COMM. MATH. SCI.
Vol. 4, No. 4, pp. 707–729

STRONG CONVERGENCE OF PROJECTIVE INTEGRATION SCHEMES FOR SINGULARLY PERTURBED STOCHASTIC DIFFERENTIAL SYSTEMS*

DROR GIVON†, IOANNIS G. KEVREKIDIS‡, AND RAZ KUPFERMAN§

Abstract. We study the convergence of the slow (or "essential") components of singularly perturbed stochastic differential systems to solutions of lower dimensional stochastic systems (the "effective", or "coarse" dynamics). We prove strong, mean-square convergence in systems where both fast and slow components are driven by noise, with full coupling between fast and slow components. We analyze a class of "projective integration" methods, which consist of a hybridization between a standard solver for the slow components, and short runs for the fast dynamics, which are used to estimate the effect that the fast components have on the slow ones. We obtain explicit bounds for the discrepancy between the results of the projective integration method and the slow components of the original system.

Key words. Dimension reduction, stochastic differential equations, scale separation, singular perturbations, projective integration

AMS subject classifications. 60H15, 60F25, 65C20

1. Introduction

Many problems in the natural sciences give rise to singularly perturbed systems of stochastic differential equations (SDEs) of the form

$$dx_t^\epsilon = a(x_t^\epsilon, y_t^\epsilon)\, dt + b(x_t^\epsilon, y_t^\epsilon)\, dU_t, \qquad x_0^\epsilon = x_0, \tag{1.1a}$$

$$dy_t^\epsilon = \frac{1}{\epsilon} f(x_t^\epsilon, y_t^\epsilon)\, dt + \frac{1}{\sqrt{\epsilon}} g(x_t^\epsilon, y_t^\epsilon)\, dV_t, \qquad y_0^\epsilon = y_0, \tag{1.1b}$$

where x_t^ϵ and y_t^ϵ are n- and m-dimensional diffusion processes. The functions $a(x,y) \in \mathbb{R}^n$ and $f(x,y) \in \mathbb{R}^m$ are the drifts, and the functions $b(x,y) \in \mathbb{R}^{n\times d_1}$ and $g(x,y) \in \mathbb{R}^{m\times d_2}$ are the diffusions; U_t and V_t are d_1- and d_2-dimensional mutually independent Wiener processes. The parameter ϵ represents the ratio between the natural time scale of the x_t^ϵ and y_t^ϵ variables. We are concerned with situations where $\epsilon \ll 1$, i.e., with a separation of scales; in such case the vector x_t^ϵ is called the "slow component" of the system, and the vector y_t^ϵ is the "fast component" of the system.

Systems of the form (1.1) arise in various situations; see [1] for a classical review with numerous applications. In many cases, one is only interested in predicting the time evolution of the slow component x_t^ϵ, yet, this cannot be done, in a direct approach, without solving the full system of equations. In biomaterials, for example, the fast dynamics may have a characteristic time of the order of picoseconds (the fastest atomic motions), while the process of interest has a characteristic time of the order of seconds. No computer can deal with such a disparity of scales.

In the past four decades, singularly perturbed systems of the form (1.1) have been the focus of extensive research, within the framework of *averaging methods*. The separation of scales is taken to advantage to derive, in the limit $\epsilon \to 0$, a reduced

*Received: March 21, 2006; accepted (in revised version): September 5, 2006. Communicated by Andrew Stuart.

†Department of Mathematics, University of California and LBNL, Berkeley, CA 94720, USA.
‡Department of Chemical Engineering, PACM and Mathematics, Princeton University, USA.
§Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel.

equation for an n-dimensional process $\bar{x}_t$, which approximates the slow component x_t^ϵ. This effective equation is of the form

$$d\bar{x}_t = \bar{a}(\bar{x}_t)\,dt + \bar{b}(\bar{x}_t)\,dW_t, \tag{1.2}$$

where $\bar{a}(x) \in \mathbb{R}^n$, $\bar{b}(x) \in \mathbb{R}^{n\times d_1}$ and W_t is a d_1-dimensional Wiener process.

The conditions under which x_t^ϵ converges, as $\epsilon \to 0$, to the solution of effective dynamics of the form (1.2) was first addressed by Khasminskii [2] in the context of Markov diffusion processes. The averaging principle in the context of SDEs is described in the monograph by Skorokhod [3], and can be summarized as follows: Assume that for every fixed x the rapid variables induce a unique invariant, ergodic measure $\mu^x(dy)$. Then, as $\epsilon \to 0$, x_t^ϵ converges in distribution, or weakly on every finite interval $[0,T]$, to the solution $\bar{x}_t$ of a closed equation of the form (1.2) where

$$\begin{aligned}\bar{a}(x) &= \int_{\mathbb{R}^m} a(x,y)\mu^x(dy),\\ \bar{b}(x)\bar{b}^T(x) &= \int_{\mathbb{R}^m} b(x,y)b^T(x,y)\mu^x(dy).\end{aligned}$$

Generalizations can be found, for example, in Papanicolaou et al. [4], Pardoux and Veretennikov [5, 6], and Freidlin and Wentzell [7], with notably extensions to convergence in probability (i.e., pathwise). Kifer [8] proved convergence in some "averaged L^2 sense" without assuming the non-degeneracy condition on the diffusion coefficients, which is an essential condition in [2] and its sequels, thus generalizing Anosov's theorem for deterministic systems. A survey on the existence of an effective dynamics for the case where $b(x,y) = 0$ in (1.1a) (i.e., x_t^ϵ satisfies an ordinary differential equation (ODE) coupled to a fast stochastic process) can be found in Papanicolaou and co-workers [9, 4] where they improve a result of Khasminskii [10] and show weak convergence to a diffusion Markov process for a broad class of situations. For the case where y_t^ϵ satisfies an Itô stochastic equation Kifer [11] proves convergence in the sup-norm and E et al. [12] derive estimates on the rate of strong (L^1) convergence to the solution of an effective ordinary differential system (see also Vanden Eijnden [13]).

While the averaging principle and its resulting effective dynamics (1.2) provide a substantial simplification of the original system (1.1), it is often impossible, or impractical, to obtain the reduced equations in closed form (for example, because the invariant measure μ^x is unknown, or because integrations cannot be performed analytically). This has motivated the development of algorithms such as projective and coarse projective integration [14, 15, 16] within the so-called equation-free framework [17, 18]. In this framework, short bursts of appropriately initialized "fine scale" simulations are used to estimate on demand the numerical quantities required to perform scientific computing task with coarse-grained models (time derivatives, residuals, the action of (slow) Jacobians, and, for the case of *stochastic* coarse-grained models, the local effective noise drift and diffusivity, e.g. [19]). When a stochastic problem effectively closes at a deterministic level (e.g. in terms of the expectations of some slow observables), traditional ODE integration algorithms, whether explicit or implicit, can be wrapped around on-demand estimates of the (slow) time-derivatives of these observables to accelerate the simulation of the effective equation (e.g. [14, 16, 20]). Here we extend the idea of projective integration for a (deterministic) effective model to the case where the effective model is a stochastic one.

In its simplest formulation, this extended, stochastic projective integration scheme can be described as follows: Let Δt be a fixed time step, and X_n be the numerical approximation to the coarse variable, $\bar{x}$, at time $t_n = n\Delta t$. Inspired by the limiting equation (1.2), X_n is evolved in time by an Euler-Maruyama step,

$$X_{n+1} = X_n + A(X_n)\,\Delta t + B(X_n)\,\Delta W_n, \tag{1.3}$$

where ΔW_n are Brownian displacements over a time interval Δt. We refer to (1.3) as the *macro-solver* (or, macro integrator).

The functions $A(X)$ and $B(X)$ approximate the functions $\bar{a}(x)$ and $\bar{b}(x)$, which result from the averaging (1.3) over an ergodic measure. The ergodic property implies that instead of ensemble averaging we can use averaging over paths of the rapid variables with fixed x. Since, by assumption, these averages cannot be performed analytically, they are approximated by an empirical average over short runs of the fast dynamics. These "short runs" are over time intervals that are sufficiently long for empirical averages to be sufficiently close to their limiting ensemble averages, yet sufficiently short for the entire procedure to be efficient compared to the direct solution of the coupled system.

Thus, given the coarse variable at the n-th time step, X_n, we take some initial value for the fast component Y_0^n, and solve (1.1b) numerically with step size δt and $x = X_n$ fixed. We denote the discrete variables associated with the fast dynamics at the n-th coarse step by Y_m^n, $m = 0, 1, \ldots, M$. The numerical solver used to generate the sequence Y_m^n is called the *micro-solver* (or micro-integrator). The simplest choice is again the Euler-Maruyama scheme,

$$Y_{m+1}^n = Y_m^n + \frac{1}{\epsilon} f(X_n, Y_m^n)\,\delta t + \frac{1}{\sqrt{\epsilon}} g(X_n, Y_m^n)\Delta V_m^n, \tag{1.4}$$

where ΔV_m^n are Brownian displacements over a time interval δt. Since we assume that the y dynamics is ergodic, we may choose, among other choices, $Y_0^n = y_0$.

Having generated the trajectories Y_m^n, the functions $\bar{a}$ and $\bar{b}$ are estimated by

$$\begin{aligned} A(X_n) &= \frac{1}{M} \sum_{m=1}^{M} a(X_n, Y_m^n), \\ B(X_n)B^T(X_n) &= \frac{1}{M} \sum_{m=1}^{M} b(X_n, Y_m^n) b^T(X_n, Y_m^n). \end{aligned} \tag{1.5}$$

$B(X_n)$ can then be extracted from $B(X_n)B^T(X_n)$ through a Cholesky decomposition. Finally, to reduce the statistical noise, several independent realizations of the micro-solver can be carried out, in which case expressions (1.5) for $A(X)$ and $B(X)$ involve an additional averaging over these independent realizations. Equations (1.3), (1.4), and (1.5) define the projective integration scheme.

