Model Misspecification and Uncertainty Quantification for Drift Estimation in Multiscale Diffusion Processes

Assyr Abdulle [*]   Giacomo Garegnani [*]   Grigorios A. Pavliotis [†]

Andrew M. Stuart [‡]   Andrea Zanoni [*]

Abstract. Copy-paste of SIAM UQ abstract, modify. We present a novel technique for estimating the drift function of a diffusion process possessing two separated time scales. Our aim is fitting a homogenized diffusion model to a continuous sample path coming from the full multiscale process, thus dealing with an issue of model misspecification. We consider a Bayesian framework and study the asymptotic limit of posterior distributions over the drift function. In this setting, we show on the one hand that if the continuous multiscale data are not pre-processed, then the posterior distribution concentrates asymptotically on the wrong value of the drift function. On the other hand, we show that data can be treated ahead of the inference procedure in order to obtain the desired posterior. In particular, we prove that there exists a family of transformations which are linear on the space of continuous sample paths and which, when applied to multiscale data, allow the posterior distribution to be asymptotically correct. We present a series of numerical examples on test cases which corroborate our theoretical findings.

AMS subject classifications.

Keywords.

1   Introduction

Add motivating introduction. In [2] (is there other literature?), the authors prove that inference of the parameters of a homogenized model has to be performed carefully. In this work, the analysis contained in [2] is widened with respect to the following aspects:

1) a more general form of the drift function is considered, thus allowing a more flexible framework for applications and hinting possible extensions to the infinite-dimensional case,

2) the inference procedure is reinterpreted from a Bayesian perspective, which guarantees more complete uncertainty quantification on the inference result. Moreover, given the nature of the problem, posterior distributions follow a Gaussian law which can be analytically determined, thus guaranteeing computationally fast inference,

3) we extend the sub-sampling technique introduced in [2], which can be applied to discrete sequences, by introducing theoretical tools which allow the treatment of continuous streams of data.

Let $\varepsilon > 0$ and let us consider the one-dimensional multiscale stochastic differential equation (SDE)

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right) + \sqrt{2\sigma}\,\mathrm{d}W_t, \tag{1.1}$$

[*]Institute of Mathematics, École Polytechnique Fédérale de Lausanne
[†]Department of Mathematics, Imperial College London
[‡]Department of Computing and Mathematical Sciences, Caltech

where $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, N$ and $\sigma > 0$ are the drift and diffusion coefficients respectively and W_t is a standard one-dimensional Brownian motion. The functions $V \colon \mathbb{R} \to \mathbb{R}^N$ and $p \colon \mathbb{R} \to \mathbb{R}$ are slow and fast potentials driving the dynamics of the solution X_t^ε. Given $N \in \mathbb{N}_{>0}$, we assume the slow potential to be of the form

$$V(x) = \begin{pmatrix} V_1(x) & V_2(x) & \cdots & V_N(x) \end{pmatrix}^\top, \tag{1.2}$$

for smooth functions $V_i \colon \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, N$. Moreover, we assume p to be smooth and periodic of period L. Theory of homogenization [1] guarantees the existence of an SDE of the form

$$dX_t^0 = -A \cdot V'(X_t^0) \, dt + \sqrt{2\Sigma} \, dW_t, \tag{1.3}$$

where W_t is the same Brownian motion and where the fast dynamics have been eliminated, such that $X_t^\varepsilon \to X_t^0$ for $\varepsilon \to 0$ in law as random variables in $\mathcal{C}^0((0,T), \mathbb{R})$. In particular, we have $A_i = K\alpha_i$ and $\Sigma = K\sigma$, where the coefficient K is given by the formula

$$K = \int_0^L (1 + \Phi'(y))^2 \, \mu(dy),$$

with

$$\mu(dy) = Z^{-1} e^{-p(y)/\sigma} \, dy, \quad \text{where} \quad Z = \int_0^L e^{-p(y)/\sigma} \, dy,$$

and where the function Φ is the solution of the elliptic partial differential equation

$$-p'(y)\Phi'(y) + \sigma\Phi''(y) = p'(y), \quad 0 \le y \le L,$$

endowed with periodic boundary conditions.

1.1 Failure without pre-processing rewrite

Let us consider a trajectory $X^\varepsilon := (X_t^\varepsilon, 0 \le t \le T)$ coming from the multiscale equation (1.1), and the corresponding posterior distribution over the parameter A, which we denote by $\mu^\varepsilon(A \mid X^\varepsilon)$. The following result holds.

Theorem 1.1. *Under assumption 2.1 and if $T = \varepsilon^{-\gamma}$ for $\gamma > 0$, then the posterior distribution $\mu^\varepsilon(A \mid X^\varepsilon) = \mathcal{N}(\bar{A}_T^\varepsilon, \sigma_T^2)$ satisfies*

$$\lim_{\varepsilon \to 0} \bar{A}_T^\varepsilon = \alpha,$$
$$\lim_{\varepsilon \to 0} \sigma_T^2 = 0.$$

Proof. Proposition 2.2 guarantees that

$$\lim_{\varepsilon \to 0} \left| \bar{A}_T - \hat{A}(X_{0:T}^\varepsilon) \right| = 0,$$

and that $\sigma_T^2 \to 0$ for $\varepsilon \to 0$. Moreover, [2, Theorem 3.4] yields

$$\lim_{\varepsilon \to 0} \hat{A}(X_{0:T}^\varepsilon) = \alpha,$$

which completes the proof. $\qquad\square$

The result above implies that the posterior distribution over the drift coefficient concentrates asymptotically on an undesired value.

