Bayesian inference of multiscale differential equations

Assyr Abdulle   Giacomo Garegnani

École polytechnique fédérale de Lausanne

EPFL

Caltech – 22 August 2019

Short bio

Education

- M.Mus. in Piano at Conservatorio Vivaldi (2012)
- B.Sc. in Mathematical Engineering at Politecnico di Milano (2014)
- M.Sc. in Computational Science & Engineering at EPFL (2017)
- PhD student in Assyr Abdulle's group at EPFL (since 2017)

Work experience

- STMicroelectronics – R&D intern (Grenoble, 2015)
- MindMaze – Software Development intern (Lausanne, 2016)

Research interests

- Probabilistic solvers for differential equations
- Bayesian inference of multiscale differential equations

Outline

1. Probabilistic solvers for differential equations
   - Ordinary differential equations (ODEs)
   - Elliptic partial differential equations (PDEs)
2. Bayesian inference of multiscale differential equations
   - Elliptic PDEs
   - Diffusion processes

Probabilistic solvers for differential equations

Main idea: Design numerical solvers for differential equations which account for errors in a probabilistic/statistical manner

Some contributions:

ODEs: Two families of methods

- From Runge–Kutta methods: Conrad et al. (2017), Lie et al. (2017), Abdulle and Garegnani (2018), (. . .)
- From filtering methods: Chkrebtii et al. (2016), Schober et al. (2014), Kersting and Hennig (2016), Kersting et al. (2018), (. . .)

PDEs: Conrad et al. (2017), Cockayne et al. (2017b,a) (. . .)

Probabilistic solvers for differential equations

Main idea: Design numerical solvers for differential equations which account for errors in a probabilistic/statistical manner

Motivations:

Uncertainty quantification of chaotic equations

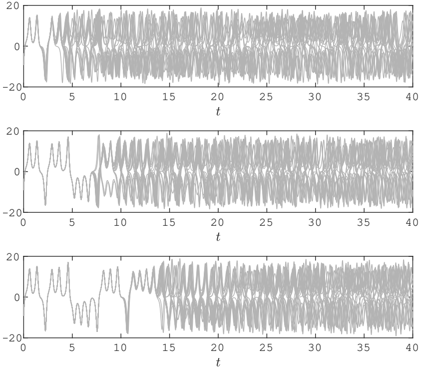


Figure : Solution of the Lorenz system with different perturbations

Probabilistic solvers for differential equations

Main idea: Design numerical solvers for differential equations which account for errors in a probabilistic/statistical manner

Motivations:

A posteriori error estimators

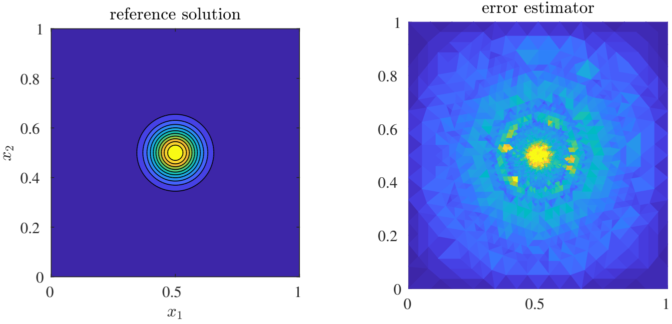


Figure : Probabilistic error estimator for a simple elliptic PDE

Probabilistic solvers for differential equations

Main idea: Design numerical solvers for differential equations which account for errors in a probabilistic/statistical manner

Motivations:

Bayesian inverse problems

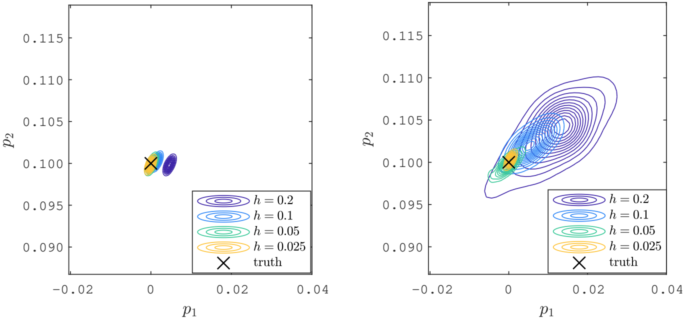


Figure : Probabilistic correction of posterior distributions.

Probabilistic solvers for ODEs

Notation: function $f : \mathbb{R}^d \to \mathbb{R}^d$, IC $y_0 \in \mathbb{R}^d$ and

$$y' = f(y), \quad y(0) = y_0.$$

Exact and numerical (Runge–Kutta) flows

$$y(t) = \varphi_t(y_0) \xrightarrow{\text{discretization}} y_{n+1} = \Psi_h(y_n)$$

Probabilistic solvers for ODEs

Notation: function $f : \mathbb{R}^d \to \mathbb{R}^d$, IC $y_0 \in \mathbb{R}^d$ and

$$y' = f(y), \quad y(0) = y_0.$$

Exact and numerical (Runge–Kutta) flows

$$y(t) = \varphi_t(y_0) \overset{\text{discretization}}{\longrightarrow} y_{n+1} = \Psi_h(y_n)$$

Additive noise method (AN-RK) (Conrad et al. (2017))
Stochastic process $\{Y_n\}_{n=1,2,\dots}$ with recurrence

$$Y_{n+1} = \underbrace{\Psi_h(Y_n)}_{\text{deterministic}} + \underbrace{\xi_n(h)}_{\text{random}}.$$

