AN AVERAGING PRINCIPLE FOR DIFFUSIONS IN FOLIATED SPACES

BY IVAN I. GONZALES-GARGATE[1] AND PAULO R. RUFFINO[2]

Universidade Estadual de Campinas

Consider an SDE on a foliated manifold whose trajectories lay on compact leaves. We investigate the effective behavior of a small transversal perturbation of order ε. An average principle is shown to hold such that the component transversal to the leaves converges to the solution of a deterministic ODE, according to the average of the perturbing vector field with respect to invariant measures on the leaves, as ε goes to zero. An estimate of the rate of convergence is given. These results generalize the geometrical scope of previous approaches, including completely integrable stochastic Hamiltonian system.

1. Introduction and set-up. Generally speaking, the original heuristic idea of an averaging principle refers to an intertwining of two dynamics where one of them is, in some sense, much slower and is affected somehow by the other faster dynamics. An averaging principle in this case refers to the possibility of approximating, in some topology, the slow dynamics considering only the average action or perturbation which the fast motion induces on it. These ideas appeared long ago, and as mentioned by Arnold ([1], page 287), they were implicitly contained in the works of Laplace, Lagrange and Gauss on celestial mechanics; literature on the matter can be found, for example, [1], Sanders, Verhulst and Murdoch [17] and references therein, among many others. Presently, regarding stochastic systems, averaging has been quite an active research field on which there is also a vast literature on the topic. Interesting quick historical overviews can be found in Li ([13], page 806), Kabanov and Pergamenshchikov ([9], Appendix), ([17], Appendix A). Among many other works somehow related to the topic, we refer to Khasminski and Krylov [11], Bakhtin and Kifer [4], Sowers [18], Namachchvaya and Sowers [14], Borodin and Freidlin [5], Kabanov and Pergamenshchikov [9] and references therein.

The specific problem that we address in this article is a perturbation of a diffusion in a foliated manifold M such that the unperturbed random trajectories lay on the leaves. The perturbations are taken transversally to the leaves of the foliation.

Here, the slow system is the transversal component, and the fast system is given by the rescaled $y_{t/\varepsilon}^{\varepsilon}$, where y_t^{ε} is the solution of the original SDE perturbed by a vector field εK.

Our results generalize the recent approach by Li [13] on an averaging principle for a completely integrable stochastic Hamiltonian system. In this article, as in the classical approach (see, e.g., [1]), Li has explored the benefits of well-structured geometrical coordinates in the state space given by the coordinates of the Liouville torus; these benefits include vanishing Itô–Stratonovich correction terms besides also vanishing covariant derivative of Hamiltonian vector fields in tangent directions to the leaves. We prove here that an averaging principle also holds in a generalized geometrical scope, so that this averaging phenomenon occurs independently of symplectic structures (but with possibly slower rates of convergence). Compared to Li's previous result ([13], Lemma 3.2), where the estimates contain a term of order $1/\sqrt{t}$, our corresponding estimates in Lemma 3.1 are continuous at $\varepsilon = t = 0$. Some of the rates of convergence of [13], Lemma 3.1, are recovered as particular cases in Corollaries 2.2 and 2.3. In the main result, we show that in the average, the approximation goes to zero with order $\varepsilon^{\alpha} + \eta(|\ln \varepsilon|^{-\beta/p})$, where η is the rate of convergence of the unperturbed system to the ergodic average on the leaves, $\beta \in (0, 1/2)$ and $\alpha \in (0, 1)$, so that, depending on η, convergences rate can be much faster or much slower than in [13].

The set-up. Let M be a smooth Riemannian manifold with an n-dimensional smooth foliation; that is, M is endowed with an integrable regular distribution of dimension n; for a definition and further properties of foliated spaces, see, for example, the initial chapters of Tondeur [19], Walczak [20], among others. We denote by L_x the leaf of the foliation passing through a point $x \in M$. For simplicity, we shall assume that the leaves are compact and that each leaf L_x has a tubular neighborhood $U \subset M$ where U is diffeomorphic to $L_x \times V$, where $V \subset \mathbf{R}^d$ is an open bounded neighborhood of the origin and d is the codimension of the foliation. We shall assume an SDE in M whose solution flow preserves the foliation; that is, we consider a Stratonovich equation

$$(1) \qquad dx_t = X_0(x_t)\,dt + \sum_{k=1}^{r} X_k(x_t) \circ dB_t^k,$$

where the smooth vector fields X_k are foliated in the sense that $X_k(x) \in T_x L_x$, for $k = 0, 1, \ldots, r$. Here $B_t = (B_t^1, \ldots, B_t^r)$ is a standard Brownian motion in $\mathbf{R}^r$ with respect to a filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{F}, \mathbf{P})$. For an initial condition x_0, the trajectories of the solution x_t in this case lay on the leaf L_{x_0} a.s. Moreover, there exists a (local) stochastic flow of diffeomorphisms $F_t : M \to M$ which, restricted to the initial leaf, is a flow in the compact L_{x_0}.

For a smooth vector field K in M, we shall denote the perturbed system by y_t^{ε} which satisfies the SDE

$$(2) \qquad dy_t^{\varepsilon} = X_0(y_t^{\varepsilon})\,dt + \sum_{k=1}^{r} X_k(y_t^{\varepsilon}) \circ dB_t^k + \varepsilon K(y_t^{\varepsilon})\,dt,$$

with the same initial condition $y_0^\varepsilon = x_0$ of the unperturbed system x_t.

Our main result, Theorem 4.1, says that locally the transversal behavior of $y_{t/\varepsilon}^\varepsilon$ can be approximated in the average by an ordinary differential equation in the transversal space whose coefficients are given by the average of the transversal component of the perturbation K with respect to the invariant measure on the leaves for the original dynamics of equation (1). The reader will notice by the end of the proofs that compactness of the leaves in fact can be substituted by some other boundedness conditions, added also to some rather technical adjustments which we will not address here. In Sections 2 and 3, we present the main lemmas. The main result appears in Section 4, where we also present a simple illustrative example. In particular, under some symmetry hypotheses on a foliated system embedded in an Euclidean space, we use the main theorem to conclude that Lyapunov exponents in the transversal direction must tend to nonpositive values as ε goes to zero; cf. Proposition 4.2.

2. Preliminary results. *The coordinate system.* Given an initial condition $x_0 \in M$, let $U \subset M$ be a bounded neighborhood of the corresponding leaf L_{x_0} such that there exists a diffeomorphism $\varphi : U \to L_{x_0} \times V$, where $V \subset \mathbf{R}^d$ is a connected open set containing the origin. The neighborhood U can be considered small enough such that the closure $\bar{U} \subset M$ and the derivative of φ is bounded. For a fixed diffeomorphism φ, the space V will be called the *vertical space*. For simplicity, the second (vertical) coordinate of a point $p \in U$ will be called the vertical projection $\pi(p) \in V$, that is, $\varphi(p) = (u, \pi(p))$ for some $u \in L_{x_0}$. Hence for any fixed $v \in V$, the inverse image $\pi^{-1}(v)$ is the compact leaf L_x, where x is any point in U such that the vertical projection $\pi(x) = v$. In Section 3 below we need the components of the vertical projection, which shall be denoted by

$$\pi(p) = (\pi_1(p), \dots, \pi_d(p)) \in V \subset \mathbf{R}^d$$

for any $p \in U$. Natural examples of these coordinates systems appear if we consider compact foliation given by the inverse image of submersions: values in the image space provide local coordinates for the vertical space V.

The next lemma gives information on the order of which the perturbed trajectories y_t^ε approach the unperturbed x_t when one varies ε and t in equation (2); it will be used to prove that the dynamics of the rescaled system $y_{(t/\varepsilon)}^\varepsilon$ is such that its time average for any function g in M approximates the time average of the spacial average of g on the leaves; see Lemma 3.1.

LEMMA 2.1. *Let τ^ε be the exit time of the process y_t^ε from the neighborhood U of our coordinate system as above. For any locally Lipschitz continuous function $f : M \to \mathbf{R}$ and $2 \le p < \infty$, we have*

$$\left[\mathbb{E} \left(\sup_{s \le t \wedge \tau^\varepsilon} |f(y_s^\varepsilon) - f(x_s)|^p \right) \right]^{1/p} \le K_1 \varepsilon t e^{K_2 t^p},$$

where K_1, $K_2 \geq 0$ are constants depending on upper bounds of the norms of the perturbing vector field K, on the Lipschitz coefficients of f and on the derivatives of $X_0, X_1, \ldots, X_r$ with respect to the coordinate system.

