Parameter Estimation for Multiscale Diffusions

G.A. Pavliotis
INRIA-MICMAC Project Team
and
Department of Mathematics Imperial College London

LECTURE 1: 14/02/2012
LECTURE 2: 21/02/2012

CONTENTS

1. Motivation/Examples.
2. Maximum Likelihood Estimation.
3. Thermal Motion in a two-scale potential.
4. Maximum likelihood estimation for general fast/slow systems of SDEs.
5. Semiparametric drift and diffusion estimation for multiscale diffusions.
6. Inverse problems for multiscale PDEs.

- 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.
  - Oceanic Transport.

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 lecture: 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.

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 singularly perturbed SPDE

$$\partial_t u = \frac{1}{\varepsilon^2}\mathcal{L}u + \frac{1}{\varepsilon}B[u,u] + \mathcal{J}u + \frac{1}{\varepsilon}\xi, \tag{1}$$

- where $\mathcal{N} := \mathcal{N}(\mathcal{L})$ with $\dim(\mathcal{N}) = 1$, $H = \mathcal{N} \oplus \mathcal{N}^{\perp}$ and $\mathbb{P}_{\mathcal{N}}\xi = 0$.
- Then for $\varepsilon \ll 1$, $\mathbb{P}_{\mathcal{N}}u \approx X(t) \cdot e(x)$ where $X(t)$ is the solution of the amplitude (homogenized) equation

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

- There exist formulas for the constants A, B, σ_a^2, σ_b^2 but they involve knowledge of the spectrum of $\mathcal{L}$, $\mathcal{J}$, the covariance operator of the noise and the nonlinearity $B[\cdot,\cdot]$.
- The form of the amplitude equation (2) 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

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

- Consider the SDE

$$dx^{\varepsilon}(t) = -V'\left(x^{\varepsilon}(t), \frac{x^{\varepsilon}(t)}{\varepsilon}; \alpha\right) dt + \sqrt{2\sigma}\, dW(t), \tag{3}$$

- 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:

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

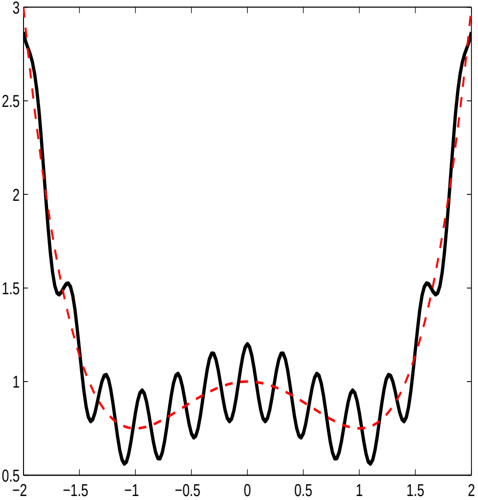


Figure: Bistable potential with periodic fluctuations

- The coefficients A, Σ are given by the standard homogenization formulas.
- Goal: fit a time series of $x^{\varepsilon}(t)$, the solution of (3), to the homogenized SDE.
- Problem: the data is not compatible with the homogenized equation at small scales.
- Model misspecification.

Deriving dynamical models from paleoclimatic records

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

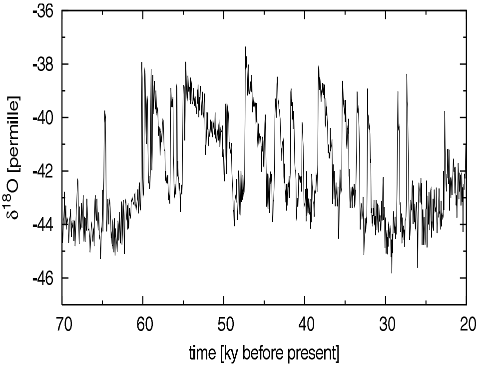


FIG. 1. δ^{18}O record from the NGRIP ice core during the last glacial period.

- Fit this data to a bistable SDE

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

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

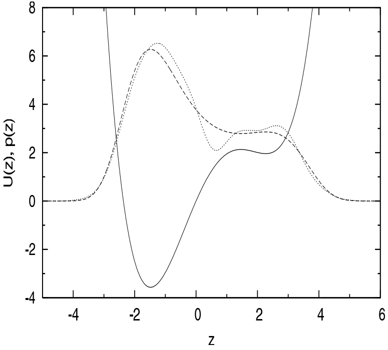


FIG. 8. Potential derived by least-squares fit from the probability density of the ice-core data (solid) together with probability densities of the model (dashed) and the data (dotted).

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

$$\frac{dx}{dt} = v(x,t), \tag{5}$$

- 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:

$$\frac{dX}{dt} = \sqrt{2\mathcal{K}}\frac{dW}{dt} \tag{6}$$

- We are given a time series of noise observations:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{7}$$

- Goal: estimate the **Eddy Diffusivity** $\mathcal{K}$ from the noisy Lagrangian data (7).

Econometrics: Market Microstructure Noise

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

- Observed process Y_t:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i = i\Delta t, \quad i = 0, \dots N-1. \tag{8}$$

- Where X_t is the solution of

$$dX_t = (\mu - \nu_t/2)\, dt + \sigma_t dB_t, \quad d\nu_t = \kappa\, (\alpha - \nu_t)\, dt + \gamma \nu_t^{1/2} dW_t, \tag{9}$$

- 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.

The Maximum Likelihood Estimator

- We are given a one dimensional SDE model

$$dX_t = b(X_t;\theta)\,dt + \sqrt{2\sigma}\,dW_t, \quad X_0 = x. \tag{10}$$

- where the constant diffusion coefficient σ and the parameters $\theta \in \Theta \in \mathbb{R}^N$ are unknown.
- We are given a set of discrete equidistant observations $\{X_{t_j}\}_{j=0}^{J-1}$ with $\delta = t_{j+1} - t_j$.
- Our goal is to estimate the diffusion coefficient and the parameters θ.

- Consider first the estimation of the diffusion coefficient. We will use the *quadratic variation* of the process:

$$\widehat{\sigma}_J = \frac{1}{2J\delta} \sum_{j=0}^{J-1} \left(X_{j+1} - X_j\right)^2, \tag{11}$$

- where $X_j := X_{(j-1)\delta}$.
- The diffusion coefficient is a local property of the path. The quadratic variation estimator (11) converges to the diffusion coefficient in the limit as the distance between subsequent observations goes to 0 while the number of observations becomes infinite, while keeping the window of observation $[0, (J-1)\delta = T]$ fixed.
- This is called the *high frequency limit*. We have (Prakasa Rao 1999)

$$\lim_{N \to +\infty} \sum_{j=1}^{2^N} \left(X_{jT2^{-N}} - X_{(j-1)T2^{-N}}\right)^2 = 2\sigma T \quad \text{a.s.} \tag{12}$$

- More generally, when the diffusion coefficient is not constant, we have (Karatzas and Sreeve)

$$\lim_{\Delta t \to 0} \sum_{t_k \leqslant t} \left(X_{t_{k+1}} - X_{t_k}\right)^2 = 2 \int_0^t \sigma(X_s)\, ds,$$

- in probability.
- We will prove a (much) weaker version of (12).

Proposition

Let $\{X_j\}_{j=0}^{J-1}$ *be a sequence of equidistant observations of* (10) *with* $\Delta t = \delta$ *and* $(J-1)\delta = T$ *fixed. Assume that the drift* $b(x;\theta)$ *is bounded. Then*

$$|\mathbb{E}\widehat{\sigma}_J - \sigma| \leqslant C\left(\delta + \delta^{1/2}\right). \tag{13}$$

In particular,

$$\lim_{J \to +\infty} |\mathbb{E}\widehat{\sigma}_J - \sigma| = 0. \tag{14}$$

Proof. We have

$$X_{j+1} - X_j = \int_{j\delta}^{(j+1)\delta} b(X_s;\theta)\,ds + \sqrt{2\sigma}\,\Delta W_j,$$

where $\Delta W_j = W_{(j+1)\delta} - W_{j\delta}$. We substitute this into (11) to obtain

$$\widehat{\sigma}_J \;=\; \sigma \frac{1}{\delta J}\sum_{j=0}^{J-1}(\Delta W_j)^2 + \frac{1}{\delta J}\sum_{j=0}^{J-1} I_j M_j + \frac{1}{\delta J}\sum_{j=0}^{J-1} I_j^2,$$

where

$$I_j := \int_{j\delta}^{(j+1)\delta} b(X_s;\theta)\,ds$$

and $M_j := \sqrt{2\sigma}\,\Delta W_j$. We note that $\mathbb{E}(\Delta W_n)^2 = \delta$. Furthermore, from the boundedness of $b(x;\theta)$ and using the Cauchy-Schwarz inequality we get

$$\mathbb{E} I_j^2 \leqslant \delta \int_{j\delta}^{(j+1)\delta} \mathbb{E}\big(b(X_s;\theta)\big)^2\,ds \leqslant C\delta^2.$$

Consequently:

$$
\begin{aligned}
\left|\mathbb{E}\widehat{\sigma}_J - \sigma\right| &\leqslant \frac{1}{2\delta}\mathbb{E}I_j^2 + \frac{1}{\delta}\mathbb{E}\left|I_j M_j\right| \\
&\leqslant C\delta + \frac{C}{\delta}\left(\frac{1}{\alpha}\mathbb{E}I_j^2 + \alpha\mathbb{E}M_j^2\right) \\
&\leqslant C\left(\delta + \delta^{1/2}\right).
\end{aligned}
$$

In the above we used Cauchy's inequality with $\alpha = \delta^{1/2}$. □

- From now on we will assume that we have already estimated the diffusion coefficient. To simplify the notation, we will set $\sqrt{2\sigma} = 1$:

$$dX_t = b(X_t; \theta)\, dt + dW_t. \tag{15}$$

- Now we estimate the unknown parameters in the drift $\theta \in \Theta$ from discrete observations. We denote the true value by θ_0.
- We will use the **maximum likelihood estimator**(MLE).
- The Likelihood function is defined as the Radon-Nikodym derivative of the law of the process X_t with respect to the Wiener measure, i.e. the law of Brownian motion:

$$\begin{aligned} \frac{d\mathbb{P}_X}{d\mathbb{P}_W} &= \exp\left(\frac{1}{2}\int_0^T (b(X_s;\theta))^2\, ds - \int_0^T b(X_s;\theta)\, dX_s\right) \\ &=: L\big(\{X_t\}_{t\in[0,T]}; \theta, T\big) \end{aligned} \tag{16}$$

