Stochastic homogenization for an energy conserving multi-scale toy model of the atmosphere

Jason E. Frank[a], Georg A. Gottwald[1]

[a] *Centrum Wiskunde & Informatica*
P.O. Box 94079, 1090 GB Amsterdam
The Netherlands
jason@cwi.nl

[b] *School of Mathematics and Statistics*
University of Sydney
NSW 2006
Australia
georg.gottwald@sydney.edu.au

Abstract

We study a Hamiltonian toy model for a Lagrangian fluid parcel in the semi-geostrophic limit which exhibits slow and fast dynamics. We first reinject unresolved fast dynamics into the deterministic equation through a stochastic parametrization that respects the conservation of the energy of the deterministic system. In a second step we use stochastic singular perturbation theory to derive an effective reduced stochastic differential equation for the slow dynamics. We verify the results in numerical simulations.

Keywords: Homogenization; Multi-scale systems; Stochastic parametrizations

1. Introduction

The dynamics of the atmosphere and the oceans is inherently complex. There are active entangled processes running on spatial scales from millimetres to thousands of kilometres, and temporal scales from seconds to millennia. To capture the whole range of spatial and temporal scales is impossible given current computer power. Any numerical forecaster has to make a decision, depending on the specific objectives, as to what spatial and

temporal scales to resolve. A corollary of this decision is that each numerical scheme inevitably fails to resolve so called subgrid scales. Usually the interesting information is carried by the slow and large scales. For example, for weather forecasts we want to resolve large scale high and low pressure fields rather than small scale fast oscillations of the stratification surfaces, or for climate prediction in a coupled ocean-atmosphere model we may want to learn about the slow dynamics of the ocean, which is constantly kicked by more rapidly evolving weather systems swirling over the surface.

The crucial question is: can one employ a coarse spatial grid and a large time step in a numerical discretization tailored to the large and slow scales of interest (i.e. large and slow weather systems or climate)? And further, can this be achieved while still accounting for the vital interactions with the unresolved processes which are smaller than the coarse spatial grid and faster than the slow time step used in the integration?

In the context of long-range weather forecasting the small unresolved scales are embodied in fast small-scale inertia-gravity waves. The dynamics on the coarse grid is therefore coupled to a collection of fast wave motion. The phases of these fast oscillators are randomized by the chaotic interactions of these waves. It has been shown that a large collection of oscillators with randomized phases may be modelled by a stochastic process [29, 21]. This approach of modelling fast small scale chaotic processes by a stochastic process is intuitive: provided the fast processes decorrelate rapidly enough, the slow variables experience during one slow time unit the sum of uncorrelated events of the fast dynamics, which according to the Central Limit Theorem corresponds to approximate Gaussian noise. In the mathematical community a method whereby many fast degrees of freedom are replaced by a stochastic process is called *stochastic model reduction*.

In climate modelling, the idea of modelling fast chaotic dynamics by stochastic processes and thereby reducing the effective dimension of the full system goes back to the seminal work by Hasselmann [27]. Prior to this, the approach to unresolved scales was to average over the measure induced by the fast variables [1] which are assumed to be known. Averaging renders an effective deterministic equation for the slow variables, and cannot describe rapid regime switches essential for important phenomena such as El Niño, atmospheric blocking and climate change. In [27] it was suggested to study climatic regime switches by introducing in an ad-hoc way a stochastic driver for the slow dynamics. Such an approximation describes the deviations from an averaged climatological system. Of course, it is natural to

expect such behaviour only if the fast variables (such as weather in a coupled climatic ocean-atmosphere model) are chaotic or approximately random. In the case of multi-scale systems rigorous mathematical theorems [32, 45] can be employed to devise systematic stochastic subgrid parametrizations using averaging and homogenization techniques [36, 37, 39].

Scientists have recently realized that these methods can be applied to many complex systems [21, 11, 28, 48, 38, 49, 52, 53]. The effective dimension reduction achieved if all fast equations are replaced by one stochastic process, and the associated computational advantage of such a reduction is a huge driving force behind this research. These ideas have been used to simulate, for example, coupled ocean atmosphere models [42], urban air-pollution [3], and in a different field, macromolecular systems [20].

In this work we will perform such a programme for a simple finite-dimensional caricature model of large-scale motion of the atmosphere. Our particular angle here will be that the stochastic subgrid scale parametrization should not destroy the energy conservation of the large scale deterministic core. In a first step we will introduce a stochastic parametrization of a multi-scale system which conserves the energy of the deterministic core. In a second step we will use homogenization techniques to construct a reduced stochastic model which effectively describes the statistics of the slow degrees of freedom.

2. A toy model for the large-scale dynamics of the atmosphere

Since the atmosphere is shallow with an approximate height of 10km (troposphere) and a longitudinal extent of roughly $40,000$km in the midlatitudes, a good approximation for the dynamics is given by the 2-dimensional rotating shallow water equations

$$\begin{aligned} \frac{\mathrm{D}u}{\mathrm{D}t} + f_0\, Ju + g\,\nabla h &= 0\,, \\ \frac{\mathrm{D}h}{\mathrm{D}t} + h\,\nabla\cdot u &= 0\,, \end{aligned}$$

where

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

denotes the canonical symplectic matrix. Here $u = u(x,t)$ denotes the velocity field on the horizontal plane $x \in \mathbb{R}^2$, $h = h(x,t)$ the layer depth, ∇

the horizontal gradient, $\mathrm{D}/\mathrm{D}t = \partial_t + u \cdot \nabla$ the material derivative, g the constant of gravity, and $f_0/2$ the ambient angular velocity in the so-called f-plane approximation [47, 54]. If we non-dimensionalize by introducing a typical velocity U, a typical horizontal length scale L and a typical layer depth H, we can rewrite the rotating shallow water equations as

$$\epsilon \frac{\mathrm{D}u}{\mathrm{D}t} + Ju - \frac{B}{\epsilon} \nabla h = 0\,, \tag{2a}$$

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h \nabla \cdot u = 0\,. \tag{2b}$$

Here we introduced the Rossby number $\epsilon = U/f_0 L$ and the Burgers number $B = (L_R/L)^2$ with the Rossby radius of deformation $L_R = \sqrt{gH}/f_0$.

The Hamiltonian Particle-Mesh method was originally introduced as a particle-based discretization of the shallow water equations [14, 15], and later extended to other geometry [16] and other fluid models [6, 12, 51]. In the context of the shallow water equations (2a)-(2b) the method is given by

$$\begin{aligned} \varepsilon m_k \dot{q}_k &= p_k \\ \dot{p}_k &= -\frac{1}{\varepsilon^2} J p_k - \frac{B}{\varepsilon} \nabla h_\alpha(x,t)\big|_{x=q_k} \end{aligned} \tag{3}$$

where $\varepsilon = \sqrt{\epsilon}$, $m_k \in \mathbb{R}$ and $q_k, p_k \in \mathbb{R}^2$ are the mass, position and momentum of the kth discrete fluid particle, and $h_\alpha(x,t)$ is a continuum approximation of the shallow water layer depth, regularized on a length scale α. To facilitate computation of the layer depth and ensure only local interactions, an auxiliary numerical mesh is introduced with grid points x_i and layer depth approximations

$$h_i(t) = \sum_k m_k \psi(x_i - q_k(t)), \qquad \tilde{h}_i(t) = \sum_j (S_\alpha)_{ij} h_j(t), \tag{4}$$

where ψ is a symmetric basis function with compact support in $\mathbb{R}^2$, and S_α approximates the action of a smoothing kernel $S_\alpha \approx (1 - \alpha\Delta)^{-2}$ with length scale α. With these definitions, the continuous layer depth approximation is

$$h_\alpha(x,t) = \sum_i \tilde{h}_i(t) \psi(x - x_i). \tag{5}$$

By construction, the Hamiltonian particle-mesh method conserves the Hamiltonian (symplectic) structure and satisfies a circulation theorem [15]. Local mass conservation is trivially enforced, since each particle mass is a fixed

parameter of the discretization. Convergence of the method is considered in [41]. Analysis and motivation of the layer depth regularization is given in [17]. The reader is also referred to [7] for a semi-Lagrangian variant.

Large-scale dynamics in the atmosphere and oceans of rotating planets is characterized by their temporal scales being much slower than the rotational time scale of the earth, i.e. $\epsilon \ll 1$, and their spatial scales being much larger than the Rossby radius of deformation, i.e. $B = O(\epsilon)$. This corresponds to the so-called *semigeostrophic scaling* [9, 47, 50, 54] where the fluid motion is dominated by large vortices as we know them from weather maps as high and low pressure fields. If we now assume that the layer depth is a prescribed function $h = h(x)$ (rather than being evolved through the continuity equation (2b), e.g. see [2]), and subsequently set $h_\alpha = V(q)$, we arrive at the following caricature model for the large scale motion of a distinguished particle with unit mass, introduced in [43],

$$\epsilon \ddot{q} = J\dot{q} - \nabla V\,(q)\ . \tag{6}$$

We remark that the model (6) also describes the dynamics of a single charged particle with mass ϵ in a planar potential $V(q)$ under the influence of an external magnetic field J.

