Title
Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems

Journal
Multiscale Modeling & Simulation, 6

ISSN
1540-3459

Author
Givon, Dror

Publication Date
2007

Peer reviewed

Multiscale Model. Simul.
Vol. 6, No. 2, pp. 577–594

STRONG CONVERGENCE RATE FOR TWO-TIME-SCALE JUMP-DIFFUSION STOCHASTIC DIFFERENTIAL SYSTEMS*

DROR GIVON†

Abstract. We study a two-time-scale system of jump-diffusion stochastic differential equations. The main goal is to study the convergence rate of the slow components to the effective dynamics. The convergence established here is in the strong sense, i.e., uniformly in time. For the ergodicity assumptions, we use the existence of a Lyapunov function to control the return times. This assumption is weaker than the one-sided Lipschitz condition, frequently used for deriving rates.

Key words. jump-diffusion processes, multiscale, averaging, mixing, strong convergence, stochastic differential equations

AMS subject classifications. 60H10, 60J75, 34C29, 65C30

DOI. 10.1137/060673345

1. Introduction.

Many problems in the natural sciences give rise to singularly perturbed systems of stochastic differential equations (SDEs). In the past four decades, singularly perturbed systems have been the focus of extensive research within the framework of *averaging methods*. The separation of scales is then taken to advantage to derive a reduced equation, which approximates the slow components. Conditions under which the averaging principle can be applied to this kind of system are well known in the classical literature. However, for numerical purposes the existence of the effective dynamics is not enough, and bounds on the deviation between the slow variables and between the effective dynamics have to be derived. Similar questions, like the existence of the effective dynamics and the rate of convergence to the effective dynamics, for jump-diffusion processes are not yet fully addressed. We consider two-time-scale systems of jump-diffusion SDEs, of the form

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

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

where x_t^ϵ is an n-dimensional jump-diffusion process and y_t^ϵ is an m-dimensional jump-diffusion process. The functions $a(x,y) \in \mathbb{R}^n$ and $f(x,y) \in \mathbb{R}^m$ are the drifts, the functions $b(x) \in \mathbb{R}^{n\times d_1}$ and $g(x,y) \in \mathbb{R}^{m\times d_2}$ are the diffusion coefficients, and the functions $c(x) \in \mathbb{R}^n$ and $h(x,y) \in \mathbb{R}^m$ are the jump coefficients; B_t and W_t are d_1, d_2-dimensional independent Wiener processes, P_t is a scalar simple Poisson process with intensity λ_1, and N_t^ϵ is a scalar simple Poisson process with intensity $\frac{\lambda_2}{\epsilon}$. The parameter ϵ represents the ratio between the natural time scales 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 a case the vector x_t^ϵ is called the "slow component" of the system, and the vector y_t^ϵ is called the "fast component" of the system.

*Received by the editors October 25, 2006; accepted for publication (in revised form) April 30, 2007; published electronically July 18, 2007. This work was supported in part by the National Science Foundation under grant DMS-0410110 and by the Director, Office of Science, Advanced Scientific Computing Research, U.S. Department of Energy under contract DE-AC02-05CH11231.
http://www.siam.org/journals/mms/6-2/67334.html

†Department of Mathematics, University of California, Berkeley, CA 94720 (givon@math.berkeley.edu).

In many cases, one is interested only 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. No computer can deal with such a disparity of scales. Within the framework of *averaging methods*, the separation of scales is taken to advantage to derive, in the limit $\epsilon \to 0$, a reduced equation for an n-dimensional process $\bar{x}_t$, which approximates the slow component x_t^ϵ [1, 2, 3, 4, 5, 6].

A natural generalization of the averaging principle, which will be proved in this paper, is the following. Assume that for every fixed x the rapid variables, governed by (1.1b), induce a unique invariant, ergodic measure $\mu^x(dy)$. Then, as $\epsilon \to 0$, x_t^ϵ converges on every finite interval $[0, T]$ to the solution $\bar{x}_t$ of a closed equation of the form

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

where

$$\bar{a}(x) = \int_{\mathbb{R}^m} a(x, y)\mu^x(dy). \tag{1.3}$$

We start by describing some background for multiscale processes that exhibit discontinuous sample paths. Our framework is a generalization of a homogeneous right-continuous Markov process $(x(t), y(t))$, with $y(t)$ being a step function. In the context of deriving a limit process for Markov processes which exhibit discontinuous sample paths, these processes were the first to be studied. The transition probability for such a process is determined by a collection of operators $\{A_y\}$, where for every y, A_y is the generating operator for the process $x(t)$ on the interval $[0, \tau]$, where τ is the first exit time of the component $y(t)$ from the initial state. These processes, called transport processes, were studied in [2], and the limiting dynamics, where the frequency of the jumps grows to infinity, were described. Similar Markov processes, called processes with rapidly varying discrete component, that have the same fast variables and where the slow variable evolves according to a diffusion were studied in [7].

An extensive study of singularly perturbed switching processes has been made by Yin and Zhang and coworkers [8, 9, 10] (and the references therein). The processes consist of diffusion components and continuous time finite state Markov chains. Their models involve a rapidly varying jump part, a slowly varying jump part, and a slowly varying diffusion part. Another model which they studied was a model in which the diffusions change rapidly in comparison with the jump processes.

In [11] Liu and Yin study a class of hybrid jump diffusions modulated by a finite state Markov process. Their motivation stems from insurance risk models, which include a finite set of regimes and a switching process that dictates which regime to take at any given instance. Once the configuration is determined, the dynamics of the system follow a jump-diffusion process. The fast parameter is the frequency of the change of regimes.

Another research area of jump processes with multiple time scales is the stochastic simulation of kinetic chemical reactions, also known as the Gillespie algorithm [12]. The time evolution is described as follows. A state space of the system is a vector consisting of the number of molecules of each species. The time gap between events is distributed with a Poisson distribution that depends on the state space. The event that takes place is chosen according to a rate function which depends on the state space. This model consists of no drift, only jumps. In the past five years extensive progress has been made in describing the effective dynamics for chemical kinetic systems that take place on vastly different time scales [13, 14, 15, 16].

It is often the case that one is interested not only in the existence of the effective dynamics but rather in the convergence rate of the slow variables to the effective dynamics. This convergence rate has been studied for many models. In the case where the slow variable is described by an ordinary differential equation and the fast variable satisfies an Itô diffusion process, Kifer [17] proved convergence in the sup-norm, and E, Liu, and Vanden-Eijnden [18] derive estimates for the rate of strong (L^1) convergence to the solution of an effective ordinary differential system. In [19] an estimate is given for the rate of mean square (L^2) convergence for the case where both the slow and the fast components are described by an Itô diffusion process.

This paper deals with (1.1). The motivation for such a problem stems from the financial market. In a financial market there are two kinds of securities. One kind is without risk, a bond, and is modeled by a linear ODE. The other kind is a security with risk, a stock. The total change in the stock price is assumed to be the composition of two types of changes [20]:

- The normal variations in price due to a temporary imbalance between supply and demand and other information that causes marginal changes in the stock's value. This component is modeled by a standard Wiener process with a constant variance per unit time and continuous sample paths.
- The "abnormal" variations in price due to the arrival of important new information about the stock that has more than a marginal effect on the price. These perturbations usually occur as finite discontinuities.