1.2 The subsampling approach for point estimates

1.3 Outline

2 Posterior formal computations

In a Bayesian setting, the goal is to determine the posterior distribution $\mu(A \mid X)$ given a continuous trajectory $X := (X_t, 0 \leq t \leq T)$. Choosing a Gaussian prior $\mu_0 = \mathcal{N}(A_0, C_0)$ on A, where $A_0 \in \mathbb{R}^N$ and $C_0 \in \mathbb{R}^{N \times N}$ is symmetric positive definite, the posterior distribution admits a density $p(A \mid X)$ with respect to the Lebesgue measure which satisfies

$$p(A \mid X) = \frac{1}{Z} \, p(X \mid A) \, p_0(A),$$

where Z is the normalization constant, p_0 is the density of μ_0, and where Girsanov's formula allows to write the likelihood as

$$p(X \mid A) = \exp\left(-\frac{I(X \mid A)}{2\Sigma}\right),$$

with

$$I(X \mid A) = \int_0^T A \cdot V'(X_t) \, \mathrm{d}X_t + \frac{1}{2} \int_0^T \left(A \cdot V'(X_t)\right)^2 \mathrm{d}t.$$

The log-posterior density is then given by

$$\log p(A \mid X) = -\log Z - \frac{1}{2\Sigma} \int_0^T A \cdot V'(X_t) \, \mathrm{d}X_t - \frac{1}{4\Sigma} \int_0^T \left(A \cdot V'(X_t)\right)^2 \mathrm{d}t$$
$$- \frac{1}{2} \left\| C_0^{-1/2}(A - A_0) \right\|_2^2.$$

Since the log-posterior density is quadratic in A, the posterior is Gaussian, and it is therefore sufficient to determine its mean and covariance to fully characterize it. We denote by m_T and C_T the mean and covariance matrix, respectively. Let us consider the matrix-valued function $M \colon \mathbb{R} \to \mathbb{R}^{N \times N}$ whose entries are given by

$$M_{ij} = \frac{1}{2\Sigma T} \int_0^T V_i'(X_t) \, V_j'(X_t) \, \mathrm{d}t, \quad i, j = 1, \ldots, N,$$

and the vector-valued function $h \colon \mathbb{R} \to \mathbb{R}^N$ defined by

$$h_i = \frac{1}{2\Sigma T} \int_0^T V_i'(X_t) \, \mathrm{d}X_t, \quad i = 1, \ldots, N.$$

Let us remark that we can rewrite M and h in vector notation as

$$M = \frac{1}{2\Sigma T} \int_0^T V'(X_t) \otimes V'(X_t) \, \mathrm{d}t, \quad h = \frac{1}{2\Sigma T} \int_0^T V'(X_t) \, \mathrm{d}X_t,$$

where $\otimes$ denotes the tensor product in $\mathbb{R}^N$. Therefore, we can rewrite the log-posterior density as

$$\log p(A \mid X) = -\log Z - TA \cdot h - \frac{T}{2} A \cdot MA - \frac{1}{2}(A - A_0) \cdot C_0^{-1}(A - A_0).$$

Completing the squares in the log-posterior density, it is possible to show that the posterior is a Gaussian $\mu(A \mid X) = \mathcal{N}(m_T, C_T)$, whose precision matrix and mean are formally given by

$$C_T^{-1} = C_0^{-1} + TM,$$
$$C_T^{-1} m_T = C_0^{-1} A_0 - Th.$$

We are now interested in the limit of the posterior distribution for $T \to \infty$. Let us first define the maximum likelihood estimator (MLE) $\widehat{A}(X)$ of A, which is obtained by maximizing the log-likelihood function $\log p(X \mid A)$, and which is hence formally given by

$$\widehat{A}(X) = -M^{-1}h.$$

We now introduce regularity assumptions on the SDE.

Assumption 2.1. The potentials p and V are such that

(i) for all $T > 0$, the symmetric matrix M is positive definite and its minimum eigenvalue satisfies $\lambda_{\min}(M) \geq \bar{\lambda} > 0$,

(ii) other assumptions to be added

We can now state the main result for asymptotic convergence of the posterior distribution.

Proposition 2.2. *Given a continuous stochastic process $X := (X_t, 0 \leq t, \leq T)$ which is ergodic with invariant measure μ^∞ such that $\mathbb{E}^{\mu^\infty}(V_i'(\cdot)) \leq C$ for some constant $C > 0$ and for all $i = 1, \dots, N$, the posterior $\mu(A \mid X)$ contracts to the limit of $\widehat{A}(X)$ for $T \to \infty$.*

Proof. Let us first consider the covariance matrix. Hua's identity yields

$$C_T = T^{-1}\left(M^{-1} - Q^{-1}\right),$$

where

$$Q = M + TMC_0M.$$

The eigenvalues of Q satisfy for all $i = 1, \dots, N$

$$\lambda_i(Q) = \lambda_i(M) + T\lambda_i(MC_0M) \geq \bar{\lambda} + T\lambda_{\min}(C_0),$$

and therefore we obtain

$$\lambda_{\max}(Q^{-1}) \leq \frac{1}{\bar{\lambda} + T\lambda_{\min}(C_0)}. \tag{2.1}$$

Similarly, we have a bound for the maximum eigenvalue of M^{-1} given by

$$\lambda_{\max}(M^{-1}) \leq \frac{1}{\bar{\lambda}}, \tag{2.2}$$

which implies that

$$\operatorname{tr} C_T \leq \frac{N}{T}\left(\frac{1}{\bar{\lambda}} + \frac{1}{\bar{\lambda} + T\lambda_{\min}(C_0)}\right),$$

and which therefore allows us to conclude that

$$\lim_{T \to \infty} \|C_T\|_F = 0.$$

We now consider the mean. Replacing the expression of the maximum likelihood estimator, we obtain

$$\left\|m_T - \widehat{A}(X)\right\|_2 = T^{-1}\left\|M^{-1}C_0^{-1}A_0 - Q^{-1}\left(C_0^{-1}A_0 - Th\right)\right\|_2.$$

Let us remark that (2.1) implies $\|Q^{-1}\|_F \to 0$ for $T \to \infty$ and (2.2) implies $\|M^{-1}\|_F \leq C$ for some $C > 0$ independently of T. Moreover, the ergodic theorem guarantees that $\|h\|_2 \leq C$ for T sufficiently big, and therefore the Cauchy–Schwarz and the triangle inequalities imply

$$\lim_{T \to \infty} \left\|m_T - \widehat{A}(X)\right\|_2 = 0,$$

which proves the desired result. $\qquad\square$

Remark 2.3. Proposition 2.2 guarantees that in the asymptotic limit of $T \to \infty$, it is equivalent to consider the Bayesian approach and the maximum likelihood approach, and can therefore be interpreted as a consistency result for both approaches. Nonetheless, in this Gaussian framework the Bayesian approach provides richer information on the inference result with a negligible additional cost.

