Electron. Commun. Probab. **18** (2013), no. 51, 1–12.
DOI: 10.1214/ECP.v18-1975
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Double averaging principle for periodically forced slow-fast stochastic systems

Gilles Wainrib*

Abstract

This paper is devoted to obtaining an averaging principle for systems of slow-fast stochastic differential equations, where the fast variable drift is periodically modulated on a fast time-scale. The approach developed here combines probabilistic methods with a recent analytical result on long-time behavior for second order elliptic equations with time-periodic coefficients.

Keywords: averaging principle; slow-fast; stochastic differential equation; periodic averaging; inhomogeneous Markov process.
AMS MSC 2010: 70K70; 65C30.
Submitted to ECP on April 25, 2012, final version accepted on May 10, 2013.

1 Introduction

Time-scales separation is a key property to investigate the dynamical behavior of non-linear dynamical systems, with techniques ranging from averaging principles to geometric singular perturbation theory. This property appears to be also crucial to understand the impact of noise on such systems. A multi-scale approach based on the *stochastic averaging principle* can be a powerful tool to unravel subtle interplays between noise properties and non-linearities. More precisely, consider a system of stochastic differential equations (SDEs) in $\mathbb{R}^{p+q}$:

$$dx_t^\epsilon = \frac{1}{\epsilon} g(x_t^\epsilon, y_t^\epsilon)dt + \frac{1}{\sqrt{\epsilon}}\sigma(x_t^\epsilon, y_t^\epsilon)dB_t \tag{1.1}$$

$$dy_t^\epsilon = f(x_t^\epsilon, y_t^\epsilon)dt \tag{1.2}$$

with initial conditions $x^\epsilon(0) = x_0 \in \mathbb{R}^p,\ y^\epsilon(0) = y_0 \in \mathbb{R}^q$, and where y^ϵ is called the slow variable, x^ϵ the fast variable, with f, g, σ smooth functions ensuring existence and uniqueness for the solution (x^ϵ, y^ϵ), and B_t a p-dimensional standard Brownian motion. Time-scale separation in encoded in the small parameter $\epsilon \ll 1$.

In order to approximate the behavior of (x^ϵ, y^ϵ) for small ϵ, the idea of stochastic averaging is to average out the equation for the slow variable with respect to the stationary distribution of the fast one. More precisely, one first assumes that, for each $y \in \mathbb{R}^q$ fixed, the *frozen* fast SDE:

$$dx_t = g(x_t, y)dt + \sigma(x_t, y)dB_t \tag{1.3}$$

*Université Paris 13, Laboratoire Analyse Géométrie et Applications, France.
E-mail: wainrib@math.univ-paris13.fr

is sufficiently mixing, typically exponentially, and admits a unique invariant measure, denoted $\rho^y(dx)$. Then, one defines the averaged vector field $\bar{f}$:

$$\bar{f}(y) := \int_{\mathbb{R}^m} f(x,y)\rho^y(dx) \tag{1.4}$$

and $\bar{y}$ the solution of $\frac{d\bar{y}}{dt} = \bar{f}(\bar{y})$ with initial condition $\bar{y}(0) = y_0$.

According to the regularity and dissipativity assumptions made on the coefficients of the system, several convergence results have been proven in the literature, from convergence in law [11], to convergence in probability [5] or strong convergence in L^2 [3, 14, 7]. Many other results have been developed since, extending the set-up to the case where the slow variable has a diffusion component or to infinite-dimensional settings [1] for instance, and also refining the convergence study, providing *homogenization* results concerning the limit of $\epsilon^{-1/2}(y^\epsilon - \bar{y})$ [9] or establishing large deviation principles [6].

In terms of applications, analyzing the behavior of the deterministic solution $\bar{y}$ can help to understand useful dynamical features of the stochastic process (x^ϵ, y^ϵ). In particular, observing that the averaged vector field $\bar{f}$ depends on the diffusion coefficient σ can be the starting point for the understanding of stochastic bifurcations [13, 12].

However, fewer results are available in the case of non-homogeneous SDEs, that is when the system is perturbed by an external time-dependent signal [10]. This setting is particularly relevant to study models of learning in neuronal activity, which was the original motivation for the present paper. In this class of models, neurons are interconnected and the connections strengths evolve at a slower speed to account for synaptic plasticity, leading to a high-dimensional periodically forced slow-fast SDE. An application of the main result (Theorem 2.3) is developed in the particular context of learning models in [2].

Hence we are interested in multiscale SDEs driven by an external time-periodic input. Consider (x^ϵ, y^ϵ) solution of:

$$\begin{aligned} dx^\epsilon &= \frac{1}{\epsilon}\left[g(x^\epsilon, y^\epsilon, \frac{t}{\epsilon})\right]dt + \frac{1}{\sqrt{\epsilon}}\sigma(x^\epsilon, y^\epsilon)dB_t && (1.5) \\ dy^\epsilon &= f(x^\epsilon, y^\epsilon)dt && (1.6) \end{aligned}$$

with $t \to g(x,y,t) \in \mathbb{R}^p$ a τ-periodic function and $\epsilon \in \mathbb{R}_+$. We consider the case where ϵ is small, that is a strong time-scale separation between the fast variable $x^\epsilon \in \mathbb{R}^p$ and the slow one $y^\epsilon \in \mathbb{R}^q$, and a fast periodic modulation of the fast drift $g(x,y,.)$. Notice that the case of a slow periodic modulation would be less mathematically interesting, since in this case the time variable t appearing in the fast drift $g(x,y,t)$ could be treated as an additional slow variable satisfying $\dot{t} = 1$. This case is fully covered by the classical stochastic averaging principle described above. However, in our case of a fast modulation, one needs to develop a new result, based on a fine understanding of the asymptotic behavior of inhomogeneous Markov processes.

