

# MULTISCALE INTEGRATION SCHEMES FOR JUMP-DIFFUSION SYSTEMS

DROR GIVON \* AND IOANNIS G. KEVREKIDIS †

**Abstract.** We study a two-time-scale system of jump-diffusion stochastic differential equations. We analyze a class of multiscale integration methods for these systems, which, in the spirit of [1], 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 multiscale integration method and the slow components of the original system.

**1. Introduction.** A wide variety of problems in the natural sciences give rise to singularly perturbed systems of stochastic differential equations (SDEs). In many cases, one is only interested in predicting the time evolution of some “slow component”, 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. 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 evolution of the slow components. Conditions under which the averaging principle can be applied to this kind of systems are well known in the classical literature. While the averaging principle and its resulting effective dynamics provide a substantial simplification of the original system, it is often impossible, or impractical to obtain the reduced equations in closed form. This has motivated the development of multiscale integration algorithms [2, 1]. Multiscale integration schemes along the lines described in [1] have been studied for different systems of SDEs [3, 4]. However, similar questions for jump-diffusion processes are not yet fully addressed. We consider two-time scale systems of jump-diffusion SDEs, of the form

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

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

where  $x_t^\epsilon$  is an  $n$ -dimensional jump-diffusion process and  $y_t^\epsilon$  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  $b(x, y) \in \mathbb{R}^{n \times d_1}$  and  $g(x, y) \in \mathbb{R}^{m \times d_2}$  are the diffusion coefficients, and the functions  $c(x, y) \in \mathbb{R}^n$  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  $\lambda_1$ , and  $N_t^\epsilon$  is a scalar simple Poisson process with intensity  $\frac{\lambda_2}{\epsilon}$ . The parameter  $\epsilon$  represents the ratio between the natural time scales of the  $x_t^\epsilon$  and  $y_t^\epsilon$  variables. We are concerned with situations where  $\epsilon \ll 1$ , i.e. where a separation of time scales prevails; in such case the vector  $x_t^\epsilon$  is called the “slow component” of the system, and the vector  $y_t^\epsilon$  is called the “fast component” of the system.

Within the framework of *averaging methods*, the separation of scales is taken to advantage to derive, in the limit  $\epsilon \rightarrow 0$ , a reduced equation for an  $n$ -dimensional process  $\bar{x}_t$ , which approximates the slow component  $x_t^\epsilon$  [5, 6, 7, 8, 9]. Define  $\mathcal{L}^\epsilon$  the

---

\*Department of Mathematics, UCB and LBNL, Berkeley CA 94720

†Department of Chemical Engineering, PACM and Mathematics, Princeton University

differential operator associated with (1.1) to be, for  $\mathcal{F} \in C_0^2(\mathbb{R}^{n+m})$ ,

$$\mathcal{L}^\epsilon = \mathcal{L}_1 + \frac{1}{\epsilon} \mathcal{L}_2,$$

where

$$\begin{aligned} \mathcal{L}_1 \mathcal{F} &= a(x, y) \partial_x \mathcal{F} + \frac{1}{2} \text{trace} \frac{\partial^2 \mathcal{F}}{\partial x^2} \cdot bb^\top(x, y) + \lambda_1 [\mathcal{F}(x + c(x, y)) - \mathcal{F}(x)] \\ \mathcal{L}_2 \mathcal{F} &= f(x, y) \partial_y \mathcal{F} + \frac{1}{2} \text{trace} \frac{\partial^2 \mathcal{F}}{\partial y^2} \cdot gg^\top(x, y) + \lambda_2 [\mathcal{F}(y + h(x, y)) - \mathcal{F}(y)]. \end{aligned}$$

Assume that for every fixed  $x$  the rapid variables, governed by (1.1b), induce a unique invariant, ergodic measure  $\mu^x(dy)$ . Denote

$$\begin{aligned} \bar{a}(x) &= \int_{\mathbb{R}^m} a(x, y) \mu^x(dy) \\ \bar{b}(x) \bar{b}^\top(x) &= \int_{\mathbb{R}^m} b(x, y) b^\top(x, y) \mu^x(dy). \end{aligned} \tag{1.2}$$

With this notation, it was shown in [10], that there exists a differential operator  $\bar{\mathcal{L}}$  of the averaged process for  $\mathcal{F} \in C_0^2(\mathbb{R}^n)$  which is defined by

$$\bar{\mathcal{L}} \mathcal{F} = \bar{a}(x) \partial_x \mathcal{F} + \frac{1}{2} \text{trace} \frac{\partial^2 \mathcal{F}}{\partial x^2} \cdot \bar{b} \bar{b}^\top(x) + \lambda_1 \int [\mathcal{F}(x + c(x, y)) - \mathcal{F}(x)] \mu^x(dy).$$

Using the definition of  $\bar{\mathcal{L}}$  we can define the effective dynamics  $\bar{x}_t = \bar{x}(t)$ . The drift coefficient  $\bar{a}(\bar{x})$  of the averaged operator  $\bar{\mathcal{L}}$  and the diffusion coefficient  $\bar{b}(\bar{x})$  are given in (1.2). The jumps of the averaged operator can be constructed as follows. The jump rate is  $\lambda_1$ . Let  $\{\tau_n\}$  denote the jump times. Then  $\bar{x}(\tau_n) - \bar{x}(\tau_n^-) = c(\bar{x}(\tau_n^-), \xi_n)$ , where  $\xi_n$  has distribution  $\mu^{\bar{x}(\tau_n^-)}(dy)$ . It is shown in [10] that as  $\epsilon \rightarrow 0$ ,  $x_t^\epsilon$  converges weakly, on every finite interval  $[0, T]$ , to the solution  $\bar{x}_t$  of a closed equation of the form

$$d\bar{x}_t = \bar{a}(\bar{x}_t) dt + \bar{b}(\bar{x}_t) dB_t + \int c(\bar{x}_t, \xi) P_{\bar{x}_t}(dsd\xi), \tag{1.3}$$

where  $P_x(ds d\xi)$  is a Poisson measure with jump rate  $\lambda_1$  and jump distribution  $\mu^x(\cdot)$ .

There are cases in the framework of the averaging principle when one is interested in the limit behavior of a trajectory rather than the limit behavior of the distribution of trajectories. In this case a stronger form of convergence is needed. Analysis of the strong, i.e., uniform in time, convergence for a two-time-scale jump-diffusion system is given in [11]. It is shown that the strong convergence is valid only if the coefficients  $b$  and  $c$  in (1.1a) do not depend on the fast variable, i.e.,  $b = b(x), c = c(x)$ , which implies that (1.1a) has the following form,

$$dx_t^\epsilon = a(x_t^\epsilon, y_t^\epsilon) dt + b(x_t^\epsilon) dB_t + c(x_t^\epsilon) dP_t \quad x_0^\epsilon = x_0. \tag{1.4}$$

As  $\epsilon \rightarrow 0$ ,  $x_t^\epsilon$  converges in the strong sense, on every finite interval  $[0, T]$ , to the solution  $\bar{x}_t$  of a closed equation of the form

$$d\bar{x}_t = \bar{a}(\bar{x}_t) dt + b(\bar{x}_t) dB_t + c(\bar{x}_t) dP_t. \tag{1.5}$$

While the averaging principle and its resulting effective dynamics (1.3), (1.5) 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  $\mu^x$  is unknown, or because integrations cannot be performed analytically). This has motivated the development of two types of multiscale algorithms which overcome this difficulty: coarse projective integration methods [2, 12, 13] (see also [14] for the non-“coarse”, ODE case) developed within the equation-free framework [15]; and multiscale integrators [1] which were presented in the framework of the heterogeneous multiscale methods [16]. Coarse projective integration has focused on the case where explicit knowledge of the original system is not available (for example, the available simulator is in the form of a “black box”). In that case short bursts of appropriately initialized fine scale simulations are used to estimate on demand the numerical quantities required to perform scientific computing tasks with coarse-grained models (time derivatives and, for the case of stochastic coarse-grained models, the local effective noise drift and diffusivity, e.g. [17]). The present paper deals with the case where one has full knowledge of the original, singularly perturbed, system. In this case the multiscale integrators introduced in [1], which take advantage of the explicit knowledge of the right-hand-sides of the singularly perturbed dynamical systems, are more suitable. This knowledge together with the averaging principle is the key factor of the multiscale integrators, which aim at integrating the averaged equations to get an approximation of the dynamics in the original system. Analysis for the multiscale integrators was presented in [3]. In [4] the authors extended the idea of multiscale integration for a deterministic effective model to the case where the effective model is a stochastic one. Here we extend the idea to the case where the effective dynamics are described by stochastic processes that admit jumps and discontinuous trajectories.