3 A filtering approach

In the previous section, we have shown that posterior distributions over the drift coefficient of the homogenized equation are asymptotically incorrect if multiscale data are replaced into the expression of the likelihood. Hence, the need of pre-processing the data is highlighted. In particular, we consider a filtering approach. Let $k \colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ be a kernel function and consider the process $Z^\varepsilon \coloneqq (Z_t^{\varepsilon,k}, 0 \le t \le T)$ defined by the weighted average

$$Z_t^\varepsilon \coloneqq \int_0^t k(t,s) X_s^\varepsilon \, \mathrm{d}s.$$

Let $\beta, \delta > 0$ and let us consider a family of exponential kernel functions defined as

$$k(t,s) = C_\beta \delta^{-1/\beta} e^{-(t-s)^\beta/\delta}, \tag{3.1}$$

where C_β is a normalizing constant given by

$$C_\beta = \beta \, \Gamma(1/\beta)^{-1},$$

and where $\Gamma(\cdot)$ is the gamma function. In the following, only in case $\beta = 1$ a rigorous analysis is carried on. Nonetheless, numerical experiments show that for higher values of β the performances of estimators computed employing the filter are more robust and qualitatively better.

Given a trajectory $X^\varepsilon \coloneqq (X_t^\varepsilon, 0 \le t \le T)$, it is relatively inexpensive to compute Z^ε from a computational standpoint. In particular, the process Z^ε is the convolution of the kernel with the process X^ε. Hence, computational tools based on the Fast Fourier Transform (FFT) exist and allow to compute Z^ε fast component-wise.

3.1 Ergodic properties of the filter

In the following, we consider the filter with $\beta = 1$, i.e.,

$$k(t,s) = \frac{1}{\delta} e^{-\frac{t-s}{\delta}},$$

since in this case it holds

$$\mathrm{d}Z_t^\varepsilon = k(t,t) X_t^\varepsilon \, \mathrm{d}t + \int_0^t \partial_t k(t,s) X_s^\varepsilon \, \mathrm{d}s \, \mathrm{d}t = \frac{1}{\delta} \left(X_t^\varepsilon - Z_t^\varepsilon \right) \, \mathrm{d}t.$$

Considering the processes X^ε and Z^ε together, we obtain a system of stochastic differential equations

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

$$\mathrm{d}Z_t^\varepsilon = \frac{1}{\delta} \left(X_t^\varepsilon - Z_t^\varepsilon \right) \, \mathrm{d}t.$$

Let $\rho^\varepsilon = \rho^\varepsilon(x,z)$ be the stationary distribution of the joint process $(X^\varepsilon, Z^\varepsilon)^\top$. How do we show that the couple $(X^\varepsilon, Z^\varepsilon)^\top$ is ergodic. Then ρ^ε is the solution of the two-dimensional stationary Fokker–Planck equation

$$\sigma \partial_{xx}^2 \rho^\varepsilon(x,z) + \left(\alpha \cdot V''(x) + \frac{1}{\varepsilon^2} p'' \left(\frac{x}{\varepsilon} \right) + \frac{1}{\delta} \right) \rho^\varepsilon(x,z)$$
$$+ \left(\alpha \cdot V'(x) + \frac{1}{\varepsilon} p' \left(\frac{x}{\varepsilon} \right) \right) \partial_x \rho^\varepsilon(x,z) + \frac{1}{\delta}(z-x) \partial_z \rho^\varepsilon(x,z) = 0. \tag{3.2}$$

Example 3.1. A closed form solution of (3.2) can be obtained in a simple homogenized case with the dimension of the parameter $N = 1$. Let $p(y) = 0$ and the parameters α, σ be replaced respectively by A and Σ. Then, if $V(x) = x^2/2$, equation (3.2) has the analytical solution

$$\rho^0(x,z) = \frac{1}{C_{\rho^0}} \exp \left(-\frac{A}{\Sigma} \frac{x^2}{2} - \frac{1}{\delta \Sigma} \frac{(x - (1+A\delta)z)^2}{2} \right),$$

where

$$C_{\rho^0} = \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left(-\frac{A}{\Sigma}\frac{x^2}{2} - \frac{1}{\delta\Sigma}\frac{(x-(1+A\delta)z)^2}{2}\right) \, \mathrm{d}x \, \mathrm{d}z.$$

This is the density of a multivariate normal distribution $\mathcal{N}(0,\Gamma)$, where the covariance matrix is given by

$$\Gamma = \frac{\Sigma}{A(1+A\delta)}\begin{pmatrix} 1+A\delta & 1 \\ 1 & 1 \end{pmatrix}.$$

Let us remark that this distribution can be obtained from direct computations involving Gaussian processes. In particular, it is known that $X_t \sim \mathcal{GP}(\mu_t, \mathcal{C}(t,s))$, where

$$\mu_t = X_0 e^{-At}, \quad \mathcal{C}(t,s) = \frac{\Sigma}{A}\left(e^{-A|t-s|} - e^{-A(t+s)}\right).$$

The basic properties of Gaussian processes imply that Z_t is a Gaussian process, and that the couple $(X_t, Z_t)^\top$ is a Gaussian process, too, whose mean and covariance are computable explicitly.

