phi(h(x_{j+1})) - \phi(h(x_j))|\mathcal{F}_j) = \varepsilon^2\mathbb{E}((\mathcal{L}\phi)(x_j)) + O(\varepsilon^3).$$

Summation over j completes the proof. □

5. Proof of Theorem 1.

Let $\mathbf{h}(t)$ be any limit point of $\{h_\varepsilon(t)\}$. By [13], Lemma 6.1.5, it is enough to show that for any smooth function ϕ of compact support

$$\phi(\mathbf{h}(t)) - \int_0^t (\bar{\mathcal{L}}\phi)(\mathbf{h}(s))ds$$

is martingale. To this end we have to show that for all $0 \leq s_1 < s_2 < s_p \leq t_1 < t_2 \leq T$ for all bounded functions ψ_i of compact support

$$\mathbb{E}\left(\prod_{i=1}^p \psi_i(s_i)\left[\phi(\mathbf{h}(t_2)) - \phi(\mathbf{h}(t_1)) - \int_{t_1}^{t_2}(\bar{\mathcal{L}}\phi)(\mathbf{h}(t))dt\right]\right) = 0.$$

In terms of our discrete system we need to show that

$$J = \mathbb{E}\left(R_\varepsilon\left[\phi(h(x_{t_2/\varepsilon^2})) - \phi(h(x_{t_1/\varepsilon^2})) - \varepsilon^2 \sum_{j=t_1/\varepsilon^2}^{t_2/\varepsilon^2} (\bar{\mathcal{L}}\phi)(h(x_j))\right]\right) \to 0.$$

where $R_\varepsilon = \prod_i \phi_i(h(x_{s_i/\varepsilon^2}))$. We have

$$J = \mathbb{E}\left(R_\varepsilon\left[\phi(h(x_{t_2/\varepsilon^2})) - \phi(h(x_{t_1/\varepsilon^2})) - \varepsilon^2 \sum_{j=t_1/\varepsilon^2}^{t_2/\varepsilon^2} (\mathcal{L}\phi)(x_j)\right]\right)$$

$$+\mathbb{E}\left(\varepsilon^2 R_\varepsilon\left[\sum_{j=t_1/\varepsilon^2}^{t_2/\varepsilon^2}\left[(\bar{\mathcal{L}}\phi)(h(x_j)) - (\mathcal{L}\phi)(x_j)\right]\right]\right) = I + I\!I.$$

By Lemma 4

$$I = \mathbb{E}\left(R_\varepsilon \mathbb{E}\left(\left[\phi(h(x_{t_2/\varepsilon^2})) - \phi(h(x_{t_1/\varepsilon^2})) - \varepsilon^2 \sum_{j=t_1/\varepsilon^2}^{t_2/\varepsilon^2} (\mathcal{L}\phi)(x_j)\right]\middle|\mathcal{F}_{t_1/\varepsilon^2}\right)\right) \to 0.$$

So we need to show that $I\!I \to 0$. By (ERG) given δ_1, δ_2 there exists T_0 such that if

$$\Omega = \{x : \quad \left|\frac{1}{T_0}\int_0^{T_0} (\mathcal{L}\phi)(\Phi(s)x) - (\bar{\mathcal{L}}\phi)(h(x))\right| < \delta_1\}$$

then $\mu(\Omega^c) \leq \delta_2$. Denote

$$\tau_k = \frac{t_1}{\varepsilon^2} + \frac{kT_0}{\varepsilon}, \quad \xi_k = \sum_{j=\tau_k}^{\tau_{k+1}-1}\left[(\bar{\mathcal{L}}\phi)(h(x_j)) - (\mathcal{L}\phi)(x_j)\right].$$

Then $I\!I = \varepsilon^2 \sum_k \mathbb{E}(R_\varepsilon \xi_k)$. Next

$$\mathbb{E}(R_\varepsilon\xi_k) = \mathbb{E}(R_\varepsilon\xi_k 1_\Omega(x_{\tau_k})) + \mathbb{E}(R_\varepsilon\xi_k 1_{\Omega^c}(x_{\tau_k})) = I\!I_a + I\!I_b.$$

By Corollary 3 for any δ

$$|I\!I_b| \leq \prod_i ||\psi_i||\,(C(\delta)\delta_2 + \delta)\,\frac{T_0}{\varepsilon}.$$

