

ICES REPORT 13-01

January 2013

Iterated averaging of three-scale oscillatory systems

by

Gil Ariel, Bjorn Engquist, Seong Jun Kim and Richard Tsai



The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Gil Ariel, Bjorn Engquist, Seong Jun Kim and Richard Tsai, Iterated averaging of three-scale oscillatory systems, ICES REPORT 13-01, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, January 2013.

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.

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

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

where $0 < \varepsilon \leq \varepsilon_0 \ll 1$ 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

$$\xi' = F(\xi), \quad \xi(0) = x_0, \quad (2)$$

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

$$F(\xi) = \langle f(\xi, \cdot) \rangle_s \equiv \int_0^1 f(\xi, s) ds. \quad (3)$$

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

$$\sup_{0 \leq t \leq T} |x(t) - \xi(t)| = \mathcal{O}(\varepsilon). \quad (4)$$

Gil Ariel
Bar-Ilan University, Ramat Gan, 52900, Israel e-mail: arieltg@math.biu.ac.il

Bjorn Engquist · Seong Jun Kim · Richard Tsai
Department of Mathematics and Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, TX 78712, USA e-mail: engquist@ices.utexas.edu, e-mail: skim@math.utexas.edu, e-mail: 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)-(4) to a class of highly oscillatory systems 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 points of view: (1) application of long time asymptotic and second order averaging theorems [40, 43]; (2) formal perturbation expansions applied to an associated 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 multiscale 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 motion on the ε time scale, motion of the 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 review of the near-resonance is beyond the scope of this paper, but a new approach for these general systems will be reported in a forthcoming paper.

On this account, we consider a system of ODEs written in the following form;

$$\begin{cases} x' = f_1(x, y, z) + \varepsilon f_0(x, y, z), & x(0) = x_0, \\ y' = g_1(x, y, z) + \varepsilon g_0(x, y, z), & y(0) = y_0, \\ z' = \varepsilon^{-1} h_2(x, y, z), & z(0) = z_0 \end{cases} \quad (5)$$

where f_1 and f_0 are periodic in both y and z ; $z(t) \in \mathbb{R}$ is the fast variable, and its time derivative in $\mathcal{O}(\varepsilon^{-1})$ time interval remains $\mathcal{O}(\varepsilon^{-1})$; $y(t) \in \mathbb{R}$ is the intermediate variable with derivative bounded uniformly in any $\mathcal{O}(\varepsilon^{-1})$ time intervals; $x \in \mathbb{R}^d$ is the slow variable, which effectively varies on the time scale of $\mathcal{O}(\varepsilon^{-1})$ or longer. More precisely, there exists slowly changing quantity $\xi(t)$ over $t \in [0, \varepsilon^{-1}T]$ and $C > 0$ such that

$$\sup_{0 \leq t \leq \varepsilon^{-1}T} |x(t) - \xi(t)| \leq C\varepsilon$$

where $T = \mathcal{O}(1)$. When only looking for the effective dynamics $\xi(t)$ of the slow variable $x(t)$ in the time interval $[0, T]$ which is independent of ε , one can simply truncate the εf_0 term, and average over y and z . On the other hand, in case of $\varepsilon^{-1}T$, we cannot neglect εf_0 because its effect appears in this longer time interval. We will show that the condition for existence of $\xi(t)$ is that f_1 should have a zero average in z for fixed x and y .

For convenience we rescale (5) by replacing t by $\varepsilon^{-1}t$ so that the slowest time scale of interest is independent of ε . When more than two time scales are considered as above, a new type of slow variables whose formal derivatives are not bounded along the flow of the dynamical system appears. Note that the time derivative of $x(t)$ now becomes formally unbounded for $0 < \varepsilon \leq \varepsilon_0$. Such type of slow variables does not exist for problems with only two separated time scales. We refer to [6] for more detailed discussions. In addition, we emphasize that the effective behavior described by this type of slow variables should be taken into account on building multiscale algorithms. This will be discussed later in Section 5.2.

We further assume that the evolution of $x(t)$ is essentially constant in time intervals whose length is $o(1)$, i.e.,

$$|g_1(x(t), y(t), z(t)) - g_1(x(t_0), y(t), z(t))| \leq C_1\varepsilon, \quad |h_2(x(t), y(t), z(t)) - h_2(x(t_0), y(t), z(t))| \leq C_2\varepsilon$$

hold over $\mathcal{O}(\varepsilon)$, $\mathcal{O}(\varepsilon^2)$ time intervals respectively. In this context, $y(t)$ and $z(t)$ is essentially independent of $x(t)$ and the following simplified model equation will be considered in the paper:

$$x' = \varepsilon^{-1} f_1(x, \varepsilon^{-1}t, \varepsilon^{-2}t) + f_0(x, \varepsilon^{-1}t, \varepsilon^{-2}t), \quad x(0) = x_0, \quad (6)$$

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 . It is assumed that a unique bounded solution exists in a time segment $[0, T]$. Both the bound and T are independent of ε . The dynamics of (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 the following, we shall use s_1 and s_2 exclusively for the variables which will be scaled respectively to $\varepsilon^{-1}t$ and $\varepsilon^{-2}t$. In the following, 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

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

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

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

We proceed 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 $\varepsilon \rightarrow 0$ [37, 40]. As a motivation for identifying an ODE for $w(t)$, consider the form

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

Substituting into (6) yields

$$w' = \varepsilon^{-1} \left(f_1 - \frac{\partial \phi}{\partial s_2} \right) + (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1} + f_0 + \mathcal{O}(\varepsilon). \quad (7)$$

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

$$w' = (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1} + f_0 + \mathcal{O}(\varepsilon). \quad (8)$$

Since ϕ is periodic, $\partial \phi / \partial s_1$ 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 (8) should be averaged with respect to s_1 and s_2 . This leads to the averaged equation

$$\xi' = F(\xi), \quad \xi(0) = x_0, \quad (9)$$

where

$$F = \langle (\nabla_x f_1) \phi + f_0 \rangle_{12}, \quad (10)$$

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

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

The following theorem is proved in Section 2.

Theorem 1. *Let $x(t)$ and $\xi(t)$ denote solutions of equations (6) and (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*

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

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 $\mathcal{O}(\varepsilon)$ 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 both oscillatory, stochastic and chaotic dynamics. Such generalizations are further discussed in Section 5.

