Derivative-free Bayesian Inversion Using Multiscale Dynamics

MINGuS-BMS workshop

Urbain Vaes
urbain.vaes@inria.fr

MATHERIALS – Inria Paris & CERMICS – École des Ponts ParisTech

May 23, 2024

Collaborators and references

José Carrillo

UNIVERSITY OF OXFORD

Mathematical Institute

Antonin Della Noce

Inria

Inria & École des Ponts

Grigorios Pavliotis

Imperial College London

Department of Mathematics

Andrew Stuart

CALIFORNIA INSTITUTE OF TECHNOLOGY 1891

Department of Computing +
Mathematical Sciences

References:

- J. A. Carrillo and U. Vaes. Wasserstein stability estimates for covariance-preconditioned Fokker–Planck equations. Nonlinearity, 2021
- G. A. Pavliotis, A. M. Stuart, and U. Vaes. Derivative-free Bayesian inversion using multiscale dynamics. SIAM J. Appl. Dyn. Syst., 2022
- U. Vaes. Sharp propagation of chaos for the ensemble Langevin sampler. Arxiv preprint, 2024

Outline

Inverse problems: optimization and sampling approaches

Solving inverse problems using EnKF

A multiscale derivative-free methodology

Paradigmatic inverse problem

Find an unknown parameter $u \in \mathcal{U}$ from data $y \in \mathbf{R}^K$ where

$$y = \mathcal{G}(u) + \eta, \tag{IP}$$

- $\mathcal{G}$ is the forward operator;
- η is observational noise.

Two difficulties[1] associated with this problem are the following:

- Because of the noise, it might be that $y \notin \mathrm{Im}(\mathcal{G})$;
- The problem might be underdetermined.

Additionally, in many PDE applications,

- $\mathcal{G}$ is expensive to evaluate;
- The derivatives of $\mathcal{G}$ are difficult to calculate;
- u is a function $\rightarrow$ infinite dimension.

[1] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In Handbook of uncertainty quantification. Vol. 1, 2, 3. Springer, Cham, 2017.

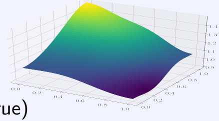
Example: inference of the thermal conductivity in a plate

Mathematical model:

$$
\begin{aligned}
-\nabla \cdot \big(u(x)\nabla T(x)\big) &= f(x), && x \in \Omega, \\
T(x) &= 0, && x \in \partial\Omega.
\end{aligned}
$$

Unknown parameter:

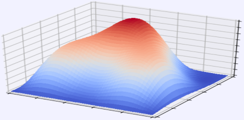
Thermal conductivity $u(x)$



(true)

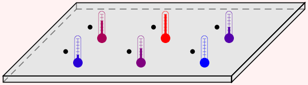
Forward problem

Solution:



Temperature field $T(x)$

Data:

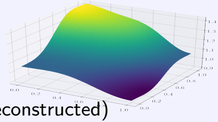


Noisy temperature measurements:

$$y = \big(T(x_1), \ldots, T(x_m)\big) + \eta.$$

Inverse problem

MAP estimator:



(reconstructed)

Optimization approach for solving "$y = \mathcal{G}u + \eta$"

Optimization approach

Find a minimizer of the regularized least-squares functional

$$u^{\dagger} = \arg\min_{u \in \mathcal{U}} \left(\frac{1}{2} |y - \mathcal{G}(u)|_{\Gamma}^{2} + R(u) \right),$$

where $|x|_A^2 := \langle x, x \rangle_A := \langle x, A^{-1} x \rangle$ and $R(u)$ is a regularization term.

- Example regularization (Tikhonov):

$$R(u) = \frac{1}{2} |u - m|_{\Sigma}^{2}.$$

- Modeling step: choice of Γ, m, Σ.

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

Linear $\mathcal{G}$ $\Rightarrow$ Quadratic Φ $\Leftrightarrow$ Gaussian $\mathrm{e}^{-\Phi}$

Probabilistic approach for solving "$y = \mathcal{G}u + \eta$"[1]

Bayesian approach to inverse problems

Modeling step:

- Probability distribution on parameter: $u \sim \pi$, encoding prior knowledge;
- Probability distribution for noise: $\eta \sim \nu$.

