**Research in
the Mathematical Sciences**

RESEARCH Open Access

CrossMark

Fluctuations in the heterogeneous multiscale methods for fast–slow systems

David Kelly[*] and Eric Vanden-Eijnden

[*]Correspondence:
dtkelly@cims.nyu.edu
Courant Institute, New York
University, New York, NY, USA

Abstract

How heterogeneous multiscale methods (HMM) handle fluctuations acting on the slow variables in fast–slow systems is investigated. In particular, it is shown via analysis of central limit theorem (CLT) and large deviation principle (LDP) that the standard version of HMM artificially amplifies these fluctuations. A simple modification of HMM, termed parallel HMM, is introduced and is shown to remedy this problem, capturing fluctuations correctly both at the level of the CLT and the LDP. All results in this article assume the HMM speedup factor λ to be constant and in particular independent of the scale parameter ε. Similar type of arguments can also be used to justify that the τ-leaping method used in the context of Gillespie's stochastic simulation algorithm for Markov jump processes also captures the right CLT and LDP for these processes.

1 Background

The heterogeneous multiscale methods (HMM) [1,21,23,24] provide an efficient strategy for integrating fast–slow systems of the type

$$\frac{\mathrm{d}X^\varepsilon}{\mathrm{d}t} = f(X^\varepsilon, Y^\varepsilon),$$
$$\frac{\mathrm{d}Y^\varepsilon}{\mathrm{d}t} = \frac{1}{\varepsilon}g(X^\varepsilon, Y^\varepsilon). \tag{1.1}$$

The method relies on an averaging principle that holds under some assumption of ergodicity and states that as $\varepsilon \to 0$ the slow variables X^ε can be uniformly approximated by the solution to the following averaged equation

$$\dot{\bar{X}} = F(\bar{X}). \tag{1.2}$$

Here $F(x) = \int f(x,y)\mu_x(\mathrm{d}y)$ is the averaged vector field, with $\mu_x(\mathrm{d}y)$ being the ergodic invariant measure of the fast variables Y_x with a frozen x variable. This averaging principle is akin to the law of large number (LLN) in the present context and it suggests to simulate the evolution of the slow variables using (1.2) rather than (1.1) when ε is small. This requires to estimate $F(x)$, which typically has to be done on-the-fly given the current value of the slow variables. To this end, note that if Euler's method with time step Δt is used as integrator for the slow variables in (1.1), we can approximate $X^\varepsilon(n\Delta t)$ by x_n satisfying the recurrence

Springer Open

$$x_{n+1}^{\varepsilon} = x_n^{\varepsilon} + \int_{n\Delta t}^{(n+1)\Delta t} f(x_n^{\varepsilon}, Y_{x_n^{\varepsilon}}^{\varepsilon}(s))\mathrm{d}s, \tag{1.3}$$

where Y_x^{ε} denotes the solution to the second equation in (1.1) with X^{ε} kept fixed at the value x. If ε is small enough that $\Delta t/\varepsilon$ is larger than the mixing time of Y_x^{ε}, the Birkhoff integral in (1.4) is in fact close to the averaged coefficient in (1.2), in the sense that

$$F(x) \approx \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} f(x, Y_x^{\varepsilon}(s))\mathrm{d}s. \tag{1.4}$$

Therefore, (1.3) can also be thought of as an integrator for the averaged equation (1.2). In fact, when ε is small, one can obtain a good approximation of $F(x)$ using only a fraction of the macro-time step. In particular, we expect that

$$\frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} f(x, Y_x^{\varepsilon}(s))\mathrm{d}s \approx \frac{\lambda}{\Delta t} \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x, Y_x^{\varepsilon}(s))\mathrm{d}s =: F_{\Delta t}(x) \tag{1.5}$$

with $\lambda \geq 1$ provided that $\Delta t/(\varepsilon\lambda)$ remains larger than the mixing time of Y_x^{ε}. This observation is at the core of HMM-type methods—in essence, they amount to replacing (1.3) by

$$x_{n+1} = x_n + \Delta t \, F_{\Delta t}(x_n). \tag{1.6}$$

Since the number of computations required to compute the effective vector field $F_n(x)$ is reduced by a factor λ, this is also the speedup factor for an HMM-type method. We note that the choice of initial condition $Y_x^{\varepsilon}(0)$ is immaterial in principle, since almost surely all choices will lead to the same statistical average by the ergodic theorem. However, the choice is of computational consequence and will be elaborated on later.

From the argument above, it is apparent that there is another, equivalent way to think about HMM-type methods, as was first pointed out in [14] (see also [2,4,22,26]). Indeed, the integral defining $F_n(x)$ in (1.5) can be recast into an integral on the full interval $[n\Delta t, (n+1)\Delta t]$ by a change in integration variables, which amount to rescaling the internal clock of the variables Y_x^{ε}. In other words, HMM-type methods can also be thought of as approximating the fast–slow system in (1.1) by

$$\begin{aligned}
\frac{\mathrm{d}\widetilde{X}^{\varepsilon}}{\mathrm{d}t} &= f(\widetilde{X}^{\varepsilon}, \widetilde{Y}^{\varepsilon}), \\
\frac{\mathrm{d}\widetilde{Y}^{\varepsilon}}{\mathrm{d}t} &= \frac{1}{\varepsilon\lambda}g(\widetilde{X}^{\varepsilon}, \widetilde{Y}^{\varepsilon}).
\end{aligned} \tag{1.7}$$

If $\varepsilon \ll 1$, we can reasonably replace ε with $\varepsilon\lambda$, provided that this product still remains small—in particular, the evolution of the slow variables in (1.7) is still captured by the limiting equation (1.2). Hence, HMM-type methods are akin to artificial compressibility [9] in fluid simulations and Car–Parrinello methods [10] in molecular dynamics.

The approximations in (1.5) or (1.7) are perfectly reasonable if we are only interested in staying faithful to the averaged equation (1.2)—that is to say, HMM-type approximations will have the correct law of large numbers (LLN) behavior. However, the fluctuations about that average will be enhanced by a factor of λ. This is quite clear from the interpretation (1.7), since in the original model (1.1), the local fluctuations about the average are of order $\sqrt{\varepsilon}$ and in (1.7) they are of order $\sqrt{\varepsilon\lambda}$. The large fluctuations about the average caused by rare events are similarly inflated by a factor of λ. This can be an issue, for example, in metastable fast–slow systems, where the large fluctuations about the average determine

the waiting times for transitions between metastable states. In particular, we shall see that an HMM-type scheme drastically decreases these waiting times due to the enhanced fluctuations.

In this article, we propose a simple modification of HMM which corrects the problem of enhanced fluctuations. The key idea is to replace the approximation (1.5) with

$$\frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t}f(x,Y_x^{\varepsilon}(s))\mathrm{d}s \approx \sum_{j=1}^{\lambda}\frac{1}{\Delta t}\int_{n\Delta t}^{(n+1/\lambda)\Delta t}f(x,Y_x^{\varepsilon,j}(s))\mathrm{d}s, \tag{1.8}$$

where each $Y_x^{\varepsilon,j}$ is an independent copy of Y_x^{ε}. By comparing (1.5) with (1.8), we see that the first approximation is *essentially* replacing a sum of λ weakly correlated random variables with one random variable, multiplied by λ. This introduces correlations that should not be there and in particular results in enhanced fluctuations. In (1.8), we instead replace the sum of λ weakly correlated random variables with a sum of λ independent random variables. This is a much more reasonable approximation to make, since these random variables are becoming less and less correlated as ε gets smaller. Since the terms appearing on the right-hand side are independent of each other, they can be computed in parallel. Thus, if one has λ CPUs available; then, the real time of the computations is identical to HMM. For this reason, we call the modification the parallelized HMM (PHMM). Note that, in analogy to (1.7), one can interpret PHMM as approximating (1.1) by the system

$$\begin{aligned}\frac{\mathrm{d}\widetilde{X}^{\varepsilon}}{\mathrm{d}t} &= \frac{1}{\lambda}\sum_{j=1}^{\lambda}f(\widetilde{X}^{\varepsilon},\widetilde{Y}^{\varepsilon,j}),\\\frac{\mathrm{d}\widetilde{Y}^{\varepsilon,j}}{\mathrm{d}t} &= \frac{1}{\varepsilon\lambda}g(\widetilde{X}^{\varepsilon},\widetilde{Y}^{\varepsilon,j})\quad\text{for } j=1,\ldots,\lambda.\end{aligned} \tag{1.9}$$

It is clear that this approximation will be as good as (1.7) in terms of the LLN, but in contrast with (1.7), we will show below that it captures the fluctuations about the average correctly, both in terms of small Gaussian fluctuations via the CLT and large fluctuations describing rare events via a LDP. A similar observation in the context of numerical homogenization was made in [7,8]. It is also worth pointing out that for many fast–slow systems it is possible to extend CLT and LDP results to longer timescales on which rare events are no longer rare. While we will not address this scenario from a theoretical perspective (all our theoretical results will be restricted to $O(1)$ timescales), we do investigate numerically what happens on much larger timescales and find that the PHMM performs quite well in the particular examples we study. We stress, however, that such an extension will not be possible in general; see the discussion for further details.

We also stress that the theoretical results of this article are all obtained under the approximation scenario above, namely that we have discretized the slow variables and worked with fast variables Y_x^{ε} that solve the exact evolution equations, but with frozen x variables. In practice, one must in general discretize the fast variables, which adds another layer of complexity to the analysis of fluctuations. We restrict ourselves to the simpler theoretical setting for the sake of simplicity and to ensure 'proof of concept.' In the numerical investigations, we find that the theoretical results derived in the above scenario are robust even with relatively crude fast integrators. It has been shown [20] that numerical schemes do not always inherit the mixing properties of the underlying evolution equation. Thus, in some situations, it is advisable to use sophisticated methods that are

known to capture longtime statistics [5]. Due to the relative simplicity of the numerical models studied in this article, we do not encounter this problem and hence can employ simple methods.

It is important to note that the averaging approximation in (1.5) still holds for a class of λ that is ε-dependent, provided that $\varepsilon\lambda \to 0$ as $\varepsilon \to 0$. This is clearly a computational benefit, with greater timescale separation leading to greater computational speedup. In this article, we will assume for simplicity that λ does not depend on ε, and will comment on the results in the ε dependent case in Appendix 1.

The outline of the remainder of this article is as follows. In Section 2, we recall the averaging principle for stochastic fast–slow systems and describe how to characterize the fluctuations about this average, including local Gaussian fluctuations and large deviation principles. In Sect. 3, we recall the HMM-type methods. In Sect. 4, we show that they lead to enhanced fluctuations. In Sect. 5, we introduce the PHMM modification, and in Sect. 6, we show that this approximation yields the correct fluctuations, both in terms of local Gaussian fluctuations and large deviations. In Sect. 7, we test PHMM for a variety of simple models and conclude in Sect. 8 with a discussion.

2 Average and fluctuations in fast–slow systems

For simplicity, we will from here on assume that the fast variables are stochastic. This assumption is convenient, but not necessary, since all the averaging and fluctuation properties stated below are known to hold for large classes of fast–slow systems with deterministically chaotic fast variables [12,17–19]. The fast–slow systems we investigate are given by

$$
\begin{aligned}
\frac{\mathrm{d}X^\varepsilon}{\mathrm{d}t} &= f(X^\varepsilon, Y^\varepsilon), \\
\mathrm{d}Y^\varepsilon &= \frac{1}{\varepsilon}g(X^\varepsilon, Y^\varepsilon)\mathrm{d}t + \frac{1}{\sqrt{\varepsilon}}\sigma(X^\varepsilon, Y^\varepsilon)\mathrm{d}W,
\end{aligned}
\tag{2.1}
$$

where $f : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^d, g : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^e, \sigma : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^e \times \mathbb{R}^e$, and W is a standard Wiener process in $\mathbb{R}^e$. We assume that for every $x \in \mathbb{R}^d$, the Markov process described by the SDE

$$
\mathrm{d}Y_x = g(x, Y_x)\mathrm{d}t + \sigma(x, Y_x)\mathrm{d}W
\tag{2.2}
$$

is ergodic, with invariant measure μ_x, and has sufficient mixing properties. For full details on the necessary mixing properties, see, for instance [15].

In this section, we briefly recall the averaging principle for stochastic fast–slow systems and discuss two results that characterize the fluctuations about the average, the central limit theorem (CLT) and the large deviations principle (LDP).

2.1 Averaging principle

As $\varepsilon \to 0$, each realization of X^ε, with initial condition $X^\varepsilon(0) = x$, tends toward a trajectory of a deterministic system

$$
\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = F(\bar{X}), \quad \bar{X}(0) = x,
\tag{2.3}
$$

where $F(x) = \int f(x,y)\mu_x(\mathrm{d}y)$ and μ_x is the invariant measure corresponding to the Markov process $\mathrm{d}Y_x = g(x, Y_x)\mathrm{d}t + \sigma(x, Y_x)\mathrm{d}W$. The convergence is in an almost sure and uniform sense:

$$\lim_{\varepsilon \to 0} \sup_{t \le T} |X^\varepsilon(t) - \bar{X}(t)| = 0$$

for every fixed $T > 0$, every choice of initial condition x and almost surely every initial condition $Y^\varepsilon(0)$ (a.s. with respect to μ_x) as well as every realization of the Brownian paths driving the fast variables. Details of this convergence result in the setting above are given in (for instance) [15, Chapter 7.2].

