SIAM J. NUMER. ANAL.
Vol. 51, No. 2, pp. 1185–1210

ANALYSIS OF AN HMM TIME-DISCRETIZATION SCHEME FOR A SYSTEM OF STOCHASTIC PDEs*

CHARLES-EDOUARD BRÉHIER†

Abstract. We consider the discretization in time of a system of parabolic stochastic partial differential equations with slow and fast components; the fast equation is driven by an additive space-time white noise. The numerical method is inspired by the averaging principle satisfied by this system and fits to the framework of heterogeneous multiscale methods. The slow and the fast components are approximated with two coupled numerical semi-implicit Euler schemes depending on two different time step sizes. We derive bounds of the approximation error on the slow component in the strong sense—approximation of trajectories—and in the weak sense—approximation of the laws. The estimates generalize the results of [W. E, D. Liu, and E. Vanden-Eijnden, *Comm. Pure Appl. Math.*, 58 (2005), pp. 1544–1585] in the case of infinite dimensional processes.

Key words. stochastic partial differential equations, heterogeneous multiscale method, time scale separation, Euler scheme, strong and weak approximation

AMS subject classifications. 60H15, 60H35, 70K70

DOI. 10.1137/110853078

1. Introduction.

In this paper, we are interested in the numerical approximation of a randomly perturbed system of reaction-diffusion equations that can be written

$$\begin{aligned} \frac{\partial x^\epsilon(t,\xi)}{\partial t} &= \frac{\partial^2 x^\epsilon(t,\xi)}{\partial \xi^2} + f(\xi, x^\epsilon(t,\xi), y^\epsilon(t,\xi)), \\ \frac{\partial y^\epsilon(t,\xi)}{\partial t} &= \frac{1}{\epsilon}\frac{\partial^2 y^\epsilon(t,\xi)}{\partial \xi^2} + \frac{1}{\epsilon} g(\xi, x^\epsilon(t,\xi), y^\epsilon(t,\xi)) + \frac{1}{\sqrt{\epsilon}}\frac{\partial \omega(t,\xi)}{\partial t} \end{aligned} \tag{1.1}$$

for $t \geq 0, \xi \in (0,1)$ with initial conditions $x^\epsilon(0,\xi) = x(\xi)$ and $y^\epsilon(0,\xi) = \xi$, and homogeneous Dirichlet boundary conditions $x^\epsilon(t,0) = x^\epsilon(t,1) = 0, y^\epsilon(t,0) = y^\epsilon(t,1) = 0$. The stochastic perturbation $\frac{\partial \omega(t,\xi)}{\partial t}$ is a space-time white noise, and $\epsilon > 0$ is a small parameter.

In the recent article [3], we have proved that an averaging principle holds for such a system, and we have exhibited an order of convergence—with respect to ϵ—in a strong and in a weak sense: the slow component x^ϵ is approximated thanks to the solution of an averaged equation. In this article, we analyze a numerical method of time discretization which reproduces this averaging effect at the discrete time level. More precisely, our aim is to build a numerical approximation of the slow component x^ϵ, taking care of the stiffness induced by the time scale separation. The heterogeneous multiscale method (HMM) procedure is used, as it is done in [14] for SDEs of the same kind. First we recall the general principle of such a method, which has been developed in various contexts, both deterministic and stochastic—see the review article [13] and the references therein, as well as [11], [12], [20]. We also mention the paper [1], which justifies the introduction of indirect methods for systems of SDEs with two time scales.

*Received by the editors October 26, 2011; accepted for publication (in revised form) January 2, 2013; published electronically April 11, 2013.
http://www.siam.org/journals/sinum/51-2/85307.html
†ENS Cachan Bretagne–IRMAR, Université Rennes 1, F-35170 Bruz, France (charles-edouard.brehier@bretagne.ens-cachan.fr).

Such systems of equations as (1.1) with two time scales may be used for the study of problems in chemistry or in biology, and in finance, when two phenomena occur at different speeds. An historical example where averaging occurs concerns celestial mechanics. The book [18] is a good reference for multiscale systems of equations. Precise assumptions on the coefficients f and g are given in section 3: they are bounded and of class $\mathcal{C}^2$ with bounded derivatives.

In system (1.1), the two components evolve at different time scales; x^ϵ is the slow component of the problem, while the fast component y^ϵ has fast variations in time. We are indeed interested in evaluating the slow component, which can be thought of as the mathematical model for a phenomenon appearing at the natural time scale of the experiment, whereas the fast component can often be interpreted as data for the slow component, taking care of effects at a faster time scale. Instead of using a direct numerical method, which might require a very small time step size because of the fast component, we use a different solver for each time scale: a macrosolver and a microsolver. The macrosolver leads to the approximation of the slow component; it takes into account data from the evolution at the fast time scale. The microsolver is then a procedure for estimating the unknown necessary data, using the evolution at the microtime scale, which also depends on the evolution at the macrotime scale. We focus on the situation where both solvers are constructed thanks to a semi-implicit Euler scheme, while in [14] the analysis is more general: first, we obtain a simple method which already contains the fundamental technical difficulties of the HMMs, and we use the results of [2] about the approximation of the invariant measure of a stochastic partial differential equation (SPDE) thanks to a numerical scheme.

In section 4.1, we state the two main theorems of this article: we show a strong convergence result—Theorem 4.1—as well as a weak convergence result—Theorem 4.2—which are similar to the available results for SDEs. Compared to [14], we propose modified and simplified proofs leading to apparently weaker error estimates; we made this choice for various reasons. First, even if apparently we get weaker estimates, under an appropriate choice of the parameters the cost of the method remains of the same order. Second, the generalization of the finite dimensional results would not yield the same bounds, due to the regularity assumptions we make on the nonlinear coefficients of the equations. Finally, we extend the weak convergence result to the situation where the fast equation only satisfies a weak dissipativity assumption.

In the case of a linear fast equation—when g is equal to 0—it is well known that the second equation in (1.1) is dissipative. In the general case, we make assumptions on g so that this property is preserved for y^ϵ—see Assumptions 3.7 and 3.8 below. The fast equation with frozen slow component—defined by (5.1) in the abstract framework—then admits a unique invariant probability measure, which is ergodic and strongly mixing, with exponential convergence to equilibrium. Under the strict dissipativity condition 3.7, we prove that the averaging principle holds in the strong and in the weak sense; moreover the fast numerical scheme has the same asymptotic behavior as the continuous time equation. If we assume only weak dissipativity of Assumption 3.8, the averaging principle holds only in the weak sense, and we can not prove uniqueness of the invariant law of the fast numerical scheme. Nevertheless, in the general setting [2] gives an approximation result of the invariant law of the continuous time equation with the numerical method which is used to prove Theorem 4.2; the order of convergence is $1/2$ with respect to the time step size—the precise result is recalled in Theorem 5.11.

The paper is organized as follows. In section 2, we give the definition of the numerical scheme. We then state the main assumptions made on the system of equa-

tions. In section 4.1 we state the two main theorems proved in this article, while in section 4.2 we compare the efficiency of the HMM scheme with a direct one in order to justify the use of a new method. Before proving the theorems, we give some useful results on the numerical schemes. Finally the last two sections contain the proofs of the strong and weak convergence theorems.

2. Description of the numerical scheme. Instead of working directly with system (1.1), we work with abstract stochastic evolution equations in Hilbert spaces H:

$$\begin{aligned} dX^\epsilon(t) &= (AX^\epsilon(t) + F(X^\epsilon(t), Y^\epsilon(t)))dt, \\ dY^\epsilon(t) &= \frac{1}{\epsilon}(BY^\epsilon(t) + G(X^\epsilon(t), Y^\epsilon(t)))dt + \frac{1}{\sqrt{\epsilon}}dW(t) \end{aligned} \tag{2.1}$$

with initial conditions $X^\epsilon(0) = x \in H$, $Y^\epsilon(0) = y \in H$.

To get system (2.1) from (1.1), we take $H = L^2(0,1)$; the linear operators A and B are Laplace operators with homogeneous Dirichlet boundary conditions and the nonlinearities F and G are Nemytskii operators—see Example 3.9. The process $(W(t))_{t\geq 0}$ is a cylindrical Wiener process on H. For precise assumptions on the coefficients, we refer to section 3.

We recall the idea of the averaging principle proved in the previous article [3]: when ϵ goes to 0, X^ϵ can be approximated by $\overline{X}$ defined by the averaged equation

$$\begin{aligned} \frac{d\overline{X}(t)}{dt} &= A\overline{X}(t) + \overline{F}(\overline{X}(t)), \\ \overline{X}(t=0) &= x \in H; \end{aligned} \tag{2.2}$$

the error is controlled in a strong sense by $C\epsilon^{1/2-r}$ and in a weak sense by $C\epsilon^{1-r}$, where $r > 0$ can be chosen as small as necessary and where C is a constant.

The averaged coefficient $\overline{F}$—see (5.4)—satisfies

$$\overline{F}(X) = \int_H F(X,y)\mu^X(dy) = \lim_{s\to+\infty} \mathbb{E}[F(X, Y_X(s,y))],$$

where μ^X is the unique invariant probability measure of the fast process Y_X with frozen slow component—more details are given in section 5.1.

To apply the HMM strategy, we need to define a macrosolver and a microsolver. We denote by Δt the macrotime step size, and by δt the microtime step size. Let also $T > 0$ be a given final time.

The construction of the macrosolver is deeply based on the averaging principle: for $n\Delta t \leq T$ $X^\epsilon(n\Delta t)$ can be approximated by $\overline{X}(n\Delta t)$. If the averaged coefficient $\overline{F}$ was known, one could build an approximation with a deterministic numerical scheme on the averaged equation; nevertheless in general it is not the case, and the idea is to calculate an approximation of this coefficient on-the-fly, by using the microsolver.

Therefore the macrosolver is defined in the following way: for any $0 \leq n \leq \lfloor \frac{T}{\Delta t} \rfloor := n_0$,

$$X_{n+1} = X_n + \Delta t A X_{n+1} + \Delta t \tilde{F}_n$$

with the initial condition $X_0 = x$. $\tilde{F}_n$ has to be defined; before that, we notice that the above definition leads to a semi-implicit Euler scheme—we use implicitness on the linear part, but the nonlinear part is explicit. If we define a bounded linear operator on H by $S_{\Delta t} = (I - \Delta t A)^{-1}$, we instead use the following explicit formula

$$X_{n+1} = S_{\Delta t} X_n + \Delta t S_{\Delta t} \tilde{F}_n. \tag{2.3}$$

We want $\tilde{F}_n$ to be an approximation of $\overline{F}(X_n)$. The role of the microsolver is to give an approximation of Y_{X_n}, the fast process with frozen slow component X_n, when n is fixed; moreover we compute a finite number M of independent replicas of the process in order to approximate theoretical expectations by discrete averages over different realizations of the random variables, in a Monte Carlo spirit. Therefore the microsolver is defined in the following way: we fix a realization index $j \in \{1, \dots, M\}$ and a macrotime step n; then for any $m \geq 0$

$$Y_{n,m+1,j} = Y_{n,m,j} + \frac{\delta t}{\epsilon} B Y_{n,m+1,j} + \frac{\delta t}{\epsilon} G(X_n, Y_{n,m,j}) + \sqrt{\frac{\delta t}{\epsilon}} \zeta_{n,m+1,j}.$$

As above we give an explicit formula

$$Y_{n,m+1,j} = R_{\frac{\delta t}{\epsilon}} Y_{n,m,j} + \frac{\delta t}{\epsilon} R_{\frac{\delta t}{\epsilon}} G(X_n, Y_{n,m,j}) + \sqrt{\frac{\delta t}{\epsilon}} R_{\frac{\delta t}{\epsilon}} \zeta_{n,m+1,j}, \tag{2.4}$$

where for any $\tau > 0$ $R_\tau = (I - \tau B)^{-1}$.

