# The averaging principle for stochastic differential equations driven by a Wiener process revisited

Charles-Edouard Bréhier

# The averaging principle for stochastic differential equations driven by a Wiener process revisited

Charles-Edouard Bréhier

Abstract. We consider a one-dimensional stochastic differential equation driven by a Wiener process, where the diffusion coefficient depends on an ergodic fast process. The averaging principle is satisfied: it is well-known that the slow component converges in distribution to the solution of an averaged equation, with generator determined by averaging the square of the diffusion coefficient.

We propose a version of the averaging principle, where the solution is interpreted as the sum of two terms: one depending on the average of the diffusion coefficient, the other giving fluctuations around that average. Both the average and fluctuation terms contribute to the limit, which illustrates why it is required to average the square of the diffusion coefficient to find the limit behavior.

## 1. Introduction

Multiscale and stochastic systems are ubiquitous in all fields of science and engineering. Averaging and homogenization techniques [**16**] are popular methods to derive lower dimensional problems, which are easier to understand and simulate. In this article, we focus on the averaging principle for the following class of stochastic differential equations (SDEs)

$$dX^\epsilon(t) = \sigma\big(X^\epsilon(t), m(t/\epsilon)\big)d\beta(t), \tag{1}$$

where $\epsilon \ll 1$ is the time scale separation parameter, $\beta$ is a standard real-valued Brownian motion, and the diffusion coefficient $\sigma$ is a smooth function. See Section 2.1 for precise assumptions. The fast component of the system is given by an ergodic Markov process $\big(m(t)\big)_{t\geqslant 0}$, evolving at the time scale $t/\epsilon$. The averaging principle states that one can eliminate the fast process when $\epsilon \to 0$, precisely the slow component $X^\epsilon$ converges (in distribution) to the solution $\overline{X}$ of an autonomous evolution equation called the averaged equation. In the case of the system (1), the averaged equation is a SDE of the type

$$d\overline{X}(t) = \Sigma(\overline{X}(t))d\beta(t), \tag{2}$$

where

$$\Sigma^2(\cdot) = \overline{\sigma^2}(\cdot) = \int \sigma(\cdot, m)^2 d\mu(m),$$

and $\mu$ denotes the invariant probability distribution of the fast ergodic process $\big(m(t)\big)_{t\geqslant 0}$.

1991 *Mathematics Subject Classification.* 60H10.

*Key words and phrases.* averaging principle; stochastic differential equations; Poisson equation.

In this article, we revisit this problem, and propose an original point of view which explains why the limit equation is not given by simply averaging the diffusion coefficient $\sigma$, which would give

$$dX(t) = \overline{\sigma}(X(t))d\beta(t).$$

Note that one has $\Sigma^2 = \overline{\sigma^2} \geqslant \overline{\sigma}^2$, thus the averaging principle may be interpreted as exhibiting enhanced diffusion. The approach used in this article can be explained as follows: we introduce a decomposition $X^\epsilon = Y^\epsilon + Z^\epsilon$ of the slow component, where

$$\begin{cases} dY^\epsilon(t) = \overline{\sigma}(Y^\epsilon(t) + Z^\epsilon(t))d\beta(t) \\ dZ^\epsilon(t) = \big(\sigma(Y^\epsilon(t) + Z^\epsilon(t), m^\epsilon(t)) - \overline{\sigma}(Y^\epsilon(t) + Z^\epsilon(t))\big)d\beta(t). \end{cases} \tag{3}$$

Observe that $Y^\epsilon$ is defined in terms of $\overline{\sigma}$, and thus may be interpreted as an average term, whereas $Z^\epsilon$ may be interpreted as a fluctuation term. The reason behind the expression of the averaged equation (2) in terms of $\overline{\sigma^2}$ is the fact that $Z^\epsilon$ converges to a non-trivial limit when $\epsilon \to 0$. Precisely, the main result of this article, Theorem 2.1, states that $(Y^\epsilon(T), Z^\epsilon(T))$ converges in distribution, when $\epsilon \to 0$, to $(Y(T), Z(T))$, for all $T \geqslant 0$, given by

$$\begin{cases} dY(t) = \overline{\sigma}(Y(t) + Z(t))d\beta_t^1 \\ dZ(t) = \langle\sigma\rangle(Y(t) + Z(t))d\beta_t^2, \end{cases} \tag{4}$$