Hence the prices per share can be modeled by a diffusion process with jumps, or a stochastic differential equation with jumps (JSDE). The drift coefficient is the instantaneous conditional expected relative change in price per unit time, and the diffusion coefficient is the instantaneous conditional variance per unit time. The jumps, which represent the arrival of new information, occur with a given mean number of arrivals per unit time.

It is often the case that the securities change over more than one time scale. A price of a stock can change in hours or days, while other stocks will change only over a time period of months or years. In [21] the authors study the pricing of defaultable derivatives. In particular, they assume an Ornstein–Uhlenbeck process for the interest rate and a two-factor diffusion model for the intensity of default. They find from empirical evaluation that the time scale of the slow factor is on the order of three months. Empirical evidence of a fast volatility factor (with a characteristic mean-reversion time of a few days) was found in the analysis of high frequency S&P 500 data in [22].

In this paper we analyze systems of the form of (1.1). This paper makes three main contributions. In the case of continuous SDEs, Freidlin and Wentzell [5] prove the existence of the effective dynamics. The convergence proved there is in a very strong form of convergence uniformly in time, i.e.,

$$\mathbb{P}\left\{\sup_{0\le t\le T}|x_t^\epsilon-\bar{x}_t|>\delta\right\}\to 0.$$

In [18, 19] rates were computed under a much weaker convergence form (L^1, L^2). We prove strong convergence of x_t^ϵ to $\bar{x}_t$ under specified conditions. In particular, we obtain an explicit estimate on the rate of the convergence of the form

$$\mathbb{E}\left(\sup_{0\le t\le T}|x_t^\epsilon-\bar{x}_t|^2\right)\le C\,(-\ln\epsilon)^{-1/2}.$$

The second contribution is as follows. Let $Y(t,y)$ denote the solution of the SDE

$$dY(t)=f(Y(t))\,dt+g(Y(t))\,dW_t,\qquad Y(0)=y.$$

The assumption on the fast dynamics which was used in [23, 24, 25, 26, 18, 19] is the one-sided Lipschitz condition,

$$[f(y_1)-f(y_2)]\cdot(y_1-y_2)\le -r\,|y_1-y_2|^2\,,$$

with $r>0$. This assumption is a strong one, and it implies that for any y_1,y_2,

$$\lim_{t\to\infty}\mathbb{E}\,|Y(t,y_1)-Y(t,y_2)|^2=0.$$

In [27] conditions are given to establish geometric convergence and other rates of convergence of a Markov process to its invariant measure, mainly the existence of a Lyapunov function to control the return times. For systems of SDEs these conditions are translated into conditions on the coefficients functions [28], and in [3] similar conditions are used to establish the averaging principle for diffusion processes. These conditions are much weaker than the conditions used in [18, 19]. In this paper the dissipative assumption we use is in the spirit of [27, 28, 3], and it is a sufficient condition to imply the geometric ergodicity of the fast variables. The third contribution is the generalization of the averaging problem, in the strong sense and with a weaker ergodicity assumption, to diffusion processes with jumps. The common technique for deriving mean square bounds for SDEs is the use of the Itô isometry. The isometry identifies stochastic integrals as time integrals, which simplifies the calculations. The jumps are chosen to be driven by a simple Poisson process, because integration over simple Poisson processes admits an Itô isometry.

Note that $b(x)$ and $c(x)$, the noise coefficient functions of the slow component, do not depend on the fast variable. In the case where there is a full coupling in the noise terms, there is no mean square convergence, as will be demonstrated by an example in section 4. In this case one can expect only weak convergence.

The rest of the paper is organized as follows. In section 2 we present our assumptions and theorems. In section 3 we present the proofs. Discussion is given in section 4.

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 strong convergence, i.e., pathwise uniform in time. We achieve this goal by estimating the strong deviation $\mathbb{E}\left(\sup_{0\le t\le T}|x_t^\epsilon-\bar{x}_t|^2\right)$ between the two processes; our main result is Theorem 2.8. 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 A1. The functions $a=a(x,y)$, $b=b(x)$, and $c=c(x)$ in (1.1a) are measurable and Lipschitz continuous and hence have linear growth bounds: specifically, there exist constants L,K such that

$$\begin{aligned}&|a(x_1,y_1)-a(x_2,y_2)|^2+\|b(x_1)-b(x_2)\|^2+|c(x_1)-c(x_2)|^2\\&\le L^2\left(|x_1-x_2|^2+|y_1-y_2|^2\right)\end{aligned}$$

and

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

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

Assumption A2. The functions $f(x,y)$, $g(x,y)$, and $h(x,y)$ 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, and h. 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 $g(x,y)$, $h(x,y)$ are bounded:

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

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

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

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

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

$$y \cdot f(x,y) \leq -\beta\, |y|^2$$

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

Existence and uniqueness of the solutions of (1.1) are guaranteed by Assumptions A1–A2.

The rest of the comments address the fast dynamics described by (1.1b) when x is viewed as a fixed parameter. Assumptions A2 and A4 ensure that the dynamics described by (1.1b) are recurrent, and A4 is called the recurrence condition. Assumption A3 ensures the nondegeneracy of the fast dynamics. Assumption A3, together with the other assumptions, imply the Doob ergodicity of the fast dynamics and hence the existence of a unique invariant probability measure $\mu^x(dy)$ (see [29]). Assumption A3 implies that the ergodic measure of

$$dy_t^\epsilon = \frac{1}{\epsilon} f(x_t^\epsilon, y_t^\epsilon)\, dt + \frac{1}{\sqrt{\epsilon}} g(x_t^\epsilon, y_t^\epsilon)\, dW_t, \qquad y_0^\epsilon = y_0, \tag{2.1}$$

has a smooth density [30, 31]. The relation between the invariant density of (1.1b) and the invariant density of (2.1) is given in [7], and hence the ergodic measure of (1.1b) has smooth density. Since the function a satisfies a Lipschitz condition, so does $\bar{a}$, and the effective dynamics (1.2) has a unique solution.

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 2.1. *The fast component y_t^ϵ satisfies*

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

where $C_1 = C_1(y_0) = |y_0|^2 + \frac{1}{\beta}\left[c_g^2 + \frac{\lambda_2^2 c_h^2}{\beta} + \lambda_2 c_h^2\right]$.

LEMMA 2.2. *The slow component x_t^ϵ satisfies*

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

where

$$C_2 = C_2(T,x_0,y_0) = |x_0|^2 \, e^{\left[1+\lambda_1+(2\lambda_1\vee 1)\,K^2\right]T} + \frac{(2\lambda_1\vee 1)K^2(1+C_1)}{\left[1+\lambda_1+(2\lambda_1\vee 1)\,K^2\right]} e^{\left[1+\lambda_1+(2\lambda_1\vee 1)\,K^2\right]T}.$$

LEMMA 2.3. *For all $0\le t_0\le t\le T$, the mean square displacement of the slow component satisfies*

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

where $C_3 = C_3(T,x_0,y_0) = c_1K^2(1+C_1+C_2)$.