1.1 A simple example

Consider the ODE system

$$x' = 2\pi\varepsilon^{-1} [x \sin(2\pi\varepsilon^{-2}t) + \cos(2\pi\varepsilon^{-2}t)], \quad x(0) = 0. \quad (11)$$

for $0 \leq t \leq 1$. By the method of variation of parameters, the exact solution of (11) is found to be

$$x(t) = 2\pi\varepsilon^{-1} e^{-\varepsilon \cos(2\pi\varepsilon^{-2}t)} \int_0^t \cos(2\pi\varepsilon^{-2}s) e^{\varepsilon \cos(2\pi\varepsilon^{-2}s)} ds. \quad (12)$$

Expanding in ε ,

$$x(t) = \pi t + \varepsilon [\sin(2\pi\varepsilon^{-2}t) - \pi t \cos(2\pi\varepsilon^{-2}t)] + \mathcal{O}(\varepsilon^2). \quad (13)$$

Hence, $x(t) = \pi t + \mathcal{O}(\varepsilon)$ 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 (11) can be approximated by integrating over the fast oscillations in the RHS of (11) while keeping x fixed. This yields

$$\bar{x}(t) = 2\pi\varepsilon^{-1} \int_0^t [X \sin(2\pi\varepsilon^{-2}s) + \cos(2\pi\varepsilon^{-2}s)] ds \Big|_{X=\bar{x}(t)} = 2\pi\varepsilon^{-1} [X \mathcal{O}(\varepsilon^2) + \mathcal{O}(\varepsilon^2)], \quad (14)$$

which implies that $\bar{x}(t) = \mathcal{O}(\varepsilon)$ in $t \in [0, 1]$.

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

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

which clearly implies that

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

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 (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 (11) at $t = 0$

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

Substituting in the expansion of $x(s)$ shown in Eq. (13), and using $\eta = n\varepsilon^2$ yields

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

which is, to leading order in ε , the correct derivative for the slow variable x .

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 stochastic differential equations [15, 35, 38, 44]. Consider SDE systems of the form

$$\begin{cases} dx_t = [\varepsilon^{-1} f_1(x_t, y_t) + f_0(x_t, y_t)] dt \\ dy_t = \varepsilon^{-2} a(x_t, y_t) dt + \varepsilon^{-1} \beta(x_t, y_t) dB_t, \end{cases} \quad (15)$$

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

$$dX_t = F(X_t)dt + b(X_t)dB_t, \quad (16)$$

where F and b can be expressed as averages with respect to the fast process y_t with X fixed [38, 39]. One of the interesting consequences of (16) is that if the dynamics of y 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 $\mathcal{O}(\varepsilon^2)$, $\mathcal{O}(\varepsilon)$ 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 $\mathcal{O}(\varepsilon^2)$ as well as the $\mathcal{O}(\varepsilon)$ scale oscillations, thus obtaining a numerical approximation for the effective equation. See [6] for further details.

The hierarchical HMM structure is illustrated in Fig. 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 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 $t_{n,m} = n\eta_0 + mh_1$.

¹ Loosely 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]

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

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

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

The 1st tier numerically approximates the effective ODE for the $\mathcal{O}(\varepsilon)$ 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 $t \in [0, \eta_0] = [0, T]$.

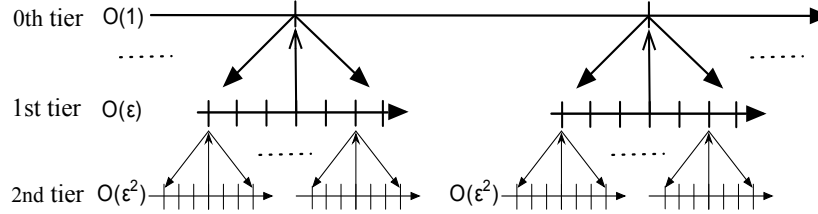


Fig. 1 An illustration of a three scale algorithm.

1.4 Layout

The layout of the paper is as follows. Section 2 details a proof of Theorem 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. Section 4 proves that the effective dynamics 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 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

$$x' = \varepsilon^{-1} f_1(x, \varepsilon^{-1}t, \varepsilon^{-2}t) + f_0(x, \varepsilon^{-1}t, \varepsilon^{-2}t), \quad x(0) = x_0, \quad (17)$$

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

$$w' = h(w(t), \varepsilon^{-1}t, \varepsilon^{-2}t) + f_0(w(t), \varepsilon^{-1}t, \varepsilon^{-2}t), \quad w(0) = x_0, \quad (18)$$

where h is defined by

$$\begin{aligned} h(x, s_1, s_2) &= (\nabla_x f_1) \phi(x, s_1, s_2) - \frac{\partial \phi}{\partial s_1}(x, s_1, s_2), \\ \phi(x, s_1, s_2) &= \int_0^{s_2} f_1(x, s_1, \tau) d\tau. \end{aligned} \quad (19)$$

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 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

$$y' = G(y, \varepsilon^{-1}t), \quad y(0) = x_0, \quad (20)$$

where G is given by

$$G(y, s_1) = \langle h(y, s_1, s_2) \rangle_2 + \langle f_0(y, s_1, s_2) \rangle_2. \quad (21)$$

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

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

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

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

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

Lemma 1. *The solutions $x(t)$ and $w(t)$ defined above satisfy*

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

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 (17) and (18), integrating with respect to time yields that

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

This leads to

$$|x(t) - w(t)| \leq \left| \int_0^t \varepsilon^{-1} f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) - h(w(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau \right| + L_{f_0} \int_0^t |x(\tau) - w(\tau)| d\tau. \quad (22)$$

where L_{f_0} is a Lipschitz constant for f_0 . We will show that the first integral in (22) is bounded by

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

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

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

It follows from Gronwall's inequality that

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

To this end, let n denote the largest integer such that $\varepsilon^2 n \leq t$. We have

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

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

$$\varepsilon \int_0^1 f_1(x(t_j + \varepsilon^2 s), \varepsilon^{-1}t_j + \varepsilon s, s) ds. \quad (23)$$

By the fundamental theorem of calculus, we write

$$\begin{aligned} x(t_j + \varepsilon^2 s) &= x(t_j) + \int_{t_j}^{t_j + \varepsilon^2 s} x'(\tau) d\tau \\ &= x(t_j) + \int_{t_j}^{t_j + \varepsilon^2 s} [\varepsilon^{-1} f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) + f_0(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau)] d\tau \\ &= x(t_j) + \int_{t_j}^{t_j + \varepsilon^2 s} \varepsilon^{-1} f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau + \mathcal{O}(\varepsilon^2) \end{aligned} \quad (24)$$

since f_0 is bounded independent of ε . After the substitution of (24) into (23) and using a Taylor expansion of $f_1(x(t_j + \varepsilon^2 s), \varepsilon^{-1}t_j + \varepsilon s, s_2)$ around $s = 0$ while keeping s_2 fixed,

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

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

$$\begin{aligned} \varepsilon^{-1} \int_0^t f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau &= \sum_{j=0}^{n-1} \left[\int_0^1 \left(\int_{t_j}^{t_j + \varepsilon^2 s} f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau \right) \nabla_x f_1(x(t_j), \varepsilon^{-1}t_j, s) ds \right. \\ &\quad \left. + \int_0^1 \varepsilon^2 s \frac{\partial f_1}{\partial s_1}(x(t_j), \varepsilon^{-1}t_j, s) ds + \mathcal{O}(\varepsilon^3) \right]. \end{aligned} \quad (25)$$