In this paper we analyze systems of the form (1.1), along with the projective integration scheme for the case where $b = b(x)$, i.e., the diffusion function of the slow variables does not depend explicitly on the fast variables. Note that in this case $\bar{b}(x) = B(x) = b(x)$ (when b depends on y there is no strong convergence, as a counter example is in the Discussion section shows).

The contribution of this paper is two-fold:

1. We prove L^2 convergence of x_t^ϵ to $\bar{x}_t$ under specified conditions. That is, we obtain strong converge for a class of systems where the limiting dynamics

is stochastic. In particular, we obtain an explicit estimate on the rate of convergence of the form

$$\sup_{0\le t\le T} \mathbb{E}\left|x_t^\epsilon - \bar{x}_t\right|^2 \le C\sqrt{\epsilon}.$$

2. We derive estimates for the L^2 (again, pathwise) error between the solution $\bar{x}_t$ of the effective dynamics (1.2) and the solution X_n of the projective integration scheme (1.3). Specifically, we obtain an error estimate of the form

$$\mathbb{E}|\bar{x}_{t_n} - X_n|^2 \le C\left(\Delta t + \frac{(\ln M\delta t)^2+1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t}\right).$$

Both results generalize the analysis in [12], which is limited to the case where the slow dynamics (and hence the limiting dynamics) are deterministic.

2. A strong limit theorem for the averaging principle

In this section we establish the convergence, under specified conditions, of x_t^ϵ, the slow component in (1.1), to $\bar{x}_t$, the solution of the effective dynamics (1.2). We prove mean-square convergence, i.e., pathwise, with W_t in (1.2) identified with U_t in (1.1a). We achieve this goal by estimating the mean-square deviation $\mathbb{E}|x_t^\epsilon - \bar{x}_t|^2$ between the two processes; our main result is Theorem 2.11. For the sake of readability we state in this section our assumptions, lemmas and theorems, deferring all proofs to the next section.

Throughout this work, the following assumptions are made:

Assumption 2.1.

A1. The functions $a = a(x,y)$ and $b = b(x)$ in (1.1a) *are measurable, Lipschitz continuous and have linear growth bounds: specifically, there exist constants L,K, such that*

$$|a(x_1,y_1) - a(x_2,y_2)|^2 + \|b(x_1) - b(x_2)\|^2 \le L^2\left(|x_1-x_2|^2 + |y_1-y_2|^2\right),$$

and

$$|a(x,y)|^2 + \|b(x)\|^2 \le K^2\left(1 + |x|^2 + |y|^2\right).$$

Here and below we use $|\cdot|$ to denote Euclidean vector norms and $\|\cdot\|$ for Frobenius matrix norms.

A2. The functions $f(x,y)$ and $g(x,y)$ in (1.1b) *are of class $\mathcal{C}^\infty$ and have bounded derivatives of any order; in particular, we can choose the Lipschitz constant L sufficiently large, such that it bounds the first derivatives of f,g. Moreover, $f(x,y)$ is assumed to be a bounded function of x for all y,*

$$\sup_x |f(x,y)| = c_f(y) < \infty,$$

and $g(x,y)$ is bounded,

$$\sup_{x,y} \|g(x,y)\| = c_g < \infty.$$

A3. There exists a constant $\alpha > 0$, independent of x, such that:

$$y^T g(x,y) g^T(x,y) y \ge \alpha\,|y|^2$$

for all $y \in \mathbb{R}^m$.

A4. There exists a constant $\beta > 0$, independent of x, such that

$$(y_1 - y_2) \cdot [f(x, y_1) - f(x, y_2)] + \|g(x, y_1) - g(x, y_2)\|^2 \leq -\beta \, |y_1 - y_2|^2$$

for all $y_1, y_2 \in \mathbb{R}^m$.

Some comments: (i) Assumption *A3* ensures the non-degeneracy of the fast dynamics (1.1b), with x viewed as a fixed parameter. (ii) Assumption *A4* is called the dissipative (or recurrence) condition; together with *A3* it guarantees the ergodicity of the fast dynamics (see Khasminskii [21, §3,4]). Assumption *A3* guarantees that the ergodic measure $\mu^x(dy)$ has a smooth density [22, 21]. (iii) Since a satisfies a Lipschitz condition, so does $\bar{a}$, and the effective dynamics (1.2) has a unique solution.

In practice, one rarely encounters systems in which the first two assumptions hold verbatim. These assumptions can be weakened at the expense of technical complications. For example, the extent of regularity can be relaxed, and the global Lipschitz constant can be replaced by a local Lipschitz continuity property. Thus, it is a matter of technicality to generalize our results to situations of practical interest. The last two assumptions are more imperative, but are satisfied in numerous (and generic) situations. Condition *A3* is satisfied, for example, in the case of fixed diffusion constant. Ergodic properties under much weaker conditions were demonstrated by Meyn and Tweedie [23]; see Mattingly et al. [24] for further elaboration and applications to various situations. Thus, both the dissipative assumption and the non-degeneracy assumptions can be partially relaxed.

Our first three lemmas provide mean-square estimates for the process $(x_t^\epsilon, y_t^\epsilon)$, with bounds independent of ϵ. The proofs are straightforward and are provided for completeness.

LEMMA 2.2. *The fast component y_t^ϵ satisfies,*

$$\sup_{0 \leq t \leq T} \mathbb{E}|y_t^\epsilon|^2 \leq C_1,$$

where $C_1 = C_1(y_0) = |y_0|^2 + c_f^2(0)/\beta^2 + c_g^2/\beta$.

LEMMA 2.3. *The slow component x_t^ϵ satisfies*

$$\sup_{0 \leq t \leq T} \mathbb{E}|x_t^\epsilon|^2 \leq C_2,$$

where

$$C_2 = C_2(T, x_0, y_0) = |x_0|^2 \, e^{(1+K^2)T} + \frac{K^2(1 + C_1)}{1 + K^2} e^{(1+K^2)T}.$$

LEMMA 2.4. *For all $0 \leq t_0 \leq t \leq T$, the mean-square displacement of the slow component satisfies*

$$\mathbb{E}|x_t^\epsilon - x_{t_0}^\epsilon|^2 \leq C_3 \, (t - t_0),$$

where $C_3 = C_3(T, x_0, y_0) = c_1 K^2 (1 + C_1 + C_2)$.

Our goal is to estimate the difference between x_t^ϵ, the slow component of (1.1), and $\bar{x}_t$, the solution of the effective dynamics (1.2). To this end we construct an auxiliary process, $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon) \in \mathbb{R}^n \times \mathbb{R}^m$: we divide the time interval $[0, T]$ into subintervals of length $\Delta = \sqrt{\epsilon}$, setting $t_k = k\Delta$, $k = 0, \ldots, \lfloor \frac{T}{\Delta} \rfloor$; for $s \in [0, T]$ we also define $t_s = \lfloor s/\Delta \rfloor \Delta$, the nearest breakpoint preceding s.

With initial conditions $(\tilde{x}_0^\epsilon, \tilde{y}_0^\epsilon) = (x_0, y_0)$, the process $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$ is governed for $t \in [t_k, t_{k+1})$ by the SDE

$$\begin{aligned} d\tilde{x}_t^\epsilon &= a(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\, dt + b(x_{t_k}^\epsilon)\, dU_t, \\ d\tilde{y}_t^\epsilon &= \frac{1}{\epsilon} f(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\, dt + \frac{1}{\sqrt{\epsilon}} g(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\, dV_t, \qquad \tilde{y}_{t_k}^\epsilon = y_{t_k}^\epsilon. \end{aligned} \tag{2.1}$$

The pair $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$ satisfies dynamics similar to (1.1), notably with the same random noise, except that the argument x in the functions a, b, f, g, is replaced by x_t^ϵ at the beginning of the sub-interval, $t = t_k$, whereas the fast component $\tilde{y}_t^\epsilon$ is reset to equal y_t^ϵ at each breakpoint t_k. The time interval Δ is selected small enough, $\Delta \ll 1$, so that $\tilde{x}_t^\epsilon$ does not deviate much from x_t^ϵ; on the other hand $\Delta \gg \epsilon$, so that the empirical distribution of $\tilde{y}_t^\epsilon$ in the k-th interval is close to the invariant distribution μ^x, with $x = x_{t_k}^\epsilon$. The introduction of the auxiliary process $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$ provides an intermediate step between the processes x_t^ϵ and $\bar{x}_t$, whose difference we need to estimate. As will be shown, $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$ remains close to $(x_t^\epsilon, y_t^\epsilon)$ because Δ is small enough (on the x-timescale) and $\tilde{y}_t^\epsilon$ is repeatedly reset to equal y_t^ϵ. On the other hand, $\tilde{x}_t^\epsilon$ remains close to $\bar{x}_t$ because Δ is large enough (on the y-timescale) so that the time average of $a(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)$ is close enough to $\bar{a}(x_{t_k}^\epsilon)$.

The next two lemmas estimate the mean-square differences between the fast and slow components of the processes $(x_t^\epsilon, y_t^\epsilon)$ and $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$:

LEMMA 2.5. *Let* $(x_t^\epsilon, y_t^\epsilon)$ *and* $(\tilde{x}_t^\epsilon, \tilde{y}_t^\epsilon)$ *be the respective solutions of* (1.1) *and* (2.1). *Then*

$$\sup_{0 \le t \le T} \mathbb{E}\, |y_t^\epsilon - \tilde{y}_t^\epsilon|^2 \le C_4 \sqrt{\epsilon},$$

where $C_4 = C_4(T, x_0, y_0) = (L^2/\beta)(2 + 1/\beta)\, C_3$.

LEMMA 2.6.

$$\sup_{0 \le t \le T} \mathbb{E}\, |x_t^\epsilon - \tilde{x}_t^\epsilon|^2 \le C_5 \sqrt{\epsilon}, \tag{2.2}$$

where $C_5 = C_5(T, x_0, y_0) = c_1\, L^2\, T\, (C_3 + C_4)$.

Having estimated the mean-square difference between x_t^ϵ and $\tilde{x}_t^\epsilon$, it remains to estimate the mean-square difference between $\tilde{x}_t^\epsilon$ and $\bar{x}_t$. The smallness of the latter is due to the mixing properties of the fast dynamics.