- The maximum likelihood estimator MLE is defined as

$$\widehat{\theta} = \text{argmax}_{\theta\in\Theta} L\big(\{X_t\}_{t\in[0,T]}; \theta, T\big). \tag{17}$$

- We assume that the diffusion process (15) is stationary.
- The MLE (17) is *asympotically unbiased*: in the limit as the window of observation becomes infinite, $T \to +\infty$, the MLE $\widehat{\theta}$ converges to the true value θ_0.
- Assume that there are N parameters to be estimated, $\boldsymbol{\theta} = (\theta_1, \dots \theta_N)$. The MLE is obtained by solving the (generally nonlinear) system of equations

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, \dots, N. \tag{18}$$

- The solution of this system of equations can be expressed in terms of functionals (e.g. moments) of the observed path $\{X_t\}_{t \in [0,T]}$

$$\widehat{\boldsymbol{\theta}} = \mathcal{F}(\{X_t\}_{t \in [0,T]}).$$

Example

(MLE for the stationary OU process). Consider the stationary OU process

$$dX_t = -\alpha X_t \, dt + dW_t \tag{19}$$

with $X_0 \sim \mathcal{N}\left(0, \frac{1}{2\alpha}\right)$. The *log Likelihood function* is

$$\log L = \frac{\alpha^2}{2} \int_0^T X_t^2 \, dt + \alpha \int_0^T X_t \, dX_t.$$

Equation (18) becomes $\frac{\partial \log L}{\partial \alpha} = 0$ from which we obtain

$$\widehat{\alpha} = -\frac{\int_0^T X_t \, dX_t}{\int_0^T (X_t)^2 \, dt} =: -\frac{B_1(\{X_t\}_{t \in [0,T]})}{M_2(\{X_t\}_{t \in [0,T]})} \tag{20}$$

where we have used the notation

$$B_n(\{X_t\}_{t \in [0,T]}) = \int_0^T (X_t)^n \, dX_t, \quad M_2(\{X_t\}_{t \in [0,T]}) := \int_0^T (X_t)^n \, dt, \quad n = 1, 2, \dots \tag{21}$$

- Given a set of discrete equidistant observations $\{X_j\}_{j=0}^{J-1}$, $X_j = X_{(j-1)\Delta t}$, $\Delta X_j = X_{j+1} - X_j$, formula (20) can be approximated by

$$\widehat{\alpha} = -\frac{\sum_{j=0}^{J-1} X_j \, \Delta X_j}{\sum_{j=0}^{J-1} |X_j|^2 \, \Delta t}. \tag{22}$$

- The MLE (20) becomes asymptotically unbiased in the *large sample limit* $J \to +\infty$, Δt fixed.

Example

Consider the following generalization of the previous example:

$$dX_t = \alpha b(X_t)\, dt + dW_t, \tag{23}$$

where $b(x)$ is such that the equation has a unique ergodic solution. The log Likelihood function is

$$\log L = \frac{\alpha^2}{2} \int_0^T b(X_t)^2 \, dt - \alpha \int_0^T b(X_t)\, dX_t.$$

The MLE is

$$\widehat{\alpha} = \frac{\int_0^T b(X_t)\, dX_t}{\int_0^T (b(X_t))^2 \, dt}.$$

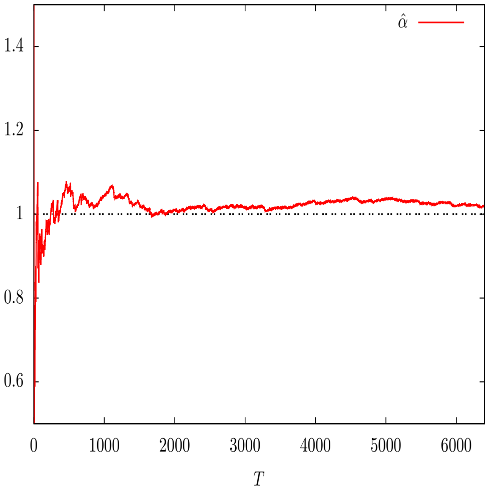


Figure: MLE for the OU processes.

Example

(MLE for a stationary bistable SDE) Consider the SDE

$$dX_t = \left(\alpha X_t - \beta X_t^3\right) dt + dW_t \tag{24}$$

This SDE is of the form $dX_t - -V'(X_t)\, dt + dW_t$ with $V(x) = \frac{\alpha}{2}x^2 - \frac{\beta}{4}x^4$ and is ergodic with invariant distribution $\rho(x) = Z^{-1}e^{-\frac{1}{2}V(x)}$. Our goal is to estimate the coefficients α and β from observations using the maximum likelihood approach. The log likelihood function reads

$$\begin{aligned} \log L &= \frac{1}{2}\int_0^T \left(\alpha X_t - \beta X_t^3\right)^2 dt - \int_0^T \left(\alpha X_t - \beta X_t^3\right) dX_t \\ &=: \frac{1}{2}\alpha^2 M_2 + \frac{1}{2}\beta^2 M_6 - \alpha\beta M_4 - \alpha B_1 + \beta B_3, \end{aligned}$$

using notation (21). Equations (18) become

$$\frac{\partial \log L}{\partial \alpha}(\widehat{\alpha}, \widehat{\beta}) = 0, \quad \frac{\partial \log L}{\partial \beta}(\widehat{\alpha}, \widehat{\beta}) = 0.$$

Example

(MLE for a stationary bistable SDE contd.) This leads to a linear system of equations

$$\begin{pmatrix} M_2 & -M_4 \\ M_4 & -M_6 \end{pmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} B_1 \\ B_3 \end{pmatrix},$$

The solution of which is

$$\widehat{\alpha} = \frac{B_1 M_6 - B_3 M_4}{M_2 M_6 - M_4^2}, \quad \widehat{\beta} = \frac{B_1 M_4 - B_3 M_2}{M_2 M_6 - M_4^2}. \tag{25}$$

When we have a discrete set of observations $\{X_j\}_{j=0}^{J-1}$ the integrals in (25) have to be approximated by sums.

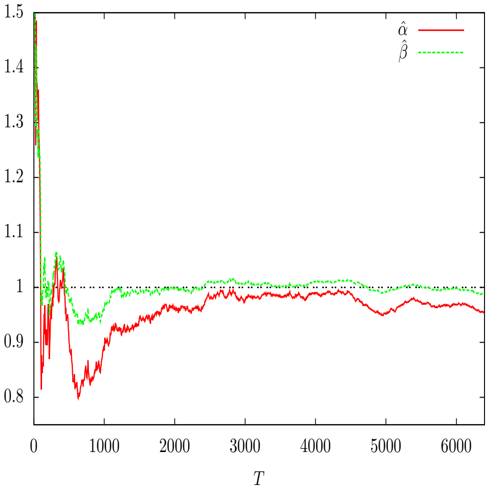


Figure: MLE estimators for a bistable potential.

- The rigorous justification of the MLE is based on Girsanov's theorem
- The proof of asymptotic consistency and normality use the law of large numbers and central limit theorems for square integrable martingales.
- Here is present a heuristic derivation of (16) which is based on the Euler-Marayama discretization of (15):

$$X_{n+1} - X_n = b(X_n; \theta)\, \Delta t + \Delta W_n, \tag{26}$$

- where $X_n = X(n\Delta t)$ and $\Delta W_n = W_{n+1} - W_n =: \xi_n$. We have that $\xi_n \sim \mathcal{N}(0, \sqrt{\Delta t})$, independent.
- Our goal is to calculate the Radon-Nikodym derivative of the law of the discrete-time process $\{X_n\}_{n=0}^{N-1}$ and the discretized Brownian motion.

- We rewrite (26) in the form

$$\Delta X_n = b_n \, \Delta + \xi_n, \tag{27}$$

- where $b_n := b(X_n; \theta)$, $\Delta := \Delta t$. The distribution function of the discretized Brownian motion is

$$p_W^N = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta}} \exp\left(-\frac{1}{2\Delta}(\Delta W_i)^2\right) \tag{28}$$

$$= \frac{1}{(\sqrt{2\pi\Delta})^{N/2}} \exp\left(-\frac{1}{2\Delta}\sum_{i=0}^{N-1}(\Delta W_i)^2\right). \tag{29}$$

- Similarly, for the law of the discretized process $\{X_n\}_{n=0}^{N-1}$ we can write

$$p_X^N = \frac{1}{(\sqrt{2\pi\Delta})^{N/2}} \exp\left(-\frac{1}{2\Delta}\sum_{i=0}^{N-1}(\Delta X_i)^2\right). \tag{30}$$

- The ratio of the laws of the two possesses is

$$\frac{d\mathbb{P}_X^N}{d\mathbb{P}_W^N} = \frac{p_X^N}{p_W^N} = \exp\left(-\frac{1}{2\Delta}\sum_{i=0}^{N-1}\left((\Delta X_i)^2 - (\Delta W_i)^2\right)\right)$$

$$= \exp\left(-\frac{1}{2\Delta}\sum_{i=0}^{N-1}\left((\Delta X_i)^2 - (\Delta X_i - b_i\Delta)^2\right)\right)$$

$$= \exp\left(\frac{1}{2}\sum_{i=0}^{N-1}(b_i)^2\Delta - \sum_{i=0}^{N-1} b_i \Delta X_i\right).$$

- Passing now (formally) to the limit as $N \to +\infty$, ($T \to +\infty$) while keeping Δ fixed we obtain (16).

- The MLE (17) depends on the path $\{X_t\}_{t \in [0,T]}$ (or, rather, on the discrete observations $\{X_j\}_{j=0}^{J-1}$).
- We want to prove that, in the large sample limit $J \to +\infty$, Δt fixed, and for appropriate assumptions on the diffusion process X_t, the MLE converges to the true value θ_0.
- We also want to obtain information about the fluctuations around the limiting value θ_0.
- Assuming that X_t is stationary we can prove that the MLE $\widehat{\theta}$ converges in the limit as $T \to +\infty$ (assuming that the entire path $\{X_t\}_{t \in [0,T]}$ is available to us) to θ_0.
- Furthermore, we can prove *asympotic normality* of the maximum likelihood estimator,

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \to \mathcal{N}(0, \sigma^2), \tag{31}$$

- for a variance σ^2 that can be calculated.

Theorem

Let X_t be the stationary OU process

$$dX_t = -\alpha X_t \, dt + dW_t, \quad X_0 \sim \mathcal{N}\left(0, \frac{1}{2\alpha}\right)$$

and let $\widehat{\alpha}$ denote the MLE (20)*. Then*

$$\lim_{T \to +\infty} \sqrt{T} |\widehat{\alpha} - \alpha| = \mathcal{N}(0, 2\alpha) \tag{32}$$

in distribution.

