

ITERATED AVERAGING OF THREE-SCALE OSCILLATORY SYSTEMS*

GIL ARIEL[†], BJORN ENGQUIST[‡], SEONG JUN KIM[§], AND RICHARD TSAI[¶]

Abstract. A theory of iterated averaging is developed for a class of highly oscillatory ordinary differential equations (ODEs) with three well separated time scales. The solutions of these equations are assumed to be (almost) periodic in the fastest time scales. It is proved that the dynamics on the slowest time scale can be approximated by an effective ODE obtained by averaging out oscillations. In particular, the effective dynamics of the considered class of ODEs is always deterministic and does not show any stochastic effects. This is in contrast to systems in which the dynamics on the fastest time scale is mixing. The systems are studied from three perspectives: first, using the tools of averaging theory; second, by formal asymptotic expansions; and third, by averaging with respect to fast oscillations using nested convolutions with averaging kernels. The latter motivates a hierarchical numerical algorithm consisting of nested integrators.

Key words. Three-scale oscillatory dynamical system, iterated averaging.

AMS subject classifications. 34E13, 65L04, 65L05.

1. Introduction

Averaging methods are some of the most widely used tools for the study and approximation of highly oscillatory ordinary differential equations (ODEs). In its most basic form, averaging means that following a change of variables that separates the system into slow coordinates and periodic fast ones, the effect of the fast oscillations can be integrated out, yielding an approximate effective equation. More precisely, consider the initial value problem

$$x' = f(x, \epsilon^{-1}t), \quad x(0) = x_0, \quad (1.1)$$

where $0 < \epsilon \leq \epsilon_0 \ll 1$ denotes the scale separation in the problem. It is assumed that a unique uniformly bounded solution $x(t) \in \mathbb{R}^d$ exists in a time segment $[0, T]$ which is independent of ϵ . Furthermore, $f(x, s)$ is sufficiently smooth and 1-periodic in s ; see, for example, [11, 40] for details. Then, $x(t)$ can be approximated by the effective averaged equation

$$\xi' = F(\xi), \quad \xi(0) = x_0, \quad (1.2)$$

where $F(\xi)$ is the average of $f(x, s)$ over one period of s :

$$F(\xi) = \langle f(\xi, \cdot) \rangle_s \equiv \int_0^1 f(\xi, s) ds. \quad (1.3)$$

Furthermore, the approximation is of order ϵ in the sup norm, i.e.

$$\sup_{0 \leq t \leq T} |x(t) - \xi(t)| = \mathcal{O}(\epsilon). \quad (1.4)$$

*Received: November 17, 2012; accepted (in revised version): June 2, 2013. Communicated by Eric Vander-Eijnden.

[†]Bar-Ilan University, Ramat Gan, 52900, Israel (arielg@math.biu.ac.il).

[‡]Department of Mathematics and Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, TX 78712, USA (engquist@ices.utexas.edu).

[§]Department of Mathematics, The University of Texas at Austin, TX 78712, USA (skim@math.utexas.edu).

[¶]Department of Mathematics and Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, TX 78712, USA (ytsai@ices.utexas.edu).

This elementary result has been generalized considerably to include, for example, chaotic and stochastic systems [7, 10, 11, 30, 31, 38, 39, 40]. From a computational point of view, averaging methods inspire efficient numerical schemes for integrating the slow components of slow-fast systems without fully resolving all fast oscillations in $[0, T]$; see, for example, [1, 2, 8, 14, 15, 16, 18, 19, 23, 25, 41, 42, 44] and references therein.

The main purpose of this paper is twofold. First, we generalize the classical two-scale averaging theory (1.1)-(1.4) to a model problem whose solutions possess slow variables or observables as well as fast oscillations with frequencies that are separated by two large spectral gaps. We shall call such systems three-scale oscillatory systems. Our generalization is along the lines of the so-called iterated homogenization problem [9, 10, 12, 29, 34]. Similarly, we develop a theory of iterated averaging which is approached from three different views: (1) application of long time asymptotic and second order averaging theorems [40, 43]; (2) formal perturbation expansions applied to an associated stochastic differential equation (SDE); and (3) a direct proof by repeated convolution with averaging kernels across different time scales. The latter is the foundation for a hierarchical numerical algorithm consisting of nested integrators, which is the second main objective of the paper. We demonstrate how tiers of two-scale numerical methods can be “stacked-up” in order to construct a consistent approximation of the effective slow dynamics in three-scale systems. To our knowledge, very few algorithms have been developed considering directly three or more scales. We develop our algorithm using the framework of the heterogeneous multi-scale method (HMM) for highly oscillatory ODEs [2, 16]. A proof of concept for the algorithm was previously suggested in [6] without the underlying mathematical theory of iterated averaging or proof of convergence.

To gain insight on the analysis of the numerical method described in [6], it is helpful to study the method when it is applied to an appropriately chosen and simpler problem which shows the similar features of the original model but is easier to analyze. We thus develop a simple model, motivated by Fermi, Pasta, and Ulam (FPU) [24], which shows the multiscale behavior in time scales of integer powers of ϵ . The FPU model involves almost-periodic dynamics on the ϵ time scale, dynamics of soft nonlinear springs on the time scale independent of ϵ , and the slow energy exchange among stiff springs on the ϵ^{-1} time scale; see e.g. [28]. We remark however that there of course exist dynamical systems whose multiscale features are not integer powers of ϵ , e.g. in weakly coupled, nearly resonant harmonic oscillators. A method for near-resonance is beyond the scope of this paper, but a new approach for these general systems will be reported in a forthcoming paper.

Consider ODE systems evolving on three well-separated time scales of the form

$$\begin{cases} x' = \epsilon^{-1} \tilde{f}_1(x, y, z) + \tilde{f}_0(x, y, z), & x(0) = x_0, \\ y' = \epsilon^{-1} \tilde{g}_1(x, y, z) + \tilde{g}_0(x, y, z), & y(0) = y_0, \\ z' = \epsilon^{-2} \tilde{h}_2(x, y, z), & z(0) = z_0, \end{cases} \quad (1.5)$$

where the separation into the three components x , y , and z is according to the time-scale on which variables evolve, ϵ^2 , ϵ , and 1, respectively. Throughout this paper we assume the existence of a unique solution for each initial condition $(x_0, y_0, z_0) \in \mathcal{D}_0 \subset \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ which is bounded independent of ϵ in a time segment $[0, T]$ also independent of ϵ . Because the time derivatives of $(x(t), y(t), z(t))$ are not uniformly bounded while solutions are, fast variables $y(t)$ and $z(t)$ are either dissipative (i.e., converge to a stable low dimensional manifold on an ϵ or ϵ^2 time scale), or highly

oscillatory. In this paper we concentrate on the highly oscillatory case. Accordingly, we assume the following:

- For fixed x and y , $z(t) \in \mathbb{R}^{d_2}$ is quasi-periodic (i.e., diffeomorphic to uniform rotations on a d_2 dimensional torus) with frequencies of order ϵ^{-2} over $t \in [0, T]$.
- For fixed x , $y(t) \in \mathbb{R}^{d_1}$ quasi-periodic with frequencies of order ϵ^{-1} over $t \in [0, T]$.
- The variable $x(t) \in \mathbb{R}^{d_0}$ is slow. More precisely, there exists a function $\xi(t)$ which is independent of ϵ such that

$$\sup_{0 < \epsilon \leq \epsilon_0} \sup_{0 \leq t \leq T} |x(t) - \xi(t)| \leq C\epsilon,$$

for some constants ϵ_0, T , and C independent of ϵ . In the following, it will be shown that a necessary condition for x to be slow is that the average of \tilde{f}_1 over z vanishes.

In many situations, one is only interested in the slowly changing quantity $x(t)$. Accordingly, our main objective is to investigate and approximate the effective dynamics $\xi(t)$, discarding the fast variables $y(t)$ and $z(t)$. When only looking for the effective dynamics $\xi(t)$ of the slow variable $x(t)$ in the time interval $[0, \epsilon T]$, one can simply truncate the \tilde{f}_0 term, and average over y and z . On the other hand, in the case of $[0, T]$, we cannot neglect \tilde{f}_0 because its effect appears in this longer time interval.

In this paper, we focus on a simplified version of (1.5) in which the fast scales associated with y and z are given explicitly as known time-dependent functions. Most of the analytical results will be proven in this simpler setup. However, in Section 1.4 we explain heuristically why this simpler form keeps the essential multiscale features of the more general system (1.5) and why our suggested numerical method applies to the full three-scale problem (1.5). We consider ODE systems of the form

$$x' = \epsilon^{-1} f_1(x, \epsilon^{-1}t, \epsilon^{-2}t) + f_0(x, \epsilon^{-1}t, \epsilon^{-2}t), \quad x(0) = x_0, \tag{1.6}$$

where $f_1(x, s_1, s_2)$ and $f_2(x, s_1, s_2)$ are sufficiently smooth (e.g. C^1) and 1-periodic in s_1 and s_2 . As before, it is assumed that a unique bounded solution exists for each $x_0 \in \mathcal{D}_1 \subset \mathbb{R}^d$ in a time segment $[0, T]$. Both the bound and T are independent of ϵ . The dynamics of (1.6) can be characterized across three well-separated time scales: a fast time scale involving time intervals with length of order ϵ^2 , an intermediate scale of order ϵ , and a slow $\mathcal{O}(1)$ scale. In Section 2 we will show that if the average of f_1 with respect to s_2 vanishes, then $x(t)$ effectively varies on the time scale of $\mathcal{O}(1)$, i.e., $x(t)$ is the slow variable. Moreover, effective equations for $x(t)$ will also be given.

In the following, we shall use s_1 and s_2 exclusively for the variables which will be scaled respectively to $\epsilon^{-1}t$ and $\epsilon^{-2}t$. Moreover, averages of a function $f(x, s_1, s_2)$ with respect to s_1 and s_2 with x fixed are denoted respectively by

$$\langle f(\xi, s_1, s_2) \rangle_1, \quad \langle f(\xi, s_1, s_2) \rangle_2.$$

Accordingly, the double average of $f(x, s_1, s_2)$ over both s_1 and s_2 is denoted

$$\langle f(x, s_1, s_2) \rangle_{12}.$$

As motivation, we begin formally. Assuming that for fixed x and s_1 the average of f_1 with respect to s_2 vanishes, $x(t)$ can be written as a sum of a smooth function $w(t)$

whose first time derivative is bounded independent of ϵ , and fast oscillations around $w(t)$ whose amplitudes tend to zero as $\epsilon \rightarrow 0$ [37, 40]. As a motivation for identifying an ODE for $w(t)$, consider the form

$$x(t) = w(t) + \epsilon \phi(w(t), \epsilon^{-1}t, \epsilon^{-2}t).$$

Substituting into (1.6) yields

$$w' = \epsilon^{-1} \left(f_1 - \frac{\partial \phi}{\partial s_2} \right) + (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1} + f_0 + \mathcal{O}(\epsilon). \quad (1.7)$$

In order to have that $w'(t)$ is bounded independent of ϵ we require that $\partial \phi / \partial s_2 = f_1$. Consequently, ϕ should be taken to be the anti-derivative of f_1 with respect to s_2 and

$$w' = (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1} + f_0 + \mathcal{O}(\epsilon). \quad (1.8)$$

Because ϕ is periodic in s_1 and s_2 , $\partial \phi / \partial s_1$ averages to zero and therefore does not contribute to the effective equation. In order to eliminate small oscillations in ϕ , the right hand side (RHS) of (1.8) should be averaged with respect to s_1 and s_2 . This leads to the averaged equation

$$\xi' = F(\xi), \quad \xi(0) = x_0, \quad (1.9)$$

where

$$F = \langle (\nabla_x f_1) \phi + f_0 \rangle_{12}, \quad (1.10)$$

$$\phi(\xi, s_1, s_2) = \int_0^{s_2} f_1(\xi, s_1, \tau) d\tau.$$

REMARK 1.1. The definition of ϕ is not unique. For example, it is straightforward to verify that taking $\phi(\xi, s_1, s_2) = \int_0^{s_2} f_1(\xi, s_1, \tau) d\tau + a(\xi, s_1)$ where, for fixed ξ , $a(\xi, s_1)$ is periodic in s_1 , one would arrive at the same averaged equation. Two canonical choices are either $a(\xi, s_1) = 0$, which is used in sections 2 and 4, or $a(\xi, s_1)$ such that $\langle \phi \rangle_2 = 0$, which is used in Section 3.

REMARK 1.2. Following the previous remark we note that in general, it is not true that $x = w + \epsilon \phi + \mathcal{O}(\epsilon^2)$.

The following theorem is proved in Section 2.

THEOREM 1.1. *Let $x(t)$ and $\xi(t)$ denote solutions of equations (1.6) and (1.9) respectively. If f_1 has a zero average with respect to s_2 , then there exists a constant $C > 0$ independent of ϵ such that*

$$\sup_{0 \leq t \leq T} |x(t) - \xi(t)| \leq C\epsilon.$$

In fact, as will be seen later, the same theorem holds under weaker assumptions on the fast and intermediate time scales s_1 and s_2 . The fast, ϵ^2 scale can be periodic, ergodic on a torus, or almost periodic with frequencies that are bounded from below. On the intermediate $\mathcal{O}(\epsilon)$ scale, the only requirement is that for fixed ξ , the effective dynamics obtained after integrating the fastest ϵ^2 time scale is ergodic and that a two-scale averaging principle [40] holds. This includes oscillatory, stochastic, and chaotic dynamics. Such generalizations are further discussed in Section 5.

1.1. A simple example. Consider the ODE system

$$x' = 2\pi\epsilon^{-1} [x\sin(2\pi\epsilon^{-2}t) + \cos(2\pi\epsilon^{-2}t)], \quad x(0) = 0. \tag{1.11}$$

for $0 \leq t \leq 1$. By the method of variation of parameters, the exact solution of (1.11) is found to be

$$x(t) = 2\pi\epsilon^{-1} e^{-\epsilon\cos(2\pi\epsilon^{-2}t)} \int_0^t \cos(2\pi\epsilon^{-2}s) e^{\epsilon\cos(2\pi\epsilon^{-2}s)} ds. \tag{1.12}$$

Expanding in ϵ ,

$$x(t) = \pi t + \epsilon [\sin(2\pi\epsilon^{-2}t) - \pi t \cos(2\pi\epsilon^{-2}t)] + \mathcal{O}(\epsilon^2). \tag{1.13}$$

Hence, $x(t) = \pi t + \mathcal{O}(\epsilon)$ for $t \in [0, 1]$.

A naive approach for obtaining an effective equation is to assume that in any sufficiently short time interval the solution of (1.11) can be approximated by integrating over the fast oscillations in the RHS of (1.11) while keeping x fixed. This yields

$$\tilde{x}(t) = 2\pi\epsilon^{-1} \int_0^t [X\sin(2\pi\epsilon^{-2}s) + \cos(2\pi\epsilon^{-2}s)] ds \Big|_{X=\tilde{x}(t)} = 2\pi\epsilon^{-1} [X\mathcal{O}(\epsilon^2) + \mathcal{O}(\epsilon^2)], \tag{1.14}$$

which implies that $\tilde{x}(t) = \mathcal{O}(\epsilon)$ only for $t \in [0, C\epsilon]$.