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 $\Delta\in(0,1)$, setting $t_k = k\Delta$, $k=0,\dots,\lfloor\frac{T}{\Delta}\rfloor$; 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 JSDE

$$\begin{aligned} d\tilde{x}_t^\epsilon &= a(x_{t_k}^\epsilon,\tilde{y}_t^\epsilon)\,dt + b(x_{t_k}^\epsilon)\,dB_t + c(x_{t_k}^\epsilon)\,dP_t, && \tilde{x}_{t_k}^\epsilon = x_{t_k}^\epsilon,\\ d\tilde{y}_t^\epsilon &= \frac{1}{\epsilon}f(x_{t_k}^\epsilon,\tilde{y}_t^\epsilon)\,dt + \frac{1}{\sqrt{\epsilon}}g(x_{t_k}^\epsilon,\tilde{y}_t^\epsilon)\,dW_t + h(x_{t_k}^\epsilon,\tilde{y}_t^\epsilon)\,dN_t^\epsilon, && \tilde{y}_{t_k}^\epsilon = y_{t_k}^\epsilon. \end{aligned} \tag{2.2}$$

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,c,f,g,h is replaced by x_t^ϵ at the beginning of the subinterval, $t=t_k$, whereas the fast component $\tilde{y}_t^\epsilon$ is reset to equal y_t^ϵ at each breakpoint t_k. The time interval $\Delta = \epsilon\left[\frac{-\ln\epsilon}{c_2}\right]^{\frac{1}{2}}$ is selected small enough (with c_2 independent of ϵ), $\Delta\ll 1$, 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 kth 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-time scale) 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-time scale) 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 differences between the fast and slow components of the processes $(x_t^\epsilon,y_t^\epsilon)$ and $(\tilde{x}_t^\epsilon,\tilde{y}_t^\epsilon)$. The first lemma estimates mean square difference, while the second lemma estimates strong difference.

LEMMA 2.4. *Let $(x_t^\epsilon,y_t^\epsilon)$ and $(\tilde{x}_t^\epsilon,\tilde{y}_t^\epsilon)$ be the respective solutions of* (1.1) *and* (2.2). *Then*

$$\sup_{0\le t\le T}\mathbb{E}\left|y_t^\epsilon - \tilde{y}_t^\epsilon\right|^2 \le C_4\epsilon\left[-\ln\epsilon\right],$$

where $C_4 = C_3/c_2$, $c_2(T,x_0,y_0) = 8L^2\left[(1+\lambda_2^2)\wedge(1+\lambda_2)\right]$.

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* (2.2). *Then*

$$\mathbb{E}\left(\sup_{0\le t\le T} |x_t^\epsilon - \tilde{x}_t^\epsilon|^2\right) \le C_5\,\epsilon\,(-\ln\epsilon)\,, \tag{2.3}$$

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

Having estimated the strong difference between x_t^ϵ and $\tilde{x}_t^\epsilon$, it remains to estimate the strong 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 JSDE

$$dz_t^k = f(x_k, z_t^k)\,dt + g(x_k, z_t^k)\,dW_t^k + h(x_k, z_t^k)\,dN_t^k, \qquad z_0^k = y_{t_k}^\epsilon, \tag{2.4}$$

where the W_t^k are independent Wiener processes, and N_t^k are independent simple Poisson processes with intensity λ_2.

LEMMA 2.6. *Let* $x(t)$ *be the solution of the equation*

$$dx(t) = \frac{1}{\epsilon} a(x(t))\,dt + \frac{1}{\sqrt{\epsilon}} b(x(t))\,dW(t) + c(x(t))\,dN^\epsilon(t),$$

where $W(t)$ *is a Wiener process and* $N^\epsilon(t)$ *is a simple Poisson process with intensity* λ/ϵ. *Then* $\breve{x}(t) = x(t\epsilon)$ *is a solution of the stochastic equation*

$$d\breve{x}(t) = a(\breve{x}(t))\,dt + b(\breve{x}(t))\,d\breve{W}(t) + c(\breve{x}(t))\,d\breve{N}(t),$$

where $\breve{W}(t) = \frac{W(t/\epsilon)}{\sqrt{\epsilon}}$, *and* $\breve{N}(t)$ *is a simple Poisson process with intensity* λ.

Lemma 2.6 implies that the process z_t^k is statistically equivalent to a shifted and rescaled version of $\tilde{y}_t^\epsilon$, that is, $z_t^k \sim \tilde{y}_{(t-t_k)/\epsilon}^\epsilon$. Menaldi and Robin [29] proved that the dynamics (2.4) is ergodic with invariant measure μ^{x_k} (Assumptions A2–A4) (see also [25]). Moreover, they prove that the process z_t^k satisfies the Doeblin condition, and hence it is exponentially mixing in the following sense. Let $P^{x_k}(t, z, E)$ denote the transition probability of (2.4). Then there are positive constants $\gamma, \alpha < 1$ such that

$$|P^{x_k}(t, z, E) - \mu^{x_k}(E)| \le \gamma\alpha^t$$

for every $E \in \mathcal{B}(\mathbb{R}^m)$.

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

LEMMA 2.7. *For small enough* ϵ,

$$\mathbb{E}\left(\sup_{0\le t\le T} |\tilde{x}_t^\epsilon - \bar{x}_t|^2\right) \le C_6\,\frac{\epsilon}{\Delta},$$

where $C_6 = C_6(T, x_0, y_0) = \frac{T^2 c_4}{c_1 L^2}\exp(18 c_1 L^2 T)$, *and* $c_4 = 2L\sqrt{1 + C_1 + C_2}\gamma$.

Combining Lemma 2.7 with Lemma 2.5 and the fact that for small enough ϵ, $\epsilon\,(-\ln\epsilon) < \frac{\epsilon}{\Delta}$, we obtain our main result.

THEOREM 2.8. *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,

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

3. Proofs for section 2. We start by establishing a number of relations, which will be used repeatedly below (for more details see [32, p. 136] for diffusion processes and [33] for jump-diffusion processes). First, recall Gronwall's inequality: if the real-valued function $v(t)$ satisfies a linear differential inequality of the form

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

then

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

Let $P(t)$ be a simple Poisson process with intensity λ and let $h(P(t), t)$ satisfy the mean square integrability condition on $0 \leq t_0 \leq t$; then the following hold:

- $\mathbb{E}\left[\int_{t_0}^t h(P(s), s)\, dP(s)\right] = \lambda \int_{t_0}^t \mathbb{E}\left[h(P(s), s)\right] ds$.
- Letting $\hat{P}(t) = P(t) - \lambda t$ be the simple mean-zero Poisson process,

$$\mathbb{E}\left[\int_{t_0}^t h(P(s), s)\, d\hat{P}(s)\right] = 0.$$

- The Itô isometry for jump stochastic integrals is

$$\mathbb{E}\left|\int_{t_0}^t h(P(s), s)\, d\hat{P}(s)\right|^2 = \lambda \int_{t_0}^t \mathbb{E}\left|h(P(s), s)\right|^2 ds.$$

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

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

such that a, b, and c are measurable and global Lipschitz continuous (note that this implies the linear growth bound). The assumption on the initial value is that $\mathbb{E}\left|z_{t_0}\right|^2$ is finite and $z(0)$ is independent of $W(t)$, $P(t)$ for all $t \geq 0$.

