An averaging principle for diffusions in foliated spaces

Iván I. Gonzáles Gargate
Paulo R. Ruffino

Departamento de Matemática, Universidade Estadual de Campinas, 13.083-859- Campinas - SP, Brazil.

Abstract

Consider an SDE on a foliated manifold whose trajectories lay on compact leaves. We investigate the effective behaviour of a small transversal perturbation of order $\epsilon$. 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 $\epsilon$ 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.

Key words: Averaging principle, foliated diffusion, rescaled stochastic systems, stochastic flows.

MSC2010 subject classification: 60H10, 58J65, 58J37.

1 Introduction and set up

Generally speaking, the original heuristic idea of an averaging principles 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 have appeared long ago and, as mentioned by V. Arnold [3, p. 287], they were implicitly contained in the works of Laplace, Lagrange and Gauss on celestial mechanics; literature on the matter can be found e.g. among many others in [3], Sanders, Verhulst and Murdoch [17] and references therein. Presently, on what regards 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 X.-M. Li [13, p.806], Kabanov and Pergamenshchikov [8, Appendix], [17, Appendix A]. Among many other works somehow related to the topic, we refer to Khasminski and Krylov [10], Bakhtin and Kifer [11], Sowers [18], Namachchivaya and Sowers [14], Borodin and Frédilin [4, 8] 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 perturbation are taken transversal to the leaves of the foliation. Here, the slow system is the transversal component and the fast system is given by the
Our results generalize the recent approach by X.M. Li [13] on an averaging principle for a completely integrable stochastic Hamiltonian system. In that article, as in the classical approach, see e.g. [3], Li has explored the benefits of a 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). Comparing 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 \(\epsilon = t = 0\). Some of the rates of convergence of [13, Lemma 3.1] is 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 \(\epsilon^\alpha + \eta(|\ln \epsilon|^{-\beta/\rho})\), where \(\eta\) 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 \(\eta\), convergence rates 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, i.e. \(M\) is endowed with an integrable regular distribution of dimension \(n\) (for definition and further properties of foliated spaces see e.g. the initial chapters of Tondeur [19], Walcak [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 neighbourhood \(U \subset M\) where \(U\) is diffeomorphic to \(L_x \times V\), where \(V \subset \mathbb{R}^d\) is an open bounded neighbourhood of the origin and \(d\) is the codimension of the foliation. We shall assume an SDE in \(M\) whose solution flow preserves the foliation, i.e. we consider a Stratonovich equation

\[
dx_t = X_0(x_t)dt + \sum_{k=1}^r X_k(x_t) \circ dB^k_t
\]

where the smooth vector fields \(X_k\) are foliated in the sense that \(X_k(x) \in T_xL_x\), for \(k = 0, 1, \ldots, r\). Here \(B_t = (B^1_t, \ldots, B^r_t)\) is a standard Brownian motion in \(\mathbb{R}^r\) with respect to a filtered probability space \((\Omega, \mathcal{F}_t, \mathcal{F}, P)\). 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 \(y^\epsilon_t\) which satisfies the SDE

\[
dy^\epsilon_t = X_0(y^\epsilon_t)dt + \sum_{k=1}^r X_k(y^\epsilon_t) \circ dB^k_t + \epsilon K(y^\epsilon_t) dt,
\]

with the same initial condition \(y^\epsilon_0 = x_0\) of the unperturbed system \(x_t\).
Our main result, Theorem 4.1, says that locally the transversal behaviour of $y^\epsilon_t$ 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 the 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 hypothesis 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 $\epsilon$ 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 $V \subset \mathbb{R}^d$ is a connected open set containing the origin. The neighbourhood $U$ can be considered small enough such that the closure $\bar{U} \subset M$ and the derivative of $\varphi$ is bounded. For a fixed diffeomorphism $\varphi$, 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 $\pi(p) \in V$, i.e. $\varphi(p) = (u, \pi(p))$ 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 denote by:

$$\pi(p) = \left(\pi_1(p), \ldots, \pi_d(p)\right) \in V \subset \mathbb{R}^d$$

for any $p \in U$. Natural examples of these coordinate 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 $V$.

Next lemma gives information on the order of which the perturbed trajectories $y^\epsilon_t$ approaches the unperturbed $x_t$ when one varies $\epsilon$ and $t$ in equation (2); it will be used to prove that the dynamics of the rescaled system $y^{\epsilon t}_{\epsilon}$ 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, Lemma 3.1.

Lemma 2.1 Let $\tau^\epsilon$ be the exit time of the process $y^\epsilon_t$ from the neighbourhood $U$ of our coordinate system as above. For any locally Lipschitz continuous function $f : M \to \mathbb{R}$ and $2 \leq p < \infty$ we have

$$\left[\mathbb{E}\left(\sup_{s \leq t \wedge \tau^\epsilon} |f(y^\epsilon_s) - f(x_s)|^p\right)\right]^{1/p} \leq K_1 \epsilon t e^{K_2 t^p}.$$
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 $y_t^\epsilon$, the solutions of equations (1) and (2) respectively, according to the coordinates given by the diffeomorphism $\varphi$: denote $(u_t, v_t) := \varphi(x_t)$ and $(u_t^\epsilon, v_t^\epsilon) := \varphi(y_t^\epsilon)$. Then

$$|f(y_t^\epsilon) - f(x_t)| = |f \circ \varphi^{-1}(u_t^\epsilon, v_t^\epsilon) - f \circ \varphi^{-1}(u_t, v_t)| \leq C |u_t^\epsilon - u_t| + C |v_t^\epsilon - v_t|,$$

for some constant $C \geq 0$, where we have used the fact that $\varphi$ 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 $v_t^\epsilon$ is given, simply, by