We will use the following result from probability theory.

Theorem

(Slutsky) *Let* $\{X_n\}_{n=1}^{+\infty}$, $\{Y_n\}_{n=1}^{+\infty}$ *be sequences of random variables such that* X_n *converges in distribution to a random variable* X *and* Y_n *converges in probability to a constant* $c \neq 0$*. Then*

$$\lim_{n \to +\infty} Y_n^{-1} X_n = c^{-1} X,$$

in distribution.

Proof of Theorem 6. First we observe that

$$\widehat{\alpha} = -\frac{\int_0^T X_t \, dX_t}{\int_0^T X_t^2 \, dt} = \alpha - \frac{\int_0^T X_t \, dW_t}{\int_0^T X_t^2 \, dt}.$$

Consequently:

$$\begin{aligned} \widehat{\alpha} - \alpha &= -\frac{\int_0^T X_t \, dW_t}{\int_0^T X_t^2 \, dt} = \frac{1}{\sqrt{T}} \frac{\frac{1}{\sqrt{T}} \int_0^T X_t \, dW_t}{\frac{1}{T} \int_0^T X_t^2 \, dt} \\ &\overset{\text{Law}}{=} -\frac{1}{\sqrt{T}} \frac{W\left(\frac{1}{T}\int_0^T X_t^2 \, dt\right)}{\frac{1}{T}\int_0^T X_t^2 \, dt}, \end{aligned}$$

where the scaling property of Brownian motion was used: we have that, in law

$$\int_0^T f(s) \, dW(s) = W\left(\int_0^T f^2(s) \, ds\right). \tag{33}$$

- The process X_t is stationary. We use the ergodic theorem for stationary Markov processes to obtain

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T X_t^2 \, dt = \mathbb{E} X_t^2 = \frac{1}{2\alpha} \quad \text{a.s.} \tag{34}$$

- Let $X = \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{1}{2\alpha}$. We can write $X = W(\sigma^2)$. We use now the Hölder continuity of Brownian motion to conclude that, almost surely,

$$\begin{aligned} \left| W\left(\frac{1}{T} \int_0^T X_t^2 \, dt \right) - X \right| &= \left| W\left(\frac{1}{T} \int_0^T X_t^2 \, dt \right) - W\left(\frac{1}{2\alpha} \right) \right| \\ &\leqslant \text{Hö l}(W) \left| \frac{1}{T} \int_0^T X_t^2 \, dt - \frac{1}{2\alpha} \right|^{\alpha}, \end{aligned}$$

- with $\alpha < 1/2$.

- We use (34) to obtain

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T X_t \, dW_t = \mathcal{N}\left(0, \frac{1}{2\alpha}\right), \tag{35}$$

in distribution. We combine (34) with (35) and use Slutsky's theorem to conclude that

$$\lim_{T \to +\infty} \sqrt{T} |\widehat{\alpha} - \alpha| = \mathcal{N}\left(0, 2\alpha\right) \tag{36}$$

- in distribution. □

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:

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

- The matrix K is given by

$$K = \int_{\mathbb{T}^d} \Big(I + \nabla_y \phi(y) \Big) \Big(I + \nabla_y \phi(y) \Big)^T \mu(dy).$$

- Here $\phi : \mathbb{T}^d \to \mathbb{R}^d$ and $\rho : \mathbb{T}^d \to \mathbb{R}^+$ solve

$$\begin{aligned} -\mathcal{L}_0 \phi &= -\nabla_y p(y), \\ -\mathcal{L}_0^* \rho &= 0. \end{aligned}$$

- The operator $\mathcal{L}_0$ is the generator of the $\mathbb{T}^d-$valued process

$$dy = -\nabla_y p(y)\, dt + \sqrt{2\sigma}\, dW.$$

- Diffusion is depleted:

$$\frac{1}{Z\widehat{Z}} |\xi|^2 \leqslant \langle \xi, K\xi \rangle \leqslant |\xi|^2 \quad \forall \xi \in \mathbb{R}^d.$$

- 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

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

- The depletion of (A, Σ) over (α, σ) is apparent from the formulae (L is the period of $p(y)$):

$$A = \frac{\alpha L^2}{Z\widehat{Z}}, \quad \Sigma = \frac{\sigma L^2}{Z\widehat{Z}}$$

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

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

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

- $\Rightarrow$ The invariant measure is the Gibbs measure

$$\mu(dx) = \rho(x)dx = \frac{1}{Z_V} e^{-\frac{\alpha}{\sigma}V(x)}\, dx, \quad Z_V = \int_{\mathbb{R}} e^{-\frac{\alpha}{\sigma}V(x)}\, dx.$$

- $\rho(x)$ is the weak–L^1 limit of the invariant distribution of the unhomogenized equation:

$$\rho^{\varepsilon}(x) = \frac{1}{Z_{\varepsilon}} e^{-\frac{\alpha}{\sigma}V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}, \quad Z^{\varepsilon} = \int_{\mathbb{R}} e^{-\frac{\alpha}{\sigma}V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}\, dx.$$

- This follows from properties of periodic functions.

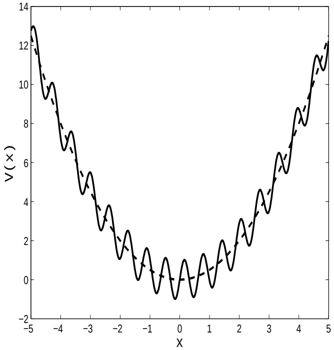


Figure: $V^{\varepsilon}(x)$ and $V(x)$.

- 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.

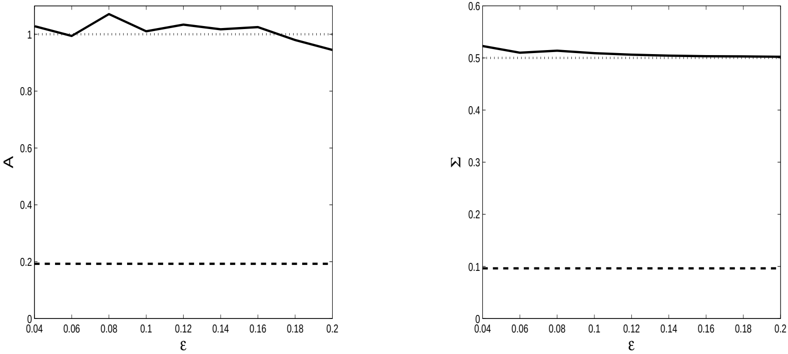


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs ε for quadratic potential.

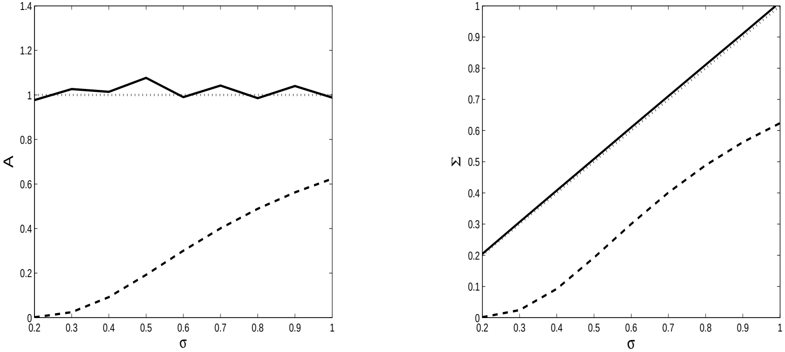


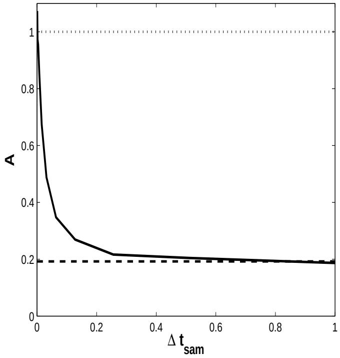
Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs σ for quadratic potential with $\varepsilon = 0.1$.

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.



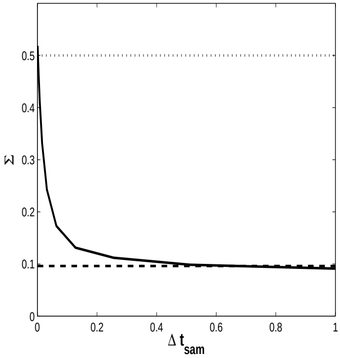


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs Δt_{sam} for quadratic potential with $\varepsilon = 0.1$.

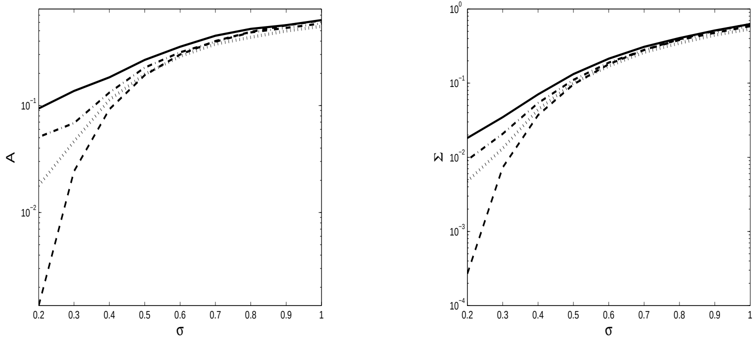


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs σ for quadratic potential with $\varepsilon = 0.1$, $\alpha = 1.0$. Dash–dotted line: $\Delta t_{sam} = 0.256$. Dotted line: $\Delta t_{sam} = 0.512$. Dashed line: homogenized coefficient.

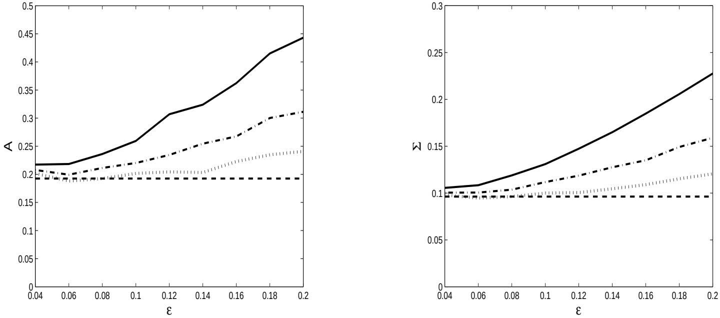


Figure: $\widehat{A}$, $\widehat{\Sigma}$ vs ε for quadratic potential with $\alpha = 1.0$, $\sigma = 0.5$. Dash–dotted line: $\Delta t_{sam} = 0.256$. Dotted line: $\Delta t_{sam} = 0.512$. Dashed line: homogenized coefficient.