where $\big(\beta_t^1\big)_{t\geqslant 0}$ and $\big(\beta_t^2\big)_{t\geqslant 0}$ are two independent standard real-valued Wiener processes, and $\langle\sigma\rangle^2 = \overline{(\sigma - \overline{\sigma})^2}$. It is then straightforward to retrieve the standard version of the averaging principle: $X^\epsilon(T) = Y^\epsilon(T) + Z^\epsilon(T) \to Y(T) + Z(T)$, and one checks that $Y(T) + Z(T)$ is equal to $\overline{X}(T)$ in distribution. That identity is due to the following observation: one has $\overline{\sigma}^2 + \langle\sigma\rangle^2 = \overline{\sigma^2}$. The decomposition into average and fluctuation terms then clearly explains the diffusion enhancement in the averaged equation (2).

The main result of this article has an elementary formulation. Even if the averaging principle has been extensively studied by many authors, to the best of our knowledge, it seems that the point of view proposed in this article is original and that Theorem 2.1 is a new result in the mathematical literature. The analysis is performed for a simple one-dimensional SDE, it may be generalized to more complicated problems.

Let us review the literature concerning the averaging principle for SDEs. The list of references is not exhaustive. We refer to the seminal article **[11]** by Hasminkskii and to the standard monograph **[8]** (in particular Chapter 7). See also **[16]** (in particular Chapter 17) for a recent overview of the averaging and homogenization techniques for SDEs. Let us also mention **[19]**, and the recent works **[17, 18]**. In the last decade, the averaging principle has been extensively studied for systems of stochastic partial differential equations, see for instance **[4, 5]**, contributions of the author **[1, 2]** and references therein. Recently Hairer and Li **[9]** have extended the averaging principle for SDE systems of the type (1) where the standard Brownian motion $\beta$ is replaced by a fractional Brownian motion $\beta^H$ with Hurst index $H > 1/2$: in Section 4 below we explain how the point of view developped in the present article is related to that generalization. Finally, numerical methods for systems of the type (1) which are efficient when $\epsilon \ll 1$ have been studied: see for instance the heterogeneous multiscale method proposed in **[7]** and the asymptotic preserving schemes proposed in **[3]**.

The proof of the main result Theorem 2.1 employs two standard tools when studying the behavior of multiscale stochastic systems: solutions of Kolmogorov and Poisson equations. We refer for instance to [**12**] and to the series of articles [**13, 14, 15**] for similar computations. See also [**1, 2**] where weak error estimates in the averaging principle for SPDEs are proved using such techniques, and [**17, 18**]. An original feature of the proof of Theorem 2.1 below is to consider the solutions of two Poisson equations (one related to the average behavior, one related to the fluctuations), whereas the standard approach to the averaging principle only requires a single Poisson equation. This may be surprising since the system (1) only depends on two time scales $t$ and $t/\epsilon$. The use of two Poisson equations is standard in homogenization or diffusion approximation problems, where three time scales $t$, $t/\epsilon$ and $t/\epsilon^2$ appear, see for instance [**6, 10**] and [**13, 14, 15**]. The list of references is not exhaustive.

This article is organized as follows. Section 2 is devoted to state the assumptions (Section 2.1) and the main result (Section 2.2) of this article. The proof of Theorem 2.1 is provided in Section 3. Concluding remarks and perspectives for future works are given in Section 4.

## 2. Setting and main result

Let $\big(\beta(t)\big)_{t\geqslant 0}$ be a standard real-valued Wiener process. Let $\epsilon \in (0,1)$ denote the time-scale separation parameter. We consider the following SDE on the one-dimensional torus $\mathbb{T}$

$$dX^\epsilon(t) = \sigma\big(X^\epsilon(t), m^\epsilon(t)\big)d\beta(t), \tag{5}$$

with initial condition $X_0^\epsilon = x_0 \in \mathbb{R}$ (assumed to be deterministic and independent of $\epsilon$ for simplicity). Assumptions for the diffusion coefficient $\sigma$ and the fast process $m^\epsilon$ are given in Section 2.1 below.

Working in the one-dimensional torus $\mathbb{T}$ simplifies the presentation, however one may replace $\mathbb{T}$ by $\mathbb{R}$ with minor modifications in the setting. Generalization to higher dimensional problems is mentioned in Section 4.