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

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

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

$$\begin{aligned}
\int_0^t h(w(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau &= \sum_{j=0}^{n-1} \left[\int_j^{j+1} \left(\int_{t_j}^{\varepsilon^2 s} f_1(w(\varepsilon^2 s), \varepsilon s, \varepsilon^{-2}\tau) d\tau \right) \nabla_x f_1(w(\varepsilon^2 s), \varepsilon s, s) ds \right. \\
&\quad \left. - \varepsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(\varepsilon^2 s), \varepsilon s, \tau) d\tau ds \right] + \mathcal{O}(\varepsilon^2).
\end{aligned} \tag{26}$$

We need to compare (25) and (26). Note that $w'(t)$ is bounded independent of ε and that $\tau \in [j, j+1]$.

$$\begin{cases} w(\varepsilon^2 s) = w(\tau) + \mathcal{O}(\varepsilon^2 s - \tau) = w(\tau) + \mathcal{O}(\varepsilon^2), \\ w(\varepsilon^2 s) = w(\varepsilon^2 j) + \mathcal{O}(\varepsilon^2 s - \varepsilon^2 j) = w(t_j) + \mathcal{O}(\varepsilon^2) \end{cases} \tag{27}$$

and

$$\begin{cases} \varepsilon s - \varepsilon^{-1}\tau = \mathcal{O}(\varepsilon), \\ \varepsilon s - \varepsilon^{-1}t_j = \mathcal{O}(\varepsilon). \end{cases} \tag{28}$$

Therefore, the first integration in (26) can be written as

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

which has the same form of the first term in (25) with an $\mathcal{O}(\varepsilon^3)$ error.

Now compare the second integrals. Applying integration by parts to the second term in (25) gives

$$\begin{aligned}
\int_0^1 \varepsilon^2 s \frac{\partial f_1}{\partial s_1}(x(t_j), \varepsilon^{-1}t_j, s) ds &= \left[\varepsilon^2 s \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \varepsilon^{-1}t_j, \tau) d\tau \right]_{s=0}^1 - \varepsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \varepsilon^{-1}t_j, \tau) d\tau ds \\
&= -\varepsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(x(t_j), \varepsilon^{-1}t_j, \tau) d\tau ds
\end{aligned} \tag{29}$$

On the other hand, in (26) using (27) and (28),

$$\begin{aligned}
-\varepsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(\varepsilon^2 s), \varepsilon s, \tau) d\tau ds &= -\varepsilon^2 \int_j^{j+1} \int_0^s \frac{\partial f_1}{\partial s_1}(w(t_j), \varepsilon^{-1}t_j, \tau) d\tau ds + \mathcal{O}(\varepsilon^3) \\
&= -\varepsilon^2 \int_0^1 \int_0^s \frac{\partial f_1}{\partial s_1}(w(t_j), \varepsilon^{-1}t_j, \tau) d\tau ds + \mathcal{O}(\varepsilon^3)
\end{aligned} \tag{30}$$

Thus, it is shown that (29) and (30) are different up to an $\mathcal{O}(\varepsilon^3)$ error. Putting all estimates together and noting $n = \mathcal{O}(\varepsilon^{-2})$, we conclude that

$$\left| \int_0^t \varepsilon^{-1} f_1(x(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) - h(w(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau \right| \leq L_h \int_0^t |x(\tau) - w(\tau)| d\tau + \mathcal{O}(\varepsilon) \quad (31)$$

where L_h is a Lipschitz constant for the function h . Substituting (31) into (22) yields the desired estimate.

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

$$|w(t) - y(t)| \leq \varepsilon C e^{L_G t}, \quad (32)$$

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

Proof. Consider $w' - y'$. Using (18) and (20) we have,

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

where

$$z(w, s_1, s_2) = [f_0(w, s_1, s_2) - \langle f_0 \rangle_2(w, s_1)] + [h(w, s_1, s_2) - \langle h \rangle_2(w, s_1)].$$

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

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

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

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

Let L_G denote a Lipschitz constant for G . Then,

$$|w(t) - y(t)| \leq L_G \int_0^t |w(\tau) - y(\tau)| d\tau + |r(t; \varepsilon)|. \quad (33)$$

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

$$r(t; \varepsilon) = \sum_{j=0}^{n-1} \int_{\varepsilon^2 j}^{\varepsilon^2(j+1)} z(w(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau + \int_{\varepsilon^2 n}^t z(w(\tau), \varepsilon^{-1}\tau, \varepsilon^{-2}\tau) d\tau. \quad (34)$$

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

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

Expanding in ε yields,

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

Therefore,

$$r(t; \varepsilon) = \mathcal{O}(n\varepsilon^3) = \mathcal{O}(\varepsilon). \quad (35)$$

Substituting into (33) yields

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

Using Gronwall's inequality we conclude Lemma 2.

2.2 Proof of Theorem 1

Lemma 1 and 2 imply that

$$\sup_{0 \leq t \leq T} |x(t) - y(t)| \leq \varepsilon C, \quad (36)$$

where $y(t)$ is the solution of

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

with

$$\begin{aligned} G(y, s_1) &= \langle h(y, s_1, s_2) \rangle_2 + \langle f_0(y, s_1, s_2) \rangle_2, \\ h(y, s_1, s_2) &= (\nabla_x f_1) \phi - \frac{\partial \phi}{\partial s_1}, \\ \phi(y, s_1, s_2) &= \int_0^{s_2} f_1(y, s_1, \tau) d\tau. \end{aligned} \quad (37)$$

Since the RHS 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 $\mathcal{O}(\varepsilon)$ time scale. Noting that since 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 $y(t)$:

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

where

$$F(\xi) = \langle (\nabla_x f_1) \phi(\xi, s_1, s_2) \rangle_{12} + \langle f_0(\xi, s_1, s_2) \rangle_{12} \quad (38)$$

and we have

$$\sup_{0 \leq t \leq T} |y(t) - \xi(t)| \leq \varepsilon C. \quad (39)$$

Combining (36) with (39) completes the proof of Theorem 1.

Remark 3. 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.