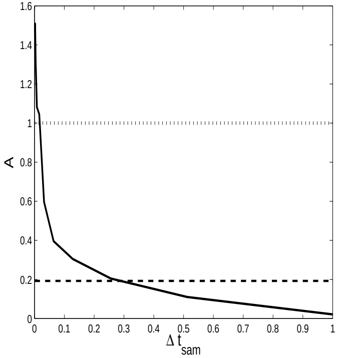


Figure: $\widehat{A}$, $\widehat{B}$ vs Δt_{sam} for bistable potential with $\sigma = 0.5$, $\varepsilon = 0.1$.

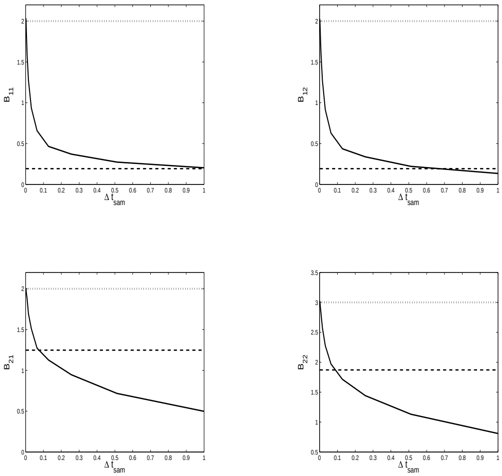


Figure: $\widehat{B}_{ij},\ i,j = 1,2$ vs Δt_{sam} for 2d quadratic potential with $\sigma = 0.5,\ \varepsilon = 0.1$.

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.

- Let $x_n^\varepsilon = x^\varepsilon(n\delta)$. We need to understand the small $\varepsilon,\ \delta$ asymptotics of x_n^ε. In particular, we can show that

$$x_{n+1} - x_n = -AV'(x_n)\delta + \sqrt{2\Sigma\delta}\xi_n + R(\varepsilon,\delta),$$

with ξ_n i.i.d. $\mathcal{N}(0,1)$ and, for $\varepsilon,\ \delta \ll 1$,

$$\|R(\varepsilon,\delta)\| \leqslant C(\varepsilon^{1/2} + \varepsilon\delta^{1/2-\iota} + \delta^{3/2}),$$

- for $\iota > 0$, arbitrarily small.
- To prove this we apply Itô's formula to the solutions of appropriate Poisson equations and estimate the various error terms.
- We also use the Dambis-Dubins-Schwarz theorem (martingales as time changed Brownian motions).

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:

$$\begin{aligned}
\frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y) + \alpha_0(x,y)\frac{dU}{dt} \\
&\quad + \alpha_1(x,y)\frac{dV}{dt}, && \text{(37a)} \\
\frac{dy}{dt} &= \frac{1}{\varepsilon^2} g_0(x,y) + \frac{1}{\varepsilon} g_1(x,y) + \frac{1}{\varepsilon}\beta(x,y)\frac{dV}{dt}. && \text{(37b)}
\end{aligned}$$

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

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(x, y) + f_1(x, y) + \alpha_0(x, y) \frac{dU}{dt} \\ &\quad + \alpha_1(x, y) \frac{dV}{dt}, \qquad (38a) \\ \frac{dy}{dt} &= \frac{1}{\varepsilon^2} g_0(x, y) + \frac{1}{\varepsilon} g_1(x, y) + \frac{1}{\varepsilon} \beta(x, y) \frac{dV}{dt}. \qquad (38b) \end{aligned}$$

- 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 $\mathbb{R}^n$.

- 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 (**??**) as the **averaging** and to (**??**) as the **homogenization** problem.
- We assume that the coefficients of SDEs (**??**) 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

$$\frac{dX}{dt} = F(X) + K(X)\frac{dW}{dt}. \tag{39}$$

- Our aim it to estimate parameters in (**??**) given $\{x(t)\}_{t\in[0,T]}$.

- The averaged/homogenized equation has the form.

$$\frac{dX}{dt} = F(X) + K(X)\frac{dW}{dt} \tag{40}$$

- The proof of (**??**) is a standard result when the state space of the fast process is compact.
- The non-compact case is more subtle. See recent work by Pardoux and Veretennikov, 2001/03/05.
- A fundamental role in the analysis is played by the Poisson equation

$$-\mathcal{L}_0\phi(y;x) = h(y;x),$$

where $\mathcal{L}_0$ is the generator of the fast process.

Homogenization

We assume that the function f_0 satisfies the centering condition

$$\int_{\mathcal{Y}} \rho(y;x) f_0(x,y)\,dy = 0.$$

Let $\Phi(y;x) \in L^2_\rho(\mathcal{Y};x)$ be the unique solution of the equation

$$-\mathcal{L}_0 \Phi(y;x) = f_0(x,y), \quad \int_{\mathcal{Y}} \rho(y;x)\Phi(y;x)\,dy = 0, \tag{41}$$

Define

$$\begin{aligned} F_0(x) &:= \int_{\mathcal{Y}} \Big(\big(\nabla_x \Phi f_0\big)(x,y) + \big(\nabla_y \Phi g_1\big)(x,y) \\ &\quad + \big(\alpha_1 \beta' : \nabla_y \nabla_x \Phi\big)(x,y)\Big) \rho(y;x)\,dy, \\ F_1(x) &:= \int_{\mathcal{Y}} f_1(x,y)\rho(y;x)\,dy \quad \text{and} \\ F(x) &= F_0(x) + F_1(x). \end{aligned}$$

Also define

$$
\begin{aligned}
A_1(x)A_1(x)' &:= \int_{\mathcal{Y}} \Big(\big(\nabla_y \Phi \beta + \alpha_1\big) \big(\nabla_y \Phi \beta + \alpha_1\big)' \Big)(x, y)\rho(y; x) dy, \\
A_0(x)A_0(x)' &:= \int_{\mathcal{Y}} \alpha_0(x, y)\alpha_0(x, y)' \rho(y; x) dy \quad \text{and} \\
K(x)K(x)' &= A_0(x)A_0(x)' + A_1(x)A_1(x)' \geqslant 0.
\end{aligned}
$$

Then $x \Rightarrow X$ in $C([0, T], \mathcal{X})$ and X solves the SDE

$$
\frac{dX}{dt} = F(X) + A(X)\frac{dW}{dt} \tag{42}
$$

where W is a standard l-dimensional Brownian motion.

- 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:

$$\frac{dX}{dt} = F(X;\theta) + K(X)\frac{dW}{dt}. \tag{43}$$

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

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

$$\mathbb{L}(\theta; z) = \int_0^T \langle F(z;\theta), dz \rangle_{a(z)} - \frac{1}{2} \int_0^T |F(z;\theta)|^2_{a(z)} dt, \tag{44}$$

- 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 (**??**) and $\mathbb{P}_0$ the path pace measure for

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

- The MLE is

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

- Choosing data from (**??**) leads to the correct estimation of drift parameters, in the limit as the length of the path $T \to \infty$:

Theorem

Assume that (**??**) *defines an ergodic Markov process with invariant distribution* $\pi(x)$ *for* $\theta = \theta_0$ *and that* $K(x)$ *is uniformly positive definite on* $\mathcal{X}$. *Let* $\{X(t)\}_{t\in[0,T]}$ *be a sample path of* (**??**) *with* $\theta = \theta_0$. *Then, in* $L^2(\Omega)$,

$$\begin{aligned} \lim_{T\to\infty} \frac{2}{T}\mathbb{L}(\theta; X) &= \mathbb{E}^{\pi}|F(X;\theta_0)|^2_{a(X)} \\ &\quad -\mathbb{E}^{\pi}|F(X;\theta) - F(X;\theta_0)|^2_{a(X)}. \end{aligned}$$

This expression is maximized by choosing $\hat{\theta} = \theta_0$, *in the limit* $T \to \infty$.

- Assume that we are given data $\{x(t)\}_{t\in[0,T]}$ from (**??**) and we want to fit it to the equation (**??**). 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 (**??**) *as well as the averaged equation* (**??**) *are ergodic. Let* $\{x(t)\}_{t\in[0,T]}$ *be a sample path of* (**??**) *and* $X(t)$ *a sample path of* (**??**) *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 $E_\infty(\theta)$.

- The generator of the slow-fast process (**??**) has the form

$$\mathcal{L}^{\varepsilon} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2.$$

- Consider the Poisson equations

$$-\mathcal{L}_0\Gamma = \langle F(x;\theta), f_0(x,y)\rangle_{a(x)}, \quad \int_{\mathcal{Y}} \rho(y;\xi)\Gamma(y;x)dy = 0 \tag{45}$$

- and

$$-\mathcal{L}_0\Phi(y;x) = f_0(x,y), \quad \int_{\mathcal{Y}} \rho(y;x)\Phi(y;x)dy = 0. \tag{46}$$

- Then E_∞ is defined as

$$E_\infty = \int_{\mathcal{X}\times\mathcal{Y}} \Big(\mathcal{L}_1\Gamma(x,y) - \langle F(x;\theta), \big(\mathcal{L}_1\Phi(x,y)\big)\rangle_{a(x)}\Big)\rho(y;x)dxdy.$$

- 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* **(??)** *and* $X(t)$ *a sample path of* **(??)** *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:*

$$\lim_{\varepsilon\to 0}\frac{1}{N\delta}\mathbb{L}^{N,\delta}(\theta;x) = \lim_{T\to\infty}\frac{1}{T}\mathbb{L}(\theta;X). \qquad (47)$$

Define

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

Then, under additional assumptions,

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

- The proof of Theorems **??** are similar to the proof of the averaging and homogenization theorems.
- The proof of Theorem **??** follows from the small ε, γ asymptotic result for the increments of the process $x(t)$:

$$x^{\varepsilon}_{n+1} - x^{\varepsilon}_{n} = F(x^{\varepsilon}_{n})\,\delta + M_n + R(\varepsilon, \delta), \tag{48}$$

- where M_n denotes a discrete martingale.