The noises $\zeta_{n,m,j}$ are defined by

$$\zeta_{n,m+1,j} = \frac{W^{(n,j)}_{(m+1)\delta t} - W^{(n,j)}_{m\delta t}}{\sqrt{\delta t}},$$

where $(W^{(n,j)})_{1 \leq j \leq M, 0 \leq n \leq n_0}$ are independent cylindrical Wiener processes on H. It is essential to use independent noises at each macrotime step. It is important to remark that this equation is well-posed in the Hilbert space H since R_τ is a Hilbert–Schmidt operator from H to H, under assumptions given in section 3.

The missing definition can now be written: $\tilde{F}_n$ is given by

$$\tilde{F}_n = \frac{1}{MN} \sum_{j=1}^{M} \sum_{m=n_T}^{n_T+N-1} F(X_n, Y_{n,m,j}). \tag{2.5}$$

n_T is the number of microtime steps that are not used in the evaluation of the average in $\tilde{F}_n$, while N is the number of microtime steps that are then used for this evaluation. Each macrotime step then requires the computation of $m_0 = n_T + N - 1$ values of the microsolver.

At each macrotime step $Y_{n,m,j}$ must be initialized at time $m = 0$. In our proofs, this is not as important as in [14], but for definiteness we use the same method; we initialize with the last value computed during the previous macrotime step:

$$\begin{aligned} Y_{n+1,0,j} &= Y_{n,m_0,j} \text{ for } n \geq 0, \\ Y_{0,0,j} &= y. \end{aligned} \tag{2.6}$$

The aim of the analysis for HMM schemes is to prove that under an appropriate choice of the parameters n_T, N, M of the scheme, we can bound the error by expressions of the following kind, where $r > 0$ is chosen as small as necessary: for $n = n_0 = \lfloor \frac{T}{\Delta t} \rfloor$, we have the strong error estimate

$$\mathbb{E}|X^\epsilon(n\Delta t) - X_n| \leq C \left(\epsilon^{1/2-r} + \Delta t^{1-r} + \left(\frac{\delta t}{\epsilon} \right)^{1/2-r} \right),$$

and if Φ is a test function of class $\mathcal{C}^2$, bounded and with bounded derivatives, we have the weak error estimate

$$|\mathbb{E}\Phi(X^\epsilon(n\Delta t)) - \mathbb{E}\Phi(X_n)| \leq C\left(\epsilon^{1-r} + \Delta t^{1-r} + \left(\frac{\delta t}{\epsilon}\right)^{1/2-r}\right).$$

The origin of the three error terms appears clearly in the proofs (see sections 6 and 7): the first one is the averaging error, the second one is the error in a deterministic scheme with the macrotime step, and the third one is the weak error in a scheme for stochastic equations with the microtime step. We recall that in the SPDE case the strong order of the semi-implicit Euler scheme for the microsolver used here is 1/4, while the weak order is 1/2, while in the SDE situation the respective orders are 1/2 and 1. The macrosolver is deterministic, so that the order is 1. Precise results for any choice of n_T, N, M are given in Theorems 4.1 and 4.2 below, while the choice of these parameters is explained in section 4.2.

3. Assumptions. As mentioned above, system (2.1) satisfies an averaging principle, and strong and weak order of convergence can be given. The HMM method relies on that idea. The natural assumptions are basically the same as the hypothesis needed to prove those results but must be strengthened sometimes.

3.1. Assumptions on the linear operators. First, we have to specify some properties of the linear operators A and B coming into the definition of system (2.1); we assume that the linear parts are of parabolic type with space variable $\xi \in (0,1)$.

We assume that A and B are unbounded linear operators with domains $D(A)$ and $D(B)$, which satisfy the following assumptions.

Assumption 3.1.

1. There exist complete orthonormal systems of H, $(e_k)_{k\in\mathbb{N}}$, and $(f_k)_{k\in\mathbb{N}}$, and $(\lambda_k)_{k\in\mathbb{N}}$ and $(\mu_k)_{k\in\mathbb{N}}$ nondecreasing sequences of real positive numbers such that

$$Ae_k = -\lambda_k e_k \text{ for all } k \in \mathbb{N},$$
$$Bf_k = -\mu_k f_k \text{ for all } k \in \mathbb{N}.$$

   We use the notation $\lambda := \lambda_0 > 0$ and $\mu := \mu_0 > 0$ for the smallest eigenvalues of A and B.
2. For every $k \in \mathbb{N}$, f_k is Lipschitz continuous and bounded on $[0,1]$ with a uniform control with respect to k: there exists $C > 0$ such that for any $\xi_1, \xi_2 \in [0,1]$

$$|f_k(\xi_1)| \leq C \quad \text{and} \quad |f_k(\xi_1) - f_k(\xi_2)| \leq C\sqrt{\mu_k}|\xi_1 - \xi_2|.$$

3. The sequences (λ_k) and (μ_k) go to $+\infty$; moreover we have some control of the behavior of (μ_k) given by

$$\sum_{k=0}^{+\infty} \frac{1}{\mu_k^\alpha} < +\infty \Leftrightarrow \alpha > 1/2.$$

In the abstract setting, powers of $-A$ and $-B$ with their domains are easily defined.

DEFINITION 3.2. *For $a, b \in [0,1]$, we define the operators $(-A)^a$ and $(-B)^b$ by*

$$(-A)^a x = \sum_{k=0}^{\infty} \lambda_k^a x_k e_k \in H \quad \text{and} \quad (-B)^b y = \sum_{k=0}^{\infty} \mu_k^b y_k f_k \in H$$

with domains

$$D(-A)^a = \left\{ x = \sum_{k=0}^{+\infty} x_k e_k \in H; |x|^2_{(-A)^a} := \sum_{k=0}^{+\infty} (\lambda_k)^{2a} |x_k|^2 < +\infty \right\};$$
$$D(-B)^b = \left\{ y = \sum_{k=0}^{+\infty} y_k f_k \in H, |y|^2_{(-B)^b} := \sum_{k=0}^{+\infty} (\mu_k)^{2b} |y_k|^2 < +\infty \right\}.$$

On $D(-A)^a$, the norm $|\,.\,|_{(-A)^a}$ and the Sobolev norm of H^{2a} are equivalent: when x belongs to a space $D(-A)^a$, the exponent a represents some regularity of the function x.

When $a \geq 0$, we also define a bounded linear operator $(-A)^{-a}$ in H with

$$(-A)^{-a} x = \sum_{k=0}^{+\infty} \lambda_k^{-a} x_k \in H,$$

where $x = \sum_{k=0}^{+\infty} x_k e_k \in H$.

Under the previous assumptions on the linear coefficients, the following stochastic integral is well defined in H for any $t \geq 0$ when $(W(t))_{t \geq 0}$ is a cylindrical Wiener process on H (see [7] for the definition and the properties of W):

$$W^B(t) = \int_0^t e^{(t-s)B} dW(s). \tag{3.1}$$

It is called a stochastic convolution, and it is the unique mild solution of

$$dZ(t) = BZ(t)dt + dW(t), \quad Z(0) = 0.$$

Under the second condition of Assumption 3.1, there exists $\delta > 0$ such that for any $t > 0$ we have $\int_0^t \frac{1}{s^\delta} |e^{sB}|^2_{\mathcal{L}_2(H)} ds < +\infty$; it is then proved that W^B has continuous trajectories—via the *factorization method*; see [7]—and that for any $1 \leq p < +\infty$, any $0 \leq \gamma < 1/4$, there exists a constant $C(p,\gamma) < +\infty$ such that for any $t \geq 0$

$$\mathbb{E}|W^B(t)|^p_{(-B)^\gamma} \leq C(p,\gamma). \tag{3.2}$$

3.2. Assumptions on the nonlinear coefficients. We now give the assumptions on the nonlinear coefficients $F, G : H \times H \to H$. First, we need some regularity properties.

Assumption 3.3. We assume that there exists $0 \leq \eta < \frac{1}{2}$ and a constant C such that the following directional derivatives are well defined and controlled:

- For any $x, y \in H$ and $h \in H$, $|D_x F(x,y).h| \leq C|h|_H$ and $|D_y F(x,y).h| \leq C|h|_H$.
- For any $x, y \in H$, $h, k \in H$, if the right-hand side is finite, we have

$$|D^2_{\cdot\cdot} F(x,y).(h,k)|_H \leq C|h|_H |k|_{(-\mathcal{D})},$$

  where $D^2_{\cdot\cdot}$ stands for a second directional derivative with respect to either x or y and with $\mathcal{D} = (-A)^\eta$ or $\mathcal{D} = (-B)^\eta$ according to the situation.

We moreover assume that F is bounded.

We also need the following.

Assumption 3.4. For η defined in the previous Assumption 3.3, we have for any $x, y \in H$ and $h, k \in H$

$$|(-A)^{-\eta} D^2_{\cdot\cdot} F(x,y).(h,k)| \leq C|h|_H |k|_H.$$

We assume that the fast equation is a gradient system: for any x the nonlinear coefficient $G(x,.)$ is the derivative of some potential U. We also assume regularity assumptions as for F.

Assumption 3.5. The function G is defined through $G(x,y) = \nabla_y U(x,y)$ for some potential $U : H \times H \to \mathbb{R}$. Moreover we assume that G is bounded and that the regularity assumptions given in Assumption 3.3 are also satisfied for G.

This condition only plays a role when we get the expression of the invariant probability measure of the fast process with frozen slow component, given by (5.3). This expression is essential to study the regularity properties of the averaged coefficient with respect to the x variable given in Proposition 5.5. Such results are required to obtain the orders of convergence in the averaging principle, given in [3]. We recall that the averaging principle still holds if this gradient assumption is removed—see [4]—but that the order of the convergence is then a priori unknown.

For G, we need a stronger hypothesis than for F in order to get Proposition 5.5. Assumption 3.4 becomes the following.

Assumption 3.6. We have for any $x, y \in H$, $h, k \in H$, $z \in L^\infty(0,1)$

$$|\langle D_{xx} G(x,y).(h,k), z\rangle_H| \leq C|h|_H |k|_H |z|_{L^\infty(0,1)}.$$

Finally, we need to assume some dissipativity of the fast equation. Assumption 3.7 is necessary to obtain strong convergence in the averaging principle, while Assumption 3.8 is weaker and leads to the weak convergence.

Assumption 3.7 (strict dissipativity). Let L_g denote the Lipschitz constant of G with respect to its second variable; then

$$L_g < \mu, \tag{SD}$$

where μ is defined in Assumption 3.1.

Assumption 3.8 (weak dissipativity). There exist $c > 0$ and $C > 0$ such that for any $y \in D(B)$

$$\langle By + G(y), y\rangle \leq -c|y|^2 + C. \tag{WD}$$

The second assumption is satisfied as soon as G is bounded, while the first one requires some knowledge of the Lipschitz constant of G.

