Statistical Inference for Multiscale Diffusions

G.A. Pavliotis
Department of Mathematics Imperial College London
and
Freie Universität Berlin Institut für Mathematik

23/03/2012
Workshop on Critical Transitions in Complex Systems
Imperial College London

- We are given data (a time-series) from a high-dimensional, multiscale deterministic or stochastic system.
- We want to fit the data to a "simple" low-dimensional, coarse-grained stochastic system.
- The available data is incompatible with the desired model at small scales.
- Many applied statistical techniques use the data at small scales.
- This might lead to inconsistencies between the data and the desired model fit.
- Additional sources of error (measurement error, high frequency noise) might also be present.
- Problems of this form arise in, e.g.
  - Molecular dynamics.
  - Econometrics.
  - Atmosphere/Ocean Science.

Data-Driven Coarse Graining

- We want to use the available data to obtain information on how to parameterize small scales and obtain accurate reduced, coarse-grained models.
- We want to develop techniques for filtering out observation error, high frequency noise from the data.
- We investigate these issues for some simple models.

- Consider a high dimensional dynamical system Z_t with state space $\mathcal{Z}$.
- Assume that the system has two-characteristic time scales, write $\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}$ with $\dim(\mathcal{X}) \ll \dim(\mathcal{Y})$.
- Assume that a coarse-grained equation for the dynamics in $\mathcal{X}$ exists:

$$dX_t = F(X_t)\, dt + \Sigma(X_t)\, dW_t.$$

- Goal: obtain $F(\cdot)$, $\Sigma(\cdot)$ from a time series of the slow variable $X_t = \mathbb{P} Z_t$, $\mathbb{P} : \mathcal{Z} \to \mathcal{X}$.
- In this talk: assume that the functional form of the coarse-grained drift and diffusion coefficients are known:

$$dX_t = F(X_t; \theta)\, dt + \Sigma(X_t; \theta)\, dW_t,$$

  with $\theta \in \Theta \subset \mathbb{R}^d$.
- Goal: estimate these parameters from observations.

Thermal Motion in a Two-Scale Potential

A.M. Stuart and G.P., J. Stat. Phys. 127(4) 741-781, (2007).

- Consider the SDE

$$dx^{\varepsilon}(t) = -V'\left(x^{\varepsilon}(t), \frac{x^{\varepsilon}(t)}{\varepsilon}; \alpha\right) dt + \sqrt{2\sigma}\, dW(t), \tag{1}$$

- Separable potential, linear in the coefficient α:

$$V(x, y; \alpha) := \alpha V(x) + p(y).$$

- $p(y)$ is a mean-zero smooth periodic function.
- $x^{\varepsilon}(t) \Rightarrow X(t)$ weakly in $C([0, T]; \mathbb{R}^d)$, the solution of the homogenized equation:

$$dX(t) = -AV'(X(t))dt + \sqrt{2\Sigma} dW(t).$$

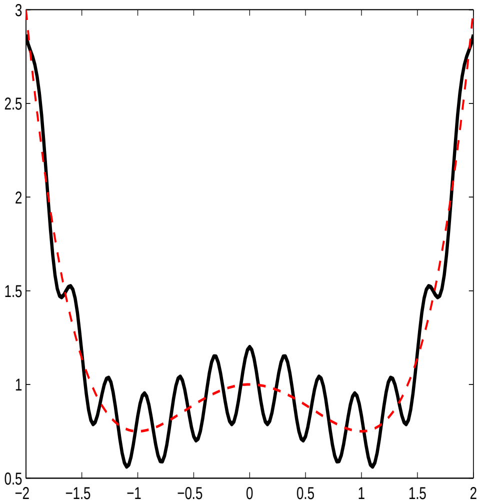


Figure: Bistable potential with periodic fluctuations

- The coefficients A, Σ are given by the standard homogenization formulas.
- Goal: fit a time series of $x^{\varepsilon}(t)$, the solution of (1), to the homogenized SDE.
- Problem: the data is not compatible with the homogenized equation at small scales.
- Model misspecification.
- Similar difficulties when studying inverse problems for PDEs with multiple scales. See: J. Nolen, G.P., A.M. Stuart *Multiscale Modelling and Inverse Problems*. in Lecture Notes in Computational Science and Engineering, Vol. 83, Springer, 2012.

Deriving dynamical models from paleoclimatic records

F. Kwasniok, and G. Lohmann, Phys. Rev. E, 80, **6**, 066104 (2009)

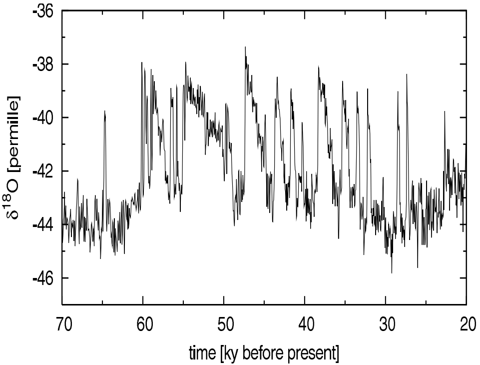


FIG. 1. δ^{18}O record from the NGRIP ice core during the last glacial period.

- Fit this data to a bistable SDE

$$dx = -V'(x;\mathbf{a})\,dt + \sigma \dot{W}, \quad V(x) = \sum_{j=1}^{4} a_j x^j. \tag{2}$$

- Estimate the coefficients in the drift from the palecolimatic data using the unscented Kalman filter.
- the resulting potential is highly asymmetric.

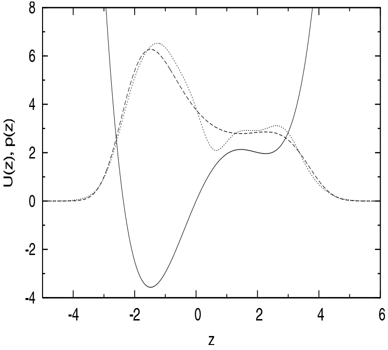


FIG. 8. Potential derived by least-squares fit from the probability density of the ice-core data (solid) together with probability densities of the model (dashed) and the data (dotted).

Estimation of the Eddy Diffusivity from Noisy Lagrangian Observations

C.J. Cotter and G.P. Comm. Math. Sci. 7(4), pp. 805-838 (2009).

- Consider the dynamics of a passive tracer

$$\frac{dx}{dt} = v(x,t), \tag{3}$$

- where $v(x,t)$ is the velocity field. We expect that at sufficiently long length and time scales the dynamics of the passive tracer becomes diffusive:

$$\frac{dX}{dt} = \sqrt{2\mathcal{K}}\frac{dW}{dt} \tag{4}$$

- We are given a time series of noisy observations:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{5}$$

- Goal: estimate the **Eddy Diffusivity** $\mathcal{K}$ from the noisy Lagrangian data (5).