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

Therefore

$$\sup_{s \leq t} |v_s^\epsilon - v_s| \leq \epsilon \sup_{s \leq t} \int_0^s |K_v(u_s^\epsilon, v_s^\epsilon)| ds \leq \epsilon t \sup_{x \in U} |K_v(x)| \leq C_1 \epsilon t,$$

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

For the horizontal summand $|u_t^\epsilon - u_t|$ in inequality (3), we consider an embedding $i : M \to \mathbb{R}^N$ of the compact submanifold $L_{x_0}$ in an Euclidean space $\mathbb{R}^N$ with a sufficiently large integer $N$. In $\mathbb{R}^N$ we have, in coordinates, $i(u_t) = (u_t^1, \ldots, u_t^N)$. We look for an equation of $i(u_t^\epsilon) \in \mathbb{R}^N$. Let $b_k$ 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 $\tilde{b}_k(u, v) = di(X_k(u, v))$ 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 neighbourhood of $i(L_{x_0}) \subset \mathbb{R}^N$ such that $b_k$ is constant in orthogonal fibres to $i(M)$ in this tubular neighbourhood, for each $k = 0, 1, \ldots, r$. Finally, in canonical coordinates in $\mathbb{R}^N$, we have, for $i = 1, \ldots, N$,

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

Again, by compactness, the image of the embedding $i(L_{x_0})$ has an induced metric from $\mathbb{R}^N$ which is uniformly equivalent to the original metric in $L_{x_0}$. Moreover note
that, by the choice of the neighbourhood $U$, 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 $i$-th component, for $s < \tau^i$:

\[
  u^{\epsilon,i}_s - u^i_s = \sum_{k=1}^r \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r)) \circ dB^k_r + 
  \int_0^s (b^i_0(u^\epsilon_r, v^\epsilon_r) - b^i_0(u_r, v_r))dr + \epsilon \int_0^s K^i_h(u^\epsilon_r, v^\epsilon_r)dr.
\]

In terms of Itô integral,

\[
  \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r)) \circ dB^k_r = \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r))dB^k_r 
  + \frac{1}{2} \int_0^s |\nabla b^i_k \cdot b_k(u^\epsilon_r, v^\epsilon_r) - \nabla b^i_k \cdot b_k(u_r, v_r)|dr.
\]

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

\[
  |u^{\epsilon,i}_s - u^i_s| \leq \sum_{k=1}^r \left| \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r))dB^k_r \right| + 
  \frac{1}{2} \sum_{k=1}^r \int_0^s \left| \nabla b^i_k \cdot b_k(u^\epsilon_r, v^\epsilon_r) - \nabla b^i_k \cdot b_k(u_r, v_r) \right|dr
  + \int_0^s |b^i_0(u^\epsilon_r, v^\epsilon_r) - b^i_0(u_r, v_r)|dr + \epsilon \int_0^s |K^i_h(u^\epsilon_r, v^\epsilon_r)|dr.
\]

Functions $b^i_0$ and ($\nabla b^i_k \cdot b_k$) are Lipschitz, hence for a common constant $C_2$,

\[
  |u^{\epsilon,i}_s - u^i_s| \leq \sum_{k=1}^r \left| \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r))dB^k_r \right| + 
  C_2 \int_0^s |v^\epsilon_r - v_r|dr + C_2 \int_0^s |u^\epsilon_r - u_r|dr + \epsilon s \sup_U |K_h|.
\]

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

\[
  |u^{\epsilon,i}_s - u^i_s| \leq \sum_{k=1}^r \left| \int_0^s (b^i_k(u^\epsilon_r, v^\epsilon_r) - b^i_k(u_r, v_r))dB^k_r \right| + C_1 C_2 \epsilon s^2
  + C_2 \int_0^s |u^\epsilon_r - u_r|dr + C_1 \epsilon s.
\]

Now, for $p \geq 1$, there exists a constant $C_3$ such that
\[
\left| u^{e,i}_s - u^i_s \right|^p \leq C_3 \sum_{k=1}^r \left| \int_0^s \left( b_k^i(u^e_r, v^e_r) - b_k^i(u_r, v_r) \right) dB_r \right|^p + C_3 \left( C_1 C_2 \varepsilon s^2 \right)^p \\
+ C_3 C_2^p \left( \int_0^s \left| u^e_r - u_r \right| \, dr \right)^p + C_3 \left( C_1 \varepsilon s \right)^p .
\]

Cauchy-Schwartz inequality yields:

\[
\left| u^{e,i}_s - u^i_s \right|^p \leq C_3 \sum_{k=1}^r \left| \int_0^s \left( b_k^i(u^e_r, v^e_r) - b_k^i(u_r, v_r) \right) dB_r \right|^p + C_3 \left( C_1 C_2 \varepsilon s^2 \right)^p \\
+ C_3 C_2^p s^{p-1} \int_0^s \left| u^e_r - u_r \right|^p \, dr + C_3 \left( C_1 \varepsilon s \right)^p .
\]

