Physica D 328–329 (2016) 34–43

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

A trajectory-free framework for analysing multiscale systems☆

Gary Froyland [a], Georg A. Gottwald [b], Andy Hammerlindl [c,*]

[a] School of Mathematics and Statistics, University of New South Wales, Australia
[b] School of Mathematics and Statistics, University of Sydney, Australia
[c] School of Mathematical Sciences, Monash University, Australia

HIGHLIGHTS

- A trajectory-free method to test for multiscale dynamics.
- Isolates the fast dynamics and determines the reduced slow dynamics.
- Based on the spectral properties of the transfer and Koopman operators.

ARTICLE INFO

Article history:
Received 18 February 2015
Received in revised form
18 December 2015
Accepted 10 April 2016
Available online 19 April 2016
Communicated by I. Melbourne

Keywords:
Multiscale dynamics
Transfer operator
Koopman operator
Model reduction

ABSTRACT

We develop algorithms built around properties of the transfer operator and Koopman operator which (1) test for possible multiscale dynamics in a given dynamical system, (2) estimate the magnitude of the time-scale separation, and finally (3) distill the reduced slow dynamics on a suitably designed subspace. By avoiding trajectory integration, the developed techniques are highly computationally efficient. We corroborate our findings with numerical simulations of a test problem.

1. Introduction

Effective numerical simulation of multiscale systems constitutes a formidable challenge. Consider a system which has slow dynamics on a time-scale of order one and fast dynamics on the scale of order $1/\epsilon$ for some parameter $\epsilon \ll 1$. To accurately simulate orbits numerically and to assure numerical stability, the time step of the integrator must be of the order of ϵ. To capture the relevant slow dynamics a total number of integration steps of the order of $1/\epsilon$ is required, making direct numerical simulations of orbits computationally impractical.

Numerical integrators are subject to two main sources of error. The first is truncation error, which is the inability of the numerical method (Runge–Kutta, Euler–Maruyama, etc.) to fully capture the actual dynamics of the system. The second is round-off error, due to implementing the numerical method on a computer with finite precision arithmetic. While truncation error decreases with a smaller time step, round-off error increases [1,2]. In a multiscale system, if the time-scale separation is large, it may be impossible to find a time step which is simultaneously small enough to avoid significant truncation error for the fast dynamics and sufficiently large to avoid detrimental accumulation of round-off error for the slow dynamics.

Even if orbits could be computed exactly, analysing a multiscale system using a time series extracted from a true orbit can still yield incorrect data about the diffusion process of the slow variables [3]. To avoid this problem, the time series must be sampled at a rate intermediate between the slow and fast variables and these rates might not be known in advance.

There exists a variety of numerical methods dealing with one or more aspects of these numerical difficulties (see [4] and references therein). These methods rely on producing trajectories of the dynamical system via some form of time-integration with some of the issues mentioned above remaining. In this paper, we

☆ This project was funded by the Australian Research Council Grant DP120104514.

* Corresponding author. Tel.: +61 3 9905 4416.
E-mail addresses: G.Froyland@unsw.edu.au (G. Froyland), georg.gottwald@sydney.edu.au (G.A. Gottwald), andy.hammerlindl@monash.edu (A. Hammerlindl).

http://dx.doi.org/10.1016/j.physd.2016.04.010
0167-2789/

develop algorithms which avoid trajectory integration altogether. Besides the advantages relating to the issues of time-integration mentioned above, the algorithm allows for a huge reduction in computational time. Our main objective is to develop numerical algorithms which, given a dynamical system,

1. test whether the system exhibits multiscale behaviour, and if so
2. determine the order of the time-scale separation, and then
3. construct effective reduced equations for the slow dynamics allowing for the application of large time steps.

The framework we adopt for this integration-free approach is based on the infinitesimal generator associated to the underlying continuous-time dynamical system.

The construction of effective reduced equations (point 3 above), requires the estimation of coordinates in which the fast and slow dynamics operate. In the situation where there is an attracting slow manifold, existing numerical methods to determine the slow manifold include [5–9] (see [10] for a recent review). Most of these methods rely on the existence of some *attracting* slow manifold towards which transient fast dynamics is approaching along fast fibres. Here we consider the situation of multiscale systems whose asymptotic behaviour does not necessarily occur on an *attracting* slow manifold. In contrast to methods which determine the fast fibres locally, we instead globally estimate the nonlinear foliation of fast fibres.

Once slow and fast coordinates are established, it is a further challenge to identify their dynamics. There are two approaches; either to devise an effective numerical method which allows for the reliable simulation of the slow coordinates or to construct closed equations for the slow coordinates. The first avenue has been successfully pursued by numerical methods such as the equation-free method [8] and the heterogeneous multiscale method [11,12] which employ short finely resolved bursts of the full dynamics to numerically estimate the averaged slow vector field which then subsequently may be propagated with a large time step. Here we tackle the second avenue of determining the slow dynamics explicitly without the need for temporally resolving the fast dynamics at each step. To compute reduced equations on the (in general, non-unique) slow coordinates, we nonlinearly project local computations along the fast fibres. Our approach does not rely on any temporal integration to estimate the reduced equations. Hence it does not suffer from possible sensitivity of these estimates to the choice of the length of the fast bursts. For example, in the case where the fast dynamics itself involves transitions between metastable states, a short temporal sampling of the full dynamics might not be sufficient to capture the fast invariant measure. This would then bias the averaged slow vector field.

In Section 2 we briefly review the notion of generators of transfer and Koopman operators. Section 3 introduces a trajectory-free test for multiscale behaviour. The degree of time-scale separation is estimated in algorithms described in Sections 4 and 5. A method to determine the reduced slow dynamics from a multiscale system without relying on statistics obtained from long time-integrations is given in Section 6. The algorithms are tested in numerical simulations in Section 7. We conclude with a discussion in Section 8.

2. Generators

We describe our methodology for Itō drift–diffusion processes, as these are a large and flexible class of dynamical systems, and the spectral properties of the corresponding transfer operators are relatively straightforward. Consider a drift–diffusion process

$$d\zeta_i = \mu_i\, dt + \sum_{k=1}^{\ell} \sigma_{ik}\, dW_k \quad \text{with } i = 1, \ldots, d \tag{1}$$

defined on a subset $\mathcal{Z}$ of $\mathbb{R}^d$ where $l \leq d$ and each W_k for $k = 1, \ldots, \ell$ represents an independent Wiener process. Given a probability density function at time $t = 0$, the density at future times is determined by the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho$$

where

$$\mathcal{L}\rho = -\sum_{i=1}^{N} \frac{\partial}{\partial z_i}\big[\mu_i\, \rho\big] + \frac{1}{2}\sum_{i,j=1}^{N} \frac{\partial^2}{\partial z_i\, \partial z_j}\big[D_{ij}\, \rho\big]. \tag{2}$$

The second order differential operator $\mathcal{L}$ is called the *Fokker–Planck operator*. At each point z in the phase space $\mathcal{Z}$, the vector $\mu(z) \in \mathbb{R}^d$ represents the drift of the process and the positive semi-definite matrix $D(z) = \sigma(z)\sigma(z)^\top \in \mathbb{R}^{d\times d}$ the diffusion. The operator $\mathcal{L}$ generates a family of operators $e^{t\mathcal{L}}$ for $t \geq 0$ such that $e^{s\mathcal{L}}e^{t\mathcal{L}} = e^{(s+t)\mathcal{L}}$ and $\lim_{t\to 0} \frac{1}{t}[e^{t\mathcal{L}} - Id] = \mathcal{L}$. If $\rho \in L^1(\mathcal{Z})$ is an initial probability density, then $e^{t\mathcal{L}}(\rho)$ is the density after time t. Thus, $e^{t\mathcal{L}}$ may be thought of as a transfer operator defined on $L^1(\mathcal{Z})$. Since the underlying system is a random dynamical system and $e^{t\mathcal{L}}$ represents an average over all possible random paths, it is an annealed transfer operator [13].

We now consider a setting where the Fokker–Planck operator has compact resolvent. Suppose the domain $\mathcal{Z}$ is a compact subset of $\mathbb{R}^d$ with piecewise smooth boundary. We also allow periodic boundary conditions, such as systems defined on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, so long as the fundamental domain is compact with smooth boundary. Under such assumptions, the Lebesgue measure of $\mathcal{Z}$ is finite, and with respect to this measure, the Hilbert space $L^2(\mathcal{Z})$ is a subset of $L^1(\mathcal{Z})$. Further assume that the operator is *uniformly elliptic*, which holds if the matrix $D(z)$ is positive definite for every $z \in \mathcal{Z}$. Results in the theory of partial differential equations then imply that the Fokker–Planck operator defined on $L^2(\mathcal{Z})$ has compact resolvent. See [14, Chapter 7] for further details and proofs. The condition of uniform ellipticity can in some cases be replaced with the weaker condition of hypo-ellipticity; see [14–16].

Assuming the resolvent is compact, the spectrum of the operator then consists of a countable set of eigenvalues $\{\lambda_k\}_{k=0}^{\infty}$ which, when ordered by the convention

$$0 = \operatorname{Re}\lambda_0 \geq \operatorname{Re}\lambda_1 \geq \operatorname{Re}\lambda_2 \geq \cdots, \tag{3}$$

satisfy $\lim_{k\to\infty} \operatorname{Re}\lambda_k \to -\infty$. As a consequence, for each $t > 0$ the operator $e^{t\mathcal{L}}$ is compact with eigenvalues $e^{t\lambda_k}$ tending to zero as $k \to \infty$. The invariant density ρ_0 of the system is an eigenfunction of $\mathcal{L}$ associated to the eigenvalue $\lambda_0 = 0$.