In a general case, it is not possible to find an explicit solution to (3.2). Nevertheless, it is possible to show some relevant properties of the solution itself, which are summarized in the following Lemma.

Lemma 3.2. *Let ρ^ε be the unique solution of (3.2) and let us write*

$$\rho^\varepsilon(x,z) = \varphi^\varepsilon(z)\psi^\varepsilon(z)R^\varepsilon(x,z), \tag{3.3}$$

where φ^ε and ψ^ε are the marginal densities, i.e.,

$$\varphi^\varepsilon(x) = \int_{\mathbb{R}} \rho^\varepsilon(x,z) \, \mathrm{d}z, \quad \psi^\varepsilon(z) = \int_{\mathbb{R}} \rho^\varepsilon(x,z) \, \mathrm{d}x.$$

Then, it holds

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp\left(-\frac{1}{\sigma}\alpha \cdot V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right), \tag{3.4}$$

where

$$C_{\varphi^\varepsilon} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\sigma}\alpha \cdot V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right) \, \mathrm{d}x.$$

Moreover, it holds

$$\sigma\delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z) \, \mathrm{d}x \, \mathrm{d}z = \mathbb{E}^{\rho^\varepsilon}[((X^\varepsilon)^2 - (Z^\varepsilon)^2)V''(Z^\varepsilon)].$$

Proof. Integrating equation (3.2) with respect to z we obtain the stationary Fokker–Planck equation for the process X^ε, i.e.

$$\sigma(\varphi^\varepsilon)''(x) + \left(\alpha \cdot V''(x) + \frac{1}{\varepsilon^2}p''\left(\frac{x}{\varepsilon}\right)\right)\varphi^\varepsilon(x) + \left(\alpha \cdot V'(x) + \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)(\varphi^\varepsilon)'(x) = 0, \tag{3.5}$$

whose solution is given by

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp\left(-\frac{1}{\sigma}\alpha \cdot V(x) - \frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)\right),$$

and which proves (3.4). By integrating equation (3.2) with respect to x and integrating by parts we obtain

$$\psi^\varepsilon(z) + z(\psi^\varepsilon)'(z) = \frac{\mathrm{d}}{\mathrm{d}z}\int_{\mathbb{R}} x\rho^\varepsilon(x,z) \, \mathrm{d}x,$$

which can be written as

$$\frac{\mathrm{d}}{\mathrm{d}z}(z\psi^\varepsilon(z)) = \frac{\mathrm{d}}{\mathrm{d}z}\int_{\mathbb{R}} x\rho^\varepsilon(x,z) \, \mathrm{d}x.$$

Now, since ψ^ε is the density of a probability distribution, this implies that

$$z\psi^\varepsilon(z) = \int_{\mathbb{R}} x\rho^\varepsilon(x, z)\, \mathrm{d}x,$$

which, replacing the decomposition (3.3), yields

$$z = \int_{\mathbb{R}} x\varphi^\varepsilon(x)R^\varepsilon(x, z)\, \mathrm{d}x. \tag{3.6}$$

Replacing the decomposition (3.3) in equation (3.2) and using equation (3.5) we obtain the following equality

$$\sigma\varphi^\varepsilon\psi^\varepsilon\partial_{xx}^2 R^\varepsilon + \frac{1}{\delta}\varphi^\varepsilon\psi^\varepsilon R^\varepsilon + \sigma(\varphi^\varepsilon)'\psi^\varepsilon\partial_x R^\varepsilon + \frac{1}{\delta}(z - x)\varphi^\varepsilon((\psi^\varepsilon)'R + \psi^\varepsilon\partial_z R^\varepsilon) = 0. \tag{3.7}$$

We now multiply the equation above by x and integrate with respect to x. Let us consider some simplifications explicitly. First, an integration by parts yields

$$\sigma\psi^\varepsilon \int_{\mathbb{R}} x\varphi^\varepsilon\partial_{xx}^2 R^\varepsilon\, \mathrm{d}x + \sigma\psi^\varepsilon \int_{\mathbb{R}} x(\varphi^\varepsilon)'\partial_x R^\varepsilon\, \mathrm{d}x = -\sigma\psi^\varepsilon \int_{\mathbb{R}} \varphi^\varepsilon\partial_x R^\varepsilon\, \mathrm{d}x. \tag{3.8}$$

Then, applying (3.6), we have

$$\frac{1}{\delta}\psi^\varepsilon \int_{\mathbb{R}} x\varphi^\varepsilon R^\varepsilon\, \mathrm{d}x = \frac{1}{\delta}z\psi^\varepsilon. \tag{3.9}$$

Moreover, again applying (3.6), we can compute

$$\frac{1}{\delta}(\psi^\varepsilon)' \int_{\mathbb{R}} (z - x)x\varphi^\varepsilon R\, \mathrm{d}x = \frac{1}{\delta}(\psi^\varepsilon)'\left(z^2 - \int_{\mathbb{R}} x^2\varphi^\varepsilon R^\varepsilon\, \mathrm{d}x\right). \tag{3.10}$$

Finally, we compute the last term always applying (3.6) obtaining

$$\frac{1}{\delta}\psi^\varepsilon \int_{\mathbb{R}} (z - x)x\varphi^\varepsilon\partial_z R\, \mathrm{d}x = \frac{1}{\delta}\psi^\varepsilon\left(z - \int_{\mathbb{R}} x^2\varphi^\varepsilon\partial_z R^\varepsilon\, \mathrm{d}x\right). \tag{3.11}$$

Replacing the equalities (3.8), (3.9), (3.10) and (3.11) into (3.7) we obtain

$$(\psi^\varepsilon)'\left(z^2 - \int_{\mathbb{R}} x^2\varphi^\varepsilon R^\varepsilon\, \mathrm{d}x\right) + \psi^\varepsilon\left(2z - \int_{\mathbb{R}} x^2\varphi^\varepsilon\partial_z R^\varepsilon\, \mathrm{d}x - \delta\sigma \int_{\mathbb{R}} \varphi^\varepsilon\partial_x R^\varepsilon\, \mathrm{d}x\right) = 0.$$