2.2 Small fluctuations: CLT

The *small* fluctuations of X^ε about the averaged system $\bar{X}$ can be understood by characterizing the limiting behavior of

$$Z^\varepsilon := \frac{X^\varepsilon - \bar{X}}{\sqrt{\varepsilon}},$$

as $\varepsilon \to 0$. It can be shown that the process Z^ε converges in distribution (on the space of continuous functions $C([0, T]; \mathbb{R}^d)$ endowed with the sup-norm topology) to a process Z defined by the SDE

$$\mathrm{d}Z = B_0(\bar{X})Z\mathrm{d}t + \eta(\bar{X})\mathrm{d}V, \quad Z(0) = 0, \tag{2.4}$$

Here $\bar{X}$ solves the averaged system in (2.3), V is a standard Wiener process, $B_0 := B_1 + B_2$ with

$$B_1(x) = \int \nabla_x f(x, y)\mu_x(\mathrm{d}y),$$

$$B_2(x) = \int_0^\infty \mathrm{d}\tau \int \mu_x(\mathrm{d}y)\nabla_y \mathbf{E}_y\Big(\tilde{f}(x, Y_x(\tau))\Big)\nabla_x g(x, y)$$

and

$$\eta(x)\eta^T(x) = \int_0^\infty \mathrm{d}\tau \, \mathbf{E}\tilde{f}(x, Y_x(0)) \otimes \tilde{f}(x, Y_x(\tau)),$$

where $\tilde{f}(x, y) = f(x, y) - F(x)$, $\mathbf{E}_y$ denotes expectation over realizations of Y_x with $Y_x(0) = y$, and $\mathbf{E}$ denotes expectation over realization of Y_x with $Y_x(0) \sim \mu_x$. We include next a formal argument deriving this limit, as it will prove useful when analyzing the multiscale approximation methods. We will replicate the argument given in [6]; a more complete and rigorous argument can be found in [15, Chapter 7.3].

First, we write a system of equations for the triple $(\bar{X}, Z^\varepsilon, Y^\varepsilon)$ in the following approximated form, which uses nothing more than Taylor expansions of the original system in (1.1):

$$\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = F(\bar{X}),$$

$$\frac{\mathrm{d}Z^\varepsilon}{\mathrm{d}t} = \frac{1}{\sqrt{\varepsilon}}\tilde{f}(\bar{X}, Y^\varepsilon) + \nabla_x f(\bar{X}, Y^\varepsilon) + O(\sqrt{\varepsilon}),$$

$$\mathrm{d}Y^\varepsilon = \frac{1}{\varepsilon}g(\bar{X}, Y^\varepsilon)\mathrm{d}t + \frac{1}{\sqrt{\varepsilon}}\nabla_x g(\bar{X}, Y^\varepsilon)Z^\varepsilon \mathrm{d}t + \frac{1}{\sqrt{\varepsilon}}\sigma(\bar{X}, Y^\varepsilon)\mathrm{d}W + O(1).$$

We now proceed with a classical perturbation expansion on the generator of the triple $(\bar{X}, Z^\varepsilon, Y^\varepsilon)$. In particular, we have $\mathscr{L}_\varepsilon = \frac{1}{\varepsilon}\mathscr{L}_0 + \frac{1}{\sqrt{\varepsilon}}\mathscr{L}_1 + \mathscr{L}_2 + \cdots$ where

$$\mathscr{L}_0 = g(x, y) \cdot \nabla_y + a(x, y) : \nabla_y^2,$$
$$\mathscr{L}_1 = \tilde{f}(x, y) \cdot \nabla_z + (\nabla_x g(x, y)z) \cdot \nabla_y,$$
$$\mathscr{L}_2 = F(x) \cdot \nabla_x + (\nabla_x f(x, y)z) \cdot \nabla_z + \text{operator in } y,$$

and $a = \sigma\sigma^T$. Let $u_\varepsilon(x, z, y, t) = \mathbf{E}_{(x,z,y)}\varphi(\bar{X}(t), Z^\varepsilon(t), Y^\varepsilon(t))$ and introduce the ansatz $u_\varepsilon = u_0 + \sqrt{\varepsilon}u_1 + \varepsilon u_2 + \cdots$. By substituting u_ε into $\partial_t u_\varepsilon = \mathscr{L}_\varepsilon u_\varepsilon$ and equating powers of ε, we obtain

$$O(\varepsilon^{-1}) : \mathscr{L}_0 u_0 = 0,$$
$$O(\varepsilon^{-1/2}) : \mathscr{L}_0 u_1 = -\mathscr{L}_1 u_0,$$
$$O(\varepsilon^{-1}) : \partial_t u_0 = \mathscr{L}_2 u_0 + \mathscr{L}_1 u_1 + \mathscr{L}_0 u_2.$$

From the $O(\varepsilon^{-1})$ identity, we obtain $u_0 = u_0(x, z, t)$, confirming that the leading order term is independent of y. By the Fredholm alternative, the $O(\varepsilon^{-1/2})$ identity has a solution u_1 which has the Feynman–Kac representation

$$u_1(x, y, z) = \int_0^\infty \mathrm{d}\tau\, \mathbf{E}_y\left(\tilde{f}(x, Y_x(\tau))\right) \cdot \nabla_z u_0(x, z),$$

where Y_x denotes the Markov process generated by $\mathscr{L}_0$, i.e., the solution of (2.2). Finally, if we average the $O(1)$ identity against the invariant measure corresponding to $\mathscr{L}_0$, we obtain

$$\partial_t u_0 = F(x)\nabla_x u_0 + \int \mu_x(\mathrm{d}y)(\nabla_x f(x, y)z) \cdot \nabla_z u_0$$
$$+ \int \mu_x(\mathrm{d}y)\int_0^\infty \mathrm{d}\tau \tilde{f}(x, y) \otimes \mathbf{E}_y\tilde{f}(x, Y_x(\tau)) : \nabla_z^2 u_0$$
$$+ \int \mu_x(\mathrm{d}y)(\nabla_x g(x, y)z)\int_0^\infty \mathrm{d}\tau\, \nabla_y\mathbf{E}_y\tilde{f}(x, Y_x(\tau))\nabla_z u_0.$$

Clearly, this is the forward Kolmogorov equation for the Markov process $(\bar{X}, Z)$ defined by

$$\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = F(\bar{X}),$$
$$dZ = B_0(\bar{X})Z\mathrm{d}t + \eta(\bar{X})\mathrm{d}V$$

with B_0 and η defined as above.

2.3 Large fluctuations: LDP

A large deviation principle (LDP) for the fast–slow system (2.1) quantifies probabilities of $O(1)$ fluctuations of X^ε away from the averaged trajectory $\bar{X}$. The probability of such events vanishes exponentially quickly and as a consequence is not accounted for by the CLT fluctuations; hence, a LDP accounts for the *rare events*.

We say that the slow variables X^ε satisfy a *large deviation principle* (LDP) with action functional $\mathscr{S}_{[0,T]}$ if for any set $\Gamma \subset \{\gamma \in C([0, T], \mathbb{R}^d) : \gamma(0) = x\}$ we have

$$
-\inf_{\gamma \in \mathring{\Gamma}} \mathscr{S}_{[0,T]}(\gamma) \le \liminf_{\varepsilon \to 0} \varepsilon \log \mathbf{P}\left(X^\varepsilon \in \Gamma\right)
$$
$$
\le \limsup_{\varepsilon \to 0} \varepsilon \log \mathbf{P}\left(X^\varepsilon \in \Gamma\right) \le -\inf_{\gamma \in \bar{\Gamma}} \mathscr{S}_{[0,T]}(\gamma), \tag{2.5}
$$

where $\mathring{\Gamma}$ and $\bar{\Gamma}$ denote the interior and closure of Γ, respectively.

An LDP also determines many important features of $O(1)$ fluctuations that occur on large timescales, such as the probability of transition from one metastable set to another. For example, suppose that X^ε is known to satisfy an LDP with action functional $\mathscr{S}_{[0,T]}$. Let D be an open domain in $\mathbb{R}^d$ with smooth boundary ∂D and let $x^* \in D$ be an asymptotically stable equilibrium for the averaged system $\dot{\bar{X}} = F(\bar{X})$. When $\varepsilon \ll 1$, we expect that a trajectory of X^ε that starts in D will tend toward the equilibrium x^* and exhibit $O(\sqrt{\varepsilon})$ fluctuations about the equilibrium—these fluctuations are described by the CLT. On very large timescales, these small fluctuations have a chance to 'pile up' into an $O(1)$ fluctuation, producing behavior of the trajectory that would be considered impossible for the averaged system. Such fluctuations are not accurately described by the CLT and require the LDP instead. For example, the asymptotic behavior of escape time from the domain D,

$$
\tau^\varepsilon = \inf\{t > 0 : X^\varepsilon(t) \notin D\},
$$

can be quantified in terms of the *quasi-potential* defined by

$$
\mathscr{V}(x, y) = \inf_{T > 0} \inf_{\gamma(0)=x, \gamma(T)=y} \mathscr{S}_{[0,T]}(\gamma). \tag{2.6}
$$

Under natural conditions, it can be shown that for any $x \in D$

$$
\lim_{\varepsilon \to 0} \varepsilon \log \mathbf{E}_x \tau^\varepsilon = \inf_{y \in \partial D} \mathscr{V}(x^*, y).
$$

Hence, the time it takes to pass from the neighborhood of one equilibrium to another may be quantified using the LDP. Details on the escape time of fast–slow systems can be found in [15, Chapter 7.6].

LDPs for fast–slow systems of the type (2.1) are well understood [15, Chapter 7.4]. First define the Hamiltonian $\mathcal{H} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by

$$
\mathcal{H}(x, \theta) = \lim_{T \to \infty} \frac{1}{T} \log \mathbf{E}_y \exp\left(\theta \cdot \int_0^T f(x, Y_x(s)) \mathrm{d}s\right), \tag{2.7}
$$

where Y_x denotes the Markov process governed by $\mathrm{d}Y_x = g(x, Y_x)\mathrm{d}t + \sigma(x, Y_x)\mathrm{d}W$. Let $\mathscr{L} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the Legendre transform of $\mathcal{H}$:

$$
\mathscr{L}(x, \beta) = \sup_\theta \left(\theta \cdot \beta - \mathcal{H}(x, \theta)\right). \tag{2.8}
$$

Then the action functional is given by

$$
\mathscr{S}_{[0,T]}(\gamma) = \int_0^T \mathscr{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s. \tag{2.9}
$$

It can also be shown that the function $u(t, x) = \inf_{\gamma(0)=x} \mathscr{S}_{[0,t]}(\gamma)$ satisfies the Hamilton–Jacobi equation

$$
\partial_t u(t, x) = \mathcal{H}(x, \nabla u(t, x)). \tag{2.10}
$$

Donsker–Varadhan theory tells us that the connection between Hamilton–Jacobi equations and LDPs is in fact much deeper. Firstly, *Varadhan's lemma* states that if a process X^ε

is known to satisfy an LDP with some associated Hamiltonian $\mathcal{H}$, then for any $\varphi : \mathbb{R}^d \to \mathbb{R}$ we have the generalized Laplace method-type result

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbf{E}_x \exp \left(\varepsilon^{-1} \varphi(X^\varepsilon(t)) \right) = S_t \varphi(x), \tag{2.11}$$

where S_t is the semigroup associated with the Hamilton–Jacobi equation $\partial_t u = \mathcal{H}(x, \nabla u)$. Conversely, if it is known that (2.11) holds for all (x, t) and a suitable class of φ, then the inverse Varadhan's lemma states that X^ε satisfies an LDP with action functional given by (2.8), (2.9). Hence, we can use (2.11) to determine the action functional for a given process.

In the next few sections, we will exploit both sides of Varadhan's lemma when investigating the large fluctuations of the HMM and related schemes. More complete discussions on Varadhan's lemma can be found in [13, Chapters 4.3, 4.4].

3 HMM for fast–slow systems

When applied to the stochastic fast–slow system (2.1), HMM-type schemes rely on the fact that the slow X^ε variables, and the coefficients that govern them, converge to a set of reduced variables as ε tends to zero. We will describe a simplest version of the method below, which is more convenient to deal with mathematically.

Before proceeding, we digress briefly on notation. When referring to continuous time variables, we will always use uppercase symbols (X^ε, Y^ε, etc.), and when referring to discrete time approximations, we will always use lowercase symbols (x_n^ε, y_n^ε, etc.). We will also encounter continuous time variables whose definition depends on the integer n for which we have $t \in [n\Delta t, (n+1)\Delta t)$. We will see below that such continuous time variables are used to define discrete time approximations. In this situation, we will use uppercase symbols with a subscript n (e.g., X_n^ε).