The Kolmogorov backward equation is given by $\frac{\partial f}{\partial t} = \mathcal{L}^* f$ where the adjoint of the Fokker–Planck operator is given by

$$\mathcal{L}^* f = \sum_{i=1}^{N} \mu_i \frac{\partial f}{\partial z_i} + \frac{1}{2}\sum_{i,j=1}^{N} D_{ij} \frac{\partial^2 f}{\partial z_i\, \partial z_j}. \tag{4}$$

This adjoint operator generates a family of operators $K_t := e^{t\mathcal{L}^*} = (e^{t\mathcal{L}})^*$ for $t \geq 0$. If $f \in L^2(\mathcal{Z})$, then $K_t(f) \in L^2(\mathcal{Z})$ is given by $K_t(f)(z) = \mathbb{E}f(\zeta(t))$ where the expectation is over all paths $\zeta(t)$ in the drift–diffusion process which satisfy $\zeta(0) = z$. The operator K_t may therefore be regarded as an annealed Koopman or composition operator.

The operators $\mathcal{L}^*$ and $\mathcal{L}$ share the same eigenvalues λ_k. Consider the eigenfunction $\psi_k \in L^2(\mathcal{Z})$ of $\mathcal{L}^*$ associated to λ_k. This function ψ_k is an observable which evolves according to

$$K_t\psi_k = e^{t\lambda_k}\psi_k$$

and therefore decays to zero at the rate given by $|e^{t\lambda_k}|$ for $t > 0$. In general, the eigenfunctions of $\mathcal{L}^*$ associated to eigenvalues with real part closest to zero are the observables of the system which

decay at the slowest speeds possible. The eigenvalues λ_k for small k therefore correspond to the speed of the slowest dynamics present in the system. These ideas have been exploited to identify almost-invariant sets [17–19] for deterministic and annealed random dynamics, and coherent sets [20] for quenched random or time-dependent dynamics.

Suppose that the system under study is a multiscale system. That is, assume there is a value $0 < \epsilon \ll 1$ representing the time-scale separation and transverse directions of fast and slow dynamics such that the components of the drift and diffusion are on the order of $O(\epsilon^{-1})$ in the fast direction and $O(1)$ in the slow direction. Further suppose that λ_k is an eigenvalue of $\mathcal{L}^*$ such that $|\operatorname{Re}\lambda_k| \ll \epsilon^{-1}$. Then, one may choose an intermediate time $t \gg \epsilon$ such that

$$|\operatorname{Re} t\lambda_k| \ll 1 \Rightarrow e^{t\lambda_k} \approx 1 \Rightarrow K_t\psi_k \approx \psi_k.$$

Since $t \gg \epsilon$, the system after time t has evolved towards equilibrium in the fast dynamics implying that $K_t\psi_k$ is nearly constant along the direction of the fast dynamics. Therefore, ψ_k should be approximately constant in this direction as well. This observation is developed in more detail in [21,22].

Given the ordering in (3), we refer to the eigenvalues λ_k for small k (and therefore small $|\operatorname{Re}\lambda_k|$) as the *leading* eigenvalues with the associated *leading* eigenfunctions ψ_k. By the above observations, in a multiscale system, the level sets of a leading eigenfunction ψ_k closely approximate *fast fibres*, submanifolds of the system along which the fast dynamics occurs. This property is used in the following sections to develop algorithms which test for multiscale behaviour.

3. An algorithm to test for multiscale behaviour

We now use these properties of the Fokker–Planck operator and its adjoint to develop an algorithm to test for multiscale behaviour. We focus on the case of stochastic differential equations (SDEs) where the slow dynamics is one-dimensional. The only information used by the algorithm are functions μ and D defining the drift and diffusion of the process. In particular, we do not assume any a priori knowledge of the slow or fast directions (if they exist) or that these directions align with the coordinates in which the system is defined. We first outline the idea behind the algorithm before providing the actual algorithm.

As established in the previous section, the level sets of the leading eigenfunctions ψ_k of the adjoint Fokker–Planck operator $\mathcal{L}^*$ contain information about the multiscale behaviour of the dynamics. If the system has multiscale behaviour, these level sets will approximate a fast fibre of the system. We use the eigenfunction ψ_1 and numerically compute a fibre F as a connected component of a level set $\psi_1^{-1}(\{s\})$ (cf. step 3 in Algorithm 1). If the leading eigenfunctions are complex valued, we may choose without loss of generality the real (or imaginary) part of the eigenfunctions to define the fibres (in Section 7 we will give such an example). There is a significant history of computing eigenfunctions of the Fokker–Planck operator or its adjoint numerically [23,18] which we will employ to determine eigenfunctions ψ_k of $\mathcal{L}^*$ associated to the eigenvalues λ_k for small k (cf. step 2 in Algorithm 1).

If the system actually exhibits multiscale dynamics and the fibre F approximates a fast fibre, then the speed of the dynamics will be much faster along the fibre than transverse to it. Therefore, one expects in this case that the drift and the diffusion along one or both of the averages tangent to the fibre F will be significantly larger than those corresponding to the direction normal to F. We therefore calculate the drift and diffusion coefficients for a set of points on the fibre F and then subsequently use their averages along the fibre to probe for fast dynamics along the fibre and slow dynamics transverse to the fibre (this is achieved in steps 4 and 5 of Algorithm 1, respectively).

Computing the drift and diffusion coefficients $\mu^{\text{nor}}(z)$, $\mu^{\tan}(z)$, $D^{\text{nor}}(z)$, and $D^{\tan}(z)$ normal and tangent to the fibre at a finite set of points Q uniformly distributed over F constitutes the most complicated step of the algorithm and is explained in further detail below in Section 3.1.

The most natural way to compute the averages of the drift and diffusion coefficients $\mu^{\text{nor}}(z)$, $\mu^{\tan}(z)$, $D^{\text{nor}}(z)$, and $D^{\tan}(z)$ along the fibre is using the invariant density along the fibre. As an approximation for the invariant density of the fast dynamics on the fibre, we use the invariant density function ρ_0 of the full multiscale system and simply restrict ρ_0 to F. The property that ρ_0 is invariant is equivalent to $\mathcal{L}\rho_0 = 0$. Therefore, we estimate ρ_0 by solving numerically for the eigenfunction of $\mathcal{L}$ whose associated eigenvalue is closest to zero (cf. step 1 of Algorithm 1). The measure on F can be approximated by a measure supported on a finite set. We choose a finite set of points Q uniformly distributed over F and define weights $w_z \in \mathbb{R}$ such that $\sum_{z\in Q} w_z\delta_z$ approximates the density ρ_0 on F. Here, δ_z is the Dirac measure at z and $w_z := c\rho_0(z)$ where the scaling constant c is defined such that $\sum_{z\in Q} w_z = 1$ (cf. step 4 of Algorithm 1).

As a criterion for the presence or absence of multiscale behaviour we then compare the averages along the fibre of the absolute values of the normal and tangential drift and diffusion coefficients; if the tangential components are much larger than the perspective normal ones, this is indicative of F being a fast fibre and the system exhibiting multiscale behaviour (cf. step 7).

3.1. Computing normal and tangent components

We now explain step 5 of Algorithm 1 in detail. To compute the values $\mu^{\text{nor}}(z)$, $\mu^{\tan}(z)$, $D^{\text{nor}}(z)$, and $D^{\tan}(z)$ one must split the dynamics at a point $z \in F$ into the dynamics along the fibre F and dynamics normal to F. There is a subtle issue caused by the diffusion. If the embedding of the fibre into the overall phase space is non-linear, then diffusion along the fibre will generate a drift term for the full system. As a simple example of this, consider the embedding of $\mathbb{R}$ into $\mathbb{R}^2$ given by $u \mapsto (u, u^2)$ and consider a (drift-free) diffusion process $du = dW$ where W is the standard Wiener process on $\mathbb{R}$. For a point (x, y) on this embedded curve, Itō's lemma [24] gives

$$dx = du = dW,$$

$$dy = d(u^2) = 2u\,du + \frac{1}{2}2\,du^2 = 2x\,dW + dt.$$

In the notation of (1), this corresponds to a drift vector $\mu = (0, 1)$ at every point and a diffusion matrix $D = \sigma\sigma^\top$ with $\sigma = (1, 2x)^\top$ at each point (x, y) on the parabola. Notice that while σ is everywhere tangent to the parabola, the drift vector μ always points in the vertical y direction and is not tangent to the parabola. To give an intuition for this, one can consider a point mass at $(x, y) = (0, 0)$ at time $t = 0$. As time progresses, the mass spreads out along the parabola $\{(x, y) : y = x^2\}$ and at each time $t > 0$ the centre of mass (as calculated in $\mathbb{R}^2$) is at the point $(0, t)$ off of the parabola. Because of this effect of diffusion along a fibre leading to drift for the full system, we must apply a non-linear change of coordinates to determine the components of the drift and diffusion both normal and tangent to the fibre.