The use of the averaging principle to compute effective jump processes was implemented in a different context: a research area where the dynamics are stochastic and consist of jumps is the stochastic simulation of kinetic chemical reactions, also known as the Gillespie stochastic simulation algorithm, SSA [18]. 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 location. The event that takes place is chosen according to a rate function which depends on the state space location. This model consists of no drift, only jumps. In the past five years extensive progress had been made in describing the effective dynamics for chemical kinetic systems that take place on vastly different time scales [19, 20, 12, 21, 22].

In its simplest formulation, the multiscale integration scheme can be described as follows: Let  $\Delta t$  be a fixed time step, and  $X_n$  be the numerical approximation to the coarse variable,  $\bar{x}$ , at time  $t_n = n\Delta t$ . 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  $\delta t$  and  $x = X_n$  fixed. We denote the discrete variables associated with the fast dynamics at the  $n$ -th coarse step by  $Y_m^n$ ,  $m = 0, 1, \dots, M$ . As in [1] the numerical solver used to generate the sequence  $Y_m^n$  is called the *micro-solver* (or micro-integrator). The micro solver as introduced in [1] is a numerical approximation of the auxiliary process which was introduced in the work of [8] and [7]. The simplest choice is the Euler scheme,

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

where  $\Delta W_m^n$  are Brownian displacements over a time interval  $\delta t$  and  $\Delta N_m^n$  are Poisson

increments with intensity  $\frac{\lambda_2}{\epsilon}$ . The  $Y_m^n$  are generated in order to give us an approximation of the ergodic measure  $\mu_x(dy)$ . This allows us to omit the  $\epsilon$  dependency from the micro solver, hence we use instead,

$$Y_{m+1}^n = Y_m^n + f(X_n, Y_m^n) \delta t + g(X_n, Y_m^n) \Delta W_m^n + h(X_n, Y_m^n) \Delta N_m^n, \quad (1.6)$$

where  $\Delta W_m^n$  are Brownian displacements over a time interval  $\delta t$  and  $\Delta N_m^n$  are Poisson increments with intensity  $\lambda_2$ . Since we assume that the  $y$  dynamics is ergodic, we may choose, among other choices,  $Y_0^n = Y_M^{n-1}$  (In Section 5 we further discuss this point). Having generated the trajectories  $Y_m^n$ , the functions  $\bar{a}, \bar{b}$  and the jump amplitude  $\xi$  are estimated by

$$A(X_n) = \frac{1}{M} \sum_{m=1}^M a(X_n, Y_m^n) \quad (1.7a)$$

$$B(X_n)B^\top(X_n) = \frac{1}{M} \sum_{m=1}^M b(X_n, Y_m^n)b^\top(X_n, Y_m^n) \quad (1.7b)$$

$$c(X_n, \xi) = c(X_n, Y_M^n). \quad (1.7c)$$

$B(X_n)$  can be extracted from  $B(X_n)B^\top(X_n)$  through a Cholesky decomposition. Finally, to reduce the statistical noise, several independent realizations of the micro-solver can be carried out, in which case expression (1.7) for  $A(X)$  and  $B(X)$  involve an additional averaging over these independent realizations.

Inspired by the limiting equation (1.3),  $X_n$  is evolved in time by an Euler step,

$$X_{n+1} = X_n + A(X_n) \Delta t + B(X_n) \Delta B_n + c(X_n, Y_n^m) \Delta P_n, \quad (1.8)$$

where  $\Delta B_n$  are Brownian displacements over a time interval  $\Delta t$  and  $\Delta P_n$  are the increments of the Poisson process over a time interval  $\Delta t$ . We refer to (1.8) as the *macro-solver* (or, macro integrator).

Equations (1.8), (1.6), and (1.7) define the multiscale integration scheme.

The functions  $A(X), B(X)$  approximate the functions  $\bar{a}(x), \bar{b}(x)$ , which result from the averaging (1.2) 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. Note that the time interval is long enough to guarantee that  $Y_x^m$  will have a distribution close enough to the distribution of the invariant measure  $\mu_x$ , which explains the approximation for the averaged jump distribution. Hence the approximation for the jump coefficient is given by  $c(x, Y_n^m)$ .

This paper deals with (1.1). The motivation for such a problem stems from attempts to model elements of a financial market that involves two kind 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 [23]:

- 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, while 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. The value of a financial instrument can change in hours or days, while the value of other financial instruments will change only over a time period of months or years. In [24] 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 [25].

In this paper we analyze the multiscale integration scheme for systems of the form (1.1). The contribution of this paper is the following. We derive estimates for the strong (pathwise) error between the solution  $\bar{x}_t$  of the effective dynamics (1.5) and the solution  $X_n$  of the multiscale integration scheme (1.6),(1.7a) and (1.8). Recall in that case the use of (1.4) instead of (1.1a). Specifically, we obtain an error estimate of the form

$$\mathbb{E} \sup_{n \leq T/\Delta t} |\bar{x}_{t_n} - X_n|^2 \leq C \left( \Delta t + \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right).$$

This convergence is uniform within the whole time interval  $[0, T]$ . We also obtain a weak convergence estimate between the solution  $\bar{x}_t$  of the effective dynamics (1.3) and the solution  $X_n$  of the multiscale integration scheme (1.6),(1.7) and (1.8). Specifically we obtain,

$$\sup_{n \leq T/\Delta t} |\mathbb{E}\mathcal{F}(\bar{X}_n) - \mathbb{E}\mathcal{F}(X_n)| \leq C \left\{ \sqrt{\frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t}} + \sqrt{\Delta t} \right\}.$$

This result extends the analysis in [3], beyond the case where the slow dynamics (and hence the limiting dynamics) are deterministic, and beyond the case where the dynamics are stochastic, yet only consist of continuous paths, which was analyzed in [4].

The paper is organized as follows. Section 2 states the lemmas and theorem for the strong case. Section 3 contains the proof for the strong case. Section 4 deals with weak convergence, and Section 5 contains our discussion.

**2. Analysis of the multiscale integration scheme.** In this section we analyze the convergence of the numerical method defined by eqs. (1.6), (1.7a) and (1.8). Specifically, we derive an estimate for the distance between the computed solution  $X_n$  and the solution  $\bar{x}_t$  of the effective dynamics (1.5) at time  $t = t_n$ . We prove strong, i.e., uniform in time, convergence. Our main result in this section is Theorem 2.10. For the sake of readability we state in this section our assumptions, lemmas and theorems, deferring all proofs to the next section. The analysis follows a structure based

on the work of [8] and [7]: an introduction of a well posed auxiliary process which is based on fixing the slow variables. The auxiliary process is compared with the effective dynamics. This leads to estimating the terms in the right hand side (the drift and diffusion coefficient, and jump components). As shown in [3], additional comparisons are needed for the numerical methods. In [3] the numerical solution of the effective equation based on a given macro-solver is compared to the evolution of the slow variable in the multiscale integrator based on the same macro-solver. We also follow this strategy.

Throughout this work, the following assumptions are made:

ASSUMPTION 2.1.

A1. The functions  $a = a(x, y)$ ,  $b = b(x)$  and  $c = c(x)$  in (1.1a) (or (1.4)) are measurable, Lipschitz continuous, and hence have linear growth bounds: specifically, there exist constants  $L, K$ , such that

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

and

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

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

A2. The functions  $f(x, y)$ ,  $g(x, y)$  and  $h(x, y)$  in (1.1b) are of class  $C^\infty$  and have bounded derivatives of any order; in particular, we can choose the globally Lipschitz constant  $L_f$  ( $L_g, L_h$ , resp.) sufficiently large, such that it bounds the first derivatives of  $f$  ( $g, h$  resp.) 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,

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

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

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

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

A4. There exists a constant  $\beta > 2\lambda_2 L_h + L_g$ , independent of  $x$ , such that for all  $y_1, y_2 \in \mathbb{R}^m$

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

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

In [11] it is proved that under these assumptions the slow component in (1.1) converges to the solution of the effective dynamics (1.5) in the strong sense.

Note that the effective dynamics does not depend on  $\epsilon$ . Also, since the discrete solution  $Y_m^n$  obtained by the micro-solver is for  $X_n$  fixed, it only depends on the

ratio  $\delta t/\epsilon$ . The first lemma describes the relation between two JSDEs which differ by rescaling time.

LEMMA 2.1. *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  $\lambda/\epsilon$ . Then  $\check{x}(t) = x(t\epsilon)$  is a solution of the stochastic equation

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

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

Thus, without loss of generality, we may take  $\epsilon = 1$  in (1.6).