An application of Bayes' theorem gives the posterior distribution $\rho^y(u) = \mathbf{P}[u|y]$ as

$$\rho^y(u) \propto \pi(u)\,\nu\big(y - \mathcal{G}(u)\big) \qquad \text{(valid in finite dimension).}$$

In the Gaussian case where $\pi = \mathcal{N}(m, \Sigma)$ and $\nu = \mathcal{N}(0, \Gamma)$,

$$\rho^y(u) \propto \exp\left(-\left(\frac{1}{2}\,|y - \mathcal{G}(u)|_\Gamma^2 + \frac{1}{2}\,|u - m|_\Sigma^2\right)\right) =: \exp\big(-\Phi_R(u)\big).$$

Two approaches for extracting information:

- Find the maximizer of $\rho^y(u)$ (maximum a posteriori estimation);
- Sample the posterior distribution $\rho^y(u)$.

---

[1] A. M. Stuart. Inverse problems: a Bayesian perspective. Acta Numer., 2010.

Outline

Inverse problems: optimization and sampling approaches

Solving inverse problems using EnKF

A multiscale derivative-free methodology

Key idea: introduce artificial dynamical system and data

Artificial state-estimation problem amenable to EnKF

Dynamical system:

$$u_{n+1} = u_n.$$

Data model:

$$y_{n+1} = \mathcal{G}(u_{n+1}) + \eta_{n+1}, \qquad \eta_{n+1} \sim \mathcal{N}(0, h^{-1}\Gamma), \qquad h = \frac{1}{N}$$

Initialization: $u_0 \sim \pi =$ prior distribution from inverse problem.

Let (μ_n) denote the associated filtering distribution:

$$\mu_n(\mathrm{d}u) =: \mathrm{Law}(u_n \mid y_1, \dots, y_n)$$

By Bayes' theorem, it holds that

$$\mu_{n+1}(\mathrm{d}u) \propto \exp\Big(-h \left|y_{n+1} - \mathcal{G}(u)\right|_\Gamma^2\Big)\, \mu_n(\mathrm{d}u)$$

Therefore, with artificial data from the inverse problem $y_{n+1} = y$ for all n,

$$\mu_N(\mathrm{d}u) \propto \exp\Big(- \left|y - \mathcal{G}(u)\right|_\Gamma^2\Big)\, \pi(\mathrm{d}u)$$ = **posterior distribution from inverse problem**

Ensemble Kalman filter for inverse problem[1,2]

Since filtering distribution at time N coincides with posterior distribution of (IP),

$\rightsquigarrow$ Filtering methods can be used to solve (IP).

In particular, application of **ensemble Kalman filter** gives

$$u_{n+1}^{(j)} = u_n^{(j)} + C^{up}(U_n)\Big(C^{pp}(U_n) + h^{-1}\Gamma\Big)^{-1}\Big(y + \eta_{n+1}^{(j)} - \mathcal{G}\left(u_n^{(j)}\right)\Big), \qquad j = 1, \dots, J,$$

which is an **interacting particle system** for $\{u^{(j)}\}_{j=1}^{J}$. Here $\left(\eta_{n+1}^{(j)}\right) \sim \mathcal{N}(0, \Gamma)$ and

$$C^{pp}(U) = \frac{1}{J}\sum_{j=1}^{J}\big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big) \otimes \big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big), \qquad \overline{u} = \frac{1}{J}\sum_{j=1}^{J} u^{(j)}, \qquad \overline{\mathcal{G}} = \frac{1}{J}\sum_{j=1}^{J}\mathcal{G}\left(u^{(j)}\right)$$

$$C^{up}(U) = \frac{1}{J}\sum_{j=1}^{J}\big(u^{(j)} - \overline{u}\big) \otimes \big(\mathcal{G}(u^{(j)}) - \overline{\mathcal{G}}\big).$$

An empirical approximation of the posterior is obtained from the ensemble at iteration N:

$$\rho^y \approx \frac{1}{J}\sum_{j=1}^{J}\delta_{u_N^{(j)}}$$

[1] Y. Chen and D. S. Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci., January 2012.

[2] A. A. Emerick and A. C. Reynolds. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. Comput. Geosci., 2013.

Continuous time limit

Letting $t = hn$, taking continuous-time limit $h \to 0$ and modifying the noise term gives

Interacting particle system for sampling: Ensemble Kalman Sampling (EKS)[1,2]

$$\dot{u}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\Big\langle \mathcal{G}(u^{(k)}) - \bar{\mathcal{G}}, \mathcal{G}(u^{(j)}) - y\Big\rangle_{\Gamma}(u^{(k)} - \bar{u})$$
$$- C(U)\Sigma^{-1}(u^{(j)} - m) + \sqrt{2C(U)}\,\dot{W}^{(j)}, \qquad j = 1, \dots, J.$$

$\rightsquigarrow$ derivative-free approximation of **interacting Langevin dynamics** (see next slide)

[1] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Ensemble Kalman methods for inverse problems. Inverse Problems, 2013.