### 2.1. Assumptions.

The diffusion coefficient $\sigma$ is assumed to satisfy the following conditions.

Assumption 1. *The mapping $\sigma : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ is of class $\mathcal{C}^4$, with bounded derivatives with respect to the second variable $m$. In addition, assume that for all $x \in \mathbb{T}$, the mapping $\sigma(x,\cdot)$ is not constant.*

In particular, note that $\sigma$ is Lipschitz continuous, this ensures the global well-posedness of (5) for all $\epsilon > 0$.

The fast process $m^\epsilon$ is assumed to satisfy the following conditions:

Assumption 2. *For all $\epsilon \in (0,1)$ and all $t \geqslant 0$, one has $m^\epsilon(t) = m(t/\epsilon)$, where $\big(m(t)\big)_{t\geqslant 0}$ is a real-valued ergodic Markov process which is independent of $\beta$. We assume that the initial condition $m(0) = m_0$ is a given deterministic real number. Assume that $\sup_{t\geqslant 0} \mathbb{E}[|m(t)|^2] < \infty$.*

*Let $\mu$ denote the unique invariant probability distribution of the process $\big(m(t)\big)_{t\geqslant 0}$, and let $\mathcal{L}$ denote its infinitesimal generator.*

*Define, for all $x \in \mathbb{T}$,*

$$
\begin{cases} \overline{\sigma}(x) = \int \sigma(x,m) d\mu(m) \\ \langle\sigma\rangle(x) = \sqrt{\int \big(\sigma(x,m) - \overline{\sigma}(x)\big)^2 d\mu(m)}. \end{cases} \tag{6}
$$

*We assume that for all $x \in \mathbb{R}$, the Poisson equations*

$$
\begin{cases} -\mathcal{L}\psi_1(x,\cdot) = \sigma(x,\cdot) - \overline{\sigma}(x) \\ -\mathcal{L}\psi_2(x,\cdot) = \big(\sigma(x,\cdot) - \overline{\sigma}(x)\big)^2 - \langle\sigma\rangle(x)^2 \end{cases} \tag{7}
$$

*admit solutions $\psi_1$, $\psi_2$ – without loss of generality one assumes that for all $x \in \mathbb{T}$ one has $\int \psi_1(x,m)d\mu(m) = \int \psi_2(x,m)d\mu(m) = 0$ – and that the solutions $\psi_1, \psi_2$ are of class $\mathcal{C}^4$ on $\mathbb{T} \times \mathbb{R}$. In addition, the derivatives are assumed to grow at most quadratically with respect to $m$.*

Note that the mappings $\overline{\sigma}$ and $\langle\sigma\rangle^2$ inherit the regularity properties from the mapping $\sigma$ with respect to the $x$-variable: in particular they are of class $\mathcal{C}^4$ on the torus $\mathbb{T}$. Recall that for all $x \in \mathbb{T}$ the mapping $\sigma(x,\cdot)$ is not constant (owing to Assumption 1), thus one has $\langle\sigma\rangle^2(x) > 0$ for all $x \in \mathbb{T}$. As a consequence, $\langle\sigma\rangle$ then inherits the regularity properties from $\langle\sigma\rangle^2$, in particular it is of class $\mathcal{C}^4$.

Note that the solvability of the Poisson equations (7) is possible since the right-hand sides satisfy the required centering conditions by definitions (6) of $\overline{\sigma}$ and $\langle\sigma\rangle^2$. Observe also that for all $x \in \mathbb{T}$ one has

$$
\overline{\sigma}(x)^2 + \langle\sigma\rangle(x)^2 = \overline{\sigma^2}(x) = \int \sigma(x,m)^2 d\mu(m). \tag{8}
$$

Let us provide a standard example for the fast process: $\big(m(t)\big)_{t\geqslant 0}$ can be the solution of the SDE

$$
dm(t) = -V'(m(t))dt + \sqrt{2}dW(t),
$$

with appropriate assumptions on the potential $V : \mathbb{R} \to \mathbb{R}$ – for instance $V(x) = x^2/2$, which gives an Ornstein-Uhlenbeck process. In that example the fast process $\big(m^\epsilon(t)\big)_{t\geqslant 0}$ solves the SDE

$$
dm^\epsilon(t) = -\frac{V'(m^\epsilon(t))}{\epsilon}dt + \frac{\sqrt{2}}{\sqrt{\epsilon}}dW(t),
$$

and the invariant distribution $\mu$ is given by

$$
d\mu(m) = Z^{-1}\exp(-V(m))dm
$$

with the normalization constant $Z = \int \exp(-V(m))dm$. In that example, it is straightforward to check that the conditions in Assumption 2 are satisfied (with appropriate regularity and growth assumptions on $V'$).