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

$$X_{n+1} = X_n + A(X_n) \Delta t + b(X_n) \Delta B_n + c(X_n) \Delta P_n, \quad X_0 = x_0, \quad (2.1)$$

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

$$A(X_n) = \frac{1}{M} \sum_{m=1}^M a(X_n, Y_m^n), \quad (2.2)$$

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

$$dz_t^n = f(X_n, z_t^n) dt + g(X_n, z_t^n) dW_t^n + h(X_n, z_t^n) dN_t^n, \quad (2.3)$$

with initial conditions  $z_0^n = Y_0^n = y_0$ , and a time step  $\delta t$  (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 (1.6) is a particular realization that uses an Euler time-stepper as well,

$$Y_{m+1}^n = Y_m^n + f(X_n, Y_m^n) \delta t + g(X_n, Y_m^n) \Delta W_m^n + h(X_n, Y_m^n) \Delta N_m^n, \quad (2.4)$$

where  $\Delta W_m^n = W_{(m+1)\delta t}^n - W_{m\delta t}^n$  are the Brownian increments associated with the SDEs (2.3), and  $\Delta N_m^n = N_{(m+1)\delta t}^n - N_{m\delta t}^n$  are the Poisson increments associated with the SDEs (2.3).

We also introduce a discrete auxiliary process,  $\bar{X}_n$ , which is the Euler solution of the effective dynamics (1.5):

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

As was proved by in [26], the Euler scheme is of order 1/2, which implies the existence of a constant  $K_0 = K_0(T, x_0)$ , such that

$$\mathbb{E} \sup_{0 \leq n \leq \lfloor T/\Delta t \rfloor} |\bar{x}(t_n) - \bar{X}_n|^2 \leq K_0 \Delta t. \quad (2.5)$$

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

The next four lemmas provide mean square estimates for the processes  $z_t^n, Y_m^n$  and  $X_n$ .

LEMMA 2.2. *The process  $z_t^n$  satisfies,*

$$\sup_{0 \leq t} \mathbb{E}|z_t^n|^2 \leq K_1,$$

where  $K_1 = K_1(y_0) = |y_0|^2 + \frac{1}{\beta} \left[ \frac{2}{\beta} c_f^2(0) + c_g^2 + \lambda_2^2 \frac{2}{\beta} c_h^2 + \lambda_2 c_h^2 \right]$ .

LEMMA 2.3. *For small enough  $\delta t$ ,*

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

where

$$K_2 = K_2(y_0) = |y_0|^2 + \frac{2}{\beta} \left( c_g^2 + \lambda_2 c_h^2 + \frac{2c_f^2(0)}{\beta} + \frac{2\lambda_2^2}{\beta} c_h^2 \right).$$

LEMMA 2.4. *For small enough  $\Delta t$ ,*

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

where

$$K_3 = K_3(T, x_0, y_0) = e^{[2(1+\lambda_1)(1+2K^2)]T} \left[ |x_0|^2 + 2 \frac{K^2(1+K_2)}{1+2K^2} \right].$$

LEMMA 2.5. *The mean square deviation between two successive iterations of the microsolver satisfies, for small enough  $\delta t$ ,*

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

where  $K_4 = 8(c_g^2 + \lambda_2 c_h^2)$ .

Lemma 2.1 implies that the process  $z_t^n$  is statistically equivalent to a shifted and rescaled version of  $y_t^\epsilon$ , with  $x$  being a parameter, that is,  $z_t^k \sim y_{(t-t_k)/\epsilon}^\epsilon$ .

Menaldi and Robin [27] proved that the dynamic (2.3) is ergodic with invariant measure  $\mu^{X_n}$  (Assumptions A2-A4) (See also [28]). Moreover, they proved that the process  $z_t^n$  satisfy the Doeblin condition hence it is exponentially mixing in the following sense. Let  $P^{X_n}(t, z, E)$  denote the transition probability of (2.3). Then there are positive constants  $\gamma, \alpha < 1$  such that

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

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

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  $\mu^{X_n}$ , which is the invariant measure induced by the process  $z_t^n$ . We denote  $z_m^n = z_{m\delta t}^n$ .

LEMMA 2.6. *For small enough  $\delta t$ , there exist a constant  $K_5$  independent of  $M, \delta t$  s.t.*

$$\mathbb{E} \left| \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) - \bar{a}(X_n) \right|^2 \leq K_5 \left[ \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} \right]. \quad (2.6)$$

The next lemma establishes the mean deviation between (2.3) and its numerical approximation (2.4).

LEMMA 2.7. *Let  $z_t^n$  be the family of processes defined by (2.3). For small enough  $\delta t$ , there exists a constant  $K_8$  such that,*

$$\max_{\substack{0 \leq n \leq \lfloor \frac{T}{\Delta t} \rfloor \\ 0 \leq m \leq M}} \mathbb{E} |Y_m^n - z_m^n|^2 \leq K_8 \sqrt{\delta t}, \quad (2.7)$$

where  $K_8 = \frac{(2L_f + \lambda_2 L_h) \sqrt{K_4}}{\beta - 2\lambda_2 L_h - L_g}$ .

LEMMA 2.8. *There exists a constant  $K_6 = K_6(T, x_0, y_0)$ , such that for all  $0 \leq n \leq \lfloor T/\Delta t \rfloor$*

$$\mathbb{E} |\bar{a}(X_n) - A(X_n)|^2 \leq K_6 \left( \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right).$$

LEMMA 2.9. *There exists a constant  $K_7 = K_7(T, x_0, y_0)$  such that,*

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

where  $K_7 = 8T^2 K_6 e^{4L^2(3T+1+\lambda_1)T}$ .

Combining this result with (2.5) our main theorem readily follows:

THEOREM 2.10. *There exists a constant  $K_7 = K_7(T, x_0, y_0)$  such that,*

$$\mathbb{E} \sup_{0 \leq n \leq \lfloor T/\Delta t \rfloor} |X_n - \bar{x}(t_n)|^2 \leq 2K_0 \Delta t + 2K_7 \left( \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right).$$

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.

**3. Proofs for Section 2.** 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  $\delta t$ , satisfy the linear inequality,

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

then

$$Z_n \leq e^{an \delta t} Z_0 + \frac{b}{a} (e^{an \delta t} - 1). \quad (3.1)$$

Young's inequality will be used repeatedly, each time however with a different constant. Let  $a, b \in \mathbb{R}^n$  then for every  $\beta > 0$ ,

$$2a \cdot b \leq \beta |a|^2 + \frac{1}{\beta} |b|^2.$$

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

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

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

hence

$$\mathbb{E} \left| \hat{P}(t + \delta t) - \hat{P}(t) \right|^2 = \lambda \delta t.$$

*Proof.* of Lemma 2.1

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

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

*Proof.* of Lemma 2.2 Applying Itô's chain rule to the process  $|z_t^n|^2$ ,

$$\begin{aligned} d|z_t^n|^2 &= 2z_t^n \cdot f(X_n, z_t^n) dt + \|g(X_n, z_t^n)\|^2 dt + 2z_t^n g(X_n, z_t^n) dW_t^n \\ &\quad + 2z_t^n \cdot h(X_n, z_t^n) dN_t^n + |h(X_n, z_t^n)|^2 dN_t^n. \end{aligned} \quad (3.2)$$

Taking expectations, and using Itô isometry,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |z_t^n|^2 &= 2\mathbb{E} z_t^n \cdot f(X_n, z_t^n) + \mathbb{E} \|g(X_n, z_t^n)\|^2 + \\ &\quad 2\lambda_2 \mathbb{E} z_t^n \cdot h(X_n, z_t^n) + \lambda_2 \mathbb{E} |h(X_n, z_t^n)|^2. \end{aligned}$$

Substituting Assumption A4 with  $y_1 = z_t^n$  and  $y_2 = 0$  gives,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |z_t^n|^2 &= -2\beta \mathbb{E} |z_t^n|^2 + 2\mathbb{E} z_t^n \cdot f(X_n, 0) + \mathbb{E} \|g(X_n, z_t^n)\|^2 \\ &\quad + 2\lambda_2 \mathbb{E} z_t^n \cdot h(X_n, z_t^n) + \lambda_2 \mathbb{E} |h(X_n, z_t^n)|^2. \end{aligned}$$

Using Young's inequality twice and substituting the bounds on  $f, g$  and  $h$  (Assumption A2),

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |z_t^n|^2 &= -2\beta \mathbb{E} |z_t^n|^2 + \frac{\beta}{2} \mathbb{E} |z_t^n|^2 + \frac{2}{\beta} \mathbb{E} |f(X_n, 0)|^2 + \mathbb{E} \|g(X_n, z_t^n)\|^2 \\ &\quad + \frac{\beta}{2} \mathbb{E} |z_t^n|^2 + \lambda_2^2 \frac{2}{\beta} \mathbb{E} |h(X_n, z_t^n)|^2 + \lambda_2 \mathbb{E} |h(X_n, z_t^n)|^2 \\ &\leq -\beta \mathbb{E} |z_t^n|^2 + \left[ \frac{2}{\beta} c_f^2(0) + c_g^2 + \lambda_2^2 \frac{2}{\beta} c_h^2 + \lambda_2 c_h^2 \right]. \end{aligned}$$