For $k = 1, 2, \dots, \lfloor T/\Delta \rfloor$, we set $x_k = x_{t_k}^\epsilon$ and define the stochastic process z_t^k which satisfies the SDE,

$$dz_t^k = f(x_k, z_t^k)\, dt + g(x_k, z_t^k)\, dV_t^k, \qquad z_0^k = y_{t_k}^\epsilon, \tag{2.3}$$

where the V_t^k are independent Wiener processes. The process z_t^k is statistically equivalent to a shifted and rescaled version of $\tilde{y}_t^\epsilon$, that is, $z_t^k \sim \tilde{y}_{(t-t_k)/\epsilon}^\epsilon$. Its introduction is only needed to simplify the notation.

The dynamics (2.3) is ergodic with invariant measure μ^{x_k} (Assumptions *A3,A4*). Moreover, the process z_t^k is exponentially mixing. Recall that if $\mathcal{F}_a^b$ is the σ-algebra generated by $\left\{z_t^k : a \le t \le b\right\}$, the strong mixing coefficient of z_t^k is defined by

$$\alpha_t = \sup_{s \ge 0} \sup_{A \in \mathcal{F}_0^s, B \in \mathcal{F}_{s+t}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\,.$$

It is a measure on the independence of events that are separated by a time interval of at least t.

The following lemma, proved by Malyshkin [25], establishes a bound on the strong mixing coefficient of the process z_t^k:

LEMMA 2.7. *There exist constants $c_2, c_3 > 0$, such that for all $\gamma \in (0, 1/2)$,*

$$\alpha_t \le c_2 \exp\left(-c_3\, t^\gamma\right).$$

Moreover, all moments associated with the invariant measure μ^{x_k} are finite.

The following relation between correlations and the mixing coefficient is proved in Billingsley [26, §20]:

LEMMA 2.8. *If ζ, η are real-valued random variables, where ζ is $\mathcal{F}_0^s$-measurable and η is $\mathcal{F}_{s+t}^\infty$-measurable, with*

$$\mathbb{E}\left|\zeta\right|^2 < \infty, \quad \mathbb{E}\left|\eta\right|^2 < \infty.$$

Then

$$\left|\mathbb{E}\left[\zeta \cdot \eta\right] - \mathbb{E}\left[\zeta\right] \cdot \mathbb{E}\left[\eta\right]\right| \le 2\alpha_t^{1/2} \left(\mathbb{E}\left|\zeta\right|^2\right)^{1/2} \left(\mathbb{E}\left|\eta\right|^2\right)^{1/2}.$$

The process z_t^k is ergodic, but not stationary because the initial condition is not drawn from its invariant distribution, μ^{x_k}. We introduce a third auxiliary process ξ_t^k, which satisfies the same dynamics as z_t^k (in a pathwise sense), but with random initial conditions, ξ_0^k, drawn from the invariant distribution μ^{x_k}; ξ_t^k is a stationary process. The process ξ_t^k is needed in order to approximate the process z_t^k, which is initialized with the value $y_{t_k}^\epsilon$, by a similar process which is stationary. Stationarity is then exploited for an explicit estimate of correlations.

The next Lemma states that the two processes z_t^k and ξ_t^k converge exponentially fast to each other in a mean-square sense (that is, the stochastic map defined by (2.3) is exponentially contracting).

LEMMA 2.9. *For $k = 0, 1, \ldots, \lfloor T/\Delta \rfloor$,*

$$\mathbb{E}\left|z_t^k - \xi_t^k\right|^2 \le 2C_1 e^{-2\beta t}.$$

Equipped with the above, we can estimate the difference between $\tilde{x}_t^\epsilon$ and $\bar{x}_t$.

LEMMA 2.10. *For small enough ϵ,*

$$\sup_{0 \le t \le T} \mathbb{E}|\tilde{x}_t^\epsilon - \bar{x}_t|^2 \le C_6 \sqrt{\epsilon},$$

where $C_6 = C_6(T, x_0, y_0) = 4T(C_3 + C_5)\exp(\frac{5}{2}c_1 L^2 T)$.

Combining Lemma 2.10 with Lemma 2.6 we obtain our main result:

THEOREM 2.11. *Let x_t^ϵ be the slow component of (1.1), and $\bar{x}_t$ be the solution of the effective dynamics (1.2). Then, for small enough ϵ,*

$$\sup_{0 \le t \le T} \mathbb{E}\left|x_t^\epsilon - \bar{x}_t\right|^2 \le 2(C_5 + C_6)\sqrt{\epsilon}.$$

3. Proofs for Section 2

We start by establishing a number of relations, which will be used repeatedly below (for more details see [27, p.136]). First, Gronwall's inequality: if the real-valued function $v(t)$ satisfies a linear differential inequality of the form

$$\frac{dv}{dt} \leq \lambda v + c, \qquad v(t_0) = v_0,$$

then

$$v(t) \leq v_0 e^{\lambda(t-t_0)} + \frac{c}{\lambda}\left(e^{\lambda(t-t_0)} - 1\right). \tag{3.1}$$

Let $z_t \in \mathbb{R}^n$, $t \in [0,T]$, be the solution of the SDE

$$dz_t = a(z_t)\,dt + b(z_t)\,dW_t,$$

such that a and b are measurable, Lipschitz continuous and have linear growth bounds. Assume also the boundedness of the second moment of the initial value z_{t_0}. Applying the Itô formula, followed by Young's inequality,

$$\begin{aligned} \frac{d}{dt}\mathbb{E}|z_t|^2 &= 2\mathbb{E}\, z_t \cdot a(z_t) + \mathbb{E}\|b(z_t)\|^2 \\ &\leq \mathbb{E}|z_t|^2 + \mathbb{E}\left[|a(z_t)|^2 + \|b(z_t)\|^2\right]. \end{aligned} \tag{3.2}$$

Alternatively, using the Itô isometry and the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}|z_t - z_{t_0}|^2 &= \mathbb{E}\Big|\int_{t_0}^t a(z_s)\,ds + \int_{t_0}^t b(z_s)\,dW_s\Big|^2 \\ &\leq 2\,\mathbb{E}\Big|\int_{t_0}^t a(z_s)\,ds\Big|^2 + 2\,\mathbb{E}\Big|\int_{t_0}^t b(z_s)\,dW_s\Big|^2 \\ &\leq 2(t-t_0)\int_{t_0}^t \mathbb{E}|a(z_s)|^2\,ds + 2\int_{t_0}^t \mathbb{E}\|b(z_s)\|^2\,ds \\ &\leq c_1 \int_{t_0}^t \mathbb{E}\left[|a(z_s)|^2 + \|b(z_s)\|^2\right]\,ds, \end{aligned} \tag{3.3}$$

where $c_1 = c_1(T) = 2\,\max(1,T)$.

Proof of Lemma 2.2. Applying the first line of (3.2) to y_t^ϵ we obtain

$$\epsilon\frac{d}{dt}\mathbb{E}|y_t^\epsilon|^2 = 2\,\mathbb{E}\,y_t^\epsilon \cdot f(x_t^\epsilon, y_t^\epsilon) + \mathbb{E}\|g(x_t^\epsilon, y_t^\epsilon)\|^2. \tag{3.4}$$

Assumption *A4* with $y_1 = y_t^\epsilon$ and $y_2 = 0$ gives,

$$y_t^\epsilon \cdot [f(x_t^\epsilon, y_t^\epsilon) - f(x_t^\epsilon, 0)] + \|g(x_t^\epsilon, y_t^\epsilon) - g(x_t^\epsilon, 0)\|^2 \leq -\beta|y_t^\epsilon|^2,$$

or

$$y_t^\epsilon \cdot f(x_t^\epsilon, y_t^\epsilon) \leq -\beta|y_t^\epsilon|^2 + y_t^\epsilon \cdot f(x_t^\epsilon, 0).$$

Using Young's inequality $2p\cdot q \leq \beta|p|^2 + \beta^{-1}|q|^2$ with $p = y_t^\epsilon/\sqrt{2}$ and $q = f(x_t^\epsilon, 0)/\sqrt{2}$, and substituting the bound on f (Assumption *A2*),

$$y_t^\epsilon \cdot f(x_t^\epsilon, y_t^\epsilon) \leq -\frac{\beta}{2}|y_t^\epsilon|^2 + \frac{c_f^2(0)}{2\beta},$$

Adding the bound on g (Assumption *A2*) and substituting into (3.4) yields the differential inequality,

$$\epsilon \frac{d}{dt}\mathbb{E}|y_t^\epsilon|^2 \le -\beta\,\mathbb{E}|y_t^\epsilon|^2 + \left[\frac{c_f^2(0)}{\beta} + c_g^2\right].$$

The desired result follows from Gronwall's inequality (3.1).

Proof of Lemma 2.3. Applying (3.2) to x_t^ϵ,

$$\frac{d}{dt}\mathbb{E}|x_t^\epsilon|^2 \le \mathbb{E}|x_t^\epsilon|^2 + \mathbb{E}\left[|a(x_t^\epsilon, y_t^\epsilon)|^2 + \|b(x_t^\epsilon)\|^2\right].$$

Substituting the linear growth bound for a, b (Assumption *A1*), it follows that

$$\begin{aligned}\frac{d}{dt}\mathbb{E}|x_t^\epsilon|^2 &\le \mathbb{E}|x_t^\epsilon|^2 + K^2(1+\mathbb{E}|x_t^\epsilon|^2 + \mathbb{E}|y_t^\epsilon|^2)\\ &\le (1+K^2)\,\mathbb{E}|x_t^\epsilon|^2 + K^2(1+C_1),\end{aligned}$$

where the last inequality follows from Lemma 2.2. The desired result follows from Gronwall's inequality (3.1).