[2] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. SIAM J. Appl. Dyn. Syst., 2020.

When $\mathcal{G}$ is linear,

$$\begin{aligned}
&\frac{1}{J}\sum_{k=1}^{J}\left\langle \mathcal{G}(u^{(k)})-\bar{\mathcal{G}},\mathcal{G}(u^{(j)})-y\right\rangle_{\Gamma}(u^{(k)}-\bar{u})\\
&\quad=\frac{1}{J}\sum_{k=1}^{J}\left\langle \mathcal{G}(u^{(k)}-\bar{u}),\mathcal{G}(u^{(j)})-y\right\rangle_{\Gamma}(u^{(k)}-\bar{u})\\
&\quad=\frac{1}{J}\sum_{k=1}^{J}\left(\nabla\Phi(u^{(j)})\cdot(u^{(k)}-\bar{u})\right)(u^{(k)}-\bar{u})=C(U)\nabla\Phi\left(u^{(j)}\right),
\end{aligned}$$

with

$$C(U)=\frac{1}{J}\sum_{j=1}^{J}(u^{(j)}-\bar{u})\otimes(u^{(j)}-\bar{u}),\qquad \Phi(u)=\frac{1}{2}\left|\mathcal{G}(u)-y\right|_{\Gamma}^{2}.$$

- EKS is a preconditioned Langevin dynamics:

$$\dot{u}^{(j)}=-\,C(U)\nabla\Phi_{R}(u^{(j)})+\sqrt{2C(U)}\,\dot{W}^{(j)},\qquad j=1,\dots,J.$$

Convergence analysis in the linear $\mathcal{G}$ setting

Convergence to invariant measure difficult for the interacting particle system

$\rightsquigarrow$ so let us first consider the formal **mean field limit** $J \to \infty$:

$$\begin{cases} \mathrm{d}\overline{u}_t = -\mathcal{C}(\rho_t)\nabla\Phi_R(\overline{u}_t)\,\mathrm{d}t + \sqrt{2\mathcal{C}(\rho_t)}\,\mathrm{d}W_t, \\ \rho_t = \mathrm{Law}(\overline{u}_t)\,. \end{cases} \qquad \text{(McKean SDE)}$$

- Here $\mathcal{C}(\rho_t)$ is the covariance under ρ_t:

$$\mathcal{C}(\rho_t) = \mathbf{E}\Big[\big(\overline{u}_t - \mathcal{M}(\rho_t)\big) \otimes \big(\overline{u}_t - \mathcal{M}(\rho_t)\big)\Big], \qquad \mathcal{M}(\rho_t) := \mathbf{E}[\overline{u}_t].$$

- The associated nonlocal Fokker–Planck equation reads

$$\begin{aligned} \frac{\partial\rho}{\partial t} &= \nabla\cdot\Big(\mathcal{C}(\rho)\,(\nabla\Phi_R\,\rho + \nabla\rho)\Big) \\ &= \nabla\cdot\Big(\mathcal{C}(\rho)\,\mathrm{e}^{-\Phi_R}\,\nabla\Big(\mathrm{e}^{\Phi_R}\,\rho\Big)\Big). \end{aligned}$$

Invariant distribution: Gaussian

$$\rho_\infty \propto \mathrm{e}^{-\Phi_R} = \text{Bayesian posterior for (IP)}.$$

but also any Dirac distribution…

Convergence analysis in the linear $\mathcal{G}$ setting: the mean field equation

If we assume initial data with nondegenerate covariance, then

- Exponential convergence of the moments[1]:

$$\left|\mathcal{M}(\rho_t) - \mathcal{M}(\rho_\infty)\right| \leqslant C\, \mathrm{e}^{-t}, \qquad \left|\mathcal{C}(\rho_t) - \mathcal{C}(\rho_\infty)\right| \leqslant C\, \mathrm{e}^{-t}.$$

- Wasserstein stability estimate[2]: any two solutions ρ^1 and ρ^2 satisfy

$$W_2\left(\rho_t^1, \rho_t^2\right) \leqslant C\ \mathrm{e}^{-t}\ W_2\left(\rho_0^1, \rho_0^2\right).$$

  - the convergence rate is independent of $\mathcal{M}(\rho_\infty)$ and $\mathcal{C}(\rho_\infty)$.
  - indicates EKS is, in a sense, self-preconditioning.
  - this is in contrast with non-preconditioned Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \Big(\nabla \Phi_R\, \rho + \nabla \rho\Big).$$

[1] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. SIAM J. Appl. Dyn. Syst., 2020.

