Inference and Inverse Problems for Multiscale Diffusions

G.A. Pavliotis
Department of Mathematics Imperial College London

03/10/2014
Stochastic and Multiscale Inverse Problems
Paris

Research supported by the EPSRC through grants EP/H034587/1 and EP/J009636/1

- 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.

- Consider a dynamical system Z_t with phase space $\mathcal{Z}$ that evolves according to the dynamics

$$\frac{dZ_t}{dt} = F(Z_t) \tag{1}$$

- $\dim(\mathcal{Z}) \gg 1$ and $F(\cdot)$ might be only partially known or unknown.
- Our basic modeling assumption is that we are only interested in the evolution of only a few selected degrees of freedom. We separate between the **resolved degrees of freedom (RDoF)** and **unresolved degrees of freedom (UDoF)**:

$$\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}, \tag{2}$$

with $\dim(\mathcal{X}) \ll \dim(\mathcal{Y})$.

- We postulate the existence of a stochastic coarse-grained equation for the RDoF:

$$dX_t = \overline{F}(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{3}$$

where W_t denotes standard Brownian motion in $\mathbb{R}^d$.

- We assume that we are given discrete noisy observations of Z_t, projected onto the space of the RDoF $\mathcal{X}$:

$$\hat{X}_{t_j} = \mathcal{P}\hat{Z}_{t_j} + \eta_j, \quad j = 1, \ldots N. \tag{4}$$

- Our goal is the derivation of the coarse-grained dynamics (3) from the noisy observations (4).
- Consider the problem in both a parametric and a nonparametric framework.

$$\overline{F} = \overline{F}(x;\theta), \quad \sigma = \sigma(x;\vartheta). \tag{5}$$

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.
- More generally: study the following problems for multiscale systems
  - Inference
  - Filtering
  - Control (W. Zhang, J.C. Latorre, G.P., C. Hartmann, to Appear, 2014)
  - Inverse problems
- We investigate these issues for some simple models.

Thermal Motion in a Two-Scale Potential

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

- Consider the SDE

$$dx^{\varepsilon}(t) = -V'\left(x^{\varepsilon}(t), \frac{x^{\varepsilon}(t)}{\varepsilon}; \alpha\right) dt + \sqrt{2\sigma}\, dW(t), \tag{6}$$

- 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:

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

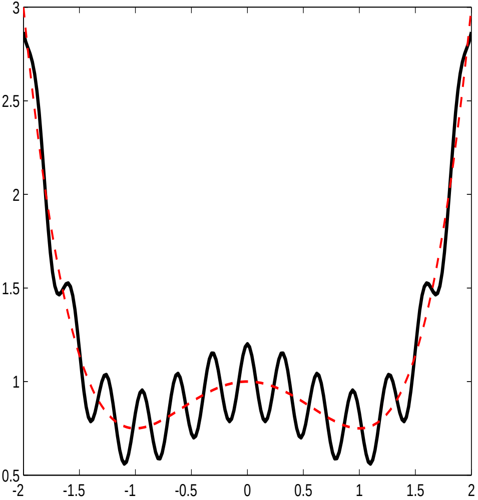


Figure : Bistable potential with periodic fluctuations

- The coefficients A, Σ are given by the standard homogenization formulas.
- Goal: fit a time series of $x^{\varepsilon}(t)$, the solution of (6), 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 a multiscale structure.

Deriving dynamical models from paleoclimatic records

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

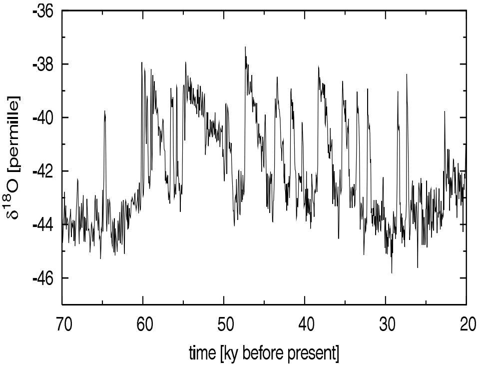


FIG. 1. δ^{18}O record from the NGRIP ice core during the last glacial period.

- Fit this data to a bistable SDE

$$dx = -V'(x; \mathbf{a})\, dt + \sigma\, dW, \quad V(x) = \sum_{j=1}^{4} a_j x^j. \tag{7}$$

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

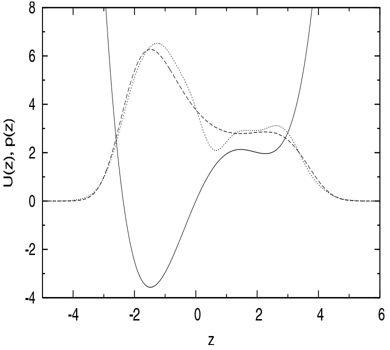


FIG. 8. Potential derived by least-squares fit from the probability density of the ice-core data (solid) together with probability densities of the model (dashed) and the data (dotted).