This may be done in two steps as depicted in Fig. 1. Let $U \subset \mathcal{Z}$ be a neighbourhood of the point $z \in F$. First, find a linear change of coordinates $A : \mathbb{R}^d \to \mathbb{R}^d$, given by a $d \times d$ matrix, so that $A(F \cap U)$ is equal to the graph $\{(v, g(v)) : v \in V\}$ of a function $g : V \to \mathbb{R}$ for some $V \subset \mathbb{R}^{d-1}$. As F is the level set of an eigenfunction, such a graph exists due to the implicit function theorem so long as the gradient of the eigenfunction is non-zero at z.

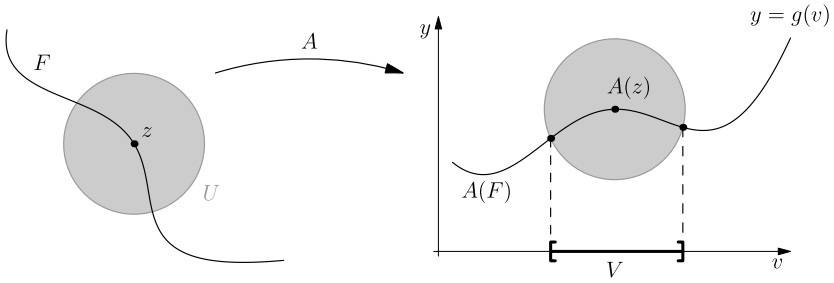

Fig. 1. Once a fast fibre F is identified, a linear transformation A is applied so that $A(F)$ in a neighbourhood of $A(z)$ may be expressed as the graph $\{(v, g(v)) : v \in V\}$ of a function $g : V \to \mathbb{R}$ where $V \subset \mathbb{R}^{d-1}$. A non-linear transformation can then be applied which takes $A(F)$ to the flat subset $V \times \{0\} \subset \mathbb{R}^d$. In these coordinates, the tangent and normal components of the dynamics can be analysed directly.

A subsequent non-linear change of coordinates $(v, y) \mapsto (v, y - g(v))$ flattens the graph out, mapping $A(F)$ to the subset $V \times \{0\} \subset \mathbb{R}^{d-1} \times \mathbb{R}$. After transforming the dynamics to lie in the hyperplane $\mathbb{R}^{d-1} \times \{0\}$, it is straightforward to isolate the normal and tangent components of the diffusion process.

For completeness, we give here the formulas for transforming the drift and diffusion. These formulas can be derived either by Itō calculus on (1) or by applying a deterministic change of variables to the linear operator $\mathcal{L}^*$. For a linear change of coordinates $A : \mathbb{R}^d \to \mathbb{R}^d$ the drift vector $\mu(z)$ at z is mapped to $A\mu(z)$ now at the point $A(z)$ and the diffusion matrix $D(z)$ is mapped to $AD(z)A^\top$. In what follows, we write μ^A and D^A for drift and diffusion functions after this linear transformation.

Now consider the change of coordinates $(v, y) \mapsto (v, y - g(v))$ which in all d coordinates may be written as

$$(z_1, \ldots, z_{d-1},\ z_d) \mapsto (z_1, \ldots, z_{d-1},\ z_d - g(z_1, \ldots, z_{d-1})).$$

This transforms the drift–diffusion process given by μ^A and D^A to one given by $\hat{\mu}$ and $\hat{D}$ where

$$\hat{\mu}_i = \mu^A_i \quad \text{for } i < d, \tag{5}$$

$$\hat{\mu}_d = \mu^A_d - \sum_{i<d} \mu^A_i g_i + \frac{1}{2} \sum_{i,j<d} D^A_{ij} g_{ij} \tag{6}$$

and

$$\hat{D}_{ij} = D^A_{ij} \quad \text{for } i, j < d, \tag{7}$$

$$\hat{D}_{id} = D^A_{id} - \sum_{j<d} D^A_{ij} g_j \quad \text{for } i < d, \text{ and} \tag{8}$$

$$\hat{D}_{dd} = D^A_{dd} - 2 \sum_{i<d} D^A_{id} g_i + \sum_{i,j<d} D^A_{ij} g_i g_j. \tag{9}$$

In the above equations, g_i and g_{ij} are the partial derivatives of g with respect to x. Since only first and second-order partial derivatives appear above, only a second-order approximation of g is necessary in order to compute the new values.

We propose the following method to realize these two changes of variables numerically. Suppose the fibre F in a neighbourhood of a point z is represented numerically by a collection of points $\{q_n\}$ near z. First, perform linear regression on this set of points to get a hyperplane $H \subset \mathbb{R}^d$ such that the set $H + z$ is tangent to F at z. Using a numerical orthogonalization process such as the QR decomposition, construct an orthonormal basis $\{u_1, \ldots, u_{d-1}, u_d\}$ of $\mathbb{R}^d$ such that $u_1, \ldots, u_{d-1} \in H$. This basis then gives the rows of an orthogonal $d \times d$ matrix representing A.

Next, write each point $A(q_n)$ in the form (v_n, y_n) with $v_n \in \mathbb{R}^{d-1}$ and $y_n \in \mathbb{R}$.[1] By polynomial regression or a similar fitting technique, find a polynomial $p_g : \mathbb{R}^{d-1} \to \mathbb{R}$ such that $p_g(v_n) \approx y_n$ and where this approximation is to at least second order. Using p_g in place of g, compute $\hat{\mu}$ and $\hat{D}$ as above.

If the fibre F truly corresponds to a fast fibre in a multiscale system, one expects the drift and diffusion terms to be large for coordinates tangent to F and small otherwise. As a test of this, one can check that the values of $\hat{\mu}_d$ are small in comparison to the values $\hat{\mu}_i$ with $1 \le i < d$ and similarly that $\hat{D}_{id}$ and $\hat{D}_{dd}$ are small in comparison to $\hat{D}_{ij}$ with $1 \le i, j < d$.

To make this comparison precise, we wish to calculate scalar quantities corresponding to the amounts of drift and diffusion both tangent and normal to the fibre. For the one-dimensional normal direction, the component $\hat{\mu}_d$ corresponds to the normal drift and $\hat{D}_{dd}$ the normal component of the diffusion. Therefore, define $\mu^{\text{nor}}(z) = |\hat{\mu}_d|$ and $D^{\text{nor}}(z) = |D_{dd}|$.

For the tangent direction, define $\mu^{\tan}(z) = \sqrt{\hat{\mu}_1^2 + \cdots + \hat{\mu}_{d-1}^2}$ which is the norm of the vector $(\hat{\mu}_1, \ldots, \hat{\mu}_{d-1})$. Define $D^{\tan}(z)$ as the largest eigenvalue of the $(d-1) \times (d-1)$ submatrix of $\hat{D}$. This is equivalent to

$$D^{\tan}(z) = \sup\{u^\top \hat{D} u :\ u \in \mathbb{R}^{d-1} \times \{0\},\ \|u\| = 1\}$$

and corresponds to the largest amount of diffusion in any direction tangent to $\mathbb{R}^{d-1} \times \{0\}$. Since $\hat{D}$ is symmetric, the value $D^{\tan}(z)$ is also the largest singular value of the submatrix and can easily be computed numerically (say by the `norm` function in MATLAB). In the special case where $d = 2$, the definitions reduce down to $\mu^{\tan}(z) = |\hat{\mu}_1|$ and $D^{\tan}(z) = |\hat{D}_{11}|$.

In general, the definitions of $\mu^{\tan}$ and $D^{\tan}$ depend on the choice of A. If we impose the restriction that A is an isometry then one can show that $\mu^{\tan}$ and $D^{\tan}$ depend only on the subspace $A^{-1}(\mathbb{R}^{d-1} \times 0)$ and not on the choice of A itself. Because of this property, we only consider the case where A is an isometry. This property holds exactly when the matrix representing A is orthogonal. The numerical methods given above for constructing this matrix ensure that it is orthogonal.

We now summarize the algorithm.

3.2. Local test for multiscale dynamics

Algorithm 1. 1. Given an Itō drift–diffusion process on a subset $\mathcal{Z}$ of $\mathbb{R}^d$, compute the invariant density ρ_0 by numerically solving $\mathcal{L}\rho_0 = 0$.

[1] Here and later in the paper, we adopt the convention that subscripts with the letters m and n refer to distinct points in the phase space $\mathcal{Z} \subset \mathbb{R}^d$ and subscripts with i and j denote the coordinates of a point. That is, $z_i, z_j \in \mathbb{R}$ for $1 \le i, j \le d$ would refer to the coordinates of a point $z = (z_1, \ldots, z_d) \in \mathcal{Z}$ whereas q_m and q_n would refer to points in $\mathcal{Z}$. Throughout the paper, the subscript k is used only to index the eigenvalues and eigenfunctions of a linear operator.