The Rossby number ϵ introduces time-scale separation. On times $t = O\,(1)$ the dynamics is slow and obtained, at lowest order, by formally setting $\epsilon = 0$ to yield

$$\dot{q} = -J\nabla V(q)\,. \tag{7}$$

Hence at the lowest order in ϵ, the slow dynamics evolves along equipotential lines of the potential V. On the fast time scale $\tau = t/\epsilon$, the dynamics is given at leading order by

$$q'' = Jq'\ ,$$

where the prime denotes derivatives with respect to τ. Thus the fast dynamics represents a harmonic oscillator.

The slow dynamics can be derived up to any order in ϵ in a variational framework assuring that the reduced equations possess geometric conservation properties analogous to those of the full parent system. In [23, 22, 8] it was shown rigorously that the slow 2-dimensional reduced dynamics converges to the full 4-dimensional dynamics. In particular, the dynamics converge on a long time scale $\mathcal{O}(1/\varepsilon)$ provided the slow dynamics is sufficiently

non-chaotic and slow-slow resonances can be excluded [22] which is certainly the case for the 2-dimensional toy model studied here. The smaller ϵ, the larger the separation between the two time-scales. This is illustrated in Figure 1, which shows projected particle trajectories as well as the approximate slow manifolds. To obtain better agreement between the reduced slow dynamics and the full system for $\varepsilon \to 0$ one would need higher-order balance equations, see [22, 23].

To extract the harmonic oscillations as the dominant process on the fast time scale we write the system (6) as in the Hamiltonian-Particle-Mesh method (3)

$$\begin{aligned} \dot{q} &= \frac{1}{\varepsilon} p \\ \dot{p} &= \frac{1}{\varepsilon^2} J p - \frac{1}{\varepsilon} \nabla V(q) \,, \end{aligned} \tag{8}$$

where again $\varepsilon = \sqrt{\epsilon}$.

A crucial ingredient of the toy model is the existence of a conserved energy,

$$H(q,p) = \frac{1}{2} |p|^2 + V(q) \,. \tag{9}$$

We assume throughout that $V(q)$ is strictly convex for large values of q, so that $q(t)$ is bounded independent of ε for all times. We also assume that V is sufficiently smooth so that all necessary derivatives exist and are continuous. The toy model (6) is a Hamiltonian system with non-canonical coordinates q and $p = \varepsilon \dot{q}$ and Hamiltonian (9). Setting $z = (q,p)$ the toy model (8) can be cast as

$$\dot{z} = \mathbb{J} \nabla_z H \tag{10}$$

with the symplectic structure matrix

$$\mathbb{J} = \frac{1}{\varepsilon} \begin{pmatrix} 0 & I \\ -I & \frac{1}{\varepsilon} J \end{pmatrix} ,$$

where $I \in \mathbb{R}^{2\times 2}$ is the identity matrix.

Assuming that the dynamics is ergodic, the Hamiltonian structure of the toy model (6) implies a particular form of the invariant measure: When projected onto the slow subspace of the q-plane, the Hamiltonian toy-model

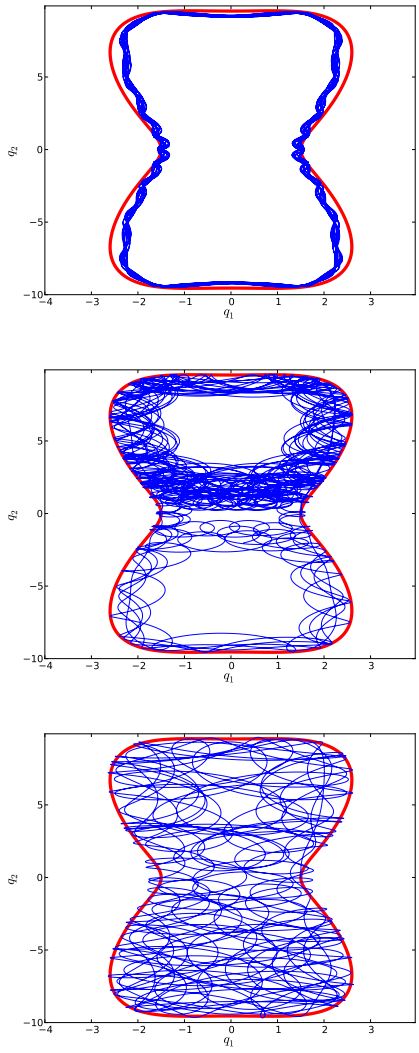

Figure 1: Trajectories of the toy model (8) with $V(q)$ given by (47) in the q-plane with $\varepsilon = 0.1115$ (top), $\varepsilon = 0.14$ (middle) and $\varepsilon = 0.25$ (bottom). The thick line (online red) denotes the approximate slow manifold given by (7). Initial conditions for the full toy model (8) were chosen to lie on the leading order slow manifold with $\dot{q}(0) = -J\nabla V(q(0))$.

(6) generates the Lebesgue measure on the energetically accessible range. This can be readily seen by applying the co-area formula (see for example [10]) to the 4-dimensional volume element $d\mu = dq_1 dq_2 dp_1 dp_2$

$$d\mu = \delta(H(q,p) - E) \frac{ds}{|\nabla_z H|} dE \,, \tag{11}$$

where ds is an infinitesimal surface element on the manifold of constant energy $H(p,q) = E$. Employing energy conservation we find

$$p_2 = \pm \sqrt{2\,(E - V(q)) - p_1^2}\,,$$

and the surface element ds can be written as

$$ds = \sqrt{1 + \left|\tilde{\nabla} p_2\right|^2}\, dq_1 dq_2 dp_1 \,,$$

with $\tilde{\nabla} = (\partial_q, \partial_{p_1})$. Hence, the 4-dimensional volume element (11) becomes

$$d\mu = \frac{1}{\sqrt{2\,(E - V(q)) - p_1^2}}\, dp_1\, dq_1\, dq_2\,. \tag{12}$$

Projecting the measure onto the 2-dimensional q-plane by integrating over p_1 requires the Cauchy formula to deal with the singularity in p_1. Substitution of $p_1 = 1/\eta$ allows for the application of the residue theorem

$$\oint \frac{1}{\sqrt{2\,(E - V) - p_1^2}}\, dp_1 = -\oint \frac{d\eta}{\eta \sqrt{2\,(E - V)\,\eta^2 - 1}} = 2\pi \,,$$

and so the invariant measure projected onto the energetically accessible q-plane is given by the Lebesgue measure

$$d\mu^q = 2\pi\, dq_1\, dq_2 \,. \tag{13}$$

In Figure 2 we show that indeed the empirical density, as estimated from a long numerical trajectory of the full deterministic toy model (8) when projected onto the q plane, is Lebesgue on the energetically accessible region. As $\varepsilon \to 0$, the dynamics is more closely constrained to the slow manifold from Equation (7), and the support of the measure becomes smaller. This effect is illustrated in Figure 2.

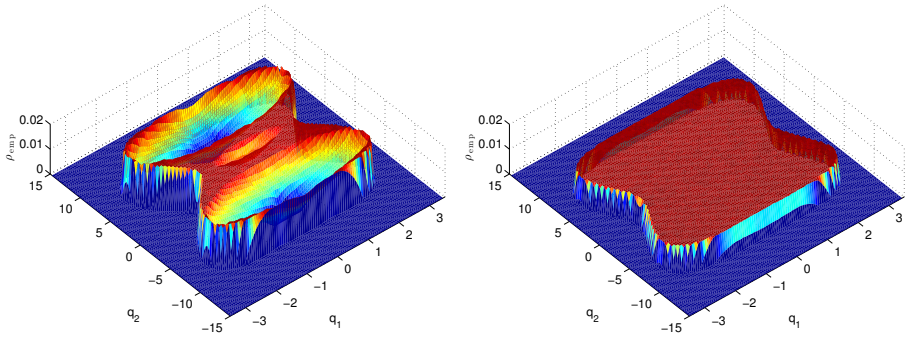

Figure 2: Empirical density $\rho_{\mathrm{emp}}(q)$ of a trajectory of the deterministic toy model (8) with $V(q)$ given by (47) with $\varepsilon = 0.15$ (left) and $\varepsilon = 0.5$ (right).