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

$$\frac{dx}{dt} = v(x,t), \tag{8}$$

- 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:

$$\frac{dX}{dt} = \sqrt{2\mathcal{K}}\frac{dW}{dt} \tag{9}$$

- We are given a time series of noisy observations:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{10}$$

- Goal: estimate the **Eddy Diffusivity** $\mathcal{K}$ from the noisy Lagrangian data (10).

Econometrics: Market Microstructure Noise

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

- Observed process Y_t:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{11}$$

- Where X_t is the solution of

$$dX_t = (\mu - \nu_t/2)\, dt + \sigma_t dB_t, \quad d\nu_t = \kappa\,(\alpha - \nu_t)\, dt + \gamma \nu_t^{1/2} dW_t, \tag{12}$$

- 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

$$\partial_t u = -(\partial_x^2 + \nu \partial_x^4) u - u \partial_x u + \tilde{\sigma} \xi, \tag{13}$$

  on 2π-domains with either homogeneous Dirichlet or Periodic Boundary Bonditions. We study the long time dynamics of (13) 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

$$dX_t = (AX_t - BX_t^3)\, dt + \sqrt{\sigma_a^2 + \sigma_b^2 X_t^2}\, dW_t. \tag{14}$$

- 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 (14) 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 $\mathbb{P}_{\mathcal{N}} u$.
- Can combine ideas from numerical analysis and statistics to develop a numerical method for solving SPDEs of the form (13):
- **Numerical Methods for Stochastic Partial Differential Equations with Multiple Scales** (with A. Abdulle). J. Comp. Phys, 231(6) pp. 2482-2497 (2012).

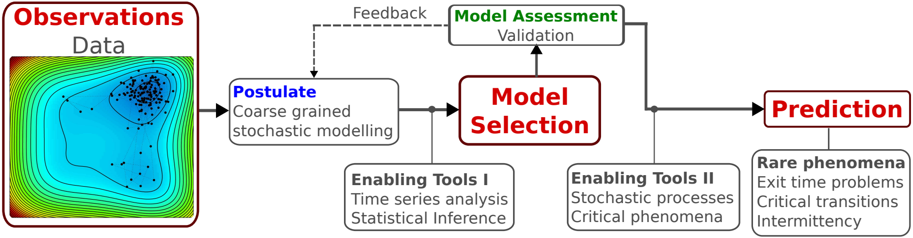


Figure : Flow chart of the data-driven modeling framework: Given observations (data) we postulate a coarse-grained stochastic parametric model which is fitted (via statistical inference and time series analysis tools) to the data and refined via a model selection process.

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):

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

- 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

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

- 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.

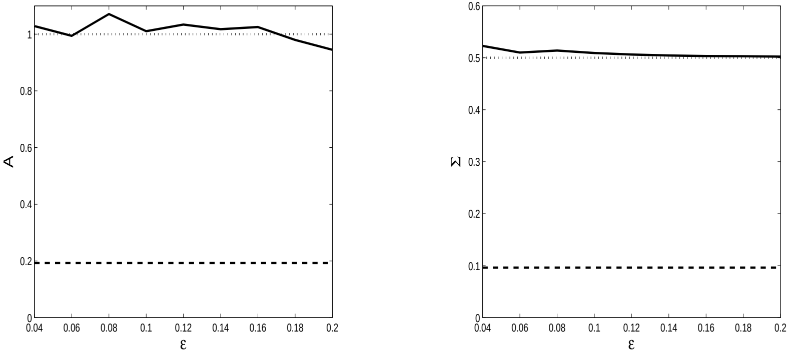


Figure : $\widehat{A}$, $\widehat{\Sigma}$ vs ε for quadratic potential.

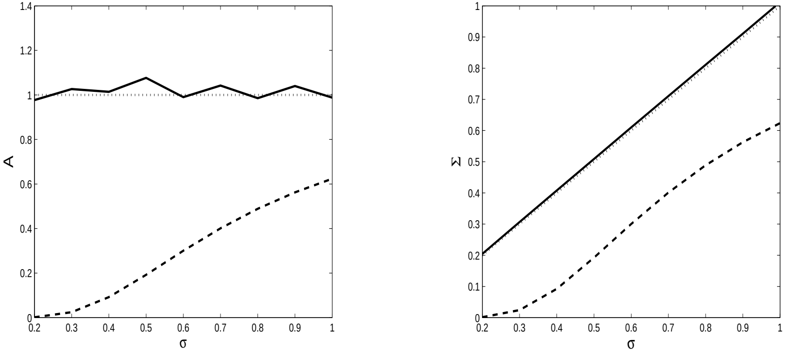


Figure : $\widehat{A}$, $\widehat{\Sigma}$ vs σ for quadratic potential with $\varepsilon = 0.1$.