Econometrics: Market Microstructure Noise

S. Olhede, A. Sykulski, G.P. SIAM J. MMS, 8(2), pp. 393-427 (2009)

- Observed process Y_t:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{6}$$

- Where X_t is the solution of

$$dX_t = (\mu - \nu_t/2)\, dt + \sigma_t dB_t, \quad d\nu_t = \kappa\,(\alpha - \nu_t)\, dt + \gamma \nu_t^{1/2} dW_t, \tag{7}$$

- Goal: Estimate the integrated stochastic volatility of X_t from the noisy observations Y_t.
- Work of Ait-Sahalia et al: Estimator fails without subsampling. Subsampling at an optimal rate+averaging+bias correction leads to an efficient estimator.
- We have developed an estimator for the integrated stochastic volatility in the frequency domain.

Homogenization for SPDEs with Quadratic Nonlinearities

D. Blomker, M. Hairer, G.P., Nonlinearity 20 1721-1744 (2007),

M. Pradas Gene, D. Tseluiko, S. Kalliadasis, D.T. Papageorgiou, G.P. Phys. Rev. Lett 106, 060602 (2011).

- Consider the noisy KS equation

$$\partial_t u = -(\partial_x^2 + \nu \partial_x^4) u - u \partial_x u + \tilde{\sigma} \xi, \tag{8}$$

  on 2π-domains with either homogeneous Dirichlet or Periodic Boundary Bonditions. We study the long time dynamics of (8) close to the instability threshold $\nu = 1 - \varepsilon^2$.
- assume that noise acts only on the stable modes (i.e on $Ker(\mathcal{L})^{\perp}$).
- Define $u(x,t) = \varepsilon v(x, \varepsilon^2 t)$.
- For $\varepsilon \ll 1$, $\mathbb{P}_{\mathcal{N}} v \approx X(t) \cdot e(x)$ where $X(t)$ is the solution of the amplitude (homogenized) equation

$$dX_t = (AX_t - BX_t^3)\, dt + \sqrt{\sigma_a^2 + \sigma_b^2 X_t^2}\, dW_t. \tag{9}$$

- There exist formulas for the constants A, B, σ_a^2, σ_b^2 but they involve knowledge of the spectrum of $\mathcal{L} = -(\partial_x^2 + \partial_x^4)$ and the covariance operator of the noise.
- The form of the amplitude equation (9) is universal for all SPDEs with quadratic nonlinearities.
- Goal: assuming knowledge of the functional form of the amplitude equation, estimate the coefficients A, B, σ_a^2, σ_b^2 from a time series of $\mathbb{P}_{\mathcal{N}} u$.

Thermal Motion in a Two-Scale Potential

- Consider the SDE

$$dx^{\varepsilon}(t) = -\nabla V\left(x^{\varepsilon}(t), \frac{x^{\varepsilon}(t)}{\varepsilon}; \alpha\right) dt + \sqrt{2\sigma}\, dW(t),$$

- Separable potential, linear in the coefficient α:

$$V(x, y; \alpha) := \alpha V(x) + p(y).$$

- $p(y)$ is a mean-zero smooth periodic function.
- $x^{\varepsilon}(t) \Rightarrow X(t)$ weakly in $C([0, T]; \mathbb{R}^d)$, the solution of the homogenized equation:

$$dX(t) = -\alpha K \nabla V(X(t))dt + \sqrt{2\sigma K} dW(t).$$

- In one dimension

$$dx^{\varepsilon}(t) = -\alpha V'(x^{\varepsilon}(t))dt - \frac{1}{\varepsilon} p'\left(\frac{x^{\varepsilon}(t)}{\varepsilon}\right) dt + \sqrt{2\sigma}\, dW(t).$$

- The homogenized equation is

$$dX(t) = -AV'(X(t))dt + \sqrt{2\Sigma}\, dW(t).$$

- (A, Σ) are given by

$$A = \frac{\alpha L^2}{Z\widehat{Z}}, \quad \Sigma = \frac{\sigma L^2}{Z\widehat{Z}} \quad Z = \int_0^L e^{-\frac{p(y)}{\sigma}}\, dy, \quad \widehat{Z} = \int_0^L e^{\frac{p(y)}{\sigma}}\, dy.$$

- A and Σ decay to 0 exponentially fast in $\sigma \to 0$.
- The homogenized coefficients satisfy (detailed balance):

$$\frac{A}{\alpha} = \frac{\Sigma}{\sigma}.$$

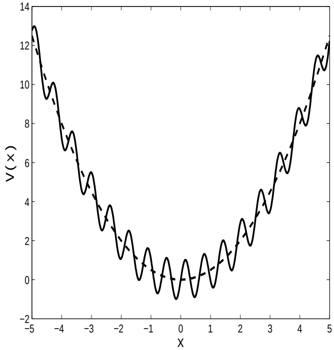


Figure: $V^{\varepsilon}(x)$ and $V(x)$.

- We are given a path of

$$dx^{\varepsilon}(t) = -\alpha V'(x^{\varepsilon}(t))\, dt - \frac{1}{\varepsilon} p' \left(\frac{x^{\varepsilon}(t)}{\varepsilon} \right) dt + \sqrt{2\sigma}\, d\beta(t).$$

- We want to fit the data to

$$dX(t) = -\widehat{A} V'(X(t)) dt + \sqrt{2\widehat{\Sigma}}\, d\beta(t).$$

- It is reasonable to assume that we have some information on the large–scale structure of the potential $V(x)$.
- We do not assume that we know anything about the small scale fluctuations.

- We fit the drift and diffusion coefficients via maximum likelihood and quadratic variation, respectively.
- For simplicity we fit scalars A, Σ in

$$dx(t) = -A\nabla V(x(t))dt + \sqrt{2\Sigma}dW(t).$$

- The Radon–Nikodym derivative of the law of this SDE wrt Wiener measure is

$$\mathbb{L} = \exp\left(-\frac{1}{\Sigma}\int_0^T A\nabla V(x)\, dx(s) - \frac{1}{2\Sigma}\int_0^T \left|A\nabla V(x(s))\right|^2\, ds\right).$$

- This is the maximum likelihood function.

- Let x denote $\{x(t)\}_{t\in[0,T]}$ or $\{x(n\delta)\}_{n=0}^{N}$ with $n\delta = T$.
- Diffusion coefficient estimated from the quadratic variation:

$$\widehat{\Sigma}_{N,\delta}(x)) = \frac{1}{dN\delta}\sum_{n=0}^{N-1}|x_{n+1}-x_n|^2,$$

- Choose $\widehat{A}$ to maximize $\log\mathbb{L}$:

$$\widehat{A}(x) = -\frac{\int_0^T \langle \nabla V(x(s)), dx(s)\rangle}{\int_0^T |\nabla V(x(s))|^2\,ds}$$

- In practice we use the estimators on discrete time data and use the following discretisations:

$$\widehat{\Sigma}_{N,\delta}(x) = \frac{1}{N\delta}\sum_{n=0}^{N-1} |x_{n+1} - x_n|^2,$$

$$\widehat{A}_{N,\delta}(x) = -\frac{\sum_{n=0}^{N-1} \langle \nabla V(x_n), (x_{n+1} - x_n) \rangle}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \, \delta},$$

$$\tilde{A}_{N,\delta}(x) = \widehat{\Sigma}_{N,\delta} \frac{\sum_{n=0}^{N-1} \Delta V(x_n) \delta}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \, \delta},$$

No Subsampling

- Generate data from the unhomogenized equation (quadratic or bistable potential, simple trigonometric perturbation).
- Solve the SDE numerically using Euler–Marayama for a single realization of the noise. Time step is sufficiently small so that errors due to discretization are negligible.
- Fit to the homogenized equation.
- Use data on a fine scale $\delta \ll \varepsilon^2$ (i.e. use all data).
- Parameter estimation fails.