The desired result follows from Gronwall's inequality.  $\square$

*Proof.* of Lemma 2.3 We rewrite (2.4) in terms of the compensated Poisson process,

$$\begin{aligned} Y_{m+1}^n &= Y_m^n + f(X_n, Y_m^n) \delta t + g(X_n, Y_m^n) \Delta W_m^n + h(X_n, Y_m^n) \Delta N_m^n \\ &= Y_m^n + f(X_n, Y_m^n) \delta t + g(X_n, Y_m^n) \Delta W_m^n + h(X_n, Y_m^n) \Delta \hat{N}_m^n \\ &\quad + \lambda_2 h(X_n, Y_m^n) \delta t. \end{aligned}$$

Squaring and taking expectations,

$$\begin{aligned} \mathbb{E} |Y_{m+1}^n|^2 &= \mathbb{E} |Y_m^n|^2 + \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 + \delta t \mathbb{E} \|g(X_n, Y_m^n)\|^2 \\ &\quad + \lambda_2 \delta t \mathbb{E} |h(X_n, Y_m^n)|^2 + \lambda_2^2 \delta t^2 \mathbb{E} |h(X_n, Y_m^n)|^2 + 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, Y_m^n) \\ &\quad + 2\lambda_2 \delta t \mathbb{E} Y_m^n \cdot h(X_n, Y_m^n) + 2\lambda_2 \delta t^2 \mathbb{E} f(X_n, Y_m^n) h(X_n, Y_m^n). \end{aligned}$$

We subtract and add  $2\delta t \mathbb{E} Y_m^n \cdot f(X_n, 0)$ ,

$$\begin{aligned} \mathbb{E} |Y_{m+1}^n|^2 &= \mathbb{E} |Y_m^n|^2 + \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 + \delta t \mathbb{E} \|g(X_n, Y_m^n)\|^2 \\ &\quad + \lambda_2 \delta t \mathbb{E} |h(X_n, Y_m^n)|^2 + \lambda_2^2 \delta t^2 \mathbb{E} |h(X_n, Y_m^n)|^2 + 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, Y_m^n) \\ &\quad - 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, 0) + 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, 0) \\ &\quad + 2\lambda_2 \delta t \mathbb{E} Y_m^n \cdot h(X_n, Y_m^n) + 2\lambda_2 \delta t^2 \mathbb{E} f(X_n, Y_m^n) h(X_n, Y_m^n). \end{aligned}$$

For each one of the last three terms on the right hand side we use Young's inequality,

$$\begin{aligned} \mathbb{E} |Y_{m+1}^n|^2 &= \mathbb{E} |Y_m^n|^2 + \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 + \delta t \mathbb{E} \|g(X_n, Y_m^n)\|^2 \\ &\quad + \lambda_2 \delta t \mathbb{E} |h(X_n, Y_m^n)|^2 + \lambda_2^2 \delta t^2 \mathbb{E} |h(X_n, Y_m^n)|^2 + 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, Y_m^n) \\ &\quad - 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, 0) + \delta t \frac{\beta}{2} \mathbb{E} |Y_m^n|^2 + \delta t \frac{2}{\beta} \mathbb{E} |f(X_n, 0)|^2 \\ &\quad + \delta t \frac{\beta}{2} \mathbb{E} |Y_m^n|^2 + \delta t \frac{2}{\beta} \lambda_2^2 \mathbb{E} |h(X_n, Y_m^n)|^2 + \lambda_2 \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 \\ &\quad + \lambda_2 \delta t^2 \mathbb{E} |h(X_n, Y_m^n)|^2. \end{aligned}$$

Using the boundedness of the functions  $f, g$  and  $h$  (Assumption A2),

$$\begin{aligned} \mathbb{E} |Y_{m+1}^n|^2 &\leq (1 + \beta \delta t) \mathbb{E} |Y_m^n|^2 + \delta t \left( c_g^2 + \lambda_2 c_h^2 + \frac{2c_f^2(0)}{\beta} + \frac{2\lambda_2^2}{\beta} c_h^2 \right) \\ &\quad + \delta t^2 (\lambda_2^2 c_h^2 + \lambda_2 c_h^2) + \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 \\ &\quad + 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, Y_m^n) - 2\delta t \mathbb{E} Y_m^n \cdot f(X_n, 0) \\ &\quad + \lambda \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2. \end{aligned}$$

For the third line on the right hand side we use Assumption  $A_4$ ,

$$\begin{aligned}
\mathbb{E} |Y_{m+1}^n|^2 &\leq (1 + \beta\delta t)\mathbb{E} |Y_m^n|^2 + \delta t \left( c_g^2 + \lambda_2 c_h^2 + \frac{2c_f^2(0)}{\beta} + \frac{2\lambda_2^2}{\beta} c_h^2 \right) \\
&\quad + \delta t^2 (\lambda_2^2 c_h^2 + \lambda_2 c_h^2) + \delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 \\
&\quad - 2\beta\delta t \mathbb{E} |Y_m^n|^2 \\
&\quad + \lambda\delta t^2 \mathbb{E} |f(X_n, Y_m^n)|^2 \\
&= (1 - \beta\delta t)\mathbb{E} |Y_m^n|^2 + \delta t \left( c_g^2 + \lambda_2 c_h^2 + \frac{2c_f^2(0)}{\beta} + \frac{2\lambda_2^2}{\beta} c_h^2 \right) \\
&\quad + \delta t^2 (\lambda_2^2 c_h^2 + \lambda_2 c_h^2) + \delta t^2 (1 + \lambda_2) \mathbb{E} |f(X_n, Y_m^n)|^2.
\end{aligned} \tag{3.3}$$

Rewriting  $f(X_n, Y_m^n) = f(X_n, Y_m^n) - f(X_n, 0) + f(X_n, 0)$  and using Assumption  $A_2$ ,

$$|f(X_n, Y_m^n)|^2 \leq 2L_f^2 |Y_m^n|^2 + 2c_f^2(0), \tag{3.4}$$

which substituted into the last term of (3.3) gives,

$$\begin{aligned}
\mathbb{E} |Y_{m+1}^n|^2 &\leq (1 - \beta\delta t + 2L_f^2(1 + \lambda_2) \delta t^2) \mathbb{E} |Y_m^n|^2 \\
&\quad + \delta t \left( c_g^2 + \lambda_2 c_h^2 + \frac{2c_f^2(0)}{\beta} + \frac{2\lambda_2^2}{\beta} c_h^2 \right) + \delta t^2 (1 + \lambda_2) [2c_f^2(0) + \lambda_2 c_h^2],
\end{aligned}$$

and the desired result follows from the discrete Gronwall inequality (3.1).  $\square$

*Proof.* of Lemma 2.4 We rewrite (2.1) in terms of the compensated Poisson process,

$$\begin{aligned}
X_{n+1} &= X_n + A(X_n) \Delta t + b(X_n) \Delta B_n + c(X_n) \Delta P_n \\
&= X_n + A(X_n) \Delta t + b(X_n) \Delta B_n + c(X_n) \Delta \hat{P}_n + \lambda_1 c(X_n) \Delta t.
\end{aligned}$$

Squaring, taking expectations and using the Itô isometry,

$$\begin{aligned}
\mathbb{E} |X_{n+1}|^2 &= \mathbb{E} |X_n|^2 + \frac{\Delta t^2}{M^2} \mathbb{E} \left| \sum_{m=1}^M a(X_n, Y_m^n) \right|^2 + \Delta t \mathbb{E} \|b(X_n)\|^2 \\
&\quad + \lambda_1 \Delta t \mathbb{E} |c(X_n)|^2 + \lambda_1^2 \Delta t^2 \mathbb{E} |c(X_n)|^2 \\
&\quad + \frac{2\Delta t}{M} \sum_{m=1}^M \mathbb{E} X_n \cdot a(X_n, Y_m^n) + 2\Delta t \lambda_1 \mathbb{E} X_n \cdot c(X_n) \\
&\quad + \frac{2\Delta t^2 \lambda_1}{M} \sum_{m=1}^M \mathbb{E} c(X_n) \cdot a(X_n, Y_m^n).
\end{aligned}$$