1. Langevin Equation in the High Friction Limit

- Consider the Langevin equation in the high friction (small mass) limit.

$$\varepsilon^2 \frac{d^2 x}{dt^2} = -\nabla V(x) - \frac{dx}{dt} + \sqrt{2\sigma}\frac{dW}{dt}. \tag{49}$$

- Write this equation as a first order system:

$$\frac{dx}{dt} = \frac{1}{\varepsilon} y, \quad \frac{dy}{dt} = -\frac{1}{\varepsilon}\nabla V(x) - \frac{1}{\varepsilon^2} y + \sqrt{\frac{2\sigma}{\varepsilon^2}}\frac{dV}{dt}.$$

- The limiting equation is

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

- Suppose that there is a set of parameters $\theta \in \Theta$ in the potential that we want to estimate,

$$\frac{dX}{dt} = -\nabla V(X;\theta) + \sqrt{2\sigma}\frac{dW}{dt}.$$

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

$$E_\infty(\theta) = -Z_V^{-1}\frac{\beta}{2}\int_{\mathbb{R}^d} |\nabla_q V(q;\theta)|^2 e^{-\beta V(q;\theta)}\, dq, \tag{50}$$

where $Z_V = \int_{\mathbb{R}^d} e^{-\beta V(q;\theta)}\, dq$. In particular, $E_\infty < 0$.

2. Thermal motion in a two-scale potential

$$\frac{dx}{dt} = -\nabla V^{\varepsilon}(x) + \sqrt{2\beta^{-1}}\frac{dW}{dt} \tag{51}$$

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

$$\frac{dX}{dt} = -K\nabla V(X) + \sqrt{2\beta^{-1}K}\frac{dW}{dt} \tag{52}$$

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

$$E_\infty(\theta) = \Big(-1 + \widehat{Z}_p^{-1} Z_p^{-1}\Big) \frac{\beta Z_V^{-1}}{2} \int_{\mathbb{R}} |\partial_x V|^2 e^{-\beta V(x;\theta)}\, dx. \tag{53}$$

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

$$\dot{x} = v(x,t) + \sqrt{2\kappa}\dot{W}. \tag{54}$$

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

$$\lim_{\varepsilon \to 0} \varepsilon x(t/\varepsilon^2) = \sqrt{2\mathcal{K}} W(t), \tag{55}$$

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

$$\dot{X} = \sqrt{2\mathcal{K}}\dot{W}. \tag{56}$$

- 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 (**??**) 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 (**??**).
- The **diffusive time scale** at which (**??**) is valid depends on the detailed properties of $v(x,t)$ and on κ.

- For the estimation of the eddy diffusivity we use

$$\mathcal{K}^{N,\delta} = \frac{1}{2N\delta} \sum_{n=0}^{N-1} \left(x_{n+1} - x_n\right) \otimes \left(x_{n+1} - x_n\right), \tag{57}$$

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

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} \left(x_{(j+1)\Delta t} - x_{j\Delta t}\right) \otimes \left(x_{(j+1)\Delta t} - x_{j\Delta t}\right) = 2\kappa I T, \quad \text{a.s.}, \tag{58}$$

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

$$\kappa \leqslant \mathcal{K}^{\xi} \leqslant \frac{C}{\kappa}, \tag{59}$$

- 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

$$\lim_{\kappa \to 0} \mathbb{E} |\mathcal{K}_\xi^{N,\delta} - \mathcal{K}_\xi|^2 = 0. \tag{60}$$

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

- Consider (**??**) in 2d with

$$v(x) = (0, sin(x)).$$

- The eddy diffusivity in the y direction is

$$\mathcal{K} = \kappa + \frac{1}{2\kappa}.$$

- The estimator is

$$K_\delta = \frac{1}{2N\delta} \sum_{n=0}^{N-1} (y_{n+1} - y_n)^2 . \tag{61}$$

- where $y_n = y(n\delta)$ and δ is the sampling rate.
- For this example Theorem **??** holds with $\gamma = \zeta = -2 - \varepsilon$ with $\varepsilon > 0$, arbitrarily small.

- 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

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

- 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

$$\mathcal{K}_{bx}^{N,\delta} = \frac{1}{2N\delta} \sum_{n=0}^{N-1} \left(\frac{1}{J} \sum_{j=1}^{J} x_{n+1}^j - \frac{1}{J} \sum_{j=1}^{J} x_n^j \right) \otimes \left(\frac{1}{J} \sum_{j=1}^{J} x_{n+1}^j - \frac{1}{J} \sum_{j=1}^{J} x_n^j \right). \tag{62}$$

- 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:

$$\mathcal{K}_{st}^{N,\delta} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{2N\delta}\sum_{n=0}^{N-1}\left(x_{n+1}^{j} - x_{n}^{j}\right)\otimes\left(x_{n+1}^{j} - x_{n}^{j}\right). \tag{63}$$

Effect of Observation Error

- Assume that the observed process is

$$Y_{t_j}^{\xi} = X_{t_j}^{\xi} + \theta \varepsilon_{t_j}^{\xi}, \quad j = 1, \dots N, \tag{64}$$

- 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

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

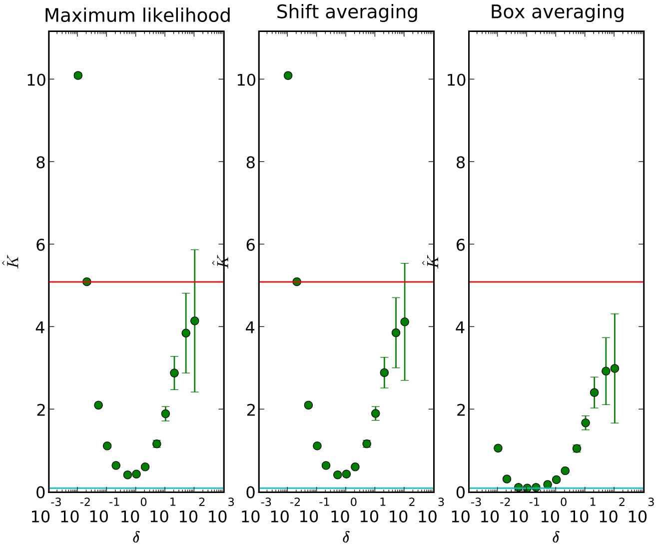


Figure: Estimated eddy diffusivity for shear flow with observation error.

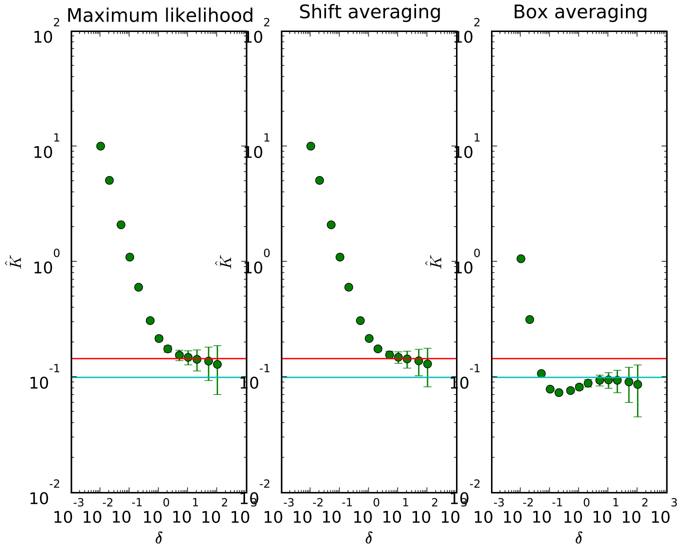


Figure: $\mathcal{K}$ for modulated shear flow with observation error.

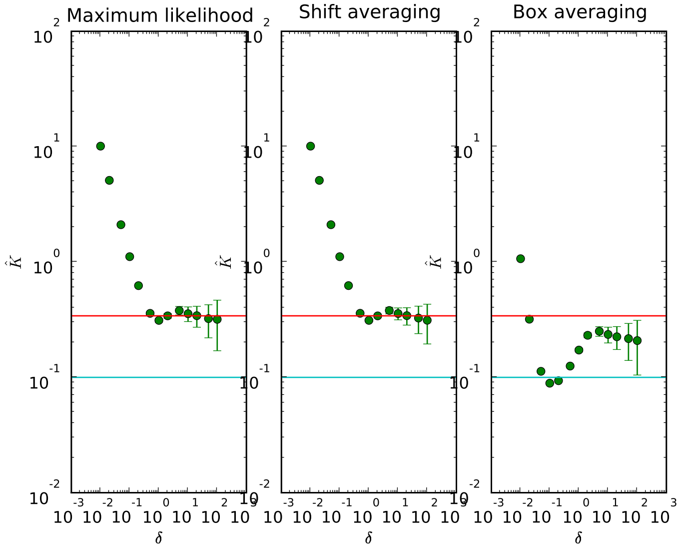


Figure: $\mathcal{K}$ for Taylor-Green flow with observation error.

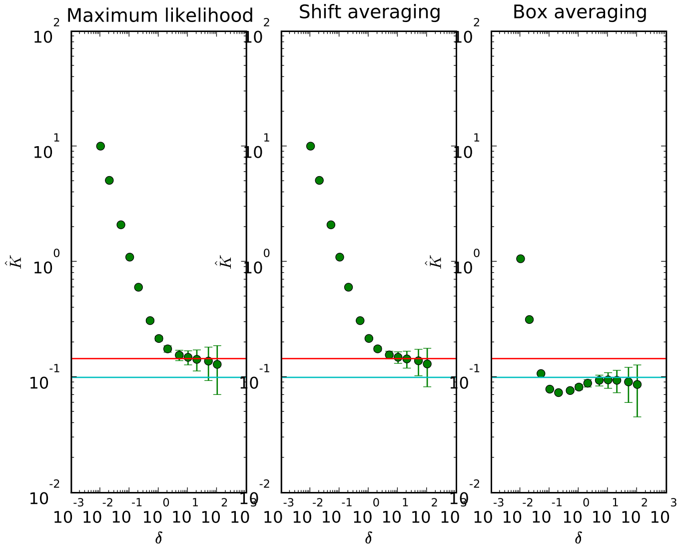


Figure: $\mathcal{K}$ for OU-modulated shear flow with observation.