PROOF. Initially write x_t and y_t^ε, the solutions of equations (1) and (2), respectively, according to the coordinates given by the diffeomorphism φ: denote $(u_t, v_t) := \varphi(x_t)$ and $(u_t^\varepsilon, v_t^\varepsilon) := \varphi(y_t^\varepsilon)$. Then

$$(3) \qquad \begin{aligned} \left| f(y_t^\varepsilon) - f(x_t) \right| &= \left| f \circ \varphi^{-1}(u_t^\varepsilon, v_t^\varepsilon) - f \circ \varphi^{-1}(u_t, v_t) \right| \\ &\leq C \left| u_t^\varepsilon - u_t \right| + C \left| v_t^\varepsilon - v_t \right|, \end{aligned}$$

for some constant $C \geq 0$, where we have used the fact that φ has bounded derivative. The inequality of the statement will be obtained from inequality (3), treating separately the norms of the summands above.

For the summand coming from the vertical components, we have that the unperturbed $v_t \equiv 0$. The perturbing vector field K in our coordinate system is given by $d\varphi(K) = K_h + K_v$ where K_h and K_v are the horizontal and the vertical components of $d\varphi(K)$ in the tangent space $TL_{x_0} \times V$. The equation for v_t^ε is given, simply, by

$$dv_t^\varepsilon = \varepsilon K_v(u_t^\varepsilon, v_t^\varepsilon) \, dt.$$

Therefore

$$(4) \qquad \begin{aligned} \sup_{s \leq t \wedge \tau^\varepsilon} \left| v_s^\varepsilon - v_s \right| &\leq \varepsilon \sup_{s \leq t \wedge \tau^\varepsilon} \int_0^s \left| K_v(u_s^\varepsilon, v_s^\varepsilon) \right| \, ds \\ &\leq \varepsilon t \sup_{x \in U} \left| K_v(x) \right| \leq C_1 \varepsilon t, \end{aligned}$$

where $C_1 = \sup_{x \in U} |K(x)|$.

For the horizontal summand $|u_t^\varepsilon - u_t|$ in inequality (3), we consider an embedding $i : M \to \mathbf{R}^N$ of the compact submanifold L_{x_0} in an Euclidean space $\mathbf{R}^N$ with a sufficiently large integer N. In $\mathbf{R}^N$ we have, in coordinates, $i(u_t^\varepsilon) = (u_t^{\varepsilon,1}, \ldots, u_t^{\varepsilon,N})$. We look for an equation of $i(u_t^\varepsilon) \in \mathbf{R}^N$. Let $\tilde{b}_k$ be the vector fields on the embedded $i(L_{x_0})$ induced by the original vector fields X_k of equation (1), for $k = 0, 1, \ldots, r$. Precisely, if $d\varphi(X_k) = X_k^h + X_k^v$ is the decomposition on the horizontal and vertical components, then $\tilde{b}_k(u, v) = di(X_k^h(u, v))$ with $u \in i(M)$ and $v \in V$. By compactness, we extend each $\tilde{b}_k$ to a vector field b_k in a tubular neighborhood of $i(L_{x_0}) \subset \mathbf{R}^N$ such that b_k is constant in orthogonal fibres to $i(M)$ in this tubular neighborhood, for each $k = 0, 1, \ldots, r$. Finally, in canonical coordinates in $\mathbf{R}^N$, we have, for $i = 1, \ldots, N$,

$$(5) \qquad du_t^{\varepsilon,i} = \sum_{k=1}^r b_k^i(u_t^\varepsilon, v_t^\varepsilon) \circ dB_t^k + b_0^i(u_t^\varepsilon, v_t^\varepsilon) \, dt + \varepsilon K_h^i(u_t^\varepsilon, v_t^\varepsilon) \, dt.$$

Again, by compactness, the image of the embedding $i(L_{x_0})$ has an induced metric from $\mathbf{R}^N$ which is uniformly equivalent to the original metric in L_{x_0}. Moreover note that, by the choice of the neighborhood U, the induced vector fields $b_0, b_1, \ldots, b_r, K_h$ and their derivatives are bounded. From equation (5) we have in each ith component, for $s < \tau^\varepsilon$,

$$
(6) \quad u_s^{\varepsilon,i} - u_s^i = \sum_{k=1}^{r} \int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) \circ dB_r^k
$$

$$
(7) \quad + \int_0^s \left(b_0^i(u_r^\varepsilon, v_r^\varepsilon) - b_0^i(u_r, v_r) \right) dr + \varepsilon \int_0^s K_h^i(u_r^\varepsilon, v_r^\varepsilon) \, dr.
$$

In terms of the Itô integral,

$$
\int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) \circ dB_r^k
$$

$$
= \int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) dB_r^k
$$

$$
+ \frac{1}{2} \int_0^s \left[\nabla b_k^i \cdot b_k(u_r^\varepsilon, v_r^\varepsilon) - \nabla b_k^i \cdot b_k(u_r, v_r) \right] dr.
$$

Hence, taking the absolute values in both sides of equation (7) we get, for each i,

$$
\left| u_s^{\varepsilon,v} - u_s^i \right| \leq \sum_{k=1}^{r} \left| \int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) dB_r^k \right|
$$

$$
(8) \quad + \frac{1}{2} \sum_{k=1}^{r} \int_0^s \left| \nabla b_k^i \cdot b_k(u_r^\varepsilon, v_r^\varepsilon) - \nabla b_k^i \cdot b_k(u_r, v_r) \right| dr
$$

$$
+ \int_0^s \left| b_0^i(u_r^\varepsilon, v_r^\varepsilon) - b_0^i(u_r, v_r) \right| dr + \varepsilon \int_0^s \left| K_h^i(u_r^\varepsilon, v_r^\varepsilon) \right| dr.
$$

Functions b_0^i and $(\nabla b_k^i \cdot b_k)$ are Lipschitz; hence for a common constant C_2,

$$
\left| u_s^{\varepsilon,i} - u_s^i \right| \leq \left| \sum_{k=1}^{r} \int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) dB_r^k \right|
$$

$$
(9) \quad + C_2 \int_0^s \left| v_r^\varepsilon - v_r \right| dr + C_2 \int_0^s \left| u_r^\varepsilon - u_r \right| dr + \varepsilon s \sup_U |K_h|.
$$

The first deterministic integral, together with inequality (4) yields

$$
\left| u_s^{\varepsilon,i} - u_s^i \right| \leq \sum_{k=1}^{r} \left| \int_0^s \left(b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r) \right) dB_r^k \right| + C_1 C_2 \varepsilon s^2
$$

$$
+ C_2 \int_0^s \left| u_r^\varepsilon - u_r \right| dr + C_1 \varepsilon s.
$$

Now, for $p \geq 1$, there exists a constant C_3 such that

$$|u_s^{\varepsilon,i} - u_s^i|^p \leq C_3 \sum_{k=1}^{r} \left| \int_0^s (b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r)) \, dB_r^k \right|^p + C_3(C_1 C_2 \varepsilon s^2)^p$$

$$+ C_3 C_2^p \left(\int_0^s |u_r^\varepsilon - u_r| \, dr \right)^p + C_3(C_1 \varepsilon s)^p.$$

The Cauchy–Schwarz inequality yields

$$|u_s^{\varepsilon,i} - u_s^i|^p \leq C_3 \sum_{k=1}^{r} \left| \int_0^s (b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r)) \, dB_r^k \right|^p + C_3(C_1 C_2 \varepsilon s^2)^p$$

$$+ C_3 C_2^p s^{p-1} \int_0^s |u_r^\varepsilon - u_r|^p \, dr + C_3(C_1 \varepsilon s)^p.$$

Hence

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^{\varepsilon,i} - u_s^i|^p$$

$$\leq C_3 \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} \sum_{k=1}^{r} \left| \int_0^s (b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r)) \, dB_r^k \right|^p + C_3(C_1 C_2 \varepsilon t^2)^p$$

$$+ C_3 C_2^p t^{p-1} \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} \int_0^s |u_r^\varepsilon - u_r|^p \, dr + C_3(C_1 \varepsilon t)^p$$

$$\leq C_4 \sum_{k=1}^{r} \mathbb{E} \left[\int_0^{t \wedge \tau^\varepsilon} (b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r))^2 \, dr \right]^{p/2} + C_3(C_1 C_2 \varepsilon t^2)^p$$

$$+ C_3 C_2^p t^{p-1} \int_0^t \mathbb{E} \left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr + C_3(C_1 \varepsilon t)^p,$$

where we have used classical L^p-inequality for martingales (e.g., Revuz and Yor [15]). Using again the Lipchitz property of each b_k for the terms in the brackets above,