Example 3.9. We give some fundamental examples of nonlinearities for which the previous assumptions are satisfied:

- Functions $F, G : H \times H \to H$ of class $\mathcal{C}^2$, bounded and with bounded derivatives, such that $G(x,y) = \nabla_y U(x,y)$ fit in the framework with the choice $\eta = 0$.
- Functions F and G can be Nemytskii operators: let $f : (0,1) \times \mathbb{R}^2 \to \mathbb{R}$ be a bounded measurable function such that for almost every $\xi \in (0,1)$ $f(\xi,.)$ is twice continuously differentiable, bounded and with bounded derivatives, uniformly with respect to ξ. Then F is defined for every $x, y \in H = L^2(0,1)$ by

$$F(x,y)(\xi) = f(\xi, x(\xi), y(\xi)).$$

For G, we assume that there exists a function g with the same properties as f above, such that $G(x,y)(\xi) = g(\xi, x(\xi), y(\xi))$. The assumption (SD) is then satisfied when

$$\sup_{\xi\in(a,b),x\in\mathbb{R},y\in\mathbb{R}} \left|\frac{\partial g}{\partial y}(\xi,x,y)\right| < \mu.$$

The conditions in Assumption 3.3 are then satisfied for F and G as soon as there exists $\eta < 1/2$ such that $D(-A)^\eta$ and $D(-B)^\eta$ are continuously embedded into $L^\infty(0,1)$.

We then deduce that the system (2.1) is well-posed for any $\epsilon > 0$ on any finite time interval $[0,T]$. Under Assumptions 3.1, 3.3, 3.5, the nonlinearities F and G are Lipschitz continuous, and the following Proposition is classical (see [7]).

PROPOSITION 3.10. *For every $\epsilon > 0$, $T > 0$, $x \in H$, $y \in H$, system (2.1) admits a unique mild solution $(X^\epsilon, Y^\epsilon) \in (L^2(\Omega, \mathcal{C}([0,T],H)))^2$: for any $t \in [0,T]$,*

$$\begin{aligned} X^\epsilon(t) &= e^{tA}x + \int_0^t e^{(t-s)A}F(X^\epsilon(s),Y^\epsilon(s))ds, \\ Y^\epsilon(t) &= e^{\frac{t}{\epsilon}B}y + \frac{1}{\epsilon}\int_0^t e^{\frac{(t-s)}{\epsilon}B}G(X^\epsilon(s),Y^\epsilon(s))ds + \frac{1}{\sqrt{\epsilon}}\int_0^t e^{\frac{(t-s)}{\epsilon}B}dW(s). \end{aligned} \tag{3.3}$$

4. Convergence results.

4.1. Statement of the theorems. We can now state the main results: the numerical process (X_n) defined by (2.3) approximates the slow component $X^\epsilon(n\Delta t)$ of system (2.1) with strong and weak error estimates given in the theorems.

If the strict dissipativity assumption, Assumption 3.7, is satisfied, we prove the following.

THEOREM 4.1 (strong convergence). *Assume that $x, y \in H$. With* (SD)*, for any $0 < r < 1/2$, $0 < \kappa < 1/2$, $T > 0$, $\epsilon_0 > 0$, $\Delta_0 > 0$, $\tau_0 > 0$, there exists $C > 0$ such that for any $0 < \epsilon \le \epsilon_0$, $0 < \Delta t \le \Delta_0$, $\delta t > 0$ such that $\tau = \frac{\delta t}{\epsilon} \le \tau_0$ and $1 < n \le n_0 = \lfloor \frac{T}{\Delta t} \rfloor$*

$$\begin{aligned} \mathbb{E}|X^\epsilon(n\Delta t) - X_n| \le{}& C\left(\epsilon^{1/2-r} + \frac{1}{n} + \Delta t^{1-r}\right) \\ &+ C\left((\frac{\delta t}{\epsilon})^{1/2-\kappa} + \frac{1}{\sqrt{N\frac{\delta t}{\epsilon}+1}}e^{-cn_T\frac{\delta t}{\epsilon}}\right) \\ &+ C\frac{\sqrt{\Delta t}}{\sqrt{M(N\frac{\delta t}{\epsilon}+1)}}. \end{aligned} \tag{4.1}$$

Under the more general weak dissipativity assumption, Assumption 3.8, we prove the following:

THEOREM 4.2 (weak convergence). *Assume that $x \in D((-A)^\theta)$ and $y \in H$ for some $\theta \in]0,1]$. Let $\Phi : H \to \mathbb{R}$ bounded, of class $\mathcal{C}^2$, with bounded first and second order derivatives. Then with* (WD)*, for any $0 < r < 1$, $\kappa < 1/2$, $T > 0$, $\epsilon_0 > 0$, $\Delta_0 > 0$, $\tau_0 > 0$, there exists $C > 0$ such that for any $0 < \epsilon \le \epsilon_0$, $0 < \Delta t \le \Delta_0$, $\delta t > 0$*

such that $\tau = \frac{\delta t}{\epsilon} \leq \tau_0$ *and* $1 < n \leq n_0 = \lfloor \frac{T}{\Delta t} \rfloor$

$$
\begin{aligned}
&|\mathbb{E}\Phi(X^\epsilon(n\Delta t)) - \mathbb{E}\Phi(X_n)| \\
&\quad \leq C\left(\epsilon^{1-r} + \frac{1}{n} + \Delta t^{1-r}\right) \\
&\quad + C\left(\left(\frac{\delta t}{\epsilon}\right)^{1/2-\kappa}\left(1 + \frac{1}{((n_T-1)\frac{\delta t}{\epsilon})^{1/2-\kappa}}\right) + \frac{1}{N\frac{\delta t}{\epsilon}+1}e^{-cn_T\frac{\delta t}{\epsilon}}\right).
\end{aligned} \tag{4.2}
$$

Remark 4.3. If we indeed assume (SD) in Theorem 4.2, the factor $(1 + \frac{1}{((n_T-1)\frac{\delta t}{\epsilon})^{1/2-\kappa}})$ can be replaced by 1. We also notice that this factor is absent in [14], where only strictly dissipative situations are considered.

If we look at the estimates of Theorems 4.1 and 4.2 at time $n = n_0$, the factor $\frac{1}{n}$ is of size $\Delta t = \mathrm{O}(\Delta t^{1-r})$.

The parameters r and κ are positive but can be chosen as small as necessary.

We remark that in both theorems we require no condition on the fast initial condition y, while some regularity of the slow initial condition x is required in the analysis of the weak error: we assume $x \in D((-A)^\theta)$ with $\theta > 0$. The reason lies in the use of the estimate of the averaging error in the weak sense, given by Theorem 1.2 in [3].

The special structure in the error bounds is a consequence of the heterogeneity in the treatment of the components in the numerical scheme.

4.2. Some comments on the convergence results.

4.2.1. Interpretation of the error terms. The proofs rely on the following decomposition, which explains the origin of the different error terms:

$$
X^\epsilon(n\Delta t) - X_n = X^\epsilon(n\Delta t) - \overline{X}(n\Delta t) + \overline{X}(n\Delta t) - \overline{X}_n + \overline{X}_n - X_n. \tag{4.3}
$$

The numerical process $(\overline{X}_n)$ is defined in (5.11) below: it is the solution of the macrosolver with a known $\overline{F}$, while (X_n) is solution of the macrosolver using $\tilde{F}_n$. The continuous processes $\overline{X}$ and X^ϵ are defined in (2.1) and (2.2).

The first term is bounded thanks to the averaging result, using strong and weak order of convergence results obtained in [3]. The second term is the error in a deterministic numerical scheme, for which convergence results are classical; we recall the estimate in Proposition 5.13. The third term is the difference between the two numerical approximations, and the main task is the control of this part: we show that an extension of the averaging effect holds at the discrete time level, where $\overline{X}_n$ plays the role of an averaged process for X_n.

When we look at the Theorems 4.1 and 4.2, we first remark that we obtain the same kind of bounds as in the finite dimensional case of [14]. However we notice some differences; they are due both to the infinite dimensional setting and to different proofs.

First, the weak order of the Euler scheme for SPDEs is only 1/2—see [9]—while it is 1 for SDEs; we remark that in (4.1) and (4.2) only the weak order of the microsolver appears, and this is one of the main theoretical advantages of the method. For completeness, we recall that the strong order is 1/4 in the SPDE case—see [19]—and 1/2 for SDEs. In fact, at least in the strictly dissipative case, we are comparing the invariant measures of the continuous fast equation with the invariant measure of its numerical approximation—see Theorem 5.11 and Corollary 5.12, obtained from [2].

In the weakly dissipative case, we use the weak error estimates where the constants do not depend on the final time.

The proofs of Theorems 4.1 and 4.2 are inspired by the ones in [14] but are different. The strong error is analyzed in a global way, as in the proof of the strong order theorem in [3]. Moreover we do not need a counterpart of Lemma 2.6 in [14]—which gives an estimate of $\tilde{F}_n - \overline{F}(X_n)$—and we thus think that our method is more natural. For the control of the weak error, we also introduce a new appropriate auxiliary function to control everything in the weak sense: as a consequence we observe that the number of independent realizations of the microsolver does not appear in (4.2).

However, we also present some simplified proofs, and a comparison of the results reveals the absence of a factor denoted by R in [14]: the difference is due to the way we use the initialization of the microsolver at each macrotime step.

The only effect we use of the definition (2.6) is the control of the moments of the fast numerical component $Y_{n,m,j}$, uniformly with respect to n, m, j; see Lemma 5.8. Precisely, two constraints are imposed on the variables $Y_{n,0,j}$: we require that for $n \geq 1$ the variable $Y_{n,0,j}$ is measurable with respect to the σ-field $\mathcal{G}_n$ defined below in (6.5) and that the estimate of Lemma 5.8 is satisfied.

4.2.2. Efficiency of HMM. We now show how the HMM used here is a more efficient direct method when the parameter ϵ goes to 0.

The fundamental remark is that ϵ only appears in the expressions with $\tau = \frac{\delta t}{\epsilon}$ and that we never have $\frac{\Delta t}{\epsilon}$ in the scheme and in the convergence estimate. The parameter τ can therefore be interpreted as the effective time step for the fast numerical scheme defined with (2.4). The reason for this absence of ϵ relies in the construction of the scheme: instead of approximating X^ϵ, we approximate $\overline{X}$ thanks to the averaging principle. As explained before, the microsolver is only used to get an approximation of the averaged coefficient $\overline{F}$; the evolution in the macrosolver uses the time step Δt and is linked with ϵ only through the definition of $\tilde{F}_n$ with only τ.

What is then important is to check that if Δt and τ are chosen small enough and M, N, and n_T are large enough, each error term due to the numerical scheme in the Theorems 4.1 and 4.2 goes to 0.

The choices must be uniform with respect to ϵ. The error term due to the application of the averaging principle is treated apart, since it only depends on ϵ.

To see this, we first choose Δt and τ.

Then we take n_T such that $n_T\tau \to +\infty$; we observe that the corresponding error terms go to 0 exponentially fast, as a consequence of the dissipation in the evolution. It is better to choose n_T before N and M, since the exponential convergence is faster than the other convergences.

Apparently, the parameters M and N play no role: indeed in (4.1) the orders of convergence with respect to Δt in the terms

$$\Delta t \quad \text{and} \quad \frac{\sqrt{\Delta t}}{\sqrt{M(N\tau+1)}}$$

are not the same: a better choice is to choose M and N such that $\frac{1}{MN\tau}$ and Δt go to 0 at the same speed. Moreover, choosing Δt after M, N, and n_T also changes the analysis.