Hence,

\[
E \sup_{s \leq t \wedge \tau^e} \left| u^{e,i}_s - u^i_s \right|^p \leq C_3 E \sup_{s \leq t \wedge \tau^e} \sum_{k=1}^r \left| \int_0^s \left( b_k^i(u^e_r, v^e_r) - b_k^i(u_r, v_r) \right) dB_r \right|^p + C_3 \left( C_1 C_2 \varepsilon t^2 \right)^p \\
+ C_3 C_2^p t^{p-1} E \sup_{s \leq t \wedge \tau^e} \int_0^s \left| u^e_r - u_r \right|^p \, dr + C_3 \left( C_1 \varepsilon t \right)^p \\
\leq C_4 \sum_{k=1}^r E \left[ \int_0^{t \wedge \tau^e} \left( b_k^i(u^e_r, v^e_r) - b_k^i(u_r, v_r) \right)^2 \, dr \right]^{p/2} + C_3 \left( C_1 C_2 \varepsilon t^2 \right)^p \\
+ C_3 C_2^p t^{p-1} \int_0^t E \left( \sup_{s \leq r \wedge \tau^e} \left| u^e_r - u_r \right|^p \right) \, dr + C_3 \left( C_1 \varepsilon t \right)^p
\]

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:

\[
\sum_{k=1}^r \int_0^{t \wedge \tau^e} \left( b_k^i(u^e_r, v^e_r) - b_k^i(u_r, v_r) \right)^2 \, dr \\
\leq 2C_2^2 \left( \int_0^{t \wedge \tau^e} |v^e_r - v_r|^2 \, dr + \int_0^{t \wedge \tau^e} |u^e_r - u_r|^2 \, dr \right) \\
\leq 2C_2^2 \left( \int_0^t C_1^2 \varepsilon^2 r^2 \, dr + \int_0^{t \wedge \tau^e} \sup_{s \leq \tau^e} |u^e_r - u_r|^2 \, dr \right) \\
\leq C_2^2 C_1^2 \varepsilon^2 t^3 + 2C_2^2 \int_0^{t \wedge \tau^e} \sup_{s \leq \tau^e} |u^e_r - u_r|^2 \, dr .
\]

We end up with:

\[
E \sup_{s \leq t \wedge \tau^e} \left| u^{e,i}_s - u^i_s \right|^p \leq C_4 \sum_{k=1}^r E \left[ C_2^2 C_1^2 \varepsilon^2 t^3 + 2C_2^2 \int_0^{t \wedge \tau^e} \sup_{s \leq \tau^e} |u^e_r - u_r|^2 \, dr \right]^{p/2}
\]
hence:

\[ +C_3 C_2^p t^{p-1} \int_0^t \mathbb{E} \left( \sup_{s \leq r \wedge \tau^*} |u^r - u_s|^p \right) dr \]

\[ +C_3 \left( C_1 C_2 \epsilon t^2 \right)^p + C_3 \left( C_1 \epsilon t \right)^p. \] (11)

For \( p \geq 2 \) one can use Cauchy-Schwartz again to conclude that there exists a constant \( C_5 \) such that the last expression is less than or equal to:

\[ C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_5 C_2^p t^{p-2} \int_0^t \mathbb{E} \sup_{s \leq r \wedge \tau^*} |u^r - u_s|^p \ dr + C_3 \left( C_1 C_2 \epsilon t^2 \right)^p \]

\[ +C_3 C_2^p t^{p-1} \int_0^t \mathbb{E} \left( \sup_{s \leq r \wedge \tau^*} |u^r - u_s|^p \right) dr + C_3 \left( C_1 \epsilon t \right)^p \]

\[ = C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_3 \left( C_1 C_2 \epsilon t^2 \right)^p + C_3 \left( C_1 \epsilon t \right)^p \]

\[ + \left( C_5 C_2^p t^{p-2} + C_3 C_2^p t^{p-1} \right) \int_0^t \mathbb{E} \left( \sup_{s \leq r \wedge \tau^*} |u^r - u_s|^p \right) dr. \]

Now, summing up over \( i \) in the inequalities above leads to:

\[ \mathbb{E} \sup_{s \leq t \wedge \tau^*} |u^r - u_s|^p \leq C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_3 \left( C_1 C_2 \epsilon t^2 \right)^p + C_3 \left( C_1 \epsilon t \right)^p \]

\[ + \left( C_5 C_2^p t^{p-2} + C_3 C_2^p t^{p-1} \right) \int_0^t \mathbb{E} \left( \sup_{s \leq r \wedge \tau^*} |u^r - u_s|^p \right) dr. \]

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

\[ \mathbb{E} \left( \sup_{s \leq t \wedge \tau^*} |u^r - u_s|^p \right) \leq C_6 \epsilon t^p (1 + t^p) \exp \{ C_7 (t^{p/2} + t^p) \} \]

\[ \leq C_8 \epsilon t^p (1 + t^p) \exp \{ C_9 t^p \}. \]

Going back to the inequality \( \mathbb{E} \) now we have:

\[ |f(y^r_t) - f(x_t)|^p \leq C_{10} |v^r_t - v_t|^p + C_{10} |u^r_t - u_t|^p \]

hence:

\[ \mathbb{E} \left( \sup_{s \leq t \wedge \tau^*} |f(y^r_s) - f(x_s)|^p \right) \leq C_{10} \mathbb{E} \sup_{s \leq t \wedge \tau^*} |v^r_s - v_s|^p + C_{10} \mathbb{E} \sup_{s \leq t \wedge \tau^*} |u^r_s - u_s|^p \]

\[ \leq C_{11} \epsilon t^p + C_{12} \epsilon t^p (1 + t^p) \exp(C_9 t^p) \]

\[ \leq C_{13} \epsilon t^p (1 + t^p) \exp(C_9 t^p). \]
From here, finally, one concludes that there exist constants $K_1$ and $K_2$ such that

$$
\mathbb{E} \left( \sup_{s \leq t \wedge \tau} |f(y_s^\epsilon) - f(x_s)|^p \right)^{\frac{1}{p}} \leq K_1 \epsilon t \exp(K_2 t^p).
$$

□

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 $\epsilon t C_1$; 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, i.e. $dy_t^\epsilon = y_t^\epsilon \circ dB_t + \epsilon dt$. The $L^p$ norm of the difference

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

which increases exponentially with respect to time $t$.