Let us now describe a 'high-level' version of HMM. Fix a step size Δt and define the intervals $I_{n,\Delta t} := [n\Delta t, (n+1)\Delta t)$. On each interval $I_{n,\Delta t}$, we update $x_n^\varepsilon \approx X^\varepsilon(n\Delta t)$ to $x_{n+1}^\varepsilon \approx X^\varepsilon((n+1)\Delta t)$ via an iteration of the following two steps:

1. (Micro-step) Integrate the fast variables over the interval $I_{n,\Delta t}$, with the slow variable frozen at $X^\varepsilon = x_n^\varepsilon$. That is, the fast variables are approximated by

$$Y_n^\varepsilon(t) = Y_n^\varepsilon(n\Delta t) + \frac{1}{\varepsilon} \int_{n\Delta t}^t g(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}s + \frac{1}{\sqrt{\varepsilon}} \int_{n\Delta t}^t \sigma(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}W(s) \quad (3.1)$$

for $n\Delta t \le t \le (n + 1/\lambda)\Delta t$ with some $\lambda \ge 1$ (that is, we do not necessarily integrate the Y_n^ε variables over the whole time window). Due to the ergodicity of Y_x, the initialization of Y_n^ε is not crucial to the performance of the algorithm. It is, however, convenient to use $Y_{n+1}^\varepsilon(0) = Y_n^\varepsilon((n + 1/\lambda)\Delta t)$, since this reinitialization leads to the interpretation of the HMM scheme given in (3.5) below.

2. (Macro-step) Use the time series from the micro-step to update x_n^ε to x_{n+1}^ε via

$$x_{n+1}^\varepsilon = x_n^\varepsilon + \lambda \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}s. \tag{3.2}$$

Note that we do not require Y_n^ε over the whole Δt time step, but only a fraction of the step large enough for Y_n^ε to mix. Indeed, if ε is small enough, we have the approximate equality

$$\frac{\lambda}{\Delta t} \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s)) \mathrm{d}s \approx \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s)) \mathrm{d}s$$

since both sides are close the ergodic mean $\int f(x_n^\varepsilon, y) \mathrm{d}\mu_{x_n^\varepsilon}(y)$.

Clearly, the efficiency of the methods comes from the fact that we do not need to compute the fast variables on the whole time interval $I_{n,\Delta t}$ but only a $1/\lambda$ fraction of it. Hence, λ should be considered the *speedup factor* of HMM. Note that λ can only take moderate values for the above method to be justifiable; in particular, we require that $\varepsilon/\lambda \gg 1$.

As already stated, the algorithm above is a high-level version, in that one must do further approximations to make the method implementable. For example, one typically must specify some approximation scheme to integrate (3.1); for instance with Euler–Maruyama, we compute the time series by

$$y_{n,m+1}^\varepsilon = y_{n,m}^\varepsilon + \frac{\delta t}{\varepsilon} g(x_n^\varepsilon, y_{n,m}^\varepsilon) + \sqrt{\frac{\delta t}{\varepsilon}} \sigma(x_n^\varepsilon, y_{n,m}^\varepsilon) \xi_{n,m}, \tag{3.3}$$

where $0 \le m \le M$ is the index within the micro-step, $\xi_{n,m}$ are i.i.d. standard Gaussians and the microscale step size δt is much smaller than the macroscale step size Δt. In the macro-step, we would similarly have

$$x_{n+1}^\varepsilon = x_n^\varepsilon + \Delta t \, F_n(x_n^\varepsilon), \tag{3.4}$$

where $F_n(x) = \frac{1}{M} \sum_{m=1}^{M} f(x, y_{n,m}^\varepsilon)$ and $M = \Delta t/(\delta t \lambda)$.

The following observation, which is taken from [14], will allow us to easily describe the average and fluctuations of the above method. On each interval $I_{n,\Delta t}$, the high-level HMM scheme described above is equivalently given by $x_{n+1}^\varepsilon = X_n^\varepsilon((n+1)\Delta t)$, where X_n^ε solves the system

$$\begin{aligned}
\frac{\mathrm{d}X_n^\varepsilon}{\mathrm{d}t} &= f(x_n^\varepsilon, \widetilde{Y}_n^\varepsilon), \\
\mathrm{d}\widetilde{Y}_n^\varepsilon &= \frac{1}{\varepsilon\lambda} g(x_n^\varepsilon, \widetilde{Y}_n^\varepsilon) \mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}} \sigma(x_n^\varepsilon, \widetilde{Y}_n^\varepsilon) \mathrm{d}B,
\end{aligned} \tag{3.5}$$

defined on the interval $n\Delta t \le t \le (n+1)\Delta t$, with the initial condition $X_n^\varepsilon(n\Delta t) = x_n^\varepsilon$. This can be checked by a simple rescaling of time. It is clear that the efficiency of HMM essentially comes from saying that the fast–slow system is not drastically changed if one replaces ε with the slightly larger, but still very small $\varepsilon\lambda$.

4 Average and fluctuations in HMM methods

In this section, we investigate whether the limit theorems discussed in Sect. 2, i.e., the averaging principle, the CLT fluctuations and the LDP fluctuations, are also valid in the HMM approximation for a fast–slow system. We will see that the averaging principle is the only property that holds, and that both types of fluctuations are *inflated* by the HMM method. It is important to note that the theory developed in this section (and likewise in Sect. 6) is to understand the averaging and fluctuation properties of the HMM approximation (3.2), where the slow variables have been discretized, but the fast variables have not. We do not make any theoretical claims about the fully discretized case. We also note that the LLN, CLT and LDP results derived for (3.2) can be used to derive the same results for the non-discretized system (3.5). In particular, the CLT and LDP of (3.5) are not the same as the original fast slow system (2.1).

4.1 Averaging

By construction, HMM-type schemes capture the correct averaging principle. More precisely, if we take $\varepsilon \to 0$, then the sequence x_n^ε converges to some $\bar{x}_n$, where $\bar{x}_n$ is a numerical approximation of the true averaged system $\bar{X}$. If this numerical approximation is well posed, the limits $\varepsilon \to 0$ and $\Delta t \to 0$ commute with one another. Hence, the HMM approximation x_n^ε is consistent, in that it features approximately the same averaging behavior as the original fast–slow system.

We will argue the claim by induction. Suppose that for some $n \geq 0$ we know that $\lim_{\varepsilon\to 0} x_n^\varepsilon = \bar{x}_n$ (the $n = 0$ claim is trivial, since they are both simply the initial condition). Then, using the representation (3.5) we know that $x_{n+1}^\varepsilon = X_n^\varepsilon((n+1)\Delta t)$ where $X_n^\varepsilon(n\Delta t) = x_n^\varepsilon$. Since (3.5) is a fast–slow system of the form (2.1) we can apply the averaging principle from Sect. 2. In particular, it follows that $X_n^\varepsilon \to \bar{X}_n$ uniformly (and almost surely) on $I_{n,\Delta t}$, where $\bar{X}_n$ satisfies the averaged ODE

$$\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} = \int f(\bar{x}_n, y)\mu_{\bar{x}_n}(\mathrm{d}y) = F(\bar{x}_n).$$

Since the right-hand side is a constant, it follows that $x_{n+1}^\varepsilon \to \bar{x}_{n+1}$ as $\varepsilon \to 0$, where

$$\bar{x}_{n+1} = \bar{x}_n + F(\bar{x}_n)\Delta t.$$

This is nothing more than the Euler approximation of the true averaged variables $\bar{X}$, which completes the induction and hence the claim.

Introducing an integrator into the micro-step will make things more complicated, as the invariant measures appearing will be those of the discretized fast variables. In [20], it is shown that discretizations of SDEs often do not possess the ergodic properties of the original system. For those situations where no such issues arise, rigorous arguments concerning this scenario, including rates of convergence for the schemes, are given in [25].

4.2 Small fluctuations

For HMM-type methods, the CLT fluctuations about the average become inflated by a factor of $\sqrt{\lambda}$. That is, if we define

$$z_{n+1}^\varepsilon = \frac{x_{n+1}^\varepsilon - \bar{x}_{n+1}}{\sqrt{\varepsilon}},$$

then as $\varepsilon \to 0$, the fluctuations described by z_{n+1}^ε are not consistent with (2.4), but rather with the SDE

$$\mathrm{d}Z = g(\bar{X})Z\mathrm{d}t + \sqrt{\lambda}\eta(\bar{X})\mathrm{d}V, \quad Z(0) = 0, \tag{4.1}$$

where $\bar{X}$ satisfies the correct averaged system.

As above, by consistency we mean that when we take $\varepsilon \to 0$, the sequence $\{z_n^\varepsilon\}_{n\geq 0}$ converges to some well-posed discretization of the SDE (4.1). Since $Z(0) = 0$, it is easy to see that the solution to this equation is simply $\sqrt{\lambda}$ times the solution of (2.4). Hence, the fluctuations of the HMM-type scheme are inflated by a factor of $\sqrt{\lambda}$.

It is convenient to look instead at the rescaled fluctuations

$$\hat{z}_n^\varepsilon = z_n^\varepsilon/\sqrt{\lambda} = \frac{x_n^\varepsilon - \bar{x}_n}{\sqrt{\varepsilon\lambda}},$$

since this allows us to reproduce the argument from Sect. 2.2, with $\varepsilon' = \varepsilon\lambda$ playing the role of ε. We will again argue by induction, assuming for some $n \geq 0$ that $\hat{z}_n^\varepsilon \to \hat{z}_n$ as $\varepsilon \to 0$ (the $n = 0$ case is trivial).

The rescaled fluctuations are given by $\hat{z}^{\varepsilon}_{n+1} = Z^{\varepsilon}_n((n+1)\Delta t)$ where $Z^{\varepsilon}_n(t) = (X^{\varepsilon}_n(t) - \bar{X}_n(t))/\sqrt{\varepsilon\lambda}$ and $X^{\varepsilon}_n(t)$ is governed by the system (3.5) with initial condition $X^{\varepsilon}_n(n\Delta t) = x^{\varepsilon}_n$ and $\bar{X}_n$ satisfies

$$\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} = F(\bar{x}_n)$$

with initial condition $\bar{X}_n(n\Delta t) = \bar{x}_n$. We can then obtain the reduced equations for the pair $(X^{\varepsilon}_n, Z^{\varepsilon}_n)$ by arguing exactly as in Sect. 2. Indeed, the triple $(\bar{X}_n, Z^{\varepsilon}_n, \widetilde{Y}^{\varepsilon}_n)$ is governed by the system

$$
\begin{aligned}
\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} &= F(\bar{x}_n), \\
\frac{\mathrm{d}\widehat{Z}^{\varepsilon}_n}{\mathrm{d}t} &= \frac{1}{\sqrt{\varepsilon\lambda}}\tilde{f}(\bar{x}_n, \widetilde{Y}^{\varepsilon}_n) + \nabla_x f(\bar{x}_n, \widetilde{Y}^{\varepsilon}_n)\hat{z}_n + O(\sqrt{\varepsilon\lambda}), \\
\mathrm{d}\widetilde{Y}^{\varepsilon}_n &= \frac{1}{\varepsilon\lambda}g(\bar{x}_n, \widetilde{Y}^{\varepsilon}_n)\mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}}\nabla_x g(\bar{x}_n, \widetilde{Y}^{\varepsilon}_n)\hat{z}_n\mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}}\sigma(\bar{x}^{\varepsilon}_n, \widetilde{Y}^{\varepsilon}_n)\mathrm{d}W + O(1).
\end{aligned}
$$

From here on, we can carry out the calculation precisely as in Sect. 2.2, with the added convenience of the vector fields no longer depending on x as a variable. In doing so, we obtain $\widehat{Z}^{\varepsilon}_n \to \widehat{Z}_n$ (in distribution) as $\varepsilon \to 0$, where

$$\mathrm{d}\widehat{Z}_n = B_0(\bar{x}_n)\hat{z}_n\mathrm{d}t + \eta(\bar{x}_n)\mathrm{d}V,$$

with the initial condition defined recursively by $\widehat{Z}_n(n\Delta t) = \hat{z}_n$. Using the fact that $\hat{z}_{n+1} = \widehat{Z}_n((n+1)\Delta t)$, we obtain

$$\hat{z}_{n+1} = \hat{z}_n + B_0(\bar{x}_n)\hat{z}_n\Delta t + \eta(\bar{x}_n)\sqrt{\Delta t}\xi_n$$

where ξ_n are i.i.d. standard Gaussians. Hence, we obtain the Euler–Maruyama scheme for the correct CLT (2.4). However, since $\hat{z}^{\varepsilon}_n$ describes the rescaled fluctuations, we see that the true fluctuations z^{ε}_n of HMM are consistent with the inflated (4.1).