Maximum Likelihood and Quadratic Variation Estimators for Multiscale Diffusions
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

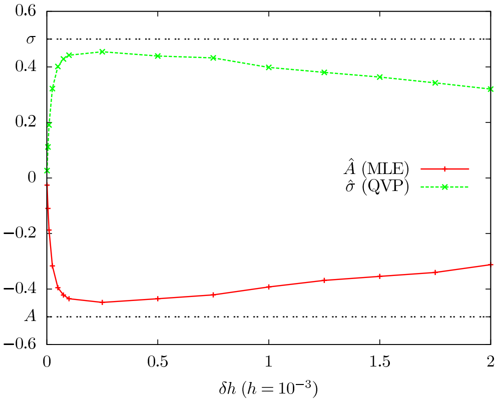
Fast/Slow System

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

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

Effective Dynamics

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



Fast OU Process Revisited

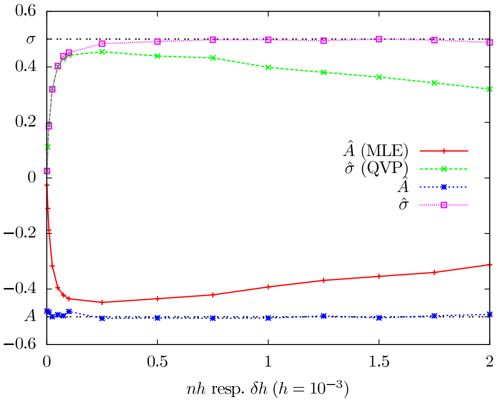
Fast/Slow System

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

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

Effective Dynamics

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



Fast OU Process II

- *Fast/slow system:*

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

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

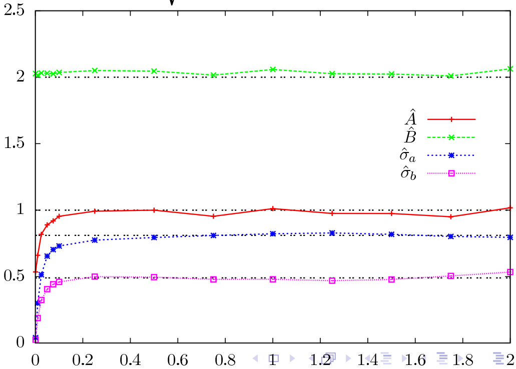
- Effective Dynamics:

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

- 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:

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

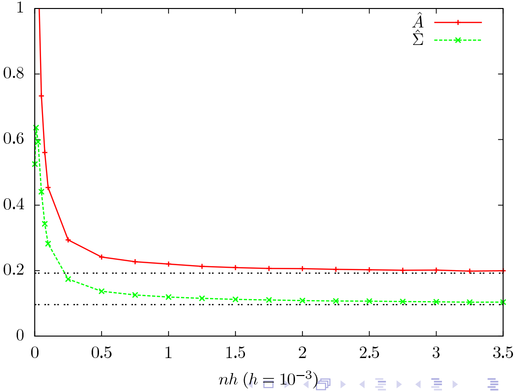
- with:

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

- 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:*

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

- Effective Dynamics: [Melbourne, Stuart '11]

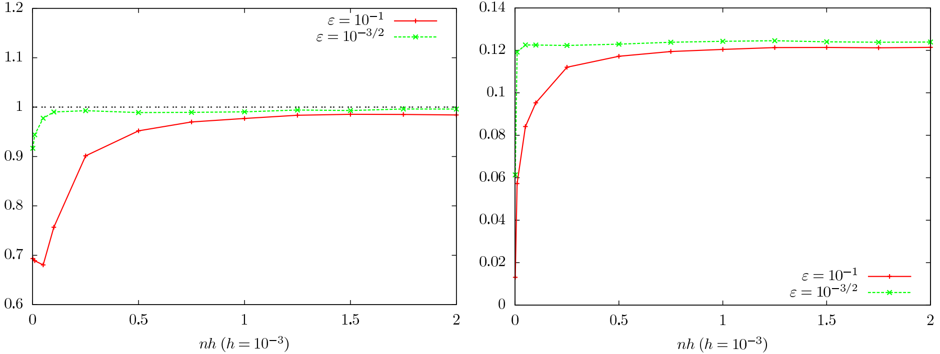
$$dX_t = A\big(X_t - {X_t}^3\big) \, dt + \sqrt{\sigma} \, dW_t$$

- 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: $\varepsilon = 10^{-1} \to \hat{\sigma} \approx 0.121$ 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

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

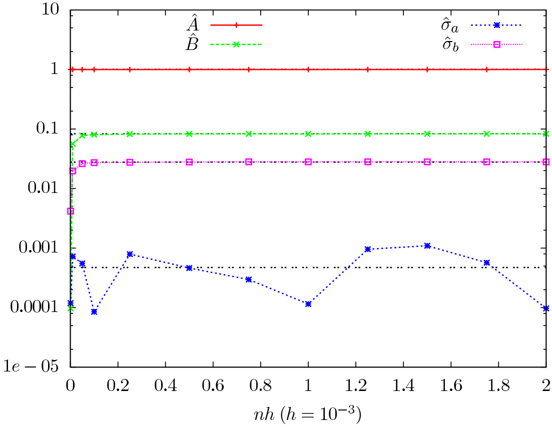
- For the three-term truncated representation the true values are:

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

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:

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

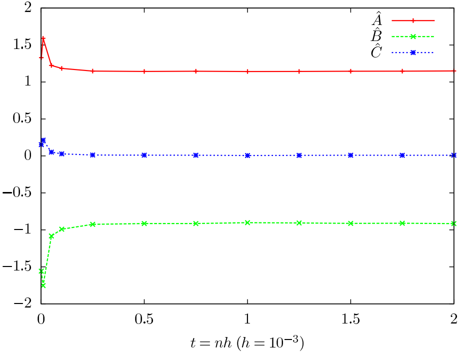
- true values ($\lambda = 2/45$):

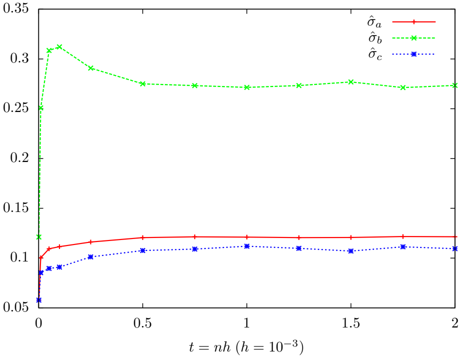
$$A = 1 + \sigma \ , \quad B = \sigma - 1 \ , \quad C = 0 \ , \quad \sigma_a = \sigma \ , \quad \sigma_b = 2\sigma \ , \quad \sigma_c = \sigma \ ,$$

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

Fast Chaotic Noise
Estimators

- $\varepsilon = 0.1$





Multiscale modeling and inverse problems

J. Nolen, A.M Stuart, G.P., in *Numerical Analysis of Multiscale Problems*, 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 multiple scales.

- We study inverse problems for PDEs with rapidly oscillating coefficients for which a homogenized equation exists.
- Unknown parameters are denoted by $u \in X$.
- We denote the data by $y \in Y$ to denote the data (usually $Y = \mathbb{R}^N$).
- z is the the predicted quantity, i.e. 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:

$$y = \mathcal{G}(u) + \xi \tag{65}$$

- for some $\xi \in \mathbb{R}^N$ quantifying model error. We will take it to be a Gaussian random variable.

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) If a homogenized solution to the inverse problem is desired, then this can be recovered from carefully designed observations of the full multiscale system.

(c) The theory of homogenization can be used to improve the estimation of homogenized parameters from observations of multiscale data.

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

$$
\begin{aligned}
\nabla \cdot v = f, &\quad x \ \text{in} \ D, \\
p = 0, &\quad x \ \text{on} \ \partial D, \\
v = -k \nabla p &
\end{aligned}
\tag{66}
$$

- 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):

$$
dx = \frac{v(x)}{\phi}\, dt + \sqrt{2\eta}\, dW, \quad x(0) = x_{\text{init}}, \tag{67}
$$

- from PDEs 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 $y \in \mathbb{R}^N$ where it is assumed that y is given by (**??**).

- 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

$$\nabla \cdot v^{\varepsilon} = f, \quad x \text{ in } D, \tag{68a}$$
$$p^{\varepsilon} = 0, \quad x \text{ on } \partial D, \tag{68b}$$
$$v^{\varepsilon} = -K^{\varepsilon} \nabla p^{\varepsilon}. \tag{68c}$$

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

$$dx^{\varepsilon} = \frac{v^{\varepsilon}(x^{\varepsilon})}{\phi} dt + \sqrt{2\eta_0 \varepsilon}\, dW, \quad x^{\varepsilon}(0) = x_{\text{init}}. \tag{69}$$

- The pressure admits the two–scale expansion

$$p^{\varepsilon}(x) \approx p_{\mathrm{a}}^{\varepsilon}(x) := p_0(x) + \varepsilon p_1(x, \frac{x}{\varepsilon}) \tag{70}$$

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

$$-\nabla_y \cdot \left(\nabla_y \chi K^T\right) = \nabla_y \cdot K^T, \quad y \in \mathbb{T}^d. \tag{71}$$

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

$$K_0(x) = \int_{\mathbb{T}^d} Q(x, y) dy, \tag{72}$$

$$Q(x, y) = K(x, y) + K(x, y) \nabla_y \chi(x, y)^T. \tag{73}$$

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

- p_0 is the solution of the homogenized PDE

$$\begin{aligned} \nabla \cdot v_0 &= f, \quad x \in D, && (74a) \\ p_0 &= g, \quad x \in \partial D, && (74b) \\ v_0 &= -K_0 \nabla p_0. && (74c) \end{aligned}$$

- and the corrector p_1 is defined by

$$p_1(x,y) = \chi(x,y) \cdot \nabla p_0(x). \tag{75}$$

- We will need the following strong convergence result.

Theorem

Let p^ε and p_0 be the solutions of (**??**) *and* (**??**). *Assume that* $f \in C^\infty(D)$ *and that* $K \in C^\infty(D; C^\infty_{\mathrm{per}}(\mathbb{T}^d))$. *Then*

$$\lim_{\varepsilon \to 0} \|p^\varepsilon - p_{\mathrm{a}}^\varepsilon\|_{H^1} = 0. \tag{76}$$

Corollary

Under the same conditions as in Theorem **??** *we have*

$$\|p^\varepsilon - p_0\|_{L^2} \to 0 \quad \text{and} \quad \|\nabla p^\varepsilon - \left(I + \chi_y(\cdot, \cdot/\varepsilon)^T\right)\nabla p_0\|_{L^2} \to 0$$

as $\varepsilon \to 0$.

Assumption

The function p^ε converges to p_0 in $L^\infty(D)$ and its gradient converges to the gradient of $p_0 + \varepsilon p_1$ in $L^\infty(D)$ so that

$$\lim_{\varepsilon \to 0} \|p^\varepsilon - p_{\mathrm{a}}^\varepsilon\|_{W^{1,\infty}} = 0.$$

- This assumption can be proved in $1d$. See also Avellaneda and Lin CPAM 40(6):803-847, 1987.
- Under this assumption we can study the asymptotic behavior of the particle trajectories.

