Learning effective models from multiscale data: filtering and Bayesian inference

ASSYR ABDULLE
(joint work with Giacomo Garegnani, Grigorios A. Pavliotis, Andrew M. Stuart, Andrea Zanoni)

1. INTRODUCTION

The problem of estimating a stochastic model from time series is important in many disciplines (e.g., chemistry, atmosphere-ocean science [5], econometrics [3], ...). Classes of models used for such inference problems are often based on stochastic differential equations (SDE) of the form

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t.$$

Inferring the drift $f(x)$ vector and the diffusion tensor $\sigma(x) = g(x)g(x)^T$ from time series is in general challenging. A major issue for such problems is that of model misspecification, when the data is not consistent with the chosen class of models. In this report we describe a new approach to learn coarse grained-models (dynamics at slow time scales) from multiscale data, based on filtering techniques. We show that robust parameter estimation can be derived and that for a class of fast/slow SDEs the theory of homogenization enables a rigorous study of the inference problem.

2. FAST SLOW SDES AND HOMOGENIZATION

We assume that the given data arise from the following class of overdamped multiscale Langevin SDEs

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t, \tag{2.1}$$

that model the motion of particles in a confining potential which has slow variations $V(x)$ with rapid oscillations superimposed $p(x/\varepsilon)$. Here $\varepsilon > 0$ represents a characteristic size of the small scales in the problem and W_t is a standard one-dimensional Brownian motion. For the rest of the paper we will assume that the fast scale is periodic. We also will assume that $\sigma > 0$, $\alpha \in \mathbb{R}^N$, $V\colon \mathbb{R} \to \mathbb{R}^N$, $V(x) = (V_1(x), V_2(x), \dots, V_N(x))^\top$, $p\colon \mathbb{R} \to \mathbb{R}$, L-periodic, with $p, V_i \in C^\infty(\mathbb{R})$. For the slow scale potential we further assume that its components V_i and $V_i^{'}$ are polynomially bounded, that V' is Lipschitz continuous and that there exist $a, b > 0$ such that $-a + bx^2 \le \alpha \cdot V'(x)x$.

Coarse-grained models. The class of models to be fitted to multiscale data are the following "homogenized models"

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t. \tag{2.2}$$

Under the assumptions above it is possible to show, via homogenization theory, that $X_t^\varepsilon \to X_t$ in law for $\varepsilon \to 0$ [4, Chapter 3]. As mentioned in the introduction, the goal is then to infer the drift coefficient A and the diffusion coefficients Σ from the multiscale data $X^\varepsilon = (X_t^\varepsilon, 0 \le t \le T)$.

3. Parameter inference, maximum likelihood estimator.

A classical way to approximate effective drift coefficients A from the coarse-grained observations X (2.2) is via path-space likelihood expressing the probability of a model X given a drift coefficient $\tilde{A}$

$$L(X \mid \tilde{A}) = \exp\left(-\frac{I(X \mid \tilde{A})}{2\Sigma}\right). \tag{3.1}$$

Maximizing the functional $L(X \mid \tilde{A})$ with respect to $\tilde{A}$ gives the maximum likelihood estimator (MLE) $\widehat{A}(X,T)$ of A defined by

$$\arg\min_{\tilde{A}\in\mathbb{R}^N} I(X \mid \tilde{A}) = -\left(\int_0^T V'(X_t) \otimes V'(X_t)\,\mathrm{d}t\right)^{-1} \int_0^T V'(X_t)\,\mathrm{d}X_t. \tag{3.2}$$

The above procedure is well understood. Our goal is however different: estimate $A \in \mathbb{R}^N$ from the multiscale observations $X^\varepsilon = (X_t^\varepsilon, 0 \le t \le T)$. As $X_t^\varepsilon \to X_t$ for $\varepsilon \to 0$ it seems reasonable to define $\widehat{A}(X^\varepsilon, T)$ for the MLE of A with multiscale data. But this turn out to be a wrong approach, indeed, under the assumptions of Section 2 this approach is shown to be biased [8, Thm. 3.4]