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

$$\begin{cases} dx_t = [\varepsilon^{-1} f_1(x_t, y_t) + f_0(x_t, y_t)] dt \\ dy_t = \varepsilon^{-2} a(x_t, y_t) dt + \varepsilon^{-1} \beta(x_t, y_t) dB_t, \end{cases} \quad (40)$$

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 $\mathcal{O}(\varepsilon)$ time scale. It is well known that under some suitable conditions on (40), $x(t)$ satisfies an effective SDE that is independent of ε ,

$$dX_t = F(X_t)dt + b(X_t)dB_t, \quad (41)$$

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 (40) can be written as

$$\begin{aligned} \mathcal{L} &= \varepsilon^{-2} \mathcal{L}_2 + \varepsilon^{-1} \mathcal{L}_1 + \mathcal{L}_0 \\ \mathcal{L}_0 &= f_0 \cdot \nabla_x \\ \mathcal{L}_1 &= f_1 \cdot \nabla_x \\ \mathcal{L}_2 &= a \cdot \nabla_y + \frac{1}{2} \beta \beta^T : \nabla_y \nabla_y^T, \end{aligned} \quad (42)$$

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

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

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

$$\mathcal{L}_2 u_0 = 0 \quad (43)$$

$$\mathcal{L}_2 u_1 = -\mathcal{L}_1 u_0 \quad (44)$$

$$\mathcal{L}_2 u_2 = \partial_t u_0 - \mathcal{L}_1 u_1 - \mathcal{L}_0 u_0. \quad (45)$$

The leading order equation (43) is automatically satisfied since 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 (44) 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, (44) 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 (45) 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

$$\partial_t u_0 = F(x) \nabla_x u_0 + B(x) : \nabla_x \nabla_x u_0, \quad (46)$$

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

where

$$\begin{aligned} F(x) &= \langle -f_1 \cdot \nabla_x \mathcal{L}_2^{-1}[f_1] + f_0 \rangle_y \\ B(x) &= \langle -f_1 \mathcal{L}_2^{-1}[f_1] \rangle_y. \end{aligned} \quad (47)$$

We identify (46) as the backwards equation associated with the effective SDE (41) 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

$$[a \cdot \nabla_y + \frac{1}{2}\beta\beta^T : \nabla_y \nabla_y]r(x, y) = f_1(x, y), \quad (48)$$

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

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 manuscripts. For two-scale systems the reader is referred to [21, 22] and 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 (40) 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, since $\beta\beta^T = 0$ the cell problem (48) 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 (42). 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 $\text{Null } \mathcal{L}_2^*$, 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

$$\phi(x, y) = \int_0^y f_1(x, \tau) d\tau - \langle \int_0^y f_1(x, \tau) d\tau \rangle_y \quad (49)$$

so that $\langle \phi \rangle_y = 0$. Substituting into (47) yields

$$\begin{aligned} F(x) &= \langle -f_1 \cdot \nabla_x \phi + f_0 \rangle_y \\ B(x) &= \langle -f_1 \phi \rangle_y. \end{aligned} \quad (50)$$

In order to identify (50) with the averaged equation (37), we need a simple Lemma.

Lemma 3. *Let $h(s)$ denote an S -periodic function with zero average and let $H(s)$ be an anti-derivative of h , $H' = h$. Then,*

$$\int_0^S h(s)H(s)ds = 0. \quad (51)$$

Proof. First, we note that since $h(s)$ 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 (51).

Since 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 is 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

$$\begin{aligned} F(x) &= \langle (\nabla_x f_1) \cdot \phi + f_0 \rangle_y, \\ B(x) &= 0. \end{aligned} \quad (52)$$

Thus, we obtain the consistent form of the effective drift as proven by Theorem 1. Note that since (40) contains only two scales $\mathcal{O}(1)$ and $\mathcal{O}(\varepsilon^2)$, $\frac{\partial \phi}{\partial s_1} = 0$ in (37).

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.,

$$\int_{[-1,1]} K(t)t^r dt = \begin{cases} 1, & r = 0, \\ 0, & 1 \leq r \leq p. \end{cases} \quad (53)$$

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 (6), the convolution approximates $F(\xi)$ of Theorem 1. The proposition below establishes the accuracy of iterated averaging with multiple kernels under the scaling condition $\varepsilon^2 \ll \eta_2 \ll \varepsilon \ll \eta_1 \ll 1$.

Proposition 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*

$$|K_{\eta_1} * \{K_{\eta_2} * (\varepsilon^{-1}f_1 + f_0)(\cdot)\}(t) - F(\xi(t))| \leq C \left(\frac{\varepsilon^{2q-1}}{\eta_2^q} + \frac{\varepsilon^q}{\eta_1^q} + \frac{\eta_2^{p+1}}{\varepsilon^{p+2}} + \eta_1^{p+1} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1 \quad (54)$$

for some constant $C > 0$, where F is given by (37) as in Theorem 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

$$\langle f \rangle_k = \int_0^1 f(x, \dots, s_k, \dots) ds_k, \quad k = 1, 2, \dots, n. \quad (55)$$

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. *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 $\eta = \mathcal{O}(\varepsilon^k)$, $k > n - 1$,*

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

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

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

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

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

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

Lemma 5. 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 $\eta = \mathcal{O}(\varepsilon^k)$ with $k > n - 1$, there exists $C > 0$ such that

$$|K_\eta * (f(\cdot, \varepsilon^{-1}\cdot, \varepsilon^{-2}\cdot, \dots, \varepsilon^{-n}\cdot) - \langle f \rangle_n)| \leq C \left(\frac{\varepsilon^{nq}}{\eta^q} + \frac{\eta^\sigma}{\varepsilon^{(n-1)\sigma}} \right) \max_{j=0\dots q} \|K^{(j)}\|_1. \quad (56)$$

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

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

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

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

where $(\mu_1, \mu_2, \dots, \mu_n)$ is in the open line segment joining $\mathbf{0}$ and $(s_n, \varepsilon^{-1}s_n, \dots, \varepsilon^{-n+1}s_n)$ 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

$$\begin{aligned} I_1 &= \sum_{j=0}^{\sigma-1} \int_{-\eta}^{\eta} K_\eta(-s_n) \cdot \frac{1}{j!} \left[\sum_{i=1}^n \frac{s_n}{\varepsilon^{i-1}} \partial_{x_i} \right]^j g(0, \dots, 0, \varepsilon^{-n}s_n) ds_n, \\ I_2 &= \int_{-\eta}^{\eta} K_\eta(-s_n) \cdot \frac{1}{\sigma!} \left[\sum_{i=1}^n \frac{s_n}{\varepsilon^{i-1}} \partial_{x_i} \right]^\sigma g(\mu_1, \varepsilon^{-1}\mu_2, \dots, \varepsilon^{-n+1}\mu_n, \varepsilon^{-n}s_n) ds_n. \end{aligned}$$

By using Lemma 4, I_1 is estimated by

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

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

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

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

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

We now compare iterated averaging of (6) with the averaged equation (9). Before we move on to the next step, we simplify our notation of the forces by writing $f(t, \varepsilon^{-1}t, \varepsilon^{-2}t)$ instead of $f(x(t), \varepsilon^{-1}t, \varepsilon^{-2}t)$. This is possible since, after solving (6) at the 1st tier, $x(t)$ is known up to a prescribed accuracy Δ . Theorem 2 shows that the error between x and x_1 (solution of the 1st tier) is $\mathcal{O}(\Delta)$.

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

Theorem 2. 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 $0 < \varepsilon < \varepsilon_0$ and a prescribed accuracy Δ_1 , there exists $C > 0$ such that

$$\sup_{t \in [t_n - \eta_1, t_n + \eta_1]} |x(t) - x_1(t)| \leq C \Delta_1. \quad (57)$$

Proof. By considering 1st and 2nd tier as the two-scale HMM solvers, we generalize the error analysis discussed in [3, 6, 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 (6) to time $\eta_1 (\ll \varepsilon)$ is given by

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

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_1 with different sources of errors. Note that the maximal possible accuracy is $\Delta_1 = \varepsilon^2$ since this is the error introduced by the averaged equation on the $\mathcal{O}(\varepsilon)$ time.