4.3 Large fluctuations

As with the CLT, the LDP of the HMM scheme is not consistent with the true LDP of the fast–slow system, but rather a rescaled version of the true LDP. In particular, define $u_{\lambda,\Delta t}$ by

$$u_{\lambda,\Delta t}(t, x) = \lim_{\varepsilon\to 0} \varepsilon \log \mathbf{E}_x \exp\left(\frac{1}{\varepsilon}\varphi(x^{\varepsilon}_{n+1})\right)$$

for $t \in I_{n,\Delta t}$. If the $O(1)$ fluctuations of HMM were consistent with those of the fast–slow system, we would expect $u_{\lambda,\Delta}$ to converge to the solution of (2.10) as $\Delta t \to 0$. Instead, we find that as $\Delta t \to 0$, $u_{\lambda,\Delta t}(t, x)$ converges to the solution to the Hamilton–Jacobi equation

$$\partial_t u_\lambda = \frac{1}{\lambda}\mathcal{H}(x, \lambda\nabla u_\lambda) \qquad u_\lambda(0, x) = \varphi(x). \tag{4.2}$$

In light of the discussion in Sect. 2.3, the reverse Varadhan lemma suggests that the HMM scheme is consistent with the wrong LDP. Before proving this claim, we first discuss some implications.

The rescaled Hamilton–Jacobi equation implies that the action functional for HMM will be a rescaled version of that for the true fast–slow system. Indeed, it is easy to see that the Lagrangian corresponding to HMM simplifies to

$$\widehat{\mathcal{L}}(x, \beta) := \sup_\theta\left(\theta\cdot\beta - \frac{1}{\lambda}\mathcal{H}(x, \lambda\theta)\right) = \frac{1}{\lambda}\mathcal{L}(x, \beta),$$

where $\mathscr{L}$ is the Lagrangian for the true fast–slow system. Thus, the action of the HMM approximation is given by $\widehat{\mathscr{T}}_{[0,T]} = \lambda^{-1}\mathscr{S}_{[0,T]}$ where $\mathscr{S}$ is the action of the true fast–slow system.

In particular, it follows immediately from the definition that the HMM approximation has quasi-potential $\widehat{\mathscr{V}}(x, y) = \lambda^{-1}\mathscr{V}(x, y)$, where $\mathscr{V}$ is the true quasi-potential. As a consequence, the escape times for the HMM scheme will be drastically faster than those of the fast–slow system. In the terminology of Sect. 2.3, if we let $\tau^{\varepsilon,\Delta t}$ be the escape time for the HMM scheme then for $\varepsilon, \Delta t \ll 1$ we expect

$$\mathbf{E}\tau^{\varepsilon,\Delta t} \asymp \exp\left(\frac{1}{\varepsilon\lambda}\mathscr{V}(x^*, \partial D)\right), \tag{4.3}$$

where $\asymp$ log-asymptotic equality. Thus, the log-expected escape times are decreasing proportionally with λ. On the other hand, since the HMM action is a multiple of the true action, the minimizers will be unchanged by the HMM approximation. Hence, the large deviation transition pathways will be unchanged by the HMM approximation.

To justify the claim for $u_{\lambda,\Delta t}$ (4.2), we first introduce some notation. Let $S_t^{(\alpha)}$ be the semigroup associated with the Hamilton–Jacobi equation

$$\partial_t v(t, x) = \mathcal{H}(\alpha, \nabla v(t, x)), \tag{4.4}$$

notice that this is the same as the true Hamilton–Jacobi equation (2.10) but with the first argument of the Hamiltonian now frozen as a parameter α. The necessity of the parameter α is due to the fact that in the system for $(X_n^\varepsilon, Y_n^\varepsilon)$, the x variable in the fast process is frozen to its value at the left endpoint of the interval and hence is treated as a parameter on each interval. We also introduce the operator $S_t\psi(x) = S_t^{(\alpha)}\psi(x)|_{\alpha=x}$ and also $S_{\lambda,t} = \lambda^{-1}S_t(\lambda\cdot)$. In this notation, it is simple to show that

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_n^\varepsilon)\right) \asymp \exp\left(\varepsilon^{-1}(S_{\lambda,\Delta t})^n\varphi(x)\right). \tag{4.5}$$

We will verify (4.5) by induction, starting with the $n = 1$ case. Since, on the interval $I_{0,\Delta t}$, the pair $(X_0^\varepsilon, \widetilde{Y}_0^\varepsilon)$ is a fast–slow system of the form (2.1) with ε replaced by $\varepsilon\lambda$, it follows from Sect. 2.3 that X_0^ε satisfies an LDP with action functional derived from the Hamiltonian–Jacobi equation (4.4), with the parameter α set to the value of X_0^ε at the left endpoint, which is $X_0^\varepsilon(0) = x$. Hence, it follows from Varadhan's lemma that for any suitable $\psi : \mathbb{R}^d \to \mathbb{R}$

$$\mathbf{E}_x \exp\left((\varepsilon\lambda)^{-1}\psi(X_0^\varepsilon(\Delta t))\right) \asymp \exp\left((\varepsilon\lambda)^{-1}S_{\Delta t}^{(\alpha)}\psi(x)|_{\alpha=x}\right).$$

Hence, since $x_1^\varepsilon = X_0^\varepsilon(\Delta t)$ with $X_0^\varepsilon(0) = x$, we have

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_1^\varepsilon)\right) = \mathbf{E}_x \exp\left((\varepsilon\lambda)^{-1}\lambda\varphi(X_1^\varepsilon(\Delta t))\right)$$
$$\asymp \exp\left((\varepsilon\lambda)^{-1}S_{\Delta t}^{(\alpha)}(\lambda\varphi)(x)|_{\alpha=x}\right) = \exp\left(\varepsilon^{-1}S_{\lambda,\Delta t}\varphi(x)\right)$$

as claimed. Now, suppose (4.5) holds for all k with $n \geq k \geq 1$, then

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_{n+1}^\varepsilon)\right) = \mathbf{E}_x\mathbf{E}_{x_1^\varepsilon} \exp\left(\varepsilon^{-1}\varphi(x_{n+1}^\varepsilon)\right). \tag{4.6}$$

By the inductive hypothesis, we have that

$$\mathbf{E}_{x_1^\varepsilon} \exp\left(\varepsilon^{-1}\varphi(x_{n+1}^\varepsilon)\right) \asymp \exp\left(\varepsilon^{-1}(S_{\lambda,\Delta t})^n\varphi(x_1^\varepsilon)\right). \tag{4.7}$$

Applying (4.7) under the expectation in (4.6) (see Remark 4.1), we see that

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_{n+1}^\varepsilon)\right) = \mathbf{E}_x\mathbf{E}_{x_1^\varepsilon} \exp\left(\varepsilon^{-1}\varphi(x_{n+1}^\varepsilon)\right) \asymp \mathbf{E}_x \exp\left(\varepsilon^{-1}(S_{\lambda,\Delta t})^n\varphi(x_1^\varepsilon)\right).$$

Now applying the inductive hypothesis with $n = 1$ and $\psi(\cdot) = (S_{\lambda,\Delta t})^n \varphi(\cdot)$

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}(S_{\lambda,\Delta t})^n \varphi(x_1^\varepsilon)\right) \asymp \exp\left(\varepsilon^{-1} S_{\lambda,\Delta t}(S_{\lambda,\Delta t})^n \varphi(x)\right),$$

which completes the induction.

By definition, we therefore have $u_{\lambda,\Delta t}(t, x) = (S_{\lambda,\Delta t})^n \varphi(x)$ when $t \in I_{n,\Delta t}$. All that remains is to argue that $u_{\lambda,\Delta t}$ converges to the solution of (4.2) as $\Delta t \to 0$. But this can be seen from the expansion of the semigroup

$$
\begin{aligned}
\frac{u_{\lambda,\Delta t}(t + \Delta t, x) - u_{\lambda,\Delta t}(t, x)}{\Delta t} &= \frac{(S_{\lambda,\Delta t})^{n+1}\varphi(x) - (S_{\lambda,\Delta t})^n \varphi(x)}{\Delta t} \\
&= \frac{S_{\lambda,\Delta t}(S_\Delta t)^n \varphi(x) - (S_{\lambda,\Delta t})^n \varphi(x)}{\Delta t} \\
&= \lambda^{-1}\mathcal{H}(\alpha, \lambda\nabla(S_{\lambda,\Delta t})^n \varphi(x))|_{\alpha=x} + O(\Delta t) \\
&= \lambda^{-1}\mathcal{H}(x, \lambda\nabla u_{\lambda,\Delta t}(t, x))) + O(\Delta t),
\end{aligned}
\tag{4.8}
$$

which yields the desired limiting equation.

Remark 4.1 Regarding the operation of taking the log-asymptotic result inside the expectation, one can find such calculations done rigorously in (for instance) [15, Lemma 4.3].

Remark 4.2 From the discussion above, it appears that the mean transition time can be estimated from HMM upon exponential rescaling; see (4.3). This is true, but only at the level of the (rough) log-asymptotic estimate of this time. How to rescale the prefactor is by no means obvious. As we will see below, PHMM avoids this issue altogether since it does not necessitate any rescaling.

5 Parallelized HMM

There is a simple variant of the above HMM-type scheme which captures the correct average behavior and fluctuations, both at the level of the CLT and LDP. In a usual HMM-type method, the key approximation is given by

$$\int_{n\Delta t}^{(n+1)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}s \approx \lambda \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}s, \tag{5.1}$$

which only requires computation of the fast variables on the interval $[n\Delta t, (n + 1/\lambda)\Delta t]$. This approximation is effective at replicating averages, but poor at replicating fluctuations. Indeed, for each j, the time series Y_n^ε on the interval $[(n + j/\lambda)\Delta t, (n + (j + 1)/\lambda)\Delta t]$ is replaced with an identical copy of the time series from the interval $[n\Delta t, (n + 1/\lambda)\Delta t]$. This introduces strong correlations between random variables that should be essentially independent. Parallelized HMM avoids this issue by employing the approximation

$$\int_{n\Delta t}^{(n+1)\Delta t} f(x_n^\varepsilon, Y_n^\varepsilon(s))\mathrm{d}s \approx \sum_{j=1}^{\lambda} \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x_n^\varepsilon, Y_n^{\varepsilon,j}(s))\mathrm{d}s,$$

where $Y_n^{\varepsilon,j}$ are for each j independent copies of the time series computed in (5.1). Due to their independence, each copy of the fast variables can be computed in parallel; hence, we refer to the method as parallel HMM (PHMM). The method is summarized below.

1. (Micro-step) On the interval $I_{n,\Delta t}$, simulate λ independent copies of the fast variables, each copy simulated precisely as in the usual HMM. That is, let

$$Y_n^{\varepsilon,j} = Y_n^{\varepsilon,j}(n\Delta t) + \frac{1}{\varepsilon}\int_{n\Delta t}^t g(x_n^\varepsilon, Y_n^{\varepsilon,j}(s))\mathrm{d}s + \frac{1}{\sqrt{\varepsilon}}\int_{n\Delta t}^t \sigma(x_n^\varepsilon, Y_n^{\varepsilon,j}(s))\mathrm{d}W_j(s) \tag{5.2}$$

for $j = 1, \ldots, \lambda$ with W_j independent Brownian motions. As with ordinary HMM, we will not require the time series of the whole interval $I_{n,\Delta t}$ but only over the subset $[n\Delta t, (n + 1/\lambda)\Delta t)$.

2. (Macro-step) Use the time series from the micro-step to update x_n^ε to x_{n+1}^ε by

$$x_{n+1}^\varepsilon = x_n^\varepsilon + \sum_{j=1}^{\lambda} \int_{n\Delta t}^{(n+1/\lambda)\Delta t} f(x_n^\varepsilon, Y_n^{\varepsilon,j}(s))\mathrm{d}s. \tag{5.3}$$

As with the HMM-type method, it will be convenient to write PHMM as a fast–slow system (when restricted to an interval $I_{n,\Delta t}$). Akin to (3.5), it is easy to verify that the parallel HMM scheme is described by the system

$$\frac{\mathrm{d}X_n^\varepsilon}{\mathrm{d}t} = \frac{1}{\lambda} \sum_{j=1}^{\lambda} f(x_n^\varepsilon, \widetilde{Y}_n^{\varepsilon,j}),$$

$$\mathrm{d}\widetilde{Y}_{n,j}^\varepsilon = \frac{1}{\varepsilon\lambda} g(x_n^\varepsilon, \widetilde{Y}_n^{\varepsilon,j})\mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}} \sigma(x_n^\varepsilon, \widetilde{Y}_n^{\varepsilon,j})\mathrm{d}W_j, \tag{5.4}$$

for $j = 1, \ldots, \lambda$ with the initial condition $X_n^\varepsilon(n\Delta t) = x_n^\varepsilon$.

6 Average and fluctuations in parallelized HMM

In this section, we check that the averaged behavior and the fluctuations in the PHMM method are consistent with those in the original fast slow system. Just as noted at the beginning of Sect. 4.1, the LLN, CLT and LDP results derived for (5.3) can be extended to the non-discretized system (3.5). In particular, the CLT and LDP of the PHMM approximation (5.4) are the same as the original fast slow system (2.1).

6.1 Averaging

Proceeding exactly as in Sect. 4.1, it follows that as $\varepsilon \to 0$ the PHMM scheme x_{n+1}^ε converges to $\bar{x}_{n+1} = \bar{X}_n((n + 1)\Delta t)$ where

$$\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} = \frac{1}{\lambda} \sum_{j=1}^{\lambda} F(\bar{x}_n) = F(\bar{x}_n) \tag{6.1}$$

with initial condition $\bar{X}_n(n\Delta t) = \bar{x}_n$. Hence, we are in the exact same situation as with ordinary HMM, so the averaged behavior is consistent with that of the original fast slow system.