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{A}(X^\varepsilon, T) = \alpha. \tag{3.3}$$

Parameter inference based on subsampling. The following MLE is introduced in [8] (written here for $N = 1$ for simplicity) based on subsampling the data with step δ

$$\widehat{A}_\delta(X^\varepsilon, T) = -\frac{\sum_{i=0}^{M-1} V'(X_{i\delta}^\varepsilon)\left(X_{(i+1)\delta}^\varepsilon - X_{i\delta}^\varepsilon\right)}{\delta \sum_{i=0}^{M-1} V'(X_{i\delta}^\varepsilon)^2}, \qquad M\delta = T. \tag{3.4}$$

It is shown in [8], again under the assumptions of Section 2, that (3.4) is an asymptotically unbiased estimator of A in the limit for $\varepsilon \to 0$: if $\delta = \varepsilon^\zeta$, $0 < \zeta < 1$ and $M = \lceil \varepsilon^{-\gamma} \rceil$ with $\gamma > \zeta$, then

$$\lim_{\varepsilon\to 0} \widehat{A}_\delta(X^\varepsilon, T) = A, \quad \text{in probability}. \tag{3.5}$$

One of the main drawbacks of this approach is its lack of robustness. Indeed for a given T and ε the error depends on the choice of ζ, and it is unknown how to quantify its optimal value (see Figure 1 and [1, Section 5.1.2]). We note that other approaches based on martingale property [7], operator eigenpairs [6] have been developed (we refer to [1] for a more comprehensive literature overview).

Parameter inference based on filtering. We note that subsampling data is a "smoothing" process, so why not directly smoothing the data ? We therefore introduce a filtered process

$$Z_t^\varepsilon = \int_0^t k(t-s) X_s^\varepsilon \, \mathrm{d}s, \tag{3.6}$$

where the filter $k(r)$ is given by

$$k(r) = C_\beta \delta^{-1/\beta} e^{-\frac{r^\beta}{\delta}}, \quad C_\beta = \beta\, \Gamma(1/\beta)^{-1}, \quad \delta, \beta > 0. \tag{3.7}$$

For the rest of the paper we assume $\delta > 0, \beta = 1$. In this case the filter has the simple expression $k(r) = \frac{1}{\delta} e^{-\frac{r}{\delta}}$ and we can derive a coupled system of SDEs

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= -\alpha \cdot V'(X_t^\varepsilon) \, \mathrm{d}t - \frac{1}{\varepsilon} p' \left(\frac{X_t^\varepsilon}{\varepsilon} \right) \mathrm{d}t + \sqrt{2\sigma} \, \mathrm{d}W_t, \\ \mathrm{d}Z_t^\varepsilon &= \frac{1}{\delta} \left(X_t^\varepsilon - Z_t^\varepsilon \right) \mathrm{d}t. \end{aligned}$$

It can then be shown that $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ is geometrically ergodic with smooth invariant density $\mu^\varepsilon(\mathrm{d}x, \mathrm{d}z) = \rho^\varepsilon(x,z) \, \mathrm{d}x \, \mathrm{d}z$ that is the solution of an explicit Fokker Planck equation. We then define the filtered MLE by

$$\widehat{A}_k(X^\varepsilon, T) = - \left(\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon) \, \mathrm{d}t \right)^{-1} \int_0^T V'(Z_t^\varepsilon) \, \mathrm{d}X_t^\varepsilon. \tag{3.8}$$

We note (see [1] for a comprehensive explanation)

- $\widehat{A}_k(X^\varepsilon, T)$ is well defined if det $\left(\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon) \, \mathrm{d}t \right) \neq 0$;
- it is essential to keep $\mathrm{d}X_t^\varepsilon$ and $V'(X_t^\varepsilon)$ to prove unbiasedness;
- $\widehat{A}_k(X^\varepsilon, T)$ has to be thought as a perturbation of $\widehat{A}(X^\varepsilon, T)$ at the level of the estimator (i.e., after the maximization process of MLE);
- $\tilde{A}_k(X^\varepsilon, T) = - \left(\int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon) \, \mathrm{d}t \right)^{-1} \int_0^T V'(Z_t^\varepsilon) \, \mathrm{d}X_t^\varepsilon$ is also a valid estimator in the non-homogenized regime (when δ depends on ε).