3. Notation

We briefly introduce some of the notation we shall use, which we adopt from [46]. All vectors are assumed to be column vectors unless otherwise indicated. The gradient operator ∇ is defined as a vector operator. Acting on a scalar $\phi \in \mathbb{R}$ we have

$$\{\nabla\phi\}_i = \partial_{q_i}\phi \ ,$$

and acting on a vector $v \in \mathbb{R}^n$ we have

$$\{\nabla v\}_{ij} = \partial_{q_j} v_i \ .$$

For phase space variables $z = (q,\, p)$ with $q,\, p \in \mathbb{R}^2$ we use subscript notation for derivatives so that $\nabla_p = (\partial_{p_1},\, \partial_{p_2})^T$. When no subscript is used we mean derivatives with respect to q. The divergence operator is defined as $\nabla\cdot = \nabla^T$, and as such acts on vector valued functions $v \in \mathbb{R}^n$ in the usual way. Acting on matrix valued functions A the divergence acts by contraction as

$$\{\nabla \cdot A\}_i = \partial_{q_j}\left(A_{ji}\right) \ ,$$

whereby we use Einstein summation convention and sum over repeating indices. For notational convenience we shall write the matrix of second order partial derivatives as $\nabla\nabla$, dropping the transpose symbol. Thus $\nabla\nabla\phi$ is the Hessian matrix of $\phi \in \mathbb{R}$. We define the inner product of matrices as

$$A : B = a_{ij}b_{ij} = Tr(AB^T) \ .$$

4. Parametrization of unresolved fast chaotic processes on a manifold

We now reinject information from the unresolved small-scale dynamics into the toy model for the large scales (8). We shall do so following the "Hasselmann" programme by parametrizing those processes by noise. Unresolved scales in the toy model (8) can be associated with two separate mechanisms. There are unresolved scales due to semi-geostrophic approximation and due to the framework of the particle method used to discretise them. In the process of the asymptotic derivation of the semi-geostrophic approximation fast small-scale wave motion has been filtered out and in turn constitute unresolved processes. In the context of the Hamiltonian Particle-Mesh method the unresolved scales are all those processes with scales below the smoothing length-scale α, including (but not exclusively) fast inertia gravity waves. Following [36], we assume that the effect of both of these temporally fast unresolved scales on the resolved scales can be represented by Gaussian noise. The heuristics behind this reasoning is that in one slow time-unit the slow resolved variables feel the sum of the uncorrelated unresolved fast variables which, by means of the Central Limit Theorem, justifies the usage of Gaussian noise. For simplicity we have restricted ourselves here to additive noise and have ignored multiplicative noise. The variance of the stochastic process can be chosen to match prior climatic information on the fast unresolved chaotic processes (see for example [37, 38]). The particular angle we take here is to require that the noise not destroy large-scale energy conservation. Local mass conservation will be unaffected by the stochastic process proposed here, but momentum balance will be modified due to diffusion.

We determine a stochastic process which leaves the dynamics constrained to the manifold of constant energy (9) (see for example [5, 35]). As the noise parametrizes unresolved fast chaotic processes we add a stochastic process only to the fast variable p and on the fast time scale t/ε, and propose the following parametrization for the toy model (8)

$$dq = \frac{1}{\varepsilon} p \, dt \tag{14}$$

$$dp = \left[\frac{1}{\varepsilon^2} \left(J - \Gamma \right) p - \frac{1}{\varepsilon} \nabla V \right] dt + \frac{1}{\varepsilon} \Sigma \, dW_t + dY_t \tag{15}$$

$$dY_t = \frac{1}{\varepsilon} S_t \, dW_t + \frac{1}{\varepsilon^2} B_t \, dt \; , \tag{16}$$

where $\Gamma, \Sigma \in \mathbb{R}^{2\times 2}$, dW is a 2-dimensional Brownian motion and dY_t is an auxiliary stochastic process with components $S_t \in \mathbb{R}^{2\times 2}$ and $B_t \in \mathbb{R}^2$. In the case $dY_t = 0$ the toy model (6) is simply augmented by an Ornstein-Uhlenbeck process. The unconstrained nature of the noise will cause the dynamics to leave the manifold of constant energy on which the deterministic model (8) evolves. The auxiliary stochastic process dY_t, parametrized by S_t and B_t, can now be chosen to impose the restriction that the energy (9) is preserved even in the stochastically driven system (14)-(16).

Using Itô calculus we evaluate the change in energy (9)

$$\begin{aligned} dH &= \nabla_q H \cdot dq + \nabla_p H \cdot dp + \frac{1}{2}\nabla_p \nabla_p H : dp\, dp^T \\ &= \mu_H\, dt + \sigma_H\, dW_t \ , \end{aligned}$$

where

$$\begin{aligned} \mu_H &= \frac{1}{\varepsilon}\nabla_q H \cdot p + \nabla_p H \cdot \left[\frac{1}{\varepsilon^2}\left(J-\Gamma\right)p - \frac{1}{\varepsilon}\nabla V + \frac{1}{\varepsilon^2}B_t\right] \\ &\quad + \frac{1}{2\varepsilon^2}\nabla_p\nabla_p H : \left(\Sigma + S_t\right)\left(\Sigma + S_t\right)^T \qquad (17) \\ \sigma_H &= \frac{1}{\varepsilon}\nabla_p H \cdot \left(\Sigma + S_t\right) \ . \qquad (18) \end{aligned}$$

For the Hamiltonian (9) we have $\nabla_q H = \nabla V$, $\nabla_p H = p$, $\nabla_p \nabla_p H = I \in \mathbb{R}^{2\times 2}$, where I is the identity matrix. The functions S_t and B_t will be chosen such that $\mu_H = 0$ and $\sigma_H = 0$ and hence $dH = 0$. Heuristically, the auxiliary process will be constructed in such a way as to force the deviations from the manifold of constant energy caused by the Ornstein-Uhlenbeck process back onto the manifold. It shall therefore only have components orthogonal to the energy manifold. To this end we introduce a projector $\mathbb{P}$ which will project onto the tangent space of the manifold.

It turns out that projecting onto the manifold of the kinetic energy $K = |p|^2/2$ is sufficient. We define the projection operator $\mathbb{P} \in \mathbb{R}^{2\times 2}$ as

$$\begin{aligned} \mathbb{P} &= I - \frac{1}{|\nabla_p H|^2}\nabla_p H \nabla_p H^T \\ &= I - \frac{1}{|p|^2} pp^T \ . \end{aligned}$$

The identities for projection operators $\mathbb{P} = \mathbb{P}^T$ and $\mathbb{P}\mathbb{P} = \mathbb{P}$ are trivially satisfied. Note that $\nabla_p H = p$ lies in the kernel of $\mathbb{P}$, and so $\mathbb{P}$ projects

onto the tangent space of the fast kinetic energy surface. Since $\mathbb{P}$ is symmetric and positive semi-definite, we obtain the following useful Cholesky decomposition of $\mathbb{P}$

$$\mathbb{P} = \frac{1}{|p|^2} \begin{pmatrix} 0 & -p_2 \\ 0 & p_1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -p_2 & p_1 \end{pmatrix} . \tag{19}$$

The auxiliary stochastic process dY_t shall not perturb the dynamics on the tangent space, and shall be constructed only to counterbalance those components of the Ornstein-Uhlenbeck process which are orthogonal to the manifold of constant energy. We therefore require

$$\mathbb{P}S_t = 0 \in \mathbb{R}^{2\times 2} \qquad \text{and} \qquad \mathbb{P}B_t = 0 \in \mathbb{R}^2 . \tag{20}$$

We now proceed with determining S_t and B_t to assure $dH = 0$. We first require $\sigma_H = 0$ in (18). Since $\nabla_p H$ is in the kernel of $\mathbb{P}$, $\sigma_H = 0$ is equivalent to $\Sigma + S_t = \mathbb{P}(\Sigma + S_t)$ and hence using (20)

$$S_t = -(I - \mathbb{P})\Sigma . \tag{21}$$

Using this expression in (17), we can determine B_t from our requirement $\mu_H = 0$ as

$$\begin{aligned} B_t &= (\mathbb{P} - I)(J - \Gamma)p - \frac{1}{2}\frac{\nabla_p \nabla_p H : \mathbb{P}\Sigma\Sigma^T\mathbb{P}}{|\nabla_p H|^2}\nabla_p H \\ &= (I - \mathbb{P})\Gamma p - \frac{1}{2}\frac{1}{|p|^2}\mathbb{P} : \left(\Sigma\Sigma^T\right) p , \end{aligned} \tag{22}$$

Substituting the expressions for S_t and B_t from (21) and (22) into (8) yields

$$dq = \frac{1}{\varepsilon} p\, dt \tag{23}$$

$$dp = \left[\frac{1}{\varepsilon^2}(J - \mathbb{P}\Gamma)p - \frac{1}{\varepsilon}\nabla V - \frac{1}{2\varepsilon^2}\frac{1}{|p|^2}\mathbb{P} : \left(\Sigma\Sigma^T\right) p\right] dt + \frac{1}{\varepsilon}\mathbb{P}\Sigma\, dW_t , \tag{24}$$

which by construction conserves the energy (9). The generator associated with (23)-(24) can be written as an expansion in ε

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 \tag{25}$$

with

$$\mathcal{L}_0 = \left[(J - \mathbb{P}\Gamma)\, p - \frac{1}{2\,|p|^2}\mathbb{P} : \left(\Sigma\Sigma^T\right)\, p \right] \cdot \nabla_p + \frac{1}{2}\mathbb{P}\Sigma\Sigma^T\mathbb{P} : \nabla_p\nabla_p \;, \tag{26}$$

$$\mathcal{L}_1 = p \cdot \nabla_q - \nabla V \cdot \nabla_p \;. \tag{27}$$

The projected stochastic Itô differential equation (23)-(24) can be alternatively formulated in the Stratonovich sense. For matrices $A, B\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$, $x \in \mathbb{R}^n$ and $W_t \in \mathbb{R}^n$ an n-dimensional Brownian motion and $dx_t = F(x) + A(x) \circ B(x) dW_t$ with $x_t \in \mathbb{R}^n$, we have

$$[A(x) \circ B(x) dW_t]_i = [A(x)B(x)dW_t]_i + \frac{1}{2}\sum_{j,k} \partial_{x_k} A_{ij}(x) \left(A(x)B(x)B^T(x)\right)_{kj}$$

for $i = 1, \ldots, n$ (see for example [44]). Hence in the Stratonovich framework (23)-(24) can be written in the more compact form as

$$\begin{aligned} dq &= \frac{1}{\varepsilon} p\, dt \\ dp &= \left[\frac{1}{\varepsilon^2} (J - \mathbb{P}\Gamma)\, p - \frac{1}{\varepsilon}\nabla V \right] dt + \frac{1}{\varepsilon}\mathbb{P} \circ \mathbb{P}\Sigma dW_t \;. \end{aligned}$$