[2] J. A. Carrillo and U. Vaes. Wasserstein stability estimates for covariance-preconditioned Fokker–Planck equations. Nonlinearity, 2021.

Convergence analysis in the linear $\mathcal{G}$ setting: from mean field to particles

Mean field convergence

When $\mathcal{G}$ is linear (so Φ_R is quadratic), solution $(\overline{u}_t)$ of

$$\begin{cases} \mathrm{d}\overline{u}_t = -\mathcal{C}(\rho_t)\nabla\Phi_R(\overline{u}_t)\,\mathrm{d}t + \sqrt{2\mathcal{C}(\rho_t)}\,\mathrm{d}W_t, \\ \rho_t = \mathrm{Law}(\overline{u}_t)\,. \end{cases}$$

converges exponentially to invariant measure.

Can we deduce anything about the interacting particle system?

$$\mathrm{d}u^{(j)} = -C(U)\nabla\Phi_R(u^{(j)})\,\mathrm{d}t + \sqrt{2C(U)}\,\mathrm{d}W^{(j)}, \qquad j = 1, \dots, J.$$

Let μ_t^J denote the associated empirical measure. Then

$$W_2(\mu_t^J, \rho_\infty) \leqslant \underbrace{W_2(\mu_t^J, \rho_t)}_{\to 0 \text{ as } J\to\infty ???} + \underbrace{W_2(\rho_t, \rho_\infty)}_{\leqslant C\,\mathrm{e}^{-t}}$$

Mean field limit for the ensemble Kalman sampler

Ensemble Kalman sampler

$$\mathrm{d}u^{(j)} = -C(U)\nabla\Phi_R(u^{(j)})\,\mathrm{d}t + \sqrt{2C(U)}\,\mathrm{d}W^{(j)}, \qquad j = 1, \dots, J.$$

- Interaction only through empirical covariance $C(U)$
- But this is a quadratic nonlinearity, in front of the noise
- Some of the usual techniques do not work[1]

Using different techniques[2,3], we can prove, assuming exchangeable initial condition,

$$\mathbf{E}\left[W_2\left(\mu_t^J, \rho_t\right)\right] \leqslant c(t) J^{-\alpha}, \qquad \text{with } c(t) \text{ growing exponentially.}$$

Corollary: (still in the linear $\mathcal{G}$ setting)

$$\lim_{t\to\infty} \lim_{J\to\infty} W_2(\mu_t^J, \rho_\infty) = 0.$$

[1] F. Bolley, J. A. Cañizo, and J. Carrillo. Mean-field limit for the stochastic Vicsek model. Appl. Math. Lett., 2012.

[2] Z. Ding and Q. Li. Ensemble Kalman sampler: mean-field limit and convergence analysis. SIAM J. Math. Anal., 2021.

[3] U. Vaes. Sharp propagation of chaos for the ensemble Langevin sampler. Arxiv preprint, 2024.

Advantages and limitations of EnKF methods for inverse problems

Main advantages of ensemble Kalman sampling and inversion (optimization variant):

- They are derivative-free;
- They are based on interacting particle systems;
- They have good convergence properties in the linear $\mathcal{G}$ setting:
  - Exponential convergence at the mean field level

$$W_2(\rho_t, \rho_\infty) \leqslant C\,\mathrm{e}^{-t}\, W_2(\rho_0, \rho_\infty), \qquad \rho_\infty : \text{Bayesian posterior.}$$

  - Rigorous mean field limit (albeit with not uniformly in time yet...)

$$W_2(\mu_t^J, \rho_t) \leqslant c(t) J^{-\alpha}.$$

- Relatedly, they are affine invariant[1];

Main limitation

Uncontrolled gradient approximation in the nonlinear case $\rightarrow$ sampling error!

[1] A. Garbuno-Inigo, N. Nüsken, and S. Reich. Affine invariant interacting Langevin dynamics for Bayesian inference. SIAM Journal on Applied Dynamical Systems, 2020.