(10)
$$\sum_{k=1}^{r} \int_0^{t \wedge \tau^\varepsilon} (b_k^i(u_r^\varepsilon, v_r^\varepsilon) - b_k^i(u_r, v_r))^2 \, dr$$

$$\leq 2C_2^2 \left(\int_0^{t \wedge \tau^\varepsilon} |v_r^\varepsilon - v_r|^2 \, dr + \int_0^{t \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^2 \, dr \right)$$

$$\leq 2C_2^2 \left(\int_0^t C_1^2 \varepsilon^2 r^2 \, dr + \int_0^{t \wedge \tau^\varepsilon} \sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^2 \, dr \right)$$

$$\leq C_2^2 C_1^2 \varepsilon^2 t^3 + 2C_2^2 \int_0^{t \wedge \tau^\varepsilon} \sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^2 \, dr.$$

We end up with

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^{\varepsilon,i} - u_s^i|^p$$

$$\leq C_4 \sum_{k=1}^r \mathbb{E}\left[C_2^2 C_1^2 \varepsilon^2 t^3 + 2C_2^2 \int_0^{t \wedge \tau^\varepsilon} \sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^2 \, dr \right]^{p/2}$$

(11)

$$+ C_3 C_2^p t^{p-1} \int_0^t \mathbb{E}\left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr$$

$$+ C_3 (C_1 C_2 \varepsilon t^2)^p + C_3 (C_1 \varepsilon t)^p.$$

For $p \geq 2$ one can use Cauchy–Schwarz again to conclude that there exists a constant C_5 such that the last expression is less than or equal to

$$C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_5 C_2^p t^{(p-2)/2} \int_0^{t \wedge \tau^\varepsilon} \mathbb{E} \sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \, dr + C_3 (C_1 C_2 \varepsilon t^2)^p$$

$$+ C_3 C_2^p t^{p-1} \int_0^t \mathbb{E}\left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr + C_3 (C_1 \varepsilon t)^p$$

$$= C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_3 (C_1 C_2 \varepsilon t^2)^p + C_3 (C_1 \varepsilon t)^p$$

$$+ (C_5 C_2^p t^{(p-2)/2} + C_3 C_2^p t^{p-1}) \int_0^t \mathbb{E}\left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr.$$

Now, summing up over i in the inequalities above leads to

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p$$

$$\leq C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_3 (C_1 C_2 \varepsilon t^2)^p + C_3 (C_1 \varepsilon t)^p$$

$$+ (C_5 C_2^p t^{(p-2)/2} + C_3 C_2^p t^{p-1}) \int_0^t \mathbb{E}\left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr.$$

We use now the integral form of Gronwall's inequality to find that

$$\mathbb{E}\left(\sup_{s \leq t \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) \leq C_6 \varepsilon^p t^p (1 + t^p) \exp\{C_7 (t^{p/2} + t^p)\}$$

$$\leq C_8 \varepsilon^p t^p (1 + t^p) \exp\{C_9 t^p\}.$$

Going back to inequality (3), now we have

$$|f(y_t^\varepsilon) - f(x_t)|^p \leq C_{10} |v_t^\varepsilon - v_t|^p + C_{10} |u_t^\varepsilon - u_t|^p.$$

Hence

$$\mathbb{E}\left(\sup_{s \leq t \wedge \tau^\varepsilon} |f(y_s^\varepsilon) - f(x_s)|^p \right) \leq C_{10} \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |v_s^\varepsilon - v_s|^p + C_{10} \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^\varepsilon - u_s|^p$$

$$\leq C_{11} \varepsilon^p t^p + C_{12} \varepsilon^p t^p (1 + t^p) \exp(C_9 t^p)$$

$$\leq C_{13} \varepsilon^p t^p (1 + t^p) \exp(C_9 t^p).$$

From here, finally, one concludes that there exist constants K_1 and K_2 such that

$$\mathbb{E}\Big(\sup_{s \leq t \wedge \tau^\varepsilon} |f(y_s^\varepsilon) - f(x_s)|^p \Big)^{1/p} \leq K_1 \varepsilon t \exp(K_2 t^p). \qquad \square$$

On the order of the estimates in the lemma. Simple examples show that the exponential order on the rate of convergence of Lemma 2.1 cannot be improved. This fact does not come from the vertical component, which is deterministic and easily bounded by $\varepsilon t C_1$; but, rather, it comes from the horizontal component. Consider the following example: a one-dimensional horizontal dynamics which locally is linear $dx_t = x_t \circ dB_t$ and a perturbing transversal vector field K which has constant unitary component on the horizontal direction, that is, $dy_t^\varepsilon = y_t^\varepsilon \circ dB_t + \varepsilon \, dt$. The L^p norm of the difference

$$\mathbf{E}[x_t - y_t^\varepsilon]^p \leq \varepsilon \int_0^t \mathbf{E}[\exp\{p(B_t - B_s)\}] \, ds$$

increases exponentially with respect to time t.

The next corollary includes the case of a completely integrable stochastic Hamiltonian system when one uses the action-angle coordinates; cf. Li [13], Lemma 3.1.

COROLLARY 2.2.   *If the vector fields $X_0, \ldots, X_r$ depend only on the vertical coordinate (null derivative in the directions of the leaves, as in the Hamiltonian case [13]), then the estimates above can be improved, and for $p \geq 1$ there exists a constant K_1 such that*

$$\Big[\mathbb{E}\Big(\sup_{s \leq t \wedge \tau^\varepsilon} |f(y_s^\varepsilon) - f(x_s)|^p \Big) \Big]^{1/p} \leq K_1 \varepsilon (t + t^2).$$

PROOF.   In this case the correction term of the Stratonovich stochastic integral in terms of the Itô integral in inequality (8) vanishes, as does the determinist integration of $|u_r^\varepsilon - u_r|$ in inequalities (9) and (10). Hence inequality (11) improves to

$$(12) \quad \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^{\varepsilon,i} - u_s^i|^p \leq C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_3 (C_1 C_2 \varepsilon t^2)^p + C_3 (C_1 \varepsilon t)^p.$$

The argument in the rest of the proof follows straightforwardly for $p \geq 1$ skipping Gronwall inequality. $\square$

The next corollary includes the case $X_0 \equiv 0$, cf. [13], Lemma 3.1(2), for stochastic Hamiltonian systems with an action-angle coordinate system.

COROLLARY 2.3.   *If, in addition to the conditions of Corollary 2.2 above, we have that the deterministic vector field X_0 is constant when represented with respect to a certain coordinate system in U (i.e., b_0 has null derivative w.r.t. u and v), then for $p \geq 1$ the estimates can be improved further to $K_1 \varepsilon (t + t^{3/2})$.*

PROOF. Besides the vanishing terms already mentioned above, the second deterministic integral on the right-hand side of inequality (8) also vanishes. Hence inequality (11) simplifies further to

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^{\varepsilon,i} - u_s^i|^p \leq C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_3 (C_1 \varepsilon t)^p. \qquad \square$$

Yet, from the proof of the Lemma 2.1, we have the following:

REMARK 2.4. For $1 \leq p < 2$ and t sufficiently small, there exist constants K_1 and K_2 such that

$$(13) \qquad \left[\mathbb{E} \left(\sup_{s \leq t \wedge \tau^\varepsilon} |f(y_s^\varepsilon) - f(x_s)|^p \right) \right]^{1/p} \leq K_1 \varepsilon t \exp (K_2 t^p).$$

PROOF. One can no longer use Cauchy–Schwarz after inequality (11). Alternatively, from (11), use that

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^\varepsilon - u_s|^p$$

$$\leq C_5 (C_2 C_1 \varepsilon t^{3/2})^p + C_5 C_2^p t^{p/2} \mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^{\varepsilon,i} - u_s^i|^p + C_3 (C_1 C_2 \varepsilon t^2)^p$$

$$+ C_3 C_2^p t^{p-1} \int_0^t \mathbb{E} \left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr + C_3 C_1^p \varepsilon^p t^p.$$