In contrast, Theorem 1.1 yields the correct averaged equation for x for $t \in [0, 1]$:

$$X' = \pi, \quad X(0) = 0,$$

which clearly implies that

$$\sup_{t \in [0, 1]} |x(t) - X(t)| \leq (1 + \pi)\epsilon.$$

On the other hand, performing averages over fast oscillations can be approximated in a convenient and computationally efficient way by convolution of x' with appropriate compactly supported kernels. Inside the convolution, the value of x is not exactly fixed but varies following the correct dynamics. *This subtle change in the values of $x(t)$ allows for the kernel to capture the correct effective change of $x(t)$ in a longer time scale.* This can be demonstrated in the case of example (1.11). Consider a cosine kernel, which is particularly convenient in this example:

$$K(t) = \frac{1}{2} \chi_{[-1, 1]}(t) [1 + \cos(\pi t)],$$

where χ_A is the indicator function of a set A . Furthermore, for $\eta > 0$, let K_η denote a scaling of $K(t)$ to $[-\eta, \eta]$,

$$K_\eta(t) = \frac{1}{\eta} K\left(\frac{t}{\eta}\right).$$

Without loss of generality, we calculate the convolution of K_η with x' in (1.11) at $t = 0$:

$$(K_\eta * x')(0) = \frac{2\pi}{\epsilon} \int_{-\eta}^\eta K_\eta(-s) [x(s)\sin(2\pi\epsilon^{-2}s) + \cos(2\pi\epsilon^{-2}s)] ds.$$

Substituting in the expansion of $x(s)$ shown in equation (1.13), and using $\eta = n\epsilon^2$ yields

$$\begin{aligned} (K_\eta * x')(0) &= \frac{\pi}{\eta\epsilon} \int_{-n\epsilon^2}^{n\epsilon^2} \left[1 + \cos\left(\frac{\pi s}{\eta}\right) \right] \left[\pi s \sin\left(\frac{2\pi n s}{\eta}\right) + \cos\left(\frac{2\pi n s}{\eta}\right) \right] ds \\ &\quad + \frac{\pi}{\eta} \int_{-n\epsilon^2}^{n\epsilon^2} \left[1 + \cos\left(\frac{\pi s}{\eta}\right) \right] \left[\sin^2\left(\frac{2\pi n s}{\eta}\right) - \frac{\pi s}{2} \sin\left(\frac{4\pi n s}{\eta}\right) \right] ds + \mathcal{O}(\epsilon) \\ &= \pi + \mathcal{O}(\epsilon), \end{aligned}$$

which is, to leading order in ϵ , the correct derivative for the slow variable x for $t \in [0, 1]$.

In Section 4 these ideas are generalized to oscillatory three-scale systems in which kernels are applied iteratively to the different time scales. This can then be exploited for construction of efficient multiscale numerical schemes.

1.2. Formal asymptotic expansions. Formal asymptotic expansions of singularly perturbed operators have been successfully applied to a wide variety of problems [10, 39]. Of particular relevance are the applications to SDEs [15, 35, 38, 44]. Consider SDE systems of the form

$$\begin{cases} dx_t = [\epsilon^{-1} f_1(x_t, y_t) + f_0(x_t, y_t)] dt, \\ dy_t = \epsilon^{-2} a(x_t, y_t) dt + \epsilon^{-1} \beta(x_t, y_t) dB_t, \end{cases} \quad (1.15)$$

where B_t is a standard Brownian motion in \mathbb{R}^d . The variable $x_t \in \mathbb{R}^d$ is a slow process that evolves according to an ODE with a fast random coefficient $y_t \in \mathbb{R}^n$. Under some ergodicity, smoothness, and growth assumptions, x_t can be approximated by an effective equation of the form

$$dX_t = F(X_t) dt + b(X_t) dB_t, \quad (1.16)$$

where F and b can be expressed as averages with respect to the fast process y_t with X_t fixed [38, 39]. One of the interesting consequences of (1.16) is that if the dynamics of y_t is mixing¹, then the effective diffusion $b(x)$ may be non-zero even if y_t is deterministic, i.e., $\beta = 0$. For details and examples, see [17, 23, 26, 27, 35, 36, 39].

Many oscillatory dynamics are not mixing, even though for fixed x_t the fast dynamics is ergodic on a low-dimensional invariant manifold. In Section 3 we show that the method of formal asymptotic expansions gives the correct vanishing effective diffusion coefficient, i.e., the effective slow dynamics is deterministic. Furthermore, the method reproduces the correct effective drift.

1.3. Numerical methods. The above discussion on applying averaging kernels across different time scales motivates a numerical method which applies our previous two-scale HMM algorithms [3, 4, 5] hierarchically to multiple (> 2) timescale systems. We consider the time scales $\mathcal{O}(\epsilon^2)$, $\mathcal{O}(\epsilon)$, and $\mathcal{O}(1)$ and assume that both the ϵ and ϵ^2 scales are oscillatory.

The HMM to be constructed should evaluate the effective rate of change of $x(t)$. For three-scale problems this requires averaging over the $\mathcal{O}(\epsilon^2)$ as well as the $\mathcal{O}(\epsilon)$

¹Loosely speaking, mixing means that for any two possible states of y_t , the occurrence of the states is independent if a sufficient amount of time t is given. For the precise definition and properties, see [45].

scale oscillations, thus obtaining a numerical approximation for the effective equation. See [6] for further details.

The hierarchical HMM structure is illustrated in figure 1.1. The downward pointing arrows depict the determination of an initial condition for a lower, fast scale from data in an upper tier working on a slower time scale. The upward pointing arrows from 2nd tier to 1st tier and 1st tier to 0th tier relate the evaluation of averages with respect to s_2 and s_1 , respectively. Below we describe in detail the equations solved in each tier.

Let η_i and h_i denote the range of integration and step size used in the i -th tier, respectively. A chosen ODE solver in the 2nd tier numerically approximates the full ODE at the initial time $t_{n,m} = nh_0 + mh_1$.

$$x'_2 = \epsilon^{-1} f_1(x_2, \epsilon^{-1}t, \epsilon^{-2}t) + f_0(x_2, \epsilon^{-1}t, \epsilon^{-2}t), \quad x_2(t_{n,m}) = X_1,$$

in a time segment $t \in [t_{n,m} - \eta_2, t_{n,m} + \eta_2]$ where $X_1(t_{n,m}; X_0(t_n))$ is an approximation of $x(t)$ at $t_{n,m}$ obtained from the 1st tier. Denote the solution by $x_2(t; X_1)$ and let

$$\mathcal{F}_1(t_{n,m}; X_1) = K_{\eta_2} * (\epsilon^{-1} f_1(x_2(\cdot; X_1), \epsilon^{-1}\cdot, \epsilon^{-2}\cdot) + f_0(x_2(\cdot; X_1), \epsilon^{-1}\cdot, \epsilon^{-2}\cdot)).$$

The 1st tier numerically approximates the effective ODE for the $\mathcal{O}(\epsilon)$ scale,

$$x'_1 = \mathcal{F}_1(t; X_1), \quad x_1(t_n) = X_0,$$

in the time interval $t \in [t_n - \eta_1, t_n + \eta_1]$ where $X_0(t_n)$ is an approximation of $x(t)$ at $t_n = nh_0$ obtained from the 0th tier. Denote the solution $x_1(t; X_0)$ and let

$$\mathcal{F}_0(t_n; X_0) = K_{\eta_1} * \mathcal{F}_1(\cdot).$$

Finally, the 0th tier numerically approximates the effective ODE for the $\mathcal{O}(1)$ scale,

$$x' = \mathcal{F}_0(x), \quad x(0) = x_0,$$

in $t \in [0, \eta_0] = [0, T]$.

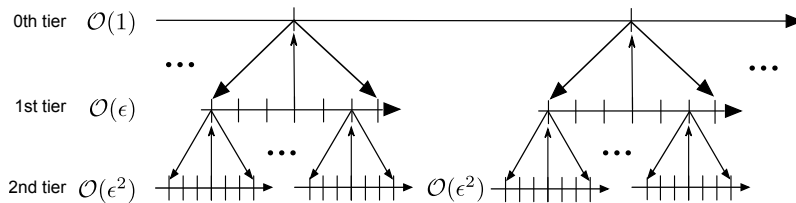


FIG. 1.1. An illustration of a three scale algorithm.

1.4. Generalizing from (1.6) to (1.5). In the following, we explain heuristically why the numerical method described above, which was motivated by the simple model (1.6), may be applied to a more general case (1.5) in which the fast and intermediate dynamics is periodic. Further generalizations to quasi and almost periodic systems are discussed in Section 5. As an analogy to the action-angle coordinates for mechanical systems, z and y can be considered as periodic angle-like coordinates with periods of order ϵ^2 and ϵ , respectively. In other words, we explicitly assume that $z(t)$

consists of a fast $\mathcal{O}(\epsilon^2)$ oscillation which is super-imposed on a slower $\mathcal{O}(\epsilon)$ oscillation and a slow $\mathcal{O}(1)$ drift. Similarly, $y(t)$ is an intermediate $\mathcal{O}(\epsilon)$ oscillation which is super-imposed on a $\mathcal{O}(1)$ drift. More precisely, recall the equation for z in (1.5),

$$z' = \epsilon^{-2} \tilde{h}_2(x, y, z).$$

On an $\mathcal{O}(\epsilon^2)$ time interval both x and y are, to leading order in ϵ , constant. Hence, $z(t)$ is oscillatory and satisfies an ODE of the form

$$z' = \epsilon^{-2} \tilde{h}_2(x(t_0), y(t_0), z) + \epsilon^{-1} \tilde{h}_1(x(t_0), y(t_0), z) + \mathcal{O}(1).$$

This implies that the solution, $z(t)$, can be generally written as

$$z(t) = \tilde{z}(t, \epsilon^{-1}t, \epsilon^{-2}t), \quad (1.17)$$

where $\tilde{z}(s_0, s_1, s_2)$ is quasi-periodic in s_2 . Similarly, recall the equation for y in (1.5),

$$y' = \epsilon^{-1} \tilde{g}_1(x, y, z) + \tilde{g}_0(x, y, z).$$

Substituting in (1.17) and using the $\mathcal{O}(\epsilon^2)$ periodicity of z , y can be approximated by an averaged equation on a time segment of length $\mathcal{O}(\epsilon)$. The averaged equation takes the form

$$\bar{y}' = \epsilon^{-1} \bar{g}_1(x, \bar{y}) + \bar{g}_0(x, \bar{y}).$$

Hence, the solution of $y(t)$ can be written as

$$y(t) = \tilde{y}_0(t, \epsilon^{-1}t) + \epsilon \tilde{y}_1(t, \epsilon^{-1}t, \epsilon^{-2}t) + \mathcal{O}(\epsilon^2), \quad (1.18)$$

where $\tilde{y}_0(s_0, s_1)$ is quasi-periodic in s_1 , and $\tilde{y}_1(s_0, s_1, s_2)$ is quasi-periodic in s_2 . Substituting (1.17) and (1.18) into the equation for x in (1.5) yields the simplified form (1.6).

Finally, we note that HMM only requires integration on reduced time segments. Specifically, computing an approximation of the effective equation for x only requires solving for y on time segments of length $\mathcal{O}(\epsilon)$. Similarly, computing an approximation of the effective equation for y only requires solving for z on time segments of length $\mathcal{O}(\epsilon^2)$. Thus (1.17) and (1.18) are consistent with the requirements of the iterated kernel, nested HMM scheme. Furthermore, the form in which the microscopic scales are given—either (1.5) or (1.6)—is not important for the HMM scheme to yield a converging approximation of the the averaged equation for the slow variable x . The only requirement is that all integrators across all time scales are stable. We conclude that the tiered HMM scheme can be applied to the full three-scale problem (1.5).

Further evidence on the applicability of the nested-HMM framework to the general (1.5) comes from Section 3, which considers formal asymptotic expansions applied directly to (1.5).

1.5. Layout. The layout of the paper is as follows. Section 2 details a proof of Theorem 1.1 using the tools of averaging theory. In Section 3, the same result is derived using formal asymptotic expansions to singular perturbations of SDEs in which the white noise is turned off. Even though this method does not constitute a rigorous proof, its scope applies to the general system given in singular perturbation form (1.5). Section 4 proves that the effective dynamics of (1.6) can be approximated using convolution with respect to averaging kernels which are applied iteratively to

the different time scales. The methods can be used in a numerical HMM scheme as described above. The HMM scheme suggests that the basic idea of iterated averaging can be applied in more general settings. Such generalizations are discussed in Section 5. A few examples of such generalizations are presented in Section 6. We conclude in Section 7.

2. A theory of iterated averaging

In this section we prove Theorem 1.1, which generalizes an averaging theorem for long time scales due to van der Burgh [40, 43]. Further generalizations are discussed in Section 5.

2.1. Basic estimate. As before, let $x(t)$ solve

$$x' = \epsilon^{-1}f_1(x, \epsilon^{-1}t, \epsilon^{-2}t) + f_0(x, \epsilon^{-1}t, \epsilon^{-2}t), \quad x(0) = x_0, \tag{2.1}$$

where $f_1(x, s_1, s_2)$ and $f_0(x, s_1, s_2)$ are sufficiently smooth and 1-periodic in s_1 and s_2 . It is further assumed that the solution $x(t)$ exists, is unique and remains bounded independent of ϵ for a time segment $[0, T]$ independent of ϵ . When $\langle f_1 \rangle_2 = 0$, then $x(t)$ can be approximated by a slow trajectory, and then our goal is to derive an approximate ODE for such a trajectory. As usual, by slow we mean that the first derivative of a time-dependent function is bounded independent of ϵ .

We consider two functions $w(t)$ and $y(t)$. Let $w(t)$ solve

$$w' = h(w(t), \epsilon^{-1}t, \epsilon^{-2}t) + f_0(w(t), \epsilon^{-1}t, \epsilon^{-2}t), \quad w(0) = x_0, \tag{2.2}$$

where h is defined by

$$\begin{aligned} h(x, s_1, s_2) &= (\nabla_x f_1)\phi(x, s_1, s_2) - \frac{\partial \phi}{\partial s_1}(x, s_1, s_2), \\ \phi(x, s_1, s_2) &= \int_0^{s_2} f_1(x, s_1, \tau) d\tau. \end{aligned} \tag{2.3}$$

Note that for fixed x , $h(x, s_1, s_2)$ and $\phi(x, s_1, s_2)$ are 1-periodic in s_1 and s_2 .

NOTATION 2.1. We use the notation $(\nabla_x f_1)\phi(x, s_1, s_2)$ for the multiplication of ϕ by the derivative of f_1 with respect to x , and both are evaluated at (x, s_1, s_2) .