Main assumption: For $p > 1$ and $Q \in \mathbb{R}^{d \times d}$

$$\xi_n(h) \overset{\text{iid}}{\sim} \mathcal{N}(0, Qh^{2p+1}).$$

Probabilistic solvers for ODEs

Notation: function $f\colon \mathbb{R}^d \to \mathbb{R}^d$, IC $y_0 \in \mathbb{R}^d$ and

$$y' = f(y), \quad y(0) = y_0.$$

Exact and numerical (Runge–Kutta) flows

$$y(t) = \varphi_t(y_0) \stackrel{\text{discretization}}{\longrightarrow} y_{n+1} = \Psi_h(y_n)$$

Random time step method (RTS-RK) (Abdulle and Garegnani (2018))

$$Y_{n+1} = \Psi_{H_n}(Y_n),$$

Main assumption: $\{H_n\}_{n=0,1,\dots}$ iid such that for $h, C > 0$ and $p > 1$

$$H_n > 0 \text{ a.s.}, \quad \mathbb{E}\, H_n = h, \quad \operatorname{Var} H_n = Ch^{2p+1}.$$

Example: $H_n \stackrel{\text{iid}}{\sim} \mathcal{U}(h - h^{p+1/2}, h + h^{p+1/2})$.

Probabilistic solvers for ODEs – Properties

Assumptions:

- RK method of order q
- variance of random perturbations $\propto h^{2p+1}$

Properties:

Common to AN-RK and RTS-RK

- Strong convergence: $\mathbb{E}\|y(hn) - Y_n\| \leq Ch^{\min\{p,q\}}$
- Weak convergence: $|\Phi\big(y(hn)\big) - \mathbb{E}\,\Phi(Y_n)| \leq Ch^{\min\{2p,q\}}$, Φ smooth
- Good qualitative behaviour for Bayesian inference problems

Only for RTS-RK: Geometric properties

- Conservation of first integrals (e.g., mass in chemical reactions)
- Good approximation for Hamiltonian problems over long time spans

Probabilistic solvers for ODEs – Numerical example

Hénon–Heiles system (celestial dynamics), energy

$$E(v, w) = \frac{1}{2}\|v\|^2 + \frac{1}{2}\|w\|^2 + w_1^2 w_2 - \frac{1}{3} w_2^3.$$

Hamiltonian ODE ($y = (v, w)^\top$)

$$\begin{aligned} v'(t) &= -\partial_w E(v, w), \quad && v(0) = v_0, \\ w'(t) &= \partial_v E(v, w), && w(0) = w_0. \end{aligned}$$

Base numerical solver: Störmer–Verlet method (symplectic)

Goal: Determine posterior distribution over (v_0, w_0) given noisy observations of $y(t)$ and prior information on the unknown

Probabilistic solvers for ODEs – Numerical example

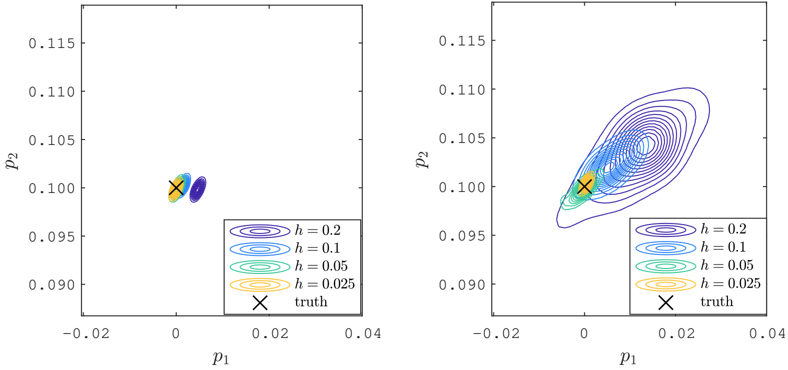


Figure : Left: deterministic Störmer–Verlet. Right: RTS-RK with Störmer–Verlet Posterior variance reflects the uncertainty due to numerical discretization.

Probabilistic solvers for PDEs

Notation: domain $\Omega \subset \mathbb{R}^d$, rhs $f \colon \Omega \to \mathbb{R}$, BC $g \colon \partial\Omega \to \mathbb{R}$, elliptic tensor $A \colon \Omega \to \mathbb{R}^{d \times d}$, equation

$$\begin{aligned} -\nabla \cdot (A \nabla u) &= f, \quad \text{in } \Omega, \\ u &= g, \quad \text{on } \partial\Omega, \end{aligned}$$

Probabilistic solvers for PDEs

Notation: domain $\Omega \subset \mathbb{R}^d$, rhs $f\colon \Omega \to \mathbb{R}$, BC $g\colon \partial\Omega \to \mathbb{R}$, elliptic tensor $A\colon \Omega \to \mathbb{R}^{d \times d}$, equation

$$\begin{aligned} -\nabla \cdot (A \nabla u) &= f, \quad \text{in } \Omega, \\ u &= g, \quad \text{on } \partial\Omega, \end{aligned}$$

Weak formulation: find $u \in V \equiv H_0^1(\Omega)$ such that

$$\int_\Omega A \nabla u \cdot \nabla v = \int_\Omega f v, \quad \forall v \in V.$$

Probabilistic solvers for PDEs

Notation: domain $\Omega \subset \mathbb{R}^d$, rhs $f\colon \Omega \to \mathbb{R}$, BC $g\colon \partial\Omega \to \mathbb{R}$, elliptic tensor $A\colon \Omega \to \mathbb{R}^{d\times d}$, equation

$$\begin{aligned} -\nabla \cdot (A\nabla u) &= f, \quad \text{in } \Omega, \\ u &= g, \quad \text{on } \partial\Omega, \end{aligned}$$

Weak formulation: find $u \in V \equiv H_0^1(\Omega)$ such that

$$\int_\Omega A\nabla u \cdot \nabla v = \int_\Omega fv, \quad \forall v \in V.$$