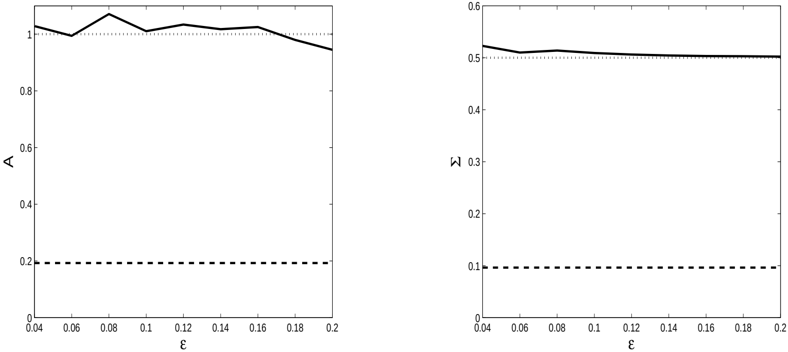


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs ε for quadratic potential.

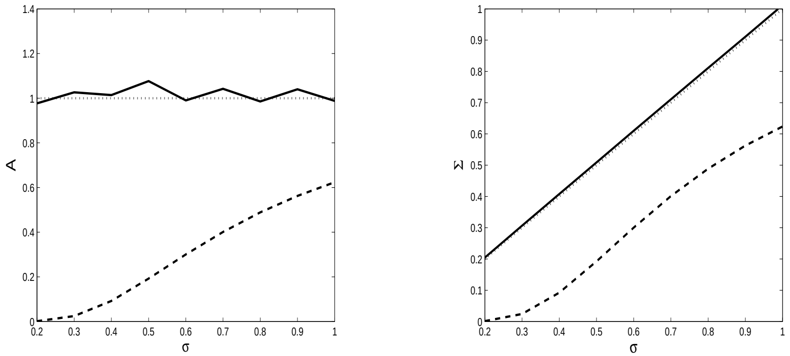


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs σ for quadratic potential with $\varepsilon = 0.1$.

Subsampling

- Generate data from the unhomogenized equation.
- Fit to the homogenized equation.
- Use data on a coarse scale $\varepsilon^2 \ll \delta \ll 1$.
- More precisely

$$\delta := \Delta t_{sam} = 2^k \Delta t, \quad k = 0, 1, \dots .$$

- Study the estimators as a function of Δt_{sam}.
- Parameter Estimation Succeeds.

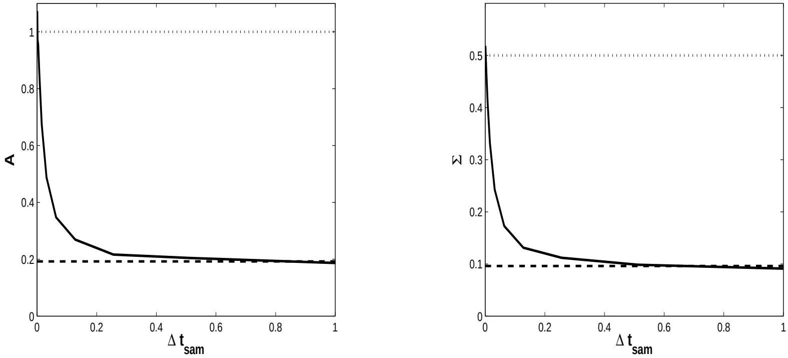


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs Δt_{sam} for quadratic potential with $\varepsilon = 0.1$.

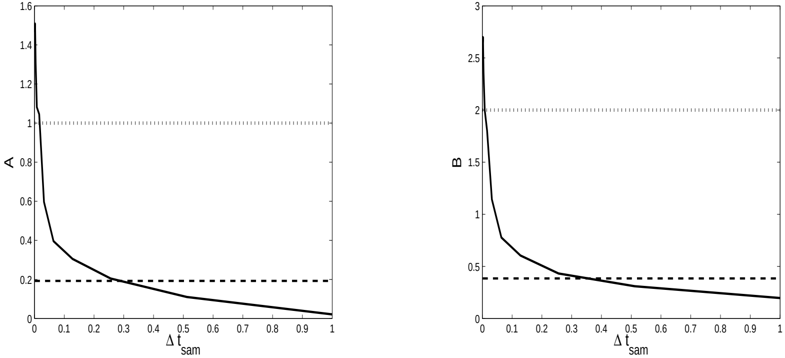


Figure: $\widehat{A}$, $\widehat{B}$ vs Δt_{sam} for bistable potential with $\sigma = 0.5$, $\varepsilon = 0.1$.

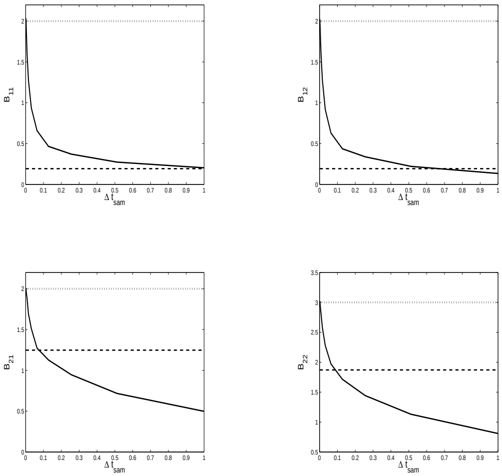


Figure: $\widehat{B}_{ij},\ i,j = 1,2$ vs Δt_{sam} for 2d quadratic potential with $\sigma = 0.5,\ \varepsilon = 0.1$.

Conclusions From Numerical Experiments

- Parameter estimation fails when we take the small–scale (high frequency) data into account.
- $\widehat{A}, \widehat{\Sigma}$ become exponentially wrong in $\sigma \to 0$.
- $\widehat{A}, \widehat{\Sigma}$ do not improve as $\varepsilon \to 0$.
- Parameter estimation succeeds when we subsample (use only data on a coarse scale).
- There is an optimal sampling rate which depends on σ.
- Optimal sampling rate is different in different directions in higher dimensions.