We rewrite the equality above as

$$\delta\sigma\psi^\varepsilon(z) \int_{\mathbb{R}} \varphi^\varepsilon(x)\partial_x R^\varepsilon(x, z)\, \mathrm{d}x = (\psi^\varepsilon)'(z)z^2 - (\psi^\varepsilon)'(z) \int_{\mathbb{R}} x^2\varphi^\varepsilon(x)R^\varepsilon(x, z)\, \mathrm{d}x$$

$$+ 2\psi^\varepsilon(x)z - \psi^\varepsilon \int_{\mathbb{R}} x^2\varphi^\varepsilon(x)\partial_z R^\varepsilon(x, z)\, \mathrm{d}x.$$

Multiplying by $V'(z)$, integrating with respect to z and integrating by parts, we obtain the following identity in $\mathbb{R}^N$

$$\sigma\delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x, z)\, \mathrm{d}x\, \mathrm{d}z = \int_{\mathbb{R}} \int_{\mathbb{R}} (x^2 - z^2)V''(z)\rho_\varepsilon(x, z)\, \mathrm{d}x\, \mathrm{d}z$$

$$= \mathbb{E}^{\rho^\varepsilon}[((X^\varepsilon)^2 - (Z^\varepsilon)^2)V''(Z^\varepsilon)],$$

which is the desired result. $\qquad\square$

Remark 3.3. It could be possible to find a closed-form solution to (3.2) in a more general framework than Example 3.1. Nonetheless, an explicit solution is not necessary for our argument.

3.1.1 Convergence of the invariant measure

It is known that the invariant distribution of the process X^ε converges weakly to the invariant distribution of the process X^0 solution to (1.3). We now consider the random variable $(X^\varepsilon, Z^\varepsilon)^\top$ and present an analogous convergence result for the couple.

Lemma 3.4. *Let μ^ε be the invariant measure of the couple $(X^\varepsilon, Z^\varepsilon)^\top$. Then, the measure μ^ε converges weakly to the measure $\mu^0(\mathrm{d}x, \mathrm{d}z) = \rho^0(x, z)\,\mathrm{d}x\,\mathrm{d}z$, whose density ρ^0 is the unique solution of the Fokker–Planck equation*

$$\Sigma \partial_{xx}^2 \rho^0(x,z) + \left(A \cdot V''(x) + \frac{1}{\delta}\right)\rho^0(x,z) + A \cdot V'(x)\partial_x \rho^0(x,z) + \frac{1}{\delta}(z-x)\partial_z \rho^0(x,z) = 0, \quad (3.12)$$

where A and Σ are the coefficients of the homogenized equation (1.3).

Proof. TO DO $\qquad\qquad\square$

We now present an analogous result to Lemma 3.2 for the limit distribution.

Corollary 3.5. *Let ρ^0 be the unique solution of (3.12) and let us write*

$$\rho^0(x,z) = \varphi^0(z)\psi^0(z)R^0(x,z),$$

where φ^0 and ψ^0 are the marginal densities, i.e.,

$$\varphi^0(x) = \int_{\mathbb{R}} \rho^0(x,z)\,\mathrm{d}z, \quad \psi^0(z) = \int_{\mathbb{R}} \rho^0(x,z)\,\mathrm{d}x.$$

Then, if A and Σ are the coefficients of the homogenized equation (1.3), it holds

$$\varphi^0(x) = \frac{1}{C_{\varphi^0}}\exp\left(-\frac{1}{\Sigma}A \cdot V(x)\right), \quad \text{where} \quad C_{\varphi^0} = \int_{\mathbb{R}} \exp\left(-\frac{1}{\Sigma}A \cdot V(x)\right)\mathrm{d}x.$$

Moreover, it holds

$$\Sigma\delta \int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\varphi^0(x)\psi^0(z)\partial_x R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z = \mathbb{E}^{\rho^0}[((X^0)^2 - (Z^0)^2)V''(Z^0)].$$

Proof. The proof is directly obtained from Lemma 3.2 replacing $p(y) = 0$ and α, σ by A, Σ respectively. $\qquad\qquad\square$

3.2 The estimator

The estimator of the homogenized drift coefficient A we analyse is given by the formula

$$\widehat{A}^\varepsilon(T) = -\widetilde{M}^{-1}\widetilde{h}, \qquad (3.13)$$

where

$$\widetilde{M} = \frac{1}{T}\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t \qquad \text{and} \qquad \widetilde{h} = \frac{1}{T}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon.$$

In practice, one evaluation of $V'(\cdot)$ is computed with the filter, while the differential and the second evaluation of $V'(\cdot)$ are maintained with the original process X^ε.

Remark 3.6. Let us remark that in the one-dimensional case $N = 1$, the estimator reduces to

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon}{\int_0^T V'(Z_t^\varepsilon)V'(X_t^\varepsilon)\,\mathrm{d}t}.$$

Theorem 3.7. *Let $\widehat{A}^\varepsilon(T)$ be defined in (3.13). Then, if Assumptions,*

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^\varepsilon(T) = A, \quad a.s., \tag{3.14}$$

where A is the drift coefficient of the homogenized equation (1.1).