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

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

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

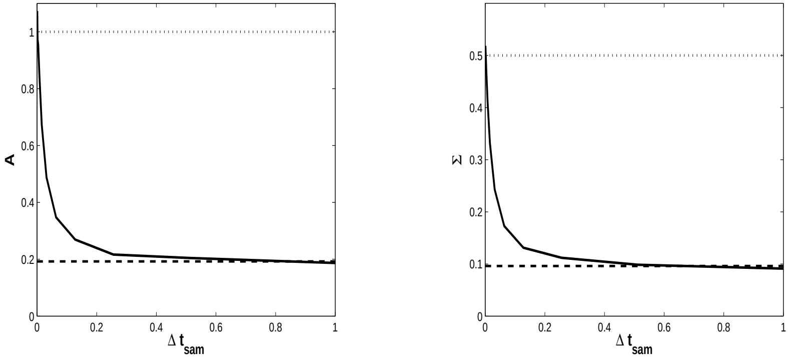


Figure : $\widehat{A}$, $\widehat{\Sigma}$ vs Δt_{sam} for quadratic potential with $\varepsilon = 0.1$.

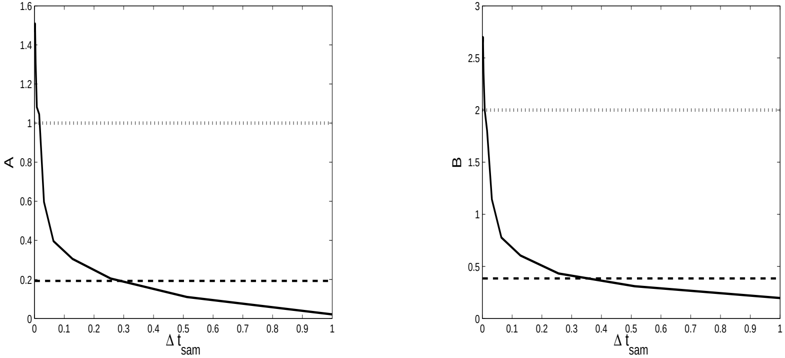


Figure : $\widehat{A}$, $\widehat{B}$ vs Δt_{sam} for bistable potential with $\sigma = 0.5$, $\varepsilon = 0.1$.

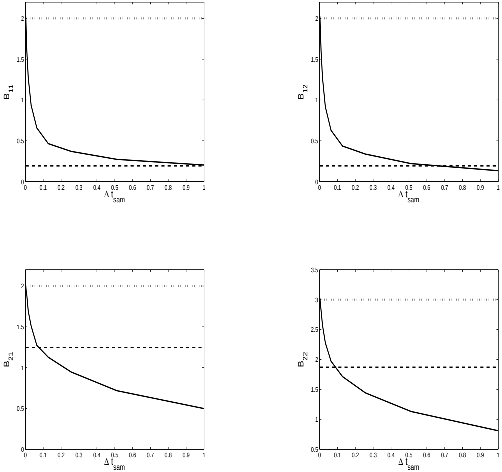


Figure : $\widehat{B}_{ij},\ i,j=1,2$ vs Δt_{sam} for 2d quadratic potential with $\sigma = 0.5,\ \varepsilon = 0.1$.

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:

$$
\begin{aligned}
\frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(x, y) + f_1(x, y) + \alpha_0(x, y) \frac{dU}{dt} \\
&\quad + \alpha_1(x, y) \frac{dV}{dt}, \qquad (15a) \\
\frac{dy}{dt} &= \frac{1}{\varepsilon^2} g_0(x, y) + \frac{1}{\varepsilon} g_1(x, y) + \frac{1}{\varepsilon} \beta(x, y) \frac{dV}{dt}. \qquad (15b)
\end{aligned}
$$

- 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 (15) as the **homogenization** problem.
- We assume that the coefficients of SDEs (15) 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

$$\frac{dX}{dt} = F(X) + K(X)\frac{dW}{dt}. \tag{16}$$

- 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 (16) 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:

$$\frac{dX}{dt} = F(X;\theta) + K(X)\frac{dW}{dt}. \tag{17}$$

- 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 (17), but using data from the slow-fast system.