Galerkin projection: find $u_h \in V_h \subset V$, $\dim V_h < \infty$ such that

$$\int_\Omega A\nabla u_h \cdot \nabla v_h = \int_\Omega fv_h, \quad \forall v_h \in V_h.$$

Linear FEM: Choose $V_h = \{\text{Piecewise linear fcts. on mesh } \mathcal{T}_h \text{ of } \Omega\}$

Probabilistic solvers for PDEs

Idea: In RTS-RK randomization of time steps
$\implies$ (controlled) randomization of the mesh $\mathcal{T}_h$: move vertices with random perturbations $\propto h^p$

Properties: (partially WIP)

- A priori convergence of FEM still applies (in a strong sense)
- New angle on a posteriori convergence of FEM: employ variability on nodes as an error indicator $\implies$ mesh adaptivity!
- Same advantages on Bayesian inference problems as in ODE case

Probabilistic solvers for PDEs – Numerical example

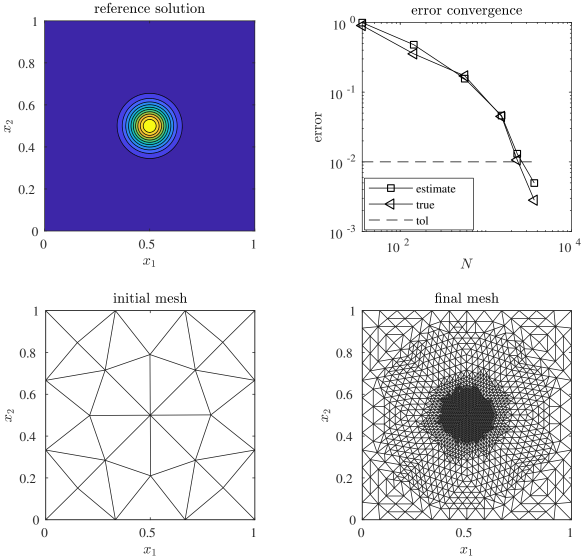


Figure : Mesh adaptivity based on probabilistic error estimators for a simple PDE

Outline

1. Probabilistic solvers for differential equations
   - Ordinary differential equations (ODEs)
   - Elliptic partial differential equations (PDEs)
2. Bayesian inference of multiscale differential equations
   - Elliptic PDEs
   - Diffusion processes

Motivation

Bayesian inference

- Fit model to data
- Full UQ approach

Differential equations

- Modelling of both deterministic and stochastic problems
- Well-developed analysis

Multiscale

- Numerous real-world applications
- Theory of homogenization applies

Multiscale elliptic PDEs

Setting: domain $\Omega \subset \mathbb{R}^d$, rhs $f\colon \Omega \to \mathbb{R}$, BC $g\colon \partial\Omega \to \mathbb{R}$, elliptic multiscale tensor $A_u^\varepsilon\colon \Omega \to \mathbb{R}^{d\times d}$

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases}$$

The function $u \in X$, X Hilbert, parametrizes the slow variations of A_u^ε.

Multiscale elliptic PDEs

Setting: domain $\Omega \subset \mathbb{R}^d$, rhs $f\colon \Omega \to \mathbb{R}$, BC $g\colon \partial\Omega \to \mathbb{R}$, elliptic multiscale tensor $A_u^\varepsilon\colon \Omega \to \mathbb{R}^{d\times d}$

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases}$$

The function $u \in X$, X Hilbert, parametrizes the slow variations of A_u^ε.

Example: One dimensional case

$$A_u^\varepsilon(x) = Ce^{u(x)} \left(2 + \sin\left(\frac{x}{\varepsilon}\right)\right),$$

where $C > 0$ and $u(x)$ could be smooth (e.g. $u(x) = \sin(x)$) or just L^2 (e.g. piecewise constant).

Multiscale elliptic PDEs

Setting: domain $\Omega \subset \mathbb{R}^d$, rhs $f\colon \Omega \to \mathbb{R}$, BC $g\colon \partial\Omega \to \mathbb{R}$, elliptic multiscale tensor $A_u^\varepsilon\colon \Omega \to \mathbb{R}^{d\times d}$

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases}$$

The function $u \in X$, X Hilbert, parametrizes the slow variations of A_u^ε.

Theory of homogenization: There exist non-oscillating A_u^0 such that

$$p^\varepsilon \rightharpoonup p^0 \text{ in } H^1(\Omega),$$

where p^0 is the solution of

$$\begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f, & \text{in } \Omega, \\ p^0 = g, & \text{on } \partial\Omega. \end{cases}$$

Multiscale elliptic PDEs

Numerical discretization: Consider multiscale and homogenized equations

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases} \qquad \begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f, & \text{in } \Omega, \\ p^0 = g, & \text{on } \partial\Omega. \end{cases}$$

Pro: Data A_u^ε is available
Con: necessary $h \ll \varepsilon$ in FEM

Pro: Cheap to solve numerically
Con: A_u^0 is unknown (only existence)

Multiscale elliptic PDEs

Numerical discretization: Consider multiscale and homogenized equations

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases} \qquad \begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f, & \text{in } \Omega, \\ p^0 = g, & \text{on } \partial\Omega. \end{cases}$$

Pro: Data A_u^ε is available
Con: necessary $h \ll \varepsilon$ in FEM

Pro: Cheap to solve numerically
Con: A_u^0 is unknown (only existence)

FE-HMM (Abdulle et al. (2012)): Numerical method for computing p^0

Main idea: Approximate A_u^0 on some points in Ω employing elliptic micro problems. Convergence properties well-established (Abdulle (2005))

Take-home message: There exist computational tools which allows to compute cheaply p^0 given the multiscale tensor A_u^ε