Proof. Replacing the expression of $\mathrm{d}X_t^\varepsilon$ into (3.13), we get for $\widetilde{h}$

$$\widetilde{h} = -\widetilde{M}\alpha - \frac{1}{T}\int_0^T \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)V'(Z_t^\varepsilon)\,\mathrm{d}t + \frac{\sqrt{2\sigma}}{T}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t.$$

Therefore, we have

$$\widehat{A}^\varepsilon(T) = \alpha + \frac{1}{T}\widetilde{M}^{-1}\int_0^T \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)V'(Z_t^\varepsilon)\,\mathrm{d}t - \frac{\sqrt{2\sigma}}{T}\widetilde{M}^{-1}\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t$$

$$=: \alpha + I_1^\varepsilon(T) - I_2^\varepsilon(T).$$

We study the terms I_1^ε and I_2^ε separately. The ergodic theorem applied to I_1^ε yields

$$\lim_{T \to \infty} I_1^\varepsilon(T) = \mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1}\,\mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right], \quad \text{a.s.} \tag{3.15}$$

Due to Lemma 3.2 and integrating by parts, we have

$$\mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right] = \int_\mathbb{R}\int_\mathbb{R} V'(z)\frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\frac{1}{C_{\varphi^\varepsilon}}e^{-\frac{1}{\sigma}\alpha \cdot V(x)}e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}\psi^\varepsilon(z)R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= -\sigma\int_\mathbb{R}\int_\mathbb{R}\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}\right)\frac{1}{C_{\varphi^\varepsilon}}e^{-\frac{1}{\sigma}\alpha \cdot V(x)}V'(z)\psi^\varepsilon(z)R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= \sigma\int_\mathbb{R}\int_\mathbb{R}\frac{1}{C_{\varphi^\varepsilon}}e^{-\frac{1}{\sigma}p\left(\frac{x}{\varepsilon}\right)}\partial_x\left(e^{-\frac{1}{\sigma}\alpha \cdot V(x)}R^\varepsilon(x,z)\right)V'(z)\psi^\varepsilon(z)\,\mathrm{d}x\,\mathrm{d}z,$$

which implies

$$\mathbb{E}^{\rho^\varepsilon}\left[\frac{1}{\varepsilon}p'\left(\frac{X^\varepsilon}{\varepsilon}\right)V'(Z^\varepsilon)\right] = -\left(\int_\mathbb{R}\int_\mathbb{R} V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon)\rho^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z\right)\alpha$$

$$+ \sigma\int_\mathbb{R}\int_\mathbb{R} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z$$

$$= -\mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]\alpha + \sigma\int_\mathbb{R}\int_\mathbb{R} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z.$$

Replacing the equality above into (3.15), we obtain

$$\lim_{T \to \infty} I_1^\varepsilon(T) = -\alpha + \mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1}\sigma\int_\mathbb{R}\int_\mathbb{R} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z, \quad \text{a.s.}$$

Due to Lemma 3.2, we therefore have

$$\lim_{T \to \infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}\mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon) \otimes V'(X^\varepsilon)]^{-1}\,\mathbb{E}^{\rho^\varepsilon}[((X^\varepsilon)^2 - (Z^\varepsilon)^2)V''(Z^\varepsilon)], \quad \text{a.s.}$$

We now pass to the limit as ε goes to zero and

$$\lim_{\varepsilon \to 0}\lim_{T \to \infty} I_1^\varepsilon(T) = -\alpha + \frac{1}{\delta}\mathbb{E}^{\rho^0}[V'(Z) \otimes V'(X)]^{-1}\,\mathbb{E}^{\rho^0}[(X^2 - Z^2)V''(Z)], \quad \text{a.s.},$$

where the function ρ^0 is the density of the limit invariant distribution for $\varepsilon \to 0$, as in Lemma 3.4. Due to Lemma 3.5, we have

$$\frac{1}{\delta}\mathbb{E}^{\rho^0}[(X^2 - Z^2)V''(Z)] = \Sigma\int_\mathbb{R}\int_\mathbb{R} V'(z)\varphi^0(x)\psi^0(z)\partial_x R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z,$$

9

and moreover, an integration by parts yields

$$
\begin{aligned}
\frac{1}{\delta}\,\mathbb{E}^{\rho^0}[(X^2-Z^2)V''(Z)] &= -\Sigma\int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)(\varphi^0)'(x)\psi^0(z)R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z \\
&= -\Sigma\int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\left(\frac{1}{C_{\varphi^0}}e^{-\frac{1}{\Sigma}A\cdot V(x)}\right)'\psi^0(z)R^0(x,z)\,\mathrm{d}x\,\mathrm{d}z \\
&= \left(\int_{\mathbb{R}}\int_{\mathbb{R}} V'(z)\otimes V'(x)\rho^0(x,z)\,\mathrm{d}x\,\mathrm{d}z\right)A \\
&= \mathbb{E}^{\rho^0}[V'(Z)\otimes V'(X)]A
\end{aligned}
$$

We can therefore conclude that

$$
\lim_{\varepsilon\to0}\lim_{T\to\infty} I_1(T) = -\alpha + A, \quad \text{a.s.} \tag{3.16}
$$

We now consider the second term I_2, and rewrite it as (Still one-dimensional, correct to N-dimensional parameter)

$$
I_2^\varepsilon(T) = -\sqrt{2\sigma}\,\frac{\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)^2\,\mathrm{d}t}{\frac{1}{T}\int_0^T V'(Z_t^\varepsilon)V'(X_t^\varepsilon)\,\mathrm{d}t}\,\frac{\int_0^T V'(Z_t^\varepsilon)\,\mathrm{d}W_t}{\int_0^T V'(Z_t^\varepsilon)^2\,\mathrm{d}t} =: -\sqrt{2\sigma}I_{2,1}^\varepsilon(T)I_{2,2}^\varepsilon(T).
$$

The ergodic theorem yields

$$
\lim_{T\to\infty} I_{2,1}^\varepsilon(T) = \frac{\mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon)^2]}{\mathbb{E}^{\rho^\varepsilon}[V'(Z^\varepsilon)V'(X^\varepsilon)]} =: M^\varepsilon,
$$

where M^ε is bounded uniformly in ε. Moreover, the strong law of large numbers for martingales implies

$$
\lim_{T\to\infty} I_{2,2}^\varepsilon(T) = 0, \quad \text{a.s.},
$$

independently of ε. Therefore

$$
\lim_{\varepsilon\to0}\lim_{T\to\infty} I_2^\varepsilon(T) = 0, \quad \text{a.s.},
$$

which, together with (3.16), proves (3.14). $\qquad\square$

3.3 The Bayesian setting

4 Numerical experiments

In this section we show numerical experiments confirming our theoretical findings and showing the potentiality of the filtering approach.