Outline

Inverse problems: optimization and sampling approaches

Solving inverse problems using EnKF

A multiscale derivative-free methodology

Uncontrolled gradient approximation for the derivative-free EKS

The derivative-free ensemble Kalman sampler is based on the approximation

$$C(U)\nabla\Phi(u^{(j)}) \approx \frac{1}{J}\sum_{k=1}^{J}\langle\mathcal{G}(u^{(k)}) - \bar{\mathcal{G}}, \mathcal{G}(u^{(j)}) - y\rangle_\Gamma (u^{(k)} - \bar{u}).$$

When the posterior is not Gaussian, this approximation can be inaccurate.

- The method produces approximate posterior samples;
- Can we correct the error?

In this section, we present a derivative free sampling method which

- can be systematically refined to produce accurate posterior samples and
- generalizes an existing derivative-free optimization method[1].

[1] E. Haber, F. Lucka, and L. Ruthotto. Never look back - A modified EnKF method and its application to the training of neural networks without back propagation. arXiv e-prints, May 2018.

A multiscale approach with small parameters σ and δ

$$\dot u = \overbrace{- \frac{1}{J\sigma^2} \sum_{j=1}^{J} \left\langle \mathcal{G}(u^{(j)}) - \mathcal{G}(u), \mathcal{G}(u) - y \right\rangle_\Gamma (u^{(j)} - u)}^{\text{EnKF approximation of } C(\Xi)\nabla\Phi_R(u)} - C(\Xi)\Sigma^{-1}(u - m) + \sqrt{2}\,\dot W,$$

$$u^{(j)} = u + \sigma\,\xi^{(j)}, \qquad j = 1, \dots, J,$$

$$\dot\xi^{(j)} = -\frac{1}{\delta^2}\,\xi^{(j)} + \sqrt{\frac{2}{\delta^2}}\,\dot W^{(j)}, \qquad \xi^{(j)}(0) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d), \qquad j = 1, \dots, J,$$

where

$$C(\Xi) = \frac{1}{J} \sum_{k=1}^{J} \xi^{(k)} \otimes \xi^{(k)},$$

- $u \in \mathbf{R}^d$: distinguished particle, provides useful information for sampling;
- $(u^{(1)}, \dots, u^{(J)})$: collection of "explorers" useful for gradient approximation;
- σ: radius of exploration around the distinguished particle u;
- δ^2: correlation time of the Ornstein–Uhlenbeck processes $\xi^{(j)}$.

$$\mathbf{E}_{X\sim\rho^y}\,\varphi(X) \approx \frac{1}{T}\int_0^T \varphi\big(u(t)\big)\,\mathrm{d}t, \qquad \rho^y : \text{Bayesian posterior.}$$

A multiscale approach: motivation

When σ is small, it holds with good accuracy that

$$\mathcal{G}(u^{(k)}) - \mathcal{G}(u) \approx \nabla \mathcal{G}(u)(u^{(k)} - u).$$

$\rightarrow$ the equation for u reduces to

$$\begin{aligned} \dot{u} &= -\frac{1}{J} \sum_{k=1}^{J} \left(\xi^{(k)} \otimes \xi^{(k)} \right) \nabla \Phi(u) - C(\Xi) \Sigma^{-1} u + \sqrt{2} \dot{W} \\ &= -C(\Xi) \nabla \Phi_R(u) + \sqrt{2} \dot{W}. \end{aligned}$$

- $C(\Xi) \nabla \Phi_R(u)$ can be viewed as a projection of $\nabla \Phi_R(u)$ onto $\operatorname{Span}\left\{\xi^{(1)}, \dots, \xi^{(J)}\right\}$.
- **Many-particle limit**: if $J \gg 1$, then

$$C(\Xi) = \frac{1}{J} \sum_{k=1}^{J} \xi^{(k)} \otimes \xi^{(k)} \approx I_d.$$

- **Averaging limit**: if $\sigma \ll 1$ and $\delta \ll 1$, then $u(t)$ approximately satisfies

$$\dot{u} = -\nabla \Phi_R(u) + \sqrt{2} \dot{W}. \qquad \text{(Overdamped Langevin dynamics)}$$

Rigorous averaging result in joint limit $\sigma \to 0$, $\delta \to 0$

Let ϑ denote the solution to

$$\dot{\vartheta} = -\nabla \Phi_R(\vartheta) + \sqrt{2}\dot{W}, \qquad \vartheta_0 = u_0.$$

Using standard tools from multiscale analysis[1], it is possible to prove

Theorem (Pathwise convergence to an overdamped Langevin dynamics)

Let $p \geqslant 1$ and assume that $\mathcal{G} \in C^2(\mathbf{T}^d, \mathbf{R}^K)$. Then for all $T > 0$ there is $C = C(T)$ such that

$$\mathbf{E}\left[\sup_{0 \leqslant t \leqslant T} |u_t - \vartheta_t|^p\right] \leqslant C\left(\frac{\delta^p}{J^{\frac{p}{2}}} + \sigma^p\right).$$

Ideas of the proof:

- convergence w.r.t. to δ^2 classical averaging approach.
- convergence w.r.t. to σ: Taylor expansions:
- convergence w.r.t. to J: Law of large numbers in L^p.