Let $y(t)$ solve

$$y' = G(y, \epsilon^{-1}t), \quad y(0) = x_0, \tag{2.4}$$

where G is given by

$$G(y, s_1) = \langle h(y, s_1, s_2) \rangle_2 + \langle f_0(y, s_1, s_2) \rangle_2. \tag{2.5}$$

We will show that for $t \in [0, T]$, there exist nonnegative constants C_0 and C_1 , independent of ϵ , such that

$$|x(t) - w(t)| \leq C_0\epsilon \quad \text{and} \quad |w(t) - y(t)| \leq C_1\epsilon.$$

Thus, we conclude by the triangle inequality for $t \in [0, T]$ that

$$|x(t) - y(t)| \leq C\epsilon.$$

We denote a generic positive constant by C whose value may change between expressions but is independent of ϵ .

LEMMA 2.1. *The solutions $x(t)$ and $w(t)$ defined above satisfy*

$$|x(t) - w(t)| \leq \epsilon C e^{(L_{f_0} + L_h)t}$$

for $t \in [0, T]$ and $T > 0$ which is independent of ϵ . L_{f_0} and L_h are Lipschitz constants for f_0 and h , respectively.

Proof. From (2.1) and (2.2), integrating with respect to time yields that

$$\begin{aligned} x(t) - x_0 &= \int_0^t \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau + \int_0^t f_0(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau, \\ w(t) - x_0 &= \int_0^t h(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau + \int_0^t f_0(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau. \end{aligned}$$

This leads to

$$|x(t) - w(t)| \leq \left| \int_0^t \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) - h(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right| + L_{f_0} \int_0^t |x(\tau) - w(\tau)| d\tau. \tag{2.6}$$

where L_{f_0} is a Lipschitz constant for f_0 . We will show that the first integral in (2.6) is bounded by

$$L_h \int_0^t |x(\tau) - w(\tau)| d\tau + \mathcal{O}(\epsilon),$$

where L_h is a Lipschitz constant for h . Then, we have

$$|x(t) - w(t)| \leq (L_{f_0} + L_h) \int_0^t |x(\tau) - w(\tau)| d\tau + \mathcal{O}(\epsilon).$$

It follows from Gronwall's inequality that

$$|x(t) - w(t)| \leq \epsilon C e^{(L_{f_0} + L_h)t}.$$

To this end, let n denote the largest integer such that $\epsilon^2 n \leq t$. We first consider $\int_0^1 \epsilon^{-1} f_1 d\tau$.

$$\begin{aligned} &\int_0^t \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \\ &= \sum_{j=0}^{n-1} \int_{\epsilon^2 j}^{\epsilon^2(j+1)} \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau + \int_{\epsilon^2 n}^t \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau. \end{aligned}$$

Because f_1 is bounded and $0 \leq t - n\epsilon^2 < \epsilon^2$, the last term is $\mathcal{O}(\epsilon)$. Denoting $t_j = \epsilon^2 j$ and using the periodicity of f_1 in s_2 , each term in the sum can be written as

$$\epsilon \int_0^1 f_1(x(t_j + \epsilon^2 s), \epsilon^{-1}t_j + \epsilon s, s) ds. \tag{2.7}$$

By the fundamental theorem of calculus, we write

$$x(t_j + \epsilon^2 s) = x(t_j) + \int_{t_j}^{t_j + \epsilon^2 s} x'(\tau) d\tau$$

$$\begin{aligned}
 &= x(t_j) + \int_{t_j}^{t_j+\epsilon^2 s} [\epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) + f_0(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau)] d\tau \\
 &= x(t_j) + \int_{t_j}^{t_j+\epsilon^2 s} \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau + \mathcal{O}(\epsilon^2), \tag{2.8}
 \end{aligned}$$

because f_0 is bounded independent of ϵ . After the substitution of (2.8) into (2.7) and using a Taylor expansion of $f_1(x(t_j + \epsilon^2 s), \epsilon^{-1}t_j + \epsilon s, s_2)$ around $s=0$ while keeping s_2 fixed,

$$\begin{aligned}
 &\epsilon \int_0^1 f_1(x(t_j + \epsilon^2 s), \epsilon^{-1}t_j + \epsilon s, s) ds \\
 &= \epsilon \int_0^1 \left\{ f_1(x(t_j), \epsilon^{-1}t_j, s) + \left(\int_{t_j}^{t_j+\epsilon^2 s} x'(\tau) d\tau \right) \nabla_x f_1(x(t_j), \epsilon^{-1}t_j, s) \right. \\
 &\quad \left. + \epsilon s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, s) \right\} ds + \mathcal{O}(\epsilon^3) \\
 &= \int_0^1 \left(\int_{t_j}^{t_j+\epsilon^2 s} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right) \nabla_x f_1(x(t_j), \epsilon^{-1}t_j, s) ds \\
 &\quad + \int_0^1 \epsilon^2 s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, s) ds + \mathcal{O}(\epsilon^3).
 \end{aligned}$$

where we used the fact that $\langle f_1 \rangle_2 = 0$. Hence, for $\epsilon^{-1} \int_0^t f_1 d\tau$,

$$\begin{aligned}
 &\epsilon^{-1} \int_0^t f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \\
 &= \sum_{j=0}^{n-1} \left[\int_0^1 \left(\int_{t_j}^{t_j+\epsilon^2 s} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right) \nabla_x f_1(x(t_j), \epsilon^{-1}t_j, s) ds \right. \\
 &\quad \left. + \int_0^1 \epsilon^2 s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, s) ds \right] + \mathcal{O}(\epsilon^1). \tag{2.9}
 \end{aligned}$$

Next, changing of variables to $s = \epsilon^{-2}\tau$ in $\int_0^t h d\tau$ gives

$$\begin{aligned}
 &\int_0^t h(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \\
 &= \sum_{j=0}^{n-1} \left[\int_{t_j}^{t_{j+1}} \left(\int_0^{\epsilon^{-2}\tau} f_1(w(\tau), \epsilon^{-1}\tau, v) dv \right) \nabla_x f_1(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right. \\
 &\quad \left. - \int_{t_j}^{t_{j+1}} \int_0^{\epsilon^{-2}\tau} \frac{\partial f_1}{\partial s_1}(w(\tau), \epsilon^{-1}\tau, \tilde{v}) d\tilde{v} d\tau \right] + \mathcal{O}(\epsilon^2) \\
 &= \sum_{j=0}^{n-1} \left[\int_j^{j+1} \epsilon^2 \left(\int_j^s f_1(w(\epsilon^2 s), \epsilon s, v) dv \right) \nabla_x f_1(w(\epsilon^2 s), \epsilon s, s) ds \right. \\
 &\quad \left. - \int_j^{j+1} \epsilon^2 \left(\int_0^s \frac{\partial f_1}{\partial s_1}(w(\epsilon^2 s), \epsilon s, \tilde{v}) d\tilde{v} \right) ds \right] + \mathcal{O}(\epsilon^2).
 \end{aligned}$$

Changing variables back to $\tau = \epsilon^2 v$ in the first integral and $\tilde{v} = \tau$ in the second integral,

$$\begin{aligned} & \int_0^t h(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \\ &= \sum_{j=0}^{n-1} \left[\int_j^{j+1} \left(\int_{t_j}^{\epsilon^2 s} f_1(w(\epsilon^2 s), \epsilon s, \epsilon^{-2}\tau) d\tau \right) \nabla_x f_1(w(\epsilon^2 s), \epsilon s, s) ds \right. \\ & \quad \left. - \epsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(\epsilon^2 s), \epsilon s, \tau) d\tau ds \right] + \mathcal{O}(\epsilon^2). \end{aligned} \quad (2.10)$$

We need to compare (2.9) and (2.10). Note that $w'(t)$ is bounded independent of ϵ and that $\tau \in [j, j+1]$,

$$\begin{cases} w(\epsilon^2 s) = w(\tau) + \mathcal{O}(\epsilon^2 s - \tau) = w(\tau) + \mathcal{O}(\epsilon^2), \\ w(\epsilon^2 j) = w(\epsilon^2 j) + \mathcal{O}(\epsilon^2 s - \epsilon^2 j) = w(t_j) + \mathcal{O}(\epsilon^2), \end{cases} \quad (2.11)$$

and

$$\begin{cases} \epsilon s = \epsilon^{-1}\tau + \mathcal{O}(\epsilon), \\ \epsilon s = \epsilon^{-1}t_j + \mathcal{O}(\epsilon). \end{cases} \quad (2.12)$$

Therefore, the first integration in (2.10) can be written as

$$\begin{aligned} & \int_j^{j+1} \left(\int_{t_j}^{\epsilon^2 s} f_1(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right) \nabla_x f_1(w(t_j), \epsilon^{-1}t_j, s) ds + \mathcal{O}(\epsilon^3) \\ &= \int_0^1 \left(\int_{t_j}^{t_j + \epsilon^2 s} f_1(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right) \nabla_x f_1(w(t_j), \epsilon^{-1}t_j, s) ds + \mathcal{O}(\epsilon^3), \end{aligned}$$

which has the same form of the first term in (2.9) with an $\mathcal{O}(\epsilon^3)$ error.

Now compare the second integrals. Applying integration by parts to the second term in (2.9) and $\langle \frac{\partial f_1}{\partial s_1} \rangle = 0$ give

$$\begin{aligned} & \int_0^1 \epsilon^2 s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, s) ds \\ &= \left[\epsilon^2 s \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, \tau) d\tau \right]_{s=0}^1 - \epsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, \tau) d\tau ds \\ &= -\epsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \epsilon^{-1}t_j, \tau) d\tau ds. \end{aligned} \quad (2.13)$$

On the other hand, in (2.10) using (2.11) and (2.12),

$$\begin{aligned} -\epsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(\epsilon^2 s), \epsilon s, \tau) d\tau ds &= -\epsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(t_j), \epsilon^{-1}t_j, \tau) d\tau ds + \mathcal{O}(\epsilon^3) \\ &= -\epsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(w(t_j), \epsilon^{-1}t_j, \tau) d\tau ds + \mathcal{O}(\epsilon^3). \end{aligned} \quad (2.14)$$

Thus, it is shown that (2.13) and (2.14) are different up to an $\mathcal{O}(\epsilon^3)$ error. Putting all estimates together and noting $n = \mathcal{O}(\epsilon^{-2})$, we conclude that

$$\left| \int_0^t \epsilon^{-1} f_1(x(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) - h(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau \right| \leq L_h \int_0^t |x(\tau) - w(\tau)| d\tau + \mathcal{O}(\epsilon), \tag{2.15}$$

where L_h is a Lipschitz constant for the function h . Substituting (2.15) into (2.6) yields the desired estimate. \square

The following lemma gives an estimate of how much $w(t)$ and $y(t)$ can be apart in $0 \leq t \leq T$.

LEMMA 2.2. *Let $w(t)$ and $y(t)$ be defined as above. Then the following estimate holds:*

$$|w(t) - y(t)| \leq \epsilon C e^{L_G t}, \tag{2.16}$$

for some constant $C > 0$ which is independent of ϵ , where L_G is a Lipschitz constant for G .

Proof. Consider $w' - y'$. Using (2.2) and (2.4) we have

$$\begin{aligned} w' - y' &= f_0(w, \epsilon^{-1}t, \epsilon^{-2}t) - \langle f_0 \rangle_2(y, \epsilon^{-1}t) + h(w, \epsilon^{-1}t, \epsilon^{-2}t) - \langle h \rangle_2(y, \epsilon^{-1}t) \\ &= [G(w, \epsilon^{-1}t) - G(y, \epsilon^{-1}t)] + z(w, \epsilon^{-1}t, \epsilon^{-2}t), \end{aligned}$$

where

$$\begin{aligned} G(w, s_1) &= \langle h(w, s_1, s_2) \rangle_2 + \langle f_0(w, s_1, s_2) \rangle_2, \\ z(w, s_1, s_2) &= [f_0(w, s_1, s_2) - \langle f_0 \rangle_2(w, s_1)] + [h(w, s_1, s_2) - \langle h \rangle_2(w, s_1)]. \end{aligned}$$

Note that for fixed w and s_1 , $z(w, s_1, s_2)$ is 1-periodic in s_2 with a zero average, $\langle z \rangle_2 = 0$. Integrating to the time t yields that

$$w(t) - y(t) = \int_0^t [G(w(\tau), \epsilon^{-1}\tau) - G(y(\tau), \epsilon^{-1}\tau)] d\tau + r(t; \epsilon),$$

where $r(t; \epsilon) = \int_0^t z(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau$. Taking absolute values,

$$|w(t) - y(t)| \leq \int_0^t |G(w(\tau), \epsilon^{-1}\tau) - G(y(\tau), \epsilon^{-1}\tau)| d\tau + |r(t; \epsilon)|.$$

Letting L_G denote a Lipschitz constant for G ,

$$|w(t) - y(t)| \leq L_G \int_0^t |w(\tau) - y(\tau)| d\tau + |r(t; \epsilon)|. \tag{2.17}$$

In order to evaluate $r(t; \epsilon)$, let n denote the largest integer such that $\epsilon^2 n \leq t$. We have

$$r(t; \epsilon) = \sum_{j=0}^{n-1} \int_{\epsilon^2 j}^{\epsilon^2(j+1)} z(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau + \int_{\epsilon^2 n}^t z(w(\tau), \epsilon^{-1}\tau, \epsilon^{-2}\tau) d\tau. \tag{2.18}$$

Because z is bounded and $0 \leq t - n\epsilon^2 < \epsilon^2$, the last term is $\mathcal{O}(\epsilon^2)$. As in the proof of Lemma 2.1, denoting $t_j = \epsilon^2 j$ and using the periodicity of z in s , each term in the sum can be written as

$$\epsilon^2 \int_0^1 z(w(t_j + \epsilon^2 s), \epsilon^{-1}t_j + \epsilon s, s) ds.$$

Expanding in ϵ yields

$$\epsilon^2 \int_0^1 z(w(t_j), \epsilon^{-1}t_j, s) ds = \epsilon^2 [\langle z \rangle_2(w(t_j), \epsilon^{-1}t_j) + \mathcal{O}(\epsilon)] = \mathcal{O}(\epsilon^3).$$

Therefore,

$$r(t; \epsilon) = \mathcal{O}(n\epsilon^3) = \mathcal{O}(\epsilon). \tag{2.19}$$

Substituting into (2.17) yields

$$|w(t) - y(t)| \leq L_G \int_0^t |w(\tau) - y(\tau)| d\tau + \mathcal{O}(\epsilon).$$

A use of Gronwall's inequality concludes the proof of Lemma 2.2. □

2.2. Proof of Theorem 1.1. Lemmas 2.1 and 2.2 imply that

$$\sup_{0 \leq t \leq T} |x(t) - y(t)| \leq \epsilon C, \tag{2.20}$$

where $y(t)$ is the solution of

$$y' = G(y, \epsilon^{-1}t), \quad y(0) = x_0,$$

with

$$\begin{aligned} G(y, s_1) &= \langle h(y, s_1, s_2) \rangle_2 + \langle f_0(y, s_1, s_2) \rangle_2, \\ h(y, s_1, s_2) &= (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1}, \\ \phi(y, s_1, s_2) &= \int_0^{s_2} f_1(y, s_1, \tau) d\tau. \end{aligned} \tag{2.21}$$

Because the right hand side of $G(y, s_1)$ is periodic in s_1 , we are in a position to apply the two-scale averaging theorem [40], integrating out the intermediate $\mathcal{O}(\epsilon)$ time scale. Noting that because ϕ is 1-periodic in s_1 , $\partial \phi / \partial s_1$ is also periodic and has zero average with respect to s_1 . This leads to an averaged equation for $y(t)$:

$$\xi' = F(\xi), \quad \xi(0) = x_0, \tag{2.22}$$

where

$$F(\xi) = \langle (\nabla_x f_1) \phi(\xi, s_1, s_2) \rangle_{12} + \langle f_0(\xi, s_1, s_2) \rangle_{12},$$

and we have

$$\sup_{0 \leq t \leq T} |y(t) - \xi(t)| \leq \epsilon C. \tag{2.23}$$

Combining (2.20) with (2.23) completes the proof of Theorem 1.1.