Theorem (No Subsampling)

Let $x^\varepsilon(t) : \mathbb{R}^+ \mapsto \mathbb{R}^d$ be generated by the unhomogenized equation. Then

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}(x^\varepsilon(t)) = \alpha, \quad \textit{a.s.}$$

Fix $T = N\delta$. Then for every $\varepsilon > 0$

$$\lim_{N \to \infty} \Sigma_{N,\delta}(x^\varepsilon(t)) = \sigma, \quad \textit{a.s.}$$

Thus **the unhomogenized parameters are estimated** – the wrong answer.

Theorem (With Subsampling)

Fix $T = N\delta$ *with* $\delta = \varepsilon^{\alpha}$ *with* $\alpha \in (0,1)$. *Then*

$$\lim_{\varepsilon \to 0} \widehat{\Sigma}_{N,\delta}(x^{\varepsilon}) = \Sigma \quad \textit{in distribution.}$$

Let $\delta = \varepsilon^{\alpha}$ *with* $\alpha \in (0,1)$, $N = [\varepsilon^{-\gamma}]$, $\gamma > \alpha$. *Then*

$$\lim_{\varepsilon \to 0} \widehat{A}_{N,\delta}(x^{\varepsilon}) = A \quad \textit{in distribution.}$$

Thus we get the right answer provided **subsampling** is used.

A Fast-Slow System of SDEs

A. Papavasiliou, G.P. A.M. Stuart, Stoch. Proc. Appl. 119(10) 3173-3210 (2009).

- Let (x, y) in $\mathcal{X} \times \mathcal{Y}$. and consider the following coupled systems of SDEs:

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y) + \alpha_0(x,y)\frac{dU}{dt} \\ &\quad + \alpha_1(x,y)\frac{dV}{dt}, \qquad (10a) \\ \frac{dy}{dt} &= \frac{1}{\varepsilon^2} g_0(x,y) + \frac{1}{\varepsilon} g_1(x,y) + \frac{1}{\varepsilon}\beta(x,y)\frac{dV}{dt}. \qquad (10b) \end{aligned}$$

- Here $f_i : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^l$, $\alpha_0 : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^{l \times n}$, $\alpha_1 : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^{l \times m}$, $g_1 : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^{d-l}$ and g_0, β and U, V are independent standard Brownian motions in $\mathbb{R}^n$.
- We will refer to (10) as the **homogenization** problem.
- We assume that the coefficients of SDEs (10) are such that, in the limit as $\varepsilon \to 0$, the slow process x converges weakly in $C([0,T], \mathcal{X})$ to X, the solution of

$$\frac{dX}{dt} = F(X) + K(X)\frac{dW}{dt}. \tag{11}$$

- This can be proved for very general classes of SDEs and formulas for $F(x)$ and $K(x)$ can be obtained (G.P. and A.M. Stuart *Multiscale Methods: Averaging and Homogenization*, Springer 2008).
- Our aim it to estimate parameters in (11) given $\{x(t)\}_{t \in [0,T]}$.

- We want to fit data $\{x(t)\}_{t\in[0,T]}$ to a limiting (homogenized or averaged) equation, but with an unknown parameter θ in the drift:

$$\frac{dX}{dt} = F(X;\theta) + K(X)\frac{dW}{dt}. \tag{12}$$

- We assume that the actual drift that is compatible with the data is given by $F(X) = F(X;\theta_0)$.
- We want to correctly identify $\theta = \theta_0$ by finding the **maximum likelihood estimator** (MLE) when using a statistical model of the form (12), but using data from the slow-fast system.

- Given data $\{z(t)\}_{t\in[0,T]}$, the log likelihood for θ satisfying (12) is given by

$$\mathbb{L}(\theta; z) = \int_0^T \langle F(z;\theta), dz\rangle_{a(z)} - \frac{1}{2}\int_0^T |F(z;\theta)|^2_{a(z)} dt, \tag{13}$$

- where

$$\langle p, q\rangle_{a(z)} = \langle K(z)^{-1}p, K(z)^{-1}q\rangle.$$

- We can define the MLE through

$$\frac{d\mathbb{P}}{d\mathbb{P}_0} = \exp\left(-\mathbb{L}(\theta; X)\right)$$

- where $\mathbb{P}$ is the path space measure for (12) and $\mathbb{P}_0$ the path pace measure for

$$\frac{dX}{dt} = K(X)\frac{dW}{dt}.$$

- The MLE is

$$\hat{\theta} = \text{argmax}_\theta \mathcal{L}(\theta; z).$$

- Assume that we are given data $\{x(t)\}_{t\in[0,T]}$ from (10) and we want to fit it to the equation (12). In this case the MLE is **asymptotically biased**, in the limit as $\varepsilon \to 0$ and $T \to \infty$. The MLE does not converge to the correct value θ_0.

Theorem

Assume that the slow-fast system (10) *as well as the averaged equation* (12) *are ergodic. Let* $\{x(t)\}_{t\in[0,T]}$ *be a sample path of* (10) *and* $X(t)$ *a sample path of* (12) *at* $\theta = \theta_0$. *Then the following limits, to be interpreted in* $L^2(\Omega)$ *and* $L^2(\Omega_0)$ *respectively, are identical:*

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty}\frac{1}{T}\mathbb{L}(\theta;x) = \lim_{T\to\infty}\frac{1}{T}\mathbb{L}(\theta;X) + E_\infty(\theta),$$

with an explicit expression for $E_\infty(\theta)$.

- In order to estimate the the parameter in the drift correctly, we need to **subsample**, i.e. use only a (small) portion of the data that is available to us.
- Assume that we are given observation of $x(t)$ at equidistant discrete points $\{x_n\}_{n=1}^N$ where $x_n = x(n\delta),\ N\delta = T$.
- The log Likelihood function has the form

$$\mathbb{L}^{\delta,N}(z) = \sum_{n=0}^{N-1} \langle F(z_n;\theta), z_{n+1} - z_n \rangle_{a(z_n)} - \frac{1}{2} \sum_{n=0}^{N-1} |F(z_n;\theta)|^2_{a(z_n)} \delta.$$

- If we choose $\delta = \varepsilon^\alpha$ appropriately, then we can estimate the drift parameter correctly.