[1] G. A. Pavliotis and A. M. Stuart. Multiscale methods. Texts in Applied Mathematics. Springer, New York, 2008. Averaging and homogenization.

Discretization in time

To discretize the multiscale system in time, we use

- the Euler–Maruyama method for u;
- the exact solution of the OU process for $\xi^{(j)}$;

$$\widehat{u}_{n+1} = \widehat{u}_n - \frac{1}{J\sigma} \sum_{j=1}^{J} \langle \mathcal{G}(\widehat{u}_n + \sigma \widehat{\xi}_n^{(j)}) - \mathcal{G}(\widehat{u}_n), \mathcal{G}(\widehat{u}_n) - y \rangle_\Gamma \, \widehat{\xi}_n^{(j)} \, \Delta$$
$$- C(\widehat{\Xi}_n) \Sigma^{-1} (\widehat{u}_n - m) \Delta + \sqrt{2\Delta}\, x_n, \qquad x_n \sim \mathcal{N}(0,1),$$
$$\widehat{\xi}_{n+1}^{(j)} = \mathrm{e}^{-\frac{\Delta}{\delta^2}} \, \widehat{\xi}_n^{(j)} + \sqrt{1 - \mathrm{e}^{-\frac{2\Delta}{\delta^2}}} \, x_n^{(j)}, \qquad x_n^{(j)} \sim \mathcal{N}(0,1), \qquad j = 1, \dots, J.$$

Theorem (Discrete strong convergence, ongoing work with A. Della Noce)

Assume that $\mathcal{G} \in C^2(\mathbf{T}^d, \mathbf{R}^K)$. Then for all T there exists $C = C(T)$ such that

$$\mathbf{E}\left[\sup_{0 \leqslant n \leqslant \lfloor T/\Delta \rfloor} |\widehat{u}_n - \vartheta_{n\Delta}|^p \right] \leqslant C \left(\frac{\delta^p}{J^{\frac{p}{2}}} + \Delta^{\frac{p}{2}} + \sigma^p \right).$$

Existence and convergence of invariant measure (fixed J)

By classical results[1], existence of a unique invariant measure $\mu_{\delta,\sigma}$ follows from

- **Lyapunov condition.** Let $\mathcal{X} = \mathbf{T}^d \times (\mathbf{R}^d)^J$ denote the state space. There exists a Lyapunov function $V : \mathcal{X} \to [1, \infty)$ and constants $\alpha > 0$ and $\beta \geqslant 0$ such that

$$\forall x \in \mathcal{X}, \qquad \mathcal{L}V(x) \leqslant -\alpha V(x) + \beta, \qquad \mathcal{L} = \text{generator}$$

- **Minorization condition.** There exists a constant $\eta \in (0,1)$ and a probability measure λ such that

$$\inf_{x \in \mathcal{C}} P_t(x, \mathrm{d}y) \geqslant \eta\, \lambda(\mathrm{d}y), \qquad P_t = \text{transition kernel}$$

Proposition (Convergence of stationary measures, ongoing work with A. Della Noce)

Assume that $\mathcal{G} \in C^\infty(\mathbf{T}^d, \mathbf{R}^K)$, and let $\nu_{\delta,\sigma}$ denote the u-marginal of $\mu_{\delta,\sigma}$. Then

$$\forall f \in W^{1,\infty}(\mathbf{T}^d), \qquad \left| \int_{\mathbf{R}^d} f(u)\, \nu_{\delta,\sigma}(\mathrm{d}u) - \int_{\mathbf{R}^d} f(u)\, \rho^y(\mathrm{d}u) \right| \leqslant C \|f\|_{W^{1,\infty}(\mathbf{T}^d)} (\delta^2 + \sigma)$$

Idea of proof: Use technique based on Poisson equation from[2]

[1] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab. Birkhäuser/Springer Basel AG, Basel, 2011.

[2] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. SIAM J. Numer. Anal., 2010.