6.2 Small fluctuations

We now show that the fluctuations

$$z_n^\varepsilon = \frac{x_n^\varepsilon - \bar{x}_n}{\sqrt{\varepsilon}}$$

are consistent with the correct CLT fluctuations, described by (2.4). As in Sect. 4.2, we instead look at the rescaled fluctuations

$$\hat{z}_n^\varepsilon = \frac{x_n^\varepsilon - \bar{x}_n}{\sqrt{\varepsilon\lambda}}.$$

In particular, we will show that these rescaled fluctuations are consistent with

$$\mathrm{d}\widehat{Z} = B_0(\bar{X})\widehat{Z}\mathrm{d}t + \lambda^{-1/2}\eta(\bar{X})\mathrm{d}V. \tag{6.2}$$

The claim for z^ε will follow immediately from the claim for $\hat{z}^\varepsilon$.

We have that $\hat{z}^\varepsilon_{n+1} = \widehat{Z}^\varepsilon_n((n+1)\Delta t)$ where

$$\widehat{Z}^\varepsilon_n(t) = \frac{X^\varepsilon_n(t) - \bar{X}_n(t)}{\sqrt{\varepsilon\lambda}},$$

with X^ε_n given by the system (5.4) and $\bar{X}_n$ given by the averaged equation (6.1). As in Sect. 4.2, we derive a system for the triple $(\bar{X}_n, \widehat{Z}^\varepsilon_n, \widetilde{Y}^\varepsilon_n)$, where now the fast process has λ independent components $\widetilde{Y}^\varepsilon_n = (\widetilde{Y}^{\varepsilon,1}_n, \ldots, \widetilde{Y}^{\varepsilon,\lambda}_n)$:

$$
\begin{aligned}
\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} &= F(\bar{x}_n), \\
\frac{\mathrm{d}\widehat{Z}^\varepsilon_n}{\mathrm{d}t} &= \frac{1}{\sqrt{\varepsilon\lambda}}\frac{1}{\lambda}\sum_{j=1}^\lambda \tilde{f}(\bar{x}_n, \widetilde{Y}^{\varepsilon,j}_n) + \frac{1}{\lambda}\sum_{j=1}^\lambda \nabla_x f(\bar{x}_n, \widetilde{Y}^{\varepsilon,j}_n)\hat{z}_n + O(\sqrt{\varepsilon\lambda}), \\
\mathrm{d}\widetilde{Y}^{\varepsilon,j}_n &= \frac{1}{\varepsilon\lambda}g(\bar{x}_n, \widetilde{Y}^{\varepsilon,j}_n)\mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}}\nabla_x g(\bar{x}_n, \widetilde{Y}^{\varepsilon,j}_n)\hat{z}_n\mathrm{d}t + \frac{1}{\sqrt{\varepsilon\lambda}}\sigma(\bar{x}_n, \widetilde{Y}^{\varepsilon,j}_n)\mathrm{d}W_j + O(1).
\end{aligned}
$$
(6.3)

With a modicum added difficulty, we can now argue as in Sect. 2.2 with $\varepsilon' = \varepsilon\lambda$ playing the role of ε. The invariant measure $\mu^\lambda_x(\mathrm{d}y)$ associated with the generator of Y^ε_n is now the product measure

$$\mu^\lambda_x(\mathrm{d}y_1, \ldots, \mathrm{d}y_\lambda) = \mu_x(\mathrm{d}y_1)\ldots\mu_x(\mathrm{d}y_\lambda),$$

where μ_x is the invariant measure associated with $\mathscr{L}_0$ from Sect. 2.2. This product structure simplifies the seemingly complicated expressions arising in the perturbation expansion of (6.3). In the setting of Sect. 2.2, we have that $u_0 = u_0(x, z, t)$ and

$$u_1(x, z, y, t) = (-\mathscr{L}^{(1)}_0 - \cdots - \mathscr{L}^{(\lambda)}_0)^{-1}\mathscr{L}_1 u_0(x, z, y, t),$$
(6.4)

where $\mathscr{L}^{(j)}_0 = g(\bar{x}_n, y_j)\nabla_{y_j} + \frac{1}{2}\sigma\sigma^T(\bar{x}_n, y_j) : \nabla^2_{y_j}$

Since

$$\mathscr{L}_1 u_0(x, z, y, t) = \frac{1}{\lambda}\sum_{j=1}^\lambda \tilde{f}(\bar{x}_n, y_j) \cdot \nabla_z u_0(x, z, t),$$

the Feynman–Kac representation of (6.4) yields

$$u_1(x, z, y, t) = \frac{1}{\lambda}\sum_{j=1}^\lambda \int_0^\infty \mathrm{d}\tau\, \mathbf{E}_{y_j}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n, j}(\tau)) \cdot \nabla_z u_0(x, z, t).$$

The equation for u_0 is now given by

$$
\begin{aligned}
\partial_t u_0 = {}& F(\bar{x}_n)\nabla_x u_0 + \int \mu_{\bar{x}_n}(\mathrm{d}y_1)\ldots\mu_{\bar{x}_n}(\mathrm{d}y_\lambda)\left(\frac{1}{\lambda}\sum_{j=1}^\lambda \nabla_x f(\bar{x}_n, y_j)\hat{z}_n\right)\nabla_z u_0 \\
&+ \int \mu_{\bar{x}_n}(\mathrm{d}y_1)\ldots\mu_{\bar{x}_n}(\mathrm{d}y_\lambda) \\
&\times \left(\int_0^\infty \mathrm{d}\tau\left(\frac{1}{\lambda}\sum_{j=1}^\lambda \tilde{f}(\bar{x}_n, y_j)\right) \otimes \left(\frac{1}{\lambda}\sum_{k=1}^\lambda \mathbf{E}_y\tilde{f}(\bar{x}_n, Y^k_{\bar{x}_n}(\tau))\right) : \nabla^2_z u_0\right. \\
&\left.+ \sum_{j=1}^\lambda(\nabla_x g(\bar{x}_n, y_j)\hat{z}_n)\int_0^\infty \mathrm{d}\tau\, \nabla_{y_j}\frac{1}{\lambda}\sum_{k=1}^\lambda \mathbf{E}_{y_k}\tilde{f}(\bar{x}_n, Y^k_{\bar{x}_n}(\tau))\nabla_z u_0\right).
\end{aligned}
$$
(6.5)

By expanding the product measure, the second term on the right-hand side of (6.5) becomes

$$\frac{1}{\lambda} \sum_{j=1}^{\lambda} \int \mu_{\bar{x}_n}(\mathrm{d}y_j)(\nabla_x f(\bar{x}_n, y_j)\hat{z}_n) \cdot \nabla_z u_0$$

$$= \int \mu_{\bar{x}_n}(\mathrm{d}y_1)(\nabla_x f(\bar{x}_n, y_1)\hat{z}_n) \cdot \nabla_z u_0 = (B_1(\bar{x}_n)\hat{z}_n) \cdot \nabla_z u_0.$$

Likewise, using the independence of Y_x^j for distinct j, the third term becomes

$$\frac{1}{\lambda^2} \sum_{j,k=1}^{\lambda} \int_0^\infty \mathrm{d}\tau \, \mathbf{E}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^j(0)) \otimes \tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^k(\tau)) : \nabla_z^2 u_0$$

$$= \frac{1}{\lambda^2} \sum_{j=1}^{\lambda} \int_0^\infty \mathrm{d}\tau \, \mathbf{E}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^j(0)) \otimes \tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^j(\tau)) : \nabla_z^2 u_0$$

$$= \frac{1}{\lambda} \int_0^\infty \mathrm{d}\tau \, \mathbf{E}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^1(0)) \otimes \tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^1(\tau)) : \nabla_z^2 u_0 = \frac{1}{\lambda} \eta(\bar{x}_n)\eta(\bar{x}_n)^T : \nabla_z^2 u_0,$$

where the expectation is taken over realizations of Y_x^j with $Y_x^j(0) \sim \mu_x$. Finally, since the $\nabla_{y_j} \mathbf{E}_{y_k}$ term vanishes on the off-diagonal, the last term in (6.5) reduces to

$$\frac{1}{\lambda} \sum_{j,k=1}^{\lambda} \int_0^\infty \mathrm{d}\tau \int \mu_{\bar{x}_n}(\mathrm{d}y_j)\mu_{\bar{x}_n}(\mathrm{d}y_k)(\nabla_x g(\bar{x}_n, y_j)\hat{z}_n) \cdot \nabla_{y_j} \mathbf{E}_{y_k}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^k(\tau)) \cdot \nabla_z u_0$$

$$= \frac{1}{\lambda} \sum_{j=1}^{\lambda} \int_0^\infty \mathrm{d}\tau \int \mu_{\bar{x}_n}(\mathrm{d}y_j)\mu_{\bar{x}_n}(\mathrm{d}y_k)(\nabla_x g(\bar{x}_n, y_j)\hat{z}_n)\nabla_{y_j} \mathbf{E}_{y_j}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^j(\tau))\nabla_z u_0$$

$$= \int_0^\infty \mathrm{d}\tau \int \mu_{\bar{x}_n}(\mathrm{d}y_1)\mu_{\bar{x}_n}(\mathrm{d}y_k)(\nabla_x g(x, y_1)\hat{z}_n)\nabla_{y_1} \mathbf{E}_{y_1}\tilde{f}(\bar{x}_n, Y_{\bar{x}_n}^1(\tau))\nabla_z u_0$$

$$= (B_2(\bar{x}_n)\hat{z}_n) \cdot \nabla_z u_0.$$

It follows immediately that the reduced equation for the pair $(\bar{X}_n, \hat{Z}_n^\varepsilon)$ is

$$\frac{\mathrm{d}\bar{X}_n}{\mathrm{d}t} = F(\bar{x}_n)$$
$$\mathrm{d}\widehat{Z}_n = B_0(\bar{x}_n)\widehat{Z}_n\mathrm{d}t + \lambda^{-1/2}\eta(\bar{x}_n)\mathrm{d}V,$$

with initial conditions $\widehat{Z}_n(n\Delta t) = \hat{z}_n$ and $\bar{X}_n(n\Delta t) = \bar{x}_n$. Hence, we see that $\hat{z}_{n+1}$ is described by

$$\hat{z}_{n+1} = \hat{z}_n + B(\bar{x}_n)\hat{z}_n\Delta t + \lambda^{-1/2}\eta(\bar{x}_n)\sqrt{\Delta t}\,\xi_n,$$

which is the Euler–Maruyama scheme for (6.2).

6.3 Large fluctuations

In this section, we show that the LDP for PHMM is consistent with the true LDP from Sect. 2.3. In particular, let

$$u_{\lambda, \Delta t}(t, x) = \lim_{\varepsilon \to 0} \varepsilon \log \mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_n^\varepsilon)\right)$$

for $t \in I_{n, \Delta t}$, where x_n^ε is the PHMM approximation. We will argue that $u_{\lambda, \Delta t}(t, x) \to u(t, x)$ as $\Delta t \to 0$, where u solves the correct Hamilton–Jacobi equation (2.10).

The argument is a slight modification of that given in Sect. 4.3. Before proceeding, we recall the notation $S_{\Delta t}^{(\alpha)}$ for the semigroup associated with the Hamilton–Jacobi equation

$$\partial_t u(t, x) = \mathcal{H}(\alpha, \nabla u(t, x)), \tag{6.6}$$

where $\mathcal{H}$ is the Hamiltonian defined by (2.7). We also define the operator $S_{\Delta t}\varphi(x) = S_{\Delta t}^{(\alpha)}\varphi(x)|_{\alpha=x}$.

As in Sect. 4.3, the claim follows from the asymptotic statement

$$\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_n^\varepsilon)\right) \asymp \exp\left(\varepsilon^{-1}(S_{\Delta t})^n\varphi(x)\right), \quad \varepsilon \to 0. \tag{6.7}$$

Given (6.7), by an identical argument to that started in Eq. (4.8), it follows from (6.7) that $u_{\lambda, \Delta t}$ is indeed a numerical approximation of the solution to (6.6) and hence $u_{\lambda, \Delta t} \to u$ as $\Delta t \to 0$.