Multiscale elliptic PDEs – Inverse problems

General formulation

$$\text{Find } u \in X \text{ given observations } y = \mathcal{G}^\varepsilon(u) + \eta \in Y,$$

- X, Y Hilbert spaces, $\dim(Y) < \infty$
- $\eta \sim \mathcal{N}(0, \Gamma)$, Γ covariance operator on Y
- $\mathcal{G}^\varepsilon \colon X \to Y$ forward operator associated to multiscale elliptic PDE

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f, & \text{in } \Omega, \\ p^\varepsilon = g, & \text{on } \partial\Omega. \end{cases}$$

Multiscale elliptic PDEs – Inverse problems

General formulation

$$\text{Find } u \in X \text{ given observations } y = \mathcal{G}^\varepsilon(u) + \eta \in Y,$$

Bayesian interpretation: Given prior μ_{pr} on X, find posterior μ^ε such that

$$\frac{\mathrm{d}\mu^\varepsilon(u \mid y)}{\mathrm{d}\mu_{\mathrm{pr}}(u)} = \frac{1}{Z^\varepsilon} \exp(-\Phi^\varepsilon(u, y)),$$

where

$$\begin{aligned} \Phi^\varepsilon(u, y) &= \frac{1}{2} \|\mathcal{G}^\varepsilon(u) - y\|_\Gamma^2, && \text{(potential)} \\ Z^\varepsilon &= \int_X \exp(-\Phi^\varepsilon(u, y)) \mu_{\mathrm{pr}}(\mathrm{d}u). && \text{(normalizing constant)} \end{aligned}$$

Multiscale elliptic PDEs – Inverse problems

General formulation

$$\text{Find } u \in X \text{ given observations } y = \mathcal{G}^{\varepsilon}(u) + \eta \in Y,$$

Idea (Nolen et al. (2012)): Replace $\mathcal{G}^{\varepsilon}$ with $\mathcal{G}^{0}$, forward operator associated to

$$\begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f, & \text{in } \Omega, \\ p^0 = g, & \text{on } \partial\Omega. \end{cases}$$

Then define "homogenized" posterior μ^0 as

$$\frac{\mathrm{d}\mu^0(u \mid y)}{\mathrm{d}\mu_{\mathrm{pr}}(u)} = \frac{1}{Z^0} \exp(-\Phi^0(u, y)),$$

where

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

Multiscale elliptic PDEs – Inverse problems

Theoretical result (Abdulle and Di Blasio (2018)): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}(\mu^{\varepsilon}, \mu^{0}) \to 0.$$

Multiscale elliptic PDEs – Inverse problems

Theoretical result (Abdulle and Di Blasio (2018)): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}(\mu^\varepsilon, \mu^0) \to 0.$$

Consequence: we can cheaply compute the posterior μ^0 instead of μ^ε and obtain a “good” approximation of μ^ε in the computationally critical case $\varepsilon \to 0$.

Multiscale elliptic PDEs – Inverse problems

Theoretical result (Abdulle and Di Blasio (2018)): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}(\mu^\varepsilon, \mu^0) \to 0.$$

Consequence: we can cheaply compute the posterior μ^0 instead of μ^ε and obtain a "good" approximation of μ^ε in the computationally critical case $\varepsilon \to 0$.

Questions:
Q1: Does approximating μ^0 spoil multiscale convergence?
Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Multiscale elliptic PDEs – Inverse problems

Posterior approximation: Employ Ensemble Kalman Filter (EnKF) to approximate the posterior (Schillings and Stuart (2017)):

- Introduce artificial dynamics on $Z = X \times Y$ as

$$\begin{cases} z_{n+1} = \Xi(z_n), \\ \quad y_n = Hz_n + \eta_n, \end{cases}$$

  where $z_n = (u_n, \, v_n)^\top \in Z$, $\Xi(z_n) = (u_n, \mathcal{G}(u_n))^\top$ and $H = (0, \, I)$.
- Evolve ensemble $\mathbf{u}_N = \{u^{(j)}\}_{j=1}^N$ with dynamics + Kalman update
- After M steps, posterior $\mu(\mathrm{d}u)$ approximated as

$$\mu(\mathrm{d}u) \approx \widehat{\mu}_N(\mathrm{d}u) = \sum_{j=1}^{N} \delta_{u^{(j)}}(\mathrm{d}u)$$

Multiscale elliptic PDEs – Inverse problems

Q1: Does approximating μ^0 spoil multiscale convergence?

Theoretical result (Abdulle et al. (2019b)): Consider EnKF results:

$\mathcal{G}^\varepsilon$ (m.s.)	$\mathcal{G}^0$ (hom.)	$\mathcal{G}_h^0$ (hom.+FEM)
$\vert$	$\vert$	$\vert$
$\mathbf{u}_N^\varepsilon = \{(u^\varepsilon)^{(j)}\}_{j=1}^N$	$\mathbf{u}_N^0 = \{(u^0)^{(j)}\}_{j=1}^N$	$\mathbf{u}_{N,h}^0 = \{(u_h^0)^{(j)}\}_{j=1}^N$
$\downarrow$	$\downarrow$	$\downarrow$
$\widehat{\mu_N^\varepsilon}$	$\widehat{\mu_N^0}$	$\widehat{\mu_{N,h}^0}$

Multiscale elliptic PDEs – Inverse problems

Q1: Does approximating μ^0 spoil multiscale convergence?