2. Compute the leading eigenvalues λ_k and eigenfunctions ψ_k of the adjoint operator $\mathcal{L}^*$.
3. For the eigenvalue $\lambda_1 \neq 0$ and associated eigenfunction ψ_1, compute a fibre F which is the connected component of a level set $\psi_1^{-1}(\{s\})$ for some $s \in \mathbb{R}$.
4. Define a finite subset $Q \subset F$ consisting of points uniformly sampled over F. Define weights $w_z := c\,\rho_0(z)$ for $z \in Q$ with $c > 0$ chosen such that $\sum_{z\in Q} w_z = 1$.
5. For each point $z \in Q$:
   (a) Using linear regression on points of F near z, find a hyperplane H such that $H + z$ is approximately tangent to F at z.
   (b) Construct an isometry $A : \mathbb{R}^d \to \mathbb{R}^d$ taking H to $\mathbb{R}^{d-1} \times \{0\}$. For a neighbourhood U of z, the image $A(U \cap F)$ is equal to a graph $\{(v, g(v)) : v \in V\}$ of a function $g : V \to \mathbb{R}$ with $V \subset \mathbb{R}^{d-1}$.
   (c) Approximate the first and second derivatives of g using polynomial regression on points in $A(F)$ near $A(z)$.
   (d) Compute the drift vector $\hat{\mu}$ and diffusion matrix $\hat{D}$ at the point $(v, g(v)) = A(z)$ using Eqs. (5)–(9). From this, compute the normal and tangent components $\mu^{\mathrm{nor}}(z)$, $\mu^{\mathrm{tan}}(z)$, $D^{\mathrm{nor}}(z)$, and $D^{\mathrm{tan}}(z)$ as explained above.
6. Compute the average values

$$\mu^{\mathrm{nor}}_{\mathrm{avg}} = \sum_{z\in Q} w_z\,\mu^{\mathrm{nor}}(z), \qquad \mu^{\mathrm{tan}}_{\mathrm{avg}} = \sum_{z\in Q} w_z\,\mu^{\mathrm{tan}}(z),$$
$$D^{\mathrm{nor}}_{\mathrm{avg}} = \sum_{z\in Q} w_z\,D^{\mathrm{nor}}(z), \qquad D^{\mathrm{tan}}_{\mathrm{avg}} = \sum_{z\in Q} w_z\,D^{\mathrm{tan}}(z).$$

7. If either of $\mu^{\mathrm{tan}}_{\mathrm{avg}}$ or $D^{\mathrm{tan}}_{\mathrm{avg}}$ is significantly larger than both $\mu^{\mathrm{nor}}_{\mathrm{avg}}$ and $D^{\mathrm{nor}}_{\mathrm{avg}}$, this is evidence that the full system has multiscale behaviour.

In the above algorithm, steps 3 through 7 may be repeated for a number of distinct fibres in order to test for multiscale behaviour throughout the domain.

4. Estimating the time-scale separation

We now propose a further test for multiscale behaviour which also estimates the *magnitude* of the time-scale separation between the slow and fast dynamics. This estimation is achieved by comparing the spectrum of the operator $\mathcal{L}$ on the full system to the spectrum of an operator defined by dynamics along the fibre F.

Recall from Section 2 that the leading eigenvalues of $\mathcal{L}$ (which are the same as those of its adjoint $\mathcal{L}^*$) correspond to the slowest rates of decay for observables under the dynamics of the system. Therefore, they give an estimate of the speed of the slow dynamics which can be computed from the full system.

We now define a Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$ for the dynamics tangent to the fast fibre F. For a multiscale system, the dynamics associated to $\widehat{\mathcal{L}}_{\tan}$ is entirely fast and does not capture any of the slow dynamics. Therefore, its leading eigenvalues $\widehat{\lambda}_k$ correspond to the speed of the fast dynamics alone.

We are assuming that the slow dynamics evolves on a time-scale of $O(1)$ and the fast dynamics on a time-scale of $O(1/\epsilon)$ for some unknown value ϵ. The leading eigenfunctions ψ_k of $\mathcal{L}^*$ are observables which decay to zero at the rate of the slow dynamics of the systems. Because of this, the real part of the associated eigenvalues λ_k are $O(1)$. The leading eigenfunctions of $\widehat{\mathcal{L}}^*_{\tan}$ are observables along the fibre F which decay at a speed associated to the fast dynamics and so the associated eigenvalues $\widehat{\lambda}_k$ are $O(1/\epsilon)$. The ratio of the real parts of the leading λ_k and the leading $\widehat{\lambda}_k$ then gives a quantitative estimate of the time-scale separation between the slow and fast variables. See also the discussion in [21,22].

We now give an algorithm for determining these values $\widehat{\lambda}_k$ numerically. Let F be a fibre as computed in Algorithm 1. Further suppose that F is such that it can be represented *globally* as the graph of a function. That is, after a linear change of coordinates $A : \mathbb{R}^d \to \mathbb{R}^d$, the image $A(F)$ is equal to the set $\{(v, g(v)) : v \in V\}$ for some function $g : V \to \mathbb{R}$ defined on a subset V of $\mathbb{R}^{d-1}$. One may then define a drift–diffusion process on V where the drift and diffusion components are given by $\hat{\mu}_i$ and $\hat{D}_{ij}$ for $1 \leq i, j \leq d-1$ as defined in the previous section. Using this data, the leading eigenvalues $\widehat{\lambda}_k$ for the operator $\widehat{\mathcal{L}}_{\tan}$ may be computed numerically.

We summarize the algorithm for the estimation of the time-scale separation as

Algorithm 2A. 1. Given an Itō drift–diffusion process on a subset $\mathcal{Z}$ of $\mathbb{R}^d$, compute the leading eigenvalues λ_k and eigenfunctions ψ_k of the adjoint $\mathcal{L}^*$ of the Fokker–Planck operator.
2. For the eigenvalue $\lambda_1 \neq 0$ and associated eigenfunction ψ_1, compute a fibre F which is the connected component of a level set $\psi_1^{-1}(\{s\})$ for some $s \in \mathbb{R}$.
3. Find a linear change of coordinates $A : \mathbb{R}^d \to \mathbb{R}^d$ such that $A(F)$ is equal to a graph of a function $g : V \to \mathbb{R}$ with $V \subset \mathbb{R}^{d-1}$.
4. Define a drift–diffusion process on V where for each point $v \in V$ the drift is given by the first $d-1$ components of $\hat{\mu}(z)$ and the diffusion by the $(d-1) \times (d-1)$ submatrix of $\hat{D}(z)$ where $z = A^{-1}(v, g(v))$.
5. Numerically compute the leading eigenvalues $\widehat{\lambda}_k$ of the Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$ (and/or its adjoint) for this drift–diffusion process on V.
6. Compare the eigenvalues $\widehat{\lambda}_k$ computed in step 5 to the eigenvalues λ_k computed in step 1. If the ratios $\mathrm{Re}(\lambda_k)/\mathrm{Re}(\widehat{\lambda}_k)$ are all of similar magnitudes and are close to zero, then this implies multiscale behaviour and the ratio gives an estimate of the time-scale separation of the slow and fast dynamics.

For this algorithm, it is not necessary to estimate the partial derivatives of g. This is because these derivatives do not occur in Eqs. (5) and (7).

Ideally, Algorithms 1 and 2A should both be performed on a given system to test for multiscale behaviour. Algorithm 1 tests that on a fibre F the tangent components of the drift and diffusion dominate the normal components. Therefore, it justifies approximating the fast dynamics by a drift–diffusion process restricted to the fibre and defined purely by these tangent components. Algorithm 2A then uses the eigenvalues of the resulting operator $\widehat{\mathcal{L}}_{\tan}$ associated to this process in order to estimate the speed of the dynamics restricted to the fibre.

Algorithm 2A assumes that the fibre F can be expressed globally as the graph of a function. If the shape of F is such that finding a graph is not possible, then one could instead estimate these eigenvalues by using methods of trajectory integration and projection to the fibre as described in [22]. This alternative method does not follow the general trajectory-free framework proposed in this work. It may therefore be slower in general and subject to the concerns listed in the introduction. In the next section, we give a variant of Algorithm 2A which is always applicable in the specific case where the slow and fast directions are one-dimensional.

5. Parameterizing by arc length

In the case that the fast fibre is one-dimensional (and the full system is therefore two-dimensional), an alternative to expressing the fibre as a graph is to parameterize the fibre by arc length. This is always possible, as topologically a one-dimensional fibre F will either be a line or a circle.

Given a representation of the curve F, compute a sequence of points $\{q_n\}$ such that each point is at a uniform distance from the

previous. To analyse the drift and diffusion at q_n, first approximate the tangent to the curve at q_n by the vector $q_{n+1} - q_{n-1}$. Apply a rotation $A : \mathbb{R}^2 \to \mathbb{R}^2$ such that $A(q_{n+1} - q_{n-1})$ is roughly horizontal. Writing $(v_n, y_n) = A(q_n)$, apply a regression method to find a polynomial p_g such that $y_m \approx p_g(v_m)$ for points (v_m, y_m) near (v_n, y_n). By adjusting the angle of the rotation A, one may further assume that the first derivative p_g' satisfies $p_g'(v_n) = 0$. Using the rotation A composed with the change of coordinates mapping (v, y) to $(v, y - p_g(v))$ compute the drift and diffusion data tangent and normal to the curve. Applying (5)–(9) in this specific case and using that $p_g'(v_n) = 0$, the resulting coefficients are

$$\hat{\mu}_1 = \mu_1^A, \qquad \hat{\mu}_2 = \mu_2^A - \frac{1}{2} D_{11}^A p_g'', \qquad \hat{D}_{11} = D_{11}^A,$$
$$\hat{D}_{12} = D_{12}^A, \qquad \hat{D}_{22} = D_{22}^A.$$

For $v > v_n$ the arc length of the graph $\{(u, p_g(u)) : u \in [v_n, v]\}$ is given by

$$\ell(v) = \int_{v_n}^{v} \sqrt{1 + [p_g'(u)]^2}\, du.$$

Using $p_g'(v_n) = 0$, the first and second derivatives satisfy

$$\left.\frac{d\ell}{dv}\right|_{v=v_n} = 1 \quad \text{and} \quad \left.\frac{d^2\ell}{dv^2}\right|_{v=v_n} = 0.$$

If we take the derivative of v with respect to itself, then

$$\left.\frac{dv}{dv}\right|_{v=v_n} = 1 \quad \text{and} \quad \left.\frac{d^2 v}{dv^2}\right|_{v=v_n} = 0$$

as well. This implies that two different parameterizations of the curve – one by arc length and the other by v – agree up to second order at v_n. Therefore, the Fokker–Planck operators coincide for the two parameterizations with the same drift and diffusion coefficients at the point (v_n, y_n). Using this data, one can find for any point on F the components tangent to the curve. One can therefore define a drift–diffusion process on the curve F as parametrized by arc length and numerically compute the leading eigenvalues of the corresponding Fokker–Planck operator as before.