REMARK 2.2. It is not difficult to generalize this result to systems with non-commensurate and widely separated frequencies.

3. Formal asymptotic expansions

In this section we analyze the multiscale structure of a system using the operator formalism firstly developed by Papanicolaou et al. as a formal asymptotic expansion for singular perturbations of SDEs [10, 38]. This approach has been further generalized and applied to many different problems, for example, in [35, 39, 44]. For the case of Hamiltonian dynamics, including integrable periodic systems, a rigorous version of formalism is presented in [21, 22] and references therein. The derivation in this section is formal; nonetheless, it is instructive and provides intuitive explanation for Theorem 1.1.

3.1. Stochastic differential equations. For completeness, we begin by reviewing singular perturbation expansions of SDEs. The resulting effective equations are then considered in the case in which all diffusion coefficients are formally set to zero.

Consider SDE systems of the form

$$\begin{cases} dx_t = [\epsilon^{-1} f_1(x_t, y_t) + f_0(x_t, y_t)] dt, \\ dy_t = \epsilon^{-2} a(x_t, y_t) dt + \epsilon^{-1} \beta(x_t, y_t) dB_t, \end{cases} \tag{3.1}$$

where $x_t \in \mathbb{R}^d$ is a slow process that evolves according to an ODE with a fast random coefficient $y_t \in \mathbb{R}^n$. We assumed that $\beta(x, y)\beta^T(x, y)$ is uniformly positive definite in $\mathbb{R}^d \times \mathbb{R}^n$. Furthermore, we assume that for fixed x_t the dynamics of y_t is ergodic on an invariant set Σ_x with a unique invariant measure $d\mu_x$. The expectations with respect to the invariant measures are denoted $\langle \cdot \rangle_y$, in which the x dependence is suppressed. A necessary condition for x to be slow is that $\langle f_1 \rangle_y = 0$. Otherwise, x exhibits non-trivial dynamics on the $\mathcal{O}(\epsilon)$ time scale. It is well known that under some suitable conditions on (3.1), $x(t)$ satisfies an effective SDE that is independent of ϵ ,

$$dX_t = F(X_t)dt + b(X_t)dB_t, \tag{3.2}$$

where F and b can be expressed as averages with respect to the fast process $y(t)$ with X fixed [10, 38, 39, 44]. We begin with a brief overview of the relevant results of [38]. For details the reader is referred to [39] and references therein.

The backwards equation that governs the evolution of a probability density, ϕ , of the initial conditions

$$\begin{cases} \partial_t u = \mathcal{L}u, \\ u(0, x, y) = \varphi(x, y), \end{cases}$$

where \mathcal{L} , the generator of (3.1), can be written as

$$\begin{aligned} \mathcal{L} &= \epsilon^{-2} \mathcal{L}_2 + \epsilon^{-1} \mathcal{L}_1 + \mathcal{L}_0, \\ \mathcal{L}_0 &= f_0 \cdot \nabla_x, \\ \mathcal{L}_1 &= f_1 \cdot \nabla_x, \\ \mathcal{L}_2 &= a \cdot \nabla_y + \frac{1}{2} \beta \beta^T : \nabla_y \nabla_y^T, \end{aligned} \tag{3.3}$$

where $A:B$ denotes formally the trace of the matrix AB^T . Next, consider a formal asymptotic expansion of u in ϵ

$$u(t, x, y) = u_0(x) + \epsilon u_1 + \epsilon^2 u_2 + \dots,$$

where we assumed that the leading order term u_0 only depends on the slow process. Substituting into the backwards equation yields

$$\mathcal{L}_2 u_0 = 0, \quad (3.4)$$

$$\mathcal{L}_2 u_1 = -\mathcal{L}_1 u_0, \quad (3.5)$$

$$\mathcal{L}_2 u_2 = \partial_t u_0 - \mathcal{L}_1 u_1 - \mathcal{L}_0 u_0. \quad (3.6)$$

The leading order equation (3.4) is automatically satisfied because u_0 only depends on x . Let \mathcal{L}_2^* be the L^2 adjoint of \mathcal{L}_2 , and assume that the Null space of \mathcal{L}_2^* is a one dimensional subspace, spanned ρ_x , the density of the invariant measure $d\mu_x$.²

Applying the Fredholm alternative, equation (3.5) has a solution if $f_1 \nabla_x u_0$ is perpendicular to the Null space of \mathcal{L}_2^* . In other words, the projection of $f_1 \nabla_x u_0$ on Null \mathcal{L}_2^* should vanish. This projection amounts to averaging with respect to the invariant measure of y (at fixed x), which is also equivalent to taking the standard L_2 inner product in Σ_x with ρ . This yields the condition

$$\langle f_1(x, y) \rangle_y = 0,$$

which implies the reasonable requirement that the average of f_1 vanishes. Otherwise, x oscillates with large amplitudes on the ϵ time scale and thus cannot be approximated by a slow variable. Then, (3.5) has a unique solution such that $\langle u_1 \rangle_y = 0$. We formally write

$$u_1 = -\mathcal{L}_2^{-1}[\mathcal{L}_1 u_0].$$

Applying again the Fredholm alternative, equation (3.6) also has a solution if the RHS is perpendicular to the Null space of \mathcal{L}_2^* , i.e.

$$\langle \partial_t u_0 - \mathcal{L}_1 u_1 - \mathcal{L}_0 u_0 \rangle_y = 0.$$

Substituting in \mathcal{L}_0 , \mathcal{L}_1 , and u_1 we obtain

$$\partial_t u_0 = F(x) \nabla_x u_0 + B(x) : \nabla_x \nabla_x u_0, \quad (3.7)$$

where

$$\begin{aligned} F(x) &= \langle -f_1 \cdot \nabla_x \mathcal{L}_2^{-1}[f_1] + f_0 \rangle_y, \\ B(x) &= \langle -f_1 \mathcal{L}_2^{-1}[f_1] \rangle_y. \end{aligned} \quad (3.8)$$

We identify (3.7) as the backwards equation associated with the effective SDE (3.2) and $2B(x) = b(x)b(x)^T$.

If $\beta\beta^T$ is uniformly positive, then inverting \mathcal{L}_2 amounts to solving a uniformly elliptic cell problem

$$[a \cdot \nabla_y + \frac{1}{2} \beta \beta^T : \nabla_y \nabla_y] r(x, y) = f_1(x, y), \quad (3.9)$$

with appropriate boundary conditions, e.g., periodic on a torus. The equation indeed has a unique solution; see [39] for further details.

²The assumptions holds for the case of ergodic rotations on a torus in which we are interested.

3.2. Periodic ODEs. We now formally set $\beta=0$ and consider a case in which for fixed x the dynamics of y is periodic. Hence, we can think of y as an fast oscillator. Assume that $a(x,y)$ is uniformly positive in $\mathbb{R}^d \times \mathbb{R}^n$, $\inf_{x \in \mathbb{R}^d, y \in \mathbb{R}^n} |a(x,y)| > 0$. Loosely speaking, this means that the system remains highly oscillatory at all times. Therefore, for fixed x trajectories are closed loops and the invariant set Σ_x is a one-dimensional manifold in \mathbb{R}^d that depends on the initial y and on x . The invariant measure $d\mu_x$, which is supported on Σ_x is absolutely continuous with respect to the Lebesgue measure on Σ_x . The ergodic assumption holds with respect to the same manifold and measure. A rigorous treatment of this problem is beyond the scope of this manuscript. For two-scale systems the reader is referred to [21, 22] and the references therein.

Instead, we continue formally trying to identify the effective equation for x_t . The main difference between the periodic and the stochastic cases can be seen from two complementary points of view. First, as a dynamical system, the fast process in the SDE (3.1) is mixing (for fixed x). This is no longer the case with periodic systems which are ergodic but not mixing. Second, as a homogenization problem, because $\beta\beta^T=0$ the cell problem (3.9) is no longer elliptic.

For fixed x , consider the periodic solution of

$$\frac{d\zeta}{dt} = a(x, \zeta(t)),$$

with a suitable initial condition on Σ_x . The period is denoted τ_x . $\zeta(t)$ transverses the exact periodic trajectory of $y(t)$ and with the correct invariant measure. Therefore, averages with respect to $d\mu_x$ can be written as the time average over a single period of ζ :

$$\langle h(x,y) \rangle_y = \frac{1}{\tau_x} \int_0^{\tau_x} h(x, \zeta(t)) dt = \frac{1}{\tau_x} \int_{\Sigma_x} h(x, \zeta(t)) \frac{1}{a(x, \zeta(t))} d\zeta.$$

Furthermore, recall the hierarchy of operators (3.3). Substituting $\beta=0$, \mathcal{L}_2 takes the form

$$\mathcal{L}_2 = a(x,y) \cdot \nabla_y.$$

We note that, for any $y \in \Sigma_x$, \mathcal{L}_2 is the directional derivative of ζ along the tangent direction to Σ_x and that $|\mathcal{L}_2\zeta|$ is inversely proportional to the density of $d\mu_x$:

$$\mathcal{L}_2 h(x,y) = a(x,y) \cdot \nabla_y h(x,y) = \left. \frac{d}{dt} h(x, \zeta(t)) \right|_{\zeta(t)=y}.$$

This implies that the inverse of \mathcal{L}_2 can be described in terms of integration with respect to time along the trajectory of $\zeta(t)$:

$$H(x,y) = \mathcal{L}_2^{-1} h(x,y) = \int_0^{t(y)} h(x, \zeta(s)) ds + C(x),$$

where $t(y)$ is the unique time in which $\zeta(t)=y$ within one period of ζ . Following the Fredholm alternative, we pick the unique inverse that is perpendicular to Null \mathcal{L}_2^* , i.e., we require

$$\langle H(x,y) \rangle_y = \frac{1}{\tau_x} \int_0^{\tau_x} H(x, \zeta(t)) dt = 0.$$

This fixes the constant

$$C(x) = -\frac{1}{\tau_x} \int_0^{\tau_x} \int_0^t H(x, \zeta(s)) ds dt.$$

In particular, we recognize that $\mathcal{L}_2^{-1}[f_1] = \phi$ as given by

$$\phi(x, y) = \int_0^y f_1(x, \tau) d\tau - \left\langle \int_0^y f_1(x, \tau) d\tau \right\rangle_y, \quad (3.10)$$

so that $\langle \phi \rangle_y = 0$. Substituting into (3.8) yields

$$\begin{aligned} F(x) &= \langle -f_1 \cdot \nabla_x \phi + f_0 \rangle_y, \\ B(x) &= \langle -f_1 \phi \rangle_y. \end{aligned} \quad (3.11)$$

In order to identify (3.11) with the averaged equation (2.22), we need a simple lemma.

LEMMA 3.1. *Let $h(s)$ denote an S -periodic function with zero average and let $H(s)$ be an anti-derivative of h , $H' = h$. Then,*

$$\int_0^S h(s) H(s) ds = 0. \quad (3.12)$$

Proof. First, we note that because $h(s)$ has zero average, $\int_0^S h(\tau) d\tau = 0$, its anti-derivative is also S -periodic:

$$H(S+s) = H(0) + \int_0^{S+s} h(\tau) d\tau = H(s).$$

Then, using integration by parts,

$$\int_0^S h(s) H(s) ds = [H^2(s)]_0^S - \int_0^S H(s) h(s) ds.$$

The first term on the right vanishes, which proves (3.12). \square

Because ϕ is the anti-derivative of f_1 , an immediate consequence is that $B(x) = -\langle f_1(x, y) \phi(x, y) \rangle_y = 0$. Hence, the effective dynamics of x_t are deterministic. More precisely, the variance of the stochastic perturbation is of order ϵ . Furthermore,

$$0 = \nabla_x \langle f_1(x, y) \phi(x, y) \rangle_y = \langle \nabla_x f_1(x, y) \phi(x, y) \rangle_y + \langle f_1(x, y) \nabla_x \phi(x, y) \rangle_y.$$

We conclude that the effective drift and diffusion coefficient can be written as

$$\begin{aligned} F(x) &= \langle (\nabla_x f_1) \phi + f_0 \rangle_y, \\ B(x) &= 0. \end{aligned} \quad (3.13)$$

Thus, we obtain the consistent form of the effective drift as proven by Theorem 1.1. Note that because the fast process in (3.1) contains only $\mathcal{O}(\epsilon^2)$ time scale, $\frac{\partial \phi}{\partial s_1} = 0$ in (2.21).

4. Iterated averaging with multiple kernels

The goal of this section is to generalize the framework of averaging kernels studied in [2, 3, 4, 5, 20] to include three or more time scales. In particular, we prove that averaging of three scale oscillatory problems can be approximated via convolution with respect to kernels with known support, differentiability properties and moments. Let $\mathbb{K}^{p,q}$ denote the space of normalized C^q functions, supported on $[-1, 1]$, that have p vanishing moments, i.e.,

$$\int_{[-1,1]} K(t)t^r dt = \begin{cases} 1, & r=0, \\ 0, & 1 \leq r \leq p. \end{cases} \tag{4.1}$$

Recall that for $\eta > 0$, $K_\eta(t)$ denotes a rescaling of K as $K_\eta(t) = \eta^{-1}K(\eta^{-1}t)$.

We will have an error estimate for approximating the double average $\langle f \rangle_{12}$ with two convolutions $K_{\eta_1} * [K_{\eta_2} * f]$. Applied to the RHS of (1.6), the convolution approximates $F(\xi)$ of Theorem 1.1. The proposition below establishes the accuracy of iterated averaging with multiple kernels under the scaling condition $\epsilon^2 \ll \eta_2 \ll \epsilon \ll \eta_1 \ll 1$.