Theoretical result (Abdulle et al. (2019b)): Consider EnKF results:

$$
\begin{array}{ccc}
\mathcal{G}^\varepsilon \text{ (m.s.)} & \mathcal{G}^0 \text{ (hom.)} & \mathcal{G}^0_h \text{ (hom.+FEM)} \\
| & | & | \\
\mathbf{u}^\varepsilon_N = \{(u^\varepsilon)^{(j)}\}_{j=1}^N & \mathbf{u}^0_N = \{(u^0)^{(j)}\}_{j=1}^N & \mathbf{u}^0_{N,h} = \{(u^0_h)^{(j)}\}_{j=1}^N \\
\downarrow & \downarrow & \downarrow \\
\widehat{\mu^\varepsilon_N} & \widehat{\mu^0_N} & \widehat{\mu^0_{N,h}}
\end{array}
$$

Then for $\varepsilon, h \to 0$

$$\widehat{\mu^\varepsilon_N} - \widehat{\mu^0_{N,h}} \xrightarrow{L^1} 0.$$

Remark: Random measures $\mu_n \xrightarrow{L^1} \mu$

$$\mathbb{E}\left| \int f \, d\mu_n - \int f \, d\mu \right| \to 0.$$

Multiscale elliptic PDEs – Inverse problems

Numerical experiment (setting from Abdulle and Di Blasio (2018))

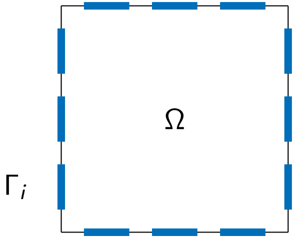
Measurements ($i = 1, \ldots, 12$)

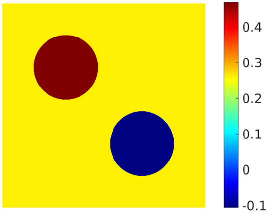
$$y_i = \int_{\Gamma_i} A^\varepsilon \nabla p_k^\varepsilon \cdot \nu \varphi_i \, \mathrm{d}s + \eta_i,$$

where φ_i test functions,
ν normal, Γ_i portions of $\partial\Omega$

Unknown: Function $u \in L^2(\Omega)$ parametrizing A^ε (amplitude of oscillations)

Goal: Verify convergence wrt ε





Multiscale elliptic PDEs – Inverse problems

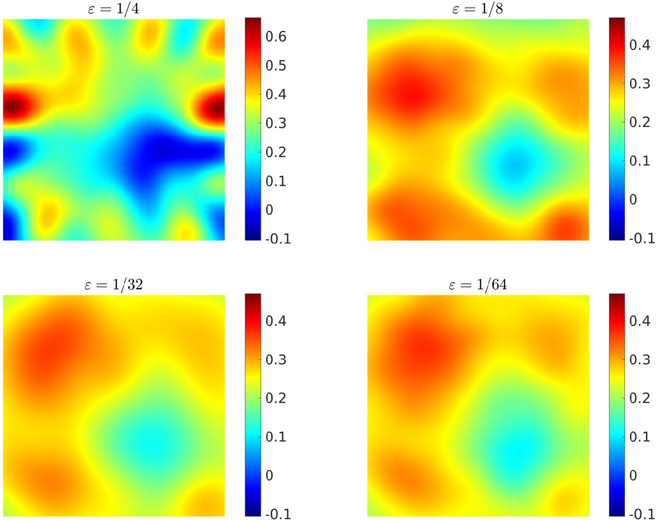


Figure : Convergence for $\varepsilon \to 0$ of the EnKF posterior estimate.

Multiscale elliptic PDEs – Inverse problems

Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Idea (Calvetti et al. (2014)): Rewrite model as

$$y = \mathcal{G}_h^0(u) + m + \eta,$$

where $m = \mathcal{G}^\varepsilon(u) - \mathcal{G}_h^0(u)$ is the modelling error. Two approaches

- Assume $m \sim \mathcal{N}(\bar{m}, \bar{\Sigma})$ independent of η, approximate offline from μ_{pr}
- Assume m independent of η and update distribution iteratively with a sequence of posterior distributions μ_l, $l = 1, \ldots, L$

Remark: Modelling error approximation is crucial if ε small but "not too small": in this case we assume that it is possible to perform a limited number of evaluations of $\mathcal{G}^\varepsilon$.

Multiscale elliptic PDEs – Inverse problems

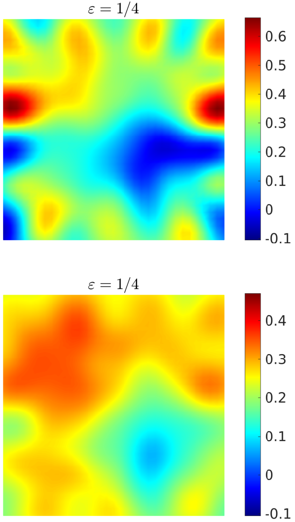


Figure : "Offline" approximation of the modelling error.

Multiscale elliptic PDEs – Inverse problems

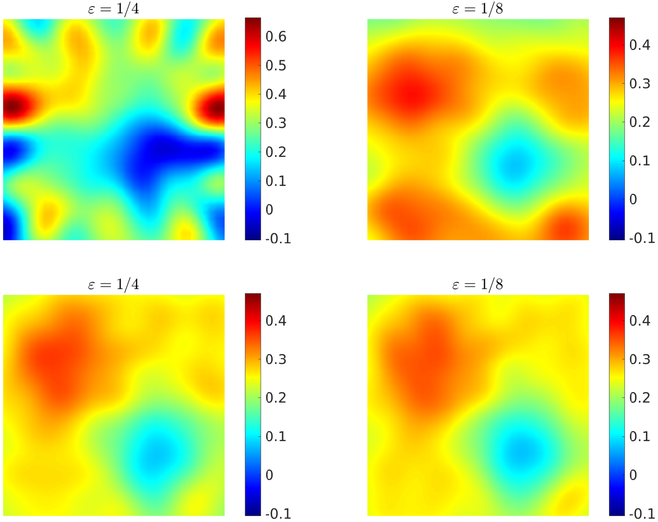


Figure : "Iterative" approximation of the modelling error.