- Given data $\{z(t)\}_{t\in[0,T]}$, the log likelihood for θ satisfying (17) is given by

$$\mathbb{L}(\theta;z) = \int_0^T \langle F(z;\theta), dz\rangle_{a(z)} - \frac{1}{2}\int_0^T |F(z;\theta)|^2_{a(z)}\,dt, \qquad (18)$$

- 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 (17) 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 (15) and we want to fit it to the equation (17). 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 (15) *as well as the averaged equation* (17) *are ergodic. Let* $\{x(t)\}_{t\in[0,T]}$ *be a sample path of* (15) *and* $X(t)$ *a sample path of* (17) *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* (15) *and* $X(t)$ *a sample path of* (17) *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:*

$$\lim_{\varepsilon\to 0} \frac{1}{N\delta}\mathbb{L}^{N,\delta}(\theta;x) = \lim_{T\to\infty} \frac{1}{T}\mathbb{L}(\theta;X). \tag{19}$$

Define

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

Then, under additional assumptions,

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

Thermal motion in a two-scale potential

$$\frac{dx}{dt} = -\nabla V^{\varepsilon}(x) + \sqrt{2\beta^{-1}}\frac{dW}{dt} \tag{20}$$

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

$$\frac{dX}{dt} = -K\nabla V(X) + \sqrt{2\beta^{-1}K}\frac{dW}{dt} \tag{21}$$

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 (20).
- The error in the asymptotic log Likelihood function is:

$$E_\infty(\theta) = \Big(-1 + \widehat{Z}_p^{-1} Z_p^{-1}\Big)\frac{\beta Z_V^{-1}}{2}\int_{\mathbb{R}} |\partial_x V|^2 e^{-\beta V(x;\theta)}\,dx. \tag{22}$$

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, $E_\infty < 0$.

Semiparametric Drift and Diffusion Estimation

S. Krumscheid, S. Kalliadasis, G.P., SIAM J. MMS, 11(2), 442-473 (2013).

- 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

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

- 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

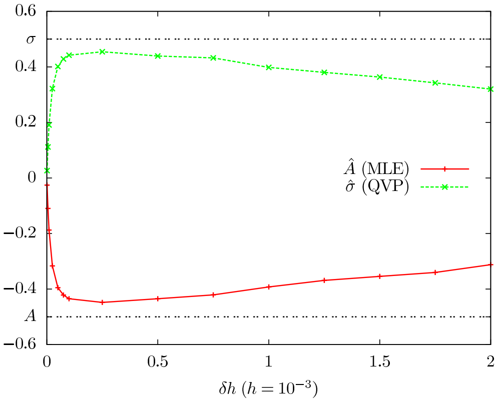
Fast/Slow System

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

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

Effective Dynamics

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



Fast OU Process Revisited

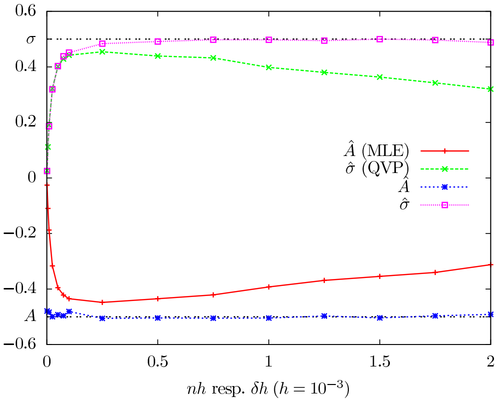
Fast/Slow System

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

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

Effective Dynamics

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



Fast OU Process II

- *Fast/slow system:*

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

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

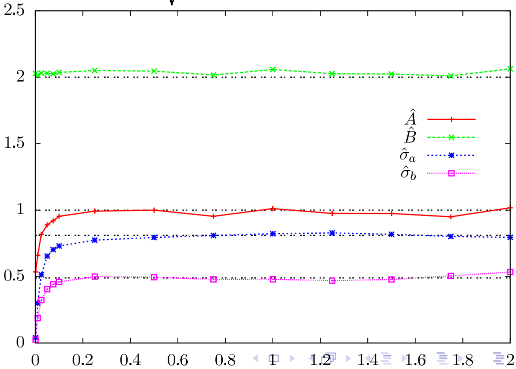
- Effective Dynamics:

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

- 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

- *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:

$$V(x) = x^2/2$$

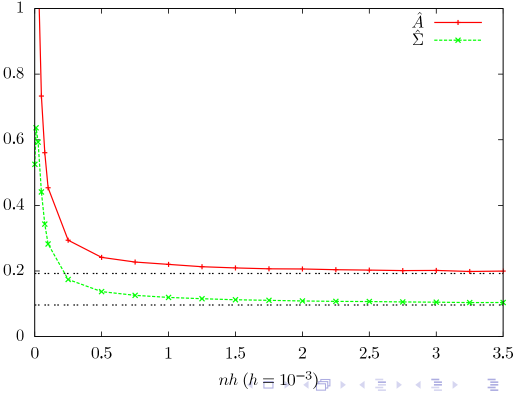
$$p(y) = \cos(y)$$

- 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]

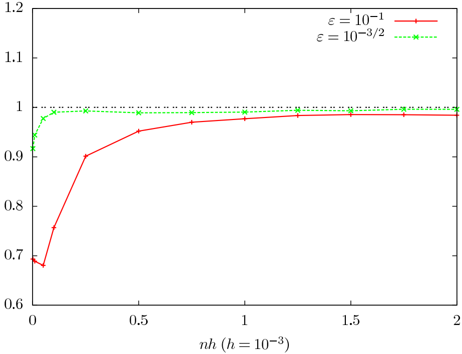
$$dX_t = A\left(X_t - {X_t}^3\right) dt + \sqrt{\sigma}\, dW_t$$

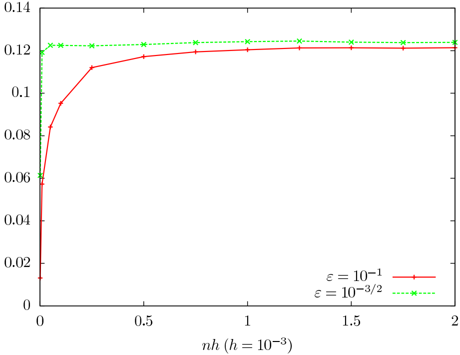
- true values:

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

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
- here: $\varepsilon = 10^{-1} \to \hat{\sigma} \approx 0.121$ and $\varepsilon = 10^{-3/2} \to \hat{\sigma} \approx 0.124$
- 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

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

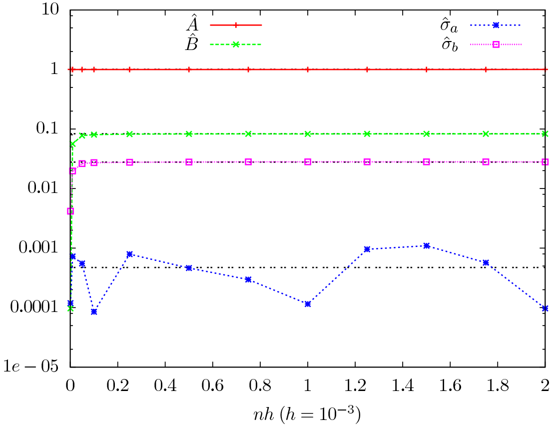
- For the three-term truncated representation the true values are:

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

Truncated Burgers Equation

Estimators

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



Fast Chaotic Noise II

- *Fast/slow system:*

$$\begin{aligned}
\frac{dx}{dt} &= x - x^3 + \frac{\lambda}{\varepsilon}(1 + x^2)y_2 \ , \\
\frac{dy_1}{dt} &= \frac{10}{\varepsilon^2}(y_2 - y_1) \ , \\
\frac{dy_2}{dt} &= \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1y_3) \ , \\
\frac{dy_3}{dt} &= \frac{1}{\varepsilon^2}(y_1y_2 - \frac{8}{3}y_3)
\end{aligned}$$

- Effective Dynamics:

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

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

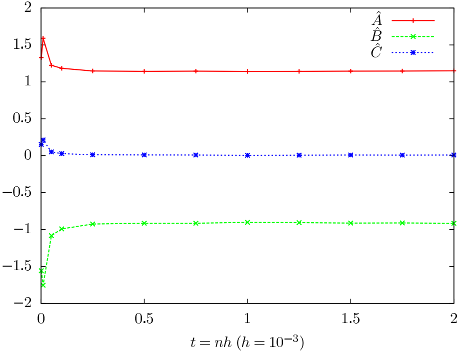
$$A = 1 + \sigma \ , \quad B = \sigma - 1 \ , \quad C = 0 \ , \quad \sigma_a = \sigma \ , \quad \sigma_b = 2\sigma \ , \quad \sigma_c = \sigma \ ,$$

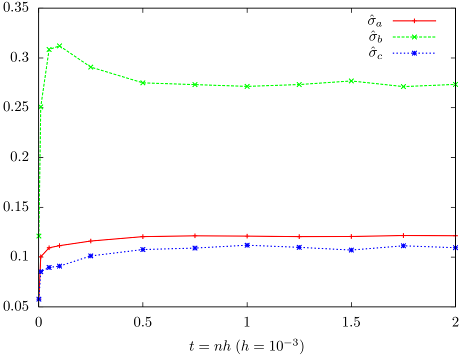
$$\sigma = 2\lambda^2 \int_0^\infty \lim_{T\to\infty} \frac{1}{T} \int_0^T \psi^s(y)\psi^{s+t}(y)\, ds\, dt$$