For simplicity we set from now on

$$\Gamma = \gamma I \qquad \text{and} \qquad \Sigma = \sigma I \;,$$

where γ and σ are scalars and $I \in \mathbb{R}^{2\times 2}$ is the identity matrix. Then the energy conserving toy model (23)-(24) can be written as

$$dq = \frac{1}{\varepsilon} p\, dt \tag{28}$$

$$dp = \left[\frac{1}{\varepsilon^2} Jp - \frac{1}{\varepsilon}\nabla V - \frac{\sigma^2}{2\varepsilon^2\,|p|^2}\, p \right] dt + \frac{1}{\varepsilon}\sigma\mathbb{P}\, dW_t \;, \tag{29}$$

with W_t being 2-dimensional Brownian motion. It is this system we will now analyze. Using the expression (19) for the projector $\mathbb{P}$, the generators (26) and (27) become

$$\mathcal{L}_0 = \left(J - \frac{\sigma^2}{2\,|p|^2} I \right) p \cdot \nabla_p + \frac{1}{2}\sigma^2 \frac{1}{|p|^2} \begin{pmatrix} p_2^2 & -p_1p_2 \\ -p_1p_2 & p_1^2 \end{pmatrix} : \nabla_p\nabla_p \;, \tag{30}$$

$$\mathcal{L}_1 = p \cdot \nabla_q - \nabla V \cdot \nabla_p \;. \tag{31}$$

4.1. The fast stochastic dynamics

Before we apply stochastic singular perturbation theory to the energy conserving stochastic system (28)-(29) to extract a reduced equation, we derive a number of properties of the fast stochastic dynamics associated with $\mathcal{L}_0$. In particular, we shall show that the fast dynamics is ergodic and explicitly compute the invariant density.

Using the Cholesky decomposition (19) for the projector, the stochastic differential equation associated with the generator of the fast dynamics $\mathcal{L}_0$ is

$$dp = Yp\,dt + \frac{\sigma}{|p|}\begin{pmatrix} -p_2 \\ p_1 \end{pmatrix} d\hat{W}_t\ , \tag{32}$$

with $\hat{W}_t \in \mathbb{R}$ being the second component of W_t, and

$$\begin{aligned} Y &= J - \frac{\sigma^2}{2\,|p|^2} I \\ &= \begin{pmatrix} -\frac{\sigma^2}{4K} & 1 \\ -1 & -\frac{\sigma^2}{4K} \end{pmatrix}\ . \end{aligned} \tag{33}$$

Energy conservation implies $\mathcal{L}H(q,p) = 0$ for all values of ε, and hence $\mathcal{L}_0 H(q,p) = 0$. Since $\mathcal{L}_0$ only involves derivatives with respect to the momentum variable p we conclude that energy conservation implies $\mathcal{L}_0\,|p|^2 = 0$, i.e. that the fast dynamics conserves the mean kinetic energy $K = |p|^2/2$. It is straightforward to show using Itô calculus that the kinetic energy is also pathwise constant under the $\mathcal{L}_0$-dynamics. Hence, the fast stochastic process (32) lives on the compact sphere with radius K. This assures the existence of an invariant measure. Moreover, since

$$\mathbb{E}[|p(t;x) - p(t;y)|^2] \leq 4\exp(-\frac{\sigma^2}{4K}t)\,|x-y|^2 \quad \forall x, y\ ,$$

where x, y are initial conditions $p(0)$, there is a unique invariant measure as the phase space cannot be decomposed in regions of non-zero measure in which the dynamics generates respective invariant measures. The fast process (32) is symmetric in p and has invariant density

$$\rho_\infty(p) = \rho_\infty(p;q) = \delta(|p|^2 - 2K)/Z\ , \tag{34}$$

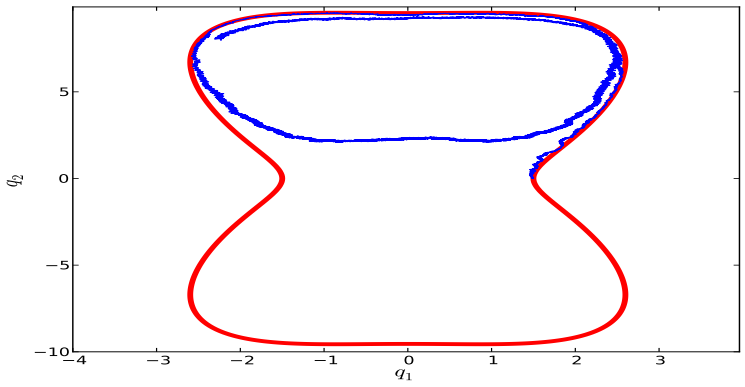

Figure 3: Trajectory of the energy conserving stochastic multi-scale system (28)-(29) with $V(q)$ given by (47), initialised on the slow manifold $\dot{q} = -J\nabla V(q)$, with $\varepsilon = 0.01$ and $\sigma = 0.5$.

with the normalization constant $Z = 1/(2\pi\sqrt{2K})$. This implies for the mean

$$\langle p \rangle_{\rho_\infty} = 0 \,, \tag{35}$$

and for the covariance

$$\langle pp^T \rangle_{\rho_\infty} = K\, I \,, \tag{36}$$

where the angular brackets denote the average over the invariant density of the fast process. Since $K = K(q)$ the invariant density is parametrized by q. This is linked to the time scale separation of our model (28)-(29). On time scales of the order $O\left(1/\varepsilon^2\right)$ the kinetic energy K is conserved since on that scale q is frozen. Hence on the fast time scale, the trajectory will follow equipotential lines before diffusing across equipotential lines, see Figure 3.

5. Stochastic Singular Perturbation Theory

In this section we perform stochastic singular perturbation theory to extract the effective slow dynamics of the position variables q for the energy conserving stochastic multi-scale system (28)-(29). In the context of the

Hamiltonian Particle Mesh method (3)–(5), this would result in a slow balanced motion, free of gravity waves, analogous to a discrete quasigeostrophic potential vorticity or point vortex model, for which the phase space is a function only of the particle positions.
We will analyse (28)-(29) in the framework of the backward Kolmogorov equation for the conditional expectation value of some sufficiently smooth observable $\phi(q,p)$ defined as

$$v(q_0,p_0,t) = \mathbb{E}\left[\phi(q(t),p(t)) \mid q(0)=q_0, p(0)=p_0\right] \ .$$

Here the expectation value is taken with respect to Brownian motion driving paths. Dropping the subscripts, we study the following Cauchy problem for $t \in [0,\infty)$

$$\begin{aligned} \frac{\partial v}{\partial t}(q,p,t) &= \mathcal{L}v(q,p,t) \\ v(q,p,0) &= \phi(q,p) \ , \end{aligned} \tag{37}$$

with the generator given in (25). Pioneered by Khasminsky [30], Kurtz [32] and Papanicolaou [45] singular perturbation theory can be formulated for a perturbation expansion according to

$$v(q,p,t) = v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \cdots \ . \tag{38}$$

A comprehensive exposition of the theory of stochastic model reductions and their implementation is given for example in [21, 46]. Substituting the series (38) into the backward Kolmogorov equation (37) yields upon equating powers of ε the following hierarchy of equations

$$\begin{aligned} O\left(\frac{1}{\varepsilon^2}\right): &\quad \mathcal{L}_0 v_0 = 0 \\ O\left(\frac{1}{\varepsilon}\right): &\quad \mathcal{L}_0 v_1 = -\mathcal{L}_1 v_0 \\ O(1): &\quad \mathcal{L}_0 v_2 = \frac{\partial v_0}{\partial t} - \mathcal{L}_1 v_1 \ . \end{aligned}$$

At lowest order, $\mathcal{O}(1/\varepsilon^2)$, we have

$$\mathcal{L}_0 v_0 = 0 \ . \tag{39}$$

The fast dynamics associated with the generator $\mathcal{L}_0$ was shown in the previous section to be ergodic with the unique invariant probability density $\rho_\infty(p)$ given by (34), i.e. ρ_∞ is the unique solution of

$$\mathcal{L}_0^\star \rho = 0 \ ,$$

where $\mathcal{L}_0^\star$ is the formal L_2-adjoint of the generator $\mathcal{L}_0$. Ergodicity implies that expectation values do not depend on initial conditions p. Hence the solution

$$v_0 = v_0(q,t)$$

is the only solution of (39).