Next, by iterated use of Lemma 5, 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 6. 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 $K \in \mathbb{K}^{p,q}$, $\varepsilon^2 \ll \eta_2 \ll \varepsilon \ll \eta_1 \ll 1$, there exists a constant $C > 0$ such that

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

Proof. Let

$$g(t, s_1, s_2) = f_0(t, s_1, s_2) - \langle f_0 \rangle_{12}(t). \quad (58)$$

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

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

We iterate the argument in Lemma 5. First, there exists $C > 0$ such that

$$|I_1(s_1) - \langle g \rangle_2(s_1, \varepsilon^{-1} s_1)| \leq C \left(\frac{\varepsilon^{2q}}{\eta_2^q} + \frac{\eta_2^\sigma}{\varepsilon^\sigma} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \quad (59)$$

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

$$I_1 = I_1 - \langle g \rangle_2(s_1, \varepsilon^{-1} s_1) + \langle g \rangle_2(s_1, \varepsilon^{-1} s_1) - \langle g \rangle_{12}. \quad (60)$$

Thus

$$\begin{aligned} & \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) \cdot I_1 ds_1 \right| \\ &= \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) (I_1 - \langle g \rangle_2(s_1, \varepsilon^{-1} s_1) + \langle g \rangle_2(s_1, \varepsilon^{-1} s_1) - \langle g \rangle_{12}) ds_1 \right| \\ &\leq C \left(\frac{\varepsilon^{2q}}{\eta_2^q} + \frac{\eta_2^\sigma}{\varepsilon^\sigma} \right) \max_{j=0 \dots q} \|K^{(j)}\|_1 + \left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) (\langle g \rangle_2(s_1, \varepsilon^{-1} s_1) - \langle g \rangle_{12}) ds_1 \right|. \end{aligned} \quad (61)$$

Second, define $\hat{g}(\hat{s}_1, s_1) = \langle g \rangle_2(\hat{s}_1, \varepsilon^{-1} s_1) - \langle g \rangle_{12}$. 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 5 yields existence of $C_2 > 0$ such that

$$\left| \int_{t-\eta_1}^{t+\eta_1} K_{\eta_1}(t-s_1) \hat{g}(\hat{s}_1, s_1) ds_1 \right| \leq C_2 \left(\frac{\varepsilon^q}{\eta_1^q} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \quad (62)$$

Hence, we can find a positive constant C such that

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

This concludes the proof of the theorem.

Recall the three scale problem (6) and its averaged equation (9):

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

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

Theorem 3. *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 suitable choice of η_1, η_2 such that $\varepsilon^2 \ll \eta_2 \ll \varepsilon \ll \eta_1 \ll 1$, the following estimate holds.*

$$|K_{\eta_1} * \{K_{\eta_2} * (\varepsilon^{-1} f_1(x(\cdot), \varepsilon^{-1}\cdot, \varepsilon^{-2}\cdot) - \langle h \rangle_{12}(\xi)) (\cdot)\} (t)| \leq C \left(\frac{\varepsilon^{2q-1}}{\eta_{2q}} + \frac{\varepsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\varepsilon^{\sigma+1}} + \eta_1^{\sigma-1} \right) \max_{j=0\dots q} \|K^{(j)}\|_1 \quad (63)$$

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 $K_{\eta_2} * (\varepsilon^{-1} f_1 - \langle h \rangle_{12})$. Lemma 1 allows ones to write $x(t)$ as

$$x(t) = w(t) + \varepsilon \psi(t) + \varepsilon \phi(t) \quad (64)$$

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

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

where we set

$$\begin{aligned} I_1 &= \int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1-s_2) \cdot (\varepsilon^{-1} f_1(w(s_2) + \varepsilon \psi(s_2), \varepsilon^{-1}s_2, \varepsilon^{-2}s_2)) ds_2 \\ I_2 &= \int_{s_1-\eta_2}^{s_1+\eta_2} K_{\eta_2}(s_1-s_2) \cdot (\nabla_x f_1(w(s_2) + \varepsilon \psi(s_2), \varepsilon^{-1}s_2, \varepsilon^{-2}s_2) \phi(s_2) - \langle h \rangle_{12}) ds_2. \end{aligned} \quad (65)$$

Each term in the integrations is estimated in a similar way. Before we move on, recall from Theorem 2 that we identify $f_1(x(s_2), \varepsilon^{-1}s_2, \varepsilon^{-2}s_2)$ with $f_1(s_2, \varepsilon^{-1}s_2, \varepsilon^{-2}s_2)$. In (65), this simplification is also allowed since by solving the 1st and 2nd tier, we know $x(t)$ over $[t - \eta_1, t + \eta_1]$, and so does $w(t) + \varepsilon \psi(t)$. First, applying Lemma 5 to I_1 with $\langle f_1 \rangle_2 = 0$ yields

$$|I_1| \leq \left(\frac{\varepsilon^{2q-1}}{\eta_2^q} + \frac{\eta_2^\sigma}{\varepsilon^{\sigma+1}} \right) \max_{j=0\dots q} \|K^{(j)}\|_1. \quad (66)$$

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 yields that

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

Then we have $\langle \nabla_x f_1 \cdot \phi - \langle h \rangle_{12} \rangle_{12} = 0 + \mathcal{O}(\varepsilon)$, and this allows us to use Lemma 6 and thus to get an estimate for $|K_{\eta_1} * I_2(\cdot)|$.

$$|K_{\eta_1} * I_2(\cdot)| \leq C \left(\frac{\varepsilon^{2q}}{\eta_2^q} + \frac{\varepsilon^q}{\eta_1^q} + \frac{\eta_2^\sigma}{\varepsilon^\sigma} + \eta_1^\sigma \right) \max_{j=0 \dots q} \|K^{(j)}\|_1. \quad (67)$$

Since 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 (66) and (67) together, Theorem 3 follows.

We conclude this section by proving Proposition 1 which is the cornerstone of our numerical method.

Proof. (Proposition 1) The RHS of the estimate in Theorem 3 dominates that of Lemma 6. Therefore, having $p (< \sigma)$ vanishing moments yields (54).

Remark 4. Theorem 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

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

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 $\mathcal{O}(\varepsilon^2)$ or $\mathcal{O}(\varepsilon)$ 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 $r(t; \varepsilon)$ in (34). In the case of a torus, $r(t; \varepsilon)$ can be written as a finite sum of periodic functions whose periods are incommensurable. Thus, estimate (35) 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, as well as while applying the expansion formalism, it was necessary to assume that the average of the singular $\mathcal{O}(\varepsilon^{-1})$ 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 $\mathcal{O}(\varepsilon)$ length, i.e.