The Itô stochastic chain rule formula for the process $Z_t = F(z_t, t)$ is

$$\begin{aligned} dZ_t &= dF(z_t, t) \\ &= \left(\partial_t F + a\partial_x F + \frac{1}{2}b^2 \partial_{xx} F\right)(z_t, t)\, dt + (b\partial_x) F(z_t, t)\, dW_t \\ &\quad + \left[F(z_t + c(z_t, t), t) - F(z_t, t)\right] dP_t. \end{aligned} \tag{3.2}$$

Applying the chain rule to $F(z) = |z|^2$, followed by Young's inequality,

$$\begin{aligned} \frac{d}{dt}\mathbb{E}|z_t|^2 &= 2\mathbb{E}\, z_t \cdot a(z_t) + \mathbb{E}\left\|b(z_t)\right\|^2 + 2\lambda \mathbb{E}\, z_t \cdot c(z_t) + \lambda \mathbb{E}\left|c(z_t)\right|^2 \\ &\leq (1+\lambda)\mathbb{E}|z_t|^2 + \mathbb{E}|a(z_t)|^2 + \mathbb{E}\left\|b(z_t)\right\|^2 + 2\lambda \mathbb{E}|c(z_t)|^2. \end{aligned} \tag{3.3}$$

Alternatively, using the definition of the simple mean-zero Poisson process $\hat{P}(t)$, followed by the inequality $(a_1 + \cdots + a_n)^2 \leq n(a_1^2 + \cdots + a_n^2)$,

$$\begin{aligned}
\mathbb{E}|z_t - z_{t_0}|^2 &= \mathbb{E}\left|\int_{t_0}^t a(z_s)\,ds + \int_{t_0}^t b(z_s)\,dW_s + \int_{t_0}^t c(z_s)\,dP_s\right|^2 \\
&= \mathbb{E}\left|\int_{t_0}^t a(z_s)\,ds + \int_{t_0}^t b(z_s)\,dW_s + \int_{t_0}^t c(z_s)\,d\hat{P}_s + \lambda\int_{t_0}^t c(z_s)\,ds\right|^2 \\
&\le 4\,\mathbb{E}\left|\int_{t_0}^t a(z_s)\,ds\right|^2 + 4\,\mathbb{E}\left|\int_{t_0}^t b(z_s)\,dW_s\right|^2 \\
&+ 4\,\mathbb{E}\left|\int_{t_0}^t c(z_s)\,d\hat{P}_s\right|^2 + 4\,\mathbb{E}\left|\lambda\int_{t_0}^t c(z_s)\,ds\right|^2.
\end{aligned}$$

Using the Itô isometry,

$$\begin{aligned}
\mathbb{E}|z_t - z_{t_0}|^2 \le 4\,\mathbb{E}\left|\int_{t_0}^t a(z_s)\,ds\right|^2 + 4\,\mathbb{E}\int_{t_0}^t \|b(z_s)\|^2\,ds \\
+ 4\lambda\,\mathbb{E}\int_{t_0}^t |c(z_s)|^2\,ds + 4\lambda^2\,\mathbb{E}\left|\int_{t_0}^t c(z_s)\,ds\right|^2.
\end{aligned}$$

Using now the Cauchy–Schwarz inequality,

$$\begin{aligned}
\mathbb{E}|z_t - z_{t_0}|^2 &\le 4(t-t_0)\int_{t_0}^t \mathbb{E}|a(z_s)|^2\,ds + 4\,\mathbb{E}\int_{t_0}^t \|b(z_s)\|^2\,ds \\
&+ 4\lambda\,\mathbb{E}\int_{t_0}^t |c(z_s)|^2\,ds + 4\lambda^2(t-t_0)\,\mathbb{E}\int_{t_0}^t |c(z_s)|^2\,ds \\
&\le c_1\int_{t_0}^t \mathbb{E}\left[|a(z_s)|^2 + \|b(z_s)\|^2 + |c(z_s)|^2\right]ds,
\end{aligned} \tag{3.4}$$

where $c_1 = c_1(\lambda, T) = 4\max(T, 1, \lambda, \lambda^2 T)$.

Finally, recall the Doob inequality for martingales $M(t) = W(t), \hat{P}(t)$,

$$\mathbb{E}\left(\sup_{0\le t\le T}\left|\int_0^t z_s\,dM_s\right|^2\right) \le 4\left|\int_0^T z_s\,dM_s\right|^2.$$

Proof of Lemma 2.1. Applying the first line of (3.3) to y_t^ϵ, we obtain

$$\begin{aligned}
\epsilon\frac{d}{dt}\mathbb{E}|y_t^\epsilon|^2 &= 2\mathbb{E}\,y_t^\epsilon\cdot f(x_t^\epsilon, y_t^\epsilon) + \mathbb{E}\,|g(x_t^\epsilon, y_t^\epsilon)|^2 \\
&+ 2\lambda_2\mathbb{E}\,y_t^\epsilon\cdot h(x_t^\epsilon, y_t^\epsilon) + \lambda_2\mathbb{E}\,|h(x_t^\epsilon, y_t^\epsilon)|^2.
\end{aligned} \tag{3.5}$$

Assumption A4 with $y = y_t^\epsilon$ gives

$$y_t^\epsilon\cdot f(x_t^\epsilon, y_t^\epsilon) \le -\beta|y_t^\epsilon|^2, \tag{3.6}$$

which gives us a bound for the left term on the right-hand side of (3.5). For the third term on the right-hand side of (3.5) we use Young's inequality $2p\cdot q \le \beta|p|^2 + \frac{1}{\beta}|q|^2$, with $p = y_t^\epsilon$ and $q = \lambda_2\,h(x_t^\epsilon, y_t^\epsilon)$:

$$2\lambda_2\,y_t^\epsilon\cdot h(x_t^\epsilon, y_t^\epsilon) \le \beta\,|y_t^\epsilon|^2 + \frac{\lambda_2^2}{\beta}\,|h(x_t^\epsilon, y_t^\epsilon)|^2. \tag{3.7}$$

Substituting into (3.5), the bound on g, h (Assumption A2) with (3.6), (3.7) yields the differential inequality

$$\epsilon \frac{d}{dt}\mathbb{E}|y_t^\epsilon|^2 \le -2\beta\,\mathbb{E}|y_t^\epsilon|^2 + c_g^2 + \beta\mathbb{E}|y_t^\epsilon|^2 + \frac{\lambda_2^2 c_h^2}{\beta} + \lambda_2 c_h^2$$
$$= -\beta\mathbb{E}|y_t^\epsilon|^2 + \left[c_g^2 + \frac{\lambda_2^2 c_h^2}{\beta} + \lambda_2 c_h^2\right].$$

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

Proof of Lemma 2.2. Applying (3.3) to x_t^ϵ,

$$\frac{d}{dt}\mathbb{E}|x_t^\epsilon|^2 \le (1+\lambda_1)\mathbb{E}|x_t^\epsilon|^2 + \mathbb{E}|a(x_t^\epsilon, y_t^\epsilon)|^2 + \mathbb{E}\,\|b(x_t^\epsilon)\|^2 + 2\lambda_1\mathbb{E}|c(x_t^\epsilon)|^2.$$

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

$$\frac{d}{dt}\mathbb{E}|x_t^\epsilon|^2 \le (1+\lambda_1)\mathbb{E}|x_t^\epsilon|^2 + (2\lambda_1 \vee 1)\,K^2(1+\mathbb{E}|x_t^\epsilon|^2 + \mathbb{E}|y_t^\epsilon|^2)$$
$$\le \left[1+\lambda_1 + (2\lambda_1 \vee 1)\,K^2\right]\mathbb{E}|x_t^\epsilon|^2 + (2\lambda_1 \vee 1)K^2(1+C_1),$$

where the last inequality follows from Lemma 2.1. The desired result follows from Gronwall's inequality (3.1). ∎

Proof of Lemma 2.3. Inequality (3.4) for x_t^ϵ reads

$$\mathbb{E}|x_t^\epsilon - x_{t_0}^\epsilon|^2 \le c_1(\lambda_1)\int_{t_0}^t \mathbb{E}\left[|a(x_s^\epsilon, y_s^\epsilon)|^2 + \|b(x_s^\epsilon)\|^2 + |c(x_s^\epsilon)|^2\right] ds.$$

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

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

where the last inequality follows from Lemmas 2.1 and 2.2. ∎

Proof of Lemma 2.4. 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 breakpoints t_k implies that $z_{t_k} = 0$ for all k.