Fast Chaotic Noise

Estimators

- $\varepsilon = 0.1$





- It is possible to use a single long trajectory rather than many short ones (S. Kalliadasis, S. Krumscheid, G.P. preprint, 2014).
- Consistency, stability and convergence of the estimators can be studied (S. Krumscheid, preprint 2014).
- This methodology can be used to analyze data from measurements, observations (S. Kalliadasis, S. Krumscheid, G.P., M. Pradas, preprint, 2014).

Climate transitions during the last glacial period

- Climate transitions during the last glacial period.
- Records covering the last glacial period, approximately from 70 ky until 20 ky before present, are dominated by repeated rapid climate shifts, the so-called Dansgaard–Oeschger (DO) events.
- It is believed that DO events are transitions between two metastable climate states, a cold stadial and a warm interstadial state.
- We want to calculate how long it takes (on average) between DO events.

- We consider the δ^{18}O isotope record (as a proxy for Northern Hemisphere temperature) during the last glacial period which was obtained from the NGRIP see Fig. 8(a).
- We observe a noisy temporal signal where the temperature increases up to a warm state until it abruptly goes down to a colder state (corresponding to the DO events), giving rise to a bimodal histogram, see Fig. 8(b).
- We consider two different parametrizations in our SDE model (drift and diffusion coefficients):

  `M1:` $f(X;\theta) = \sum_{j=0}^{3} \theta_j X^j; \;\; g(X;\theta) = \theta_4$.

  `M2:` $f(X;\theta) = \sum_{j=0}^{3} \theta_j X^j; g(X;\theta) = \begin{cases} \theta_4 & \text{if } X < \theta_6 \\ \theta_5 & \text{if } X \geqslant \theta_6 \end{cases}.$

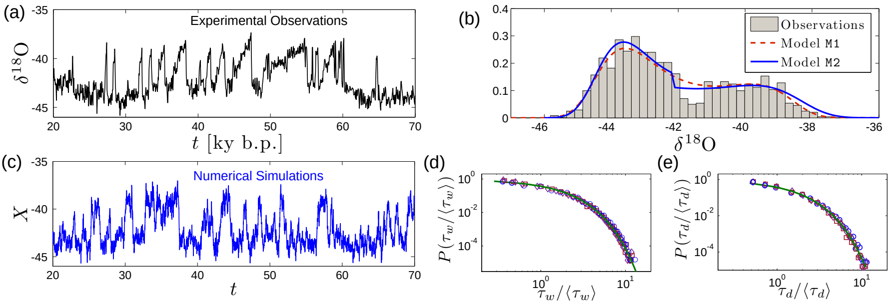


Figure : (a) Paleoclimatic record time series. (b) PDF of the experimental observations (histogram in gray) and the numerical ones obtained from model M1 and M2. (c) Time series of the fitted coarse-grained process X computed by using model M2. (d) and (e) PDF of the residence times τ_w at the cooler state and PDF of the durations τ_d of the DO events, normalized to their mean values and for different values of the threshold ($X_{th} \in [-42.5, -42]$). The solid lines correspond to $P(z) = \exp(-z)$.

Multiscale modeling and inverse problems

J. Nolen, A.M Stuart, G.P., in *Numerical Analysis of Multiscale Problems*, Lecture Notes in Computational Science and Engineering, Vol. 83, Springer, 2012

- In many applications we need to blend observational data and mathematical models.
- Parameters appearing in the model, such as constitutive tensors, initial conditions, boundary conditions, and forcing can be estimated on the basis of observed data.
- The resulting inverse problems are usually ill-posed and some form of regularization is required.
- We are interested in problems where the unknown parameters vary across scales.