4.1 Parameters of the filter

We consider $N = 1$ and the quadratic potential $V(x) = x^2/2$. In this case, the solution of the homogenized equation is an Ornstein–Uhlenbeck process. Moreover, we set the multiscale equation (1.1) with $\varepsilon = 0.1$ and the fast potential $p(y) = \cos(y)$. We generate data X^ε for $0 \le t \le T$ and $T = 10^3$ employing the Euler–Maruyama method with time step $\Delta_t = \varepsilon^3$.

As a first experiment, we consider the effect of setting different values of the parameter δ of the filter (3.1) on the estimator. As a comparison and for similarity with the subsampling approach, we consider $\delta = \varepsilon^\zeta$ and vary $\zeta \in [0,1]$. We employ the same values of δ for the subsampling, too, as the theoretical results guarantee convergence only for these values. Moreover, we repeat the experiments fixing $\beta = 1,5$ and for three different values of the diffusion, i.e., we choose $\sigma = 0.5, 0.7, 1$. We report in Figure 1 the experimental results. Let us remark that
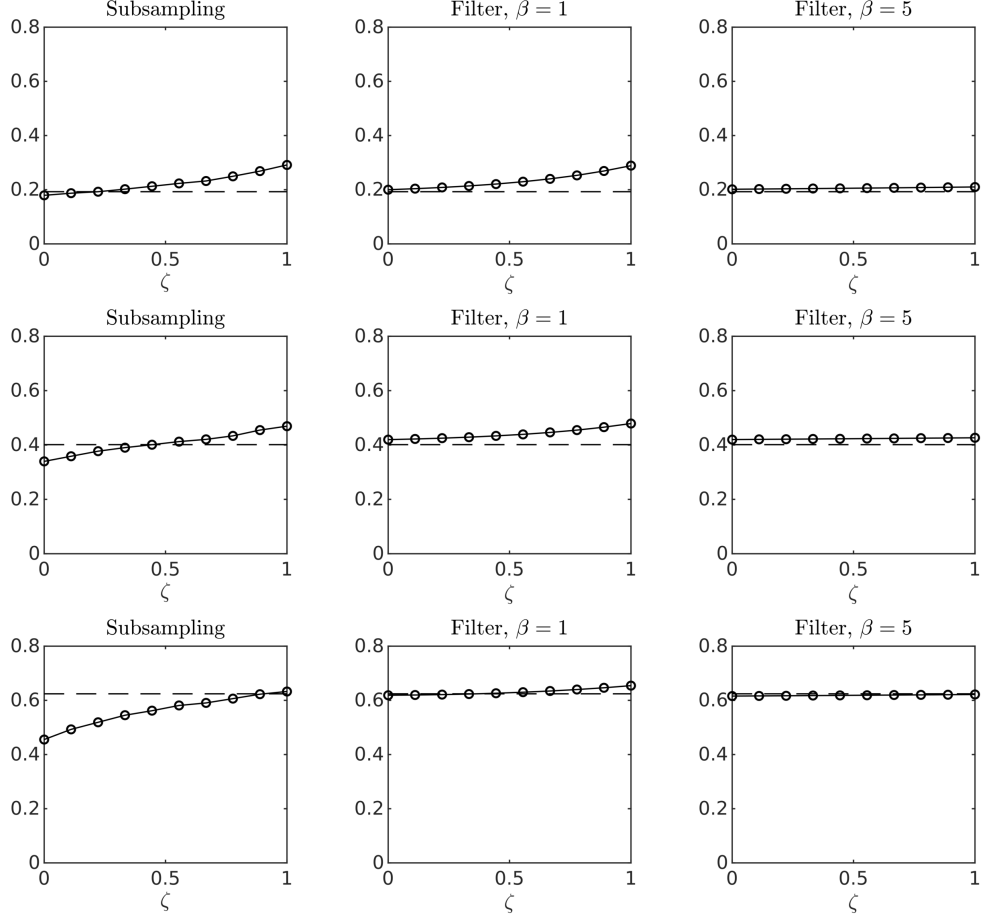
Figure 1: Results for the Ornstein–Uhlenbeck process. The axis of abscissae ζ represents the value of $\delta = \varepsilon^\zeta$ for both subsampling and the filter. The three rows correspond to $\sigma = 0.5, 0.7, 1.0$ from top to bottom.

   (i) for $\sigma = 0.5$ the results given by subsampling and by the filter with $\beta = 1$ are similar, while for higher values of σ the filtering approach seems better than subsampling,

  (ii) in general, choosing a higher value of β seems beneficial for the quality of the estimator,

 (iii) the dependence on δ of numerical results given by the filter seems relevant only in case $\beta = 1$ and for small values of σ. For $\beta = 1$ and higher values of σ, the estimator is stable with respect to this parameter, as predicted in Theorem 3.7. This can be observed for a higher value of β but we have no theoretical guarantee in this case.

As a second experiment, we therefore test the variability of the estimator with respect to β in (3.1). We repeat the numerical experiment in Figure 1 and we choose to consider $\delta = \varepsilon$, which corresponds to $\zeta = 1$ and seems to be the worst-case scenario for the filter, at least for $\beta = 1$. We consider again $\sigma = 0.5, 0.7, 1$ and vary $\beta = 1, 2, \ldots, 10$. Results, given in Figure 2, show empirically that the estimator is fast stable with respect to β.