**2.2. Main result.** The objective of this article is to propose a version of the averaging principle with an original point of view. First, recall that the standard version states that when $\epsilon \to 0$, the solution $X^\epsilon$ of (5) converges in distribution to the solution $\overline{X}$ of the averaged equation

$$d\overline{X}(t) = \Sigma(\overline{X}(t))d\beta(t) \tag{9}$$

with initial condition $\overline{X}(0) = x_0$, where

$$\Sigma(x) = \sqrt{\overline{\sigma^2(x)}}.$$

Note that $\Sigma(x) > \overline{\sigma}(x)$, owing to the identity (8) and Assumption 1. We refer for instance to [**16**, Chapter 17] (and the other references mentioned in Section 1).

The version of the averaging principle studied in this article requires to introduce two auxiliary processes $Y^\epsilon$ and $Z^\epsilon$ as follows: we consider the system

$$\begin{cases} dY^\epsilon(t) = \overline{\sigma}(Y^\epsilon(t) + Z^\epsilon(t))d\beta(t) \\ dZ^\epsilon(t) = \big(\sigma(Y^\epsilon(t) + Z^\epsilon(t), m^\epsilon(t)) - \overline{\sigma}(Y^\epsilon(t) + Z^\epsilon(t))\big)d\beta(t) \end{cases} \tag{10}$$

with initial conditions $Y^\epsilon(0) = 0$ and $Z^\epsilon(0) = x_0$. Observe that by construction, one has the identity

$$X^\epsilon(t) = Y^\epsilon(t) + Z^\epsilon(t)$$

for all $t \geqslant 0$.

The main result of this article is the convergence in distribution of $(Y^\epsilon(T), Z^\epsilon(T))$ to $(Y(T), Z(T))$, where the process $\big(Y(t), Z(t)\big)_{t\geqslant 0}$ is defined as follows:

$$\begin{cases} dY(t) = \overline{\sigma}(Y(t) + Z(t))d\beta_t^1 \\ dZ(t) = \langle\sigma\rangle(Y(t) + Z(t))d\beta_t^2, \end{cases} \tag{11}$$

where $\big(\beta_t^1\big)_{t\geqslant 0}$ and $\big(\beta_t^2\big)_{t\geqslant 0}$ are two independent standard real-valued Wiener processes.

We are now in position to state the refined version of the averaging principle.

THEOREM 2.1. *For all $T \in (0, \infty)$, one has the convergence in distribution*

$$(Y^\epsilon(T), Z^\epsilon(T)) \underset{\epsilon\to 0}{\to} (Y(T), Z(T)).$$

Note that the standard version of the averaging principle is a straightforward corollary of Theorem 2.1. On the one hand, one has the almost sure equality $X^\epsilon(T) = Y^\epsilon(T) + Z^\epsilon(T)$. On the other hand, set $X(t) = Y(t) + Z(t)$, then one has

$$dX(t) = \overline{\sigma}(X(t))d\beta_t^1 + \langle\sigma\rangle(X(t))d\beta_t^2.$$

The associated infinitesimal generator is given by

$$\frac{1}{2}\big(\overline{\sigma}(x)^2 + \langle\sigma\rangle(x)^2\big)\partial_{xx}^2 = \frac{1}{2}\overline{\sigma^2}(x)\partial_{xx}^2 = \frac{1}{2}\Sigma(x)^2\partial_{xx}^2,$$

owing to the identity (8). As a consequence $X$ and $\overline{X}$ are Markov processes with the same infinitesimal generator, and $X(0) = \overline{X}(0) = x_0$: we thus obtain the equality $X(T) = \overline{X}(T)$ in distribution. Finally, Theorem 2.1 implies

$$X^\epsilon(T) = Y^\epsilon(T) + Z^\epsilon(T) \underset{\epsilon\to 0}{\to} Y(T) + Z(T) = \overline{X}(T)$$

where the convergence and the equality are understood to hold in distribution.