Theorem

Let $\{x(t)\}_{t\in[0,T]}$ *be a sample path of* (10) *and* $X(t)$ *a sample path of* (12) *at* $\theta = \theta_0$. *Let* $\delta = \varepsilon^{\alpha}$ *with* $\alpha \in (0,1)$ *and let* $N = [\varepsilon^{-\gamma}]$ *with* $\gamma > \alpha$. *Then (under appropriate assumptions) the following limits, to be interpreted in* $L^2(\Omega')$ *and* $L^2(\Omega_0)$ *respectively, and almost surely with respect to* $X(0)$, *are identical:*

$$\lim_{\varepsilon\to 0} \frac{1}{N\delta}\mathbb{L}^{N,\delta}(\theta;x) = \lim_{T\to\infty} \frac{1}{T}\mathbb{L}(\theta;X). \tag{14}$$

Define

$$\hat{\theta}(x;\varepsilon) := \arg\max_{\theta} \mathbb{L}^{N,\delta}(\theta;x).$$

Then, under additional assumptions,

$$\lim_{\varepsilon\to 0} \hat{\theta}(x;\varepsilon) = \theta_0, \text{ in probability}.$$

$$\frac{dx}{dt} = -\nabla V^{\varepsilon}(x) + \sqrt{2\beta^{-1}}\frac{dW}{dt} \tag{15}$$

where

$$V^{\varepsilon}(x) = V(x) + p(x/\varepsilon),$$

where $p(\cdot)$ is a smooth 1-periodic function. The coarse-grained equation is The homogenized equation is

$$\frac{dX}{dt} = -K\nabla V(X) + \sqrt{2\beta^{-1}K}\frac{dW}{dt} \tag{16}$$

where

$$K = \int_{\mathbb{T}^d} (I + \nabla_y \Phi(y))(I + \nabla_y \Phi(y))^T \rho(y)\, dy.$$

- Suppose there is a set of parameters $\theta \in \Theta$ in the large-scale part of the potential

$$\frac{dX}{dt} = -K\nabla V(X;\theta) + \sqrt{2\beta^{-1}K}\frac{dW}{dt}$$

- using data from (15).
- The error in the asymptotic log Likelihood function is:

$$E_\infty(\theta) = \Big(-1 + \widehat{Z}_p^{-1} Z_p^{-1}\Big)\frac{\beta Z_V^{-1}}{2}\int_{\mathbb{R}} |\partial_x V|^2 e^{-\beta V(x;\theta)}\,dx. \tag{17}$$

where $Z_V = \int_{\mathbb{R}} e^{-\beta V(q;\theta)}\,dq$, $Z_p = \int_0^1 e^{-\beta p(y)}\,dy\, \widehat{Z}_p = \int_0^1 e^{\beta p(y)}\,dy$. In particular, $E_\infty < 0$.

Estimating the Eddy Diffusivities from Noisy Lagrangian Observations

C.J. Cotter and G.P. Comm. Math. Sci. 7(4), pp. 805-838 (2009).

- Consider the equation for the Lagrangian trajectories

$$\dot{x} = v(x,t) + \sqrt{2\kappa}\dot{W}. \tag{18}$$

- For $v(x,t)$ being either periodic or random solutions to (18) converge to an effective Brownian motion:

$$\lim_{\varepsilon \to 0} \varepsilon x(t/\varepsilon^2) = \sqrt{2\mathcal{K}} W(t), \tag{19}$$

- weakly on $C([0,T];\mathbb{R}^d)$. At long length-time scales the dynamics of the passive tracer is given by

$$\dot{X} = \sqrt{2\mathcal{K}}\dot{W}. \tag{20}$$

- We want to estimate the eddy diffusivity and other large-scale quantities (e.g. effective drift) from noisy Lagrangian observations.

- The eddy diffusivity (along the direction $\xi \in \mathbb{R}^d$) is given by

$$\mathcal{K}^{\xi} = \kappa \|\nabla_z \chi^{\xi}(z) + \xi\|^2_{L^2(\mathbb{T}^d)}$$

  where

$$-\mathcal{L}_0 \chi = v, \quad \mathcal{L}_0 = v(z) \cdot \nabla_z + \kappa \Delta_z$$

- with periodic boundary conditions.

- The homogenized equation (20) is compatible with the data only at sufficiently large scales.
- We do not know a priori what the right length and time scales are for which the dynamics can be adequately described by (20).
- The **diffusive time scale** at which (20) is valid depends on the detailed properties of $v(x, t)$ and on κ.

- For the estimation of the eddy diffusivity we use

$$\mathcal{K}^{N,\delta} = \frac{1}{2N\delta} \sum_{n=0}^{N-1} \left(x_{n+1} - x_n\right) \otimes \left(x_{n+1} - x_n\right), \tag{21}$$

- where N is the number of observations. We have that

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} \left(x_{(j+1)\Delta t} - x_{j\Delta t}\right) \otimes \left(x_{(j+1)\Delta t} - x_{j\Delta t}\right) = 2\kappa I T, \quad \text{a.s.,} \tag{22}$$

- with $\Delta t N = T$, fixed.
- The eddy diffusivity satisfies the bounds

$$\kappa \leqslant \mathcal{K}^{\xi} \leqslant \frac{C}{\kappa}, \tag{23}$$

- The estimator $\mathcal{K}^{N,\delta}$ underestimates the value of the eddy diffusivity, in particular when $\kappa \ll 1$.

Theorem

Let $v(z)$ be a smooth, divergence-free smooth vector field on $\mathbb{T}^d$. Then

$$\lim_{\kappa \to 0} \mathbb{E} |\mathcal{K}_{\xi}^{N,\delta} - \mathcal{K}_{\xi}|^2 = 0. \tag{24}$$

when $N \sim \kappa^{\zeta}$ and $\delta \sim \kappa^{\gamma}$ where the exponents γ and ζ depend on the properties of the velocity field $v(x, t)$.

- Simply subsampling is clearly not the optimal strategy since we are not using a large portion (almost all!) of the data. Note however that the data that we do not use is highly correlated.
- We combine subsampling with averaging, in order to reduce the bias of the estimator and to remove the measurement error.

- We split the data into N_B bins of size δ with $\delta N_B = N$ and to perform a local averaging over each bin. Let

$$x_n^j := x((n-1)\delta + (j-1)\Delta t), \\ n = 1, \dots N_B, \; j = 1, \dots J, \quad J N_B = N,$$

- be the j-th observation in the n-th bin. $J = \delta / \Delta t$ is the number of observations in each bin. The **box-averaged** estimator is

$$\mathcal{K}_{bx}^{N,\delta} = \frac{1}{2N\delta} \sum_{n=0}^{N-1} \left(\frac{1}{J} \sum_{j=1}^{J} x_{n+1}^j - \frac{1}{J} \sum_{j=1}^{J} x_n^j \right) \otimes \left(\frac{1}{J} \sum_{j=1}^{J} x_{n+1}^j - \frac{1}{J} \sum_{j=1}^{J} x_n^j \right). \tag{25}$$

- Compute a series of estimators, each using a different observation from each bin, and then to compute the average. This is the **shift-averaged** estimator:

$$\mathcal{K}_{st}^{N,\delta} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{2N\delta}\sum_{n=0}^{N-1}\left(x_{n+1}^{j} - x_{n}^{j}\right)\otimes\left(x_{n+1}^{j} - x_{n}^{j}\right). \tag{26}$$

Effect of Observation Error

- Assume that the observed process is

$$Y_{t_j}^{\xi} = X_{t_j}^{\xi} + \theta \varepsilon_{t_j}^{\xi}, \quad j = 1, \dots N, \tag{27}$$