4.2 The Bayesian approach: bistable potential

In this numerical experiment we consider $N = 2$ and the bistable potential, i.e., the function V defined as

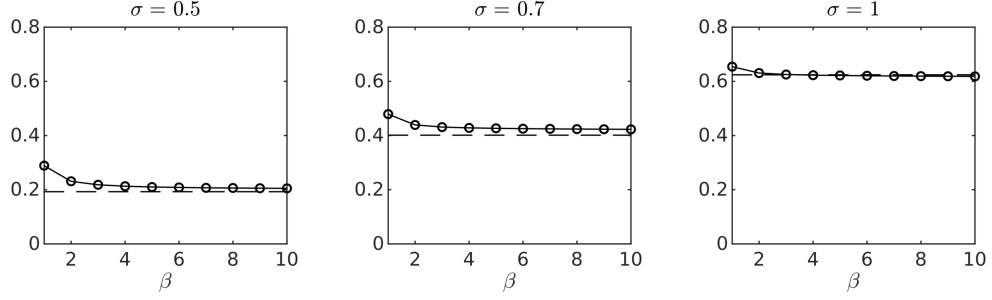$$V(x) = \left(\frac{x^4}{4} \quad \frac{x^2}{2} \right)^\top,$$

11

Figure 2: Dependence of the estimator given by the filter for the Ornstein–Uhlenbeck process on the parameter β in (3.1). From left to right we consider different values of σ.
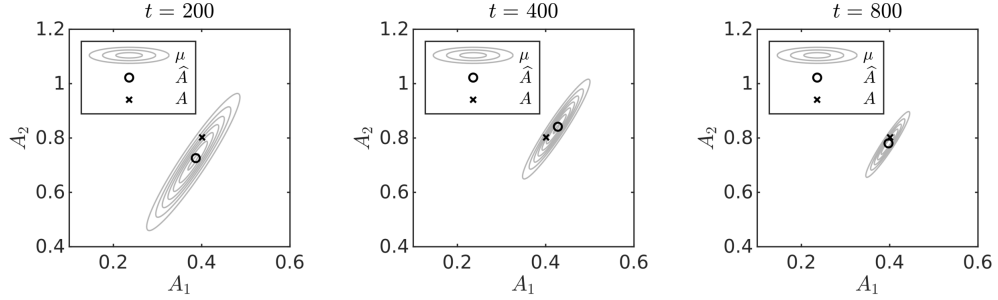

Figure 3: Posterior distributions over the parameter $A = (A_1, A_2)^\top$ for the bistable potential obtained with the filtering approach. The figures refer to time $t = 200, 400, 800$ from left to right, respectively. The MLE $\widehat{A}$ is represented with a circle, while the true value A of the drift coefficient of the homogenized equation is represented with a cross.

with coefficients $\alpha_1 = 1$ and $\alpha_2 = -2$. We then consider the multiscale equation with $\sigma = 0.8$, the fast potential $p(y) = \cos(y)$ and $\varepsilon = 0.05$, thus simulating a trajectory X^ε for time $0 \leq t \leq T$, with $T = 800$. We adopt here a Bayesian approach and compute the posterior distribution $\mu(A \mid X^\varepsilon)$ obtained with the filtering approach. The filter's parameters are set to $\beta = 1$ and $\delta = \varepsilon^{1/2}$ in (3.1). Let us remark that in order to compute the posterior covariance the diffusion coefficient Σ of the homogenized equation has to be known. In this case, we pre-compute the value of Σ via the coefficient K and the theory of homogenization, but let us remark that Σ could be estimated employing the subsampling technique of [2]. We stop computations at times $t = 200, 400, 800$ in order to observe the shrinkage of the Gaussian posterior towards the MLE with respect to time. In Figure 3, we observe that the posterior does indeed shrink towards the MLE, which in turn gets progressively closer to the true value of the drift coefficient A of the homogenized equation.

4.3 Multi-dimensional parameter

We consider $N = 4$ and the potential function $V(x)$ as in (1.2) with

$$V_i(x) = \frac{x^{2i}}{2i}, \quad i = 1, \dots, 4,$$

so that $\alpha \cdot V'(x) = \alpha_1 x + \alpha_2 x^3 + \alpha_3 x^5 + \alpha_4 x^7$, with $\alpha = (-2, -1, 1, 1/2)$. We consider $\varepsilon = 0.05$, the fast potential $p(y) = \cos(y)$ and simulate a trajectory of X^ε for $0 \leq t \leq T$ with $T = 10^3$ employing the Euler–Maruyama method with time step $\Delta_t = \varepsilon^3$. We compute the MLE of the drift coefficient $A \in \mathbb{R}^4$ of the homogenized equation (1.3) employing

- (i) the data X^ε itself,
- (ii) subsampled data with subsampling parameter $\delta = \varepsilon^{2/3}$,
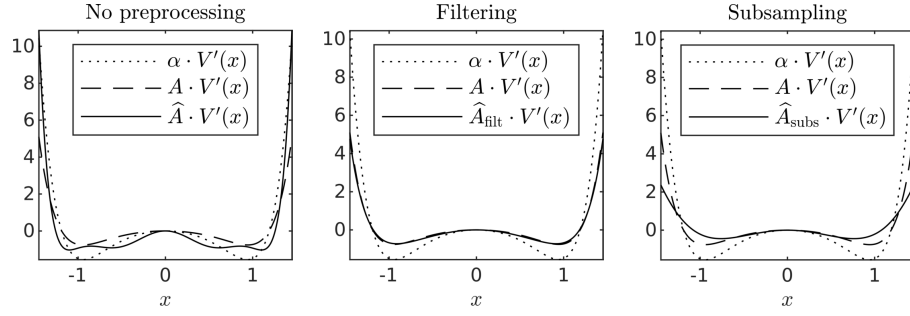
Figure 4: Results for the four-dimensional parameters. From left to right the potential function estimated with (i) the data itself, (ii) the filter, (iii) subsampled data.

(iii) filtered data Z^ε computed with $\beta = 1$ and $\delta = 0.2$.

In particular, we pick this specific value of δ for the subsampling following the optimality criterion given in [2]. Results, given in Figure 4, show that the filter-based estimation captures well the homogenized potential.

References

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

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