Next corollary includes the case of a completely integrable stochastic Hamiltonian system when one uses the action-angle coordinates, cf. X.-M. 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

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

**Proof:** In this case the correction term of the Stratonovich stochastic integral in terms of Itô integral in inequality (3) vanishes, and also so does the deterministic integration of $|u_t^\epsilon - u_t|$ in inequalities (9) and (10). Hence inequality (11) improves to

$$
\mathbb{E} \sup_{s \leq t \wedge \tau} |u_s^\epsilon - u_s|^p \leq C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_3 \left( C_1 C_2 \epsilon t^2 \right)^p + C_3 \left( C_1 \epsilon t \right)^p. \tag{12}
$$

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

□

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

**Corollary 2.3** If in addition to 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 $K_1 \epsilon (t + t^{\frac{3}{2}})$.  

8
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

\[
E \sup_{s \leq t \wedge \tau} |u^s_{\epsilon,i} - u^i_s|^p \leq C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_3 (C_1 \epsilon t)^p.
\]

\[\square\]

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

\[
\left[ E \left( \sup_{s \leq t \wedge \tau} |f(y^s) - f(x_s)|^p \right) \right]^\frac{1}{p} \leq K_1 \epsilon t \exp(K_2 t^p). \tag{13}
\]

Proof: One can no longer use Cauchy-Schwartz after inequality (11). Alternatively, from (11), use that

\[
E \sup_{s \leq t \wedge \tau} |u^s_{\epsilon,i} - u^i_s|^p \leq C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + C_5 C_2^p t^{p/2} \sup_{s \leq t \wedge \tau} |u^s_{\epsilon,i} - u^i_s|^p + C_3 (C_1 C_2 \epsilon t^2)^p + C_3 C_2^p t^{p-1} \int_0^t E \left( \sup_{r \leq t \wedge \tau} |u^r_{\epsilon} - u^r|^p \right) dr + C_3 C_1^p \epsilon^p t^p.
\]

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

\[
E \sup_{s \leq t \wedge \tau} |u^s_{\epsilon} - u_s|^p \leq \delta^{-1} C_5 \left( C_2 C_1 \epsilon t^{3/2} \right)^p + \delta^{-1} C_3 (C_1 C_2 \epsilon t^2)^p + \delta^{-1} C_3 C_2^p t^{p-1} \int_0^t E \left( \sup_{r \leq t \wedge \tau} |u^r_{\epsilon} - u^r|^p \right) dr + \delta^{-1} C_3 C_1^p \epsilon^p t^p.
\]

And one completes the calculation as before using the integral version of Gronwall 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 \(\mu_p\) for the unperturbed system (1). We shall assume that each of such \(\mu_p\) is ergodic. We define a function \(Q^g : V \subset \mathbb{R}^d \to \mathbb{R}\) 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 \(p\), i.e. \(\varphi(p) = (u, v)\), 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 concern now with this time average.

On the rate of convergence on the leaves. For deterministic dynamics it is known that there is no estimates for the speed of convergence in the ergodic theorem: it can be as slow as any prescribed decreasing rate, see e.g. Kakutani-Petersen [9] 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 $\gamma$ 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, p.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 $x$, i.e. 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 $\gamma$. 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, 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 $\gamma$ 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 is 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 [7, Chap.7 §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

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

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 neighbourhood $U$ 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 $Q^g(\pi(y_\epsilon^r))$ with respect to $r$ guarantees the convergence to zero; 2) $\alpha$-Hölder continuity guarantees the rate of convergence.

Next lemma estimates the time average of the difference between function $g$ and $Q^g$ along the perturbed system $y_\epsilon^t$. Here, again, the stopping time $\tau^\epsilon$ denotes the first exit time of the open neighbourhood $U \subset M$ 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 $K$, i.e. $g \in \{d\pi_1(K), \ldots, d\pi_d(K)\}$ and $Q^g : V \to \mathbb{R}$ be its corresponding average on the leaves. For $s, t \geq 0$ write

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

Then $\delta(\epsilon, t)$ goes to zero when $t$ or $\epsilon$ tend to zero.
Moreover, if $Q^g$ is $\alpha$-Hölder continuous then for $p \geq 2$, $0 < \alpha' < \alpha$ and any $\beta \in (0, 1/2)$ we have the following estimates:

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

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 $h(t, \epsilon)$ is continuous for $t, \epsilon > 0$ and converges to zero when $(t, \epsilon) \to 0$.

**Proof:**

The proof consists of considering a convenient partition of the interval $(\frac{s}{\epsilon} \wedge \tau^\epsilon, \frac{s + t}{\epsilon} \wedge \tau^\epsilon)$ 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 $\epsilon$, we take the following assignment of increments:

\[
\Delta t = \left( \frac{s + t}{\ln \epsilon \frac{2^\beta}{p}} \right)^{-1} \wedge \tau^\epsilon - \left( \frac{s}{\ln \epsilon \frac{2^\beta}{p}} \right)^{-1} \wedge \tau^\epsilon.
\]

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

\[
\frac{s}{\epsilon} \wedge \tau^\epsilon = t_0 < t_1 < \cdots < t_N \leq \frac{s + t}{\epsilon} \wedge \tau^\epsilon.
\]

with $N = N(\epsilon) = [\epsilon^{-1} \ln \epsilon^{-\frac{2^\beta}{p}}]$ where $[x]$ denotes the integer part of $x$.