PROPOSITION 4.1. *Let $K \in \mathbb{K}^{p,q}$ and $f_0, f_1 \in C^{\max\{q,p+1\}}$. Then, for $K \in \mathbb{K}^{p,q}$, there exists a constant $C > 0$ such that*

$$|K_{\eta_1} * \{K_{\eta_2} * (\epsilon^{-1}f_1 + f_0)(\cdot)\}(t) - F(\xi(t))| \leq C \left(\frac{\epsilon^{2q-1}}{\eta_2^q} + \frac{\epsilon^q}{\eta_1^q} + \frac{\eta_2^{p+1}}{\epsilon^{p+2}} + \eta_1^{p+1} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1 \tag{4.2}$$

for some constant $C > 0$, where F is given by (2.21) as in Theorem 1.1, $K^{(j)}(\cdot)$ denotes the j -th derivative of K , and $\|\cdot\|_1$ is the L_1 norm.

Throughout, C denotes a generic positive constant whose value may change between expressions.

4.1. Estimation of the effective force. Let $f: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ denote a scalar function of the vector argument $(x, s_1, s_2, \dots, s_n)$ and 1-periodic in s_1, \dots, s_n . To be consistent with previous notation, averaging with respect to s_k is respectively denoted by

$$\langle f \rangle_k = \int_0^1 f(x, \dots, s_k, \dots) ds_k, \quad k = 1, 2, \dots, n. \tag{4.3}$$

Motivated by the averaging techniques in [2], we approximate $\langle f \rangle_n$ using a kernel. First let us prove the following lemma.

LEMMA 4.1. *If $f(t, s_1, s_2, \dots, s_n) = a(t, s_1, s_2, \dots, s_{n-1})b(s_n)$, where $n \in \mathbb{N}$, $b(s_n + 1) = b(s_n)$, $\int_0^1 b(s_n) ds_n = 0$, $a \in C^q(\mathbb{R}^n)$, and $\max_{0 \leq j \leq q} \|a^{(j)}\|_\infty \leq M$, then for any $K \in \mathbb{K}^{p,q}$ and $\eta = \mathcal{O}(\epsilon^k)$, $k > n - 1$,*

$$|K_\eta * f(\cdot, \epsilon^{-1}\cdot, \epsilon^{-2}\cdot, \dots, \epsilon^{-n}\cdot)| \leq CM \left(\frac{\epsilon^n}{\eta} \right)^q \max_{j=0 \dots q} \|K^{(j)}\|_1.$$

Proof. Let $\tilde{K}_\eta(x, y) = K_\eta(x - y)a(y, \epsilon^{-1}y, \epsilon^{-2}y, \dots, \epsilon^{-n+1}y)$.

$$|K_\eta * f| = \left| \int \tilde{K}_\eta(t, s)b(\epsilon^{-n}s) ds \right| \leq \epsilon^{nq} \int \left| \partial_y^q \tilde{K}_\eta(t, s)b^{[q]}(\epsilon^{-n}s) \right| ds,$$

where $b^{[j]}(t) = \int_0^t b^{[j-1]}(s)ds - \int_0^1 \int_0^t b^{[j-1]}(s)dsdt$ and $\|b^{[j]}(t)\|_\infty \leq \|b\|_\infty$. For the kernel part,

$$\begin{aligned} & \int \left| \partial_y^q \tilde{K}_\eta(t, s) \right| ds \\ &= \int \left| \sum_{k_1+k_2+\dots+k_{n+1}=q} \binom{q}{k_1, k_2, \dots, k_{n+1}} \cdot (-\eta^{-1})^{k_1} \cdot K_\eta^{(k_1)}(t-s) \right. \\ & \quad \left. \cdot \prod_{i=1}^n \frac{1}{\epsilon^{k_{i+1}(i-1)}} \partial_i^{k_{i+1}} a(s, \epsilon^{-1}s, \epsilon^{-2}s, \dots, \epsilon^{-n+1}s) \right| ds \\ &\leq \frac{C}{\eta^q} \max_{j=0\dots q} \|a^{(j)}\|_\infty \max_{j=0\dots q} \|K^{(j)}\|_1, \end{aligned}$$

where $\binom{q}{k_1, k_2, \dots, k_{n+1}}$ is the multinomial coefficient and defined by $\frac{q!}{k_1!k_2!\dots k_{n+1}!}$. \square

LEMMA 4.2. For $r = (r_1, r_2, \dots, r_n) \in \mathbb{Z}_+^n$ an ordered n -tuple of nonnegative integers, assume $\partial^r f(t, s_1, \dots, s_n)$ is continuous and bounded by C_f for $r = 0, \dots, \sigma$, and $\sigma \geq 1$.³ Then, for any $K \in \mathbb{K}^{p,q}$ and $\eta = \mathcal{O}(\epsilon^k)$ with $k > n-1$, there exists $C > 0$ such that

$$|K_\eta * (f(\cdot, \epsilon^{-1}\cdot, \epsilon^{-2}\cdot, \dots, \epsilon^{-n}\cdot) - \langle f \rangle_n)| \leq C \left(\frac{\epsilon^{nq}}{\eta^q} + \frac{\eta^\sigma}{\epsilon^{(n-1)\sigma}} \right) \max_{j=0\dots q} \|K^{(j)}\|_1. \quad (4.4)$$

Proof. Let

$$g(t, s_1, s_2, \dots, s_n) = f(t, s_1, s_2, \dots, s_n) - \langle f \rangle_n.$$

g is 1-periodic with respect to s_i and $\partial^r g(t, s_1, s_2, \dots, s_n)$ are continuous and bounded for $r \in \mathbb{Z}_+^n$, $|r| = 0, \dots, \sigma$. In considering

$$K_\eta * g = \int_{t-\eta}^{t+\eta} K_\eta(t-s_n) \cdot g(s_n, \epsilon^{-1}s_n, \epsilon^{-2}s_n, \dots, \epsilon^{-n}s_n) ds_n,$$

we expand $g(s_n, \epsilon^{-1}s_n, \epsilon^{-2}s_n, \dots, \epsilon^{-(n-1)}s_n, t)$ around $s_n = 0$ while keeping t fixed. We denote by $\partial_{x_i} g$ the partial derivative with respect to the i -th component of g .

$$\begin{aligned} & g(s_n, \epsilon^{-1}s_n, \epsilon^{-2}s_n, \dots, \epsilon^{-n}s_n) \\ &= \sum_{j=0}^{\sigma-1} \frac{1}{j!} \left[\sum_{i=1}^n \frac{s_n}{\epsilon^{i-1}} \partial_{x_i} \right]^j g(0, \dots, 0, \epsilon^{-n}s_n) \\ & \quad + \frac{1}{\sigma!} \left[\sum_{i=1}^n \frac{s_n}{\epsilon^{i-1}} \partial_{x_i} \right]^\sigma g(\mu_1, \epsilon^{-1}\mu_2, \dots, \epsilon^{-n+1}\mu_n, \epsilon^{-n}s_n) \end{aligned}$$

where $(\mu_1, \mu_2, \dots, \mu_n)$ is in the open line segment joining $\bar{0}$ and $(s_n, \epsilon^{-1}s_n, \dots, \epsilon^{-n+1}s_n)$ in \mathbb{R}^n . The key idea consists of in writing the expansion as a sum of two parts. Without loss of generality, we set $t=0$ and write $K_\eta * g$ as $I_1 + I_2$, where

$$I_1 = \sum_{j=0}^{\sigma-1} \int_{-\eta}^{\eta} K_\eta(-s_n) \cdot \frac{1}{j!} \left[\sum_{i=1}^n \frac{s_n}{\epsilon^{i-1}} \partial_{x_i} \right]^j g(0, \dots, 0, \epsilon^{-n}s_n) ds_n,$$

³ $|r| = r_1 + r_2 + \dots + r_n$ and $\partial^r = (\partial_{x_1})^{r_1} (\partial_{x_2})^{r_2} \dots (\partial_{x_n})^{r_n}$.

$$I_2 = \int_{-\eta}^{\eta} K_{\eta}(-s_n) \cdot \frac{1}{\sigma!} \left[\sum_{i=1}^n \frac{s_n}{\epsilon^{i-1}} \partial_{x_i} \right]^{\sigma} g(\mu_1, \epsilon^{-1} \mu_2, \dots, \epsilon^{-n+1} \mu_n, \epsilon^{-n} s_n) ds_n.$$

By using Lemma 4.1, I_1 is estimated by

$$|I_1| \leq C \left(\frac{\epsilon^n}{\eta} \right)^q \cdot \max_{j=0 \dots q} \|K^{(j)}\|_1 \cdot \sum_{j=0}^{\sigma-1} \frac{1}{j!} \left[\sum_{i=1}^n \frac{\eta}{\epsilon^{i-1}} \right]^j.$$

Finding the leading order term in the summation, I_2 is estimated by

$$|I_2| \leq C \left(\frac{\eta}{\epsilon^{n-1}} \right)^{\sigma} \sup_{t, s_1, s_2, \dots, s_n} \sup_{|r|=\sigma} |\partial^r f(t, s_1, \dots, s_n)| \cdot \|K\|_1$$

Putting these estimates together, we find that there exists a positive constant C such that

$$|K_{\eta} * g| \leq C \left(\frac{\epsilon^{nq}}{\eta^q} + \frac{\eta^{\sigma}}{\epsilon^{(n-1)\sigma}} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1.$$

□

We now compare the iterated averaging of (1.6) with the averaged equation (1.9). Before we move on to the next step, we simplify our notation of the forces by writing $f(t, \epsilon^{-1}t, \epsilon^{-2}t)$ instead of $f(x(t), \epsilon^{-1}t, \epsilon^{-2}t)$. This is possible because, after solving (1.6) at the 1st tier, $x(t)$ is known up to a prescribed accuracy Δ . Theorem 4.3 shows that the error between x and x_1 (solution of the 1st tier) is $\mathcal{O}(\Delta)$.

THEOREM 4.3. *Let $x_1(t)$ denote an approximation of $x(t)$ in the 1st tier using a two-scale HMM in the time interval $t \in [t_n - \eta_1, t_n + \eta_1]$. Given $0 < \epsilon < \epsilon_0$ and a prescribed accuracy Δ , there exists $C > 0$ such that*

$$\sup_{t \in [t_n - \eta_1, t_n + \eta_1]} |x(t) - x_1(t)| \leq C \Delta. \tag{4.5}$$

Proof. By considering the 1st and 2nd tiers as the two-scale HMM solver, we generalize the error analysis discussed in [3, 20]. We denote the order of accuracy, step size and length of integration in i -th tier by m_i , h_i , and η_i respectively.

At the 1st tier, the global accuracy of integrating the original ODE (1.6) to time $\eta_1 (\ll \epsilon)$ is given by

$$C \max \left\{ \frac{\eta_1 h_1^{m_1}}{\epsilon^{m_1}}, \frac{\eta_1 \eta_2 h_2^{m_2}}{h_1 \epsilon^{2m_2+1}}, \frac{\eta_1 \epsilon^{2q}}{h_1 \eta_2^{q+1}} \right\}$$

for some $C > 0$. The errors from each evaluation at the 2nd tier accumulate by taking $h_1^{-1} \eta_1$ steps. Thus we can balance the required accuracy Δ with different sources of errors. Note that the maximal possible accuracy is $\Delta = \epsilon^2$ because this error is introduced by simulating the averaged equation instead of the original equation. □

Next, by iterated use of Lemma 4.2, we show that $\langle f(t, s_1, s_2) \rangle_{12}$ is well approximated by $K_{\eta_1} * [K_{\eta_2} * f](t)$.

LEMMA 4.4. *Let $f_0(t, s_1, s_2)$ be 1-periodic in s_1 and s_2 . For $r = (r_1, r_2) \in \mathbb{Z}_+^2$, assume that $\partial^r f_0(t, s_1, s_2)$ is continuous and bounded for $|r| = 0, \dots, \sigma$, and $\sigma \geq 1$. Then, for*

any $K \in \mathbb{K}^{p,q}$ and suitable choice of η_1, η_2 such that $\epsilon^2 \ll \eta_2 \ll \epsilon \ll \eta_1 \ll 1$, there exists a constant $C > 0$ such that

$$\begin{aligned} & |K_{\eta_1} * [K_{\eta_2} * (f_0(\cdot, \epsilon^{-1}\cdot, \epsilon^{-2}\cdot) - \langle f_0(t, s_1, s_2) \rangle_{12}) (\cdot)] (t)| \\ & \leq C \left(\frac{\epsilon^{2q}}{\eta_2^q} + \frac{\epsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\epsilon^\sigma} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \end{aligned}$$

Proof. Let

$$g(t, s_1, s_2) = f_0(t, s_1, s_2) - \langle f_0 \rangle_{12}(t). \quad (4.6)$$

Note that $\langle g \rangle_{12} = \int_0^1 \int_0^1 (f_0(t, s_1, s_2) - \langle f_0 \rangle_{12}) ds_2 ds_1 = 0$ and that $g(x, s_1, s_2)$ is 1-periodic with respect to s_1, s_2 and $\partial^r g(x, s_1, s_2)$ is continuous and bounded for $r \in \mathbb{Z}_+^2, |r| = 0, \dots, \sigma$. Iterated convolution with two kernels yields

$$\begin{aligned} & K_{\eta_1} * [(K_{\eta_2} * g(\cdot, \epsilon^{-1}\cdot, \epsilon^{-2}\cdot)) (\cdot)] (t) \\ & = \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) \underbrace{\left[\int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1-s_2) \cdot g(s_2, \epsilon^{-1}s_2, \epsilon^{-2}s_2) ds_2 \right]}_{=I_1(s_1)} ds_1. \end{aligned}$$

We iterate the argument in Lemma 4.2. First, there exists $C > 0$ such that

$$|I_1(s_1) - \langle g \rangle_2(s_1, \epsilon^{-1}s_1)| \leq C \left(\frac{\epsilon^{2q}}{\eta_2^q} + \frac{\eta_2^\sigma}{\epsilon^\sigma} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \quad (4.7)$$

Recalling that $\langle g \rangle_{12} = 0$,

$$I_1 = I_1 - \langle g \rangle_2(s_1, \epsilon^{-1}s_1) + \langle g \rangle_2(s_1, \epsilon^{-1}s_1) - \langle g \rangle_{12}. \quad (4.8)$$

Thus

$$\begin{aligned} & \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) \cdot I_1 ds_1 \right| \\ & = \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) (I_1 - \langle g \rangle_2(s_1, \epsilon^{-1}s_1) + \langle g \rangle_2(s_1, \epsilon^{-1}s_1) - \langle g \rangle_{12}) ds_1 \right| \\ & \leq C \left(\frac{\epsilon^{2q}}{\eta_2^q} + \frac{\eta_2^\sigma}{\epsilon^\sigma} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1 + \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) (\langle g \rangle_2(s_1, \epsilon^{-1}s_1) - \langle g \rangle_{12}) ds_1 \right|. \end{aligned} \quad (4.9)$$

Second, define $\hat{g}(\hat{s}_1, s_1) = \langle g \rangle_2(\hat{s}_1, \epsilon^{-1}s_1) - \langle g \rangle_{12}$. Hence, $\hat{g}(\hat{s}_1, s_1)$ is 1-periodic in the second variable and the average over s_1 is zero. A second application of Lemma 4.2 yields existence of $C_2 > 0$ such that

$$\left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) \hat{g}(\hat{s}_1, s_1) ds_1 \right| \leq C_2 \left(\frac{\epsilon^q}{\eta_1^q} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \quad (4.10)$$

Hence, we can find a positive constant C such that

$$\begin{aligned} & |K_{\eta_1} * [K_{\eta_2} * (f_0(\cdot, \epsilon^{-1}\cdot, \epsilon^{-2}\cdot) - \langle f_0(t, s_1, s_2) \rangle_{12}) (\cdot)] (t)| \\ & \leq C \left(\frac{\epsilon^{2q}}{\eta_2^q} + \frac{\epsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\epsilon^\sigma} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \end{aligned}$$

This concludes the proof of the theorem. \square

Recall the three scale problem (1.6) and its averaged equation (1.9):

$$\begin{aligned} x' &= \epsilon^{-1} f_1(x, \epsilon^{-1}t, \epsilon^{-2}t) + f_0(x, \epsilon^{-1}t, \epsilon^{-2}t), \quad x(0) = x_0, \\ \xi' &= \langle (\nabla_x f_1) \phi(\xi, s_1, s_2) \rangle_{12} + \langle f_0(\xi, s_1, s_2) \rangle_{12}, \quad \xi(0) = x_0. \end{aligned}$$

Key to the following theorem is the vanishing of $\langle f_1(x(t), s_1, s_2) \rangle_2$.