Multiscale diffusion processes

Multiscale SDE – first order Langevin

$$\mathrm{d}x^\varepsilon(t) = -\underbrace{\alpha \nabla V_0\big(x^\varepsilon(t)\big)\,\mathrm{d}t}_{\text{large-scale potential}} - \underbrace{\frac{1}{\varepsilon}\nabla V_1\left(\frac{x^\varepsilon(t)}{\varepsilon}\right)\,\mathrm{d}t}_{\text{fluctuating potential}} + \underbrace{\sqrt{2\sigma}\,\mathrm{d}W(t)}_{\text{diffusion}}.$$

Multiscale diffusion processes

Multiscale SDE – first order Langevin

$$\mathrm{d}x^\varepsilon(t) = -\underbrace{\alpha\nabla V_0\big(x^\varepsilon(t)\big)\,\mathrm{d}t}_{\text{large-scale potential}} - \underbrace{\frac{1}{\varepsilon}\nabla V_1\left(\frac{x^\varepsilon(t)}{\varepsilon}\right)\,\mathrm{d}t}_{\text{fluctuating potential}} + \underbrace{\sqrt{2\sigma}\,\mathrm{d}W(t)}_{\text{diffusion}}.$$

Homogenized SDE

$$\mathrm{d}x^0(t) = -A\,\nabla V_0(x^0(t))\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W(t), \quad A = K\alpha,\ \Sigma = K\sigma.$$

Homogenization result (Bensoussan et al. (1978))

$$x^\varepsilon \Rightarrow x^0 \text{ in } \mathcal{C}^0\big((0,T),\mathbb{R}^d\big) \text{ for } \varepsilon \to 0.$$

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^{\varepsilon}(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_{\eta}.$$

Posterior distribution $\mu^{\varepsilon}(\theta \mid \mathbf{y})$ with density

$$p^{\varepsilon}(\theta \mid \mathbf{y}) = \frac{1}{Z^{\varepsilon}} \underbrace{p(\theta)}_{\text{prior}} \underbrace{p^{\varepsilon}(\mathbf{y} \mid \theta)}_{\text{likelihood}}, \quad Z^{\varepsilon} \text{ s.t. } \int p^{\varepsilon}(\theta \mid \mathbf{y}) \, \mathrm{d}\theta = 1.$$

Prior: Easy to evaluate (e.g. Gaussian), independent of ε

Likelihood: Needs more work

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^{\varepsilon}(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_{\eta}.$$

Likelihood: Needs more work $\Rightarrow$ marginalization

$$p^{\varepsilon}(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p^{\varepsilon}(\mathbf{y} \mid \mathbf{x}, \theta)\, p^{\varepsilon}(\mathbf{x} \mid \theta)\, \mathrm{d}\mathbf{x}.$$

where (observation independence)

$$p^{\varepsilon}(\mathbf{y} \mid \mathbf{x}, \theta) = \prod_{k=1}^{N} p^{\varepsilon}(y_k \mid x_k, \theta).$$

Observation density: $p(y_k \mid x_k, \theta) = \rho_{\eta}^{(k)}(y_k - x_k)$

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^\varepsilon(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_\eta.$$

Likelihood: Needs more work $\Rightarrow$ marginalization

$$p^\varepsilon(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p(\mathbf{y} \mid \mathbf{x}, \theta)\, p^\varepsilon(\mathbf{x} \mid \theta)\, \mathrm{d}\mathbf{x}.$$

where (observation independence)

$$p(\mathbf{y} \mid \mathbf{x}, \theta) = \prod_{k=1}^{N} p(y_k \mid x_k, \theta).$$

Observation density: $p(y_k \mid x_k, \theta) = \rho_\eta^{(k)}(y_k - x_k) \Rightarrow$ independent of ε.

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^\varepsilon(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_\eta.$$

Likelihood: Needs more work $\Rightarrow$ marginalization

$$p^\varepsilon(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p(\mathbf{y} \mid \mathbf{x}, \theta)\, p^\varepsilon(\mathbf{x} \mid \theta)\, \mathrm{d}\mathbf{x}.$$

where (Markov property)

$$p^\varepsilon(\mathbf{x} \mid \theta) = p(x_0) \prod_{k=1}^{N} p^\varepsilon(x_k \mid x_{k-1}, \theta).$$

Transition density: $p^\varepsilon(x_k \mid x_{k-1}, \theta) \Rightarrow$ only "ingredient" depending on ε.

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^\varepsilon(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_\eta.$$

Idea: Replace $p^0(\mathbf{x} \mid \theta) \approx p^\varepsilon(\mathbf{x} \mid \theta) \Rightarrow$ cheaper!

Result: Homogenized posterior $\mu^0(\theta \mid \mathbf{y})$ with density

$$p^0(\theta \mid \mathbf{y}) = \frac{1}{Z^0} p(\theta)\, p^0(\mathbf{y} \mid \theta), \quad Z^0 \text{ s.t. } \int p^0(\theta \mid \mathbf{y})\,\mathrm{d}\theta = 1,$$

with

$$p^0(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p(\mathbf{y} \mid \mathbf{x}, \theta)\, p^0(\mathbf{x} \mid \theta)\,\mathrm{d}\mathbf{x}.$$

Warning: High-dimensional integral!

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^\varepsilon(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_\eta.$$

Idea: Replace $p^0(\mathbf{x} \mid \theta) \approx p^\varepsilon(\mathbf{x} \mid \theta) \Rightarrow$ cheaper!