Initially we represent the left hand side as the sum:

\[
\epsilon \int_{\frac{s}{\epsilon} \wedge \tau^\epsilon}^{\frac{s + t}{\epsilon} \wedge \tau^\epsilon} g(y^\epsilon_r) dr = \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g(y^\epsilon_r) dr + \epsilon \int_{t_N}^{\frac{s + t}{\epsilon} \wedge \tau^\epsilon} g(y^\epsilon_r) dr.
\]

Denote by $\theta_t$ 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

\[
|\delta(\epsilon, t)| \leq |A_1| + |A_2| + |A_3| + |A_4|, \quad (14)
\]

where

\[
A_1 = \epsilon \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \left[ g(y^\epsilon_r) - g(F_{r-t_n}(y^\epsilon_{t_n}, \theta_{t_n}(\omega))) \right] dr,
\]

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

12
Lemma 3.2

Process goes to infinity when $\epsilon$ except when $t$ by $\gamma$gence: For any $\epsilon$ goes to zero. More precisely, we have the following estimate on the rate of convergence: For any $\epsilon \in (0,1)$, there exists a function $h_1$ such that

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

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

Proof: Initially note that by triangular inequality, and putting the supremum inside the integral we get

$$
\left( \mathbb{E} \sup_{s \leq t} |A_1|^p \right)^{\frac{1}{p}} \leq \epsilon \sum_{n=0}^{N-1} \left( \mathbb{E} \left[ \int_{t_n}^{t_{n+1}} \sup_{t_n \leq s \leq r} \left| g(y_s^\epsilon) - g(F_{s-t_n}(y_{t_n}^\epsilon, \theta_{t_n}(\omega))) \right|^p \, dr \right] \right)^{\frac{1}{p}}.
$$

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

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

In terms of the increments $\Delta t$, last inequality is again bounded by

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

$$
\leq \epsilon \Delta t \sum_{n=0}^{N-1} \left( \mathbb{E} \left[ \sup_{t_n \leq s \leq t_{n+1}} \left| g(y_s^\epsilon) - g(F_{s-t_n}(y_{t_n}^\epsilon, \theta_{t_n}(\omega))) \right|^p \right] \right)^{\frac{1}{p}}.
$$
Lemma 2.1 (or its corollaries, in whose case 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 \( y^n_t \) satisfies:

\[
\mathbb{E} \sup_{t_n \leq s \leq t_{n+1}} \left| g(y^n_s) - g(F_{s-t_n}(y^n_{t_n}, \theta_{t_n}(\omega))) \right|^p \leq K_1 \epsilon \Delta t e^{K_2(\Delta t)^p}.
\]

Hence, using that \( N \) has order \([\epsilon^{-1} \ln \epsilon]^{-\beta/2}\) and that \( \Delta t \leq t |\ln \epsilon|^{2\beta/\rho} \),

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

\[
\leq K_1 \epsilon^2[\epsilon^{-1} |\ln \epsilon|^{\frac{2\beta}{\rho}}] t^2 |\ln \epsilon|^{2\beta} e^{K_2(t |\ln \epsilon|^{2\beta})^p}
\]

\[
= K_1 t \epsilon^\gamma h_1(\epsilon, t)
\]

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

\[
h_1(\epsilon, t) = t \epsilon^\frac{1-\gamma}{2} |\ln \epsilon|^{\frac{2\beta}{\rho}} \exp \left\{ \left( \frac{1-\gamma}{2} \right) \ln \epsilon + K_2 t^p |\ln \epsilon|^{2\beta} \right\},
\]

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:

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

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

\[
\left[ \mathbb{E} \sup_{s \leq t} |A_2|^p \right]^{\frac{1}{p}} \leq \epsilon \left[ \mathbb{E} \sum_{n=0}^{N-1} \left| \int_{t_n}^{t_{n+1}} g(F_{t-t_n}(y^n_{t_n}, \theta_{t_n}(\omega))) dr - \Delta t Q^g(\pi(y^n_{t_n})) \right|^p \right]^{\frac{1}{p}}
\]

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

\[
= \epsilon \Delta t \sum_{n=0}^{N-1} \mathbb{E} \left[ \left| \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} g(F_{t-t_n}(y^n_{t_n}, \theta_{t_n}(\omega))) dr - Q^g(\pi(y^n_{t_n})) \right|^p \right]^{\frac{1}{p}}.
\]
For all \( n = 0, \ldots, N - 1 \), by the ergodic theorem, the two terms inside the modulus converges to each other when \( \Delta t \) goes to infinity with rate \( \eta(\Delta t) \). Hence, for small \( \epsilon \) we have

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

\[
\leq \epsilon \left[ \epsilon^{-1} |\ln \epsilon|^{-\frac{2\alpha}{p}} \right] t |\ln \epsilon|^{\frac{2\alpha}{p}} \eta \left( t |\ln \epsilon|^{\frac{2\alpha}{p}} \right)
\]

\[
= t \eta \left( t |\ln \epsilon|^{\frac{2\alpha}{p}} \right).
\]

\[ \square \]

**Lemma 3.4** \( A_3 \) converges to zero when \( t \) or \( \epsilon \) go to 0. Moreover, if \( Q^\alpha \) is \( \alpha \)-Hölder continuous then the rate of convergence satisfies

\[
\left( \mathbb{E} \sup_{s \leq t} |A_3|^p \right)^{\frac{1}{p}} \leq C \epsilon^\alpha t^{1+\alpha} \ln \epsilon^{\frac{2\alpha}{p}},
\]

for a positive constant \( C \).