If we fix an $0 < \delta < 1$ and take t sufficiently small such that $1 - C_5 C_2^p t^{p/2} > \delta$, then

$$\mathbb{E} \sup_{s \leq t \wedge \tau^\varepsilon} |u_s^\varepsilon - u_s|^p$$

$$\leq \delta^{-1} C_5 (C_2 C_1 \varepsilon t^{3/2})^p + \delta^{-1} C_3 (C_1 C_2 \varepsilon t^2)^p$$

$$+ \delta^{-1} C_3 C_2^p t^{p-1} \int_0^t \mathbb{E} \left(\sup_{s \leq r \wedge \tau^\varepsilon} |u_r^\varepsilon - u_r|^p \right) dr + \delta^{-1} C_3 C_1^p \varepsilon^p t^p,$$

and the calculation is completed as before using the integral version of Gronwall's inequality. $\square$

3. Averaging functions on the leaves. Consider a differentiable function $g : M \to \mathbb{R}$. The leaf L_p passing through a point $p \in M$ contains the support of an invariant measure μ_p for the unperturbed system (1). We shall assume that each of such μ_p is ergodic. We define a function $Q^g : V \subset \mathbf{R}^d \to \mathbf{R}$ such that $Q^g(v)$ is the average of g with respect to the invariant measure on the corresponding leaf $\pi^{-1}(v)$. Namely, if v is the vertical coordinate of p, that is, $\varphi(p) = (u, v)$, then

$$Q^g(v) = \int_{L_p} g(x) \, d\mu_p(x).$$

By the ergodic theorem, the function Q^g will also be recovered as the time average along the unperturbed trajectories. We are now concerned with this time average.

On the rate of convergence on the leaves. For deterministic dynamics it is known that there are no estimates for the speed of convergence in the ergodic theorem: it can be as slow as any prescribed decreasing rate; see, for example, Kakutani and Petersen [10] and Krengel [12]. The speed of convergence, in general, depends both on the dynamics and on the function whose average is considered. In particular, the main theorem in [12] states that given an ergodic invertible transformation in the interval [0, 1], and a prescribed rate of convergence (given in terms of a sequence converging to zero which, by comparison is the prescribed rate of convergence), there exists a continuous function whose rate of convergence of the ergodic theorem is slower than the convergence to zero of the prescribed sequence. The result on the rate of convergence holds almost surely and in the L^p norm.

With this result in mind, one easily construct an example of a deterministic continuous dynamics in the torus such that for any prescribed (no matter slow) rate of convergence, there exists a function in the torus such that the rate of convergence of the ergodic theorem is slower than the order of convergence prescribed. Indeed, consider the deterministic transformation in the interval [0, 1] given by a rotation by an irrational angle γ in the circle S^1 after the identification of extremal points of the interval [alternatively, $x \mapsto x + \gamma \pmod 1$]. Given a prescribed (slow) rate of convergence, Krengel's theorem guarantees that there exists a continuous function $f(x)$ in [0, 1] such that the rate of convergence for this function is slower than the prescribed speed. Remarks in [12], page 6, say that f can be taken continuous in S^1. Represent the flat torus in the square $T^2 = [0, 1] \times [0, 1]$ with the appropriate identifications. Let $0 = x_0 < x_1 < x_2 < x_3 < x_4 = 1$ be a partition of the (horizontal) interval [0, 1], whose subintervals will be denoted by I_1, I_2, I_3 and I_4. Consider the dynamics in T^2 described in terms of its velocities as: horizontal component is unitary (constant); it is purely horizontal when $x \in I_1 \cup I_2 \cup I_3$ and makes a rotation in the y-coordinate (vertical) which depends smoothly in x, that is, twisting the torus vertically when $x \in I_4$, such that the total rotation along the interval I_4 is given by the irrational angle γ. Consider the continuous function $\tilde{f}(x, y)$ in the torus T^2 given by $\tilde{f}(x, y) = f(y)$, if $x \in I_2$, $|\tilde{f}(x, y)|$ decay uniformly in y to zero along $x \in I_3$, $\tilde{f}(x, y)$ vanishes if $x \in I_4$ and $|\tilde{f}(x, y)|$ increase uniformly in y when $x \in I_1$. Here, by "increase or decrease uniformly on y," we mean that the function $f(y)$ is multiplied by a factor in [0, 1] which depends only on x. The function $\tilde{f}$ vanishes when the flow performs the vertical rotation of angle γ in the torus. The rate of convergence of the ergodic theorem for this continuous deterministic system in T^2 is as slow as the prescription for the original pure rotation in S^1.

The deterministic example above shows that, in our context (which is not necessarily an elliptic diffusion as in [13]), there are no optimal estimates on the rate of convergence of the ergodic theorem to the averaging function Q^g. We proceed our calculations on the estimates on the rate of convergence of the averaging principle

considering a certain rate of convergence $\eta(t)$ which will depend on the unperturbed system and on the perturbing vector field. We remark that this dependence of an averaging principle on a certain rate of convergence also had to be used in the classical approach as in Freidlin and Wentzell ([8], Chapter 7, Section 9), where they deal with an averaging principle with convergence in probability rather than in L^p.

The functions which we are interested in averaging on the leaves are the components of the vertical coordinate of the perturbing vector field K. Hence function g is going to denote one of the real differentiable functions $d\pi_1(K), \ldots, d\pi_d(K)$.

For a fixed initial condition for the unperturbed system $x_0 \in U$, denote the rate of convergence in L^p by $\eta(x_0, t)$. More precisely, let $\eta(x_0, t)$ be a positive, asymptotically decreasing to zero real function such that

$$\left[\mathbb{E} \left| \frac{1}{t} \int_0^t g\big(F_r(x_0)\big)\, dr - Q^g\big(\pi(x_0)\big) \right|^p \right]^{1/p} \leq \eta(x_0, t),$$

for all $g = d\pi_i(K)$, with $i = 1, \ldots, d$. Continuity of the infinitesimal generator on the transversal direction and compactness of the leaves implies that taking supremum, there exists a rate of converge $\eta(t)$ in the neighborhood U such that

$$\eta(x, t) \leq \eta(t),$$

for any initial condition $x \in U$. Obviously, we have that estimate $\eta(t)$ is bounded and goes to zero when t goes to infinity. Examples in our context include: (1) uniformly elliptic diffusions as in the Hamiltonian stochastic systems (cf. [13]), where we have $\eta(t) = K/\sqrt{t}$; (2) rotations in S^1 (see Section 4.1) $\eta(t) = K/t$; (3) constant averaged function g, then $\eta(t) \equiv 0$.

We assume that $Q^g(v)$ has some degree of continuity with respect to v. This assumption appears in two levels in Lemma 3.4: (1) Riemann integrability of $Q^g(\pi(y_r^\varepsilon))$ with respect to r guarantees the convergence to zero; (2) α-Hölder continuity guarantees the rate of convergence.

The next lemma estimates the time average of the difference between function g and Q^g along the *perturbed* system y_t^ε. Here again, the stopping time τ^ε denotes the first exit time of the open neighborhood $U \subset M$ which is diffeomorphic to $L_{x_0} \times V$. We have the following estimates for the difference of the averages of functions g and Q^g.

LEMMA 3.1. *Let g be one of the functions given by the vertical coordinates of the perturbing vector field K; that is, let $g \in \{d\pi_1(K), \ldots, d\pi_d(K)\}$ and $Q^g : V \to \mathbf{R}$ be its corresponding average on the leaves. For $s, t \geq 0$ write*

$$\delta(\varepsilon, t) = \int_{s \wedge \varepsilon \tau^\varepsilon}^{(s+t) \wedge \varepsilon \tau^\varepsilon} g\big(y_{r/\varepsilon}^\varepsilon\big) - Q^g\big(\pi\big(y_{r/\varepsilon}^\varepsilon\big)\big)\, dr.$$

Then $\delta(\varepsilon, t)$ goes to zero when t or ε tend to zero.