- We study inverse problems for PDEs with rapidly oscillating coefficients for which a homogenized equation exists.
- We want to estimate unknown parameters $u \in X$ from noisy data $y \in Y$ (usually $Y = \mathbb{R}^N$).
- z is the solution of the PDE.
- The map $\mathcal{G} : X \to \mathbb{R}^N$ denotes the mapping from the unknown parameter to the data (*observation operator*)
- The map $\mathcal{F} : X \to Z$ denotes the mapping from the parameter to the prediction (*prediction operator*).
- The mapping $G : X \to P$ mapping $u \in X$ to the solution $G(u) \in P$ of a (PDE), is the *solution operator*.
- We assume that we are given noisy data:

$$y = \mathcal{G}(u) + \xi, \quad \xi \sim \mathcal{N}(0, \Gamma). \tag{23}$$

The main conclusions are:

(a) The choice of the space or set in which to seek the solution to the inverse problem is intimately related to whether a low-dimensional “homogenized" solution or a high-dimensional “multiscale" solution is required for predictive capability. This is a choice of regularization.
(b) The regularisation procedure is a part of the modelling strategy.
(c) If a homogenized solution to the inverse problem is desired, then this can be recovered from carefully designed observations of the full multiscale system.
(d) Homogenization theory can be used to improve the estimation of homogenized parameters from observations of multiscale data.

- Example: Dirichlet problem for the pressure (groundwater flow)

$$\begin{aligned} \nabla \cdot v = f, &\quad x \text{ in } D, \\ p = 0, &\quad x \text{ on } \partial D, \\ v = -k\nabla p & \end{aligned} \tag{24}$$

- where $D \subset \mathbb{R}^d$.
- The permeability tensor field $k(x) = \exp\big(u(x)\big)$, $u(x)$ positive definite is assumed to be unknown and must be determined from data.
- Equation for Lagrangian trajectories (ϕ is the porosity):

$$dx = \frac{v(x)}{\phi}\, dt + \sqrt{2\eta}\, dW, \quad x(0) = x_{\text{init}}, \tag{25}$$

- from PDE theory we know that we may define $G : X \to H_0^1(D)$ by $G(u) = p$.
- Consider a set of real-valued continuous linear functionals

$$\ell_j : H^1(D) \to \mathbb{R}$$

- and define

$$\mathcal{G} : X \to \mathbb{R}^N \quad \text{by} \quad \mathcal{G}(u)_j = \ell_j(G(u)).$$

- **Inverse problem:** determine $u \in X$ from the noisy observations $y \in \mathbb{R}^N$ (23).

- Assume that the permeability tensor has two characteristic length scales $k = K^{\varepsilon}(x) = K(x, x/\varepsilon)$, periodic in the second argument, and $\varepsilon > 0$ a small parameter.
- Family of problems

$$
\begin{aligned}
\nabla \cdot v^{\varepsilon} &= f, \quad x \text{ in } D, && \text{(26a)}\\
p^{\varepsilon} &= 0, \quad x \text{ on } \partial D, && \text{(26b)}\\
v^{\varepsilon} &= -K^{\varepsilon} \nabla p^{\varepsilon}. && \text{(26c)}
\end{aligned}
$$

- Family of SDEs (we set $\eta = \varepsilon \eta_0$)

$$
dx^{\varepsilon} = \frac{v^{\varepsilon}(x^{\varepsilon})}{\phi} dt + \sqrt{2\eta_0 \varepsilon}\, dW, \quad x^{\varepsilon}(0) = x_{\text{init}}. \tag{27}
$$

- The pressure admits the two–scale expansion

$$p^{\varepsilon}(x) \approx p_{\mathrm{a}}^{\varepsilon}(x) := p_0(x) + \varepsilon p_1(x, \frac{x}{\varepsilon}) \tag{28}$$

- The *cell problem* for $\chi(x,y)$ is:

$$-\nabla_y \cdot \left(\nabla_y \chi K^T\right) = \nabla_y \cdot K^T, \quad y \in \mathbb{T}^d. \tag{29}$$

- We can now define for each $x \in D$ the effective (homogenized) permeability tensor K_0

$$K_0(x) = \int_{\mathbb{T}^d} Q(x,y)dy, \tag{30}$$

$$Q(x,y) = K(x,y) + K(x,y)\nabla_y \chi(x,y)^T. \tag{31}$$

- We write $K_0 = \exp(u_0)$.

- p_0 is the solution of the homogenized PDE

$$\nabla \cdot v_0 = f, \quad x \in D, \tag{32a}$$

$$p_0 = g, \quad x \in \partial D, \tag{32b}$$

$$v_0 = -K_0 \nabla p_0. \tag{32c}$$

- and the corrector p_1 is defined by

$$p_1(x, y) = \chi(x, y) \cdot \nabla p_0(x). \tag{33}$$

Large Data Limits

- We study inverse problems where a single scalar parameter is sought and we study whether or not this parameter is correctly identified when a large amount of noisy data is available.
- We consider the problem of estimating a single scalar parameter $u \in \mathbb{R}$ in the elliptic PDE

$$\begin{aligned} \nabla \cdot v = f, &\quad x \in D, \\ p = 0, &\quad x \in \partial D, \\ v = -\exp(u) A \nabla p & \end{aligned} \tag{34}$$