THEOREM 4.5. *Let $f_1(x(t), s_1, s_2)$ be 1-periodic in s_1 and s_2 , and have a zero average with respect to s_2 . For $r = (r_1, r_2) \in \mathbb{Z}_+^2$, assume $\partial^r f_1(x(t), s_1, s_2)$ are continuous and bounded for $|r| = 0, \dots, \sigma + 1$, and $\sigma \geq 1$. Then, for any $K \in \mathbb{K}^{p,q}$ and $\epsilon^2 \ll \eta_2 \ll \epsilon \ll \eta_1 \ll 1$, the following estimate holds:*

$$\begin{aligned} &|K_{\eta_1} * [K_{\eta_2} * (\epsilon^{-1} f_1(x(\cdot), \epsilon^{-1}\cdot, \epsilon^{-2}\cdot) - \langle h \rangle_{12}(\xi))(\cdot)](t)| \\ &\leq C \left(\frac{\epsilon^{2q-1}}{\eta_{2q}} + \frac{\epsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\epsilon^{\sigma+1}} + \eta_1^{\sigma-1} \right) \max_{j=0\dots q} \|K^{(j)}\|_1, \end{aligned}$$

where

$$\begin{aligned} h(\xi, s_1, s_2) &= (\nabla_x f_1) \phi(\xi, s_1, s_2), \\ \phi(\xi, s_1, s_2) &= \int_0^{s_2} f_1(\xi, s_1, \tau) d\tau. \end{aligned}$$

Proof. We begin with the first convolution $K_{\eta_2} * (\epsilon^{-1} f_1 - \langle h \rangle_{12})$. Lemma 2.1 allows ones to write $x(t)$ as

$$x(t) = w(t) + \epsilon \psi(t) + \epsilon \phi(t), \tag{4.11}$$

where $\psi(t) = \psi(w(t), \epsilon^{-1}t, \epsilon^{-2}t)$, $\phi(t) = \phi(w(t), \epsilon^{-1}t, \epsilon^{-2}t)$, and $\psi(t)$ is bounded independent of ϵ .

$$\begin{aligned} &K_{\eta_2} * (\epsilon^{-1} f_1 - \langle h \rangle_{12})(s_1) \\ &= \int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1 - s_2) \cdot (\epsilon^{-1} f_1(w(s_2) + \epsilon \psi(s_2) + \epsilon \phi(s_2), \epsilon^{-1}s_2, \epsilon^{-2}s_2) - \langle h \rangle_{12}) ds_2 \\ &= I_1 + I_2 + \mathcal{O}(\epsilon), \end{aligned}$$

where we set

$$\begin{aligned} I_1 &= \int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1 - s_2) \cdot (\epsilon^{-1} f_1(w(s_2) + \epsilon \psi(s_2), \epsilon^{-1}s_2, \epsilon^{-2}s_2)) ds_2, \\ I_2 &= \int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1 - s_2) \cdot (\nabla_x f_1(w(s_2) + \epsilon \psi(s_2), \epsilon^{-1}s_2, \epsilon^{-2}s_2) \phi(s_2) - \langle h \rangle_{12}) ds_2. \end{aligned} \tag{4.12}$$

Each term in the integrations is estimated in a similar way. Before we move on, recall from Theorem 4.3 that we identify $f_1(x(s_2), \epsilon^{-1}s_2, \epsilon^{-2}s_2)$ with $f_1(s_2, \epsilon^{-1}s_2, \epsilon^{-2}s_2)$. In (4.12), this simplification is also allowed because by solving the 1st and 2nd tier, we know $x(t)$ over $[t - \eta_1, t + \eta_1]$, and thus $w(t) + \epsilon \psi(t)$ as well. First, applying Lemma 4.2 to I_1 with $\langle f_1 \rangle_2 = 0$ yields

$$|I_1| \leq \left(\frac{\epsilon^{2q-1}}{\eta_2^q} + \frac{\eta_2^\sigma}{\epsilon^{\sigma+1}} \right) \max_{j=0\dots q} \|K^{(j)}\|_1. \tag{4.13}$$

Second, for I_2 , we now return to consider convolving with K_{η_1} and K_{η_2} . Note that $\langle h \rangle_{12}$ is a function of $\xi(\cdot)$. Theorem 1.1 yields that

$$\sup_{0 \leq t \leq T} |w(t) - \xi(t)| \leq \sup_{0 \leq t \leq T} (|w(t) - x(t)| + |x(t) - \xi(t)|) \leq C\epsilon.$$

Then we have $\langle (\nabla_x f_1)\phi - \langle h \rangle_{12} \rangle_{12} = 0 + \mathcal{O}(\epsilon)$, and this allows us to use Lemma 4.4 and thus to get an estimate for $|K_{\eta_1} * I_2(\cdot)|$:

$$|K_{\eta_1} * I_2(\cdot)| \leq C \left(\frac{\epsilon^{2q}}{\eta_2^q} + \frac{\epsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\epsilon^\sigma} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \tag{4.14}$$

Because we differentiate f_1 with respect to x , $\partial^r \nabla_x f_1 \cdot \phi$ are continuous and bounded for $|r|=0, \dots, \sigma$. Putting estimates (4.13) and (4.14) together, Theorem 4.5 follows. \square

We conclude this section by proving Proposition 4.1, which is the cornerstone of our numerical method.

Proof. (Proposition 4.1) The RHS of the estimate in Theorem 4.5 dominates that of Lemma 4.4. Therefore, having $p (< \sigma)$ vanishing moments yields (4.2). \square

REMARK 4.2. Theorem 1.1 is only valid up to times T independent of ϵ . However, in special cases in which additional cancellation or self averaging occurs, iterated averaging with kernels may give an consistent approximation for the effective behavior of ODEs for longer time intervals.

5. Generalizations

This paper is focused on three scale problems modeled by

$$x' = \epsilon^{-1} f_1(x, \epsilon^{-1}t, \epsilon^{-2}t) + f_0(x, \epsilon^{-1}t, \epsilon^{-2}t), \quad x(0) = x_0,$$

restricting $\langle f_1 \rangle_2 = 0$. However, the discussions at the preceding sections suggest several possible generalizations.

5.1. Almost-periodic dynamics. The three-scale averaging theorem can be generalized to include dynamics in which the fast $\mathcal{O}(\epsilon^2)$ or $\mathcal{O}(\epsilon)$ time scales are not necessarily periodic but rather ergodic on a torus. The periodicity of s_2 is only taken into account when evaluating the remainder term $r(t; \epsilon)$ in (2.18). In the case of a torus, $r(t; \epsilon)$ can be written as a finite sum of periodic functions whose periods are incommensurate. Thus, estimate (2.19) still holds. A similar generalization can be obtained for almost-periodic functions whose spectrum is bounded away from zero; see, for example, [13, 40].

5.2. The 3-tier HMM using slow charts. In the proof of Theorem 1.1, as well as while applying the expansion formalism, it was necessary to assume that the average of the singular $\mathcal{O}(\epsilon^{-1})$ part in x' vanishes, $\langle f_1 \rangle_2 = 0$. Nonetheless, it can be shown that the estimate of Section 2.1 does hold, but only on a short time segment of $\mathcal{O}(\epsilon)$ length, i.e.

$$\sup_{0 \leq t \leq \epsilon T} |x(t) - \xi(t)| = \mathcal{O}(\epsilon).$$

In addition, the HMM procedure which utilizes the iterated averaging estimate indicates that in fact, in order to generate a consistent approximation of a slow variable, one only needs to evaluate its derivative on a short time segment of order ϵ . Then, if

the dynamics on the intermediate $\mathcal{O}(\epsilon)$ is again oscillatory (i.e. periodic or as above), additional averaging on the ϵ time scale may average this divergence out. This self averaging property can be captured by the iterated averaging procedure.

To this end, we first need to define slow variables. Formally, slow variables of a dynamical system involving three or more time scales are defined as below [6].

DEFINITION 5.1. A smooth time dependent function $\alpha: [0, T] \mapsto \mathbb{R}$ is said to evolve on the ϵ^k time scale in $[0, T]$ for some integer k and for $0 < \epsilon \leq \epsilon_0$, if there exists a smooth function $\beta: [0, T] \mapsto \mathbb{R}$ and constants C_0 and C_1 such that

$$\sup_{t \in [0, T]} \left| \frac{d}{dt} \beta(t) \right| \leq C_0 \epsilon^{-k} \quad \text{and} \quad \sup_{t \in [0, T]} |\alpha(t) - \beta(t)| \leq C_1 \epsilon.$$

DEFINITION 5.2. A function $\xi(x)$ is said to evolve on the ϵ^k time scale along the trajectories of (1.6) in $[0, T]$ if the time dependent function $\xi(x(t; \epsilon, x_0))$ evolves on the ϵ^k time scale in $[0, T]$. For brevity, we will refer to variables that evolve on the ϵ^0 time scale as slow.

The 3-tier HMM using slow variables shares a similar strategy described in Section 1.3, which implements a recursive two-level solver, but the fast oscillations we need to average over are not explicitly given. As [3], with the need for identifying hidden slow variables, we approximate an averaged equation for slow variables by time averaging the microscopic evolution using a suitable kernel. Recall that because the time scales $\mathcal{O}(\epsilon^2)$, $\mathcal{O}(\epsilon^1)$, and $\mathcal{O}(1)$ are considered, we need to identify three sets of variables which evolve on each time scale respectively.

Suppose we obtain such a coordinate system using, e.g., the method described in [6]. We denote this system of coordinates $\xi = (\xi_0, \xi_1, \xi_2)$, where $\xi_i = (\xi_i^1, \dots, \xi_i^{d_i})$ are the variables evolving on the ϵ^i time scale, and $\sum_{i=0}^2 d_i = d$. One should take the coordinates of ξ as slow as possible, i.e., if ϕ evolves on both ϵ^0 and ϵ^1 time scales, then $\phi \in \xi_0$. In terms of the new coordinates the ODE system takes the form

$$\begin{aligned} \xi_0' &= \epsilon^{-1} f_1(\xi_0, \xi_1, \xi_2) + f_0(\xi_0, \xi_1, \xi_2), \quad \langle f_1 \rangle_2 = 0, \\ \xi_1' &= \epsilon^{-1} g_1(\xi_0, \xi_1, \xi_2) + g_0(\xi_0, \xi_1, \xi_2), \quad \langle g_1 \rangle_2 > C_1 > 0, \\ \xi_2' &= \epsilon^{-2} h_2(\xi_0, \xi_1, \xi_2), \end{aligned} \tag{5.1}$$

where $\langle \cdot \rangle_2$ denotes averaging with respect to the invariant measure for ξ_2 on fixed ξ_0 and ξ_1 and C_1 is independent of ϵ , i.e., $\langle g_1 \rangle_2$ is bounded away from 0 independent of ϵ . Note that (5.1) is of a form similar to (1.5). We assume that no resonances, passage through resonances, or turning points exist as these may cause hidden slow variables and the decomposition of states into time scales may not be trivial, as discussed in [6, 32].

Then outline of the 3-tier HMM with slow variables is as follows. For simplicity of notation, we suppress the superscript in ξ_i . As before, we denote the discretized time $t_{n,m} = nh_0 + mh_1$ and $t_n = nh_0$. We concentrate on the forward Euler and a symmetric kernel.

1. Determination of slow variables:

Find a coordinate system $\xi(x) = (\xi_0(x), \xi_1(x), \xi_2(x))$ where ξ_i are the variables evolving on the ϵ^i time scale; see [6] for details. Set $n = 0$.

2. Multilevel evolution:

- (0th tier) At $t = t_n$, set $\tilde{X}_0 = x_{1,0} = x_n$.

- (a) (1st tier) For $m=0$ to k ($=\eta_1/h_1$),
 set $t=t_{n,m}=nh_0+mh_1$,
 i. (2nd tier) Solve the full ODE (1.6) in $t \in [t_{n,m}-\eta_2, t_{n,m}+\eta_2]$
 with initial conditions \tilde{X}_0 .
 ii. (2nd tier) Force estimation in $\mathcal{O}(\epsilon^1)$ scale: approximate $\xi'_i(t_{n,m})$,
 $i=0,1$ by

$$\langle \xi'_i \rangle_{\eta_2}(t_{n,m}) = (K_{\eta_2} * \xi'_i)(t_{n,m}). \quad (5.2)$$

- (b) (1st tier) $x_{1,m+1} = x_{1,m} + h_1 \delta x_1$, where δx_1 is the least squares solution to the linear system

$$\delta x_1 \cdot \nabla \xi_i = \langle \xi'_i \rangle_{\eta_2}, \quad (5.3)$$

for all $i=0,1$. Redefine $\tilde{X}_0 = x_{1,m+1}$.

- (c) End FOR

- (d) (1st tier) Force estimation in $\mathcal{O}(1)$ scale: approximate $\xi'_0(t_n)$ by

$$\langle \xi'_0 \rangle_{\eta_1}(t_n) = K_{\eta_1} * (\langle \xi'_0 \rangle_{\eta_2}(\cdot))(t_n). \quad (5.4)$$

- (0th tier) $x_{n+1} = x_n + h_0 \delta x_0$, where δx_0 is the least squares solution to the linear system

$$\delta x_0 \cdot \nabla \xi_0 = \langle \xi'_0 \rangle_{\eta_1}. \quad (5.5)$$

Set $\tilde{X}_0 = x_{n+1}$.

3. $n=n+1$ and repeat 2.

A rigorous proof of Section 5.2 is beyond the scope of the current paper. However, we refer the reader to [2, 3] for designing multiscale algorithms that compute the effective behavior of two-scale highly oscillatory dynamical systems by using slow variables. A related example is presented in Section 6.2.