Moreover, if Q^g is α-Hölder continuous, then for $p \geq 2$, $0 < \alpha' < \alpha$ and any $\beta \in (0, 1/2)$ we have the following estimates:

$$\left(\mathbb{E} \sup_{s \leq t} |\delta(\varepsilon, s)|^p \right)^{1/p} \leq t\varepsilon^{\alpha'} h(t, \varepsilon) + t\eta \big(t |\ln \varepsilon|^{(2\beta)/p} \big),$$

where $\eta(t)$ is the rate of convergence in L^p of the ergodic averages in the unperturbed trajectories on the leaves as defined above, and $h(t, \varepsilon)$ is continuous for $t, \varepsilon > 0$ and converges to zero when $(t, \varepsilon) \to 0$.

PROOF. The proof consists of considering a convenient partition of the interval $(\frac{s}{\varepsilon} \wedge \tau^\varepsilon, \frac{s+t}{\varepsilon} \wedge \tau^\varepsilon)$ where we can get the estimates by comparing in each subinterval the average of the flow of the original system (on the corresponding leaf) with the average of the perturbed flow (possibly transversal to the leaves). These estimates in each subinterval are obtained using Lemma 2.1, so, a key point in the proof is a careful choice of the increments of such a convenient partition.

For sufficiently small ε, we take the following assignment of increments:

$$\Delta t = \left(\frac{s+t}{|\ln \varepsilon|^{-2\beta/p}} \right) \wedge \tau^\varepsilon - \left(\frac{s}{|\ln \varepsilon|^{-2\beta/p}} \right) \wedge \tau^\varepsilon.$$

Hence, the partition $t_n = \frac{s}{\varepsilon} \wedge \tau^\varepsilon + n\Delta t$, for $0 \leq n \leq N$, is such that

$$\frac{s}{\varepsilon} \wedge \tau^\varepsilon = t_0 < t_1 < \cdots < t_N \leq \frac{s+t}{\varepsilon} \wedge \tau^\varepsilon,$$

with $N = N(\varepsilon) = \lfloor \varepsilon^{-1} |\ln \varepsilon|^{-2\beta/p} \rfloor$ where $\lfloor x \rfloor$ denotes the integer part of x.

Initially we represent the left-hand side as the sum

$$\varepsilon \int_{s/\varepsilon \wedge \tau^\varepsilon}^{((s+t)/\varepsilon) \wedge \tau^\varepsilon} g(y_r^\varepsilon)\, dr = \varepsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(y_r^\varepsilon)\, dr + \varepsilon \int_{t_N}^{((s+t)/\varepsilon) \wedge \tau^\varepsilon} g(y_r^\varepsilon)\, dr.$$

Denote by θ_t the canonical shift operator on the probability space. Let $F_t(\cdot, \omega)$ with $t \geq 0$ be the flow of the original unperturbed system in M. Triangular inequality splits our calculation into four parts,

$$(14) \qquad\qquad |\delta(\varepsilon, t)| \leq |A_1| + |A_2| + |A_3| + |A_4|,$$

where

$$A_1 = \varepsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \big[g(y_r^\varepsilon) - g(F_{r-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega))) \big]\, dr,$$

$$A_2 = \varepsilon \sum_{n=0}^{N-1} \left[\int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega)))\, dr - \Delta t\, Q^g(\pi(y_{t_n}^\varepsilon)) \right],$$

$$A_3 = \sum_{n=0}^{N-1} \varepsilon \Delta t\, Q^g(\pi(y_{t_n}^\varepsilon)) - \int_{s \wedge \varepsilon \tau^\varepsilon}^{(s+t) \wedge \varepsilon \tau^\varepsilon} Q^g(\pi(y_{r/\varepsilon}^\varepsilon))\, dr,$$

$$A_4 = \varepsilon \int_{t_N}^{((s+t)/\varepsilon) \wedge \tau^\varepsilon} g(y_r^\varepsilon) \, dr.$$

We proceed by showing that each of the processes A_1, A_2, A_3 and A_4 above tends to zero uniformly on compact intervals. In what follows, we will explore many times the fact that for $a > 0$ and $b \in \mathbf{R}$, $\varepsilon^a |\ln \varepsilon|^b$ goes to zero as $\varepsilon \searrow 0$. Hence, by construction, except when $\tau^\varepsilon \leq s$ (where, restricted to which the lemma is trivial) both Δt and N go to infinity when ε tends to zero. The increment Δt is a random variable bounded by $t|\ln \varepsilon|^{2\beta/p}$. Using that the exit time τ^ε has a lower bound estimate $\tau^\varepsilon \geq \frac{C}{\varepsilon}$ for a positive constant C independently of ω [due to equation (4)], one sees that, in fact, the increment $\Delta t = t|\ln \varepsilon|^{2\beta/p}$ when ε is sufficiently small, independently of ω.

LEMMA 3.2. *Process A_1 converges to zero uniformly on compact intervals when ε goes to zero. More precisely, we have the following estimates on the rate of convergence: For any $\gamma \in (0, 1)$, there exists a function h_1 such that*

$$\left(\mathbb{E} \sup_{s \leq t} |A_1|^p \right)^{1/p} \leq K_1 t \varepsilon^\gamma h_1(t, \varepsilon),$$

where h_1 is continuous in t, $\varepsilon > 0$ and converges to zero when $(t, \varepsilon) \mapsto (0, 0)$.

PROOF. Initially, note that by triangular inequality, and putting the supremum inside the integral, we get

$$\left(\mathbb{E} \sup_{s \leq t} |A_1|^p \right)^{1/p}$$

$$\leq \varepsilon \sum_{n=0}^{N-1} \left(\mathbb{E} \left[\int_{t_n}^{t_{n+1}} \sup_{t_n \leq s \leq r} |g(y_s^\varepsilon) - g(F_{s-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega)))| \, dr \right]^p \right)^{1/p}.$$

If $\frac{1}{p} + \frac{1}{q} = 1$, by Hölder's inequality, we have that the estimate above is again bounded by

$$\varepsilon \sum_{n=0}^{N-1} \left(\mathbb{E} \left[\left(\int_{t_n}^{t_{n+1}} dr \right)^{1/q} \right. \right.$$

$$\left. \left. \times \left(\int_{t_n}^{t_{n+1}} \sup_{t_n \leq s \leq r} |g(y_s^\varepsilon) - g(F_{s-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega)))|^p \, dr \right)^{1/p} \right]^p \right)^{1/p}.$$

In terms of the increments Δt, the last inequality is again bounded by

$$\leq \varepsilon (\Delta t)^{1/q} \sum_{n=0}^{N-1} \left(\mathbb{E} \left[\Delta t \sup_{t_n \leq s \leq t_{n+1}} |g(y_s^\varepsilon) - g(F_{s-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega)))|^p \right] \right)^{1/p}$$

$$\leq \varepsilon \Delta t \sum_{n=0}^{N-1} \left(\mathbb{E} \left[\sup_{t_n \leq s \leq t_{n+1}} |g(y_s^\varepsilon) - g(F_{s-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega)))|^p \right] \right)^{1/p}.$$

Lemma 2.1 (or its corollaries, where the result holds for $p \geq 1$) says that for all $0 \leq n \leq N - 1$ above, the function g evaluated along trajectories of the perturbed system compared with g evaluated along the unperturbed trajectories, both starting at $y_{t_n}^\varepsilon$, satisfies

$$\left[\mathbb{E} \sup_{t_n \leq s \leq t_{n+1}} \left| g(y_s^\varepsilon) - g(F_{s-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega))) \right|^p \right]^{1/p} \leq K_1 \varepsilon \Delta t e^{K_2 (\Delta t)^p}.$$

Hence, using that N has order $[\varepsilon^{-1} |\ln \varepsilon|^{-2\beta/p}]$ and that $\Delta t \leq t |\ln \varepsilon|^{2\beta/p}$,

$$\left[\mathbb{E} \sup_{s \leq t} |A_1|^p \right]^{1/p} \leq K_1 N \varepsilon^2 (\Delta t)^2 e^{K_2 (\Delta t)^p}$$

$$\leq K_1 \varepsilon^2 [\varepsilon^{-1} |\ln \varepsilon|^{-2\beta/p}] t^2 |\ln \varepsilon|^{4\beta/p} e^{K_2 (t|\ln \varepsilon|^{2\beta/p})^p}$$

$$= K_1 t \varepsilon^\gamma h_1(\varepsilon, t)$$

for any $\gamma \in (0, 1)$ where

$$h_1(\varepsilon, t) = t \varepsilon^{(1-\gamma)/2} |\ln \varepsilon|^{2\beta/p} \exp \left\{ \left(\frac{1-\gamma}{2} \right) \ln \varepsilon + K_2 t^p |\ln \varepsilon|^{2\beta} \right\},$$

which satisfies the required properties for $\beta \in (0, 1/2)$. □

LEMMA 3.3.   *Process A_2 in equation* (14) *goes to zero with the following rate of convergence*:

$$\left[\mathbb{E} \sup_{s \leq t} |A_2|^p \right]^{1/p} \leq t \eta(t |\ln \varepsilon|^{2\beta/p}),$$

where $\eta(t)$ is the rate of convergence in L^p of the ergodic averages in the unperturbed trajectories on the leaves.