- where $D \subset \mathbb{R}^d$ is bounded and open, and $f \in H^{-1}$ as well as the constant symmetric matrix A are assumed to be known.
- We let $G : \mathbb{R} \to H_0^1(D)$ be defined by $G(u) = p$.
- The observation operator $\mathcal{G} : \mathbb{R} \to \mathbb{R}^N$ is defined by

$$\mathcal{G}(u)_j = \ell_j(G(u)).$$

- Our aim is to solve the inverse problem of determining u given y satisfying (23).
- We assume that $\xi \sim N(0, \gamma^2 I)$ i.e. that the observational noise on each linear functional is i.i.d. $N(0, \gamma^2)$.
- u is finite dimensional, so we can minimize the least squares functional and no regularization is needed.
- Since the solution p of (34) is linear in $\exp(-u)$, we can write $G(u) = \exp(-u)p^\star$ where

$$\begin{aligned} \nabla \cdot v = f, \quad x \in D, \\ p^\star = 0, \quad x \in \partial D. \\ v = -A \nabla p^\star \end{aligned} \tag{35}$$

- Note that $\mathcal{G}(u)_j = \exp(-u)\ell_j(p^\star)$ so that the least squares functional has the form

$$\Phi(u) = \frac{1}{2\gamma^2}\sum_{j=1}^{N}|y_j - \mathcal{G}_j(u)|^2 = \frac{1}{2\gamma^2}\sum_{j=1}^{N}|y_j - \exp(-u)\ell_j(p^\star)|^2.$$

- We can prove that Φ has a unique minimizer $\overline{u}$ satisfying

$$\exp(-\overline{u}) = \frac{\sum_{j=1}^{N} y_j \ell_j(p^\star)}{\sum_{j=1}^{N} \ell_j(p^\star)^2}. \tag{36}$$

- We ask whether, for large N, the estimate $\overline{u}$ is close to the desired value of the parameter. We study two situations:
  - The data is generated by the model which is used to fit the data.
  - The data is generated by a multiscale model whose homogenized limit gives the model which is used to fit the data.
- We define $p_0 = \exp(-u_0)p^\star$ so that p_0 solves (34) with $u = u_0$.

Assumption

We assume that the data y is given by noisy observations generated by the statistical model:

$$y_j = \ell_j(p_0) + \xi_j$$

where $\{\xi_j\}$ form an i.i.d. sequence of random variables distributed as $N(0, \gamma^2)$.

Theorem

Let the above assumption hold and assume that $\liminf_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \ell_j(p^\star)^2 \geqslant L > 0$ as $N \to \infty$. Then ξ-almost surely

$$\lim_{N\to\infty} \left| \exp(-\overline{u}) - \exp(-u_0) \right| = 0.$$

Data from the multiscale problem

- We consider the situation where the data is taken from a multiscale model whose homogenized limit falls within the class used in the statistical model to estimate parameters.
- We define $p_0 = \exp(-u_0)p^\star$ and we let p^ε be the solution of (26) with K^ε chosen so that the homogenized coefficient associated with this family is $K_0 = \exp(u_0)A$.

Assumption

We assume that the data y is generated from noisy observations of a multiscale model:

$$y_j = \ell_j(p^\varepsilon) + \xi_j$$

with p^ε as above and the $\{\xi_j\}$ an i.i.d. sequence of random variables distributed as $N(0, \gamma^2)$.

Theorem

Let Assumptions 7 hold and assume that that the linear functionals ℓ_j *are chosen so that*

$$\lim_{\varepsilon\to 0}\limsup_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}|\ell_j(p^\varepsilon - p_0)|^2 = 0 \tag{37}$$

and $\liminf_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\ell_j(p^\star)^2 \geqslant L > 0$ *as* $N\to\infty$. *Then* $\xi-$ *almost surely*

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}|\exp(-\overline{u}) - \exp(-u_0)| = 0.$$

Remarks

1. *Assumption (37) encodes the idea that, for small ε, the linear functionals used in the observation process return nearby values when applied to the solution p^ε of the multiscale model or to the solution p_0 of the homogenized equation.*
2. *If $\{\ell_j(p)\}_{j=1}^{\infty}$ is a family of bounded linear functionals on $L^2(D)$, uniformly bounded in j, then (37) will hold.*
3. *On the other hand, we may choose linear functionals that are bounded as functionals on $H^1(D)$ yet unbounded on $L^2(D)$. In this case (37) may not hold and the correct homogenized coefficient may not be recovered, even in the large data limit.*
4. *This is analogous to the situation in the problem of parameter estimation for multiscale diffusions.*