The refined version is an explanation of the well-known fact that $\Sigma(x) > \overline{\sigma}(x)$ – which is often justified by the observation that one needs to average the infinitesimal generator of the process instead of its coefficients. It also illustrates why the convergence only holds in distribution. To the best of our knowledge, Theorem 2.1 is a new result.

Let us present a simplified case to illustrate Theorem 2.1: assume that $\sigma(x,m) = \sigma(m)$ only depends on $m$. In that case, $\overline{\sigma}$ and $\langle\sigma\rangle$ are constants, the system (10) is rewritten as

$$\begin{cases} dY^\epsilon(t) = \overline{\sigma} d\beta(t) \\ dZ^\epsilon(t) = \big(\sigma(m^\epsilon(t)) - \overline{\sigma}\big) d\beta(t). \end{cases}$$

In particular, the distribution of $Y^\epsilon(T)$ is $\mathcal{N}(0,\overline{\sigma}^2)$ does not depend on $\epsilon$. Owing to Theorem 2.1, $Z^\epsilon(T)$ converges in distribution to $Z(T) \sim \mathcal{N}(0,\langle\sigma\rangle^2 T)$. In fact, more precisely $(Y^\epsilon(T), Z^\epsilon(T))$ converges in distribution to the non-degenerate Gaussian distribution $\mathcal{N}(0,Q)$ with diagonal covariance matrix $Q$, such that $Q_{11} = \overline{\sigma}^2 T$, $Q_{22} = \langle\sigma\rangle^2 T$. Finally, $Y^\epsilon(T) + Z^\epsilon(T)$ converges in distribution to $\mathcal{N}(0,\overline{\sigma^2}T)$, since $\overline{\sigma}^2 + \langle\sigma\rangle^2 = \overline{\sigma^2}$. This confirms how Theorem 2.1 is a refinement of the standard averaging principle in the simplified case

## 3. Proof of Theorem 2.1

The objective of this section is to give the proof of Theorem 2.1. Before proceeding, let us first introduce some of the main arguments of the proof.

Assume that $\varphi : \mathbb{T}^2 \to \mathbb{R}$ is a mapping of class $\mathcal{C}^4$. We prove below that the weak error satisfies

$$\Big|\mathbb{E}[\varphi(Y^\epsilon(T), Z^\epsilon(T))] - \mathbb{E}[\varphi(Y(T), Z(T))]\Big| \leqslant C(T,\varphi,x_0)\epsilon \tag{12}$$

for some $C(T,\varphi,x_0) \in (0,\infty)$. By a standard approximation argument, the weak error estimate (12) implies that one has

$$\mathbb{E}[\varphi(Y^\epsilon(T), Z^\epsilon(T))] \underset{\epsilon\to 0}{\to} \mathbb{E}[\varphi(Y(T), Z(T))]$$

for all continuous mappings $\varphi : \mathbb{T}^2 \to \mathbb{R}$, which means the convergence in distribution stated in Theorem 2.1. It thus suffices to establish the weak error estimate (12) to prove Theorem 2.1.

To prove the weak error estimate (12), it is convenient to introduce two auxiliary mappings $u$ and $\Phi$ from $[0,T] \times \mathbb{T}^2 \times \mathbb{R}$ to $\mathbb{R}$. First, $u$ is the solution of the Kolmogorov equation associated with the SDE system (11) for $\big(Y(t), Z(t)\big)_{t\geqslant 0}$:

$$\partial_t u(t,y,z) = \frac{1}{2}\overline{\sigma}(y+z)^2 \partial_{yy}^2 u(t,y,z) + \frac{1}{2}\langle\sigma\rangle(y+z)^2 \partial_{zz}^2 u(t,y,z), \tag{13}$$

with initial condition $u(0,y,z) = \varphi(y,z)$ for all $(y,z) \in \mathbb{T}^2$. Using Assumption 1 and (6), one checks that $u$ is of class $\mathcal{C}^4$ with respect to $(y,z)$ and of class $\mathcal{C}^1$ with respect to $t$.