5.3. Stochastic effects. The theory of asymptotic expansions and the nested-HMM integrators approach can be extended to a setting in which the intermediate scale is stochastic. In fact, the only requirements for applying the numerical method 1.3 is that the fastest $\mathcal{O}(\epsilon^2)$ scale is oscillatory - thus ensuring that the effective slow scale is deterministic, and that the intermediate $\mathcal{O}(\epsilon)$ time scale is ergodic. Consider (5.1). If the dynamics of ξ_1 is stochastic, then, on the $\mathcal{O}(1)$ time scale averaging with kernels needs to be replaced by an alternative method such as stochastic HMM [17, 44]. Such an example is presented in Section 6.3.

6. Numerical examples

In this section, we numerically apply the iterated HMM approach described in Section 1.3 to deterministic and stochastic systems with three scales. A basic example which has the form of (1.6) is studied in Section 6.1. Following that, we concentrate on examples which demonstrate the applicability of our method to the generalizations discussed in the previous section. The classical example of two coupled harmonic oscillators in resonance is illustrated in Section 6.2. In this example, one of the slow variables has formally unbounded derivatives as $\epsilon \rightarrow 0$, but it evolves on the ϵ^0 time scale due to a zero-average of ϵ^{-1} term. Section 6.3 is a deterministic example for the generalization discussed in Section 5. The period of the fast oscillator on the $\mathcal{O}(\epsilon^2)$

time scale changes according to the ϵ scale variable. Lastly, an interesting stochastic system whose period of the fastest oscillator changes randomly on the ϵ scale is given in Section 6.4.

Our multiscale algorithm is constructed as a family of multilevel (>2) solvers which resolve the different time scales and use kernels to estimate the effective force of the slower scales. In sections 6.1 through 6.3, we use a symmetric C^∞ kernel. We see that for a smooth kernel the computational cost is independent of ϵ ; see [6] for discussions about accuracy and efficiency. In the stochastic example 6.4, a $\mathbb{K}^{2,7}$ kernel is used.

6.1. Example 1. We begin with a simple example of a three-scale system, which is similar to (1.12)

$$x' = \epsilon^{-1} [x \sin(\epsilon^{-2}t) + \cos(\epsilon^{-2}t) \sin^2(\epsilon^{-1}t)] + \cos^2(x) + \cos(\epsilon^{-1}t), \quad x(0) = 1. \quad (6.1)$$

Applying Theorem 1.1 to (6.1), we have an averaged equation for $x(t)$,

$$\bar{x}' = 1/4 + \cos^2(\bar{x}), \quad \bar{x}(0) = 1. \quad (6.2)$$

The different three-time scales of (6.1) are illustrated in figures 6.1 and 6.2. The solution $x(t)$ undergoes small-amplitude fast oscillations around the slow trajectory over the interval $[0,10]$. As proved in sections 2 and 4, the 3-tier HMM approximates $\bar{x}(t)$, which remains close (of order ϵ) to the slow trajectory of $x(t)$. We apply an exponential kernel $K^{\text{exp}} \in \mathbb{K}^{1,\infty}([-1,1])$; see [3, 4, 20] for details. In figure 6.2, we compare the results of 3-tier HMM with both $\bar{x}(t)$ and $x(t)$ obtained by the explicit Runge-Kutta 4th order method. HMM is about 12 times faster than RK4 applied to (6.1) directly with the step size $h = \epsilon^2/5$. The computational effort of HMM is independent of ϵ once η_i and h_i are fixed. However, for classical numerical methods moving from $\epsilon = 10^{-3}$ to $\epsilon = 10^{-4}$ multiplies the computational effort by 100.

$\epsilon = 10^{-3}$	η_i	h_i	Method	Kernel
2nd tier	$18\epsilon^2$	$\epsilon^2/5$	RK4	$K^{\text{exp}} \in \mathbb{K}^{1,\infty}([-1,1])$
1st tier	18ϵ	$\epsilon^1/5$	RK2	$K^{\text{exp}} \in \mathbb{K}^{1,\infty}([-1,1])$
0th tier	10	$1/3$	RK2	-

TABLE 6.1. (Section 6.1) Parameters for the 3-tier HMM of example 1.

6.2. Example 2. Consider the following system describing two coupled harmonic oscillators in resonance [6]:

$$\begin{cases} x_1' &= -\epsilon^{-2}y_1 + \epsilon^{-1}y_2^2 - 3x_1x_2^2, \\ y_1' &= \epsilon^{-2}x_1 + y_1/2, \\ x_2' &= -(\epsilon^{-2} + \epsilon^{-1})y_2 - x_2, \\ y_2' &= (\epsilon^{-2} + \epsilon^{-1})x_2 - y_2 + 2x_1^2y_2, \end{cases} \quad (6.3)$$

with initial conditions $(x_1(0), y_1(0), x_2(0), y_2(0)) = (0, 1, 0, 1)$. As depicted in figure 6.3, all four state variables oscillate with a period which is of the order of ϵ^2 . Hence, x_1 , y_1 , x_2 , and y_2 evolve on the ϵ^2 time scale.

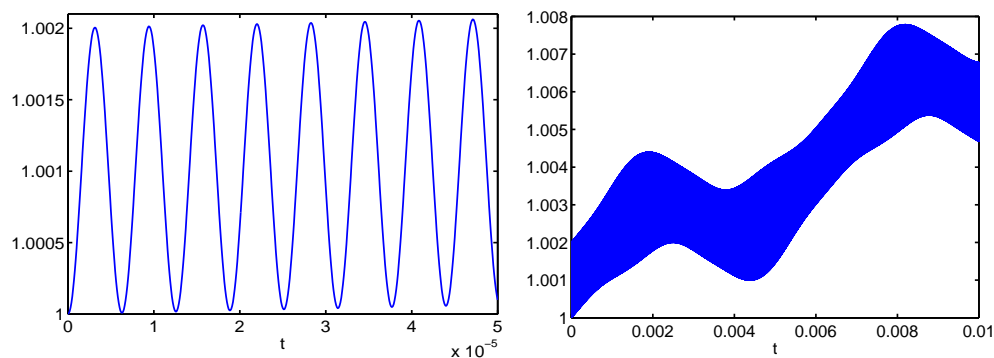


FIG. 6.1. (Section 6.1) The dynamics of (6.1) on the (Left) ϵ^2 time scale and (Right) ϵ^1 time scale ($\epsilon = 10^{-3}$). Plots are obtained by RK_4 with $h = \epsilon^2/100$.

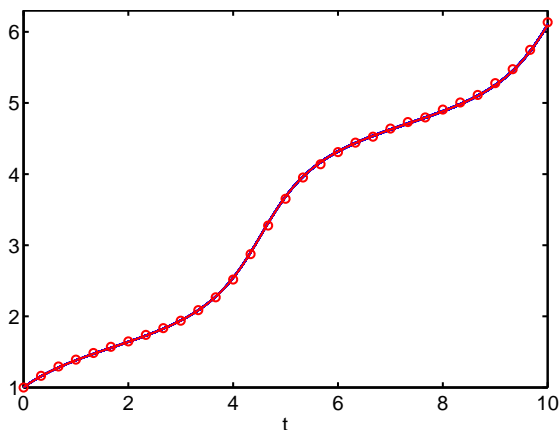


FIG. 6.2. (Section 6.1) The dynamics of (6.1) on the ϵ^0 time scale. $x(t)$ and $\bar{x}(t)$ are represented by a full line where both are almost indistinguishable. The results of 3-tier HMM are indicated by circles.

In order to find a slow coordinate system, we change to polar coordinates $(x_i, y_t) \mapsto (I_i, \varphi_i)$, $i = 1, 2$ and introduce a polynomial variable θ that describes the 1:1 resonance between the oscillators:

$$\begin{aligned} I_1 &= x_1^2 + y_1^2, \\ I_2 &= x_2^2 + y_2^2, \\ \theta &= x_1 x_2 + y_1 y_2, \\ \cos \varphi_1 &= x_1 / \sqrt{I_1}. \end{aligned}$$

The corresponding time derivatives are

$$\begin{aligned} I_1' &= 2\epsilon^{-1} x_1 y_2^2 - 6x_1^2 x_2^2 + y_1^2, \\ I_2' &= -2I_2 + 4x_1^2 y_2^2, \\ \theta' &= \epsilon^{-1} (x_2 y_2^2 + y_1 x_2 - x_1 y_2) + (-0.5 y_1 y_2 - x_1 x_2 - 3x_1 x_2^3 + 2x_1^2 y_1 y_2), \\ \varphi_1' &= \epsilon^{-2}. \end{aligned}$$

It appears as if $(I_1, I_2, \theta, \varphi_1)$ is a chart in which φ_1 evolves on the ϵ^2 time scale, I_1 and θ evolve on the ϵ time scale, while I_2 , which has a bounded derivative, evolves on the $\mathcal{O}(1)$ scale. The dynamics of the three slow variables I_1 , I_2 , and θ on the $\mathcal{O}(\epsilon)$ scale is depicted on the right in figure 6.3. The figure suggests that both I_1 and I_2 are practically constant on the ϵ scale. Indeed, it can be shown that the average of $x_1 y_2^2$ on any segment of length $\mathcal{O}(\epsilon)$ and larger is of order ϵ^2 . Therefore, the ϵ^{-1} term in I_1' has a zero average. As a result, the averaged I_1' is bounded independent of ϵ and I_1 evolves on the $\mathcal{O}(1)$ time scale, rather than the expected $\mathcal{O}(\epsilon)$.

The time evolution of I_1 and I_2 on the slowest $\mathcal{O}(1)$ time scale is depicted in figure 6.4. In addition, the figure shows the results of the 3-tier HMM integrator described in Section 1.3. We refer to the solver integrating the ϵ^i scale as the i -th tier. The step-size and length of integration of the i -th tier are denoted h_i, η_i , respectively. The HMM algorithm approximates the slow $\mathcal{O}(1)$ dynamics using macroscopic steps which are independent of ϵ . The integration is done using a fourth order method (in the macroscopic step size) and its efficiency is essentially independent of ϵ . Simulation parameters are detailed in table 6.2.

$\epsilon = 10^{-3}$	η_i	h_i	Method	Kernel
2nd tier	$70.1\epsilon^2$	$\epsilon^2/10$	RK4	$K^{exp} \in \mathbb{K}^{1,\infty}([-1,1])$
1st tier	$70.1\epsilon^1$	$\epsilon^1/10$	RK4	$K^{exp} \in \mathbb{K}^{1,\infty}([-1,1])$
0th tier	10	1/3	RK4	-

TABLE 6.2. (Section 6.2) Parameters for the 3-tier HMM of Example 2.

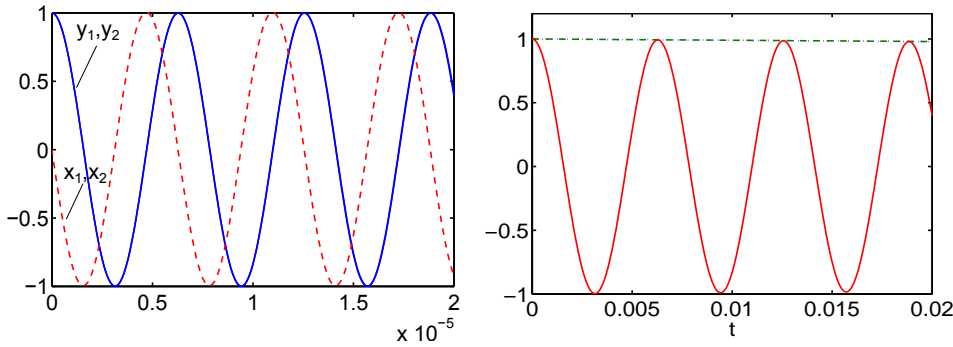


FIG. 6.3. (Section 6.2) The dynamics of (6.3) on the (Left) ϵ^2 time scale and (Right) ϵ^1 time scale. $\epsilon = 10^{-3}$.

6.3. Example 3. Consider the following deterministic system describing two coupled fast harmonic oscillators and a slow dependent mode:

$$\begin{cases}
 x_1' = -\epsilon^{-2}(1 + 0.5 \sin y_2)x_2 + (1 - z)(x_1^2 + x_2^2)^{-1}x_1, \\
 x_2' = \epsilon^{-2}(1 + 0.5 \sin y_2)x_1, \\
 y_1' = -\epsilon^{-1}y_2 - 0.5(1 + x_1^2 - z)y_1, \\
 y_2' = \epsilon^{-1}y_1, \\
 z' = -(1 + 0.5x_1^2)z + y_2^2,
 \end{cases} \tag{6.4}$$

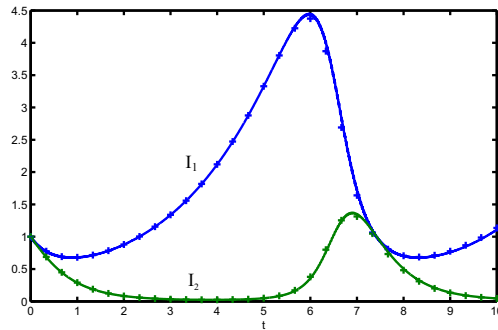


FIG. 6.4. (Section 6.2) The dynamics of (6.3) on the ϵ^0 time scale. Plus signs are results of the 3-tier HMM.

with initial conditions $(x_1(0), x_2(0), y_1(0), y_2(0), z(0)) = (1, 0, 1, 1.5, 0.5)$. The system describes two coupled harmonic oscillators (x_1, x_2) and (y_1, y_2) with $\mathcal{O}(\epsilon^2)$ and $\mathcal{O}(\epsilon)$ periods, respectively. However, the period of the fastest $\mathcal{O}(\epsilon^2)$ oscillator depends on y_2 and is therefore changing on the slower ϵ scale. Figure 6.5 (Left) demonstrates the different period of x_1 and x_2 over 2.5×10^{-7} duration with $\epsilon = 10^{-4}$. This is an example for the first generalization suggested in Section 5 in which the fastest oscillation exhibits non-trivial dynamics of the intermediate ϵ scale.

The system admits three slow variables that evolve on the ϵ^0 scale: z , and the square amplitudes of the harmonic oscillators, $I_1 = x_1^2 + x_2^2$ and $I_2 = y_1^2 + y_2^2$. A numerical algorithm for identifying polynomial slow variables is described in [3]. Hence, we have a coordinate system (ξ_0, ξ_1, ξ_2) in which ξ_i evolves on the ϵ^i time scale:

$$\begin{aligned}\xi_0 &= \{I_1, I_2, z\}, \\ \xi_1 &= \{y_2\}, \\ \xi_2 &= \varphi \in S^1.\end{aligned}$$

As before, we refer to the solver integrating the ϵ^i scale as the i -th tier. The step-size and length of integration of the i -th tier are denoted h_i , η_i , respectively. The dynamics of the slow variables I_1 , I_2 , and z , as well as the 3-tier HMM approximation is depicted in figure 6.5 (Right). See table 6.3 for simulation parameters.