At the next order, $\mathcal{O}(1/\varepsilon)$, we obtain

$$\mathcal{L}_0 v_1 = -\mathcal{L}_1 v_0 \; . \tag{40}$$

To assure boundedness of v_1 (and thereby of the asymptotic expansion (38)) a solvability condition prescribed by a Fredholm alternative has to be satisfied. Equation (40) is solvable only provided the right-hand-side lies in the space orthogonal to the (one-dimensional) null space of the adjoint $\mathcal{L}_0^\star$, i.e. if

$$\langle \mathcal{L}_1 v_0 \rangle_{\rho_\infty} = -\langle p \rangle_{\rho_\infty} \cdot \nabla v_0(q,t) = 0 \; .$$

Since we showed in the previous Section that $\langle p \rangle_{\rho_\infty} = 0$ and that the fast stochastic process is ergodic, the solvability condition is trivially satisfied and there exists a bounded solution of (40). Since v_0 is independent of p, the right side of (40) is $p \cdot \nabla_q v_0$. Furthermore, since from (27) the only differential operator in $\mathcal{L}_1$ is ∇_p, we write v_1 as,

$$v_1 = r(q,p) \cdot \nabla_q v_0 + R(q),$$

where $R(q)$ is a kernel function of $\mathcal{L}_0$. The function $r(q,p)$, inserted into (40), satisfies

$$\left(J - \frac{\sigma^2}{2\,|p|^2} I \right) p \cdot \nabla_p r + \frac{1}{2}\sigma^2 \mathbb{P} \colon \nabla_p \nabla_p r = -p \; ,$$

which is solved by

$$r = -\left(J - \frac{\sigma^2}{2\,|p|^2} I \right)^{-1} p \; .$$

Here we used the conservation of the kinetic energy by the fast dynamics implying $\mathcal{L}_0 \left(f\left(|p|^2 \right) \right) = 0$ for sufficiently smooth f. Hence we obtain for the $\mathcal{O}(1/\varepsilon)$ contribution

$$v_1 = -\left(J - \frac{\sigma^2}{2\,|p|^2} I \right)^{-1} p \cdot \nabla_q v_0 + R(q)$$

$$= -Y^{-1} p \cdot \nabla_q v_0 + R(q) \,, \tag{41}$$

where Y^{-1} is given by

$$Y^{-1} = -\frac{4K}{\sigma^4 + 16K^2} \begin{pmatrix} \sigma^2 & 4K \\ -4K & \sigma^2 \end{pmatrix} \,. \tag{42}$$

with $K = E - V(q)$.

At the next order, $\mathcal{O}(1)$, we obtain

$$\mathcal{L}_0 v_2 = \frac{\partial}{\partial t} v_0 - \mathcal{L}_1 v_1 \,.$$

Again a solvability condition has to be imposed which yields the desired evolution equation for v_0

$$\frac{\partial v_0}{\partial t} = \langle \mathcal{L}_1 v_1(q,t) \rangle_{\rho_\infty} \,. \tag{43}$$

Substituting the expression (41) for v_1 into (43) yields

$$\mathcal{L}_1 v_1 = -p \cdot \nabla_q \left(Y^{-1} p \cdot \nabla_q v_0 \right) + \nabla V \cdot \nabla_p \left(Y^{-1} p \cdot \nabla_q v_0 \right) + p \cdot \nabla_q R \,.$$

Energy conservation, and in particular the conservation of kinetic energy K under the $\mathcal{L}_0$-dynamics, implies that Y is a constant matrix with respect to the fast dynamics, and we have

$$\mathcal{L}_1 v_1 = -Y^{-1} p p^T \colon \nabla_q \nabla_q v_0 - p \cdot ((p \cdot \nabla_q) Y^{-T}) \nabla_q v_0 + Y^{-1} \nabla V \cdot \nabla_q v_0 + p \cdot \nabla_q R \,.$$

Here and in the following the gradient operator only acts on functions directly following it; hence $p \cdot ((p \cdot \nabla_q) Y^{-T}) \nabla_q = p_i p_l \left(\partial_i Y^{-1}_{kl} \right) \partial_k$. Averaging with respect to the invariant density of the $\mathcal{L}_0$-dynamics and using the expressions for the mean and the variance of p given by (35) and (36), respectively, we arrive at

$$\langle \mathcal{L}_1 v_1 \rangle_{\rho_\infty} = -K\, Y^{-1} \colon \nabla_q \nabla_q v_0 - K \nabla_q \cdot Y^{-T} \nabla_q v_0 + Y^{-1} \nabla_q V \cdot \nabla_q v_0 \,.$$

We summarise and arrive at the reduced slow backward Kolmogorov equation for v_0 as

$$\frac{\partial v_0}{\partial t} = \left(Y^{-1} \nabla_q V \cdot \nabla_q - K \nabla_q \cdot Y^{-T} \nabla_q \right) v_0 - K\, Y^{-1} \colon \nabla_q \nabla_q v_0 \,. \tag{44}$$

Note that $R(q)$ does not contribute to the dynamics. We can therefore choose $R(q) = 0$ in order to assure that $\langle v \rangle_{\rho_\infty} = v_0 + \mathcal{O}(\varepsilon^2)$.

Evaluating and using that $\nabla_q \nabla_q v_0$ is symmetric and Y^{-1} is antisymmetric the reduced backward Kolmogorov equation (44) can be simplified to

$$\frac{\partial v_0}{\partial t} = \left(\frac{2\sigma^4}{\sigma^4 + 16K^2} Y^{-1} I - \frac{16K^2}{\sigma^4 + 16K^2} J \right) \nabla V \cdot \nabla v_0 + \frac{4\sigma^2 K^2}{\sigma^4 + 16K^2} \Delta v_0 \,, \tag{45}$$

where we now drop the subscripts to denote differentiation with respect to q. The slow reduced Langevin equation associated with the reduced backward Kolmogorov equation (45) is then

$$dq = \frac{2\sigma^4}{\sigma^4 + 16K^2} Y^{-1} \nabla V \, dt - \frac{16K^2}{\sigma^4 + 16K^2} J \nabla V \, dt + \frac{2\sqrt{2}\sigma K}{\sqrt{\sigma^4 + 16K^2}} \, dU_t \,, \tag{46}$$

where $K = K(q) = (H - V(q))$ and U_t is a 2-dimensional Wiener process.

Note that for $\sigma = 0$ we have $Y^{-1} = -J$ and we recover the zeroth-order deterministic balance equation (7). For $\sigma = 0$ the slow dynamics evolves along equipotential lines, for small diffusion a trajectory slowly diffuses away from equipotential lines and will explore the whole energetically available phase-space as seen in Figure 4 where we show a trajectory for a (short) simulation of the slow reduced Langevin equation (46) with small diffusion $\sigma = 0.5$ (eventually the trajectory will fill the whole energetically available phase-space).

6. Numerical Results

In this Section we will numerically verify the results from the previous sections. We consider a non-convex quartic potential of the form

$$V(q) = \frac{1}{4} \left(6q_1^4 + 0.1 q_2^4 \right) - \frac{9}{4} \left(q_1^2 + q_2^2 \right) \,. \tag{47}$$

We employ splitting methods (see [40, 24, 33] and in the context of SDEs [26, 34, 55]) to numerically propagate the full projected system (28)-(29) and the reduced system (46). The numerical integrator for the full projected system was constructed to exactly conserve the energy to machine precision (see

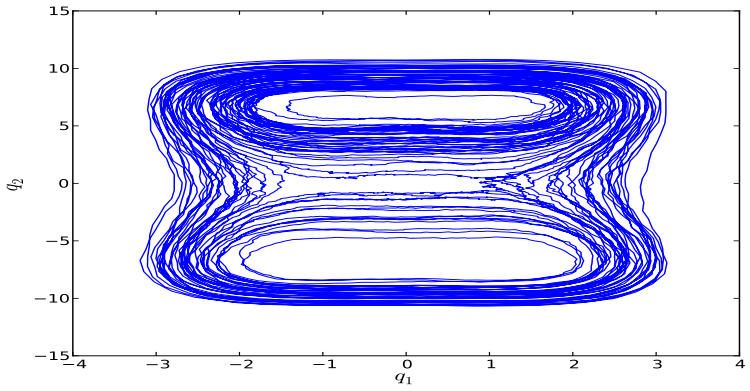

Figure 4: Trajectories of the reduced SDE (46) with $V(q)$ given by (47) and $\sigma = 0.5$.

Figure 5). The numerical challenge in integrating the full projected system is the stiff nature of this multi-scale problem. A detailed description of the numerical algorithms developed has been included in the Appendix.

We have performed stochastic singular perturbation theory for the multi-scale projected system (28)-(29) to derive an effective reduced equation (46) for the slow q-variables. We examine now how the dynamics of the reduced Langevin equation (46) approximates the full stochastically driven projected system (28)-(29). In Figure 6 we show the empirical densities obtained from a simulation of a long trajectory for both systems. The empirical densities are close (when integrated over the whole energetically accessible region the norm of the difference is of the order of 10^{-2}). Furthermore, the empirical densities are uniform on the accessible phase space region, and hence the invariant measure is in agreement with the Lebesgue measure (13) of the deterministic toy model (8). We note that densities in Figure 2 cannot be directly compared with those of Figure 6 because, due to the balanced initial condition in the former, the total energies are different.