We present this as a variant of Algorithm 2A to estimate the time-scale separation:

Algorithm 2B. 1. Given an Itō drift–diffusion process on a subset $\mathcal{Z}$ of $\mathbb{R}^2$, compute the leading eigenvalues λ_k and eigenfunctions ψ_k of the adjoint $\mathcal{L}^*$ of the Fokker–Planck operator.
2. For the eigenvalue $\lambda_1 \neq 0$ and associated eigenfunction ψ_1, compute a fibre F which is the connected component of a level set $\psi_1^{-1}(\{s\})$ for some $s \in \mathbb{R}$.
3. Define a uniformly spaced finite sequence of points $\{q_n\}$ along F.
4. Using the techniques in this section, calculate for each $\{q_n\}$ the components $\hat{\mu}_1$ and $\hat{D}_{11}$ which are tangent to F. These values define a drift–diffusion process on F.
5. Numerically compute the leading eigenvalues $\widehat{\lambda}_k$ of the Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$ (and/or its adjoint) for this drift–diffusion process on F.
6. Compare the eigenvalues $\widehat{\lambda}_k$ computed in step 5 to the eigenvalues λ_k computed in step 1. If the ratios $|\lambda_k|/|\widehat{\lambda}_k|$ are all of similar magnitudes and are close to zero, then this implies that multiscale behaviour exists and the ratio gives an estimate of the timescale separation of the slow and fast dynamics.

6. Determining the reduced dynamics

Once the existence of multiscale dynamics has been established, one further step is to compute the reduced slow dynamics. We give here a simple technique to perform the reduction to a lower dimensional SDE. As with the other methods developed in this paper, the algorithm does not rely on trajectory integration.

First, consider a multiscale system where model reduction is applicable. That is, following the formalism given in [25,22], there is a projection $\mathcal{P} : \mathcal{Z} \to \mathcal{X}$ from the domain of the full system $\mathcal{Z}$ to a smaller dimensional space $\mathcal{X}$ and an SDE on $\mathcal{X}$ such that orbits of the reduced system on $\mathcal{X}$ resemble projected orbits of the full system. Our goal is to reconstruct this reduced dynamics.

As before, let $\mathcal{L}$ denote the Fokker–Planck operator for the full SDE on $\mathcal{Z}$ and let (λ_k, ψ_k) be the eigenpairs for the adjoint $\mathcal{L}^*$. For ease of notation, we assume here that all eigenvalues are simple. Similar reasoning holds in the more general case. Let $\tilde{\mathcal{L}}$ denote the Fokker–Planck operator for the reduced SDE on $\mathcal{X}$ and let $(\tilde{\lambda}_k, \tilde{\psi}_k)$ be the eigenpairs of $\tilde{\mathcal{L}}^*$. Assume both sequences are indexed by decreasing real part as in (3).

Recall that the leading eigenvalues λ_k of $\mathcal{L}^*$ are associated to the speed of the slow dynamics and the eigenfunctions ψ_k are approximately constant along fast fibres. In fact, using operator approximation theory, one may show that for ϵ sufficiently small, the approximations $\lambda_k \approx \tilde{\lambda}_k$ and $\psi_k \approx \tilde{\psi}_k \circ \mathcal{P}$ hold [21].

Suppose $C \subset \mathcal{Z}$ is a smooth curve transverse to the fast fibre which intersects each fibre exactly once (see Fig. 2). Define a projection $P_c : \mathcal{Z} \to C$ by letting $P_c(z)$ be the unique point on C which is on the same fast fibre as z. Then C may be identified with $\mathcal{X}$ and P_c with $\mathcal{P}$. The above approximations then imply that $(\lambda_k, \psi_k|_C)$ is an approximation of the eigenpair $(\tilde{\lambda}_k, \tilde{\psi}_k)$. Moreover, the objects λ_k, ψ_k, and C can all be computed numerically from the operator $\mathcal{L}^*$ for the full system. This means that even though we have no direct way of representing the operator $\tilde{\mathcal{L}}^*$ associated to the reduced dynamics on $\mathcal{X}$, we still have an indirect method of computing its leading eigenpairs.

There are a number of techniques for reconstructing the drift and diffusion of a system from its spectral data. For instance, see the survey [26] and more recent work in [27,21]. We use a method attributed in [26] to S.G. Demoura. Let $\tilde{\mu}$ and $\tilde{D}$ be the (as yet unknown) coefficients of the drift and diffusion for the reduced SDE on C. Consider the formula (4) for the adjoint of the Fokker–Planck operator in the one-dimensional setting along the curve C. One sees that each eigenpair $(\tilde{\lambda}_k, \tilde{\psi}_k)$ of $\tilde{\mathcal{L}}^*$ satisfies the equation

$$\tilde{\lambda}_k \tilde{\psi}_k = \tilde{\mu} \tilde{\psi}_k' + \tfrac{1}{2} \tilde{D} \tilde{\psi}_k'' \tag{10}$$

where the prime denotes derivatives along the curve C. Using $\psi_k|_C$ as an approximation for $\tilde{\psi}_k$, one may estimate the derivatives $\tilde{\psi}_k'$ and $\tilde{\psi}_k''$ numerically. By considering (10) for two distinct eigenpairs, say $(\tilde{\lambda}_1, \tilde{\psi}_1)$ and $(\tilde{\lambda}_2, \tilde{\psi}_2)$, one may solve the linear system of equations

$$\begin{bmatrix} \tilde{\psi}_1' & \frac{1}{2}\tilde{\psi}_1'' \\ \tilde{\psi}_2' & \frac{1}{2}\tilde{\psi}_2'' \end{bmatrix} \begin{bmatrix} \tilde{\mu} \\ \tilde{D} \end{bmatrix} = \begin{bmatrix} \tilde{\lambda}_1 \tilde{\psi}_1 \\ \tilde{\lambda}_2 \tilde{\psi}_2 \end{bmatrix} \tag{11}$$

pointwise along C in order to solve for the drift $\tilde{\mu}$ and diffusion $\tilde{D}$ of the reduced system.

This technique requires an accurate estimate of the first and second derivatives of the eigenfunctions. In some settings, accurate derivative estimation is not possible and more sophisticated techniques must be applied, such as solving a quadratic programming problem based on the eigenfunction equations [28,27,21].

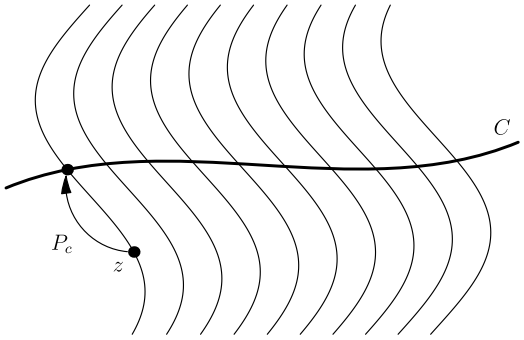

Fig. 2. A depiction of the curve C, which intersects each fast fibre in a single point and the resulting projection $P_c : \mathcal{Z} \to C$ defined by letting $P_c(z)$ be the unique point on C which is on the same fast fibre as z. We define reduced slow dynamics on the curve C.

7. Numerical example

We now apply these techniques to an example studied in [27] and given by

$$\dot{x} = \sin y + \sqrt{1 + \frac{1}{2}\sin y}\,\dot{W}_x, \tag{12}$$

$$\dot{y} = \frac{1}{\epsilon}\left[-y + \sin x\right] + \frac{1}{\sqrt{\epsilon}}\,\dot{W}_y \tag{13}$$

where W_x and W_y are independent Wiener processes. This SDE is defined on $[0, 2\pi] \times \mathbb{R}$ with periodic boundary conditions on x. In order to have an example with nonlinear fast fibres, we consider the system after a change of coordinates taking (x, y) to $(x + \sin(y), y)$. After this change, the system is given by the (admittedly much uglier) equations

$$\dot{x} = \left[\sin y + \frac{\cos y}{\epsilon}\big(\sin(x - \sin y) - y\big) - \frac{\sin y}{2\epsilon}\right] + \sqrt{1 + \frac{1}{2}\sin y}\,\dot{W}_x + \frac{1}{\sqrt{\epsilon}}\cos y\,\dot{W}_y, \tag{14}$$

$$\dot{y} = \frac{1}{\epsilon}\big[\sin(x - \sin y) - y\big] + \frac{1}{\sqrt{\epsilon}}\,\dot{W}_y. \tag{15}$$

Note that terms involving ϵ now occur in the formulas for both $\dot{x}$ and $\dot{y}$ and so these equations have no obvious slow–fast splitting. As in [27] we fix $\epsilon = 10^{-3}$. We now apply the steps of Algorithm 1 to test for multiscale behaviour.