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

\[
|A_3| \leq \sum_{n=0}^{N-1} \epsilon \Delta t \sup_{\epsilon t_n < r \leq \epsilon t_{n+1}} |Q^\alpha(\pi(y^\epsilon_{t_n})) - Q^\alpha(\pi(y^\epsilon_{t}))| \]

\[
\leq \epsilon \Delta t \sum_{n=0}^{N-1} C_1 \left( \sup_{\epsilon t_n < r \leq \epsilon t_{n+1}} |\pi(y^\epsilon_{t_n}) - \pi(y^\epsilon_{r})| \right)^\alpha. \tag{15}
\]

We use now that

\[
|\pi(y^\epsilon_{t}) - \pi(y^\epsilon_{r})| \leq \left( \sup_{x \in U} |K(x)| \right) |s - t|
\]

for all \( s, t \geq 0 \), such that the right hand side is independent of \( \epsilon \) (we are going to use this fact again, which follows either by equation (14) or see more details at the very beginning of proof of Theorem 4.1). Hence, continuing the estimates for \( |A_3| \), Inequality (15) above implies that

\[
|A_3| \leq C_2 N (\epsilon \Delta t)^{(1+\alpha)}
\]

\[
\leq C_2 \left( \epsilon^{-1} |\ln \epsilon|^{-\frac{2\alpha}{p}} \right) \left( \epsilon t |\ln \epsilon|^{\frac{2\alpha}{p}} \right)^{1+\alpha}
\]

\[
\leq C_2 \epsilon^\alpha t^{1+\alpha} |\ln \epsilon|^{\frac{2\alpha}{p}}
\]

for a constant \( C_2 \). \[ \square \]
Lemma 3.5 Process $A_4$ converges to zero with
\[ \left( \mathbb{E} \sup_{s \leq t} |A_4|^p \right)^{1/p} \leq C t \epsilon |\ln \epsilon|^{2/\beta}. \]

Proof: Denote
\[ C = \sup_{x \in U} |g(x)|. \]
The result follows straightforward since
\[ \epsilon \left| \int_{t_N}^{t \wedge \tau^\epsilon} g(y^\epsilon_r) dr \right| \leq C \epsilon \Delta t = C t \epsilon |\ln \epsilon|^{2/\beta}. \]

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). \qed

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 behaviour 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 $K$, i.e. $g \in \{d\pi_1(K), \ldots, d\pi_d(K)\}$. We assume that its corresponding average on the leaves $Q^g : V \to \mathbb{R}$ is Lipschitz.

This hypothesis holds naturally if the invariant measures $\mu_p$ for the unperturbed foliated system have 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 $V \subset \mathbb{R}^g$
\[ \frac{dv}{dt} = (Q^{d\pi_1(K)}, \ldots, Q^{d\pi_d(K)})(v(t)) \]
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
\[
\left[ \mathbb{E}\left( \sup_{s \leq t} \left| \pi \left( y^\varepsilon_{\gamma} \right) - v(s) \right|^p \right) \right]^{1/p} \leq C_1 \epsilon^\alpha + C_2 \eta \left( t |\ln \epsilon|^{-\frac{2\beta}{p}} \right),
\]
where $\eta(t)$ is the rate of convergence in $L^p$ of the ergodic averages of the un-perturbed trajectories on the leaves.

(2) For $\gamma > 0$, let
\[
T_\gamma = \inf \{ t > 0 \mid \text{dist}(v(t), \partial V) \leq \gamma \}.
\]

The exit times of the two systems satisfy the estimates
\[
\mathbb{P}(\varepsilon T^\varepsilon < T_\gamma) \leq \gamma^{-p} \left[ C_1(T_\gamma) \epsilon^\alpha + C_2(T_\gamma) \eta \left( T_\gamma |\ln \epsilon|^{-\frac{2\beta}{p}} \right) \right]^p.
\]

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

**Proof:** The gradient of each real function $\pi_i$ is orthogonal to the leaves, hence by Itô formula, for $i = 1, 2, \ldots, d$, we have that
\[
\pi_i \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) = \int_0^{t \wedge \varepsilon T^\varepsilon} d\pi_i(K) \left( y^\varepsilon_{\gamma} \right) ds.
\]

Lemma 3.1 for the function $d\pi_i(K)$ in $M$, triangular inequality and hypothesis (H) imply that
\[
\left| \pi_i \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) - v_i(t) \right| \leq \int_0^{t \wedge \varepsilon T^\varepsilon} \left| Q^{d\pi_i(K)} \left( \pi \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) \right) - Q^{d\pi_i(K)}(v(s)) \right| ds + |\delta_i(\varepsilon, t)|,
\]
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 lemma we have, for a constant $C$:
\[
\left| \pi \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) - v(t) \right| \leq e^{Ct} \sum_{i=1}^{n} |\delta_i(\varepsilon, t)|.
\]

The first part of the theorem follows by Lemma 3.1.