Result: Homogenized posterior $\mu^0(\theta \mid \mathbf{y})$ with density

$$p^0(\theta \mid \mathbf{y}) = \frac{1}{Z^0} p(\theta)\, p^0(\mathbf{y} \mid \theta), \quad Z^0 \text{ s.t. } \int p^0(\theta \mid \mathbf{y})\,\mathrm{d}\theta = 1,$$

Theoretical result (Abdulle et al. (2019a) – unpublished): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}\big(\mu^\varepsilon(\cdot \mid \mathbf{y}), \mu^0(\cdot \mid \mathbf{y})\big) \to 0.$$

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^{\varepsilon}(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_{\eta}.$$

Theoretical result (Abdulle et al. (2019a) – unpublished): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}\big(\mu^{\varepsilon}(\cdot \mid \mathbf{y}), \mu^{0}(\cdot \mid \mathbf{y})\big) \to 0.$$

Questions (Recall):

Q1: Does approximating μ^0 spoil multiscale convergence?

Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Multiscale diffusion processes – Inference problem

Inference problem

$$\text{Find } \theta = (\alpha, \sigma) \text{ given } \mathbf{y} = \mathbf{x}^\varepsilon(\theta) + \boldsymbol{\eta},\ \boldsymbol{\eta} \sim \rho_\eta.$$

Theoretical result (Abdulle et al. (2019a) – unpublished): For $\varepsilon \to 0$

$$d_{\mathrm{Hell}}\big(\mu^\varepsilon(\cdot \mid \mathbf{y}), \mu^0(\cdot \mid \mathbf{y})\big) \to 0.$$

Questions ($\approx$ Recall):

Q1: How do we sample from μ^0?

Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Multiscale diffusion processes – Inference problem

Q1: How do we sample from μ^0?

Recall: Density of $\mu^0(\theta \mid \mathbf{y})$

$$p^0(\theta \mid \mathbf{y}) = \frac{1}{Z^0} p(\theta)\, p^0(\mathbf{y} \mid \theta), \quad Z^0 \text{ s.t. } \int p^0(\theta \mid \mathbf{y})\, \mathrm{d}\theta = 1,$$

with

$$p^0(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p(\mathbf{y} \mid \mathbf{x}, \theta)\, p^0(\mathbf{x} \mid \theta)\, \mathrm{d}\mathbf{x}.$$

Multiscale diffusion processes – Inference problem

Q1: How do we sample from μ^0?

Recall: Density of $\mu^0(\theta \mid \mathbf{y})$

$$p^0(\theta \mid \mathbf{y}) = \frac{1}{Z^0} p(\theta)\, p^0(\mathbf{y} \mid \theta), \quad Z^0 \text{ s.t. } \int p^0(\theta \mid \mathbf{y})\, \mathrm{d}\theta = 1,$$

with

$$p^0(\mathbf{y} \mid \theta) = \int_{\mathbb{R}^{Nd}} p(\mathbf{y} \mid \mathbf{x}, \theta)\, p^0(\mathbf{x} \mid \theta)\, \mathrm{d}\mathbf{x}.$$

Solution: Employ a particle filter (PF) to obtain an unbiased estimator

$$\widehat{p^0}(\theta \mid \mathbf{y}) \approx p^0(\theta \mid \mathbf{y}),$$

then run a particle MCMC (PMCMC) algorithm (Andrieu et al. (2010))

Multiscale diffusion processes – Inference problem

Q1: How do we sample from μ^0?

PMCMC: Given $\theta^{(0)}$, $M \in \mathbb{N}$, proposal q

1. compute $\widehat{p}^{(0)} = \widehat{p}(\theta^{(0)} \mid \mathbf{y})$
2. For $k = 0, \dots, M$
   2.1 sample $\theta^\star \sim q(\cdot \mid \theta^{(k)})$
   2.2 compute $\hat{p}^\star = \hat{p}(\theta^\star \mid \mathbf{y})$
   2.3 compute
   $$\alpha\big(\theta^*, \theta^{(k)}\big) = \min\left\{1, \frac{\hat{p}^\star}{\hat{p}^{(k)}} \frac{q(\theta^{(k)} \mid \theta^\star)}{q(\theta^\star \mid \theta^{(k)})}\right\};$$
   2.4 with probability $\alpha\big(\theta^\star, \theta^{(k)}\big)$ set $\theta^{(k+1)} = \theta^\star$, $\hat{p}^{(k+1)} = \hat{p}^\star$, otherwise set $\theta^{(k+1)} = \theta^{(k)}$, $\hat{p}^{(k+1)} = \hat{p}^{(k)}$

Property: PF likelihood estimator unbiased $\implies$ PMCMC targets the correct posterior

Multiscale diffusion processes – Inference problem

Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Modelling error: Same idea as in the PDE case

$$\mathbf{y} = \mathbf{x}^0(\theta) + \mathbf{m} + \eta,$$

where $\mathbf{m} := \mathbf{x}^\varepsilon - \mathbf{x}^0$. We can assume $\mathbf{m}$ independent of η and θ, and compute the likelihood consequently with offline or dynamic procedures.

Multiscale diffusion processes – Inference problem

Q2: What if $\varepsilon \ll 1$ (expensive) but ε far from asymptotic limit $\varepsilon \to 0$?

Modelling error: Same idea as in the PDE case

$$\mathbf{y} = \mathbf{x}^0(\theta) + \mathbf{m} + \eta,$$

where $\mathbf{m} := \mathbf{x}^\varepsilon - \mathbf{x}^0$. We can assume $\mathbf{m}$ independent of η and θ, and compute the likelihood consequently with offline or dynamic procedures.