Using the inequality  $(a_1 + \dots + a_n)^2 \leq n(a_1^2 + \dots + a_n^2)$  together with Young's inequality,

$$\begin{aligned} \mathbb{E} |X_{n+1}|^2 &\leq \mathbb{E} |X_n|^2 + \frac{\Delta t^2}{M} \mathbb{E} \sum_{m=1}^M |a(X_n, Y_m^n)|^2 + \Delta t \mathbb{E} \|b(X_n)\|^2 \\ &\quad + \lambda_1 \Delta t \mathbb{E} |c(X_n)|^2 + \lambda_1^2 \Delta t^2 \mathbb{E} |c(X_n)|^2 + \Delta t \mathbb{E} |X_n|^2 \\ &\quad + \frac{\Delta t}{M} \sum_{m=1}^M \mathbb{E} |a(X_n, Y_m^n)|^2 + \Delta t \lambda_1 \mathbb{E} |X_n|^2 + \Delta t \lambda_1 \mathbb{E} |c(X_n)|^2 \\ &\quad + \Delta t^2 \lambda_1 \mathbb{E} |c(X_n)|^2 + \frac{\Delta t^2 \lambda_1}{M} \sum_{m=1}^M \mathbb{E} |a(X_n, Y_m^n)|^2. \end{aligned}$$

Using Assumption A1, we get,

$$\begin{aligned} \mathbb{E} |X_{n+1}|^2 &\leq (1 + \Delta t + \lambda_1 \Delta t) \mathbb{E} |X_n|^2 + 2(1 + \lambda_1) K^2 (1 + \mathbb{E} |X_n|^2) \Delta t \\ &\quad + 2(1 + \lambda_1) K^2 \sup_{m \geq 0} [\mathbb{E} |Y_m^n|^2] \Delta t + 2\Delta t^2 (1 + \lambda^2) K^2 (1 + \mathbb{E} |X_n|^2) \\ &\quad + 2\Delta t^2 (1 + \lambda^2) K^2 \sup_{m \geq 0} [\mathbb{E} |Y_m^n|^2]. \end{aligned}$$

Using Lemma 2.3, we get, for small enough  $\Delta t$ ,

$$\mathbb{E} |X_{n+1}|^2 \leq [1 + 2(1 + \lambda_1)(1 + 2K^2)\Delta t] \mathbb{E} |X_n|^2 + 4(1 + \lambda_1)K^2(1 + K_2) \Delta t.$$

and the desired result follows from the discrete Gronwall inequality (3.1).  $\square$

*Proof.* of Lemma 2.5 Eq. (2.4) together with Assumption A2 implies,

$$\begin{aligned} \mathbb{E} |Y_{m+1}^n - Y_m^n|^2 &= \mathbb{E} |f(X_n, Y_m^n) \delta t + g(X_n, Y_m^n) \Delta W_m^n + h(X_n, Y_m^n) \Delta \hat{N}_m^n \\ &\quad + \lambda_2 h(X_n, Y_m^n) \delta t|^2 \\ &\leq 4 \mathbb{E} |f(X_n, Y_m^n)|^2 \delta t^2 + 4 \mathbb{E} \|g(X_n, Y_m^n)\|^2 \delta t \\ &\quad + 4\lambda_2 \mathbb{E} |h(X_n, Y_m^n)|^2 \delta t + 4\lambda_2^2 \mathbb{E} |h(X_n, Y_m^n)|^2 \delta t^2 \\ &\leq 8L^2 \delta t^2 \mathbb{E} |Y_m^n|^2 + 8c_f^2(0) \delta t^2 + 4c_g^2 \delta t \\ &\quad + 4\lambda_2 c_h^2 \delta t + 4\lambda_2^2 c_h^2 \delta t^2, \end{aligned}$$

where we have used (3.4). Lemma 2.3 implies that for  $\delta t$  small enough,

$$\mathbb{E} |Y_{m+1}^n - Y_m^n|^2 \leq 8(c_g^2 + \lambda_2 c_h^2) \delta t.$$

$\square$

*Proof.* of Lemma 2.6 Expanding (2.6),

$$\begin{aligned}
& \mathbb{E} \left| \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) - \bar{a}(X_n) \right|^2 \\
&= \frac{1}{M^2} \sum_{m=1}^M \sum_{p=1}^M \{ \mathbb{E} [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot [a(X_n, z_p^n) - \bar{a}(X_n)] \} \\
&= \frac{2}{M^2} \sum_{m=1}^M \sum_{p=m+1}^M \mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot [a(X_n, z_p^n) - \bar{a}(X_n)] \} \\
&+ \frac{1}{M^2} \sum_{m=1}^M \mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot [a(X_n, z_m^n) - \bar{a}(X_n)] \}.
\end{aligned}$$

We estimate the summand using the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot [a(X_n, z_p^n) - \bar{a}(X_n)] \} \\
&= \mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot \mathbb{E} [a(X_n, z_p^n) - \bar{a}(X_n) | z_m^n] \} \\
&= \mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot \mathbb{E}_{z_m^n} [a(X_n, z_{p-m}^n) - \bar{a}(X_n)] \} \\
&\leq \max_m \left\{ \mathbb{E} [a(X_n, z_m^n) - \bar{a}(X_n)]^2 \right\}^{\frac{1}{2}} \left\{ \mathbb{E} [\mathbb{E}_{z_m^n} [a(X_n, z_{p-m}^n) - \bar{a}(X_n)]^2] \right\}^{\frac{1}{2}}.
\end{aligned}$$

For the left hand term we use the linear growth bound of the functions  $a, \bar{a}$  and Lemmas 2.2, 2.4,

$$\begin{aligned}
\mathbb{E} [a(X_n, z_m^n) - \bar{a}(X_n)]^2 &\leq 2\mathbb{E} |a(X_n, z_m^n)|^2 + 2\mathbb{E} |\bar{a}(X_n)|^2 \\
&\leq 4K^2(1 + K_1 + K_3).
\end{aligned}$$

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

$$\mathbb{E} \{ [a(X_n, z_m^n) - \bar{a}(X_n)] \cdot [a(X_n, z_p^n) - \bar{a}(X_n)] \} \leq 2L\sqrt{1 + K_1 + K_3}\gamma\alpha^{(p-m)\delta t}.$$

Set  $k_1 = 2L\sqrt{1 + K_1 + K_3}\gamma$ ,

$$\mathbb{E} \left| \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) - \bar{a}(X_n) \right|^2 \leq \frac{2k_1}{M^2} \sum_{m=1}^M \sum_{p=m+1}^M \alpha^{(p-m)\delta t} + \frac{k_1}{M} \alpha^0.$$

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,

$$\begin{aligned}
\mathbb{E} \left| \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) - \bar{a}(X_n) \right|^2 &\leq \frac{2k_1}{M^2} \sum_{m=1}^M \sum_{p=m+1}^{m+l} \alpha^{(p-m)\delta t} \\
&+ \frac{2k_1}{M^2} \sum_{m=1}^M \sum_{p=m+l+1}^M \alpha^{(p-m)\delta t} + \frac{k_1}{M} \\
&\leq \frac{2k_1 l}{M} + 2k_1 \alpha^{l\delta t} + \frac{k_1}{M}.
\end{aligned}$$

Set  $l = \frac{1}{\delta t} (-\log_\alpha(M\delta t))$ . Thus, there exist constants  $k_2, k_3, k_4$  such that,

$$\mathbb{E} \left| \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) - \bar{a}(X_n) \right|^2 \leq k_2 \frac{-\log_\alpha(M\delta t)}{M\delta t} + k_3 \frac{1}{M\delta t} + k_4 \frac{1}{M}.$$

This concludes the proof.  $\square$

*Proof.* of Lemma 2.7 Define  $t_{\delta t} = \lfloor t/\delta t \rfloor \delta t$ . Let  $Y_t^n$  be the Euler approximation  $Y_m^n$ , interpolated continuously by

$$Y_t^n = \int_0^t f(X_n, Y_{n,s_{\delta t}}) ds + \int_0^t g(X_n, Y_{n,s_{\delta t}}) dW_{n,s} + \int_0^t h(X_n, Y_{n,s_{\delta t}}) dN_{n,s}.$$

Define

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

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

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &= 2\mathbb{E} v_t \cdot [f(X_n, Y_{n,t_{\delta t}}) - f(X_n, z_t^n)] + \mathbb{E} \|g(X_n, Y_{n,t_{\delta t}}) - g(X_n, z_t^n)\|^2 \\ &\quad + 2\lambda_2 \mathbb{E} v_t \cdot [h(X_n, Y_{n,t_{\delta t}}) - h(X_n, z_t^n)] + \lambda_2 \mathbb{E} |h(X_n, Y_{n,t_{\delta t}}) - h(X_n, z_t^n)|^2 \\ &= 2\mathbb{E} v_t \cdot [f(X_n, Y_{n,t_{\delta t}}) - f(X_n, Y_t^n)] + 2\mathbb{E} v_t \cdot [f(X_n, Y_t^n) - f(X_n, z_t^n)] \\ &\quad + \mathbb{E} \|g(X_n, Y_{n,t_{\delta t}}) - g(X_n, z_t^n)\|^2 + 2\lambda_2 \mathbb{E} v_t \cdot [h(X_n, Y_{n,t_{\delta t}}) - h(X_n, Y_t^n)] \\ &\quad + 2\lambda_2 \mathbb{E} v_t \cdot [h(X_n, Y_t^n) - h(X_n, z_t^n)] + \lambda_2 \mathbb{E} |h(X_n, Y_{n,t_{\delta t}}) - h(X_n, z_t^n)|^2. \end{aligned}$$