In Figure 7 we illustrate the convergence of the full model to the reduced model in the limit $\varepsilon \to 0$. To do so, we compute the solutions to both models for a short interval $t \in [0, 0.2]$ using step size $\Delta t = 10^{-7}$. The solutions are averaged over 10^3 realisations of the Wiener processes W_t and U_t. In Figure 7 we plot the maximum error on the time interval $\max_{t\in[0,0.2]} \|q_{\text{full}}(t) -$

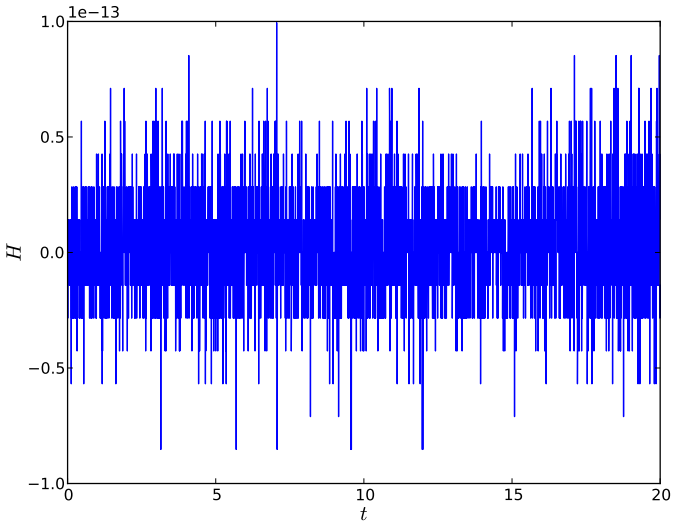

Figure 5: Error in Hamiltonian $H - H_0$ (9) as a function of time during a simulation of the full energy conserving stochastic multi-scale system (28)-(29). Parameters are $\varepsilon = 0.05$ and $\sigma = 0.5$.

$q_{\text{reduced}}(t)\|$. We observe approximately first order convergence in ε.

Temporal characteristics such as the decay time of the auto-correlation function are often not well reproduced by the homogenised models (see for example [18, 19]). Figure 8 shows that remarkably our reduced model reproduces the correlations of the variables q_1 and q_2 reasonably well up to approximately time $t = 0.5$, which we identify as corresponding roughly to the mean period of motion around a potential well. The correlations are defined by

$$C(t) = \int_{\mathcal{S}} (\Phi^t q)\, q^T \, dq_1 \, dq_2,$$

where $\mathcal{S} = \{q \in \mathbb{R}^2 \,|\, V(q) < H_0\}$, and Φ^t denotes the time-t solution map of the respective SDE under the associated realizations of the Wiener processes. Figure 8 compares the matrix elements of $C(t)$, i.e. $c_{11}(t)$, $c_{22}(t)$ and $c_{12}(t)$, for the full and reduced models (note that $c_{21}(t) = -c_{12}(t)$ for reversible systems). For the full model we chose $\varepsilon = 0.01$; no further convergence was observed for smaller ε. To produce the curves, a uniformly distributed, 10^5-member ensemble was integrated to time $T = 2$ using stepsize $\Delta t = 5 \times 10^{-3}$ for the reduced model and $\Delta t = 5 \times 10^{-5}$ for the full model. Note that change

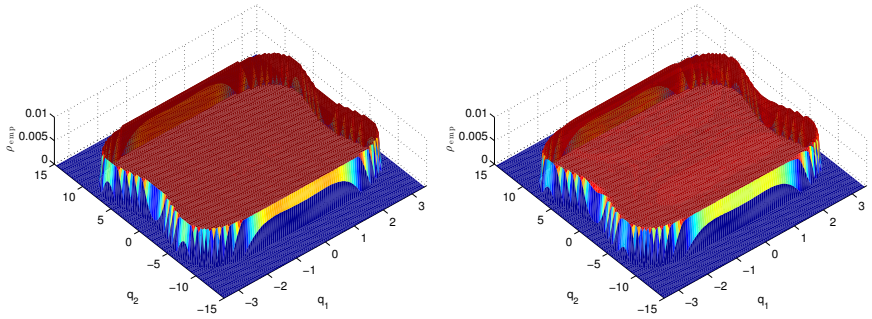

Figure 6: Empirical density $\rho_{\text{emp}}(q)$ generated by the stochastically driven systems. Top: density of the full projected system (28)-(29) with $\varepsilon = 0.05$. Middle: density of the reduced system (46). Both systems were simulated with $\sigma = 0.5$. A time step of $\Delta t = 10^{-3}$ was used, and the simulations were performed for 10^7 time units starting from the initial condition $q(0) = (1.5, 10^{-4})$ and $H = 93.65$.

of sign in q_2 is a rare event on this time scale, leading to the apparent nonzero mean in $c_{22}(t)$.

7. Summary

We have performed a programme for stochastic subgrid scale parametrization for conservative multi-scale systems. In a first step we augment a Hamiltonian slow-fast system by a stochastic process driving the fast dynamics. The stochastic process is chosen so as to conserve the energy of the deterministic core. This choice was motivated by the observation that in numerical weather forecasting the large scales behave like an ideal fluid [50]. However, depending on the application one may choose different conserved quantities such as mass or energy fluxes to be preserved by introducing one auxiliary stochastic process for each conserved quantity. We do not address here the important question whether forcing a conservation law via constraints as done in this work can lead to dynamically inconsistent states. Furthermore, we have not addressed here the issue of how to determine the value of the diffusion coefficients for the energy preserving noise. This is, of course, necessary for real world applications to achieve consistency with observed variability, and the reader is referred to [37, 38] and references therein.

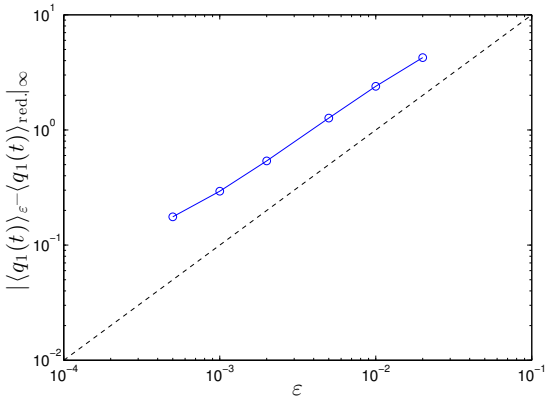

Figure 7: Convergence of the full and reduced models on a short time interval $t \in [0, 0.2]$. The maximum error $\max_t \|q_{\mathrm{full}}(t) - q_{\mathrm{reduced}}(t)\|$ is plotted as a function of ε, for fixed step size $\Delta t = 10^{-7}$. Trajectories were averaged over 10^3 realizations of the Wiener process. The dashed line indicates a first order convergence rate in ε.

In the small diffusion limit, most relevant to stochastic subgrid scale parametrization, we showed that the effect of diffusion combined with energy conservation is to allow for cross-diffusion of the slow dynamics. In a second step the projected stochastic 4-dimensional slow-fast system was reduced to a slow 2-dimensional stochastic differential equation using stochastic singular perturbation theory. This subsequent homogenization and reduction to an effective reduced slow equation has proven to be of practical value. Whereas the full 4-dimensional projected system (29) is highly stiff, the reduced 2-dimensional slow equation (46) can be simulated with significantly larger time steps.

We do not claim that the actual results will have any meaning in interpreting real processes in the atmosphere and oceans. However, we believe that the general strategy to attack the problem may be useful for more realistic models of the atmosphere. To apply the techniques presented here to the Hamiltonian Particle-Mesh method for the rotating shallow water equation is planned for further research. The unresolved scales which are below the smoothing length would then be modelled by Gaussian noise on the Lagrangian particle side, the hope being that one can run a coarse-resolution particle simulation achieving comparable accuracy to high-resolution Eulerian methods.

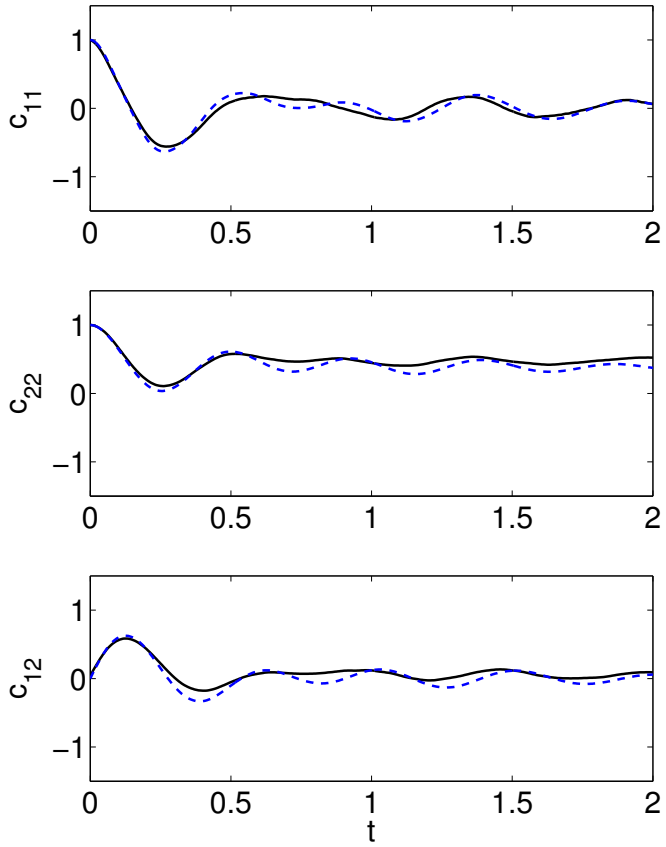

Figure 8: Comparison of the correlation functions of q_1 an q_2 under the full (solid) and reduced (dashed) models. In the full model $\varepsilon = 0.01$. Top: auto-correlation of q_1. Middle: auto-correlation of q_2. Bottom: cross-correlation.

Acknowledgments

GAG thanks Ben Goldys for fruitful discussions. GAG acknowledges support from the Australian Research Council. JEF acknowledges support from the Netherlands Organization for Scientific Research under project number 613.001.009.