- where $\theta > 0$ and $\varepsilon_{t_j}^{\xi}$ is collection of i.i.d $\mathcal{N}(0,1)$ random variables , independent from the Brownian motion driving the Lagrangian dynamics.
- We have

$$\begin{aligned} \mathbb{E}\left|\mathcal{K}_{N,\delta}^{\xi}(Y_t) - \mathcal{K}^{\xi}\right|^2 &= \mathbb{E}\left|\mathcal{K}_{N,\delta}^{\xi}(X_t) - \mathcal{K}^{\xi}\right|^2 + 3\frac{\theta^4}{\delta^2} \\ &\quad + 2\theta^2\left(\frac{1}{\delta} + \frac{2}{N\delta}\right)(\mathcal{K}^{\xi} + R), \end{aligned}$$

- Provided that , N, $\delta \to \infty$ at an appropriate rate, then

$$\lim_{\kappa \to 0} \mathbb{E}\left|\mathcal{K}_{N,\delta}^{\xi}(Y_t) - \mathcal{K}^{\xi}\right|^2 = 0.$$

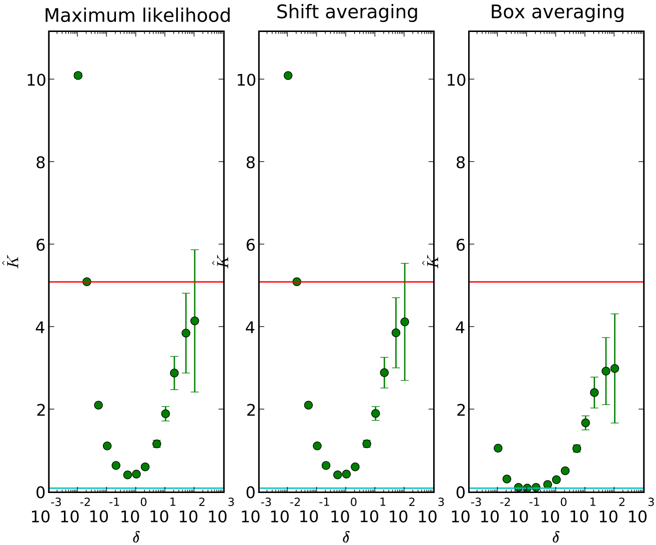


Figure: Estimated eddy diffusivity for shear flow with observation error.

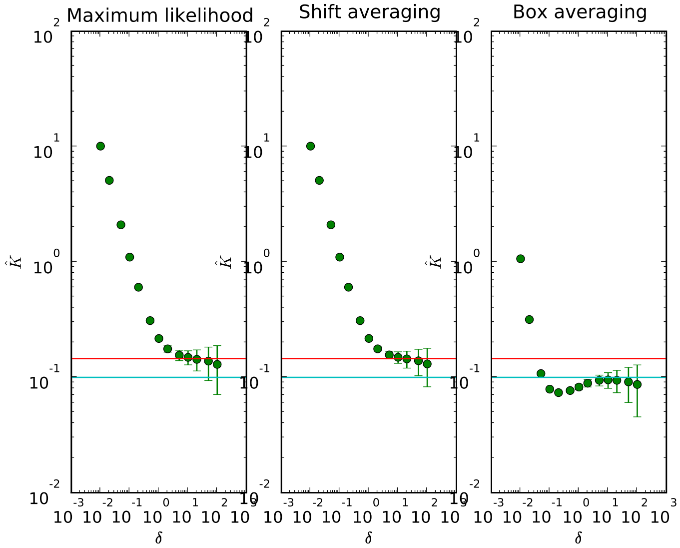


Figure: $\mathcal{K}$ for modulated shear flow with observation error.

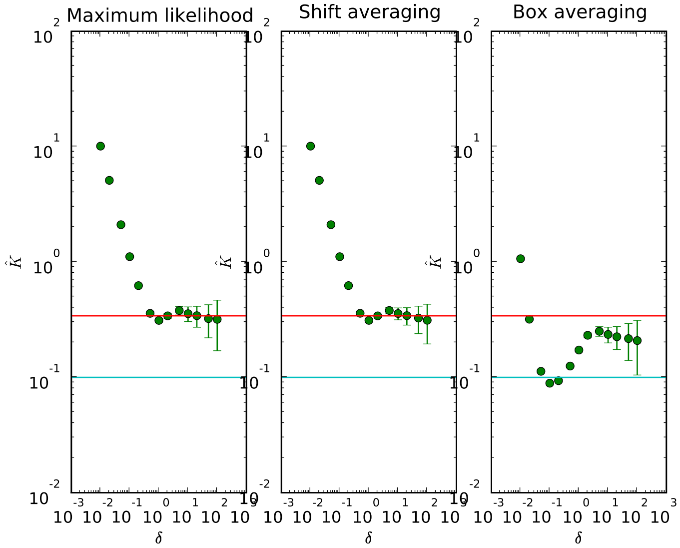


Figure: $\mathcal{K}$ for Taylor-Green flow with observation error.

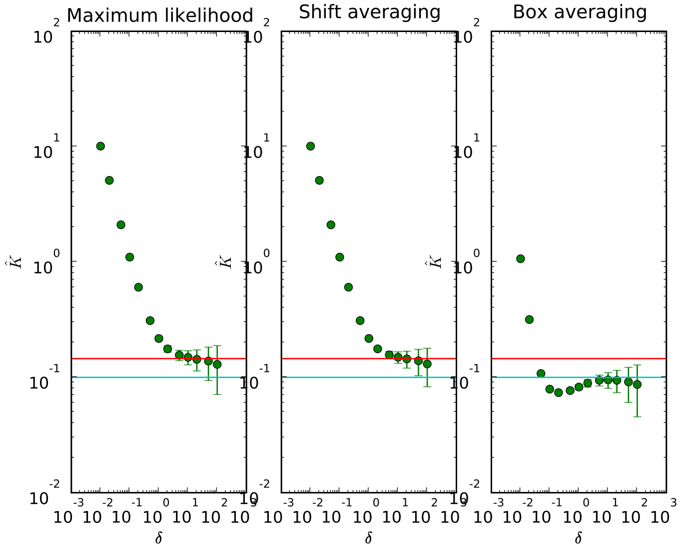


Figure: $\mathcal{K}$ for OU-modulated shear flow with observation.

Semiparametric Drift and Diffusion Estimation

S. Krumscheid, S. Kalliadasis, G.P., Preprint 2011

- Optimal subsampling rate and estimator curves generally unknown
- MLE only feasible for drift parameters.
- QVP only applicable for constant diffusion coefficients.
- We propose new estimators that are applicable in a semiparametric framework and for non-constant diffusion coefficients.