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

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 $\mathcal{O}(\varepsilon)$ 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 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 $0 < \varepsilon \leq \varepsilon_0$, if there exists a smooth function $\beta : [0, T] \mapsto \mathbb{R}$ and constants C_0 and C_1 such that

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

Definition 2. A function $\xi(x)$ is said to *evolve on the ε^k time scale along the trajectories of (6) in $[0, T]$* if the time dependent function $\xi(x(t; \varepsilon, x_0))$ 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 since the time scales $\mathcal{O}(\varepsilon^2)$, $\mathcal{O}(\varepsilon^1)$, 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

$$\begin{aligned} \xi_0' &= \varepsilon^{-1} f_1(\xi_0, \xi_1, \xi_2) + f_0(\xi_0, \xi_1, \xi_2), \quad \langle f_1 \rangle_2 = 0, \\ \xi_1' &= \varepsilon^{-1} g_1(\xi_0, \xi_1, \xi_2) + g_0(\xi_0, \xi_1, \xi_2), \quad \langle g_1 \rangle_2 > C_1 > 0, \\ \xi_2' &= \varepsilon^{-2} h_2(\xi_0, \xi_1, \xi_2), \end{aligned} \tag{68}$$

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 both (68) and (??) are identical. We assume that no resonances, passage through resonances or turning points exists 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 (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 $\mathcal{O}(\varepsilon^1)$ scale: approximate $\xi_i'(t_{n,m})$, $i = 0, 1$ by

$$\langle \xi_i' \rangle_{\eta_2}(t_{n,m}) = (K_{\eta_2} * \xi_i')(t_{n,m}). \tag{69}$$

- 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

$$\delta x_1 \cdot \nabla \xi_i = \langle \xi_i' \rangle_{\eta_2}, \tag{70}$$

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

$$\langle \xi_0' \rangle_{\eta_1}(t_n) = K_{\eta_1} * (\langle \xi_0' \rangle_{\eta_2}(\cdot))(t_n). \tag{71}$$

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

$$\delta x_0 \cdot \nabla \xi_0 = \langle \xi_0' \rangle_{\eta_1}. \tag{72}$$

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 $\mathcal{O}(\varepsilon^2)$ scale is oscillatory - thus ensuring that the effective slow scale is deterministic, and that the intermediate $\mathcal{O}(\varepsilon)$ time scale is ergodic. Consider (68). 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 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 (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 $\varepsilon \rightarrow 0$, 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 $\mathcal{O}(\varepsilon^2)$ 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 Section 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 (12)

$$x' = \varepsilon^{-1} [x \sin(\varepsilon^{-2}t) + \cos(\varepsilon^{-2}t) \sin^2(\varepsilon^{-1}t)] + \cos^2(x) + \cos(\varepsilon^{-1}t), \quad x(0) = 1. \quad (73)$$

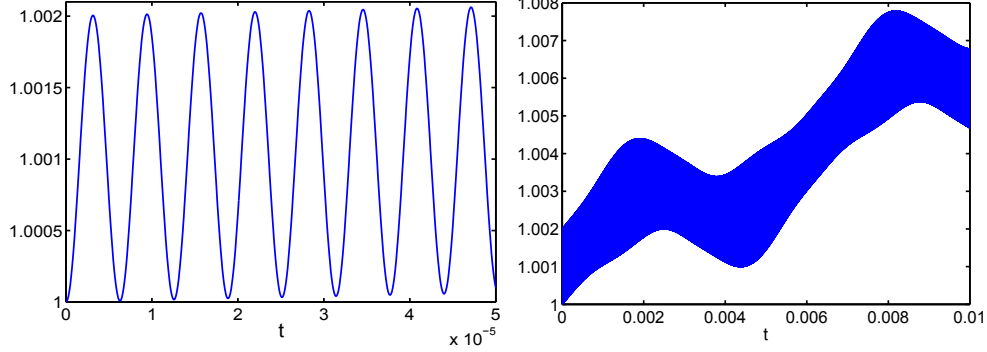
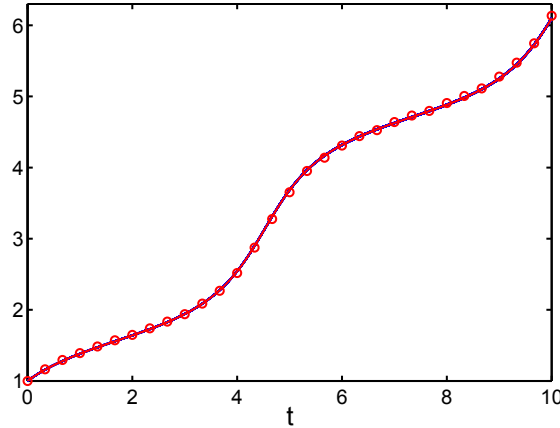
Applying Theorem 1 to (73), we have an averaged equation for $\bar{x}(t)$,

$$\bar{x}' = 1/4 + \cos^2(\bar{x}), \quad \bar{x}(0) = 1. \quad (74)$$

The different three-time scales of (73) are illustrated in Figures 2 and 3. The solution $x(t)$ undergoes small-amplitude fast oscillations around the slow trajectory over the interval $[0, 10]$. As proved in Section 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 3, 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 (73) directly with the step size $h = \varepsilon^2/5$. The computational effort of HMM is independent of ε once η_i and h_i are fixed. However, for classical numerical methods moving from $\varepsilon = 10^{-3}$ to $\varepsilon = 10^{-4}$ multiplies the computational effort by 100.

Table 1 Section 6.1. Parameters for the 3-tier HMM of example 1.

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

**Fig. 2** Section 6.1. The dynamics of (73) on the (Left) ε^2 time scale and (Right) ε^1 time scale ($\varepsilon = 10^{-3}$). Plots are obtained by RK4 with $h = \varepsilon^2/100$.**Fig. 3** Section 6.1. The dynamics of (73) 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.

6.2 Example 2

Consider the following system describing two coupled harmonic oscillators in resonance [6].

$$\begin{cases} x_1' &= -\varepsilon^{-2}y_1 + \varepsilon^{-1}y_2^2 - 3x_1x_2^2, \\ y_1' &= \varepsilon^{-2}x_1 + y_1/2, \\ x_2' &= -(\varepsilon^{-2} + \varepsilon^{-1})y_2 - x_2, \\ y_2' &= (\varepsilon^{-2} + \varepsilon^{-1})x_2 - y_2 + 2x_1^2y_2. \end{cases} \quad (75)$$

As depicted in Figure 4, 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.