Appendix A. Numerical methods

In this appendix we describe the numerical methods used to compute the full projected system (28)-(29) and the reduced system (46). In both cases we make use of splitting methods (see [40, 24, 33] and in the context of SDEs [26]). To construct a splitting method, the vector fields of the SDE are decomposed as a sum of simpler (usually integrable) vector fields, for each of which either an explicit solution or a numerical approximation with desirable properties is available. Numerical methods of various orders can be defined via composition of these flow maps.

For convenience we recall the stochastic full projected model (28)-(29)

$$\begin{aligned} dq &= \frac{1}{\varepsilon} p \, dt \\ dp &= \left[\frac{1}{\varepsilon^2} Jp - \frac{1}{\varepsilon} \nabla V - \frac{\sigma^2}{2\varepsilon^2 \left|p\right|^2} p \right] dt + \frac{1}{\varepsilon} \sigma \mathbb{P} \, dW_t \, , \end{aligned}$$

which we split into a Hamiltonian part:

$$dq = \varepsilon^{-1} p dt, \tag{A.1}$$

$$dp = \left[\frac{1}{\varepsilon^2} Jp - \frac{1}{\varepsilon} \nabla V \right] dt. \tag{A.2}$$

and a projected Langevin perturbation:

$$\begin{aligned} dq &= 0, \\ dp &= -\frac{\sigma^2}{2\varepsilon^2 |p|^2} p \, dt + \frac{\sigma}{\varepsilon} \mathbb{P} \, dW_t. \end{aligned} \tag{A.3}$$

To solve the Hamiltonian part (A.1)–(A.2), we make use of an averaged vector field (AVF) method [4]. The Hamiltonian $H = \frac{1}{2}\|p\|^2 + V(q)$ is

quartic due to the potential (47). An exact quartic conserving method for a Hamiltonian system of the form $dy/dt = J\nabla H(y)$ is given by

$$y^{n+1} = y^n + \Delta t J \left[\nabla H(y^-) + \nabla H(y^+)\right], \tag{A.4}$$

where

$$y^- = \alpha y^n + (1-\alpha)y^{n+1}, \qquad y^+ = (1-\alpha)y^n + \alpha y^{n+1},$$

and $\alpha = (1-\sqrt{3})/2$ corresponds to a Gaussian quadrature point.

Since the Hamiltonian part (A.1)–(A.2) clearly conserves energy, and the whole system is energy conserving, it follows that the flow of (A.3) must also conserve the kinetic energy $K = \frac{1}{2}|p|^2$. This can also be checked by computing the Itô derivative of K and substituting the above derivative dp into the result. Hence, we introduce polar coordinates

$$\gamma = |p| = \sqrt{2K}, \qquad \theta = \tan^{-1}\left(\frac{p_2}{p_1}\right),$$

and compute the Itô derivative of θ as follows:

$$d\theta = \frac{\partial\theta}{\partial p_1}\,dp_1 + \frac{\partial\theta}{\partial p_2}\,dp_2 + \frac{1}{2}\left(\frac{\partial^2\theta}{\partial p_1^2}\,dp_1^2 + \frac{\partial^2\theta}{\partial p_2^2}\,dp_2^2 + 2\frac{\partial^2\theta}{\partial p_1\partial p_2}\,dp_1\,dp_2\right), \tag{A.5}$$

where one can check that

$$\frac{\partial\theta}{\partial p_1} = -\frac{\sin\theta}{\gamma}, \qquad \frac{\partial\theta}{\partial p_2} = \frac{\cos\theta}{\gamma},$$
$$\frac{\partial^2\theta}{\partial p_1^2} = \frac{2\cos\theta\sin\theta}{\gamma^2}, \qquad \frac{\partial^2\theta}{\partial p_2^2} = -\frac{2\cos\theta\sin\theta}{\gamma^2}, \qquad \frac{\partial^2\theta}{\partial p_1\partial p_2} = \frac{\sin^2\theta - \cos^2\theta}{\gamma^2}.$$

Substituting the above relations and the differentials dp_1 and dp_2 into (A.5) and simplifying yields

$$\begin{aligned}
d\theta = -\frac{\sin\theta}{\gamma}&\left\{-\frac{\sigma^2}{2\varepsilon^2\gamma^2}\gamma\cos\theta\,dt + \frac{\sigma}{\varepsilon\gamma^2}(p_2^2\,dW_1 - p_1p_2\,dW_2)\right\}\\
&+\frac{\cos\theta}{\gamma}\left\{-\frac{\sigma^2}{2\varepsilon^2\gamma^2}\gamma\sin\theta\,dt + \frac{\sigma}{\varepsilon\gamma^2}(-p_1p_2\,dW_1 + p_1^2\,dW_2)\right\}\\
&+\frac{\cos\theta\sin\theta}{\gamma^2}\frac{\sigma^2}{\varepsilon^2\gamma^4}(p_2^4 + p_1^2p_2^2)\,dt\\
&-\frac{\cos\theta\sin\theta}{\gamma^2}\frac{\sigma^2}{\varepsilon^2\gamma^4}(p_1^2p_2^2 + p_1^4)\,dt\\
&+\frac{\sin^2\theta - \cos^2\theta}{\gamma^2}\frac{\sigma^2}{\varepsilon^2\gamma^4}(-p_1p_2^3 - p_1^3p_2)\,dt,
\end{aligned}$$

$$= \frac{\sigma}{\varepsilon\gamma}\left(-\sin\theta\, dW_1 + \cos\theta\, dW_2\right),$$

the right side of which is a linear combination of independent normally distributed random increments. This linear combination is again normally distributed with variance $\sin^2\theta + \cos^2\theta = 1$. Therefore we can replace the right side with a new Wiener process dW_3, i.e.

$$d\theta = \frac{\sigma}{\varepsilon\gamma} dW_3.$$

It follows that the flow of (A.3) is simply a rotation of p through an angle proportional to dW_3. We combine this symmetrically with the flow map (A.4) to obtain the composite method:

$$\tilde{p}^n = \exp(d\theta(\Delta t/2)J)p^n, \tag{A.6}$$

$$q^{n+1} = q^n + \Delta t\left(\frac{\tilde{p}^{n+1} + \tilde{p}^n}{2}\right), \tag{A.7}$$

$$\tilde{p}^{n+1} = \tilde{p}^n + \Delta t\left\{\frac{1}{\varepsilon^2} J\,\frac{\tilde{p}^{n+1} + \tilde{p}^n}{2} - \frac{1}{2\varepsilon}\left[\nabla V(q^-) + \nabla V(q^+)\right]\right\}, \tag{A.8}$$

$$\tilde{p}^{n+1} = \exp\left(d\theta(\Delta t/2)J\right)\tilde{p}^{n+1}, \tag{A.9}$$

where $d\theta(\Delta t) \sim \sqrt{\Delta t/2}\,\mathcal{N}(0,1)$. We use (A.6)–(A.9) to solve (28)-(29) numerically in this paper.

For the reduced model (46), we split into a potential energy conserving flow

$$dq = B(K)\nabla V(q)\, dt, \qquad B(K) = -\left(\frac{32K^2\sigma^4}{(\sigma^4 + 16K^2)^2} + \frac{16K^2}{\sigma^4 + 16K^2}\right) J \tag{A.10}$$

and a stochastic-gradient flow

$$dq = -b(K)\nabla V(q)\, dt + d(K)\, dU_t, \tag{A.11}$$

where

$$b(K) = \frac{8K\sigma^6}{(\sigma^4 + 16K^2)^2}, \quad d(K) = \frac{2\sqrt{2}\sigma K}{\sqrt{\sigma^4 + 16K^2}}.$$

Note that since the flow of (A.10) conserves the potential energy $V(q)$, it follows that $K = H - V(q)$ is constant along the solution. Therefore, we

can fix $B(K) = const.$ and apply the AVF method (A.4) to (A.10) to obtain an energy preserving integrator.

We symmetrically apply the Euler-Maruyama method (see for example [31]) to (A.11) and the AVF method (A.4) to (A.10), to obtain:

$$K^n = H - V(q^n), \tag{A.12}$$

$$\tilde{q}^n = q^n - \frac{\Delta t}{2} b(K^n)\nabla V(q^n) + d(K^n)\Delta U(\Delta t/2), \tag{A.13}$$

$$\tilde{K}^n = H - V(\tilde{q}^n), \tag{A.14}$$

$$\tilde{q}^{n+1} = \tilde{q}^n + \frac{\Delta t}{2} B(\tilde{K}^n)\left(\nabla V(q^-) + \nabla V(q^+)\right), \tag{A.15}$$

$$\tilde{K}^{n+1} = \tilde{K}^n, \tag{A.16}$$

$$q^{n+1} = \tilde{q}^{n+1} - \frac{\Delta t}{2} b(\tilde{K}^{n+1})\nabla V(\tilde{q}^{n+1}) + d(\tilde{K}^{n+1})\Delta U(\Delta t/2). \tag{A.17}$$

We use (A.12)–(A.17) to solve (46) numerically in this paper.

[1] Arnold, L., Hasselmann's program revisited: The analysis of stochasticity in deterministic climate models. In: P. Imkeller, J.-S. von Storch, eds., Stochastic Climate Models, (Birkhäuser, Boston, 2001).