Using Assumption A4,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &\leq 2\mathbb{E} v_t \cdot [f(X_n, Y_{n,t_{\delta t}}) - f(X_n, Y_t^n)] - 2\beta \mathbb{E} |v_t|^2 \\ &\quad + \mathbb{E} \|g(X_n, Y_{n,t_{\delta t}}) - g(X_n, z_t^n)\|^2 + 2\lambda_2 \mathbb{E} v_t \cdot [h(X_n, Y_{n,t_{\delta t}}) - h(X_n, Y_t^n)] \\ &\quad + 2\lambda_2 \mathbb{E} v_t \cdot [h(X_n, Y_t^n) - h(X_n, z_t^n)] + \lambda_2 \mathbb{E} |h(X_n, Y_{n,t_{\delta t}}) - h(X_n, z_t^n)|^2, \end{aligned}$$

followed by Cauchy-Schwarz inequality, and Assumption A2,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &\leq 2L_f \sqrt{\mathbb{E} |v_t|^2} \sqrt{\mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2} - 2\beta \mathbb{E} |v_t|^2 \\ &\quad + L_g \mathbb{E} |Y_{n,t_{\delta t}} - z_t^n|^2 + 2\lambda_2 L_h \sqrt{\mathbb{E} |v_t|^2} \sqrt{\mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2} \\ &\quad + 2\lambda_2 L_h \mathbb{E} |v_t|^2 + \lambda_2 L_h \mathbb{E} |Y_{n,t_{\delta t}} - z_t^n|^2 \\ &= 2(L_f + \lambda_2 L_h) \sqrt{\mathbb{E} |v_t|^2} \sqrt{\mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2} \\ &\quad - 2(\beta - \lambda_2 L_h) \mathbb{E} |v_t|^2 + (L_g + \lambda_2 L_h) \mathbb{E} |Y_{n,t_{\delta t}} - z_t^n|^2. \end{aligned}$$

Using Young's inequality,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &\leq 2(L_f + \lambda_2 L_h) \sqrt{\mathbb{E} |v_t|^2} \sqrt{\mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2} \\ &\quad - 2(\beta - \lambda_2 L_h) \mathbb{E} |v_t|^2 + 2(L_g + \lambda_2 L_h) \mathbb{E} |Y_t^n - z_t^n|^2 \\ &\quad + 2(L_g + \lambda_2 L_h) \mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2 \\ &= 2(L_f + \lambda_2 L_h) \sqrt{\mathbb{E} |v_t|^2} \sqrt{\mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2} \\ &\quad - 2(\beta - 2\lambda_2 L_h - L_g) \mathbb{E} |v_t|^2 \\ &\quad + 2(L_g + \lambda_2 L_h) \mathbb{E} |Y_{n,t_{\delta t}} - Y_t^n|^2. \end{aligned}$$

Using Lemma 2.5,

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &\leq 2(L_f + \lambda_2 L_h) \sqrt{\mathbb{E} |v_t|^2} \sqrt{K_4 \delta t} \\ &\quad - 2(\beta - 2\lambda_2 L_h - L_g) \mathbb{E} |v_t|^2 + 2(L_g + \lambda_2 L_h) K_4 \delta t. \end{aligned} \quad (3.5)$$

Let  $T_1 = \inf \{t : \mathbb{E} |v_t|^2 = 1\}$ . Since  $v_0 = 0$ , the line segment  $[0, T_1]$  is not empty. We first solve (3.5) 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

$$\begin{aligned} \frac{d}{dt} \mathbb{E} |v_t|^2 &\leq 2(L_f + \lambda_2 L_h) \sqrt{K_4 \delta t} \\ &\quad - 2(\beta - 2\lambda_2 L_h - L_g) \mathbb{E} |v_t|^2 + 2(L_g + \lambda_2 L_h) K_4 \delta t. \end{aligned}$$

For  $\delta t$  small enough, the third term on the right hand side is smaller than the first term. Hence

$$\frac{d}{dt} \mathbb{E} |v_t|^2 \leq -2(\beta - 2\lambda_2 L_h - L_g) \mathbb{E} |v_t|^2 + 4(L_f + \lambda_2 L_h) \sqrt{K_3 \delta t}.$$

Gronwall's inequality implies,

$$\mathbb{E} |v_t|^2 \leq \frac{(2L_f + \lambda_2 L_h) \sqrt{K_4 \delta t}}{\beta - 2\lambda_2 L_h - L_g}. \quad (3.6)$$

With  $\delta t$  small enough, the right hand side of (3.6) can be made smaller than one, hence (3.6) is valid for all  $t$ .  $\square$

*Proof.* of Lemma 2.8 By definition,

$$\begin{aligned} \mathbb{E} |\bar{a}(X_n) - A(X_n)|^2 &= \mathbb{E} \left| \int a(X_n, y) \mu^{X_n}(dy) - \frac{1}{M} \sum_{m=1}^M a(X_n, Y_m^n) \right|^2 \\ &\leq I_1^n + I_2^n, \end{aligned} \quad (3.7)$$

where

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

where  $z_t^n$  is the family of processes defined by (2.3).  $I_1^n$  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  $M$  equi-distanced sample points.  $I_2^n$  is the difference between empirical averages of  $a(X_n, \cdot)$  over  $M$  equi-distanced sample points, once for the process  $z_t^n$ , and once for its Euler approximation  $Y_m^n$ .

The estimation of  $I_1^n$ , is given in Lemma 2.6,

$$\begin{aligned} I_1^n &= 2\mathbb{E} \left[ \int a(X_n, y) \mu^{X_n}(dy) - \frac{1}{M} \sum_{m=1}^M a(X_n, z_m^n) \right]^2 \\ &\leq 2K_5 \left[ \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} \right]. \end{aligned} \quad (3.8)$$

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

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

Using Lemma 2.7 we get,

$$I_2^n \leq 2L^2 K_8 \sqrt{\delta t}. \quad (3.9)$$

Combining (3.8) and (3.9),

$$\mathbb{E} |\bar{a}(X_n) - A(X_n)|^2 \leq K_6 \left[ \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right], \quad (3.10)$$

which is uniform in  $n \leq T/\Delta t$   $\square$

*Proof.* of Lemma 2.9 Set  $E_n = \mathbb{E} \sup_{l \leq n} |\bar{X}_l - X_l|^2$ , then

$$\begin{aligned} E_n &= \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(\bar{X}_i) - A(X_i)] \Delta t + \sum_{i=0}^{l-1} [b(\bar{X}_i) - b(X_i)] \Delta W_i \right. \\ &\quad \left. + \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta P_i \right|^2 \\ &= \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(\bar{X}_i) - A(X_i)] \Delta t + \sum_{i=0}^{l-1} [b(\bar{X}_i) - b(X_i)] \Delta W_i \right. \\ &\quad \left. + \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta \hat{P}_i + \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta t \right|^2 \\ &\leq 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(\bar{X}_i) - A(X_i)] \Delta t \right|^2 + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [b(\bar{X}_i) - b(X_i)] \Delta W_i \right|^2 \\ &\quad + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta \hat{P}_i \right|^2 + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta t \right|^2. \end{aligned}$$

We split the first term on the right hand side,

$$\begin{aligned} E_n &\leq 8 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(\bar{X}_i) - \bar{a}(X_i)] \Delta t \right|^2 + 8 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(X_i) - A(X_i)] \Delta t \right|^2 \\ &\quad + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [b(\bar{X}_i) - b(X_i)] \Delta W_i \right|^2 \\ &\quad + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta \hat{P}_i \right|^2 + 4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta t \right|^2. \end{aligned} \quad (3.11)$$

The first and fifth terms on the right hand side are estimated using the Lipschitz continuity of  $\bar{a}$  and  $c$ ,

$$\begin{aligned}
8 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(\bar{X}_i) - \bar{a}(X_i)] \Delta t \right|^2 &\leq 8 \mathbb{E} \sup_{l \leq n} L^2 l \sum_{i=0}^{l-1} |\bar{X}_i - X_i|^2 \Delta t^2 \\
&\leq 8L^2 T \sum_{i=0}^{n-1} \mathbb{E} |\bar{X}_i - X_i|^2 \Delta t \leq 8L^2 T \sum_{i=0}^{n-1} E_i \Delta t \\
4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta t_i \right|^2 &\leq 4 \mathbb{E} \sup_{l \leq n} L^2 l \sum_{i=0}^{l-1} \mathbb{E} |\bar{X}_i - X_i|^2 \Delta t^2 \\
&\leq 4L^2 T \sum_{i=0}^{n-1} \mathbb{E} |\bar{X}_i - X_i|^2 \Delta t \leq 4L^2 T \sum_{i=0}^{n-1} E_i \Delta t.
\end{aligned} \tag{3.12}$$