In order to find a slow coordinate system, we change to polar coordinates $(x_i, y_i) \mapsto (I_i, \varphi_i)$, $i = 1, 2$ 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

$$\begin{aligned}
I_1' &= 2\varepsilon^{-1}x_1y_2^2 - 6x_1^2x_2^2 + y_1^2, \\
I_2' &= -2I_2 + 4x_1^2y_2^2, \\
\theta' &= \varepsilon^{-1}(x_2y_2^2 + y_1x_2 - x_1y_2) + (-0.5y_1y_2 - x_1x_2 - 3x_1x_2^3 + 2x_1^2y_1y_2), \\
\varphi_1' &= \varepsilon^{-2}.
\end{aligned}$$

It appears as if $(I_1, I_2, \theta, \varphi_1)$ is a chart in which φ_1 evolves of 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 $\mathcal{O}(\varepsilon)$ scale is depicted on the right in Figure 4. 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 $x_1y_2^2$ on any segment of length $\mathcal{O}(\varepsilon)$ 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 $\mathcal{O}(\varepsilon)$.

The time evolution of I_1 and I_2 on the slowest $\mathcal{O}(1)$ time scale is depicted in Figure 5. 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 2.

Table 2 Section 6.2. Parameters for the 3-tier HMM of example 2.

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

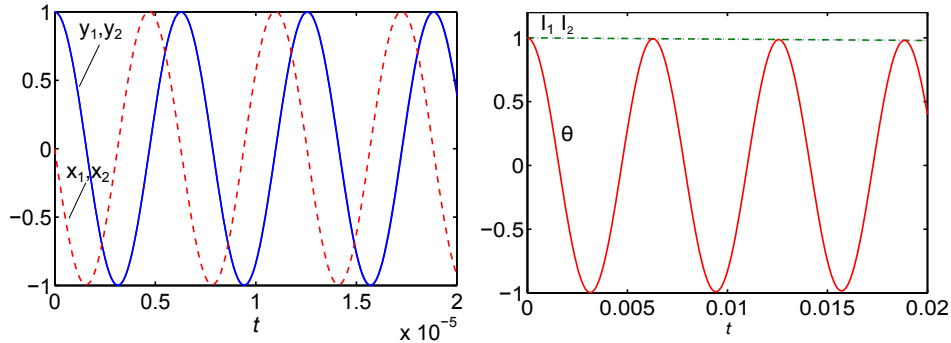


Fig. 4 Section 6.2. The dynamics of (75) on the (Left) ε^2 time scale and (Right) ε^1 time scale. $\varepsilon = 10^{-3}$.

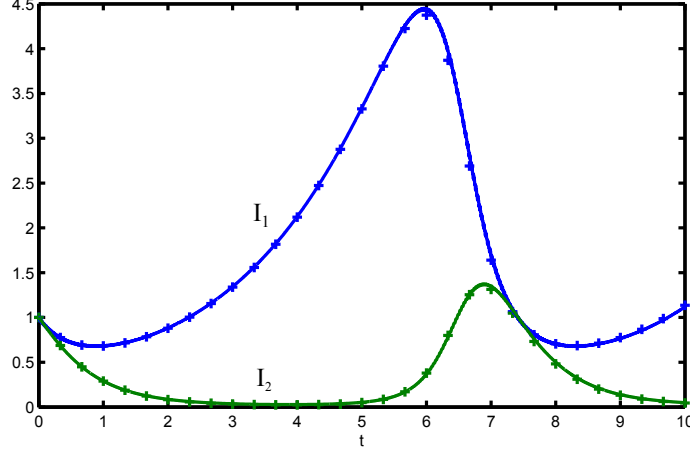


Fig. 5 Section 6.2. The dynamics of (75) on the ε^0 time scale. Plus signs are results of the 3-tier HMM.

6.3 Example 3

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

$$\begin{cases} x_1' = -\varepsilon^{-2}(1 + 0.5 \sin y_2)x_2 + (1 - z)(x_1^2 + x_2^2)^{-1}x_1, \\ x_2' = \varepsilon^{-2}(1 + 0.5 \sin y_2)x_1, \\ y_1' = -\varepsilon^{-1}y_2 - 0.5(1 + x_1^2 - z)y_1, \\ y_2' = \varepsilon^{-1}y_1, \\ z' = -(1 + 0.5x_1^2)z + y_2^2 \end{cases} \quad (76)$$

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 $\mathcal{O}(\varepsilon^2)$ and $\mathcal{O}(\varepsilon)$ periods, respectively. However, the period of the fastest $\mathcal{O}(\varepsilon^2)$ oscillator depends on y_2 and is therefore changing on the slower ε scale. Figure 6 (Left) demonstrates the different period of x_1 and x_2 over 2.5×10^{-7} duration with $\varepsilon = 10^{-4}$. 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 Fig. 6 (Right). See Table 2 for simulation parameters.

Table 3 Section 6.3. Parameters for the 3-tier HMM of example 3.

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

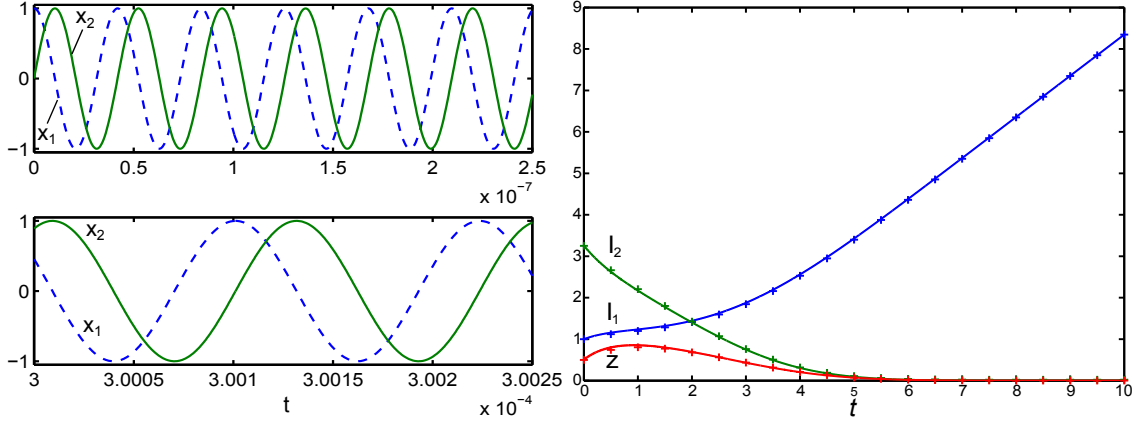


Fig. 6 Section 6.3. (Left) The period of the fastest $\mathcal{O}(\varepsilon^2)$ oscillator is changing on the slower ε scale. (Right) The dynamics of (76) on the ε^0 time scale. Plus signs: 3-tier HMM. Solid line: a reference solution using the RK4 method with step size $h = \varepsilon^2/50$. HMM runs about 1150 times faster. $\varepsilon = 10^{-4}$.

6.4 Example 4

Consider the following system in which a fast harmonic oscillator has a randomly changing period.

$$\begin{cases} dx_1 = -(\varepsilon^{-2}(1 + 0.5 \sin y)x_2 + x_1(1 - z))dt, \\ dx_2 = \varepsilon^{-2}(1 + 0.5 \sin y)x_1 dt, \\ dy = -\varepsilon^{-1}ydt + \varepsilon^{-1/2}zdB_t, \\ dz = -((1 + x_1^2)z - y)dt \end{cases} \quad (77)$$

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 $\mathcal{O}(\varepsilon^2)$ period. However, the period changes randomly through a random variable y which is an Ornstein–Uhlenbeck process evolving on the $\mathcal{O}(\varepsilon^1)$ time scale. The system has two slow variables that evolve on the $\mathcal{O}(\varepsilon^0)$ 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 $\mathcal{O}(\varepsilon^i)$ scale.