Now

$$I\!I_a = \mathbb{E}\left(1_\Omega(x_{\tau_k}) R_\varepsilon \mathbb{E}\left(\xi_k|\mathcal{F}_{\tau_k}\right)\right).$$

To estimate the inner expectation we split it as follows

$$\mathbb{E}(\xi_k|\mathcal{F}_{\tau_k}) = \mathbb{E}\left(\left[\sum_{j=\tau_k}^{\tau_{k+1}-1} (\mathcal{L}\phi)(x_j) - \frac{1}{\varepsilon}\int_0^{T_0} (\mathcal{L}\phi)(\Phi(t)x_{\tau_k})dt\right]\middle|\mathcal{F}_{\tau_k}\right) +$$

$$\left[\frac{1}{\varepsilon}\int_0^{T_0}(\mathcal{L}\phi)(\Phi(t)x_{\tau_k})dt-\frac{T_0}{\varepsilon}(\bar{\mathcal{L}}\phi)(h(x_{\tau_k}))\right]+$$

$$\mathbb{E}\left(\sum_{j=\tau_k}^{\tau_{k+1}-1}\left[(\bar{\mathcal{L}}\phi)(h(x_j))-(\bar{\mathcal{L}}\phi)(h(x_{\tau_k}))\right]|\mathcal{F}_{\tau_k}\right)=r_1+r_2+r_3.$$

For small ε we have $|r_1|\le\delta_1 T_0/\varepsilon$ by Averaging Principle (see e.g [9]), $|r_2|\le\delta_1 T_0/\varepsilon$ if $x_{\tau_k}\in\Omega$ by the definition of Ω and $|r_3|\le \text{Const}T_0^{3/2}/\sqrt{\varepsilon}$ by Lemma 1 and Cauchy–Schwartz. Hence

$$\mathbb{E}(R_\varepsilon\xi_k)\le\prod_i||\psi_i||\left(C(\delta)\delta_2+\delta+2\delta_1+\text{Const}\sqrt{T_0\varepsilon}\right)\frac{T_0}{\varepsilon}.$$

Summation over k gives

$$|I\!I|\le\prod_i||\psi_i||\left(C(\delta)\delta_2+\delta+2\delta_1+\text{Const}\sqrt{T_0\varepsilon}\right)T.$$

Since δ,δ_1 and δ_2 are arbitrary the result follows. □

Appendix A. Moments.

Proof of Lemma 1.

$$\tilde{S}_{n_2}-\tilde{S}_{n_1}=\varepsilon\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j)+\varepsilon^2\sum_{j=n_1}^{n_2-1}\tilde{Q}(x_j,\xi_j)+\varepsilon^3\sum_{j=n_1}^{n_2-1}\tilde{R}(x_j,\xi_j)$$

where we denoted $\tilde{Z}(x_j,\xi_j)=Z(x_j,\xi_j)1_{j<\tau_\varepsilon}$, $Z\in\{P,Q,R\}$. Hence by the triangle inequality it suffices to show that

$$\tilde{E}\left(\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j)\right)=0, \tag{12}$$

$$\tilde{E}\left(\left(\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j)\right)^2\right)=O(n_2-n_1), \tag{13}$$

$$\tilde{E}\left(\left(\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j)\right)^4\right)=O\left((n_2-n_1)^2\right) \tag{14}$$

where $\tilde{E}$ denotes $\mathbb{E}(\cdot|\mathcal{F}_{n_1})$. Since the estimates are done componentwise it suffices to consider the case when S and P are scalar functions. Then

$$\tilde{E}\left(\sum_{j=n_1}^{n_2-1}\tilde{P}\left(x_j,\xi_j\right)\right)=\sum_j\tilde{E}\left(\mathbb{E}(\tilde{P}(x_j,\xi_j)|\mathcal{F}_j)\right)=0$$

in view of (10). Next,

$$\tilde{E}((\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j))^2)=\sum_{j_1,j_2}\tilde{E}(\tilde{P}(x_{j_1},\xi_{j_1})\tilde{P}(x_{j_2},\xi_{j_2}))$$