$\epsilon = 10^{-4}$	η_i	h_i	Method	Kernel
2nd tier	$75.1\epsilon^2$	$\epsilon^2/10$	RK4	$K^{exp} \in \mathbb{K}^{1,\infty}([-1,1])$
1st tier	$75.1\epsilon^1$	$\epsilon^1/10$	RK2	$K^{exp} \in \mathbb{K}^{1,\infty}([-1,1])$
0th tier	10	1/2	RK2	-

TABLE 6.3. (Section 6.3) Parameters for the 3-tier HMM of Example 3.

6.4. Example 4. Consider the following system in which a fast harmonic oscillator has a randomly changing period:

$$\begin{cases} dx_1 = -(\epsilon^{-2}(1 + 0.5 \sin y)x_2 + x_1(1 - z))dt, \\ dx_2 = \epsilon^{-2}(1 + 0.5 \sin y)x_1 dt, \\ dy = -\epsilon^{-1}y dt + \epsilon^{-1/2}z dB_t, \\ dz = -((1 + x_1^2)z - y)dt, \end{cases} \tag{6.5}$$

with initial conditions: $(x_1(0), x_2(0), y(0), z(0)) = (2, 0, 1, 1)$. In this example, (x_1, x_2) is a harmonic oscillator with an $\mathcal{O}(\epsilon^2)$ period. However, the period changes randomly through a random variable y which is an Ornstein–Uhlenbeck process evolving on the $\mathcal{O}(\epsilon^1)$ time scale. The system has two slow variables that evolve on the $\mathcal{O}(\epsilon^0)$ scale: z and $I_1 = x_1^2 + x_2^2$. Thus, we find a coordinate system; $\xi_0 = \{I_1, z\}$, $\xi_1 = \{y\}$, and $\xi_2 = \varphi \in S^1$ in which ξ_i evolve on the $\mathcal{O}(\epsilon^i)$ scale.

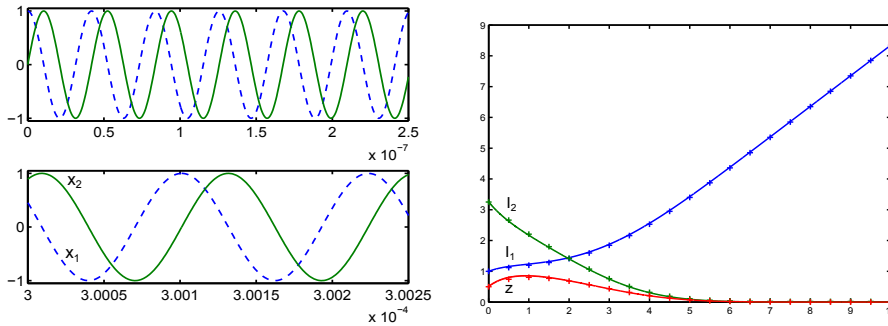


FIG. 6.5. (Section 6.3) (Left) The period of the fastest $\mathcal{O}(\epsilon^2)$ oscillator is changing on the slower ϵ scale. (Right) The dynamics of (6.4) on the ϵ^0 time scale. Plus signs: 3-tier HMM. Solid line: a reference solution using the RK4 method with step size $h = \epsilon^2/50$. HMM runs about 1150 times faster. $\epsilon = 10^{-4}$.

$\epsilon = 10^{-4}$	η_i	h_i	Method	Kernel
2nd tier	$50\epsilon^2$	$\epsilon^2/10$	semi-implicit Euler	$K^{2,7} \in \mathbb{K}^{2,7}([0,1])$
1st tier	$50\epsilon^1$	$\epsilon^1/10$	Euler	$K^{2,7} \in \mathbb{K}^{2,7}([0,1])$
0th tier	10	1/10	RK2	-

TABLE 6.4. (Section 6.4) Parameters for the 3-tier HMM of Example 4.

In order to demonstrate that the effective dynamics of z and I_1 is indeed deterministic, figure 6.6 (Left) shows the standard deviations of I_1 and z as a function of ϵ . As expected, it is of order $\sqrt{\epsilon}$.

Figure 6.6 (Right) compares the results computed by the proposed HMM with those by the semi-implicit Euler method [33]. The sample averages of I_1 and z against t are plotted with a solid line (Euler) and plus signs (HMM). We estimate the errors of the method by comparing the standard deviation of sample paths. Taking $\epsilon = 10^{-4}$, for the semi-implicit Euler, we take 1,000 paths over $[0, 4]$ and decrease step size until the desired accuracy is reached, $(\max\{\sigma(I_1)\} + \max\{\sigma(z)\})/2 = 0.1$. This requires

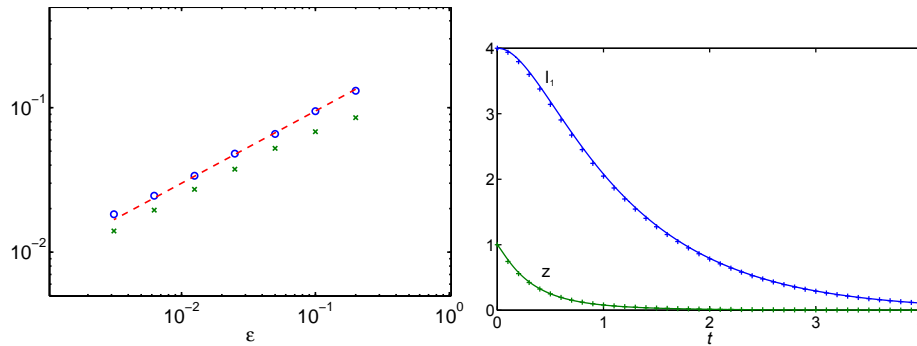


FIG. 6.6. (Section 6.4) (Left) \circ -markers are $\max_{t \in [0,4]} \{\sigma(I_1)\}$. \times -markers are $\max_{t \in [0,4]} \{\sigma(z)\}$. The dashed line is a guide for the eye with slope $1/2$. (Right) The evolution of the slow variables in Example 3. Plus signs: HMM. The solid line is a reference solution calculated as explained in the text.

$h = \epsilon^2/100$. For the 3-tier HMM, we compute 100 independent paths with $h_0=0.5$ and calculate the standard deviation for each. The kernel was constructed from Chebyshev polynomials to have exactly seven continuous derivatives and two vanishing moments:

$$K^{2,7}(t) = 4157010 \chi_{[0,1]}(t)(42t^2 - 44t + 11)t^8(t-1)^8,$$

where $\chi_{[0,1]}$ is the characteristic function of the interval $[0,1]$. HMM parameters are shown in table 6.4. With these parameters, HMM achieves $(\max\{\sigma(I_1)\} + \max\{\sigma(z)\})/2 = 0.025$. Even if HMM has four times less standard deviation, it runs about 1,000 times faster than the semi-implicit Euler. In addition, we note that the dominant error of 3-tier HMM comes from h_0 and decreases with smaller h_0 .

7. Summary

We developed an iterated averaging theory for oscillatory dynamical systems involving three widely separated time scales and the relevant multiscale method for computing the effective behavior. In such multiple time scale problems, we identified a new type of slow variables which do not have formally bounded derivatives. The effective behavior for such variables are studied intensively in two ways: one is a formal approach via the tools of averaging theory, and the other involves homogenization techniques based on singular perturbation expansions and consequent matching of variables. We showed that the results of the developed averaging theory can be efficiently approximated computationally via convolutions of the dynamical system's solutions with a smooth compactly supported kernel. With the developed averaging strategies, we proposed an HMM which is built hierarchically from our previously developed HMMs for two-scale problems. Several numerical examples were presented that demonstrate the efficacy of the proposed algorithms.

Acknowledgments. Kim and Tsai are partially supported by NSF grants DMS-0914840 and DMS-0914465, and Tsai is partially supported by a Moncrief Grand Challenge Award. Engquist is partially supported by NSF grants DMS-1217203 and DMS-0714612.

REFERENCES

- [1] G. Ariel, B. Engquist, S. Kim, Y. Lee, and R. Tsai, *A multiscale method for highly oscillatory dynamical systems using a poincaré map type technique*, J. Sci. Comput., 54(2-3), 247–268, 2013.
- [2] G. Ariel, B. Engquist, H.-O. Kreiss, and R. Tsai, *Multiscale computations for highly oscillatory problems*, Lect. Notes Comput. Sci. Eng., Springer, Berlin, 66, 237–287, 2009.
- [3] G. Ariel, B. Engquist, and R. Tsai, *A multiscale method for highly oscillatory ordinary differential equations with resonance*, Math. Comput., 78, 929–956, 2009.
- [4] G. Ariel, B. Engquist, and R. Tsai, *Numerical multiscale methods for coupled oscillators*, Multi. Mod. Simul., 7, 1387–1404, 2009.
- [5] G. Ariel, B. Engquist, and R. Tsai, *A reversible multiscale integration method*, Comm. Math. Sci., 7, 595–610, 2009.
- [6] G. Ariel, B. Engquist, and R. Tsai, *Oscillatory systems with three separated time scales – analysis and computation*, Lect. Notes Comput. Sci. Eng., Springer, Berlin, 82, 23–45, 2011.
- [7] Z. Artstein, *A Young measures approach to averaging*, in Differential Equations, Chaos and Variational Problems, Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser Verlag, Basel, 75, 15–28, 2007.
- [8] Z. Artstein, J. Linshiz, and E. S. Titi, *Young measure approach to computing slowly advancing fast oscillations*, Multiscale Model. Simul., 6(4), 1085–1097, 2007.
- [9] M. Avellaneda, *Iterated homogenization, differential effective medium theory and applications*, Comm. Pure Applied Math., 150, 527–554, 1987.
- [10] A. Bensoussan, J.L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, Studies in Mathematics and its Applications, North-Holland, Amsterdam, New York, Oxford, 5, 1978.
- [11] N.N. Bogoliubov and Yu. A. Mitropolski, *Asymptotic Methods in the Theory of Nonlinear Oscillations*, Gordon and Breach, New York, 1961.
- [12] A. Braides and . Lukkassen, *Reiterated homogenization of integral functionals*, Math. Models Meth. Appl. Sci., 10(01), 47–71, 2000.
- [13] I. Bright, *Tight estimates for general averaging applied to almost-periodic differential equations*, J. Diff. Equ., 246, 2922–2937, 2009.
- [14] Y. Cao, D.T. Gillespie, and L.R. Petzold, *The slow-scale stochastic simulation algorithm*, J. Chem. Phys., 122, 014116, 2005.
- [15] W. E., *Principles of Multiscale Modeling*, Cambridge University Press, Cambridge, UK, 2011.
- [16] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *Heterogeneous multiscale methods: A review*, Comm. Comput. Phys., 2, 367–450, 2007.
- [17] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. on Pure and Applied Math., 58, 1544–1585, 2005.
- [18] W. E, D. Liu, and E. Vanden-Eijnden, *Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales*, J. Comput. Phys., 221, 158–180, 2007.
- [19] B. Engquist, A. Fokas, E. Hairer, and A. Iserles, *Highly Oscillatory Problems*, London Mathematical Society Lecture Notes Series, Cambridge University Press, Cambridge, UK, 366, 2009.
- [20] B. Engquist and Y.-H. Tsai, *Heterogeneous multiscale methods for stiff ordinary differential equations*, Math. Comput., 74(252), 1707–1742, 2005.
- [21] L.C. Evans and D. Gomes, *Effective Hamiltonians and averaging for Hamiltonian dynamics i*, Arch. Rational Mech. Anal., 157, 1–33, 2001.
- [22] L.C. Evans and D. Gomes, *Effective Hamiltonians and averaging for Hamiltonian dynamics ii*, Arch. Rational Mech. Anal., 161, 271–305, 2002.
- [23] I. Fatkullin and E. Vanden-Eijnden, *A computational strategy for multiscale chaotic systems with applications to Lorenz 96 model*, J. Comput. Phys., 200, 605–638, 2004.
- [24] E. Fermi, J. Pasta, and S. Ulam, *Studies of the Nonlinear Problems*, i, Los Alamos Report LA-1940, 1955.
- [25] 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.
- [26] D. Givon, I.G. Kevrekidis, and R. Kupferman, *Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems*, Comm. Math. Sci., 4, 707–729, 2006.
- [27] D. Givon, R. Kupferman, and A.M. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms*, Nonlin., 17, R55–R127, 2004.

- [28] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration*, Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 31, 2002.
- [29] A. Holmbom, N. Svanstedt, and N. Wellander, *Multiscale convergence and reiterated homogenization of parabolic problems*, Appl. Math., 50, 131–151, 2005.
- [30] J. Kevorkian and J.D. Cole, *Multiple Scale and Singular Perturbation Methods*, Appl. Math. Sci., Springer-Verlag, New York, Berlin, Heidelberg, 114, 1996.
- [31] Y. Kifer, *Random Perturbations of Dynamical Systems*, Progress in Probability and Statistics, Birkhäuser Boston Inc., Boston, 16, 1988.
- [32] M. Krupa, N. Popović, and N. Kopell, *Mixed-mode oscillations in three time-scale systems: A prototypical example*, SIAM J. Appl. Dyn. Syst., 7(2), 361–420, 2008.
- [33] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 14, 2004.
- [34] D. Lukkassen and G.W. Milton, *On hierarchical structures and reiterated homogenization*, Function Spaces, Interpolation Theory and Related Topics, de Gruyter, Berlin, 355–368, 2002.
- [35] A.M. Majda, I. Timofeyev, and E. Vanden-Eijnden, *Stochastic models for selected slow variables in large deterministic systems*, Nonlin., 19, 769–794, 2006.
- [36] I. Melbourne and A.M. Stuart, *A note on diffusion limits of chaotic skew product flows*, Nonline., 24, 1361, 2011.
- [37] J.A. Murdock, *Perturbations. Theory and Methods*, Wiley-Interscience, New York, 1991.
- [38] G. Papanicolaou, *Introduction to the asymptotic analysis of stochastic equations*, Modern modeling of Continuum Phenomena, Lectures in Applied Mathematics, Amer. Math. Soc., Providence, RI, 16, 47–109, 1977.
- [39] G.A. Pavliotis and A.M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts in Applied Mathematics, Springer-Verlag, New York, 53, 2008.
- [40] J.A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems*, Appl. Math. Sci., Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 59, 1985.
- [41] J.M. Sanz-Serna, *Modulated Fourier expansions and heterogeneous multiscale methods*, IMA J. Numer. Anal., 29(3), 595–605, 2009.
- [42] M. Tao, H. Owhadi, and J. Marsden, *Nonintrusive and structure preserving multiscale integration of stiff odes, sdes, and hamiltonian systems with hidden slow dynamics via flow averaging*, Multi. Model. Simul., 8, 1269–1324, 2010.
- [43] A.H.P. van der Burgh, *On the higher order asymptotic approximations for the solutions of the equations of motion of an elastic pendulum*, J. Sound and Vibrations, 42, 463–475, 1975.
- [44] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Comm. Math. Sci., 1, 385–391, 2003.
- [45] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics. Springer, October, 79, 2000.