Second, for all $t \geqslant 0$, $(y,z) \in \mathbb{T}^2$ and $m \in \mathbb{R}$, set

$$\Phi(t,y,z,m) = \overline{\sigma}(y+z)\partial_{yz}^2 u(T-t,y,z)\psi_1(y,z,m) + \frac{1}{2}\partial_{zz}^2 u(T-t,y,z)\psi_2(y,z,m). \tag{14}$$

One checks that $\Phi$ is of class $\mathcal{C}^1$ with respect to $t$, and of class $\mathcal{C}^2$ with respect to $(y,z,m)$, with at most quadratic growth with respect to $m$.

Proof of Theorem 2.1. Expressing the weak error in terms of the solution $u$ of the Kolmogorov equation (13), and applying Itô's formula, one obtains

$$\begin{aligned}
&\mathbb{E}[\varphi(Y_T^\epsilon, Z_T^\epsilon)] - \mathbb{E}[\varphi(Y_T, Z_T)]\\
&\quad= \mathbb{E}[u(0, Y_T^\epsilon, Z_T^\epsilon)] - \mathbb{E}[u(T, Y_0^\epsilon, Z_0^\epsilon)]\\
&\quad= \int_0^T \mathbb{E}\big[-\partial_t u(T-t, Y_t^\epsilon, Z_t^\epsilon)\big]dt\\
&\quad+ \frac{1}{2}\int_0^T \mathbb{E}\big[\overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)^2 \partial_{yy} u(T-t, Y_t^\epsilon, Z_t^\epsilon)\big]dt\\
&\quad+ \int_0^T \mathbb{E}\big[\overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\big(\sigma(Y_t^\epsilon + Z_t^\epsilon, m_t^\epsilon) - \overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\big)\partial_{yz}^2 u(T-t, Y_t^\epsilon, Z_t^\epsilon)\big]dt\\
&\quad+ \frac{1}{2}\int_0^T \mathbb{E}\big[\big(\sigma(Y_t^\epsilon + Z_t^\epsilon, m_t^\epsilon) - \overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\big)^2 \partial_{zz}^2 u(T-t, Y_t^\epsilon, Z_t^\epsilon)\big]dt\\
&\quad= \int_0^T \mathbb{E}\Big[\overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\Big(\sigma(Y_t^\epsilon + Z_t^\epsilon, m_t^\epsilon) - \overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\Big)\partial_{yz}^2 u(T-t, Y_t^\epsilon, Z_t^\epsilon)\Big]dt\\
&\quad+ \frac{1}{2}\int_0^T \mathbb{E}\Big[\Big(\big(\sigma(Y_t^\epsilon + Z_t^\epsilon, m_t^\epsilon) - \overline{\sigma}(Y_t^\epsilon + Z_t^\epsilon)\big)^2 - \langle\sigma\rangle(Y_t^\epsilon + Z_t^\epsilon)^2\Big)\partial_{zz}^2 u(T-t, Y_t^\epsilon, Z_t^\epsilon)\Big]dt,
\end{aligned}$$

where the last line comes from replacing $\partial_t u$ using the Kolmogorov equation (13).

Observe that the two terms in the right-hand side above have a nice form, since the factors in parenthesis are centered with respect to the invariant distribution $\mu$ in the $m$ variable, and the other factors do not depend on $m$. Recall that the auxiliary functions $\psi_1$ and $\psi_2$ are defined as solutions of the Poisson equations (7). As a consequence, by the definition (14) of the auxiliary function $\Phi$, the weak error satisfies the identity

$$\mathbb{E}[\varphi(Y_T^\epsilon, Z_T^\epsilon)] - \mathbb{E}[\varphi(Y_T, Z_T)] = -\int_0^T \mathbb{E}[\mathcal{L}\Phi(t, Y_t^\epsilon, Z_t^\epsilon, m_t^\epsilon)]dt. \tag{15}$$

Applying Itô's formula, one has

$$\begin{aligned}
\mathbb{E}\big[\Phi(T, Y_T^\epsilon, Z_T^\epsilon, m_T^\epsilon)\big] &= \mathbb{E}[\Phi(0, Y_0^\epsilon, Z_0^\epsilon, m_0^\epsilon)]\\
&\quad+ \int_0^T \mathbb{E}\big[\mathcal{A}\Phi(t, Y_t^\epsilon, Z_t^\epsilon, m_t^\epsilon)\big]dt + \frac{1}{\epsilon}\int_0^T \mathbb{E}[\mathcal{L}\Phi(t, Y_t^\epsilon, Z_t^\epsilon, m_t^\epsilon)]dt,
\end{aligned}$$

where the auxiliary differential operator $\mathcal{A}$ is given by

$$\mathcal{A} = \partial_t + \frac{1}{2}\overline{\sigma}(y+z)^2\partial_{yy}^2 + \overline{\sigma}(y+z)\big(\sigma(y+Z,m) - \overline{\sigma}(y+z)\big)\partial_{yz}^2 + \frac{1}{2}\big(\sigma(y+Z,m) - \overline{\sigma}(y+z)\big)^2\partial_{zz}^2.$$