To obtain an averaging principle, one needs to understand the long time behavior of the rescaled periodically forced SDE, for any y_0 fixed :

$$dx = g(x, y_0, t)dt + \sigma(x, y_0)dB(t)$$

Recently, in an important contribution [8], a precise description of the long time behavior of inhomogeneous Markov diffusion processes has been obtained, using analytical methods. In particular, conditions ensuring the existence of a periodic family of probability measures $\mu(t, dx)$ to which the law of x converges as time grows have been identified, together with a sharp estimation of the speed of mixing. These results are at the heart of the extension of the classical stochastic averaging principle, that we present here, to the case of periodically forced slow-fast SDEs. As a result, we obtain a reduced equation describing the slow evolution of variable y in the form of an ordinary differential equation:

$$\frac{d\bar{y}}{dt} = \bar{f}(\bar{y})$$

where $\bar{f}$ is constructed as an average of f with respect to a specific probability measure, that is precisely the time-average over one period of the periodic family $\mu(t,dx)$. We prove the strong convergence in L^2 over finite time intervals of the slow variable y^ϵ to $\bar{y}$.

The paper is organized as follows. In the next section, we first recall the key theorem from [8], before stating our main convergence result in Theorem 2.3. In Section 3, we give the proof of our main result.

2 Main result

2.1 Preliminary : long-time behavior of inhomogeneous Markov diffusion processes

We recall here the recent result from [8]. Consider $X_t^{s,x}$ solution of the SDE:

$$dX_t^{s,x} = g(X_t^{x,s},t)dt + \sigma(X_t^{x,s})dW_t, \quad t > s \tag{2.1}$$

$$X_s = x \tag{2.2}$$

where g, σ are Lipschitz-continuous functions, and with the property that $t \to g(x,t)$ is a τ-periodic function of time.
Under the following assumptions:

Assumptions 2.1. (i) *The diffusion matrix σ is bounded:*

$$\exists M_\sigma > 0 \text{ s.t } \forall x,\ ||\sigma(x)|| < M_\sigma \tag{2.3}$$

and uniformly non-degenerate:

$$\exists \eta_0 > 0 \text{ s.t } \forall x\ < \sigma(x)\sigma(x)'\xi, \xi > \geq \eta_0 ||\xi||^2,\ \forall \xi \in \mathbb{R}^p \tag{2.4}$$

(ii) *There exists $r_0 < 0$ such that for all $t \geq 0$ and for all $x \in \mathbb{R}^p$:*

$$< \nabla_x g(x,t)\xi, \xi > \leq r_0 ||\xi||^2,\ \forall \xi \in \mathbb{R}^p \tag{2.5}$$

The following result holds:

Theorem 2.1. *(cf. [8], Theorem 3.15)*
There exist a unique τ-periodic family of probability measures $\{\mu(s,.), s \in \mathbb{R}\}$, such that:

$$\int_{x\in\mathbb{R}^p} \mathbb{E}\left[\phi(X_t^{s,x})\right]\mu(s,dx) = \int_{x\in\mathbb{R}^p} \phi(x)\mu(t,dx) \tag{2.6}$$

Such a family is called an evolution system of measures.
Furthermore, under the strong dissipativity condition (ii), the convergence of the law of X to μ is exponentially fast. More precisely, for any $r \in (1,+\infty)$ there exist $M > 0$ and $\omega < 0$, such that for all $\phi \in L^r(\mathbb{R}^p, \mu(t,.))$:

$$\int_{x\in\mathbb{R}^p} ||\mathbb{E}\left[\phi(X_t^{s,x})\right] - \int_{x'\in\mathbb{R}^p} \phi(x')\mu(t,dx')||^r \mu(s,dx) \leq Me^{\omega(t-s)} \int_{x\in\mathbb{R}^p} ||\phi(x)||^r \mu(t,dx) \tag{2.7}$$

2.2 Main result : averaging principle

We consider the following system (S^ϵ) of inhomogeneous stochastic differential equations, with fast periodic forcing:

$$dx_t^\epsilon = \frac{1}{\epsilon} g\left(x_t^\epsilon, y_t^\epsilon, \frac{t}{\epsilon}\right) dt + \frac{1}{\sqrt{\epsilon}} \sigma(x_t^\epsilon, y_t^\epsilon) dW_t \tag{2.8}$$

$$dy_t^\epsilon = f(x_t^\epsilon, y_t^\epsilon) dt \tag{2.9}$$

with initial conditions $(x_0^\epsilon, y_0^\epsilon) = (x_0, y_0) \in \mathbb{R}^p \times \mathbb{R}^q$, and where the function g is τ-periodic in time.
We are interested in describing the asymptotic behavior of (x^ϵ, y^ϵ) when $\epsilon \to 0$.
We make the following assumptions.

Assumptions 2.2.

1. *Existence and uniqueness of a strong solution: We make standard Lipschitz and linear growth assumptions on the coefficients, ensuring global existence and uniqueness for (S^ϵ) (cf. Thm 2.9 Ch. 5 in [4]). Note that one can replace the linear growth assumption by a condition on the drift which prevents explosion of the solution.*
2. *Asymptotic periodic behavior: for all $y \in \mathbb{R}^q$ fixed, denote $P^y_{t_0,x_0}(t,x)$ the transition density for the time-inhomogeneous diffusion process X_t^y solution of :*

$$dX_t^y = g(x_t, y, t)dt + \sigma(x_t, y)dW_t \tag{2.10}$$

starting at x_0 at $t = t_0$. We assume that there exists a τ-periodic family of probability measures $\mu^y(., dx)$ such that the law of X_t^y becomes close to $\mu^y(., dx)$ exponentially fast. This condition is ensured as soon as Assumptions 2.1 are satisfied by g and σ uniformly in the variable $y \in \mathbb{R}^q$. More precisely we assume:

$$\exists M_\sigma > 0 \text{ s.t } \forall (x,y) \in \mathbb{R}^{p+q} \,,\; ||\sigma(x,y)|| < M_\sigma \tag{2.11}$$

$$\exists \eta_0 > 0 \text{ s.t } \forall (x,y) \in \mathbb{R}^{p+q} \,,\; < \sigma(x,y)\sigma(x,y)'\xi, \xi > \geq \eta_0 ||\xi||^2,\; \forall \xi \in \mathbb{R}^p \tag{2.12}$$

and finally, there exists $r_0 < 0$ such that for all $t \geq 0$ and for all $(x,y) \in \mathbb{R}^{p+q}$:

$$< \nabla_x g(x,y,t)\xi, \xi > \leq r_0 ||\xi||^2, \forall \xi \in \mathbb{R}^p. \tag{2.13}$$