For the second part we have the following estimates
\[
\mathbb{P}(\varepsilon T^\varepsilon < T_\gamma) \leq \mathbb{P} \left( \sup_{s \leq T_\gamma \wedge \varepsilon T^\varepsilon} \left| v(s) - \pi \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) \right| > \gamma \right)
\leq \gamma^{-p} \mathbb{E} \left( \sup_{s \leq T_\gamma \wedge \varepsilon T^\varepsilon} \left| v(s) - \pi \left( \frac{y^\varepsilon_{\gamma}}{\varepsilon} \right) \right|^p \right)
\leq \gamma^{-p} \left[ C_1(T_\gamma) \epsilon^\alpha + C_2(T_\gamma) \eta \left( T_\gamma |\ln \epsilon|^{-\frac{2\beta}{p}} \right) \right]^p.
\]
4.1 A detailed example:

The following simple example illustrates the framework where the averaging principle described in this section holds. Consider \( M = \mathbb{R}^3 - \{(0,0,z), z \in R\} \) 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 linear SDE on \( M \) consisting of random rotations:

\[
\begin{pmatrix}
0 & -1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} x_t \end{pmatrix} \left( \lambda_1 dt + \lambda_2 \circ dB_t \right).
\]

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 neighbourhood \( U = \mathbb{R}^3 \setminus \{(x,0,z); x \leq 0; z \in \mathbb{R}\} \) given by cylindrical coordinates. Hence, using the same notation as before \( \varphi = (u,v) \) will be defined by \( \varphi : U \subset M \to (-\pi,\pi) \times \mathbb{R}^\times \mathbb{R} \), where \( u \in (-\pi,\pi) \) is angular and \( v = (r,z) \in \mathbb{R}^\times \times \mathbb{R} \) such that \( \varphi^{-1} : (u,v) \mapsto (r \cos u, r \sin u, z) \in M \). In this coordinates system, the transversal projections \( \pi_1 \) and \( \pi_2 \) correspond to the radial \( r \)-component and the \( z \)-coordinate, respectively.

For \( \lambda_1, \lambda_2 \in \mathbb{R} \) with \( |\lambda_1| + |\lambda_2| > 0 \), the invariant measures \( \mu_p \) 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 effective behaviour of a small transversal perturbation of order \( \epsilon \):

\[
\begin{pmatrix}
0 & -1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} y_t \end{pmatrix} \left( \lambda_1 dt + \lambda_2 \circ dB_t \right) + \epsilon K(y_t) dt.
\]

with initial condition \( x_0 = (1,0,0) \). Typically, due to uniformly ellipticity, for any perturbing vector field, the average of its transversal component on the leaves converges 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 vector field which is constant \( K = (k_1,k_2,k_3) \) with respect to Euclidean coordinates 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, i.e. \( Q^{dz}K = 0 \), but also, by the geometrical symmetry of \( K \) with respect to the invariant measure, the average radial \( r \)-component also vanishes, i.e. \( Q^{dr}K = 0 \). 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, concerns only the the difference between the initial radius \( r(0) = 1 \)
and \( r(\frac{t}{\epsilon} \wedge \tau^\epsilon) = \pi_1(y_{z, \tau^\epsilon}^\epsilon) \). Precisely, we have that

\[
\left[ \mathbb{E} \left( \sup_{s \leq t} \left| r \left( \frac{s}{\epsilon} \wedge \tau^\epsilon \right) - 1 \right|^p \right) \right]^{\frac{1}{p}}
\]

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

(B) Vertical perturbations. Now, for constant and vertical \( K = (0, 0, k_3) \), the radial average \( Q^d_{Z_1} K \) is null but \( Q^d_{Z_2} K \) equals \( k_3 \) for every leaf in \( M \). Hence the averaged system \( v(t) = (r(t), k_3 t) \) is constant in the radial component and increases linearly in the \( z \)-coordinate. The perturbed systems has the simple solution

\[
y_{z}^\epsilon = \begin{pmatrix}
\cos \left( \frac{\Delta t}{\epsilon} + \lambda_2 B \right) \\
\sin \left( \frac{\Delta t}{\epsilon} + \lambda_2 B \right) \\
k_3 t
\end{pmatrix}
\]

Hence, the comparison

\[
|\pi_2 \left( y_{z, \tau^\epsilon}^\epsilon \right) - v(t)| \equiv 0
\]

for all \( t \geq 0 \) 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 \( v(t) = (e^{\pi_1 r(t)}, z(0)) \). Hence the result guarantees that the radial part of \( y_{z, \tau^\epsilon}^\epsilon \) must have a behaviour close to the exponential \( e^{\frac{t}{\epsilon}} \) in the sense that

\[
\left[ \mathbb{E} \left( \sup_{s \leq t} \left| r \left( \frac{s}{\epsilon} \wedge \tau^\epsilon \right) - e^{\frac{t}{\epsilon}} \right|^p \right) \right]^{\frac{1}{p}}
\]

goes to zero when \( \epsilon \) goes to zero.

The fundamental solution of the linear perturbed Stratonovich systems \( v_t^\epsilon \) is given by the exponential of the matrix for each fixed \( t \)

\[
\begin{pmatrix}
et & - (\lambda_1 t + \lambda_2 B_t) & 0 \\
\lambda_1 t + \lambda_2 B_t & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

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

\[
\lambda_{1,2} = \frac{et \pm \sqrt{e^{2t^2} - 4(\lambda_1 t + \lambda_2 B_t)^2}}{2}.
\]
whose real part is given by $\epsilon t/2$ with probability increasing to 1 as $\epsilon$ goes to zero. This exponential rates coincides with that one above guaranteed by the Theorem 4.1.