The question of an optimal joint choice of the parameters is related to the analysis of the cost of the scheme. Even if the multiscale method seems more efficient than a

direct one when ϵ goes to 0, we have to check that for a fixed value of the parameter ϵ the approximation of the slow component at a fixed time T by our method does not require too much computational time. Following the approach in [14], we consider as a unit of time the computation of one realization of a step of the microsolver, i.e., the computation of $Y_{n,m+1,j}$ knowing $Y_{n,m,j}$. We consider that the computation of $\tilde{F}_n$ in (2.5) requires a negligible amount of time and then that one step of the macrosolver also requires 1 unit of time.

For each step of the macrosolver, $m_0.M$ steps of the microsolver are necessary; then we require $\frac{T}{\Delta t}$ macrosteps, so that we define the cost with $T = 1$:

$$\text{cost(HMM)} = \frac{m_0.M}{\Delta t};$$

for a direct scheme with time step δt, the corresponding cost is defined by

$$\text{cost(direct)} = \frac{1}{\delta t} = \frac{1}{\epsilon}\frac{1}{\delta t/\epsilon}.$$

Working as in [14], we introduce a tolerance in the discretization error *tol*, and we choose the parameters in order to bound the error with

$$C\epsilon^{\Lambda} + tol$$

with $\Lambda = 1/2 - r$ for the strong estimate (4.1) and $\Lambda = 1 - r$ for the weak estimate (4.2) with $r > 0$ chosen as small as necessary. In other words, we only focus on the difference between $\overline{X}(n\Delta t)$ and X_n, and consider that the averaging error is negligible. The costs are then functions of *tol*.

The computation of parameters satisfying the above condition can be done like in [14] so we do not develop it here. The result is that for an appropriate choice of the parameters, for a fixed value of *tol*, we have when $\epsilon \to 0$

$$\frac{\text{cost(HMM)}}{\text{cost(direct)}} \to 0.$$

Finally, we have not treated here the discretization in space, since the main issue is the presence of ϵ in the evolution in time of the system of SPDEs. To complete the analysis, we need to take into account the cost of the discretization in space of one step of the microsolver with an additional factor; then one realization of the microsolver would not require one unit of time but a number of units depending on the size of the discretization. The same remark also holds for a direct method, and we see that the comparison of the costs of time discretization is sufficient to prove the better efficiency of the HMM.

5. Preliminary results.

5.1. Known results about the fast equation and the averaged equation.

In this section, we just recall without proof the main results on the fast equation with frozen slow component and on the averaged equation, defined below. Proofs are found in [5] for the strict dissipative case, and the extension to the weakly dissipative situation relies on arguments explained below.

If $x \in H$, we define an equation on the fast variable where the slow variable is fixed and equal to x:

$$\begin{aligned} dY_x(t,y) &= (BY_x(t,y) + G(x, Y_x(t,y)))dt + dW(t), \\ Y_x(0,y) &= y. \end{aligned} \tag{5.1}$$

This equation admits a unique mild solution, defined on $[0, +\infty[$.

Since Y^ϵ is involved at time $t > 0$, heuristically we need to analyze the properties of $Y_x(\frac{t}{\epsilon}, y)$ with $\epsilon \to 0$, and by a change of time we need to understand the asymptotic behavior of $Y_x(., y)$ when time goes to infinity.

Under the strict dissipativity assumption, Assumption (3.7), we obtain a contractivity of trajectories issued from different initial conditions and driven by the same noise.

PROPOSITION 5.1. *With* (SD)*, for any $t \geq 0$, $x, y_1, y_2 \in H$ we have*

$$|Y_x(t, y_1) - Y_x(t, y_2)|_H \leq e^{-\frac{(\mu - L_g)}{2}t} |y_1 - y_2|_H.$$

Under the weak dissipativity assumption, Assumption 3.8, we obtain such an exponential convergence result for the laws instead of trajectories. This result is based on a coupling argument; a detailed proof inspired from [15] and [17] is written in [10]—see also [2] for further references.

PROPOSITION 5.2. *With* (WD)*, there exist $c > 0$, $C > 0$ such that for any bounded test function ϕ, any $t \geq 0$, and any $y_1, y_2 \in H$*

$$|\mathbb{E}\phi(Y_x(t, y_1)) - \mathbb{E}\phi(Y_x(t, y_2))| \leq C\|\phi\|_\infty (1 + |y_1|^2 + |y_2|^2) e^{-ct}. \tag{5.2}$$

The proofs of Propositions 5.1 and 5.2 are only based on the respective dissipativity assumption; the gradient structure of the equation has not been used yet.

As a consequence, there exists a unique invariant probability measure μ^x associated with Y_x, and the convergence to equilibrium is exponentially fast.

PROPOSITION 5.3. *If we assume* (SD) *or* (WD)*, the fast process Y_x with frozen slow component x admits a unique invariant probability measure μ^x, and there exist constants $C, c > 0$ such that for any bounded function $\phi : H \to \mathbb{R}$ or $\phi : H \to H$, $t \geq 0$ and $x, y \in H$, we have*

$$\left| \mathbb{E}\phi(Y_x(t, y)) - \int_H \phi(z) \mu^x(dz) \right| \leq C\|\phi\|_\infty (1 + |y|_H^2) e^{-ct}.$$

This result is not sufficient to obtain the orders of convergence in the averaging principle of [3] and to prove Theorems 4.1 and 4.2. Thanks to the existence of a potential U such that $G = D_y U$, we are moreover able to give an explicit formula (5.3) for μ^x.

Let $\nu = \mathcal{N}(0, (-B)^{-1}/2)$ be the centered Gaussian probability measure on H with the covariance operator $(-B)^{-1}/2$—which is positive and trace-class, thanks to Assumption 3.1. Then μ^x satisfies

$$\mu^x(dy) = \frac{1}{Z(x)} e^{2U(x,y)} \nu(dy), \tag{5.3}$$

where $Z(x) \in]0, +\infty[$ is a normalization constant.

Now we define the averaged equation. First we define the averaged nonlinear coefficient $\overline{F}$.

DEFINITION 5.4. *For any $x \in H$,*

$$\overline{F}(x) = \int_H F(x, y) \mu^x(dy). \tag{5.4}$$

Using Assumptions 3.3 and 3.5 and the expression of μ^x, it is easy to prove that $\overline{F}$ is bounded and Lipschitz continuous.

Under Assumptions 3.3 and 3.5 and thanks to the expression (5.3) of μ^x, we have the following properties on $\overline{F}$.

PROPOSITION 5.5. *There exists $0 \leq \eta < 1$ and a constant C such that the following directional derivatives of $\overline{F}$ are well defined and controlled:*

- *For any $x \in H$, $h \in H$, $|D\overline{F}(x).h| \leq C|h|_H$.*
- *For any $x \in H$, $h \in H$, $k \in D(-A)^\eta$, $|D^2\overline{F}(x).(h,k)| \leq C|h|_H|k|_{(-A)^\eta}$.*
- *For any $x \in H$, $h, k \in H$, $|(-A)^{-\eta}D^2\overline{F}(x).(h,k)| \leq C|h|_H|k|_H$.*

Moreover, $\overline{F}$ is bounded and Lipschitz continuous.

The last estimate above is a consequence of Assumptions 3.4 and 3.6 and of the following fact: we have almost surely $W^B(t) \in L^\infty(0,1)$ for any $t \geq 0$, and

$$\int_H |z|_{L^\infty(0,1)}\nu(dz) < +\infty. \tag{5.5}$$

We notice that if $\eta > 1/4$—which is the right condition in the case of linear Laplace operators and of nonlinear Nemytskii operators—we have $\int_H |z|_{(-A)^\eta}\nu(dz) = +\infty$. We thus need the restrictive condition in Assumption 3.6.

Remark 5.6. Even when F and G are Nemytskii operators, $\overline{F}$ is not such an operator in general.

Then the averaged equation—see (2.2) in the introduction—can be defined:

$$\frac{d\overline{X}(t)}{dt} = A\overline{X}(t) + \overline{F}(\overline{X}(t))$$

with initial condition $\overline{X}(0) = x \in H$. For any $T > 0$, this deterministic equation admits a unique mild solution $\overline{X} \in \mathcal{C}([0,T],H)$.

5.2. Estimates on the numerical solutions. We give uniform estimates on X_n and $Y_{n,m,j}$, defined by 2.3 and 2.4.

LEMMA 5.7. *There exists $C > 0$ such that we have $\mathbb{P}$-almost surely*

$$|X_n| \leq C(1 + |x|)$$

for any $0 \leq n \leq n_0$.

Proof. The linear operator $S_{\Delta t}$ satifies $|S_{\Delta t}|_{\mathcal{L}(H)} \leq \frac{1}{1+\lambda\Delta t}$; moreover F is bounded, so that by (2.5) we almost surely have for any $n \geq 0$ $|\tilde{F}_n| \leq \|F\|_\infty$. The end of the proof is then straightforward; we also notice that C does not depend on the final time T. ∎

LEMMA 5.8. *There exists $C > 0$—which does not depend on $T > 0$, on N, on n_T, or on M—such that for any $\Delta t > 0$, $\tau = \frac{\delta t}{\epsilon} > 0$, $0 \leq n \leq n_0$, $0 \leq m \leq m_0$, and $1 \leq j \leq M$, we have*

$$\mathbb{E}|Y_{n,m,j}|^2 \leq C(|y|^2 + 1).$$

Proof. We introduce $\omega_{n,m,j}$ defined by the fast numerical scheme with no nonlinear coefficient—see (2.4)—with the notation $\tau = \frac{\delta t}{\epsilon}$: for any $0 \leq j \leq M$, $0 \leq n \leq n_0$, and $0 \leq m \leq m_0$

$$w_{n,m+1,j} = R_\tau w_{n,m,j} + \sqrt{\tau}R_\tau\zeta_{n,m+1,j} \tag{5.6}$$

with the initial condition $w_{n,0,j} = 0$. It is a classical result that for any $\tau_0 > 0$, there exists a constant $C(\tau_0)$ such that for any $0 \leq \tau < \tau_0$, $0 \leq n \leq n_0$, $1 \leq j \leq M$, and $0 \leq m \leq m_0$ we have

$$\mathbb{E}|\omega_{n,m,j}|^2 \leq C(\tau_0).$$

Now for any $0 \leq m \leq m_0$ we define $D_{n,m,j} = Y_{n,m,j} - w_{n,m,j}$; it is enough to control $|D_{n,m,j}|^2$, $\mathbb{P}$-almost surely. By (2.4), we have the following expression: for any $0 \leq m \leq m_0$

$$D_{n,m+1,j} = R_\tau D_{n,m,j} + \tau R_\tau G(X_n, Y_{n,m,j}).$$

Since G is bounded, and using the inequality $|R_\tau|_{\mathcal{L}(H)} \leq \frac{1}{1+\mu\tau}$, we get

$$|D_{n,m+1,j}| \leq \frac{1}{1+\mu\tau}|D_{n,m,j}| + C\tau.$$

Therefore we have for any $0 \leq m \leq m_0$

$$(1+\mu\tau)^m |D_{n,m,j}| \leq |D_{n,0,j}| + C[(1+\mu\tau)^m - 1].$$

But $D_{n,0,j} = D_{n-1,m_0,j}$. So we get

$$(1+\mu\tau)^{m_0} |D_{n+1,0,j}| \leq |D_{n,0,j}| + C[(1+\mu\tau)^{m_0} - 1].$$

Therefore

$$(1+\mu\tau)^{nm_0} |Z_{n,0,j}| \leq |D_{0,0,j}| + C(1+\mu\tau)^{nm_0} = |y| + C(1+\mu\tau)^{nm_0}.$$

As a consequence, we get for any $0 \leq n \leq n_0$ and any $0 \leq m \leq m_0$, $|D_{n,0,j}| \leq C(1+|y|)$, and then $|Z_{n,m,j}| \leq 2C(1+|y|)$. □

We remark that Lemma 5.8 is a consequence of the choice of the initialization of the microsolver at each macrotime step (2.6) through the equality $D_{n,0,j} = D_{n-1,m_0,j}$ appearing at the end of the proof. However, as mentioned earlier other choices for $Y_{n,0,j}$ could lead to the same kind of estimate.