The third and fourth terms are estimated using the Doob inequality for martingales followed by the Itô isometry, and the Lipschitz continuity of  $b$  and  $c$ ,

$$\begin{aligned}
4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [b(\bar{X}_i) - b(X_i)] \Delta W_i \right|^2 &= 16 \sum_{i=0}^{n-1} \mathbb{E} \|b(\bar{X}_i) - b(X_i)\|^2 \Delta t \\
&\leq 16L^2 \sum_{i=0}^{n-1} \mathbb{E} |\bar{X}_i - X_i|^2 \Delta t \leq 16L^2 \sum_{i=0}^{n-1} E_i \Delta t \\
4 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [c(\bar{X}_i) - c(X_i)] \Delta \hat{P}_i \right|^2 &= 16\lambda_1 \sum_{i=0}^{n-1} \mathbb{E} |c(\bar{X}_i) - c(X_i)|^2 \Delta t \\
&\leq 16\lambda_1 L^2 \sum_{i=0}^{n-1} \mathbb{E} |\bar{X}_i - X_i|^2 \Delta t \leq 16\lambda_1 L^2 \sum_{i=0}^{n-1} E_i \Delta t.
\end{aligned} \tag{3.13}$$

The second term on the right hand side of (3.11) can be bounded as follows:

$$8 \mathbb{E} \sup_{l \leq n} \left| \sum_{i=0}^{l-1} [\bar{a}(X_i) - A(X_i)] \Delta t \right|^2 \leq 8T^2 \max_{i < n} \mathbb{E} |\bar{a}(X_i) - A(X_i)|^2. \tag{3.14}$$

Combining eqs. (3.11)–(3.14) and Lemma 2.8, we obtain a discrete linear integral inequality,

$$\begin{aligned}
E_n &\leq 4L^2(2T + 4 + 4\lambda_1 + T) \sum_{i=1}^{n-1} E_i \Delta t \\
&\quad + 8T^2 K_6 \left( \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right),
\end{aligned}$$

with initial condition  $E_0 = 0$ . It follows that for sufficiently small  $\Delta t$ ,

$$\begin{aligned}
E_n &\leq 8T^2 K_6 \left( \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right) \{1 + 4L^2(2T + 4 + 4\lambda_1 + T)\Delta t\}^n \\
&\leq 8T^2 K_6 \left( \frac{-\log_\alpha M \delta t + 1}{M \delta t} + \frac{1}{M} + \sqrt{\delta t} \right) e^{4L^2(3T+4+4\lambda_1)T}.
\end{aligned}$$

This estimate proves the theorem with  $K_7 = 8T^2 K_6 e^{4L^2(3T+4+4\lambda_1)T}$ .  $\square$

**4. Weak Convergence.** In the previous section we discussed the convergence of the multiscale scheme to the averaged system. We analyzed the averaged system in the strong sense. In this case the Itô isometry assures that one can not average the diffusion coefficient or the jump coefficient. In this section we deal with the weak form of the effective dynamics. We analyze the convergence of the numerical method defined by eqs. (1.6), (1.7) and (1.8) to the solution  $\bar{x}_t$  of the effective dynamics (1.3) at time  $t = t_n$ . The Poisson processes in (1.1) are simple Poisson processes, i.e., the jump coefficients depend only on the current state just like the drift and the diffusion coefficients. However, in the effective dynamics, while the drift and the diffusion coefficients still only depend on the current position of the effective variable, the jump coefficient depends also on an additional random variable which is distributed like the invariant measure. Hence the Poisson process of the effective dynamics is not a simple Poisson process.

In this section we change Assumption A1 to the following:

A1. The functions  $a = a(x, y)$ ,  $b = b(x, y)$  and  $c = c(x, y)$  in (1.1a) are four times continuously differentiable with all partial derivatives having polynomial growth.

**THEOREM 4.1.** *Let  $\mathcal{F}(x)$  be four times continuously differentiable with polynomial growth. Let  $X_n$  be the solution of (1.8), let  $\bar{x}_t$  be the solution of (1.3), and let  $\bar{X}_n$  denote the Euler approximation for  $\bar{x}_t$ . Then*

$$\sup_{n \leq T/\Delta t} |\mathbb{E}\mathcal{F}(\bar{X}_n) - \mathbb{E}\mathcal{F}(X_n)| \leq C \left\{ \sqrt{\frac{-\log_\alpha M\delta t + 1}{M\delta t}} + \frac{1}{M} + \delta t^{1/2} + \Delta t^{1/2} \right\}.$$

*Proof.* of Theorem 4.1. Let  $N \leq T/\Delta t$ . Define an auxiliary function  $u(k, x)$  for  $k \leq N$  as follows:

$$u(N, x) = \mathcal{F}(x), \quad u(k, x) = \mathbb{E} \left[ u(k+1, x + \bar{a}(x)\Delta t + \bar{b}(x)\Delta W) \right].$$

The function  $u(k, x)$  shares the same properties as the functions  $\mathcal{F}, a, b, c$ . Note that  $u(0, x) = \mathbb{E}\mathcal{F}(\bar{X}_N)$  and  $u(N, X_N) = \mathcal{F}(X_N)$  which implies that we need to estimate,

$$\begin{aligned} |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| &= |\mathbb{E}(u(N, X_N) - u(0, x))| \\ &= \left| \mathbb{E} \left( \sum_{n=0}^{N-1} u(n+1, X_{n+1}) - u(n, X_n) \right) \right|. \end{aligned}$$

Add and subtract equal terms (the terms in the brackets  $\{\}$ )

$$\begin{aligned} &|\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| \\ &= \left| \mathbb{E} \left( \sum_{n=0}^{N-1} u(n+1, X_{n+1}) - u(n, X_n) - \left\{ u(n+1, \bar{X}_{n+1}^{n, X_n}) - u(n, X_n) \right\} \right) \right| \\ &= \left| \mathbb{E} \left( \sum_{n=0}^{N-1} u(n+1, X_{n+1}) - u(n+1, X_n) - \left\{ u(n+1, \bar{X}_{n+1}^{n, X_n}) - u(n+1, X_n) \right\} \right) \right|, \end{aligned}$$

where in the third line we exchange the second and fourth summands. In view of the

differentiability of  $u$ ,

$$\begin{aligned} & |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| \\ &= \left| \mathbb{E} \left( \sum_{n=0}^{N-1} \frac{\partial u}{\partial x}(n+1, X_n) \left\{ X_{n+1} - X_n - \left( \bar{X}_{n+1}^{n, X_n} - X_n \right) \right\} \right. \right. \\ & \left. \left. + \frac{\partial^2 u}{2\partial x^2}(n+1, X_n) \left\{ (X_{n+1} - X_n)^2 - \left( \bar{X}_{n+1}^{n, X_n} - X_n \right)^2 \right\} + R_n(X_{n+1}) + R_n(\bar{X}_{n+1}^{n, X_n}) \right) \right|, \end{aligned}$$

where

$$R_n(z) = \frac{\partial^3 u}{3!\partial x^3}(n+1, X_n + \theta_{n,z}(z - X_n))(z - X_n)^3,$$

and  $0 \leq \theta_{n,z} \leq 1$ . To bound this remainder we use the following: first the polynomial growth of  $u(n, x)$ , which allows us to use the bound  $\mathbb{E}|X_{n+1}|^p + \mathbb{E}|\bar{X}_{n+1}^{n, X_n}|^p < \infty$  [29]. Second, we use the moment time deviation estimate [29]  $\mathbb{E}|X_{n+1} - X_n|^3 + \mathbb{E}|\bar{X}_{n+1}^{n, X_n} - X_n|^3 \leq C_1 \Delta t^{3/2}$ . Hence we get,

$$\begin{aligned} & |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| \\ & \leq \sum_{n=0}^{N-1} \mathbb{E} \left\{ \left| \frac{\partial u}{\partial x} \right| \left| \mathbb{E}_{X_n} \left[ X_{n+1} - X_n - \left( \bar{X}_{n+1}^{n, X_n} - X_n \right) \right] \right| \right\} \\ & \quad + \mathbb{E} \left\{ \left| \frac{\partial^2 u}{2\partial x^2} \right| \left| \mathbb{E}_{X_n} \left[ (X_{n+1} - X_n)^2 - \left( \bar{X}_{n+1}^{n, X_n} - X_n \right)^2 \right] \right| + C_2 \Delta t^{3/2} \right\}. \end{aligned}$$