We also assume that the initial condition x_0 belongs to the support of $\mu^{y_0}(s,.)$ for all $s \in [0,\tau)$.
3. *Moment conditions. We further assume the following moment bounds:*

$$\sup_{y \in \mathbb{R}^q,\ s \in [0,\tau)} \int_{x \in \mathbb{R}^p} ||f(x,y)||^2 \mu^y(s, dx) < \infty \tag{2.14}$$

$$\sup_{t \in [0,T], \epsilon > 0} \mathbb{E}\left[||f(x_t^\epsilon, y_t^\epsilon)||\right] < \infty. \tag{2.15}$$

Before stating the main result of this paper, we need to introduce the following definition.

Definition 2.2. *We define the averaged vector field:*

$$\bar{f}(y) := \int_{x \in \mathbb{R}} f(x,y) \bar{\mu}^y(dx) \tag{2.16}$$

where

$$\bar{\mu}^y(dx) := \frac{1}{\tau} \int_0^\tau \mu^y(t, dx) dt. \tag{2.17}$$

Accordingly, we introduce $\bar{y}$ the solution of :

$$\frac{d\bar{y}}{dt} = \bar{f}(\bar{y}) \tag{2.18}$$

with initial condition $\bar{y}(0) = y_0$.
Then, under Assumptions 2.2, we have the following averaging principle:

Theorem 2.3. *The following convergence result holds :*

$$\lim_{\epsilon \to 0} \mathbb{E}\left[\sup_{t\in[0,T]} ||y_t^\epsilon - \bar{y}_t||^2\right] = 0 \tag{2.19}$$

As a consequence, the convergence also holds in probability : for all $T > 0$ and $\delta > 0$:

$$\lim_{\epsilon \to 0} \mathbf{P}\left[\sup_{t\in[0,T]} ||y_t^\epsilon - \bar{y}_t||^2 > \delta\right] = 0 \tag{2.20}$$

3 Proof of the averaging principle

The idea of the proof is to decompose the interval $[0, t]$ into many disjoint subintervals of size $\Delta = 1/n$. In each subinterval, the slow variable y^ϵ is almost constant and the fast variable x^ϵ, by a time change, is well described by the long time behavior of the frozen variable X_t^y. As we will see below the idea, introduced in [5], is to chose a subinterval size that depends on ϵ, and that is small enough to control the discrepancy between the fast variable and the frozen variable, yet large enough so that the frozen variable can be described by asymptotically periodic measure μ.
We start by splitting $[0, t]$ as the union of $L_k = [kt/n, (k+1)t/n]$ for $k = 0, ..., n-1$. Within each L_k we define $\hat{x}^\epsilon$ the strong solution of:

$$\begin{aligned} \text{For } kt/n < s \le (k+1)t/n \quad &: \quad d\hat{x}_s^\epsilon = \frac{1}{\epsilon} g(\hat{x}_s^\epsilon, y_{kt/n}^\epsilon, \frac{s}{\epsilon})ds + \frac{1}{\sqrt{\epsilon}}\sigma(x_s^\epsilon, y_{kt/n}^\epsilon)dW_s \\ \text{At } s = kt/n \quad &: \quad \hat{x}_{kt/n}^\epsilon = x_{kt/n}^\epsilon \end{aligned}$$

where W_t is the same Brownian path used in the definition of x^ϵ as the strong solution x^ϵ of (S^ϵ).
We write the difference $y_t^\epsilon - \bar{y}_t$ as a sum:

$$\begin{aligned} y_t^\epsilon - \bar{y}_t &= \int_0^t \left(f(x_s^\epsilon, y_s^\epsilon) - \bar{f}(\bar{y}_s)\right) ds \\ &= \sum_{k=0}^{n-1} (I_{1,k} + I_{2,k}) + \int_0^t \left(f(\hat{x}_s^\epsilon, y_s^\epsilon) - f(\hat{x}_s^\epsilon, \bar{y}_s)\right) ds \end{aligned}$$

with

$$\begin{aligned} I_{1,k} &:= \int_{kt/n}^{(k+1)t/n} \left(f(x_s^\epsilon, y_s^\epsilon) - f(\hat{x}_s^\epsilon, y_s^\epsilon)\right) ds \\ I_{2,k} &:= \int_{kt/n}^{(k+1)t/n} \left(f(\hat{x}_s^\epsilon, \bar{y}_s) - \bar{f}(\bar{y}_s)\right) ds \end{aligned}$$

We will show in Lemma 3.1 how to control the term $I_{1,k}$ in terms of n and ϵ, studying the difference $x^\epsilon - \hat{x}^\epsilon$ and using the Lipschitz property of f. To estimate the second term $I_{2,k}$, we will apply Lemma 3.2 below that shows how to take advantage of the results of [8] (cf. section 2.1) so that $I_{2,k}$ will be of order $O(\sqrt{\epsilon/n})$. From those estimates we will be able to chose $n(\epsilon)$ to control the growth of $y^\epsilon - \bar{y}$.