If $j_1\neq j_2$ then taking $\mathbb{E}(\cdot|\mathcal{F}_{\max(j_1,j_2)})$ we obtain that the above term is 0. So there are only n_2-n_1 non-zero terms and (13) follows. Similarly

$$\tilde{E}\left(\left(\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j)\right)^4\right)=\sum_{j_1,j_2,j_3,j_4}\tilde{E}(\tilde{P}(x_{j_1},\xi_{j_1})\tilde{P}(x_{j_2},\xi_{j_2})\tilde{P}(x_{j_3},\xi_{j_3})\tilde{P}(x_{j_4},\xi_{j_4}))$$

Renumber the indices so that $j_1\leq j_2\leq j_3\leq j_4$. Then the previous argument shows that terms with $j_3\neq j_4$ are zero. Hence

$$\tilde{E}((\sum_{j=n_1}^{n_2-1}\tilde{P}(x_j,\xi_j))^4)$$

$$=6\sum_{j_3=n_1}^{n_2-1}\sum_{m_1=n_1}^{j_3-1}\sum_{m_2=n_1}^{j_3-1}\tilde{E}(\tilde{P}(x_{m_1},\xi_{m_1})\tilde{P}(x_{m_2},\xi_{m_2})\tilde{P}^2(x_{j_3},\xi_{j_3}))+O((n_2-n_1)^2)$$

where the second part counts terms with $j_2=j_3$. The first part equals

$$6\sum_{j_3=n_1}^{n_2-1}\tilde{E}\left(\left(\sum_{m=n_1}^{j_3-1}\tilde{P}(x_m,\xi_m)\right)^2\tilde{P}^2(x_{j_3},\xi_{j_3})\right)\leq \text{Const}\sum_{j_3=n_1}^{n_2-1}\tilde{E}\left(\left(\sum_{m=n_1}^{j_3-1}\tilde{P}(x_m,\xi_m)\right)^2\right)$$

$$\leq \text{Const}\sum_{j_3=n_1}^{n_2-1}j_3\leq \text{Const}(n_2-n_1)^2$$

where the second inequality follows from (13). This completes the proof of Lemma 1. □

Proof of Lemma 2. This Lemma follows from Lemma 1 and [2], Theorem 12.3. □

References

[1] Anosov D. V. *Averaging in systems of ODE with rapidly oscillating solutions,* Izvestiya Math. **24** (1960) 721–742.

[2] Billingsley P. *Convergence of probability measures,* John Wiley & Sons (1968) New York-London-Sydney.

[3] Bogachev V. I., Krylov N. V. & Rockner M. *On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions,* Comm. PDE **26** (2001) 2037–2080.

[4] Borodin A. N. *A limit theorem for the solutions of differential equations with a random right-hand side,* Th. Prob. Appl. **22** (1977) 482–497.

[5] Borodin A. N. & Freidlin, M. I. *Fast oscillating random perturbations of dynamical systems with conservation laws,* Ann. Inst. H. Poincare Probab. Statist. **31** (1995) 485–525.

[6] Cogburn R. & Ellison J. A. *A stochastic theory of adiabatic invariance,* Comm. Math. Phys. **149** (1992) 97–126.

[7] Cogburn R. & Ellison J. A. *A four-thirds law for phase randomization of stochastically perturbed oscillators and related phenomena,* Comm. Math. Phys.**166** (1994) 317–336.

[8] Freidlin M. I. & Wentzell A. D. *Averaging Principle for Stochastic Perturbations of Multifrequency Systems,* Stoch. Dyn. **3** (2003) 393–408.

[9] Khasminskii R. Z. *Stochastic processes defined by differential equations with a small parameter,* Th. Prob. Appl. **11** (1966), 211–228.

[10] Khasminskii R. Z. *A limit theorem for solutions of differential equations with a random right hand part,* Teor. Verojatnost. i Primenen **11** (1966) 444–462.

[11] Kifer Yu. *Stochastic versions of Anosov's and Neistadt's theorems on averaging,* Stoch. Dyn. **1** (2001) 1–21.

[12] Lyubich M. *The quadratic family as a qualitatively solvable model of chaos,* Notices AMS **47** (2000) 1042–1052.

[13] Stroock D. W. & Varadhan S. R. S. *Multidimensional diffusion processes,* Grundlehren der Mathematischen Wissenschaften **233,** (1979) Springer-Verlag, Berlin-New York.