PROOF.   We have

$$\left[\mathbb{E} \sup_{s \leq t} |A_2|^p \right]^{1/p}$$

$$\leq \varepsilon \left[\mathbb{E} \left| \sum_{n=0}^{N-1} \left[\int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega))) \, dr - \Delta t \, Q^g(\pi(y_{t_n}^\varepsilon)) \right] \right|^p \right]^{1/p}$$

$$\leq \varepsilon \sum_{n=0}^{N-1} \left[\mathbb{E} \left| \int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega))) \, dr - \Delta t \, Q^g(\pi(y_{t_n}^\varepsilon)) \right|^p \right]^{1/p}$$

$$= \varepsilon \Delta t \sum_{n=0}^{N-1} \left[\mathbb{E} \left| \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} g(F_{r-t_n}(y_{t_n}^\varepsilon, \theta_{t_n}(\omega))) \, dr - Q^g(\pi(y_{t_n}^\varepsilon)) \right|^p \right]^{1/p}.$$

For all $n = 0, \ldots, N - 1$, by the ergodic theorem, the two terms inside the modulus converge to each other when Δt goes to infinity with rate $\eta(\Delta t)$. Hence, for small ε, we have

$$\left[\mathbb{E}\sup_{s \leq t} |A_2|^p\right]^{1/p} \leq \varepsilon N(\Delta t)\eta(\Delta t)$$

$$\leq \varepsilon\big[\varepsilon^{-1}|\ln\varepsilon|^{-2\beta/p}\big]t|\ln\varepsilon|^{2\beta/p}\eta\big(t|\ln\varepsilon|^{2\beta/p}\big)$$

$$= t\eta\big(t|\ln\varepsilon|^{2\beta/p}\big). \qquad \square$$

LEMMA 3.4. *A_3 converges to zero when t or ε go to 0. Moreover, if Q^g is α-Hölder continuous then the rate of convergence satisfies*

$$\left(\mathbb{E}\sup_{s \leq t} |A_3|^p\right)^{1/p} \leq C\varepsilon^\alpha t^{1+\alpha}|\ln\varepsilon|^{2\beta\alpha/p},$$

for a positive constant C.

PROOF. Consider the partition given by the sequence $(\varepsilon t_n)_{0 \leq n \leq N}$ of the interval $(s \wedge \varepsilon\tau^\varepsilon, (s + t) \wedge \varepsilon\tau^\varepsilon)$. The convergence to zero here corresponds to the existence of the Riemann integral. Moreover, assuming that Q^g is α-Hölder continuous, then

(15)
$$|A_3| \leq \sum_{n=0}^{N-1} \varepsilon\Delta t \sup_{\varepsilon t_n < r \leq \varepsilon t_{n+1}} \big|Q^g\big(\pi\big(y_{t_n}^\varepsilon\big)\big) - Q^g\big(\pi\big(y_{r/\varepsilon}^\varepsilon\big)\big)\big|$$

$$\leq \varepsilon\Delta t \sum_{n=0}^{N-1} C_1\Big(\sup_{\varepsilon t_n < r \leq \varepsilon t_{n+1}} \big|\pi\big(y_{\varepsilon t_n/\varepsilon}^\varepsilon\big) - \pi\big(y_{r/\varepsilon}^\varepsilon\big)\big|\Big)^\alpha.$$

We use now that

$$\big|\pi\big(y_{s/\varepsilon}^\varepsilon\big) - \pi\big(y_{t/\varepsilon}^\varepsilon\big)\big| \leq \Big(\sup_{x \in U}\big|K(x)\big|\Big)|s - t|$$

for all $s, t \geq 0$, such that the right-hand side is independent of ε [we are going to use this fact again, which follows by equation (4); for more details, see the very beginning of proof of Theorem 4.1]. Hence, continuing the estimates for $|A_3|$, inequality (15) above implies that

$$|A_3| \leq C_2 N(\varepsilon\Delta t)^{(1+\alpha)}$$

$$\leq C_2\big(\varepsilon^{-1}|\ln\varepsilon|^{-2\beta/p}\big)\big(\varepsilon t|\ln\varepsilon|^{2\beta/p}\big)^{1+\alpha}$$

$$\leq C_2\varepsilon^\alpha t^{1+\alpha}|\ln\varepsilon|^{2\beta\alpha/p}$$

for a constant C_2. $\square$

LEMMA 3.5. *Process A_4 converges to zero with*

$$\left(\mathbb{E}\sup_{s\leq t}|A_4|^p\right)^{1/p} \leq Ct\varepsilon|\ln\varepsilon|^{2\beta/p}.$$

PROOF.   Denote

$$C = \sup_{x\in U}|g(x)|.$$

The result follows straightforwardly since

$$\varepsilon\left|\int_{t_N}^{((s+t)/\varepsilon)\wedge\tau^\varepsilon} g(y_r^\varepsilon)\,dr\right| \leq C\varepsilon\Delta t = Ct\varepsilon|\ln\varepsilon|^{2\beta/p}. \qquad\square$$

Now, going back to the proof of Lemma 3.1, the statement follows by inequality (14) and adding up the estimates of the last four lemmas (3.2 up to 3.5).   □

4. An averaging principle.   We state the averaging principle in the next theorem. To use Lemma 3.1 of the previous section we have to assume regularity in the average function Q^g, which naturally depends on g, on the foliated coordinate system and on the transversal behavior of the invariant measures on the leaves of the original foliated system. We are going to assume the following condition:

HYPOTHESIS (H).   *Let g be one of the functions given by the vertical coordinates of the perturbing vector field K, that is, $g \in \{d\pi_1(K), \ldots, d\pi_d(K)\}$. We assume that its corresponding average on the leaves $Q^g : V \to \mathbf{R}$ is Lipschitz.*

This hypothesis holds naturally if the invariant measures μ_p for the unperturbed foliated system has a sort of weakly continuity on p. For deterministic systems it corresponds to a certain regularity in the sense that there is no bifurcation with respect to the vertical parameter $v \in V$.

Let $v(t)$ be the solution of the deterministic ODE in the transversal component $V \subset \mathbf{R}^n$

(16)
$$\frac{dv}{dt} = \left(Q^{d\pi_1(K)}, \ldots, Q^{d\pi_d(K)}\right)\left(v(t)\right)$$

with initial condition $v(0) = \pi(x_0) = 0$. Let T_0 be the time that $v(t)$ reaches the boundary of V.

THEOREM 4.1.   *Assuming Hypothesis (H) above, we have*:

(1) *For any $0 < t < T_0$, $\beta \in (0, 1/2)$, $\alpha \in (0, 1)$ and $2 \leq p < \infty$ ($1 \leq p$ if in the conditions of Corollaries 2.2, 2.3 or Remark 2.4), there exist functions $C_1 = C_1(t)$ and $C_2 = C_2(t)$ such that*

$$\left[\mathbb{E}\left(\sup_{s\leq t}|\pi(y_{(s/\varepsilon)\wedge\tau^\varepsilon}^\varepsilon) - v(s)|^p\right)\right]^{1/p} \leq C_1\varepsilon^\alpha + C_2\eta(t|\ln\varepsilon|^{2\beta/p}),$$

where $\eta(t)$ is the rate of convergence in L^p of the ergodic averages of the unperturbed trajectories on the leaves.

(2) *For $\gamma > 0$, let*

$$T_\gamma = \inf\{t > 0 | \operatorname{dist}(v(t), \partial V) \le \gamma\}.$$

The exit times of the two systems satisfy the estimates

$$\mathbb{P}(\varepsilon\tau^\varepsilon < T_\gamma) \le \gamma^{-p}\big[C_1(T_\gamma)\varepsilon^\alpha + C_2(T_\gamma)\eta(T_\gamma|\ln\varepsilon|^{2\beta/p})\big]^p.$$

The second part of the theorem above guarantees the robustness of the averaging phenomenon in the transversal direction.