Multiply and divide by  $\Delta t$  and use the conditional infinitesimal moments of (1.3) to get

$$\begin{aligned} & |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| \\ &= \sum_{n=0}^{N-1} \Delta t \mathbb{E} \left\{ \left| \frac{\partial u}{\partial x} \right| \left| \mathbb{E}_{X_n} [A(X_n) + \lambda c(X_n, Y_n^m) - (\bar{a}(X_n) + \lambda_1 c(X_n, \xi_{X_n}))] \right| \right\} \\ &+ \sum_{n=0}^{N-1} \Delta t \mathbb{E} \left\{ \left| \frac{\partial^2 u}{2\partial x^2} \right| \left| \mathbb{E}_{X_n} [BB^T(X_n) + \lambda_1 c^2(X_n, Y_n^m) - (\bar{b}b^T(X_n) + \lambda_1 c^2(X_n, \xi_{X_n}))] \right| \right\} \\ &+ C_2 \Delta t^{1/2}. \end{aligned}$$

Using again the polynomial growth of the function  $u$  and its derivatives together with the moment estimates for JSDEs [29] we get,

$$\begin{aligned} & |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| \\ & \leq C_3 \sum_{n=0}^{N-1} \Delta t \mathbb{E} \{ |A(X_n) - \bar{a}(X_n) + \lambda_1 \mathbb{E}_{X_n} [c(X_n, Y_n^m) - c(X_n, \xi_{X_n})]| \} \\ & + C_4 \sum_{n=0}^{N-1} \Delta t \mathbb{E} \{ |BB^T(X_n) - \bar{b}b^T(X_n) + \lambda_1 \mathbb{E}_{X_n} [c^2(X_n, Y_n^m) - c^2(X_n, \xi_{X_n})]| \} \\ & + C_2 \Delta t^{1/2}. \end{aligned}$$

Using Lemma 2.8 we get

$$\begin{aligned} |\mathbb{E}\mathcal{F}(X_N) - \mathbb{E}\mathcal{F}(\bar{X}_N)| &\leq C_5 \sum_{n=0}^{N-1} \Delta t \left\{ \sqrt{\frac{-\log_\alpha M\delta t + 1}{M\delta t} + \frac{1}{M}} + \delta t^{1/2} \right\} + C_2 \Delta t^{1/2} \\ &= C \left\{ \sqrt{\frac{-\log_\alpha M\delta t + 1}{M\delta t} + \frac{1}{M}} + \delta t^{1/2} + \Delta t^{1/2} \right\}. \end{aligned}$$

□

**5. Discussion.** The main aim of this work is to show that multiscale integration schemes can be applied to jump-diffusion systems. We proved strong convergence uniformly in time for the case (1.6), (1.7a), (1.8) and (1.5), and we proved weak convergence for the case (1.6), (1.7), (1.8) and (1.3). Our analysis focuses on the simplest case, where both the macro- and micro-solvers use an Euler scheme, but the analysis can be extended to other schemes like [30, 26].

All along this work we used the fact that the fast dynamics is ergodic, and thus paid no attention to the way we choose the initial value of the fast dynamics. The main assumption behind this is that we have a separation of time scales and we explicitly know which the fast and which the slow variables are. Furthermore, as in [1], the formulas for the various terms in the equations are known, so that the fast dynamics can be simulated at fixed values of the slow variables, and averages over the computationally approximated invariant measures can be evaluated. It would be interesting to analyze the case where the right-hand-sides of the equations are not explicitly known, and we only have data from the full “black box” simulation (the “equation-free” case). Estimating the local effective dynamics [31, 32] from dynamic data becomes then an important consideration.

The situation becomes more complicated (and interesting) when separation of time scales is known (or hypothesized) to exist, but its parameterization is not known, i.e. we do not know which of the model variables or combination of variables is slow. It is also conceivable that the “micro simulator” is not the implementation of an explicit SDE, but rather a different stochastic process that can effectively be modeled as an SDE; in this case the formulas for its right-hand-side are not available, and have to be estimated “on the fly” from short bursts of direct stochastic simulation. One must then use simulation data to detect good parameterizations of the slow variables (see, for example, the diffusion map approach in [33]), and invent procedures for consistently initializing the full simulator given the present value of such a parameterization. This is the subject of present and future research.

**Acknowledgments.** This work was supported in part by the National Science Foundation under Grant DMS 04-32710, and by the Director, Office of Science, Computational and Technology Research, U.S. Department of Energy under Contract No. DE-AC03-76SF000098.

#### REFERENCES

- [1] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.*, 1(2):385–391, 2003.
- [2] C.W. Gear, I. Kevrekidis, and C. Theodoropoulos. “Coarse” integration/bifurcation analysis via microscopic simulators: Micro-Galerkin methods. *Comp. Chem. Engr.*, 26:941–963, 2002.

- [3] W. E, D. Liu, and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.*, 58(11):1544–1585, 2005.
- [4] D. Givon, I. G. Kevrekidis, and R. Kupferman. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.*, 4(4):707–729, 2006.
- [5] R. Z. Khas'minskiĭ. On the principle of averaging the Itô's stochastic differential equations. *Kybernetika (Prague)*, 4:260–279, 1968.
- [6] 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., Durham, N.C., 1976)*, Paper No. 6, pages ii+120 pp. Duke Univ. Math. Ser., Vol. III. Duke Univ., Durham, N.C., 1977.
- [7] A. Yu. Veretennikov. On an averaging principle for systems of stochastic differential equations. *Mat. Sb.*, 181(2):256–268, 1990.
- [8] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szűcs.
- [9] Y. Kifer. Stochastic versions of Anosov's and Neistadt's theorems on averaging. *Stoch. Dyn.*, 1(1):1–21, 2001.
- [10] H. J. Kushner. *Weak convergence methods and singularly perturbed stochastic control and filtering problems*, volume 3 of *Systems & Control: Foundations & Applications*. Birkhäuser Boston Inc., Boston, MA, 1990.
- [11] D. Givon. Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. *Multiscale Modeling & Simulation*, 6(2):577–594, 2007.
- [12] R. Rico-Martinez, C. W. Gear, and Ioannis G. Kevrekidis. Coarse projective KMC integration: forward/reverse initial and boundary value problems. *Journal of Computational Physics*, 196:474–489, 2004.
- [13] 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. *The Journal of Chemical Physics*, 118(22):10149–10156, 2003.
- [14] 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(4):1091–1106 (electronic), 2003.
- [15] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 2003.
- [16] W. E and B. Engquist. The heterogeneous multi-scale methods. *Comm. Math. Sci.*
- [17] G. Hummer and I. G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations. *The Journal of Chemical Physics*, 118(23):1076210773, 2003.
- [18] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.
- [19] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The Journal of Chemical Physics*, 117(15):6959–6969, 2002.
- [20] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics*, 118:4999–5010, 2003.
- [21] Y. Cao, D. T. Gillespie, and L. R. Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1):014116, 2005.
- [22] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J. Comput. Phys.*, 221(1):158–180, 2007.
- [23] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.
- [24] E. Papageorgiou and R. Sircar. Multiscale intensity models for single name credit derivatives. *Applied Mathematical Finance*, 2006.
- [25] J. Fouque, G. Papanicolaou, R. Sircar, and K. Solna. Multiscale stochastic volatility asymptotics. *Multiscale Model. Simul.*, 2(1):22–42 (electronic), 2003.
- [26] D. J. Higham and P. E. Kloeden. Convergence and stability of implicit methods for jump-diffusion systems. *Int. J. Numer. Anal. Model.*, 3(2):125–140, 2006.
- [27] J.-L. Menaldi and M. Robin. Invariant measure for diffusions with jumps. *Appl. Math. Optim.*, 40(1):105–140, 1999.

- [28] C. Yuan and J. Lygeros. Invariant measure of stochastic hybrid processes with jumps. *43th Conference of Decision and Control*, 3:3209–3214, 2004.
- [29] Y. Maghsoodi. Mean square efficient numerical solution of jump-diffusion stochastic differential equations. *Sankhyā Ser. A*, 58(1):25–47, 1996.
- [30] D. J. Higham and P. E. Kloeden. Numerical methods for nonlinear stochastic differential equations with jumps. *Numer. Math.*, 101(1):101–119, 2005.
- [31] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.
- [32] G. A. Pavliotis and A. M. Stuart. Parameter estimation for multiscale diffusions. *J. Stat. Phys.*, 127(4):741–781, 2007.
- [33] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Appl. Comput. Harmon. Anal.*, 21(1):113–127, 2006.