Using the first line of (3.4) for the real-valued process $|z_t|^2$,

$$\mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2 \le 4(t-t_k)\frac{1}{\epsilon^2}\int_{t_k}^t \mathbb{E}\left|\left(f(x_t^\epsilon, y_t^\epsilon) - f(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\right|^2 ds$$
$$+ \frac{4}{\epsilon}\,\mathbb{E}\int_{t_k}^t \left|\left(g(x_t^\epsilon, y_t^\epsilon) - g(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\right|^2 ds$$
$$+ 4\frac{\lambda_2}{\epsilon}\,\mathbb{E}\int_{t_k}^t \left|\left(h(x_t^\epsilon, y_t^\epsilon) - h(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\right|^2 ds$$
$$+ 4\frac{\lambda_2^2}{\epsilon^2}(t-t_k)\,\mathbb{E}\int_{t_k}^t \left|\left(h(x_t^\epsilon, y_t^\epsilon) - h(x_{t_k}^\epsilon, \tilde{y}_t^\epsilon)\right)\right|^2 ds.$$

Using Assumption A2,

$$
\begin{aligned}
\mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2 &\le 4(t-t_k)\frac{L^2}{\epsilon^2}\,\mathbb{E}\int_{t_k}^t |x_t^\epsilon - x_{t_k}^\epsilon|^2 + |y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds \\
&\quad + 4\frac{L^2}{\epsilon}\,\mathbb{E}\int_{t_k}^t |x_t^\epsilon - x_{t_k}^\epsilon|^2 + |y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds \\
&\quad + 4\frac{\lambda_2}{\epsilon}L^2\,\mathbb{E}\int_{t_k}^t |x_t^\epsilon - x_{t_k}^\epsilon|^2 + |y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds \\
&\quad + 4\frac{\lambda_2^2}{\epsilon^2}L^2(t-t_k)\,\mathbb{E}\int_{t_k}^t |x_t^\epsilon - x_{t_k}^\epsilon|^2 + |y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds \\
&\le 4L^2\left[\frac{t-t_k}{\epsilon^2}(1+\lambda_2^2) + \frac{1+\lambda_2}{\epsilon}\right]\mathbb{E}\int_{t_k}^t |x_t^\epsilon - x_{t_k}^\epsilon|^2 + |y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds.
\end{aligned}
$$

Using Lemma 2.3,

$$
\begin{aligned}
\mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2 &\le 4L^2\left[\frac{t-t_k}{\epsilon^2}(1+\lambda_2^2) + \frac{1+\lambda_2}{\epsilon}\right]\int_{t_k}^t C_3(t-t_k) + \mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds \\
&\le 4L^2\left[\frac{\Delta}{\epsilon^2}(1+\lambda_2^2) + \frac{1+\lambda_2}{\epsilon}\right]C_3\Delta^2 \\
&\quad + 4L^2\left[\frac{\Delta}{\epsilon^2}(1+\lambda_2^2) + \frac{1+\lambda_2}{\epsilon}\right]\int_{t_k}^t \mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2\,ds.
\end{aligned}
$$

Applying Gronwall's inequality (3.1) upon integrating from t_k to t,

$$
\mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2 \le C_3\Delta^2\, e^{4L^2\left[\frac{\Delta}{\epsilon^2}(1+\lambda_2^2)+\frac{1+\lambda_2}{\epsilon}\right]\Delta}.
$$

Set $c_2 = 8L^2\left[(1+\lambda_2^2)\vee(1+\lambda_2)\right]$. Using the definition of $\Delta = \epsilon\left[\frac{-\ln\epsilon}{c_2}\right]^{\frac{1}{2}}$,

$$
\begin{aligned}
\mathbb{E}|y_t^\epsilon - \tilde{y}_t^\epsilon|^2 &\le C_3\Delta^2\, e^{c_2\frac{\Delta^2}{\epsilon^2}} \\
&= C_3\epsilon\left[-\frac{\ln\epsilon}{c_2}\right]. \qquad \square
\end{aligned}
$$

Proof of Lemma 2.5. By (3.4) with $z_t = x_t^\epsilon - \tilde{x}_t^\epsilon$,

$$
\begin{aligned}
&\mathbb{E}\sup_{t\in[0,T]}|x_t^\epsilon - \tilde{x}_t^\epsilon|^2 \le 3\mathbb{E}\sup_{t\in[0,T]}\left|\int_0^t a(x_s^\epsilon, y_s^\epsilon) - a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon)\,ds\right|^2 \\
&+ 3\mathbb{E}\sup_{t\in[0,T]}\left\|\int_0^t b(x_s^\epsilon) - b(x_{t_s}^\epsilon)\,dB_s\right\|^2 + 3\mathbb{E}\sup_{t\in[0,T]}\left|\int_0^t c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\,dP_s\right|^2 \\
&\le 3\mathbb{E}\sup_{t\in[0,T]}\left|\int_0^t a(x_s^\epsilon, y_s^\epsilon) - a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon)\,ds\right|^2 + 3\mathbb{E}\sup_{t\in[0,T]}\left\|\int_0^t b(x_s^\epsilon) - b(x_{t_s}^\epsilon)\,dB_s\right\|^2 \\
&+ 6\mathbb{E}\sup_{t\in[0,T]}\left|\int_0^t c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\,d\hat{P}_s\right|^2 + 6\mathbb{E}\sup_{t\in[0,T]}\left|\lambda_1\int_0^t c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\,ds\right|^2.
\end{aligned}
$$

Using the Doob inequality for the two martingale integrals and the Cauchy–Schwarz inequality for the time integrals,

$$\begin{aligned}
&\mathbb{E} \sup_{t\in[0,T]} |x_t^\epsilon - \tilde{x}_t^\epsilon|^2 \\
&\le 3T\mathbb{E}\int_0^T \left|a(x_s^\epsilon, y_s^\epsilon) - a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon)\right|^2 ds + 12\mathbb{E}\left\|\int_0^T b(x_s^\epsilon) - b(x_{t_s}^\epsilon)\, dB_s\right\|^2 \\
&+ 24\mathbb{E}\left|\int_0^T c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\, d\hat{P}_s\right|^2 + 6T\lambda_1^2\mathbb{E}\int_0^T \left|c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\right|^2 ds.
\end{aligned}$$