Importance of preconditioning

Simplified setting:

- Φ_R is quadratic:

$$\Phi_R = \frac{1}{2}\left|u\right|_C^2, \qquad C \succ 0.$$

- Explicit Euler for equation $\dot{u} = -\nabla\Phi_R(u) = -C^{-1}u$:

$$u_{n+1} = (I - \Delta t\, C^{-1})u_n$$

Stability requires $\Delta t < \lambda_{\min}(C)$. When $\Delta t = \frac{1}{2}\lambda_{\min}(C)$,

$$|u_n| \leqslant \left|1 - \frac{1}{2}\frac{\lambda_{\min}(C)}{\lambda_{\max}(C)}\right|^n |u_0|.$$

- Slow convergence when $\lambda_{\min}(C) \ll \lambda_{\max}(C)$!
- Need for preconditioning:

$$\dot{u} = -K\nabla\Phi_R(u), \qquad \text{Optimal preconditioner: } K = C = \mathrm{Cov}\left(\frac{1}{2}\,\mathrm{e}^{-\Phi_R(u)}\right)$$

Improving convergence of the multiscale method with preconditioning

The method can be preconditioned with an appropriate matrix $K \succ 0$.

$$\dot{u} = -\frac{1}{J\sigma^2} \sum_{j=1}^{J} \langle \mathcal{G}(u^{(j)}) - \mathcal{G}(u), \mathcal{G}(u) - y \rangle_\Gamma \, (u^{(j)} - u) - C_K(\Xi)\Sigma^{-1} u + \sqrt{2K}\, \dot{W}$$

$$u^{(j)} = u + \sigma \sqrt{K} \xi^{(j)}, \qquad j = 1, \dots, J,$$

$$\dot{\xi}^{(j)} = -\frac{1}{\delta^2} \xi^{(j)} + \sqrt{\frac{2}{\delta^2}}\, \dot{W}^{(j)}, \qquad \xi^{(j)}(0) \sim \mathcal{N}(0, I_d), \qquad j = 1, \dots, J,$$

where $C_K(\Xi) := \sqrt{K}\, C(\Xi) \sqrt{K}$.

Formal justification: For small σ,

$$\dot{u} \approx -C_K(\Xi) \nabla \Phi_R + \sqrt{2K}\, \dot{W},$$

which, in the limit $\delta \to 0$, converges to

$$\dot{u} \approx -K \nabla \Phi_R + \sqrt{2K}\, \dot{W}.$$

In practice, we set $K \approx \mathrm{Cov}\left(\frac{1}{Z}\, \mathrm{e}^{-\Phi_R(u)}\right)$ approximated by ensemble Kalman sampling.

Illustration: effect of preconditioning

Here we use the multiscale method to find the minimizer of

$$\Phi(u) = \frac{1}{2}\left(|u_1 - 1|^2 + k^2|u_2 - 1|^2 + k^4|u_3 - 1|^2\right), \qquad k = 5.$$

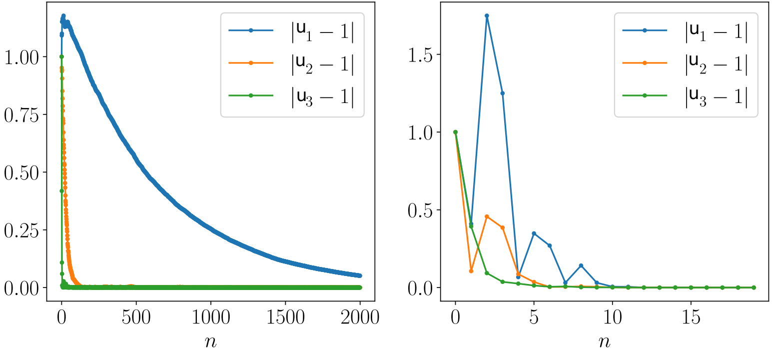


Figure: Error between the iterates and the MAP estimator, without (left) and with (right) preconditioning.

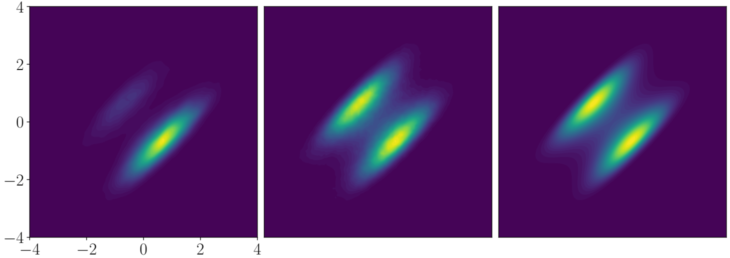
Example 1: Bimodal target distribution

Inverse problem with bimodal posterior

Find $u = (u_1, u_2) \in \mathbf{R}^2$ from

$$y = |u_1 - u_2|^2 + \eta, \qquad \eta \sim \mathcal{N}(0,1).$$

Prior distribution $u \sim \mathcal{N}(0, I_2)$. Below $y = 2$.