For step 1 of the algorithm, we compute the invariant density ρ_0 by solving numerically for the leading eigenfunction of the Fokker–Planck operator $\mathcal{L}$ using a collocation method [23]. As x is periodic and y is not, we restrict the domain to $[0, 2\pi] \times [-L, L]$ with $L = 5$ and use as basis the tensor product of a Fourier basis in x and a Chebyshev basis in y. A 50×50 grid of points is used for the collocation. Since one expects the density to decay to zero as $|y| \to \infty$, we impose a Dirichlet boundary condition that ρ_0 is zero on $[0, 2\pi] \times \{-L\}$ and $[0, 2\pi] \times \{L\}$.

From a theoretical standpoint, the restriction to $[0, 2\pi] \times [-L, L]$ also ensures that the Fokker–Planck operator has compact resolvent as discussed in Section 2.

For step 2, we solve for the leading eigenpairs (λ_k, ψ_k) of the adjoint operator $\mathcal{L}^*$ using the same domain and basis. However, since the eigenfunctions of $\mathcal{L}^*$ corresponding to the eigenvalue $\lambda_0 = 0$ are given by the constant functions, we impose the boundary condition $\frac{\partial \psi}{\partial y} = 0$ on $[0, 2\pi] \times \{-L\}$ and $[0, 2\pi] \times \{L\}$. The leading eigenvalues are given in the first column of Table 1. After the eigenvalue $\lambda_0 = 0$ (numerically computed as 2.0971×10^{-11}), the eigenvalues with greatest real part are $\lambda_1 = -0.6467 + 0.1097i$ and its conjugate $\lambda_2 = \bar{\lambda}_1$. The eigenfunction ψ_1 associated to λ_1 is complex valued and the real part of this function is plotted in Fig. 3.

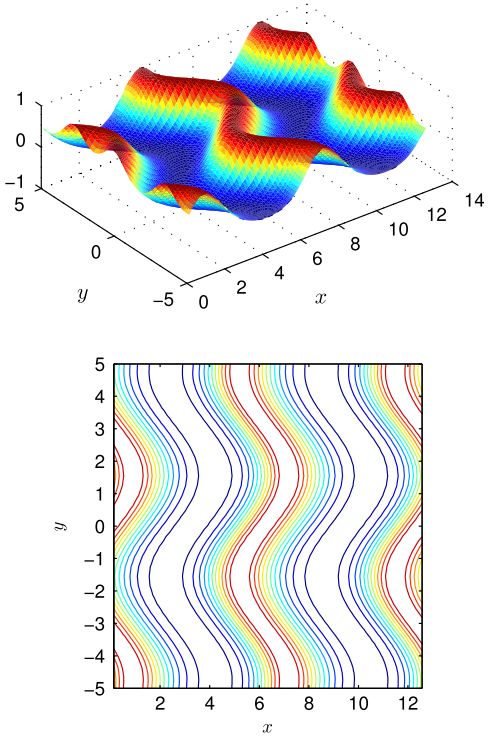

Fig. 3. The real part of the eigenfunction ψ_1 corresponding to the eigenvalue $\lambda_1 = -0.6467 + 0.1097i$ for the adjoint $\mathcal{L}^*$ of the Fokker–Planck operator for the SDE given by (14)–(15) with $\epsilon = 10^{-3}$. The function is plotted for $x \in [0, 4\pi]$ using the 2π-periodicity of the x variable. Also plotted are the resulting level sets, used as a numerical approximation of the fast fibres.

This real part is used to compute level sets which closely approximate fast fibres of the system when multiscale behaviour is present (cf. step 3). For concreteness, we use a fibre F defined by the level set of $\mathrm{Re}\,\psi_1$ which passes through the point $(x, y) = (5, 0)$. Other choices of level sets yield similar results.

Once this fibre F is computed, we define a finite subset Q by sampling points along F where each point is at a distance 0.1 from the previous. Taking weights w_z proportional to $\rho_0(z)$, this defines a measure on the fibre (cf. step 4).

For each point in Q, we construct a change of coordinates (cf. step 5). This change of coordinates yields at each point a new drift vector $\hat{\mu}$ and diffusion matrix $\hat{D}$. The components of the computed $\hat{\mu}$ and $\hat{D}$ are plotted in Figs. 4 and 5. Note that the components $\hat{\mu}_1$ and $\hat{D}_{11}$ are the largest in magnitude, and these correspond to the dynamics tangent to the fibre F.

From the original equations (12) and (13), one sees that the dynamics on a fast vertical fibre in these original coordinates are given by a process with drift $(-y + \sin x)/\epsilon$ and diffusion $1/\sqrt{\epsilon}$. These drift and diffusion formulas can be transformed under the change of coordinates $(x, y) \mapsto (x + \sin(y), y)$ to give exact analytical formulas for the drift $\hat{\mu}_1$ and diffusion $\hat{D}_{11}$ along the fibre when it is parametrized by arc length. These exact functions for $\hat{\mu}_1$ and $\hat{D}_{11}$ are also plotted in Figs. 4 and 5 and agree closely with the computed values. Since these analytical formulas are long and not elucidating, we do not include them here.

To compare the tangent and normal components quantitatively, we apply step 6 of Algorithm 1, computing the weighted averages over the fibre. The resulting averages for the drift are $\mu^{\tan}_{\text{avg}} = 704.27$ and $\mu^{\text{nor}}_{\text{avg}} = 25.165$ and for the diffusion are $D^{\tan}_{\text{avg}} = 1.4809 \times 10^3$ and $D^{\text{nor}} = 1.4216$. Note that $\mu^{\tan}_{\text{avg}}$ and $D^{\tan}_{\text{avg}}$ are both

Table 1
Leading eigenvalues λ_k computed for the adjoint of the Fokker–Planck operator for the SDE given by (14)–(15) with $\epsilon = 10^{-3}$. Also computed are the leading eigenvalues $\widehat{\lambda}_k$ of the Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$ on a fast fibre F. The ratios $\mathrm{Re}(\lambda_k)/\mathrm{Re}(\widehat{\lambda}_k)$ give a computable estimate of the time-scale separation ϵ. Analytically, $\widehat{\lambda}_k$ should be equal to $-k/\epsilon = -k\times 10^3$ and the relative error in computation, as defined in (16), is also given.

k	λ_k	$\widehat{\lambda}_k$	$\mathrm{Re}(\lambda_k)/\mathrm{Re}(\widehat{\lambda}_k)$	err_k
0	2.0971×10^{-11}	-3.3707×10^{-6}	–	–
1	$-0.6467 + 0.1097i$	-9.9275×10^{2}	6.5139×10^{-4}	7.2535×10^{-3}
2	$-0.6467 - 0.1097i$	-2.0325×10^{3}	3.1816×10^{-4}	1.6268×10^{-2}
3	$-2.0508 + 0.2465i$	-3.0931×10^{3}	6.6303×10^{-4}	3.1044×10^{-2}
4	$-2.0508 - 0.2465i$	-4.0126×10^{3}	5.1110×10^{-4}	3.1492×10^{-3}
5	$-4.4543 + 0.3912i$	-4.9614×10^{3}	8.9780×10^{-4}	7.7173×10^{-3}
6	$-4.4543 - 0.3912i$	-5.9946×10^{3}	7.4306×10^{-4}	8.9551×10^{-4}

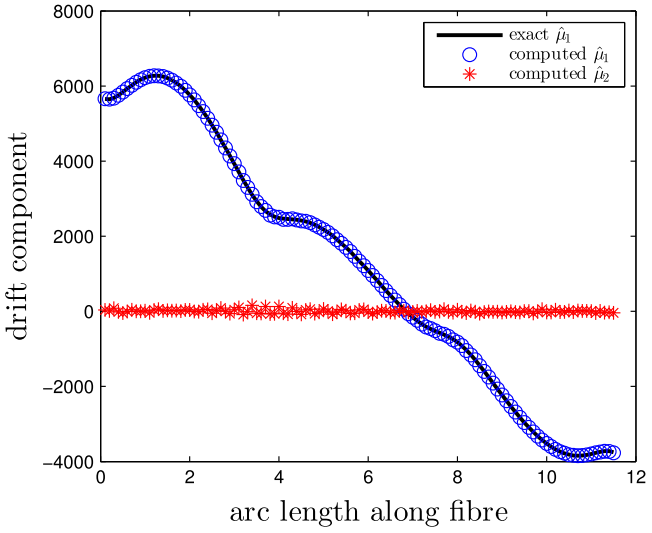

Fig. 4. The drift components computed in coordinates adapted to the fibre. The $\hat{\mu}_1$ component corresponds to drift along the fibre and $\hat{\mu}_2$ to drift normal to the fibre.