Using the Itô isometry,

$$\begin{aligned}
&\mathbb{E} \sup_{t\in[0,T]} |x_t^\epsilon - \tilde{x}_t^\epsilon|^2 \\
&\le 3T\mathbb{E}\int_0^T \left|a(x_s^\epsilon, y_s^\epsilon) - a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon)\right|^2 ds + 12\mathbb{E}\int_0^T \left\|b(x_s^\epsilon) - b(x_{t_s}^\epsilon)\right\|^2 ds \\
&+ 24\mathbb{E}\int_0^T \left|c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\right|^2 ds + 6T\lambda_1^2\mathbb{E}\int_0^T \left|c(x_s^\epsilon) - c(x_{t_s}^\epsilon)\right|^2 ds.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E} \sup_{t\in[0,T]} |x_t^\epsilon - \tilde{x}_t^\epsilon|^2 &\le (3T + 36 + 6T\lambda_1^2)L^2\mathbb{E}\int_0^T \left(|x_s^\epsilon - x_{t_s}^\epsilon|^2 + |y_s^\epsilon - \tilde{y}_s^\epsilon|^2\right) ds \\
&\le (3T + 36 + 6T\lambda_1^2)L^2\left[\int_0^T C_3\,(s - t_s)\, ds + TC_4\epsilon\,(-\ln\epsilon)\right] \\
&\le (3T + 36 + 6T\lambda_1^2)\, L^2\, T\,[C_3\Delta + C_4\epsilon\,(-\ln\epsilon)] \\
&\le (3T + 36 + 6T\lambda_1^2)\, L^2\, T(C_3 + C_4)\epsilon\,[-\ln\epsilon]\,,
\end{aligned}$$

where the bound on $\mathbb{E}|x_s^\epsilon - x_{t_s}^\epsilon|^2$ follows from Lemma 2.3, and the bound on $\mathbb{E}|y_s^\epsilon - \tilde{y}_s^\epsilon|^2$ follows from Lemma 2.4. ∎

Proof of Lemma 2.6.

$$\begin{aligned}
\breve{x}(t) - \breve{x}(s) = x(t\epsilon) - x(s\epsilon) &= \frac{1}{\epsilon}\int_{s\epsilon}^{t\epsilon} a(x(u))\, du + \frac{1}{\sqrt{\epsilon}}\int_{s\epsilon}^{t\epsilon} b(x(u))\, dW(u) \\
&+ \int_{s\epsilon}^{t\epsilon} c(x(u))\, dN(u) \\
&= \frac{1}{\epsilon}\int_s^t a(x(u\epsilon))\, d(u\epsilon) + \frac{1}{\sqrt{\epsilon}}\int_s^t b(x(u\epsilon))\, dW(u\epsilon) \\
&+ \int_s^t c(x(u\epsilon))\, dN(u\epsilon) \\
&= \int_s^t a(\breve{x}(u))\, du + \int_s^t b(\breve{x}(u))\, d\breve{W}(u) \\
&+ \int_s^t c(\breve{x}(u))\, d\breve{N}(u),
\end{aligned}$$

where we used the facts that $\breve{W}(t) = \frac{W(\epsilon t)}{\sqrt{\epsilon}}$ is a Wiener process and $\breve{N}(t) = N(\epsilon t)$ is a simple Poisson process with intensity $\lambda = \epsilon\frac{\lambda}{\epsilon}$. ∎

Proof of Lemma 2.7. For any $0 \le T_1 \le T$,

$$\begin{aligned}
\mathbb{E} \sup_{t\in[0,T_1]} |\tilde{x}_t^\epsilon - \bar{x}_t|^2 &= \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(\bar{x}_s))\, ds + \int_0^t (b(x_{t_s}^\epsilon) - b(\bar{x}_s))\, dB_s \right. \\
&\qquad \left. + \int_0^t (c(x_{t_s}^\epsilon) - c(\bar{x}_s))\, dP_s \right|^2 \\
&\le 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon))\, ds \right|^2 + 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (\bar{a}(x_{t_s}^\epsilon) - \bar{a}(\tilde{x}_s^\epsilon))\, ds \right|^2 \\
&+ 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (\bar{a}(\tilde{x}_s^\epsilon) - \bar{a}(\bar{x}_s))\, ds \right|^2 + 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (b(x_{t_s}^\epsilon) - b(\tilde{x}_s^\epsilon))\, dB_s \right|^2 \\
&+ 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (b(\tilde{x}_s^\epsilon) - b(\bar{x}_s))\, dB_s \right|^2 + 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(x_{t_s}^\epsilon) - c(\tilde{x}_s^\epsilon))\, d\hat{P}_s \right|^2 \\
&+ 9\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(\tilde{x}_s^\epsilon) - c(\bar{x}_s))\, d\hat{P}_s \right|^2 + 9\lambda_1^2\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(x_{t_s}^\epsilon) - c(\tilde{x}_s^\epsilon))\, ds \right|^2 \\
&+ 9\lambda_1^2\mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(\tilde{x}_s^\epsilon) - c(\bar{x}_s))\, ds \right|^2,
\end{aligned}$$

where we have added and subtracted equal terms. Using the Doob inequality for the martingale integrals and then the Itô isometry,

$$\begin{aligned}
\mathbb{E} \sup_{t\in[0,T_1]} |\tilde{x}_t^\epsilon - \bar{x}_t|^2 &\le 9\, \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon))\, ds \right|^2 \\
&+ 9\, \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (\bar{a}(x_{t_s}^\epsilon) - \bar{a}(\tilde{x}_s^\epsilon))\, ds \right|^2 + 9\, \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (\bar{a}(\tilde{x}_s^\epsilon) - \bar{a}(\bar{x}_s))\, ds \right|^2 \\
&+ 36 \int_0^{T_1} \mathbb{E} \left\| b(x_{t_s}^\epsilon) - b(\tilde{x}_s^\epsilon) \right\|^2 ds + 36 \int_0^{T_1} \mathbb{E} \left\| b(\tilde{x}_s^\epsilon) - b(\bar{x}_s) \right\|^2 ds \\
&+ 36\lambda_1 \int_0^{T_1} \mathbb{E} \left| c(x_{t_s}^\epsilon) - c(\tilde{x}_s^\epsilon) \right|^2 ds + 36\lambda_1 \int_0^{T_1} \mathbb{E} \left| c(\tilde{x}_s^\epsilon) - c(\bar{x}_s) \right|^2 ds \\
&+ 9\lambda_1^2\, \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(x_{t_s}^\epsilon) - c(\tilde{x}_s^\epsilon))\, ds \right|^2 + 9\lambda_1^2\, \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t (c(\tilde{x}_s^\epsilon) - c(\bar{x}_s))\, ds \right|^2.
\end{aligned}$$

Using now the Cauchy–Schwarz inequality, we get

$$\mathbb{E} \sup_{t\in[0,T_1]} |\tilde{x}_t^\epsilon - \bar{x}_t|^2 \le 9I_1 + 9c_1(T_1)(I_2 + I_3), \tag{3.8}$$

where

$$\begin{aligned}
I_1 &= \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^{T_1} (a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon))\, ds \right|^2, \\
I_2 &= \int_0^{T_1} \mathbb{E}|\bar{a}(x_{t_s}^\epsilon) - \bar{a}(\tilde{x}_s^\epsilon)|^2\, ds + \int_0^{T_1} \mathbb{E}\|b(x_{t_s}^\epsilon) - b(\tilde{x}_s^\epsilon)\|^2\, ds \\
&\quad + 2\int_0^{T_1} \mathbb{E}|c(x_{t_s}^\epsilon) - c(\tilde{x}_s^\epsilon)|^2\, ds,
\end{aligned} \tag{3.9}$$

$$I_3 = \int_0^{T_1} \mathbb{E}|\bar{a}(\tilde{x}_s^\epsilon) - \bar{a}(\bar{x}_s)|^2\, ds + \int_0^{T_1} \mathbb{E}\|b(\tilde{x}_s^\epsilon) - b(\bar{x}_s)\|^2\, ds$$
$$+ 2\int_0^{T_1} \mathbb{E}|c(\tilde{x}_s^\epsilon) - c(\bar{x}_s)|^2\, ds.$$