Left: Approximate posterior using EKS.
Middle: Approximate posterior using multiscale method.
Right: True Bayesian posterior.

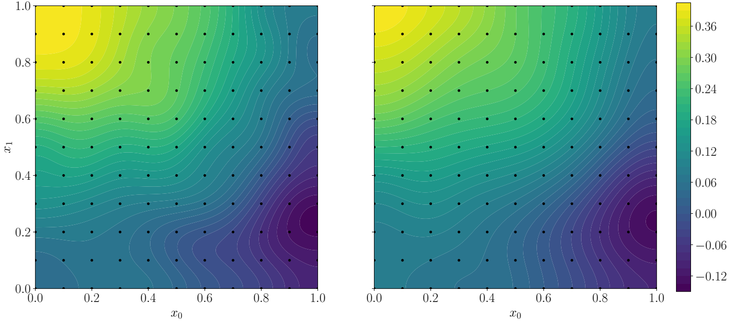
Inference of the conductivity in a plate

Find $u(x)$ from 100 noisy measurements of the temperature $T(x)$ where

$$-\nabla \cdot \left(\mathrm{e}^{u(x)} \nabla T(x) \right) = \mathrm{cst} \qquad x \in D = [0,1]^2, \quad + \text{ homogeneous Dirichlet BC.}$$

Model: $u(x) \sim \mathcal{N}(0, \mathcal{C})$ with $\mathcal{C} = (-\Delta + \tau^2 \mathcal{I})^{-\alpha}$:

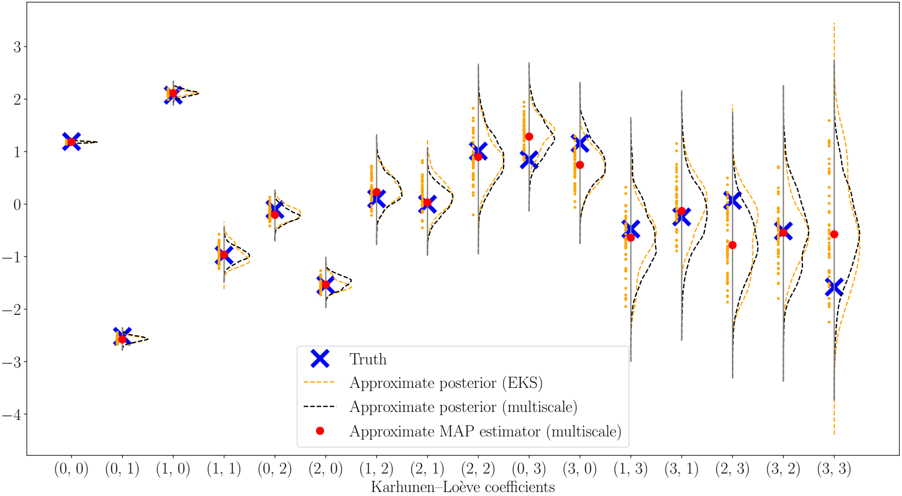
$$\text{KL expansion :} \quad u(x) = \sum u_i \sqrt{\lambda_i} \varphi_i(x), \qquad u_i \sim \mathcal{N}(0,1), \qquad \mathcal{C}\varphi_i = \lambda_i \varphi_i.$$



True (left) and reconstructed (right) log-conductivity ($\delta = \sigma = 10^{-5}$, $J = 8$)

Example 2: two-dimensional elliptic boundary value problem – Sampling

Approximate posterior from 10,000 iterations of the multiscale method:



Conclusions and perspectives for future work

In this presentation, we presented a novel method for sampling and optimization which

- is derivative-free and based on a system of interacting particles;
- is provably refineable over finite time intervals;
- can be preconditioned using information from EnKF methods for efficiency.

Many interesting questions remain open:

- Adaptive σ for computational efficiency;
- uniform-in-time weak error estimate (ongoing);
- Alternative (e.g. semi-implicit) time discretizations.
- Alternative derivative-free methodologies.

Thank you for your attention!