PROOF OF THEOREM 4.1. The gradient of each real function π_i is orthogonal to the leaves, hence by Itô's formula, for $i = 1, 2, \ldots, d$, we have that

$$\pi_i\big(y^\varepsilon_{(t/\varepsilon)\wedge\tau^\varepsilon}\big) = \int_0^{t\wedge\varepsilon\tau^\varepsilon} d\pi_i(K)\big(y^\varepsilon_{s/\varepsilon}\big)\,ds.$$

Lemma 3.1 for the function $d\pi_i(K)$ in M, triangular inequality and Hypothesis (H) imply that

$$\begin{aligned}
&\big|\pi_i\big(y^\varepsilon_{(t/\varepsilon)\wedge\tau^\varepsilon}\big) - v_i(t)\big| \\
&\quad \le \int_0^{t\wedge\varepsilon\tau^\varepsilon} \big|Q^{d\pi_i(K)}\big(\pi\big(y^\varepsilon_{(s/\varepsilon)\wedge\tau^\varepsilon}\big)\big) - Q^{d\pi_i(K)}\big(v(s)\big)\big|\,ds + \big|\delta_i(\varepsilon, t)\big| \\
&\quad \le C_i \int_0^t \big|\pi\big(y^\varepsilon_{(s/\varepsilon)\wedge\tau^\varepsilon}\big) - v(s)\big|\,ds + \big|\delta_i(\varepsilon, t)\big|,
\end{aligned}$$

where each C_i is the Lipschitz constant of $Q^{d\pi_i(K)}$, and $\delta_i(\varepsilon, t)$ is defined in Lemma 3.1. Summing up the i's and using Gronwall's lemma, we have, for a constant C,

$$\big|\pi\big(y^\varepsilon_{(t/\varepsilon)\wedge\tau^\varepsilon}\big) - v(t)\big| \le e^{Ct}\sum_{i=1}^n\big|\delta_i(\varepsilon, t)\big|.$$

The first part of the theorem follows by Lemma 3.1.

For the second part we have the following estimates:

$$\begin{aligned}
\mathbb{P}(\varepsilon\tau^\varepsilon < T_\gamma) &\le \mathbb{P}\Big(\sup_{s \le T_\gamma \wedge \varepsilon\tau^\varepsilon} \big|v(s) - \pi\big(y^\varepsilon_{(s/\varepsilon)\wedge\tau^\varepsilon}\big)\big| > \gamma\Big) \\
&\le \gamma^{-p}\mathbb{E}\Big(\sup_{s \le T_\gamma \wedge \varepsilon\tau^\varepsilon} \big|v(s) - \pi\big(y^\varepsilon_{(s/\varepsilon)\wedge\tau^\varepsilon}\big)\big|^p\Big) \\
&\le \gamma^{-p}\big[C_1(T_\gamma)\varepsilon^\alpha + C_2(T_\gamma)\eta(T_\gamma|\ln\varepsilon|^{2\beta/p})\big]^p. \qquad \square
\end{aligned}$$

4.1. *A detailed example.*   The following simple example illustrates the frame-
work where the averaging principle described in this section holds. Consider
$M = \mathbf{R}^3 - \{(0, 0, z), z \in R\}$ with the 1-dimension horizontal circular foliation
of M where the leaf passing through a point $p = (x, y, z)$ is given by the circle
$L_p = \{(\sqrt{x^2 + y^2} \cos\theta, \sqrt{x^2 + y^2} \sin\theta, z), \theta \in [0, 2\pi]\}$. Consider the foliated lin-
ear SDE on M consisting of random rotations

$$dx_t = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} x_t (\lambda_1 \, dt + \lambda_2 \circ dB_t).$$

For an initial condition $p_0 = (x_0, y_0, z_0)$, say with $x_0 \geq 0$, consider the local
foliated coordinates in the neighborhood $U = \mathbf{R}^3 \setminus \{(x, 0, z); x \leq 0; z \in \mathbf{R}\}$ given
by cylindrical coordinates. Hence, using the same notation as before, $\varphi = (u, v)$
will be defined by $\varphi : U \subset M \to (-\pi, \pi) \times \mathbf{R}_{>0} \times \mathbf{R}$, where $u \in (-\pi, \pi)$ is angular
and $v = (r, z) \in \mathbf{R}_{>0} \times \mathbf{R}$ such that $\varphi^{-1} : (u, v) \mapsto (r \cos u, r \sin u, z) \in M$. In this
coordinates system, the transversal projections π_1 and π_2 correspond to the radial
r-component and the z-coordinate, respectively.

For $\lambda_1, \lambda_2 \in \mathbf{R}$ with $|\lambda_1| + |\lambda_2| > 0$, the invariant measures μ_p in the leaves L_p
passing through points $p \in M$ are given by normalized Lebesgue measures in L_p,
which here corresponds to the normalized angle 1-form. Note that Hypothesis (H)
is satisfied. We investigate the following effective behavior of a small transversal
perturbation of order ε:

$$dy_t^\varepsilon = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_t^\varepsilon (\lambda_1 \, dt + \lambda_2 \circ dB_t) + \varepsilon K(y_t^\varepsilon) \, dt,$$

with initial condition $x_0 = (1, 0, 0)$. Typically, due to uniform ellipticity, for any
perturbing vector field, the average of its transversal component on the leaves con-
verges with rates at least $\frac{1}{\sqrt{t}}$. Further, if $\lambda_1 \neq 0$ and $\lambda_2 = 0$, due to periodicity, the
rate of convergence improves to order $\eta(t) \sim \frac{1}{t}$. Here we consider two classes of
perturbing vector field K.

Constant perturbation. (A) Assume that the perturbation is given by a vec-
tor field which is constant $K = (k_1, k_2, k_3)$ with respect to Euclidean coordi-
nates in M. In this case we have that the convergence of the time average on
the leaves to the ergodic average is given by $\eta(t) \equiv 0$. Initially, to fix the ideas,
assume that $k_3 = 0$. Then, not only the average on the z-component vanishes,
that is, $Q^{d\pi_2 K} = 0$, but also, by the geometrical symmetry of K with respect
to the invariant measure, the average radial r-component also vanishes, that is,
$Q^{d\pi_1 K} = 0$. Hence the transversal component in the main theorem is constant
$v(t) = (r(0), z(0))$ for all $t \geq 0$.

The vertical components in the statement of the first formula in Theorem 4.1, in
this example, concern only the difference between the initial radius $r(0) = 1$ and

$r(\frac{t}{\varepsilon} \wedge \tau^\varepsilon) = \pi_1(y^\varepsilon_{(t/\varepsilon) \wedge \tau^\varepsilon})$. Precisely, we have that

$$\left[\mathbb{E}\left(\sup_{s \le t} \left| r\left(\frac{s}{\varepsilon} \wedge \tau^\varepsilon \right) - 1 \right|^p \right) \right]^{1/p}$$

goes to zero, for a fixed t, with order ε^α, with $\alpha \in (0, 1)$.

(B) *Vertical perturbations.* Now, for constant and vertical $K = (0, 0, k_3)$, the radial average $Q^{d\pi_1 K}$ is null, but $Q^{d\pi_2 K}$ equals k_3 for every leaf in M. Hence the averaged system $v(t) = (r(0), k_3 t)$ is constant in the radial component and increases linearly in the z-coordinate. The perturbed systems have the simple solution

$$y^\varepsilon_{t/\varepsilon} = \begin{pmatrix} \cos\left(\dfrac{\lambda_1 t}{\varepsilon} + \lambda_2 B_{t/\varepsilon} \right) \\ \sin\left(\dfrac{\lambda_1 t}{\varepsilon} + \lambda_2 B_{t/\varepsilon} \right) \\ k_3 t \end{pmatrix}.$$

Hence, the comparison

$$\left| \pi_2\left(y^\varepsilon_{(t/\varepsilon) \wedge \tau^\varepsilon} \right) - v(t) \right| \equiv 0$$

for all $t \ge 0$ and the convergence to zero is trivial.