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

$$\begin{aligned} I_3 &\le 2L^2 \int_0^{T_1} \mathbb{E}\left|\tilde{x}_s^\epsilon - \bar{x}_s\right|^2 ds \\ &\le 2L^2 \int_0^{T_1} \mathbb{E} \sup_{r\in[0,s]} \left|\tilde{x}_r^\epsilon - \bar{x}_r\right|^2 ds. \end{aligned} \tag{3.10}$$

Similarly, we have for I_2,

$$\begin{aligned} I_2 &\le 2L^2 \int_0^{T_1} \mathbb{E}|x_{t_s}^\epsilon - \tilde{x}_s^\epsilon|^2\, ds \\ &\le 4L^2 \left(\int_0^{T_1} \mathbb{E}|x_{t_s}^\epsilon - x_s^\epsilon|^2\, ds + \int_0^{T_1} \mathbb{E}|x_s^\epsilon - \tilde{x}_s^\epsilon|^2\, ds \right) \\ &\le 4L^2 \left(\int_0^{T_1} C_3(s - t_s)\, ds + \int_0^{T_1} C_5\epsilon\,(-\ln\epsilon)\, ds \right) \\ &\le 4L^2 T\,(C_3 + C_5)\,\epsilon\,(-\ln\epsilon)\,, \end{aligned} \tag{3.11}$$

where we have used Lemmas 2.3 and 2.5.

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

$$\begin{aligned} I_1 &= \mathbb{E} \sup_{t\in[0,T_1]} \left| \int_0^t \left(a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon) \right) ds \right|^2 \\ &= \mathbb{E} \max_{0\le l\le \lfloor T_1/\Delta \rfloor} \left| \sum_{k=0}^{l} \int_{t_k}^{t_{k+1}} \left(a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon) \right) ds \right|^2 \\ &\le \mathbb{E} \max_{0\le l\le \lfloor T_1/\Delta \rfloor} (l+1) \sum_{k=0}^{l} \left| \int_{t_k}^{t_{k+1}} \left(a(x_{t_s}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_s}^\epsilon) \right) ds \right|^2 \\ &\le (\lfloor T_1/\Delta \rfloor + 1) \sum_{k=0}^{\lfloor T_1/\Delta \rfloor} \mathbb{E} \left| \int_{t_k}^{t_{k+1}} \left(a(x_{t_k}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_k}^\epsilon) \right) ds \right|^2 \\ &\le \frac{{T_1}^2}{\Delta^2} \max_{k\le T_1/\Delta} \mathbb{E} \left| \int_{t_k}^{t_{k+1}} \left(a(x_{t_k}^\epsilon, \tilde{y}_s^\epsilon) - \bar{a}(x_{t_k}^\epsilon) \right) ds \right|^2, \end{aligned} \tag{3.12}$$

where the time integral has been split into a sum of integrals over time intervals Δ (except for the last one, which has upper limit t).

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

$$I_1 \le \frac{{T_1}^2}{\Delta^2} \epsilon^2 \max_{k\le T_1/\Delta} I_1^k,$$

where

$$I_1^k = \mathbb{E}\left|\int_0^{\Delta/\epsilon} \big(a(x_k, z_s^k) - \bar{a}(x_k)\big)\, ds\right|^2. \tag{3.13}$$

To bound I_1^k,

$$\begin{aligned} I_1^k &= \int_0^{\Delta/\epsilon}\int_0^{\Delta/\epsilon} \mathbb{E}\left\{\left[a(x_k, z_s^k) - \bar{a}(x_k)\right]\cdot\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\right\} ds\, ds' \\ &= 2\int_0^{\Delta/\epsilon}\int_{s'}^{\Delta/\epsilon} \mathbb{E}\left\{\left[a(x_k, z_s^k) - \bar{a}(x_k)\right]\cdot\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\right\} ds\, ds'. \end{aligned} \tag{3.14}$$

We estimate the integrand using the Cauchy–Schwarz inequality:

$$\begin{aligned} &\mathbb{E}\left\{\left[a(x_k, z_s^k) - \bar{a}(x_k)\right]\cdot\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\right\} \\ &= \mathbb{E}\left\{\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\cdot\mathbb{E}\left[a(x_k, z_s^k) - \bar{a}(x_k)|z_{s'}^k\right]\right\} \\ &= \mathbb{E}\left\{\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\cdot\mathbb{E}_{z_{s'}^k}\left[a(x_k, z_{s-s'}^k) - \bar{a}(x_k)\right]\right\} \\ &\le \sup_{s'}\left\{\mathbb{E}\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]^2\right\}^{\frac{1}{2}}\left\{\mathbb{E}\left[\mathbb{E}_{z_{s'}^k}\left[a(x_k, z_{s-s'}^k) - \bar{a}(x_k)\right]\right]^2\right\}^{\frac{1}{2}}. \end{aligned}$$

For the left-hand term we use the linear growth bound of the functions $a, \bar{a}$ and Lemmas 2.1 and 2.2:

$$\begin{aligned} \mathbb{E}\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]^2 &\le 2\mathbb{E}\left|a(x_k, z_{s'}^k)\right|^2 + 2\mathbb{E}\left|\bar{a}(x_k)\right|^2 \\ &\le 4L^2(1 + C_1 + C_2). \end{aligned}$$

Combining this with the bound on the mixing rate, $\gamma\alpha^t$, we get

$$\mathbb{E}\left\{\left[a(x_k, z_s^k) - \bar{a}(x_k)\right]\cdot\left[a(x_k, z_{s'}^k) - \bar{a}(x_k)\right]\right\} \le 2L\sqrt{1 + C_1 + C_2}\,\gamma\alpha^{s-s'}.$$

Inserting back into (3.14),

$$I_1^k \le 4L\sqrt{1 + C_1 + C_2}\,\gamma\int_0^{\Delta/\epsilon}\int_{s'}^{\Delta/\epsilon} \alpha^{s-s'}\, ds\, ds'.$$

Thus, there exists a constant $c_4 = 2L\sqrt{1 + C_1 + C_2}\,\gamma$ such that

$$I_1^k \le c_4\,\frac{\Delta}{\epsilon}.$$

Hence,

$$I_1 \le \frac{T^2}{\Delta^2}\epsilon^2 c_4\,\frac{\Delta}{\epsilon} = T^2 c_4\frac{\epsilon}{\Delta}. \tag{3.15}$$

Combining (3.8), (3.15), (3.11), and (3.10),

$$\begin{aligned} \mathbb{E}\sup_{t\in[0,T_1]}\left|\tilde{x}_t^\epsilon - \bar{x}_t\right|^2 &\le 9\,T^2 c_4\,\frac{\epsilon}{\Delta} + 9c_1 4L^2 T(C_3 + C_5)\epsilon\,(-\ln\epsilon) \\ &\quad + 18c_1L^2\int_0^{T_1}\mathbb{E}\sup_{r\in[0,s]}\left|\tilde{x}_r^\epsilon - \bar{x}_r\right|^2 ds, \end{aligned}$$

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

4. Discussion. In this paper we proved a strong averaging principle for a system of JSDEs in which slow and fast dynamics are driven both by Brownian noise and by Poisson noise; as a result, the limiting dynamics are stochastic as well. Our results thus generalize the analysis of [18, 19] in which only mean square convergence is shown, and the noise in the original (and in the effective) dynamics is driven only by Brownian noise and hence have continuous sample paths. We were also able to prove that the strong averaging principle holds under dissipative assumptions which are given in [27, 28, 3]. These assumptions are much weaker than those used in [18, 19].