Proof of Lemma 2.4. Inequality (3.3) for x_t^ϵ reads

$$\mathbb{E}|x_t^\epsilon - x_{t_0}^\epsilon|^2 \le c_1 \int_{t_0}^t \mathbb{E}\left[|a(x_s^\epsilon, y_s^\epsilon)|^2 + \|b(x_s^\epsilon)\|^2\right]\, ds.$$

Using the linear growth bound for a, b (Assumption *A1*),

$$\mathbb{E}|x_t^\epsilon - x_{t_0}^\epsilon|^2 \le c_1 \int_{t_0}^t K^2(1+\mathbb{E}|x_s^\epsilon|^2 + \mathbb{E}|y_s^\epsilon|^2)\, ds \le c_1 K^2 (1 + C_1 + C_2)(t-t_0),$$

where the last inequality follows from Lemmas 2.2 and 2.3.

Proof of Lemma 2.5. Define $z_t = y_t^\epsilon - \tilde{y}_t^\epsilon$, fix $t \in [0,T]$ and set k such that $t \in [t_k, t_{k+1})$. The resetting of the auxiliary process at the break points t_k implies that $z_{t_k} = 0$ for all k.

Using the first line of (3.2) for the real-valued process $\epsilon|z_t|^2$,

$$\begin{aligned}\epsilon\frac{d}{dt}\mathbb{E}\,|z_t|^2 &= 2\,\mathbb{E}\, z_t\cdot\left(f\left(x_t^\epsilon, y_t^\epsilon\right) - f\left(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon\right)\right) + \mathbb{E}\left\|g(x_t^\epsilon, y_t^\epsilon) - g(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right\|^2\\ &= 2\,\mathbb{E}\, z_t\cdot\left(f(x_t^\epsilon, y_t^\epsilon) - f(x_t^\epsilon, \tilde{y}_t^\epsilon)\right) + 2\,\mathbb{E}\, z_t\cdot\left(f(x_t^\epsilon, \tilde{y}_t^\epsilon) - f(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\\ &\quad + \mathbb{E}\|g(x_t^\epsilon, y_t^\epsilon) - g(x_t^\epsilon, \tilde{y}_t^\epsilon) + g(x_t^\epsilon, \tilde{y}_t^\epsilon) - g(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\|^2\\ &\le 2\,\mathbb{E}\, z_t\cdot\left(f(x_t^\epsilon, y_t^\epsilon) - f(x_t^\epsilon, \tilde{y}_t^\epsilon)\right) + 2\,\mathbb{E}\, z_t\cdot\left(f(x_t^\epsilon, \tilde{y}_t^\epsilon) - f(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\\ &\quad + 2\,\mathbb{E}\,\|g(x_t^\epsilon, y_t^\epsilon) - g(x_t^\epsilon, \tilde{y}_t^\epsilon)\|^2 + 2\,\mathbb{E}\left\|g(x_t^\epsilon, \tilde{y}_t^\epsilon) - g(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right\|^2,\end{aligned} \tag{3.5}$$

where we have used $\|a+b\|^2 \le 2\,\|a\|^2 + 2\,\|b\|^2$.

By the dissipative assumption *A4*, the sum of the first and third terms on the right hand side are bounded by $-2\beta\,\mathbb{E}|z_t|^2$. The global Lipschitz continuity of f, g (Assumption *A2*) implies

$$\begin{aligned}\epsilon\frac{d}{dt}\mathbb{E}\,|z_t|^2 &\le -2\beta\,\mathbb{E}\,|z_t|^2 + 2L\,\mathbb{E}\,|z_t|\left|x_t^\epsilon - x_{t_k}^\epsilon\right| + 2L^2\,\mathbb{E}\left|x_t^\epsilon - x_{t_k}^\epsilon\right|^2\\ &\le -\beta\,\mathbb{E}\,|z_t|^2 + L^2(2+1/\beta)\,\mathbb{E}\left|x_t^\epsilon - x_{t_k}^\epsilon\right|^2,\end{aligned}$$

where we have used Young's inequality $2pq \le \beta p^2 + \frac{1}{\beta} q^2$ with $p = |z_t|$ and $q = L\left|x_t^\epsilon - x_{t_k}^\epsilon\right|$. By Lemma 2.4

$$\mathbb{E}\left|x_t^\epsilon - x_{t_k}^\epsilon\right|^2 \le C_3\,(t - t_k) \le C_3\,\sqrt{\epsilon},$$

thus, we obtain a linear differential inequality

$$\frac{d}{dt}\mathbb{E}\,|z_t|^2 \le -\frac{\beta}{\epsilon}\mathbb{E}\,|z_t|^2 + \frac{L^2(2+1/\beta)C_3}{\sqrt{\epsilon}}.$$

The desired result follows from Gronwall's inequality (3.1) upon integrating from t_k to t.

Proof of Lemma 2.6. By (3.3) with $z_t = x_t^\epsilon - \tilde{x}_t^\epsilon$,

$$\mathbb{E}\,|x_t^\epsilon - \tilde{x}_t^\epsilon|^2 \le c_1 \int_0^t \mathbb{E}\left[\left|a(x_s^\epsilon, y_s^\epsilon) - a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon)\right|^2 + \left\|b(x_s^\epsilon) - b(x_{t_s}^\epsilon)\right\|^2\right] ds.$$

Using the Lipschitz continuity of a, b (Assumption *A1*),

$$\begin{aligned}
\mathbb{E}\,|x_t^\epsilon - \tilde{x}_t^\epsilon|^2 &\le c_1 L^2 \mathbb{E}\int_0^t \left(|x_s^\epsilon - x_{t_s}^\epsilon|^2 + |y_s^\epsilon - \tilde{y}_s^\epsilon|^2\right)\,ds \\
&\le c_1 L^2 \int_0^t C_3\,(s - t_s)\,ds + c_1 L^2 T C_4\,\sqrt{\epsilon} \\
&\le c_1\,L^2\,T\sqrt{\epsilon}\,(C_3 + C_4)\,,
\end{aligned}$$

where the bound on $\mathbb{E}|x_s^\epsilon - x_{t_s}^\epsilon|^2$ follows from Lemma 2.4 , and the bound on $\mathbb{E}|y_s^\epsilon - \tilde{y}_s^\epsilon|^2$ follows from Lemma 2.5.

Proof of Lemma 2.9. First, we note that Lemma 2.2 implies that

$$\mathbb{E}|z_0^k|^2 = \mathbb{E}|y_{t_k}^\epsilon|^2 \le C_1.$$

Since ξ_t^k is a stationary version of z_t^k, we also have

$$\mathbb{E}|\xi_t^k|^2 = \lim_{t\to\infty} \mathbb{E}|z_t^k|^2,$$

which is easily found to be bounded by C_1 as well.

Using the first line of (3.2), followed by the dissipativity assumption (Assumption *A2*),

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left|z_t^k - \xi_t^k\right|^2 &= 2\,\mathbb{E}\left(z_t^k - \xi_t^k\right)\cdot\left[f(x_k, z_t^k) - f(x_k, \xi_t^k)\right] + \mathbb{E}\left\|g(x_k, z_t^k) - g(x_k, \xi_t^k)\right\|^2. \\
&\le -2\beta\,\mathbb{E}\left|z_t^k - \xi_t^k\right|^2.
\end{aligned}$$

By Gronwall's inequality (3.1),

$$\mathbb{E}\left|z_t^k - \xi_t^k\right|^2 \le \mathbb{E}\left|z_0^k - \xi_0^k\right|^2 e^{-2\beta t} = 2C_1\,e^{-2\beta t}.$$

Proof of Lemma 2.10. Start with

$$
\begin{aligned}
\mathbb{E}|\tilde{x}_t^\epsilon - \bar{x}_t|^2 &= \mathbb{E}\Big| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(\bar{x}_s))\, ds + \int_0^t (b(x_{t_s}^\epsilon) - b(\bar{x}_s))\, dU_s \Big|^2 \\
&\le 5\,\mathbb{E}\Big| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon))\, ds \Big|^2 \\
&\quad + 5\,\mathbb{E}\Big| \int_0^t (\bar{a}(x_{t_s}^\epsilon) - \bar{a}(\tilde{x}_s^\epsilon))\, ds \Big|^2 + 5\,\mathbb{E}\Big| \int_0^t (b(x_{t_s}^\epsilon) - b(\tilde{x}_s^\epsilon))\, dU_s \Big|^2 \\
&\quad + 5\,\mathbb{E}\Big| \int_0^t (\bar{a}(\tilde{x}_s^\epsilon) - \bar{a}(\bar{x}_s))\, ds \Big|^2 + 5\,\mathbb{E}\Big| \int_0^t (b(\tilde{x}_s^\epsilon) - b(\bar{x}_s))\, dU_s \Big|^2, \quad (3.6)
\end{aligned}
$$

where we have added an subtracted equal terms, and used the Cauchy-Schwarz inequality. Using the Itô isometry for the third and fifth terms on the right hand side, and the Cauchy-Schwartz inequality for the second and fourth terms, we get

$$
\mathbb{E}|\tilde{x}_t^\epsilon - \bar{x}_t|^2 \le 5 I_1 + \frac{5}{2} c_1 (I_2 + I_3), \tag{3.7}
$$

where

$$
\begin{aligned}
I_1 &= \mathbb{E}\Big| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon))\, ds \Big|^2, \\
I_2 &= \int_0^t \mathbb{E}|\bar{a}(x_{t_s}^\epsilon) - \bar{a}(\tilde{x}_s^\epsilon)|^2\, ds + \int_0^t \mathbb{E}\|b(x_{t_s}^\epsilon) - b(\tilde{x}_s^\epsilon)\|^2\, ds, \\
I_3 &= \int_0^t \mathbb{E}|\bar{a}(\tilde{x}_s^\epsilon) - \bar{a}(\bar{x}_s)|^2\, ds + \int_0^t \mathbb{E}\|b(\tilde{x}_s^\epsilon) - b(\bar{x}_s)\|^2\, ds. \quad (3.8)
\end{aligned}
$$

I_3 is readily estimated using the Lipschitz continuity of $\bar{a}, b$,

$$
I_3 \le L^2 \int_0^t \mathbb{E}|\tilde{x}_s^\epsilon - \bar{x}_s|^2\, ds. \tag{3.9}
$$

Similarly, we have for I_2,

$$
\begin{aligned}
I_2 &\le L^2 \int_0^t \mathbb{E}|x_{t_s}^\epsilon - \tilde{x}_s^\epsilon|^2\, ds \\
&\le 2L^2 \left(\int_0^t \mathbb{E}|x_{t_s}^\epsilon - x_s^\epsilon|^2\, ds + \int_0^t \mathbb{E}|x_s^\epsilon - \tilde{x}_s^\epsilon|^2\, ds \right) \\
&\le 2L^2 \left(\int_0^t C_3 (s - t_s)\, ds + \int_0^t C_5 \sqrt{\epsilon}\, ds \right) \\
&\le 2L^2 T \,(C_3 + C_5)\, \sqrt{\epsilon}, \quad (3.10)
\end{aligned}
$$

where we have used Lemmas 2.4 and 2.6.