Theorem

Let $x^{\varepsilon}(t)$ *and* $x_0(t)$ *be the solutions to equations* (**??**) *and*

$$\frac{dx_0}{dt} = \frac{v_0(x_0)}{\phi}, \quad x_0(0) = x_{\mathrm{init}}, \tag{77}$$

with velocity fields extended from $D = (L\mathbb{T})^d$ *to* $\mathbb{R}^d$ *by periodicity, and assume that Assumption* **??** *holds. Assume also that* $f \in C^{\infty}(D)$ *and that* $K \in C^{\infty}(D; C^{\infty}_{\mathrm{per}}(\mathbb{T}^d))$. *Then*

$$\lim_{\varepsilon \to 0} \mathbb{E} \sup_{0 \leqslant t \leqslant T} \|x^{\varepsilon}(t) - x_0(t)\| = 0.$$

- If the length scale ε is small, the data generated from K^{ε} and K_0 may appear very similar due to homogenization effects. Therefore, when trying to infer parameters from data, it is difficult to distinguish between K^{ε} and K_0 without some form of regularization or prior assumptions about the form of the parameter.

Regularization of Inverse Problems

- Objective: determine u, given $y \in \mathbb{R}^N$ and $\xi \sim N(0, \Gamma)$.
- Least squares functional

$$\Phi(u) = \frac{1}{2} |y - \mathcal{G}(u)|_{\Gamma}^2 \tag{78}$$

- with $|\cdot| = |\Gamma^{-1/2} \cdot|$.
- Inverse problems are ill-posed/hard to solve.
- Some kind of regularization is needed.

Regularization by Minimization Over a Compact Set

- E is a reflexive Banach space compactly embedded into X.
- Let $E_{\mathrm{ad}} = \{u \in E : \|u\|_E \leqslant \alpha\}$. Then E_{ad} is a closed bounded set in E and any sequence in E_{ad} must contain a weakly convergent subsequence with limit in E_{ad}.
- Consider the minimization problem

$$\overline{\Phi} = \inf_{u \in E_{\mathrm{ad}}} \Phi(u). \tag{79}$$

Theorem

*Any minimizing sequence $\{u^n\}_{n \in \mathbb{Z}^+}$ for (**??**) contains a weakly convergent subsequence in E with limit $\overline{u} \in E_{\mathrm{ad}}$ which attains the infimum: $\Phi(\overline{u}) = \overline{\Phi}$.*

Tikhonov Regularization

- We consider the minimization problem

$$\bar{I} = \inf_{u \in E} I(u), \tag{80}$$

- where

$$I(u) = \frac{\lambda}{2} \|u\|_E^2 + \Phi(u). \tag{81}$$

Theorem

*Any minimizing sequence $\{u^n\}_{n \in \mathbb{Z}^+}$ for (**??**) contains a weakly convergent subsequence in E with limit $\overline{u}$ which attains the infimum: $I(\overline{u}) = \bar{I}$.*

Bayesian Regularization

- In many cases it may be interesting to find a large class of solutions to the inverse problem, and to give relative weights to their importance.
- The Bayesian approach to regularization does this by adopting a probabilistic framework in which the solution to the inverse problem is a probability measure on X, rather than a single element of X.
- We think of $(u, y) \in X \times \mathbb{R}^N$ as a random variable.
- **Goal**: find the distribution of u given y, denoted by $u|y$.

- Assume that u and ξ appearing in (**??**) are independent mean zero Gaussian random variables, supported on X and $\mathbb{R}^N$ respectively, with covariance operator $\frac{1}{\lambda}\mathcal{C}$ and covariance matrix Γ respectively.
- The distribution of y given u, denoted $y|u$, is Gaussian $N(\mathcal{G}(u), \Gamma)$.
- The measure $\mu_0 = N(0, \frac{1}{\lambda}\mathcal{C})$ is known as the *prior* measure.
- When solving the inverse problem, the aim is to find the posterior measure $\mu^y(du) = \mathbb{P}(du|y)$, and to obtain information about likely candidate solutions to the inverse problem from it.

- Informal application of Bayes' theorem gives

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mu_0(u). \tag{82}$$

- The probability density function for $\mathbb{P}(y|u)$ is, using the property of Gaussians, proportional to

$$\exp\Big(-\frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2\Big) = \exp\big(-\Phi(u)\big).$$

- The infinite dimensional analogue of this result is to show that μ^y is absolutely continuous with respect to μ_0 with Radon-Nikodym derivative relating posterior to prior as follows:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp\big(-\Phi(u)\big). \tag{83}$$

- Here $\Phi(u)$ is given by **(??)** and $Z = \int_X \exp\big(-\Phi(u)\big)\mu_0(du)$.

- The meaning of the formula (**??**) is that expectations under the posterior measure μ^y can be rewritten as weighted expectations with respect to the prior: for a function $\mathcal{F}$ on X we may write

$$\int_X \mathcal{F}(u)\mu^y(du) = \int_X \frac{1}{Z}\exp\big(-\Phi(u)\big)\mathcal{F}(u)\mu_0(du).$$

- These problems have been studied rigorously in Dashti, Stuart 2011.
- The choice of prior μ_0, relates directly to the regularization of the inverse problem.

- Priors which charge functions with a multiscale character can be built in this Gaussian context.
- The formula (**??**) shows how regularization works in the Bayesian context: the main contribution to the expectation will come from places where Φ is close to its minimum value and where μ_0 is concentrated; thus minimizing Φ is important, but this minimization is regularized through the properties of the measure μ_0.
- In the Bayesian context the solution of the Tikhonov regularized problem is known as the Maximum A Posteriori estimator (MAP estimator)
- Choosing the correct regularization is part of the overall modelling scenario in which the inverse problem is embedded.

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

$$\begin{aligned} \nabla \cdot v &= f, \quad x \in D, \\ p &= 0, \quad x \in \partial D, \\ v &= -\exp(u) A \nabla p \end{aligned} \tag{84}$$

- 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 $G : \mathbb{R} \to H^1_0(D)$ 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 (**??**).
- 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 (**??**) is linear in $\exp(-u)$, we can write $G(u) = \exp(-u)p^\star$ where

$$\begin{aligned} \nabla \cdot v &= f, \quad x \in D, \\ p^\star &= 0, \quad x \in \partial D. \\ v &= -A\nabla p^\star \end{aligned} \tag{85}$$

- 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

$$\exp(-\overline{u}) = \frac{\sum_{j=1}^{N} y_j \ell_j(p^\star)}{\sum_{j=1}^{N} \ell_j(p^\star)^2}. \tag{86}$$

- It is now natural to 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 (**??**) 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

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

Proof.

Substituting the assumed expression for the data from Assumption **??** into the formula (**??**) gives

$$\exp(-\overline{u}) = \exp(-u_0) + I_1$$

where

$$I_1 = \frac{\frac{1}{N}\sum_{j=1}^{N} \xi_j \ell_j(p^\star)}{\frac{1}{N}\sum_{j=1}^{N} \ell_j(p^\star)^2}.$$

Therefore,

$$\mathbb{E}[I_1^2] = \frac{\gamma^2}{\sum_{j=1}^{N} \ell_j(p^*)^2} \leqslant \frac{2\gamma^2}{NL} \tag{87}$$

for N sufficiently large. Since I_1 is Gaussian we deduce that $\mathbb{E} I_1^{2p} = \mathcal{O}(N^{-p})$ as $N \to \infty$. Application of the Borel-Cantelli lemma shows that I_1 converges almost surely to zero as $N \to \infty$.

□

Remarks

1. *The condition that $L > 0$ prevents additional observation noise from overwhelming the information obtained from additional measurements as $N \to \infty$.*
2. *Theorem* **??** *is a example of what is known as* posterior consistency *in the theory of statistics.*

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 (**??**) 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 **??** hold and assume that that the linear functionals* ℓ_j *are chosen so that*

$$\lim_{\varepsilon\to 0}\limsup_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}|\ell_j(p^\varepsilon - p_0)|^2 = 0 \tag{88}$$

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*

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

Proof.

Notice that the solution of the homogenized equation is $p_0 = \exp(-u_0) p^\star$. We write

$$
\begin{aligned}
y_j &= \ell_j(p_0) + \ell_j(p^\varepsilon - p_0) + \xi_j \\
&= \exp(-u_0)\ell_j(p^\star) + \ell_j(p^\varepsilon - p_0) + \xi_j.
\end{aligned}
$$

Substituting this into the formula (**??**) gives

$$
\exp(-\overline{u}) = \exp(-u_0) + I_1 + I_2^\varepsilon
$$

where I_1 is as defined in the proof of Theorem **??** and is independent of ε, and

$$
I_2^\varepsilon = \frac{\sum_{j=1}^N \ell_j(p^\varepsilon - p_0)\ell_j(p^\star)}{\sum_{j=1}^N \ell_j(p^\star)^2}.
$$

The Cauchy-Schwarz inequality gives

$$|I_2^\varepsilon| \leqslant \frac{\left(\sum_{j=1}^N |\ell_j(p^\varepsilon - p_0)|^2\right)^{1/2}}{\left(\sum_{j=1}^N \ell_j(p^\star)^2\right)^{1/2}} \leqslant \left(\frac{2}{NL}\sum_{j=1}^N |\ell_j(p^\varepsilon - p_0)|^2\right)^{1/2}$$

for N sufficiently large. As in the proof of Theorem **??** we have, ξ-almost surely,

$$\lim_{N\to 0} |\exp(-\overline{u}) - \exp(-u_0) - I_2^\varepsilon| = 0.$$

From this and (**??**) the desired result now follows. □

Remarks

1. *The assumption (**??**) 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. *In particular, Corollary **??** implies that if $\{\ell_j(p)\}_{j=1}^{\infty}$ is a family of bounded linear functionals on $L^2(D)$, uniformly bounded in j, then (**??**) 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 Theorem **??** shows that (**??**) 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.*

Exploiting Multiscale Properties Within Inverse Estimation

- We consider the situation where the unknown parameter has small-scale random fluctuations.
- If we attempt to recover the homogenized parameter the error ξ appearing in (**??**) is affected by the model mismatch.
- This is because the simplified, low-dimensional parameter used to fit the data is different from the true unknown coefficient.
- Even when there is no observational noise, the error ξ has a statistical structure.
- We can use homogenization theory to predict the (universal) statistical structure of the discrepancy between $G(u)$ and y.
- This structure can be exploited in the inverse problem, as we now describe.