**Lyapunov exponents.** For foliated manifolds embedded in $\mathbb{R}^N$, 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 $Q^{dK}$ vanishes. This phenomenon also appears in a couple of other examples where the leaves are not only diffeomorphic to each other, but also has this symmetry in the sense that the integration of a constant perturbation $K \in \mathbb{R}^N$ with respect to the Lebesgue measure is zero. To mention a couple of simple examples: the spherical foliation of $\mathbb{R}^n \setminus \{0\}$, nested torus (increasing the smaller radius) foliation of the solid torus minus the central circle $S^1 \times D^2 \setminus S^1 \subset \mathbb{R}^3$ or more generally (when they exists) tubular foliation of $\mathbb{R}^n \setminus \{C\}$, with $C$ a compact set (this context also includes the Hamiltonian case with the Lyouville foliation of the symplectic structure of $\mathbb{R}^{2n}$ 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 for instance, as in [6]) the averages of $Q^{dK}$ vanishes, hence our main theorem says that on the average, the trajectories of the perturbed system stay somehow close to the initial leaf as $\epsilon$ decreases to zero.

Lyapunov exponent of the system in the direction of a tangent vector $v \in T_{x_0} M$ contains the long time behaviour of points close to $x_0$ in the direction of $v$, for details on the definition, properties, existence conditions, multiplicative ergodic theory, etc see e.g. among many others L. Arnold [2] and the references therein. In particular, under the symmetric geometrical circumstances above, if there exists the Lyapunov exponents of the perturbed system $y^t_\epsilon$, Theorem 4.1 will imply that in transversal directions the Lyapunov exponent can not 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 [1, 5, 16], and Lyapunov exponents) do exist. In short:

**Proposition 4.2 (Continuity of Lyapunov exponents)** Assume that the perturbed system $y^t_\epsilon$ does have Lyapunov exponents a.s. at the assigned initial condition. If the averaged perturbation in the leaves vanishes, i.e. $Q^{dK} = 0$, then a number, given by the codimension of the foliation, of Lyapunov exponents in the spectrum go to nonpositive values as $\epsilon$ goes to zero.

**Proof:** In fact, Theorem 4.1 says that the perturbed system increases in the transversal coordinates with order $|\pi(y^t_\epsilon)| \lesssim C_1(\epsilon t) e^{t \alpha} + C_2(\epsilon t) \eta(\epsilon t |\ln \epsilon|^{\beta/p})$. Hence, any exponential behaviour, if it exists, tends to nonpositive values as $\epsilon$ decreases to zero.

**Acknowledgments:** This article has been written while the second author was in a sabbatical visit to University of Humboldt. He would like to express his gratitude
to Prof. Peter Imkeller and his research group for the nice and friendly hospitality. Author I. G.-G. has been supported by CNPq 555241/2009-2. and P. R. has been partially supported by FAPESP 11/50151-0 and 12/03992-1.

References

[1] L. Arnold and P. Imkeller – Rotation numbers for linear stochastic differential equations. Ann. Probab. 27, 130–149 (1999).

[2] L. Arnold – Random Dynamical Systems, Springer-Verlag (1998).

[3] V. Arnold – Mathematical Methods in Classical Mechanics. 2nd ed. Berlin, Springer (1989).

[4] A. Borodin and M. Freidlin – Fast oscillating random perturbations of dynamical systems with conservation laws. Ann. Inst. H. Poincaré. Prob. Statist., 31 (1995) 485-525.

[5] P. Catuogno, D. Ledesma and P. Ruffino – Relative rotation number for stochastic systems: dynamical and topological applications. Dynamical Systems, 23 (2008) 425-435.

[6] P. Catuogno, D. Ledesma and P. Ruffino – Harmonic measures in embedded foliated manifolds. Submitted. (2012) ArXiv 1208.0629.

[7] M. I. Freidlin and A. D. Wentzell – Random Perturbations of Dynamical Systems. Springer-Verlag, 1991.

[8] Y. Kabanov and S. Pergamenshchikov – Two-Scale Stochastic Systems: asymptotic analysis and control. Springer-Verlag, 2003.

[9] S. Kakutani and K. Petersen – The speed of convergence in the ergodic theorem. Monat. Mathematik 91, 11-18 (1981).

[10] R. Khasminski and N. Krylov – On averaging principle for diffusion processes with null-recurrent fast component. Stoch Proc Appl. 93 (2001) 229-240.

[11] V. Bakhtin and Y. Kifer – Nonconvergence examples in averaging. Geometric and probabilistic structures in dynamics, 1717, Contemp. Math. 469, 2008.

[12] U. Krengel – On the speed of convergence of the ergodic theorem. Monat. Mathematik 86, 3-6 (1978).

[13] Xue-Mei Li – An averaging principle for a completely integrable stochastic Hamiltonian systems. Nonlinearity, 21 (2008) 803-822.

[14] S. Namachchivaya and R. Sowers – Rigorous stochastic averaging at a center with additive noise. Meccanica, 37 (2002) 85-114.
[15] D. Revuz and M. Yor – *Brownian Motion and Continuous Martingales*. Springer-Verlag, 3rd ed. Berlin 1999.

[16] P. R. Ruffino – Decomposition of stochastic flow and rotation matrix. *Stochastics and Dynamics*, 2 (2002) 93-108.

[17] J.A. Sanders, F. Verhulst and J. Murdock – *Averaging Methods in Nonlinear dynamical Systems*. 2nd Ed. Springer, 2007.

[18] R. Sowers – Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere). *Trans. Am. Math. Soc*, 354 (2002) 853-900.

[19] P. Tondeur. *Foliations on Riemannian manifolds*. Universitext, Springer Verlag, Berlin-Heidelberg-New York, 1988.

[20] P. Walcak – *Dynamics of foliations, groups and pseudogroups*. Birkhäuser Verlag 2004.