It remains to estimate I_1, which we decompose as follows,

$$
\begin{aligned}
I_1 &= \mathbb{E}\Big| \int_0^t \left(a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon) \right)\, ds \Big|^2 \\
&\le (\lfloor t/\Delta \rfloor + 1) \sum_{k=0}^{\lfloor t/\Delta \rfloor} \mathbb{E}\Big| \int_{t_k}^{t_{k+1}} \left(a(x_{t_k}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_k}^\epsilon) \right)\, ds \Big|^2 \\
&\le \frac{T^2}{\Delta^2} \max_{k \le T/\Delta} \mathbb{E}\Big| \int_{t_k}^{t_{k+1}} \left(a(x_{t_k}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_k}^\epsilon) \right)\, ds \Big|^2, \quad (3.11)
\end{aligned}
$$

where the time integral has been split into a sum of integrals over time intervals $\Delta = \sqrt{\epsilon}$ (except for the last one which has upper limit t).

Setting as before $x_k = x^\epsilon_{t_k}$, we stretch the time variables by a factor of ϵ, and using the fact that z^k_t is statistically equivalent to $\tilde{y}^\epsilon_{(t-t_k)/\epsilon}$,

$$I_1 \leq T^2 \epsilon \max_{k \leq T/\Delta} I^k_1,$$

where

$$I^k_1 = \mathbb{E}\left| \int_0^{1/\Delta} \big(a(x_k, z^k_s) - \bar{a}(x_k)\big)\, ds \right|^2. \tag{3.12}$$

To estimate I^k_1 we introduce the stationary process ξ^k_t,

$$I^k_1 \leq J^k_1 + J^k_2,$$

where

$$\begin{aligned} J^k_1 &= 2\,\mathbb{E}\left| \int_0^{1/\Delta} \big(a(x_k, z^k_s) - a(x_k, \xi^k_s)\big)\, ds \right|^2, \\ J^k_2 &= 2\,\mathbb{E}\left| \int_0^{1/\Delta} \big(a(x_k, \xi^k_s) - \bar{a}(x_k)\big)\, ds \right|^2. \end{aligned}$$

To estimate J^k_1 we use the Lipschitz continuity of a (Assumption *A1*), and Lemma 2.9,

$$J^k_1 \leq 2L^2 \int_0^{1/\Delta} \mathbb{E}|z^k_s - \xi^k_s|^2\, ds \leq 4L^2 C_1 \int_0^\infty e^{-2\beta s}\, ds = \frac{2L^2}{\beta} C_1. \tag{3.13}$$

To bound J^k_2 we use the fact that ξ^k_t is a stationary process with invariant distribution μ^{x_k}. Thus, $\mathbb{E}a(x_k, \xi^k_t) = \bar{a}(x_k)$, and

$$\begin{aligned} J^k_2 &= 2 \int_0^{1/\Delta} \int_0^{1/\Delta} \left[\mathbb{E}\left(a(x_k, \xi^k_s) \cdot a(x_k, \xi^k_{s'})\right) - \bar{a}(x_k) \cdot \bar{a}(x_k)\right]\, ds\, ds' \\ &\leq 4 \int_0^\infty \int_{s'}^\infty \left[\mathbb{E}\left(a(x_k, \xi^k_{s-s'}) \cdot a(x_k, \xi^k_0)\right) - \bar{a}(x_k) \cdot \bar{a}(x_k)\right]\, ds\, ds' \\ &\leq 8\,\mathbb{E}\,|\bar{a}(x_k)|^2 \int_0^\infty \int_{s'}^\infty \alpha^{1/2}_{s-s'}\, ds\, ds', \end{aligned}$$

where we have used the stationarity of ξ^k_t and Lemma 2.8. Using the bound on the mixing rate, α_t, the fact that $\bar{a}(x)$ is Lipschitz and Lemma 2.3, we get that J^k_2 is bounded. Thus, there exists a constant c_4, such that $I_1 \leq c_4 \epsilon$, and in particular, for ϵ sufficiently small, $I_1 \leq I_2$.

Combining (3.7), (3.10), and (3.9),

$$\mathbb{E}|\tilde{x}^\epsilon_t - \bar{x}_t|^2 \leq 10 c_1 L^2 T (C_3 + C_5)\sqrt{\epsilon} + \frac{5}{2} c_1 L^2 \int_0^t \mathbb{E}|\tilde{x}^\epsilon_s - \bar{x}_s|^2\, ds,$$

which by the integral version of Gronwall's inequality yields the desired result.

4. Analysis of the projective integration scheme

In this section we analyze the convergence of the numerical method defined by eqs. (1.3)–(1.5). Specifically, we derive an estimate for the distance between the computed solution X_n and the solution $\bar{x}_t$ of the effective dynamics at time $t = t_n$. Note that the effective dynamics does not depend on ϵ. Also, since the discrete solution Y_m^n obtained by the micro-solver is for X_n fixed, it only depends on the ratio $\delta t/\epsilon$. Thus, without loss of generality, we may take $\epsilon = 1$. This observation is at the heart of projective integration. The separation of scales is exploited to break the coupled system into two systems that operate each on its separate time scale. The fast variables are solved for fixed values of the slow variables; as a result, there remains a single time scale (which may well be rescaled to one), and the integration time has to be sufficiently long only with respect to the mixing time of the fast dynamics—not with respect to the evolution time of the slow dynamics.

Our projective integration scheme consists of a macro-solver: an Euler-Maruyama time-stepper,

$$X_{n+1} = X_n + A(X_n)\,\Delta t + b(X_n)\,\Delta W_n, \qquad X_0 = x_0, \tag{4.1}$$

where $A(X_n)$ is estimated by an empirical average

$$A(X_n) = \frac{1}{M}\sum_{m=1}^{M} a(X_n, Y_m^n), \tag{4.2}$$

and Y_m^n are numerically generated discrete solutions of the family of SDEs

$$dz_t^n = f(X_n, z_t^n)\,dt + g(X_n, z_t^n)\,dV_t^n, \tag{4.3}$$

with initial conditions $z_0^n = Y_0^n = y_0$, and a time step δt (the choice of a fixed Y_0^n for all n simplifies our estimates; in practice, one could take $Y_0^n = Y_M^{n-1}$ for $n > 0$). Our micro-solver (1.4) is a particular realization that uses an Euler-Maruyama time-stepper as well,

$$Y_{m+1}^n = Y_m^n + f(X_n, Y_m^n)\,\delta t + g(X_n, Y_m^n)\,\Delta V_m^n, \tag{4.4}$$

where $\Delta V_m^n = V_{(m+1)\delta t}^n - V_{m\delta t}^n$ are the Brownian increments associated with the SDEs (4.3). Later on, we will use the auxiliary processes ξ_t^n, which differ from z_t^n by the choice of initial conditions (ξ_t^n is the "stationary version" of z_t^n with invariant measure μ^{X_n}). The estimate

$$\mathbb{E}\,|z_t^n - \xi_t^n|^2 \le 2C_1 e^{-2\beta t}$$

was established in lemma 2.9. With a slight abuse of notation we denote by z_m^n and ξ_m^n the equally distanced samples of $z_{m\delta t}^n$ and $\xi_{m\delta t}^n$.

We also introduce a discrete auxiliary process, $\bar{X}_n$, which is the Euler-Maruyama solution of the effective dynamics (1.2):

$$\bar{X}_{n+1} = \bar{X}_n + \bar{a}(\bar{X}_n)\,\Delta t + b(\bar{X}_n)\,\Delta W_n.$$

As is well-known [27, §10.2.2, §10.6.3], the Euler-Maruyama scheme is of order 1/2, which implies the existence of a constant $K_1 = K_1(T, x_0, y_0)$, such that

$$\sup_{0\le n\le \lfloor T/\Delta t\rfloor} \mathbb{E}|\bar{x}(t_n) - \bar{X}_n|^2 \le K_1\,\Delta t. \tag{4.5}$$

Thus, it remains to estimate the difference between the outcome of the projective integration scheme, X_n, and the numerical solution of the effective dynamics, $\bar{X}_n$, both being discrete-time processes.

The next two lemmas are analogous to Lemmas 2.2 and 2.3.

LEMMA 4.1. *For small enough δt,*

$$\sup_{\substack{0\le n\le \lfloor \frac{T}{\Delta t}\rfloor \\ 0\le m\le M}} \mathbb{E}\,|Y_m^n|^2 \le K_2$$

where

$$K_2 = K_2(y_0) = |y_0|^2 + 2\left(\frac{c_f^2(0)}{\beta^2} + \frac{c_g^2}{\beta}\right).$$

LEMMA 4.2. *For small enough Δt,*

$$\sup_{0\le n\le T/\Delta t} \mathbb{E}|X_n|^2 \le K_3,$$

where

$$K_3 = K_3(T, x_0, y_0) = e^{(1+2K^2)T}\left[|x_0|^2 + 2K^2(1+K_2)\right].$$

LEMMA 4.3. *The mean square deviation between two successive iterations of the microsolver satisfies, for small enough δt,*

$$\sup_{\substack{0\le n\le \lfloor \frac{T}{\Delta t}\rfloor \\ 0\le m\le M}} \mathbb{E}\left|Y_{m+1}^n - Y_m^n\right|^2 \le K_4\,\delta t$$

where $K_4 = 4c_g^2$.