The Estimators

- Scalar-valued Itô SDE

$$dx_t = f(x_t)\,dt + \sqrt{g(x_t)}\,dW_t\ , \quad x(0) = x_0$$

- Parameterization of drift and diffusion coefficient

$$f(x) \equiv f(x;\vartheta) := \sum_{j\in J_f} \vartheta_j x^j \quad \text{and} \quad g(x) \equiv g(x;\theta) := \sum_{j\in J_g} \theta_j x^j$$

Goal

Determine $\vartheta \equiv (\vartheta_j)_{j\in J_f} \in \mathbb{R}^p$ and $\theta \equiv (\theta_j)_{j\in J_g} \in \mathbb{R}^q$, with $J_f, J_g \subset \mathbb{N}_0$

- By the Martingale property of the stochastic integral we find

$$\mathbb{E}(x_t - x_0) = \mathbb{E}\Big(\int_0^t f(x_s)\,ds\Big) = \sum_{j\in J_f} \vartheta_j \int_0^t \mathbb{E}(x_s{}^j)\,ds\ , \text{ for } t>0 \text{ fixed}$$

- This can be rewritten as

$$b_1(x_0) = a_1(x_0)^T\vartheta$$

with $b_1(\xi) := \mathbb{E}_\xi(x_t - \xi) \in \mathbb{R}$ and $a_1(\xi) := \Big(\int_0^t \mathbb{E}_\xi(x_s{}^j)\,ds\Big)_{j\in J_f} \in \mathbb{R}^p$

- Equation $a_1(x_0)^T\vartheta = b_1(x_0)$ is *ill-posed*
- Since the equation is valid for each initial condition, we can overcome this shortcoming by considering *multiple initial conditions* $(x_{0,i})_{1\leqslant i\leqslant m}$, $m \geqslant p$, and obtain

$$A_1\vartheta = b_1$$

with $A_1 := \big(a_1(x_{0,i})^T\big)_{1\leqslant i\leqslant m} \in \mathbb{R}^{m\times p}$, $b_1 := \big(b_1(x_{0,i})\big)_{1\leqslant i\leqslant m} \in \mathbb{R}^m$

- Define estimator to be the *best approximation*

$$\hat{\vartheta} := A_1^+ b_1$$

- Assume now that drift f is already estimated, hence known
- By Itô Isometry and the parameterization of g we find

$$\mathbb{E}\Big(\big(x_t - x_0 - \int_0^t \hat{f}(x_s)\,ds\big)^2\Big) = \mathbb{E}\Big(\int_0^t g(x_s)\,ds\Big) = \sum_{j\in J_g} \theta_j \int_0^t \mathbb{E}(x_s{}^j)\,ds$$

- Provides the *same structure* as for ϑ
- Thus, we can follow the *same steps* as before: Rewriting, considering multiple initial conditions, and taking the best approximation to obtain

$$\hat{\theta} := A_2^+ b_2$$

with A_2 and b_2 defined appropriately

Summary: Two Step Estimation Procedure

1. Estimate drift coefficient via $\hat{\vartheta} := A_1^+ b_1$
2. Based on $\hat{\vartheta}$ estimate diffusion coefficient via $\hat{\theta} := A_2^+ b_2$

Further Approximations

- Discrete Time Data: Approximate integrals via trapezoidal rule
- Approximate expectations via Monte Carlo experiments

Fast OU Process Revisited

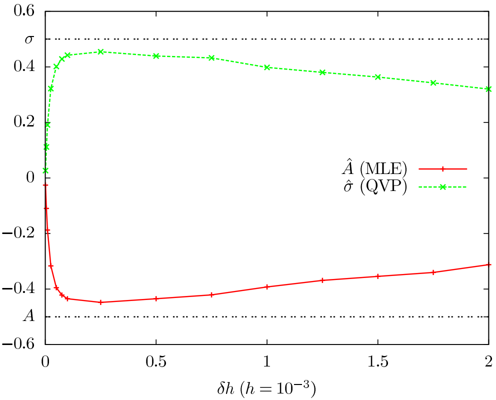
Fast/Slow System

$$dx_t = \left(\frac{\sigma}{\varepsilon} y_t + A x_t\right) dt \,,$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t \, dt + \frac{\sqrt{2}}{\varepsilon} \, dV_t$$

Effective Dynamics

$$dX_t = A X_t \, dt + \sqrt{2\sigma} \, dW_t$$



Fast OU Process Revisited

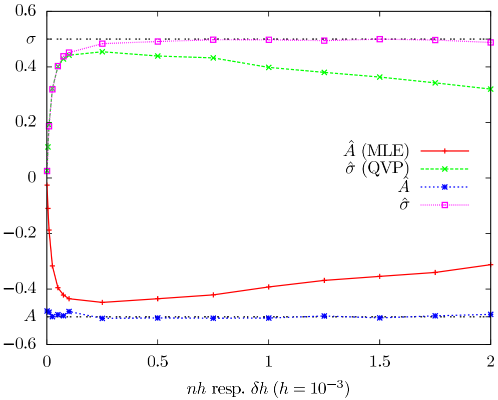
Fast/Slow System

$$dx_t = \left(\frac{\sigma}{\varepsilon} y_t + A x_t\right) dt \,,$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t \, dt + \frac{\sqrt{2}}{\varepsilon} \, dV_t$$

Effective Dynamics

$$dX_t = A X_t \, dt + \sqrt{2\sigma} \, dW_t$$



Fast OU Process II

- *Fast/slow system:*

$$dx_t = \left(\frac{y_t}{\varepsilon}\sqrt{\sigma_a + \sigma_b x_t^2} + (A - \sigma_b)x_t - B{x_t}^3\right) dt \;,$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t \, dt + \frac{\sqrt{2}}{\varepsilon} \, dV_t$$

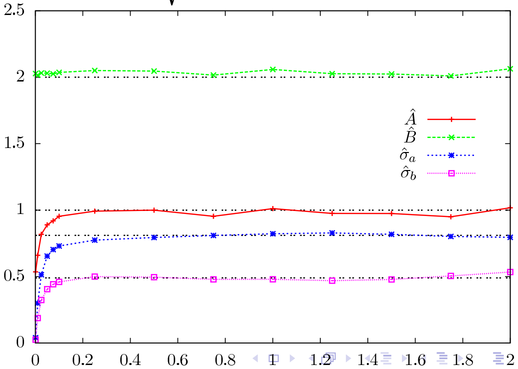
- Effective Dynamics:

$$dX_t = (AX_t - B{X_t}^3)\, dt + \sqrt{2(\sigma_a + \sigma_b {X_t}^2)}\, dW_t$$

- True values:

$$A = 1 \;, \quad \sigma_a = 0.81$$
$$B = 2 \;, \quad \sigma_b = 0.49$$

- $\varepsilon = 0.1$



Brownian Motion in two-scale Potential

[Pavliotis, Stuart '07]

- *Fast/slow system:*

$$dx_t = -\frac{d}{dx} V_\alpha\left(x_t, \frac{x_t}{\varepsilon}\right) dt + \sqrt{2\sigma}\, dU_t$$