The Model

- We consider the two-point boundary value problem

$$-\frac{d}{dx}\left(\exp(u(x))\frac{dp}{dx}\right) = f(x), \quad x \in [-1,1], \tag{89a}$$

$$p(-1) = p(1) = 0. \tag{89b}$$

- The coefficient $k(x) = \exp(u(x))$ is a single realization of a stationary, ergodic and mixing random field $k(x,\omega)$, where (k_0 is deterministic)

$$\frac{1}{k(x,\omega)} = \frac{1}{k_0(x)} + \sigma\mu\left(\frac{x}{\varepsilon},\omega\right), \tag{90}$$

- and $\mu(x,\omega)$ is a stationary, mean zero random field with covariance

$$R(x) = \mathbb{E}(\mu(x+y)\mu(y)).$$

- We assume that $R(0) = 1$ and $\int_{\mathbb{R}} R(x)\,dx = 1$. Thus, σ^2 and ε are the (given) variance and correlation length of the fluctuations.
- We are interested in the case where $\varepsilon \ll 1$.

- In the limit as $\varepsilon \to 0$, $p_\varepsilon(x,\omega)$ coverges to $p_0(x)$ which is the solution of the homogenized Dirichlet problem

$$
\begin{aligned}
-\frac{d}{dx}\left(k_0(x)\frac{d}{dx}p_0\right) &= f(x), \quad x \in [-1,1], && \text{(91a)}\\
p_0(-1) = p_0(1) &= 0. && \text{(91b)}
\end{aligned}
$$

- Furthermore, the fluctuations of p_ε are Gaussian (Bal, Garnier, Motsch, Perrier 2008):

$$
\frac{p_\varepsilon(x,\omega) - p_0(x)}{\varepsilon^{1/2}} \to \sigma \int_D Q(x,y;k_0)v_0(y;k_0)\,dW_y(\omega) \tag{92}
$$

- in distribution as $\varepsilon \to 0$, where $W_y(\omega)$ is a Brownian random field.
- Here $v_0(x;k_0) = k_0(x)p_0(x)$, and the kernel $Q(x,y;k_0)$ is related to the Green's function for the one dimensional system.

Enhanced Estimation

- The inverse problem is to identify the parameter $k_0(x)$ in the model

$$
\begin{aligned}
-\frac{d}{dx}\left(k_0(x)\frac{d}{dx}p_0\right) &= f(x), \quad x \in [-1,1], &\text{(93a)}\\
p_0(-1) = p_0(1) &= 0. &\text{(93b)}
\end{aligned}
$$

- We assume that the data actually come from observations of $p_\varepsilon(x,\omega)$, which is the solution of the multiscale model (**??**) with $k(x,\omega)$ given by (**??**).
- Consequently, there is a discrepancy between the model used to fit the data and the true model which generates the data.
- The modelling issue is the choice of statistical model for the error ξ in (**??**).

- Suppose we make noisy observations of $p_\varepsilon(x_j)$ at points $\{x_j\}_{j=1}^N$ distributed throughout the domain. Then the measurements are

$$y_j = p_\varepsilon(x_j, \omega) + \xi_j, \quad j = 1, \dots, N$$

- where $\xi_j \sim N(0, \gamma^2)$ are mutually independent, representing observation noise. The limit (**??**) tells us that for ε small, these measurements are approximated well by

$$y_j \approx p_0(x_j) + \xi_j',$$

- where $\{\xi_j'\}_{j=1}^N$ are Gaussian random variables with mean zero and covariance

$$C_{j,\ell}(k_0, \varepsilon) = \mathbb{E}[\xi_j' \xi_\ell'] = \gamma^2 \delta_{j,\ell} + \varepsilon \sigma^2 \int_D Q(x_j, y; k_0) v_0(y; k_0)^2 Q(x_\ell, z; k_0)\, dy \tag{94}$$

- Therefore, we model the observations as

$$y_j \approx \mathcal{G}(k_0) + \xi_j', \quad j = 1, \dots, N$$

- where $\mathcal{G}(k_0) = p_0(x_j; k_0)$ with p_0 being the solution of (**??**).

- The modified statistical error ξ' has two components:
  - The first term $\gamma^2\delta_{j,\ell}$ is due to observation error.
  - The second term comes from the asymptotic theory and is associated with the random microstructure in the true parameter $k(x,\omega)$.
- If ε is very small, relative to γ^2, then the observation noise dominates (**??**) and we can ignore the error associated with the model mismatch.
- If γ^2 is small relative to ε then the statistical error ξ' is dominated by the model mismatch.

- Using the covariance (**??**), we make the approximation

$$\mathbb{P}(y|k_0) \approx \frac{1}{\sqrt{2\pi|C(k_0;\varepsilon)|}} \exp\Big(-\frac{1}{2}\big(y-\mathcal{G}(k_0)\big)^T C(k_0;\varepsilon)^{-1}\big(y-\mathcal{G}(k_0)\big)\Big).$$

- The parameter $k_0(x)$ is a function and we may place a Gaussian prior μ_0 on $u_0(x) = \log k_0(x)$. Application of Bayes' theorem (**??**) (with k_0 replacing u) gives that

$$\begin{aligned}\mathbb{P}(k_0|y) \quad &\propto \quad \frac{1}{\sqrt{2\pi|C(k_0;\varepsilon)|}} \exp\Big(-\frac{1}{2}\big(y-\mathcal{G}(k_0)\big)^T C(k_0;\varepsilon)^{-1}\big(y-\mathcal{G}(k_0)\big)\Big) \\ &\times \quad \mu_0(\log k_0).\end{aligned}$$

- The maximum a posteriori estimator (MAP) is then found as the function $k_0(x)$ which maximizes $\mathbb{P}(k_0|y)$ which is the same as minimizing $I(k_0) = -\ln\big(\mathbb{P}(k_0|y)\big)$.
- The key contribution of homogenization theory is to correctly identify the noise structure which has covariance $C(k_0;\varepsilon)$ depending on $k_0(x)$, the parameter to be estimated.

Numerical Results

- Given noisy observations of $p_\varepsilon(x_j)$ we may compute the MAP estimator $\hat{k}_1$ using the covariance $C(k_0;\varepsilon)$ given by (**??**):

$$\begin{aligned}\hat{k}_1 = \quad & \text{argmax}_{k_0} \frac{1}{\sqrt{2\pi|C(k_0;\varepsilon)|}} \\ & \exp\Big(-\frac{1}{2}\big(y-\mathcal{G}(k_0)\big)^T C(k_0;\varepsilon)^{-1}\big(y-\mathcal{G}(k_0)\big)\Big)\mu_0(\log k_0),\end{aligned}$$

- The other option is to ignore the effect of the random microstructure and simply use $C=\gamma^2 I$, accounting only for observation noise:

$$\hat{k}_2 = \text{argmax}_{k_0} \frac{1}{\sqrt{2\pi|\gamma^2 I|}} \exp\Big(-\frac{1}{2}\gamma^{-2}|y-\mathcal{G}(k_0)|^2\Big)\mu_0(\log k_0). \quad (95)$$

- Both estimates $\hat{k}_1$ and $\hat{k}_2$ are random variables, depending on the random data observed.
- It is expected that $\hat{k}_1$ gives us a better approximation of k_0, since it makes use of the true covariance (**??**).
- The numerical results are consistent with the expectation that approximation of the true covariance (through homogenization theory) yields a MAP estimator that has smaller variance, relative to the estimate that makes no use of the homogenization theory (see Figure **??**).

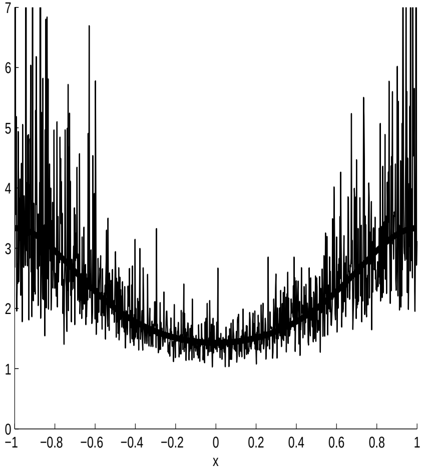


Figure: The thin erratic curve is one realization of the true coefficient $k^\varepsilon(x,\omega)$. The thick curve is the slowly-varying harmonic mean $k_0(x)$. This realization was used to generate the data.

- The data was generated as follows:
  - using one realization of $k(x, \omega)$ and given forcing f, we solve the Dirichlet boundary value problem (**??**).
  - The observation data involves point-wise evaluation of $p^{\varepsilon}(x_j)$ at points $\{x_j\}_{j=1}^{N}$ spaced uniformly across the domain, plus independent observation noise $N(0, \gamma^2)$ at each point of observation.
  - Using this data, we compute estimates $\hat{k}_1$ and $\hat{k}_2$ by minimizing (**??**) and (**??**), respectively.
  - For the computation shown the function $k_0(x)$ is parameterized by the first three coefficients in a Fourier series expansion. Consequently, computing $\hat{k}_1$ and $\hat{k}_2$ involves an optimization in $\mathbb{R}^3$.
  - To evaluate $\mathbb{P}(k_0|y)$ at each step in the minimization algorithm, we must solve the forward problem (**??**) with the current estimate of k_0, and in the case of $\hat{k}_1$ we must also compute $C(k_0, \varepsilon)$.

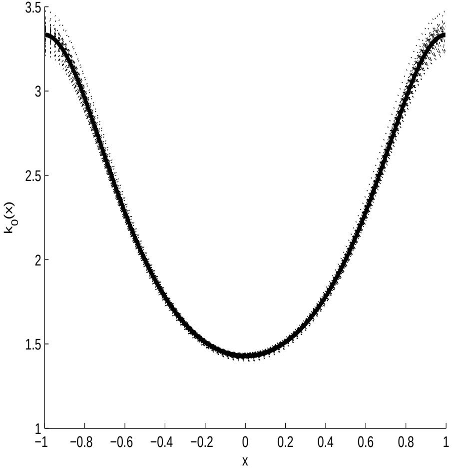


Figure: The thick curve is the true k_0. The dashed series represent 100 independent realizations of the estimate $\hat{k}_1$.

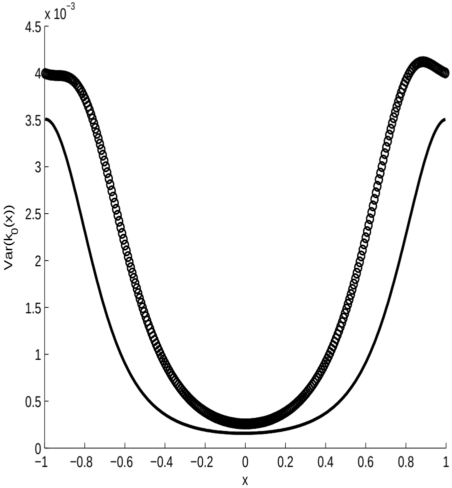


Figure: The upper series (o) is the empirical variance Var$[\hat{k}_2(x)]$. The lower series (-) is Var$[\hat{k}_1(x)]$. Both quantities were computed using 500 samples.