5.3. Asymptotic behavior of the "fast" numerical scheme. At the continuous time level, the averaging principle proved in [3] comes from the asymptotic behavior of the fast equation with frozen slow component (5.1), as described in section 5.1. The underlying idea of the HMM method in our setting is to prove a similar averaging effect at the discrete time level: we therefore study the asymptotic behavior of the fast numerical scheme which defines the microsolver with frozen slow component—in other words we are looking at the evolution of the microsolver during one fixed macrotime step.

In section 5.1, we have seen that under the weak dissipativity assumption, Assumption 3.8, the fast equation with frozen slow component admits a unique invariant probability measure μ^x. At the discrete time level, this assumption only yields the existence of invariant laws; to get a unique invariant law $\mu^{x,\tau}$, we need (SD), and we obtain the following.

THEOREM 5.9. *Under Assumption* 3.7, *for any* $\tau > 0$ *and any* $x \in H$, *the numerical scheme* 5.8 *admits a unique ergodic invariant probability measure* $\mu^{x,\tau}$. *Moreover, we have convergence to equilibrium in the following sense: for any* $\tau_0 > 0$, *there exist* $c > 0$ *and* $C > 0$ *such that for any* $0 < \tau \leq \tau_0$, $x \in H$, $y \in H$, *any Lipschitz continuous function* ϕ *from* H *to* R, *and* $m \geq 0$, *we have*

$$\left| \mathbb{E}(\phi(Y_m^x(y))) - \int_H \phi(z)\mu^{x,\tau}(dz) \right| \leq C(1+|y|)[\phi]_{Lip} e^{-cm\tau}. \tag{5.7}$$

We recall the notation $\tau = \frac{\delta t}{\epsilon}$ for the effective time step; the noise is defined with a cylindrical Wiener process $\tilde{W}$: $\tilde{\zeta}_{m+1} = \frac{\tilde{W}_{(m+1)\tau} - \tilde{W}_{m\tau}}{\sqrt{\tau}}$. If we fix the slow component $x \in H$, we define

$$Y^x_{m+1}(y) = R_\tau Y^x_m(y) + \tau R_\tau G(x, Y^x_m(y)) + \sqrt{\tau} R_\tau \tilde{\zeta}_{m+1} \tag{5.8}$$

with the initial condition $Y^x_0(y) = y$.

The proof of Theorem 5.9 is divided into two steps. First in section 5.3.1 we focus on the existence, which is obtained under (WD) thanks to the Krylov–Bogoliubov criterion—see [8]. Second, in section 5.3.2 we show that uniqueness holds if (SD) is satisfied.

In the case when (WD) is satisfied while (SD) is not true, it seems that there is no general uniqueness result at the discrete time level.

5.3.1. Existence of an invariant law. With (5.8), we associate the transition semigroup $(P^{x,\tau}_m)$: if ϕ is a bounded measurable function from H to $\mathbb{R}$, $y \in H$, and $m \geq 0$,

$$P^{x,\tau}_m \phi(y) = \mathbb{E}[\phi(Y^x_m(y))]. \tag{5.9}$$

We also denote by $\nu^{x,\tau}_{m,y}$ the law of $Y^{x,\tau}_m(y)$; then

$$P^{x,\tau}_m \phi(y) = \mathbb{E}[\phi(Y^{x,\tau}_m(y))] = \int_H \phi(z) \nu^{x,\tau}_{m,y}(dz).$$

We notice that the semigroup (P_m) satisfies the Feller property: if ϕ is bounded and continuous, then $P_m \phi \in$ is bounded and continuous.

The required tightness property for the use of the Krylov–Bogoliubov criterion is a consequence of the following estimate, which is proved thanks to regularization properties of the semigroup $(R^m_\tau)_m$: for any $0 < \gamma < 1/4$, $\tau > 0$, there exists $C(\gamma, \tau) > 0$ such that for any $m \geq 1$ and $\tau \leq \tau_0$

$$\mathbb{E}|Y_m(y)|^2_{(-B)^\gamma} \leq C(\gamma, \tau).$$

Moreover if $0 < \gamma < 1/4$, the embedding of $D(-B)^\gamma$ in H is compact.

We then see that for any $y \in H$ the family of probability measures $(\frac{1}{m} \sum_{k=1}^m \nu^{x,\tau}_{k,y})$ is tight.

5.3.2. Uniqueness under strict dissipativity. The key estimate to prove uniqueness is the following contractivity property, which holds thanks to Assumption 3.7.

PROPOSITION 5.10. *For any $\tau_0 > 0$, there exists $c > 0$ such that for any $0 < \tau \leq \tau_0$, $m \geq 0$, $y_1, y_2 \in H$, $x \in H$ we have $\mathbb{P}$-almost surely*

$$|Y^x_m(y_1) - Y^x_m(y_2)| \leq e^{-cm\tau} |y_1 - y_2|. \tag{5.10}$$

Proof. If we define $r_m = Y_m(y_1) - Y_m(y_2)$, then we have the equation

$$\begin{aligned} r_{m+1} &= r_m + \tau B R_{m+1} + \tau (G(x, Y_m(y_1)) - G(x, Y_m(y_2))), \\ r_0 &= y_1 - y_2. \end{aligned}$$

If we take the scalar product in H of this equation with r_{m+1}, we get

$$\begin{aligned} |r_{m+1}|^2 - \langle r_m, r_{m+1} \rangle &= \tau \langle B r_{m+1}, r_{m+1} \rangle \\ &\quad + \tau \langle G(x, Y_m(y_1)) - G(x, Y_m(y_2)), r_{m+1} \rangle. \end{aligned}$$

The left-hand side is equal to $\frac{1}{2}(|r_{m+1}|^2-|r_m|^2)+\frac{1}{2}|r_{m+1}-r_m|^2$, and we get

$$\begin{aligned}\frac{1}{2}(|r_{m+1}|^2-|r_m|^2) &\leq -\tau|(-B)^{1/2}r_{m+1}|^2+\tau L_g|r_m||r_{m+1}|\\ &\leq -\mu\tau|r_{m+1}|^2+\frac{1}{2}\tau L_g(|r_{m+1}|^2+|r_m|^2).\end{aligned}$$

Therefore we have $(1+\tau(2\mu-L_g))|r_{m+1}|^2\leq(1+\tau L_g)|r_m|^2$. We remark that (SD) implies that for any $\tau_0>0$, there exists $c>0$ such that if $\tau\leq\tau_0$ we have $\rho=\frac{1+\tau L_g}{1+\tau(2\mu-L_g)}\leq e^{-2c\tau}$; therefore

$$|r_m|^2\leq\rho^m|y_1-y_2|^2\leq e^{-2cm\tau}|y_1-y_2|^2. \qquad \square$$

As a consequence, there exists a unique ergodic invariant probability measure $\mu^{x,\tau}$, which is strongly mixing. Moreover since there exists $C>0$ such that for any $\tau>0$ and any $x\in H$ $\int_H|y|\mu^{x,\tau}(dy)\leq C$, we get

$$\begin{aligned}|\mathbb{E}(\phi(Y_m^x(y)))-\int_H\phi(z)\mu^{x,\tau}(dz)| &= |\mathbb{E}(\phi(Y_m^x(y)))-\int_H\mathbb{E}\phi(Y_m^x(z))\mu^{x,\tau}(dz)|\\ &\leq\int_H\mathbb{E}|\phi(Y_m^x(y))-\phi(Y_m^x(z)|\mu^{x,\tau}(dz)\\ &\leq[\phi]_{\mathrm{Lip}}\int_H e^{-cm\tau}|y-z|\mu^{x,\tau}(dz)\\ &\leq C(1+|y|)[\phi]_{\mathrm{Lip}}e^{-cm\tau}.\end{aligned}$$

5.3.3. Approximation of the invariant law μ^x by the fast numerical scheme. We recall that μ^x denotes the invariant law of the continuous time fast equation with frozen slow component (5.1). Thanks to the fast numerical scheme, we have an approximation result, which is proved in [2]: with test functions of class $\mathcal{C}_b^2$, we control the weak error for any time with a convergence of order $1/2$ with respect to the time step τ. Moreover the estimate is easily seen to be independent from the slow component x.

We define for any Φ of class $\mathcal{C}_b^2$

$$\|\Phi\|_{(2)}=\sup_{y\in H}|\Phi(y)|+\sup_{y\in H,h\in H,|h|=1}|D_y\Phi(y).h|+\sup_{y\in H,h,k\in H,|h|=|k|=1}|D_{yy}^2\Phi(y).(h,k)|.$$

For the convenience of the reader, we recall the following result from [2].

THEOREM 5.11. *With* (WD)*, for any $0<\kappa<1/2$, any $\tau_0>0$, there exists $C,c>0$ such that for any Φ of class $\mathcal{C}_b^2$, any $x,y\in H$, any $0<\tau\leq\tau_0$, and any integer $2\leq m<+\infty$*

$$|\mathbb{E}[\Phi(Y_x(m\tau,y))]-\mathbb{E}[\Phi(Y_m^x(y))]|\leq C\|\Phi\|_{(2)}(1+|y|^3)(((m-1)\tau)^{-1/2+\kappa}+1)\tau^{1/2-\kappa}.$$

We remark that this theorem is proved without requiring the gradient structure of the fast equation with frozen slow component.

As explained in section 5.3, the existence of invariant probability measures for the numerical scheme is true with only (WD), while uniqueness is a priori only satisfied when the strict dissipativity assumption, Assumption 3.7, holds; the unique invariant law is then denoted by $\mu^{x,\tau}$.

COROLLARY 5.12. *Under the assumptions of Theorem* 5.11
(i) *we have for any* $m \geq 2$

$$\left| \int_H \Phi d\mu^x - \mathbb{E}[\Phi(Y_m^x)] \right| \leq C\|\Phi\|_{(2)}(1+|y|^3)(((m-1)\tau)^{-1/2+\kappa}+1)\tau^{1/2-\kappa} + CN(\Phi)(1+|y|^2)e^{-cm\tau},$$

(ii) *if moreover* (SD) *is satisfied,*

$$\left| \int_H \Phi d\mu^x - \int_H \Phi d\mu^{x,\tau} \right| \leq C\|\Phi\|_{(2)}\tau^{1/2-\kappa}.$$

The result (i) is sufficient for the proof of Theorem 4.2, while the result (ii) is necessary to obtain the strong convergence in Theorem 4.1.

We recall that in the case of Euler scheme for SDEs this kind of results holds with the order of convergence 1.