We will verify (6.7) by induction, starting with the $n = 1$ case. Since $(X^\varepsilon, \widetilde{Y}_{0,1}^\varepsilon, \ldots, \widetilde{Y}_{0,\lambda}^\varepsilon)$ is a fast–slow system of the form (2.1) with ε replaced by $\varepsilon\lambda$, it follows from Sect. 2.3 (Varadhan's lemma) that

$$\mathbf{E}_x \exp\left((\varepsilon\lambda)^{-1}\psi(X_1^\varepsilon(\Delta t))\right) \asymp \exp\left((\varepsilon\lambda)^{-1}\widehat{S}_{\Delta t}^{(\alpha)}\psi(x)|_{\alpha=x}\right),$$

where $\widehat{S}_{\Delta t}^{(\alpha)}$ is the semigroup associated with $\partial_t v(t, x) = \widehat{\mathcal{H}}(\alpha, \nabla v(t, x))$ and

$$\widehat{\mathcal{H}}(\alpha, \theta) = \lim_{T \to \infty} T^{-1} \log \mathbf{E} \exp\left(\theta \cdot \int_0^T d\tau \frac{1}{\lambda} \sum_{j=1}^\lambda f(\alpha, Y_\alpha^j(\tau))\right).$$

Hence, we have

$$\begin{aligned}
\mathbf{E}_x \exp\left(\varepsilon^{-1}\varphi(x_1^\varepsilon)\right) &= \mathbf{E}_x \exp\left((\varepsilon\lambda)^{-1}\lambda\varphi(X_0^\varepsilon(\Delta t))\right) \\
&\asymp \mathbf{E}_x \exp\left((\varepsilon\lambda)^{-1}\widehat{S}_{\Delta t}^{(\alpha)}(\lambda\varphi)(x)|_{\alpha=x}\right).
\end{aligned} \tag{6.8}$$

But since Y_α^j are i.i.d. for distinct j, the Hamiltonian $\widehat{\mathcal{H}}$ reduces to

$$\begin{aligned}
\lim_{T \to \infty} &T^{-1} \log \mathbf{E} \exp\left(\theta \cdot \int_0^T d\tau \frac{1}{\lambda} \sum_{j=1}^\lambda f(\alpha, Y_\alpha^j(\tau))\right) \\
&= \lambda \lim_{T \to \infty} T^{-1} \log \mathbf{E} \exp\left(\frac{\theta}{\lambda} \cdot \int_0^T d\tau f(\alpha, Y_\alpha^1(\tau))\right) = \lambda \mathcal{H}\left(\alpha, \frac{\theta}{\lambda}\right).
\end{aligned}$$

It follows that

$$\partial_t \left(\lambda^{-1}\widehat{S}_t^{(\alpha)}(\lambda\varphi)\right) = \lambda^{-1}\widehat{\mathcal{H}}(\alpha, \nabla(\widehat{S}_t^{(\alpha)}(\lambda\varphi))) = \mathcal{H}(\alpha, \lambda^{-1}\nabla(\widehat{S}_t^{(\alpha)}(\lambda\varphi)))$$

and hence $\lambda^{-1}\widehat{S}_{\Delta t}^{(\alpha)}(\lambda\varphi) = S_{\Delta t}^{(\alpha)}\varphi$. Combining this with (6.8) completes the claim for $n = 1$. The proof of the inductive step for arbitrary $n \geq 1$ follows identically to Sect. 4.3.

7 Numerical evidence

In this section, we investigate the performance of the standard HMM and PHMM methods for systems with well-understood fluctuations and metastability properties. These simple experiments confirm that HMM amplifies fluctuations, which can drastically change the system's metastable behavior, and that the PHMM succeeds in avoiding these problems. In Sect. 7.1, we investigate simple CLT fluctuations for a simple quadratic potential systems; in Sect. 7.2, we look at large deviation fluctuations for a quartic double-well potential.

Finally in Sect. 7.3, we look at fluctuations for a non-diffusive double-well potential, which has large deviation properties that cannot be captured by a so-called 'small noise' diffusion.

In all of the experiments below, we use the numerical approximation (3.3), (3.4) with macro-step Δt and micro-step δt as specified for each experiment. The number of micro-steps is accordingly $M = \lfloor \Delta t/(\lambda \delta t) \rfloor$. At the start of each micro-integration, the fast variables are initialized using their final value at the previous micro-step. As stated in the introduction, with the specific choice of Δt and δt for which $M = 1$, this initialization corresponds to performing an Euler–Maruyama approximation of the inflated system of the type (1.7). This choice is used for the experiment in Sect. 7.3.

We also note that the Euler–Maruyama scheme was chosen due to the relative simplicity of the underlying fast–slow system. In general, to ensure that numerical CLT and LDP results are faithful to the original fast–slow system, it may be advisable to use more sophisticated integrators for the fast variables, such as Störmer–Verlet-type methods [5].

7.1 Small fluctuations

We examine the small CLT-type fluctuations by looking at the following fast–slow system

$$
\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t} &= Y - X, \\
\mathrm{d}Y &= \frac{\theta}{\varepsilon}(\mu X - Y)\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\mathrm{d}W.
\end{aligned}
\tag{7.1}
$$

It is simple to check that the averaged system is given by

$$
\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = (\mu - 1)\bar{X}.
$$

Hence, for $\mu < 1$ the averaged system is a gradient flow in a quadratic potential centered at the origin.
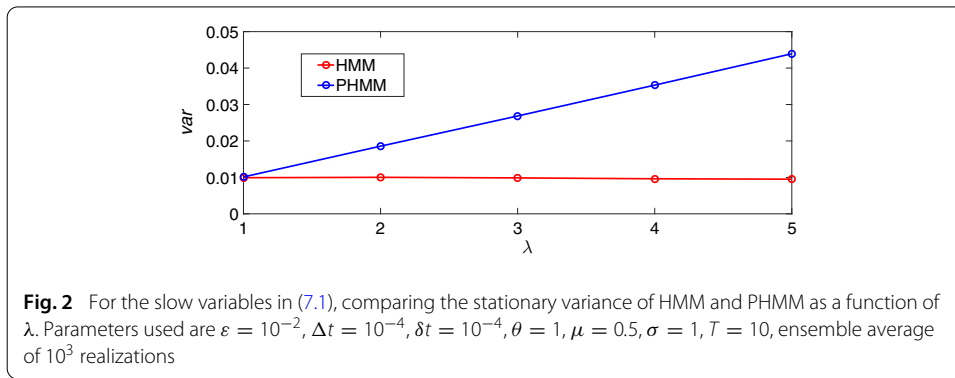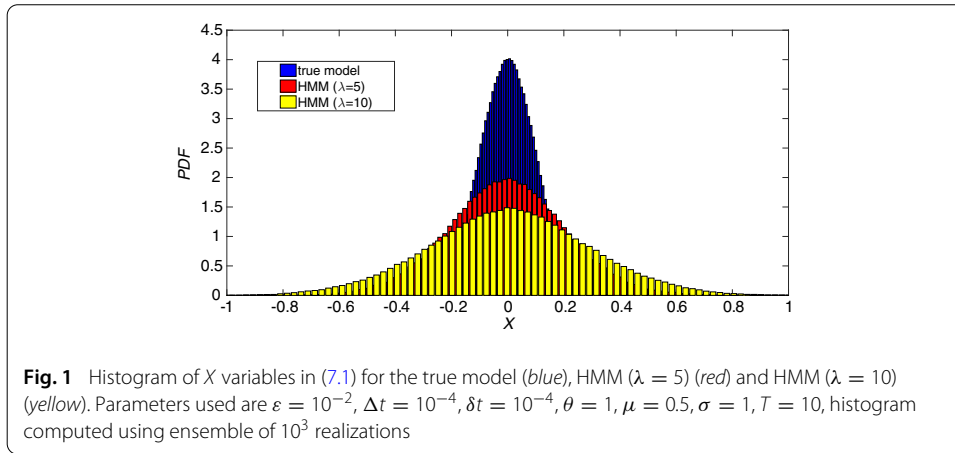
We will first illustrate that the HMM-type method described in Sect. 3 inflates the $O(\sqrt{\varepsilon})$ fluctuations about the average by a factor of $\sqrt{\lambda}$. In Fig. 1, we plot histograms of the slow variable X for different speedup factors λ. It is clear that the spread of the invariant distribution is increasing with λ. The profile remains Gaussian, but the variance is greatly inflated. In Fig. 2, we plot the variance of the stationary time series for X as a function of λ. The blue line is computed using HMM, and the red line is computed using PHMM. As predicted by the theory in Sect. 4.2, in the case of HMM the variance is increasing linearly with λ and in the case of PHMM the variance is approximately constant. Note that in this example, the CLT captures the large deviations as well. This is because, to leading order in ε, the fluctuations above the limiting behavior can be captured at all times $t > 0$ by the SDE

$$
dX = (\mu - 1)X\mathrm{d}t + \frac{\sigma\sqrt{\varepsilon}}{\theta}\mathrm{d}W
$$

PHMM is consistent with this SDE, whereas HMM is not (it is consistent with an SDE where the noise is inflated by $\sqrt{\lambda}$. The variance of the solution of the SDE above at time T is

$$
\sigma^2\varepsilon 2\theta^2(\mu - 1)(1 - e^{-2(\mu-1)T}) \to \sigma^2\varepsilon 2\theta^2(\mu - 1) \quad \text{as } T \to \infty,
$$

which is already very close to its asymptotic value $\sigma^2\varepsilon/(2\theta^2(\mu - 1))$ for the parameter value reported in Fig. 2. As can also be seen in this figure, the variance of HMM is λ times the one above.

Fig. 1 Histogram of *X* variables in (7.1) for the true model (*blue*), HMM ($\lambda = 5$) (*red*) and HMM ($\lambda = 10$) (*yellow*). Parameters used are $\varepsilon = 10^{-2}, \Delta t = 10^{-4}, \delta t = 10^{-4}, \theta = 1, \mu = 0.5, \sigma = 1, T = 10$, histogram computed using ensemble of 10^3 realizations

Fig. 2 For the slow variables in (7.1), comparing the stationary variance of HMM and PHMM as a function of λ. Parameters used are $\varepsilon = 10^{-2}, \Delta t = 10^{-4}, \delta t = 10^{-4}, \theta = 1, \mu = 0.5, \sigma = 1, T = 10$, ensemble average of 10^3 realizations

7.2 Large fluctuations

To investigate the effect of parallelization on $O(1)$ deviations not captured by the CLT, we will look at a fast–slow system which exhibits metastability. Hence, it is natural to take

$$
\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t} &= Y - X^3, \\
\mathrm{d}Y &= \frac{\theta}{\varepsilon}(\mu X - Y)\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\mathrm{d}W.
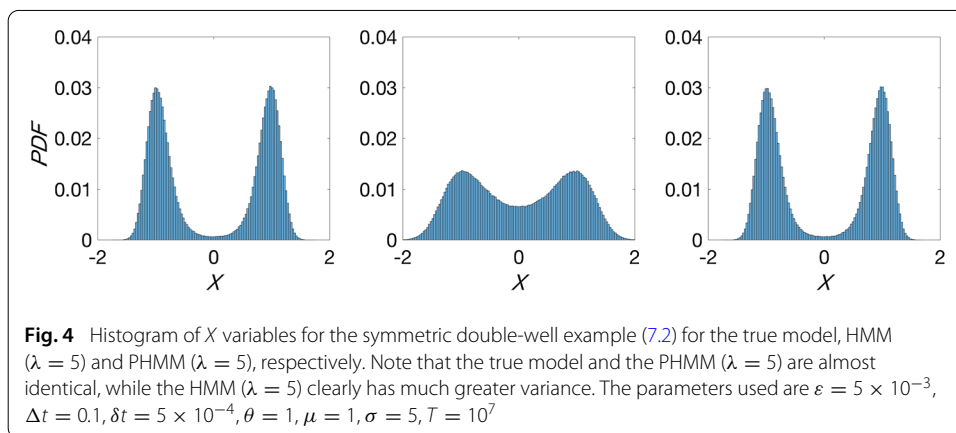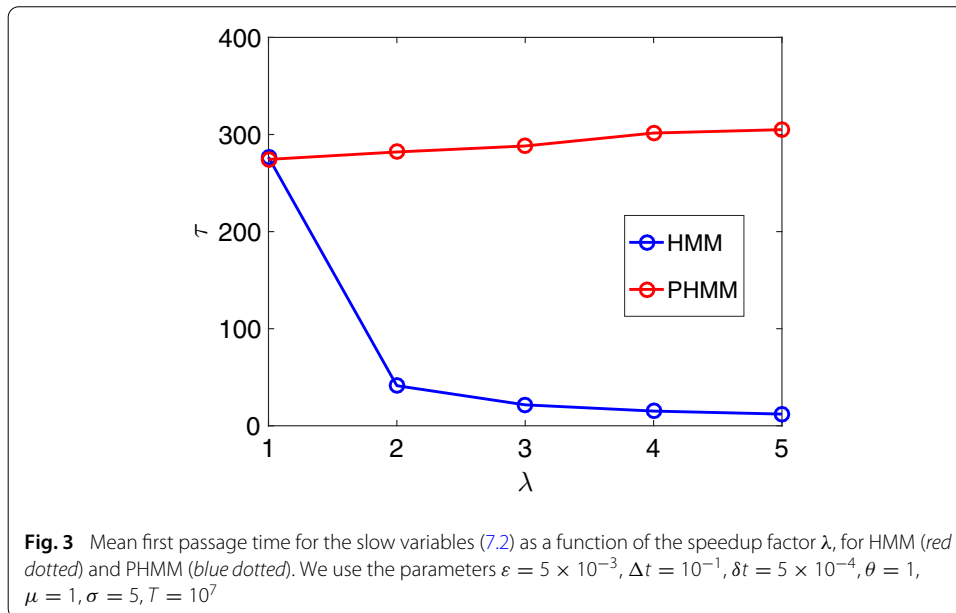\end{aligned}
\tag{7.2}
$$

It is simple to check that the averaged system is

$$
\frac{\mathrm{d}\bar{X}}{\mathrm{d}t} = \mu\bar{X} - \bar{X}^3.
$$

Hence, for any $\mu > 0$ the averaged system is a gradient flow in a symmetric double-well potential, with stable equilibria at $\pm\sqrt{\mu}$ and a saddle point at the origin. The large fluctuations of the fast–slow system can be investigated by looking at the first passage time for transitions from a neighborhood of one stable equilibrium to the other.