The next lemma establishes the mixing properties of the auxiliary processes z_t^n. Recall that $\bar{a}(X_n)$ is the average of $a(X_n, y)$ with respect to μ^{X_n}, which is the invariant measure induced by the process z_t^n.

LEMMA 4.4. *For small enough δt, there exist a constant K_5 independent of $M, \delta t$ s.t.*

$$\mathbb{E}\Big|\frac{1}{M}\sum_{m=1}^{M} a\left(X_n, z_m^n\right) - \bar{a}(X_n)\Big|^2 \le K_5\left[\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t} + \frac{1}{M}\right]. \tag{4.6}$$

The next lemma establishes the mean deviation between (4.3) and its numerical approximation (4.4).

LEMMA 4.5. *Let z_t^n be the family of processes defined by* (4.3). *For small enough δt, there exists a constant K_8 s.t.*

$$\max_{\substack{0\le n\le \lfloor \frac{T}{\Delta t}\rfloor \\ 0\le m\le M}} \mathbb{E}\,|Y_m^n - z_m^n|^2 \le K_8\sqrt{\delta t}. \tag{4.7}$$

LEMMA 4.6. *There exists a constant* $K_6 = K_6(T, x_0, y_0)$*, such that for all* $0 \le n \le \lfloor T/\Delta t \rfloor$

$$\mathbb{E}\,|\bar{a}(X_n) - A(X_n)|^2 \le K_6 \left(\frac{(\ln M\delta t)^{1/\gamma} + 1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t} \right).$$

LEMMA 4.7. *There exists a constant* $K_7 = K_7(T, x_0, y_0)$ *such that,*

$$\sup_{0 \le n \le \lfloor T/\Delta t \rfloor} \mathbb{E}|X_n - \bar{X}_n|^2 \le K_7 \left(\frac{(\ln M\delta t)^2 + 1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t} \right),$$

where $K_7 = 6T^2 K_6/(1 - 2L^2(2T+1)\,\Delta t)$.

Combining this result with (4.5) our main theorem readily follows:

THEOREM 4.8. *There exists a constant* $K_7 = K_7(T, x_0, y_0)$ *such that,*

$$\sup_{0 \le n \le \lfloor T/\Delta t \rfloor} \mathbb{E}|X_n - \bar{x}(t_n)|^2 \le 2K_1\,\Delta t + 2K_7 \left(\frac{(\ln M\delta t)^2 + 1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t} \right).$$

Note the sources of the various terms: The first term arises from the truncation error of the macro-solver. The second and third terms are the deviation of the ensemble average from the empirical average. The last term is the truncation error of the micro-solver.

5. Proofs for Section 4

Throughout this section we will need a discrete version of Gronwall's inequality. Let Z_n be a sequence of positive numbers which, for small enough δt, satisfy the linear inequality,

$$Z_{n+1} \le (1 + a\,\delta t) Z_n + b\,\delta t,$$

then

$$Z_n \le e^{an\,\delta t} Z_0 + \frac{b}{a}\left(e^{an\,\delta t} - 1\right). \tag{5.1}$$

Proof of Lemma 4.1. Squaring (4.4) and taking expectations,

$$\mathbb{E}\,|Y^n_{m+1}|^2 = \mathbb{E}\,|Y^n_m|^2 + \delta t\,\mathbb{E}\,\|g(X_n, Y^n_m)\|^2 + 2\delta t\,\mathbb{E}Y^n_m \cdot f(X_n, Y^n_m) + \delta t^2\,\mathbb{E}\,|f(X_n, Y^n_m)|^2.$$

Using Assumptions *A2,A4* (cf. Lemma 2.2),

$$\mathbb{E}\left|Y^n_{m+1}\right|^2 \le (1 - \beta\delta t)\,\mathbb{E}\,|Y^n_m|^2 + \delta t \left(\frac{c_f^2(0)}{\beta} + c_g^2 \right) + \delta t^2\,\mathbb{E}\,|f(X_n, Y^n_m)|^2.$$

By Assumption *A2*,

$$|f(X_n, Y^n_m)|^2 \le 2L^2\,|Y^n_m|^2 + 2c_f^2(0), \tag{5.2}$$

which substituted into the last term gives,

$$\mathbb{E}\left|Y_{m+1}^n\right|^2 \le \left(1-\beta\delta t+2L^2\,\delta t^2\right)\mathbb{E}\left|Y_m^n\right|^2+\delta t\left(\frac{c_f^2(0)}{\beta}+c_g^2\right)+2\delta t^2\,c_f^2(0),$$

and the desired result follows from the discrete Gronwall inequality (5.1).

Proof of Lemma 4.2. Squaring (4.1) and taking expectations,

$$\begin{aligned}\mathbb{E}\left|X_{n+1}\right|^2 &= \mathbb{E}\ \left|X_n\right|^2+\frac{2\Delta t}{M}\sum_{m=1}^{M}\mathbb{E}\,X_n\cdot a(X_n,Y_m^n)+\Delta t\mathbb{E}\ \left\|b(X_n)\right\|^2\\ &\quad+\frac{\Delta t^2}{M^2}\mathbb{E}\left|\sum_{m=1}^{M}a(X_n,Y_m^n)\right|^2\\ &\le \mathbb{E}\ \left|X_n\right|^2+\frac{\Delta t}{M}\sum_{m=1}^{M}\mathbb{E}\left|X_n\right|^2+\frac{\Delta t}{M}\sum_{m=1}^{M}\mathbb{E}\left|a(X_n,Y_m^n)\right|^2+\Delta t\,\mathbb{E}\left\|b(X_n)\right\|^2\\ &\quad+\frac{\Delta t^2}{M}\sum_{m=1}^{M}\mathbb{E}\left|a(X_n,Y_m^n)\right|^2.\end{aligned}$$

Using Assumption *A1* and Lemma 4.1, we get, for small enough Δt,

$$\begin{aligned}\mathbb{E}\left|X_{n+1}\right|^2 &\le (1+\Delta t)\mathbb{E}\left|X_n\right|^2+2K^2(1+\mathbb{E}\left|X_n\right|^2)\,\Delta t+2\,K^2\sup_{m\ge 0}\left[\mathbb{E}\left|Y_m^n\right|^2\right]\,\Delta t\\ &\le \left[1+(1+2K^2)\Delta t\right]\mathbb{E}\left|X_n\right|^2+2K^2(1+K_2)\,\Delta t,\end{aligned}$$

and the desired result follows from the discrete Gronwall inequality (5.1).

Proof of Lemma 4.3. Eq. (4.4) together with Assumption *A2* implies,

$$\begin{aligned}\mathbb{E}\left|Y_{m+1}^n-Y_m^n\right|^2 &= 2\,\mathbb{E}\left|f(X_n,Y_m^n)\right|^2\ \delta t^2+2\,\mathbb{E}\left\|g(X_n,Y_m^n)\right\|^2\delta t\\ &\le 4L^2\delta t^2\,\mathbb{E}\left|Y_m^n\right|^2+4c_f^2(0)\delta t^2+2c_g^2\,\delta t,\end{aligned}$$

where we have used (5.2). Lemma 4.1 implies that for δt small enough,

$$\mathbb{E}\left|Y_{m+1}^n-Y_m^n\right|^2 \le 4c_g^2\delta t.$$

Proof of Lemma 4.4. The proof follows the lines of the proof of estimate (3.11). By inserting ξ_m^n, which is the stationary version of z_m^n, we find

$$\begin{aligned}\mathbb{E}\Big|\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,z_m^n\right)-\bar{a}(X_n)\Big|^2 &\le 2\,\mathbb{E}\ \Big|\frac{1}{M}\sum_{m=1}^{M}\left[a(X_n,z_m^n)-a(X_n,\xi_m^n\right]\Big|^2\\ &\quad+2\,\mathbb{E}\Big|\frac{1}{M}\sum_{m=1}^{M}\left[a(X_n,\xi_m^n)-\bar{a}(X_n)\right]\Big|^2\\ &\equiv J_1^n+J_2^n.\end{aligned}$$

To bound J_1^n we use sequentially the Cauchy-Schwarz inequality, the Lipschitz continuity of $a(x,y)$, and Lemma 2.9,

$$\begin{aligned} J_1^n &\le \frac{2}{M}\sum_{m=1}^{M}\mathbb{E}\Big|a(X_n,z_m^n)-a(X_n,\xi_m^n)\Big|^2 \\ &\le \frac{2L^2}{M}\sum_{m=1}^{M}\mathbb{E}\,|z_m^n-\xi_m^n|^2 \\ &\le \frac{2L^2}{M}2C_1\sum_{m=1}^{M}e^{-2\beta m\delta t}. \end{aligned}$$

The last sum is bounded by its corresponding integral,

$$J_1^n \le \frac{2L^2}{M\delta t}2C_1\sum_{m=1}^{M}e^{-2\beta m\delta t}\delta t \le \frac{4L^2}{M\delta t}C_1\frac{1}{2\beta}. \tag{5.3}$$