[2] Bokhove, O. and Oliver, M., Parcel Eulerian-Lagrangian fluid dynamics for rotating geophysical flows, *Proc. Roy. Soc. A* **462** (2006) 2575–2592.

[3] Brasseur, G. P., Orlando, J. J. and Tyndall, G.S., (Eds.). Atmospheric Chemistry and Global Change, Oxford University Press, (1999).

[4] Celledoni, E., McLachlan, R. I., McLaren, D. I., Owren, B., Quispel, G. R. W. and Wright, W. M., Energy-preserving Runge-Kutta methods, *ESAIM: Mathematical Modelling and Numerical Analysis* **43** (2009) 645–649.

[5] Ciccotti, G., Lelievre, T. and Vanden-Eijnden, E., Projection of diffusions on submanifolds: Application to mean force computation, *Commun. Pure Appl. Math.* **61** (2008) 371–408.

[6] Cotter, C. J., Frank, J. and Reich, S. , Hamiltonian Particle-Mesh Method for Two-Layer Shallow-Water Equations Subject to the Rigid-Lid Approximation. *SIAM J. Appl. Dyn. Syst.* **3** (2004), 69–83.

[7] Cotter, C. J., Frank, J. and Reich, S. , A remapped particle-mesh semi-Lagrangian advection scheme. *Q. J. R. Meteorol. Soc.* **133** (2007), 251–260.

[8] Cotter, C. and Reich, S., Semi-geostrophic particle motion and exponentially accurate normal forms, *SIAM Multi-scale Model. Simul.* **5** (2006), 476–496.

[9] Cullen, M. J. P., Large-Scale Atmosphere/Ocean Flow, Imperial College Press, London, 2006.

[10] Daners, D., Krahn's proof of the Rayleigh conjecture revisited, *Archiv der Mathematik*, **96** (2011), 187-199.

[11] Duan, J., and Nadiga, B. T., Stochastic parameterization for large eddy simulation of geophysical flows. *Proc. Ameri. Math. Soc.* **135** (2006), 1187–1196.

[12] Dubinkina, S. and Frank, J., Statistical relevance of vorticity conservation with the Hamiltonian particle-mesh method. *J. Comput. Phys.* **229** (2010), 2634–2648.

[13] Frank, J. and Gottwald, G. A., The Langevin equation limit of the Nosé-Hoover-Langevin thermostat. *J. Stat. Phys.* **143** (2011), 715–724.

[14] Frank, J., Gottwald, G. A. and Reich, S., A Hamiltonian Particle-Mesh Method for the rotating shallow water equations, in M. Griebel and M.A. Schweitzer, eds., *Meshfree Methods for Partial Differential Equations, Lecture Notes in Computational Science and Engineering*, Vol. 26, pp. 131–142, Springer, 2002.

[15] Frank, J. and Reich, S., Conservation properties of smoothed particle hydrodynamics applied to the shallow water equations. *BIT* **43** (2003), 40–54.

[16] Frank, J. and Reich, S., The Hamiltonian Particle-Mesh Method for the spherical shallow water equations. *Atmospheric Science Letters* **5** (2004), 89–95.

[17] Frank, J., Reich, S., Staniforth, A., White, A. and Wood, N., Analysis of a regularized, time-staggered discretization method and its link to the semi-implicit method. *Atmospheric Science Letters* **6** (2005), 97–104.

[18] Franzke, C., Majda, A. J. and Vanden-Eijnden, E., Low-order stochastic mode reduction for a realistic barotropic model climate. *Journal of the Atmospheric Sciences* **62** (2005), 1722–1745.

[19] Franzke, C. and Majda, A. J., Low-order stochastic mode reduction for a prototype atmospheric GCM. *Journal of the Atmospheric Sciences* **63** (2006), 457–479.

[20] Frenkel, D. and Smit, B., Understanding molecular simulations. In: From Algorithms to Applications, *Academic Press, San Diego* (1996).

[21] Givon, D., Kupferman, R. and Stuart, A. M., Extracting macroscopic dynamics: model problems and algorithms, *Nonlinearity* **17** (2004), R55–R127.

[22] Gottwald, G. A., Oliver, M. and Tecu, N., Long-time accuracy for approximate slow manifolds in a finite dimensional model of balance. *Journal of Nonlinear Science* **17** (2007), 383–407.

[23] Gottwald, G. A. and Oliver, M., Slow dynamics via degenerate variational asymptotics. submitted

[24] Hairer, E., Lubich, C. and Wanner, G., Geometric Numerical Integration: Structure Preserving Methods for Ordinary Differential Equations, Springer (2006).

[25] Hall, P. and Heyde, C. C., Martingale Limit Theory and Its Application, Academic Press, New York (1980).

[26] Halley, W., Malham, S. J. A. and Wiese, A., Positive stochastic volatility simulation. Preprint (2008) arXiv:0802.4411v1.

[27] Hasselmann, K., Stochastic climate models: Part I. Theory. *Tellus* **28** (1976), 473–485.

[28] Just, W., Kantz, H., Rödenbeck, C. and Helm, M., Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A* **34** (2001), 3199–3213.

[29] Kahane, J. P., Some Random Series of Functions, Cambridge University Press, Cambridge (1985).

[30] Khasminsky, R. Z., On stochastic processes defined by differential equations with a small parameter. *Theory Prob. Applications* **11** (1966), 211–228.

[31] Kloeden, P. E. and Platen, E., Numerical Solution of Stochastic Differential Equations, Springer Verlag, New-York (1992).

[32] Kurtz, T. G., A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis* **12** (1973), 55–67.

[33] Leimkuhler, B. and Reich, S., Simulating Hamiltonian Dynamics, Cambridge University Press (2005).

[34] Leimkuhler, B. and Matthews, C., Rational Construction of Stochastic Numerical Methods for Molecular Sampling, *Appl. Math. Res. Express* (2012) doi: 10.1093/amrx/abs010.

[35] Leliévre, T., Le Bris, C. and Vanden-Eijnden, E., Analyse de certains schémas de discrétisation pour des équations différentielles stochastiques contraintes. *C. R. Acad. Sci. Paris, Ser. I* **346** (2008), 471-476.

[36] Majda, A. J., Timofeyev, I. and Vanden Eijnden, E., Models for stochastic climate prediction, *PNAS* **96** (1999), 14687–14691.

[37] Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E., A mathematics framework for stochastic climate models. *Commun. Pure Appl. Math.* **54** (2001) 891-974.

[38] Majda, A. J., Timofeyev, I. and Vanden Eijnden, E., A priori tests of a stochastic mode reduction strategy. *Physica D* **170** (2002), 206–252.

[39] Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E., Systematic strategies for stochastic mode reduction in climate, *Journal of the Atmospheric Sciences* **60** (2003), 1705–1722.

[40] McLachlan, R. I. and Quispel, G. R. W., Splitting methods, *Acta Numerica* (2002), 341–434.

[41] Molchanov, V. and Oliver, M., Convergence of the Hamiltonian Particle-Mesh Method for barotropic fluid flow, *Math. Comp.* (2012), in press.

[42] Nadiga, B. T. and Luce, B., Global bifurcation of Shilnikov type in a double gyre ocean. *J. Phys.Ocean.* **31** (2001), 2669–2690.

[43] Oliver, M., Variational asymptotics for rotating shallow water near geostrophy: A transformational approach, *J. Fluid. Mech.* **551** (2006), 197–234.

[44] Oksendal, B., Stochastic Differential Equations, Springer-Verlag, New York, 1995.

[45] Papanicolaou, G. C. , Introduction to the asymptotic analysis of stochastic equations, *Modern Modeling of Continuum Phenomena* ed R C DiPrima (Providence, RI: AMS), (1974).

[46] Pavliotis, G. A. and Stuart, A. M., multi-scale Methods – Averaging and Homogenization. *Texts in Applied Mathematics* **53**, Springer, New York, (2008).

[47] Pedlosky, J. P. , Geophysical Fluid Dynamics, Springer-Verlag, New York, 1987.

[48] Riegert, A., Baba, N., Gelfert, K., Just, W. and Kantz, H., Hamiltonian chaos acts like a finite energy reservoir: Accuracy of the Fokker-Planck approximation. *Phys. Rev. Lett.* **94** (2005), 054103.

[49] Rödenbeck, C., Beck, C. and Kantz, H., Dynamical Systems with Time Scale Separation: Averaging, Stochastic Modelling, and Central Limit Theorem. In: P. Imkeller, J.-S. von Storch, eds., Stochastic Climate Models, (Birkhäuser, Boston, 2001).

[50] Salmon, R., Lectures on Geophysical Fluid Dynamics, Oxford University Press, New York, 1998.

[51] Shin, S., Reich, S. and Frank, J., Hydrostatic Hamiltonian particle-mesh (HPM) methods for atmospheric modeling. *Q. J. R. Meteorol. Soc.* **138** (2012), 1388–1399.

[52] Stinis, P., A comparative study of two stochastic mode reduction methods. *Physica D* **213** (2006), 197–213.

[53] Stinis, P., Higher order Mori-Zwanzig models for the Euler equations. *Multi-scale Modeling and Simulation* **6** (2007), 741–760.

[54] Vallis, G. K., Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, Cambridge, 2006.

[55] Zygalakis, K. C., On the Existence and Applications of Modified Equations for Stochastic Differential Equations, *SIAM J. Sci. Comput.*, **33** (2011) 102–130.