Finally, the weak error estimate satifies

$$\begin{aligned}
\mathbb{E}[\varphi(Y_T^\epsilon, Z_T^\epsilon)] - \mathbb{E}[\varphi(Y_T, Z_T)] &= \epsilon\big(\mathbb{E}[\Phi(0, Y_0^\epsilon, Z_0^\epsilon, m_0^\epsilon)] - \mathbb{E}[\Phi(T, Y_T^\epsilon, Z_T^\epsilon, m_T^\epsilon)]\big)\\
&\quad+ \epsilon\int_0^T \mathbb{E}\big[\mathcal{A}\Phi(t, Y_t^\epsilon, Z_t^\epsilon, m_t^\epsilon)\big]dt\\
&= \mathrm{O}(\epsilon)
\end{aligned}$$

using the regularity properties of $\Phi$ and the moment estimate

$$\sup_{\epsilon\in(0,1)} \sup_{t\geqslant 0} \mathbb{E}[|m^\epsilon(t)|^2] = \sup_{t\geqslant 0} \mathbb{E}[|m(t)|^2] < \infty$$

owing to Assumption 2.

This concludes the proof of the weak error estimate (12) and of Theorem 2.1. □

Observe that the proof of Theorem 2.1 requires to exploit the solutions $\psi_1$ and $\psi_2$ of two auxiliary Poisson equation. On the one hand, the proof of the standard averaging principle exploits the solution $\psi$ of a single Poisson equation, namely

$$-\mathcal{L}\psi(x,\cdot) = \sigma^2(x,\cdot) - \overline{\sigma^2}(x).$$

On the other hand, using the solutions of two Poisson equations is standard in homogenization theory, where the infinitesimal generator has an expansion of the form $\mathcal{L}^\epsilon = \mathcal{L}_0 + \epsilon^{-1}\mathcal{L}_1 + \epsilon^{-2}\mathcal{L}$ – whereas it is of the form $\mathcal{L}^\epsilon = \mathcal{L}_0 + \epsilon^{-1}\mathcal{L}$ in the averaging regime we consider. The two Poisson equation appears to deal with different scales $\epsilon^0$ and $\epsilon^{-1}$ in that problem.

## 4. Discussion

In this article, we have revisited the averaging principle for the class of stochastic differential equations given by (1). Contrary to the standard approach, we propose to decompose $X^\epsilon = Y^\epsilon + Z^\epsilon$ (see (10)), where $Y^\epsilon$ is defined in terms of the average $\overline{\sigma}$ (with respect to the fast variable) of the diffusion coefficient, and $Z^\epsilon$ represents fluctuations around the average. Our main result, Theorem 2.1, states that $(Y^\epsilon, Z^\epsilon)$ converges in distribution to a non-trivial limit $(Y, Z)$. The key observation is that $Z$ is not equal to 0, this explains why the limit $\overline{X}$ for $X^\epsilon$ is defined in terms of the average $\overline{\sigma^2}$ of the square of the diffusion coefficient. Note that $\overline{\sigma^2} \geqslant \overline{\sigma}^2$ by the Cauchy-Schwarz inequality (see (8)), *i.e.* diffusion is enhanced in the averaging procedure, and the behavior of the fluctuation term $Z^\epsilon$ quantifies the increase in the diffusion.

The approach to prove Theorem 2.1 is based on a classical strategy when studying multiscale stochastic systems: weak error estimates are proved using solutions of the Kolmogorov equation associated with the limit, and of Poisson equations associated with the behavior of the fast component. The solvability of the Poisson equations requires centering conditions to be satisfied, which identify limit. The proof of Theorem 2.1 is original since we employ the solutions of two Poisson equations, instead of only one in the standard proof of the averaging principle.