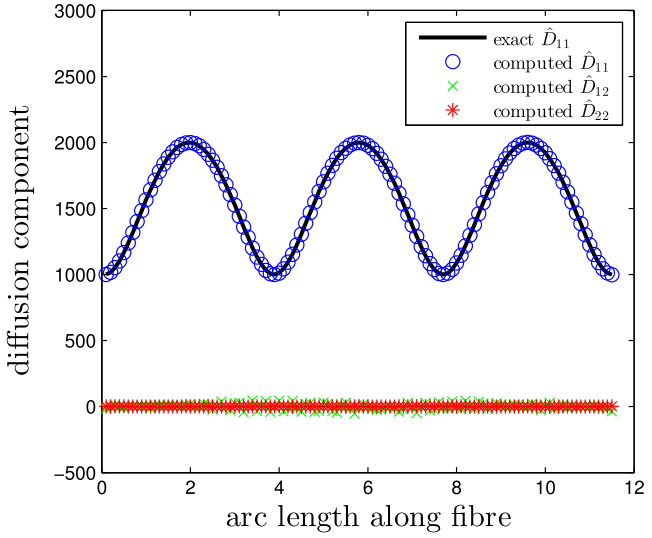

Fig. 5. The diffusion components computed in coordinates adapted to the fibre. The $\hat{D}_{11}$ component corresponds to diffusion along the fibre and $\hat{D}_{12}, \hat{D}_{22}$ to diffusion normal to the fibre.

significant larger than $\mu^{\mathrm{nor}}_{\mathrm{avg}}$ and $D^{\mathrm{nor}}_{\mathrm{avg}}$. By step 7, we conclude that the system has multiscale behaviour.

We now apply Algorithms 2A and 2B to the system. Before discussing the numerics, first note that from the original equations (12) and (13), one sees that the dynamics on a fast vertical fibre in these original coordinates is given by an Ornstein–Uhlenbeck process with drift $(-y + \sin x)/\epsilon$ and diffusion $1/\sqrt{\epsilon}$. For such processes, the eigenvalues of the Fokker–Planck operator can been determined analytically [29]. For (13), the eigenvalues are given exactly by $\lambda_k = -k/\epsilon$ for integers $k \geq 0$. Since the spectrum of the operator is not affected by a change of coordinates, these are also the exact values for the eigenvalues of the operator $\widehat{\mathcal{L}}_{\tan}$ associated to the fast dynamics of the system given by (14)–(15).

In the application of Algorithm 1 above, we computed the values $\hat{\mu}_1$ and $\hat{D}_{11}$ at points uniformly distributed along the fast fibre F. These define a drift–diffusion process along F with associated Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$. Applying Algorithm 2B, we calculate leading eigenvalues $\widehat{\lambda}_k$ of this operator $\widehat{\mathcal{L}}_{\tan}$ numerically. A fourth-order central finite difference method was used. The computed eigenvalues are given in the second column of Table 1. Note that the eigenvalues $\widehat{\lambda}_k$ are all of order 10^3 and the eigenvalues λ_k are of order 1. Applying the final step of Algorithm 2B, the ratios $\mathrm{Re}(\lambda_k)/\mathrm{Re}(\widehat{\lambda}_k)$ are all between 10^{-4} and 10^{-3} and give an estimate of the time-scale separation. This agrees with the value $\epsilon = 10^{-3}$ used in defining the system.

For each computed eigenvalue $\widehat{\lambda}_k$, the relative error

$$\mathrm{err}_k = \frac{|\widehat{\lambda}_k - (-k/\epsilon)|}{|-k/\epsilon|} \tag{16}$$

is also given in the table.

We now consider Algorithm 2A and perform eigenvalue calculations by expressing the fibre globally as a graph. A fast fibre F is of the form $\{(x, y) : x = \sin(y) + c\}$ for some constant c. Therefore, if A is a rotation matrix

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

with θ between 45° and 135°, then $A(F)$ is the graph of a function $g : \mathbb{R} \to \mathbb{R}$. For values of θ at every 5° between 55 and 125, we applied Algorithm 2A with the corresponding A and computed seven leading eigenvalues $\widehat{\lambda}^{(\theta)}_0 > \widehat{\lambda}^{(\theta)}_1 > \cdots > \widehat{\lambda}^{(\theta)}_6$ for the associated Fokker–Planck operator. This computation was achieved by first determining the values $\hat{\mu}_1(z_n)$ and $\hat{D}_{11}(z_n)$ for points $z_n := (v_n, g(v_n)) \in A(F)$ where the v_n are given by 200 evenly spaced numbers across an interval in $\mathbb{R}$. Then, these sampled values of $\hat{\mu}_1$ and $\hat{D}_{11}$ are used to approximate the Fokker–Planck operator $\widehat{\mathcal{L}}_{\tan}$ and the eigenvalues are solved using a fourth order central finite difference method. As noted above, the leading eigenvalues for the actual system are given by $-k/\epsilon$ and the computed values $\widehat{\lambda}^{(\theta)}_k$ agreed with these theoretical values with a relative error less than one percent for all computed θ and $1 \leq k \leq 6$. This shows that Algorithm 2A produces estimates of the eigenvalues $\widehat{\lambda}^{(\theta)}_k$ for the fibre dynamics that are robust with respect to the parameterization the fibre.

Algorithms 1 and 2 have established the presence of multiscale behaviour for the system and determined the time-scale separation. As a final step, we compute the drift and diffusion data for the

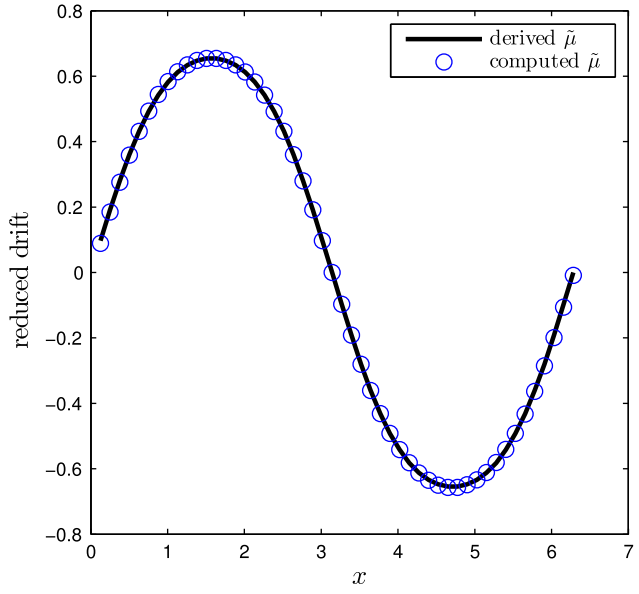

Fig. 6. The drift component $\tilde{\mu}$ computed for the reduced slow dynamics along the line $C = [0, 2\pi] \times \{0\}$. Plotted as circles is the value determined by computation for (14)–(15) with $\epsilon = 10^{-3}$. Plotted as a solid line is the value derived by homogenization theory in the limit $\epsilon \to 0$.

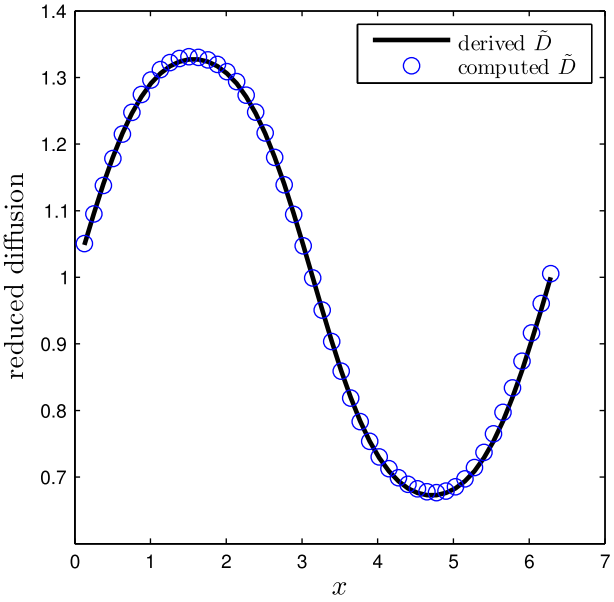

Fig. 7. The diffusion component $\tilde{D}$ computed for the reduced slow dynamics along the line $C = [0, 2\pi] \times \{0\}$. Plotted as circles is the value determined by computation for (14)–(15) with $\epsilon = 10^{-3}$. Plotted as a solid line is the value derived by homogenization theory in the limit $\epsilon \to 0$.

reduced slow system using the technique described in Section 6. We restrict the leading eigenfunctions ψ_1 and ψ_2 of the full system to the line $C = [0, 2\pi] \times \{0\}$ which is transverse to the fast fibres. Each eigenfunction ψ_k on the full space was solved numerically by collocation and the result is represented by the values that ψ_k takes on a 50×50 grid on $[0, 2\pi] \times [-L, L]$. Restriction of this eigenfunction to C may therefore be performed simply by taking the appropriate row from the matrix representing the computed function. The restriction $\psi_k|_C$ is then given by the values that the function takes on a set $X \times \{0\}$ consisting of 50 points uniformly spaced along C. At each point $x \in X$, we look at a window of points consisting of x and the ten other points in X closest to x. Using these eleven data points, we fit a cubic polynomial to $\psi_k|_C$. This polynomial is used to estimate the first and second derivatives of $\tilde{\psi}_k \approx \psi_k|_C$ at the point x which are then used to solve pointwise the linear equations given by (11). The computed values for the drift $\tilde{\mu}$ and diffusion $\tilde{D}$ of the reduced system are plotted in Figs. 6 and 7. Using homogenization techniques (see [27]) the reduced slow system of the multiscale system (12)–(13) can be determined as

$$dX = e^{-\frac{1}{4}} \sin(\sin(X))\, dt + \left(1 + \tfrac{1}{2} e^{-\frac{1}{4}} \sin(\sin(X))\right) dW_t. \qquad (17)$$

The drift and diffusion functions of this reduced slow equation are also plotted in the figures and the values determined by computation agree closely with those given by homogenization theory.