Lemma 3.1. *There exists a constant $C > 0$ such that:*

$$\sup_{s\in[0,t]} \mathbb{E}\left[||x^\epsilon_s - \hat{x}^\epsilon_s||^2\right] \le C\left(\frac{1}{\epsilon^2 n^3} + \frac{1}{\epsilon n^2}\right)\exp\left[C\left(\frac{1}{\epsilon^2 n^2} + \frac{1}{\epsilon n}\right)\right] \tag{3.1}$$

Proof. Let $s \in [0,t]$. There exists $k = k(s)$ such that $s \in L_k$ and we have:

$$\begin{aligned} x^\epsilon_s - \hat{x}^\epsilon_s &= \frac{1}{\epsilon}\int_{kt/n}^{s}\Big(g(x^\epsilon_u, y^\epsilon_u, u/\epsilon) - g(\hat{x}^\epsilon_u, y^\epsilon_{kt/n}, u/\epsilon)\Big)du \\ &+ \frac{1}{\sqrt{\epsilon}}\int_{kt/n}^{s}\Big(\sigma(x^\epsilon_u, y^\epsilon_u) - \sigma(\hat{x}^\epsilon_u, y^\epsilon_{kt/n})\Big)dW_u \end{aligned}$$

Using Cauchy-Schwartz inequality for the deterministic integral and Ito isometry for the stochastic one, we obtain:

$$\begin{aligned} \mathbb{E}\left[||x^\epsilon_s - \hat{x}^\epsilon_s||^2\right] &\le \frac{1}{\epsilon^2}\frac{1}{n}\int_{kt/n}^{s}\mathbb{E}||g(x^\epsilon_u, y^\epsilon_u, u/\epsilon) - g(\hat{x}^\epsilon_u, y^\epsilon_{kt/n}, u/\epsilon)||^2 du \\ &+ \frac{1}{\epsilon}\int_{kt/n}^{s}\mathbb{E}||\sigma(x^\epsilon_u, y^\epsilon_u) - \sigma(\hat{x}^\epsilon_u, y^\epsilon_{kt/n})||^2 du \end{aligned}$$

Now, using the Lipschitz property of g and σ, there exist $K, K' > 0$ such that:

$$\begin{aligned} \mathbb{E}\left[||x^\epsilon_s - \hat{x}^\epsilon_s||^2\right] &\le \frac{K}{\epsilon^2}\frac{1}{n}\int_{kt/n}^{s}\mathbb{E}||x^\epsilon_u - \hat{x}^\epsilon_u||^2 + \mathbb{E}||y^\epsilon_u - y^\epsilon_{kt/n}||^2 du \\ &+ \frac{K'}{\epsilon}\int_{kt/n}^{s}\mathbb{E}||x^\epsilon_u - \hat{x}^\epsilon_u||^2 + \mathbb{E}||y^\epsilon_u - y^\epsilon_{kt/n}||^2 du \end{aligned}$$

Since $\mathbb{E}||y^\epsilon_u - y^\epsilon_{kt/n}||^2 \le K''|u - kt/n|$ (by Assumption 2.2 (3.)) , we have:

$$\begin{aligned} \mathbb{E}\left[||x^\epsilon_s - \hat{x}^\epsilon_s||^2\right] &\le K''\left(\frac{K}{\epsilon^2}\frac{1}{n} + \frac{K'}{\epsilon}\right)\int_{kt/n}^{s}(u - kt/n)du \\ &+ \left(\frac{K}{\epsilon^2}\frac{1}{n} + \frac{K'}{\epsilon}\right)\int_{kt/n}^{s}\mathbb{E}||x^\epsilon_u - \hat{x}^\epsilon_u||^2 du \\ &\le C\left[\left(\frac{1}{\epsilon^2 n^3} + \frac{1}{\epsilon n^2}\right) + \left(\frac{1}{n\epsilon^2} + \frac{1}{\epsilon}\right)\int_{kt/n}^{s}\mathbb{E}||x^\epsilon_u - \hat{x}^\epsilon_u||^2 du\right] \end{aligned}$$

We conclude by applying Gronwall Lemma. □

The previous Lemma will help us to chose n large enough such that the frozen variable and the original fast variable would stay close. However, one is not allowed to take n too large (i.e the interval spacing too small) since the ergodic mixing needs some time to occur. The aim of next Lemma is to quantify this statement.
For $t_0 > 0$, we define $X^{\epsilon,x,y}_s$ solution of:

$$dX^{\epsilon,x,y}_s = \frac{1}{\epsilon}g(X^{\epsilon,x,y}_s, y, s/\epsilon)ds + \frac{1}{\sqrt{\epsilon}}\sigma(X^{\epsilon,x,y}_s, y)dW_s \tag{3.2}$$

for $s > t_0$ and initial condition $X_{t_0} = x$.

Lemma 3.2. *There exists a constant M such that for any $\xi > 0$, $y \in \mathbf{R}^q$ and $z \in [0,\tau]$, one can find a subsequence ϵ_k going to zero as $k \to \infty$ such that :*

$$\int_{x\in\mathbf{R}^p}\mathbf{E}\left[\left|\left|\frac{1}{\xi}\int_{t_0}^{t_0+\xi}(f(X^{\epsilon_k,x,y}_s, y) - \bar{f}(y))ds\right|\right|^2\right]\mu(z,dx) \le \frac{M\epsilon_k}{\xi} \tag{3.3}$$

Proof. In this proof we will make an essential use of the convergence rate of the law of the frozen process to its asymptotic time-periodic limit (cf. Section 2.1).

First, by a time change, we observe that $X_s^{\epsilon,x,y}$ has the same law as $X_{s/\epsilon}^{x,y}$ with $X^{x,y}$ solution of:

$$dX_s^{x,y} = g(X_s^{x,y}, y, s)ds + \sigma(X_s^{x,y}, y)dW_s \tag{3.4}$$

for $s > \frac{t_0}{\epsilon} := t_\epsilon$ and initial condition $X_{t_\epsilon} = x$. Then denoting $T = \xi/\epsilon$ we have:

$$\begin{aligned}
& \mathbb{E}\left[\left|\left|\frac{1}{\xi}\int_{t_0}^{t_0+\xi} \left(f(X_s^{\epsilon,x,y}, y) - \bar{f}(y)\right) ds\right|\right|^2\right] \\
= \; & \mathbb{E}\left[\left|\left|\frac{1}{T}\int_{t_\epsilon}^{t_\epsilon+T} \left(f(X_s^{x,y}, y) - \bar{f}(y)\right) ds\right|\right|^2\right] \\
= \; & \frac{1}{T^2}\int_{t_\epsilon}^{t_\epsilon+T}\int_{t_\epsilon}^{t_\epsilon+T} \mathbb{E}\left[\left(f(X_s^{x,y}, y) - \bar{f}(y)\right) . \left(f(X_r^{x,y}, y) - \bar{f}(y)\right)\right] dsdr \\
= \; & \frac{1}{T^2}\int_{t_\epsilon}^{t_\epsilon+T}\int_{t_\epsilon}^{t_\epsilon+T} \left(\mathbb{E}\left[f(X_s^{x,y}, y) f(X_r^{x,y}, y)\right] - \mathbb{E}\left[f(X_s^{x,y}, y)\right]\mathbb{E}\left[f(X_r^{x,y}, y)\right]\right) dsds' \\
+ \; & \left[\frac{1}{T}\int_{t_\epsilon}^{t_\epsilon+T} \left(\mathbb{E}\left[f(X_s^{x,y}, y)\right] - \bar{f}(y)\right) ds\right]^2
\end{aligned}$$

Let us denote Λ_1 and Λ_2 respectively the first and second term of the above sum. Using a change of variable, and conditioning w.r.t X_s, we bound Λ_1 as:

$$\begin{aligned}
||\Lambda_1|| \; &= \; \frac{2}{T^2}\left|\left|\int_{t_\epsilon}^{t_\epsilon+T}\int_0^{T+t_\epsilon-s} \mathbb{E}\left[\left(\mathbb{E}\left[f(X_{s+z}^{x,y}, y)|X_s^{x,y}\right] - \mathbb{E}\left[f(X_{s+z}^{x,y}, y)\right]\right) f(X_s^{x,y}, y)\right] dsdz\right|\right| \\
&\leq \; \frac{2}{T^2}\int_{t_\epsilon}^{t_\epsilon+T}\int_0^{T+t_\epsilon-s} \left|\left|\mathbb{E}\left[\left(\mathbb{E}\left[f(X_{s+z}^{x,y}, y)|X_s^{x,y}\right] - \mathbb{E}\left[f(X_{s+z}^{x,y}, y)\right]\right) f(X_s^{x,y}, y)\right]\right|\right| dsdz
\end{aligned}$$

Using Cauchy-Schwartz inequality to the integrand, we obtain:

$$\begin{aligned}
& \left|\left|\mathbb{E}\left[\left(\mathbb{E}\left[f(X_{s+z}^{x,y}, y)|X_s^{x,y}\right] - \mathbb{E}\left[f(X_{s+z}^{x,y}, y)\right]\right) f(X_s^{x,y}, y)\right]\right|\right| \\
\leq \; & \left(\mathbb{E}\left[||f(X_s^{x,y}, y)||^2\right]\mathbb{E}\left[\left|\left|\mathbb{E}\left[f(X_{s+z}^{x,y}, y)|X_s^{x,y}\right] - \mathbb{E}\left[f(X_{s+z}^{x,y}, y)\right]\right|\right|^2\right]\right)^{1/2}
\end{aligned}$$

From Assumption 2.2 (3.), we know that $\mathbb{E}\left[||f(X_s^{x,y}, y)||^2\right]$ will be uniformly bounded by a constant C_1. Furthermore, we deduce from Theorem 2.1 (cf. [8]) that

$$h_z(x) := \mathbb{E}\left[\left|\left|\mathbb{E}\left[f(X_{s+z}^{x,y}, y)|X_s^{x,y}\right] - \mathbb{E}\left[f(X_{s+z}^{x,y}, y)\right]\right|\right|^2\right]$$

goes to zero exponentially fast in z. Indeed, we first use Eq. (2.6) which implies:

$$\int_{x\in\mathbb{R}^p} h_z(x)\mu(t_\epsilon, dx) = \int_{x'\in\mathbb{R}^p}\left|\left|\mathbb{E}_{x',s}\left[f(X_{s+z}^y, y)\right] - \int_{x\in\mathbb{R}^p} f(x,y)\mu(s+z, dx)\right|\right|^2 \mu(s, dx') \tag{3.5}$$

Then from Eq. (2.7), we conclude that there exist constants $M, \kappa > 0$ such that:

$$\int_{x\in\mathbb{R}^p} h_z(x)\mu(t_\epsilon, dx) \leq Me^{-\kappa z}\int_{x\in\mathbb{R}^p} ||f(x,y)||^2\mu(s+z, dx) \leq MC_1e^{-\kappa z} \tag{3.6}$$

By Cauchy-Schwartz inequality, and since $h_z(x) \geq 0$:

$$\int_{x\in\mathbb{R}^p} h_z(x)^{1/2}\mu(t_\epsilon, dx) \leq \left(\int_{x\in\mathbb{R}^p} h_z(x)\mu(t_\epsilon, dx)\right)^{1/2} \tag{3.7}$$

so that:

$$\begin{aligned}
\int_{x\in\mathbb{R}^p} ||\Lambda^1||\mu(t_\epsilon, dx) &\leq C_1^{1/2}\frac{2}{T^2}\int_{t_\epsilon}^{T+t_\epsilon}\int_0^{T+t_\epsilon-s}\int_{x\in\mathbb{R}^p}(h_z(x))^{1/2}\mu(t_\epsilon, dx)dzds \\
&\leq C_1^{1/2}\frac{2}{T^2}\int_{t_\epsilon}^{T+t_\epsilon}\int_0^{T+t_\epsilon-s}\left(\int_{x\in\mathbb{R}^p} h_z(x)\mu(t_\epsilon, dx)\right)^{1/2} dzds \\
&\leq C_1 M^{1/2}\frac{2}{T^2}\int_{t_\epsilon}^{T+t_\epsilon}\left(\int_0^{T+t_\epsilon-s} e^{-\kappa z/2}dz\right)ds \\
&= 4C_1 M^{1/2}\left[\frac{1}{\kappa T} - \frac{2}{\kappa^2 T^2}(1-e^{-\kappa T/2})\right] = O(\frac{1}{T}) = O(\epsilon/\xi)
\end{aligned}$$