Our study is limited to one-dimensional SDEs. It is expected that generalizing the result to higher-dimensional SDEs and SPDEs is possible. This may be studied in future works. Note also that it would be straightforward to include drift terms in the SDE (1): since for those terms one would only need to average the drift term, one would only need to modify the definition of the average term $Y^\epsilon$, whereas the definition of the fluctuation term $Z^\epsilon$ would not be modified.

To conclude this article, let us mention that recently the averaging principle was proved for stochastic differential equations driven by a fractional Brownian motion with Hurst index $H > 1/2$, see [**9**]:

$$dX_t^\epsilon(t) = \sigma\big(X^\epsilon(t), m(t/\epsilon)\big)d\beta^H(t).$$

The expression of the averaged equation is different from (9): it is of the type

$$d\overline{X}^H(t) = \overline{\sigma}(\overline{X}^H(t))d\beta^H(t)$$

*i.e.* one simply needs to average the diffusion coefficient. In that case, the decomposition $X^\epsilon = Y^\epsilon + Z^\epsilon$ would give $Z^\epsilon \to 0$ when $\epsilon \to 0$, *i.e.* the fluctuation term does not contribute to the limit if $H > 1/2$ – in the same way as it does not contribute for drift terms. Our result thus illustrates the differences in the averaging principle between the standard and fractional Brownian motion cases. Note that, to the best of our knowledge, the validity and expression of the averaging principle if the Hurst index satisfies $H < 1/2$ is not known. The approach introduced in this article may be suitable to investigate this challenging question in future works.


## Acknowledgments

The author would like to thank Greg Pavliotis and Andrew Stuart for the suggestions of some references.


## References


[1] C.-E. Bréhier. Strong and weak orders in averaging for SPDEs. *Stochastic Process. Appl.*, 122(7):2553–2593, 2012.

[2] C.-E. Bréhier. Orders of convergence in the averaging principle for SPDEs: the case of a stochastically forced slow component. *Stochastic Process. Appl.*, 130(6):3325–3368, 2020.

[3] C.-E. Bréhier and S. Rakotonirina-Ricquebourg. On asymptotic preserving schemes for a class of stochastic differential equations in averaging and diffusion approximation regimes. *Preprint*, 2020.

[4] S. Cerrai. A Khasminskii type averaging principle for stochastic reaction-diffusion equations. *Ann. Appl. Probab.*, 19(3):899–948, 2009.

[5] S. Cerrai and M. Freidlin. Averaging principle for a class of stochastic reaction-diffusion equations. *Probab. Theory Related Fields*, 144(1-2):137–177, 2009.

[6] M. Coti Zelati and G. A. Pavliotis. Homogenization and hypocoercivity for Fokker-Planck equations driven by weakly compressible shear flows. *IMA J. Appl. Math.*, 85(6):951–979, 2020.

[7] W. E, D. Liu, and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.*, 58(11):1544–1585, 2005.

[8] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.

[9] M. Hairer and X.-M. Li. Averaging dynamics driven by fractional Brownian motion. *Ann. Probab.*, 48(4):1826–1860, 2020.

[10] M. Hairer and G. A. Pavliotis. Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.*, 117(1-2):261–279, 2004.

[11] R. Z. Hasminskii. On the principle of averaging the Itô's stochastic differential equations. *Kybernetika (Prague)*, 4:260–279, 1968.

[12] R. Z. Khasminskii and G. Yin. Limit behavior of two-time-scale diffusions revisited. *J. Differential Equations*, 212(1):85–113, 2005.

[13] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.

[14] E. Pardoux and A. Y. Veretennikov. On Poisson equation and diffusion approximation. II. *Ann. Probab.*, 31(3):1166–1192, 2003.

[15] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. III. *Ann. Probab.*, 33(3):1111–1133, 2005.

[16] G. A. Pavliotis and A. M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[17] M. Röckner, X. Sun, and L. Xie. Strong and weak convergence in the averaging principle for sdes with Hölder coefficients. *arXiv preprint arXiv:1907.09256*, 2019.

[18] M. Röckner and L. Xie. Averaging Principle and Normal Deviations for Multiscale Stochastic Systems. *Comm. Math. Phys.*, 383(3):1889–1937, 2021.

[19] A. Y. Veretennikov. On an averaging principle for systems of stochastic differential equations. *Mat. Sb.*, 181(2):256–268, 1990.

Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France

*E-mail address*: `brehier@math.univ-lyon1.fr`