Idea (latest): Use PF for estimation of modelling error on state

$$\mathbf{X} := (\mathbf{x}^\varepsilon, \mathbf{m})^\top,$$

for which we have dynamics

$$\begin{cases} X_{k+1} \sim \begin{pmatrix} p(x^\varepsilon \mid x_k^\varepsilon) \\ p(m \mid m_k) \end{pmatrix}, & \text{(transition)}, \\ y_{k+1} = H X_{k+1} + \eta_{k+1}, & \text{(observation)} \end{cases}$$

where $H = (I, 0)^\top$.

Multiscale diffusion processes – Inference problem

Numerical experiment (setting from Pavliotis and Stuart (2007)):

Consider one-dimensional multiscale SDE

$$\mathrm{d}x^\varepsilon(t) = -\alpha x^\varepsilon(t)\,\mathrm{d}t + \frac{1}{\varepsilon}\sin\left(\frac{x^\varepsilon}{\varepsilon}\right) + \sqrt{2\sigma}\mathrm{d}W(t),$$

with $\varepsilon = 0.1$ (not so small) and $\alpha = 1$, $\sigma = 0.5$. In this case

$$\mathrm{d}x^0(t) = -Ax^0(t)\,\mathrm{d}t + \sqrt{2\Sigma}\mathrm{d}W(t),$$

where

$$\Sigma = \frac{4\sigma\pi^2}{Z\widehat{Z}}, \quad A = \frac{4\alpha\pi^2}{Z\widehat{Z}},$$

and

$$Z = \int_0^{2\pi} e^{-\cos(y)/\sigma}\,\mathrm{d}y, \quad \widehat{Z} = \int_0^{2\pi} e^{\cos(y)/\sigma}\,\mathrm{d}y.$$

Goal: Determine posterior over (α, σ) employing homogenized model.

Multiscale diffusion processes – Inference problem

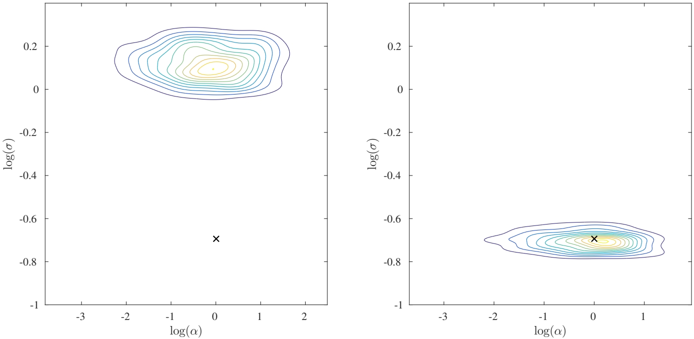


Figure : Parameter estimation without and with (iterative) estimation of modelling error.

References I

Abdulle, A. (2005). On a priori error analysis of fully discrete heterogeneous multiscale FEM. *Multiscale Model. Simul.*, 4(2):447–459.

Abdulle, A. and Di Blasio, A. (2018). A Bayesian numerical homogenization method for elliptic multiscale inverse problems. Submitted to SIAM UQ.

Abdulle, A., E, W., Engquist, B., and Vanden-Eijnden, E. (2012). The heterogeneous multiscale method. *Acta Numer.*, 21:1–87.

Abdulle, A. and Garegnani, G. (2018). Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration. arXiv preprint arXiv:1801.01340.

Abdulle, A., Garegnani, G., and Pavliotis, G. (2019a). Bayesian parameter inference of multiscale diffusion processes. Unpublished.

Abdulle, A., Garegnani, G., and Zanoni, A. (2019b). Ensemble kalman filter for multiscale inverse problems. arXiv preprint arXiv:1908.05495.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, pages 269 – 342.

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

Calvetti, D., Ernst, O., and Somersalo, E. (2014). Dynamic updating of numerical model discrepancy using sequential sampling. *Inverse Problems*, 30(11):114019, 19.

References II

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., and Girolami, M. A. (2016). Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.*, 11(4):1239–1267.

Cockayne, J., Oates, C., Sullivan, T., and Girolami, M. (2017a). Bayesian probabilistic numerical methods. arXiv preprint arXiv:1702.03673.

Cockayne, J., Oates, C., Sullivan, T., and Girolami, M. (2017b). Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. *AIP Conference Proceedings*, 1853(1):060001.

Conrad, P. R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. (2017). Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Stat. Comput.*, 27(4):1065–1082.

Kersting, H. and Hennig, P. (2016). Active uncertainty calibration in Bayesian ODE solvers. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016)*, pages 309–318. AUAI Press.

Kersting, H., Sullivan, T. J., and Hennig, P. (2018). Convergence rates of Gaussian ODE filters. arXiv preprint arXiv:1807.09737.

Lie, H. C., Stuart, A. M., and Sullivan, T. J. (2017). Strong convergence rates of probabilistic integrators for ordinary differential equations. arXiv preprint arXiv:1703.03680.

Nolen, J., Pavliotis, G. A., and Stuart, A. M. (2012). Multiscale modeling and inverse problems. In *Numerical analysis of multiscale problems*, volume 83 of *Lect. Notes Comput. Sci. Eng.*, pages 1–34. Springer, Heidelberg.

References III

Pavliotis, G. A. and Stuart, A. M. (2007). Parameter estimation for multiscale diffusions. *J. Stat. Phys.*, 127(4):741–781.

Schillings, C. and Stuart, A. M. (2017). Analysis of the ensemble Kalman filter for inverse problems. *SIAM J. Numer. Anal.*, 55(3):1264–1290.

Schober, M., Duvenaud, D., and Hennig, P. (2014). Probabilistic ODE solvers with Runge–Kutta means. In *Advances in Neural Information Processing Systems 27*, pages 739–747. Curran Associates, Inc.