5.4. Error in the deterministic scheme (5.11). We define a scheme based on the macrosolver, for theoretical purposes, in the situation when $\overline{F}$ is known:

$$\begin{aligned} \overline{X}_{n+1} &= S_{\Delta t}\overline{X}_n + \Delta t S_{\Delta t}\overline{F}(\overline{X}_n), \\ \overline{X}_0 &= x. \end{aligned} \tag{5.11}$$

We look at the error between $\overline{X}_n$, defined by (5.11), and $\overline{X}(n\Delta t)$, defined by (2.2). Here quantities are deterministic, and the following result is classical—see [16], [6], or the details of the proofs in [19].

PROPOSITION 5.13. *For any* $0 < r < 1$, $\Delta t_0 > 0$, *and* $T > 0$, *there exists* $C > 0$, *such that for any* $0 < \Delta t \leq \Delta t_0$ *and* $1 \leq n \leq \lfloor \frac{T}{\Delta t} \rfloor$

$$|\overline{X}_n - \overline{X}(n\Delta t)| \leq \frac{C}{n} + C(1+|x|)\Delta t^{1-r}.$$

6. Proof of the strong convergence Theorem 4.1. The final time T is fixed and we recall the notation $n_0 = \lfloor \frac{T}{\Delta t} \rfloor$.

To simplify notation, we do not always make precise the range of summation in the expressions below: the indices j, j_1, j_2 belong to $\{1, \ldots, M\}$, and m, m_1, m_2 belong to $\{n_T, \ldots, n_T + N - 1 = m_0\}$.

We recall that according to the decomposition of the error (4.3), we have to control

$$\begin{aligned} \mathbb{E}|X^\epsilon(n\Delta t) - X_n| \leq{}& \mathbb{E}|X^\epsilon(n\Delta t) - \overline{X}(n\Delta t)| \\ &+ |\overline{X}(n\Delta t) - \overline{X}_n| \\ &+ \mathbb{E}|\overline{X}_n - X_n|. \end{aligned} \tag{6.1}$$

The first part is controlled thanks to the strong order theorem of [3]: for any $0 < r < 1/2$, we have for any $0 \leq n \leq n_0$

$$\mathbb{E}|X^\epsilon_{n\Delta t} - \overline{X}_{n\Delta t}| \leq C\epsilon^{1/2-r}.$$

The second part is deterministic and is controlled thanks to Proposition 5.13:

$$|\overline{X}_n - \overline{X}(n\Delta t)| \leq \frac{C}{n} + C(1+|x|)\Delta t^{1-r},$$

where C depends on T, r, x, y.

It remains to focus on the third part $e_n = \overline{X}_n - X_n$. Instead of analyzing the local error like in [14], we adopt a global point of view, and we follow the idea of the proof of Theorem 1.1 in [3]: for any $0 \leq n \leq n_0$

$$X_n - \overline{X}_n = S_{\Delta t}^n x + \Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k} \tilde{F}_k - S_{\Delta t}^n x - \Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k} \overline{F}(\overline{X}_k). \tag{6.2}$$

The averaged coefficient $\overline{F}$ is Lipschitz continuous, and $|S_{\Delta t}|_{\mathcal{L}(H)} \leq 1$; moreover we define the averaged coefficient $\overline{F}^\tau$ with respect to the invariant measure $\mu^{x,\tau}$ of the fast numerical scheme, which is unique since we assume strict dissipativity (SD): for any $x \in H$

$$\overline{F}^\tau(x) = \int_H F(x,y)\mu^{x,\tau}(dy). \tag{6.3}$$

The error in (6.2) is then decomposed in the following way—the idea of looking at the square of the norm in the second expression is an essential tool of the proof:

$$\begin{aligned} \mathbb{E}|X_n - \overline{X}_n| \leq C\Delta t \sum_{k=0}^{n-1} \mathbb{E}|X_k - \overline{X}_k| \\ + \left(\mathbb{E}|\Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k} \tilde{F}_k - S_{\Delta t}^{n-k} \overline{F}^\tau(X_k)|^2 \right)^{1/2} \\ + \Delta t \sum_{k=0}^{n-1} \mathbb{E}|S_{\Delta t}^{n-k} \overline{F}(X_k) - S_{\Delta t}^{n-k} \overline{F}^\tau(X_k)|. \end{aligned} \tag{6.4}$$

If we can control the two last terms by a certain quantity Q, by a discrete Gronwall Lemma we get for any $0 \leq k \leq n_0$ $\mathbb{E}|X_n - \overline{X}_n| \leq e^{CT} Q$.

First, the third term in (6.4) is linked to the distance between the invariant measures μ^x and $\mu^{x,\tau}$—since we assume strict dissipativity for this strong estimate—which is evaluated thanks to Theorem 5.11 and Corollary 5.12 for test functions of class $\mathcal{C}_b^2$. Since by regularization properties of the semigroups we have $|(-A)^\eta S_{\Delta t}^{n-k}|_{\mathcal{L}(H)} \leq \frac{C}{((n-k)\Delta t)^\eta}$, we apply Corollary 5.12 with the regular test function $S_{\Delta t}^{n-k}\overline{F}$, which thanks to Assumption 3.4 satisfies for some constant $C > 0$ and for any $x, h, k \in H$

$$\begin{aligned} |(S_{\Delta t}^{n-k}\overline{F})(x)|_H \leq C, |D_x(S_{\Delta t}^{n-k}\overline{F})(x).h|_H \leq C|h|, \\ |D_{xx}^2(S_{\Delta t}^{n-k}\overline{F})(x).(h,k)|_H = |(-A)^\eta S_{\Delta t}^{n-k}(-A)^{-\eta} D_{xx}^2(S_{\Delta t}^{n-k}\overline{F})(x).(h,k)| \\ \leq \frac{C}{((n-k)\Delta t)^\eta}|h||k|. \end{aligned}$$

We thus obtain

$$|S_{\Delta t}^{n-k}\overline{F}(X_k) - S_{\Delta t}^{n-k}\overline{F}^\tau(X_k)| \leq \frac{C}{((n-k)\Delta t)^\eta}\tau^{1/2-\kappa},$$

and summing we get

$$\Delta t \sum_{k=0}^{n-1} \mathbb{E}|S_{\Delta t}^{n-k}\overline{F}(X_k) - S_{\Delta t}^{n-k}\overline{F}^\tau(X_k)| \leq C\tau^{1/2-\kappa}.$$

The control of the other term is more complicated; in order to get a precise estimate, we expand the square of the norm of the sum. We then use some conditional expectations, which allow to use exponential convergence to equilibrium via Theorem 5.9. Therefore we obtain the following expansion and we treat separately each term:

$$\begin{aligned}
&\mathbb{E}\left|\Delta t\sum_{k=0}^{n-1} S_{\Delta t}^{n-k}\tilde{F}_k - S_{\Delta t}^{n-k}\overline{F}^{\tau}(X_k)\right|^2 \\
&\quad = \Delta t^2\sum_{k=0}^{n-1}\mathbb{E}|S_{\Delta t}^{n-k}(\tilde{F}_k - \overline{F}^{\tau}(X_k))|^2 \\
&\qquad + 2\Delta t^2\sum_{0\le k_1<k_2\le n-1}\mathbb{E}\langle S_{\Delta t}^{n-k_1}(\tilde{F}_{k_1} - \overline{F}^{\tau}(X_{k_1})), S_{\Delta t}^{n-k_2}(\tilde{F}_{k_2} - \overline{F}^{\tau}(X_{k_2}))\rangle \\
&\quad =: \Sigma_1 + \Sigma_2.
\end{aligned}$$

(i) We first treat Σ_1.

We introduce the following notation: $\mathbb{E}_n$ denotes conditional expectation with respect to the σ-field

$$\mathcal{G}_n = \sigma(\zeta_{k,m,j}, 0 \le k \le n-1, 1 \le m \le m_0, 1 \le j \le M). \tag{6.5}$$

We notice that X_n is $\mathcal{G}_n$-measurable but that $\tilde{F}_n$ is not.

From (2.5), for any $0 \le k \le n-1$ we have

$$\tilde{F}_k = \frac{1}{MN}\sum_{j=1}^{M}\sum_{m=n_T}^{n_T+N-1} F(X_k, Y_{k,m,j});$$

therefore we see that

$$\begin{aligned}
&\mathbb{E}|S_{\Delta t}^{n-k}(\tilde{F}_k - \overline{F}^{\tau}(X_k))|^2 \\
&\quad = \frac{1}{M^2N^2}\sum_{j_1,j_2}\sum_{m_1,m_2}\mathbb{E}\mathbb{E}_k\langle S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_1,j_1}) - \overline{F}^{\tau}(X_k)), \\
&\qquad\qquad S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_2,j_2}) - \overline{F}^{\tau}(X_k))\rangle
\end{aligned}$$

with the conditional expectation $\mathbb{E}_k$ with respect to $\mathcal{G}_k$; see (6.5).

When $j_1 \ne j_2$, $\zeta_m^{(j_1)}$ and $\zeta_m^{(j_2)}$ are independent, so that if we treat differently the cases $j_1 = j_2$ and $j_1 \ne j_2$ in the above summation we obtain

$$\begin{aligned}
&M^2N^2\mathbb{E}|S_{\Delta t}^{n-k}(\tilde{F}_k - \overline{F}^{\tau}(X_k))|^2 \\
&\quad = \sum_{j_1\ne j_2}\mathbb{E}\Big\langle\sum_{m_1}\mathbb{E}_k S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_1,j_1}) - \overline{F}^{\tau}(X_k)), \\
&\qquad\qquad \sum_{m_2}\mathbb{E}_k S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_2,j_2}) - \overline{F}^{\tau}(X_k))\Big\rangle \\
&\qquad + \sum_{j=1}^{M}\sum_{m_1,m_2}\mathbb{E}\langle S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_1,j}) - \overline{F}^{\tau}(X_k)), S_{\Delta t}^{n-k}(F(X_k, Y_{k,m_2,j}) - \overline{F}^{\tau}(X_k))\rangle.
\end{aligned}$$

For the first part, we directly use the exponential convergence to equilibrium result of (5.7), on each factor, to get a bound with

$$\left(\frac{1}{N}\sum_{m=n_T}^{m_0} e^{-cm\tau}\right)^2 \leq \left(\frac{Ce^{-cn_T\tau}}{N\tau+1}\right)^2.$$

For the second part, with no loss of generality we only treat the case $m_1 \leq m_2$, and we introduce the conditional expectation $\mathbb{E}_{k,m_1,j}$ with respect to the σ-field generated by $\mathcal{G}_k$ and $(\zeta^{(j)}_{km_0+m})_{0\leq m\leq m_1-1}$ when $m_1 \leq m_2$. The current time appearing in the exponential convergence estimate is $(m_2-m_1)\tau$, and we get a bound with

$$\frac{2}{MN^2}\sum_{n_T\leq m_1\leq m_2\leq m_0} e^{-c(m_2-m_1)\tau} \leq \frac{C}{M(N\tau+1)}.$$

We therefore get

$$\Sigma_1 \leq C\Delta t\left(\left(\frac{e^{-cn_T\tau}}{N\tau+1}\right)^2 + \frac{1}{M(N\tau+1)}\right). \tag{6.6}$$

(ii) We now consider Σ_2, which corresponds to the cross-terms in the expansion of the square of the norm of the quantity $\sum_k S^{n-k}_{\Delta t}(\tilde{F}_k - \overline{F}^\tau(X_k))$. By the definition of $\tilde{F}_k$, the general term with indices $k_1 < k_2$ in $|\Sigma_2|$ is bounded by

$$\begin{aligned}
&|\mathbb{E}\langle S^{n-k_1}_{\Delta t}(\tilde{F}_{k_1} - \overline{F}^\tau(X_{k_1})), S^{n-k_2}_{\Delta t}(\tilde{F}_{k_2} - \overline{F}^\tau(X_{k_2}))\rangle| \\
&\quad \leq \frac{\Delta t^2}{M^2N^2}\left|\sum_{m_i,j_i}\mathbb{E}\langle S^{n-k_1}_{\Delta t}(F(X_{k_1},Y_{k_1,m_1,j_1}) - \overline{F}^\tau(X_{k_1})),\right. \\
&\qquad\qquad \left. S^{n-k_2}_{\Delta t}(F(X_{k_2},Y_{k_2,m_2,j_2}) - \overline{F}^\tau(X_{k_2}))\rangle\right| \\
&\quad \leq C\frac{\Delta t^2}{MN}\sum_{m=n_T}^{m_0}\mathbb{E}|\mathbb{E}_{k_2}[S^{n-k_2}_{\Delta t}(F(X_{k_2},Y_{k_2,m,j}) - \overline{F}^\tau(X_{k_2})]|,
\end{aligned}$$

using conditional expectation $\mathbb{E}_{k_2}$ and the boundedness of F.