Next, we bound J_2^n. We use the fact that ξ_t^n is a stationary process with invariant distribution μ^{X_n}. Thus, $\mathbb{E}a(x,\xi_t^n)=\bar{a}(x)$, and so

$$\begin{aligned} J_2^n &= \frac{2}{M^2}\sum_{m=1}^{M}\sum_{p=1}^{M}\left\{\mathbb{E}\left[a(X_n,\xi_m^n)\cdot a(X_n,\xi_p^n)\right]-\bar{a}(X_n)\cdot\bar{a}(X_n)\right\} \\ &= \frac{4}{M^2}\sum_{m=1}^{M}\sum_{p=m+1}^{M}\left\{\mathbb{E}\left[a(X_n,\xi_{m-p}^n)\cdot a(X_n,\xi_0^n)\right]-\bar{a}(X_n)\cdot\bar{a}(X_n)\right\} \\ &\quad+\frac{2}{M^2}\sum_{m=1}^{M}\left\{\mathbb{E}a(X_n,\xi_m^n)\cdot a(X_n,\xi_m^n)-\bar{a}(X_n)\cdot\bar{a}(X_n)\right\}, \end{aligned}$$

where we have used the stationarity of ξ_t^n. The summands on the right hand side are bounded by Lemma 2.7, which establishes the mixing coefficient α_t of the process ξ_t^n, and by Lemma 2.8, with $\zeta=a(X_n,\xi_0^n)$, $\eta=a(X_n,\xi_{m-p}^n)$,

$$\left|\mathbb{E}\left[a(X_n,\xi_{m-p}^n)\cdot a(X_n,\xi_0^n))\right]-\bar{a}(X_n)\cdot\bar{a}(X_n)\right| \le 2\left(\alpha_{(m-p)\delta t}\right)^{1/2}\mathbb{E}|a(X_n,\xi_0^n)|^2.$$

Note, also, that Assumption *A1* and Lemmas 4.1, 4.2, imply that for all small enough $\Delta t>0$,

$$\sup_{n\le T/\Delta t}\mathbb{E}|a(X_n,\xi_0^n)|^2<\infty.$$

Denote $k_1=4\sup_{n\le T/\Delta t}\mathbb{E}|a(X_n,\xi_0^n)|^2$. Hence,

$$J_2^n \le \frac{2k_1}{M^2}\sum_{m=1}^{M}\sum_{p=m+1}^{M}\left(\alpha_{(m-p)\delta t}\right)^{1/2}+\frac{k_1\,\alpha_0^{1/2}}{M}.$$

We split the upper triangular sum into two summands. One is of terms which are near the diagonal, and the second summand is for terms which are far from the diagonal.

Hence,

$$
\begin{aligned}
J_2^n &\le \frac{2k_1}{M^2}\sum_{m=1}^{M}\sum_{p=m+1}^{m+l}\left(\alpha_{(m-p)\delta t}\right)^{1/2} \\
&\quad + \frac{2k_1}{M^2}\sum_{m=1}^{M}\sum_{p=m+l+1}^{M}\left(\alpha_{(m-p)\delta t}\right)^{1/2} + \frac{k_1\,\alpha_0^{1/2}}{M} \\
&\le \frac{2k_1}{M^2}\sum_{m=1}^{M}\sum_{p=m+1}^{m+l} c_2^{1/2} \\
&\quad + \frac{2k_1}{M^2}\sum_{m=1}^{M}\sum_{p=m+l+1}^{M}\left(c_2^{1/2}\exp\left(-\frac{c_3}{2}((l+1)\delta t)^{\gamma}\right)\right) + \frac{k_1\,c_2^{1/2}}{M} \\
&\le \frac{2k_1\,l(c_2)^{1/2}}{M} + 2k_1\,c_2^{1/2}\,\exp(-\frac{c_3}{2}((l+1)\delta t)^{\gamma}) + \frac{k_1\,c_2^{1/2}}{M},
\end{aligned}
$$

where we have used lemma 2.7 three times. Setting $l = \lfloor \frac{1}{\delta t}\left(\frac{2\ln(M\delta t)}{c_3}\right)^{1/\gamma}\rfloor$, there exist constants k_2, k_3, k_4 such that

$$
J_2^n \le k_2 \frac{(\ln M\delta t)^{1/\gamma}}{M\delta t} + k_3\frac{1}{M\delta t} + k_4\frac{1}{M}.
$$

Combined with the bound (5.3) on $J_1(x)$, this concludes the proof.

Proof of Lemma 4.5. Let Y_t^n be the Euler-Maruyama approximation Y_m^n, interpolated continuously by

$$
Y_t^n = \int_0^t f(X_n, Y_{n,t_s})\,ds + \int_0^t g(X_n, Y_{n,t_s})\,dV_{n,s}.
$$

Define

$$
v_t = Y_t^n - z_t^n.
$$

Applying the Itô formula for $\mathbb{E}\,|v_t|^2$,

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\,|v_t|^2 &= 2\mathbb{E}v_t\cdot[f(X_n,Y_{n,t_s}) - f(X_n,z_t^n)] + \mathbb{E}\,\|g(X_n,Y_{n,t_s}) - g(X_n,z_t^n)\|^2 \\
&= 2\mathbb{E}v_t\cdot[f(X_n,Y_{n,t_s}) - f(X_n,Y_t^n)] + 2\mathbb{E}v_t\cdot[f(X_n,Y_t^n) - f(X_n,z_t^n)] \\
&\quad + \mathbb{E}\,\|g(X_n,Y_{n,t_s}) - g(X_n,Y_t^n) + g(X_n,Y_t^n) - g(X_n,z_t^n)\|^2 \\
&\le 2\mathbb{E}v_t\cdot[f(X_n,Y_{n,t_s}) - f(X_n,Y_t^n)] + 2\mathbb{E}v_t\cdot[f(X_n,Y_t^n) - f(X_n,z_t^n)] \\
&\quad + 2\mathbb{E}\,\|g(X_n,Y_{n,t_s}) - g(X_n,Y_t^n)\|^2 + 2\mathbb{E}\,\|g(X_n,Y_t^n) - g(X_n,z_t^n)\|^2
\end{aligned}
$$

Using Assumption *A4*,

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\,|v_t|^2 &\le 2\mathbb{E}v_t\cdot[f(X_n,Y_{n,t_s}) - f(X_n,Y_t^n)] - 2\beta\mathbb{E}\,|v_t|^2 \\
&\quad + 2\mathbb{E}\,\|g(X_n,Y_{n,t_s}) - g(X_n,Y_t^n)\|^2,
\end{aligned}
$$

followed by Assumption *A2*, and Lemma 4.3,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\,|v_t|^2 &\leq 2L\sqrt{\mathbb{E}\,|v_t|^2}\sqrt{\mathbb{E}\,|Y_{n,t_s}-Y_t^n|^2}-2\beta\mathbb{E}\,|v_t|^2+2L^2\mathbb{E}\,|Y_{n,t_s}-Y_t^n|^2\\ &\leq 2L\sqrt{\mathbb{E}\,|v_t|^2}\sqrt{K_3\delta t}-2\beta\mathbb{E}\,|v_t|^2+2L^2K_3\delta t. \end{aligned}\tag{5.4}$$

Let $T_1=\inf\left\{t:\mathbb{E}\,|v_t|^2=1\right\}$. Since $v_0=0$, the line segment $[0,T_1)$ is not empty. We first solve (5.4) up until time T_1, and then show that $T_1=\infty$, which implies that the bound for $\mathbb{E}\,|v_t|^2$ is everywhere true. For $t<T_1$ we have,

$$\frac{d}{dt}\mathbb{E}\,|v_t|^2 \leq 2L\sqrt{K_3\delta t}-2\beta\mathbb{E}\,|v_t|^2+2L^2K_3\delta t.$$

For δt small enough ($\sqrt{\delta t}<\frac{1}{\sqrt{K_3}L}$), the third term on the right hand side is smaller than the first term. Hence

$$\frac{d}{dt}\mathbb{E}\,|v_t|^2 \leq -2\beta\mathbb{E}\,|v_t|^2+4L\sqrt{K_3\delta t}.$$

Gronwall's inequality implies,

$$\mathbb{E}\,|v_t|^2 \leq \frac{2L\sqrt{K_3\delta t}}{\beta}.\tag{5.5}$$

The right hand side of (5.5) can be made smaller than one ($\delta t<\frac{\beta^2}{4L^2K_3}$), hence (5.5) is valid for all t.

Proof of Lemma 4.6. By definition,

$$\begin{aligned}\mathbb{E}\,|\bar{a}(X_n)-A(X_n)|^2 &= \mathbb{E}\left|\int a(X_n,y)\mu^{X_n}(dy)-\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,Y_m^n\right)\right|^2\\ &\leq I_1^n+I_2^n,\end{aligned}\tag{5.6}$$

where

$$\begin{aligned}I_1^n &= 2\mathbb{E}\left|\int a(X_n,y)\mu^{X_n}(dy)-\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,z_m^n\right)\right|^2,\\ I_2^n &= 2\mathbb{E}\left|\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,z_m^n\right)-\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,Y_m^n\right)\right|^2,\end{aligned}$$

where z_t^n is the family of processes defined by (4.3). I_1^n is the difference between the ensemble average of $a(X_n,\cdot)$ with respect to the (exact) invariant measure of z_t^n, and its empirical average over M equidistant sample points. I_2^n is the difference between empirical averages of $a(X_n,\cdot)$ over M equidistant sample points, once for the process z_t^n, and once for its Euler-Maruyama approximation Y_m^n.

The estimation of I_1^n, is given in Lemma 4.4,

$$\begin{aligned}I_1^n &= 2\mathbb{E}\left[\int a(X_n,y)\mu^{X_n}(dy)-\frac{1}{M}\sum_{m=1}^{M}a\left(X_n,z_m^n\right)\right]^2\\ &\leq 2K_5\left[\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t}+\frac{1}{M}\right].\end{aligned}\tag{5.7}$$

We proceed with estimating I_2^n using Assumption *A1*,

$$\begin{aligned}
I_2^n &= 2\mathbb{E}\left|\frac{1}{M}\sum_{m=1}^{M} a\left(X_n, Y_m^n\right) - \frac{1}{M}\sum_{m=1}^{M} a\left(X_n, z_m^n\right)\right|^2 \\
&\le \frac{2}{M^2} M \sum_{m=1}^{M} \mathbb{E}\left|a\left(X_n, Y_m^n\right) - a\left(X_n, z_m^n\right)\right|^2 \\
&\le 2L^2 \max_{m\le M} \mathbb{E}\left|Y_m^n - z_m^n\right|^2 .
\end{aligned}$$

Using Lemma 4.5 with $\delta t < \min(\frac{1}{K_3 L^2}, \frac{\beta^2}{4L^2 K_3})$ we get,

$$I_2^n \le \frac{2L^3\sqrt{K_3\delta t}}{\beta}. \tag{5.8}$$

Combining (5.7) and (5.8),

$$\mathbb{E}\left[\left|\bar{a}(X_n) - A(X_n)\right|^2\right] \le K_6\left[\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t}\right], \tag{5.9}$$

which is uniform in $n \le T/\Delta t$.