In all the numerical results above, the system (14)–(15) was considered with $\epsilon = 10^{-3}$ as this was the value of the time separation used in [27]. We also performed the same numerical computations for the system with different values of ϵ. The results for (14)–(15) with $\epsilon = 10^{-6}$ were qualitatively similar to those listed above. In particular, the algorithms correctly identified the slow and fast directions, isolated the fast dynamics, and determined the reduced slow dynamics. For values of $\epsilon < 10^{-6}$, the diffusion matrix D becomes ill-conditioned and numerical instabilities prevent a reliable computation of the reduced slow dynamics. However, in the case of the original undistorted system (12)–(13) the diffusion is aligned with the axes, meaning D is diagonal, and the algorithms perform acceptably for values of ϵ as small as 10^{-10}.

8. Discussion

To conclude, we look at the computational overhead of the techniques introduced in this paper in comparison to other methods for analysing multiscale systems. In the example in Section 7, building the matrices used in steps 1 and 2 of Algorithm 1 required evaluating the functions for the drift and diffusion at every point of a 50×50 grid. Other steps in Algorithms 1, 2A, and 2B consider the drift and diffusion on lower-dimensional fibres and require even fewer evaluations. Overall, the computation of the drift and diffusion data at various points do not significantly contribute to the run time of the algorithms. In fact, the construction of the matrices representing the partial derivatives are more time consuming. Since there were no severe bottlenecks, all of the computations in Section 7 were performed with a total runtime of 25 s on a desktop computer.[2] This running time is small in comparison to techniques which rely on trajectory integration to analyse a system.

A time series long enough to accurately capture the statistics of the slow dynamics of (12)–(13) could easily be of length 10^6 or longer. Assuming the time step used to compute this orbit at least on the order of the time-scale separation $\epsilon^{-1} = 10^3$, this corresponds to 10^9 individual time steps of the Euler–Maruyama method.

This paper is a continuation of the work developed in [22]. There, we relied solely on trajectory integration to analyse a system and used Ulam's method to compute the eigenfunctions of $e^{t\mathcal{L}}$ for a dynamical system. To apply Ulam's method to the multiscale system, the length of orbits computed need to be at least of the order of the slow dynamics and many individual segments of the orbits need to be computed. For instance, for the numerical examples given in [22], each orbit segment was computed using 2×10^4 time steps and 10^4 individual orbits were computed in each square of a 200×200 grid. This gives a total count of 8×10^{12} steps of the Euler–Maruyama method where each step must compute

[2] The computer used was an Intel Core i7-4770 with four CPUs at 3.40 GHz and 16 GB of RAM running Ubuntu 12.04 with 64-bit Linux kernel 3.13.0-39-generic and MATLAB R2013b.

the values of the drift and diffusion coefficients at a point in the phase space. In [22], this computation was achieved by splitting the construction of the matrix over many computers in parallel. Methods based on the infinitesimal generator avoid these issues and can thus lead to large gains in computation speed.

That said, the infinitesimal generator method is not always applicable. If the diffusion in the actual reduced slow process is very small, it may be dominated by numerical diffusion in the full space arising when computing the eigenfunctions of the Fokker–Planck operator. In this case, other less efficient methods, such as Ulam's method with a large enough time step, must be used to yield accurate results.

The analysis and numerical experiments of this paper consider multiscale systems with one-dimensional slow dynamics. The techniques likely also apply to systems with higher-dimensional slow dynamics and we give an outline of how the algorithms might be extended. The application of such techniques is left as an area of future research. See also the Discussion section of [22].

Assume now that the slow dynamics has dimension r where $1 < r < d$. In Algorithm 1, several leading eigenfunctions ψ_k are computed. Instead of just considering a fibre defined as $\psi_1^{-1}(\{s\})$ for some $s \in \mathbb{R}$, a fibre F is now defined as an intersection $\psi_1^{-1}(\{s\}) \cap \cdots \psi_r^{-1}(\{s\})$ of the level sets of several eigenfunctions. Polynomial regression may still be used to compute the normal and tangential components of the drift and diffusion. In Algorithm 2A, the function g is now a function $g : V \to \mathbb{R}^r$ where $V \subset \mathbb{R}^{d-r}$, but otherwise the algorithm is unchanged. In Section 6, the drift and diffusion of the reduced slow dynamics are determined from several eigenpairs associated to the Fokker–Planck operator of the reduced system. When the reduced system is one-dimensional this can be readily achieved with a small number of eigenfunctions, but techniques exist which also apply in the higher-dimensional setting [21].

References

[1] Peter Henrici, Discrete Variable Methods in Ordinary Differential Equations, John Wiley & Sons, Inc., New York-London, 1962.
[2] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations. I. Nonstiff Problems, second ed., in: Springer Series in Computational Mathematics, vol. 8, Springer-Verlag, Berlin, 1993.
[3] G.A. Pavliotis, A.M. Stuart, Parameter estimation for multiscale diffusions, J. Stat. Phys. 127 (4) (2007) 741–781.
[4] E. Weinan, Principles of Multiscale Modeling, Cambridge University Press, 2011.
[5] A.J. Roberts, Appropriate initial conditions for asymptotic descriptions of the long term evolution of dynamical systems, J. Aust. Math. Soc. Ser. B. Appl. Math. 31 (01) (1989) 48–75.
[6] Marc R. Roussel, Simon J. Fraser, On the geometry of transient relaxation, J. Chem. Phys. 94 (11) (1991) 7106–7113.
[7] Ulrich Maas, Stephen B. Pope, Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space, Combust. Flame 88 (3) (1992) 239–264.
[8] C.W. Gear, I. Kevrekidis, Projective methods for stiff differential equations: Problems with gaps in their eigenvalue spectrum, SIAM J. Sci. Comput. 24 (4) (2003) 1091–1106.
[9] Anthony J. Roberts, Normal form transforms separate slow and fast modes in stochastic dynamical systems, Physica A 387 (1) (2008) 12–38.
[10] Christian Kuehn, Multiple Time Scale Dynamics, in: Applied Mathematical Sciences, vol. 191, Springer, Cham, 2015.
[11] E. Weinan, Björn Engquist, The heterogeneous multiscale methods, Commun. Math. Sci. 1 (1) (2003) 87–132.
[12] Eric Vanden-Eijnden, On HMM-like integrators and projective integration methods for systems with multiple time scales, Commun. Math. Sci. 5 (2) (2007) 495–505.
[13] Viviane Baladi, Correlation spectrum of quenched and annealed equilibrium states for random expanding maps, Comm. Math. Phys. 186 (3) (1997) 671–700.
[14] Grigoris A. Pavliotis, Andrew M. Stuart, Multiscale Methods: Averaging and Homogenization, Springer, New York, 2008.
[15] L.C.G. Rogers, David Williams, Diffusions, Markov Processes, and Martingales. Vol. 1. Foundations, in: Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Reprint of the second (1994) edition.
[16] L.C.G. Rogers, David Williams, Diffusions, Markov Processes, and Martingales. Vol. 2. Foundations, in: Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Reprint of the second (1994) edition.
[17] Michael Dellnitz, Oliver Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal. 36 (2) (1999) 491–515.
[18] Gary Froyland, Oliver Junge, Péter Koltai, Estimating long-term behavior of flows without trajectory integration: the infinitesimal generator approach, SIAM J. Numer. Anal. 51 (1) (2013) 223–247.
[19] Gary Froyland, Cecilia González-Tokman, Anthony Quas, Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools, J. Comput. Dyn. 1 (2) (2014) 249–278.
[20] Gary Froyland, Simon Lloyd, Naratip Santitissadeekorn, Coherent sets for nonautonomous dynamical systems, Physica D 239 (16) (2010) 1527–1541.
[21] D. Crommelin, E. Vanden-Eijnden, Diffusion estimation from multiscale data by operator eigenpairs, Multiscale Model. Simul. 9 (4) (2011) 1588–1623.
[22] Gary Froyland, Georg A. Gottwald, Andy Hammerlindl, A computational method to extract macroscopic variables and their dynamics in multiscale systems, SIAM J. Appl. Dyn. Syst. 13 (4) (2014) 1816–1846.
[23] John P. Boyd, Chebyshev and Fourier Spectral Methods, second ed., Dover Publications Inc., Mineola, NY, 2001.
[24] Bernt Øksendal, Stochastic Differential Equations. An Introduction with Applications, Universitext. Springer-Verlag, Berlin, 1985.
[25] Dror Givon, Raz Kupferman, Andrew Stuart, Extracting macroscopic dynamics: model problems and algorithms, Nonlinearity 17 (6) (2004) R55–R127.
[26] Lars Peter Hansen, José Alexandre Scheinkman, Nizar Touzi, Spectral methods for identifying scalar diffusions, J. Econometrics 86 (1) (1998) 1–32.
[27] D.T. Crommelin, E. Vanden-Eijnden, Reconstruction of diffusions using spectral data from timeseries, Commun. Math. Sci. 4 (2006) 651–668.
[28] Emmanuel Gobet, Marc Hoffmann, Markus Reiß, Nonparametric estimation of scalar diffusions based on low frequency data, Ann. Statist. 32 (5) (2004) 2223–2253.
[29] Peter Sjögren, Operators associated with the Hermite semigroup—a survey, in: Proceedings of the Conference Dedicated to Professor Miguel de Guzmán (El Escorial, 1996), Vol. 3, 1997, pp. 813–823.