For this estimator, under the assumptions of Section 2, we can prove [1]

Theorem 3.1 (homogenization regime)**.** *If δ is independent of ε*

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

Theorem 3.2 (multiscale regime)**.** *If $\delta = \varepsilon^\zeta, \quad \zeta \in (0,2)$*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}_k(X^\varepsilon, T) = A, \quad \textit{in probability.}$$

The value $\zeta = 2$ is critical, indeed

Theorem 3.3 (switch to biasedness)**.** *If $\delta = \varepsilon^\zeta, \quad \zeta > 2$*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}_k(X^\varepsilon, T) = \alpha, \quad \textit{in probability.}$$

For the diffusion coefficient, the estimator $\widehat{\Sigma}_k(X^\varepsilon, T) := \frac{1}{\delta T} \int_0^T \left(X_t^\varepsilon - Z_t^\varepsilon \right)^2 \mathrm{d}t$, for Σ based on filtering can be employed and proved to be unbiased.

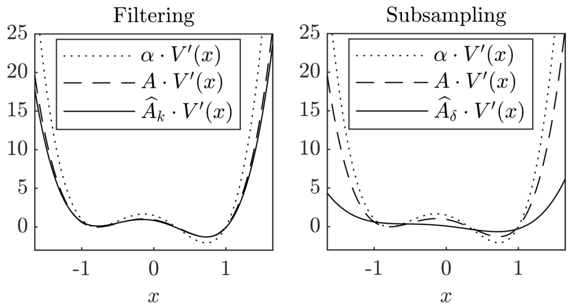

FIGURE 1. Multi-dimensional problem $N = 4$, $V_i(x) = T_i(x)$ i-th Chebyshev polynomials, $\alpha = (-1, -1/2, 1, 1/2)$, $\varepsilon = 0.05$, $T = 10^3$, subsampling $\delta = \varepsilon^{2/3}$, taken from [1].

Discussion. MLE based on filtering are robust in practice with respect to the parameter of the filter in contrast to estimators based on subsampling (see e.g. Figure 1, where the subsampling size is not optimal, but hard to find for this example). The MLE based on filtering has also been extended to the Bayesian setting to allow for a probability distribution for the effective drift A and uncertainty quantification. Finally, we note that in many applications only discrete measurements of the diffusion process are available. Recently, using the filtering approach developed in this paper and martingale estimating functions a new estimator for learning homogenised SDEs from noisy discrete data has been introduced [2].

REFERENCES

[1] A. Abdulle, G. Garegnani, G. A. Pavliotis, A. M. Stuart, and A. Zanoni, *Drift estimation of multiscale diffusions based on filtered data* (2020), arXiv preprint arXiv:2009.13457.

[2] A. Abdulle, G. A. Pavliotis, and A. Zanoni, *Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions* (2021), arXiv preprint arXiv:2104.10587.

[3] Y. Aït-Sahalia and J. Jacod, *High-frequency financial econometrics*, Princeton University Press (2014).

[4] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam (1978).

[5] C. J. Cotter and G. A. Pavliotis, *Estimating eddy diffusivities from noisy Lagrangian observations*, Commun. Math. Sci. **7** (2009), 805–838.

[6] D. Crommelin and E. Vanden-Eijnden, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Model. Simul. **9** (2011), 1588–1623.

[7] S. Kalliadasis, S. Krumscheid, and G. A. Pavliotis, *A new framework for extracting coarse-grained models from time series with multiscale structure*, J. Comput. Phys. **296** (2015), 314–328.

[8] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys. **127** (2007), 741–781.