Using the exponential convergence result of (5.7) and Lemma 5.8, we get the bound

$$\mathbb{E}|\mathbb{E}_{k_2}[S^{n-k_2}_{\Delta t}(F(X_{k_2},Y_{k_2,m,j}) - \overline{F}^\tau(X_{k_2})]| \leq Ce^{-m\tau},$$

so that the previous quantity is bounded by

$$C\frac{\Delta t^2}{MN}\sum_{m=n_T}^{m_0}\sum_{j=1}^{M} e^{-cm\tau} \leq C\Delta t^2\frac{e^{-cn_T\tau}}{N\tau+1}.$$

Summing on $k_1 < k_2$, we can now conclude that

$$\Sigma_2 \leq C\frac{e^{-cn_T\tau}}{N\tau+1}; \tag{6.7}$$

then by (6.6) and (6.7)

$$\mathbb{E}\left|\Delta t\sum_{k=0}^{n-1} S^{n-k}_{\Delta t}\tilde{F}_k - S^{n-k}_{\Delta t}\overline{F}^\tau(X_k)\right|^2 \leq C\left(\frac{e^{-cn_T\tau}}{N\tau+1} + \Delta t\left(\frac{e^{-cn_T\tau}}{N\tau+1}\right)^2 + \frac{\Delta t}{M(N\tau+1)}\right),$$

and the result of Theorem 4.1 now follows from (6.4).

7. Proof of the weak convergence Theorem 4.2. In order to get a better bound for the weak error than for the strong error, we use an auxiliary function which is a solution of a Kolmogorov equation.

The proof below only requires the weak dissipativity assumption, Assumption 3.8.

We divide the proof in two parts: the first one contains the elements of the proof, while the second one is devoted to two technical lemmas.

7.1. Proof of the theorem. According to the decomposition (4.3), we want to control for any $0 \leq n \leq n_0$

$$\begin{aligned} |\mathbb{E}\Phi(X^\epsilon(n\Delta t)) - \mathbb{E}\Phi(X_n)| &\leq |\mathbb{E}\Phi(X^\epsilon(n\Delta t)) - \mathbb{E}\Phi(\overline{X}(n\Delta t))| \\ &\quad + |\Phi(\overline{X}(n\Delta t)) - \Phi(\overline{X}_n)| \\ &\quad + |\Phi(\overline{X}_n) - \mathbb{E}\Phi(X_n)|. \end{aligned} \tag{7.1}$$

Thanks to the averaging theorem of [3], which is proved under (WD), the first term above can be controlled by $C_r\epsilon^{1-r}$, where C_r is a constant, depending on r, Φ, x, y, T for any $0 < r < 1$.

For the second term, since we look at the error made by using a deterministic scheme to approximate a deterministic equation, there is no difference between the strong and the weak orders since the test function Φ is Lipschitz continuous; so we again use Proposition 5.13.

For the third term, we see that we have to control some error between two different numerical schemes, in a weak sense. The usual strategy is to decompose this error by means of an auxiliary function satisfying some kind of Kolmogorov equation.

More precisely, we use the deterministic scheme defining $\overline{X}_k$ in order to define for any $0 \leq k \leq n$

$$u_n(k,x) = \Phi(\overline{X}_{n-k}(x)), \tag{7.2}$$

where we explicitly mention dependence of the numerical solution $\overline{X}_k$ on the initial condition x.

Remark 7.1. We can easily prove that we have

$$\begin{aligned} &u_n(n,x) = \Phi(x), \\ &u_n(k,x) = u_n(k+1, S_{\Delta t}x + \Delta t S_{\Delta t}\overline{F}(x)) \text{ for any } k < n. \end{aligned}$$

This is the way this function is defined in [14].

We now analyze the error by identifying a telescoping sum:

$$\begin{aligned} |\Phi(\overline{X}_n) - \mathbb{E}\Phi(X_n)| &= |u_n(0,x) - \mathbb{E}u_n(n, X_n)| \\ &= \left| \sum_{k=0}^{n-1} (\mathbb{E}u_n(k, X_k) - \mathbb{E}u_n(k+1, X_{k+1})) \right| \\ &\leq \sum_{k=0}^{n-1} |\mathbb{E}u_n(k, X_k) - \mathbb{E}u_n(k+1, X_{k+1})|. \end{aligned} \tag{7.3}$$

According to Lemma 7.2 below, u_n is of class $\mathcal{C}_b^2$, and we can control the first and second order derivatives.

LEMMA 7.2. *For any $0 < T < +\infty$, there exists $C_T > 0$ such that for any $0 \leq n \leq n_0 = \lfloor \frac{T}{\Delta t} \rfloor$ and $0 \leq k \leq n$, we have for any $x \in H$, $h \in H$, $h_1, h_2 \in H$*

$$\begin{aligned} |D_x u_n(k,x).h| &\leq C_T|h|, \\ |D^2_{xx} u_n(k,x).(h_1,h_2)| &\leq C_T|h_1||h_2|. \end{aligned}$$

Moreover for any $0 \leq k \leq n-1$ *and any* $x \in H$, $h \in H$,

$$|D_x u_n(k,x).h| \leq C_T \left(\Delta t |h| + \frac{|h|_{(-A)^{-\eta}}}{((n-k)\Delta t)^\eta} \right),$$

where η *is defined in Assumption* 3.3.

Since the auxiliary functions u_n are linked to the deterministic discrete time process $(\overline{X}_n)_{n\geq 0}$, the proof does not use stochastic tools.

Moreover the second and the third estimates of this lemma reveal some smoothing effect in the equation, due to the semigroup $(S_{\Delta t}^n)_{n\in\mathbb{N}}$. The necessity for such results is specific to the infinite dimensional framework; we can also remark that a control of the second derivative with $C|h_1||h_2|_{(-A)^\eta}$ is not sufficient.

To make the proof of Theorem 4.2 clearer, we postpone the proof of Lemma 7.2 in section 7.2.

In the general term of (7.3), we proceed with a Taylor expansion, and using the estimates of Lemma 7.2 we get

$$\begin{aligned} &|\mathbb{E}u_n(k,X_k) - \mathbb{E}u_n(k+1,X_{k+1})| \\ &\quad = |\mathbb{E}u_n(k+1, S_{\Delta t}X_k + \Delta t S_{\Delta t}\overline{F}(X_k)) - \mathbb{E}u_n(k+1, S_{\Delta t}X_k + \Delta t S_{\Delta t}\tilde{F}_k)| \\ &\quad \leq \Delta t |\mathbb{E}D_x u_n(k+1, S_{\Delta t}X_k + \Delta t S_{\Delta t}\overline{F}(X_k)).(S_{\Delta t}\overline{F}(X_k) - S_{\Delta t}\tilde{F}_k)| \\ &\qquad + C\Delta t^2 \mathbb{E}|\tilde{F}_k - \overline{F}(X_k)|^2. \end{aligned} \tag{7.4}$$

Since F is bounded, the last term is of order $\mathrm{O}(\Delta t^2)$, and when we sum over $0 \leq k \leq n-1$, we get a $\mathrm{O}(\Delta t)$ term, which is already dominated in the final estimate.

When $k = n-1$, since $u_n(n,.) = \Phi$,

$$|\mathbb{E}u_n(n-1,X_{n-1}) - \mathbb{E}u_n(n,X_n)| \leq C\Delta t. \tag{7.5}$$

In the rest of the proof, we focus on the general case $0 \leq k < n-1$.

For the first term, we do not exactly follow the proof of [14]; we rather define auxiliary functions for $0 \leq k \leq n$ in order to keep on looking at a weak error term:

$$\Psi_n(k,x,y) = D_x u_n(k, S_{\Delta t}x + \Delta t S_{\Delta t}\overline{F}(x)).(S_{\Delta t}F(x,y)). \tag{7.6}$$

Then if we define $\overline{\Psi}_n(k+1,x) = \int_H \Psi_n(k+1,x,y)\mu^x(dy)$ we have

$$\begin{aligned} &|\mathbb{E}D_x u_n(k+1, S_{\Delta t}X_k + \Delta t S_{\Delta t}\overline{F}(X_k)).(S_{\Delta t}\overline{F}(X_k) - S_{\Delta t}\tilde{F}_k)| \\ &\quad \leq \frac{1}{MN} \sum_{m=n_T}^{n_T+N-1} \sum_{j=1}^{M} |\mathbb{E}\Psi_n(k+1, X_k, Y_{k,m,j}) - \mathbb{E}\overline{\Psi}_n(k+1,X_k)|. \end{aligned}$$

By using conditional expectation $\mathbb{E}_k$ with respect to $\mathcal{G}_k$ defined by (6.5),

$$\mathbb{E}\Psi_n(k+1,X_k,Y_{k,m,j}) = \mathbb{E}\mathbb{E}_k\Psi_n(k+1,X_k,Y_{k,m,j})$$

does not depend on j, so that in the sequel we fix $j \in \{1,\dots,M\}$.

The following Lemma gives the regularity results for the auxiliary functions.

LEMMA 7.3. *For any* $0 < T < +\infty$ *and* $\Delta t_0 > 0$, *there exists a constant* C, *such that for any* $0 < \Delta t \leq \Delta t_0$, *any* $1 \leq n \leq n_0 = \lfloor \frac{T}{\Delta t} \rfloor$, *and any* $0 \leq k \leq n-1$, *the following derivatives exist and are controlled: for any* $x, y \in H$, $h, k \in H$,

$$\begin{aligned} &|D_y\Psi_n(k,x,y).h| \leq C|h|_H, \\ &|D^2_{yy}\Psi_n(k,x,y).h| \leq C\left(1 + \frac{C}{((n-k)\Delta t)^\eta}\right)|h|_H|k|_H. \end{aligned}$$

Like in Lemma 7.2, the proof relies on the smoothing effect of the semigroup $(S^n_{\Delta t})_{n\in\mathbb{N}}$. The lemma is proved below in section 7.2.