In order to demonstrate that the effective dynamics of z and I_1 is indeed deterministic, Figure 7 (Left) shows the standard deviations of I_1 and z as a function of ε . As expected, it is of order $\sqrt{\varepsilon}$.

Fig. 7 (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 $\varepsilon = 10^{-4}$, 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 $h = \varepsilon^2/100$. 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 a single vanishing moment:

$$K^{2.7}(t) = 4157010 \chi_{[0,1]}(t)(42t^2 - 44t + 11)t^8(t-1)^8$$

where $\chi_{[0,1]}$ is the characteristic function of the interval $[0, 1]$. HMM parameters are shown in Table 3. 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 .

Table 4 Section 6.4. Parameters for 3-tier HMM

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

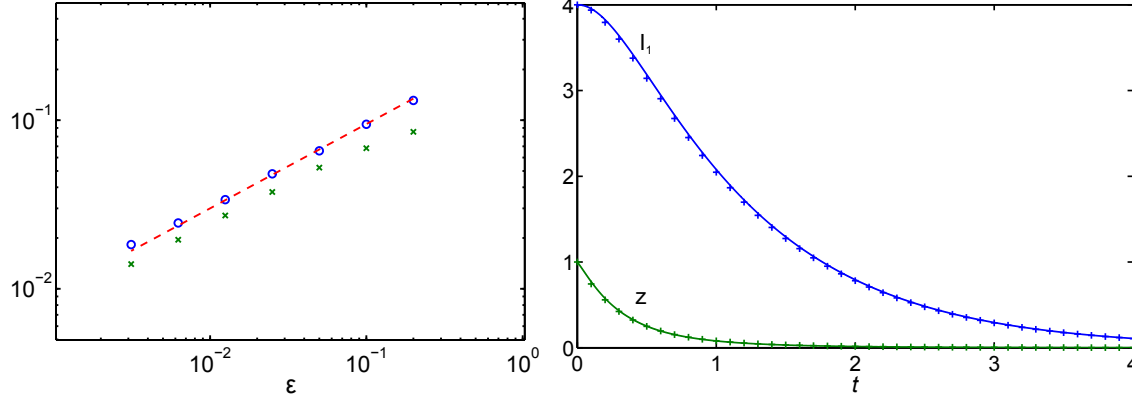


Fig. 7 Section 6.4. (Left) o-markers are $\max_{r \in [0,4]} \{\sigma(I_1)\}$, x-markers are $\max_{r \in [0,4]} \{\sigma(z)\}$. The dashed line is a guide for the eye with slope 1/2. (Right) The evolution of the slow variables in example 3. Plus signs: HMM. The solid line is a reference solution calculated as explained in the text.

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. Engquist was partially supported by NSF grant 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. To appear in *Journal of Scientific Computing*.
2. G. Ariel, B. Engquist, H.-O. Kreiss, and R. Tsai. Multiscale computations for highly oscillatory problems. In *Multiscale modeling and simulation in science*, volume 66 of *Lect. Notes Comput. Sci. Eng.*, pages 237–287. Springer, Berlin, 2009.
3. G. Ariel, B. Engquist, and R. Tsai. A multiscale method for highly oscillatory ordinary differential equations with resonance. *Math. Comp.*, 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. In *Numerical analysis of multi scale computations*, volume 82 of *Lect. Notes Comput. Sci. Eng.*, pages 23–45. Springer, Berlin, 2011.
7. Z. Artstein. A young measures approach to averaging, in differential equations, chaos and variational problems. volume 75 of *Progress in Nonlinear Differential Equations and Their Applications*, pages 15–28. Birkhäuser Verlag, Basel, 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*, volume 5 of *Studies in mathematics and its applications*. North-Holland, Amsterdam, New York, Oxford, 1978.
11. N. N. Bogoliubov and Yu. A. Mitropolski. *Asymptotic Methods in the Theory of Nonlinear Oscillations*. Gordon and Breach, New York, 1961.
12. Andrea Braides and Dag Lukkassen. Reiterated homogenization of integral functionals. *Mathematical Models and Methods in Applied Sciences*, 10(01):47–71, February 2000.
13. I. Bright. Tight estimates for general averaging applied to almost-periodic differential equations. *J. Diff. Eqn.*, 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. *Commun. 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, editors. *Highly Oscillatory Problems*, volume 366 of *London Mathematical Society Lecture Notes Series*. Cambridge University Press, Cambridge, UK, 2009.
20. B. Engquist and Y.-H. Tsai. Heterogeneous multiscale methods for stiff ordinary differential equations. *Math. Comp.*, 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. Comp. 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. Later published in *Collected Papers of Enrico Fermi*, ed. E. Segre, Vol. II (University of Chicago Press, 1965) p.978.
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. *Nonlinearity*, 17:R55–R127, 2004.
28. E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2002. Structure-preserving algorithms for ordinary differential equations.
29. A. Holmbom, N. Svanstedt, and N. Wellander. Multiscale convergence and reiterated homogenization of parabolic problems. *Applications of Math.*, 50:131–151, 2005.
30. J. Kevorkian and J. D. Cole. *Multiple Scale and Singular Perturbation Methods*, volume 114 of *Applied Mathematical Sciences*. Springer-Verlag, New York, Berlin, Heidelberg, 1996.
31. Y. Kifer. Random perturbations of dynamical systems. volume 16 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, 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*, volume 14 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2004.
34. D. Lukkassen and G.W. Milton. On hierarchical structures and reiterated homogenization. In *Function spaces, interpolation theory and related topic*, pages 355–368. de Gruyter, Berlin, 2002.

35. A.M. Majda, I. Timofeyev, and E. Vanden-Eijnden. Stochastic models for selected slow variables in large deterministic systems. *Nonlinearity*, 19:769–794, 2006.
36. I. Melbourne and A. M. Stuart. A note on diffusion limits of chaotic skew product flows. *Nonlinearity*, 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. In *Modern modeling of continuum phenomena*, volume 16 of *Lectures in Applied Mathematics*, pages 47–109. Amer. Math. Soc., Providence, RI, 1977.
39. G. A. Pavliotis and A. M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Number 53 in Texts in Applied Mathematics. Springer-Verlag, New York, 2008.
40. J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*, volume 59 of *Applied Mathematical Sciences*. Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1985.
41. J. M. Sanz-Serna. Modulated fourier expansions and heterogeneous multiscale methods. *IMA J. Numer. Anal.* to appear.
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. of 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. Peter Walters. *An Introduction to Ergodic Theory*. Number 79 in Graduate Texts in Mathematics. Springer, October 2000.