Furthermore, one remarks that the above bound holds for any ϵ sufficiently small, say $\epsilon < \epsilon_0$, in particular any $\epsilon \in [\epsilon_1, \epsilon_0]$ where ϵ_1 is such that $\frac{t_0}{\epsilon_1} = \frac{t_0}{\epsilon_0} + \tau$. Because of the time periodicity of μ, one concludes that the obtained bound is also valid for any $\int_{x\in\mathbb{R}^p} ||\Lambda^1||\mu(z, dx)$, with $z \in [0, \tau)$. Notice that in the bound we have obtained, ϵ is such that $z = \frac{t_0}{\epsilon}[\tau]$.
The second term Λ_2 also goes to 0, as $O(1/T^2)$ because $\mathbb{E}\left[f(X_s^{x,y}, y)\right]$ is asymptotically periodic, which ends the proof. □

In the following Lemma, we establish a link between the law of x_t^ϵ, denoted P_t^ϵ and the law $\mu^{\bar{y}_t}(z,.)$, for some $z \in [0, \tau)$. To this end, we need first to introduce an appropriate distance between two probability measures, namely the Kantorovich distance here.

Definition 3.3. *If P and Q are two probability measures on $\mathbb{R}^p$, one defines:*

$$d(P, Q) := \sup_{Lip(h)\leq 1}\left|\int hdP - \int hdQ\right| \tag{3.8}$$

where the supremum is taken over all Lipschitz functions $h : \mathbb{R}^p \to \mathbb{R}$ such that $|h(x) - h(y)| \leq ||x - y||$.

Lemma 3.4. *With the above notations, for any $t > 0$ and any $\epsilon > 0$, there exists a constant $C = C(t) > 0$ (independent of ϵ) such that*

$$d(P_t^\epsilon, \mu^{\bar{y}_t}(z,.)) \leq C\left(\mathbb{E}\left[\int_0^t ||y_s^\epsilon - \bar{y}_s||\, ds\right] + e^{-\kappa t/\epsilon}\right) \tag{3.9}$$

with $z \equiv \frac{t}{\epsilon}\,[\tau]$.

Proof. The idea of the proof is to decompose the distance $d(P_t^\epsilon, \mu^{\bar{y}_t}(z,.))$ using the triangular inequality as follows:

$$d(P_t^\epsilon, \bar{\mu}^{\bar{y}_t}) \leq d(P_t^\epsilon, \tilde{P}_t^\epsilon) + d(\tilde{P}_t^\epsilon, \mu^{\bar{y}_t}(z,.)) \tag{3.10}$$

where $\tilde{P}_t^\epsilon$ is the law of $\tilde{x}_t^\epsilon$ solution of the SDE:

$$d\tilde{x}_s^\epsilon = \frac{1}{\epsilon}g(\tilde{x}_s^\epsilon, \bar{y}_s, s/\epsilon)ds + \frac{1}{\sqrt{\epsilon}}\sigma(\tilde{x}_s^\epsilon, \bar{y}_s)dB_s \tag{3.11}$$

for $s > 0$, with initial condition $\tilde{x}_0^\epsilon = x$. Let us study both terms of the sum:

- First, from Theorem 2.1 (cf. [8]), we deduce that there exists a constant $M > 0$ such that:
$$d(\tilde{P}^\epsilon_t, \mu^{\bar{y}_t}(z,.)) \leq M e^{-\kappa t/\epsilon} \tag{3.12}$$
Indeed, by a time-change $u = s/\epsilon$, one remarks that $\tilde{P}^\epsilon_t$ is also the law of $X_{\frac{t}{\epsilon}}$ solution of:
$$dX_u = g(X_u, \bar{y}_t, u)du + \sigma(X_u, \bar{y}_t)dB_u$$
The long-time behavior of the above SDE (corresponding $\epsilon \to 0$ and t fixed) is described by Theorem 2.1, implying that the law of X_u converges exponentially fast to $\mu^{\bar{y}_t}(u[\tau],.)$ as u becomes large. Therefore, the law $\tilde{P}^\epsilon_t$ also converges exponentially fast to $\mu^{\bar{y}_t}(z[\tau],.)$. Notice that, since the left hand side of inequality (2.7) involves an average over μ, the constant M appearing in (3.12) shall depend on the initial condition x.
- Second, the distance $d(P^\epsilon_t, \tilde{P}^\epsilon_t)$ can be controlled by the spread between y^ϵ and $\bar{y}$:
$$d(P^\epsilon_t, \tilde{P}^\epsilon_t) \leq C\mathbb{E}\left[\int_0^t ||y^\epsilon_s - \bar{y}_s||\, ds\right] \tag{3.13}$$
Indeed,
$$\begin{aligned} |\mathbb{E}\left[h(\tilde{x}^\epsilon_t) - h(x^\epsilon_t)\right]| &\leq \mathbb{E}\left[||\tilde{x}^\epsilon_t - x^\epsilon_t||\right] & (3.14)\\ &\leq K\left[\int_0^t \mathbb{E}\,||y^\epsilon_s - \bar{y}_s||\, ds + \int_0^t \mathbb{E}\,||\tilde{x}^\epsilon_s - x^\epsilon_s||\, ds\right] & (3.15)\end{aligned}$$
Applying Gronwall Lemma to $\mathbb{E}\left[||\tilde{x}^\epsilon_t - x^\epsilon_t||\right]$ gives:
$$\mathbb{E}\left[||\tilde{x}^\epsilon_t - x^\epsilon_t||\right] \leq Ke^{Kt}\int_0^t \mathbb{E}\,||y^\epsilon_s - \bar{y}_s||\, ds$$
implying (3.13) with $C = Ke^{Kt}$.

□

We are now able to conclude the proof of the averaging principle:

Proof of Theorem 2.3

Proof. Equipped with Lemmas 3.1 and 3.2, we are now able to select the value of $n(\epsilon)$ so that the subintervals size $\Delta(\epsilon)$ would be:

- sufficiently small to be able to approximate x^ϵ by $\hat{x}^\epsilon$ during a time $\Delta(\epsilon)$, that is we want the right hand size of Eq. (3.1) goes to 0 when $\epsilon \to 0$
- sufficiently large for the mixing to occur: each $I_{2,k}$ is of order $O(\sqrt{\epsilon/n(\epsilon)})$ (from Lemma 3.2 with $\xi = 1/n(\epsilon)$) and we have to sum them $n(\epsilon)$ times, that is we want $\sqrt{\epsilon n(\epsilon)} \to 0$.