Linear perturbation. Consider a linear perturbation of the form $K(x, y, z) = (x, 0, 0)$. In this case, again, the z-coordinate average vanishes trivially. For the radial component, we have that $d\pi_1 K = r_0 \cos^2 u$, where u is the angular coordinate of p whose distance to the z-axis (r-coordinate) is r_0. Hence the average with respect to the invariant measure on the leaves is given by $Q^{d\pi_1 K} = r/2$ for leaves with radius r. The transversal system stated in the theorem is then $v(t) = (e^{t/2} r(0), z(0))$. Hence the result guarantees that the radial part of $y^\varepsilon_{(t/\varepsilon) \wedge \tau^\varepsilon}$ must have a behavior close to the exponential $e^{t/2}$ in the sense that

$$\left[\mathbb{E}\left(\sup_{s \le t} \left| r\left(\frac{t}{\varepsilon} \right) - e^{t/2} \right|^p \right) \right]^{1/p}$$

goes to zero when ε goes to zero.

The fundamental solution of the linear perturbed Stratonovich systems y^ε_t is given by the exponential of the matrix for each fixed t

$$\begin{pmatrix} \varepsilon t & -(\lambda_1 t + \lambda_2 B_t) & 0 \\ \lambda_1 t + \lambda_2 B_t & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where the eigenvalues corresponding to the first two coordinates (horizontal plane) are

$$\lambda_{1,2} = \frac{\varepsilon t \pm \sqrt{\varepsilon^2 t^2 - 4(\lambda_1 t + \lambda_2 B_t)^2}}{2},$$

whose real part is given by $\varepsilon t/2$ with probability increasing to 1 as ε goes to zero. This exponential rates coincides with that one above, guaranteed by the Theorem 4.1.

Lyapunov exponents. For foliated manifolds embedded in $\mathbf{R}^N$, the symmetry of the perturbing vector field K with respect to the geometry of the leaves, hence also with respect to Lebesgue invariant measure, as presented in the case of constant K, has implied that the transversal average $Q^{d\pi K}$ vanishes. This phenomenon also appears in a couple of other examples where the leaves are not only diffeomorphic to each other, but also have this symmetry in the sense that the integration of a constant perturbation $K \in \mathbf{R}^N$ with respect to the Lebesgue measure is zero. Here we mention a couple of simple examples: the spherical foliation of $\mathbf{R}^n \setminus \{0\}$, nested torus (increasing the smaller radius) foliation of the solid torus minus the central circle $S^1 \times D^2 \setminus S^1 \subset \mathbf{R}^3$ or more generally (when they exists) tubular foliation of $\mathbf{R}^n \setminus \{C\}$, with C a compact set (this context also includes the Hamiltonian case with the Lyouville foliation of the symplectic structure of $\mathbf{R}^{2n}$ as in [13]). In these symmetric geometrical configuration, if the invariant measure on the leaves are the Lebesgue measures (taking gradient Brownian motion on the leaves, e.g., as in [6]) the averages of $Q^{d\pi}$ vanish. Hence our main theorem says that on the average, the trajectories of the perturbed system stay somehow close to the initial leaf as ε decreases to zero.

The Lyapunov exponent of the system in the direction of a tangent vector $v \in T_{x_0}M$ contains the long time behavior of points close to x_0 in the direction of v; for details on the definition, properties, existence conditions, multiplicative ergodic theory, etc., see, for example, Arnold [2] and the references therein, among many others. In particular, under the symmetric geometrical circumstances above, if there exists the Lyapunov exponents of the perturbed system y_t^ε, Theorem 4.1 will imply that in transversal directions the Lyapunov exponent cannot be too far from zero. This vanishing property must happen with multiplicity given by the codimension of the foliation, as in the examples of the paragraphs above, where the asymptotic relevant parameters (rotation number [3, 7, 16], and Lyapunov exponents) do exist. In short:

PROPOSITION 4.2 (Continuity of Lyapunov exponents). *Assume that the perturbed system y_t^ε does have Lyapunov exponents a.s. at the assigned initial condition. If the averaged perturbation in the leaves vanishes, that is, $Q^{d\pi K} = 0$, then a number, given by the codimension of the foliation, of Lyapunov exponents in the spectrum go to nonpositive values as ε goes to zero.*

PROOF.   In fact, Theorem 4.1 says that the perturbed system increases in the transversal coordinates with order $|\pi(y_t^\varepsilon)| \lesssim C_1(\varepsilon t)\varepsilon^\alpha + C_2(\varepsilon t)\eta((\varepsilon t)|\ln \varepsilon|^{2\beta/p})$. Hence, any exponential behavior, if it exists, tends to nonpositive values as ε decreases to zero.   □

Acknowledgments. This article was written while the second author's was in a sabbatical visit to University of Humboldt. He would like to express his gratitude to Professor Peter Imkeller and his research group for their nice and friendly hospitality.

REFERENCES

[1] ARNOLD, V. I. (1989). *Mathematical Methods of Classical Mechanics*, 2nd ed. *Graduate Texts in Mathematics* **60**. Springer, New York. MR0997295

[2] ARNOLD, L. (1998). *Random Dynamical Systems*. Springer, Berlin. MR1723992

[3] ARNOLD, L. and IMKELLER, P. (1999). Rotation numbers for linear stochastic differential equations. *Ann. Probab.* **27** 130–149. MR1681134

[4] BAKHTIN, V. and KIFER, Y. (2008). Nonconvergence examples in averaging. In *Geometric and Probabilistic Structures in Dynamics*. *Contemp. Math.* **469** 1–17. Amer. Math. Soc., Providence, RI. MR2478463

[5] BORODIN, A. N. and FREIDLIN, M. I. (1995). Fast oscillating random perturbations of dynamical systems with conservation laws. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 485–525. MR1338450

[6] CATUOGNO, P., LEDESMA, D. and RUFFINO, P. (2012). Harmonic measures in embedded foliated manifolds. Preprint. Available at arXiv:1208.0629.

[7] CATUOGNO, P. J., LEDESMA, D. S. and RUFFINO, P. R. C. (2008). Relative rotation number for stochastic systems: Dynamical and topological applications. *Dyn. Syst.* **23** 425–435. MR2473215

[8] FREIDLIN, M. I. and WENTZELL, A. D. (1991). *Random Perturbations of Dynamical Systems*. Springer, Berlin.

[9] KABANOV, Y. and PERGAMENSHCHIKOV, S. (2003). *Two-Scale Stochastic Systems*: *Asymptotic Analysis and Control*. *Applications of Mathematics* (*New York*) **49**. Springer, Berlin. MR1942740

[10] KAKUTANI, S. and PETERSEN, K. (1981). The speed of convergence in the ergodic theorem. *Monatsh. Math.* **91** 11–18. MR0612450

[11] KHASMINSKII, R. and KRYLOV, N. (2001). On averaging principle for diffusion processes with null-recurrent fast component. *Stochastic Process. Appl.* **93** 229–240. MR1828773

[12] KRENGEL, U. (1978). On the speed of convergence in the ergodic theorem. *Monatsh. Math.* **86** 3–6. MR0510630

[13] LI, X.-M. (2008). An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity* **21** 803–822. MR2399826

[14] NAMACHCHVAYA, S. and SOWERS, R. B. (2002). Rigorous stochastic averaging at a center with additive noise. *Meccanica* **37** 85–114. MR1927747

[15] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* **293**. Springer, Berlin. MR1725357

[16] RUFFINO, P. R. C. (2002). Decomposition of stochastic flows and rotation matrix. *Stoch. Dyn.* **2** 93–107. MR1897354

[17] SANDERS, J. A., VERHULST, F. and MURDOCK, J. (2007). *Averaging Methods in Nonlinear Dynamical Systems*, 2nd ed. *Applied Mathematical Sciences* **59**. Springer, New York. MR2316999

[18] SOWERS, R. B. (2002). Stochastic averaging with a flattened Hamiltonian: A Markov process on a stratified space (a whiskered sphere). *Trans. Amer. Math. Soc.* **354** 853–900 (electronic). MR1867364

[19] TONDEUR, P. (1988). *Foliations on Riemannian Manifolds*. Springer, New York. MR0934020

[20] WALCZAK, P. (2004). *Dynamics of Foliations*, *Groups and Pseudogroups. Mathematics Institute of the Polish Academy of Sciences*. *Mathematical Monographs* (*New Series*) **64**. Birkhäuser, Basel. MR2056374

DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE ESTADUAL DE CAMPINAS
13.083-859- CAMPINAS - SP
BRAZIL
E-MAIL: ivangargate@utfpr.edu.br
            ruffino@ime.unicamp.br