In Fig. 3, we compare the mean first passage time for HMM and PHMM as a function of λ. Even for $\lambda = 2$, the distinction between the two methods is vast, with the mean first passage time for HMM rapidly dropping off and for PHMM staying approximately constant.

In Fig. 4, we compare, respectively, the stationary distributions of the true fast–slow system, HMM ($\lambda = 5$) and PHMM ($\lambda = 5$). In the case of HMM, the energy barrier separating the two metastable states is now overpopulated, which explains the rapid fall

Fig. 3 Mean first passage time for the slow variables (7.2) as a function of the speedup factor λ, for HMM (*red dotted*) and PHMM (*blue dotted*). We use the parameters $\varepsilon = 5 \times 10^{-3}$, $\Delta t = 10^{-1}$, $\delta t = 5 \times 10^{-4}$, $\theta = 1$, $\mu = 1$, $\sigma = 5$, $T = 10^7$

Fig. 4 Histogram of X variables for the symmetric double-well example (7.2) for the true model, HMM ($\lambda = 5$) and PHMM ($\lambda = 5$), respectively. Note that the true model and the PHMM ($\lambda = 5$) are almost identical, while the HMM ($\lambda = 5$) clearly has much greater variance. The parameters used are $\varepsilon = 5 \times 10^{-3}$, $\Delta t = 0.1$, $\delta t = 5 \times 10^{-4}$, $\theta = 1$, $\mu = 1$, $\sigma = 5$, $T = 10^7$
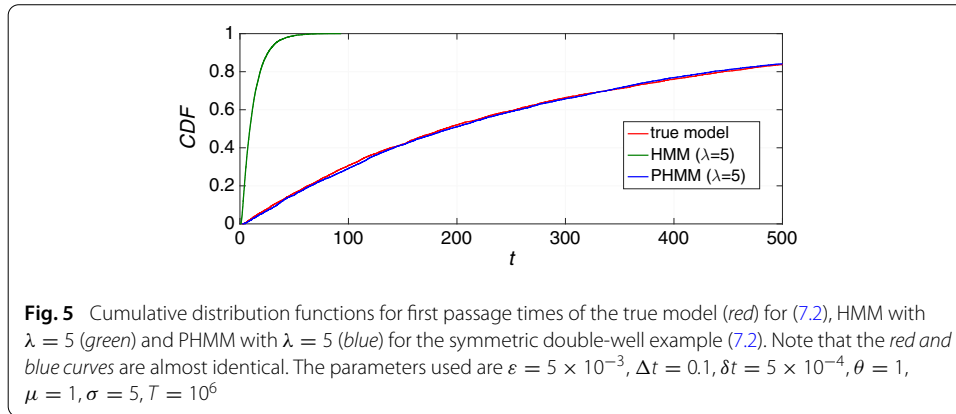
in mean first passage time. In the case of PHMM, the histogram is indistinguishable from the true stationary distribution.

In Fig. 5, we plot the cumulative distributions function (CDF) for the first passage time, comparing that of the true fast–slow system, with HMM ($\lambda = 5$) and PHMM ($\lambda = 5$). We see that the HMM first passage times are supported on a much faster timescale than that of the true fast–slow system. In contrast, the CDF of PHMM is practically indistinguishable from that of the true fast–slow system. Hence, PHMM is not just replicating the mean first passage time, but also the entire distribution of first passage times.

7.3 Asymmetric, non-diffusive fluctuations

We now compare HMM and PHMM for a multiscale model that also displays metastability, but in which the large fluctuations cannot be characterized by a 'small noise' Ito diffusion. In particular, the Hamiltonian describing the LDP of the system is non-quadratic, as opposed the previous systems. The system has been used [6] to illustrate the ineffec-

Fig. 5 Cumulative distribution functions for first passage times of the true model (*red*) for (7.2), HMM with $\lambda = 5$ (*green*) and PHMM with $\lambda = 5$ (*blue*) for the symmetric double-well example (7.2). Note that the *red and blue curves* are almost identical. The parameters used are $\varepsilon = 5 \times 10^{-3}, \Delta t = 0.1, \delta t = 5 \times 10^{-4}, \theta = 1, \mu = 1, \sigma = 5, T = 10^{6}$

tiveness of diffusion-type approximations for fast–slow systems. The fast–slow system is given by

$$\frac{dX}{dt} = Y^2 - \nu X,$$
$$dY = -\frac{1}{\varepsilon}\gamma(X)Y\,dt + \frac{\sigma}{\sqrt{\varepsilon}}dW, \tag{7.3}$$

where $\gamma(x) = x^4/10 - x^2 + 3$. The averaged equation for this system reads

$$\frac{d\bar{X}}{dt} = \frac{\sigma^2}{2\gamma(\bar{X})} - \nu\bar{X}.$$

For $\nu = 1$ and $\sigma = \sqrt{3}$, this averaged equation possesses two stable fixed points at $x \approx 0.555$ and $x \approx= 2.459$ and one unstable fixed point at $x \approx 2.459$. The rates of transition between these stable fixed points are captured by the LDP. By an elementary calculation [6], the Hamiltonian of this LDP is found to be non-quadratic and given by
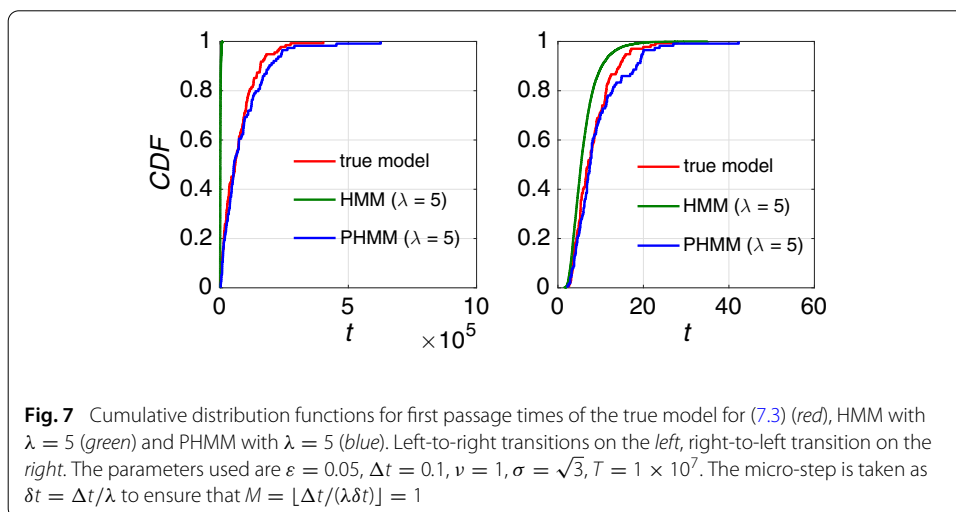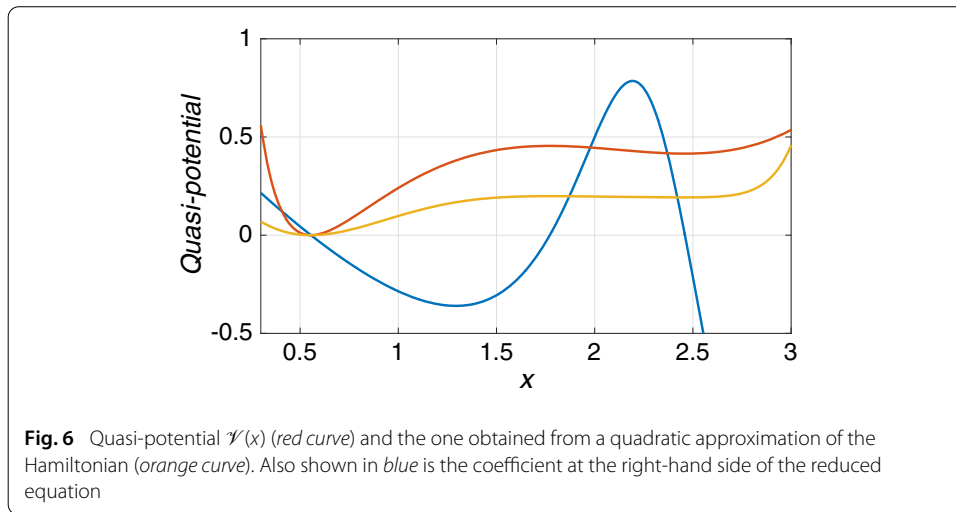
$$\mathcal{H}(x,\theta) = -\nu x\theta + \frac{1}{2}\left(\gamma(x) - \sqrt{\gamma^2(x) - 2\sigma^2\theta}\right).$$

The quasi-potential associated with this Hamiltonian satisfies $0 = \mathcal{H}(x, \mathcal{V}')$, i.e.,

$$\mathcal{V}'(x) = \frac{\nu x\gamma(x) - \frac{1}{2}\sigma^2}{\nu^2 x^2},$$

and is displayed in Fig. 6. While there is a significant barrier corresponding to left-to-right transitions, there is almost no barrier corresponding to right-to-left transitions.

In Fig. 7, we plot CDFs of the first passage times: Due to the asymmetry, we plot separately the transitions from the left-to-right and right-to-left. For left-to-right transitions, the HMM procedure drastically speeds up transitions because it enhances fluctuations: As is the case with the previous experiment, the HMM transitions are supported on a timescale several orders of magnitude faster than those of the true fast slow system. The PHMM method does not experience this problem and the distribution of first passage times agrees quite well with the true model. For right-to-left transitions, PHMM shows similarly good agreement with the true fast–slow system, but in contrast HMM is not too far off either. This can be accounted for by the 'flatness' of the right potential well, meaning that increasing the amplitude of fluctuations will only decrease the escape time by a linear multiplicative factor. We note that the noise appearing in the CDF plots is due to the scarcity of transitions occurring in the model (7.3).

Fig. 6 Quasi-potential $\mathscr{V}(x)$ (*red curve*) and the one obtained from a quadratic approximation of the Hamiltonian (*orange curve*). Also shown in *blue* is the coefficient at the right-hand side of the reduced equation

Fig. 7 Cumulative distribution functions for first passage times of the true model for (7.3) (*red*), HMM with $\lambda = 5$ (*green*) and PHMM with $\lambda = 5$ (*blue*). Left-to-right transitions on the *left*, right-to-left transition on the *right*. The parameters used are $\varepsilon = 0.05$, $\Delta t = 0.1$, $\nu = 1$, $\sigma = \sqrt{3}$, $T = 1 \times 10^7$. The micro-step is taken as $\delta t = \Delta t / \lambda$ to ensure that $M = \lfloor \Delta t / (\lambda \delta t) \rfloor = 1$

8 Discussion

We have investigated HMM methods for fast–slow systems, in particular their ability (or lack thereof) to capture fluctuations, both small (CLT) and large (LDP). We found, both theoretically (Sect. 4) and numerically (Sect. 7), that the amplitude of fluctuations is enhanced by an HMM-type method. In particular with an HMM speedup factor λ, in the CLT the variance of Gaussian fluctuations about the average is increased by a factor λ as well. In the LDP, the quasi-potential is decreased by a factor λ, leading to the first passage times being supported on a timescale λ orders of magnitude smaller than in the true fast slow system. This inability to correctly capture fluctuations about the average suggests that HMM can be a poor approximation of fast–slow systems, particularly when metastable behavior is important. As noted in Sect. 4.3, although the fluctuations of HMM are enhanced, the large deviation transition *pathways* remain faithful to the true model. Thus, we stress that, typically, HMM is a reliable method of finding transition pathways in metastable systems, but not for simulating their dynamics.

We have introduced a simple modification of HMM, called parallel HMM (PHMM), which avoids these fluctuation issues. In particular, the PHMM method yields fluctuations

that are consistent with the true fast slow system for any speedup factor λ (provided that we still have $\varepsilon\lambda \ll 1$), as was shown both theoretically (Sect. 6) and numerically (Sect. 7). The HMM method relies on computing one short burst of the fast variables, and inferring the statistical behavior of the fast variables by extrapolating this short burst over a large time window. PHMM on the other hand computes an ensemble of λ short bursts and infers the statistics of the fast variables using the ensemble. Since the ensemble members are independent, they can be computed in parallel. Hence, if one has λ CPUs available, then the real computational time required in PHMM is identical to that in HMM.