To this end, we set:
$$n(\epsilon) = \frac{1}{\epsilon \ln(1/\epsilon)^h}$$
and plugging this expression in Eq. (3.1), we obtain that: if $2h - 1 < 0$ then the right hand side of Eq. (3.1) goes to 0 when $\epsilon \to 0$. So we choose $h = 1/4$ for instance. Obviously, the second requirement $\epsilon n(\epsilon) \to 0$ is also satisfied. So with this choice of $n(\epsilon)$, by Lemma 3.1 and using the Lipschitz property of f, we deduce that
$$\lim_{\epsilon \to 0} \sup_{t\in[0,T]} \mathbb{E}\left[||\int_0^t (f(x_s f^\epsilon, y^\epsilon_s) - f(\hat{x}^\epsilon_s, y^\epsilon_s)ds||\right] = 0$$

We want to to control each:

$$\tilde{I}_{2,k} = \int_{kt/n}^{(k+1)t/n} (f(\hat{x}_s^\epsilon, \bar{y}_{kt/n}) - \bar{f}(\bar{y}_{kt/n}))ds$$

In fact, Lemma 3.2 does not control exactly $\mathbb{E}\left[\left|\left|\tilde{I}_{2,k}\right|\right|^2\right]$ since the initial value $\hat{x}_{kt/n}^\epsilon$, which is equal to $x_{kt/n}^\epsilon$, is not exactly distributed according to $\mu^{\bar{y}_{kt/n}}(z,x)$ for $z \in [0,\tau)$, nor according to $\bar{\mu}^{\bar{y}_{kt/n}}(x)$. So one cannot directly apply Lemma 3.2. However, the law of $x_{kt/n}^\epsilon$ is shown in Lemma 3.4 to be close in some sense to $\mu^{\bar{y}_{kt/n}}(z,x)$ for some $z \in [0,\tau)$. More precisely we obtain:

$$\mathbb{E}\left[\left|\left|\tilde{I}_{2,k}\right|\right|^2\right] = \mathbb{E}_\nu\left[\left|\left|\tilde{I}_{2,k}\right|\right|^2\right] + R_{kt/n}^\epsilon \leq M\epsilon/n + R_{kt/n}^\epsilon$$

where $\nu = \mu^{\bar{y}_{kt/n}}(z,.)$ with $z = \frac{kt}{\epsilon n}[\tau]$ and with

$$R_{kt/n}^\epsilon \leq C\left(\mathbb{E}\left[\int_0^t ||y_s^\epsilon - \bar{y}_s||^2 ds\right] + e^{-\frac{\kappa kt}{n\epsilon}}\right)$$

where each $e^{-\frac{\kappa kt}{n\epsilon}}$ goes to zero as $\epsilon \to 0$ since n is chosen such that $n\epsilon \to 0$.
As for $s \in L_k$ we have $||\bar{y}_s - \bar{y}_{kt/n}|| \leq K/n$ so that $||I_{2,k} - \tilde{I}_{2,k}||$ is of order $1/n^2$ since $\int_0^{1/n} sds = 1/2n^2$. Using Gronwall Lemma, we deduce that

$$\lim_{\epsilon\to 0} \sup_{t\in[0,T]} \mathbb{E}\left[||y_t^\epsilon - \bar{y}_t||^2\right] = 0 \tag{3.16}$$

Finally, one applies Prop A.2 (Appendix) to conclude the proof. Indeed, $\Delta y_t^\epsilon = y_t^\epsilon - \bar{y}_t$ is continuous in the sense that there exists $K > 0$ such that:

$$\sup_{\epsilon>0} \mathbb{E}\left[\sup_{t\in[0,T-h]} \sup_{r,s\in[t,t+h]} |\Delta y_r^\epsilon - \Delta y_s^\epsilon|\right] \leq Kh$$

□

A Strong convergence and regularity

Let $(x_\epsilon(t))_{t\in[0,T]}$ a family of real valued stochastic processes, with ϵ a positive parameter. We are looking for a condition such that

$$\lim_{\epsilon\to 0} \sup_{t\in[0,T]} \mathbb{E}[|x_\epsilon(t)|] = 0 \quad \Rightarrow \quad \lim_{\epsilon\to 0} \mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] = 0 \tag{A.1}$$

In general, it is not true. Indeed, consider the following counter-example. Let j be a random integer chosen uniformly from $0, ..., n-1$. Let $x_n(t)$ be a piecewise linear function on $[0,1]$ as follows:

- $x_n(t) = 0$ if $t \notin J_n$ where $J_n = [j/n, (j+1)/n]$
- If $t \in J_n$, then the graph of $x_n(t)$ has a "tent shape": it vanishes at each endpoint and increases linearly with slope $2n$ as we move toward the midpoint so that it takes value 1 at the midpoint.

The resulting function $x_n(t)$ is piecewise linear on $[0,1]$, bounded by 1 and the slope of any linear segment is bounded by $2n$. Then, for each t, $\mathbb{E}[|x_n(t)|] \leq 1/n$ since $x_n(t) \neq 0$ only with probability $1/n$ and $0 \leq x_n(t) \leq 1$. However, $\mathbb{E}[\sup_{t\in[0,1]} |x_n(t)|] = 1$ since $\sup_{t\in[0,1]} |x_n(t)| = 1$ for every outcome.

We show below that if one controls the continuity of x^ϵ uniformly on $[0,T]$ then the implication becomes true.