We can then apply Theorem 5.11, using the conditional expectation $\mathbb{E}_k$: for any $n_T \le m \le m_0$ and $k < n-1$

$$|\mathbb{E}\Psi_n(k{+}1, X_k, Y_{k,m,j}){-}\mathbb{E}\overline{\Psi}_n(k{+}1, X_k)| \le Ce^{-cm\tau}{+}C\left(1+\frac{1}{((m-1)\tau)^{1/2-\kappa}}\right)\tau^{1/2-\kappa}.$$

Therefore

$$\begin{aligned}
&|\mathbb{E}D_x u_n(k+1, S_{\Delta t}X_k + \Delta t S_{\Delta t}\overline{F}(X_k)).(S_{\Delta t}\overline{F}(X_k) - S_{\Delta t}\tilde{F}_k)| \\
&\quad \le \frac{1}{MN}\sum_{m=n_T}^{n_T+N-1}\sum_{j=1}^{M}|\mathbb{E}\Psi_n(k+1, X_k, Y_{k,m,j}) - \mathbb{E}\overline{\Psi}_n(k+1, X_k)| \\
&\quad \le C\frac{e^{-cn_T\tau}}{N\tau+1} + C\tau^{1/2-\kappa} + C\frac{\tau^{1/2-\kappa}}{((n_T-1)\tau)^{1/2-\kappa}}.
\end{aligned}$$

To conclude, it remains to use (7.3) and (7.1).

Remark 7.4. When the strict dissipativity assumption is satisfied, we can indeed obtain a bound without $\frac{1}{((m-1)\tau)^{1/2-\kappa}}$: we can control the distance between the invariant measures of the continuous and discrete time processes, thanks to the second part of Corollary 5.12.

7.2. Proof of the auxiliary Lemmas 7.2 and 7.3.

Proof of Lemma 7.2. We use the following expression for $\overline{X}_n(x)$:

$$\overline{X}_n(x) = S^n_{\Delta t}x + \Delta t\sum_{k=0}^{n-1} S^{n-k}_{\Delta t}\overline{F}(\overline{X}_k(x)).$$

By definition, for any $0 \le k \le n$ we have $u_n(k,x) = \Phi(\overline{X}_{n-k}(x))$; we see that the derivatives in directions $h, h_1, h_2 \in H$ are given by

$$D_x u_n(k,x).h = D\Phi(\overline{X}_{n-k}(x)).\left(\frac{d}{dx}\overline{X}_{n-k}(x).h\right),$$

and

$$\begin{aligned}
D^2_{xx}u_n(k,x).(h_1,h_2) &= D\Phi(\overline{X}_{n-k}(x)).\left(\frac{d^2}{dx^2}\overline{X}_{n-k}(x).(h_1,h_2)\right) \\
&\quad + D^2\Phi(\overline{X}_{n-k}(x)).\left(\frac{d}{dx}\overline{X}_{n-k}(x).h_1, \frac{d}{dx}\overline{X}_{n-k}(x).h_2\right).
\end{aligned}$$

Φ is of class $\mathcal{C}^2$ on H with bounded derivatives; therefore we just need to control $\frac{d}{dx}\overline{X}_{n-k}(x).h$ and $\frac{d^2}{dx^2}\overline{X}_{n-k}(x).(h_1,h_2)$. We use the following estimates of the derivatives of $\overline{F}$, given in Proposition 5.5: for any $x \in H$, $h \in H$, $h_1, h_2 \in H$,

$$\begin{aligned}
|D\overline{F}(x).h| &\le C|h| \\
|(-A)^{-\eta}D^2\overline{F}(x).(h_1,h_2)| &\le C|h_1||h_2|.
\end{aligned}$$

(i) For any $0 \le n \le n_0$, we can write

$$\frac{d}{dx}\overline{X}_n(x).h = S^n_{\Delta t}h + \Delta t\sum_{k=0}^{n-1} S^{n-k}_{\Delta t}D\overline{F}(\overline{X}_k(x))\left(\frac{d}{dx}\overline{X}_k(x).h\right);$$

therefore

$$\left|\frac{d}{dx}\overline{X}_n(x).h\right| \leq |h| + C\Delta t \sum_{k=0}^{n-1} \left|\frac{d}{dx}\overline{X}_k(x).h\right|,$$

and a discrete Gronwall lemma then yields

$$\left|\frac{d}{dx}\overline{X}_n(x).h\right| \leq |h|e^{Cn\Delta t} \leq |h|e^{CT}.$$

(ii) For any $0 \leq n \leq n_0$, we can write

$$\begin{aligned}
&\frac{d^2}{dx^2}\overline{X}_n(x).(h_1,h_2) \\
&\quad = \Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k}(-A)^{\eta}(-A)^{-\eta} D^2\overline{F}(\overline{X}_k(x)).\left(\frac{d}{dx}\overline{X}_k(x).h_1, \frac{d}{dx}\overline{X}_k(x).h_2\right) \\
&\qquad + \Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k} D\overline{F}(\overline{X}_k(x)).\left(\frac{d^2}{dx^2}\overline{X}_k(x).(h_1,h_2)\right).
\end{aligned}$$

Since $|S_{\Delta t}^{n-k}(-A)^{\eta}| \leq \frac{C}{((n-k)\Delta t)^{\eta}}$ when $k < n$, and thanks to the previous estimates on $\frac{d}{dx}\overline{X}_k(x).h$ and $D^2\overline{F}$, we get

$$\begin{aligned}
&\left|S_{\Delta t}^{n-k}(-A)^{\eta}(-A)^{-\eta} D^2\overline{F}(\overline{X}_k(x)).\left(\frac{d}{dx}\overline{X}_k(x).h_1, \frac{d}{dx}\overline{X}_k(x).h_2\right)\right| \\
&\quad \leq \frac{C}{((n-k)\Delta t)^{\eta}}|h|_1|h|_2.
\end{aligned}$$

Therefore

$$\begin{aligned}
\left|\frac{d^2}{dx^2}\overline{X}_n(x).(h_1,h_2)\right| &\leq C|h_1||h_2| \\
&\quad + C\Delta t \sum_{k=0}^{n-1} \left|\frac{d^2}{dx^2}\overline{X}_k(x).(h_1,h_2)\right|,
\end{aligned}$$

and a discrete Gronwall Lemma then yields

$$\left|\frac{d^2}{dx^2}\overline{X}_n(x).(h_1,h_2)\right| \leq C|h_1||h_2|.$$

(iii) To prove the last estimate of the lemma, we write

$$\begin{aligned}
\left|\frac{d}{dx}\overline{X}_n(x).h\right| &= \left|S_{\Delta t}^n h + \Delta t \sum_{k=0}^{n-1} S_{\Delta t}^{n-k} D\overline{F}(\overline{X}_k(x))\left(\frac{d}{dx}\overline{X}_k(x).h\right)\right| \\
&\leq \frac{C}{(n\Delta t)^{\eta}}|h|_{(-A)^{-\eta}} + C\Delta t \sum_{j=0}^{n-1} \left|\frac{d}{dx}\overline{X}_k(x).h\right|.
\end{aligned}$$

We obtain

$$(n\Delta t)^\eta \left| \frac{d}{dx}\overline{X}_n(x).h \right| \le C|h|_{(-A)^{-\eta}} + C\Delta t(n\Delta t)^\eta |h|_H + C(n\Delta t)^\eta \Delta t \sum_{j=1}^{n-1} \frac{1}{(j\Delta t)^\eta} (j\Delta t)^\eta \left| \frac{d}{dx}\overline{X}_j(x).h \right|.$$

To conclude, we use a Gronwall lemma to get

$$(n\Delta t)^\eta \left| \frac{d}{dx}\overline{X}_n(x).h \right| \le C_T(|h|_{(-A)^{-\eta}} + \Delta t(n\Delta t)^\eta). \qquad \square$$

Proof of Lemma 7.3.

(i) The first derivative with respect to y is easy to control: we have for any $h \in H$

$$D_y\Psi_n(k,x,y).h = D_x u_n(k, S_{\Delta t}x + \Delta t S_{\Delta t}\overline{F}(x)).(S_{\Delta t}D_yF(x,y).h),$$

and we get $|D_y\Psi_n(k,x,y).h| \le C|h|_H$.

(ii) When we look at the second order derivative, we see that

$$D^2_{yy}\Psi_n(k,x,y).(h,k) = D_x u_n(k, S_{\Delta t}x + \Delta t S_{\Delta t}\overline{F}(x)).(S_{\Delta t}D^2_{yy}F(x,y).(h,k)).$$

Thanks to the last estimate of Lemma 7.2, we can control the expression by

$$C\left(\Delta t|S_{\Delta t}D^2_{yy}F(x,y).(h,k)|_H + \frac{|S_{\Delta t}D^2_{yy}F(x,y).(h,k)|_{(-A)^{-\eta}}}{((n-k)\Delta t)^\eta}\right).$$

We then notice that

$$\begin{aligned}\Delta t|S_{\Delta t}D^2_{yy}F(x,y).(h,k)|_H &\le C\Delta t^{1-\eta}|(-A)^{-\eta}D^2_{yy}F(x,y).(h,k)| \\ &\le C|h|_H|k|_H,\end{aligned}$$

since $\eta < 1$ and Δt is bounded, and thanks to Assumption 3.4; the other part is controlled thanks to Assumption 3.4. $\square$

Acknowledgments. The author is glad to thank Arnaud Debussche and Erwan Faou for their comments during the preparation of this article, and the referees who helped improve the exposition with their remarks.

REFERENCES

[1] A. Abdulle, W. E, and T. Li, *Effectiveness of implicit methods for stiff stochastic differential equations*, Comm. Comput. Phys., 3 (2008), pp. 295–307.

[2] C.-E. Bréhier, *Approximation of the invariant measure via an Euler scheme for stochastic pde's driven by space-time white noise*, submitted.

[3] C.-E. Bréhier, *Strong and weak orders in averaging for spdes*, Stochastic Process. Appl., 122 (2012), pp. 2553–2593.

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

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

[6] M. Crouzeix and V. Thomée, *On the discretization in time of semilinear parabolic equations with nonsmooth initial data*, Math. Comp., 49 (1987), pp. 359–377.

[7] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 44, Cambridge University Press, Cambridge, UK, 1992.
[8] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, London Math. Soc. Lecture Note Ser. 229, Cambridge University Press, Cambridge, UK, 1996.
[9] A. Debussche, *Weak approximation of stochastic partial differential equations: The nonlinear case*, Math. Comp., 80 (2011), pp. 89–117.
[10] A. Debussche, Y. Hu, and G. Tessitore, *Ergodic BSDEs under weak dissipative assumptions*, Stochastic Process. Appl., 121 (2011), pp. 407–426.
[11] W. E, *Analysis of the heterogeneous multiscale method for ordinary differential equations*, Commun. Math. Sci., 1 (2003), pp. 423–436.
[12] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.
[13] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *Heterogeneous multiscale methods: A review*, Comm. Comput. Phys., 2 (2007), pp. 367–450.
[14] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.
[15] S. Kuksin and A. Shirikyan, *A coupling approach to randomly forced nonlinear PDE's.* I, Comm. Math. Phys., 221 (2001), pp. 351–366.
[16] M.-N. Le Roux, *Semidiscretization in time for parabolic problems*, Math. Comp., 33 (1979), pp. 919–931.
[17] J. C. Mattingly, *Exponential convergence for the stochastically forced Navier-Stokes equations and other partially dissipative dynamics*, Comm. Math. Phys., 230 (2002), pp. 421–462.
[18] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts Appl. Math. 53, Springer, New York, 2008.
[19] J. Printems, *On the discretization in time of parabolic stochastic partial differential equations*, Monte Carlo Methods Appl., 7 (2001), pp. 359–368.
[20] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 385–391.