- Two-scale potential: $V_\alpha(x,y) = \alpha V(x) + p(y)$, with $p(\cdot)$ periodic
- Effective Dynamics:

$$dX_t = -AV'(X_t)\, dt + \sqrt{2\Sigma}\, dW_t$$

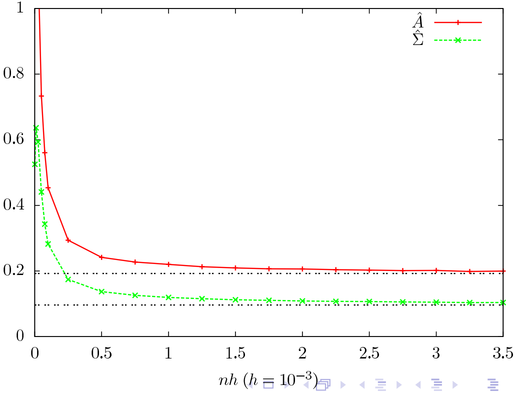
- with:

$$V(x) = x^2/2$$
$$p(y) = \cos(y)$$

- True values:

$$\alpha = 1\,, \quad A \approx 0.192$$
$$\sigma = \frac{1}{2}\,, \quad \Sigma \approx 0.096$$

- $\varepsilon = 0.1$



Fast Chaotic Noise

- *Fast/slow system:*

$$\begin{aligned}\frac{dx}{dt} &= x - x^3 + \frac{\lambda}{\varepsilon} y_2 \,, \\ \frac{dy_1}{dt} &= \frac{10}{\varepsilon^2}(y_2 - y_1) \,, \\ \frac{dy_2}{dt} &= \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3) \,, \\ \frac{dy_3}{dt} &= \frac{1}{\varepsilon^2}(y_1 y_2 - \frac{8}{3} y_3)\end{aligned}$$

- Effective Dynamics: [Melbourne, Stuart '11]

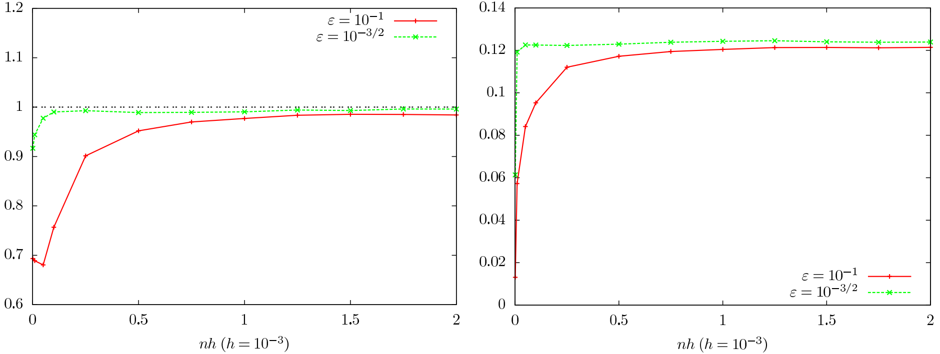
$$dX_t = A\left(X_t - {X_t}^3\right) dt + \sqrt{\sigma}\, dW_t$$

- true values:

$$A = 1 \,, \quad \lambda = \frac{2}{45} \,, \quad \sigma = 2\lambda^2 \int_0^\infty \lim_{T\to\infty} \frac{1}{T} \int_0^T \psi^s(y)\psi^{s+t}(y)\, ds\, dt$$

Fast Chaotic Noise

Estimators



- Values for σ reported in the literature ($\varepsilon = 10^{-3/2}$)
  - 0.126 ± 0.003 via Gaussian moment approx.
  - 0.13 ± 0.01 via HMM

[Givon, Kupferman, Stuart '04]

- here: $\varepsilon = 10^{-1} \rightarrow \hat{\sigma} \approx 0.121$ and $\varepsilon = 10^{-3/2} \rightarrow \hat{\sigma} \approx 0.124$
- But we estimate also $\hat{A}$

Truncated Burgers Equation

- Diffusively time rescaled variant of Burgers' equation

$$du_t = \left(\frac{1}{\varepsilon^2}(\partial_x^2 + 1)u_t + \frac{1}{2\varepsilon}\partial_x u_t^2 + \nu u_t\right) dt + \frac{1}{\varepsilon} Q\, dW_t$$

on an open interval equipped with homogeneous Dirichlet boundary conditions

- Effective dynamics for dominant mode

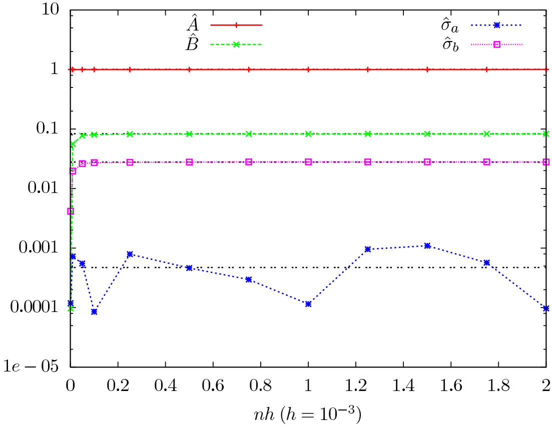
$$dX_t = \left(AX_t - B{X_t}^3\right) dt + \sqrt{\sigma_a + \sigma_b {X_t}^2}\, dW_t$$

- For the three-term truncated representation the true values are:

$$A = \nu + \frac{{q_1}^2}{396} + \frac{{q_2}^2}{352}, \quad B = \frac{1}{12}, \quad \sigma_a = \frac{{q_1}^2 {q_2}^2}{2112}, \text{ and } \quad \sigma_b = \frac{{q_1}^2}{36}$$

Truncated Burgers Equation
Estimators

- $\nu = 1$, $q_1 = 1 = q_2$ and $\varepsilon = 0.1$



Fast Chaotic Noise II

- *Fast/slow system:*

$$\frac{dx}{dt} = x - x^3 + \frac{\lambda}{\varepsilon}(1 + x^2)y_2 \,,$$
$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1) \,,$$
$$\frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3) \,,$$
$$\frac{dy_3}{dt} = \frac{1}{\varepsilon^2}(y_1 y_2 - \frac{8}{3}y_3)$$

- Effective Dynamics:

$$dX_t = \left(AX_t + BX_t{}^3 + CX_t{}^5\right) dt + \sqrt{\sigma_a + \sigma_b X_t{}^2 + \sigma_c X_t{}^4}\, dW_t$$

- true values ($\lambda = 2/45$):

$$A = 1 + \sigma \,, \quad B = \sigma - 1 \,, \quad C = 0 \,, \quad \sigma_a = \sigma \,, \quad \sigma_b = 2\sigma \,, \quad \sigma_c = \sigma \,,$$
$$\sigma = 2\lambda^2 \int_0^\infty \lim_{T\to\infty} \frac{1}{T} \int_0^T \psi^s(y)\psi^{s+t}(y)\, ds\, dt$$

Fast Chaotic Noise

Estimators

- $\varepsilon = 0.1$