Proof of Lemma 4.7. Set $E_n = \mathbb{E}|X_n - \bar{X}_n|^2$, then

$$\begin{aligned}
E_n &= \mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(\bar{X}_i) - A(X_i)\right]\Delta t + \sum_{i=0}^{n}\left[b(\bar{X}_i) - b(X_i)\right]\Delta W_i\right|^2 \\
&\le 2\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(\bar{X}_i) - A(X_i)\right]\Delta t\right|^2 + 2\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[b(\bar{X}_i) - b(X_i)\right]\Delta W_i\right|^2 \\
&\le 4\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(\bar{X}_i) - \bar{a}(X_i)\right]\Delta t\right|^2 + 4\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(X_i) - A(X_i)\right]\Delta t\right|^2 \\
&\quad + 2\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[b(\bar{X}_i) - b(X_i)\right]\Delta W_i\right|^2 .
\end{aligned} \tag{5.10}$$

The first and third sums on the right hand side are easily estimated using the Lipschitz continuity of $\bar{a}$ and b, and the Itô isometry,

$$\begin{aligned}
4\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(\bar{X}_i) - \bar{a}(X_i)\right]\Delta t\right|^2 &\le 4L^2 n \sum_{i=0}^{n-1} E_i\,\Delta t^2 = 4L^2 T \sum_{i=0}^{n-1} E_i\,\Delta t, \\
2\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[b(\bar{X}_i) - b(X_i)\right]\Delta W_i\right|^2 &= 2\sum_{i=0}^{n-1}\mathbb{E}\left\|b(\bar{X}_i) - b(X_i)\right\|^2\Delta t = 2L^2\sum_{i=0}^{n-1} E_i\,\Delta t,
\end{aligned} \tag{5.11}$$

whereas the middle term can be bounded as follows:

$$4\,\mathbb{E}\left|\sum_{i=0}^{n-1}\left[\bar{a}(X_i) - A(X_i)\right]\Delta t\right|^2 \le 4T^2 \max_{i<n}\mathbb{E}\left|\bar{a}(X_i) - A(X_i)\right|^2 . \tag{5.12}$$

Combining eqs. (5.10)–(5.12) and Lemma 4.6, we obtain a discrete linear integral inequality,

$$E_n \le 2L^2(2T+1)\sum_{i=1}^{n-1} E_i\,\Delta t + 4T^2K_6\left(\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t}\right),$$

with initial condition $E_0 = 0$. It follows by a discrete version of Gronwall's inequality that for sufficiently small Δt,

$$\begin{aligned} E_n &\le \left(\sum_{i=0}^{n-1}(6L^2\Delta t)^i\right) 4T^2K_6\left(\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t}\right) \\ &\le \frac{4T^2K_6}{1-2L^2(2T+1)\Delta t}\left(\frac{(\ln M\delta t)^{1/\gamma}+1}{M\delta t} + \frac{1}{M} + \sqrt{\delta t}\right). \end{aligned}$$

This is valid for all $\gamma \in (0,1/2)$. Since the right hand side is a continuous function of γ, this inequality holds also for $\gamma = 1/2$. This estimate is independent of n, which proves the theorem with $K_7 = 4T^2K_6/(1-2L^2(2T+1)\Delta t)$.

6. Discussion

In this paper we proved a strong averaging principle for a system of SDEs in which slow and fast dynamics are driven by Brownian noise; as a result, the limiting dynamics are stochastic as well. Our results thus generalize the analysis of E et al. in which the slow (and effective) dynamics are deterministic. Note that the rate of convergence scales like $\epsilon^{1/4}$, in contrast with the $\epsilon^{1/2}$ rate obtained when the slow dynamics are deterministic.

We then proceeded to show that under the same conditions, a stochastic extension of projective integration schemes strongly converges to the $\epsilon \to 0$ effective dynamics. Our analysis focuses on the simplest case, where both the macro- and micro-solvers use an Euler-Maruyama scheme, but the analysis is easily extended to higher-order schemes.

We have limited ourselves to the case where the diffusion function of the slow dynamics b does not depend on the fast component y. It is easy to see that when $b = b(x,y)$ strong convergence does not hold (although weak convergence does, see [8]). Indeed, take for example the case of $x_t^\epsilon, y_t^\epsilon \in \mathbb{R}$,

$$\begin{aligned} dx_t^\epsilon &= \sin(y_t^\epsilon)\,dU_t, & x_0^\epsilon &= x_0 \\ dy_t^\epsilon &= -\frac{1}{\epsilon}y_t^\epsilon\,dt + \frac{\sqrt{2}}{\sqrt{\epsilon}}\,dV_t, & y_0^\epsilon &= y_0, \end{aligned}$$

where y_t^ϵ is an Ornstein-Uhlenbeck process, independent of x_t^ϵ. If a strong averaging principle were to hold, the effective dynamics could be determined analytically as the invariant distribution of y_t^ϵ is a standard normal distribution,

$$d\bar{x}_t = \gamma\,dU_t,$$

where

$$\gamma = \frac{1}{2\pi}\int \sin^2 y\, e^{-y^2/2}\,dy.$$

However by the Itô isometry,

$$\begin{aligned}\mathbb{E}\left|x_t^\epsilon - \bar{x}_t\right|^2 &= \mathbb{E}\left|\int_0^t (\sin y_s^\epsilon - \gamma)\, dU_s\right|^2 \\ &= \int_0^t \mathbb{E}\left|\sin y_s^\epsilon - \gamma\right|^2\, ds \\ &= \frac{T}{2\pi}\int (\sin y - \gamma)^2 e^{-y^2/2}\, dy,\end{aligned}$$

which is independent of ϵ, i.e.,

$$\lim_{\epsilon\to 0}\mathbb{E}\left|x_t^\epsilon - \bar{x}_t\right|^2 \neq 0.$$

REFERENCES

[1] N. G. van Kampen, *Elimination of fast variables*, Phys. Rep., 124, 69-160, 1985.
[2] R. Z. Khasminskii, *On the averaging principle for Itô stochastic differential equations*, Kybernetica (Prague), 4, 260-279, 1968.
[3] A. V. Skorokhod, *Asymptotics Method in the Theory of Stochastic Differential Equations*, AMS, Providence, RI, 1987.
[4] G. C. Papanicolaou, D. Strook and S. R. S. Varadhan, *Martingale approach to some limit theorem*, Stat. Mech. Dynam. Syst., Duke Univ. Math. Ser., Durham (N.C.), 3, 20, 1977.
[5] A. Yu. Veretennikov, *On the averaging principle for systems of stochastic differential equations*, Math. USSR Sborn., 69, 271-284, 1991.
[6] E. Pardoux and A. Yu. Veretennikov, *On Poisson equation and diffusion approximation II*, Ann. Probab., 31, 1166-1192, 2003.
[7] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.
[8] Y. Kifer, *Stochastic versions of Anosov's and Neistadt's theorems on averaging*, Stoch. Dyn., 1, 1-21, 2001.
[9] G. C. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math., XXVII, 641-668, 1974.
[10] R. Z. Khasminskii, *A limit theorem for the solutions of differential equations with random right-handed sides*, Theory Prob. Applics., 11, 390-406, 1966.
[11] Y. Kifer, *Averaging and climate models*, Prog. Prob., 49, 171-188, 2001.
[12] W. E, D. Liu and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic dynamical equations*, Comm. Pure Appl. Math., 58, 1544-1585, 2005.
[13] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Comm. Math. Sci., 1, 377-384, 2003.
[14] C. W. Gear, I. G. Kevrekidis and C. Theodoropoulos, *"Coarse" integration/bifurcation analysis via microscopic "coarse" integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods*, Comp. Chem. Eng., 26, 941-963, 2002.
[15] C. W. Gear and I. G. Kevrekidis, *Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum*, SIAM J. Sci. Comp, 24, 1091-1106, 2003.
[16] R. Rico-Martinez, C. W. Gear and I. G. Kevrekidis, *Coarse projective KMC integration: forward/reverse initial and boundary value problems*, J. Comp. Phys., 196, 474-489, 2004.
[17] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg and K. Theodoropoulos, *Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks*, Comm. Math. Sci., 1, 715-762, 2003.
[18] I. G. Kevrekidis, C. W. Gear and G. Hummer, *Equation-free: the computer assisted analysis of complex multiscale systems*, AIChE J., 50, 1346-1354, 2004.
[19] G. Hummer and I. G. Kevrekidis, *Coarse molecular dynamics of a peptide fragment: free energy, kinetics and long time dynamics computations*, J. Chem. Phys., 118, 10762-10773, 2003.
[20] C. I. Siettos, M. D. Graham and I. G. Kevrekidis, *Coarse Brownian dynamics for nematic liquid crystals: bifurcation, projective integration, and control via stochastic simulation*, J. Chem. Phys., 118, 10149-10156, 2003.
[21] R. Z. Khasminskii, *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, 1981.

[22] D. Talay, *Second order discretization schemes of stochastic differential systems for the computation of the invariant law*, Stochastics and Stochastic Reports, 29, 13-36, 1990.
[23] S. Meyn and R. L. Tweedie, *Stochastic Stablity of Markov Chains*, Springer-Verlag, New York, 1992.
[24] J. Mattingly, A. M. Stuart and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stoch. Proc. and Applics., 101, 185-232, 2002.
[25] M. N. Malyshkin, *Subexponential estimates of the rate of convergence to the invariant measure for stochastic differential equations*, Theory Probab. Appl., 45, 466-479, 2001.
[26] P. Billingsley, *Convergence of Probability Measures*, John Wiley and Sons, New York, 2nd edition, 1999.
[27] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.