Note that the rate of convergence scales like $(-\ln\epsilon)^{-1/2}$. As noted in the introduction the slow/fast time-scale ratio in the financial markets is months/days, which suggests that if the logarithmic rate is optimal, then the averaging might not be so applicable. However, when we used the strong assumption of ergodicity, i.e., the one-sided Lipschitz condition, we were able to obtain an algebraic convergence rate of $\epsilon^{1/4}$.

We have limited ourselves to the case where the noise coefficients of the slow dynamics do not depend on the fast component, that is, $b(x,y)=b(x)$ and $c(x,y)=c(x)$. In [19] a simple example was constructed to show that for systems of SDEs (i.e., $c=0$, $h=0$), strong convergence does not hold when $b=b(x,y)$. We use a similar example here to show that strong convergence does not hold when $c=c(x,y)$. Indeed, take, for example, the case of $x_t^\epsilon, y_t^\epsilon \in \mathbb{R}$,

$$\begin{aligned} dx_t^\epsilon &= \sin(y_t^\epsilon)\, d\hat{P}_t, & x_0^\epsilon &= x_0,\\ dy_t^\epsilon &= -\frac{1}{\epsilon} y_t^\epsilon\, dt + \frac{\sqrt{2}}{\sqrt{\epsilon}}\, dW_t, & y_0^\epsilon &= y_0, \end{aligned}$$

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

$$d\bar{x}_t = \gamma\, d\hat{P}_t,$$

where γ is independent of y. However, by the Itô isometry,

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

which is independent of ϵ, i.e.,

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

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 [34, 35] within the so-called equation-free framework. In an ongoing work the coarse projective integration is applied to the system described by (1.1) [36].

REFERENCES

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

[2] G. C. PAPANICOLAOU, D. STROOCK, AND S. R. S. VARADHAN, *Martingale approach to some limit theorems*, in Papers from the Duke Turbulence Conference, Duke Univ. Math. Ser. 3, Duke University, Durham, NC, 1977, Paper No. 6.

[3] A. YU. VERETENNIKOV, *On an averaging principle for systems of stochastic differential equations*, Mat. Sb., 181 (1990), pp. 256–268 (in Russian); Math. USSR-Sb., 69 (1991), pp. 271–284 (in English).

[4] J. GOLEC AND G. LADDE, *Averaging principle and systems of singularly perturbed stochastic differential equations*, J. Math. Phys., 31 (1990), pp. 1116–1123.

[5] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, 2nd ed., Springer-Verlag, New York, 1998.

[6] Y. KIFER, *Stochastic versions of Anosov's and Neistadt's theorems on averaging*, Stoch. Dyn., 1 (2001), pp. 1–21.

[7] A. V. SKOROKHOD, *Asymptotic Methods in the Theory of Stochastic Differential Equations*, AMS, Providence, RI, 1989.

[8] G. G. YIN AND Q. ZHANG, *Discrete-Time Markov Chains. Two-Time-Scale Methods and Applications*, Appl. Math. (N.Y.) 55, Springer-Verlag, New York, 2005.

[9] G. YIN, *On limit results for a class of singularly perturbed switching diffusions*, J. Theoret. Probab., 14 (2001), pp. 673–697.

[10] A. M. ILIN, R. Z. KHASMINSKII, AND G. YIN, *Asymptotic expansions of solutions of integro-differential equations for transition densities of singularly perturbed switching diffusions: Rapid switchings*, J. Math. Anal. Appl., 238 (1999), pp. 516–539.

[11] Y. LIU AND G. YIN, *Asymptotic expansions of transition densities for hybrid jump-diffusions*, Acta Math. Appl. Sin. Engl. Ser., 20 (2004), pp. 1–18.

[12] D. T. GILLESPIE, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.

[13] E. L. HASELTINE AND J. B. RAWLINGS, *Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics*, J. Chem. Phys., 117 (2002), pp. 6959–6969.

[14] C. V. RAO AND A. P. ARKIN, *Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm*, J. Chem. Phys., 118 (2003), pp. 4999–5010.

[15] R. RICO-MARTÍNEZ, C. W. GEAR, AND I. G. KEVREKIDIS, *Coarse projective kMC integration: Forward/reverse initial and boundary value problems*, J. Comput. Phys., 196 (2004), pp. 474–489.

[16] Y. CAO, D. T. GILLESPIE, AND L. R. PETZOLD, *The slow-scale stochastic simulation algorithm*, J. Chem. Phys., 122 (2005), 014116.

[17] Y. KIFER, *Averaging and climate models*, in Stochastic Climate Models (Chorin, 1999), Progr. Probab. 49, Birkhäuser, Basel, 2001, pp. 171–188.

[18] 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.

[19] D. GIVON, I. G. KEVREKIDIS, AND R. KUPFERMAN, *Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems*, Commun. Math. Sci., 4 (2006), pp. 707–729.

[20] R. C. MERTON, *Option pricing when underlying stock returns are discontinuous*, J. Financial Econom., 3 (1976), pp. 125–144.

[21] E. PAPAGEORGIOU AND R. SIRCAR, *Multiscale intensity models for single name credit derivatives*, Appl. Math. Finance, to appear.

[22] J.-P. FOUQUE, G. PAPANICOLAOU, R. SIRCAR, AND K. SOLNA, *Multiscale stochastic volatility asymptotics*, Multiscale Model. Simul., 2 (2003), pp. 22–42.

[23] C. YUAN AND X. MAO, *Asymptotic stability in distribution of stochastic differential equations with Markovian switching*, Stochastic Process. Appl., 103 (2003), pp. 277–291.

[24] C. YUAN AND X. MAO, *Stability in distribution of numerical solutions for stochastic differential equations*, Stochastic Anal. Appl., 22 (2004), pp. 1133–1150.

[25] C. YUAN AND J. LYGEROS, *Invariant measure of stochastic hybrid processes with jumps*, in Proceedings of the 43rd IEEE Conference on Decision and Control, Vol. 3, 2004, pp. 3209–3214.

[26] X. MAO AND C. YUAN, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London, 2006.

[27] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Comm. Control Engrg. Ser., Springer-Verlag, London, 1993.

[28] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232.

[29] J. L. Menaldi and M. Robin, *Invariant measure for diffusions with jumps*, Appl. Math. Optim., 40 (1999), pp. 105–140.

[30] D. Talay, *Second order discretization schemes of stochastic differential systems for the computation of the invariant law*, Stochastics Stochastics Rep., 29 (1990), pp. 13–36.

[31] R. Z. Hasminski, *Stochastic Stability of Differential Equations*, Monographs Textbooks Mech. Solids Fluids: Mech. Anal. 7, Sijthoff & Noordhoff, Groningen, The Netherlands, Alphen aan den Rijn, Germantown, MD, 1980.

[32] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Appl. Math. (N.Y.) 23, Springer-Verlag, Berlin, 1992.

[33] F. B. Hanson, *Applied Stochastic Processes and Control for Jump Diffusions: Modeling, Analysis, and Computation*, Adv. Des. Control 13, SIAM, Philadelphia, 2007.

[34] C. W. Gear and I. G. Kevrekidis, *Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum*, SIAM J. Sci. Comput., 24 (2003), pp. 1091–1106.

[35] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 385–391.

[36] D. Givon and I. G. Kevrekidis, *Multiscale integration schemes for jump-diffusion systems*, Multiscale Model. Simul., submitted.