Interestingly, one can draw connections between the parallel method introduced here and the tau-leaping method used in stochastic chemical kinetics [16]. The tau-leaping method is an approximation used to speedup simulation of stochastic fast–slow systems of the type

$$X^\varepsilon(t) = X^\varepsilon(0) + \sum_{k=1}^m \varepsilon \mathcal{N}_k \left(\varepsilon^{-1} \int_0^t a_k(X^\varepsilon(s)) \mathrm{d}s \right) v_k, \tag{8.1}$$

where $\mathcal{N}_k$ are independent unit rate Poisson processes, v_k are vectors in $\mathbb{R}^d$ and $a_k : \mathbb{R}^d \to \mathbb{R}$. The system (8.1) can be solved exactly by the stochastic simulation algorithm (SSA), but when ε is small this can be extremely expensive, due to the Poisson clocks being reset each time a jump occurs. The tau-leaping procedure avoids this issue by chopping the simulation window into subintervals of size τ and on each subinterval fixing the Poisson clocks to their value at the left endpoint. The speedup is a result of the fact that one can simulate the Poisson jumps in parallel, since their clocks are fixed over the τ interval. As a consequence of this analogy, one can check (using calculations similar to those found above) that the tau-leaping method also captures the fluctuations correctly, both at the level of the CLT and that of the LDP. The former observation was made in [3]; to the best of our knowledge, the second one is new.

As a final note, we stress that there are non-dissipative fast–slow systems for which the LDP does not adequately describe the metastability of the system. For this reason, the PHMM method cannot be expected to model the metastability correctly, even though it does capture the LDP correctly. These are systems for which the CLT and LDP hold on $O(1)$ timescale, but they either cannot be extended to longer timescale (in the case of the CLT) or leads to trivial prediction on these timescales (in the case of the LDP). To clarify this point, take, for example, the fast–slow Langevin system

$$\begin{aligned}
&\dot{q}_1 = p_1 \quad \dot{p}_1 = q_1 - q_1^3 + (q_2 - q_1), \\
&\dot{q}_2 = \varepsilon^{-1} p_2 \quad \dot{p}_2 = \varepsilon^{-1}(q_1 - q_2) - \varepsilon^{-1}\gamma p_2 + \sqrt{2\varepsilon^{-1}\beta^{-1}\gamma}\,\eta,
\end{aligned} \tag{8.2}$$

where $\gamma > 0$ and $\beta > 0$ are parameters. For any value of ε, γ, this system is invariant with respect to the Gibbs measure with Hamiltonian

$$H(q_1, q_2, p_1, p_2) = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{4}q_1^4 - \frac{1}{2}q_1^2 + \frac{1}{2}(q_1 - q_2)^2.$$

As $\varepsilon \to 0$, it is easy to check that the slow variables (q_1, q_2) converge to the averaged system

$$\dot{\bar{q}}_1 = \bar{p}_1, \quad \dot{\bar{p}}_1 = -G'(\bar{q}_1), \tag{8.3}$$

where the averaged vector field is the gradient of the free energy

$$G(q_1) = \frac{1}{4}q_1^4 - \frac{1}{2}q_1^2 + \frac{1}{2}q_1^2 = -\beta^{-1}\log \int \exp(-\beta U(q_1, q_2))\mathrm{d}q_2,$$

with $U(q_1, q_2) = \frac{1}{4}q_1^4 - \frac{1}{2}q_1^2 + \frac{1}{2}(q_1 - q_2)^2$. Likewise, if we introduce

$$\eta_1 = \frac{q_1 - \check{q}_1}{\sqrt{\varepsilon}}, \quad \zeta_1 = \frac{p_1 - \check{p}_1}{\sqrt{\varepsilon}},$$

the CLT indicates that the evolution of these variables is captured by

$$\dot{\eta}_1 = \zeta_1, \quad \mathrm{d}\zeta_1 = \sqrt{2\beta^{-1}\gamma}\,\mathrm{d}B \tag{8.4}$$

and we can also derive an LDP for (8.2) with action

$$\mathscr{S}_{[0,T]}(q_1) = \frac{\beta}{4\gamma}\int_0^T |\ddot{q}_1 - q_1 + q_1^3|^2 \mathrm{d}t. \tag{8.5}$$

However, neither (8.4) nor (8.5) captures the longtime behavior of the solution to (8.2). The problem stems from the fact that the averaged equation in (8.3) is Hamiltonian, hence non-dissipative. As a result, fluctuations accumulate as time goes on. Eventually, the CLT stops being valid, and the LDP becomes trivial—in particular, it is easy to see that the quasi-potential associated with the action in (8.5) is flat. For examples of this type, other techniques will have to be employed to describe their longtime behavior including, possibly, their metastability (which, in the case of (8.2) is controlled by how small β^{-1} is, rather than ε). These questions will be investigated elsewhere.

Appendix 1: Non-constant speedup factor

In this section, we briefly discuss the fluctuations of HMM and PHMM under the generalized setting with λ being allowed to depend on ε. Firstly, under the assumptions made in Sect. 4.1, but with $\lambda = \lambda(\varepsilon)$, it is clear that the averaging results for HMM and PHMM still hold provided that $\varepsilon\lambda(\varepsilon) \to 0$ as $\varepsilon \to 0$. It is intuitively clear that the enhancement of fluctuations will be far more pronounced if $\lambda(\varepsilon) \to \infty$ as $\varepsilon \to 0$. To investigate the fluctuations, we consider the example studied in Sect. 7.2, in particular, the semi-discretized approximation

$$\begin{aligned}
x_{n+1}^\varepsilon &= x_n^\varepsilon + \int_0^{\Delta t} Y^\varepsilon(s)\mathrm{d}s - (x_n^\varepsilon)^3 \Delta t, \\
\mathrm{d}Y^\varepsilon &= \frac{\theta}{\varepsilon}(\mu x_n^\varepsilon - Y^\varepsilon)\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\mathrm{d}W.
\end{aligned} \tag{8.6}$$

For HMM, the slow variables are approximated by

$$x_{n+1}^\varepsilon = x_n^\varepsilon + \lambda \int_0^{\Delta t/\lambda} Y^\varepsilon(s)\mathrm{d}s - (x_n^\varepsilon)^3 \Delta t.$$

For PHMM, the slow variables are approximated by

$$x_{n+1}^\varepsilon = x_n^\varepsilon + \sum_{j=1}^{\lambda} \int_0^{\Delta t/\lambda} Y^{\varepsilon,j}(s)\mathrm{d}s - (x_n^\varepsilon)^3 \Delta t,$$

where the $Y^{\varepsilon,j}$ are independent copies of Y^ε. For this example, we can explicitly compute the contribution from the fast scales in the equation for the slow variables. In particular, by explicitly solving for Y^ε we see that the contribution to the slow variables in the original system is

$$\int_0^{\Delta t} Y^\varepsilon(s)\mathrm{d}s = \mu x_n^\varepsilon \Delta t + \frac{\varepsilon}{\theta}(1 - e^{-\frac{\theta\Delta t}{\varepsilon}})(Y_0 - 1) + \frac{\sigma\sqrt{\varepsilon}}{\theta}\int_0^{\Delta t}(1 - e^{-\frac{\theta(\Delta t - r)}{\varepsilon}})\mathrm{d}W_r,$$

whereas for HMM we have (after rescaling the stochastic integral)

$$\lambda \int_0^{\Delta t/\lambda} Y^\varepsilon(s)\mathrm{d}s = \mu x_n^\varepsilon \Delta t + \frac{\varepsilon\lambda}{\theta}(1 - e^{-\frac{\theta\Delta t}{\varepsilon\lambda}})(Y_0 - 1)$$
$$+ \frac{\sigma\sqrt{\varepsilon\lambda}}{\theta}\int_0^{\Delta t}(1 - e^{-\frac{\theta(\Delta t - r)}{\varepsilon\lambda}})\mathrm{d}W_r.$$

And finally, for PHMM we have (after replacing the sum of independent stochastic integrals with a single rescaled stochastic integral)

$$\sum_{j=1}^\lambda \int_0^{\Delta t/\lambda} Y^{\varepsilon,j}(s)\mathrm{d}s = \mu x_n^\varepsilon \Delta t + \frac{\varepsilon\lambda}{\theta}(1 - e^{-\frac{\theta\Delta t}{\varepsilon\lambda}})(Y_0 - 1)$$
$$+ \frac{\sigma\sqrt{\varepsilon}}{\theta}\int_0^{\Delta t}(1 - e^{-\frac{\theta(\Delta t - r)}{\varepsilon\lambda}})\mathrm{d}W_r.$$

From these three calculations, it can be seen that, at the level of the CLT, PHMM will have the same fluctuations as the original system as $\varepsilon \to 0$, for any $\lambda = \lambda(\varepsilon)$ with $\varepsilon\lambda \to 0$ as $\varepsilon \to 0$. For HMM on the other hand, the variance of the fluctuations will be enhanced by a factor of λ, just as we found for the λ constant regime. This simple example indicates that PHMM may be useful for capturing fluctuations even with $\lambda = \lambda(\varepsilon)$, but further investigation is needed before such claims can be made definitive. A similar question concerning rare events in particle systems has been asked in [11].

Dedication
Dedicated with admiration and friendship to Bjorn Engquist on the occasion of his 70th birthday.

Publisher's Note

References
1. Abdulle, A., Weinan, E., Engquist, B., Vanden-Eijnden, E.: The heterogeneous multiscale method. Acta Numer. **21**, 1–87 (2012)
2. Ariel, G., Engquist, B., Kim, S., Lee, Y., Tsai, R.: A multiscale method for highly oscillatory dynamical systems using a Poincaré map type technique. J. Sci. Comput. **54**(2–3), 247–268 (2013)
3. Anderson, D.F., Ganguly, A., Kurtz, T.G.: Error analysis of tau-leap simulation methods. Ann. Appl. Probab. **21**(6), 2226–2262 (2011)
4. Ariel, G., Sanz-Serna, J., Tsai, R.: A multiscale technique for finding slow manifolds of stiff mechanical systems. Multiscale Model. Simul. **10**(4), 1180–1203 (2012)
5. Brünger, A., Brooks, C.L., Karplus, M.: Stochastic boundary conditions for molecular dynamics simulations of ST2 water. Chem. Phys. Lett. **105**(5), 495–500 (1984)
6. Bouchet, F., Grafke, T., Tangarife, T., Vanden-Eijnden, E.: Large deviations in fast–slow systems. J. Stat. Phys. **162**(4), 793–812 (2016)
7. Bal, G., Jing, W.: Corrector theory for MsFEM and HMM in random media. Multiscale Model. Simul. **9**, 1549–1587 (2011)
8. Bal, G., Jing, W.: Corrector analysis of a heterogeneous multi-scale scheme for elliptic equations with random potential. M2AN **48**(2), 387–409 (2014)
9. Chorin, A.: A numerical method for solving incompressible viscous flow problems. J. Comput. Phys **2**, 12–26 (1967)
10. Car, R., Parrinello, M.: Unified approach for molecular dynamics and density functional theory. Phys. Rev. Lett. **55**(22), 2471–2475 (1985)
11. Del Moral, P., Garnier, J., et al.: Genealogical particle analysis of rare events. Ann. Appl. Probab. **15**(4), 2496–2534 (2005)
12. Dolgopyat, D.: Limit theorems for partially hyperbolic systems. Trans. Am. Math. Soc. **356**(4), 1637–1689 (2004)
13. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, vol. 38. Springer, Berlin (2009)
14. Fatkullin, I., Vanden-Eijnden, E.: A computational strategy for multiscale systems with applications to Lorenz 96 model. J. Comput. Phys. **200**(2), 605–638 (2004)
15. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems, vol. 260. Springer, Berlin (2012)
16. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reaction systems. J. Chem. Phys. **115**(4), 1716–1733 (2000)
17. Kifer, Y.: Averaging in dynamical systems and large deviations. Invent. Math. **110**(1), 337–370 (1992)
18. Kelly, D., Melbourne, I.: Deterministic homogenization for fast-slow systems with chaotic noise. J. Funct. Anal **272**(10), 4063–4102 (2017)

19. Kelly, D., Melbourne, I.: Smooth approximations of stochastic differential equations. Ann. Probab. **44**, 479–520 (2016)
20. Mattingly, J.C., Stuart, A.M., Higham, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. Stoch. Process. Appl. **101**(2), 185–232 (2002)
21. Vanden-Eijnden, E.: Numerical techniques for multi-scale dynamical systems with stochastic effects. Commun. Math. Sci. **1**(2), 385–391 (2003)
22. Vanden-Eijnden, E.: On hmm-like integrators and projective integration methods for systems with multiple time scales. Commun. Math. Sci. **5**(2), 495–505 (2007)
23. E, W., Engquist, B.: The heterogeneous multiscale methods. Commun. Math. Sci. **1**(1), 87–132 (2003)
24. E, W., Engquist, B., Li, X., Ren, W., Vanden-Eijnden, E.: Heterogeneous multiscale methods: a review. Commun. Comput. Phys. **2**(3), 367–450 (2007)
25. E, W., Liu, D., Vanden-Eijnden, E.: Analysis of multiscale methods for stochastic differential equations. Commun. Pure Appl. Math. **58**(11), 1544–1585 (2005)
26. Weinan, E., Ren, W., Vanden-Eijnden, E.: A general strategy for designing seamless multiscale methods. J. Comput. Phys. **228**(15), 5437–5453 (2009)