Definition A.1. *For any $h > 0$, we define the modulus of continuity of a trajectory $(x(t))_{t\in[0,T]}$ by:*

$$\omega_x(h) := \sup_{t\in[0,T-h]} \sup_{r,s\in[t,t+h]} |x(r) - x(s)| \tag{A.2}$$

The main result is then the following:

Proposition A.2. *Suppose that:*

1. *there exists $\phi(\epsilon)$ such that $\sup_{t\in[0,T]} \mathbb{E}[|x_\epsilon(t)|] \le \phi(\epsilon) \to 0$ as $\epsilon \to 0$*
2. *$\sup_{\epsilon>0} \mathbb{E}\left[\omega_{x_\epsilon}(h)\right] \le c(h)$ with $\lim_{h\to 0} c(h) = 0$*

Then:

$$\mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] \le \inf_{\delta\in(0,1)} \left\{\phi(\epsilon)^\delta + c(T\phi(\epsilon)^{1-\delta})\right\} \tag{A.3}$$

Proof. From Lemma A.5, we have for all $n \in \mathbb{N}^*$:

$$\mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] \le (n+1)\phi(\epsilon) + c(T/n) \tag{A.4}$$

Choosing n dependent on ϵ as $n(\epsilon) + 1 = \phi(\epsilon)^{\delta-1}$ for a constant $\delta \in (0,1)$ we obtain:

$$\mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] \le \phi(\epsilon)^\delta + c(T\phi(\epsilon)^{1-\delta}) \tag{A.5}$$

so that:

$$\mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] \le \inf_{\delta\in(0,1)} \left\{\phi(\epsilon)^\delta + c(T\phi(\epsilon)^{1-\delta})\right\} \tag{A.6}$$

□

Remark A.3. *Under assumptions 1. and 2., we have in particular*

$$\lim_{\epsilon\to 0} \mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] = 0$$

Remark A.4. *In the case $c(h) = O(h^\alpha)$ and $\phi(\epsilon) = O(\epsilon^\gamma)$ one finds that*

$$\mathbb{E}[\sup_{t\in[0,T]} |x_\epsilon(t)|] = O(\epsilon^\mu) \text{ with } \mu = \frac{\gamma\alpha}{1+\alpha}$$

In fact we have found a bound μ on the convergence rate of $\sup_{t\in[0,T]} |x_\epsilon(t)|$ in L^1 as a function of the averaged uniform Hölder exponent α and the uniform convergence rate γ of $\mathbb{E}[|x_\epsilon(t)|]$. If the trajectories are very smooth, then α is large and μ is close to γ. If the trajectories are only uniformly Lipschitz-continuous (in the sense $\alpha = 1$), then the bound μ is half of γ.

To establish Prop. A.2 we have used the following inequality:

Lemma A.5. *For all $n \in \mathbb{N}^*$, we have the following inequality:*

$$\mathbb{E}[\sup_{t\in[0,T]} |x(t)|] \le (n+1) \sup_{t\in[0,T]} \mathbb{E}[|x(t)|] + \mathbb{E}\left[\omega_x(T/n)\right] \tag{A.7}$$

Proof. Denote $S := \sup_{t\in[0,T]} |x(t)|$. For $n \in \mathbb{N}^*$, consider a sequence of $n+1$ points $t_k = kT/n$ for $0 \le k \le n$ in the inverval $[0,T]$, and denote $S_n := \max_{0\le k\le n} |x(t_k)|$.
Then $S - S_n \ge 0$ and

$$\mathbb{E}[S - S_n] \le \mathbb{E}\left[\omega_x(T/n)\right] \tag{A.8}$$

Finally:

$$\begin{aligned} \mathbb{E}[S] &= \mathbb{E}[S_n] + \mathbb{E}[S - S_n] && \text{(A.9)} \\ &\le (n+1) \max_{0\le k\le n} \mathbb{E}[|x(t_k)|] + c(T/n) && \text{(A.10)} \end{aligned}$$

□

References

[1] S. Cerrai, *A khasminskii type averaging principle for stochastic reaction–diffusion equations*, The Annals of Applied Probability **19** (2009), no. 3, 899–948. MR-2537194

[2] M. Galtier and G. Wainrib, *Multiscale analysis of slow-fast neuronal learning models with noise.*, The Journal of Mathematical Neurosciences, to appear. MR-3029813

[3] D. Givon, *Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems*, Multiscale Modeling & Simulation **6** (2007). MR-2338495

[4] I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, Springer, 1991. MR-1121940

[5] R.Z. Khas' minskii, *The averaging principle for stochastic differential equations*, Problemy Peredachi Informatsii **4** (1968), no. 2, 86–87.

[6] Y. Kifer, *Large deviations and adiabatic transitions for dynamical systems and markov processes in fully coupled averaging*, Memoirs of the AMS **201** (2009). MR-2547839

[7] D. Liu, *Strong convergence of principle of averaging for multiscale stochastic dynamical systems*, Commun. Math. Sci. **8** (2010), 999–1020. MR-2744917

[8] L. Lorenzi, A. Lunardi, and A. Zamboni, *Asymptotic behavior in time periodic parabolic problems with unbounded coefficients*, Journal of Differential Equations **249** (2010), no. 12, 3377–3418. MR-2737435

[9] G.C. Papanicolaou, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain J. Math **6** (1976), no. 4, 653–674. MR-0431378

[10] A.Y. Veretennikov, *On large deviations in the averaging principle for stochastic differential equations with periodic coefficients. ii*, Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya **55** (1991), no. 4, 691–715. MR-1137583

[11] ———, *On the averaging principle for systems of stochastic differential equations*, Mathematics of the USSR-Sbornik **69** (1991), 271. MR-1046602

[12] G. Wainrib, *Noise-controlled dynamics through the averaging principle for stochastic slow-fast systems.*, Phys. Rev. E (2011).

[13] G. Wainrib, M. Thieullen, and K. Pakdaman, *Reduction of stochastic conductance based neuron models with time-scales separation.*, Journal of Computational Neurosciences. (2011). MR-2904337

[14] E. Weinan, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Communications on Pure and Applied Mathematics **58** (2005), no. 11, 1544–1585. MR-2165382

Acknowledgments. The author thanks the anonymous reviewer for pointing out several improvements of this article.